ISSN 1682-8356
ansinet.org/ijps

INTERNATIONAL JOURNAL OF POULTRY SCIENCE

ANSI*net*
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan
Mob: +92 300 3008585, Fax: +92 41 8815544
E-mail: editorijps@gmail.com

International Journal of Poultry Science 1 (4): 63-73, 2002
© Asian Network for Scientific Information 2002

Characterization of pMGA Genes from the F - Strain (Vaccine Strain) of *Mycoplasma Gallisepticum*

G. T. Pharr[1]*, S. L. Branton[2], L. A. Hanson[1], F. C. Minion[3], M. B. Hughlett[1] and X. Wan[1]
[1]Department of Basic Sciences, College of Veterinary Medicine, Mississippi State University, Mississippi State, MS 39762, USA
[2]USDA/ARS, South Central Poultry Disease Unit, Mississippi State, MS 39762, USA
[3]Veterinary Medical Research Institute, Iowa State University, 1802 Elwood Drive, Ames, IA 50011, USA
E-mail: pharr@cvm.msstate.edu

Abstract: Our long term goal is to determine the role that pMGA hemagglutinins play in *Mycoplasma gallisepticum* F-strain (termed as F-strain for brevity) persistence in poultry. To do this however, we first had to identify the pMGA genes expressed by an isolate of the F-strain. To identify F-strain pMGA genes, a genomic library was constructed in bacteriophage lambda and screened with polyclonal antisera raised against F-strain surface antigens capable of inhibiting agglutination of chicken red blood cells. One phage clone was identified, designated λ 9, which remained positive through additional rounds of screening and was sequenced. Sequence data analysis predicted one complete and two partial open reading frames (ORF) arranged in tandem within the 4.5 kb genomic insert, termed ORF 9.1, ORF 9.2 and ORF 9.3, which showed significant homology to the pMGA gene family of *M. gallisepticum*. The number of GAA trinucleotide repeats in the intergenic region of 9.3 ORF suggested that this ORF is expressed. The pMGA cDNAs amplified from the F-strain showed significant homology to the 9.3 ORF. One set of cDNAs however, differed from the 9.3 ORF sequence by a repeat of the sequence "AACCAAA" in the 5' end. We conclude that the F-strain vaccine isolate expresses pMGA gene variants, similar to what has been described in field strains of *M. gallisepticum*.

Key Words: *Mycoplasma gallisepticum* F - strain, pMGA-like cDNA, genomic library

Introduction

Species comprising the genus Mycoplasma are among the smallest living prokaryotes and are characterized by the absence of a cell wall and reduced genome size (Samuelsson and Boren, 1992). Mycoplasmas have limited biosynthetic capabilities, and exist as commensals or parasites of plants, vertebrates, and arthropods (Razin, 1992). Specifically in the chicken, virulent strains of *M. gallisepticum* cause a chronic respiratory disease syndrome in chickens and severe sinusitis in turkeys (Ley and Yoder, 1997). Chronic respiratory disease in chickens is characterized by exfoliation of ciliated epithelial cells, accumulation of an inflammatory exudate in the trachea, and a severe air-saculitis (Dykstra *et al*., 1985; Ley and Yoder, 1997). In the reproductive system, virulent *M. gallisepticum* induce inflammatory lesions leading to the formation of large caseous plugs in the oviduct (Pruthi and Kharole, 1981). Primary control of disease is through vaccination using live attenuated *M. gallisepticum* (Whithear, 1996).

The F strain of *M. gallisepticum* (termed F-strain for brevity) is used as a live vaccine; it protects the respiratory tract of vaccinates (Abd-El-Motelib and Kleven, 1993), and alleviates egg production decreases (Branton *et al*., 1997). In addition, the F-strain displaces field strains of *M. gallisepticum* from infected animals (Kleven *et al*., 1998). Our long term goal is to understand the mechanism of host adaptation of the F-strain to poultry, which may be related to the efficacy of the F-strain vaccine. We postulate the mechanism of host adaptation is due to surface proteins expressed by the F-strain.

One prominent class of surface proteins expressed by *M. gallisepicum* is encoded by the pMGA gene family. The pMGA gene family encodes hemagglutinin lipoproteins, initially identified using monoclonal antibodies (mAbs) which inhibited the hemagglutination of chicken erythrocytes by *M. gallisepticum* (Markham *et al*., 1992). The mAbs recognized a 67 kDa lipoprotein from the virulent field strains S6 and R and a 75 kDa lipoprotein expressed by the F-strain. Peptides derived from the 67 kDa protein were sequenced and degenerate oligodeoxynucleotides were deduced from the amino acid sequence for use as hybridization probes. The clones derived from screening a genomic library prepared from *M. gallisepticum* S6 DNA revealed a number of open reading frames in tandem, separated by intergenic regions of about 350 bp (Markham *et al*., 1993). Additional studies identified 5 pMGA genes, termed pMGA1.1 - pMGA1.5 (Markham *et al*., 1994). A genomic survey of several *M. gallisepticum* strains indicated that the number of pMGA genes varied from 32 genes in F-strain, to 33 and 70 genes in the virulent S6 and R strains, respectively (Baseggio *et al*., 1996).

We propose that adaptation of the F-strain to poultry is due to the spectrum of pMGA hemagglutinin genes expressed, which may play an important role in the binding of F-strain to host epithelia. Thus, it is critical to identify the pMGA genes expressed by the F-strain isolate that has been used in vaccine studies for a number of years (Branton *et al*., 1984). In this report we identify pMGA mRNAs expressed in the F-strain. This information will be used in the future to evaluate pMGA gene translation and function of the products in F-strain host colonization.

Materials and Methods

Mycoplasma: F-strain (gift of S. H. Kleven, University of Georgia) was propagated as described (Frey *et al*., 1968). F-strain cells after 24 h culture, were washed three times in phosphate buffered saline (8.12 g Na_2HPO_4, 8.7 g NaCl, and 1.88 g NaH_2PO_4 per liter) and used in hemagglutination inhibition experiments and for genomic DNA extraction (below).

Genomic library construction: Genomic DNA was isolated from 24 h F-strain cultures as described (Keeler *et al*.,1996). Cells were digested using Proteinase K (Sigma Chemical Co.), and DNA extracted using phenol and chloroform, followed by ethanol

precipitation. Genomic DNA (200 ug) was partially digested with restriction enzyme Tsp 509 I (0.25 units/ug DNA, New England Biolabs, Inc., Beverly, MA), and then purified by phenol/chloroform followed by ethanol precipitation (Minion *et al*., 1995). The digested DNA was then size - fractionated by centrifugation on sucrose gradients as described (Luthe, 1983). DNA fragments (5 - 8 kb) were ligated into the *Eco* RI restriction enzyme site of the λ ZAPII, bacteriophage λ vector, followed by packaging to produce infectious particles with Gigapack Gold III packaging extracts according to the manufacturer's protocol (Stratagene, La Jolla, CA). The unamplified library had a titer of 3 x 106 pfu/ml, with 1% nonrecombinants and an average insert size of 5 kb.

Library screening: The library was plated with an equal mixture of *E. coli* strains ISM612 and XL1-blue. ISM612 has two UGA suppressors, a prfB3 mutation blocking release factor function, and *Trp T* under IPTG-inducible control on a plasmid which will allow expression of mycoplasma genes containing the UGA codon (Smiley and Minion, 1993; Minion *et al*., 1995). Plates containing bacteria infected with recombinant phage (50,000 pfu) were overlaid with nitrocellulose filters previously soaked with IPTG (to induce expression from the *lacZ* promoter). For immunoscreening, a rabbit anti - F-strain antiserum with hemmagglutination - inhibition activity was used (May *et al*., 1988). Duplicate filters from each plate were incubated with a 1:1024 dilution of the rabbit antiserum, washed five times in TBST (20 mM Tris-Cl pH 8.0, 150 mM NaCl, and 0.05% Tween 20), and then incubated with goat anti-rabbit immunoglobulin (conjugated with alkaline phosphatase). Positive signals were visualized by incubation of the filters with nitroblue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate according to the manufacturer 's protocol (Stratagene).

RT-PCR: The expression of pMGA genes at the mRNA level was evaluated as described (Glew *et al*., 1995). Total RNA was prepared from 24 h cultures of F-strain using Trizol (Invitrogen, Carlsbad CA), treated with RNase-free DNAse I at 37 C (Waldo *et al*., 1999). DNAse I was removed by incubation with a protein-binding reagent (Ambion Inc, Austin TX). Total cDNA was synthesized from RNA primed with random hexamers and reverse transcribed with SuperScript II reverse transcriptase (Invitrogen). Control samples for the reverse transcription were;

1. DNAse I, RNAse A, and reverse transcriptase only,
2. RNA only,
3. RNA with DNAse I, RNAse A, and reverse transcriptase,
4. RNA with DNAse I only, and
5. 100 ng of F-strain genomic DNA with or without DNAse I.

The pMGA cDNAs were amplified using the following primers (Genosys Biotechnologies, The Woodlands, TX). Forward primer (pMGA1), 5'-AGTCCCGGGAGTGAAAAGAAAAAACATTTTAAAG-3' complementary to the conserved 5' region of all pMGA genes analyzed to date, nucleotide positions 967 - 990 and 3313 - 3336 of the 9.2 and 9.3 ORF, respectively (GenBank: AF210770); underlined region is a *Sma* I restriction enzyme site. The reverse primer was a degenerate pMGA primer (T23), 5'-GTTAGAATTCTTYTWGCRGCWGYTAAT-3' (Glew *et al*., 1995) where Y=C+T, W=A+T, and R=A+G; underlined region is an *Eco* RI restriction enzyme site. T23 is complementary to nucleotide positions 1197 - 1213 and 3543 - 3559 of the 9.2 and 9.3 orf, respectively (GenBank: AF210770). An additional reverse primer (128A), 5' -CGTCTTTTCGCTTATTAGTT-3', complementary to nucleotide positions 3994 - 4014 of the 9.3 orf sequence of the phage clone λ9 (GenBank: AF210770). To amplify the tuf gene of *M. gallisepticum* (Inamine *et al*., 1989); forward primer 5'-GACACCACGCCATGGCAAAAGAAAGGTTC-3' and the reverse primer 5'-AGCACCCTCGAGAGCACCTACAGTTCTACCACC-3'; underlined regions are *Nco* I and *Xho* I restriction enzyme sites respectively. The tuf gene forward and reverse primers are complementary to nucleotide positions 29 - 109 and 1229 - 1249 of the tuf gene sequence, respectively (Inamine *et al*., 1989). The tuf gene - specific primers were used to assay cDNA synthesis quality and removal of contaminating genomic DNA from RNA. The PCR consisted of 2 mM magnesium chloride, 0.2 mM each dNTP, 0.2 mM each primer, and 0.3 units of Taq Gold polymerase (Applied Biosystems, Foster City CA) in water for a 50 ul final volume. PCR was conducted using a Hybaid Omn-E thermocycler (Scientific Consultants Inc, Baton Rouge LA), conditions: 94 C, 10 min; 30 cycles [94 C, 30 sec; 55 C, 30 sec, ; 72 C, 3 min; final extension at 72 C, 5 min]. The PCR products were analyzed on 1% agarose gels stained with ethidium bromide for visualization of DNA fragments.

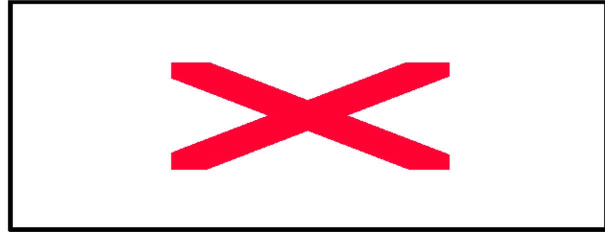

Fig. 1: Structure of the λ 9 phage clone compared to pMGA genes in the genome. The pMGA genes are arranged in tandem on the *M. gallisepticum* chromosome, separated by intergenic regions containing promoter elements (Markham *et al*., 1994). The direction of transcription is denoted by arrows.

Nucleotide sequence analysis of library phage and cDNA: The insert from the phage clone, remaining positive through additional rounds of screening with the rabbit antisera, was sequenced using dye terminator cycle sequencing using pBluescript phagemid promoter primers (Stratagene, La Jolla, CA) or internal primers (Genosys, The Woodlands TX). Sequencing reaction products were analyzed on an ABI Prism 310 Genetic Analyzer (Applied Biosystems). The cDNAs obtained from the PCR experiments were ligated into pT7Blue vector and sequenced with the T7 and U19 vector primers (Novagen, Inc. Madison, WI). The phage insert and the PCR amplified products were completely sequenced on both strands. Sequence data were analyzed with MacDNASIS Pro V3.1 software (Hitachi Software Engineering America, Ltd, San Francisco, CA) and using the GenBank database (National Center for Biotechnology Information). Alignments of predicted amino acid sequence were created with the Clustal W version 1.7 program (Thompson *et al*., 1994).

Results

Genomic library screening: To detect F-strain pMGA genes, a rabbit ant - F-strain antiserum with hemagglutination - inhibition activity was used in immunoscreening. Immunoscreening of phage-library-lifts with the antiserum yielded a phage clone, termed λ9, which remained positive through additional rounds of screening, and was characterized by nucleotide sequencing (Genbank accession number AF210770). Analysis of sequence data predicted three ORF arranged in tandem within the 4.5 kb insert, termed 9.1, 9.2, and 9.3 (Fig. 1).

```
pMGA1.1     IYDDIFGNSVTTENNRTIISVDALNGYSLASDWSTFIAEYSGAGLTLNDQAKPNEKYYLI
pMGA1.2     IYDDIFGNSVTTKNNRTIISVDALNGYSLASDWSTYIAEYSGAGLTLNDQAKPNEKYYLI
M9          IYDDIFGNSVTTENNRTIISVDALNGYSLASDWSTYIAEYSGAGLTLNDQAKPNEKYYLI
9.3 orf     IENSIFGNSVTTENNITKISVDTLSAYSLASDWSTFIGQYSSDSLTLNGNRMSDQKYYLI
pMGA1.2HS   IENSIFGNSVTTKNNITKISVDTLSAYSLASDWSTFIGQYSSDSLTLNGNRISDQKYYLI
9.1 orf     VYNDIFSNILDKQDEPNSVTVNLLKGYSLAADHSTYFYQFS--SSN---RMNESNPTYLV
pMGA1.1HS   VYNDIFSNILDKQDEPTSVTVNLLKGYSLAADHSTYFYQFS--SSN---GMNESNPTYLV
9.2 orf     VYNSIFGNVDNSSEASTYVTVDILKGYSLATNWSTYVTRFM--NLT-N-SMPENATTYLV
pMGA1.4     IYDDLFGNNSVQQDNQTAVTVDLLKGYSLATNYSIFVRRFM--GLQEN-SMTRTDPIYLV
            : :.:*.*    .:  . ::*: *..****:: * :. .:   .          **:

pMGA1.1     GYVGG-TGARNDMMVPK--NNVQKFPLANNTSNRNYVFYVNAPREGDYYIKGVFASGVG-
pMGA1.2     GYVGG-TGARNDMMVPK--NNVQKFPLANNTSNRNYVFYVNAPREGDYYIKGVFASGVG-
M9          GYVGG-TGARNDMMVPK--NNVQKFPLANNTSNRNYVFYVNAPKAGDYYIKGVFASGVH-
9.3 orf     GYVGGNTGQRDITMVAN--TNQQRLPTASNQNTRSYTLYVNAPKAGAYYIKGVFASEVR-
pMGA1.2HS   GYVGGNTGQRNITMVAN--TNQQRLPTASNQNTRSYTFYVNAPKAGAYYIKGVFTSEVR-
9.1 orf     GFIGG-HGNRRNNLNSSNTTNNEVASPSVQT-SNRTLTIYVNAPKDGQYYIKGSYLTSNN-
pMGA1.1HS   GFIGG-RGNRRNNLNSSVTADNKVASPSFQT-SSRTLTIYVNAPKDGQYYIKGSYLTSDN-
9.2 orf     GFIGG-QLARTTVGS---IPNRNNFPIMNN-ENRTFTLYVNAPKAGDYHISGSYLTRNT-
pMGA1.4     GYIGG-SLDRLPRANRSKVQNFNNSPTQNN-NTRTFTIYVNAPVEGNYYVSGSYLFSSSQ
            *::**    *            *    *   .. ..*. .:*****  * *::.* :

pMGA1.1     SDLKFSTGDMS--SNNVTVKQLFTGNLTTTLRTFDTSATT---ESTRVTTDPTNK-KTLT
pMGA1.2     SDLKFSTGDMS--SNNVTVKQLFTGNLTTTLRTFDTSATT---ESTRVTTDPTNK-KTLT
M9          SDLKFSTGDMS--SNNVTVKQLFTGNLTTTLRTFDTSATT---ESTRVTTDPTNK-KTLT
9.3 orf     RDLKFSTGDMS--SNNVTIRQLSTGNLTT-LKTFDTSAIT---GPTQVTTVDTNR-KTLT
pMGA1.2HS   RDLKFSTGDMS--SNNVTIQQLTTGNLTT-LKTFDTSATE---GPTRVTTVDTNR-KTLT
9.1 orf     RNLKFTTTATA--NNSITFTVKGKNNWST-LGTFNTANNNDIETSGSSSSGQPNESKTIK
pMGA1.1HS   RNLKFTTTATA--NNSITFTVKGKNNWST-LGTFNTANNNDIETSGSSSSGQANESKTIK
9.2 orf     RGLKLTVTDTTDKNNSITITTSGKNNWNT-LGHFDTSKAN--NSNGNDGSVENNK-ASLT
pMGA1.4     TNTQRGLKFLIDGNNAVSITVQRQVDWNT-LGAFDTSKTN--NQDGNSSSVVGNI-KTLR
             . :          .* ::.       : .* *  *:*:              :   *   ::

pMGA1.1     LVEGLNKIVVSGTTENIG---APNFGYLEFILN------ETQPETTNVSNPS--
pMGA1.2     LVEGLNKIVVSGTTENIG---APNFGYLEFILN------ETQPETTNVSNPS--
M9          LVEGLNKIVVSGTTENIG---APNFGYLEFILN------ETQPETTNVSNPS--
9.3 orf     LVEGLNKIVVSGATADNGN--APNFGYLEFILN------ETQPETT--------
pMGA1.2HS   LVKGLNKIVVSGATANNGN--APNFGYLEFILN------ETQS-----------
9.1 orf     LNKGLNKVVITSVMMDNKNPGAPYIGNLTFTLM-----SPTMMEAKK-------
pMGA1.1HS   LNKGLNKVVIGSVMINNRYPGAPYIGNLKFTLM-----SPTMMEAKK-------
9.2 orf     LKEGLNKIVIAGGTQDGKN--APYIGNLTFTLN----NSSTNASQDSSST----
pMGA1.4     LGKGLNKIIISGGTQDGTN--APYIGNLTFKLMTTTSNSETNTPAEGTSTEHAK
            * :****::: .    :      ** :* * * *         *
```

Fig. 2a: Alignment of predicted amino acid sequences of the 9.1 orf and the C-terminus of the 9.2 and 9.3 orf with published pMGA proteins.

```
9.3 orf        VKRKNILKFVSLLGIGSFVMVAAASCTSAATPTP----NPEPN--PNPEPKPDPMPNPPS
pMGA1.2HS      MKRKNILKFVSLLGIGSFVMLAAASCTSATTPTP----NPEPKPTPNPEPKPDPMPNPPS
pMGA1.2        MKKKNILKFVSLLGIGSFVMLAAASCT---TPTP----NPTPN------------PNPPS
M9             MKRKNILKFVSLLGIGSFVMLAAASCT---TPTP--------N------------PNPPS
pMGA1.1        MKRKNILKFISLLGIGSFVMLAAASCT---TPTP----SPAPN------------PNPPS
9.2 orf        VKRKNILKFVSLLGIGSFVMLAAASCTSATTPTP----NPEPKP--------DPMPNPPS
pMGA1.4        MKRKNILKFVSLLGIGSFVMLAAASCTTPVNPTPKPKPNPEPKPNPAP--KPDPMPNPGG
               :*:******:**********:******   .***        :            *** .

9.3 orf        ---GGMNGGDINPGGGQNMMDSAAQELTAARTALTSLLASKNANIEMYSDYAKIQNTLIA
pMGA1.2HS      ---GGMNGGDINPGGGQNMMDSAAQELTAARTALTSLLASKNANVEMYSDYAKIQNTLIA
pMGA1.2        ---GGMNGGDTNPGDGQGMMNAASQELAAARMGLTTIFDSKAKNLGLYVDYKKTQNTLTK
M9             ---GGMNGGDTNPGDGQGMMNAASQELAAARMGLTTVFDSKAKNLGLYVDYKKTQDTLTK
pMGA1.1        N--GGMNGGNINPGDGQGMMNAAAQELAAARMGLTTVFDSKAKNLGLYVDYKKTQNTLTK
9.2 orf        ---GDMNGGDTNPGND-GGMENSAQQLAAAKKELSDLLATQSSNLAKYADYTNIQNTLTA
pMGA1.4        GMMGGMNGGNTNPGNG-GGTDNAAQQLAAAKKELSDLLATQNSNLSTYADYANIQNTLTA
                  *.****: ***.. .  : ::*:*:**:  *: :: ::  *:  * ** : *:**

9.3 orf        AYTTAEQTSQNSSATLEQVKNATSALQTAINTANSNKQKFDQDHSNLLMSYKNLMATLAK
pMGA1.2HS      AYTTAEQTSQNSSATLEQVKNATSALQTAINTANSNKQKFDQDHSNLLMSYKNLMATLAK
pMGA1.2        AYDAAKTVLDNSSSTTQNLNEAKTRLETAIRTAATSKQTFDEQHAELVKVYKELKTTLSN
M9             AYDAAKTVLDNSSSTTQNLNEAKTRLETAIRTAATSKQTFDEQHAELVKVYEELKTTLSN
pMGA1.1        AYDAAKTVLDNSSSTTQNLNEAKTRLETAIRTAATSKQTFDEQHAELVKVYKELKTTLSN
9.2 orf        AYTTAKSTSDNTSVTLEQVKSATSTLQAAIDTAASSKTSFDEKNPELIKAYYALKETLKN
pMGA1.4        AYTTAKSTSDNTSATLEQVKSATSTLQTAIDTAASSKTSFDEKNPELIKAYNALKETLKK
               ** :*: . :*:* * ::::.*.: *::** ** :.* .**:.:.:*:  *  *  ** :

9.3 orf        KETTVMTLKDPKYSAILDQINGVSCKGEELVQHTLDPVSGIVPAANTITEEITKIEEVIS
pMGA1.2HS      KETAVMTLKDPKYSAILDQINGVSSKGEELVQHTLDPVSGIVPAANTITEEITKIEEVIS
pMGA1.2        ETATLAPYADAQYAGIKMHLSGLYDAGKAITTKTLEPVEGDPLTASAVMMANTKIVEAIK
M9             ETATLAPYAAAQYAGIKMHLSGLYDAGKAITTKTLEPVEGDPLTADVVMMANTKIVEAIK
pMGA1.1        ETATLAPYADAQYAGIKMHLSGLYDAGKAITTKTLEPVEGDPLTAGAVTMANTKIVEAIK
9.2 orf        EETVLSGLTDSNFATIKTNLTALYQSGKDFVKATLDPVSGNAPQIADITKADKDIADAVS
pMGA1.4        WRNSLSGLTDSNFATIKTNLTALYQSGKDIVTKTLDPLMGTAINLSAVSQANTNISNAVS
                   :     .::: *  ::..:   *: :.  **:*: *       :    ..* :.:.

9.3 orf        EKTLQDQKNNADQFAN-YQSFTLDKTKLENVEDAKKMGQPANYSFVGYSVDVTGTSGQET
pMGA1.2HS      EKTLQDQKNNADQFAN-YQSFTLDKTKLENVEDAKKMGQPANYSFVGYSVDVTGTSGQET
pMGA1.2        DEVLNPQKENATKLADSFVKQVLVKEKITGVEEAHNKAQPANYSFVGYSVDITGTTTGQT
M9             DEVLNPQKENATKLADSFVKQVLVKEKITGVEEAHNKAQPANYSFVGYSVDITGTVTGQT
pMGA1.1        DEVLNPKKENATKLADSFVKQVLVKEKITGVEEAHNKAQPANYSFVGYSVDITGTANGQT
9.2 orf        K--LETWKTNANTLATSFVKEVLVKNKLTGIDTTNNREQPGNYSFVGYSVNATN----NN
pMGA1.4        K--LETWKTNATVLATSFVKEVLVKNKLTGIDTTNNQEQPGNYSFVGYSVDVTTGS--DN
               .  *:  * **  :*  : . .* * *: .:: :::  **.*********: *     :.

9.3 orf        TIPNWNFAQRAIFTSGNQPTKVTATTTGEDQSTAKPLSDVSWIYSLAGTGAKYTLEFTYY
pMGA1.2HS      TIPNWNFAQRAIFTSGNQPTKVTATTTGEDQSTAKPLSDVSWIYSLAGTGAKYTLEFTYY
pMGA1.2        SIPNWDYAQRTIFTNSDEPRSISNTPADG-QTMAQPLSNVSWIYSLAGTGAKYTLEFTYY
M9             SIPNWDYAQRTIFTNGDEPRSISNTPADG-QTMVQPLSNVSWIYSLAGTGAKYTLEFTYY
pMGA1.1        SIPNWNYAQRTIFTNGDEPRSVSNTPVDG-QTMAQPLSNVSWIYSLAGTGAKYTLEFTYY
9.2 orf        EIPNWNFAQRKVWTSDNGRTSLISSTS----DNSSTLTEVSWIYSLSGAGTKYSLTFNYY
pMGA1.4        ARPNWSFAQRKVWTSN---TDILSQPQPAEGENQQSAPDVSWIYNLTGMGAKYSLTFNYY
                 ***.:*** ::*..    .:   .        .. .:*****.*:* *:**:* *.**
```

```
9.3 orf     GPSTGWLYFPYKLVKAND--DVGLQYKLNSN---------ETLTPIIFGEGTTTNG----
pMGA1.2HS   GPSTGWLYFPYKLVKAND--DVGLQYKLNSN---------ETLTPIIFGEGTTTNG----
pMGA1.2     GPSTGYLYFPYKLVNTSDQVKLGLEYKLNDA---------TKPSAITFGSDQTMNG----
M9          GPSTGYLYFPYKLVNTSDQMKLGLEYKLNDA---------TEPSAITFGSEQTMNG----
pMGA1.1     GPSTGYLYFPYKLVNTSDQMKLGLEYKLNDA---------TEPSAITFGNEQTMNG----
9.2 orf     GPSTGYLYFPYKLVKEGDENNVALQYTLNSGSAQEVNFAPTVKTSVSADSSGDSNNQTES
pMGA1.4     GPSTGFLYFPYKLVNSSDSDKVALEYKLNESAVKTIDFSPSQTSPVASDATRE-NN--RS
            *****:********: .*  .:.*:*.**.               :.:  .     *.

9.3 orf     ------------PAATVENINVAKVRLTGLAFGKNTIEFSAPM------SKVAPMIGNMY
pMGA1.2HS   ------------PAATVENINVAKVRLTGLAFGKNTIEFSVPM------SKVAPMIGNMY
pMGA1.2     ------------KTPTVNDINVAKVTLANLNFGSNKIEFSVPA------EKVSPMIGNMY
M9          ------------KTPTVNDINVAKVTLANLKFGSNKIEFSVPA------EKVSPMIGNMY
pMGA1.1     ------------KTPTVNDINVAKVTLANLIFGSNKIEFSVPA------EKVSPMIGNMY
9.2 orf     AAETMPVTSDLNPAPTVSDINIAKLTLSNLKFGSNTIEFSVPT---EPSNKVAPMIGNMY
pMGA1.4     TAAPAQGSTEINPAPTLDDIKIAKVTLSNLKFGSNTIEFSVPTTAKEGTSKVAPMIGNMY
                         :.*:.:*::**: *:.* **.*.****.*       .**:*******

9.3 orf     LTSSDTETNKQNIENSIFGNSVTTENNITKISVDTLSAYSLASDWSTFIGQYSSDSLTLN
pMGA1.2HS   ITSSDTETNKQNIENSIFGNSVTTKNNITKISVDTLSAYSLASDWSTFIGQYSSDSLTLN
pMGA1.2     LSSSPNNWNK--IYDDIFGNSVTTKNNRTIISVDALNGYSLASDWSTYIAEYSGAGLTLN
M9          LSSSPNNWNK--IYDDIFGNSVTTENNRTIISVDALNGYSLASDWSTYIAEYSGAGLTLN
pMGA1.1     LSSSPNNWNK--IYDDIFGNSVTTENNRTIISVDALNGYSLASDWSTFIAEYSGAGLTLN
9.2 orf     LTSN--IANEAKVYNSIFGNVDNSSEASTYVTVDILKGYSLATNWSTYVTRFMNLT----
pMGA1.4     LTSSDRDVNKNKIYDDLFGNNSVQQDNQTAVTVDLLKGYSLATNYSIFVRRFMGLQE---
            ::*.     *:  : :.:***     .:  * ::** *..****:::* :: .: .

9.3 orf     GNRMSDQKYYLIGYVGGN---TGQRDITMVANTNQQRLPTASNQNTRSYTLYVNAPKAGA
pMGA1.2HS   GNRISDQKYYLIGYVGGN---TGQRNITMVANTNQQRLPTASNQNTRSYTFYVNAPKAGA
pMGA1.2     DQAKPNEKYYLIGYVGG----TGARNDMMVPKNNVQKFPLANNTSNRNYVFYVNAPREGD
M9          DQAKPNEKYYLIGYVGG----TGARNDMMVPKNNVQKFPLANNTSNRNYVFYVNAPKAGD
pMGA1.1     DQAKPNEKYYLIGYVGG----TGARNDMMVPKNNVQKFPLANNTSNRNYVFYVNAPREGD
9.2 orf     NSMPENATTYLVGFIGG---QLARTTVGSIPNRN--NFPIMNN-ENRTFTLYVNAPKAGD
pMGA1.4     NSMTRTDPIYLVGYIGGSLDRLPRANRSKVQNFN--NSPTQNN-NTRTFTIYVNAPVEGN
             ..        **:*::**                : : *  . *  .* ..*.:.:*****  *

9.3 orf     YYIKGVFASE-----VRRDLKFSTGD--MSSNNVTIRQLSTGNLTT-LKTFDTSAITGPT
pMGA1.2HS   YYIKGVFTSE-----VRRDLKFSTGD--MSSNNVTIQQLTTGNLTT-LKTFDTSATEGPT
pMGA1.2     YYIKGVFASG-----VGSDLKFSTGD--MSSNNVTVKQLFTGNLTTTLRTFDTSATTEST
M9          YYIKGVFASG-----VHSDLKFSTGD--MSSNNVTVKQLFTGNLTTTLRTFDTSATTEST
pMGA1.1     YYIKGVFASG-----VGSDLKFSTGD--MSSNNVTVKQLFTGNLTTTLRTFDTSATTEST
9.2 orf     YHISGSYLTR----NT-RGLKLTVTDTTDKNNSITITTSGKNNWNT-LGHFDTSKANNSN
pMGA1.4     YYVSGSYLFSSSQTNTQRGLKFLI----DGNNAVSITVQRQVDWNT-LGAFDTSKTNNQD
            *::.* :           .  .**:         .* :::        : .* *  ****

9.3 orf     -QVTTVDTNRKTLTLVEGLNKIVVSGATADNGNAPNFGYLEFILN----ETQPETT----
pMGA1.2HS   -RVTTVDTNRKTLTLVKGLNKIVVSGATANNGNAPNFGYLEFILN----ETQS-------
pMGA1.2     -RVTTDPTNKKTLTLVEGLNKIVVSGTTENIG-APNFGYLEFILN----ETQPETTNVSN
M9          -RVTTDPTNKKTLTLVEGLNKIVVSGTTENIG-APNFGYLEFILN----ETQPETTNVSN
pMGA1.1     -RVTTDPTNKKTLTLVEGLNKIVVSGTTENIG-APNFGYLEFILN----ETQPETTNVSN
9.2 orf     GNDGSVENNKASLTLKEGLNKIVIAGGTQDGKNAPYIGNLTFTLN----NSSTNASQDSS
pMGA1.4     GNSSSVVGNIKTLRLGKGLNKIIISGGTQDGTNAPYIGNLTFKLMTTTSNSETNTPAEGT
             .  :    *   :* * :*****:::* * :    ** :* * * *       ::..
```

Fig 2b: Alignment of predicted amino acid sequences of the complete 9.2 and 9.3 orf with published pMGA proteins. The pMGA1.1, pMGA1.2, and pMGA1.4 proteins are from the *M. gallisepticum* S6 strain (Markham *et al.*, 1994). The pMGA1.1 and pMGA1.2 protein from the *M. gallisepticum* HS strain (GenBank accession number AF275312) are denoted as pMGA1.1HS and pMGA1.2HS in the figure. The pMGA M9 protein is from the *M. gallisepticum* PG31 strain Liu *et al* . (1998). Identical amino acids are designated by an "*". Conservative amino acid substitutions are designated by "." and : ":" Dashes indicate gaps introduced to maximize homology.

```
9.3 orf     TGTAAAACTCATAAAATCTTTTGTTTTC-----------GAAGAAGAAGAAGAAGAAGAA
M9          -ATAAAAACTGATAAAATCTTTTGTTTTC----------GAAGAAGAAGAAGAAGAAGAA
pMGA1.1     -TTAAAAACATACAAAACCTTTATTTAGC----------GAAGAAGAAGAAGAAGAAGAA
9.2 orf     TATAAATATCAACAAAACCTTTTGTTTTCCGAAGAAGAAGAAGAAGAAGAAGAAGAAGAA
              **** *   * **** ****  **  *          *********************


9.3 orf     GAAGAAGAAGAAGAAGTTCTTAGAAGTTTTGGGGTTTGGGAATCCTGTGATCAGCGAAAA
M9          GAAGAAGAAGAAGAAGTTCTTAGGAGTTCTGGGGTTTGGGGCTGGTTTGATCAGTGAAAA
pMGA1.1     GAAGAAGAAGAAGAAGTTCTTAGGAGTTCTGGGGTTTTGGTTTGGCTTGATGAGCGAAAA
9.2 orf     GAAGAAGAAGAAGAAGTTCTTAGGAGTTCTGTGGTCTGGGGTTGGTTTGATCAGCGAAAA
            *********************** **** ** *** * **  *    **** ** *****


9.3 orf     TTAAGCCGATTTATTCCTTACT-GAACTTTATATATTC-TTTAGATAATAA-TAGACGTG
M9          TTAAGCAGATTTATTACTTACT-GAACTTTATATATTC-TTATATTAATAA-TAGACGTG
pMGA1.1     TAAACCCGATTTATTACTTACT-GAACTTTATATATTC-TTAAATTAATAA-TAGACGTG
9.2 orf     TAAACCCGATTTATTACTTAATTGAACTTTATATATTCCTTAACTTGATGAGTATGTATT
            * ** * ******** **** * *************** **    * ** * **    *

                                                   (GAA)n
9.3 orf     GTGAACGTAAGTT-ATTGA-TTAACTTTAAGTG        12
M9          TTTAACGTAAGTT-ATTGGCTTAACTTTAAGTG        12
pMGA1.1     TTTAACGTAAGTT-ATTGGCTTAACTTTAAGTG        12
9.2 orf     TTTCATACAAATTCATCGATTCAACTTTAAGTG        15
             *  *   ** ** ** *  * ***********
```

Fig. 3: Nucleotide sequence alignment of the 9.2 and 9.3 intergenic regions with the intergenic regions of pMGA genes pMGA1.1 and M9 which are expressed at the protein level in M. gallisepticum strains S6 (Markham *et al*., 1994) and PG31 (Liu *et al*., 1998), respectively. Identical nucleotides are designated by an "*". Dashes indicate gaps introduced to maximize homology.

```
            3337                                                    3396
9.3 orf     TTTGTTAGTTTATTAGGTATTGGTTCGTTTGTAATGGTAGCTGCAGCTAGTTGTACTTCA
1-1 cDNA    TTTGTTAGTTTATTAGGTATTGGTTCGTTTGTAATGGTAGCTGCAGCTAGTTGTACTTCA
1-2 cDNA    TTTGTTAGTTTATTAGGTATTGGTTCGTTTGTAATGGTAGCTGCAGCTAGTTGTACTTCA
1-3 cDNA    TTTGTTAGTTTATTAGGTATTGGTTCGTTTGTAATGGTAGCTGCAGCTAGTTGTACTTCA
1-4 cDNA    TTTGTTAGTTTATTAGGTATTGGTTCGTTTGTAATGGTAGCTGCAGCTAGTTGTACTTCA
            ************************************************************

            3397                                                    3450
9.3 orf     GCAGCTACACCAACTCCAAACCCTGAACCAAA------TCCAAACCCTGAACCAAAACCA
1-1 cDNA    GCAGCTACACCAACTCCAAACCCTGAACCAAA------TCCAAACCCTGAACCAAAACCA
1-2 cDNA    GCAGCTACACCAACTCCAAACCCTGAACCAAAACCAAATCCAAACCCTGAACCAAAACCA
1-3 cDNA    GCAGCTACACCAACTCCAAACCCTGAACCAAA------TCCAAACCCTGAACCAAAACCA
1-4 cDNA    GCAGCTACACCAACTCCAAACCCTGAACCAAAACCAAATCCAAACCCTGAACCAAAACCA
            ********************************      **********************

            3451                                                    3510
9.3 orf     GATCCAATGCCAAACCCTCCTAGTGGTGGTATGAATGGTGGAGATATTAATCCAGGTGGT
1-1 cDNA    GATCCAATGCCAAACCCTCCTAGTGGTGGTATGAATGGTGGAGATATTAATCCAGGTGGT
1-2 cDNA    GATCCAATGCCAAACCCTCCTAGTGGTGGTATGAATGGTGGAGATATTAATCCAGGTGGT
1-3 cDNA    GATCCAATGCCAAACCCTCCTAGTGGTGGTATGAATGGTGGAGATATTAATCCAGGTGGT
1-4 cDNA    GATCCAATGCCAAACCCTCCTAGTGGTGGTATGAATGGTGGAGATATTAATCCAGGTGGT
            ************************************************************

            3511                         3543
9.3 orf     GGTCAAAATATGATGGATTCTGCAGCTCAAGA
1-1 cDNA    GGACAAAATATGATGGATTCTGCAGCTCAAGA
1-2 cDNA    GGACAAAATATGATGGATTCTGCAGCTCAAGA
1-3 cDNA    GGACAAAATATGATGGATTCTGCAGCTCAAGA
1-4 cDNA    GGACAAAATATGATGGATTCTGCAGCTCAAGA
            ** ******************************
```

```
          3337                                                        3396
9.3  orf  TTTGTTAGTTTATTAGGTATTGGTTCGTTTGTAATGGTAGCTGCAGCTAGTTGTACTTCA
13A3 cDNA TTTGTTAGTTTATTAGGTATTGGTTCGTCTGTAATGGTAGCTGCAGCTAGTTGTACTTCA
13-5 cDNA TTTGTTAGTTTATTAGGTATTGGTTCGTTTGTAATGGTAGCTGCAGCTAGTTGTACTTCA
13   cDNA TTTGTTAGTTTATTAGGTATTGGTTCGTTTGTAATGGTAGCTGCAGCTAGTTGTACTTCA
14   DNA  TTTGTTAGTTTATTAGGTATTGGTTCGTTTGTAATGGTAGCTGCAGCTAGTTGTACTTCA
14-5 DNA  TTTGTTAGTTTATTAGGTATTGGTTCGTTTGTAATGGTAGCTGCAGCTAGTTGTACTTCA
14A3 DNA  TTTGTTAGTTTATTAGGTATTGGTTCGTTTGTAATGGTAGCTGCAGCTAGTTGTACTTCA
          **************************** *******************************

          3397                                                        3450
9.3  orf  GCAGCTACACCAACTCCAAACCCTGAACCAAA------TCCAAACCCTGAACCAAAACCA
13A3 cDNA GCAGCTACACCAACTCCAAACCCTGAACCAAAACCAAATCCAAACCCTGAACCAAAACCA
13-5 cDNA GCAGCTACACCAACTCCAAACCCTGAACCAAAACCAAATCCAAACCCTGAACCAAAACCA
13   cDNA GCAGCTACACCAACTCCAAACCCTGAACCAAA------TCCAAACCCTGAACCAAAACCA
14   DNA  GCAGCTACACCAACTCCAAACCCTGAACCAAA------TCCAAACCCTGAACCAAAACCA
14-5 DNA  GCAGCTACACCAACTCCAAACCCTGAACCAAA------TCCAAACCCTGAACCAAAACCA
14A3 DNA  GCAGCTACACCAACTCCAAACCCTGAACCAAAACCAAATCCAAACCCTGAACCAAAACCA
          ********************************      **********************

          3451                                                        3510
9.3  orf  GATCCAATGCCAAACCCTCCTAGTGGTGGTATGAATGGTGGAGATATTAATCCAGGTGGT
13A3 cDNA GATCCAATGCCAAACCCTCCTAGTGGTGGTATGAATGGTGGAGATATTAATCCAGGTGGT
13-5 cDNA GATCCAATGCCAAACCCTCCTAGTGGTGGTATGAATGGTGGAGATATTAATCCAGGTGGT
13   cDNA GATCCAATGCCAAACCCTCCTAGTGGTGGTATGAATGGTGGAGATATTAATCCAGGTGGT
14   DNA  GATCCAATGCCAAACCCTCCTAGTGGTGGTATGAATGGTGGAGATATTAATCCAGGTGGT
14-5 DNA  GATCCAATGCCAAACCCTCCTAGTGGTGGTATGAATGGTGGAGATATTAATCCAGGTGGT
14A3 DNA  GATCCAATGCCAAACCCTCCTAGTGGTGGTATGAATGGTGGAGATATTAATCCAGGTGGT
          ************************************************************

          3511                                                        3570
9.3  orf  GGTCAAAATATGATGGATTCTGCAGCTCAAGAATTAACTGCCGCTAGAACTGCTTTAACT
13A3 cDNA GGACAAAATATGATGGATTCTGCAGCTCAAGAATTAACTGCCGCTAGAACTGCTTTAACT
13-5 cDNA GGACAAAATATGATGGATTCTGCAGCTCAAGAATTAACTGCCGCTAGAACTGCTTTAACT
13   cDNA GGACAAAATATGATGGATTCTGCAGCTCAAGAATTAACTGCCGCTAGAACTGCTTTAACT
14   DNA  GGACAAAATATGATGGATTCTGCAGCTCAAGAATTAACTGCCGCTAGAACTGCTTTAACT
14-5 DNA  GGACAAAATATGATGGATTCTGCAGCTCAAGAATTAACTGCCGCTAGAACTGCTTTAACT
14A3 DNA  GGACAAAATATGATGGATTCTGCAGCTCAAGAATTAACTGCCGCTAGAACTGCTTTAACT
          ** *********************************************************

          3571                                                        3630
9.3  orf  AGTTTGTTAGCTAGCAAGAACGCAAATATTGAGATGTACTCTGACTATGCAAAAATTCAG
13A3 cDNA AGTTTGTTAGCTAGCAAGAACGCAAATATTGAGATGTACTCTGACTATGCAAAAATTCAG
13-5 cDNA AGTTTGTTAGCTAGCAAGAACGCAAATATTGAGATGTACTCTGACTATGCAAAAATTCAG
13   cDNA AGTTTGTTAGCTAGCAAGAACGCAAATATTGAGATGTACTCTGACTATGCAAAAATTCAG
14   DNA  AGTTTGTTAGCTAGCAAGAACGCAAATATTGAGATGTACTCTGACTATGCAAAAATTCAG
14-5 DNA  AGTTTGTTAGCTAGCAAGAACGCAAATATTGAGATGTACTCTGACTATGCAAAAATTCAG
14A3 DNA  AGTTTGTTAGCTAGCAAGAACGCAAATATTGAGATGTACTCTGACTATGCAAAAATTCAG
          ************************************************************
```

```
           3631                                                       3690
9.3   orf  AATACTCTAATAGCTGCATACACAACGGCAGAACAAACTTCTCAAAATTCATCTGCTACA
13A3 cDNA  AATACTCTAATAGCTGCATACACAACGGCAGAACAAACTTCTCAAAATTCATCTGCTACA
13-5 cDNA  AATACTCTAATAGCTGCATACACAACGGCAGAACAAACTTCTCAAAATTCATCTGCTACA
13   cDNA  AATACTCTAATAGCTGCATACACAACGGCAGAACAAACTTCTCAAAATTCATCTGCTACA
14   DNA   AATACTCTAATAGCTGCATACACAACGGCAGAACAAACTTCTCAAAATTCATCTGCTACA
14-5 DNA   AATACTCTAATAGCTGCATACACAACGGCAGAACAAACTTCTCAAAATTCATCTGCTACA
14A3 DNA   AATACTCTAATAGCTGCATACACAACGGCAGAACAAACTTCTCAAAATTCATCTGCTACA
           ************************************************************

           3691                                                       3750
9.3   orf  CTAGAACAAGTTAAAAACGCTACATCAGCTTTACAAACAGCTATTAATACGGCTAATAGT
13A3 cDNA  CTAGAACAAGTTAAAAACGCTACATCAGCTTTACAAACAGCTATTAATACGGCTAATAGT
13-5 cDNA  CTAGAACAAGTTAAAAACGCTACATCAGCTTTACAAACAGCTATTAATACGGCTAATAGT
13   cDNA  CTAGAACAAGTTAAAAACGCTACATCAGCTTTACAAACAGCTATTAATACGGCTAATAGT
14   DNA   CTAGAACAAGTTAAAAACGCTACATCAGCTTTACAAACAGCTATTAATACGGCTAATAGT
14-5 DNA   CTAGAACAAGTTAAAAACGCTACATCAGCTTTACAAACAGCTATTAATACGGCTAATAGT
14A3 DNA   CTAGAACAAGTTAAAAACGCTACATCAGCTTTACAAACAGCTATTAATACGGCTAATAGT
           ************************************************************

           3751                                                       3809
9.3   orf  AATAAACAAAAA-TTTGATCAAGATCATAGTAATCTATTAATGTCATATAAAAACTTAAT
13A3 cDNA  AATAAACAAAAA-TTTGATCAAGATCATAGTAATCTATTAATGTCATATAAAAACTTAAT
13-5 cDNA  AATAAACAAAAA-TTTGATCAAGATCATAGTAATCTATTAATGTCATATAAAAACTTAAT
13   cDNA  AATAAACAAAAAATTTGATCAAGATCATAGTAATCTATTAATGTCATATAAAAACTTAAT
14   DNA   AATAAACAAAAA-TTTGATCAAGATCATAGTAATCTATTAATGTCATATAAAAACTTAAT
14-5 DNA   AATAAACAAAAA-TTTGATCAAGATCATAGTAATCTATTAATGTCATATAAAAACTTAAT
14A3 DNA   AATAAACAAAAA-TTTGATCAAGATCATAGTAATCTATTAATGTCATATAAAAACTTAAT
           ************ ***********************************************

           3810                                                       3868
9.3   orf  GGCTACTCTTGCTAAAAAA-GAAACTACTGTAATGACATTAAAAGATCCAAAATATAGCG
13A3 cDNA  GGCTACTCTTGCTAAAAAAAGAAACTACTGTAATGACATTAAAAGATCCAAAATATAGCG
13-5 cDNA  GGCTACTCTTGCTAAAAAA-GAAACTACCGTAATGACATTAAAAGATCCAAAATATAGCG
13   cDNA  GGCTACTCTTGCTAAAAAA-GAAACTACTGTAATGACATTAAAAGATCCAAAATATAGCG
14   DNA   GGCTACTCTTGCTAAAAAA-GAAACTACTGTAATGACATTAAAAGATCCAAAATATAGCG
14-5 DNA   GGCTACTCTTGCTAAAAAA-GAAACTACTGTAATGACATTAAAAGATCCAAAATATAGCG
14A3 DNA   GGCTACTCTTGCTAAAAAA-GAAACTACTGTAATGACATTAAAAGATCCAAAATATAGCG
           ******************* ******** *******************************

           3869                                                       3928
9.3   orf  CAATTTTGGATCAAATTAATGGTGTATCTTGTAAAGGTGAAGAGCTTGTTCAACATACAT
13A3 cDNA  CAATTTTGGATCAAATTAATGGTGTATCTTGTAAAGGTGAAGAGCTTGTTCAACATACAT
13-5 cDNA  CAATTTTGGATCAAATTAATGGTGTATCTTGTAAAGGTGAAGAGCTTGTTCAACATACAT
13   cDNA  CAATTTTGGATCAAATTAATGGTGTATCTTGTAAAGGTGAAGAGCTTGTTCAACATACAT
14   DNA   CAATTTTGGATCAAATTAATGGTGTATCTTGTAAAGGTGAAGAGCTTGTTCAACATACAT
14-5 DNA   CAATTTTGGATCAAATTAATGGTGTATCTTGTAAAGGTGAAGAGCTTGTTCAACATACAT
14A3 DNA   CAATTTTGGATCAAATTAATGGTGTATCTTGTAAAGGTGAAGAGCTTGTTCAACATACAT
           ************************************************************
```

```
               3929                                                        3988
9.3   orf     TAGATCCAGTTAGTGGAATTGTTCCTGCTGCTAATACAATCACAGAAGAGATTACTAAGA
13A3  cDNA    TAGATCCAGTTAGTGGAATTGTTCCTGCTGCTAATACAATCACAGAAGAGATTACTAAGA
13-5  cDNA    TAGATCCAGTTAGTGGAATTGTTCCTGCTGCTAATACAATCACAGAAGAGATTACTAAGA
13    cDNA    TAGATCCAGTTAGTGGAATTGTTCCTGCTGCTAATACAATCACAGAAGAGATTACTAAGA
14    DNA     TAGATCCAGTTAGTGGAATTGTTCCTGCTGCTAATACAATCACAGAAGAGATTACTAAGA
14-5  DNA     TAGATCCAGTTAGTGGAATTGTTCCTGCTGCTAATACAATCACAGAAGAGATTACTAAGA
14A3  DNA     TAGATCCAGTTAGTGGAATTGTTCCTGCTGCTAATACAATCACAGAAGAGATTACTAAGA
              ************************************************************

                   3993
9.3   orf         TTGAA
13A3  cDNA        TTGAA
13-5  cDNA        TTGAA
13    cDNA        TTGAA
14    DNA         TTGAA
14-5  DNA         TTGAA
14A3  DNA         TTGAA
                  *****

9.3 orf     VKRKNILKFVSLLGIGSFVMVAAASCTSAATPTPNPEP--NPNPEPKPDPMPNPPS
cDNA        VKRKNILKFVSLLGIGSFVMVAAASCTSAATPTPNPEPKPNPNPEPKPDPMPNPPS
cDNA        VKRKNILKFVSLLGIGSFVMVAAASCTSAATPTPNPEP--NPNPEPKPDPMPNPPS
DNA         VKRKNILKFVSLLGIGSFVMVAAASCTSAATPTPNPEPKPNPNPEPKPDPMPNPPS
DNA         VKRKNILKFVSLLGIGSFVMVAAASCTSAATPTPNPEP--NPNPEPKPDPMPNPPS
            **************************************  ****************
```

Fig. 4: Nucleotide sequence alignment of pMGA gene cDNA clones from *M. gallisepticum* F-strain using the 9.3 genomic library clone sequence as the prototype.

A. The cDNA clones (1-1, 1-2, 1-3, and 1-4) were PCR amplified with primers pMGA1 and T23.

B. The cDNA (13A3, 13-5, and 13) and genomic (14, 14-5, and 14A3) clones were obtained with primers pMGA1 and 128A from *M. gallisepticum* F-strain.

C. N-terminal predicted amino acid sequence derived from the PCR amplified cDNA and genomic clones. The sequence from the 9.3 ORF is used as the prototype. Identical nucleotides are designated by an "*".

Nucleotide sequence analysis: The 9.1 ORF represents the C-terminal region of a pMGA gene and the predicted amino acid sequence shows an identity of 92% in the C-terminal region with the pMGA1.1 protein of *M. gallisepticum* HS strain (GenBank: AF275312) (Fig. 2A). The 9.2 ORF is the largest ORF (2.0kb) contained in the λ9 insert, and is the full coding sequence of a pMGA gene (Fig. 1 and Fig. 2B). The 9.2 ORF has an overall amino acid identity of 66% with the pMGA1.4 protein of *M. gallisepticum* S6 strain, and 50% identity to the 9.1 ORF in the C-terminal region. At the opposite (or T7) end of the insert , the 9.3 ORF encodes the N-terminal part of a pMGA gene (Fig. 1). The complete coding sequence of the 9.3 ORF was obtained by PCR amplification of genomic DNA. The predicted amino acid sequence of the complete 9.3 ORF has 96% identity with the pMGA1.2 protein of *M. gallisepticum* HS strain (Fig. 2B).

The λ 9 ORF are separated by two intergenic regions of 332 bp and 333 bp located 5' to the 9.2 and 9.3 orf, respectively, as illustrated in the nucleotide sequence alignment of intergenic regions (Fig. 3). The intergenic regions 5' to the 9.2 and 9.3 ORF contain 15 and 12 copies of the GAA trinucleotide repeats which are characteristic of the intergenic regions of pMGA family members (Baseggio *et al*., 1996; Markham *et al*., 1994). Variation in the number of GAA repeats is a mechanism of transcriptional regulation for the pMGA genes; (GAA)12 are present in the intergenic regions of those pMGA genes that are expressed (Glew *et al*., 1998; Liu *et al*., 1998). (GAA)12 in the intergenic region 5' to the 9.3 ORF suggests that the 9.3 ORF is expressed in the F-strain.

Expression of pMGA gene mRNA: The PCR primers pMGA1 and T23 designed complementary to regions conserved among all pMGA family members and were used to amplify cDNA derived from F-strain total RNA. The pMGA1 and T23 primers resulted in a 247 bp cDNA product from which the nucleotide sequence of 4 cDNAs was obtained. The nucleotide sequence of the cDNA clones denoted as 1-1, 1-2, 1-3, and 1-4 excluding the primer sequences (Fig. 4A). The cDNA sequences are most similar to the 9.3 ORF sequence; except for a nucleotide change at position 3513 where the 9.3 ORF sequence differs from the PCR amplified sequences. Our cDNA sequences were divided into two groups; one group having a repeat of the 6 base pairs "AACCAAA" at the N-terminus of the cDNA, and the other group lacking the repeat

(as with the 9.3 ORF genomic library clone sequence, nucleotide positions 3423 - 3428 of the 9.3 ORF, Fig. 4A). To confirm that the cDNAs arose from the 9.3 ORF in the genome of the F-strain, cDNA and genomic DNA were prepared from a fresh culture of the F-strain and amplified using pMGA1 and 128A (complementary to a region of the 9.3 ORF which differs from other reported pMGA genes, except for the pMGA1.2 gene of *M. gallisepticum* HS strain). Three cDNA and three genomic clones were sequenced from the 702 bp product (Fig. 4B). The PCR amplified cDNA sequences only differed from the 9.3 ORF sequence by a T to C change at nucleotide positions 3366 and 3838, and an additional base pair (A) at nucleotide positions 3762 and 3829, probably Taq polymerase fidelity errors. The genomic sequences were identical to the 9.3 ORF except for a change from T to A at position 3513. Two of the cDNA clones and one genomic clone showed the repeat sequence "AACCAAA" at nucleotide positions 3423 - 3428 (Fig. 4B). The presence of the nucleotide sequence repeat predicts two additional amino acids (Proline and Lysine) in the protein at residue positions 39 and 40 (Fig. 4C).

Discussion

The F-strain is used as a live vaccine by the poultry egg layer industry (Branton *et al*., 2000), and has the desirable quality of long term persistence (Kleven, 1981). The pMGA genes occupy a significant portion of the genome in various *M. gallisepticum* strains (Baseggio *et al*., 1996) and the pMGA ORFs characterized to date lack in-frame stop codons suggesting they should be capable of expression (Markham *et al*., 1994). The importance of pMGA gene products in host colonization and persistence has been proposed (Markham *et al*., 1993), and homologs of pMGA genes have been identified in two other species of poultry mycoplasmas (Noormohammadi *et al*., 1998; Markham *et al*., 1999). Therefore, we considered it critical to identify and characterize the pMGA hemagglutinin genes expressed by F-strain cells to understand the persistence of this vaccine in chickens. To identify pMGA genes, we constructed a λ ZAPII F-strain genomic library and immunoscreened this with rabbit polyclonal antibody raised against *M. gallisepticum* F-strain surface antigens (May *et al*., 1988). Rabbit antiserum with hemagglutination - inhibition activity identified a genomic library clone with two partial, and one complete, ORFs with significant homology to the pMGA hemagglutinin gene family of *M. gallisepticum*.

A specific number of GAA repeats, (GAA) n = 12, is found in the promoter region of expressed pMGA genes (Baseggio *et al*., 1996; Markham *et al*., 1998; Glew *et al*., 1998). Recent studies of pMGA transcription propose that the length of the repeat region affects the distance between two DNA sequences that bind a protein responsible for pMGA gene transactivation (Liu *et al*., 2000, 2002). Sequencing of the intergenic regions 5' of our 9.2 and 9.3 ORF revealed a series of GAA trinucleotide repeats 5' of the coding regions, and the 12 GAA repeats 5' of the 9.3 ORF suggested that the pMGA gene is expressed in F-strain.

The cDNAs obtained with our RT-PCR studies were, not surprisingly, more similar to the 9.3 ORF sequence we obtained than to other pMGA gene sequences reported to date. The N-terminal region of the predicted amino acid sequence of our 9.3 ORF cDNA sequence differs from the N-terminal sequence reported by peptide sequencing of a pMGA protein of the F-strain by others (Glew *et al*., 1995). It is possible that the 9.3 ORF - like cDNAs we detected represent a variant pMGA gene(s) expressed in the F-strain population. Noteably, Glew *et al*. (1995), identified 3 different pMGA genes in addition to the predominantly expressed pMGA1.1 mRNA species the *M. gallisepticum* S6 strain population. Another possibility is that the adaptation of the F-strain used in our studies to in vitro culture (99 passages above the unknown level) has selected for expression of the variant 9.3 ORF - like cDNAs in place of the dominant pMGA gene product detected in the F-strain by Glew *et al*. (1995).

In the evaluation of pMGA gene expression in various *M. gallisepticum* strains obtained from infected chickens, Berlic *et al*. (2000) showed that two types of pMGA1.1/1.2 genes could be distinguished in different strains based the presence or absence of a repeat sequence in the N-terminal proline-rich region. The 9.3 ORF - like cDNAs we identified differed from each other by a short sequence of 6 bp in the proline-rich region. Identical sequences to both variants were also amplified from genomic DNA. It is not clear whether two highly similar genes are present in the F-strain genome, or whether only one 9.3 ORF - like gene is present which shows heterogeneity in the N-terminal region in the F-strain population.

Our data predicts two additional amino acids are encoded in the proline rich region (Fig. 4B and 4C). Proline rich regions of proteins can form elongated structures that may function in protein - protein interactions (Kay *et al*., 2000). Because pMGA proteins are probably attached to the mycoplasmal cell membrane at the N-terminus by a lipid moiety (Markham *et al*., 1993), we speculate that the N-terminal proline rich region of pMGA proteins are involved in protein - protein interactions at the cell surface. One possible function would be to sequester host proteins at the mycoplasmal cell surface for digestion by cell-associated proteases for subsequent amino acid transport into the mycoplasma cell for the nutrition. Experimental evidence for peptide binding by a mycoplasma lipoprotein adhesin is reported (Henrich *et al*., 1999). In this regard, if both cDNAs identified in this study are translated, then such an auxiliary function of pMGA adhesins may play a role in the persistence of F-strain in the respiratory tract of chickens.

Acknowledgments

We are grateful to Dr. Stanley H. Kleven (University of Georgia) for the gift of the *M. gallisepticum* F-strain, and to Jerry Drott and Alvah Strickland for assistance in growing mycoplasma cultures and preparation of genomic DNA (USDA/ARS); to Dr. Dawn Luthe (Department of Biochemistry, Mississippi State) for sharing sucrose gradient protocols, and to Suzana Marinovic, Renuka Shivaji, Mary Rudis, and Hai Chen for assistance in protocols and helpful discussions of this work. We also thank Drs. Shane Burgess, Mark Lawrence, and Lesya Pinchuk for their helpful suggestions and critical review of the manuscript. This project was supported in part by a USDA, REE, ARS Cooperative Agreement and by a USDA-NRI, grant number 98-35208-6543. Contribution No. J-10177 from the Mississippi Agriculture and Forestry Experiment Station.

References

Abd-El-Motelib, T. Y. and S. H. Kleven, 1993. A comparative study of *Mycoplasma gallisepticum* vaccines in young chickens. Avian Dis., 37: 981-987.

Baseggio, N., M. D. Glew, P. F. Markham, K. G. Whithear and G. F. Browning, 1996. Size and genomic location of the pMGA multigene family of *Mycoplasma gallisepticum*. Microbiology., 142: 1429-1435.

Berlic, T. M., D. Bencina and P. Dovc, 2000. Sequence polymorphisms within the pMGA genes and pMGA antigenic variants in *Mycoplasma gallisepticum*. FEMS Microbiology Letters., 184: 133-139.

Branton, S. L., H. Gerlach and S. H. Kleven, 1984. *Mycoplasma gallisepticum* isolation in layers. Poult. Sci., 63: 1917-1919.

Branton, S. L., B. D. Lott, J. D. May, W. R. Maslin, C. R. Boyle and G. T. Pharr, 1997. The effects of F strain *Mycoplasma gallisepticum*, Mycoplasma synoviae, and dual infection in commercial layer hens over a 44-week laying cycle when challenged before beginning of lay. I. Egg production and selected egg quality parameters. Avian Dis., 41: 832-837.

Branton, S. L., B. D. Lott, J. D. May, W. R. Maslin, G. T. Pharr, S. D. Bearson, S. D. Collier and D. L. Boykin, 2000. The effects of ts-11 strain *Mycoplasma gallisepticum* vaccination in commercial layers on egg production and selected egg quality parameters. Avian Dis., 44: 618-623.

Dykstra, M.J., S. Levisohn, O. J. Fletcher and S. H. Kleven, 1985. Evaluation of cytopathologic changes induced in chicken tracheal epithelium by *Mycoplasma gallisepticum in vivo* and *in vitro*. Am. J. Vet. Res., 46: 116-122.

Frey, M. C., R. P. Hanson and D. P. Anderson, 1968. A medium for the isolation of avian Mycoplasma. Am. J. Vet. Res., 29: 2164-2171.

Glew, M. D., P. F. Markham, G. F. Browning and I. D. Walker, 1995. Expression studies on four members of the pMGA multigene family in *Mycoplasma gallisepticum* S6. Microbiology., 141: 3005-3014.

Glew, M. D., N. Baseggio, P. F. Markham, G. F. Browning and I. D. Walker, 1998. Expression of the pMGA genes of *Mycoplasma gallisepticum* is controlled by variation in the GAA trinucleotide repeat lengths within the 5' noncoding regions. Infect. Immun., 66: 5833-5841.

Henrich, B., M. Hopfe, A. Kitzerow and U. Hadding, 1999. The adherence-associated lipoprotein P100, encoded by an opp operon structure, functions as the oligopeptide-binding domain OppA of a putative oligopeptide transport system in Mycoplasma hominis. J. Bacteriol., 181: 4873-4878.

Inamine, J. M., S. Loechel and P. C. Hu, 1989. Nucleotide sequence of the tuf gene from *Mycoplasma gallisepticum*. Nucl. Acids. Res., 23: 10126.

Kay, B. K., M. P. Williamson and M. Sudol, 2000. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. FASEB J., 14: 231-241.

Keeler, C. L., L. L. Hnatow, P. L. Whetzel and J. E. Dohms, 1996. Cloning and characterization of a putative cytadhesin gene (mgc1) from *Mycoplasma gallisepticum*. Infect. Immun., 64: 1541-1547.

Kleven, S. H., 1981. Transmissibility of the F strain of *Mycoplasma gallisepticum* in leghorn chickens. Avian Dis., 25: 1005-1018.

Kleven, S. H., H. H. Fan and K. S. Turner, 1998. Pen trial studies on the use of live vaccines to displace virulent *Mycoplasma gallisepticum* in chickens. Avian Dis., 42: 300-306.

Ley, D. H. and H. W. Yoder, Jr., 1997. Mycoplasmosis. *Mycoplasma gallisepticum* infection. In: Calnek, B. W., Barnes, H. J., Beard, C. W., McDougal, L. R, and Saif, Y. M. (Eds.), Diseases of Poultry. 10th ed. Iowa State University Press, Ames, IA. pp: 194-207.

Liu, L., M. Payne, V. L. Van Santen, K. Dybvig and V. S. Panangala, 1998. A protein (M9) associated with monoclonal antibody-mediated agglutination of *Mycoplasma gallisepticum* is a member of the pMGA family. Infect. Immun., 66: 5570-5575.

Liu, L., K. Dybvig, V. S. Panangala, V. L. van Santen and C. T. French, 2000. GAA trinucleotide repeat region regulates M9/pMGA gene expression in *Mycoplasma gallisepticum*. Infect. Immun., 68: 871-876.

Liu, L., V. S. Panangala and K. Dybvig, 2002. Trinucleotide GAA repeats dictate pMGA gene expression in *Mycoplasma gallisepticum* by affecting spacing between flanking regions. J. Bacteriol., 184: 1335-1339.

Luthe, D.S., 1983. A simple technique for the preparation and storage of sucrose gradients. Anal. Biochem., 135: 230-232.

Markham, P.F., M. D. Glew, M. R. Brandon, I. D. Walker and K. G. Whithear, 1992. Characterization of a major hemagglutinin protein from *Mycoplasma gallisepticum*. Infect. Immun., 60: 3885-3891.

Markham, P. F., M. D. Glew, K. G. Whithear and I. D. Walker, 1993. Molecular cloning of a member of the gene family that encodes pMGA, a hemagglutinin of *Mycoplasma gallisepticum*. Infect. Immun., 61: 903-909.

Markham, P. F., M. D. Glew, J. E. Sykes, T. R. Bowden, T. D. Pollocks, G. F. Browning, K. G. Whithear and I. D. Walker, 1994. The organization of the multigene family which encodes the major cell surface protein, pMGA, of *Mycoplasma gallisepticum*. FEBS Letters., 352: 347-352.

Markham, P. F., M. D. Glew, G. F. Browning, K. G. Whithear and I. D. Walker, 1998. Expression of two members of the pMGA gene family of *Mycoplasma gallisepticum* oscillates and is influenced by pMGA - specific antibodies. Infect. Immun., 66: 2845-2853.

Markham,P. F., M. F. Duffy, M. D. Glew and G. F. Browning, 1999. A gene family in Mycoplasma imitans closely related to the pMGA family of *Mycoplasma gallisepticum*. Microbiology., 145: 2095-2103.

May, J. D., S. L. Branton and M. A. Cuchens, 1988. Identification of *Mycoplasma gallisepticum* and M. synoviae by flow cytometry. Avian Dis., 32: 513-516.

Minion, F. C., C. VanDyk and B. K. Smiley, 1995. Use of an enhanced *Escherichia coli* opal suppressor strain to screen a Mycoplasma hyopneumoniae library. FEMS Microbiology Letters., 131: 81-85.

Noormohammadi, A. H., P. F. Markham, M. F. Duffy, K. G. Whithear and G. F. Browning, 1998. Mutigene families encoding the major hemagglutinins in phylogenetically distinct mycoplasma. Infect. Immun., 66: 3470-3475.

Pruthi, A. K. and M. U. Kharole, 1981. Sequential pathology of genital tract in chickens experimentally infected with *Mycoplasma gallisepticum*. Avian Dis., 25: 768-778.

Razin, S., 1992. Mycoplasma taxonomy and ecology. In: Mycoplasma. Molecular biology and pathogenesis. Published by The American Society of Microbiology, Washington, DC.

Samuelsson, T. and T. Boren. 1992. Evolution of macromolecule synthesis. In: Mycoplasmas. Molecular Biology and Pathogenesis. Published by The American Society of Microbiology, Washington, DC.

Smiley, B. K. and F. C. Minion, 1993. Enhanced read through of opal (UGA) stop codons and production of Mycoplasma pneumoniae P1 epitopes in *Escherichia coli*. Gene., 134: 33-40.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res., 22: 4673-4680.

Waldo, R. H., P. L. Popham, C. E. Romero-Arroyo, E. A. Mothershed, K. K. Lee and D. C. Krause, 1999. Transcriptional Analysis of the hmw Gene Cluster of Mycoplasma pneumoniae. J. Bacteriol., 181: 4978-4985.

Whithear, K. G., 1996. Control of avian mycoplasmosis by vaccination. Rev. Sci. Tech. Off. Int. Epiz., 15: 1527-1553.