Open Archive TOULOUSE Archive Ouverte (OATAO)
OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: http://oatao.univ-toulouse.fr/
Eprints ID: 3943

To cite this version:
Nouvel, Laurent- Xavier and Marenda, Marc and Sirand-Pugnet, Pascal and Sagné, Eveline and Glew, Michelle and Mangenot, Sophie and Barbe, Valérie and Barré, Aurélien and Claverol, Stéphane and Citti, Christine (2009) Occurrence, plasticity and evolution of the vpma gene family, a genetic system devoted to high frequency surface variation in Mycoplasma agalactiae. Journal of Bacteriology, vol. 191 (n° 13). pp. 4111-4121. ISSN 0021-9193

Any correspondance concerning this service should be sent to the repository administrator: staff-oatao@inp-toulouse.fr.
Occurrence, plasticity and evolution of the \textit{vpma} gene family, a genetic system devoted to high frequency surface variation in \textit{Mycoplasma agalactiae}

Laurent-Xavier Nouvel$^{1,2}$, Marc Marenda$^{1,2a}$, Pascal Sirand-Pugnet$^{3,4}$, Eveline Sagné$^{1,2}$, Michelle Glew$^{1,2,b}$, Sophie Mangenot$^5$, Valérie Barbe$^6$, Aurélien Barré$^6$, Stéphane Claverol$^7$, and Christine Citti$^{2,1\ast}$

$^1$Université de Toulouse, ENVT, UMR 1225, F-31076 Toulouse, France
$^2$INRA, UMR 1225, F-31076 Toulouse, France
$^3$Université de Bordeaux, UB2, UMR1090, F-33076 Bordeaux, France
$^4$INRA, UMR 1090, F-33883 Villenave d’Ornon, France
$^5$CEA-IG, Genoscope, F-91057 Evry Cedex, France
$^6$Centre de Bioinformatique de Bordeaux, UB2, F-33076 Bordeaux, France
$^7$Pole Protéomique, Plateforme Génomique Fonctionnelle Bordeaux, UB2, 33076 Bordeaux, France

$^a$ Current address. School of Veterinary Science, 250 Princes Highway, Werribee, Victoria 3030, Australia.
$^b$ Current address. Bio21 Institute, Level 3 North, Rm 375, Melbourne Dental School, 30 Flemington Rd, Parkville 3010, Australia.

Corresponding author
Christine Citti
UMR 1225 INRA, ENVT
Ecole Nationale Vétérinaire de Toulouse
23 Chemin des Capelles BP 87614 31076 Toulouse Cedex 3
France
Tel: +33.(0)5.61.19.38.56
Fax: +33.(0)5.61.19.32.73
E-mail: c.citti@envt.fr

Running title: The \textit{vpma} locus: occurrence, plasticity and evolution
Abstract

*Mycoplasma agalactiae*, an important pathogen of small ruminants, exhibits a very versatile surface architecture by switching ON-OFF multiple, related lipoproteins (Vpmas). In the type strain, PG2, Vpma-phase variation is generated by a cluster of 6 *vpm*a genes that undergoes frequent DNA rearrangements via site-specific recombination. To further comprehend the degree of diversity that can be generated at the *M. agalactiae* surface, the *vpm*a gene repertoire of a field strain, 5632, was analyzed and shown to contain an extended repertoire of 23 *vpm*a genes distributed onto two loci located 250 kbp apart. Loci I and II include 16 and 7 *vpm*a genes, respectively, with all *vpm*a of locus II being duplicated at locus I. Several Vpmas displayed a chimeric structure suggestive of homologous recombination and a global proteomic analyses further indicate that at least 13 out of the 16 Vpmas can be expressed by the 5632 strain. Because a single promoter is present in each *vpm*a locus, concomitant Vpma expression can occur in a strain with duplicated loci. Consequently, the number of possible surface combinations is much higher in 5632 compared to the type strain. Finally, our data suggested that insertion sequences are likely to be involved in 5632 *vpm*a locus duplication at a remote chromosomal position. The role of such mobile genetic element in chromosomal shuffling of genes encoding major surface components may have important evolutionary and epidemiological consequences for pathogens such as mycoplasmas that have a reduced genome and no cell wall.
Introduction

Bacteria of the *Mycoplasma* genus belong to the Class *Mollicutes* and represent a remarkable group of organisms that derived from the Firmicutes lineage by massive genome reduction (41, 51). Consequent to this regressive evolution, modern mycoplasmas have been left with a small genome (580 to 1,400 kb), a limited number of metabolic pathways and no cell wall. Due to these particularities, members of the *Mycoplasma* genus have often been portrayed as "minimal self-replicating organisms". Despite this apparent simplicity, a large number of mycoplasma species are successful pathogens of man and a wide range of animals in which they are known to cause diseases that are often chronic and debilitating (1, 33). The surface of their single membrane is considered as a key interface in mediating adaptation and survival in the context of a complex, immunocompetent host (10, 13, 34, 40). Indeed, mycoplasmas possess a highly versatile surface architecture due to a number of sophisticated genetic systems that promote the intraclonal variation in expression and structure of abundant surface lipoproteins (9, 50). Usually, these systems combine a set of contingency genes with a molecular switch for turning expression ON or OFF that is either based (i) on spontaneous mutation (slipped strand mispairing), (ii) on gene conversion or (iii) on specific DNA rearrangements (9). While high-frequency phenotypic variation using the two first mechanisms has been thoroughly described in other bacteria (47), switching of surface components by shuffling silent genes at a particular single expression locus has mainly been studied in mycoplasmas (3, 8, 14, 16, 23, 39, 43).

*Mycoplasma agalactiae*, an important pathogen responsible for contagious agalactia in small ruminants (listed by the world organization for animal health), possesses a family of lipoproteins encoded by the *vpma* genes which phase variation in expression is driven
by a “cut and paste” mechanism involving a tyrosine site-specific recombinase designated as Xer1 (16). Data previously gathered in the PG2 type strain had identified a single vpma cluster (42) composed of 6 vpma genes adjacent to one xer1 gene (Fig. 1A). Based on fine genetic analyses, Xer1 was further shown to mediate frequent site-specific DNA rearrangements by targeting short DNA sequences located upstream each vpma (8, 16). While some vpma rearrangements can be phenotypically silent, others result in Vpma ON/OFF switching by linking a silent vpma gene sequence immediately downstream to the unique vpma promoter. Because site-specific recombination can be reciprocal, the initial vpma configuration can be restored without loss of genetic information.

The vsa family of the murine pathogen M. pulmonis (3, 39), the vsp family of the bovine pathogen M. bovis (2, 24), and the vpma family of M. agalactiae, all generate intraclonal surface diversity using a very similar molecular switch (23) although their overall coding sequences seem to be specific of one Mycoplasma species. DNA rearrangements also govern phase variation of the 38 mpl genes of the human pathogen, M. penetrans (27, 35, 38). However, in this mycoplasma species the molecular switch is slightly different since each mpl gene possesses its own invertible promoter (19). In M. penetrans, the individual expression of each mpl gene can then be switched ON/OFF in a combinatory manner resulting in a large number of possible Mpl surface configurations. Since M. pulmonis, M. bovis and M. agalactiae all belong to the Mycoplasma hominis phylogenetic cluster (48) and are relatively closely related while M. penetrans belongs to the distinct Mycoplasma pneumoniae phylogenetic cluster (30, 48), it is tempting to speculate that the vsa, the vsp and the vpma systems have all been inherited from a common ancestor and that the bulk of their coding sequences has evolved
independently in their respective host while the molecular switch mechanism has been retained.

In so-called “minimal” bacteria, the occurrence of relatively large genomic portions dedicated to multigene families, each encoding phase variable, related, surface proteins, suggests that they serve an important function(s). Data accumulated over the years in several mycoplasma species tend to indicate that one general purpose of these systems is to provide the mycoplasma with a variable shield that modulates surface accessibility in order to escape the host response and to adapt to rapidly changing environments (10, 11, 13, 40, 50). On the other hand, the sequences of phase variable proteins are relatively conserved within one species but divergent between species, suggesting a more specific role for these molecules.

The role of the Vpma family of \textit{M. agalactiae} has yet to be elucidated but it was recently shown that Vpma switches in expression occur at a remarkably high rate \textit{in vitro}, \(10^{-2}\) to \(10^{-3}\) per cell per generation (8, 17). The vpma system described in PG2 (16) and in another \textit{M. agalactiae} strain isolated in Israel (in which Vpma were designated as Avg (14)), both revealed a repertoire of 6 \textit{vpma} genes and only one promoter, suggesting that in \textit{M. agalactiae} the number of Vpma configurations is limited to 6. This contrast with the situation commonly found in other \textit{Mycoplasma} variable systems that can offer larger mosaic of surface architecture because of the concomitant switches of several surface related proteins and/or because of a higher number of phase variable genes.

To further understand the degree of diversity that can be generated at the surface of \textit{M. agalactiae}, we analyzed the \textit{vpma} gene content of a field strain, 5632, whose genome has recently been sequenced by our group (unpublished data). The present study shows that 5632 contains a total of 23 \textit{vpma} genes distributed in two distinct loci.
that both contain a recombinase gene. Further genomic and proteomic analyses indicate that the capacity of 5632 to vary its Vpma surface architecture is far more complex than that described for the type strain. Unlike PG2, both 5632 vpmA loci are associated with several mobile genetic elements (IS) that could play an evolutionary role in the dynamics of vpmA repertoires as suggested by data presented here. One 5632 vpmA locus contains ORFs that are highly conserved in both M. bovis, a closely related bovine mycoplasma, and in the phylogenetically distant mycoplasmas of the mycoides cluster that are also important ruminant pathogens. Whether these have been acquired through evolution or through horizontal transfer will be discussed. The present study reveals an additional degree of complexity for the Vpma system and further suggests that some field strains might have a more dynamic genome and a more variable surface than was first estimated (42).
Materials and methods

Bacterial strains, culture conditions and DNA isolation

*M. agalactiae* type strain PG2, clone 55.5 (16) and strain 5632, clonal variant C1 (26), used in this study have been previously described. These strains have been independently isolated from goat in Spain. Experiments reported in this manuscript have all been performed with these clonal variants but for simplicity, we will further refer as the PG2 and the 5632 only. *M. agalactiae* field isolates were kindly provided by Dr. Poumarat (AFSSA Lyon, France) (see Table S1 in the supplemental material). Mycoplasmas were propagated in SP4 liquid medium (46) at 37°C and genomic DNA was extracted as described elsewhere, (7, 37).

Identification of *M. agalactiae* vpma and associated loci

Whole genome sequencing of strain 5632 (clonal variant C1) was performed as follows. A library of 3 kb inserts (A) was generated by mechanical shearing of the DNA followed by cloning of the blunt end fragments into the pcDNA2.1 (Invitrogen) *E. coli* vector. Two libraries of 25 kb (B) and 80 kb (C) inserts were generated by *HindIII* partial digestion and cloning into the pBeloBAC11 (CALTECH) modified *E. coli* vector. The plasmid inserts of 10752, 3072 and 768 clones picked from the libraries A, B and C, respectively, were end-sequenced by dye-terminator chemistry on ABI3730 sequencers. The PHRED/PHRAP/CONSED software package was used for sequence assemblies. Gap closure and quality assessment were made according to the Bermuda rules with 10307 additional sequences. The *vpma* loci of 5632 were detected by DNA homologies to that previously described of PG2 (positions of the loci on the 5632 genome are given in the text with reference to the first nucleotide of the dnaA gene as nucleotide 1) and were
annotated using the CAAT-Box platform (15) with the aid of the Artemis software (36) and the ACT software (5). The BLAST program suite was used for sequence homology searches in non-redundant databases (http://www.ncbi.nlm.nih.gov/blast/blast.cgi). In order to determine the extent of sequence similarity, alignments between sequences were performed using softwares Needle (Needleman-Wunsch global alignment algorithm) and Water (Smith-Waterman local alignment algorithm) (http://www.ebi.ac.uk/Tools/emboss/align/). Designation of 5632 vpma genes was based on local and global alignment scores using all vpma gene sequences including those previously described in PG2. A same name was attributed to vpma presenting an amino-acid identity > 50% in global alignment (Needle) and an amino-acid identity >70% in local alignment (Water). Using these rules, strain 5632 was found to contain the vpmaW and vpmaX genes also present in PG2. As well, a distinct name was given to vpmaY and vpmaF although they display a global identity >50% because their local identity is <70%. vpma products with highly similar sequences that only differed in the number of their C-terminal repeats were considered as allelic versions and designated with a same letter followed by a different number (i.e. VpmaD1 and VpmaD2 and, VpmaF1 and VpmaF2). Finally, vpmaI was not considered as an allelic version of vpmaD because the C-terminal repeated region of its corresponding product displayed a local identity of 64.7% (<70%) with VpmaD1 and D2 counterparts.

Direct sequencing of PCR products and of genomic DNA (18, 20) was performed using specific primers (see Table S2 in the supplemental material) at the sequencing facility of UMR 5165 (CNRS, UPS, CHU Purpan, Toulouse, France).
Phylogenetic analyses were performed using MEGA 3.1 (21) and the Neighbor-Joining tree method. The reliability of the tree nodes was tested by performing 500 bootstrap replicates.

Accession number for locus I\textsubscript{5632} and locus II\textsubscript{5632} are, respectively, FP245515 and FP245514

**PCR assays**

PCR assays were performed on an Eppendorf Mastercycler ep-Gradient thermocycler using 5 ng *M. agalactiae* DNA as template with specific primers (see Table S2 in the supplemental material). PCR assays were performed in 25 μL reaction mixtures containing 0.4 mM of each primer, PCR reaction buffer (with MgSO\textsubscript{4}, New England Biolabs [NEB]) at 1X final concentration, 200 mM dNTPs and 2.5 U *Taq* DNA polymerase (NEB). Reaction mixtures were subjected to 2 min at 94°C, 30 cycles of 30s at 94 °C, 30s at 55°C, 30s at 72 °C, and a final elongation step of 5 min at 72°C. All PCR assays were performed at the unique annealing temperature of 55°C using primer pair xerF/phydR for specific amplification of locus I; primer pairs xerF/agpR or Mag2F/agpR for locus II; primer pairs aip1F/aip1R for the *abiGI* gene; primer pairs aip2F/aip2R for the *abiGII* gene. PCR products were analyzed by gel electrophoresis in 1% agarose.

**Colony immunoblotting and proteomic analyses**

Colony immunoblottings were performed as previously described (8). Briefly, nitrocellulose membranes were placed on mycoplasma colonies freshly grown on the agar medium, then removed and rinsed three times in TS buffer (10 mM Tris, 154 mM NaCl, pH 7.4). Membranes were then incubated overnight at 4°C with rabbit Vpma-
specific antibodies previously described (8), washed three times in TS buffer containing 0.05% Tween® 20 (Roth) and then incubated for 1 h at 25°C in 1:2000 dilution of swine anti-rabbit IgG conjugated to horseradish peroxidase (Dako). After three washes, the colony blots were developed for 15–30 min in 4-chloro-1-naphtol and 0.02% hydrogen peroxide. The reaction was stopped by washing the blots in sterile distilled water.

Proteins that partitioned into Triton-Tx114 were extracted from 5632 as previously described (4), precipitated overnight at -70°C after addition of 9 volumes of cold MeOH, centrifugated 10 min at 12,000xg and subjected to SDS-PAGE. The gel was sliced into 16 sections which were subjected to trypsin digestion. Peptides were further analyzed by nano liquid chromatography coupled to a MS/MS ion-trap mass spectrometer (LC-MS/MS).

Peptides were identified with SEQUEST through the Bioworks 3.3.1 interface (Thermo-Finnigan, Torrence, CA, USA) against a database consisting of both direct and reverse sense *Mycoplasma agalactiae* strain 5632 entries (1652 entries). Using the following criteria (ΔCN ≥ 0.1, Xcorr vs Charge State ≥ 1.5 (+1), 2.0 (+2), 2.5 (+3), Peptide Probability ≤ 0.001 and Number of Different Peptides ≥ 2) as validation filters, the False Positive rate is null.
Results and discussion voir PG2-vpma et the PG2 vpma

Genome sequencing of strain 5632 reveals an extended vpma repertoire

The fully sequenced genome of *M. agalactiae* strain 5632 revealed a total of 23 *vpma* genes (Fig. 1). In contrast to the PG2 type strain, this extended repertoire is distributed onto two distinct chromosomal loci, locus I and II, that contain 16 and 7 *vpma* genes, respectively. Blast analyses only identified two having significant similarity with those previously described in PG2 (16, 42), namely *vpmaW* and *vpmaX*, the others (*vpmaA* to *vpmaG*) only shared blocks of high similarity with PG2 *vpmas* (Fig. 2).

More precisely, locus I<sub>5632</sub> is 19453 bp long and is the counterpart of the PG2 *vpma* locus (Fig. 1A) because of their flanking CDS being nearly identical. In addition to 16 *vpma* genes, locus I<sub>5632</sub> contains two CDS (MAGa8140 and MAGa8130) that have no homology with the Vpma family but were found to have a high amino-acid similarity with *Streptococcus agalactiae* *abiGI* and *abiGII* gene products (45). The two corresponding genes are lacking in PG2 and were designated here as *abiGI* and *abiGII*, respectively. Locus I<sub>5632</sub> further differs from its PG2 counterpart by the presence of an Insertion Sequence (IS) corresponding to the previously described IS<sub>Mag1</sub> (31). This IS element is immediately adjacent to the 3’ end of the integrase-recombinase xer1 gene, whose product is identical to that of PG2.

The second *vpma* locus of 5632, locus II<sub>5632</sub>, is 8462 kb long and contains only 7 *vpma* genes clustered between two IS<sub>Mag1</sub> copies, themselves flanked by a 16S rRNA gene at the 5’ end and by a CDS annotated as a hypothetical protein (HP, MAGa5900) at the 3’ end. In PG2, this region is highly conserved except that it does not contain any IS element or *vpma* or any other gene (locus II<sub>PG2</sub>, Fig. 1C and below). DNA blast analyses
of locus II_{5632} demonstrated that all seven \textit{vpma} coding sequences are identical to seven \textit{vpma} genes of locus I_{5632} (see asterisks in Fig. 1B). Interestingly, an identical \textit{xer1} gene also occurs at both \textit{vpma} loci of 5632. Since a single functional gene would be sufficient to generate DNA rearrangements in each locus, the duplication was further confirmed by PCR to rule out any possible artefact of sequence assembling. This was performed using a \textit{xer1}-specific primer in combination with a primer specific of the 3' region of locus I, phydR, or of locus II, agpR (Fig. 1D).

**The Vpma repertoire is composed of common and distinct structures**

Detailed analyses of the 23 \textit{vpma} coding regions of 5632 showed that (i) two are allelic versions of PG2 \textit{vpma} genes (\textit{vpmaW} and \textit{vpmaX}), (ii) 7 are duplicated (\textit{vpmaA}, \textit{B}, \textit{C}, \textit{D1}, \textit{D2}, \textit{E} and \textit{F1}) and distributed in two \textit{vpma} loci that are 250 kbp apart, and (iii) two are allelic copies of \textit{vpmaD} (\textit{D1} and \textit{D2}) and \textit{vpmaF} (\textit{F1} and \textit{F2}) that mainly vary in the number of their C-terminal repeated motifs. Consequently, strain 5632 contains 12 new, distinct Vpma products when compared to PG2 (Fig. 2). All \textit{vpma} gene products encode a conserved amino-acid signal sequence at their N-termini followed by a lipobox and by a 11 amino-acids conserved sequence (41). Within 5632, some mature \textit{vpma} gene products are unique (VpmaA, G, W) while others share blocks of amino-acids between them (VpmaB, C, D1, D2, E, F1, F2, H, I, J, K, L, X). Also, most Vpmas contains amino-acid blocks that are directly repeated (VpmaA, B, D1, D2, I, F1, F2, G, W). These data indicate that the diversity of the Vpma family is represented by a mosaic of structures that is suggestive of homologous recombination events between \textit{vpma} genes, gene duplication and/or insertion-deletion of repeated motifs. For instance, VpmaF1 and VpmaF2 share a common block of 104 amino-acids with VpmaK and VpmaX and, a
common block of 85 amino-acids with VpmaY not found in VpmaK or VpmaX. Some Vpma genes such as VpmaD1 and D2 or VpmaF1 and F2 represent size variants of a same Vpma, with a variable number of repeated motifs at their C-termini (i.e. VpmaD1 and D2 contain respectively 20 and 2 repeated motifs of 15 amino-acids). Additionally, VpmaI may have a common ancestor with VpmaD1 and D2 and can be seen as the result of combined events such as gene duplication followed by sequence drift and/or repeat expansion at the 3' end.

Anti-VpmaW and anti-VpmaY rabbit antibodies previously produced against Vpma products of strain PG2 (8) also reacted with 5632 surface exposed epitopes in colony immunoblotting (Fig. 1) and revealed the presence of sectored colonies characteristic of high-frequency variation in expression. Most likely, 5632 Vpmas recognised by these antibodies correspond to VpmaW and to VpmaF1 and/or F2 that share blocks of amino acids with PG2 VpmaY (Fig. 2).

To better define the Vpma products expressed by 5632, proteins that partitioned in Triton X-114 detergent were subjected to SDS PAGE followed by in-gel trypsin digestion and by LC-MS/MS. Results showed that at least two specific peptides were detected for 13 Vpma (VpmaA, B, C, D1, E, F1, F2, G, H, I, J, X, and W) indicating that these were expressed by 5632 during propagation (see Table S3 in the supplemental material). While D1-specific peptides were detected, the expression of VpmaD2 could not be assessed because all VpmaD2 trypsin-peptides being also encoded in VpmaD1 or in VpmaI. No peptide specific of VpmaL or VpmaK was detected although each product possesses a long specific sequence of 225 and 129 amino acids, respectively (Fig. 2). Whether VpmaL or VpmaK are expressed at low level or under different conditions can not be ruled out. Finally, all Vpmas encoded by locus II_{5632} are duplicated at locus I_{5632}
and consequently, their expression can not be monitored by this method or by any other such as Northern blot analyses because their sequences are also identical at the DNA level.

The global proteomic approach taken above is qualitative and does not reflect the level of expression of each Vpma, two of which are predicted to be expressed at high level by the clonal population while others should only reflect a minority of back-switchers. In PG2, the vpma locus contains a unique promoter which drives the expression of the vpma localized immediately downstream while other vpma genes remain silent (16). This promoter region was more precisely defined by Flitman-Tene et al. (14) using primer extension and a similar sequence was detected in each locus of the strain 5632. Therefore, Vpma expression in 5632 is most likely driven at each locus by a single promoter sequence which is highly conserved between the two loci (99% nucleotide identity). In locus I_{5632}, the promoter sequence is located upstream of the vpmaB gene while in locus II_{5632} it is upstream of vpmaA suggesting that these two Vpmas would be predominantly co-expressed by the clonal population. However, in the absence of specific-antibodies this hypothesis can not be tested. One interesting observation is that the length of a polyT tract, located between the putative -35 and -10 boxes of the vpma promoters, varies among the three vpma promoters sequenced and is of 14-, 15- and 12-nt respectively for loci I of PG2 and 5632 and for locus II of 5632 (see Fig. S1 in the supplemental material). In other variable mycoplasma systems, i.e. the Vlp system of *M. hyorhinis*, fluctuation in the length of the polynucleotide tract between the -35 and the -10 boxes has been shown to affect transcription (12). Whether this is also the case for the *M. agalactiae* species and for strain 5632 in particular, cannot be currently tested because of the complexity of the 5632 vpma system. In PG2, DNA rearrangements are
mediated by Xer1 which one recognises a specific Recombination Site (RS) sequence located upstream of each vpma. Indeed, analysis of the 5632 and PG2 vpma loci revealed that a rigorously identical RS sequence is located upstream of each of the 29 vpma genes (23 in 5632 and 6 in PG2). This suggests that each vpma loci of 5632 can independently undergo DNA rearrangements and that the concomitant expression of two distinct Vpmas is most likely to occur frequently in this strain. If so, 5632 could theoretically display 91 different Vpma surface configurations against only 6 in the type strain PG2. In parallel, additional Vpma size variations by insertion or deletion of direct repeats can further participate in altering surface architecture.

**Evidence for horizontal gene transfer at the vpma locus of 5632**

As mentioned earlier, *M. bovis* is a close relative of *M. agalactiae* and generates surface variation using a very similar genetic system designated as the vsp gene family. Both, the vsp and the vpma systems, have several common features: (i) vsp or vpma related genes are clustered in close proximity to a conserved recombinase gene (32), (ii) the recombination site involved in DNA rearrangements is identical except for a single nucleotide polymorphism that is conserved in each species, and (iii) vpma or vsp genes code for proteins having a signal peptide that is highly conserved between the two systems. Unlike PG2, vpma loci of 5632 are associated with IS elements as in *M. bovis* (23) and one, locus l_{5632}, contains two non-vpma genes, abiGI and abiGII, that have 86.97 and 85.15 % overall DNA identity with two *M. bovis* ORFs, ORF4 and ORF5, (23), located in vsp locus. Searching the databases revealed that abiGI and abiGII from 5632 also match with gene products annotated as conserved hypothetical products in *Mycoplasma mycoides* subsp. mycoides biotype LC strain GM12 (22) and *M. mycoides*
subsp. mycoides SC strain PG1 (49), with more than 50% of amino-acid similarity in global alignment (Table 1). While abiG homologs are present in several Gram positive bacteria such as Streptococcus agalactiae and S. suis (6, 45), none was found in other Mollicutes sequences available in the databases, which included 21 complete genomes. The abiGI and abiGII genes were first described in Lactococcus lactis subsp. cremoris UC653 (28) where they occur as an operon carried by a conjugative plasmid and confer resistance to phage infection in this species (29). Detection by LC/MS-MS of AbiGII-like specific peptides in the Triton-X114 detergent phase of 5632 indicates that the corresponding abiGII gene is expressed although there is no hydrophobic domain that could account for this partitioning. Previous attempts to define the function of abiGI and abiGII of Streptococcus species failed (44) and whether abiG homologues found in mycoplasma have a role in protecting them against phage attack remains to be addressed. Nevertheless, comparison of AbiGI and AbiGII amino-acid sequences by constructing Neighbour-Joining trees (500 bootstraps) indicates a strong relationship between Mycoplasma and Streptococcus sequences suggesting that their corresponding genes might have undergone horizontal gene transfer (see Fig. S2 in the supplemental material).

To further define the occurrence of the two abi genes in M. agalactiae, a collection of 92 strains representative of (i) various geographical areas and (ii) various histories (see Table S1 in the supplemental material) was screened by PCR for the presence of abiG genes using specific primer pairs, aip1F/aip1R and aip2F/aip2R. PCR products corresponding to each gene were obtained for 11 strains out of 92. Further PCR assays showed that, like in Streptococcus agalactiae, the two abiG genes always occur next to each other. Interestingly, direct genome sequencing of six abiG positive strains using
aip2F2 as primer shows that the abiGII gene is always associated with vpma genes. Because strains displaying abiG genes in their vpma loci have very different geographical origin (Ivory cost, Spain and France), this results raised the question of whether these genes have been vertically inherited from a *M. agalactiae/M. bovis* common ancestor or whether the vpma locus is a hot spot for their insertion. Since the vpma locus is frequently subjected to DNA rearrangements, deletion of abiG genes by intra-recombination, involving for instance vpma-RS sequences, may have easily occurred resulting in abiG negative strains.

**Plasticity of the *M. agalactiae* vpma loci**

To define whether the occurrence of the two vpma loci observed in 5632 is common among *M. agalactiae* strains, a PCR assay was performed with primer pairs xerF/phydR, xerF/agpR and Mag2F/agpR that yield amplicons corresponding to either locus I or II of PG2 and/or 5632. More specifically, the two pairs xerF/phydR and xerF/agpR target the 3' ends of the loci which are not affected by vpma rearrangements while Mag2F/agpR were designed to amplify a more or less empty locus II under our PCR conditions. The assay was conducted with the panel of 92 strains (see above). Results showed that 89 strains produced an amplification profile identical to that obtained with PG2 while three showed new profiles none of which was identical to those obtained with 5632 (Fig. 3). Sequencing of PCR products or selected genomic regions showed that (i) strain 13628, like 5632, contains an IS element at locus I while displaying a locus II similar to PG2, and (ii) strains 13375 and 4025, like PG2, displayed no IS at locus I while locus II contain one or more IS but no vpma, with strain 4025 having an IS30-like inserted into an ISMag1. These data suggest that the occurrence of a vpma cluster at locus II is a
rare event and has so far only been observed in 5632, while the occurrence of \textit{vpmas} at locus I is shared by all \textit{M. agalactiae} strains tested so far.

The presence of an IS\textit{Mag1} in close proximity to the \textit{vpma} locus I in 5632 is not an isolated event since it is shared by at least one other strain, 13628, which, unlike 5632, has been isolated in France in 2003 from a goat ear canal. IS\textit{Mag1} belongs to the IS30 family (31) that generates, when inserting itself into a host genome, a direct repeat that flanks the IS (25). Indeed, IS\textit{Mag1} of locus I is flanked in 5632 and in 13628 by an identical 14-nt directly repeated sequence ("b" in Fig 3) suggesting that its insertion most likely occurred in an ancestor common to the two strains. Interestingly, the 14-nt element found on each side of the IS of locus II\textsubscript{13375} ("a" in Fig. 3) is also repeated in locus II\textsubscript{5632} as if flanking the whole locus including the IS. Because no other 14-nt directed repeat is found in the vicinity of the two IS, this suggested the introduction of the \textit{vpma} cluster at locus II\textsubscript{5632} via a duplication-insertion mechanism involving a circular intermediate carrying a set of \textit{vpma} genes together with an IS\textit{Mag1} element (Fig. 4). In this scenario, the duplicated \textit{vpma} genes would have first clustered at one end of locus I via a series of DNA recombination. In a second step, part of locus I including the \textit{xer1} and IS\textit{Mag1} would have been duplicated and circularized to finally recombine at locus II via a single cross-over between the IS elements. One argument in favour of the occurrence of the circular form is that a PCR product was obtained with 5632 using outwardly oriented primers pv1R and xerF which partial sequencing revealed the presence of an IS\textit{Mag1} next to a sequence "c" usually found next to the promoter (Fig. 4). Whether this result really reflects a duplication-excision event of part of the \textit{vpma} locus I and whether the circular form can re-insert itself elsewhere in the genome have still to be formally demonstrated. Nevertheless, the 100\% nucleotide identity observed
between locus I and II of 5632 indicates that this duplication event as occurred relatively recently.

**Conclusion**

A number of sophisticated genetic systems devoted to high frequency surface variation have been described in mycoplasma species. Many of these systems appear as a small set of species-specific genes, which are nearly identical in all the isolates and are clustered at one locus. The clustering of these genes has been proposed to facilitate recombinations which generate diversity, and theoretically a wide surface diversity can be produced with only few genes, particularly as they are prone to internal rearrangements of their sequences. Our characterization of a new set of *vpma* genes in a *M. agalactiae* field isolate (5632) illustrates this point. This new repertoire is far more complex than the one first described in the PG2 type strain and represents a mosaic of structures that is suggestive of homologous recombination events between *vpma* genes, gene duplication and/or insertion-deletion of repeated motifs (see Fig. S3 in the supplemental material). This opens the question of the true diversity in equivalent variable systems from mycoplasmas with apparently few variable genes, most of which have been sequenced from type strains. Field or clinical isolates have the potential to reveal a greater than suspected diversity of variable gene repertoires, which could be useful to design molecular probes for epidemiology studies.

One of the most interesting finding of our study is the occurrence of duplicated *vpma* loci in strain 5632. Because one *vpma* locus can only express a single *vpma* gene per cell, the consequence of such event is tremendous in terms of variability because it allows concomitant expression and, in turns, multiplies the number of possible surface
combinations. Our data suggests that this situation is most likely resulting from the duplication of one 5632 vpmas region at a remote chromosomal position (250 kbp away) and that it involves IS elements. The role of such mobile genetic elements in chromosomal shuffling of genes encoding major surface components may have important evolutionary and epidemiological consequences for pathogens such as mycoplasmas that have a reduced genome and no cell wall.

Finally, whether a successful vaccine strategy based on surface antigens could be developed for mycoplasmas is not known but this would have to take into account the dynamics and scope of any variable antigen gene repertoires.

Acknowledgment

This work was supported by the French Ministry of Agriculture and Fisheries and by the National Institute for Agricultural Research (INRA). We thank Dr. Poumarat and the French National Network for Surveillance of Mycoplasmosis (AFSSA, Lyon, and VIGIMYC, France) for providing the strains. We also thank Dr. Blanchard for helpful discussion.

References


FIG. 1. Comparison of the *M. agalactiae* *vpma* loci between the type strain PG2 and the strain 5632.

Schematics represent the organization of the *vpma* loci in a clonal variant 55.5 derived from PG2 (16, 42) (Panel A), and in a clonal variant c1 derived from strain 5632 (Panel B). Panel C presents the counterpart of locusII 5632 in PG2 and shows the absence of *vpma* genes in this region. The presence of two distinct loci in 5632 was confirmed by PCR using primers xerF-phydR or xerF-agpR and the resulting amplicons are shown in Panel D. The location of the primers is indicated by arrowheads in panels A and B. White large arrows labelled by a letter represent Vpma CDSs. The position of the promoters is represented by a black arrowhead labelled with a “P”. The two non-Vpma related CDSs (*abiGI* and *abiGII*) are indicated by large arrows filled with a dotted pattern. Insertion sequence elements IS*Mag1* are indicated by hatched boxes. Recombination sites downstream of each *vpma* gene are indicated by a black dot. An asterisk (*) indicates that the corresponding *vpma* gene is present at two distinct loci. Schematics were approximately drawn to scale. HP: hypothetical protein; CHP: conserved hypothetical protein. Small letters and bars indicate the position of short particular sequences mentioned in the text and in Fig. 3 and 4.

The pictures on the left side of Panel A and B illustrate the variable surface expression of Vpma as previously described (8, 17). These correspond to colony immunoblot using Vpma-specific pAb recognizing PG2 VpmaW (α W) and VpmaY (α Y) epitopes.

FIG. 2. Structural features and comparison of *vpma* gene products in *M. agalactiae* strain PG2 and 5632.
Predicted Vpma proteins are schematically represented by boxes and begin with a homologous 25-aa leader sequence (black boxes) followed by regions that have homology between \textit{vpma} gene products or that are repeated within a same product (coloured boxes). Two boxes of a same colour display an amino acid identity > 30%. White boxes represent unique sequences. Numbers below boxes indicate the number of amino-acids. For 5632, detection by MS/MS of expressed Vpma specific peptides is noted by +; – indicates that no specific peptides were detected for the corresponding Vpma. For VpmaD2, +\# indicates that VpmaD2 peptides detected are not specific because all are shared with D1 or I (see the list of detected peptides in Table S3 in the supplemental material). An asterisk (*) indicates that the corresponding \textit{vpma} gene is present in 5632 at both \textit{vpma} loci.

**FIG. 3. Comparison of \textit{vpma} loci or their counterparts in \textit{M. agalactiae} strains PG2 55.5, 5632 c1, 13628, 13375 and 4025.**

Pictures represent the amplification of locus I and/or II using primers xerF/phydR, xerF/agpR and Mag2F/agpR which positions are indicated in the schematics below by arrow heads. Schematics represent the \textit{vpma} locus I and the \textit{vpma} locus II or counterpart. Brackets indicate the \textit{vpma} genes cluster localization next to the \textit{xer1} gene (filled black arrow). Insertion elements (\textit{ISMag1} or \textit{IS30-like}) are indicated by shaded boxes. Lines traced between primers symbolise PCR amplicons obtained with an asterisk (*) indicating those which have been directly sequenced. White bars below the loci indicate regions that were directly sequenced using genomic DNA. Small letters (a, b, c and d) represent the 14-nt sequences that flanked the IS.
FIG. 4. Scenario of duplication-insertion of 5632 *vpma* genes, from locus I to locus II via a circular intermediate

Schematics illustrate the putative scenario in which the *vpma* cluster of 5632 at locus II originates from locus I by a mechanism of duplication-insertion involving a circular intermediate and IS elements.

Small bars with a letter below represent 14-nucleotides sequences (a, b, c) flanking the IS*Mag1* either after insertion event (a, b) or after homologous recombination event (c).

The dotted arrow indicates the clustering process of seven *vpma* gene at one end of the locus I starting from the current configuration. It represents a series of an unknown number of possible DNA rearrangements. Asterisk (*) indicates duplicated *vpmas*.

The picture shows the PCR amplification of the circular intermediate obtained when using primers xerF and pv1R and 5632 crude DNA extract as template.
**TABLE 1. Analyses of *abiG* related genes products in mollicutes and other bacteria**

<table>
<thead>
<tr>
<th></th>
<th>AbiG I-like (MAGa8140)</th>
<th>AbiG II-like (MAGa8130)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mycoplasma agalactiae</strong></td>
<td><strong>5632</strong></td>
<td><strong>5632</strong></td>
</tr>
<tr>
<td><em>M. agalactiae</em> PG2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><em>M. bovis</em> PG45</td>
<td>+ (84.8% ; 93.6%)</td>
<td>+ (81.1% ; 91.3%)</td>
</tr>
<tr>
<td><em>M. mycoides mycoides SC</em> PG1</td>
<td>+ (40.8% ; 59.2%)</td>
<td>+ (41.8% ; 64.6%)</td>
</tr>
<tr>
<td><em>M. mycoides mycoides LC GM12</em></td>
<td><strong>Mmycm_04185</strong> + (41.7% ; 59.2%)</td>
<td><strong>Mmycm_04180</strong> + (31.9% ; 51.6%)</td>
</tr>
<tr>
<td></td>
<td><strong>Mmycm_00130</strong> + (40.1% ; 58.5%)</td>
<td><strong>Mmycm_00135</strong> + (32.6% ; 50.9%)</td>
</tr>
<tr>
<td><strong>Other Mollicutes</strong></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><em>Streptococcus agalactiae</em> H36B</td>
<td>+ (40.2% ; 58.3%)</td>
<td>+ (41.2% ; 65.3%)</td>
</tr>
<tr>
<td><em>Streptococcus suis</em> 05ZYH33</td>
<td>+ (41.2% ; 58.3%)</td>
<td>+ (40.4% ; 66.3%)</td>
</tr>
<tr>
<td><strong>Other than Streptococcus sp.</strong></td>
<td>+ (&lt;40% ; na)</td>
<td>+ (&lt;40%)</td>
</tr>
</tbody>
</table>

(+) = presence of homologous sequence; (-) = absence of homologous sequence in public databases (June 2008); identity and similarity percentages between amino-acid sequences are respectively indicated in brackets and were obtained using the Needle alignment program (http://mobyle.pasteur.fr/Mobyle/programs/needle.xml).
Figure 1
Figure 2

Identified as expressed by MS/MS
Locus I

PG2

5632

13628

13375

4025

Locus II

PG2

5632

13628

13375

4025

Figure 3
1. Clustering of duplicated vpma genes

2. Duplication and circularisation

3. Insertion by homologous recombination at locus II

locus I_{5632}

locus II_{5632}

Figure 4
FIG. S1. Multiple sequence alignment of *vpma* promoter regions in PG2 and 5632 loci I and II
Promoter region sequences, from -35 box to ATG start codon, of the three PG2 and 5632 *vpma* loci were aligned using ClustalW (http://mobyle.pasteur.fr/Mobyle/programs/clustalw.xml) Blue shading indicates the variable polyT stretches existing between -35 and -10 boxes shaded in dark grey.
FIG. S2. AbiGI-like and AbiGII-like phylogenetic trees
Neighbour-joining tree based on AbiGI-like (A) and AbiGII-like (B) homologous gene products among Mycoplasma and Streptococcus genus. AbiEI (A) and AbiEII (B) sequences encoded by pNP40 plasmid genes of Lactococcus lactis serve as the outgroup. Two AbiG-like pair of homologues present in Mycoplasma mycoides subsp. mycoides LC strain GM12, (a) and (b) annotations differentiate each pair. Bootstrap supporting values after 500 replicates are indicated. Sequence accession numbers: (A), NP_975179, ZP_02512914, ZP_02511961, AAK94955, YP_001198332, ZP_00783155, YP_001966453; (B), NP_975178, ZP_02512913, ZP_02511962, AAK94954, YP_001198333, ZP_00783153, YP_001966452.
FIG. S3. Imaging of whole sequence comparison between PG2 *vpma* locus and 5632 *vpma* locus I

Screenshot image obtain with Artemis comparison tool. Homologies between the two genome portions are represented by coloured boxes correspond to single match, with red representing a good match between positive strands and blue between negative strands. An example of the redundancy of the recombination site (RS) is illustrated in yellow, and shows the identity of PG2 RS with those of locus I of 5632. As well, the match between the *xerI* genes is also shown in yellow.