Development and Evaluation of Sequence Typing Assays for investigating the Epidemiology of *Mycoplasma synoviae* Outbreaks in Poultry

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Mohamed El-Gazzar

Graduate Program in Comparative and Veterinary Medicine

The Ohio State University

2014

Dissertation Committee:

Richard D. Slemons, Advisor

Päivi J. Rajala-Schultz

Daral J. Jackwood
Abstract

*Mycoplasma gallisepticum* (MG) and *Mycoplasma synoviae* (MS) are the two most pathogenic avian mycoplasma species. In birds Mycoplasma can be transmitted both horizontally and vertically, and understanding the epidemiology of the infections is an integral part of controlling the disease. Several molecular assays can be used for mycoplasma genotyping. Random Amplified Polymorphic DNA (RAPD) and Amplified Fragment Length Polymorphism (AFLP) are two fingerprinting assays commonly used for this purpose. But sequence typing assays are the preferred method for avian mycoplasma genotyping. Sequence typing is preferred because it can be performed directly on clinical samples without the need to isolate the microorganism in a pure culture.

Multiple genes have been identified for sequence typing of MG, while the partial *vlhA* gene sequence is the only target being used for MS sequence typing. Relying on a single gene sequence creates ambiguity in the genetic relatedness between MS strains. In an effort to enlarge the repertoire of MS sequence typing assays, we explore expansion of the currently described *vlhA* assay segment and we investigate one new genomic target for sequence typing of MS. We also report the development of a new MLST assay for MS.
The first chapter is a literature review summarizing the main features of class *Mollicutes* which is the class to which avian mycoplasma belong. In addition we also outline the characteristics of avian mycoplasma diseases and their diagnosis, control and molecular epidemiology.

The second chapter of this document describes the attempt to expand the targeted area of the *vlhA* gene used for MS sequence typing in an effort to maximize its genotypic potential. The expansion was successful and the target segment was extended from 420 bp to 900 bp. The gene conversion recombination process was discovered to occur immediately downstream of the original 420 bp. Hence the sequence information generated beyond 420 bp was found not to be useful for MS sequence typing purposes. This study identified the maximum potential of *vlhA* gene for MS sequence typing.

The third chapter describes an attempt to amplify another DNA segment that can be used for MS sequence typing by sequencing. Cluster Regularly Interspaced Short Palindromic Repeats (CRISPR) system in prokaryotes. This system serves as an acquired immune system of prokaryotes against invading phage infections. Amplification and sequencing of the CRISPR segment was successful in a few MS samples, and CRISPR sequences showed variations between MS samples. However, amplification of the CRISPR segment was not successful in the rest of the tested MS samples using conventional PCR techniques at hand. Whole genome sequencing can be one of the ways to harness the discriminatory power and the genotypic potential of this region. While for the time being the whole genome sequencing is not one of the used techniques in the
regular investigation of mycoplasma outbreaks in the poultry industry, but probably will be in the future.

The fourth chapter describes the development Multilocus Sequence Typing (MLST) assay for MS. Twenty four housekeeping genes were studied and seven internal gene segments were finally selected. The amplification and sequencing of seven gene segments was successful in all 58 tested MS strains, field isolates and clinical samples. Out of the tested 58 samples MLST generated 30 sequence types. The vlhA sequence typing assay also generated 30 sequence types for the same population. In spite of having the same apparent discriminatory power, MLST provided links between the related outbreaks that is more congruent with the available epidemiological information. Therefore using both MLST and vlhA in tandem is recommended for a more refined understanding of the epidemiological picture of mycoplasma outbreaks.

In conclusion, the maximum potential of vlhA sequence typing assay was determined and a new MS sequence typing assay (MLST) was developed. As a result a tenfold increase in the size of genomic sequence used for MS sequence typing was achieved. The author recommends to use both MLST and vlhA assays in tandem for more accurate and refined epidemiological analysis of MS outbreaks.
Acknowledgments

I’m privileged to be in this position by the help and the support of many individuals along the journey who encouraged me to pursue my dreams when they seemed impossible and assured me when I was in doubt. For that I am forever indebted to them and forever grateful.

I first want to pay tribute to both of my parents who are no longer with us in this life. I want to let them know that the love of science and knowledge they instilled in me and their insistence that I should always follow my passions has paid dividends.

It goes without saying that my advisors and committee members are the pillar without which I would have not been able to rise to this new peak in my career. However, I’m beholden to Dr. Richard Slemons for accepting to be my advisor and for his sincere guidance. I would like also to thank Dr. Päivi Rajala-Schultz and Dr. Daral Jackwood for their valuable contributions to the success of this program. Among my committee member I want to especially thank to Dr. Mo Saif for his priceless input and continuous encouragement.

I want to acknowledge Dr. Ziv Raviv who afforded me the opportunity to come to The Ohio State University, and put me on the right track. Especial thanks Dr. William J.
A. Saville our department chair who made my graduation possible by his unflinching support through uncertain times.

I also want to take the opportunity to recognize other individual who had profound effect on my life. Thanks to Dr. Alaa Abdo of Cairo University for his guidance that immensely helped me to push forward and reach for my goals. I thank him not only for his mentorship but also for being the role model that I aspired to follow on both the professional and personal levels. I would like to thank Dr. Nicolas Dorko of Aviagen who opened for me the horizon of better poultry education and pointed me in the right direction.

With that I’m ready to close one chapter of my life and excited to start a new one.
Vita

1995..........................................................Ismail El-Kabbany High School
2000..........................................................B.S. Veterinary Medicine, Cairo University
2000 to 2006 ............................................Technical Service Veterinarian, Cairo Poultry Group
2006 to 2008 .............................................Avian Range Product Specialist, Ceva Santé Animale – Egypt Ltd.
2009..........................................................Master of Avian Medicine, The University of Georgia
2010 to present ............................................Graduate Research Associate, Department of Veterinary Preventive Medicine, The Ohio State University

Publications


Fields of Study

Major Field: Veterinary Preventive Medicine
Table of Contents

Abstract............................................................................................................................................... ii
Acknowledgments................................................................................................................................. v
Vita........................................................................................................................................................ vii
Table of Contents.................................................................................................................................. ix
List of Tables ........................................................................................................................................... xiv
List of Figures ......................................................................................................................................... xv
Introduction............................................................................................................................................ 1

Chapter 1: Literature Review................................................................................................................ 5
  1.1 History and evolution of Class Mollicutes.................................................................................. 5
    1.1.1 History and significance ........................................................................................................ 5
    1.1.2 Evolution and phylogeny....................................................................................................... 7
  1.2 Biology of Class Mollicutes........................................................................................................... 10
    1.2.1 Lack of cell wall.................................................................................................................... 10
    1.2.2 Small genome....................................................................................................................... 12
1.2.3 Scavenging metabolism ................................................................. 16
1.2.4 Central dogma of Mollicutes life .................................................. 17

1.3 Avian Mycoplasma ........................................................................... 20
1.3.1 Mycoplasma species and hosts range .......................................... 20
1.3.2 Diseases ....................................................................................... 22
1.3.3 Pathogenicity ............................................................................... 25
1.3.3 Diagnosis and control .................................................................. 26
1.3.5 Molecular epidemiology .............................................................. 29

Chapter 2: The Genotyping Potential of the Mycoplasma synoviae vlhA Gene ... 59

2.1 Abstract .......................................................................................... 60
2.2 Introduction ...................................................................................... 61
2.3 Materials and Methods ................................................................. 63
2.3.1 Mycoplasma synoviae field samples selection ............................. 63
2.3.2 Mycoplasma synoviae strains and culture method ...................... 64
2.3.3 Serial in vitro passages of MS-K1968 experiment ....................... 64
2.3.4 Genomic DNA extraction ............................................................. 65
2.3.5 PCR design and amplification of extended vlhA gene segments .................................................. 65
2.3.6 Sequencing and sequence analysis of PCR products .......... 66

2.4 Results.................................................................................................................. 67

2.4.1 Amplification and sequencing of extended vlhA gene segment ........................................................................... 67

2.4.2 Sequence analysis of extended vlhA gene amplicons .......... 68

2.4.3 Sequence analysis of sequential generations of MS culture .. 70

2.5 Discussion............................................................................................................... 71

2.6 Acknowledgement ................................................................................................ 76

2.7 References............................................................................................................. 76

Chapter 3: Mycoplasma synoviae Genotyping by Sequencing the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) ........ 92

3.1 Abstract ................................................................................................................. 93

3.2 Introduction ............................................................................................................ 93

3.3 Materials and Methods ....................................................................................... 98

3.3.1 Mycoplasma synoviae strains, field isolates and clinical samples ............................................................................. 98

3.3.2 Genomic DNA extraction ............................................................................... 98

3.3.3 Primer design .................................................................................................... 99

3.3.4 PCR design and amplification ......................................................................... 99
3.3.5 Sequencing................................................................. 100
3.3.6 Sequencing analysis..................................................... 101
3.4 Results.................................................................................. 101
  3.4.1 Original direct amplification of CRISPR array and sequencing .............................................. 101
  3.4.2 Upstream and downstream amplification and sequencing.......................................................... 101
  3.4.3 Final direct amplification of CRISPR array and sequencing ......................................................... 102
  3.4.4 Sequence analysis ............................................................... 102
  3.4.5 Spacer sequence identification............................................................... 103
3.5 Discussion............................................................................. 103
3.6 References............................................................................ 106

Chapter 4: Multilocus Sequence Typing of *Mycoplasma synoviae* ....................... 124
  4.1 Abstract................................................................................ 125
  4.2 Introduction.......................................................................... 126
  4.3 Materials and Methods.......................................................... 129
    4.3.1 *Mycoplasma synoviae* strains and field isolates........... 129
    4.3.2 *Mycoplasma synoviae* clinical samples ..................... 130
    4.3.3 Genomic DNA extraction ............................................. 130
4.3.4 Selection of the housekeeping genes ........................................ 130
4.3.5 PCR design and amplification .............................................. 132
4.3.6 Sequencing ......................................................................... 133
4.3.7 Sequence analysis of PCR products ...................................... 134
4.3.8 Allelic profiles and sequence types ...................................... 134
4.3.9 MLST data analysis ............................................................. 134
4.3.10 Comparison of MLST and \( vlhA \) genotyping ................. 135

4.4 Results .................................................................................. 135
4.4.1 Gene selection .................................................................... 135
4.4.2 Amplification and sequencing ........................................... 136
4.4.3 Allelic profiles and sequence types ................................... 136
4.4.4 MLST data analysis .............................................................. 137
4.4.5 Comparing MLST to \( vlhA \) genotyping ............................ 139

4.5 Discussion ........................................................................... 141

4.6 Acknowledgement ................................................................ 144

4.7 References ........................................................................... 144

Chapter 5: Concluding Remarks and Future Research .................... 173

Bibliography .............................................................................. 177
List of Tables

Table 2.1: Summary of selected samples and \textit{vlhA} larger segment and smaller segment PCR amplification and sequencing results ................................................................. 81

Table 2.2: Summary of MS \textit{vlhA} PCR assays primer sequences and amplicon size ...... 83

Table 3.1: A list of Twenty four \textit{Mycoplasma synoviae} strains, field isolates and clinical samples in which CRISPR amplification was attempted ........................................ 113

Table 3.2: A list of nineteen primers designed and used for CRISPR amplification .... 114

Table 3.3: Seventeen different primer set combinations used for CRISP Amplification, with their amplicon expected size according to MS53 genome ................................. 115

Table 4.1: List of all 58 MS strains, field isolates and clinical samples tested for MLST .......................................................................................................................... 151

Table 4.2: List of studied 24 genes as potential targets of MLST and their size, position, direction and the reason they were not included in the final 7 genes ....................... 153

Table 4.3: Final 7 Genes for the MLST assay forward and a reverse primer, amplicon size, final segment used for MLST after sequencing, and the point where the final segments starts and ends ......................................................................................................... 154

Table 4.4: A list of all alleles for each of the selected 7 genes and allelic profiles of each of the 30 sequence types generated by MLST assay ......................................................... 155
List of Figures

Figure 2.1: A schematic representation of the \textit{vlhA} gene family............................................. 84

Figure 2.2: Steps performed for the serial in vitro passages......................................................... 86

Figure 2.3: Agarose gel pictures showing the successful amplification using the larger and the smaller segment assays........................................................................................................... 87

Figure 2.4: Example of three \textit{vlhA} genotypes............................................................................. 88

Figure 2.5: Amino acid sequence deduced from nucleotide sequence of partially amplified \textit{vlhA} gene of sequential generations derived from two clones of a single MS strain K1968 colony............................................................................................................................ 89

Figure 2.6: Alignment of expressed \textit{vlhA} gene against 8 of the largest pseudogenes .... 91

Figure 3.1: Schematic representation of CRISPR/cas system......................................................... 116

Figure 3.2: CRISPR sequence and putative cas genes identified in MS53 whole genome.............................................................................................................................................. 117

Figure 3.3: MS53 CRISPR sequenc showing the position of primers used for amplification attempts........................................................................................................................................ 118

Figure 3.4: Agarose gel picture showing one of the original attempts to amplify MS CRISPR array............................................................................................................................................. 120
Figure 3.5: Agarose gel picture of the Upstream and Downstream amplification attempt. ................................................................. 121

Figure 3.6: Agarose gel picture showing the amplification results of the CRISPR final amplification. ......................................................................................................................................... 122

Figure 3.7: Sequence alignment of the 3 generated MS CRISPR sequences against the MS53 CRISPR sequence obtained from the GenBank. ............................................................ 123

Figure 4.1: A schematic representation of MS whole genome showing the positions of selected genes ............................................................................................................................................. 157

Figure 4.2: Nine different MS strains amplified for each of the selected MLST genes and visualized on agarose gel ........................................................................................................................................ 158

Figure 4.3: Sequence alignment of 14 different adk alleles .................................................. 159

Figure 4.4: Sequence alignment of 17 different atpG alleles ............................................. 160

Figure 4.5: Sequence alignment of 13 different efp alleles ................................................ 161

Figure 4.6: Sequence alignment of 13 different gmk alleles ............................................. 162

Figure 4.7: Sequence alignment of 18 different nagC alleles ............................................ 163

Figure 4.8: Sequence alignment of 15 different ppa alleles ............................................. 164

Figure 4.9: Sequence alignment of 15 different recA alleles ........................................... 165

Figure 4.10: Sequence alignment of 30 different vlhA alleles ........................................... 166

Figure 4.11: Graph showing number of alleles, number of polymorphic sites and the percent DNA variability for the 7 MLST final segments compared to the partial vlhA gene segment using 58 MS strains, field isolates and clinical samples. .............................. 168
Figure 4.12: Phylogenetic tree of all 58 MS strains, field isolates and clinical sample using MLST assay

Figure 4.13: Arkansas subpopulation: Phylogenetic tree of 14 MS strains, field isolates and clinical sample using MLST assay

Figure 4.14: Ohio-Pensylvania subpopulation: Phylogenetic tree of 14 MS field isolates and clinical sample using MLST assay

Figure 4.15: Phylogenetic tree of 54 MS strains, field isolates and clinical samples using vlhA assay
Introduction

With the size and density of the poultry industry, even mild infectious diseases can inflict significant economic losses. Mycoplasma is an example of a mildly pathogenic infectious agent that can severely impact performance. More than twenty species are considered avian mycoplasma, but only four of them are pathogenic to poultry. *Mycoplasma gallisepticum* (MG) and *Mycoplasma synoviae* (MS) are pathogenic for both chickens and turkeys. *Mycoplasma iwoae* (MI) and *Mycoplasma meleagrisidis* (MM) are pathogenic only for turkeys. MG has been traditionally considered the most pathogenic mycoplasma species in poultry. But recently an increase in the pathogenicity of MS has been witnessed, both in the US and in Europe; and some poultry disease specialists think that MS is surpassing MG as a cause of diseases in poultry.

The mycoplasma infections are only one of the many challenges faced by the poultry industry; however, mycoplasma is unique in many aspects. They are unique as a microorganism, lacking the typical cell wall of other bacteria and thriving with the smallest amount of genetic material among all independently replicating organisms. *Mycoplasma* is also unique in the typically long incubation period between the infection and onset of clinical signs in chickens and turkeys. While some of the more pathogenic strains can produce significant respiratory disease, high mortality, lameness, embryo
mortality and deformity, the more common form of the disease is subclinical. As a primary pathogen, avian mycoplasma typically require the presence of secondary pathogens to produce significant disease. In spite of their apparent benign consequences, mycoplasma infections are among the costliest diseases facing the poultry industry in terms of both impact on production and the cost of disease prevention and control.

The outcome of a mycoplasma outbreak could be devastating to the poultry producer in the form of increased mortality, reduced feed conversion rate, loss in egg production and the especially expensive carcass condemnations in the processing plant. Antibiotic treatments might mitigate the impact of the infection on performance. However, if the source of the infection is vertical transmission from breeder flocks, the nagging losses will continue in the processing plant leaving the producers with very little choice but to eliminate the breeder flock from the system. Even if the breeder flock is young and early in the production cycle, the industry is willing to eliminate the flock and take the losses up front rather than the multiplied losses on several orders of magnitude in the processing plant. In some cases the breeder flocks are eliminated due to mycoplasma infections even before the onset of egg production.

This system of prevention and control requires high levels of biosecurity in addition to efficient monitoring systems. The monitoring used by the poultry industry is aimed at detecting infections in the breeders as early as possible in order to prevent them from passing the agent on to their progeny and amplify the losses later in the production system. This is typically achieved by periodical serologic testing of breeders. Serum plate agglutination test (SPA), enzyme linked immunosorbent assay (ELISA) and
heamagglutination inhibition test (HI) are commonly used tests for this purpose. Detecting the infection is just one step in controlling the disease, another required step is identifying the source of the infection. This information is useful in identifying weaknesses in biosecurity and helps with corrective action in order to avoid the introduction of the agent in the future.

There are multiple molecular epidemiology assays developed specifically for genotyping and differentiating between mycoplasma strains for epidemiology purposes. These assays organize in two broad categories of techniques. First is DNA fingerprinting, including random amplified polymorphic DNA (RAPD) and amplified fragment length polymorphism (AFLP). DNA fingerprinting requires isolating the microorganism in a pure culture. The second category of molecular epidemiology tools is sequence typing. This is the preferred method in avian mycoplasma genotyping because it can be applied to clinical samples directly without the need for isolating the microorganism in a pure culture, something that is required for fingerprinting.

Unlike Mycoplasma gallisepticum (MG) which has 5 genomic targets identified for sequence typing, there is only one genomic target for MS sequence typing. That is the 5’ end partial sequence of the Variable Lipoprotein Hemagglutinin A (vlhA) gene. The vlhA assay is widely used to identify and differentiate between MS strains, field isolates and clinical samples. However, due to the size of the targeted segment and its variable nature, there is a degree of ambiguity in the inferred relations between MS outbreaks. If two MS strains are identical on the partial vlhA sequence, clonality is inferred, given that it is supported by epidemiological information. But if their sequences different at this
locus it is difficult to identify the degree of relatedness between them, even if they are epidemiologically connected.

In order to overcome these gaps in the current MS sequence typing system a three step research plan was designed, with 3 specific aims. First aim was to explore the vlhA gene to see if the maximum potential of its genotypic potential has been realized. The second aim was to investigate the MS Clustered Regularly Interspaced Palindromic Repeats (CRISPR) as an additional genomic target for MS sequence typing besides the currently available vlhA target. The third aim was to develop and validate a Multilocus Sequence Typing (MSLT) scheme as an additional sequence typing assay that would provide more refined information.

This document is organized in 5 chapters, the first is a literature review, summarizing history and a knowledge about Mycoplasma as a pathogen. The second through the fourth respectively address the 3 research aims listed in the previous paragraph. The research projects are in the format for peer reviewed journal publications. And the final chapter (the fifth) is the conclusion from the research reported in this document and recommendations for future research.
Chapter 1- Literature Review

1.1 History and evolution of Class *Mollicutes*

1.1.1 History and significance: Class *Mollicutes* refers to a group of eubacteria that are very small in size, lack the typical cell wall of other bacteria and possess the smallest genomes among the all independently replicating organisms. Contagious Bovine Pleuropneumonia (CBPP) is one of the most devastating animal diseases. It is the only bacterial disease listed on the prioritized communicable animal diseases A-list of the World Organization for Animal Health (http://www.oie.int) along with fourteen viral diseases. CBPP was investigated by Lois Pasteur himself along with some of the very first described microbial diseases. However, Pasteur couldn’t isolate the causative agent of the bovine pleuropneumonia or see it under the microscope due to its small size. The agent was mistaken first as a virus but then was successfully isolated in 1889 in a cell free media by E. I. E. Nocard and P. P. E. Roux and found to be a small wall-less bacteria (111).

In spite of the relatively long time since the first isolation of *Mollicutes* but their nature remained elusive. In the 1950 and 1960 scientific literature debated whether *Mollicutes* are wall-less L-forms of bacteria. The debate wasn’t over until the late 1960
when the genomic DNA hybridization excluded any relation between *Mollicutes* and stable L-Forms of bacteria and classified them as a separate group of microorganisms (132).

Class *Mollicutes* represents a very diverse group of microorganisms that inhabit a wide range of environments including human, animals, insects and plants. In most cases they live as commensals on their host tissue, and symbiotic relations have been reported in many insects (133). However, many *Mollicutes* can cause human, animal and plant diseases with significant losses in lives and economy. *Mycoplasma mycoides*, *M. pneumonia*, *M. genitalium*, *M. gallisepticum* and *M. synoviae* are just some prominent examples of the disease casing *Mollicutes*. They have also been known to be devastating cell culture contaminants.

In addition to the typical importance that comes with holding some of the major human and animal pathogens, class *Mollicutes* were the focus of the scientific community due to some of their unusual properties. One of the quests NASA commissioned since the early beginnings of space exploration was the search of extra-terrestrial life. That led the scientists to try to identify and study the smallest and most simple form of life here on earth. Morwetz and Tortellote with the help of NASA launched the first meeting on molecular biology of mycoplasma. This meeting along with a publication in *Scientific American* spurred interests and many research projects focusing on the study of *Mycoplasma* in the 1960’s. They were successful in understanding the cell wall of mycoplasma and use it as a model to study all biological membranes. However, they
were not successful in their initial goal of disassembling and reassembling the simplest living cell (132, 111, 130).

The concept of defining the simplest life form or the minimum requirements of life didn’t go away. In the 1980 the focus to understand the concept of minimum requirement of life was shifted towards DNA rather than the cell organelles (109). And with the advent of whole genome sequencing of Mollicutes (47, 63, 64); this goal became more achievable. By deleting 100 dispensable genes out of the 482 gene set of M. genitalium, the potential minimum gene set required for life was proposed by Craig Venter and his group in The J. Craig Venter Institute, Rockville, Maryland. They even went further and chemically synthesized the whole genome of M. genitalium. They also transferred naked DNA chromosome from M. mycoides into DNA free Mycoplasma capricolum cell which then exhibited the phenotypic characters of M. mycoides. Finally they repeated the same experiment but with a purely synthetic M. mycoides chromosome. These experiments attracted the attention not only of the scientific community but also the main stream media, as they reported the first successful attempts of creating at least partially synthetic life (51, 52, 53, 88).

1.1.2 Evolution and phylogeny: The original use of 16S ribosomal RNA (16S rRNA) genes led to the addition of archeabacteria as a third domain of life. The three suggested lineages seem to be at an equal distance from each other (166). The reorganization of all living things into three domains was further supported by other biological characteristics. Cell wall structure, plasma membrane structure, ribosomal
structure and protein synthesis are all fundamental differences at the basis of biology between the two apparently similar domains archaebacteria and eubacteria (97).

It was first hypothesized that based on the genome size gap between the walled bacteria and *Mollicutes* that the later must represent a primitive cell line that emerged before the evolution of cell wall (110, 159). But another opposing theory suggested that evolution of *Mollicutes* was due to the gradual loss of ancestral DNA. The accumulation of 16S rRNA sequence data supported the second model of evolution concluding that mollicutes represent a late evolutionary event rather than earlier primitive cells (97). The 16S rRNA enabled the positioning of most mollicutes in the eubacterial domain of life. However, at least one group of mollicutes (*Thermoplasma*) was found to have evolved within the archaebacterial domain suggesting the evolution of mollicutes happened independently more than once in two different domains of life (97).

Using the available 16S rRNA sequence data at the time in 1996 Maniloff (98) proposed a hypothetical scenario for the evolution of mollicutes. This scenario suggests that the first mollicutes evolved from the *streptococcus* branch of the low GC content gram positive bacteria around 600 million years ago. The estimated genome size of that ancestral mollicute microorganism was estimated to be 2 million base pairs. Then a further division on the phylogenetic tree occurred about 450 million years ago producing two branches. The first branch later evolved into *Acheloplasmatales* and *Anaeroplasmatales*. While the second branch contains the other two orders in class *Mollicutes* which are *Mycoplasmatales* and *Entemoplasmatales*. Both lines later proceed separately to evolve smaller species with genomes of 600 to 1200 kbp. According to this
model of evolution older *Mollicutes* have larger genome size than more recent *Mollicutes*. More recent genomic data supported this model and added that the loss of cell wall might have happened more than once within the class *Mollicutes* (35).

*Mollicutes* genome sizes are different not only between genera or families but even among strains of the same species (22, 85, 139). This could be at least partially due to the higher rate of genetic variation exhibited by *Mollicutes*. This in turn is thought to be due to their reduced capacity and accuracy of DNA replication and repair (36, 62, 66, 156). The high genomic variability could have been one of the tools available to mycoplasmas to explore different environmental conditions when adapting to new niches of habitat. In this regard, mycoplasmas could be considered one of the most evolved bacterial groups (165).

The enormous usefulness and importance of 16S rRNA sequences in *Mollicutes* phylogeny classification compelled The International Committee on Systematics of Prokaryotes Subcommittee on the Taxonomy of *Mollicutes* to consider the 16S rRNA sequence the primary basis for ranking any new mollicute species (20). It also could also be used as a proof of novelty of the proposed species; however, this proof needs to be substantiated by serologic and phenotypic evidence (161). In a broad respect the class *Mollicutes* hosts eight genera (*Acholeplasma*, *Anaeroplasma*, *Asteroleplasma*, *Entemoplasma*, *Mesoplasma*, *Mycoplasma*, *Spiroplasma*, and *Ureaplasma*), with generic definition depending on morphology, nutritional requirements and genetic phylogeny.
1.2 Biology of Class *Mollicutes*

1.2.1 Lack of cell wall: *Mollicutes* are unique amongst eubacteria in many aspects. However, most of their unique biological and morphological characteristics can be traced back to their lack of cell wall. The importance of such feature is reflected in their name (*mollis*, soft; *cutis*, skin, in Latin). It is expected that the spherical form should be the dominant morphology form of morphology in *Mollicutes* as they are bounded only by plasma membrane. But many species exhibit different morphologies including pear-shape, terminal tip element, filamentous or spiral shape (133). Some species are even capable of producing gliding movement. A group of mycoplasmas including two human pathogens *M. pneumonia* and *M. genitalium*, in addition to *M. gallisepticum*, an important poultry pathogen, possess a terminal protruding structure. This structure is used as an attachment organelle to help hold the mycoplasma against the host cell surface (7).

Due to the total lake of cell wall, it was not clear for years whether mollicutes replicated by binary fission (132). The controversy ended when microcinematographic evidence of mycoplasma binary fission was obtained. However, the mechanism of such process was not clear (19).

All of these features suggest the existence of cytoskeleton elements supporting many structural and biological functions performed by cell wall in other bacteria (131). However, the nature of these cytoskeleton elements was not clear and not defined with contradicting reports (29, 163). With the generation of more genomic and proteomic data, we are still far from understanding the breadth and complexity of the *Mollicutes*
cytoskeleton, but the composition and functions of these elements are starting to be resolved (21, 23, 95, 135).

Cell membrane is the only cell boundary the *Mollicutes* have since they lack the cell wall. The ease by which *Mollicutes* membranes can be isolated, manipulated and altered facilitated their use as a model to study structure and function of all plasma membranes. They have unusually high percentage of lipoproteins which are the dominant surface antigens in *Mollicutes*. Most notably is the high rate under which they undergo antigenic variations using remarkable molecular mechanisms, in order to evade the host immunity. Surface lipoproteins often play a major role in attachment and pathogenicity of *Mollicutes* to host cells (131,133). The immunogenicity and the high variation rate make these surface antigens a prime target for strain differentiation assays (45, 61, 69, 128).

*Mycoplasmas* and *Ureaplasmas* totally or partially lack the ability to synthesize cholesterol or fatty acids. This resulted in their inability to regulate the membrane fluidity through the incorporation of specific fatty acid in membrane. To overcome this problem *Mycoplasmas* incorporate large amounts of host cell cholesterol into their membrane, this serves as an effective membrane viscosity buffer (131). The requirement of cholesterol has long been considered an important criterion for Mollicutes taxonomic classification (161). Non cholesterol requiring groups like *Acheloplasmas* incorporate relatively low level of cholesterol their cell wall when it is present in the media (≈ 10%). *Mycoplasmas* on the other hand incorporate cholesterol from the media to levels (≈ 30%) very comparable to their eukaryotic host cell membranes (141). This is thought to play a major
role in facilitating the intracellular penetration or cell fusion between mycoplasmas and their host cells (150).

### 1.2.2 Small genome: Another prominent feature of Mollicutes is their small genome size. Even before having access to complete genomic sequencing information, accurate data about Mollicutes genome size was available (115, 125). The genome size of Mollicutes is significantly smaller when compared to other bacteria. But within mollicutes themselves there is a range of size, starting with the more recent Mycoplasmas with small genome from 580 kb in M. genitalium (47) to the phylogenetically more ancient Spiroplasmas up to 2,220 kb in S. ixodetis (15, 22). This still is significantly smaller when compared to the 4,639 of E. coli (13). The small size mycoplasmal genome is thought to have formed through the process degenerative evolution from their gram positive bacterial ancestors (98, 35).

The field of comparative genomics started to work immediately after the availability of the first few bacterial full genome sequences. The first available full genome sequence belonged to the gram negative bacteria Haemophilus influenza (46), but soon after two Mycoplasma species (M. pneumonia and M. genitalium) were sequenced almost simultaneously (47, 63, 64). The comparisons between the gram negative and Mycoplasma genomes produced some interesting insights about the lost parts of the Mollicutes genomes (65, 83, 151). A more recent study comparing 35 Mollicutes genomic information was almost congruent with the previous study (60). Most expectedly a significant loss of genes was observed in the genes encoding for cell wall.
However, they retained the capacity to produce membrane lipoproteins and transmembrane proteins in agreement with the fact that high portion of Mollicutes cellular proteins are associated with the cell membrane. Mollicutes lost almost all genes involved in amino acid synthesis and Cofactor biosynthesis dictating the addition of all amino acids and necessary vitamins to the artificial culture media (48).

Also most Mollicutes lost fatty acid synthesis genes resulting in reduced capacity of regulating membrane fluidity (150). This necessitates the incorporation of host cholesterol in their plasma membranes to relatively high levels. Mollicutes also suffer from scarcity of genes involved in the synthesis of nucleic acid precursors. This explains the abundance of potent nucleases that many Mollicutes posses and use to digest host nucleic acids (150). Gene groups involved in other cellular processes including protein secretion systems, signal transduction, transcription regulation, cell division DNA replication and repair are either reduced, insufficient or completely inexistent (60, 65, 83, 151).

On the other hand when we look at genes involved in DNA replication, transcription and translation the situation seems to be different. As expected the process of economizing the genes that accompanied Mollicutes evolution have to be restricted when it comes to genes involved in the central dogma of life. Parasitic life style can work in acquiring necessary small molecules from the environment, but it can’t help in obtaining fully functional macromolecules like proteins or DNA. This means that Mollicutes must have the ability to produce such molecules if they were to survive. Relative to their genome size Mollicutes have larger proportions of genes involved in
DNA replication, transcription and translation, compared to other walled bacteria. But we have to point out that this higher percentage is a function of their small genome size, and the absolute number of genes is still smaller compared to other bacteria. Most prominently is the very smaller number of genes involved in the DNA repair and the SOS response in Mollicutes (36, 62, 66, 156).

One of the interesting features of Mollicutes genome is the low G.C content. While the G.C content in Mollicutes ranges between 20% and 33% it still is lower than the related Gram-positive bacteria like Bacillus subtilis (42%), Lactobacillus viridescens (40%), Staphylococcus aureus (33%) or Clostridium perfringens (38%). Other Gram-positive bacteria have much higher G.C content Micrococcus luteus (75%), Streptomyces griseus (73%), and Mycobacterium tuberculosis (67%). While Gram-negative bacteria have values that are in between, like Escherichia coli (50%), Serratia marcescens (58%), Salmonella typhimurium (51%) (113, 160). The low G.C content of Mollicutes might be explained by the reduced activity of uracil–DNA glycosylase enzyme (dUTPase) and its reduced capacity to remove uracil residues from DNA. This process would lead to the gradual substitution of G.C positions with A.T positions leading eventually to the high A.T content of Mollicutes. This is supported by the fact that G.C content is not uniform within Mollicutes genomes. Intergenic regions have lower G.C content than the coding regions due to the selective pressure on the coding regions (38). The inefficiency of dUTPase might also explain at least partially along with the limited DNA replication and repair machinery the high mutation rates and the fast paced or tachylytic evolution characterizing the mollicutes. (36, 38, 62, 66, 156).
In spite of being small *Mollicutes* genomes are dynamic and constantly changing through processes like rearrangements, insertions, and deletions of genes or entire genomic segments (38). One of the striking examples of this was revealed when comparing the two genome sequences of *M. pneumonia* and *M. genitalium*. The genomes of these two phylogenetically related mycoplasmas could be divided into six genomic sections. The organization of orthologous genes within each of the segments was preserved, but the organization of the six segments within the genome was different. (65)

Repetitive elements are one of the hall marks of *Mollicute* genomes. Insertion segments like (IS-like) elements, integrated viral sequences, or repetitive endogenous gene families are all common in *Mollicutes* genomes. They represent hot spots for homologous recombination and genetic rearrangements and play an essential role producing the antigenic variations of *Mycoplasma* cell surface antigens and help them to escape host immune response (122, 158, 160). Surface antigen gene families with repetitive sequences occupy significant portions their genomes. These families provide a reservoir for constant variation in surface antigens that appears to be spontaneous within uniform population. This phenomenon happens in single colonies grown on artificial media without any apparent triggering mechanisms (4, 118). This seems at odds with the extreme frugality with which these microorganisms are economizing their genomes. However, the evolutionary benefits afforded by theses repetitive segments by allowing manipulations and variations using a minimum number of genes to produce endless possibilities of epitopes on the cell surface can easily outweigh the cost of caring these extended stretches of repeats in their genomes (28, 38, 96, 171).
1.2.3 Scavenging metabolism: As demonstrated the previous section *Mollicutes* have rather limited coding capacity due their small genomes. During their evolutionary history they evidently discarded almost all the genes associated with the biosynthesis of precursor molecules like amino acids, fatty acids, nucleic acids, cofactors, and vitamins. This makes them entirely reliant on the host metabolic machinery to provide them with the full spectrum of biochemical precursors (123, 124, 140). For this purpose *Mollicutes* produce a large variety of lytic enzymes including nucleases and proteases and use them to digest macromolecules and scavenge required nutrients from their environment or host cells (38).

In addition to the absence of most synthetic pathways, *Mollicutes* lack many of the essential energy generating pathways (38, 99). This explains the fastidious nature of many member of class *Mollicutes*. Perhaps one of the most surprising departures from the norms in *Mollicutes* biology is their inability generate ATP through the process of oxidative phosphorylation. All *Mollicutes* have deficient tricarboxylic acid cycle and have no cytochromes or quinones (104, 123, 133). Instead they resort to generate their ATP through low energy yield of substrate level phosphorylation process leading to inefficient consumption of substrate and large amounts of waste products. This can lead to destruction of host tissue due to depletion of substrate or accumulation of toxic substances.

According to their ability to utilize carbohydrates *Mollicutes* are divided into fermenters and nonfermenters. Most *Mollicutes* generate energy via the process of
glycolysis. It is possible to further metabolized pyruvate to produce lactate by lactate dehydrogenase or to acetyl coenzyme A by pyruvate dehydrogenase, reducing the pH of the environment and potentially leading to some tissue damage (133). Nonfermenters on the other hand produce energy through the arginine hydrolase pathway (5). In this pathway arginine is hydrolyzed into ornithine, ATP, CO2, and ammonia, which leads to increased pH of the environment and tissue damage as well. Some Mollicutes contain both glycolysis and arginine hydrolase pathways, giving them more metabolic versatility (120, 32). A few mycoplasmas including M. agalactiae, M. bovigenitalium, and M. bovis don’t possess the necessary enzymes for neither sugars nor arginine metabolism, but they can oxidize organic acids (lactate, pyruvate) to acetate and CO2 (104, 152). Ureaplasma represent a special case among all bacteria in that it relies in the hydrolysis of urea in generating energy (12, 116, 144, 151).

So far parallel evidence from genomic and biochemical data are congruent and show that Mollicutes have limited metabolic capacities hunting for their required nutrients in their environment relying mainly on host cell to provide them. Additionally they lack efficient pathway to utilize energy from substrate they acquire. This can explain the fastidious nature of Mollicutes in artificial media and their host cell associated status.

1.2.4 Central dogma of Mollicutes life: In most Mollicutes there is only one DNA polymerase compared to three different types of polymerases in other bacteria. The nature of DNA polymerase in Mollicutes eluded scientists for some time. Most notable is
the lack of the 5’ → 3’ proofreading exonuclease property of *Mollicutes* polymerase. The availability of whole genome sequences showed that the genes encoding for the α subunit of the enzyme responsible for the proofreading activity is missing (6, 64, 84). Many *Mollicutes* genomes failed to show genes encoding for other factors important in DNA replication process. Initiation, elongation and termination factors in addition to Pol I (Pol A) are all examples for these missing components. As mentioned before DNA repair components and efficiency in *Mollicutes* seems to be greatly reduced. This along with the lack of proofreading activity of the polymerase might be the reason of the increased mutation rates in *Mollicutes*. On its face this might be considered a disadvantage, however, along with the dynamic genomic rearrangements it can help *Mollicutes* explore new environmental niches and adapt to them and survive (36, 62, 66, 156).

No major differences have been found between the DNA-dependent RNA polymerase of mollicutes and that of other bacteria. Most notably is the resistance of its B subunit to binding with rifamycins antibiotics. This leads to the resistance of *Mollicutes* to this group of antibiotics, resembling in that some of the closely related gram positive bacterial like *Clostridium ramosum*, *C. innocuum*, and *C. acidurici* as a further evidence of having a common ancestor (14, 143). Another sign of gene saving strategy of *Mollicutes* is that they contain only one sigma factor suggesting that the response to environmental conditions are not performed at the transcription level.

Translation machinery probably is the most conserved system in *Mollicutes*. Their ribosomes are similar to the typical bacterial ribosomes in shape, size, and structure
Ribosomal proteins found in Mollicutes represent almost the same set of ribosomal proteins found in both gram negative and gram positive genomes and in similar operonal organization (63).

The typical organization of rRNA genes in bacteria is 16S-23S-5S working together as a single operon (129). However, some variations on this organization have been recorded. Most importantly from the scope of this document is the *M. gallisepticum* (MG) rRNA gene arrangement. MG carries two copies of ribosomal operons, one with the typical arrangement of 16S-23S-5S genes while the second copy have the 16S rRNA separated from the 23S-5S rRNA genes (58). This property has been taped later to be used for the MG intraspecific differentiation (128). As expected tRNA genes are kept near the absolute minimum number needed for the translation process (2, 160).

The highly conserved nature of the translational machinery is reflected in the fact that they are considered the basis of phylogeny. However, *Mollicutes* seem to always find a way to be different. The codon usage in *Mollicutes* is almost identical to the universal codone usage with one exception. This exception is the reassignment of UGA from a stop codon to a tryptophan codon (38). All *Mollicutes* have low G.C content due to an A.T directional mutational pressure (113). The outcome of this phenomenon is the usage biases towards synonymous codons with A or T. One of the outcomes of this bias is the use of UGA as a tryptophan codon. While this has no apparent effect on the survival of *Mollicutes* it complicates the expressing cloned *Mollicutes* genes in usual expression systems.
The cost to the drastic savings in the structures needed for protein synthesis didn’t go unnoticed in Mollicutes biology. It is reflected in the fact that protein synthesis and ultimately cell replication in Mollicutes are much slower than other bacteria (142). This in turn might explain the rather long incubation period and the mild and chronic nature characteristic of many Mollicutes diseases, aided of course with many other survival mechanisms that help them escape the host immunity.

1.3 Avian Mycoplasma

1.3.1 Mycoplasma species and host range: There has been more than 20 Mollicute species isolated from a wide variety of different avian hosts so far. Most of the species belong to genus Mycoplasma with two exceptions (Acholeplasma ladlawii and Ureaplasma gallorale). Chicken is the most frequent host of isolated mycoplasma followed by turkey, pigeon and geese (77). Only four mycoplasma species are considered pathogenic for birds. The rest of avian Mollicutes species are commensals. Several reports (1, 49, 81, 114, 134, 146, 147) indicated the involvement of some of the avian commensal mycoplasma species in clinical diseases. However, no conclusive evidence exists to indicate that any of them is the etiological agent of any naturally occurring disease in birds. Two mycoplasma species, M. gallinarum and M. gallinaceum are often isolated during the attempts to isolate avian pathogenic species (77). They are considered
contaminants and they are relatively fast growing often competing with the much slower growing pathogenic species interfering with their growth and obscuring the results.

The four pathogenic mycoplasma strains are *Mycoplasma gallisepticum* (MG), *Mycoplasma synoviae* (MS), *Mycoplasma iowae* (MI) and *Mycoplasma meleagris* (MM). Two of these species, namely MG and MS, are pathogenic for both chicken and turkey, while the other two species, MI and MM, are pathogenic only for turkey. MG is often regarded as the most pathogenic and the most economically significant *Mycoplasma* disease facing the poultry industry, followed by MS. MI and MM are of minor importance even for the turkey industry as the losses caused by them are often negligible (25).

Avian *Mycoplasma* species typically have a narrow range of hosts, similar to other *Mollicutes* in that respect. Occasional reports suggest that some of the avian pathogenic mycoplasma species can be isolated from multiple wild and domestic bird species; however, most of these reports don’t provide solid evidence on the successful colonization or adaption of new hosts (8, 9, 10, 16, 39, 68, 93, 94). One exception to this is the cross species outbreak of MG that started in 1994 in house finch (*Haemorhous* (formerly *Carpodacus*) *mexicanus*) on the Mid Atlantic east coast in the United States (92). This outbreak spread into the house finch population across the United States and reached the west coast (91). This outbreak led to minimizing their overwinter survival (43) and up to 60% declines in populations (66). Recent studies suggest that this population shift might have resulted from multiple introductions but seemingly one MG lineage has been able to establish itself into the newly exploited population. This event attracted a lot
of attention in an attempt to study and understand the evolution of emerging diseases in a new host species (36, 62, 66, 156).

1.3.2 Diseases: *Mycoplasma gallisepticum* (MG) is often considered the most pathogenic avian mycoplasma species (90). It was first described in England in 1905, but wasn’t successfully isolated and correctly characterized as *Mycoplasma* until the 1950’s (37, 100, 157). Chronic respiratory disease (CRD) is often used synonymously with MG induced clinical signs in chicken. It’s a strict respiratory pathogen that attaches itself to respiratory mucous membranes. The disease is often slow progressing and the incubation period can extend up to 21 days. Younger birds are more susceptible than older adult birds (25).

MG can be transmitted both horizontally and vertically. Horizontal transmission occurs when the infection is delivered to birds probably from another infected source flock. This can be aided by fomites and contaminated material including dust, feathers, or organic matter some of which can be airborne. Probably the biggest culprit in the horizontal transmission is human vector coupled with poor biosecurity measures and improper personal protective equipments (90). The vertical transmission from parents to progeny tends to occur intermittently and at a very low rate. This might in turn prolong the incubation period as it takes longer time for the disease to be detected at the flock level (54, 55, 89).

MG is primarily a respiratory pathogen. Clinical signs include respiratory noise and discharges. Reduction in feed consumption, slower growth rate and reduced egg
production are all reported clinical signs. However, the disease might be silent in a subclinical mode of infection particularly with adult birds that have been exposed or immunized before. This is the reason why MG outbreaks might go unnoticed in the field leading to much greater loses later due to higher condemnation rates in the broiler processing plant (40). Turkeys are considered more susceptible to MG infection with more severe clinical signs. Swelling of the infraorbital sinuses can be severe leading to cessation of feed consumption and weight loss (90). In addition to the respiratory disease, reports about potential role of MG in brain infections and inflammation in turkey have been published (26, 31, 154, 167).

MG as a single pathogen in an outbreak almost never occurs in the field, complications with secondary pathogens always happen. *E. coli*, Newcastle Disease and Infectious Bronchitis are common agents associated with MG infections (42, 59, 145). This complicated chronic respiratory disease is the main reason for high mortality. Some highly virulent strains, however, seem be capable of producing high mortality on their own (40). The biggest losses faced by the industry may be in the processing plants both in chicken and turkey. Partial and total condemnations can reach as high as 30% and constitute the primary reason of why the poultry industry prefers the mycoplasma free status particularly in breeder flocks (90).

*Mycoplasma synoviae* (MS) is the only other *Mycoplasma* species that is pathogenic for both turkey and chicken. Features of the disease caused by MS are very similar to that of MG in many respects with a few differences. Horizontal and vertical disease transmission patterns are very similar; however, MS seem to spread faster
through a population (79). MS is capable of exploring more tissues than MG, including respiratory, synovial and most recently the genital epithelium (44). MS can produce lameness and respiratory disease. Typically lameness and joint infections are the more prominent form of the disease and responsible for most of the losses, particularly in turkey. However, recently MS is causing more respiratory disease in the US and egg shell softness in Europe as the pathogenicity of MS appears to be on the rise (87).

*Mycoplasma meleagridis* (MM) is exclusively a turkey pathogen (27). Similar to other avian mycoplasmas it can be transmitted vertically and horizontally. It’s a much milder disease compared to MG and MS. It produces airsacculits in vertically infected young turkey poults. It can also produce bone deformities and reduced performance if complicated with other diseases (35, 36, 103). Artificial insemination and vertical transmission seems to play a critical role in sustaining the infection (107, 108). Currently with the availability of MM free turkey breeders the disease doesn’t represent a problem for the industry anymore.

*Mycoplasma iwoae* (MI) is the other exclusive *Mycoplasma* turkey pathogen (17). And similar to MM the venereal and vertical route of transmission is critical in its epidemiology (73, 170). Horizontal transmission in the hatchery might play a role in spreading the disease, as MI persist in the meconium due to its predilection to the intestinal tract (106). It causes increased embryonic mortality and in turn reduced hatchability. Very little is know or reported on its clinical presentation in young birds, but poor performance and leg problems are common symptoms of infection (155). Unlike MM not all primary breeding companies are free from MI. According to the number of
clinical cases submitted to our lab MI is another mycoplasma pathogen that seems to be on the rise. Increased mortality and joint arthritis are common, and MI has been isolated from affected joints (unpublished data).

1.3.3 Pathogenicity: Diseases produced by mycoplasma species are often mild and chronic. An argument could be made that mycoplasma pathogens come close to be “ideal parasites” living in harmony with their host causing minimal damage. It has been noted that the molecular mechanisms of Mollicutes pathogenicity have not been elucidated. However, it is suggested that their pathogenicity is likely due to the close interaction with the host tissue and due to the host immune response (133). In that respect avian mycoplasmas are no exception to the rule. No potent toxins have been identified for avian mycoplasma species. Accumulation or mycoplasma metabolic byproducts and changing the pH of the local environment and depletion of cell resources have all been accused of producing pathogenicity (133).

*Mycoplasma gallisepticum* and *Mycoplasma synoviae* have both been reported to perform extensive surface antigenic variations. Both of these organisms have expansive gene families including long stretches of repetitive elements used in elaborate genetic mechanisms to produce such variability (11, 57, 100, 72, 117, 118, 119). This variability in both organisms occur spontaneously even within the same colony (4, 118). These gene families occupy significant proportions in their genomes which reflect that these mycoplasmas rely on antigenic variation to a great extent in their survival. A model could be proposed that on a population level a pathogen can utilize such advantage as a basic survival
apparatus. These spontaneous variations can produce diverse population within the host, so that when the immune system can eliminate a dominant population, others can survive and extend the infection. This could be the main reason of the inability of the immune system to clear the infection (133). The fact that small organisms with limited genomes have such an impressive arsenal of tools devoted to the deception of host immune system should place mycoplasmas in the higher ranks of advanced bacterial pathogens (126, 138).

Another potential mechanism of survival is the intracellular position of mycoplasma which has been reported in MG (86, 112, 164). Intracellular position might provide a safe haven for MG organisms to escape the effect of antibiotic treatments and extend the chronicity of the infection.

1.3.4 Diagnosis and control: The word “gold standard” is often associated with *Mycoplasma* isolation as a mean of confirmatory diagnosis. Isolation, however, with such fastidious organisms is not an easy task. Particularly out of clinical samples that are often contaminated with other organisms that can hinder their growth. As discussed before for other *Mollicutes*, avian *Mycoplasmas* are not capable of synthesizing most of their precursor molecules and require them to be supplied in the isolation media. MS may be the most fastidious in all four avian pathogenic species. MS colonies were observed as satellites next to *Micrococcus* culture which enabled the identification of nicotinamide adenine dinucleotide (NAD) as requirement for their growth (24).

A typical mycoplasma media is rich in protein requiring the addition animal serum and the incorporation of yeast extract. Avian mycoplasmas are naturally resistant
to Penicillin and Thallium Acetate making these products valuable additives to the media
to control contamination. Phenol red is used as an indicator for mycoplasma growing and
changing the pH of media due to fermenting sugars or hydrolyzing arginine. A commonly
used media with great rate of success is the Frey’s modified media (48). Best results are
achieved at 37˚ C incubation and it takes 3 to 5 days for enough growth to occur to change
the phenol red color. Colony morphology of *Mycoplasma* on solid agar takes the
characteristic form of “fried egg” appearance which has been attributed to the lack of cell
wall. Colony morphology cannot be used for species identification (79). Direct
fluorescent antibody staining on agar surface is the most commonly used technique for
(33, 149) species identification.

Due to its difficulty and limitations and the great deal of expertise required for
successful avian mycoplasma isolation a plethora of molecular identification methods
have been developed. DNA probes and Polymerase Chain Reactions (PCR) have been
specifically developed to detect and differentiate between avian Mycoplasma species
(78). Real time PCR reactions have been developed as well with the added benefit of
multiplexing more than one test in the same reaction (127). These techniques have the
ability to detect mycoplasma DNA in the clinical samples directly and bypassing the
need of isolation and identification. These advancements have reduced the definitive
diagnosis turn around time in mycoplasma cases from the range of a few days to a few
weeks down to the range of a few hours.

Serological tests are extremely useful tool in monitoring the infection. Avian
mycoplasma monitoring and eradication programs are built as part of the National
Poultry Improvement Plan (NPIP) (3). Serological tests are fast, accurate and inexpensive assays (78). Assays like Serum Plate Agglutination (SPA), Heamagglutination Inhibition (HI) and Enzyme Linked Immunosorbent Assay (ELISA) tests are instrumental in executing these control and eradication plans (3). SPA is easy to perform and reagents are commercially available and are particularly sensitive and can detect the infection very early. But SPA is known for its low specificity and false positive results (56). It is typically used as a screening test to detect any potential infection. HI on the other hand is time and labor consuming and the required reagents are not commercially available, but exquisitely specific. It is used as a confirmatory test for SPA positive reactors, to avoid any false positive results leading to unnecessary condemnations (78). ELISA test is reasonably sensitive and specific with the added benefit of automation (78).

Prevention is the most relied upon arm in avian mycoplasma control. Due to its vertical transmission potential the industry elected to have all the breeders free from mycoplasma (3). All the primary breeders are free from MG, MS and MM. But not all of them are free from MI. At the parent stock level flocks are mostly free, however, occasional flocks get infected due to some gaps in the biosecurity. Several antibiotics can be used to treat mycoplasma outbreaks (18, 70). Antibiotic treatment can mitigate the clinical signs, reduce mortality and prevent drops in egg production (71, 121). Antibiotic treatment can reduce the rate of shedding (149, 168), but typically under field conditions it is not successful in eliminating the vertical transmission. In many situations when the mycoplasma is endemic and the challenge is overwhelming antibiotic treatments are used to control the losses (76).
Inactivated and live vaccines are commercially available for both MG and MS (79, 90). They have been used with variable degrees of success. There have been reports that vaccinated layer flocks in endemic areas have better egg production (169). Vaccines can be used to augment the preventive measures and be part of a control plan including biosecurity, but they cannot be relied upon to prevent the infection on their own.

1.3.5 Molecular epidemiology: The industry choice to maintain *Mycoplasma* free status at least for breeders require the elimination of infected flocks before they pass the infection to their progeny. But the control of the disease requires more than stamping out of the infection at its point of entry. It requires the identification of the source of the infection and pointing out the biosecurity breaches that lead to the infection in the first place. This is where the molecular epidemiology tools are most useful. Multiple assays have been described for the intraspecific strain differentiation of avian mycoplasma. Protein profile analysis, Random Amplified Polymorphic DNA (RAPD), Restriction Fragment Length Polymorphism (AFLP), ribotyping, strains specific DNA probes and Gene Targeted Sequencing (GTS) (45, 74, 75, 82, 172). Two categories of techniques, however, are commonly used, DNA finger printing (RAPD and AFLP) and sequence typing. Sequence typing by direct amplification and sequencing of DNA specific targets is the most commonly used method for avian mycoplasma molecular epidemiology. It is preferred by laboratories as it doesn’t require the isolation of avian mycoplasma in pure culture as do the DNA finger printing methods. It is also more reproducible and more scalable.
Out of all four pathogenic avian mycoplasmas sequence typing assay have been developed for only MG and MS. At least 5 genomic targets have been identified and used for MG sequence typing and in most cases multiple targets are used simultaneously for greater discriminatory power and more reliable epidemiological conclusions (45). For MS, on the other hand, there is only one target identified for such purpose, the partial sequence of the *vlhA* gene (41, 61, 69, 162). Using the *vlhA* sequence typing assay, unrelated MS strains were shown to be identical or very closely related strains were shown to be different. This indicates that there is a degree of ambiguity still in sequence typing methods and there is a great room for improvement. Next generation sequencing and comparative genomics are starting to be used for molecular epidemiologic and population evolution studies (36, 62, 66, 156).
References


Fiorentin, G. R. Franco, N. S. Freitas, D. Fri´as, T. B. Grangeiro, E. C. Grisard, C. T.
Lopes, E. L. Loreto, H. M. Madeira, G. P. Manfio, A. Q. Maranhaˇo, C. T. Martinkovics,
Oliveira, R. F. Paixaˇo, F. O. Pedrosa, S. D. Pena, M. Pereira, L. Pereira-Ferrari, I. Piffer,
L. S. Pinto, D. P. Potrich, A. C. Salim, F. R. Santos, R. Schmitt, M. P. Schneider, A.
P. Urmenyi, M. H. Vainstein, L. W. Zucherato, A. J. Simpson, and A. Zaha. Swine and
poultry pathogens: the complete genome sequence of two strains of *Mycoplasma


160. Wei, S., Z. Guo, T. Li, T. Zhang, X. Li, Z. Zhou, Z. Li, M. Liu, R. Luo, D. Bi,
H. Chen, R. Zhou, and H. Jina. Genome sequence of *Mycoplasma iwoae* strain 695, an

Mandelco, J. Sechrest, T. G. Lawrence and J. Van Etten. A phylogenetic analysis of the


Chapter 2: The Genotyping Potential of the *Mycoplasma synoviae* *vlhA* Gene

Mohamed M. El-Gazzar, Amy N. Wetzel, and Ziv Raviv

A Slightly different version has been published in Avian Dis. 56:711–719, 2012

Received 12 April 2012; Accepted 25 July 2012;

Published ahead of print 13 August 2012
2.1 Abstract

*Mycoplasma synoviae* (MS) continues to cause significant losses to poultry producers, and studying the epidemiology of infection is an important component of MS control. The partial variable lipoprotein hemagglutinin A (*vlhA*) gene is the only genomic target identified so far for MS sequence typing. The *vlhA* gene codes for two variable cell surface proteins, lipoprotein and hemagglutinin, and the proposed mechanism for the variation is gene conversion between a single expressed gene and an array of pseudogenes. The upstream portion of the *vlhA* gene is present in the genome in a single copy (not present in the pseudogenes), and it is the only part of the gene that can be used for targeted sequence typing. However, the 3’ end of the *vlhA* ‘‘single copy’’ as well as this region’s discriminatory potential for genotyping purposes has not been established.

The purpose of this study was to identify the exact limit and the genotyping potential of the *vlhA* single copy region. New PCR assays were developed to amplify the entire conserved region and part of the variable region of the *vlhA* gene. Amplification and sequencing were performed on a variety of MS samples and on in vitro sequential generations of a standard MS strain. Sequence analyses determined the site and composition of the most proximal sequence variation that could be attributed to a gene conversion event, and they predicted the end point of the *vlhA* single copy region. The results indicated that a currently available ‘‘revised Hammond’’ PCR spans the whole single copy of the *vlhA* gene and exploits the full genotyping potential of this MS genomic target. In addition, this study allows interesting insight into the gene conversion
mechanism of MS and offers the opportunity for further investigation this mechanism in mycoplasmas.

2.2 Introduction

Mycoplasmas belong to a unique group of bacteria that lack a cell wall and possess the smallest genome among all independently replicating organisms (19). *Mycoplasma synoviae* (MS) is one of four pathogenic avian mycoplasma species that can be transmitted both horizontally and vertically. MS infection can be associated with upper respiratory disease, airsacculitis, synovitis and tenosynovitis, and keel bursitis. Disease severity can be influenced by other respiratory pathogens (e.g., Newcastle disease virus and infectious bronchitis virus) and host species susceptibility (e.g., turkeys are more susceptible than chickens). MS can be a very costly disease to the poultry industry particularly due to reduced live-production performances, increased condemnation rates, and implementation of control measures (13). Eradication of MS has been traditionally the preferred method of control in commercial poultry flocks (11). However, the recent increase in the number and severity of MS outbreaks in certain geographic areas has lead to increased interest in vaccination (MS-H live vaccine strain) as a preventive measure. Studying MS epidemiology is an essential part of outbreak investigation and control and often requires intraspecific differentiation (genotyping) of involved MS field and vaccine strains. The commonly used molecular epidemiology methods for MS genotyping are random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), and target sequence typing (TS) (12). Unlike
RAPD and AFLP, TS can be performed directly on clinical samples and does not require the isolation of the organism in a pure culture, a difficult task in many field situations.

The variable lipoprotein hemagglutinin A (\textit{vlhA}) gene is the sole gene that has been identified as a genomic target for MS sequence typing (1). The \textit{vlhA} gene encodes a single polypeptide chain that undergoes posttranslational cleavage to produce two surface proteins: MSPB, a surface lipoprotein encoded by the 5’ end of the gene, and MSPA, a hemagglutinin encoded by the 3’ end of the gene (15,16,17). The \textit{vlhA} is part of a gene family. In the whole genome sequence of MS strain 53 (MS53) the \textit{vlhA} gene family comprises 64 genes that occupy more than 8% of the genome (22) (GenBank accession AE017245). Only one member of the \textit{vlhA} gene family contains a promoter, while the rest of the genes in the family are promotorless and are referred to as pseudogenes. The \textit{vlhA} gene is both size and phase variable. The size and phase variation of the \textit{vlhA} gene has been proposed to occur mainly via the mechanism of gene conversion between the expressed and the pseudogenes (16). Based on the degree of sequence variation the \textit{vlhA} gene can be divided into three main regions: conserved, semivariable, and highly-variable (Figure 2.1).

The initial upstream 410–530 bp segment of the conserved region was demonstrated in the genome in a single copy (herein ‘‘single copy’’), while the rest of the conserved region and the semivariable and highly variable regions were detected in multiple copies within the pseudogenes (16) (Figure 2.1). Proline rich repeats (PRR) and a polymorphic region (RIII) are two variable regions detected within the conserved ‘‘single copy’’ of the \textit{vlhA} gene and were not associated with variation related to
pseudogenes conversion events (1). Target sequence typing assays to the single copy of the expressed \textit{vlhA} gene were developed and utilized for MS genotyping applications (7,9). These assays were further revised and modified for better discriminatory power, sensitivity, and specificity (23). Nevertheless, these TS assays were demonstrated to have imperfect discriminatory power and failed to discriminate between several unrelated MS cases analyzed in our laboratory.

Although the single copy of the \textit{vlhA} gene is the only genomic target currently available for MS sequence typing the full genotyping potential of this genomic target has not been established. Herein we investigate the \textit{vlhA} gene single copy 3’ endpoint and the initial sequence variations that could be attributed to putative gene conversion events.

2.3 Materials and Methods

2.3.1 \textit{Mycoplasma synoviae} field samples selection. Sample types were tracheal, choanal cleft, or joint swabs from clinical and subclinical MS infections of turkey or chicken flocks. Field samples represented in this study were submitted to our lab over a 2-year period from Ohio, Arkansas, and the Netherlands (Table 2.1.). Most of the field samples were selected from groups of cases that were 100% identical by the ‘‘revised Hammond’’ \textit{vlhA} PCR (23) product (data not included), and were selected to evaluate whether sequence differences could be detected downstream to the revised Hammond PCR amplicon. The selected samples belonged to eight different \textit{vlhA} genotypes. Several
other laboratory and field strains and field samples were selected to increase the repertoire of tested samples (Table 2.1.).

### 2.3.2 Mycoplasma synoviae strains and culture method

Two laboratory adopted MS strains, MS-WVU1853 and MS-F10-2AS; one field strain, MS-K1968; and the MS-H live vaccine were all received from the University of Georgia. The field strain 2008-062-17 was received from North Carolina State University. All MS strains were grown in modified Frey’s broth and solid agar and were incubated at 37 C (12).

### 2.3.3 Serial in vitro passages of MS-K1968 experiment

Fifteen in vitro serial passages of MS strain K1968 were performed in order to investigate the extent of the \(vlhA\) single copy region and the initial site at which putative gene conversion with pseudogenes may occur. Broth media (30 ml) were streaked on a modified Frey’s agar and the plate was incubated at 37 C. Two single colonies from this plate were picked and designated clone 1 and clone 2. Clones 1 and 2 were separately resuspended in 1 ml of modified Frey’s broth. Immediately after resuspension, 200 ml from each broth was assigned for genomic DNA extraction; these 200-ml aliquots were designated “clone 1, generation 1” and “clone 2, generation 1.” The remaining 800 ml of the resuspended media were incubated at 37 C until yellow color change was detected. Subsequently, 30 ml of incubated culture from clones 1 and 2 was streaked separately on modified Frey’s agar and the plates were incubated at 37 C. A single colony was picked from each of the two clones’ plates and resuspended in 1 ml of modified Frey’s broth. Immediately after
resuspension, 200 ml from each clone was assigned for genomic DNA extraction; these 200-ml aliquots were designated “clone 1, generation 2” and “clone 2, generation 2”. For clone 1 the above process was repeated to obtain 13 generations of progeny from the original picked colony, and for clone 2 the above process was repeated to obtain 15 generations of progeny from the original picked colony (Figure 2.2). Genomic DNA of all generations were extracted, amplified by the smaller segment PCR, and sequenced.

2.3.4 Genomic DNA extraction. Dry swabs were pooled up to five per sample and were resuspended in 1 ml of phosphate-buffered saline (PBS). Genomic DNA was extracted from 200 ml of PBS resuspension solution or 200 ml of modified Frey’s broth culture using the QIAamp DYNAMini Kit (QIAGEN, Valencia, CA) following the manufacturer’s recommendations.

2.3.5 PCR design and amplification of extended vlhA gene segments. May and Brown (pers. comm., 14) sequenced the whole vlhA gene of 10 MS strains and revealed highly conserved islands within the hypervariable parts of the gene. Based on this sequence data a first set of primers was designed to amplify an approximately 1290-bp segment, herein “larger segment,” of the 5’ part of the vlhA gene (Figure 2.1). A second set of primers was designed to amplify a smaller segment of the vlhA gene by targeting another highly conserved island at the 900-bp level, herein “smaller segment” (Figure 2.1). This second set of primer design was based on May and Brown sequence data as well as the sequence information obtained from the larger segment amplicons of several
MS samples in this study. Both primer sets utilized the forward primer published by Wetzel et al. (23) and newly designed reverse primers. Primer sequences and the approximate size of targeted sequences are summarized in Table 2.2.

All of the study’s PCRs were performed with the Roche FastStart High Fidelity kit (Roche Diagnostics, Indianapolis, IN) following the manufacturer’s recommendations with some modifications. Briefly, the PCR was performed in the MJ-Mini thermocycler (BioRad Laboratories, Hercules, CA) in a total volume of 50 ml reaction mix containing 5 ml of 10X FastStart High Fidelity reaction buffer (1.8 mM MgCl2), 1 ml of 10 mM deoxynucleotides, 4 ml of each 5 mM primers, 0.5 ml of 5 U/ml FastStart High Fidelity enzyme, and 5.0 ml of DNA template solution. All reactions were performed using a thermocycler program of a hot start at 95 C for 3 min, 40 cycles of denaturing at 94 C for 30 sec, annealing at 56 C for 30 sec, extension at 72 C for 60 sec, and a final extension at 72 C for 5 min. PCR amplicons were separated by electrophoresis on a 1.5% agarose gel containing 0.53 Tris-borate-ethylenediaminetetraacetic acid buffer and 0.5 mg/ml ethidium bromide and visualized by ultraviolet trans-illumination.

2.3.6 Sequencing and sequence analysis of PCR products. PCR products were purified using the QIAquickHPCR Purification Kit (350) (QIAGEN) according to manufacturer’s recommendations. In PCR products with more than one band, the specific band was excised from the agarose gel after electrophoresis and in these cases the DNA was purified using the QIAquickHPCR Gel Extraction Kit (50) (QIAGEN) according to manufacturer’s recommendations.
Sequencing was performed by using dye termination sequencing method on the purified PCR products in both forward and reverse directions (Genewize Inc., South Plainfield, NJ). Sequence data were analyzed for complete overlapping of complementary sequences, editing, and consensus construction with the SeqMan program (in Lasergene; DNASTAR, Inc., Madison, WI). Alignments of sequences were constructed by the Clustal V method with a gap penalty of 10 using the MegAlign program (in Lasergene; DNASTAR, Inc.). All the obtained sequences were trimmed in accordance with the smaller segment amplicon sequence to appropriately align in the same MegAlign file.

2.4 Results

2.4.1 Amplification and sequencing of extended *vlhA* gene segments. The amplification of the larger segment of the *vlhA* gene was successful in all tested samples. A total of 36 samples were amplified and each produced a specific intense band accompanied by a few variable nonspecific faint bands (Figure 2.3). Size variation was observed among the *vlhA* PCR products of this assay. Out of the 36 PCR products only six were successfully sequenced while the rest of the samples failed to produce good quality sequence data. We assume that the failure to sequence most of those PCR products might be due to the length of the amplicon and/or its level of homogeneity as well as the requirement for gel purification in PCR products with multiple bands. The amplification of the smaller segment of the *vlhA* gene was successful in all the tested samples. A total of 38 samples were amplified and each produced a specific highly
intense band (Figure 2.3). Twenty-four of the smaller segment PCR products generated good quality sequences. Most of the samples were tested by both PCR assays, but some of the samples were tested only by either the larger segment or smaller segment assays due to limited quantity of genomic DNA template. A list of all tested samples and their epidemiological relatedness are presented in Table 2.1.

2.4.2 Sequence analysis of extended vlhA gene amplicons. The recovered sequences provided new data about the downstream region of the vlhA gene from a variety of field cases, a vaccine strain, and two laboratory strains. The extended vlhA sequence data showed differences in the gene downstream region among samples that shared the same revised Hammond vlhA sequences (Table 2.1). However, the sequence variations within the same vlhA genotype (Table 2.1) were not detected upstream to nucleotide number 421 (counting from the ATG start codon) of the reference sequence published by Noormohammadi et al. (15) (GenBank accession AF035624) herein ‘‘reference sequence.’’ The extended vlhA sequences of each vlhA genotype were identical until the

reference sequence nucleotide number 421, where a triple A codon (nucleotide positions 421, 422, and 423) changes to a GCT or vice versa (Figure 2.4). Two of the vlhA genotypes, designated genotypes 4 and 7, started the variation at later points, but in all the other vlhA genotypes the variation started at the aforementioned nucleotide (421) and never upstream. In vlhA genotypes 4 and 7 the sequence variation started at nucleotides 512 and 426 of the reference sequence, respectively.
The findings of the AAA/GCT nucleotide triplets as the first variation in the *vlhA* gene of supposedly same MS strains were in agreement with the results of the in vivo study conducted by Slavec et al. (21) and with our study’s in vitro MS-K1968 serial passages experiment. The observed sequence variations could be attributed to gene conversion events that occur instantly between the expressed gene and the pseudogenes within the same strain. Hence the observed sequence variations were of limited genotyping significance and not useful for MS strain differentiation purposes.

It was further observed that highly conserved short sequences (22 bp) were located immediately upstream to the initial sequence variations attributed to putative gene conversion events. All the analyzed *vlhA* sequences of this experiment in addition to many GenBank sequences (total of 62 analyzed sequences) had the same identical sequence at this segment (5’ 399 AGC ATT AAG CGG ATC GGT TAC T 420 3’). This sequence appeared to be universal to all MS *vlhA* GenBank sequences, and it was utilized for the reverse primer sequence of the Hammond et al. (7) and revised Hammond MS *vlhA* PCRs (23). Furthermore, in silico analysis of MS53 whole genome sequence demonstrated the presence of 63 *vlhA* pseudogenes (22) (GenBank accession AE017245). These 63 *vlhA* pseudogenes had different sizes ranging from the largest being 1961 bp to the smallest being 215 bp. Out of those 63 pseudogenes only the largest eight had homologous sequences extending into the 700-bp conserved region of the *vlhA* gene (Figure 2.1). A perfect homology between the eight largest pseudogenes and the expressed gene conserved region was observed immediately upstream to nucleotide 421 according to the reference sequence, and extended for 16 identical nucleotides (5’ 406
AAG CGG ATC GGT TAC T 420 3’) (Figure 2.6). This observation locates the end of the \textit{vlhA} gene single copy of MS53 strain at nucleotide 405 of the reference sequence. Initiation of sequence variation at nucleotide 512 of the reference sequence was detected in genotype 4 (Table 2.1.) and clone 2 of the in vitro passage experiment. Five nucleotides upstream to this point (nucleotide 507), a highly conserved region was detected. This region is 13 bp long and is conserved among all 62 studied sequences with two nucleotide variations (5’ 495 AAA AT\textit{W} MAC AGC A 507 3’).

\textbf{2.4.3 Sequence analysis of sequential generations of MS culture.} Two single colonies from MS strain K1968 culture were designated clone 1 and clone 2 and were grown for sequential cycled passages, named “generations” (13 generations for clone 1 and 15 generations for clone 2), in order to observe the in vitro occurrence of gene conversion events as well as the sequence stability of the \textit{vlhA} gene. A 21-bp insertion in our MS-K1968 strain \textit{vlhA} sequence relative to the reference sequence resulted in seven amino acids insertion (Figure 2.5A), and hence the differences in residue numbering between the cloned strains sequences and the reference sequence. In clone 1 sequence change appeared at the 12th generation with 14% sequence variation and persisted onto the 13th generation, when the experiment was concluded. The variation started at nucleotides 442/443/444 (AAA.GCT) or amino acid residue 148 (K.A) (Figure 2.5B). In clone 2 a variation of 4% occurred at the ninth generation and persisted with no further changes through generation 15, when the experiment was concluded. This variation
began at position 533, which was the second nucleotide of the codon encoding for amino acid residue 178 (Figure 2.5C).

2.5 Discussion

*Mycoplasma synoviae* infections continue to cause considerable losses to poultry producers. Moreover, the introduction of virulent MS strains to new geographical areas and the association of MS with new disease syndromes (5) increase the tendency to use the MS-H live vaccine. These developments emphasize the need for improved MS genotyping methods as well as capabilities to differentiate between the vaccine and field strains.

The partial *vlhA* gene is the only genomic target identified so far for genotyping by TS (1). As elaborated in the introduction, the *vlhA* gene is both size and phase variable and the main mechanism for its variation formation was proposed (16). The recombination of pseudogenes into the expressed *vlhA* gene through the gene-conversion mechanism makes only the single copy of the gene (Figure 2.1) relevant for strain differentiation purposes (genotyping). Sequence variations beyond the single copy region could be attributed to gene conversion events within the same MS strain, or even within the same colony (17), and have no genotyping relevancy. Noormohammadi et al. (16) have indicated that the *vlhA* gene single copy ends at or after nucleotide 410 and before nucleotide 530, of the reference sequence, while the currently *vlhA* TS assays are limited to nucleotide 420, accordingly (23). Our objectives were to identify the endpoint of the
*vlhA* gene single copy region and explore possible improvements to the *vlhA* TS assay’s discriminatory power.

Until recently, extended amplification of the *vlhA* gene was impractical due to the lack of sequence data of the whole *vlhA* gene from different MS strains and the assumption that the downstream region of the gene was too variable and unsuitable for species-specific primer design. In fall 2010 Dr. Meghan May shared with us the whole *vlhA* gene sequences of 10 MS strains; data were later published by May and Brown (14). This MS *vlhA* sequence data revealed highly conserved islands among strains within the semivariable and highly variable regions of the *vlhA* gene (Figure 2.1). The identified conserved islands opened the opportunity for amplification and sequencing of extended segments of the *vlhA* gene. Two extended MS *vlhA* PCRs were developed to study the genetic variation of different MS field samples and laboratory, field, and vaccine strains. Out of the two assays the smaller segment PCR product (Table 2.2) appeared to be more suitable for sequencing purposes.

Sequence analysis of extended PCR products of samples that were grouped in the same *vlhA* genotype based on the sequence of the revised Hammond PCR product (Table 2.1) and 15 in vitro serial passages of the MS strain K1968 indicated the first *vlhA* sequence variation that could be attributed to a gene conversion event. The first sequence variation was identified at nucleotide 421 according to the reference sequence, where a triple A codon changes to a GCT or vice versa. Our findings of the AAA/GCT nucleotide triplets as the initial variation attributed to a putative gene conversion event in the *vlhA* gene were in agreement with the in vivo study conducted by Slavec et al. (21). The
observations of a highly conserved short sequence of 22 bp immediately upstream to the initial sequence variation attributed to gene conversion and the presence of portions of this sequence in the lipoprotein edge of \( vlhA \) pseudogenes suggest that the endpoint of the single copy of the \( vlhA \) gene lies within this 22-bp region. Sequence analysis of the MS53 \( vlhA \) eight largest pseudogenes indicates 16 identical nucleotides within the highly conserved 22 nucleotides of the MS53 \( vlhA \) expressed gene (Figure 2.6). This observation locates the end of the \( vlhA \) single copy of this strain at nucleotide 405 of the reference sequence. The highly conserved 22-bp sequence immediately upstream to nucleotide 421 of the reference sequence is actually the annealing site of the revised Hammond PCR reverse primer. This study’s findings and observations conclude that the revised Hammond PCR covers the whole single copy of the \( vlhA \) gene and exploits the full genotyping potential of this MS genomic target.

The identification of the putative MS \( vlhA \) initial gene conversion site opens the opportunity to study this gene conversion mechanism and allows comparisons with gene conversion processes in other bacteria. \textit{Neisseria gonorrhoeae}, \textit{Borrelia burgdorferi}, and \textit{T. pallidum} are some examples of microorganisms in which antigenic variations via gene conversion have been studied (2,18). Similar to the MS \( vlhA \) gene family, a surface variable lipoprotein (\( vls \)) has been identified in \textit{B. burgdorferi} (24). The \( vlsE \) is the single expressed gene, while there are 15 additional unexpressed silent cassettes arranged in an opposite orientation that act as a donor reservoir for sequence variation in the single expressed gene via gene conversion. The \( vls \) silent genes are flanked by a 17-bp conserved segment proposed to be the point of recombination during the process of gene
conversion (24). This organization of *B. burgdorferi* *vls* genes resembles the highly conserved sequences identified in MS *vlhA* pseudogenes immediately upstream to the *vlhA* putative gene conversion sites. In *T. pallidum* (3,4) and *N. gonorrhoeae* (6,10) areas of homology as short as 4–5 bp can act as additional recombination sites between the expressed gene and its family of pseudogenes to enable the process of gene conversion. Indications of additional MS *vlhA* gene conversion sites downstream to the initial variation attributed to putative gene conversion event (nucleotide 421 of the reference sequence) were detected in this study. A highly conserved region with two variations was detected between nucleotides 495–507 of the reference sequence in all 62 analyzed sequences of this study (see Results; sequence analysis of extended *vlhA* gene amplicons). Moreover, sequence aligning of the MS53 expressed *vlhA* gene with its eight largest pseudogenes revealed highly conserved regions downstream to nucleotide 421 of the reference sequence. These downstream conserved regions varied from 5 to 35 bp long (data not shown). We hypothesized that these areas of intraspecific homology, similarly to other bacteria, might be involved in the putative *vlhA* gene conversion mechanism.

RecBCD and RecF are the two identified recombination pathways in *Escherichia coli*. In *N. gonorrhoeae* surface antigen gene conversion is thought to be dependant on RecA protein and RecF-like pathway but not the RecBCD pathway (8). A list of proteins and enzymes deemed essential for pilin antigenic variation of *N. gonorrhoeae* was reviewed by Cahoon and Seifert et al. (2). In order to verify whether MS possesses the required machinery for gene conversion, we surveyed the predicted proteome of MS53 (GenBank accession AE017245). Apparently, MS53 encodes homologues of some gene
conversion required genes including recA, recR, ruvAB, and recU (homologue of recC in other bacteria). However, other genes that code for gene conversion protein factors were not found in MS53, including recX, recJ, recQ, and rdgC. None of the two identified recombination pathways in E. coli (RecBCD and RecF) is completely fulfilled in the genomes of any mycoplasma species and data are lacking on the specifications and details of mycoplasma recombination pathways (20).

It is interesting to note that intraspecific genetic variations can be limited to smaller segments of the gene as observed in clone 2 of the in vitro sequential passages experiment or can be quite extensive as in clone 1 of that experiment (Figure 2.5B,C), an observation that suggests the potential plasticity of the MS-K1968 conversion process. Nevertheless, further studies will be required to elucidate the overall MS vlhA gene conversion mechanism.

Herein we defined the endpoint site of the single copy segment of the MS vlhA gene, which is the only gene segment relevant for intraspecific differentiation (genotyping) purposes. Our results and observations indicate that the revised Hammond PCR (23) covers the whole single copy of the vlhA gene and exploits the maximum genotyping potential of this MS genomic target. Apparently, additional genomic targets will be required to augment the discriminatory power of genotyping by sequencing of MS strains. In addition, the development of expanded MS vlhA PCRs and the sequence analysis of these amplicons allow interesting insight into the gene conversion mechanism in MS, and open opportunities to expand our knowledge about this mechanism in mycoplasmas.
2.6 Acknowledgements

We are thankful for Dr. Meghan May for providing the \textit{vlhA} sequence data that inspired the idea for this work. We also thank Dr. Naola Ferguson-Noel at the University of Georgia and Dr. David Ley at North Carolina State University for providing several \textit{Mycoplasma synoviae} strains.

2.6 References


variable major membrane antigens, one of which is a putative hemagglutinin. Infect. Immun. 65:2542–2547. 1997.


Table 2.1: Summary of selected samples and \(vlhA\) larger segment and smaller segment PCR amplification and sequencing results.

<table>
<thead>
<tr>
<th>Sample #</th>
<th>Description</th>
<th>Amplification</th>
<th>Sequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td>All samples were from the same company and from a relatively small geographical area. All cases bore 100% sequence similarity within the revised Hammond (vlhA) segments.</td>
<td>Yes</td>
<td>ZR 13-F51 ZR 20-F2</td>
<td>ZR 54</td>
</tr>
<tr>
<td>ZR 74-504</td>
<td></td>
<td>ZR 20-F2</td>
<td>ZR 46</td>
</tr>
<tr>
<td>ZR 74-505</td>
<td>Case ZR13 and ZR15 were part of an epidemiologic study in meat-type turkeys. The two (vlhA) samples included here were from the same flock. Cases ZR13, ZR31, and ZR71 were from meat-type turkeys that were geographically related.</td>
<td>Yes</td>
<td>None</td>
</tr>
<tr>
<td>ZR 14-F61-B1</td>
<td>Case ZR14 F61-B1 received from a backyard flock suspected to be the source of infection to a breeder breeder flock (case ZR14 F61). Case ZR66 received from a separate group of four samples originated from breeder and breeder breeder flocks in the same geographical area.</td>
<td>Yes</td>
<td>ZR 14-F61-B1 ZR66-3908 ZR66-3802</td>
</tr>
<tr>
<td>ZR 14-F61-C1</td>
<td>Two (vlhA) segments from different samples with no geographical connection between them.</td>
<td>Yes</td>
<td>ZR 14-F61-C1</td>
</tr>
<tr>
<td>ZR 66-3908</td>
<td>Samples originated from chickens in the Netherlands located on three farms: two neighboring farms and a third farm 100 km away. The samples from the two neighboring farms (74-504, 505, 520, 524, 74-504 and 505) were identical to each other and different from the neighboring farms.</td>
<td>Yes</td>
<td>ZR 100-ZR 54</td>
</tr>
<tr>
<td>ZR 66-3908</td>
<td>Additional field cases with no particular epidemiological reasons were amplified and sequenced to increase the repertoire of tested samples.</td>
<td>No</td>
<td>ZR 100-ZR 54</td>
</tr>
<tr>
<td>ZR 66-3909</td>
<td>Additional field cases with no particular epidemiological reasons were amplified and sequenced to increase the repertoire of tested samples.</td>
<td>No</td>
<td>ZR 100-ZR 54</td>
</tr>
<tr>
<td>ZR 66-3111</td>
<td>Additional field cases with no particular epidemiological reasons were amplified and sequenced to increase the repertoire of tested samples.</td>
<td>No</td>
<td>ZR 100-ZR 54</td>
</tr>
</tbody>
</table>

---

(Continued)
**Table 2.1:** (Continued)

The *vlhA* genotypes represent 100% identical strains using the gene segment amplified by the revised Hammond assay. ZR13-FS1, ZR 20-FF2, and ZR 46 sequences were obtained only from the larger segment PCR. ZR43, MS-WVU1853, and MS-2008-017-62 were obtained from sequencing of both the larger and smaller segment PCRs. The rest (21 sequences) were obtained only from sequencing of the smaller segment.

N/A = not applicable.
Table 2.2: Summary of MS *vlhA* PCR assays primer sequences and amplicon size.

| MS *vlhA* PCR       | Forward primer sequence (5’ to 3’)
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GGC CAT TGC TCC TRC TGT TAT&lt;sup&gt;C&lt;/sup&gt;</td>
</tr>
<tr>
<td>Revised Hammond&lt;sup&gt;B&lt;/sup&gt;</td>
<td>AGT AAC CGA TCC GCT TAA TGC</td>
</tr>
<tr>
<td>Larger segment</td>
<td>GGC CAT TGC TCC TRC TGT TAT</td>
</tr>
<tr>
<td>Smaller segment</td>
<td>GGC CAT TGC TCC TRC TGT TAT</td>
</tr>
</tbody>
</table>

<sup>A</sup>The same forward primer in all three assays starts at nucleotide 45 from the start codon.

<sup>B</sup>Wetzel et al. (23).

<sup>C</sup>Underlined letters represent International Union of Pure and Applied Chemistry/International Union of Biochemistry code of base substitutions.

<sup>D</sup>Sizes are calculated with reference to the start codon of the WVU1853 strain *vlhA* gene sequence by Noormohammadi et al. (15) (GenBank accession AF035624).
Figure 2.1: A schematic representation of the *vlhA* gene family. The MSPB protein (lipoprotein) is encoded at the 5’ end and the MSPA protein (hemagglutinin) is encoded at the 3’ end of the expressed gene. The conserved, semivariable, and highly variable with conserved 3’ end regions of the gene are presented by solid, dashed, and dotted arrows respectively. The gene’s single copy and its PRRs and RIII are represented in a black rectangle and white squares, respectively. The parts of the gene that are present in the expressed and pseudogenes (multiple copies) are represented in gray.

...........................................................................................................................................................................(Continued)
Figure 2.1 (Continued)

The (FP) arrow represents the same forward primer used for all the assays; revised Hammond, the smaller segment, and the larger segment PCRs. The (RH) arrow represents the reverse primer of the revised Hammond PCR; the (RS) arrow represents the reverse primer of the smaller segment PCR; the (RL) arrow represents the reverse primer of the larger segment PCR.

The pseudogenes exist in the genome in different sizes and spanning different parts of the gene. Only a limited number of large pseudogenes cross into the conserved region and overlap the gene’s single copy. Most of the pseudogenes align with parts of the semivariable or the highly variable regions. The pseudogenes in the graph are just for illustration of size differences and do not represent any particular pseudogenes.
Figure 2.2: Steps performed for the serial in vitro passages. These steps were repeated to produce 13 generations of clone 1 and 15 generations of clone 2.
Figure 2.3: Agarose gel pictures showing the successful amplification using the larger and the smaller segment assays. (A) Five different amplified samples in one, two, or three replicates (labeled by numbers 1 to 5) using the larger segment assay; size variation is observed among the tested samples. (B) Eleven different amplified samples (labeled by numbers 6 to 16) using the smaller segment assay.
Figure 2.4: Example of three \( vlhA \) genotypes. These genotypes are determined by the revised Hammond assay (Table 2.1). Ruler on the top and numbers on the right of sequences indicate the position of nucleotides relative to the top reference sequence.

Sequence similarity extends all the way up to the nucleotide 420, after which sequence variation within the same genotype starts to be noticed. Black boxes represent residues that differ from the consensus. Gray shaded box marks the base triplet where the sequence variation starts.
Figure 2.5: Amino acid sequence deduced from nucleotide sequence of partially amplified \textit{vhlA} gene of sequential generations derived from two clones of a single MS strain K1968 colony. The sequences are demonstrated in an amino acid format to allow extended presentation. (A) Difference in amino acid sequence between the reference sequence published by Noormohammadi et al. (15) and the sequence of MS strain K1968 used for the cloning experiment: seven amino acid insertion highlighted in gray-shaded box.  

(Continued)
Figure 2.5 (Continues)

(B) Generations from clone 1 continued to have identical sequences through the 11th generation, and variation started on the 12th generation at nucleotide 421 of the reference sequence leading to change in amino acid residue 148. (C) Variations in clone 2 started at ninth generation at nucleotide 512 of the reference sequence leading to change in amino acid residue 178. Sequence variations are marked with thin line boxes, and the initial variation position is marked with the thick line box; this amino acid residue was variable only in clone 1. Ruler on the top and numbers on the right of sequences indicate the position of residues. Clone and generations numbers are mentioned on the left.
**Figure 2.6:** Alignment of expressed *vlhA* gene against 8 of the largest pseudogenes.

Sequences obtained from GenBank according to Noormohammadi et al. (15) reference sequence and MS53 with the eight largest MS53 pseudogenes sequences. In black boxes residues that differ from consensus. Blue box highlights the 16 nucleotides identical among all studied sequences and in the pseudogenes lying immediately next to the proposed initial gene conversion site AAA/GCT nucleotides (solid border box). The 16-bp segment indicates the end of the single copy region at nucleotide 405. All eight largest pseudogenes aligned with additional two locations in the conserved regions. One of the two is a 4-bp homology (highlighted by dotted border box) located 30 bp upstream to the conserved segment (highlighted by dashed border box). Another segment of 6 bp exists upstream to this position (not shown). The significance of these two homologies is not known. All sequences obtained from GenBank accession AF035624 and AE017245.
Chapter 3: *Mycoplasma synoviae* Genotyping by Sequencing the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs)

Mohamed El-Gazzar and Ziv Raviv

This chapter has been prepared in a manuscript format to be submitted for publication in a peer reviewed journal.
3.1 Abstract

The partial vlhA gene sequence is the only genomic target currently used for *Mycoplasma synoviae* sequence typing. In this study we investigate the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) as a potential additional target of MS sequence typing. Nineteen different primers were designed and seventeen different primer set combinations were used to attempt to directly amplify the MS CRISPR sequence of the surrounding regions in 24 different MS strains, field isolates or clinical samples. None of the used primer set combinations was successful in consistently amplifying the targeted sequence in all the tested samples. However, a few CRISPR sequences were generated, allowing us some insights into the nature, the function and the potential use of this target for MS sequence typing. The CRISPR sequence region seems to be too variable to directly amplify using conventional PCR techniques. The whole genome sequence might be a more proper tool to explore and harness the genotypic power of this region.

3.2 Introduction

*Mycoplasma synoviae* (MS) is one of two pathogenic mycoplasma species that can cause disease in both chicken and turkey. Significant losses are faced by the poultry industry due to mycoplasma infections. In addition to that, the infection can be transmitted both vertically and horizontally which complicates the epidemiology of the disease (17). As a result, the poultry industry elected to use eradication as the method of
choice to combat the disease. Eliminating the infected breeder flocks is the strategy used to stamp out the infectious agent at the point of entry to prevent the infection from trickling down the production system (15).

This system of control requires high levels of biosecurity in addition to efficient monitoring for early detection and timely corrective actions. The monitoring used by the poultry industry aims to detect the infection in the breeders as early as possible in order to prevent the transmission of the agent onto their progeny, which will amplify the losses down the production system. This is typically achieved by periodical serological testing of breeders using serum plate agglutination test (SPA), enzyme linked immunosorbent assay (ELISA) and haemagglutination inhibition test (HI) (16). Detecting an infection and taking corrective actions is just one step in controlling the disease. Identifying the source of the infection is another important part of the control program. Sourcing the agent is used in identifying the weaknesses in the biosecurity and help correcting these weaknesses in order to avoid reintroduction of the infectious agent in the future.

There are multiple molecular epidemiology tools used for differentiating between mycoplasma strains, understanding the epidemiology of the disease and identifying the source of the infection. Two broad categories of molecular techniques are used for avian mycoplasma molecular epidemiology. First is DNA finger printing, including random amplified polymorphic DNA (RAPD) and amplified fragment length polymorphism (AFLP) (16), while the second category is sequence typing. Of these two categories sequence typing is the preferred method in avian mycoplasma genotyping because it can
be applied to clinical samples directly without the need for isolating the microorganism in a pure culture, as required for fingerprinting.

Unlike *Mycoplasma gallisepticum* (MG) which has 5 genomic targets identified for genotyping by sequencing (8, 25), there is only one genomic target for MS sequence typing. That is the 5’ end partial sequence of the variable lipoprotein hemagglutinin A (*vlhA*) gene (10, 11). In the second chapter of this document (7, 31) we identified the maximum potential for *vlhA* as a sequence typing target as being the original 420 bp from 5’end of the conserved region of this gene.

In this chapter we are exploring Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) as a genomic target for MS sequence typing. Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) system is regarded as a prokaryotic defensive mechanism against invading genetic elements (1, 2). CRISPR immune function is performed using small interfering RNA-like (siRNA) principles (29). CRISPR sequence was first discovered by chance next to an isozyme of alkaline phosphatase (iap) gene on E. coli chromosome. Five direct repeats were found with 32 nucleotides spacer sequence. A dyad symmetry was also found in the direct repeats, but the biological significance of these sequences was unknown at the time (12).

With the accumulation of whole genome data from multiple bacteria, it was later possible to identify CRISPR sequences in 40% of bacterial and 90% of Archeal genomes (9, 19). CRISPR sequences typically are formed of 28–40 bp repeats with dyad symmetry alternating with unique spacer sequences which can be 26–72 bp. The number of CRISPR motifs is highly variable between species and ranges from 2 up to 249 repeats.
within the same array (19). The number of CRISPR arrays per genome is variable as well, and up to 18 arrays within the same genome were documented (5).

CRISPR sequences are flanked from the 5’ end by a 21 leader sequence that acts as a promoter for transcription (13). In addition to the repeat arrays, CRISPR associated sequences (cas genes) were identified only in CRISPR containing prokaryotes and were not found in CRISPR negative prokaryotes. The cas genes are constantly located in close proximity to CRISPR sequences suggesting these genes have a function role in the system (13). The cas genes are integral part of CRISPR/cas system, where they encode for proteins and enzymes that are involved in the biology of the system (1, 26) (figure 3.1).

CRISPR biological function remained unknown for several years after its original discovery, with speculations on their role in gene regulation, DNA repair and other biological processes (20, 21, 27). But later, observations were made about the similarity between the spacer sequences and mobile DNA elements including bacteriophages and plasmids (3, 22, 23). These observations lead to proposing the CRISPPs to be a sort of prokaryotic acquired immune system against invading DNA, and since that time scientific evidence has accumulated proving CRISPRs to be indeed an immune system (2).

In van der Oost et al. 2009 (29), a model was proposed on how CRISPR/cas system performs immune function where 3 stages are involved in such process: (i) acquisition of invading DNA sequence and inserting it as a new spacer sequence immediately after the leader sequence with the insertion of a new repeat as well forming
together a newly added CRISPR unit to the array. There appear to be deletion of the older CRISPR units to avoid the indefinite growth of the array. (ii) Expression of CRISPR and posttranscriptional processing to produce guide RNA directed against the invading DNA molecules. (iii) The use of guide RNA in identification and interference with the invading DNA in a siRNA-like mechanism. cas proteins are involved in all the stages performing variable functions.

The most relevant feature for our purpose is the addition and deletion of CRISPR units making the CRISPR arrays a continuously evolving sequence in any genome and could be used to differentiate between isolates within the same species (4, 23). Even before their biological role was discovered CRISPER sequences were used for *Mycobacterium tuberculosis* strain differentiation via a technique known as Spacer Oligotyping or spoligotyping (14). In *Mycoplasma synoviae* only one CRISPR loci and 17 putative cas genes have been identified in-silico (CRISPRdb; http://crispr.u-psud.fr/crispr/) (9) (figure 3.2). The array contains 11 unique spacer sequences (30 bp long) and 12 (36 bp long) identical repeats with a total size of 759 bp.

The purpose of this study is to design a PCR assay to amplify the MS CRISPR array and evaluate its sequence as a new genomic target for genotyping. More than twenty primer sets have been designed in the attempt to consistently amplify and sequence the MS CRISPR array. However, only a few successful amplifications were achieved and only three sequences were generated. It appears that the region surrounding the MS CRISPR array is unstable and direct amplification of the CRISPR sequence from isolates and clinical samples is not possible in our hands. However, the few sequences we
generated provided some insights about the structure of the CRISPR system in MS. Based on the results we obtained in this experiment and in order to harness the sequence typing potential of the MS CRISPR, a whole genome sequencing approach might be a better tool for such a purpose.

3.3 Materials and Methods

3.3.1 *Mycoplasma synoviae* strains, field isolates and clinical samples: A total of 23 samples were used in this experiment including MS laboratory adapted strains (N=2), the MS-H vaccine strain, field isolates (N=2) and MS DNA positive clinical samples (N=19). In addition, the CRISPR sequence was obtained from the full length genomic sequence of the MS53 strain from Brazil, the only available MS whole genome sequence on GenBank (30) (table 3.1). Clinical samples included tracheal, choanal cleft and joint swabs from clinical or subclinical MS cases in commercial turkey or chicken flocks that were confirmed positive by real-time PCR (24). The clinical samples tested in this study were from submissions to our lab over a period of 4 years (from 2009 to 2013), mainly from Ohio but also from Arkansas and Indiana (table 3.1).

3.3.2 Genomic DNA extraction: Clinical samples were pooled (5 swabs per pool) into 1 ml phosphate-buffered saline (PBS) sample pools. For MS strains and field isolates, a total volume of 200 µl were inoculated in 3 ml of modified Frey’s liquid media and incubated at 37 ºC until the phenol red indicator in the media changed color.
indicating mycoplasma growth. The liquid culture was then centrifuged at 1500 rpm for
30 minute. The pellets were resuspended in 200 ml of modified Frey’s broth. Genomic
DNA was extracted from 200 ml of PBS (clinical samples) resuspension solution or 200
ml of resuspended culture (strains of filed isolates) using the QIAamp DNA mini Kit
(QIAGEN, Valencia, CA) following the manufacturer’s recommendations.

3.3.3 Primer design: originally we aimed to directly amplify the CRISPR arrays.
Primers were designed to be placed in the upstream and the down stream regions of the
CRISPR array. Initial attempts to amplify failed; so we decided to separately amplify and
study both the upstream and downstream regions aiming to find conserved sequences in
those areas that can be used to place new primer. Based on the sequence information we
obtained from the upstream and down stream regions; new primers were designed for the
purpose to identify primers that will universally amplify the CRISPR array in all MS
samples for the purposed of sequence typing. All primers and primer combinations that
were used in this study are summarized in tables 3.2 and 3.3.

3.3.4 PCR design and amplification: Nineteen different primers were designed
for this study specifically of the upstream and the downstream regions of the CRISPR
(table 3.2 and figure 3.3). The designed primers were used in 17 different combinations.
Some of the combinations were tested with multiple annealing temperatures in an attempt
to optimize the PCR reaction (table 3.3). In addition, for some of the primer
combinations the number of replications cycles was increased to 45 or 50 cycles. But in
general PCRs in this study were performed using the Roche FastStart High Fidelity kit (Roche Diagnostics, Indianapolis, IN) according to manufacturer’s recommendations with some modifications. PCRs were carried out using MJ-Mini thermocycler (BioRad Laboratories, Hercules, CA) in a total volume of 50 ml containing 1.0 ml of 10 mM deoxynucleotides, 5.0 ml of 10X FastStart High Fidelity reaction buffer (1.8 mM MgCl2), 5 ml of 5 mM primers, 0.5 ml of 5 U/ml FastStart High Fidelity enzyme, and 5.0 ml of DNA extract solution.

All reactions were performed using a thermocycler program of a hot start at 95 °C for 3 min, 40 - 50 cycles of denaturing at 94 °C for 30 sec, annealing at 50 - 56 °C for 30 sec, extension at 72 °C for 90 sec, and a final extension at 72 °C for 5 minutes. Annealing temperatures and the number of cycles were altered depending on the primer combination used (table 3.3). Electrophoreses was performed on PCR products on a 1% agarose gel with 0.53 Tris-borate-ethylenediaminetetraacetic acid buffer and 0.5 mg/ml ethidium bromide, at 90V for 45 minutes. The PCR products were visualized by ultraviolet trans-illumination (figure 3.4).

**3.3.5 Sequencing.** PCR products purification was performed using the QIAquickHPCR Purification Kit (350) (QIAGEN) following manufacturer’s recommendations. The purified PCR products were sequenced using Sanger sequencing method (Dye termination chemistry) for forward and reverse directions. The sequencing process was performed in (Genewize Inc., South Plainfield, NJ).
3.3.6 **Sequence analysis:** The analysis of sequence data was performed with the SeqMan program (in Lasergene; DNASTAR, Inc., Madison, WI). Alignments of sequences were constructed by the Clustal V method using the MegAlign program (in Lasergene; DNASTAR, Inc.).

### 3.4 Results

#### 3.4.1 Original direct amplification of CRISPR array and sequencing:
Out of all the MS positive clinical samples that were tested originally to directly amplify and sequence the CRISPR array only ZR15 was successfully amplified and sequenced. The F102AS strain amplification and sequencing resulted in a sequence that didn’t match with the CRISPR array. Instead the sequence matched a gene sequence that is immediately downstream from the CRISPR array on the MS53 whole genome. This might be suggestive of a different arrangement of sequences in this area which might be an indication of a variable region.

#### 3.4.2 Upstream and downstream amplification and sequencing:
After many failed attempts of directly amplifying the MS CRISPR array, we aimed to study the amplify, sequence and study the areas upstream and downstream form the CRISPR array in order to find a conserved sequences that can be used to design primers. The amplification and sequencing of this region was relatively more successful than the CRISPR array (figure 3.5). At the end of this experiment we had 10 upstream and 7
downstream sequences (data not shown). Using the generated sequence information, 3 more sets of primers were designed in another attempt to directly amplify the CRISPR sequence.

**3.4.3 Final direct amplification of CRISPR array and sequencing:** Several MS positive clinical samples were successfully amplified using different combinations of the 3 different primer sets designed for the final amplification (table 3.3 and figure 3.6). However, only 2 additional CRISPR arrays were successfully sequenced (ZR133 and WVU1853). An attempt to sequence some of the generated bands using a primer walking technique, but it was also unsuccessful.

**3.4.4 Sequence analysis:** In total we had 4 MS CRISPR sequences to analyze. Three of those sequences were generated in this study (ZR15, ZR133 and WVU1853). The fourth sequence was obtained from the MS53 whole genome sequence from GenBank. The ZR15 CRISPR sequence showed 3 additional CRISPR units compared to MS53 genome sequence (3 additional repeats and 3 additional unique spacers) as the sequence of this MS clinical sample showed 15 palindromic repeats and 14 unique spacer sequences (figure 3.7). ZR133 had one additional CRISPR unit compared to MS53. WVU1853 seem to have two CRISPR units less than MS53, 4 of the units didn’t align in the proper position and were missing, but two of them aligned downstream. However, the WVU1853 sequence we obtained seems to be truncated and there could potentially be additional CRISPR units that our sequence didn’t reach far enough to include (figure 3.7).
Other observations included that two of the spacer sequences were 29 bp and one spacer sequence was 31 bp long instead of the standard 30 bp. Both of shorter spacer sequences were in the MS53 sequences that were obtained from GenBank. The longer spacer sequence was in WVU1853 spacer sequence. It was also noticed that the palindrome repeat sequence ended in T instead of C in the first 5 repeats, but starting from the 6th repeat the last nucleotide became C similar to the default sequence of the MS CRISPR palindromic repeats. This is an indication that there could be some variations in the repeat sequences (try to find a similar incidence in other species).

3.4.5 Spacer sequence identification: A total of 46 spacer sequences were blasted against the GenBank data base. None of them had a 100% match with any nucleotide sequence in the GenBank data base, except the spacer sequences of MS53 which are expected to have 100% match with the MS53 whole genome sequence. Other less congruent matches included primates, other animals including insects, protozoa, bacteria and viruses. However, many of these matches had a high probability of being just by chance as the length of the investigated spacer sequences is relatively short.

3.5 Discussion

Understanding the epidemiology of mycoplasma is an integral part of controlling the disease in the poultry industry (15). The partial sequence of the variable lipoprotein hemagglutinin A (vIhA) gene is the only genomic target for MS sequence typing. The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) sequence is used
by prokaryotic species as a form of acquired immunity against the invading parasites (2). A model was proposed on how CRISPR/cas system performs its function (29). Following this model CRISPR sequences might provide a biological record as new spacer sequences are added and old ones are dropped. In this paper we explored the CRISPR sequence as a potential second sequence typing target for MS.

Identified one CRISPR array by the web site and many cas genes (figure 3.2). Seventeen different primer set combinations were tested on 24 different MS strains, field isolates and clinical samples, targeting the amplification of MS CRISPR and the surrounding sequences. Almost all the primer set combinations were tested with multiple amplification conditions (table 3.3). However, none of the tested primer sets produced consistent amplification of the CRISPR sequence in all of the tested MS strains (figures 3.3 – 3.7).

In spite of the inconsistent amplification we still successfully generated a few CRISPR sequences in this study. We observed differences in the number of CRISPR units between different strains, ranging from 10 units in WVU1853 to 15 units in ZR15. We also observed variations in the length of some of the unique spacer sequences, and that the repeat sequence was different in one position in some of the units in one of the CRISPR sequences. Examination of the spacer sequences revealed that none of the spacer sequences matched any potential bacteriophage sequences. It is still possible that these spacer sequences belong to bacteriophages that have not yet been identified, there aren’t many studies defining the phage population parasitizing on genus Mycoplasma (32, 28). The amount of variation we observed in the CRISPR units numbers and features in the
very limited number of sequences generated in this study could indicate that the CRISPR/cas system is active in MS.

The amplified sequence out of F102AS using CRISPR-F1 and CRISPR-R2 primers didn’t align with the CRISPR sequence of MS53 obtained form GenBank. Many of the tested strains successfully generated sequences of the upstream and downstream regions of the CRISPR region. However, those same strains did not amplify when tested with the final amplification primer set designed to fit their own upstream and downstream sequences. These two pieces of information together further indicate that the CRISPR area is a variable region and direct amplification of the Mycoplasma synoviae CRISPR sequence problematic. In order to develop a useful sequence typing assay for MS it has to consistently apply to all MS strains and field isolates and preferably clinical samples. Something we failed to achieve with our attempt to directly amplify MS CRISPR sequence. The whole genome approach might be one way to harness the genotypic power of CRISPR assay. In 2012 Delaney et. all (6) sequenced the whole genome of sequence of 12 Mycoplasma gallisepticum isolates spanning 13 years form the naturally occurring house finch outbreak. They found that the MG CRISPR sequence was useful in tracing this naturally occurring outbreak. This approach was applied to other bacterial species as well (18). However, the whole genome sequence still is not a commonly used technique in investigating mycoplasma poultry outbreaks. There is only one available MS whole genome sequence (30), but it is possible that when there are more whole genome sequences we discover more than one array in MS which may offer better epidemiologic information and better understanding of CRISPR/cas system function in MS.
In this study we investigated the CRISPR sequence as a potential genomic target for MS sequence typing and that the direct amplification of CRISPR sequence was not consistently successful. The most obvious conclusion would be that the CRISPR region is a variable region that will be difficult to directly amplify. This might be circumvented by full length genomic sequencing of MS genomes. This should be pursued as the few sequences we obtained showed a potentially active CRISPR array and the potential use for this region as a sequence typing tool for MS.

3.6 References


<table>
<thead>
<tr>
<th>Samples</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 MS53</td>
<td>Isolated from Brazil, Chicken, Unknown date</td>
</tr>
<tr>
<td>2 MS-H</td>
<td>Isolated from Australia, Chicken, 1986</td>
</tr>
<tr>
<td>3 K1968</td>
<td>Isolated from Colorado, Turkey, 1983</td>
</tr>
<tr>
<td>4 F102AS</td>
<td>Isolated from North Carolina, Chicken, 1970</td>
</tr>
<tr>
<td>5 2008-062</td>
<td>Isolated from North Carolina, Unknown, 2008</td>
</tr>
<tr>
<td>6 ZR13</td>
<td>Clinical samples Ohio, Turkey, 2009</td>
</tr>
<tr>
<td>7 ZR15</td>
<td>Clinical samples Ohio, Chicken, 2009</td>
</tr>
<tr>
<td>8 ZR19</td>
<td>Clinical samples Ohio, Turkey, 2009</td>
</tr>
<tr>
<td>9 ZR20</td>
<td>Clinical samples Ohio, Turkey, 2009</td>
</tr>
<tr>
<td>10 ZR35</td>
<td>Clinical samples Ohio, Turkey, 2010</td>
</tr>
<tr>
<td>11 ZR43</td>
<td>Clinical samples Ohio, Turkey, 2010</td>
</tr>
<tr>
<td>12 ZR46</td>
<td>Clinical samples Ohio, Turkey, 2010</td>
</tr>
<tr>
<td>13 ZR100</td>
<td>Isolated from Arkansas, Turkey, 2011</td>
</tr>
<tr>
<td>14 ZR133</td>
<td>Isolated from Arkansas, Chicken, 2011</td>
</tr>
<tr>
<td>15 ZR140</td>
<td>Clinical samples Ohio, Turkey, 2011</td>
</tr>
<tr>
<td>16 ZR152</td>
<td>Clinical samples Arkansas, Chicken, 2012</td>
</tr>
<tr>
<td>17 ZR156</td>
<td>Clinical samples Arkansas, Chicken, 2012</td>
</tr>
<tr>
<td>18 ZR161</td>
<td>Clinical samples Ohio, Turkey, 2012</td>
</tr>
<tr>
<td>19 ZR165</td>
<td>Clinical samples Indiana, Turkey, 2011</td>
</tr>
<tr>
<td>20 ZR170</td>
<td>Clinical samples Arkansas, Chicken, 2012</td>
</tr>
<tr>
<td>21 ZR176-1</td>
<td>Clinical samples Indiana, Turkey 2012</td>
</tr>
<tr>
<td>22 ZR176-2</td>
<td>Clinical samples Indiana, Turkey 2012</td>
</tr>
<tr>
<td>23 ZR178</td>
<td>Clinical samples Ohio, Turkey, 2013</td>
</tr>
<tr>
<td>24 ZR185</td>
<td>Clinical samples Ohio, Turkey 2013</td>
</tr>
</tbody>
</table>

Table 3.1: A list of twenty four *Mycoplasma synoviae* strains, field isolates and positive clinical samples in which CRISPR amplification was attempted.
<table>
<thead>
<tr>
<th>Primer Sequence</th>
<th>Primer name</th>
</tr>
</thead>
<tbody>
<tr>
<td>GTACGATATTTATAGCGTTGATG</td>
<td>CRISPR-UP-STREAM-F1</td>
</tr>
<tr>
<td>TGTCACGACACGATTCA</td>
<td>CRISPR-F2</td>
</tr>
<tr>
<td>TGAACATGCAAAAATCATCTCAAT</td>
<td>CRISPR-UP-STREAM-F2</td>
</tr>
<tr>
<td>CTTCAATAGATAAAATTCAAACC</td>
<td>CRISPR-UP-STREAM-R1</td>
</tr>
<tr>
<td>GGGTTGGAATTTGTATCTATTTGAAG</td>
<td>CRISPR-FINAL-F3</td>
</tr>
<tr>
<td>AATTGGATTATATAGAAAGYGTTGCT</td>
<td>CRISPR-FINAL-F1</td>
</tr>
<tr>
<td>GCGTTGCTTTTACAAATGAAAG</td>
<td>CRISPR-R1</td>
</tr>
<tr>
<td>AATTCGCGTTTGGGGTTGTA</td>
<td>CRISPR-FINAL-F2</td>
</tr>
<tr>
<td>CGTTAACAATAATTGTACAACCCCA</td>
<td>CRISPR-UP-STREAM-R2</td>
</tr>
<tr>
<td>TGGGGTTGTACAATTTTTGTAAAGT</td>
<td>CRISPR-DOWN-STREAM-F2</td>
</tr>
<tr>
<td>AAGTAAACATAGCCCCACAA</td>
<td>CRISPR-R1</td>
</tr>
<tr>
<td>GGTAAAWATAATTGTGGGGCTAA</td>
<td>CRISPR-FINAL-R1</td>
</tr>
<tr>
<td>ACAATTATWTATACTCCCTCCCAT</td>
<td>CRISPR-DOWN-STREAM-F1</td>
</tr>
<tr>
<td>AATTGGATCYCTAATGGATTTACAT</td>
<td>CRISPR-FINAL-R2</td>
</tr>
<tr>
<td>CGCCTATTITTRGAATCTGACA</td>
<td>CRISPR-FINAL-R3</td>
</tr>
<tr>
<td>CATALGCGCTTTATTTRGAA</td>
<td>CRISPR-R3</td>
</tr>
<tr>
<td>GTGCATTATTGCCATCTTACATAAGC</td>
<td>CRISPR-DOWN-STREAM-R2</td>
</tr>
<tr>
<td>GAAATTGTCATAGAAGATGGAA</td>
<td>CRISPR-R2</td>
</tr>
<tr>
<td>TGTGTTATTCCAATGCTACA</td>
<td>CRISPR-DOWN-STREAM-R1</td>
</tr>
</tbody>
</table>

**Table 3.2:** A list of nineteen primers designed and used for CRISPR amplification. Primers are color-coded; forward primers are written in red and reverse primers are written in blue. Numbers to the left of the primer sequences coordinate with figure 3.3.
<table>
<thead>
<tr>
<th>Primer Combination</th>
<th>Size</th>
<th>Temp</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 CRISPR-F1: GCGTTGCTTTTACAAATGAA</td>
<td>961 bp</td>
<td>50 - 55 °C</td>
</tr>
<tr>
<td>CRISPR-R1: AAGTAGACATCGGGCTCACAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 CRISPR-F1: GCGTTGCTTTTACAAATGAA</td>
<td>1,467 bp</td>
<td>50 - 53 °C</td>
</tr>
<tr>
<td>CRISPR-R2: GAAATTTGTGAATAAGGATTGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 CRISPR-F2: TGTTCACACACAAATTTGATCA</td>
<td>1,724 bp</td>
<td>52 °C</td>
</tr>
<tr>
<td>CRISPR-R1: AAGTAGACATCGGGCTCACAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 CRISPR-F2: TGGTTACACACAAATTTGATCA</td>
<td>2,230 bp</td>
<td>50 - 52 °C</td>
</tr>
<tr>
<td>CRISPR-R2: GAAATTTGTGAATAAGGATTGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 CRISPR-F2: TGGTTACACACAAATTTGATCA</td>
<td>2,010 bp</td>
<td>50 - 52 °C</td>
</tr>
<tr>
<td>CRISPR-R3: CATAGCCTTATTTGAGAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 CRISPR-UP-STREAM-F1: GTACGATATTTATAGCGTTGATG</td>
<td>637 bp</td>
<td>52 °C</td>
</tr>
<tr>
<td>CRISPR-UP-STREAM-R1: CTCAATAGATACAAATTCAAACC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 CRISPR-DOWN-STREAM-F1: CACAAATATTTATACGAATTC</td>
<td>875 bp</td>
<td>53 °C</td>
</tr>
<tr>
<td>CRISPR-DOWN-STREAM-R1: TGGGTTATCCAATGCCTACA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 CRISPR-DOWN-STREAM-F2: TGGGTTATCCAATTTTTGTTAAGT</td>
<td>703 bp</td>
<td>54 °C</td>
</tr>
<tr>
<td>CRISPR-DOWN-STREAM-R2: GTGAATATTTGACATCTTAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 CRISPR-Final-F1: AATTITGATTAATATAGAAAYGTTGCT</td>
<td>991 bp</td>
<td>51 - 55 °C</td>
</tr>
<tr>
<td>CRISPR-Final-R1: GGTTAAWATAATTGTGGGGCTAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 CRISPR-Final-F2: AATTITGATTAATATAGAAAYGTTGCT</td>
<td>791 bp</td>
<td>52 - 56 °C</td>
</tr>
<tr>
<td>CRISPR-Final-R1: GGTTAAWATAATTGTGGGGCTAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 CRISPR-Final-F3: GGGTTTGAATTTGTATCTAT</td>
<td>1,394 bp</td>
<td>53 °C</td>
</tr>
<tr>
<td>CRISPR-Final-R2: AAATTTGTGAATAAGGATTGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 CRISPR-Final-F3: GGGTTTGAATTTGTATCTAT</td>
<td>1,732 bp</td>
<td>53 °C</td>
</tr>
<tr>
<td>CRISPR-R1: AAGTAGACATCGGGCTCACAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 CRISPR-Final-F3: GGGTTTGAATTTGTATCTAT</td>
<td>1,226 bp</td>
<td>53 °C</td>
</tr>
<tr>
<td>CRISPR-R2: GAATTTGTGAATAAGGATTGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 CRISPR-Final-F3: GGGTTTGAATTTGTATCTAT</td>
<td>1,530 bp</td>
<td>53 °C</td>
</tr>
<tr>
<td>CRISPR-R3: CATAGCCTTATTTGAGAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 CRISPR-Final-F3: GGGTTTGAATTTGTATCTAT</td>
<td>1,512 bp</td>
<td>53 °C</td>
</tr>
<tr>
<td>CRISPR-DOWN-STREAM-R2: GTGCAATATTTGCACTTACATAAGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 CRISPR-Final-F3: GGGTTTGAATTTGTATCTAT</td>
<td>1,506 bp</td>
<td>53 °C</td>
</tr>
<tr>
<td>CRISPR-Final-R3: CGCCTTATTTTGRGAATCGTACAA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 3.3:** Seventeen different primer set combinations used for CRISP Amplification, with their amplicon expected size according to MS53 genome. Also provided the annealing temperatures tried with each combination. The forward primers are written in red and reverse primers in blue.
**Figure 3.1**: Schematic representation of CRISPR/cas system. Figure adapted and slightly modified from Al-Attar et al. 2011. (1)
**Figure 3.2:** CRISPR sequence and putative cas genes identified in MS53 whole genome. (A) details of CRISPR loci identified in the only available whole genome sequence MS 53 (43). (B) a schematic representation of the MS 53 genome indicating the position of the identified CRISPR loci. (C) Putative cas genes identified in the same genome.

CRISPRdb; [http://crispr.u-psud.fr/crispr/](http://crispr.u-psud.fr/crispr/) (9).
Figure 3.3: MS53 CRISPR sequenc showing the position of primers used for
Figure 3.3: (Continued)

amplification attempts. The sequence shows CRISPR array with two genes upstream colored differently (Green and yellow) and two genes down stream colored differently (yellow and Green). CRISPR sequence is also color coded (repeats in grey and spacers in multiple colors). Sequence showing the positions of the 19 primers designed and used for this study in relation to the CRISPR sequence. Numbers on top of the primers coordinate with table 3.2.
Figure 3.4: Agarose gel picture showing one of the original attempts to amplify MS CRISPR array. This amplification attempt used CRISPR-F1 and CRISPR-R2 primers with annealing temp of 50°C and 50 amplification cycles. Bands were generated with all three tested samples, but sequencing was successful for only F10 MS strain and ZR15 clinical sample. ZR146 clinical sample failed to sequence using the same amplification primers.
Figure 3.5: Agarose gel picture of the Upstream and Downstream amplification attempt. Amplification results show the relatively successful amplification of upstream and downstream region of some of the tested MS samples. Amplification was performed using CRISPR-UPSTREAM-F1 and CRISPR-UPSTREAM-R1 at 52°C and 45 cycles of amplification of the upstream reaction. Amplification was performed using CRISPR-DOWNSTREAM-F1 and CRISPR-DOWNSTREAM-R1 at 53°C and 45 cycles of amplification of the upstream reaction. Note that still one of the tested samples didn’t successfully amplify for the downstream segment.
Figure 3.6: Agarose gel picture showing the amplification results of the CRISPR final amplification. Multiple samples of MS Showing successful amplification using CRISPR-FINAL-F3 and CRISPR-DOWNSTREAM-R2 primer at 53°C and 50 amplification cycle. But also some of the tested sample didn’t amplify any bands. Out of amplified bands we could only successfully sequence WVU1853.
Figure 3.7: Sequence alignment of the 3 generated MS CRISPR sequences against the MS53 CRISPR sequence obtained from the GenBank. Numbers in black indicate the 12 palindromic repeats of the MS53 CRISPR sequence. ZR15 has 3 additional CRISPR units (blue vertical arrows). ZR133 has one additional CRISPR unit (red vertical arrow). WVU1853 seem to have two CRISPR units less than MS53, 4 of the units didn’t align in the proper position and were missing (horizontal green arrows), but two of them aligned downstream (vertical green arrows). However, the WVU1853 sequence we obtained is truncated and there could potentially be additional CRISPR units that our sequence didn’t reach far enough to include.
Chapter 4: Multilocus Sequence Typing of *Mycoplasma synoviae*

Mohamed El-Gazzar, Kristina McDonald, and Richard D. Slemons

This chapter has been prepared in a manuscript format to be submitted for publication in a peer reviewed journal.
4.1 Abstract

*Mycoplasma synoviae* (MS) is an important mycoplasma pathogen of poultry that could require more attention in the coming years. A maximum of 420 bp segment of the variable Lipoprotein Hemagglutinin A (*vlhA*) gene is currently the only target that could be used for MS sequence typing. There is an apparent limitation, if the *vlhA* target sequences of two MS strains are identical, clonality is inferred; however, if their sequences are different at this locus, there appear to be no clonality and the degree of relatedness between them is unknown. Some of the MS strains that are different in their partial *vlhA* sequence are temporally and spacially closely related. In this study we propose the Multi-Locus Sequence typing (MLST) assay as an additional tool for MS sequence typing. After initial screening of 24 housekeeping genes as potential targets, 7 genes were selected for the MLST assay. An internal segment between 450 bp and 710 pb from each of the 7 genes was successfully amplified and sequenced for 58 different MS strains and field isolates (N=30) or positive clinical samples (N=28). The collective sequence of all 7 gene segments was used for MS sequence typing. Using MLST, the 58 tested MS samples were organized into 30 different sequence types. Coincidently the same group of samples organized also into 30 sequence types when typed using the *vlhA* assay. Multiple positive clinical samples obtained from MS outbreaks in the same geographical vicinity differed in *vlhA* sequence. The same samples were identical on all 7 MLST gene sequences, showing that they belong to the same sequence type. Other samples were identical on *vlhA* but different on MLST. The phylogenetic tree generated by MLST was more congruent to the epidemiological information than the tree generated
by \textit{vlhA} assay. MLST identifies more coherent epidemiological relationships between outbreaks. We suggest that MLST and \textit{vlhA} could be used in tandem. In our view MLST represents a valuable tool for MS sequence typing, providing better understanding of the source of infection and the epidemiology of the disease.

4.2 Introduction

The sciences of preventive medicine and epidemiology are concerned with understanding the determinants, distribution and deterrents of diseases. Understanding the source and the distribution of the disease are essential to developing effective prevention, control, and eradication strategies for mycoplasmosis in poultry. \textit{Mycoplasma gallisepticum} (MG) and Mycoplasma synoviae (MS) are significant pathogens of poultry and they can cause clinical disease in both chickens and turkeys. Mycoplasma can be transmitted vertically from breeders to progeny and horizontally between individuals within the same flock or between flocks (11). While the clinical disease doesn’t cause high mortality it can produce high carcass condemnation rates in the processing plant due to synovitis, tenosynovitis and/or airsacculitis (11). Mycoplasma can also cause loss of egg production which minimizes the profit margin of egg layers flocks.

Due to these production losses, the poultry industry is quite keen on early diagnosis and control of mycoplasma outbreaks and eradicating mycoplasma in breeder flocks. Antibiotic treatments might mitigate the impact of the infection on performance, however, if the source of the infection is vertical transmission from breeder flocks, the continuous nagging production losses leave the producers with very little choice but to
eliminate the breeder flock from the system, even if the breeder flock is young and early in the production cycle. In some cases the breeder flocks are eliminated due to mycoplasma infections even before the onset of egg production (10). The outbreak investigation doesn’t end by controlling the disease, minimizing the losses or eradicating the infectious agent from the breeding population. Identifying the source of the agent and the route of the agent spread are integral parts of preventing future reintroductions.

Classical diagnostics for avian mycoplasma including serological testing, isolation and fluorescent antibody identification represent very useful tools for detecting the infection. However, they have limitations when it comes to differentiating between strains and identifying the source of the infectious agent. Molecular assays are the more suitable diagnostic tools for differentiating between strains within the same bacterial species and identifying the source of the infectious agent and describing the epidemiology of the disease. For avian mycoplasma, genotyping by DNA fingerprinting and/or sequencing are two main ways to precisely differentiate between strains. Genotyping by sequencing is a preferred method for mycoplasma because it can be performed on clinical samples directly without the need for isolating the microorganism in a pure colony which can be difficult to achieve. Sequence typing for MG is much more advanced than for MS. There are at least 5 genomic targets commonly used for MG sequence typing with approximately a total of 2800 base pair of sequence information that can be used for genotyping (4, 25). For MS there is only one identified genome target used for genotyping purposes. This target is the partial sequence of the \(\text{vlhA}\) gene with a maximum genotyping potential of 420 bp (2, 5, 6, 31).
*vlhA* sequence typing assay is a high discriminatory power assay that can easily differentiate between unrelated strains. However, in many cases it is difficult to establish how closely related the strains are. Many spacially and temporally closely related strains are different the partial sequence of the *vlhA* gene. Due to the variable nature of the gene, clonality and evolutionary relatedness of such strains is difficult to identify. One solution to this problem could be the use of Multilocus Sequence Typing (MLST) as a sequence typing assay which relies on the cumulative discriminatory power of multiple housekeeping genes (12).

The purpose of this study is to develop an MLST assay and examine its potential use for MS sequence typing. Twenty four different housekeeping genes were studied and 7 final internal gene segments, ranging from 450 bp to 710 pb, were selected for the MLST assay. Fifty eight MS strains, field isolates and MS positive clinical samples were tested by the newly developed MLST. All tested samples were successfully amplified for the selected 7 gene segments. All amplified sequences were sequenced and database was established. A total of 30 sequence types were generated with the MLST and also 30 sequence types generated by the *vlhA* assay.

While MLST retained similar discriminatory power to the *vlhA* assay, the sequence types generated by MLST were more congruent with the epidemiologic information available about the tested samples compared to *vlhA* sequence types. We propose the multilocus sequence typing assay as an additional tool for MS genotyping. The use of both MLST and *vlhA* in tandem could provide a more comprehensive
understanding of MS outbreaks in the poultry industry, not only on the local, but also on the national and international level of investigation.

4.3 Materials and Methods

4.3.1 *Mycoplasma synoviae* strains and field isolates: A total of 30 MS strains and field isolates were tested by MLST. The list included laboratory adapted strains, the vaccine strain, and field isolates from 11 states in the USA, plus isolates from Canada, Australia, Israel and Argentina. In addition to that the sequence for the MS53 strain from Brazil which was obtained from the GenBank from the only available MS whole genome sequence (30). The oldest isolate on the list was collected in 1955 and the most recent in 2003, spanning a time period of 48 years. Most of the MS strains and field isolates were provided by Dr. Noel-Ferguson from the mycoplasma lab in the Poultry Diagnostic and Research Center, at the University of Georgia (table 4.1). The group of field isolates and strains is a subset that was selected from the same list of field isolates used to evaluate the original *vlhA* assay in 2004 (6) to maximize the diversity of the populated tested in this study. Four strains namely (WVU 1853, MSH, F102AS and K1968 were tested twice from two different sources. One source is the depository of the University of Georgia, the other is the depository in our lab in Ohio State University. The field isolates in the depository in our lab was originally obtained for UGA a few years ago and it was passed in culture media separately since then.
4.3.2 *Mycoplasma synoviae* clinical samples: A total of 28 MS PCR positive clinical samples were tested by MLST. MS positive samples are tracheal or choanal cleft or joint swabs from clinical or subclinical MS cases of commercial turkey or chicken flocks that were confirmed positive by real-time PCR (24). The group of samples tested in this study were from submissions to our lab over a period of 3 years (from 2010 to 2013), mainly from Ohio but also from Arkansas, Indiana, Pennsylvania and Missouri, in addition to samples submitted from the Netherlands (table 4.1). The clinical cases selected for this study to represent cases that were identical on vlhA sequence but have no apparent epidemiological links or cases that were different on vlhA but have apparent epidemiological link.

4.3.3 Genomic DNA extraction: For clinical samples, all swabs were pooled into sample pools (5 swabs per pool) of 1 ml of phosphate-buffered saline (PBS). For MS field isolates, a total volume of 3 ml to 5 ml of modified Frey’s broth culture was centrifuged at 1500 rpm for 30 minute. The pellets were resuspended in 200 ml of modified Frey’s broth. Genomic DNA was extracted from 200 ml of PBS resuspension solution or 200 ml of resuspended culture using the QIAamp DNA Mini Kit (QIAGEN, Valencia, CA) following the manufacturer’s recommendations.

4.3.4 Selection of the housekeeping genes: A total of 24 housekeeping genes were studied as potential targets for MLST assay (table 2). The list of studied genes was generated by combining the studded genes from the only MLST assays developed for
mycoplasma species (13, 14). In addition to that, several genes were added to the list by selecting them from the MS whole genome sequence of MS53 (30) (GenBank accession AE017245). Not all genes were attempted to be amplified, some of the listed genes were too large to be considered (table 4.2). For each attempted gene a set of forward and reverse primers were designed, and initial amplification of the gene was attempted on 4 MS strains. If the gene was not successfully amplified out of any of the test strains, the amplification attempt was repeated.

Two criteria were used to decide if the studied gene will be kept on or dropped off the list. First criterion was the successful amplification of the gene in all tested strains. Second criterion was the level of variation in the studied gene (table 4.2). The final 7 genes were selected to represent the most variable genes that were successfully amplified in all tested strains. Another standard that was taken in consideration was the distribution of the genes along the whole genome sequence of MS53 (30) (GenBank accession AE017245) (figure 4.1) to capture all potential recombination.

Originally all primers were place outside the targeted genes. As we expanded the group of tested MS strains and field isolates, amplification was not successful on all genes. We hypothesized that the reason for amplification failure was due to recombination that would lead to have different sequences around the genes from the single published whole genome sequence of MS53. Primers sets were moved inside the gene sequence for 11 different genes. The position of the primers inside the gene was dictated by the nucleotide sequences. But we attempted to put the primers as close as
possible to the start and stop codon of the gene, to be able to amplify the largest possible internal gene fragment.

At this point several genes had been dropped due to not enough variations or proximity to other selected genes on the MS53 whole genome. Still out of the 11 genes with primer inside, 3 didn’t amplify consistently in all strains. Additional new primer sets were designed for two of the three genes that didn’t amplify consistently. A list of 8 pre-final genes was generated. The pre-final 8 genes continued to be tested against more MS strains and field isolates. Eventually one gene was dropped due to not having enough variation, which lead to the list of the final 7 selected genes (table 4.2 and figure 4.2).

4.3.5 PCR design and amplification: A total of 66 different primers have been designed comprising different 33 primer sets (data not shown) during the process of studying and selected the final MLST genes. The primers of the final 7 genes are listed in table 4.3. The amplified segments range from 522bp to 847 bp. Due to Sanger sequencing limitations, not the whole amplified segment was used for the MLST. A 50 to 100 bp segment was usually lost during the sequencing process. The size of each amplicon and the size of the actual gene segment used for the MLST are summarized in table 4.3.

PCRs in this study were carried out using the Roche FastStart High Fidelity kit (Roche Diagnostics, Indianapolis, IN) according to manufacturer’s recommendations with some modifications. Typically for conventional PCR amplification in our lab we used a total volume of 50 ml of PCR reaction mix and DNA template solution. But for the purpose of this study we modified the volume of the PCR reaction to be 25 ml. The
reduced volume allowed us to use less DNA template in each reaction. This was done due to the fact that each of the MS strains had to be tested against at least 7 genes and some of the strains were tested against 18 different genes, while the amount of DNA extracts are ultimately limited. The PCR was performed using the MJ-Mini thermocycler (BioRad Laboratories, Hercules, CA) in a total volume of 25 ml reaction mix containing 0.5 ml of 10 mM deoxynucleotides, 2.5 ml of 10X FastStart High Fidelity reaction buffer (1.8 mM MgCl2), 2 ml of each 5 mM primers, 0.25 ml of 5 U/ml FastStart High Fidelity enzyme, and 2.0 ml of DNA template.

To compensate for the limited amount of DNA template used the number of cycles was increased from 40 to 45 cycles. All reactions were performed using a thermocycler program of a hot start at 95 C° for 3 min, 45 cycles of denaturing at 94 C° for 30 sec, annealing at 54 C° for 30 sec, extension at 72 C° for 90 sec, and a final extension at 72 C° for 5 minutes. Electrophoreses was performed on PCR products on a 1% agarose gel with 0.53 Tris-borate-ethylenediaminetetraacetic acid buffer and 0.5 mg/ml ethidium bromide, at 90V for 45 minutes. The PCR products were visualized to confirm the success of amplification by ultraviolet trans-illumination (figure 4.2).

4.3.6 Sequencing. Purification of all PCR products was carried out using the QIAquickHPCR Purification Kit (350) (QIAGEN) following manufacturer’s recommendations. Sequencing the purified PCR products was performed using the Dye termination sequencing method in both directions. The sequencing was performed in (Genewize Inc., South Plainfield, NJ).
4.3.7 **Sequence analysis of PCR products:** Sequence data were analyzed and aligned for overlapping of complementary sequences, and consensus was constructed with the SeqMan program (in Lasergene; DNASTAR, Inc., Madison, WI). Alignments of sequences were constructed by the Clustal V method using the MegAlign program (in Lasergene; DNASTAR, Inc.). Sequences for each of the tested genes were compiled in a separate MegAlign data base file to allow for the assessment of individual genes potential.

4.3.8 **Allelic profiles and sequence types:** Allelic numbers were assigned arbitrary for each new sequence that was added to the data base of each gene out of the selected 7 genes. And for each MS strain an allelic profile was generated according to the numbers that reflect their 7 genes sequence (table 4). Hence the allelic profile for each strain consisted of 7 numbers that corresponded to the 7 allelic numbers of its 7 genes. For example the MS53 was used as the reference sequence in all the genes since it is the only available whole genome sequence available on GenBank that was used for the development of this assay. And its allelic profile was expressed as (1,1,1,1,1,1,1). While the allelic profile for MS-H the vaccine strain was (2,2,2,2,1,2). For each new allelic profile an arbitrary Sequence Type (ST) number was assigned.

4.3.9 **MLST data analysis:** Analysis of MLST data was performed using Sequence Type Analysis and Recombinational Tests, version 2 (START2) software (8).
The original step of the analysis is loading all the different allelic sequences for each of the selected gene segments. The next step is to load all the created allelic profiles and their assigned sequence types. The final data loading step is to enter the name of each sample and its MLST sequence type. The START2 software will use the sequence types, allelic profiles and allelic sequences to perform the analysis. The analysis included polymorphism, phylogeny, testing for recombination and testing for selection.

4.3.10 Comparison of MLST and vhlA genotyping: vhlA is the single gene that is currently used for MS genotyping by sequencing. Partial vhlA sequences were generated for all tested 58 strains and clinical samples using primers from Wetzel et al 2010. In order to facilitate the comparison between MLST and vhlA sequence types allelic numbers were assigned arbitrary for each new sequence added to the vhlA data base. Since it’s a single gene locus, vhlA allelic numbers became sequence types numbers. Sequence types generated by MLST and vhlA assays were compared as a mean of validation of the newly developed assay. Levels of variation and other parameters were also compared between the two assays.

4.4 Results

4.4.1 Gene selection: Seven final gene segments were selected for the MLST assay (figure 4.1). These gene segments ranged from 450 bp to 710 bp in size and named
after the genes they originated form as follow: adk, atpG, efp, gmk, nagC, ppa and recA (table 4.3). The distribution of the genes along the only published genome was taken in consideration. The 7 selected genes are represented in all 4 quarters of the genome. When we placed the primers outside the genes according to the structure of the MS53 genome, some of the genes failed to amplify. But when we moved the primers inside the genes all the genes were successfully amplified for all the strains. This suggests that the space outside the genes are not similar in all strains. This is why the distribution of the selected genes along the genome is important, it allows for capturing any variability due to potential recombination on the genome level.

For each of the four strains repeated from two different sources, all genes (MLST and vhlA) were found to be identical from the two sources except for WVU1853. It appears that the WVU1853 strain was incorrectly labeled in our lab, as all genes were identical to the isolate labeled K1968 (table 4.1 and figures 4.11).

4.4.2 Amplification and sequencing: A total of 836 segment amplification attempts and 1,230 single strand sequencing were performed for the purpose of developing this assay. Fifty eight tested MS strains and clinical samples were successfully amplified and sequenced for the selected 7 genes. In total the 58 MS samples generated 406 DNA segments and 812 single strand sequences.

4.4.3 Allelic profiles and sequence types: sequence analysis of 58 tested strains and clinical samples resulted in 14 different alleles for adk (figure 4.3), 17 alleles for
137 alleles for \textit{gmk} (figure 4.5), 13 alleles for \textit{efp} (figure 4.6), 18 alleles for \textit{nagC} (figure 4.7), 15 alleles for \textit{ppa} (figure 4.8) and 15 alleles for \textit{recA} (figure 4.9) (table 4). The allelic profiles of all 7 genes resulted in 30 sequence types referred to as ST-1 to ST-30. The exclusion of \textit{atpG} and \textit{recA} did not change the sequence types for any of the tested 58 strains. However, we observed significant variation (figure 4.11) within these two genes, and we expect that with the expansion of their data bases these two genes will contribute to the discriminatory power of the assay.

\textbf{4.4.4 MLST data analysis:} The number of polymorphic positions per gene was calculated (table 4.4 and figure 4.11) using START2 (Ref). \textit{ppa} was found to contain the highest percentage of polymorphic positions (3.75%), while \textit{efp} was found to have the lowest percentage of polymorphic positions (2.22%). Out of these polymorphic sites \textit{ppa} was found to have the highest percentage of the non synonymous mutations (2.65%), followed by \textit{nagC} (2.26%) and \textit{recA} (2.11%), while the lowest non synonymous mutation percentage was in \textit{atpG} (1.35%). When the dN/dS ratios were calculated by START2 using Nei and Gojobori method of estimating synonymous substitution (18) for all the selected genes, they were all found to be under purifying selective pressure (9). Additionally START2 also calculated the Index of Association ($I_A$) according to Maynard Smith method (28). And for the 7 tested genes the $I_A$ was found to be 2.7916 which was significantly different from zero indicating linkage disequilibrium. This means that pattern of alleles distribution in the population is suggesting clonal structure of the population with no or minimal recombination. The $I_A$ was calculated for two
subpopulations of isolates that are temporally and geographically close. The samples from Arkansas 2000 – 2013 were considered one subpopulation and the samples from Ohio-Pennsylvania 1993 – 2013 as another subpopulation. For Arkansas (14 clinical samples or field isolates) (figure 4.13) and Ohio-Pennsylvania (10 clinical samples or field isolates) (figure 4.14) the $I_A$ was found to be 4.4194 and 3.7817 respectively. In both cases the $I_A$ was significantly different from zero indicating clonal structure of the population at the local level. This is suggesting that the lack of recombination suggested was not due to geographical isolation and it reflects a true clonal manner of replication at all levels with no or minimal role of recombination in gene exchange (28).

The Phylogenetic tree for the 58 tested strains, field isolates and clinical samples was constructed by START2 software using the neighbor-joining method (figure 4.12). Based on previous results of dN/dS ratios and the calculated $I_A$ values we elected to construct the tree based on sequences rather than allelic profiles. A high degree of correlation between clustering and the epidemiological information was observed. The international isolates clustered separately from the USA isolates with the exception of the Canadian isolates. The Canadian isolates, however, clustered with the US isolates particularly close to isolates from the northern Border States like Main, Illinois and Ohio. Other groups clustered broadly according to geographical location; North Carolina isolates close to West Virginia isolates, Ohio close to Indiana and Pennsylvania, also Colorado and California. Arkansas isolates seems to be the most diverse population and the most wide spread over all parts of the generated tree.
4.4.5 Comparing MLST to \textit{vlhA} genotyping: The number of polymorphic sites in \textit{vlhA} segment was 158 sites (35.83%), which was much higher compared to MLST variation level (figure 4.11). The higher levels of variation were mainly due to deletions and insertions in two proline rich hyper variable regions within the single copy of the \textit{vlhA} gene (figure 4.10) (1). The level of variation observed in the \textit{vlhA} gene segment resulted in a higher discriminatory power and generated the same number of sequence types generated by 7 genes of the MLST assay. The 58 MS tested strains, field isolates and clinical samples were organized in 30 \textit{vlhA} sequence types (table 4.4 and figures 4.10 and 4.11).

There was a general agreement between \textit{vlhA} and MLST sequence types but there were differences. Some epidemiologic relations among the 58 tested strains were better explained by MLST (figure 4.12). The relations among some of the clinical samples with close epidemic ties were better explained by MLST (figures 4.12 and 4.15). Clinical samples (ZR176-2, ZR250, ZR251-1 and ZR251-3) are 4 samples from the same production company submitted in 3 different dates and all 4 were different on \textit{vlhA} and 100% identical on MLST (table 4.1). Also Four MS positive DNA extracts (ZR34-18 to ZR34-21) submitted from the same production complex from broiler breeders and broiler farms from Arkansas. One of the samples (ZR34-18) was different from the rest of the samples on \textit{vlhA}. However, they were all identical on MLST. Interestingly all samples from the two groups from Indiana and Arkansas are identical on MLST, and some of them are identical of \textit{vlhA}. This is suggesting that they belong to the same MS strains but there are no apparent epidemiological links between the two cases. Similarly, 3 DNA
extracts (ZR74-505, ZR74-520 and ZR74-525) submitted from the 3 different farms with close geographical proximity in the Netherlands, these clinical samples were identical on MLST but on vlhA ZR74-505 was different from the other samples. Three isolates from Colorado (K1968, K4862I and K4862 D) were identical on MLST but different on vlhA. These were all groups of samples where samples within group were identical on MLST but not on vlhA. And epidemiologically, samples within each group have close epidemiological ties and it is more likely for them to have been originated from the same MS strain which is better exhibited by MLST than vlhA.

On the other hand there were cases that were identical on vlhA but different on MLST. K5542A isolated from layer chicken from Georgia in 2003 was a 100% match on vlhA to 3 clinical samples (ZR71, ZR141 and ZR142) from meat turkey in 2011, while it was different in only one gene (ppa) on MLST. Also 2 chicken isolates from California from 2001 were identical on vlhA to a California turkey isolate from 2001, but there were different from the turkey isolate on 5 different genes (atpG, gmk, nagC, ppa and recA) on MLST. Additionally, 2 isolates from Arkansas and one clinical sample (ZR100, ZR133 and ZR152) were all identical on vlhA but different on MLST. Epidemiologic relations between these cases could be also better explained by MLST.

Unlike MLST phylogenetic tree the international isolates did not cluster away from the isolates and cases from the USA. Within the USA clustering of geographically related isolates and samples was less apparent on the vlhA compared to MLST phylogenetic trees.
4.5 Discussion:

Until now a \textit{vlhA} partial gene sequence is the only target that can be used for MS sequence typing, \cite{5, 6, 31} and it is established that the current assay is harnessing the maximum potential of the gene as a sequence typing target \cite{2}. Identical \textit{vlhA} sequence types were thought to infer clonality between MS strains. Unlike MG which has multiple genomic targets for sequence typing \cite{4, 25}, we didn’t have any other sequence typing method to verify this assumption for MS sequence types. Also if two MS sequence types were not identical on \textit{vlhA} it was not possible to infer any relation between strains.

In this study we developed and examined an MLST assay specific for \textit{Mycoplasma synoviae} (MS). A total of 58 MS strains, field isolates and clinical samples were tested using the MLST assay and the resulted sequence types were compared to \textit{vlhA}. All tested strains and clinical samples were successfully amplified for all 7 targets on MLST. This means that the developed MLST assay can be applied on clinical samples. This is significant as MLST doesn’t require the isolation of MS when it comes to the use of the assay in investigating outbreaks.

Some epidemiologic relations among the 58 tested strains were better explained by MLST. Also the topology of the MLST phylogenetic tree (figure 4.12) was more congruent with the epidemiological data when compared to \textit{vlhA} phylogenetic tree (figure 4.15). This could be attributed to the nature of the genes used in the assays. The \textit{vlhA} assay uses a single gene which is part of a variable gene encoding for surface proteins
while MLST uses 7 housekeeping genes which are wide spread along the genome and more stable than *vlhA*. In addition to that cumulative size of the 7 genes (3,960 bp) is more than 10 times the size of the *vlhA* gene segment used for MS sequence typing. In spite the differences in the sequence types between the two assays they both generated the same number of sequence types (30 sequence types) out of the 58 strains and clinical samples used in this study. This makes the two assays equal in their discriminatory powers. But we predict that with the expansion of the data sets *vlhA* might have higher discriminatory power due to its variable nature.

Traditionally multiple phenotypic characteristics were used in classification of bacterial strains. Techniques including but not limited to serotyping, phage typing and carbohydrate profiling were used for such purpose. With the advancement of molecular technique the characterization of isolates became more sophisticated. Multilocus Enzyme Electrophoresis (MLEE) uses the electrophoretic mobility of relatively conserved metabolic enzymes. This technique was first applied in bacteria in 1973 (15) and it became commonly used in studying prokaryotic populations during the 1980’s (26). Herbert John Webber was first to use the expression “clone” to define “a population in which all members have been derived from one and the same progenitor by nonsexual multiplication” (22). Originally, MLEE studies showed evidence of strong clonality in bacteria (16, 17, 23, 27). Later, other bacterial populations showed less clonality suggesting that recombination might play a larger role than previously thought (7, 21). Maiden *et. al* in 1998 development MLST based on the concept of MLEE but applied directly to the DNA sequence (12) as a powerful tool for studying the bacterial
population and epidemiology. The availability of MLST data for a large number of species allowed for a better studying of the role of recombination in population. However, this role is still undefined and strongly debated (3, 29). By calculating the $I_A$ (28) using the data set generated in this study MS appears to be relatively clonal with a significant linkage disequilibrium suggesting a minimal role for recombination in genetic distribution in the population. This favors MLST as a tool of studying the structure of MS population and generating reliable epidemiologic relations between strains and outbreaks. Unlike surface antigens which are typically under selective positive pressure, we found that all MLST genes after calculating $dN/dS$ ratio to be under purifying negative pressure.

In our opinion the best way to use MLST and $vlhA$ sequence typing assays in studying the epidemiology of MS outbreaks is to use both assays in tandem. MLST should be used as the first step of sequence typing, anchoring the stain or clinical sample to its position in the population structure. Then followed by $vlhA$ with the predicted higher discriminatory power as a higher resolution step to differentiate the closely related outbreaks. If two isolates or clinical samples are similar on MLST and different on $vlhA$ this might suggest the same origin strain have evolved separately in two different geographical or temporal incidences. It might also suggest it is the same strain with different entry points. If the two sequence types are identical on both MLST and $vlhA$ that could be considered as a stronger inference of clonality than $vlhA$ alone.

In this study we propose MLST as an additional molecular epidemiology tool to be used in sequence typing of MS and the investigation of outbreaks. This tool can be applied on strains or isolates or even clinical sample. Its discriminatory power is very
comparable to \textit{vlhA} partial gene sequence assay according to the current data set. However, MLST assay utilizes 10 times as much nucleotide sequence as does \textit{vlhA} assay in 7 different loci wide spread along the whole MS genome, and it is more likely to provide better epidemiologic relations between outbreaks. We also suggest that MLST could be used in tandem with \textit{vlhA} assay.

\textbf{4.6 Acknowledgement}

We would like to express our thanks to Dr. Naola Ferguson-Noel, Mrs. Ruth Wooten and all the member of the Mycoplasma lab in the Poultry Diagnostic and Research Center at the University of Georgia for providing most of the \textit{Mycoplasma synoviae} strains and field isolates used in this study. We would like to thank them also for their welcoming cooperation and their kind generosity.

\textbf{4.7 References}


Table 4.1: List of all 58 MS strains, field isolates and clinical samples tested for MLST. *vlhA and MLST sequence types are listed for each entry. *Four repeated isolates from two sources and notice that WVU# was mislabeled in our lab.

<table>
<thead>
<tr>
<th>Tested Sample</th>
<th>Information</th>
<th>MLST ST</th>
<th>vlhA ST</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 MS33</td>
<td>Isolated from Brazil, Chicken, Unknown date</td>
<td>1</td>
<td>2</td>
<td>Not tested, reference sequence, available whole genome sequence on GenBank.</td>
</tr>
<tr>
<td>2 MS-H GA*</td>
<td>Isolated from Australia, Chicken, 1986</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>3 MS-H</td>
<td>Isolated from Australia, Chicken, 1986</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4 WVU GA*</td>
<td>Isolated from West Virginia, Chicken, 1955</td>
<td>3</td>
<td>1</td>
<td>For 4 strains (MS-H, WVU, K1968 and F10), the same strain was tested twice from two different sources. The isolate with GA next to its name is collected from the University of Georgia in 2013. Other isolates from the depository in our lab, originally from the University of Georgia but has been passed separately in culture media.</td>
</tr>
<tr>
<td>5 K1968 GA*</td>
<td>Isolated from Colorado, Turkey, 1983</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>6 K1968</td>
<td>Isolated from Colorado, Turkey, 1983</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>7 K1968</td>
<td>Isolated from Colorado, Turkey, 1983</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>8 F10 GA*</td>
<td>Isolated from North Carolina, Chicken, 1970</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>9 F10</td>
<td>Isolated from North Carolina, Chicken, 1970</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>10 FMT</td>
<td>Isolated from Mississippi, Chicken, 1979</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>11 K870</td>
<td>Isolated from Maine, Chicken, 1976</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>12 K3344</td>
<td>Isolated from Ohio, Chicken, 1992</td>
<td>7</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>13 K5091</td>
<td>Isolated from Argentina, Unknown species or date</td>
<td>8</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>14 K5230</td>
<td>Isolated from Israel, Unknown species or date</td>
<td>9</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>15 K4643</td>
<td>Isolated from Minnesota, Turkey, 1997</td>
<td>10</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>16 K4862I</td>
<td>Isolated from California, Turkey, Unknown date</td>
<td>4</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>17 K4862D</td>
<td>Isolated from Colorado, Turkey, 1999</td>
<td>4</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>18 K5079</td>
<td>Isolated from California, Chicken, 2001</td>
<td>11</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>19 K5095</td>
<td>Isolated from California, Chicken, 2001</td>
<td>11</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>20 K5416</td>
<td>Isolated from Georgia, Chicken, 2002</td>
<td>12</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>21 K5420</td>
<td>Isolated from California, Turkey, 2001</td>
<td>13</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>22 K5581</td>
<td>Isolated from Illinois, Chicken, 1988</td>
<td>6</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>23 K4822B</td>
<td>Isolated from Canada, Turkey, 1999</td>
<td>14</td>
<td>7</td>
<td>Four MS positive DNA extracts (ZR34-18 to ZR34-21) submitted from broiler breeders and broiler farms from Arkansas. Different on vlhA and identical on MLST. They were also similar to 4 submissions from Indiana, with no apparent epidemiological links.</td>
</tr>
<tr>
<td>24 K4822D</td>
<td>Isolated from Canada, Turkey, 1999</td>
<td>14</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>25 K5901</td>
<td>Isolated from Arkansas, Chicken, 2000</td>
<td>10</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>26 K5173</td>
<td>Isolated from Arkansas, Chicken, 2001</td>
<td>15</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>27 K5190</td>
<td>Isolated from Arkansas, Chicken, 2001</td>
<td>10</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>28 K5542A</td>
<td>Isolated from Georgia, Chicken, 2003</td>
<td>16</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>29 ZR34-18</td>
<td>Clinical samples Arkansas, Chicken, 2010</td>
<td>17</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>30 ZR34-19</td>
<td>Clinical samples Arkansas, Chicken, 2010</td>
<td>17</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>31 ZR34-20</td>
<td>Clinical samples Arkansas, Chicken, 2010</td>
<td>17</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>32 ZR34-21</td>
<td>Clinical samples Arkansas, Chicken, 2010</td>
<td>17</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>33 ZR43</td>
<td>Clinical samples Ohio, Turkey, 2010</td>
<td>18</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>34 ZR55</td>
<td>Clinical samples Ohio, Chicken, 2010</td>
<td>18</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Tested Sample</td>
<td>Information</td>
<td>MLST ST</td>
<td>vlhA ST</td>
<td>Remarks</td>
</tr>
<tr>
<td>---------------</td>
<td>--------------------------------</td>
<td>---------</td>
<td>---------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>35 ZR59</td>
<td>Clinical samples Ohio, Turkey, 2010</td>
<td>19</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>36 ZR60</td>
<td>Clinical samples Ohio, Turkey, 2010</td>
<td>19</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>37 ZR71</td>
<td>Clinical samples Ohio, Turkey, 2011</td>
<td>20</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>38 ZR74-505</td>
<td>Clinical samples Netherlands, Chicken, 2011</td>
<td>21</td>
<td>21</td>
<td>MS positive DNA extracts submitted from broiler breeders and broiler farms from the Netherlands. Different on vlhA and identical on MLST</td>
</tr>
<tr>
<td>39 ZR74-520</td>
<td>Clinical samples Netherlands, Chicken, 2011</td>
<td>21</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>40 ZR74-525</td>
<td>Clinical samples Netherlands, Chicken, 2011</td>
<td>21</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>41 ZR100</td>
<td>Isolated from Arkansas, Turkey, 2011</td>
<td>22</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>42 ZR133</td>
<td>Isolated from Arkansas, Chicken, 2011</td>
<td>23</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>43 ZR141</td>
<td>Clinical samples Ohio, Turkey, 2011</td>
<td>20</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>44 ZR142</td>
<td>Clinical samples Ohio, Turkey, 2011</td>
<td>20</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>45 ZR152</td>
<td>Clinical samples Arkansas, Chicken, 2012</td>
<td>24</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>46 ZR156</td>
<td>Clinical samples Arkansas, Chicken, 2012</td>
<td>25</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>47 ZR170</td>
<td>Clinical samples Arkansas, Chicken, 2012</td>
<td>26</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>48 ZR176-1</td>
<td>Clinical samples Indiana, Turkey 2012</td>
<td>27</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>49 ZR176-2</td>
<td>Clinical samples Indiana, Turkey 2012</td>
<td>17</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>50 ZR197</td>
<td>Clinical samples Arkansas, Turkey 2013</td>
<td>28</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>51 ZR200</td>
<td>Clinical samples Arkansas, Turkey 2013</td>
<td>28</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>52 ZR206</td>
<td>Clinical samples Arkansas, Turkey 2013</td>
<td>28</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>53 ZR207</td>
<td>Clinical samples Arkansas, Turkey 2013</td>
<td>28</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>54 ZR233</td>
<td>Clinical samples Ohio, Turkey 2013</td>
<td>29</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>55 ZR242</td>
<td>Clinical samples Pennsylvania, Turkey 2013</td>
<td>30</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>56 ZR250</td>
<td>Clinical samples Indiana, Turkey 2013</td>
<td>17</td>
<td>30</td>
<td>4 strains (ZR176-2, ZR250, ZR251-1 and ZR251-3) from the same production company submitted in 3 different dates and all 4 were different on vlhA and 100% identical on MLST. Also 100% similar to ZR34 from Arkansas 2010.</td>
</tr>
<tr>
<td>57 ZR251-1</td>
<td>Clinical samples Indiana, Turkey 2013</td>
<td>17</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>58 ZR251-3</td>
<td>Clinical samples Indiana, Turkey 2013</td>
<td>17</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Beginning</td>
<td>End</td>
<td>Size</td>
<td>Direction</td>
</tr>
<tr>
<td>------</td>
<td>-----------</td>
<td>-------</td>
<td>-------</td>
<td>-----------</td>
</tr>
<tr>
<td>dnaA</td>
<td>57</td>
<td>1427</td>
<td>1371</td>
<td></td>
</tr>
<tr>
<td>glpQ</td>
<td>22038</td>
<td>24893</td>
<td>2856</td>
<td>Complement</td>
</tr>
<tr>
<td>glkX</td>
<td>100880</td>
<td>102280</td>
<td>1401</td>
<td>Complement</td>
</tr>
<tr>
<td>glkX</td>
<td>125501</td>
<td>126097</td>
<td>597</td>
<td>Complement</td>
</tr>
<tr>
<td>ppa</td>
<td>154803</td>
<td>155348</td>
<td>546</td>
<td></td>
</tr>
<tr>
<td>tktA</td>
<td>194710</td>
<td>196650</td>
<td>1941</td>
<td>Complement</td>
</tr>
<tr>
<td>nagC</td>
<td>218237</td>
<td>219097</td>
<td>861</td>
<td>Complement</td>
</tr>
<tr>
<td>gapA</td>
<td>294593</td>
<td>295597</td>
<td>1005</td>
<td></td>
</tr>
<tr>
<td>uvrC</td>
<td>295849</td>
<td>297597</td>
<td>1749</td>
<td></td>
</tr>
<tr>
<td>pdhC</td>
<td>299788</td>
<td>300672</td>
<td>885</td>
<td></td>
</tr>
<tr>
<td>metG</td>
<td>336705</td>
<td>338261</td>
<td>1557</td>
<td></td>
</tr>
<tr>
<td>ugpA</td>
<td>374750</td>
<td>376840</td>
<td>2091</td>
<td></td>
</tr>
<tr>
<td>tpx</td>
<td>418238</td>
<td>418729</td>
<td>492</td>
<td>Complement</td>
</tr>
<tr>
<td>atpG</td>
<td>458197</td>
<td>459066</td>
<td>870</td>
<td>Complement</td>
</tr>
<tr>
<td>efp</td>
<td>487039</td>
<td>487599</td>
<td>561</td>
<td>Complement</td>
</tr>
<tr>
<td>pgiB</td>
<td>562313</td>
<td>563605</td>
<td>1293</td>
<td>Complement</td>
</tr>
<tr>
<td>rpoC</td>
<td>563681</td>
<td>568159</td>
<td>4479</td>
<td>Complement</td>
</tr>
<tr>
<td>rpoB</td>
<td>568152</td>
<td>571763</td>
<td>3612</td>
<td>Complement</td>
</tr>
<tr>
<td>tpi</td>
<td>571937</td>
<td>572668</td>
<td>732</td>
<td>Complement</td>
</tr>
<tr>
<td>recU</td>
<td>572774</td>
<td>573262</td>
<td>489</td>
<td></td>
</tr>
<tr>
<td>recA</td>
<td>667630</td>
<td>668619</td>
<td>990</td>
<td>Complement</td>
</tr>
<tr>
<td>adk</td>
<td>682056</td>
<td>682691</td>
<td>636</td>
<td>Complement</td>
</tr>
<tr>
<td>tufA</td>
<td>772379</td>
<td>773563</td>
<td>1185</td>
<td></td>
</tr>
<tr>
<td>gyrB</td>
<td>793173</td>
<td>795125</td>
<td>1953</td>
<td></td>
</tr>
</tbody>
</table>

**Table 4.2:** List of studied 24 genes as potential targets of MLST and their size, position, direction and the reason they were not included in the final 7 genes.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward primer sequence (5' to 3')</th>
<th>Reverse primer sequence (5' to 3')</th>
<th>Amplicon Size</th>
<th>Final Segment Size</th>
<th>Final Segment Start</th>
<th>Final Segment End</th>
</tr>
</thead>
<tbody>
<tr>
<td>adk</td>
<td>GCTTTGGATTAGATTCWGAGCTA&lt;sup&gt;&lt;a style='color:black;'&gt;H&lt;/a&gt;&lt;/sup&gt;</td>
<td>TTCTTATGGGAATGCCAGGT</td>
<td>585</td>
<td>498</td>
<td>73</td>
<td>570</td>
</tr>
<tr>
<td>atpG</td>
<td>GCTACAATTCGTTATTTCTTGAG</td>
<td>ATGCTATGCAGYGTTTCTACTT&lt;sup&gt;&lt;a style='color:black;'&gt;H&lt;/a&gt;&lt;/sup&gt;</td>
<td>784</td>
<td>669</td>
<td>82</td>
<td>750</td>
</tr>
<tr>
<td>efp</td>
<td>CGCATTATTTACGGTTTACGG</td>
<td>CTGGAATTACATTTCAAGAT</td>
<td>522</td>
<td>450</td>
<td>52</td>
<td>501</td>
</tr>
<tr>
<td>gmk</td>
<td>TCAATGTCTAACTCCTTGGAAGA</td>
<td>TTACAGGTCCATCAGGTGT</td>
<td>563</td>
<td>471</td>
<td>49</td>
<td>519</td>
</tr>
<tr>
<td>nagC</td>
<td>CCGATTATTCGGCGTTATT</td>
<td>ATYGGGGCCATTCTATTAAAT&lt;sup&gt;&lt;a style='color:black;'&gt;H&lt;/a&gt;&lt;/sup&gt;</td>
<td>803</td>
<td>708</td>
<td>73</td>
<td>780</td>
</tr>
<tr>
<td>ppa</td>
<td>AGTAATTGAAATTCAAAAAGGTCA</td>
<td>AACTATATTCTCTGGATTTTCTT</td>
<td>530</td>
<td>453</td>
<td>58</td>
<td>510</td>
</tr>
<tr>
<td>recA</td>
<td>CTTTACCTTGCGCTACGTATT</td>
<td>TCGGAAAGAATCTATTAGGTT</td>
<td>847</td>
<td>711</td>
<td>130</td>
<td>840</td>
</tr>
</tbody>
</table>

**Table 4.3:** Final 7 Genes for the MLST assay forward and a reverse primers, amplicon size, final segment used for MLST after sequencing, and the point the final segments starts and ends.

<sup>A</sup> Number of nucleotide from the start codon where the final segment begins and ends

<sup>B</sup> Underlined letters represent International Union of Pure and Applied Chemistry/International Union of Biochemistry code of base substitutions.
<table>
<thead>
<tr>
<th>Sequence Types</th>
<th>adk</th>
<th>atpG</th>
<th>efp</th>
<th>gmk</th>
<th>nagC</th>
<th>ppa</th>
<th>recA</th>
<th>Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1,1,1,1,1,1</td>
</tr>
<tr>
<td>ST-2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2,2,2,2,1,2</td>
</tr>
<tr>
<td>ST-3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3,3,3,3,3,3</td>
</tr>
<tr>
<td>ST-4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4,4,3,2,3,4</td>
</tr>
<tr>
<td>ST-5</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3,3,2,4,3,3</td>
</tr>
<tr>
<td>ST-6</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5,5,5,5,5,5</td>
</tr>
<tr>
<td>ST-7</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>5,5,2,5,6,5</td>
</tr>
<tr>
<td>ST-8</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>5</td>
<td>6</td>
<td>6,6,6,7,5,6</td>
</tr>
<tr>
<td>ST-9</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>6</td>
<td>6,7,4,7,7,5,6</td>
</tr>
<tr>
<td>ST-10</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td>8</td>
<td>8</td>
<td>6</td>
<td>7</td>
<td>7,5,2,8,8,6,7</td>
</tr>
<tr>
<td>ST-11</td>
<td>8</td>
<td>8</td>
<td>5</td>
<td>9</td>
<td>9</td>
<td>4</td>
<td>4</td>
<td>8,8,5,9,9,4,4</td>
</tr>
<tr>
<td>ST-12</td>
<td>9</td>
<td>9</td>
<td>5</td>
<td>10</td>
<td>10</td>
<td>7</td>
<td>8</td>
<td>9,9,5,10,10,7,8</td>
</tr>
<tr>
<td>ST-13</td>
<td>8</td>
<td>10</td>
<td>6</td>
<td>1</td>
<td>11</td>
<td>8</td>
<td>4</td>
<td>8,10,6,1,11,8,4</td>
</tr>
<tr>
<td>ST-14</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5,5,7,5,5,4,5</td>
</tr>
<tr>
<td>ST-15</td>
<td>7</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>6</td>
<td>7</td>
<td>7,5,8,8,8,6,7</td>
</tr>
<tr>
<td>ST-16</td>
<td>9</td>
<td>11</td>
<td>9</td>
<td>11</td>
<td>12</td>
<td>4</td>
<td>9</td>
<td>9,11,9,11,12,4,9</td>
</tr>
<tr>
<td>ST-17</td>
<td>10</td>
<td>12</td>
<td>10</td>
<td>5</td>
<td>13</td>
<td>4</td>
<td>10</td>
<td>10,12,10,5,13,4,10</td>
</tr>
<tr>
<td>ST-18</td>
<td>4</td>
<td>13</td>
<td>10</td>
<td>5</td>
<td>13</td>
<td>4</td>
<td>4</td>
<td>4,13,10,5,13,4,4</td>
</tr>
<tr>
<td>ST-19</td>
<td>9</td>
<td>13</td>
<td>10</td>
<td>5</td>
<td>13</td>
<td>4</td>
<td>4</td>
<td>9,13,10,5,13,4,4</td>
</tr>
<tr>
<td>ST-20</td>
<td>9</td>
<td>11</td>
<td>9</td>
<td>11</td>
<td>12</td>
<td>9</td>
<td>9</td>
<td>9,11,9,11,12,9,9</td>
</tr>
<tr>
<td>ST-21</td>
<td>11</td>
<td>6</td>
<td>11</td>
<td>1</td>
<td>14</td>
<td>10</td>
<td>11</td>
<td>11,6,11,1,14,11,11</td>
</tr>
<tr>
<td>ST-22</td>
<td>12</td>
<td>14</td>
<td>5</td>
<td>10</td>
<td>15</td>
<td>11</td>
<td>12</td>
<td>12,14,5,10,15,11,12</td>
</tr>
<tr>
<td>ST-23</td>
<td>8</td>
<td>15</td>
<td>5</td>
<td>12</td>
<td>16</td>
<td>12</td>
<td>4</td>
<td>8,15,5,12,16,12,4</td>
</tr>
<tr>
<td>ST-24</td>
<td>12</td>
<td>14</td>
<td>5</td>
<td>10</td>
<td>17</td>
<td>11</td>
<td>12</td>
<td>12,14,5,10,17,11,12</td>
</tr>
<tr>
<td>ST-25</td>
<td>5</td>
<td>16</td>
<td>12</td>
<td>13</td>
<td>10</td>
<td>12</td>
<td>13</td>
<td>5,16,12,13,10,12,13</td>
</tr>
<tr>
<td>ST-26</td>
<td>5</td>
<td>14</td>
<td>5</td>
<td>9</td>
<td>15</td>
<td>13</td>
<td>13</td>
<td>5,14,5,15,13,13</td>
</tr>
<tr>
<td>ST-27</td>
<td>13</td>
<td>14</td>
<td>13</td>
<td>1</td>
<td>17</td>
<td>3</td>
<td>14</td>
<td>13,14,13,1,17,3,14</td>
</tr>
<tr>
<td>ST-28</td>
<td>14</td>
<td>17</td>
<td>12</td>
<td>9</td>
<td>18</td>
<td>4</td>
<td>15</td>
<td>14,17,12,9,18,4,15</td>
</tr>
<tr>
<td>ST-29</td>
<td>4</td>
<td>13</td>
<td>10</td>
<td>5</td>
<td>13</td>
<td>14</td>
<td>4</td>
<td>4,13,10,5,13,14,4</td>
</tr>
<tr>
<td>ST-30</td>
<td>4</td>
<td>13</td>
<td>10</td>
<td>5</td>
<td>13</td>
<td>15</td>
<td>4</td>
<td>4,13,10,5,13,15,4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of Alleles</th>
<th>14</th>
<th>17</th>
<th>13</th>
<th>13</th>
<th>18</th>
<th>15</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polymorphic Sites</td>
<td>16</td>
<td>22</td>
<td>10</td>
<td>16</td>
<td>26</td>
<td>17</td>
<td>20</td>
</tr>
<tr>
<td>% DNA Variability</td>
<td>3.20%</td>
<td>3.28%</td>
<td>2.22%</td>
<td>3.40%</td>
<td>3.66%</td>
<td>3.78%</td>
<td>2.82%</td>
</tr>
<tr>
<td>Silent Polymorphic sites</td>
<td>9</td>
<td>3</td>
<td>3</td>
<td>9</td>
<td>10</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>% Protein Variability</td>
<td>1.40%</td>
<td>1.43%</td>
<td>1.56%</td>
<td>1.49%</td>
<td>2.26%</td>
<td>2.65%</td>
<td>2.11%</td>
</tr>
<tr>
<td>dN/dS</td>
<td>0.1215</td>
<td>0.1177</td>
<td>0.4683</td>
<td>0.1094</td>
<td>0.4257</td>
<td>0.6639</td>
<td>0.6761</td>
</tr>
</tbody>
</table>

**Table 4.4:** A list of all alleles for each of the selected 7 genes and allelic profiles of each of the 30 sequence types generated by MLST assay. ........................................(Continued)
**Table 4.4 (Continued)**

The numbers at the bottom of the table show the number of different alleles identified for each gene, number of polymorphic Sites, percent DNA variability, number of silent polymorphic sites and the predicted percent protein variability for each the final gene segments.
Figure 4.1: A schematic representation of MS whole genome showing the positions of selected genes. MS whole genome sequence (approximately 800 kbp), the red circles represent the position of the 7 selected genes. The blue circle represents the position \(vlhA\) expressed gene. The sizes of the circles are not to the sizes of the genes scale. Notice that the selected genes are represented in 4 quarters of the genome.
Figure 4.2: Nine different MS strains amplified for each of the selected MLST genes and visualized on agarose gel. For each gene the expected amplicon size is shown.
**Figure 4.3:** Sequence alignment of 14 different *adk* alleles showing 16 polymorphic sites (9 of them are silent) along the 498 bp segment with a 3.20% DNA variability.
Figure 4.4: Sequence alignment of 17 different \textit{atpG} alleles showing 22 polymorphic sites (13 of them are silent) along the 669 bp segment with a 3.28\% DNA variability.
Figure 4.5: Sequence alignment of 13 different efp alleles showing 10 polymorphic sites (3 of them are silent) along the 450 bp segment with a 2.22% DNA variability.
Figure 4.6: Sequence alignment of 13 different gmk alleles showing 16 polymorphic sites (9 of them are silent) along the 471 bp segment with a 3.40% DNA variability.
Figure 4.7: Sequence alignment of 18 different nagC alleles showing 26 polymorphic (10 of them are silent) sites along the 708 bp segment with a 3.66% DNA variability.
Figure 4.8: Sequence alignment of 15 different ppa alleles showing 17 polymorphic sites (5 of them are silent) along the 453 bp segment with a 3.78% DNA variability.
Figure 4.9: Sequence alignment of 15 different *recA* alleles showing 20 polymorphic sites (5 of them are silent) along the 711 bp segment with a 2.82% DNA variability.
Figure 4.10: Sequence alignment of 30 different vlhA alleles .......(Continued)
Figure 4.10 (Continued)

showing 158 polymorphic (5 of them are silent) sites along the 441 bp segment with a
35.83% DNA variability. Notice a significant portion of the variation is due
insertions/deletions in hyper variable Proline rich repeats (1)
Figure 4.11: Graph showing number of alleles, number of polymorphic sites and the percent DNA variability for the 7 MLST final segments compared to the partial \( vlhA \) gene segment using 58 MS strains, field isolates and clinical samples. Percent DNA variability is according to the secondary vertical axis.
**Figure 4.12:** Phylogenetic tree of all 58 MS strains, field isolates and clinical samples using MLST assay. This tree was generated using neighbor-joining method for the cumulative sequence of the 7 MLST final gene segments. Each entry is labeled with the MLST sequence type, allelic profile and the available epidemiologic information.
Figure 4.13: Arkansas subpopulation: Phylogenetic tree of 14 MS strains, field isolates and clinical samples using MLST assay. This tree was generated using neighbor-joining method for the cumulative sequence of the 7 MLST final gene segments. Each entry is labeled with the MLST sequence type, allelic profile and the available epidemiologic information. This population was used to study the index of association ($I_A$) of the 7 MLST genes to determine if MS has a clonal structure at the local geographical level or not.
**Figure 4.14:** Ohio-Pennsylvania subpopulation: Phylogenetic tree of 14 MS field isolates and clinical samples using MLST assay. This tree was generated using neighbor-joining method for the cumulative sequence of the 7 MLST final gene segments. Each entry is labeled with the MLST sequence type, allelic profile and the available epidemiologic information. This population was used to study the index of association ($I_A$) of the 7 MLST genes to determine if MS has a clonal structure at the local geographical level or not.
Figure 4.15: Phylogenetic tree of 54 MS strains, field isolates and clinical samples using \( vlhA \) assay. The 4 repeated isolates are not represented. This tree was generated using neighbor-joining method for the partial \( vlhA \) gene sequence. Each entry is labeled with the MLST sequence type, allelic profile and the available epidemiologic information.
Chapter 5: Concluding Remarks and Future Research

*Mycoplasma gallisepticum* (MG) is considered the most pathogenic and economically significant amongst the four pathogenic avian mycoplasmas, and therefore the most studied and understood species. A clear example of this is the number of assigned genomic targets for MG sequence typing compared to the other avian mycoplasma. There are 5 targets for MG, one for *Mycoplasma synoviae* (MS) and there is none for *Mycoplasma iwoae* (MI) or *Mycoplasma meleagridis* (MM). Another example is that at least 11 whole genome sequences for MG are publicly accessible and there will be probably more to come in the near future, while there is only one for MS, one for MI, and none for MM. This enables more refined investigations examining the characters of MG, its biological processes, ecology and epidemiology. This makes MG the best studied avain mycoplasma, and makes it one of the best studied *Mollicutes* in general. Furthermore, at the commercial level, in the USA there are 3 available live MG vaccines providing a gradient of attenuation and effectiveness, but there is only one live vaccine for MS and it is not currently available in the United States and no licensed vaccines availbale for MI or MM.

In the recent years an apparent increase in the pathogenicity of non-MG avian mycoplasmal pathogens has been observed. Earlier this year Landman published a paper where he asked the question “Is *Mycoplasma synoviae* out running *Mycoplasma gallisepticum*?” indicating in the paper his observation of the increased pathogenicity of
MS and the declining importance of MG particularly in countries with more aggressive MG eradication programs. This situation presents avian disease specialists and epidemiologists with quite a knowledge gap and limited diagnostic tools, thus hindering their ability to address an emerging health challenge in poultry. In this research dissertation we attempted to fill some of these gaps in knowledge and diagnostic tools to assist with the prevention and control of the avian mycoplasma diseases increasing in importance. In the first project we studied the Variable Lipoprotein Hemagglutinin A (vlhA) gene as the single target for MS sequence typing to determine if it could be further improved. We designed new PCR primers in an effort to expand the targeted segment of the gene. We also performed a sequential generation study to identify the point at which the gene performs the spontaneous gene conversion process to produce the antigenic variation of the surface protein. By defining this point we defined the maximum genotypic potential for that genomic target. In the second study we targeted Clustered Regularly Interspaced Palindromic Repeats (CRISPR) region as an additional genomic target to be used for MS sequence typing. Twenty four different MS strains, field isolates and clinical samples were included in this study, but the amplification of the targeted region was not consistent in all strains. We obtained a few CRISPR sequences and the region showed a promising potential for sequence typing. This was in agreement with another study which showed through the whole genome sequencing of 12 MG isolates that the CRISPR region was useful in tracing the temporal and geographical patterns in MG house finch outbreak. But the direct amplification and sequencing of this region in MS using the convention PCR technique is inconsistent. In the third and final study we
developed a new Multilocus Sequence Typing (MLST) assay for MS. Seven housekeeping genes were selected to be incorporated in the final MLST scheme. Specific primers targeting the selected gene segments were designed and the selected gene segments were successfully amplified and sequenced in 58 MS strains, field isolates and clinical samples. The sequence profiling and analysis of the newly developed MLST showed a better correlation with the epidemiological information than did the \(vlhA\) assay and therefore appears to be a promising tool for more refined epidemiological investigations of MS outbreaks.

There is a need for further research to continue filling the knowledge gap. First there is an urgent need to have more available whole genome sequence for at least all the four pathogenic species, if not some of the closely non-related species as well. And with advent of next generation sequencing this shouldn’t be a difficult goal to achieve. This will facilitate the \(in-silico\) studying and extracting useful information about mycoplasma. Comparative genomics can be a very powerful tool to study evolution, strain differentiation and pathogenicity of these organisms.

Another mycoplasma that seems to be on the rise is \(M. iwoae\), at least in our experience. And currently there is no genomic targets for MI sequence typing. The only practical genotyping method available for MI is DNA finger printing, which is cumbersome and requires the isolation of the MI in pure culture. The development of sequence typing assay for MI would be a great help for the poultry industry.

Next generation sequencing is knocking of the doors of avian mycoplasma epidemiology. In 2012 and 2013 a series of papers has been published showing the use of
next generation sequencing in studying the evolution and epidemiology of MG over the course of 12 years of the MG house finch outbreak. They followed MG and its evolution as an emerging pathogen in a new population. In this series of studies the group has been able to produce 12 whole genome sequences of 12 different field isolates which is 4 times more the number of available genomes before conducting this study. This is just showing the sheer power of this technology and its potential to change our understanding of biology. Genomics can be extremely useful for a much better understanding of mycoplasma pathogenicity and epidemiology.
Bibliography


178


Studies of pathogenicity of turkey *Ureaplasma*. Avian Pathol. 7:577—582. 1978.


