The ostrich mycoplasma Ms02

Partial genome assembly, bioinformatic analysis and the development of three DNA vaccines

Marliz Strydom

Thesis presented in fulfillment of the requirements for the degree of

Master of Science (Biochemistry)

At the University of Stellenbosch

Supervisor: Dr. A. Botes

Co-Supervisor: Prof. D.U. Bellstedt

Department of Biochemistry

University of Stellenbosch

March 2013

Declaration

By submitting this thesis/dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly stated otherwise), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: March 2013

Abstract

The South African ostrich industry is under enormous threats due to diseases contracted by the ostriches. H5N2 virus (avian influenza) outbreaks the past two years have resulted in thousands of ostriches having to be culled. However, the more silent respiratory infectious agents of ostriches are the three ostrich-specific mycoplasmas. Named Ms01, Ms02, and Ms03, these three mycoplasmas are responsible for dramatic production losses each year, due to their intrusive nature and the fact that no vaccines are currently available to prevent mycoplasma infections in ostriches. The use of antibiotics does not eradicate the disease completely, but only alleviates symptoms. The ostrich industry commissioned investigations into the development of three specific vaccines using the relatively novel approach of DNA vaccination.

The concept of DNA vaccine development is based on the availability of complete genome sequences of the pathogen against which the vaccine is to be developed. This is necessary in order to identify vaccine candidate genes through comparative genomic studies. The Ms02 genome has previously been sequenced, resulting in 28 large contiguous sequences. This thesis used the technique of **T**hermal **A**symmetric **I**nterlaced **P**olymerase **C**hain **R**eaction (TAIL-PCR) to attempt assembly of these 28 contiguous sequences. The number was reduced to 14 large contiguous sequences, which were then subjected to repetitive sequence analysis and open reading frame analysis. Bioinformatic software was also used to predict the origin of replication. The extent of repeats in the Ms02 genome is illustrated, as well as the problems with genome assembly when dealing with repetitive-rich and A+T-rich genomes as those of mycoplasmas.

Previous studies determined the mycoplasma *oppA* gene to be a good vaccine candidate gene, due to its cytadherent properties. This thesis describes the development of three DNA vaccines containing the Ms02 *oppA* gene, and a preliminary attempt to prove expression of one of these vaccines in a cell culture-based system. The DNA vaccine vectors pCI-neo, VR1012, and VR1020 were chosen for the vaccine development. The Ms02 *oppA* gene was also cloned into the prokaryotic expression vector pGEX-4T-1 in order to express the OppA protein for purification. The purified protein may be used in future serological tests in ostrich vaccination trials. In this study the protein was used to elicit anti-OppA rabbit antibodies, which were used to attempt detection of the pCI-neo-driven OppA protein expression in an MDA cell line in a transfection study. However, preliminary findings could not detect expression, but did indicate that the currently used colorimetric western blot technique may not be sensitive enough. It is suggested that different cell lines need to be investigated. Further optimisations are also required to decrease the observed non-specific binding.

Opsomming

Die Suid-Afrikaanse volstruisbedryf is onder geweldige druk vanweë siektes wat die volstruise bedreig. Die epidemie van die H5N2 virus (voëlgriep) in die afgelope twee jaar het veroorsaak dat duisende volstruise van kant gemaak moes word. Daar is egter nog 'n bedreiging wat tot geweldige produksie verliese lei elke jaar: die respiratoriese infeksies wat versoorsaak word deur die drie volstruis mikoplasmas, genoem Ms01, Ms02 en Ms03. Geen entstowwe is tans beskibaar om die infeksies te voorkom nie, en behandeling met behulp van antibiotikas is nie effektief in die genesing van infeksie nie, maar help net om die simptome te verlig. Weens die erns van die saak, het die Suid-Afrikaanse volstruisbedryf 'n ondersoek geloods na die ontwikkeling van enstowwe teen elkeen van die drie volstruis mikoplasmas. Die relatiewe nuwe benadering van DNA-entstof ontwikkeling was die strategiese keuse.

Die beginsel van DNA-entstof ontwikkeling berus op die beskikbaarheid van die genoomvolgordes van die siekte-veroorsakende organisme waarteen die enstof ontwikkel word. Geskikte kandidaat entstof gene word so opgespoor met behulp van vergelykende studies met ander beskikbare genome. Die Ms02 genoomvolgorde is voorheen bepaal en word verteenwoordig deur 28 groot geenvolgorde fragmente. Die tegniek van ***T****hermal* ***A****symmetric* ***I****nterlaced* ***P****olymerase* ***C****hain* ***R****eaction (TAIL-PCR)* is gebruik om van die 28 fragmente aan mekaar te las. Die aantal fragmente is verminder na 14 groot geenvolgorde fragmente, wat vervolgens gebruik was om die omvang van herhalende volgordes in die genoom te bepaal, om nuwe leesrame te ondersoek, asook om die oorsprong van DNA replikasie op te spoor met behulp van bioinformatika sagteware. Die omvang van die herhalende aard van die Ms02 genoom word geïllustreer, asook die gepaardgaande probleme met die las van geenvolgorde fragmente wanneer met genome van veelvuldige herhalende volgordes, wat boonop A+T-ryk is, gewerk word, soos die van mikoplasmas.

Vorige studies het die mikoplasma *oppA* geen geïdentifiseer as 'n geskikte kandidaat entstof geen as gevolg van sy selaanhegting-eienskappe. Hierdie studie behels die invoeging van die Ms02 *oppA* geen in drie DNA-enstof vektore, naamlik pCI-neo, VR1012, en VR1020, asook die voorlopige poging om bewys van uitdrukking van een van die entstowwe in 'n selkultuursisteem te bewerkstellig. Die geen is ook gekloneer in die prokariotiese ekspressie vektor pGEX-4T-1, ten einde die Ms02 OppA proteïen te isoleer. Die geïsoleerde proteïen kan in serologiese toetse in toekomstige volstruis enstof proewe gebruik word. In hierdie studie is die proteïen gebruik om konyn teenliggame teen dit op te wek, wat dan gebruik was om vir die pCI-neo-gedrewe ekspressie van die *oppA* geen te toets in 'n selkultuur omgewing deur 'n MDA sellyn te transfekteer. Die voorlopige resultate toon nie ekspressie van die OppA proteïen aan nie, maar het wel uitgelig dat die western blot tegniek wat tans gebruik word, dalk nie sensitief genoeg is nie. Dit kan belowend wees om ander tipes selle te toets. Verdere optimisering is ook nodig om die nie-spesifieke binding wat waargeneem is, te verlaag.

Acknowledgements

There are so many people who helped to make this thesis possible. I would like to thank all of you, whom have supported and inspired me. My sincerest gratitude.

I thank the following in particular:

Dr Annelise Botes, my supervisor, for her support and direction.

Prof. Dirk U. Bellstedt, for his leadership and never-ending well of wisdom.

Mrs Coral de Villiers, for her help, both emotional and technical.

The South African Ostrich Business Chamber for financial support.

The Bellstedt laboratory, for interesting times and stimulating conversation, and more recently, the Botes laboratory, for company in solitude (2010-2012).

My family, friends and loved-ones, for their encouragement, love, and advice, even when not asked for.

Abbreviations

A	adenine
ABC	ATP-binding cassette
AD	arbitrary degenerate
Ag	antigen
ampR	ampicillin resistance gene
APC	antigen presenting cell
BGH	bovine growth hormone
BM	basal membrane
bp	base pair(s)
BSA	bovine serum albumin
C	cytosine
CAT	chloramphenicol acetyltransferase
CpG	cytidine-phosphate-guanosine
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
dTTP	deoxythymidine triphosphate
EDTA	ethylene diamine tetra-acetic acid di-sodium salt
ELISA	enzyme-linked immunosorbent assay
EMC	extracellular matrix components
EU	European Union
G	guanine
G+C	guanine and cytosine
GOI	gene(s) of interest
GST	glutathione S-transferase
HIV	human immunodeficiency virus
hCMV	human cytomegalovirus
HMM	hidden Markov models
IBV	infectious bronchitis virus
IDT	Integrated DNA Technologies
IFN	interferon
Ig	immunoglobulin

IL	interleukin
ISS	immunostimulatory sequence
kanR	kanamycin resistance gene
kb	kilo base pairs
kDa	kilo dalton
LB	Luria-Bertani
MCS	multiple cloning site
MG	*Mycoplasma gallisepticum*
MHC	major histo-compatability complex
mRNA	messenger ribonucleic acid
MS	*Mycoplasma synoviae*
NCBI	National Center for Biotechnology Information
NDV	Newcastle Disease virus
NKC	natural killer cells
Opp	oligopeptide permease
ORF	open reading frame
ori	origin of replication
PBS	phosphate buffered saline
PCR	polymerase chain reaction
RBS	ribosome-binding site
RE	restriction endonuclease
rRNA	ribosomal ribonucleic acid
SDM	site-directed mutagenesis
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SV40	Simian virus 40
T	thymidine
TAIL-PCR	Thermal Asymmetric Interlaced Polymerase Chain Reaction
TLR	toll-like receptor
T_M	melting temperature
TNF	tumour necrosis factor
UV	ultraviolet

TABLE OF CONTENTS

Chapter 1

1.1. INTRODUCTION

Commercial ostrich farming in South Africa started in the 1850s when pioneering farmers in the Oudtshoorn district saw the economic potential in farming for ostrich feathers. Up until the early 20th century, horse-drawn carriages enabled elaborate feather hats to be fashionable and were a sign of the wealthy and elite. The feather industry came to a halt with the mass production of automobiles, the roof making the wearing of feather hats impractical. Soon afterwards, the start of World War I led to the collapse of the feather industry.

In the 1950s the Oudtshoorn farmers started to rebuild ostrich farming. Increasing demand for ostrich meat led to the first abattoir being erected in the mid-1960s. After that, the industry boomed with demand for fresh and dried ostrich meat locally, as well as demand for ostrich leather internationally, especially in the USA. Years later South Africa is still considered the leader in the ostrich industry, although there are ostrich farms in the USA and Australia as well.

In modern day society, where people are becoming increasingly aware of their health and the nutrient contents of their food, ostrich meat has received attention: it tasting similar to lean beef and being regarded as the healthiest red meat, with almost no fat, low cholesterol, and high in calcium, protein and iron (Canadian Ostrich Association, 2008). Total ostrich export revenue for 2005 amounted to R1.2 billion, with the Western Cape contributing 75% of total ostrich production (SAOBC, 2010). Furthermore, the industry provides work opportunities in rural areas for unskilled workers, amounting to 20 000 jobs (South African Ostrich Industry, 2004). Being one of the top agricultural industries in South Africa, the ostrich industry not only has local socio-economic benefits, but draws international acclaim as well.

1.2. THREATS TO THE SOUTH AFRICAN OSTRICH INDUSTRY

Almost all ostrich products are exported, including meat, leather, fertile eggs and even live ostriches (Verwoerd, 2000). This places a tremendous strain on the industry in terms of quality and disease control, especially for export into the European Union (EU), which imports 90% of SA ostrich meat. This brings to mind the ban in August 2004 when the EU banned the import of South African live ostriches, meat and ostrich eggs due to an outbreak of avian influenza. The ban resulted in millions of rands of loss for the industry and was lifted only a year later. The possibility of disease outbreak and resulting trade limitations and quarantine could cripple the socio-economic ostrich environment. In recent times, another outbreak of avian influenza is the biggest threat yet to the ostrich industry. More than 50 000 ostriches have been culled during the last outbreak, and more than R60 million has been spent to combat the spread of the virus (Die Burger, 6-9-2012). Contaminated meat can be made safe for human consumption, by heating for only 3 sec at 70°C. However, the 'precooked' meat is not as popular and commands only a fraction of the price of raw meat. In addition to the pressure of avian influenza, mycoplasmas cause respiratory diseases in feedlot ostriches and can result in production loss of up to 30% (Dr A. Olivier, in personal communication with B. Pretorius, 2009). The South African Ostrich Industry is as a result almost on the brink of collapse.

1.3. OSTRICH MYCOPLASMA INFECTION

There are three mycoplasmas that only infect ostriches. Infection causes loss in production, downgrading of the meat, as well as susceptibility to secondary infections. The pressure of feedlot farming creates an ideal environment for these opportunistic bacteria to infect the ostriches. Ostriches are especially in danger with seasonal changes, when placed under stress when moved or transported, and due to poor hygiene practices. Although certain antibiotics mitigate the symptoms of infection, they do not eradicate the infection completely. This, together with the fact that there is currently no methods available to prevent infection, impress upon the need for vaccine development. The South African ostrich industry commissioned an investigation into the development of vaccines against ostrich mycoplasmas. A previous study determined that a species-specific approach is needed, since existing poultry vaccines do not provide adequate protection. Thus far a DNA vaccine against the ostrich mycoplasma Ms01 was developed (Pretorius, 2009) and tested (Brandt, 2012).

A similar approach was employed to develop an Ms02 DNA vaccine. The gene encoding the antigenic OppA protein was identified as a candidate gene for DNA vaccine development. The gene was cloned and modified to enable subcloning and expression in the relevant DNA vaccine vectors (Steenmans, 2010).

1.4. OBJECTIVES OF THE STUDY

This study focused firstly on filling in the gaps of the previously sequenced Ms02 genome. In the second part, the modified *oppA* gene was cloned into suitable vectors for protein expression and vaccine development. This is to fulfil the objectives of the study, which are:

i. To assemble partially the contiguous sequences generated from previous Ms02 genome sequencing and to subject the assembled sequence to bioinformatic analysis.
ii. To subclone the gene into a suitable prokaryotic expression vector in order to express and purify the OppA protein to be used as antigen in serological tests.
iii. To subclone the gene into three different DNA vaccine vectors and test the efficacy of the DNA vaccine by evaluating its expression in a eukaryotic cell system.

To aid in understanding mycoplasmas, Chapter 2 starts with a discussion on the history of mycoplasmas: their origin, classification, biochemistry and genomic details. The second part of Chapter 2 is devoted to the rise and development of vaccines, with the focus on DNA vaccines as a next-generation vaccine. The method of **T**hermal **A**symmetric **I**nterlaced **P**olymerase **C**hain **R**eaction (TAIL-PCR) and its use in assembling the contiguous sequences (contigs) of the Ms02 genome, followed by bioinformatic analysis of the assembled sequence, are discussed in Chapter 3. The subsequent chapters focus on the DNA vaccine development. Firstly, in Chapter 4, cloning of the *oppA* gene into three suitable DNA vaccine vectors, cloning and expression of the gene in a prokaryotic expression vector, as well as antibody production, are discussed. Thereafter, the use and efficacy of one DNA vaccine vector is examined in Chapter 5 to determine if eukaryotic cells will transcribe the plasmid *in vivo* in a tissue culture study. Chapter 6 covers the possibilities of further research, based on the study's conclusions. Lastly, a reference list and relevant addenda are attached.

Chapter 2

In this literature review, an overview of mycoplasmas will be given, including origin, phylogeny, genome features, biochemistry and pathogenicity of these organisms. Ostrich mycoplasmas will be discussed as an integrated theme throughout the general discussion on mycoplasmas. The next section will include a general discussion of vaccines, with the focus on DNA vaccine development and strategies.

2.1. MYCOPLASMAS: FEATURES, METABOLISM AND PATHOGENICITY

2.1.1. INTRODUCTION

The name mycoplasma was first coined by Albert Bernhard Frank in 1889, from the Greek *mykes* (fungus) and *plasma* (formed), due to fungus-like characteristics (Krass and Gardner, 1973). Later it was discovered that only *M. mycoides* exhibits fungus-like growth. Since the first successful cultivation in 1898 by Nocord and Roux, mycoplasmas have been mostly misunderstood. Due to their small size, mycoplasmas are able to pass through bacteria-retaining filters and were thus thought to be viruses. In the 1930s the belief arose that mycoplasmas are stable L-phase variants of common bacteria. Genomic analysis in the 1960s and 70s, together with advancements in the knowledge of cell structure, genome, and metabolic pathways of mycoplasmas, finally led to the current paradigm where mycoplasmas are known to be cell wall-less descendants of Gram-positive bacteria.

Today mycoplasmas are known to be the smallest cellular organisms capable of self-reproduction (Razin, 1987). Their lack of a cell wall and lack of many metabolic pathways render mycoplasmas dependent on a host for their required nutrients and protection. Mycoplasmas have been isolated from hosts such as humans, other mammals, reptiles, fish, arthropods, and plants, where they cause diseases such as contagious bovine and caprine pleuropneumonia in cattle, contagious agalactia in small ruminants, calf pneumonia, enzootic pneumonia in pigs, and chronic respiratory diseases in poultry. Three mycoplasmas specific to ostriches have been associated with respiratory diseases in the South African ostrich. The economic impact of this disease is considerable, as respiratory diseases in ostriches have many direct and indirect effects. Significant production losses are caused by the cost of treatment, wearing down of ostriches, downgrading of carcasses and the ostriches' increased susceptibility to opportunistic infections. Mortality may even occur. This has far-reaching socio-economic implications, since ostrich farming is a significant contributor to agricultural exports and provider of job opportunities.

Despite the severity of the situation in general, with regards to all hosts, few vaccines are available to combat the illnesses: mostly whole-cell vaccines, some semi-virulent, that only provide partial immunity and can have negative side effects. Attempts at vaccine development and improvement often lead to exacerbation of diseases. It is becoming even more pertinent to develop reliable vaccines, because antibiotics only control mycoplasma infections partially, and their efficacy is decreasing. In addition to the problems associated with vaccine development, the antibodies generated against either ostrich mycoplasma proteins show little or no

cross-reactivity with the other ostrich mycoplasma proteins, emphasising the urgent need to develop mycoplasma-specific vaccines.

2.1.2. MYCOPLASMA IDENTIFICATION AND CLASSIFICATION

2.1.2.1. Taxonomy

The mycoplasma's most distinguishing feature is its lack of a cell wall (Razin, 1985). Mycoplasmas evolved by degenerate evolution from low G+C, Gram-positive bacteria, although they stain negative in the Gram test (Herrmann, 1992). In the process they lost the genes encoding for a cell wall and many other metabolic processes (Bradbury, 2005; Razin *et al.,* 1998; Weisburg *et al.,* 1989). Their lack of a cell wall classifies mycoplasmas into a distinct class of prokaryotes, the *Mollicutes* (Latin *mollis*, soft; *cutes*, skin), the only class within the division *Tenericutes* (wall-less bacteria) of the kingdom Procaryotae (Freundt, 1973). Three other divisions in the Kingdom exist, namely the Gram-positive bacteria *Firmicutes*, the Gram-negative bacteria *Gracilicutes* and the archaebacteria *Mendosicutes* (Razin *et al.,* 1998). Furthermore, differences in morphology, unusually small genome size, low G+C content and nutrient requirements, subdivides the Mollicutes in five orders, namely *Acholeplasmatales* (*Acholeplasma*), *Anaeroplasmatales* (*Anaeroplasma*, *Asteroleplasma*), *Entomoplasmatales* (*Entomoplasma, Mesoplasma*, *Spiroplasma*), *Haloplasmatales (Haloplasma)* and *Mycoplasmatales* (*Mycoplasma*, *Ureaplasma*) (Bradbury, 2005; Razin *et al.,* 1998; Weisburg *et al.,* 1989). Currently, six families with a total of twelve genera are recognised in the class *Mollicutes* (Ludwig *et al*., 2009).

2.1.2.2. Origin

Two schools of thought used to exist on mycoplasma evolution. One model, as proposed by Neimark and London (1982), presumed that mycoplasmas originated by degenerate evolution after the diversification of bacteria. Thus different species of mycoplasmas would have originated from different branches of bacteria and would therefore not be phylogenetically related.

The other model, proposed by Morowitz and Wallace (1973), theorised that mycoplasmas are bacteria precursors, because of their small genome size and simplicity. Thus the assumption is that the small genome must have duplicated numerous times to evolve into the ancestral bacteria.

Today it has been proven that neither of these is correct, but that mycoplasmas originated by degenerate evolution from a low G+C content, Gram-positive, walled branch of bacteria. Firstly, Woese *et al.* (1980) showed a phylogenetic relationship between bacteria and mycoplasmas based on comparative analysis of 16S rRNA sequences. The findings established a relationship between mycoplasmas and a low G+C branch of a clostridial phenotype (*Clostridium innocuum* and *C. ramosum*) (Razin, 1985; Weisburg *et al.,* 1989). Woese (1987) also proved the mycoplasma phylogenetic tree to be monophyletic, thus disproving the first-mentioned model. Secondly, nucleic acid hybridisation and sequencing studies showed that mycoplasmas are related to Gram-positive bacteria, although they stain negative in the Gram test.

2.1.2.3. Phylogeny

The mycoplasma genus consists of four phylogenetic groups: (a) the anaeroplasma group, (b) the spiroplasma group, (c) the pneumonia group, and (d) the hominis group (Dybvig and Voelker, 1996). In 2005 three distinct mycoplasmas were isolated from ostriches in the Klein Karoo, Central Karoo and Garden Route districts (Botes *et al*., 2005a). Based on 16S rRNA analysis, the mycoplasmas were identified as three unique and ostrich-specific mycoplasmas that fall within the hominis group. Mycoplasmas are usually named according to their host, and therefore the three mycoplasmas were provisionally named *Mycoplasma struthiolus*(Ms)01, 02, and 03, after their host *Struthio camelus,* until formally described. Further studies named Ms01 as *Mycoplasma struthionis* sp. nov. and Ms03 as *Mycoplasma nasistruthionis* sp. nov. (Langer, 2009). However, for the remainder of this thesis, they will be referred to as Ms01 and Ms03. The three mycoplasmas are phylogenetically divergent and fall into two different clades, with Ms01 in a separate clade (Figure 2.1).

2.1.3. MOLECULAR AND BIOPHYSICAL CHARACTERISTICS

2.1.3.1. General characteristics of the mycoplasma genome

At the start of the 1990s the first large scale attempts were made to sequence entire genomes of mycoplasmas (Razin *et al*., 1998). The mycoplasma cell is typical of prokaryotes, containing a plasma membrane, ribosomes, and a circular double-stranded DNA genome. However, that is where the similarity ends. Mycoplasma genomes are significantly smaller than any other bacteria and have a lower G+C content than typical self-replicating cellular organisms (Razin, 1985, 1997). Mycoplasma genome sizes vary between 580 kb for *M. genitalium* and 1 380 kb for *M. mycoides* subsp. *mycoides* (Dybvig and Voelker, 1996; Fraser *et al.,* 1995; Lucier *et al.,* 1994; Razin *et al.,* 1998). Genome size may even vary between strains of the same species (Razin *et al*., 1998).

2.1.3.2. Repetitive DNA sequences

The mycoplasma genome is filled with repetitive elements, thought to be either insertion elements or segments of genes that encode for surface antigens, and may be the cause of variability in the genome size of mycoplasmas (Razin *et al.,* 1998). Repetitive DNA sequences are not expected in genomes as small as those of mycoplasmas, and the high prevalence of these elements is somewhat surprising. Many repeats are homologous to genes that encode major surface antigens and are also often found in these genes, possibly contributing to antigenic variation through DNA rearrangement (Dybvig and Voelker, 1996; Razin *et al*., 1998). Bacteria use repeats to enable gene recombination and to change the number and size of repeats. This alters protein expression (Shapiro, 2002), and some repeats may even encode small proteins themselves (Reed and Altman, 1983). Mycoplasma genome repeats will be further discussed in Chapter 3, Section 3.1.3.

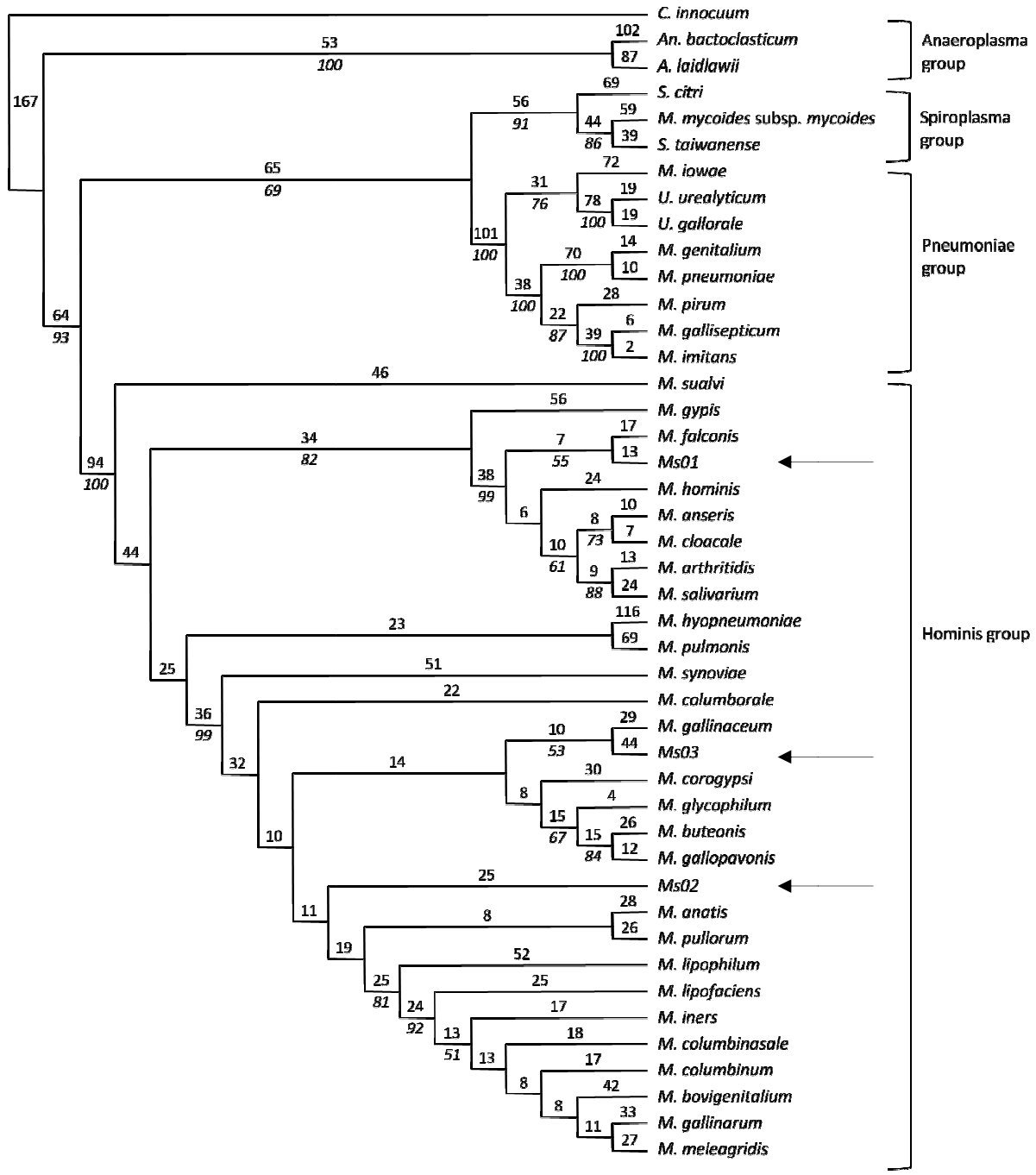

Figure 2.1 A phylogenetic tree of mycoplasmas based on the analysis of 16S rRNA gene sequences. The tree represents one of eight of the shortest trees retrieved by doing a heuristic search (CI = 0.381, RI = 0.696). The arrows indicate the branches that collapse in the strict consensus tree. Branch lengths and bootstrap values are indicated above and below the line respectively. Long arrows point to the three ostrich mycoplasmas.

2.1.3.3. Base composition and codon usage

The G+C content of mycoplasma genomes range from 24 – 33 mol%, with the exception of *M. pneumoniae*, which has the highest value of 41% (Bové, 1993; Razin, 1992; Rottem and Barile,1993). The unique base composition is a result of ineffective uracil-DNA glycosylation activity in mycoplasmas, which causes preferential usage of adenine- and thymine-rich codons (Dybvig and Voelker, 1996; Razin, 1985). Up to 90% of codons contain an A or T in the third nucleotide (wobble) position. As a consequence, mycoplasmas employ an alternative genetic code, known as the mold mitochondrial code, where the TGA codon encodes for tryptophan rather than signalling termination, as in the universal genetic code. Mycoplasmas utilise the codons UAA and UAG as stop codons, with preference given to UAA (Bové, 1993; Marin and Oliver, 2003; Razin *et al.*, 1998). This has practical implications when mycoplasma genes are to be expressed in bacteria that employ the universal genetic code (Dybvig and Voelker, 1996; Razin *et al.,* 1998, Söll and RajBhandary, 2006). Due to premature truncation caused by the TGA codon, shorter and dysfunctional products will form, hindering experimental studies of the gene of interest (Dybvig and Voelker, 1996; Razin *et al.,* 1998). Codon bias is not only apparent in the third nucleotide position, but also in the first and second, with considerable implications for the amino acid composition of a protein (Fadiel *et al.*, 2005; Razin *et al.*, 1998). Compared to *Escherichia coli*, an organism frequently used in laboratory studies, mycoplasmas have fewer GGN, CCN, GCN and CGN codons and as a result fewer glycine, proline, alanine and arginine residues in their proteins. On the other hand, the A+T-rich environment of the mycoplasma genome causes a high percentage of AAN, TTY, TAY and ATN codons to occur, resulting in an abundance of asparagine, lysine, phenylalanine, tyrosine, and isoleusine amino acid residues in their proteins. Furthermore, highly conserved proteins often have lysine residues (AAA, AAG), where the corresponding protein in another organism will have arginine residues (AGA, AGG, CGN) (Dybvig and Voelker, 1996).

2.1.3.4. Rare codons and codon bias in E. coli

E. coli is a popular heterologous expression system, but due to its high codon bias, genes that contain so-called rare codons will cause translational problems when expressed.

There are 64 codons in the universal genetic code, with three being stop codons. The amount of anticodons (tRNA) seems to be directly proportional to the codon usage frequency (Kane, 1995). Low-usage codons have a usage frequency lower than that of the non-degenerate codons AUG (Met) and UGG (Trp), as well as that of optimal codons. Optimal codons have two or more degenerate codons, each with its own specific tRNA (Chen and Texada, 2006). Rare codons are not only used seldom or infrequently, but also have low-abundant tRNAs to decode them. tRNA availability causes common codons to be more easily translated than rare codons. Rare codons are thus distinguished from low usage codons by the tRNA availability and tend to be in genes expressed at a low level, while major codons are used in genes with a high expression level. Therefore, if a gene with many rare codons is desired in high amounts, as is the case with recombinant protein expression, the translational machinery may not be able to keep up with production. This will lead to mistranslation or reduced yield (Kane, 1995).

Low usage codons, such as the Cys codons UGU and UGC, Thr codons ACU and ACG, or His codons CAC and CAU, are not rare codons of *E. coli*. However, rare codons are always low usage codons. *E. coli* has about 30 low usage codons, of which 20 are rare codons. Seven of them (AGG, AGA, CGA, CUA, AUA, CCC and CGG) are used at a frequency lower than 0.5% (Chen and Texada, 2006). The rare arginine codons AGG and AGA were the first illustrated to have a detrimental effect on expression (Brinkmann *et al*., 1989; Spanjaard and Van Duin, 1988; Spanjaard *et al.,* 1990). These codons occur at a frequency of 0.14% and 0.21% respectively. CUA and AUA is the next rare codon pair, with CGG, CGA and CCC codons with frequencies close to those (Kane, 1995). Rare codons and rare codon clusters have been shown to cause quantitative and qualitative expression problems in *E. coli* and other organisms, occurring mainly on the translational level (Gurvich *et al*., 2005; Kane, 1995; Makrides, 1996).

2.1.3.5. Gene arrangement

The structure and organisation of important genes is usually highly conserved between different species. Rottem and Barile (1993) found groups of genes conserved within the genomes of different mycoplasma species, while Rocha and Blanchard (2002) stated gene order in the genome to be poorly conserved. This finding was supported by Van der Merwe (2006), who found that the genes in the operons encoding cytadhesin proteins GapA, CrmA, CrmB and CrmC are found in *M. gallisepticum*, *M. hyopneumoniae* and *M. pulmonis*, but that their position in the respective genomes is not conserved.

Genome reduction can occur either by reducing the amount of genes or clustering genes together. In mycoplasmas, both processes were employed to reduce genome size (Siqueira *et al*., 2011). Essential genes are generally clustered together, such as those involved in nucleotide transport and metabolism (Souza *et al*., 2007). In twelve studied mycoplasma genomes, glycolysis was found to be the most conserved pathway. Genes involved with DNA replication, recombination and repair, and especially ribosomal proteins, are also conserved within different clusters. Although there are numerous conserved genes and clusters among mycoplasma genomes, the position of these gene clusters are generally not in the same relative position (Yotoko and Bonatto, 2007). Furthermore, comparison of gene organisation between pathogenic and non-pathogenic mycoplasmas seems to indicate that gene organisation is not related to pathogenesis (Siqueira *et al*., 2011).

2.1.3.6. Cell size and shape

The most distinctive feature of mycoplasmas is the absence of a cell wall. Their cell size is unusually small, ranging from 0.3 – 0.8 µm in diameter (Prescott *et al*., 2002; Weisburg *et al*., 1989). All mycoplasma cultures have a basic coccal shape, although elongated and filamentous forms have been observed under certain conditions. The growth medium's nutritional qualities, osmotic pressure, as well as the growth phase of the culture, all have an influence on the cell shape. The absence of a cell wall means no restrictions are placed on the cell shape, since the plasma membrane is not rigid. As such, mycoplasma cells are pleomorphic and vary in shape from spherical, pear-shaped organisms to branched or helical filaments, with a diameter of up to 150 µm (Carson *et al.,* 1992; Edward and Fitzgerald, 1951; Klainer and Pollack, 1973; Razin, 1978; Razin *et al.,* 1998).

2.1.3.7. Motility

Although mycoplasmas lack many metabolic pathways, they do have complex systems to enable their parasitic lifestyle, for example gliding motility, adhesion to host cells, internalisation into the host cells, membrane transport, and antigenic variation (Miyata and Seto, 1999). Their cell morphology revealed mycoplasmas to be grouped as either highly polar or non-polar. The polar mycoplasmas exhibit a terminal tip structure, shown to assist in gliding motility on liquid-covered solid surfaces, as demonstrated in *M. pulmonis, M. pneumoniae* and *M. gallisepticum*. It would seem that only the pathogenic mycoplasmas contain the tip structure. Mycoplasmas do not move as a mass, but individually either in a circular pattern or narrow bends (Razin, 1987). Even so, the gliding motility assists in defending the mycoplasma from the host's defense mechanisms and helps it to reach its target (Bradbury, 2005).

Mycoplasma motility also assists with adherence to host cells. Not all mycoplasmas are capable of motility. Absence of these structures in the non-polar mycoplasmas contributes to the observed irregular and often branched morphology (Miyata and Seto, 1999). Although attachment and adhesion to cell surfaces represent the major requirement for pathogenesis, commensal mycoplasmas are also capable of cell adhesion, even in the absence of established adhesion proteins. Rather, adhesion is achieved through specific glycolipid and/or glycoprotein receptors on the host cell surface (Razin and Jacobs, 1992).

2.1.4. METABOLISM

2.1.4.1. Gene expression

Gene expression and protein production in mycoplasmas occur mostly as for Gram-positive bacteria. The only difference in mRNA translation between mycoplasmas and Gram-positive bacteria is the TGA codon usage. Mycoplasmas use the mold mitochondrial genetic code where TGA codes for tryptophan, whereas Gram-positive bacteria employ the universal genetic code where the TGA codon signals termination of translation. TGG was retained as a second tryptophan codon (Jukes and Osawa, 1993). Most genes in the mycoplasma genome start with the usual initiation codon ATG, but GTG and TTG can also serve as alternative start codons (Dybvig and Voelker, 1996). ATG is the preferred start codon in genes that are intermittently expressed, as it interacts more strictly with the initiation transcript RNA (tRNA) than the alternative start codons (Sakai *et al.,* 2001).

2.1.4.2. Protein production and post-translational modifications

Since mycoplasmas do not have a cell wall and therefore no periplasmic space, proteins are either cytoplasmic, membrane-bound or secreted (Dybvig and Voelker, 1996). Mycoplasmas possess a high amount of lipoproteins, many of which possess a consensus signal peptidase II recognition site, followed by a cysteine residue. Lipoprotein processing often involves cleavage of the signal peptide in front of the cysteine residue, resulting in cysteine being the first amino acid of the lipoprotein.

The typical eubacterial secretion signal sequence (-4)-VAASC-(+1) that is responsible for directing proteins toward a secretory pathway for transport across the plasma membrane, is also present in mycoplasmas (Henrich *et al*., 1999). In Figure 2.2, a comparison of the conserved sequence that forms part of the OppA protein is shown for *M. hominis* (M. hom), Ms01, Ms02 and Ms03. The signal peptide directs the protein to the outside of

the membrane, where the protein then attaches via the cysteine amino acid. A signal peptidase II enzyme is responsible for the cleavage of the signal peptide form the lipoprotein precursor.

M. hom	22'	LVAAS***C***KIDPA	32'
Ms01	23'	LVAAA***C***NSKSA	33'
Ms02	23'	SVAVS***C***ANNVS	33'
Ms03	18'	ALALS***C***KWTSS	28'

Figure 2.2 Signal peptidase II recognition site of the OppA protein. Comparison of the typical secretion signal amino sequence of the signal peptidase II recognition sequence as part of the OppA protein amino sequence between *M. hominis* (M. hom) and the three ostrich mycoplasmas. Underlined is the conserved sequence, where the *C* represents the cysteine amino acid responsible for the attachment of the protein to the membrane.

Although simpler than eukaryotic proteins, prokaryotic proteins do require certain post-translational modifications, such as phosphorylation, acylation and isoprenylation in mycoplasmas. Data about post-translational modification of bacterial proteins remain scarce, and especially little is known about glycosylation (Demina *et al*., 2011). Proteins have a cysteine residue in a position that is susceptible to covalent fatty acid modifications. Mycoplasmas are known for their large amount of membrane proteins that contain bound fatty acids (Demina *et al*., 2011).

Covalent modifications to proteins are a major pathway of control and regulation in many important cellular processes, for instance protein phosphorylation is a known regulatory switch involved in intracellular signalling. It is possible that phosphorylation of cytoskeletal proteins in mycoplasmas may regulate cellular activities such as cytadherence, gliding motility and cell division.

The function and mechanism of action of isoprenylation is not entirely clear (Dybvig and Voelker, 1996, Razin *et al.,* 1998). Isoprenylation refers to the addition of an isoprenoid group to a cysteine residue at the C-terminal of a protein. It has been well recognised that the incorporation of isoprenoids to proteins as a post-translational method is an important process for facilitating protein-protein and protein-membrane interactions through specialised prenyl binding domains (Kloog and Cox, 2004; Marshal, 1993; Novelli and D'Apice, 2012). It is generally assumed that the isoprenyl groups enable attachment to the cell membrane, even though direct proof is lacking. Studies have shown that certain proteins, such as G-proteins whose function is critically dependant on their localisation, require this post-translational modification in order to anchor the protein to the membrane (Magee and Marshall, 1999). Thus prenylation is seen as necessary for regulating the localisation and function of proteins. Even though there are numerous studies on protein prenylation, there is still a lot to discover and this remains a field of active investigation (McTaggart, 2006).

2.1.4.3. Biosynthetic pathways

Mycoplasmas have not only lost the genes for cell wall synthesis during evolution, but also many genes involved in catabolic pathways and metabolite transport. This reduction in metabolic genes accounts for the mycoplasma's obligatory parasitic lifestyle. Their diminutive gene reservoir renders mycoplasmas dependant on their hosts for a variety of nutrients, such as fatty acids, cholesterol, some amino acids, and purines and

pyrimidines (Dybvig and Voelker, 1996; Henrich *et al.,* 1999; Prescott *et al*., 2002). Mycoplasmas are unable to synthesise sterols, the same as other prokaryotes. However, unlike other prokaryotes, most mycoplasmas are dependent on exogenous cholesterol for growth and incorporate large quantities into their cell membranes (Razin, 1975, 1982; Rottem, 2002).

The fermentative mycoplasmas generate ATP through glycolysis and the pyruvate dehydrogenase pathway, producing lactic acid. Non-fermentative species, such as *M. hominis* and *M. arthritidis*, and some fermentative species, produce ATP, ammonia, ornithine and CO_2 through the arginine dihydrolase pathway (Dybvig and Voelker, 1996; Prescott *et al*., 2002; Razin *et al*., 1998). Mycoplasmas have no cytochromes or quinones and lack complete tricarboxylic acid (TCA) cycle enzymes, therefore ATP is synthesised by substrate level phosphorylation where the electron transport chain is terminated by flavin (Pollack, 2002). This is a less effective mechanism than oxidative phosphorylation, an unfortunate consequence of gene reduction. In some mycoplasmas at least, the pentose phosphate pathway seems functional (Dybvig and Voelker, 1996; Prescott *et al*., 2002; Schuster *et al*., 2002).

2.1.4.4. ABC transporters

Due to the genome reduction and accompanied loss of genes, the mycoplasma is dependent on its host for a variety of nutrients and metabolites. Therefore, it is not surprising that a substantial number of genes in the mycoplasma genome encode for transport proteins. There are three types of membrane transport systems in mycoplasmas, namely the ATP-binding cassette (ABC) transport system, the phosphotransferase transport system, and facilitated diffusion by transmembrane proteins that function as specific carriers. The ABC transporter is involved in the import and export of a large number of substances including sugars, peptides, proteins and toxins, and due to its broad functionality, mycoplasmas depend the most on this transporter system (Razin *et al*., 1998). ABC transporters are found in prokaryotes and eukaryotes, although eukaryotes only have the ABC exporter and not importer. A possible reason for this could be that the function was attributed to a different organelle, due to the sub-cellular compartmentalisation of eukaryotes.

The ABC transporters can be differentiated into different classes, families and subfamilies, based on their phylogenetic analysis (Garmory and Titball, 2004). ABC transporters are one of the largest protein families in living organisms. In bacteria, the number of ABC transporters correlates with the genome size and the physiological niche of residence. This is due to their importance regarding the competitive growth of bacteria in these various niches (Nicolás *et al*., 2007). Thus, the combination of transporters in a given organism can allow one to predict its lifestyle (Ren and Paulsen, 2005). An understanding and in-depth knowledge of the transport proteins of an organism can greatly improve the understanding of its metabolic capabilities. Completely sequenced mycoplasma genomes have indicated that more than 50% of all membrane transporters are ATP-dependent, and of these, 59% to 92% are ABC transporters (Nicolás *et al*., 2007). The interest in ABC transporters increased after realising their potential to be targets in the development of antitumor agents, antibacterial vaccines and antimicrobials. Studies have demonstrated the ABC transporters' role in drug resistance due to its associated outer membrane proteins in Gram-negative bacteria, as well as direct or indirect roles in the virulence of the bacteria, for instance virulence associated with the uptake of nutrients and metal ions (Garmory and Titball, 2004).

2.1.4.5. The oligopeptide permease system of M. hominis as an example of an ABC transporter

The oligopeptide permease (Opp) system is one of the ABC transporters in bacteria and is responsible for the import of oligopeptides (Henrich *et al*., 1999). To discuss the structure and functionality of this system, the mycoplasma *M. hominis* will be used as an example. For a visual demonstration, refer to Figure 2.3. Four core domains (OppBCDF), together with a cytadherence-associated lipoprotein known as the P100 protein in *M. hominis*, comprise the Opp system. The species-specific P100 protein is a cysteine-anchored lipoprotein, expressed as a precursor polypeptide that functions as a substrate-binding site, also known as the OppA domain. The OppB and OppC subunits each contain six membrane-spanning α-helices, which together form the integral membrane-spanning domain for the transport of oligopeptides across the membrane. These two subunits possess the conserved hydrophobic motif that is characteristic to bacterial permeases: RTAKKGLXXXI/VZXXHZLR in the OppB domain and AAXXZGAXXXRXIFXHILP in the OppC domain. The OppD and OppF subunits form the peripheral ATPase domains: binding and hydrolysing ATP, to provide energy for the active transport of an oligopeptide (Henrich *et al.,* 1999; Hopfe and Henrich, 2004). The OppA domain of *M. hominis* contains the highly conserved Walker A and Walker B motifs, usually associated with the OppD and OppF ATP-binding domains. Thus it can be deduced that the OppA domain is also associated with ATP hydrolysis. As such, the OppA domain of *M. hominis* is sometimes described as the ecto-ATPase domain. However, due to the vague understanding of its specific function, a number of hypotheses arose to explain its physiological function. Firstly, it is possible that it hydrolyses extracellular ATP to protect the cell against the cytolytic effect of extracellular ATP (Redegeld *et al*., 1991). It might also regulate the ecto-kinase substrate concentration (Plesner, 1995). Another hypothesis speculates as to its involvement in signal transduction (Clifford *et al*., 1997; Margolis *et al*., 1990; Najjar *et al*., 1993), and lastly it could also be involved in cell-adhesion (Kirley, 1997; Knowles, 1995; Stout *et al*., 1995). Hopfe *et al.* (2011) recently demonstrated that the OppA-mediated cytadherence of *M. hominis* depends on the functionality of the ecto-ATPase activity. No mutants without this domain have been identified. This suggests that the domain does play an essential role in the survival of the organism (Hopfe and Henrich, 2004).

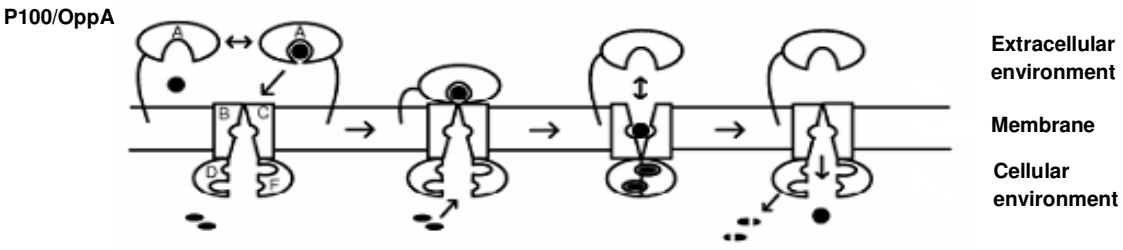

Figure 2.3 Schematic illustration of the basic components of the oligopeptide permease system of *M. hominis* (Doeven *et al*., 2008).

2.1.5. MYCOPLASMA PATHOGENICITY

Despite their small and rudimentary nature, many mycoplasmas cause adverse effects in their hosts. The virulence and pathogenicity of mycoplasmas is still poorly understood, but is sure to entail a complex and multi-factorial process (Kleven, 2008; Lockaby *et al.,* 1998). Different factors influence the pathogenicity of an

organism, such as virulence, tropism, number of organisms (dosage) and route of infection (Jordan, 1979). Pathogenesis may also result from the nutrient uptake and substrate sequestering (Garmory and Titball, 2004).

2.1.5.1. Distribution

Not all mycoplasmas are parasites. Some live in nature as saphrophytes, growing on and obtaining nutrients from decaying organic matter. Others are commensals, living in or on another organism without any effect on the host. The parasitic mycoplasmas have been isolated from a broad range of hosts (Razin and Freundt, 1984; Razin *et al.*, 1998). Mycoplasmas are quite host-specific, even organ- and tissue-specific, possibly a reflection on their nutritionally fastidious nature and parasitic lifestyle (Kleven, 1997; Razin *et al.*,1998). However, there are numerous reports of isolation from sites other than their normal specified niches and of mycoplasmas crossing species barriers. This demonstrates their ability to adapt to different environments (Dybvig and Voelker, 1996; Erickson *et al.*, 1988; Pitcher and Nicholas, 2005; Razin *et al.*, 1998; Tully, 1993). Despite these reports, mycoplasmas are still considered to be very specific to their hosts. They are found mostly in mucous surfaces such as the respiratory and urogenital tract, the eyes, alimentary canal, mammary glands, as well as joints (Razin *et al.*, 1998; Rocha and Blanchard, 2002).

In a study done by Botes *et al.* (2005b), the three ostrich mycoplasmas were found to be mostly located in the upper respiratory system. In addition to tissue specificity, the ostrich mycoplasmas are also seasonal, with a higher occurrence during the change of the seasons. Ms01 was detected at the start of the winter months and again at the start of summer. Ms02 and Ms03 were found to occur throughout the year, although Ms02 at much lower incidence. Ms03 had the highest occurrence, followed by Ms01. Of all the samples that tested positive, Ms02 occurred only at a frequency of 11.6%, with Ms03 the highest at 57.4%, not taking into account simultaneous mycoplasma infections (Botes *et al.*, 2005b).

2.1.5.2. Host cell attachment

Adherence of the mycoplasma to the host cell is the foremost and critical step in pathogenesis. It is a complex, multifactorial process that involves the adhesins of the pathogen, the host cell receptors, extracellular matrix components (EMC) and host cell molecules that act as a bridge between the pathogen and the host receptors (DeBey and Ross, 1994; Duensing *et al.*, 1999; Zhang *et al.*, 1994, 1995).

Most of the pathogenic mycoplasmas possess a specialised attachment tip structure that mediates attachment to and interaction with the host cell (Dybvig and Voelker, 1996; Henrich *et al.*, 1993; Razin *et al.*, 1998). The attachment and colonisation of mycoplasmas to the epithelial cells of the host is regarded as the first critical step in pathogenesis (Razin *et al.*, 1998), especially where the mycoplasma is confined to the mucosal surfaces of the host (Kleven, 2008). The mycoplasma cytadherence proteins are large integral membrane proteins with regions that are exposed on the cell surface (Dybvig and Voelker, 1996; Evans *et al.*, 2005; Henrich *et al.*, 1993; Razin *et al.*, 1998). Loss of cytadherence prevents colonisation, and consequently infection and disease. Therefore, membrane cytadherent proteins are deemed crucial regarding mycoplasma pathogenicity. One of the most abundant transporters in prokaryotes, the ABC transporter, is located on the outside of the cell membrane. Since the cell membrane of the mycoplasma is the outer barrier, ABC transporters may possibly play a role in

the pathogenicity of a mycoplasma and represent an immunogenic target to exploit in vaccine development (Henrich *et al.,* 1993; Lockaby *et al.,* 1998).

The attachment tip structure contains a network of interactive proteins, adhesins, and adherence-accessory proteins (Baseman, 1993; Baseman *et al*., 1996). It could also have a motile function: certain mycoplasmas can glide along plastic and glass surfaces if the tip is pointed in the right direction (refer to Motility, Section 2.1.3.7). Mycoplasmas that do not have this tip are still able to attach to the host cell (Dybvig and Voelker, 1996; Henrich *et al*., 1993). For instance, *M. hominis* and *M. fermentans* lack such a tip structure. However, *M. hominis* is able to adhere via two cytadhesins: the membrane proteins P50 and P100 (Henrich *et al*., 1993). *M. penetrans* has a tip structure containing a 65 kDa fibronectin-binding protein that is able to bind selectively to immobilised fibronectin, forming a molecular bridge (so to speak) between the mycoplasma and the host cell's surface proteins (Giron *et al*., 1996). This is followed by internalisation of the mycoplasma (Dimitrov *et al*., 1993; Duensing *et al*., 1999). After adherence, some mycoplasmas are able to enter the host cell. Refer to Intracellular location, Section 2.2.3.2 for more information.

Well-known cytadhesins include the cell surface-exposed P1 and P30 proteins of *M. pneumoniae* (Dybvig and Voelker, 1996), both clustered together at the attachment organelle (Krause, 1998). Without P1, *M. pneumoniae* would be unable to attach to its host and would thereby lose its virulence (Chaudhry *et al*., 2005; Krause, 1998; Krause and Balish, 2001; Razin and Jacobs, 1992; Razin *et al*., 1998). Other well-known adhesins include the MgPa adhesin of *M. genitalium:* an analogue of P1 with apparent similar roles; the GapA adhesin and cytadherence-related molecule CrmA of *M. gallisepticum*; and a P1-like adhesin in *M. pirum* (Papazisi *et al*., 2000).

Attachment is not only dependent on the mycoplasma, but also on receptors on the host cell surface. Razin *et al*. (1998) demonstrated the involvement of two different types of receptors in ligand-receptor interactions, namely sialoglycoproteins and sulfated glycolipids. Sialoglycoconjugates are the receptors for *M. pneumoniae*, *M. genitalium*, *M. gallisepticum*, and *M. synoviae* (Razin and Jacobs, 1992). Some mycoplasmas, such as *M. pneumoniae*, have more than one type of receptor (Razin, 1999; Rottem, 2003). Sometimes the receptors function as a directional infection point. For instance, at the primary site of *M. pneumoniae* infection, i.e. the optical micro-villar border and cilia, sialoglycoconjugates function as the receptors, while the secretory cells and mucus lack them. Thus attachment of *M. pneumoniae* to the ciliated cells is favoured (Loveless and Feizi, 1989). *M. hyopneumoniae*, in turn, binds to glycosaminoglycans, plasminogen, fibronectin, and glycolipid receptors on respiratory ciliated cells (Zhang *et al*., 1994).

In addition to receptors on the host cell surface, the basal membrane (BM) surrounding the host cells also present attractive targets for microorganisms to adhere to. The BM contains the EMC laminin and collagen, which are targeted by microorganisms to adhere to and invade the host tissue (Diacovich and Gorvel, 2010). Some microorganisms may then further damage the tissue with their secretory or membrane-bound proteases, and the damaged and exposed EMC present an expanded target for adhering microorganisms (Vanlaere and Libert, 2009).

2.1.5.3. Virulence causal factors

In addition to the cytadherence and internalisation of the mycoplasma, other factors may also contribute to the pathogenicity of mycoplasma infections. Possible factors include cell damage and disruption due to oxidative stress, concurrent infections and environmental factors. Mycoplasma infection often causes opportunistic infections to occur as well. Thus, in many cases, two or more simultaneous infections may exist within the same host, exacerbating the effect of one another and leading to aggravated infections and severe illness (Kleven, 1998). Concurrent infections caused by bacteria such as *E. coli, Pseudomonas aeruginosa*, *Avibacterium paragallinarum*, *Bordetella avium* and *Pasteurella* species have been associated with mycoplasmal clinical conditions in South African feedlot ostriches (Pretorius, 2009). As such, even mild mycoplasma infections pose a serious risk. Kleven (1998) also reported synergism between different mycoplasmas, such as *M. synoviae* with *M. gallisepticum* or *M. meleagridis*.

Respiratory diseases are known to be aggravated by environmental factors. Mycoplasma infections associated with respiratory disease often arise as a result of the temperature fluctuations that occur with seasonal change. Respiratory diseases are also exacerbated by stress (transportation and handling, change in feed, increase in population density) and environmental conditions (dust, high levels of ammonia, temperature, ventilation) (Allwright, 1995). These environmental factors greatly increase the severity of the infection (Kleven, 1998). Infections may lie dormant until any of these debilitating factors occur.

2.1.5.4. Interaction with the host organism

Mycoplasma infections are rarely of the fulminant type, but rather follow a chronic course (Razin, 1992). That these cell wall-less, fragile and simple organisms can resist the host immune response and cause chronic infection, seems surprising. Although the molecular basis of mycoplasma pathogenicity still remains largely unknown, the mycoplasma has developed numerous ways in which it is able to avoid the immune attacks of its host, leading to chronic infection and resistance to treatment. In addition to evasion, the clinical symptoms associated with mycoplasma diseases are often more indicative of damage to the immune and inflammatory responses of the host, rather than to the direct toxic effects of the mycoplasma cell components (Bradbury, 2005; Razin *et al*., 1998).

2.1.5.4.1. Competition for precursors

Mycoplasmas depend on their host to supply amino acids, fatty acids, co-factors and vitamins, competing with the host for these biological precursors (Pollack *et al*., 1997; Rottem and Barile, 1993). This presents a danger to the host cell's integrity and function, due to the depletion of these critical biochemicals (Baseman and Tully, 1997; Simecka *et al*., 1992). Non-fermentative mycoplasmas rapidly deplete the host's arginine reserves to generate ATP, adversely affecting the host's protein synthesis, cell division and growth (Rottem and Barile, 1993). Arginine utilisation by the mycoplasmas is suggested to inhibit histone synthesis and cause chromosomal damage of the host, due to the arginine-rich nature of histones. Certain strains cause further chromosomal abnormalities such as breakage, multiple translocations, reduction in chromosome number and possible new chromosome varieties (Rottem and Barile, 1993).

2.1.5.4.2. Damage due to cytadherence

Cytadherence of the mycoplasma to the host cell and fusion with its membrane may interfere with the membrane receptors and transport mechanisms of the host cell (Rottem, 2003). Adhering mycoplasmas also release toxic materials and the close proximity of the two cells may result in the hydrolysis of host cell phospholipids, caused by phospholipases present in many mycoplasma cell membranes (Shibata *et al*., 1995).

2.1.5.4.3. Damage due to fusion

Fusion implies the transference of the mycoplasma's cellular components to the host cell, such as hydrolytic enzymes (Razin *et al*., 1998; Shibata *et al*., 1995) and nucleases that may degrade the host cell's DNA (Paddenberg *et al*., 1996, 1998). Mycoplasma membrane compounds will be introduced to the host cell membrane, possibly altering receptor recognition sites and thus affecting induction and expression of cytokines, influencing cross-talk between the cells in the infected tissue (Rottem, 2003).

2.1.5.4.4. Damage due to internalisation

Internalisation of the mycoplasma into the host cell may disrupt its normal cellular function, causing cell damage, disruption and death. Cell damage is thought to occur due to the toxic effects of metabolic byproducts, such as hydrogen peroxide and superoxide radicals, as well as toxic components on the mycoplasma's cell surface (Lockaby *et al*., 1998). The oxidative stress damages the cell membrane and disrupts cellular function (Rottem and Naot, 1998; Rottem. 2003).

2.1.5.5. Clinical signs and symptoms

Mycoplasma infection in the South African ostrich is associated with respiratory disease in feedlot birds. Rhino-tracheatis (inflammation of the upper respiratory tract) and airsacculitis (an extension of nasal infection) are observed as a result of infection (Huchzermeyer, 1994). Other symptoms include nasal exudates (rhinitis), swollen sinuses (may be filled with thick yellow-white exudates), foamy eyes (conjunctivitis), rattling sounds of the throat due to post-nasal drip (sinusitis), shaking of the head and excessive swallowing. Laboured breathing may indicate air sac infection. The air sacs have a wet appearance and clear to white foamy exudates. Although growth is retarded, the ostrich's appetite generally remains unaffected, causing an increase in the cost of feeding (Allwright, 1995; Huchzermeyer, 1994). In the case of concurrent infections, the ostrich may exhibit increased pus in the sinuses and air sacs, fever, pneumonia and septic infection (Botes *et al*., 2005a).

2.1.5.6. Diagnosis

The variety of clinical symptoms, especially those that overlap, complicates the diagnosis of specific mycoplasmas. The clinical signs and lesions are not pathognomonic for any one mycoplasma, but can only be regarded as an indication of general mycoplasma infection (Jordan, 1990). Mycoplasma infections can be diagnosed using *in vitro* cultivation methods. Due to the different places of infection and maturity level, mycoplasmas can be isolated from different tissues. Tracheal swabs, and to a lesser extent, air sac -, sinus - or conjunctival swabs, are taken from the ostrich and introduced directly onto a suitable medium, either agar or broth. Different techniques are available to identify and confirm the presence of a mycoplasma, for example PCR-based methods and antibody-based procedures (Ley and Yoder, 1997).

2.1.5.7. Prevention

The first step in keeping a flock free of infection should always be prevention. Currently, the only way to prevent ostrich mycoplasma infections is with good farming and bio-security practises. However, this is complicated by the fact that ostrich carrier conditions exist without any clinical signs. Mycoplasmas have also developed ways of evading the host's immune system, further contributing to the difficulty of their eradication.

Bio-security programs are the most effective, but with the increase in population density of feedlot farming, this has become problematic to maintain (Kleven, 2008). Further methods of prevention for breeding stock are listed below:

i. Minimum contact between the host and the pathogen (Jordan, 1979).
ii. Hatching eggs can be treated to reduce transmission, by manual injection into the air sac or by dipping them in a suitable drug solution (Jordan, 1990, 1996; Ley and Yoder, 1997).
iii. Isolate the progeny in flocks of smaller groups, separated from the other flocks (Jordan, 1990; Ley and Yoder, 1997; Yoder, 1984).
iv. Frequent monitoring of the flock for infection and eradication of the whole group if a member shows signs of infection (Jordan, 1990).

Vaccination presents another method of prevention, but an important requirement for effective vaccination is that the flock has to be vaccinated before the field challenge occurs.

2.1.5.8. Treatment and control

Since mycoplasmas do not have a cell-wall, they are unaffected by antibiotics that target cell-wall synthesis, for example penicillin. They are, however, susceptible to antibiotics such as fluoroquinolones, macrolides and tetracyclines (Levisohn and Kleven, 2000). Fluoroquinolones target DNA gyrase, blocking DNA replication (Roberts, 1992). It kills bacteria quite rapidly, but when mycoplasmas are present at high concentrations, a decrease in killing is observed. Tetracyclines are broad-spectrum antibiotics, effective against almost all of the mycoplasma species, and work by binding to ribosomes and preventing their normal functioning.

Even though antibiotics are widely used to reduce the severity of mycoplasma diseases, it is still preferable to keep the flock disease free. Antibiotics such as tylosin, oxytetracycline, doxycycline and danofloxacin are used to control infection, but these antibiotics are not effective in eradicating the infection, they only manage the clinical symptoms (Dr. A. Olivier, in personal communication with B. Pretorius, 2009). Tylosin is a mycoplasma-specific antibiotic that can be administrated orally, either through the feed, dosed, or by injection (Botes *et al.*, 2005b). Unfortunately tylosin is one of the antibiotics that have been banned by the European Union since 1999, in addition to bacitracin, spiramycin, and virginiamycin (Casewell *et al.*, 2003). Antibiotics cannot be used for young birds and their usage should in general be kept to an absolute minimum (Huchzermeyer, 1998).

If used as a long term solution against microbial infections, resistance may build up in the pathogenic organism, and they predispose birds to fungal infections of the mouth and digestive tract (Perelman, 1999). Resistance against antibiotic treatment has led to more and more focus being placed on effective vaccine development.

There are no commercially available vaccines for ostrich inoculation and the need for such vaccines is imperative.

2.1.5.9. Current poultry vaccines

There are numerous live attenuated *M. gallisepticum* (MG) and *M. synoviae* (MS) vaccines commercially available that show a high degree of protection against the spread of MG or MS infection. There are also commercially available inactivated, oil-emulsified bacterins against MG and MS. Bacterins have been demonstrated to cause a consistent and reliable immune response (Droual *et al*., 1990). Oil-emulsified bacterins often cause tissue damage and cause abscesses to form under the skin, making it an unfeasible approach to follow for ostriches.

Van der Merwe (2006) demonstrated that MS and MG vaccines elicit an immune response in ostrich chicks, but protection against challenge was not assessed. Pretorius (2009) tested the hypothesis that MS and MG vaccines could provide protection against mycoplasma infections in ostriches. In chickens and turkeys, the MS and MG vaccines only provided protection against MS and MG respectively, indicating that cross-reactivity between these two mycoplasmas is too low to provide simultaneous protection. Pretorius (2009) noted the same result in ostriches, where the antibodies elicited by the MS and MG vaccines were specific to the MS and MG antigen and did not provide protection against the three ostrich specific mycoplasmas. This demonstrates the need to develop species-specific vaccines for each of the three ostrich mycoplasmas. Preliminary investigations into a DNA vaccine against Ms01 showed promising results, although these were compromised by co-infection with avian influenza (Brandt, 2012).

2.2. MYCOPLASMA VACCINE DEVELOPMENT

Different types of mycoplasmas vaccines exist, as well as different administrating routes, but none of these have been developed for ostrich-specific mycoplasmas. Due to the high antibody specificity towards individual mycoplasma species and the low cross-reactivity between different mycoplasmas, the need to develop a vaccine for each ostrich mycoplasma is imperative. To follow is an introduction to immunology, as well as a discussion of the immune responses elicited by mycoplasmas and their immune-response evasion mechanisms. This is followed by a series of discussions of the history of vaccines and their current applications. These discussions will focus on DNA vaccines: the immune responses that they elicit; their advantages; a vaccine development strategy, and its promise and application in ostrich-specific mycoplasma vaccine development.

2.2.1. INTRODUCTION TO VACCINES AND IMMUNOLOGY

Vaccines could possibly be regarded as one of the greatest achievements of medicine. The first illustration of the principle of vaccination was in 1798 when Edward Jenner discovered that prior exposure to cow pox could prevent later infection of small pox (Stern and Markel, 2005). In less than 200 years later, the World Health Organization launched a campaign that led to the complete eradication of the pox virus. Jenner's experiments formed the basis on which vaccines against almost all infectious agents were developed, many of which today have been successfully controlled or even eliminated altogether (Babiuk, 2002; Gurunathan *et al*., 2000; Oshop

et al., 2002). Despite the success, there are still many infections hazardous to human and animal health, where an effective vaccination is unattainable, mostly because of the need for cellular immunity. Traditional vaccines have been successful due to the extended time of antibody response, but only live attenuated vaccines provide the much-needed cellular response. Due to manufacturing and safety implications, the use of live attenuated organisms is restricted. However, these traditional methods provide insufficient protection against cell-invading pathogens such as *Plasmodium*, *Ehrlichia ruminantium,* and possibly mycoplasmas. These organisms require a comprehensive immune response in order to protect the host. The need for long lasting cellular immunity instigated the research on DNA vaccination, as DNA vaccines elicit both humoral and cellular immunity (Oshop *et al.*, 2002).

2.2.2. BASIC PRINCIPLES OF IMMUNOLOGY

2.2.2.1. Immunology and the immune response

To understand how a vaccine works, it is important to understand the immune system and its defence mechanisms. The immune system consists of innate and specific acquired immunity. Innate immunity refers to an existing immunity, not requiring prior exposure, which is present before the contact to the infectious agent. Due to the focus on vaccination, only specific acquired immunity will be discussed further. Specific acquired immunity requires contact with the live pathogenic organism or a vaccine against this organism. The immune response launched is a delayed response, but provides long term protection. Actively acquired immunity consists of two branches: antibody-mediated immunity (humoral (B cells) immunity) and cell-mediated immunity (cellular (T cells) immunity). The humoral branch of the immune system produces antibodies to protect the host against extracellular pathogens, while the cellular branch protects against intracellular pathogens by killing the infected cells (cytotoxic T cells) or helping the cells to recover (helper T cells). Different types of helper T (Th) cells have different effects on the immune response that is launched through the effect of secreted cytokines and lymphokines. Th-1 cells secrete lymphokines that activate macrophages, while Th-2 cells stimulate B cell development and activate cytotoxic T (Tc) cells (Rabson *et al.*, 2005).

Foreign molecules are recognised in the external environment by naive B-cells and dendritic cells. These antigen-presenting cells (APCs) have surface receptors that internalise proteins and digest them. Some of the resulting fragments are then exposed on the cell surface of the APCs the major histo-compatibility complex (MHC) class II. All cells of the body can process intracellular organisms such as viruses and present peptide fragments on MHC I. MHC class I presentation is recognised by the cytotoxic $CD8^+$ T cells, which proceed to attack the relevant infected cell (cellular immunity), killing it before the pathogen replicates. Class II is recognised by helper $CD4^+$ T cells that induce differentiation and proliferation of pre-B cells to produce antibodies, as well as release a variety of cytokines that in some cases activate macrophages, enabling them to kill intracellular parasites (Rabson *et al.*, 2005).

The pre-B cells mature into antibody-secreting cells, replicating and eventually resulting in the production of long-lived memory B cells (humoral immunity). The maturation of B-cells is important when looking at immunity; it is responsible for producing high-affinity antibodies and creating memory of the initial exposure. When the

memory B cell is then re-exposed to the antigen, it quickly transforms into an antibody-secreting plasma cell (Rabson *et al*., 2005).

2.2.2.2. Vaccination and the immune response

In the first response, B cells are activated by the interaction of the antibodies they express on the cell surface and other activation signals. The antigen binds to the antibody, activating a process that results in the formation of memory B cells, known as priming. When the host is exposed to the pathogen for a second time, known as the secondary or memory response, the response is far quicker and greater in magnitude. This is the basis for the success of vaccination: by priming the immune response, vaccination presents a harmless method to prepare the host for eventual infection (Rabson *et al*., 2005).

Vaccination works on the principle that exogenous antigens provided by the inactivated or killed pathogens, or protein from live vaccines, are taken up by the APCs and presented on the MHC class II, stimulating $CD4^+$ T cells. The MHC class I is used when antigens that are produced in the cell's cytoplasm are presented, a response generally elicited by live or DNA vaccines (Rabson *et al*., 2005). Immunologically, live vaccines provide the most effective protection through their wide range of effects, but from a safety point of view, the use of live vaccines raises many concerns (Babiuk, 2002). This emphasises the advantage of DNA vaccines: eliciting both types of immunity in the same manner as live vaccines, but without the associated risks.

2.2.2.3. Mucosal immunity

The surface of the respiratory system, eyes, intestinal and urinary tract are exposed to the external pathogen-containing environment. These tissues are covered by epithelial cells that provide a protective barrier against this environment (Jeurissen *et al*., 1989). The first line of defence is non-specific mechanisms (epithelial cells, mucus, and extreme pH) that prevent pathogens from adhering to the surface (Neutra *et al*., 1996). The epithelial cells contain lymphoid tissue that is able to respond specifically to invading pathogens (Jeurissen *et al*., 1989). Mucosal immunisation has two major advantages: ease of administration and induction of local immunity. The mucosal surfaces are the target tissues of the ostrich-specific mycoplasmas, demonstrated by the fact that the majority of the mycoplasmas were isolated from the ostriches' trachea (Botes *et al*., 2005b). Mucosal surfaces represent the primary portal of entry for animal pathogens (Oshop, 2002). Inducing local mucosal immunity is of particular importance in the case of respiratory pathogen infections.

2.2.2.4. Immune responses elicited by mycoplasmas

Mycoplasma infections cause specific responses in the host, including production of local and systemic antibodies and stimulation of cellular phagocytosis of the organism. Unfortunately, as necessary as these protective mechanisms are, they also play a role in exacerbating issue damage and mycoplasma-induced illnesses. This contributes to the mycoplasma's pathogenicity and enables it to avoid the host's immune defences, leading to chronic infection (Razin, 1992b; Razin *et al*., 1998).

Immune effects caused by mycoplasma infections include either suppression or stimulation of B and T lymphocytes, induction of cytokine production, cytotoxicity of macrophages, natural killer cells (NKC) and T cells, as well as enhanced expression of cell receptors and complement activation. Mycoplasmas are able to avoid

phagocytosis, but still interact with mononuclear and polymorphonuclear phagocytes to stimulate the synthesis of pro-inflammatory cytokines (Rottem and Naot, 1998; Salman *et al*., 1994). Mycoplasmas can have a synergistic or antagonistic influence on immune cells (Rottem, 2003). Different mycoplasmas are able to stimulate either B cells or T cells, or both. They can be potent stimulators of T cell-derived cytokines interleukin (IL)-2, interferon (IFN)-α and IL-4. Mycoplasmas have an effect on phagocytes and lymphocytes, thus influencing the balance between Th1 and Th2 populations of $CD4^+$ T cells, and also enhancing NKC activity of the innate immune system (Razin *et al*., 1998; Rottem and Naot, 1998).

Although the precise mechanism by which mycoplasmas elicit immune responses is not completely understood, it is accepted that the transfection and presentation by dentritic cells (DC) plays a significant role in the development of immunity brought about by DNA vaccines (Casares *et al*., 1997; Porgador *et al*., 1998; Condon *et al*., 1996). DC presentation may also occur by uptake of secreted protein from other transfected cells, or uptake of the antigen, known as cross-presentation (Corr *et al*., 1996).

2.2.3. MYCOPLASMA EVASIVE TECHNIQUES

2.2.3.1. Lipoproteins and antigenic variation

Antigenic variation contributes to the pathogenicity of mycoplasmas by enabling the mycoplasma to evade the host immunity and adapt to the environment. By generating a versatile surface coat that constantly changes via high frequency phenotypic switching, the antigenic profile is modified too often for the host to recognise it and synthesise antibodies (Dybvig and Voelker, 1996). Antigenic variation is thought to occur primarily in the lipoproteins. Bacterial lipoproteins are potent modulators of the immune system and lipoproteins represent the preferential target of the humoral immune response in mycoplasma infections (Chambaud *et al*., 1999). Studies with *M. gallisepticum, M. mycoides* subsp. *mycoides* and *M. bovis* have demonstrated that lipoproteins are the most active immunogens in experimental chicken and cattle infections (Abdo *et al*., 1998; Behrens *et al*., 1996; Jan *et al*., 1995). The P53 lipoprotein of *M. penetrans* is the preferentially recognised antigen of the host's humoral immune system (Grau *et al*., 1995; Neyrolles *et al*., 1999). Mycoplasmas have an unusually high number of lipoproteins with unknown function, often expressing several at once. Lipoproteins account for a significant portion of the membrane-protein mass (Chambaud *et al*., 1999). The lipoproteins are fixed in the cell membrane via a lipid moiety or hydrophobic amino acids, with part of the protein exposed to the outer surface of the cell. Mycoplasmas are able to modify the expression of these immunogenic surface lipoproteins and so avoid antibody attacks (Tryon and Baseman, 1992).

Two methods of lipoprotein variation in mycoplasmas have been described. In the first method, the number of tandem repeats at the carboxy terminal is varied, creating lipoproteins of different sizes, and in the second method, phase variation is employed, which entails the reversible and inheritable switching between two phenotypes. Certain gene families contain multiple copies of a gene, and by varying expression of these alternative genes, different antigenic moieties are encoded (Dybvig and Voelker, 1996; Kleven, 1998). Another advantage of phase variation entails masking the epitope accessibility of another surface antigen. Known as differential masking, this increases the possible variation even further (Theiss *et al*., 1996). Genes that encode

surface antigens are often found to contain repetitive elements, which further contribute to antigenic variation, based on sequence variation (Dybvig and Voelker, 1996; Razin *et al*.,1998).

It should be noted that although the repertoire of lipoprotein variation is vast, it does not necessarily change during infection. A long-term study on *M. hominis* infected patients showed no antigenic variation on the mycoplasma isolates over time in the same patient (Jensen *et al*., 1998). However, when isolates from different patients were compared, a high degree of antigenic variation was observed. The study concluded that antigenic variation in *M. hominis* is a result of the specific host, rather than a reaction to the development of an immune response. Bergonier *et al*. (1996) demonstrated that geographical differences may also account for lipoprotein variation. They found different epitopes between *M. agalactiae* isolates from Europe and those from Africa. Taking all this into account, surface variation of mycoplasmas can be considered more of a general mechanism that allows the organism to adapt to its fluctuating environment. This is further supported by the fact that microorganisms that are not exposed to host immune systems also have surface variation (Deitsch *et al*., 1997).

2.2.3.2. Intracellular location

Although most mycoplasmas are non-invasive surface pathogens, species such as *M. fermentans*, *M. genitalium*, *M. hominis* and *M. penetrans* have the ability to infiltrate and survive inside their host's cells (Evans *et al*., 2005; Razin *et al*., 1998; Rottem, 2003). *M. penetrans* and *M. genitalium* both contain tip structures that enable penetration into the host cell, while other mycoplasmas, such *as M. fermentans* and *M. hominis,* enter via internalisation, where the mycoplasma is engulfed by the cell membrane. *M. penetrans* localises in the cytoplasm and perinuclear regions (Borovsky *et al*., 1998; Giron *et al*., 1996). *M. fermentans, M. pneumoniae*, *M. genitalium* and *M. gallisepticum* are surface pathogens, but under certain circumstances are able to reside in non-phagocytic cells (Rottem, 2003). Inside the cell the mycoplasmas can exist freely or inside a vesicle. The intracellular location of mycoplasmas not only protects them against the humoral immune response of the host, but also against antibiotic exposure. This contributes to the difficulty with which mycoplasmas are eradicated, thus promoting chronic infection (Kleven, 2008; Razin *et al.,* 1998). The process of invasion is based on the ability of bacteria to bind to fibronectin (adhesive glycoprotein) or sulfated polysaccharides (Duensing *et al*., 1999; Dziewanowska *et al*., 1999). Some mycoplasmas have fusogenic activity due to the absence of a cell wall, as well as the close proximity of the mycoplasma membrane and the cytoplasmic membrane of the host. Rottem and Naot (1998) showed that mycoplasmas that require unesterified cholesterol for growth have fusogenic activity. To illustrate the importance of unesterified cholesterol in fusogenic activity: *M. capricolum* adapted to survive in the absence of cholesterol, and with the dramatic reduction of cholesterol in its membrane, the organism lost its fusogenic activity (Tarshis *et al*., 1993). The ability of mycoplasmas to multiply within the host cell has not been proven completely (Rottem, 2003).

2.2.3.3. Immunomodulins

Henderson *et al*. (1996) proposed the term 'modulin' to describe molecular moieties that induce cytokine production with pathological consequences. Mycoplasmas produce at least three of the 16 classes of modulins currently listed: lipoproteins and lipids (Razin *et al*., 1998; Rottem and Naot, 1998), as well as a polypeptide super-antigen only produced in *M. arthritidis* (Cole *et al*., 1996). The role of lipoproteins in antigenic variation

has been discussed (refer to Lipoproteins and antigenic variation, Section 2.2.3.1), but lipoproteins also stimulate monocytes and induce the secretion of pro-inflammatory cytokines (Rawadi and Roman-Roman, 1996; Razin *et al*., 1998). All membrane-anchored lipoproteins contain a lipoylated amino-terminal cysteinyl residue, which is responsible for the modulin activity (Brenner *et al*., 1997). In some cases this residue can be *N*-acylated. Lack of this terminal acylation of specific mycoplasma lipoproteins, such as the 2 kDa macrophage activating lipopeptide 2 (MALP-2) from *M. fermentans*, stimulates macrophages considerably more than triacylated lipoproteins (Mühlradt *et al*., 1997, 1998). Interestingly, mycoplasma infections are not necessarily associated with a strong inflammatory response, demonstrated by the fact that some mycoplasmas colonise respiratory and urogenital tracts without any clinical symptoms. Indeed, it has been shown that lipoproteins of *M. fermentans* induce IL-10 in isolated human monocytes, which is one of the cytokines that down-regulates the inflammatory response (Rawadi *et al*., 1996). There is still much to learn about the nature and kinetics of cytokine production following mycoplasmas infection, but it appears that cross-talk between cytokines determines the intensity of inflammatory reactions (Chambaud *et al*., 1999).

2.2.3.4. Molecular mimicry and phenotypic plasticity

Another immune evasion tactic that hinders the host's recognition of mycoplasmas involves molecular mimicry and phenotypic plasticity. Some antigenic epitopes are shared between different mycoplasmas and their host cells, referred to as molecular mimicry. In addition to evasion of the host immune system, auto-antibodies have been observed during certain mycoplasma infections (Rottem, 2003). Phenotypic plasticity refers to the ability of a single genotype to produce more than one morphology, physiological state and/or behaviour, either due to environmental signals or random changes in the expression profile of single or multiple genes (Rottem, 2003). Surface compounds such as flagella, pilli, outer membranes or capsules, are major immune targets. Lacking most of these, mycoplasmas' major source of antigenic variation falls on membrane proteins (refer to Lipoproteins and antigenic variation, Section 2.2.3.1).

2.2.4. VACCINES

The ideal vaccine has many specific characteristics and requirements. Firstly, a vaccine must be immunogenic, which is characterised by Ada (1994) as:

i. activation of APCs to stimulate antigen processing and interleukin production,
ii. activation of B and T cells to yield memory cells,
iii. generation of Th and Tc cells to several epitopes, thus overcoming MHC polymorphism in populations,
iv. and continuous presence of antibodies.

Furthermore, other factors that should be noted when developing a vaccine are cost-effectiveness, safety, ease of manufacturing and administration. In the case of a live vaccine, it should be genetically and phenotypically stable and should not cause disease, nor should it spread to neighbouring populations and cause disease there. The vaccine itself should not cause any toxic reactions or negative immune effects (Ellison *et al*., 1992). Ideally, a single dose would provide life-long immunity (Whithear, 1996).

2.2.4.1. Conventional approaches

Conventional vaccines are known as first generation vaccines. First generation vaccines refer to whole-organism vaccines, which can either be killed (chemical or heat inactivation), attenuated (live organism with disabled virulence) or a close relative of the pathogenic strain without the virulence. Live attenuated vaccines are able to elicit Tc and Th cell responses, as well as antibody immunity. Live vaccines have the advantage of providing long-lasting cellular and humoral immunity, due to their close resemblance to natural infection. However, the possibility of reverting back to its virulent form and causing disease in the immune-compromised inoculated host does remain a concern and one of the biggest disadvantages.

Killed vaccines, also called bacterins, provide a high and extended level of immunity. In comparison with live vaccines, however, they have a lower immunogenic effect and are unable to generate specific Tc cell responses (Lechmann and Liang, 2000). Bacterins are non-infectious, an advantage over live vaccines, since there is no risk of reverting back to virulence or risk of cross-infection. Certain disadvantages make this a non-ideal vaccine approach for ostriches. Firstly, the vaccine is quite expensive and the birds need to be vaccinated individually. In addition, oil emulsion adjuvants cause large abscesses and granulomas beneath the skin in ostriches. With reference to mycoplasmas, synergistic effects may occur between the mycoplasma and another infectious agent, possibly causing several diseases, especially if the bird is vulnerable due to environmental and physiological stress (Whithear, 1996).

2.2.4.2. Novel approaches

The availability of bioinformatic tools enabled *in silico* analysis of genome sequences and identification of active proteins, leading to more innovative approaches to vaccine development. Possible antigenic proteins can be identified, regardless of their abundance, and the genes encoding biologically active proteins can be cloned, expressed and analysed under controlled conditions. Vaccine development that starts with *in silico* analysis of the genomic information, in contrast with the conventional approach that starts by cultivating the pathogen, is termed reverse vaccinology (Rappuoli, 2000). These novel and innovative approaches include conjugate and recombinant approaches. Conjugate vaccines contain a bacterial capsular polysaccharide, generally a poor antigen, joined to a carrier protein to enhance the immunogenicity. Some immunogenic compounds, such as the surface polysaccharides of encapsulated bacteria, are immunogenic but not able to induce immune memory. Protein antigens are conjugated to these components to improve elicitation of the T cell response. The immunological properties of the carrier proteins are thereby conferred to the attached antigen. The technique is mostly applied to provide protection against invasive bacterial disease.

There are three categories of recombinant vectors, as discussed below (Ellis, 1999).

i. Live genetically modified organisms

   Live genetically modified organisms have one or more genes deleted, usually a double knock-out to prevent regression (Uzzau *et al*., 2005). The vaccine can also be an organism that is transfected with a gene from the pathogen, thus functioning as a vaccine vector. This requires knowledge of the pathogenic organism's genes and care should be taken that the gene(s) or resultant protien(s) do not harm the viability or immunogenicity of the organism.

ii. Recombinant inactivated vaccines

Recombinant inactivated vaccines include subunit vaccines (containing part of a whole organism), synthetic peptides (representing the most basic part of the immunogenic protein), virus-like particles (which look like a virus on the outside but do not contain genetic material and thus cannot replicate) and whole proteins extracted or expressed from cloned genes. Subunit vaccines are injected with an adjuvant; a substance that stimulates the immune system, enhances response by protecting the vaccine from degradation, attracts APCs (which process and present the antigen to the immune system), stimulates immune cells directly, stimulates cytokine secretion, or any combination of these. These are known as second generation vaccines. Second generation vaccines are able to elicit the humoral and helper T cell responses, but not cytotoxic T cell responses. They overcome the risks and problems associated with first generation vaccines; however, they too have limitations. *In vivo* immunogenic proteins are not necessarily protective antigens. They may also be too variable in their coding sequence, difficult to express, and have high production costs. In addition, only the most abundant antigens can be identified and as such there are many diseases for which vaccines cannot be developed (Mora *et al*., 2003).

iii. Genetic vaccines

Genetic vaccines are known as the third generation vaccines and are currently one of the largest growing fields in vaccine development. A DNA vaccine consists of a plasmid vector, carrying a foreign gene that encodes for an immunogenic (or pathogenic) protein. The plasmids are multiplied and maintained in a bacterial host, usually *E. coli*, before it is purified and injected directly into the host as 'naked DNA'. The host will then express the proteins and an immune response is launched against the foreign protein.

2.2.4.3. Comparison of the essential steps of the conventional vaccine approach and reverse vaccinology

Conventional vaccine development requires the pathogen to be cultivated on a large scale. The steps involved are cumbersome and time-consuming: first the pathogen is grown in laboratory conditions to search for protective antigens using biochemical, serological and microbial methods; then it needs to be mass produced. This approach works well for organisms with abundant antigens, but fails to provide a solution for pathogens that do not have obvious immunodominant protective antigens. With reverse vaccinology, all that is required is the genome sequence. Computer analyses are used to predict possible antigens without relying on their expression profiles or on the cultivation of the pathogen. The concept of DNA vaccines will now be discussed in greater detail, focussing on the development strategy.

2.2.4.4. DNA vaccines

The concept of DNA vaccination were originally illustrated by Wolff *et al*. (1990), when they inoculated mice with a plasmid that expressed the human growth hormone. Instead of causing growth as they predicted, antibodies were produced against the hormone. Thus they demonstrated that direct immunisation with naked DNA resulted in the *in vivo* expression of the protein within the muscle cells of mice. Their unexpected results sparked interest in the use of DNA plasmids that encode for an antigenic protein as a means of inducing immune responses, a principle known as DNA immunisation.

Today, DNA immunisation is employed in a wide range of animal models to provide antibody and cell-mediated protection against different bacterial, fungal and parasitic diseases (Dhama *et al.*, 2008; Donnelly *et al.*, 1997). Traditional vaccines make use of whole organisms, either killed, attenuated or a close non-pathogenic relative of the pathogen, whereas DNA vaccines circumvent the associated problems by employing a gene encoding for the immunogenic protein of the pathogen. The DNA vaccine provides the gene(s) for encoding the antigen, rather than consisting of the antigen itself. Recombinant DNA technology and bioinformatic tools are used to develop and construct a DNA vaccine. A typical DNA vaccine consists of a bacterial plasmid backbone that provides the frame for inserting the gene(s) of interest (GOI). Since the plasmid does not contain an origin of replication that is active in eukaryotic cells, it will not be able to replicate itself or integrate into the chromosomal DNA of the host (Garmory *et al.*, 2003; Webster and Robinson, 1997).

2.2.4.4.1. Principle and mechanism of DNA vaccination

In order to develop a DNA vaccine against the selected pathogen, a gene or genes must be identified that will possibly evoke an immune response. For a schematic illustration of the process of DNA vaccine development, refer to Figure 2.4 (Dufour, 2001). Sequence analysis and gene comparison between the genomes of different pathogenic organisms may be used to identify immunogenic proteins. When an appropriate gene is identified, it is then isolated and cloned into a suitable eukaryotic expression vector. The vector has to contain a bacterial origin of replication to allow mass production of the plasmid in a bacterial host, such as *E. coli*. After large-scale production and purification, the plasmid (DNA vaccine) can be inoculated directly into the host animal, usually by intramuscular (IM) injection. The host cell will then internalise the plasmid and produce the encoded protein. Following the process of antigen production and processing, the foreign protein or derived peptides are suggested to be presented on both MHC class I and MHC class II molecules of local APCs, activating $CD8^+$ T cells and $CD4^+$ T cells, inducing both humoral and cellular immunity (Gurunathan *et al.*, 2000).

The precise mechanism by which the antigen is produced, processed and presented, is less clear. The two hypotheses entail firstly that the *in vivo* uptake of DNA occurs through specific receptors (Bennet *et al.*, 1985), and secondly, rather non-specifically, similar to phago- or pinocytosis (Lewis and Babiuk, 1999). The amount of antigen produced is only up to the nanogram range, but despite the low quantity, a broad, sustained immune response is induced. Presentation of the foreign peptides mimics natural infection, because they are presented in their native form, similar to the presentation that occurs during natural infection. The basic immune responses are thought to occur similarly in avian hosts as in mammals, although unique organs in the avian lymphatic system, such as the bursa of Fabricius and the Harderian gland, may account for some of the differences in the early immune responses to DNA vaccination (Oshop *et al.*, 2002).

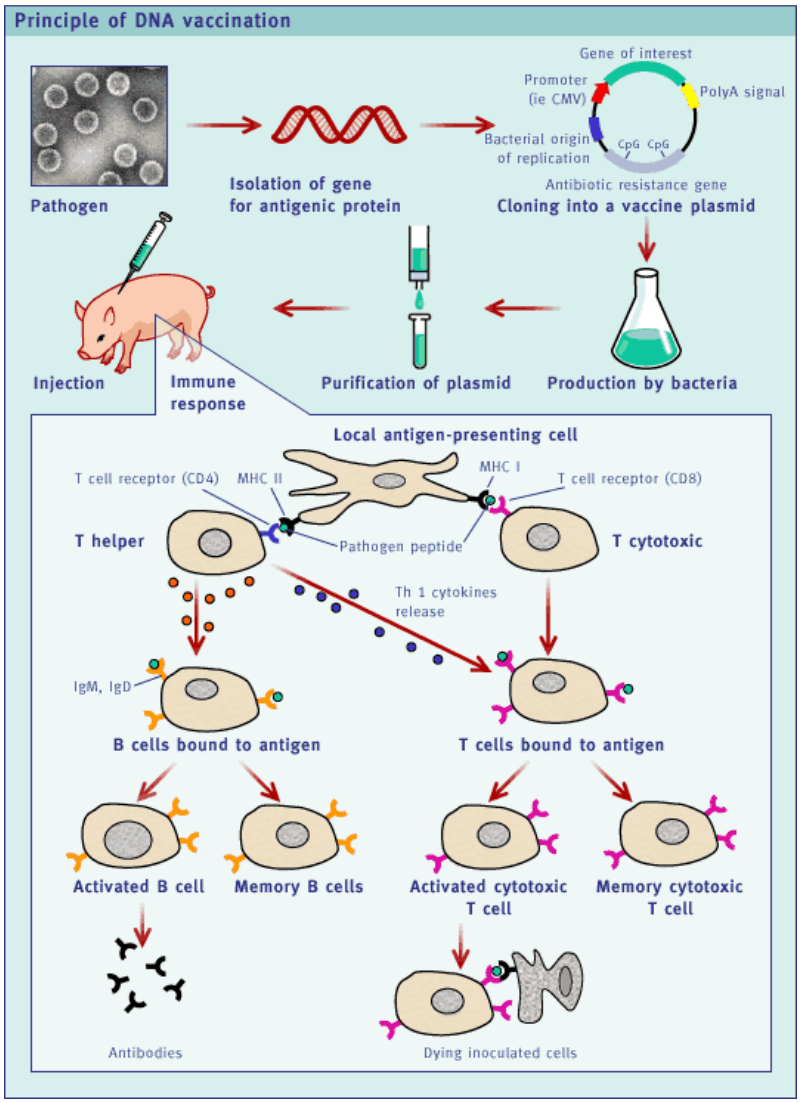

Figure 2.4 The principle of DNA vaccination (Dufour, 2001). A candidate gene is identified and isolated from the pathogenic organism's DNA. The gene is cloned into a eukaryotic DNA expression vector. After large-scale production and purification, the DNA vaccine is then delivered directly into the host animal. The host animal will produce, process and present the antigenic peptides to the immune system on MHC I and MHC II molecules, inducing both cellular and humoral immune responses.

2.2.4.4.2. Immune responses elicited by DNA vaccination

Different features will have an effect on the efficacy of the DNA vaccine, including the choice of expression vector, the method of DNA delivery, and even the DNA itself. Certain bacterial DNA motifs, such as unmethylated cytidine-phosphate-guanosine (CpG) motifs, have been shown to be immunogenic, and polynucleotides are known for their ability to induce cytokines (Krieg *et al.*, 1994). These motifs activate monocytes, NKC, as well as DC and B cells, by acting as an adjuvant and enhancing the immune response when added into the plasmid (Weeratna *et al.*, 1998). The presence of a well-known consensus sequence, known as the Kozak sequence, influences the recognition of the start codon by eukaryotic ribosomes, thus influencing translation efficiency. The method of delivery, as well as the delivery site, will influence the type of

immune response launched: for example Th1 helper cells are activated following IM injection, but when DNA is delivered by the gene-gun method, the emphasis shifts to Th2 helper cell responses (Donnelly *et al*., 1997). The reason for this is that with needle injection the plasmid DNA is taken up by dendritic cells in the extracellular space, proceeding to differentiate for Th1 cytokine production, while gene-gun delivery bypasses this mechanism with direct entry into the cell (Jakob *et al*., 1998).

The immune response raised by a DNA vaccine can be divided into three different sections:

i. Humoral immune responses
   Antibody responses induced by DNA vaccination have been demonstrated for a wide array of proteins, in some cases detectable years afterwards, thereby providing a continued antigen stimulus. Humoral immune responses were launched against various viral, bacterial, parasitic, tumour and eukaryotic proteins (Donnelly *et al*., 1997).

ii. Cytotoxic T cell responses
   MHC class I-restricted $CD8^+$ cytotoxic T lymphocyte (CTL) responses have been demonstrated in mice that have been inoculated previously. Antigens and viral infections have been used to challenge the mice lymph node or spleen cells both *in vitro* and *in vivo*. In some cases, the CTL responses persisted for up to 15 weeks after the final immunisation (Donnelly *et al*., 1997).

iii. Helper T cell responses
   Following antigen activation, the helper T cells secrete cytokines, signalling macrophages and B and T lymphocytes to facilitate the differentiation and development of memory cells. Mice, humans, chickens, ferrets and many more animal models have been shown to launch helper T cell responses to a wide variety of antigens. The comprehensive results suggest that the generation of helper T cells may be a general characteristic of DNA vaccines (Donnelly *et al*., 1997).

These responses constitute one of the many advantages of DNA vaccination: its ability to evoke both branches of the immune system, the cellular and humoral response.

2.2.4.4.3. Administration / delivery methods

Many different methods of delivery into the animal tissue are available, namely: IM, intradermal (ID), subcutaneous (SC), ocular, intrabursally (IB), intravenous (IV), intraperitoneal (IP), oral, vaginal, intranasal (IN), or by gene-gun delivery to the skin. The route of administration of the vaccine is believed to influence the cells involved with protein production. For example, following IM injection, the plasmid can occur in the monocytes or in the muscle cells (Babiuk *et al*., 2000), while for intra-dermal injection, it is believed that the skin is important in generating an immune response (Bos, 1997). Babiuk *et al*. (2003) established that gene-gun delivery delivers the DNA directly into the cell and appears to be more efficient in transfecting Langerhans cells or dendritic APCs. The skin is known to contain these cells, which can induce MHC class II responses, thus inducing strong $CD4^+$ T cell responses (Akbari *et al*., 1999). Recently several new methods have been explored to increase the transfection efficiency of DNA vaccines, such as high-pressure delivery, dermal patches, forming DNA in particles such as liposomes, and electroporation. A variety of enhancing agents can be co-administrated with the DNA vaccine, such as sucrose, cardiotoxin, bupivacaine, hypertonic solutions or adjuvants such as calcium

phosphate, lipids, CpG oligonucleotides and cytokines (Oshop *et al*., 2002). Most often IM and ID is the preferred delivery choice, often formulated without any adjuvants, due to their simplicity and effectiveness (Oshop *et al*., 2002). Other methods include aerosol delivery onto mucosal surfaces and topical administration to the eye and vaginal mucosa (Lewis and Babiuk, 1999).

Another possibility of DNA vaccine delivery is the use of carriers. DNA vaccines can be delivered in attenuated intracellular bacteria such as *Salmonella*, *Shigella* and *Listeria* (Du and Wang, 2005; Fairweather *et al*., 1990; Yang *et al*., 2009). Non-viable strains are employed as carriers of the DNA vaccine plasmid and administrated orally. Success has been achieved against Newcastle Disease virus, hepatitis B virus and human immunodeficiency virus (HIV). Fagan *et al*. (2001) demonstrated oral immune responses in pigs against a number of genes of *M. hyopneumoniae*. Numerous other studies demonstrated oral immunity when recombinant plasmids transformed in *Salmonella typhimurium* aroA SL3261 were used as a delivery system (Chen *et al*., 2003, 2006, 2008; Fagan *et al*., 2001). The results indicated that using *Salmonella* carriers as a way of inoculation should be considered as a means of evoking immune responses in ostriches against the ostrich mycoplasmas. It is suspected that the three ostrich-specific mycoplasmas might be combated with mucosal immunity, as would be provided with oral administration. Eliciting mucosal immunity may prevent infection altogether, instead of preventing disease after infection. Another advantage of using a carrier as delivery method is that it provides protection against degradation of the DNA by extracellular deoxyribonucleases (Gurunathan *et al*., 2000).

2.2.4.4.4. The effect of the route of delivery on dosage requirements

The dosage is dependent on the particular antigen and model system that is used, as well as the method of administration. Intramuscular or subcutaneous injection usually requires 100 - 200 µg of plasmid DNA, while gene-gun immunisation requires 250 to 2 500 times less DNA to induce antibody or CTL responses (Fynan *et al*., 1993). Fynan *et al*. (1993) demonstrated that gene-gun immunisation achieved similar protection with up to 2 500 times less DNA than used with saline DNA injections. As a dose example, IM/ID immunisation of mice requires between 1 and 100 µg DNA, while monkeys and calves require between 10 µg and 1 mg to provide the same level of immune response. If mice are inoculated using the gene-gun delivery system, the amount falls to between 10 ng and 10 µg (Robinson and Pertmer, 2000). Some are of the opinion that the dose required is independent of the host animal size (Cox *et al*., 1993), even though others argue, as an example, that mice require a tenth of the DNA than primates do (Robinson and Pertmer, 2000). In poultry, a single ballistic injection delivery required from 0.25 µg DNA to induce a measurable immune response (Kodihalli *et al*., 1997), while an IM injection required at least 750 µg (Triyatni *et al*., 1998). Different doses resulted in a clear difference in immune responses (Kodihalli *et al*., 1997; Song *et al*., 2000; Suarez and Schultz-Cherry, 2000).

2.2.4.5. Advantages of DNA vaccination

DNA vaccines have several advantages over the use of conventional vaccines, including mimicry of live vaccination, but without the possibility of reverting back to virulence and causing disease. DNA vaccines induce cytotoxic T cell responses, an advantage over protein vaccines, which elicit a poor T cell response when immunised without adjuvants (Donnelly *et al*., 1997). Although live vaccines also generate both types of immune

responses, they have the disadvantage of having a risk of infection if reverting back to their virulent form. DNA vaccines do not pose this security risk (Gurunathan *et al*., 2000; Robinson and Torres, 1997). Recombinant-based vaccines are costly to produce and difficult to purify, they could cause incorrect folding of the antigen and poor induction of $CD8^+$ T cells, all of which DNA vaccines do not suffer from. It is also possible to direct the type of immune response, either toward a Th1 cell or Th2 cell response (Robinson and Pertmer, 2000). With DNA vaccination concurrent administering is possible, the plasmid is genetically stable and the immune response can be modified by manipulation of the genetic sequence (Oshop *et al*., 2002; Webster, 1998).

DNA vaccines obviate the need for peptide synthesis, expression and purification of recombinant proteins, and eliminate the need for toxic adjuvants (Sedegah *et al*., 1994). The need for pathogen cultivation is circumvented, making it possible to develop vaccines against difficult-to-cultivate pathogens, as well as for pathogens that are dangerous to work with. This makes DNA vaccination a feasible approach for mycoplasma vaccine development, due to the latter's notoriously difficult cultivation properties. DNA vaccines also present a tool to identify protective antigens without knowing the function of the protein.

2.2.4.6. Limitations of DNA vaccination

Concerns regarding DNA vaccines include the possibility of tolerance to the antigen produced, host genome integration, induction of auto-immunity and anti-DNA antibodies, as well as the risk of affecting the genes that control cell-growth (Oshop *et al*., 2002; Webster, 1998). A limitation of the DNA vaccine approach is the inability to detect and therefore utilise any non-protein antigens. For instance, bacterial polysaccharides represent potent vaccine targets, due to their extracellular exposure, and are an important component of many successful vaccines. CD1-restricted antigens such as glycolipids may be also present promising vaccine candidate targets, but will not be detected using the reverse vaccinology approach (Rappuoli, 2001). Another limitation is the possibility of atypical processing of bacterial and parasitic proteins, since the expression occurs in a eukaryotic organism (Robinson and Pertmer, 2000).

2.2.4.7. Safety of DNA vaccines

In contrast with conventional vaccines, DNA vaccines are stable. Originally it was feared that the plasmid could potentially integrate into the host's genome, and possibly lead to the development of anti-DNA immune responses. The DNA vaccine is non-live, non-replicating and non-spreading, adding to their list of advantages. Research into these safety concerns found little evidence of genomic integration, with a significantly lower risk for integration than the risk associated with natural mutations (Ledwith *et al*., 2000; Sheets *et al*., 2006; Wang *et al*., 2004). Numerous studies and clinical trials were done to monitor for possible anti-DNA responses, but no evidence was found (Tavel *et al*., 2007). It seems that the platform of DNA vaccination is safe and well-tolerated in hosts (Ferraro *et al*., 2011).

There are already licensed DNA vaccines available for commercial use (Table 2.1), such as for the West Nile virus that causes encephalitis in horses, infectious haematopoietic necrosis virus that affects salmonid fish, and melanoma in dogs (Bergman *et al*., 2006).

Table 2.1 Currently licensed DNA therapies (Kutzler and Weiner, 2008).

Vaccine target	Product name	Date licensed and country	Target organisms
West Nile virus	West Nile Innovator	2005, USA	Horses
Infectious haematopoietic necrosis virus	Apex-IHN	2005, Canada	Salmon
Growth hormone releasing hormone	Life-Tide-SW5	2007, Australia	Swine and food animals
Melanoma	Canine Melanoma Vaccine	2007, USA, conditional license	Dogs

2.2.4.8. Candidate genes for DNA vaccine development

Vaccine candidate genes are genes encoding for proteins with good immunogenic properties, such as the proteins involved in pathogen attachment and interaction with the host, as well as those associated with the virulence of the pathogen (Henrich *et al*. 1993). The membrane proteins of mycoplasmas that mediate the adherence to the host cells, as well as ABC transporters that are essential for nutrient uptake, are thus considered candidate genes (Garmory and Titball, 2004). Proteins that are secreted or located on the outer membrane surface are more accessible to the antibodies produced by the host than those inside the cell.

In order to identify such a gene, the genome sequence of the pathogen is required. Publicly available genetic databases, such as GenBank, provide genetic information on an ever-increasing list of organisms. Partial or whole genomes are annotated and can be used in similarity searches to identify possible antigenic genes. Similarity searches were applied on the whole genome, using bioinformatic tools, to identify the *oppA* gene as a vaccine candidate gene for Ms01. The decision was based on the high homology of the putative protein with the *M. hominis* P100 protein, also coded by the *oppA* gene. In a similar manner, Steenmans (2010) identified the Ms02 *oppA* gene as a candidate gene for an Ms02 DNA vaccine.

This study will reduce the 28 contiguous sequences of the Ms02 genome, produced in a previous round of genome sequencing, through linkage of the contigs using a technique called **T**hermal **A**symmetric **I**nterlaced **P**olymerase **C**hain **R**eaction (TAIL-PCR). This, together with bioinformatic analysis on the Ms02 genome, such as repetitive sequence analysis, will be described in Chapter 3.

Chapter 3

This chapter is presented as a short review on TAIL-PCR, the technique used to assemble the Ms02 genome partially, and on mycoplasma genomes in general: their repetitive sequences; how this influences the assembly of mycoplasma genomes; the order and arrangement of genes; and also gene distribution on the leading and lagging strands of the genome. The experimental procedures and results specific to Ms02 follow thereafter in an article format.

3.1. INVESTIGATION INTO MYCOPLASMA GENOME REPEATS, GENE ARRANGEMENT, ASSEMBLY AND ANNOTATION

3.1.1. INTRODUCTION

Genome sequencing methods have improved rapidly in recent years, reaching a point where genome analysis and annotation cannot keep up with the amount of newly released sequences. In order to understand and use the raw information contained in the genome sequence, it is necessary to identify and annotate genes, the proteins they code for, and the metabolic pathways the proteins form part of. Understanding the genetic information contained in the genomes of bacteria could lead to the development of novel vaccines, identification of drug targets to develop antimicrobial compounds, and modification of the bacteria for bioremediation applications (Barker, 2006; Binnewies *et al*., 2006; Stothard and Wishart, 2006).

In order to develop a DNA vaccine, it is necessary to have the DNA sequence of a vaccine candidate gene of the pathogenic organism. Advances in molecular genetics and bioinformatics resulted in a large number of genomes that have either been sequenced completely or for which contiguous sequences exist. The genome sequence of an organism provides insight about possible proteins, without having to clone and express every hypothetical gene of interest. Comparisons between different genomes can lead to the identification of conserved domains, thus identifying vaccine candidate genes through *in silico* analysis of the genome, circumventing laborious and time-consuming experiments. Identification of essential genes in mycoplasmas not only contributed to the discovery of the minimal set of genes required for self-replicating life, but also aid in drug development, because essential genes represent attractive targets for developing antibiotics and DNA vaccines. Identification of essential genes will also enable quicker annotation of unknown genome sequences by using homology searches and genome comparative studies (Lin and Zhang, 2011). The comparative genomic approach is thus limited by the availability of annotated, complete genomes of close and distantly related sets of species.

3.1.2. THERMAL ASYMMETRIC INTERLACED POLYMERASE CHAIN REACTION

In 1995, Liu and Whittier developed a simple, powerful technique to recover unknown genome sequences adjacent to known sequences. Based on the strategy that primers of different lengths and base pair compositions will have different annealing temperatures, **T**hermal **A**symmetric **I**nter**l**aced **P**olymerase **C**hain **R**eaction (TAIL-PCR) employs cycles with interlaced high and low temperatures, thus favouring both long and

short primers to function with near equal annealing efficiency. Three long, specific, nested primers are designed on the known DNA sequence, and the DNA is amplified in three consecutive PCRs with a non-specific, shorter, arbitrary degenerate (AD) primer. Because of the difference in annealing temperatures, the relative amplification of products can be controlled to favour the target product over the non-specific products. Performing three consecutive TAIL-PCRs dilutes non-specific products to negligible concentrations so that predominantly only the target product remains (Figure 3.1).

In the primary reaction three types of product may form. Firstly, the target product primed by the specific forward and degenerate reverse primer will form (type I). A detectable amount of product amplified by the specific primer alone will also form (type II), but is usually diluted out by the secondary reaction due to the nested effect of the specific primers. Due to the indiscriminate nature of the degenerate primers, fragments amplified on both sides by these primers will also form (type III) and constitutes the major source of background, since they cannot be diluted out by the nested specific primers during normal PCR cycling. The strategy of TAIL-PCR is designed to favour the amplification of the target product (type I) and suppress formation of type III product. An initial low stringency cycle creates the binding sites for the AD primer by facilitating initial base miss-match annealing. In the following high stringency cycle only the long specific primer can effectively anneal, and type I product is amplified, while little or no type III product will form. A reduced stringency cycle then ensures annealing of both primers, replicating the single stranded DNA product produced in the high stringency cycle. Thus the template of type I is greatly increased. Cycling of the high and reduced stringency cycles are then used to increase the target product, while suppressing the formation of non-specific products. Unfortunately, the high stringency cycle also favours mispriming of the specific primer, and the product primed on both ends by the specific primer will also be amplified with even higher efficiency. However, due to the use of nested primers, the effect of such mispriming is diluted out in the second reaction. The third reaction usually does not require further TAIL-cycling and only serves to amplify the target sequence.

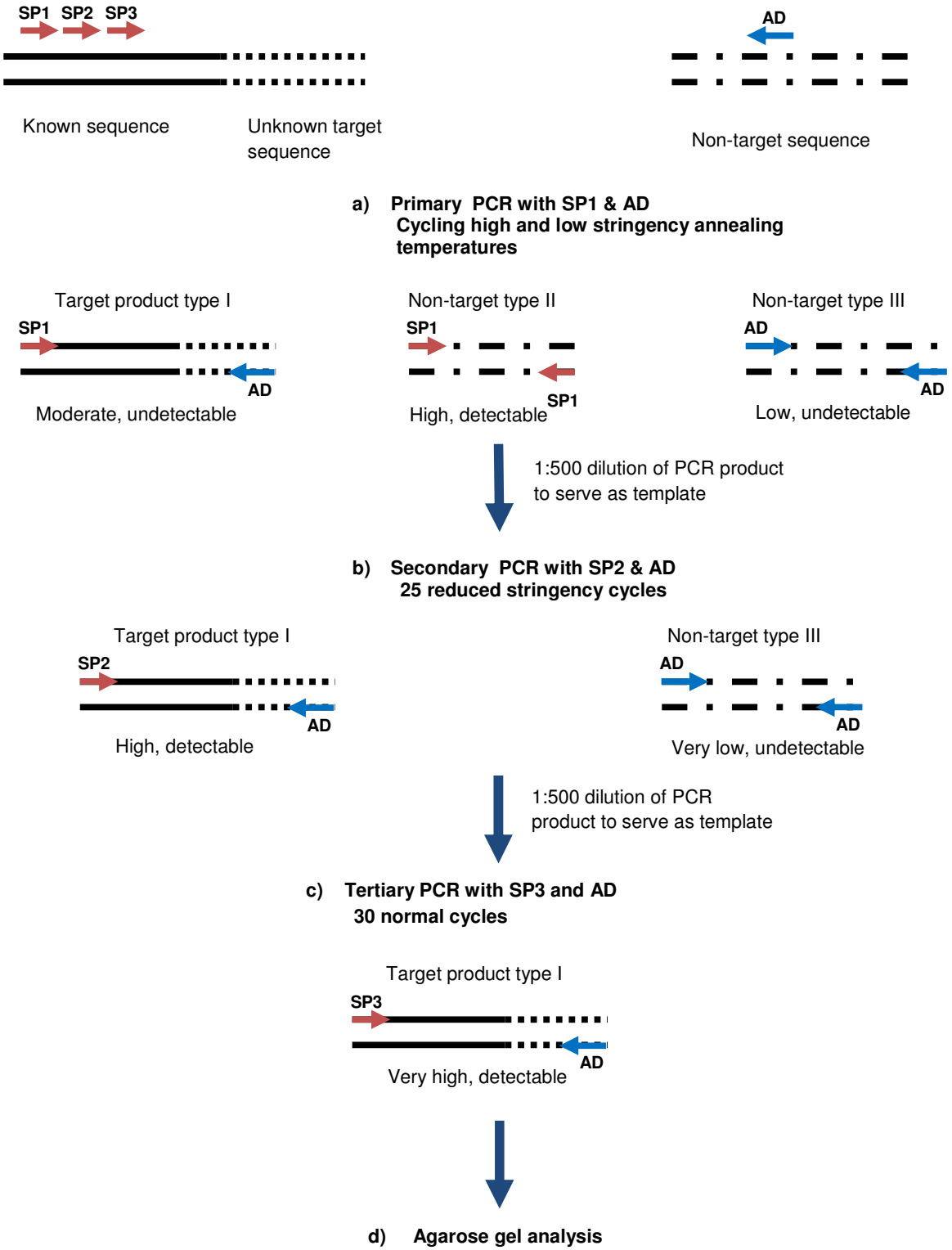

Figure 3.1 Schematic diagram of TAIL-PCR indicating the preferential amplification of the target sequence over non-target sequences. (a) indicates the primary PCR, (b) the secondary PCR, and (c) the tertiary PCR. Analyses of the PCR products are performed with (d) agarose gel analysis, a simple and cost-effective method. Specific primers are denoted by SP and arbitrary degenerate primers denoted by AD.

3.1.2.1. Advantages of TAIL-PCR

TAIL-PCR has proven effective in a wide range of genomes, including the hexaploid wheat genome, a genome about five times larger than the human genome (Liu and Whittier, 1995). TAIL-PCR is a method that can be used for chromosome walking, genome physical mapping, developing sequence-tagged sites, serial gene-walking, as well as analysis of genomic sequences that flank T-DNA, transposon or retrovirus insertions (Liu and Whittier, 1995). It is especially useful when dealing with a large number of samples. The advantages include:

i. Simplicity: TAIL-PCR does not require complex DNA manipulations, either prior or post PCR. The resultant DNA can be analysed easily on an agarose gel. After a simple purification step, the DNA can be used directly in sequencing reactions.
ii. Sensitivity: Only nanogram quantities of DNA are required and crude DNA or cell lysate can be used. Thus TAIL-PCR is an effective, sensitive method to use when only a small amount of template is available.
iii. Specificity: Due to the cycling process and consecutive PCR reactions, the amount of non-specific product is so low that the TAIL-PCR products can be used as a sequencing template or as hybridisation probes without any complicated preparation steps.
iv. Efficiency: Using four different AD primers, it is almost certain to obtain a specific product with very little background. Product sizes range from 100 to 2 000 bp.
v. Speed: The three reactions, together with agarose gel analysis, can all be completed in one day. Due to the direct sequencing of the template, the process from amplification to sequencing data is streamlined.

3.1.2.2. Limitations of TAIL-PCR

Although TAIL-PCR is easy to optimise and generally has a high success rate, it is dependent upon the quality of the specific primers and the nature of the genome. The specific primer needs to bind to a unique site in the target sequence of the genome. In certain cases, such as genomes with high numbers of repeats, one specific product is not amplified, but rather an array of different products. The high background prevents purification and sequencing of a specific product.

Genomes that contain multiple runs, or that are A+T-rich such as mycoplasmas, make the design of optimal specific primers difficult. It is generally difficult to obtain a long specific primer that adheres to the primer guidelines, i.e. no secondary structures or base pair runs, and a certain T_M (melting temperature), to prevent non-specific binding. AD primers with too high a degree of degeneracy will also result in too many small undesirable products. TAIL-PCR has proved inefficient when applied to complex genomes such as rice, with lower efficiencies and amplification products that rarely exceeded 600 bp (Nakazaki *et al.*, 2000; Sato *et al.*, 2001).

3.1.3. REPETITIVE SEQUENCES IN MYCOPLASMA GENOMES

Repetitive elements are most often found in eukaryotes and to a lesser extent in prokaryotes, but the prevalence of these elements are unexpected in genomes as small as those of mycoplasmas. Repetitive elements present a problem in sequence alignments and assembly by creating ambiguities, which in turn produces biases and

errors when the results are interpreted. While it is tempting to just ignore repeats, this leads to further problems and may cause important biological information to be missed. If the repeats are longer than the read lengths generated by the sequencing method, gaps will be created during *de novo* assembly (Treangen and Salzberg, 2012). Worse than gaps, the repeats may be collapsed on top of one another, resulting in a lower representation of repeats than what actually occurs in the genome, as well as misassembly (Treangen and Salzberg, 2012). The main problem with repeats is that the assembler cannot recognise and distinguish between them, leading to a common error where two chromosomal regions are joined that do not belong together. At first glance, the partially assembled genome would seem almost complete with very few gaps, but when attempts are made to complete the assembly using chromosome walking techniques, or bioinformatic analysis is performed on the partial genome, it is very quickly realised that the presence of repeats caused missassembly and complicated further down-stream processes.

3.1.3.1. Simple sequence repeats

Simple sequence repeats (SSRs) refer to perfect, uninterrupted tandem repeats of a single nucleotide or a short oligomer in a DNA sequence, also called microsatellites, and have been found in all eukaryotic and prokaryotic genomes that have been studied (Katti *et al.*, 2001). They are powerful genetic markers that have become useful tools for high density genetic mapping, molecular tagging of genes, phenotype mapping and genetic diversity studies (Jewell *et al.*, 2006). These repeats are caused by DNA polymerase slippage during replication, causing insertion or deletion of base pairs and are therefore quite diverse with respect to their lengths (Kashi and King, 2006; Toth *et al.*, 2000). This is seen as a common source of phase variation in bacteria and antigenic variation in pathogens, aiding the pathogenic organism in evasion of the host immune system, by increasing the genetic diversity of the population (Groisman and Casadesus, 2005; Rocha, 2003; Van der Woude and Baumler, 2004).

3.1.3.1.1. SSRs and their function and gene associations

Long SSRs are generally rare in prokaryotes and mostly restricted to host-adapted pathogens that do not readily survive outside of their host (Mrázek *et al.*, 2007). Mrázek (2006) investigated SSRs in mycoplasma genomes and found that they are not just involved in phase variation, but also in the organisation of the chromosome, as well as protein structure and function. Depending on their length, sequence, and location, various functions have been attributed to SSRs in the genome. This include involvement in gene expression, gene regulation and function, as well as significance even when present in non-coding regions (Gupta *et al.*, 1994; Kashi *et al.*, 1997). They provide recombination hot-spots, facilitate genome rearrangements, have been found to bind nuclear proteins, and function as transcriptional activating elements (Li *et al.*, 2002). They may also affect protein structure and thus protein-protein interactions (Mrázek, 2006). Some of the repeats act as contingency loci that are associated with surface antigens. Obligate parasites benefit from this due to the mutations that aid in avoidance of the host immune response, or in general the adaptive mutations promote the survival of the microbes that live in rapidly changing environments (Moxon *et al.*, 1994).

Analysis of over 300 prokaryotic species showed a large variance in the SSR content (Mrázek *et al.*, 2007). Trivedi (2010) concluded that the highest amount of SSRs in the 13 mycoplasmas she studied, are found in

methylase genes, followed by DNA polymerase, exonuclease and topoisomerase genes. Guo and Mrázek (2008) detected SSRs that are associated with housekeeping genes such as ribosomal protein genes, RNA genes and chaperones. This is somewhat surprising, since these genes are essential for the survival of the organism, and inactivation due to DNA polymerase slippage and recombination could be lethal.

M. capricolum and *M. mycoides* contain two SSR-associated genes for ABC transporters (Guo and Mrázek, 2008). Both the genes in *M. mycoides* have a 14 bp run of T-nucleotides (T_n) in the upstream flanking regions. Taking into account that components of mycoplasma ABC transporters can be recognised as antigens, this is consistent with the role of SSRs in antigenic variation. Some SSRs in *M. capricolum* encode for putative membrane proteins that include three ABC transporters, five genes of the *vmc* cluster, and several proteins of the Lpp family. The *vmc* cluster contains TA_n repeats in the promoter regions, thus possibly influencing expression of these phase variable genes (Wise *et al*., 2006). In *M. pulmonis* there are five SSR-associated genes encoding DNA methylases, which might indirectly contribute to antigenic variation by changing the methylation patterns in distant chromosome regions and influencing the expression levels of other genes (Rocha and Blanchard, 2002). Approximately 16% of *M. gallisepticum*'s genome seems to be a reservoir of sequences that exists for the variation of a single haemagglutinin antigen (Baseggio *et al*., 1996).

3.1.3.1.2. SSRs and immune evasion

One of the mechanisms of adaptive immune evasion in *M. pulmonis* concerns phase variation in the Vsa proteins. These are surface-bound lipoproteins that are both phase and size variable (Denison *et al*., 2005). A tandem repeat region of 40 or more repeat units confers resistance to complement lysis of the mycoplasma, while those with a repeat of 5 or fewer are readily killed when dispersed, but able to form a bio-film. This led to the development of the shield hypothesis: the Vsa protein with many repeats may shield the surface of the mycoplasma and prevent the complement membrane attack complex from inserting into the mycoplasma membrane (Simmons *et al*., 2007). Shaw *et al.* (2012) showed the VsaA protein to be the major protein involved in this phenomenon. Irrespective of the Vsa isotype, mycoplasmas that produce a long VsaA protein are able to resist binding to macrophages (in this case specifically alveolar macrophages), and once bound, are then able to resist being killed.

The above-mentioned findings were also observed in other mycoplasmas. Proteins with extensive repeats that are similar to the Vsa proteins were found in *M. hominis*, *M. agalactiae*, *M. arthritidis*, *M. bovis* and *M. hyorhinis* (Citti *et al*., 1997; Flitman-Tene *et al*., 2003; Glew *et al*., 2002; Jensen *et al*., 1995; Sachse *et al*., 2000; Washburn *et al*., 1998; Yogev *et al*., 1991). In *M. hyorhinis*, the tandem repeats of Vlp proteins protect the mycoplasma from serum antibodies (Citti *et al*., 1997). *M. hominis* produces a Lmp1 protein containing tandem repeats, which has been shown to protect the organism from antibodies (Jensen *et al*., 1995).

3.1.3.1.3. SSRs and antibiotic resistance

It is possible that SSRs aid the mycoplasma in evading or resisting antibiotics. Guo and Mrázek (2008) have reported antigenic variations in host-adapted pathogens that are facilitated by SSRs. This is supported by the increasing resistance of prokaryotes, including mycoplasmas, to the broad-spectrum fluoroquinolones that target proteins in the topoisomerase II family, DNA gyrase, and topoisomerase IV. The resistance of some prokaryotes

is mediated by mutations that occurred in the target regions of these enzymes (Bébéar *et al*., 1998). As such, enzymes that target two or more proteins or genes would be more effective, since the odds of a mutation occurring in both of these targeted proteins/genes would be rare, as shown by studies on *Staphylococcus aureus* (Fournier *et al*., 2000). Antibiotics that target ribosomes have also proved ineffective against certain mutants (Pereyre *et al*., 2002). SSRs in the rRNA genes may provide the genetic variability that enables the mycoplasma to survive in the altered environment, living with the antibiotics that target their ribosomes (Guo and Mrázek, 2008).

The focus thus far has been on the ability of SSRs to enable the mycoplasma to avoid attacks by its host, as well as avoid the workings of antibiotics, therefore it should be noted that the pathogenic mycoplasma *M. hyopneumoniae* have fewer repeats when compared to other mycoplasmas. How this mycoplasma then evades the host immune system and is able to lead to chronic infection even in the absence of the help of these repeats, remains an open question (Minion *et al*., 2004). Even the repeats that do exist, mainly A_n and T_n repeats as discovered between three different strains, may rarely be involved with genome rearrangements (Mrázek, 2006).

3.1.3.2. Repetitive elements in different mycoplasma species

A large number of close repeats were found in *M. pneumoniae*. Ten of these have well-conserved motifs, and most of these repeats are associated with lipoproteins, UFOs or RMSs. Complete analysis of the *M. pneumoniae* genome revealed multiple copies of four large repetitive elements, designated RepMP2/3, RepMP4 and RepMP5, with homology to large repeats also found in *M. gallisepticum* and *M. genitalium* (Musatovova *et al*., 2012). In contrast with this, the repeat sequence designated RepMP1 seems to be *M. pneumoniae* specific, with no homologues found in other sequenced mycoplasma species, and with numerous copies in its own genome (Musatovova *et al*., 2012). These repeats vary in size from 0.3 kb to 2.2 kb, and illustrate the extent to which repeats are present in bacterial genomes.

M. pulmonis contains three sets of large repeats, corresponding to sets of lipoproteins, restriction and modification systems (RMSs) and unknown function open reading frames (UFOs). In addition, 33% of the direct or inverse two-copy repeats were found within UFOs, followed by 25% in lipoproteins, and 11% in intergenic positions (Rocha and Blanchard, 2002). The most relevant large repeats found in *M. genitalium* represent multiple copies of fragments of the gene in the *mgpABC* operon that encode for an adhesin. Only one of *M. genitalium*'s lipoproteins has been shown to contain a clearly defined close repeat (Rocha and Blanchard, 2002). This might be due to its small genome, it being the smallest of the mycoplasmas identified thus far.

Parts of the P1, P40, P90 and P110 proteins in *M. pneumoniae* and *M. genitalium* are subjected to homologous recombination due to repetitive sequences such as RepMP and MgPar (Sluijter *et al*., 2009). *M. synoviae* contains amino acid repeats in its haemaglutinins that may affect the antigenic response. Haemaglutinins are highly immunogenic, variably expressed, surface proteins (Bencina, 2002). Repetitive sequences (DR-1 and DR-2) in the putative cytadhesin gene *pvpA* of *M. gallisepticum*, consisting of 30 or more proline residue repeats and 7-10 repeats of the tetrapeptide motif, affect the functionality of the gene (Jiang *et al*., 2009). *M. bovis* contains insertion sequences, IS3, IS4, and IS30, whereas *M. hyopneumoniae*, *M. agalactiae* and *M. mycoides*

subsp. *mycoides* contain variable number tandem repeats (VNTRs), which are associated with coding sequences and may provide genetic diversity. The insertion sequences and VNTRs possibly code for amino acid residues that affect interaction with the host immune system, as well as host cell attachments (de Castro *et al.*, 2006; McAuliffe *et al.*, 2007, McAuliffe *et al.*, 2008). Tandem amino acid repeats in the genes of *M. pneumoniae* that contain the RepMP1 repeat, and Vsa proteins of *M. pulmonis,* may have certain regulatory functions (Musatovova *et al.*, 2008; Simmons and Dybvig, 2007). Almost 9% of the genes of a Californian strain of *M. capricolum* subsp. *capricolum* contain repeats, just under *M. pulmonis* with 9.3% of its genes, and above *M. synoviae* with 8.3% of its genes (Trivedi, 2010).

3.1.4. PROKARYOTIC GENE ARRANGEMENT

Mycoplasma genomes exhibit a poor conservation of gene order. Exceptions are the more recently diverged *M. pneumoniae* and *M. genitalium*, whose gene order tends to be more conserved (Glass *et al.*, 2000). In general, enzymatic pathways are the most conserved and are often organised in gene clusters to better regulate and optimise protein production. An example is the arginine deiminase pathway (ADI) of *M. penetrans*, which consists of three enzymes and a specific membrane-associated arginine permease that is organised in a gene cluster. The most conserved pathway is that of glycolysis, followed by the genes involved with DNA replication, recombination, repair, and ribosomal proteins, all conserved within clusters (Souza *et al.*, 2007). Essential genes are generally clustered together, such as those involved with nucleotide transport and metabolism (Souza *et al.*, 2007). Gene order and localisation tend to become fixed over time and occur in such an order and position to best support the fitness of the organism (Képès *et al.*, 2012).

3.1.5. STRAND BIAS, BACTERIAL ORIGIN OF REPLICATION AND RELATED SURROUNDING GENES

It is generally accepted that bacterial chromosomes have a bias of G_n towards their leading strands, leading to the so-called G+C skew, but for some genomes, including many Firmicutes, there is also a weaker A+T strand bias, with the leading strand containing more T_n than A_n (Rocha *et al.*, 2000). G+C content has been related to numerous properties, such as genome size, oxygen and nitrogen exposure, and specific habitats. On that note, intracellular bacteria on average have smaller genomes and are more A+T-rich (Wassenaar *et al.*, 2009). The higher A+T content in small, intracellular bacteria has been attributed to a loss of repair genes, leading to an increase in the mutation rates from cytosine to thymine (Moran, 2002; Rocha and Danchin, 2002).

Bacterial chromosome replication starts from a single origin, the origin of replication (*ori*). Most bacteria show G+C and A+T skews around the origin of replication. However, two mycoplasma species were shown to be C+T-rich in the third codon position (McLean *et al.*, 1998). Uneven distribution of nucleotides on the leading and lagging strands cannot always be reduced to a nucleotide skew. Circular chromosomes tend to have the *ori* and terminus of replication on exact opposite sides, which evens out the G+C skew to the extent that the bias is filtered out to near zero, resulting in the number of G_n being equal on both strands (Bohlin *et al.*, 2010). Replication mechanisms may influence the strand's symmetry patterns, and as such, a lot can be learned by the examination thereof. However, horizontally transferred genes may also influence the strand symmetry, leading to changes in the polarity of skew plots (Xia, 2012).

The direction of the nucleotide skew switches at the origin of replication, so that the leading strand contains more G_n than C_n. However, in *M. genitalium* and *M. pneumoniae*, the direction of the skew was found to be opposite to the direction for codon position 3 (McLean *et al.*, 1998). Their genes are also arranged in such a way resulting in the transcriptional and replicational directions to be the same, a tendency that is more pronounced for highly expressed genes such as ribosomal protein genes (Brewer and Fangman, 1988). According to Karlin *et al.* (2003) most bacterial genomes contain a large cluster close to the *ori* that contains between 20% and 40% of all ribosomal protein genes. Many bacterial genomes' rRNA operons are in the same direction as replication (Price *et al.*, 2005; Rocha, 2002). Other genes that are involved in protein synthesis are encoded within or proximal to this large cluster, such as the *tuf*, *fus*, *rpoA*, *rpoB* and *rpoC* genes.

All of the replication origins that have been experimentally confirmed occur in intergenic regions (Gao and Zhang, 2008; Mackiewicz *et al.*, 2004). The position of the *ori* is conserved among bacterial species, mostly found in close proximity to the *dnaA* gene. The *dnaA* gene encodes for the universally conserved DNA replication initiation protein DnaA (Ogasawara *et al.*, 1991). The gene cluster *rnpA-rpmH-dnaA-dnaN-recF-gyrB-gyrA* that surrounds the *ori* and the *dnaA* gene is often conserved, with the *ori* located adjacent to the *dnaA* gene in an intergenic region (Briggs *et al.*, 2012). The *ori* of *M. genitalium* and *M. pneumoniae* reside in an A+T-rich area between the *dnaA* and *dnaN* genes. The low G+C content Firmicutes generally have two unique essential genes, the *dnaD* and *dnaB* genes, in addition to the *dnaA* gene (Leonard and Grimwade, 2011; Soultanas, 2011). Several Mollicutes have only a single gene, the *dnaD*-like gene, combining both of the functions of the *dnaD* and *dnaB* genes (Briggs *et al.*, 2012). Genomes with more A_n than T_n were shown to contain both a *polC* and *dnaE* homologue, while the other genomes lack the *polC* homologue (Worning *et al.*, 2006). The direction of the A+T skew is determined by the DNA polymerase-α subunit that replicates the leading strand (Worning *et al.*, 2006).

3.1.6. DEFINITION OF ESSENTIAL GENES AND THE MINIMAL GENOME

Essential genes (EG) form one of the three groups of functional genes, with highly expressed genes (HEG) and horizontally transferred genes (HGT) being the other two, that play an important role in the survival and infectious ability of pathogens (Gao and Chen, 2010). EG represent the minimal gene set that is required for sustaining life of the simplest, self-reproducing bacteria. For some researchers they represent the gene set absolutely required for cell viability, under any conditions. Others use EG to refer to the gene set that contributes to the fitness of the organism and its competitive growth under favourable conditions, for instance in the absence of environmental stress and in the presence of essential nutrients (Mushegian and Koonin, 1996a). Most of the metabolic genes' essentiality is relative to the metabolic context or specific conditions (Gerdes *et al.*, 2003; Koonin, 2003). These genes therefore represent the part of the bacterial genome that one would suspect to be relatively constant, since a change in the expression profile or change in the phase variability would cause functional differences and lead to a loss of essentiality. Some of the known functional genes are drug targets, therefore further identification and investigation into important EG might provide further drug targets, lead to development of broad-spectrum antibiotics, and aid in developing species-specific antibiotics (Gao and Chen, 2010). It will further enable quicker annotation of unknown genome sequences through comparative studies and homology searches (Lin and Zhang, 2011).

One would expect EG to be conserved, as is the case in bacteria (Gerdes *et al*., 2003; Jordan *et al*., 2002). EG are more conserved than non-EG, however, conserved genes are not necessarily essential (Fang *et al*., 2005). In addition to this, the fact that certain genes are conserved among bacteria, does not necessarily imply that the genes' essentiality is also conserved, i.e. extrapolation of essentiality is not possible among bacteria (Zalacain *et al*., 2003).

3.1.7. GENOME STRAND GENE DISTRIBUTION AND REPLICATION

Mycoplasmas indicate extreme strand distribution, with 80% of genes occurring on the leading strand, compared to the more evenly spread 55% of *E. coli* (Rocha and Blanchard, 2002). Due to the nature of replication, where the leading strand is continuously replicated while the lagging strand is replicated in fragments, bacterial chromosomes exhibit strong strand-specific bias, both in terms of oligonucleotide sequence and gene orientation (Rocha, 2004). For Firmicutes, the PolC polymerase and the *dnaE* gene are responsible for replication in the leading and lagging strand respectively (Dervyn *et al*., 2001; le Chatelier *et al*., 2004). In addition, for Firmicutes in general, the polymerase that replicates the leading strand with proof-reading capacity resides in one polypeptide chain, while the proof-reading capabilities resides in a complex of two polypeptides for the lagging strand.

Necşulea and Lobry (2007) found that all rRNA genes and most of the ribosomal protein coding genes are transcribed more often than not in the same direction as DNA replication, this being the leading strand. The same is true for most of the EG, such as those involved with transport and metabolism. Using this information, one would expect to be able to predict the leading and lagging strands in an incomplete and partially annotated genome. An annotated contiguous sequence can be studied to determine if most of the genes, especially EG, occur in one direction, thus predicted to be the leading strand.

3.1.8. GENOME ANNOTATION

Annotation refers to the extraction of biological information from raw nucleotide sequences (Médigue and Moszer, 2007). Two different methods of genome annotation currently preside: first, the static level and second, the dynamic level. While the static level is only concerned with the basic, automatic annotation that predicts the location and function of genes and proteins, including such features as the chemical and structural properties, subcellular location, and modular organisation, the dynamic level includes the biological context, by taking into account the interactions between the genomic components (Médigue and Moszer, 2007). These interactions refer mainly to protein-protein interactions, regulatory interactions and metabolite transformations. The dynamic level is more focussed on the networks and the biological contexts, using the information obtained from the first, static round of annotation. This leads to a more complex interpretation of the dimensions of annotation (Reed *et al*., 2006).

Correct, successful annotation relies on similarity analysis, which requires numerous data collections, archives and correctly annotated genomes. Different databases exist for this purpose, including DNA, RNA and protein databases. Protein domains are used to determine the function of predicted genes. A problem that arose with the expanding number of databases available is the issue of integration. One solution is the Interpro project that

integrates different domain signatures to create a unique characterisation for a given protein family, domain or functional site (Apweiler *et al.*, 2000). Annotation and comparison between large numbers of genomes require a practical, user-friendly interface in order to make sense of and organise the currents of data. Therefore, in addition to the annotation software available, there are also different software solutions from which to present the annotation results visually.

In the not too distant past, annotation was done manually; a cumbersome and time-consuming process requiring careful curation and review. However, today with high-throughput and low-cost sequencing methods, such attention to individual sequences is not feasible. Thus, the focus shifted to automated annotation, with very little 'human' influence. The upshot is that some genomes remain poorly, or incorrectly, annotated. Reannotation provides a solution to this: genomes need to be updated regularly. As manual curation is still performed, albeit on a much smaller scale, certain putative proteins or other proteins with unfamiliar functions may be incorporated into the automatic database, thus integrating experimental and predicted data. In support to this feature, integrated databases will provide even better annotation. Armengaurd (2009) indicated that an alliance between proteomics and genomics, which can be carried out in the earliest phase of genome sequencing, produces almost complete and accurate gene catalogues for small microbial genomes, as has been illustrated with the *M. mobile* and *Deinococcus deserti* genomes. Proteomics, transcriptomics, metabolomics and interactomics present the opportunity to integrate genomic and functional information. Integration is important in more than one aspect: it enables standardisation of experimental results, simultaneous and complex correlations between different pieces of information and coordinated data mining is possible, and genes and proteins that behave similarly under various conditions can be grouped (Swertz and Jansen, 2007).

3.2. THE PARTIAL ASSEMBLY, REPETITIVE SEQUENCE AND BIOINFORMATIC ANALYSIS OF THE CONTIGUOUS SEQUENCES OF THE Ms02 GENOME

3.2.1. INTRODUCTION

By January 2012, the number of complete bacterial genome sequences deposited into the **N**ational **C**enter for **B**iotechnology **I**nformation (NCBI) reached 1743. This rapid increase in the availability of numerous prokaryotic sequences has revolutionised new fields of research: that of comparative genomics. We are able to scrutinise pathogenic organisms based on their genome sequence alone. Comparative studies between different annotated genomes enable identification of possible pathogenic proteins, immunogenic targets, and other important metabolic features of prokaryotes in general. The advantage of this bioinformatic approach is especially relevant to pathogens. Different approaches regarding vaccines and other immune therapies against pathogens can be investigated, without having to cultivate these dangerous organisms. The role and importance of bioinformatics thus lies in it enabling the extraction of meaningful biological information from the flood of sequence data. Raw sequence data need to be interpreted, which involves the identification and annotation of genes, proteins, and regulatory and metabolic pathways, in order to exploit fully the enormous possibilities therein. Genomes also need to be reprocessed or reannotated on a regular basis to incorporate newly characterised functions.

Liu and Whittier (1995) developed a PCR strategy using interlaced cycles of high and low stringency in a multistep thermal cycling program, called TAIL-PCR. This strategy enables the isolation of flanking regions of unknown DNA to a known DNA sequence, irrespective of whether it is a coding sequence or not, as is the case with library screening methods. Mazars *et al*. (1991) devised the first PCR strategy based on thermal asymmetry. However, the concept of interlaced cycles was illustrated by Liu and Whittier (1995). Instead of just having low and high stringency cycles, their concept is based on intertwining the thermal asymmetric cycles: as such allowing for better formation of specific product, as well as amplifying longer products.

Mycoplasma species represent an interesting organismal group to study from an evolutionary viewpoint, given their adaptation to parasitic lifestyles and interaction with their host cells. In order to explore the organism to its full extent, it is necessary to obtain the genome sequence. The genome sequence of an organism may reveal novel biological information. With the fast expanding availability of bioinformatic tools, the amount of information that can be extracted by purely investigating the DNA sequence of an organism is remarkable. Knowledge about its lifestyle, adaptive capabilities, metabolism, evolution and phylogenetics is just the tip of the iceberg. The Ms02 genome was previously sequenced using GS FLX 454 sequencing technology, which generated read lengths of an average 300 bp. The short sequences were assembled into 28 larger contiguous sequences, amounting to an estimated size of 900 kbp (895 119 bp). In the present study, an attempt was made to decrease the number of contigs using TAIL-PCR. After assembling a certain amount of the initial contiguous sequences, the newly generated consensus sequence was subjected to bioinformatic analysis, including repetitive sequence analysis and predicting the location of the origin of replication.

3.2.2. EXPERIMENTAL PROCEDURES

3.2.2.1. Genomic DNA isolation

Ms02 cultures were cultivated by Mr. Johan Gouws at the Veterinary Institute in Onderstepoort, Pretoria. Mycoplasma colonies were routinely visible after 72 hours' growth on modified Hayflick's medium plates, although it tends to vary, especially between species. The genomic DNA was isolated from a phosphate buffered saline (PBS) suspension of the colonies, using a slightly modified method that was first described by Hempstead (1990). In short, 25 ml Ms02 culture was centrifuged at 12 000 x g (4°C) in a sterile JA-20 centrifuge tube for 60 min (Avanti® J-E Centrifuge; Beckman Coultier, USA). Care was taken to decant the supernatant as completely as possible, after which the pellet was resuspended in 1 ml TE buffer (50 mM Tris-HCl and 10 mM EDTA, pH 7.5), followed by centrifugation at 16 000 x g (4°C) for 30 min in a microcentrifuge (Spectrafuge, 24D, Labnet International, Inc.). After resuspending the pellet in 100 µl TE buffer, the sample was stored overnight at -20°C. The frozen cell suspension was incubated at 55°C until just thawed, after which 1 ml TE buffer, containing 1% (v/v) SDS, was added. To this, 44 µl proteinase K (20 mg/ml, Roche) was added and incubated for 30 min in a 55°C water bath with occasional mixing. Thereafter, 14.3 µl pancreatic RNase A (20 mg/ml) was added and incubated for 1 h at 37°C. To precipitate the SDS (indicated by the clearing of the suspension), 100 µl (5 M) potassium acetate (untitrated) was added and incubated on ice for 30 min. After centrifuging the sample at 16 000 x g (room temperature, RT) for 10 min, the supernatant was transferred to two sterile 2 ml microcentrifuge tubes. An equal volume of chloroform:isoamylalcohol (24:1) was added and the samples mixed

by inverting the tubes by hand for 15 min at RT. After 2 min centrifugation at 16 000 x g (RT), the upper phase was removed and the process repeated twice more. After the third centrifugation, the supernatant was briefly mixed with a tenth volume of 3 M sodium acetate (pH 4.8), before adding 2 volumes 95% ethanol. The tubes were then inverted a number of times to ensure a homogenous mixture. Overnight incubation at -20°C was performed to allow for maximum precipitation of the DNA. Centrifugation for up to 40 min at 16 000 x g was performed to maximise pellet formation, after which the pellet was washed twice with 70% ethanol. To ensure complete evaporation of the ethanol, the pellet was air-dried overnight at room temperature. When the pellet was completely dry, it was resuspended in 50 µl TE buffer (10 mM Tris-HCl and 1 mM EDTA, pH 8.0) and left overnight at 4°C to dissolve the DNA.

3.2.2.2. Quantity and quality determination of the genomic DNA

The concentration and quality of the isolated Ms02 genomic DNA samples were analysed spectrophotometrically with the Novell® Nanodrop ND-1000 (v 3.5.1) at 260 nm. Concentration (ng/µl) and purity (A_{260}/A_{230} and A_{260}/A_{280}) were measured.

3.2.2.3. Confirmation of Ms02 genomic DNA identity

To ensure the culture was free from contamination from Ms01 and Ms03, and to confirm the identity of the DNA preparations that were isolated from the Ms02 culture, a standardised mycoplasma PCR was performed using species-specific primer pairs (Botes *et al*., 2005a), given in Table 3.1.

Table 3.1 Nucleotide sequences of the species-specific primer pairs to confirm the purity of the isolated Ms02 DNA.

Primer pair	Primer name	Primer sequence (5'→3')	Bp-position	T_m (°C)
Ms01	Ms01Z (F)	5'-AACATTAGTTAATGCCGGATACGC-3'	114	75
	Ms01D (R)	5'-GCCAGTATCCAAAGCGAGCC-3'	613	75
Ms02	Ms02H (F)	5'-AATATAAAAGGAGCGTTTGC-3'	160	70
	Ms02A (R)	5'-AAGGCAATAGCATTTCCTCTACT-3'	447	70
Ms03	Ms03A (F)	5'-AGTGCTAATGCCGGATACTTATAC-3'	118	70
	Ms03C (R)	5'-CGTTAACCTCTATACAATTCTAGCG-3'	639	70

Three positive control reactions were performed; each mycoplasma DNA with its own primer set. The Ms02 DNA was tested with each of the three primer pairs. Briefly, the reaction contained 2 µl of 10x reaction buffer (RB; JMR-Holdings, USA), 0.8 µl of 5 mM of each deoxynucleotide (dATP, dCTP, dGTP and dTTP; Advanced Biotechnologies Ltd., UK), 0.4 µl of each primer (20 pmol/µl), 1.6 µl of 25 mM $MgCl_2$, 0.1 µl of Taq DNA polymerase (5 U/µl Super-Therm Taq polymerase, JMR-Holdings), 12.7 µl Milli-Q® water and 2 µl of purified DNA.

The amplifications were carried out in a 2720 Thermal Cycler (Applied Biosystems Ltd., Warrington, UK). The program consisted of 30 cycles of 30 s at 94°C, 15 s at 57°C, and 1 min at 72°C, followed by a final extension reaction for 6 min at 72°C. The amplified DNA was subsequently cooled and analysed by gel electrophoresis.

3.2.2.4. PCR product agarose gel analysis

Ten microliters of the PCR product, mixed with 1 µl of gel loading buffer (50% (v/v) glycerol, 0.1% (v/v) bromophenol blue, 50 mM EDTA, and 100 mM Tris-acetate, pH 8.0), were electrophoresed (100 V,~60 min) on a 1% (w/v) agarose gel (Molecular Grade Agarose D1-LE, Whitehead Scientific) in 1 x TAE buffer (40mM Tris-acetate and 2 mM EDTA, pH 8.0). GelRed™ Nucleic Acid Gel Stain (Cat #41003, Biotium, Hayward, CA) (0.1 µl/ml) was included in the gel for ultraviolet (UV) visualisation of the DNA.

3.2.2.5. Assembly of the Ms02 contiguous sequences by TAIL-PCR

3.2.2.5.1. Oligonucleotide primer design

All primers were designed using the computer software package Primer Designer Version 1.01, and synthesised by Integrated DNA Technologies (IDT), Iowa City, USA. Four arbitrary degenerate primers were synthesised, according to Liu and Whittier (1995). These primers have an average T_m of 47-48°C, as calculated by the method of Liu and Whittier (1995). Contiguous sequences were chosen at random and three long, sequence-specific nested primers were designed on the 3' end of a contig. All the primers were either 26 or 28 bp long. These three primers were designed to have a T_m within 5°C of one another and to range between 68°C and 76°C. The T_m of the specific primers was calculated by the Primer Designer software. Where deemed necessary, further primer pairs were designed to confirm results obtained with the TAIL-PCRs. These primers were designed to be 24 bp in length.

3.2.2.5.2. TAIL-PCR procedure

Three rounds of TAIL-PCRs were carried out, using a Veriti™ 96-well Thermal Cycler (Applied Biosystems Ltd.). The product of the previous PCR served as the template for the next. The PCR conditions for all three procedures can be seen in Table 3.2. The PCR mixture for the primary TAIL-PCR contained 4 µl of a 5x RB, 0.4 µl of 10 mM of each deoxynucleotide (dATP, dCTP, dGTP, dTTP), 0.8 µl of 25 mM $MgCl_2$, 0.2 µl of the forward primer (primer a) (20 pmol/µl), 2 µl of the degenerate primer (primer AD) (20 pmol/µl), 0.1 µl of 5 U/µl Kapa2G Robust Taq DNA polymerase, all from Kapa Biosystems (South Africa), approximately 1 ng purified DNA, and Milli-Q® water up to 20 µl. The reaction mixture for the secondary TAIL-PCR contained 2 µl of 10x RB, 0.8 µl of 5 mM of each deoxynucleotide (dATP, dCTP, dGTP and dTTP), 0.2 µl of the forward primer (primer b) (20 pmol/µl), 2 µl of the degenerate primer (primer AD) (20 pmol/µl), 2 µl of 25 mM $MgCl_2$, 0.1 µl of Taq DNA polymerase (5 U/µl; Super-Therm Taq polymerase), 11.9 µl Milli-Q® water and 1 µl of DNA (1:500 diluted from the primary reaction). For the final tertiary PCR reaction, the mixture contained the same final concentrations as for the secondary reaction, except for equal amounts of the specific and arbitrary primer: 1 µl (20 pmol/µl) each. The reactions were scaled up to 50 µl to ensure ample amount of DNA was available for down-stream use.

Table 3.2 TAIL-PCR procedures for the three reactions.

Reaction	Number of cycles	Thermal settings
Primary		
	5	94℃-30s; 68℃-30s
	1	94℃-30s; 25℃-1s; 10.4% ramp rate to 68℃; 68℃-30s
	24	94℃-30s; 68℃-30s; 94℃-30s; 68℃-30s; 94℃-30s; (AD1)41℃/(AD2)39℃/(AD3,4)45℃-30s; 68℃-30s
	1	68℃-5min; 15℃-hold
Secondary		
	25	94℃-30s; 68℃-30s; 94℃-30s; 68℃-30s; 94℃-30s; (AD1)41℃/(AD2)39℃/(AD3,4)45℃-30s; 68℃-30s
	1	68℃-5min; 15℃-hold
Tertiary	30	94℃-30s; (AD1)41℃/(AD2)39℃/(AD3,4)45℃-15s; 68℃-15s
	1	68℃-5min; 15℃-hold

The amplified PCR products were analysed on a 1% (w/v) agarose gel as previously discussed in Section 3.2.2.4. Only 10 µl of the 50 µl were loaded onto the gel. Fragments were selected based on their size and concentration (degree of visibility). Those with only one fragment visible on the gel were purified directly from the PCR product, while the PCR reactions with more than one fragment were again loaded completely on a gel and the chosen band excised. Prior to sequencing, the generated products were purified with illustra GFX™ PCR DNA and Gel Band Purification kit (GE Healthcare, UK), according to the manufacturer's instruction, but with one difference: the elution of the DNA was performed with 10-50 µl pre-heated (60℃) Milli-Q® water. Concentration of the purified DNA fragments was determined spectrophotometrically as discussed in Section 3.2.2.2.

3.2.2.5.3. Sequencing and analysis of PCR products

Sequencing of the TAIL-PCR products was done with the ABI BigDye® Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, UK). Each sequencing reaction consisted of 5 µl of 5x Sequencing buffer, 2 µl Terminator mix, 1 µl (8 pmol/µl) primer (primer c), purified DNA (amount as determined by the ABI BigDye® Terminator instructions), and Milli-Q® water to add up to a final volume of 10 µl. Amplifications were performed in a 2720 Thermal Cycler programmed to perform 35 cycles of 10 s at 96℃, 30 s at 52℃, 4 min at 60℃, and a final extension of 10 min at 60℃.

The sequencing PCR products were sequenced and analysed with an ABI® 3130 Genetic Analyser (Applied Biosystems) at the Central Analytical Facility (CAF), the DNA sequencing facility of the University of Stellenbosch. The resulting sequences were used as a query sequence in a **B**asic **L**ocal **A**lignment **S**earch (BLAST) against a database containing all the contigs of the Ms02 genome in CLC Combined Workbench

(v. 4.0.1). The contigs with significant hits were then manually aligned in BioEdit® Sequence Alignment Editor (v 7.0.5.2, Hall), with the generated sequence, as well as with the sequence the specific primers were originally designed on.

3.2.2.6. Investigation of the Ms02 genome repeats

3.2.2.6.1. Large repetitive sequences in the Ms02 genome

Large repetitive units were defined as being at least 80 bp long and occurring at least twice in the genome with 80% and higher identity (based on the definition by Koressaar and Remm, 2012). To find these repetitive sequences, a BLAST was performed on all of the Ms02 contigs. Each contig was used as a query sequence in a BLASTn (nucleotide BLAST) against all 28 large Ms02 contigs using the 4-mold mitochondrial code with the word size set at 7 and a match:mismatch setting of 1:-3, all performed in CLC Combined Workbench. All the hits above 80 bp and with 80% or higher identity were investigated to determine whether the repeats were involved in coding sequences as annotated previously (Steenmans, 2010). If one copy of the repeat falls within an ORF, the repeat was regarded as gene-associated.

3.2.2.6.2. Simple sequence repeats in the Ms02 genome

In order to search for SSRs, the contigs of the Ms02 genome were subjected to the repeat search program SPUTNIK (http://espressosoftware.com/sputnik/index.html). It searches for di-, tri-, tetra- and penta-nucleotide repeats, using its own scoring mechanism. Insertions, mismatches and deletions were tolerated, although they affected the overall score. The program does not search against a library of known microsatellites, but instead reads through the provided DNA sequence, assuming the existence of a repeat at every nucleotide, comparing subsequent nucleotides and allocating a score to the significance. If the score rises above a preset threshold value, the repeat is accepted. If it remains below the threshold, the search is abandoned and a new search started at the next nucleotide. Nucleotides that match the predicted value, assuming a repeat, add to the score, while those that do not match, subtracts from the score. Three possible kinds of error exist: insertions, deletions, or mismatches. These are revisited and comparisons may result in a score above the cutoff threshold, and the best score of the three may be accepted.

The length of an SSR was measured in base pairs, rather than as the number of repetitive units. This allows for detection of partial copies and also facilitates comparison among SSRs of different lengths. In addition, a repeat that formed part of more than one oligomer, was deemed to be the shortest SSR and counted as such. This means that a sequence such as ATATATATATAT qualifies as a dinucleotide (AT), as well as a tetranucleotide (ATAT) and a hexanucelotide (ATATAT), but is counted as a dinucleotide.

3.2.2.6.3. Tandem repeats in the Ms02 genome

The Tandem Repeat Finder (TRF) software available at http://tandem.bu.edu/trf/trf.html was used to identify tandem repeats (Benson, 1999). The program uses the technique of independent Bernoulli trials (coin-tosses) to compare the alignment of two tandem copies of a pattern of length *n,* by a sequence of n independent “coin-tosses”. Each head then represents a match between the aligned nucleotides, while tails represent a mismatch, insertion or deletion. It also calculates the probability which specifies the average percentage of insertions and deletions between the aligned copies. Two components are employed in the process of repeat finding: detection

and analysis. First the detection component finds candidate repeats based on a set of criteria. Following this, the analysis component attempts to produce an alignment for each candidate tandem repeat. Successful alignments are further analysed to determine statistics about the alignment, such as percentage identity and indels (insertions and deletions), and the nucleotide sequence, including the nucleotide composition and entropy measure. The period size is the pattern size of the tandem repeat. If the repeat contained numerous copies, then this repeat will be detected at various period sizes. The software noted the period size and the three best scoring sizes were chosen. The following parameters were chosen: match (2), mismatch (-3), indel (-5), a minimum alignment score of 50 to report a repeat, and a maximum period size of 2 000. These represent the least stringent parameters available on the program. To analyse the repeats, the 28 contigs were linked to one another numerically. This enabled ease of assessment of the repeats, without having to search manually between contigs.

3.2.2.6.4. Single base pair runs in the Ms02 genome

In order to determine the amount and extent of single nucleotide repeats in the Ms02 genome, the contigs were searched manually as individual sequences for base pair runs of A, T, C and G. The contigs were imported into the BioEdit® Sequence Alignment Editor and searched for base pair runs of 6 bp and higher (Gur-Arie *et al.*, 2000).

3.2.2.7. Bioinformatic analysis of the Ms02 contiguous sequences

The sequences that were newly assembled, as well as the sequences that could not be assembled, were first subjected to a series of analyses to improve the preliminary annotation. This meant taking into account the orientation and grouping of essential genes to decide in which direction a contig is probably positioned, as well as using the genes located on the ends of the contigs to try and predict whether certain contigs can be provisionally linked to one another. In this study the terms ‘plus’ and ‘minus’ are not necessarily used to indicate leading or lagging, but rather indicative of the direction of the genes on the specific contig, as well as to distinguish between the orientation of a contig after being linked to another, from its original orientation.

3.2.2.7.1. Directional orientation of contiguous sequences based on predicted gene orientation

Using the preliminary annotations (Steenmans, 2010), the contigs were organised to be orientated in a specific direction. This was determined by arranging the individual contigs according to their annotated gene content. The plus strand was predicted to contain the majority of essential genes and also predicted to contain the most ORFs. The rRNA genes and ribosomal genes were analysed in order to determine their direction, since they are mostly transcribed in the direction of DNA replication, i.e. the leading strand.

3.2.2.7.2. Provisional linkage of contigs based on the gene content on the contig ends

The non-linked contigs were investigated to determine if the genes occurring on their ends could be used to predict whether specific contigs could be linked by using a focused PCR. The genes that occur on the ends of the contigs were noted and the contigs analysed, resulting in a prediction as to which other contig it likely occurs next to. Specific primers were then designed to test the prediction. The PCR was performed as for the Ms02 identity confirmation (refer to Section 3.2.2.3).

3.2.2.7.3. Assembling the single contigs into one large contiguous sequence

The contigs that were linked with the PCR-based methods, the contigs with the additional sequences that did not align with another contig, as well as the rest of the contigs, were used to generate one large contiguous sequence. The information of the directional predictions, based on the gene distribution and gene orientation on the individual contigs, and the provisional linkage based on the gene content on the ends of the contigs, were used to link the individual contigs in one single sequence.

3.2.2.7.4. Searching for the origin of replication

Different programs were used to search for the *ori*, but only after the contigs were aligned and orientated in their respective predicted directions.

Ori-Finder (http://tubic.tju.edu.cn/Ori-Finder/) finds *ori*s in bacterial genomes based on an integrated method that comprises the analysis of base composition symmetry (using the Z-curve method), the analysis of the distribution of DnaA boxes, and the occurrence of genes that occur frequently close to *ori*s (Gao and Zhang, 2008). Unannotated genome sequences' *ori* is routinely found by using the integration of the gene-finding program ZCURVE (v. 1.02). The Z-curve constitutes a unique representation of a DNA sequence, shown as a 3D curve, commonly used for replication origin identification (Worning *et al.*, 2006; Zhang and Zhang, 2005), gene-prediction (Guo *et al.*, 2003), isochore identification (Zhang and Zhang, 2004a), genomic island identification (Zhang and Zhang, 2004b), and comparative genomics (Zhang *et al.*, 2003).

DoriC (v. 5.0) (http://tubic.tju.edu.cn/doric/blast1.html) is an updated database of bacterial replication origins, allowing one to search against known *ori*s. The software's default settings were used to search the Ms02 genome, namely the BLOSUM62 matrix, with an Expect value (e-value) of $1.0e^{-10}$. The genome was also analysed in GraphDNA (http://athena.bioc.uvic.ca/virology-ca-tools/graphdna/), a DNA Skew Graphing tool that allows the user to generate graphical representations of raw DNA sequences (Thomas *et al.*, 2007). The G+C skew is calculated as (C-G)/(C+G) and changes sign when crossing the o*ri* region (Lobry, 1996a; Lobry 1996b). The A+T skew [(A-T)/(A+T)], keto-amino skew [(A+C)-(G+T)] and purine-pyrimidine [(A+G)-(C+T)] skew were also calculated. A sliding window size of 70 bp was chosen, based on the software's recommendation.

3.2.2.7.5. Gene prediction of the Ms02 genome

The TAIL-PCR assembled contigs, together with the findings on the genome strand analysis and the provisional linkage of the contigs, were used to create one large contiguous sequence. The contiguous sequence was then submitted to four different automatic gene prediction programs (Table 3.3), the results compared, with the most likely ORFs identified. The lower limit cutoff size used for ORFs was 70 bp. This was decided based upon the previous annotations, where the smallest group of genes were the tRNA genes, falling in the range of 70-90 bp. In NCBI, the ORF Finder is limited to 50 bp as the smallest ORF. The 4-mold, mitochondrial and mycoplasma code was chosen as the genetic code used to identify the open reading frames. Another gene prediction program available from the NCBI website, GLIMMER, was also used. GLIMMER (**G**ene **L**ocator and **I**nterpolated **M**arkov **M**odel**ER**) (v. 3.02), uses interpolated Markov models to identify coding regions in the genomes of bacteria and archaea (Delcher *et al.*, 1999; Salzberg *et al.*, 1998). The Ms02 genome was run through the algorithm with the (4) mycoplasma/spiroplasma genetic code. The final software used, GeneMark,

required a reference genome, because the Ms02 genome is too short for self-training. For this purpose the genome of *M. synoviae* 53 was chosen, because it is the closest to Ms02 phylogenetically in comparison with other available genomes. GeneMark.hmm 2.4 (v. 2.10f) was run using the genetic code 4. The program integrates the GeneMark models into a hidden Markov model framework, where gene boundaries are modeled as transitions between hidden states, using the ribosome binding site model to make the gene-start predictions more accurate. Protein-coding potential is analysed within a sliding window using species-specific parameters of the Markov models of coding and non-coding regions. The final application used to predict ORFs (the earlier introduced ZCURVE), utilises the Z-curve, a bioinformatics algorithm widely used in genome analysis, including *ab initio* gene prediction (Guo *et al*., 2003).

The ORFs that were identified to overlap the linked contigs, i.e. new ORFs that did not occur in the previous annotations, were submitted to a BLASTx search (where a translated nucleotide query is searched against the protein database) to determine the identity of the ORF. The mold mitochondrial (4) genetic code was chosen for translation to search the mycoplasma (taxid: 31969) organism database. Significant hits were chosen based on the E-value (an indication of the statistical significance of the alignment between the query and the target sequence), and the bit score values. The bit score is an indication of how good the alignment is: a higher score means a better alignment.

Table 3.3 Software applications and databases that were used to search for open reading frames in the Ms02 genome.

Name	Comments	URL/Availability
BLAST	Software for sequence database searching	http://www.ncbi.nlm.nih.gov/blast/
GenBank	Annotated collection of publicly available DNA sequences	http://www.ncbi.nlm.nih.gov/GenBank/
GLIMMER	Protein gene prediction software for microbial genomes	www.ncbi.nlm.nih.gov/genomes/MICROBES/-glimmer_3.cgi
GeneMark	Self-training protein gene prediction software	http://exon.gatech.edu/-gmhmm2_prok.cgi
ORF Finder	Finds open reading frames of selectable minimum size	http://www.ncbi.nlm.nih.gov/projects/gorf/
ZCURVE	Genome analysis that includes gene prediction	http://tubic.tju.edu.cn/Zcurve_B/

3.2.2.7.6. PCR of the linkages between the predicted contig alignments

A standard mycoplasma PCR was performed on the contigs as they were linked, where primers were available to test the alignments (Section 3.2.2.3). The reactions were scaled down to contain only 10 µl final reaction volume. The products were separated on a 1% agarose gel (Section 3.2.2.4).

3.2.3. RESULTS

3.2.3.1. Isolation of Ms02 genomic DNA

Genomic DNA was isolated successfully with the modified Hempstead method and was free of contamination from Ms01 or Ms03. The quality and quantity of the DNA were sufficient to use it as a DNA template for the TAIL-PCRs. The DNA was confirmed to be Ms02, producing a product of the correct size in the Ms02 DNA sample, and the absence of any product in the DNA sample tested with the Ms01 and Ms03 primer sets. This indicated that there was no Ms01 or Ms03 DNA present.

3.2.3.2. Linkage of contiguous sequences by TAIL-PCR

Contigs were chosen at random and three specific primers were designed on the 3' end. It was found that primers designed on the 5' end tended to result in band sizes of insufficient length to justify sequencing. Certain cases required optimisation: in these cases the annealing temperatures were adjusted to be approximately 5°C above the T_m of each individual specific primer. When the sequenced PCR product proved linkage between two contigs, a new set of primers was designed on the 3' end of the linked contig. In this manner, a total of 22 sets of specific TAIL primers on 16 contigs were designed. This resulted in 15 linkages (Table 3.4). Another four contigs were linked (contig 3 with contig 16 and contig 7 with contig 17), but this was not done using TAIL-PCR and will be discussed later. The net result was the linkage of 19 contigs in four separate contiguous sequences, named A, B, C and D for ease of further reference. Thus the initial 28 contigs of the Ms02 genome was reduced to 14. However, it was soon realised that the sequencing results were not always reliable, and it was necessary to design specific forward and reverse primers, to corroborate the linkage sequence. In addition, the presence of repetitive units in the genome hindered the definite linkage of the sequences.

In Figure 3.2, an image of a typical TAIL-PCR agarose gel can be seen, following electrophoretic separation and UV visualisation of the DNA. In this instance, specific primers were designed on the 3' end of contig 15. The nested effect due to the use of three specific primers in the three consecutive PCR reactions is clearly visible in the image. Also prevalent is the absence of any visible product in the primary TAIL reaction, as can be seen for AD_1. The degenerate primers AD_3 and AD_4 did not result in any usable products. This was found to be the norm rather than the exception for the TAIL-PCRs. As can be seen in Table 3.4, 76% of the products that were used in successful sequencing resulted from the AD_1 and AD_2 primer mixes. An exception was the linkage sequence between contigs 32 and 1, and contigs 32 and 33.

Table 3.4 The TAIL-PCR linked contigs, additional base pairs that were added and the E-values of the linkage.

Contig alignment	AD primer	Additional base pairs	E-value
1_32	AD_4	71	$6.57e^{-123}$
4_38	AD_1	173	$7.04e^{-21}$
17_31	AD_1	10	0
31_37	AD_1	374	$5.84e^{-153}$
32_33	AD_4	8	0
33_34	AD_1	0	$2.35e^{-72}$
34_33	AD_1 & AD_2	8	0
37_39	AD_1 & AD_2	214	$4.89e^{-24}$
37_7	AD_1	370	$1.08e^{-49}$
38_15	AD_2	0	$4.20e^{-86}$
26_1	AD_2	overlap 5 bp	$1.02e^{-85}$
15_8	AD_2	430	$1.86e^{-31}$
41_4	AD_1	171	$2.56e^{-43}$
10_41	AD_1 & AD_2	160	0
8_33	AD_2	50	0

The linkage between contigs 1 and 32 was first detected using a focused PCR, after which a TAIL-PCR product was also sequenced. The linkage between contigs 32 and 33 was sequenced a few times afterwards with the other AD primers as well, to confirm the sequence alignments. In general, the resulting sequences produced a surprisingly large number of additional sequences. Also evident in Table 3.4, is the linkage of one contig to more than two other contigs (contig 37 links to contig 7, contig 31, and contig 39). This in itself is an indication of the presence of repeats, although it was not expected for the repetitive sequences to be amplified, when doing a nested PCR such as with a TAIL-PCR. Contig 33 and contig 34 also present ambiguous results. The 3' end of contig 33 linked with the 5' end of contig 34, whose 3' end in turn linked with the 5' end of contig 33, effectively forming a circle. Numerous attempts were made to try and establish which of these linkages are correct, but definitive results remained elusive. Another interesting observation was that the additional base pairs between these three contig linkages represent the largest amount of extra sequences that were produced during any contig linkage. The only other (largest) additional sequence is the linkage between contig 15 and contig 8. TAIL-PCRs performed on the 3' end of contig 15 did not produce a long enough sequence to enable linkage to another contig. It necessitated the design of a new set of specific primers on the generated sequence from the first TAIL-PCR, on the 3' end of the contig 15-generated sequence, before a linkage was established with contig 8.

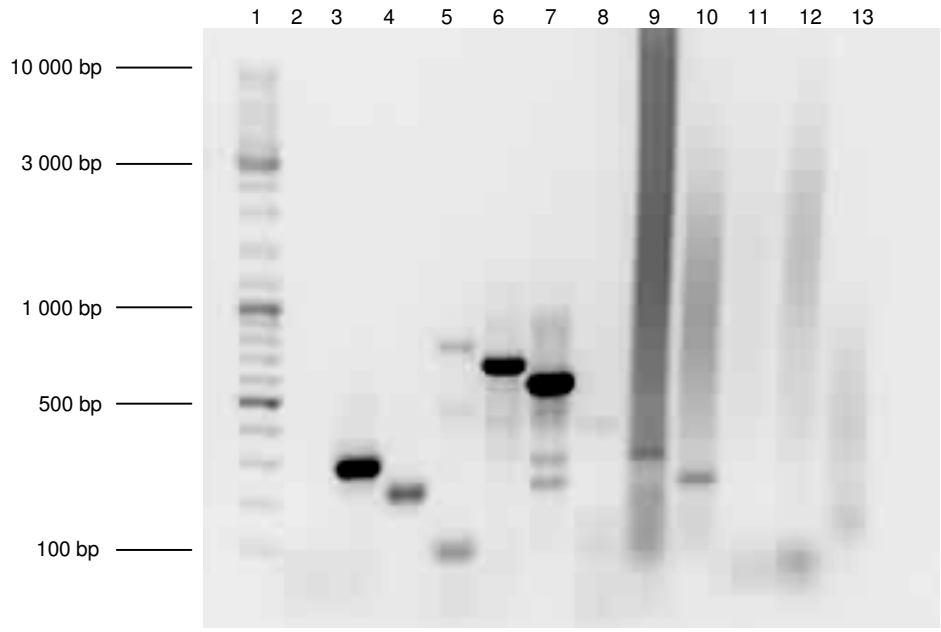

Figure 3.2 Image of the TAIL-PCR products of contig 15 obtained after electrophoretic separation. The PCR products were separated on a 1% (w/v) agarose gel with 0.1 µl/ml GelRed™ Nucleic Acid Gel Stain for UV visualisation of the DNA. Lane 1: 1 KB DNA GeneRuler™ DNA ladder mix (Fermentas). Lane 2-4: TAIL1-3 with AD_1. Lane 5-7: TAIL 1-3 with AD_2. Lane 8-10: TAIL 1-3 with AD_3. Lane 11-13: TAIL 1-3 with AD_4. To sequence a product in order to determine the contig linkage, the product of TAIL 3 from AD_2 was loaded onto a gel, excised, purified and sequenced.

3.2.3.2.1. Linkage of Ms02 contig A

With a test PCR, the primers contig_3d and contig_39x produced a band of approximately 400 bp. Although the general appearance of the agarose gel was poor, it was decided to purify the PCR product and subject it to sequencing from both ends. However, sequencing with the contig_39x primer resulted in sequences that did not align with contig 39. In this instance, the 39x sequence aligned with contig 3, together with the 3d sequence, but unexpectedly did not align with contig 39. Instead it aligned with contig 16 with an additional 16 bp, consisting of a CTn repeat. In Table 3.5, the results of the local BLAST with the 3d sequence and the 39x sequence can be seen. Thus contig 3 (3' end) was linked with contig 16 (5' end).

Table 3.5 Local BLAST performed on the sequences 3d and 39x. The result was that contig 3 links to contig 16, but no hits were achieved with contig 39.

Contig hit	Contig length (bp)	Sequence	E-value	Hit range	Query range	Hit length (bp)
Contig 16	809	3d	$1.04e^{-114}$	17-296	98-376	280
		39x	$2.38e^{-99}$	17-271	266-14	255
Contig 3	1204	3d	$3.90e^{-15}$	1147-1202	10-64	56
		39x	$3.70e^{-43}$	1110-1202	392-300	93

3.2.3.2.2. Linkage of the Ms02 contig B

Contig B consisted of 5 linked contigs in the following sequence: 7-17-31-37-39. A large repeat between contig 7 and contig 31 also linked contig 7 with contig 37. Sequences from contig 7 and sequences from contig 31 aligned for approximately 120 bp, causing the confusing alignment results. Primers that were designed to determine the correct linkage resulted in unrepeatable, numerous different results. Finally, it was decided that, based on the E-values (Table 3.4), the linkage was finalised as described above. In addition, the 3' end of contig 7 aligned perfectly for 130 bp with the 5' end of contig 17. It is therefore possible that contig 7 links to contig 17. In order to confirm whether it is a feasible assumption, a standard PCR was performed. It has been noted that the binding site for primer 7c is located inside the repeat, so only the two products formed by primer 7a and primer 7b were sequenced (Figure 3.3).

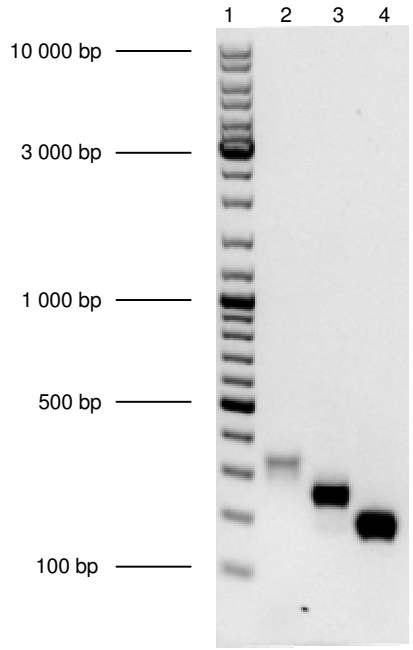

Figure 3.3 Image of the PCR products between contigs 7 and 17 obtained after electrophoretic separation. The PCR products were separated on a 1% (w/v) agarose gel with 0.1 µl/ml GelRed™ Nucleic Acid Gel Stain for UV visualisation of the DNA. Lane 1: 1 KB DNA GeneRuler™ DNA ladder mix. Lane 2: product formed with the primer pair contig_7a and contig_17R. Lane 3: product formed with the primer pair contig_7b and contig_17R. Lane 4: product formed with the primer pair contig_7c and contig_17R. The products that resulted from primer pairs contig_7a and contig_17R and contig_7b and contig_17R were purified and subjected to sequencing.

3.2.3.2.3. Linkage of the Ms02 contig C

Named contig C, this newly generated sequence consisted of the following linked contigs: 8-15-39-4-41-10. Assembling these contigs was accomplished using the TAIL-PCR method, but specific primers needed to be designed to confirm the linkages. The linkage between contig 4 and contig 41 represents a rare success when using reverse primers in the TAIL-PCRs. Three specific primers for use in a TAIL-PCR were designed on the end of contig 10. The resultant sequence had a BLAST hit of $1.19e^{-26}$ with contig 39, but the alignment was not

significant throughout. Although numerous attempts were made, it was not possible to establish a link between contig 10 and contig 39. It was decided that the linkage was not correct.

For contig 15, three specific primers were designed on the 3' end for use in TAIL-PCR, which resulted in a clear sequence of over 500 bp, but in spite of the length, did not align to another contig. To address this problem, three more specific primers were designed on the generated sequence. This resulted in a large enough sequence (total 680 bp) and aligned with the 5' end of contig 8, with an overlap of 70 bp and an additional 430 bp sequence in between. Three specific primers were also designed on the 3' end of contig 8 and the resultant sequences aligned perfectly with contig 33. This will be discussed further together with contig D, since the presence of a repeat on contig 33 hindered a confident prediction of the alignments.

3.2.3.2.4. Linkage of the Ms02 contig D

Contig D was one of the most difficult sequences to assemble, consisting of the following contigs: 34-33-32-1-26. While at first every contig aligned perfectly with long overhangs with the next contig, later the presence of a large repetitive unit was detected. This enabled a linkage with perfect alignments of contig 32, contig 34, as well as contig 8, with the 5' end of contig 33 (Figure 4.3). A series of nested and specific PCRs finally indicated that contig 8 did not link to contig 33: although bands formed in each PCR reaction, the bands were of incorrect size. However, the formation of products was erratic between contig 32 and contig 33: a number of times the product size was extremely large (in excess of 3 000 bp).

The additional base pairs between these contigs and contig 33 are runs of An (refer to Section 3.2.3.3.4 for more information on this matter). For none of these contigs are there other possibilities of alternative linkages. For this reason, it was decided that contig 33 links to the 5' end of contig 34, and that contig 32 links to the 5' end of contig 33. Further support for this decision resides in the reannotation of contig 8 linked to contig 33, which was established in the course of an Honours project in this laboratory (Bezuidenhout, 2012). No difference in the ORF distribution was detected, especially near the connected ends.

```
contig 33
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~AAAAAAATAACCCTCCGTGGGTTTTGATCTTTGAAAACTGAATAGT
32c_contig32
TAAGTTCTGTAAATTAAGTTTATGCACTAAAAAATAA......~.......................................

8C_contig8
CCTTAAGGTGTTAATTAGGAATGCATCAAAAAAATAA..............................................

34c_contig34
AGTTAACTTTTTTGAGTATTTTAACGTTGAAAAAATA..............................................
```

Figure 3.4 Alignment of contig 33 with the sequences that were generated from contig 32, contig 8, and contig 34 respectively. Alignment was done manually in BioEdit® Sequence Alignment Editor (v 7.0.5.2, Hall).

3.2.3.3. Investigation into the Ms02 genome repetitive elements

3.2.3.3.1. Large genome repeats

The single contigs were used as query sequences in a local BLAST against all of the Ms02 genome contigs. Repeats were noted only when the identity was 80% or higher. Sizes ranged from 18 to 396 bp. The total size of these repetitive sequences amounted to 11 310 bp. This is not an indication of the total amount and extent of the repeats, since the true number of copies was not determined, but rather just the presence of a repeat. The repeats accumulated to 116 different units, i.e. not taking into account the number of times a single repeat occurs in the genome.

Repeats of 80 bp or more in length, with 80% or more identity to at least one other copy in the genome, were defined as being "large". Only 17 such large repeats were found, ranging from 82 to 396 bp in length. The repeats were found to be mostly gene duplications. Only one repeat was not associated with any coding sequence, and one repeat was associated with a hypothetical protein coding sequence. The repeats appear to be mostly aminopeptidases and Type I Restriction Modification systems (RMS). Only one large repeat is associated with a tRNA gene (Table 3.6).

Reasons for the small number of large repeats include the mechanism of analysis. Numerous repeats were found to occur adjacent to one another. The program recognised these repeats as single segments, while the large comprehensive repetitive sequence is a more correct representation of the repeats. Manual analysis of these regions identified repeats of more than 1 000 kb. Generally, large repeats are associated with lipoproteins, RMSs, and unknown function open reading frames (UFOs). The Ms02 genome adheres to this generalisation, with five of the genes associated with large repeats being Type I Restriction Modification System genes.

3.2.3.3.2. Simple sequence repeat searches using SPUTNIK

Simple sequence repeats (SSRs) are defined as being repeating units of 2-6 bp in length, also known as microsatellites, and are a type of variable number tandem repeat (VNTR). Thirteen contigs were found to contain microsatellites using the SPUTNIK program. Each SSR was treated as a separate repeat, i.e. the complements were not grouped together. For instance, CA was not grouped with GT repeats. Each repeat was also treated separately per contig. In Table 3.7, the type of SSRs is indicated, together with their respective lengths. In the first group, the di-nucleotide repeats, the microsatellites TA and GA occur on two different contigs with different lengths, and they are therefore treated separately. There are 30 different types of SSRs in the Ms02 genome, distributed as follows: two di-nucleotides, twelve tri-nucleotides, seven tetra-nucleotides and nine penta-nucleotides. The total number of these repeats is 39, with a total length of 545 nucleotides, ranging from 11 to 21 bp.

Table 3.6 Large repetitive sequences occurring in the Ms02 genome and their gene association.

Contig	Repeat contig	Size (bp)	E-value	% Identity	Gene association
3	10	108	$1.02e^{-15}$	84	HsdS
	44	99	$9.82e^{-13}$	84	/
4	30	115	$1.14e^{-53}$	98	tRNA-Gly-TCC
7	17	131	$1.25e^{-55}$	95	Cellobiose-related proteins
	30	152	$8.58e^{-26}$	85	Aminopeptidase
	30	141	$3.15e^{-19}$	84	Aminopeptidase
	30	103	$2.89e^{-20}$	87	Aminopeptidase
	12	114	$1.85e^{-03}$	80	ATP synthase alpha chain
8	14	88	$2.78e^{-09}$	84	Aminopeptidase
9	44	359	$7.56e^{-48}$	82	Type I restriction modification system
	11	341	$2.88e^{-44}$	82	Type I restriction modification system
	11	82	$8.44e^{-05}$	82	Type I restriction modification system
11	44	396	$1.92e^{-70}$	83	Type I restriction modification system
	39	116	$2.99e^{-32}$	90	Type I restriction modification system
13	1	194	$1.62e^{-17}$	81	Hypothetical protein
30	41	92	$4.48e^{-16}$	87	DNA-cytosine methyltransferase
	41	87	$2.56e^{-08}$	84	DNA-cytosine methyltransferase

Table 3.7 The sequences, distribution and lengths of Ms02 genome simple sequence repeats.

Repeat type							
Di		Tri		Tetra		Penta	
Repeat group	Repeat length	Repeat group	Repeat length	Repeat group	Repeat length	Repeat group	Repeat length
GA	20 + 11	TTC	12 + 12 + 13	ACAA	13	CTTTT	14
TA	20 + 17	TCC	13	TTAT	13	TTTTG	21 + 14
		ACC	12	TTTG	13	ATGAA	14
		TAA	12 + 12	TAAA	13	ATAAT	12
		GTT	12	TATT	13	TTGTT	14
		CTG	12	TTGA	14	AAATG	15 + 14
		GCT	13 + 15	AATT	21	TGTTT	14
		TGT	13			AGAAA	15
		ACT	15			AAAAT	14
		AAT	12 + 12				
		GAA	12				
		CAG	12				

3.2.3.3.3. Tandem Repeat Finder

Tandem repeats are patterns of two or more nucleotides that form a series of repetitions adjacent to one another. Tandem repeats shorter than 10 nucleotides are called microsatellites, while repeats between 10 and 60 nucleotides are known as minisatellites. The Tandem Repeat Finder (TRF) was used to detect these repeats in the Ms02 genome. Doing the analysis on all of the contigs at once by linking them numerically resulted in a total of 1 218 repeats over the 900 kbp sequence. Copy numbers of individual repeats range from 1.9 to 34.8, and the sizes of the repeats range from 14 to 99 bp. Although the definition of a repeat implies at least two copies of a given sequence, this program did not work on absolute values. The possibility of insertions and deletions was also analysed and taken into account, and as such, the lowest copy number found was 1.9. All of the repeats have an alignment score of 50 or higher. The size of the repeats represents 8.2% of the genome. The distribution of nucleotides in the repeats supports the A+T-rich nature of mycoplasma genomes: 39% of the repetitive sequence are T_n, 35% are A_n, and C_n and G_n are 13% each.

3.2.3.3.4. Single base pair runs in the Ms02 genome

The contiguous sequences of the Ms02 genome were subjected to manual searches for single nucleotide repeats. A minimum starting value of 6 bp was used (Gur-Arie *et al*., 2000), but due to the A+T-richness of the mycoplasma genome, this was increased to 8 bp for A_n and T_n repeats. True to the A+T-rich nature of mycoplasma genomes, G and C runs are underrepresented. There are only two repeats of G_n^6 (6 repeating guanine nucleotides), and 3 repeats of C_n^6. This is in sharp contrast with A_n and T_n runs.

There are a total of 50 repeats of A_n runs of 8 bp or more. The Ms02 genome contains 5 repeat sequences with more than 8 bp A_n runs: A_n^9, A_n^{12}, A_n^{17}, and two A_n^{16}. The repeat sequence of the two 16 bp runs is similar, excluding one base pair (AGCAAAAAAAAAAAAAAACA-A/G). In addition to these repeats on the single contigs, the linkage of contig 32, 33, 34, and 8 also exposed A_n runs, which is suspected to contribute to the difficulty of finalising these linkages. Between the contigs 8 and 33 part of the additional sequence that was produced, an A_n^7 run occurred, i.e. AAAAAAATAA. This sequence was directly followed by the 5' end of contig 33, which contained the exact repeat. However, the additionally produced sequence contained one less A both before and after the T_n, between the alignment of contig 32 with 33, i.e. AAAAAATA. The same occured on the alignment between contigs 34 and 33.

The T_n nucleotide repeats of 8 bp or more amounted to 57, with two T_n^{15}, one T_n^{13}, three T_n^9, and four T_n^{14}. Many of the larger repeats contained the same three base pairs at the end of the run, and differed only in the number of T_n (as well as A_n, as mentioned above). The following sequence serves to demonstrate the extent of the runs, and reminds one of the problems when attempting to assemble the genome: CTGTTTTTTTTTTTTTTGTTTTTTGCAGTTTTTTT. Runs such as these complicate the assembly of reads during *de novo* sequencing, as well as complicate primer design for use in PCR-based chromosome walking techniques, such as TAIL-PCR, to assemble individual contigs.

The repeats that were 9 bp or higher were investigated to determine their gene association. Most were found to occur in intergenic regions, some close to the genes, both upstream and downstream. Two repeats are associated with hypothetical proteins (T_n^9 and A_n^{17}), one repeat occurs in a 5S RNA gene (T_n^9), and one occurs

near the end of a maltodextrin ABC transporter permease protein MalC (A_n^9). According to Guo and Mrázek (2008), both *M. mycoides* and *M. capricolum* contain T_n runs upstream from their SSR-associated ABC transporter genes. The Ms02 genome contains five T_n runs upstream from ABC transporter genes, located between 190 and 534 bp upstream, and one T_n run inside an ABC transporter gene, the ABC transporter xylose-binding lipoprotein.

3.2.3.4. Bioinformatic analysis of the contiguous sequences

The gene predictions were performed on contigs A, B, C, and D (that resulted from the TAIL-PCRs), as well as on the sequences that were force-linked by analysis of their gene distribution patterns. Thus the gene prediction was only done after the linkage of the contigs. First, the contigs were assembled into one contiguous sequence, based on this information. The contiguous sequence was then submitted to gene prediction programs. Only the ORFs that were located across two contigs were subjected to identity searches using the BLASTx tool.

3.2.3.4.1. Directional orientation of contigs based on predicted gene orientation

The individual contiguous sequences were analysed based on the previous ORF predictions and annotation results (Steenmans, 2010). The total amount of ORFs, the location of essential genes, and the gene distribution on the plus and minus strand, were taken into consideration in order to predict the direction of the contig. The essential gene categories (as shown in Table 3.8) were chosen based on the essential gene analysis of twelve mollicute genomes, as done by Souza *et al.* (2007). To categorise the contigs, both essential genes and total ORFs were taken into account (Addendum A). Only contigs with 10 or more ORFs were considered. From these, if the percentage genes on each strand were too close to one another, and if the comparison between essential gene and total gene percentages cancelled out, it was deemed that not enough information existed to make a reasonably confident prediction of the orientation of the contig. Of the 25 annotated contigs, only 11 contigs' direction could be predicted confidently. Contigs 16, 37, and 39 were not annotated previously, because of their small size. Three of the 11 contigs remained in the current direction, the other 8 were predicted to occur in the opposite direction. However, this method of prediction is not infallible, and to test its validity, the contigs that were already linked were investigated. Three sets of newly generated sequences, categorically named contig B (linked contig 7-17-31-37-39), contig C (linked contig 8-15-38-4-41-10), and contig D (linked contig 34-33-32-1-26), were analysed to determine if there was any merit in the predictions. Contig A was deemed too short to be of significance (2 027 bp). Of contig C, all of the individual contigs were linked in the same direction as their respective predictions. Certain other factors were taken into account, such as the orientation of specific genes. The Opp operon, containing the OppA protein, occurs on the negative strand of contig 8. The analysis of total genes and essential genes with regard to their direction predicted contig 8 to be in the opposite direction. Thus, this was taken as enough evidence to orientate the contig opposite to what it was, currently correctly indicated in Section 3.2.3.2.3. In addition, contig 8 is part of a linked contig, where the total direction of the combined contig also indicated the validity of the prediction. Contig B could not be validated, since 2 of the linked contigs were not annotated, due to their small size, and the other 3 could not be predicted as there was not enough information available. Contig D provided mixed results. Contig 34, part of contig D, contains a ribosomal cluster with 31% of the total ribosomal genes. According to Necşulea and Lobry (2007),

most of the ribosomal protein coding genes are transcribed in the same direction as DNA replication, this being the leading strand. Due to the fact that 20 ribosomal genes, forming a gene cluster, are transcribed on the plus strand of contig 34, it was decided to uphold this contig's orientation. Taking this into account, the rest of contig D adhered to the method of prediction, except for contig 1. This was taken as a good starting point of predicting strand orientation. Only two contigs' strand orientation differed after taking into account the orientation of the ribosomal genes and rRNA genes, namely contig 13 and contig 33, both changed from a "no prediction" to a "plus prediction".

Table 3.8 Distribution of the different categories of essential genes in the Ms02 genome.

Category	Total genes	Genes on plus strand	Genes on minus strand
Amino acid transport and metabolism	18	7	11
tRNA genes and related metabolism	62	27	35
Cell division	24	6	18
Nucleotide transport and metabolism	18	7	11
Ribosomal proteins and metabolism	68	45	23
Lipid metabolism	11	0	11
Phosphate metabolism	4	1	3
Inorganic transport and metabolism	8	1	7
Glycolysis and gluconeogenesis	9	5	4
Carbohydrate transport and metabolism	18	10	8
DNA repair	9	4	5
Pentose phosphate pathway	2	2	0
Total:	251	115	136
Percentage:		46%	54%

3.2.3.4.2. Provisional linkage of contigs based on the gene content on the contig ends

Using the assumption that certain genes must occur adjacent to one another, the remainder of the contigs were investigated to determine if the genes could be used to predict the alignments. PCR primers were designed on those contigs for which primers had not been designed yet. Three 'gene sets' were detected on the contig ends. Five contigs contained a Type I Restriction Modification system, or subunit, regarding DNA specificity (contigs 2, 7, 9, 11, and 44). Three contigs contained a gene coding for phosphotransferase or cellobiose components (contigs 7, 9, and 17). Two other contigs contained genes involved in deoxyribose and deoxynucleoside metabolism (contigs 8 and 14). However, it was not possible to prove a linkage between any of these contigs. This is in line with the observation that mycoplasma genes are not conserved in terms of their position or relative order in the genome. In addition to this, it is suspected that the repetitive nature of the genome further complicated positive results. Figure 3.5 illustrates this point. A standard amplification PCR was performed using forward and reverse primers designed on contig 14 in combination with primer contig_8a. Numerous products resulted in both reactions.

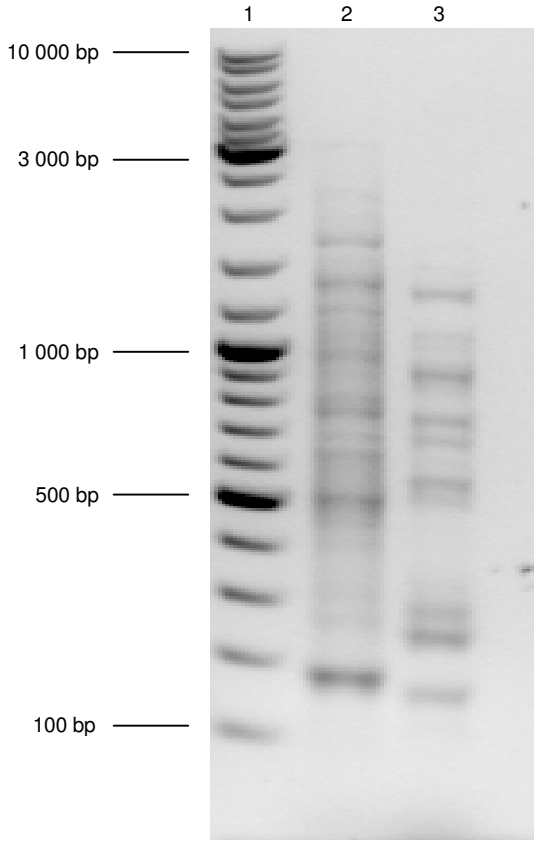

Figure 3.5 Image of the PCR products obtained after electrophoretic separation depicting the repeats between contig 8 and contig 14. The PCR products were separated on a 1% (w/v) agarose gel with 0.1 µl/ml GelRed™ Nucleic Acid Gel Stain for UV visualisation of the DNA. Lane 1: 1 KB DNA GeneRuler™ DNA ladder mix. Lane 2: product formed with the primer pair contig_8a and contig_14F. Lane 3: product formed with the primer pair contig_8a and contig_14R.

3.2.3.4.3. Assembling the single contigs into one contiguous sequence

Aftre the linkage of the contigs using TAIL-PCR, and the linkage that resulted from focused amplification PCR, the contig number was reduced to 14 large contigs. The 14 contigs were assembled into one large contiguous sequence in order to allow gene prediction as a whole. In determining the sequence in which to link the contigs, the information of the above findings was taken into account, i.e. not to link contigs that has preliminary been proven not to occur adjacent to each other. The contigs whose direction was predicted with confidence to be in the minus direction were inverted to give the reverse complement (RS) sequences. Additional sequences (AS) that were produced, even though no significant hits were found with other contigs, were also included. These occurrences were unexpected, since most of the additional sequences were quite long. For instance, the produced sequence on contig 6 was 240 bp, on contig 38 it was 350 bp, and on contig 13 it exceeded 500 bp, but no alignment with other contigs was detected. Table 3.9 indicates the sizes of the old contigs as well as the newly linked contigs from smallest to largest. For the above-mentioned reasons, the following sequence was used for the final linkage: A-2(AS)-6(AS)-B-C-9-11-12-13(AS)-14(RC)-29(RC)-30(RC)-D-44.

Table 3.9 The 14 large contigs obtained from whole-genome GS FLX sequencing of Ms02, as well as the linked contigs, using TAIL-PCR.

Contig ID	Contig length (bp)
Contig00002_AS	1 347
Contig00011	1 800
Contig_Ms02_A	2 027
Contig00044	5 319
Contig00009	5 851
Contig00006_AS	25 801
Contig00012	39 046
Contig00014_RC	67 524
Contig00013_AS	72 293
Contig_Ms02_D_RC	105 763
Contig00029_RC	111 665
Contig_Ms02_B	119 099
Contig00030_RC	127 294
Contig_Ms02_C_RC	214 166

3.2.3.4.4. Searching for the origin of replication

The Ori-Finder detected an *ori*, including three DNA box motifs, located in the region of 866 359 – 866 821 bp in the Ms02 genome, which corresponds to the end of contig 32 just before the alignment with contig 1. The Ms02 DnaA box motifs detected are: TTTTCCAAA, TTTGGATAA, and TGTTGAAAA. The A+T content of the *ori* is 68%.

DoriC is a database containing numerous other *ori*s from a vast array of genomes. The *ori* hit most related to Ms02 was with *M. hyorhinis*, with an Expect value of $8e^{-10}$. The sequences have 84% identity and a score of 89.7. This sequence occurs in the region of 561 223 – 561 339 bp in the Ms02 genome, which corresponds to contig 29, about 10 000 bp into the contig.

Sequences for both predicted *ori*s can be seen in Addendum B. The distance between the two predicted *ori* locations gives an indication of the nature of bioinformatic analysis: although largely useful, manual curation and experimental proof are still required to finalise the predictions.

3.2.3.4.5. Nucleotide skews of the Ms02 genome

The GraphDNA program was used to generate a graphical representation of the raw Ms02 DNA sequence. This allows the graphical analysis of different types of genomic skew, including purine-pyrimidine skew, keto-amino skew, G+C skew, and A+T skew (Figure 3.6a – Figure 3.6d).

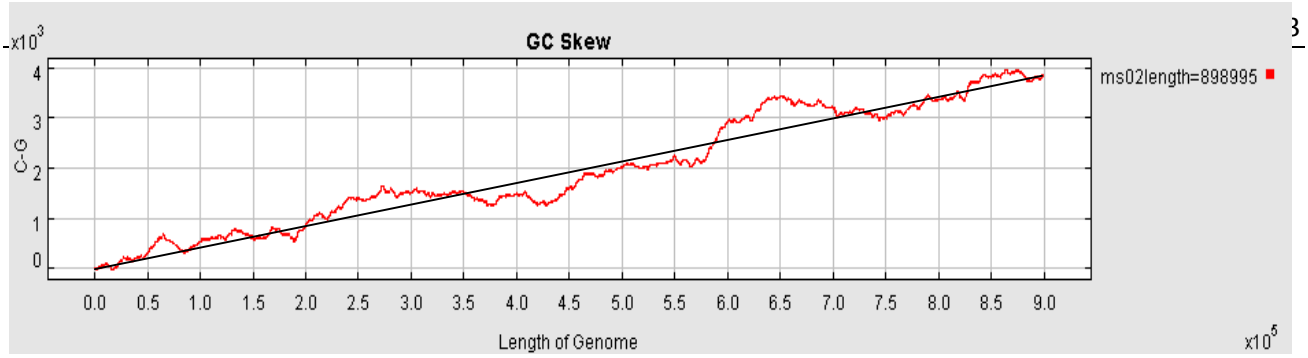

Figure 3.6(a) Graphical representation of the differential G+C distribution along the Ms02 genome as calculated by GraphDNA.

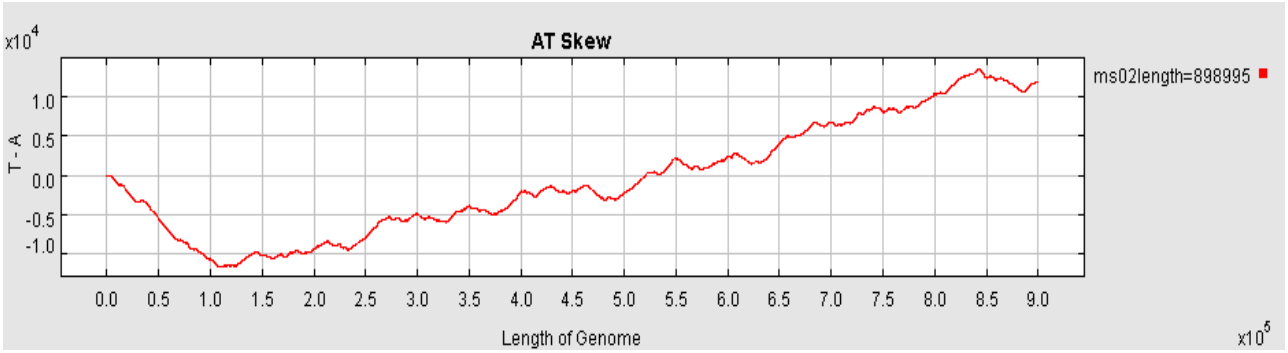

Figure 3.6(b) Graphical representation of the differential A+T distribution of the Ms02 genome as calculated by GraphDNA.

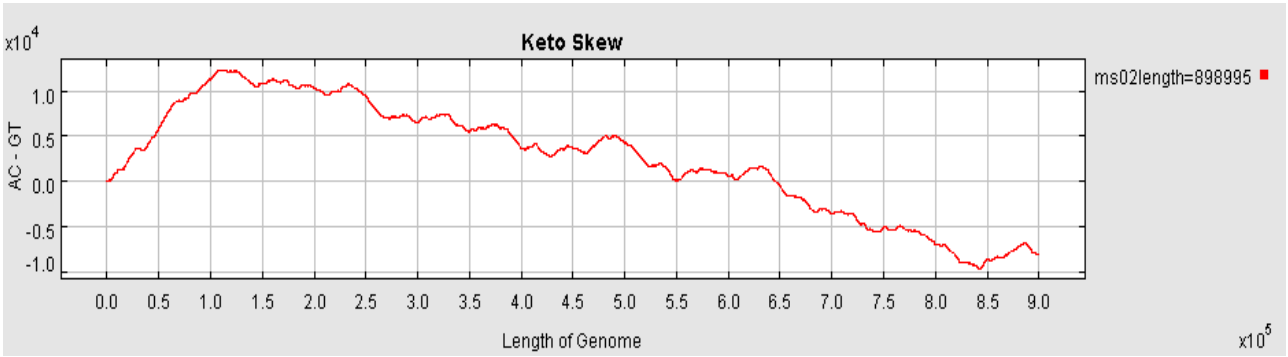

Figure 3.6(c) Graphical representation of the keto-amino distribution of the Ms02 genome as calculated by GraphDNA.

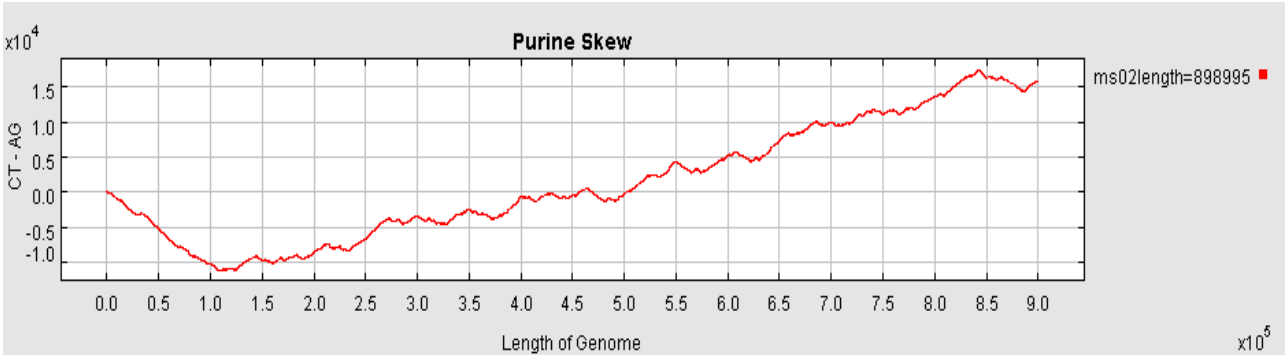

Figure 3.6(d) Graphical representation of the purine-pyrimidine distribution of the Ms02 genome as calculated by GraphDNA.

A steady increase in the G+C skew is observed (Figure 3.6a). In this case the G+C skew is indicated by a general trend line that increases linearly. No clear point is detected where a change in the direction of the G+C distribution occurred. The A+T skew is a better demonstration of a change in direction in the distribution of nucleotides across the genome, as can be seen in Figure 3.6b. The predicted *ori* occurs just as the A+T skew starts to decrease (change in direction). At the start of the graph, the A+T skew starts higher, decreases, and then rises again. If the vantage point of the graph is changed, it will form a clear peak around the *ori*, with a definite change in the direction of the A+T skew. From these graphs it is clear that the A+T skew is a better representation of the predicted *ori*, as happens in A+T-rich genomes. The same trend is visible in the purine-pyrimidine skew and the keto-amino skew, where a turning point is observed at 866 000 bp. The Ori-Finder program predicted the *ori* to be located around 866 000 bp. DoriC predicted the *ori* to be located at 561 000 bp. None of the graphs present a turning point at this location, possibly indicating that this is not the correct *ori*, but instead the one predicted by Ori-Finder.

3.2.3.4.6. Gene predition of the Ms02 genome

Four different automated gene prediction programs were used to search for ORFs in the newly generated contiguous sequence of the Ms02 genome: the ORF Finder available from the NCBI website; GLIMMER from the same website; GeneMark; and ZCURVE, an algorithm that forms part of *ori* finder software. The results were compared and all the ORFs that were located at the joining point between two contigs, or as part of the newly generated sequence from the TAIL-PCRs, were submitted to a BLASTx search to determine any significant alignments with a known or predicted gene function. The individual results can be seen in Addendum C. In this manner, 18 ORFs were analysed. The largest group of ORFs was associated with Type I Restriction Modification systems (8/18 ORFs). The ORFs were analysed as a means of proving the linkages, as determined by the TAIL-PCRs. Linkages that were confirmed with the ORF analysis are: 7-17, 37-39, 8-15, 1-26, and 32-1. Additional ORFs that spanned across two contigs that were provisionally linked (but not experimentally) include: 16-2, 6-7, 11-12, and 26-44. However, these cannot be seen as a definite indication of a linkage, because of the predictive nature of the ORF finding software, and due to the small size of some of these contigs, such as contig 2 (1347 bp), where an ORF that spans the contig over the ends is somewhat expected.

3.2.3.4.7. PCR of the final linked contigs

A PCR was performed, where primers were available, to indicate the true alignments of the final linked contigs. As can be seen in Figure 3.7, not many reactions did in fact form a product. From the 7 reactions that did amplify DNA, one was not experimentally proven prior to this. In lane 11 a clear and intense band formed using the primer pair contig_10c and contig_9R. The size of the amplified product indicates that there is most likely another contig between contigs 10 and 9. The same can be said for the linkage between contigs 32 and 33. The size of the amplified product here is well over 3 000 bp, which is odd, since the TAIL-PCR results aligned the two contigs with only 8 additional base pairs. A similar result occured when trying to finalise which contig links to contig 33 (Section 3.2.3.2.3). This might be an indication that this alignment is in fact incorrect, since the problem with contig 33 affects so many other alignments, it is possible that another contig is supposed to align between contig 32 and contig 33. The results of this PCR correspond well with the results from the additional ORFs detected on the contig ends. Unfortunately the contigs that contained an ORF located across the ends

and that were not previously linked, did not contain primers to determine if the ORF distribution proved the existence of a linkage. An ORF was detected across contigs 16 and 2, although, as can be seen in Lane 2, no bands were formed in the corresponding PCR reaction. This possibly occurred due to the small size of both contig 16 and contig 2: 809 bp and 1 347 bp long respectively. Contig 16 was not annotated previously, due to this fact. Another example is the ORF spanning contigs 6 and 7, although no product was formed in a PCR between these two contigs to corroborate the ORF, as can be seen in Lane 3. Two other incidences like this occurred, but without available primers to verify: a slight overlap of an ORF occurs on contig 11 onto contig 12, and an ORF that links contig 26 with contig 44.

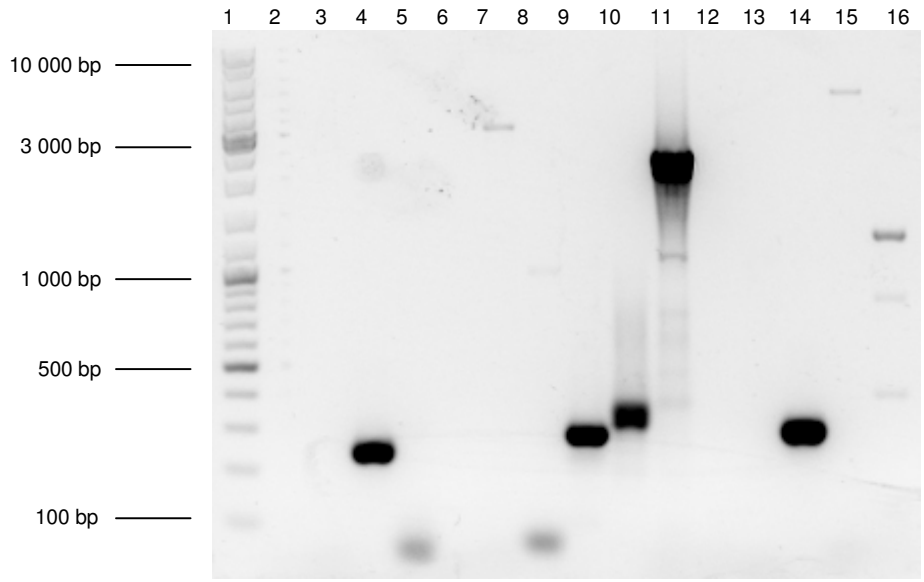

Figure 3.7 Image of the PCR products obtained after electrophoretic separation depicting the alignment between the different contigs linked into one. The PCR products were separated on a 1% (w/v) agarose gel with 0.1 µl/ml GelRed™ Nucleic Acid Gel Stain for UV visualisation of the DNA. Lane 1: 1 KB DNA GeneRuler™ DNA ladder mix. Lane 2: primer pair contig_16c and contig_2c. Lane 3: primer pair contig_6c and contig_7R. Lane 4: primer pair contig_7b and contig_17R. Lane 5: primer pair contig_31c and contig_37c. Lane 6: primer pair contig_37R and contig_39c. Lane 7: primer pair contig_39z and contig_8Rc. Lane 8: primer pair contig_8b and contig_15Fc. Lane 9: primer pair contig_15R and contig_38c. Lane 10: primer pair contig_4z and contig_41F. Lane 11: primer pair contig_10c and contig_9R. Lane 12: primer pair contig_9F and contig_11R. Lane 13: primer pair contig_14F and contig_13Fc. Lane 14: primer pair contig_33c and contig_34R. Lane 15: primer pair contig_32c and contig_33R. Lane 16: primer pair contig_1c and contig_32z.

3.2.4. DISCUSSION

The Ms02 genome was previously subjected to GS FLX 454 sequencing that resulted in 28 large contiguous sequences, ranging from 635 bp to 126 294 bp (Steenmans, 2010). The estimated genome size was 900 kbp (895 119 bp) and was sequenced with 99.78% accuracy. It was decided to employ the technique of TAIL-PCR to attempt assembly of as many as possible of the 28 contigs into larger contigs. While the TAIL-PCR technique proved very effective initially, it soon became apparent that the presence of repeats greatly complicated the prediction and alignment of the contigs. Even though 18 contigs have been linked, resulting in a new total of 14 contigs, many of the alignments contained contradictory results, and some aligned with more than two other

contigs. This is thought to be due to the difficulty associated with primer design in A+T-rich genomes such as mycoplasmas. It is difficult to design primers without runs or secondary structures. However, in spite of the difficulties that were faced with the TAIL-PCRs in this study, the Ms02 genome's contiguous sequences were reduced by half.

Attempts to link the contigs based on gene ends proved ineffective, in line with the observation that gene order in mycoplasma genomes are not necessarily conserved (Mushegian and Koonin, 1996b; Rocha and Blanchard, 2000). Essential genes are predicted to occur on the leading strand, and this proved to be a good starting point for predicting the orientation of single contigs in the Ms02 genome (Necşulea and Lobry, 2007). When this method of prediction regarding essential genes and ORF distribution was applied to the contigs that were already linked, it correctly predicted the relative directions of the linked contigs of contig A and contig C. The distribution of ribosomal genes was also compatible with these predictions, further supporting this as a preliminary mechanism of predicting strand orientation. Therefore, the distribution of ORFs and essential genes on the individual contigs were deemed to be useful in predicting the relative direction of the contigs. This will allow the design of more focussed primers, quickening the assembly process of contiguous sequences.

The 14 new contiguous sequences now range from the smallest 1 347 bp (contig 2 plus additional sequence) to the largest 214 116 bp (contig A). These sequences were then analysed with different programs to determine the extent of the repetitive nature of the Ms02 genome. Large genome repeats were searched using the CLC Combined Workbench, which detected repeats up to 400 bp. However, manual sequence analysis during the alignment and assembly of contigs, and sequences generated with TAIL-PCR, indicated that this might be an underestimation of the size of large repeats present in the Ms02 genome. Further studies and analysis will have to be done in order to obtain a better representation of the size of the repeats, as well as the true distribution thereof, since the programs did not allow for detection of iterations of the same repeating unit.

Simple sequence repeats (SSRs) are generally rare in most bacterial genomes, but quite common among mycoplasmas (Mrázek, 2006; Mrázek *et al.*, 2007). SSRs (repeating sequences of 2-6 bp) were searched for by using SPUTNIK, which detected 39 different SSRs. These repeats ranged from 11 bp to 21 bp in final size, and consisted of di-, tri- tetra-, and penta-nucleotide repeats. Mononucleotide repeats were searched for manually on the individual contigs. The mononucleotide repeats of A_n and T_n were extensive, even with an inclusion start value of 8 bp, a repeat size that was deemed to be underrepresented by Metzgar *et al.* (2002) due to the tendency of SSRs to contract in the absence of selection. Guo and Mrázek (2008) found T_n runs upstream of ABC transporter genes, both in *M. mycoides* and *M. capricolum*. In the Ms02 genome, T_n^6 runs were found to occur within 190-540 bp of ABC transporter genes. *M. hyopneumoniae* and *M. pulmonis* contain short AT_n runs that are periodically spaced at 11 bp and 12 bp respectively. This may likely affect the supercoiling ability of the DNA molecule (Mrázek, 2006). The roles of these runs appear to be independent of up- or downstream location of genes.

The mononucleotide repeats contributed to the difficulty of assembly: where contigs were linked with a number of bp runs, such as A_n in the instance of the linkages with contig 33, the exact number of single nucleotide repeats is difficult to determine, and when these repeats occur inside a gene, frameshifts may occur. In order to assemble the Ms02 genome more accurately, a good idea would be to resequence using paired-end technology,

which would give an indication of the direction and sequence of the contigs, allowing for better linkages and focused searches using TAIL-PCR and direct PCRs.

Nucleotide skews have long been used as a method of predicting the origin of replication in bacterial genomes (Lobry, 1996a, 1996b). Analysis of the different skews of the Ms02 genome indicated the applicability of using such skews in *ori* predictions. The A+T skew, keto-amino skew, and purine-pyrimidine skew, supported the *ori* prediction as calculated by the Ori-Finder. In A+T-rich genomes such as mycoplasmas, the A+T skew is often a better indication of the *ori* location than G+C skew, and this has been demonstrated with the Ms02 genome as well. An interesting analysis could result from using Oriloc, which only examines the third codon position to determine the skew. It has been shown that in some species the direction of the skew is opposite to that for the third codon position in genes, observed in *B. subtilis* for the A+T skew and in *M. genitalium* and *M. pneumoniae* for the A+T and G+C skew (McLean *et al*., 1998).

Analysing the ORFs over the contig ends, for those that were linked provisionally and those that were linked with the TAIL-PCRs, firstly confirmed the results of the TAIL-PCRs (that resulted in a reduced number of 14 contigs), and secondly predicted the additional linkage between contigs 10 and 9, contigs 6 and 7, and contigs 26 and 44. There probably exists another contig in between contigs 10 and 9, due to the large product that was formed from a focused PCR. The PCR also exposed a large product formed between contigs 32 and 33, which indicates the probability that the predicted linkage is incorrect, and that another contig should be linked to contig 33. Further analysis is therefore imperative to prove and establish the correct linkages, not just for this group of contigs, but for all of the contigs. Although this is not a definite indication of a linkage, it is worth exploring to determine if the ORFs that were detected, are correct. However, only the ORFs that occured over two newly linked contigs were anlaysed, and not the ORFs that were detected inside the contigs. Therefore, direct comparisons with the previous annotation results cannot be drawn. The previous annotations were performed on the ORFs that were detected on individual contigs, thus no ORFs would have occured across two contigs' ends.

3.2.5. CONCLUSION

The Ms02 genome was previously sequenced using GS FLX Roche 454 sequencing technology, which resulted in 28 large contiguous sequences. This study used a technique called TAIL-PCR to reduce the amount of contigs to 14 large contigs, contributing to the completion of the genome. These contigs were then subjected to bioinformatic analysis, such as different repeat searches, searching for the origin of replication, and analysing the additional open reading frames that were located on the overlapping sequences of the linked contigs. The study contributed by providing a richer knowledge of the content of the Ms02 genome by emphasising the difficulties that repetitive sequences can cause in assembly and annotation; and illustrating the techniques that can be used for assembling genomes.

Chapter 4

This chapter starts with a discussion on DNA vaccines and how the vector components of the vaccine may influence the expression of the recombinant gene. This is used as an motivation for choosing the three DNA vaccine vectors developed in this study. This is followed by an explanation of the different conditions and optimisation strategies used to express recombinant Ms02 OppA protein and the difficulties that accompanied the purification of this protein. Lastly, the experimental procedures and results will be given in an article format.

4.1. DNA VACCINE VECTORS AND THE CLONING, EXPRESSION AND PURIFICATION OF THE Ms02 OppA PROTEIN

4.1.1. INTRODUCTION

Mycoplasmas infect a wide range of animal hosts, often colonising the mucosal surfaces of mammals and birds. The South African ostrich is susceptible to infection by three distinct ostrich mycoplasmas, provisionally named Ms01, Ms02 and Ms03 (Botes *et al.,* 2005a). These three mycoplasmas cause infection of the upper respiratory tract, leading to retarded growth, downgrading of the carcasses, increased treatment costs, and ultimately a loss in production. In addition to the local impact on the South African Ostrich Industry, concerns regarding transmission of mycoplasmas to other countries through contaminated meat products, may lead to export constraints, resulting in substantial financial losses (Olivier, 2006). Current methods of controlling mycoplasma infections in feedlot ostriches include strict biosecurity practices and a limited range of antibiotics. Biosecurity practices are difficult to maintain successfully and antibiotics should be used sparingly and do not eradicate infection (Kleven, 2008).

Vaccines against the poultry mycoplasmas *M. gallisepticum* and *M. synoviae* were tested to determine if a response could be elicited in ostriches that would provide protection against the ostrich mycoplasmas (Pretorius, 2009). Antibodies that were elicited only reacted against the specific mycoplasma it was vaccinated against, and not against the ostrich mycoplasmas. The conclusion was that due to the low cross-reactivity of the antibodies, separate vaccines would have to be developed for the three ostrich mycoplasmas. Mycoplasmas are notoriously difficult to cultivate, and along with other practical problems, make it unfeasible to develop a conventional vaccine. To circumvent the disadvantages associated with traditional vaccines, a new method of vaccination was chosen to use in this study: that of DNA vaccination. DNA vaccination is based on the principle that a gene of the pathogenic organism can be cloned into an expression vector and directly injected into the host animal. This results in expression of the antigenic protein and then elicits a protective immune response against this foreign protein.

DNA vaccine development starts with the identification of a vaccine candidate gene. Proteins involved in membrane transport and host cell attachment have been proposed as good vaccine candidate genes, due to their strong antigenic properties (Garmory and Titball, 2004). Sequencing of the Ms02 genome and comparative analysis of the predicted open reading frames (ORFs), resulted in the identification of the *oppA* gene as a potential vaccine candidate gene (Steenmans, 2010). The gene was isolated by PCR from the genomic DNA

and modified by site-directed mutagenesis (Steenmans, 2010). In this study, the modified *oppA* gene was cloned into four different vectors; one prokaryotic expression vector and three different DNA vaccine vectors. The choice of a suitable eukaryotic expression vector is the first and foremost critical step in the development of a DNA vaccine (Gurunathan, 2000), as the efficacy of the vaccine is dependent on the different components of the expression vectors.

4.1.2. DNA VACCINES: BASIC VECTOR REQUIREMENTS, COMPONENTS THAT INFLUENCE EXPRESSION, AND THE THREE DNA VACCINE VECTORS CHOSEN

DNA vaccines are part of next-generation vaccines, an example of the reverse vaccinology approach. The shift from conventional vaccines toward next-generation vaccines ocurred in parallel with advances in bioinformatics. By using the resources available with genome databases, it is now possible to compare and analyse possible ORFs within existing annotated genomes, leading to identification and functional annotation of unknown hypothetical genes. In this manner, possible antigens can be identified based on homology searches with other known antigens, instead of laboriously expressing and analysing each protein from organisms with incomplete and/or unannotated genomes. Thus, reverse vaccinology refers to the development of a vaccine where one starts with identification of antigens *in silico*, and works backwards toward cultivation *in vitro* (Rappuoli, 2001).

4.1.2.1. Basic requirements for a DNA vaccine vector

The different components of the DNA vaccine vector will determine the level of expression, and subsequently the immune response elicited. Most plasmids that are used as vaccines have a shared set of characteristics, including: (i) a bacterial origin of replication with a high copy number for high production of plasmid DNA in transformed bacteria; (ii) a selection marker to allow for plasmid selection during bacterial growth; (iii) a strong eukaryotic promoter for expression in mammalian cells; (iv) a cloning site downstream of this promoter for insertion of the gene(s) of interest (GOI); (v) and lastly a polyadenylation sequence to stabilise and direct expression of the mRNA (Garmory *et al.,* 2003; Gurunathan *et al.,* 2000; Webster and Robinson, 1997). Sometimes introns are also included in the vector, since a lot of mammalian gene expression is dependent on, or increased by, the presence of an intron. Multicistronic vectors may be created to express more than one antigen, or to express an antigen together with an immunostimulatory element (Lewis and Babiuk, 1999).

4.1.2.2. Vector components that influence expression

The structure of a vector can be divided into two conceptual parts: the plasmid backbone unit and the transcription complex unit.

4.1.2.2.1. The plasmid backbone unit

The plasmid backbone unit consists of prokaryotic elements such as an origin of replication, multiple cloning site (MCS), a selectable marker, as well as optional immunostimulatory sequences with adjuvant activity. DNA vaccine vectors may contain an origin of replication from viral origin, such as from the Simian virus 40 (SV40) and the polyoma virus. The MCS is designed with unique and convenient restriction sites to allow insertion of the GOI, and designed to avoid hairpin formation in the 5' end of the mRNA, since this reduces the level of translation in higher eukaryotes (Fu *et al.,* 1991; Kozak, 1986; Kozak, 1989; Rao *et al.,* 1988). The selectable

marker can either be a drug-resistant marker, enabling cells to detoxify an exogenously added toxic substance, or be an auxotrophic marker, enabling cells to synthesise an essential compound not present in the media.

Krieg *et al.* (1994) demonstrated that certain DNA sequences induce cytokine secretion and lymphocyte activation. Certain bacterial species-dependant unmethylated cytidine-phosphate-guanosine (CpG) motifs are particularly immunostimulatory. These sequences were found to activate B cells *in vitro* and act as an adjuvant *in vivo* (Donnelly *et al.,*1997), as well as to activate monocytes, natural killer cells (NKC) and dentritic cells (DC) (van Drunen Littel-van den Hurk *et al.,* 2001). Elimination of the CpG motifs from the plasmid backbone reduced the immunogenicity of the vaccine, an effect that was reversed by co-administering exogenous CpG-containing DNA, although not to the magnitude of including the motifs in the plasmid backbone (Klinman, 1998). Even the selectable marker can have an influence. The human cytomegalovirus (hCMV)-based vectors containing the ampicillin resistance gene (*ampR*) instead of the kanamycin resistance gene (*kanR*), produced a stronger immune response (Sato *et al.,* 1996). This was thought to be the effect of a certain immunostimulatory sequence (ISS), a palindromic CpG hexamer 5' AACGTT 3', which is present twice in the *ampR* gene and absent in the *kanR* gene.

4.1.2.2.2. The transcription complex unit

The transcription complex unit (that drives the synthesis of the GOI) contains a promoter/enhancer region, an intron with functional splicing sites, the GOI, and a polyadenylation signal.

Mammalian expression plasmids mostly carry immediate early promoter/enhancer regions from pathogenic viruses, due to their high transcription initiation ability in mammalian tissues (Harms and Splitter, 1995). The hCMV enhancer/promoter is used most often, due to its induction of strong and consecutive expression in various cell types (Boshart *et al.,* 1985; Schmidt *et al.,* 1990; Thomson *et al.,* 1984). Other promoter/enhancement regions include those from the SV40, the Rous Sarcoma Virus (RSV), the murine leukemia virus SL3-3, the mouse mammary tumour virus (MMTV), and the human immunodeficiency virus (HIV). A study done by Galvin *et al.* (2000) concluded that the level of humoral and cellular immunity induced by a DNA vaccine is directly correlated to the strength of its promoter.

Intervening sequences (introns) have a positive effect on antigen expression, possibly due to an enhanced rate of RNA polyadenylation, or nuclear transport, that is linked to RNA splicing (Huang and Gorman, 1990). Most gene vaccination vectors contain the intron A of hCMV. Chimeric introns, constructed from different donor and acceptor sites, can be optimised at the branchpoint site to match consensus sequences of splicing ,and lead to an increase in the level of expression (Senapathy *et al.,* 1990).

High levels of gene expression are dependent on efficient termination and polyadenylation of RNA transcripts. Polyadenylation, the addition of 200-250 adenosine residues to the 3' end of the RNA transcript, contributes to the stability and translation of the transcript (Bernstein and Ross, 1989; Jackson and Standart, 1990).

For DNA vaccine development, the GOI should be inserted into the vaccine vector to have the maximum immune response. Boyle *et al.* (1998) showed that both cellular and humoral immune responses were enhanced when using an antigen-targeting strategy, by vaccinating with DNA than encodes for an antigen-ligand fusion protein. Another factor to take into consideration is the target of the gene product. Different cellular

compartments will elicit different immune responses, for instance secreted or membrane-bound proteins induce antibodies more effectively than cytosolic antigens. On the other hand, cytotoxic T cell responses are greatly improved by cytosolic degradation and subsequent presentation on the MHC class I (Robinson and Pertmer, 2000).

4.1.2.3. The DNA vaccine vectors pCI-neo, VR1012 and VR1020

In this study the three mammalian expression vectors: pCI-neo, and two Vical vectors, VR1012 and VR1020 were used for DNA vaccine development. All three contain an expression cassette under control of the hCMV-IE enhancer/promoter, an intron, and a polyadenylation signal to regulate expression. To date only VR1020 has been approved by the FDA for human use.

In the pCI-neo vector (Figure 4.1), the chimeric intron is composed of the 5' donor site of the first intron from the human β-globin gene, as well as the branch and 3' acceptor site from an intron of an immunoglobulin heavy chain variable region gene (Bothwell *et al.,* 1981). The sequences were modified to match the consensus sequences for splicing (Senapathy *et al.,* 1990). Both the Vical vectors contain intron A from hCMV.

The pCI-neo vector contains the *ampR* gene, conferring resistance to ampicillin, while the two Vical vectors contain the *kanR* gene, conferring resistance to kanamycin. This allows for plasmid selection during growth in bacteria. Based on a study by Sato *et al.* (1996), this difference could cause the pCI-neo vector to produce a stronger immune response than the Vical vectors.

The VR1012 and VR1020 vectors contain a hCMV untranslated region, which contains intron A and a bovine growth hormone terminator element. VR1020 also contains a secretion signal, the human tissue plasminogen activator (tPA) signal peptide sequence. This is situated at the amino terminus of the expressed protein to facilitate secretion of the expressed recombinant protein to the outside of the cell, when the foreign gene is inserted downstream of and in-frame with the tPA signal. Smooker *et al.* (1999) demonstrated that their VR1020 construct, and not the VR1012 construct, directed expression so that the protein was exported from the cell. They found that when immunising mice with a *Fasciola hepatica* glutathione-S-transferase, an increased humoral response was achieved when the VR1020 vector was used, if compared to the VR1012 vector.

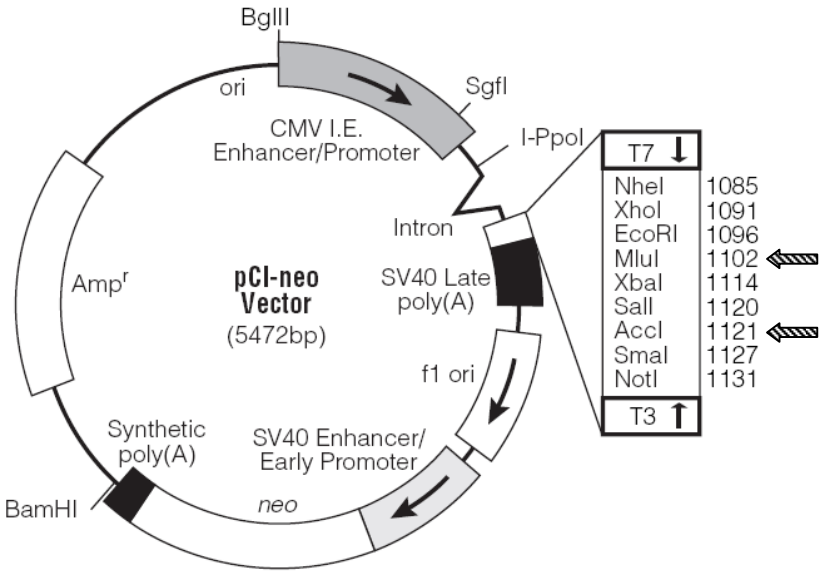

Figure 4.1 The mammalian expression vector pCI-neo (Promega), as an example to illustrate the important vector components. In addition, the two restriction sites that were used for insertion of the *oppA* gene are indicated with striped arrows.

4.1.3. Ms02 OppA PROTEIN EXPRESSION AND PURIFICATION

The *oppA* gene, part of the oligopeptide permease system, was chosen as a possible DNA vaccine candidate gene in previous studies for the other two ostrich mycoplasmas. The three DNA vaccine vectors pCI-neo, VR1012 and VR1020, were chosen to test the antigenicity of the *oppA* gene in different DNA plasmid vectors. For Ms02, expression of the plasmid was to be determined by using a cell culture system (see Chapter 5). In other studies performed in this laboratory, the Ms01 and Ms03 vaccine vectors were tested, or are still being tested, for their efficacy in eliciting an immune response in ostriches. Purified OppA protein is needed to test the serological responses of the ostriches. It is also needed to produce rabbit antibodies against the recombinant protein which can be used to test for protein expression in a cell culture system. Expression of the OppA protein was performed by using a GST-tag expression system to allow for affinity chromatography purification that utilises immobilised glutathione.

After realising that the expression and purification conditions previously used for the Ms01 and Ms03 GST-OppA proteins in *Escherichia coli* did not work for the Ms02 GST-OppA protein, an attempt was made to optimise the expression conditions, with the objective of improving the production of soluble protein. Different approaches towards soluble protein expression were tested; on the one hand to maximise the expression of soluble protein, and on the other hand to solubilise the protein after expression. In order to maximise the expression of soluble protein, different cell types were used for expression, and different inducer concentrations, growth temperatures and time points for inducing and harvesting were tested. Before induction, an additional heat-shock incubation step was included, based on the postulate that inducing the expression of heat-shock proteins to chaperone the OppA protein into correct folding might prevent protein aggregation.

4.1.3.1. Expression of the Ms02 GST-OppA protein

All of the expression optimisations were performed in *E. coli* BL21(DE3) cells (Novagen) and Single Step (KRX) cells (Promega). The KRX cells are inducable with rhamnose and subjected to catabolite repression by glucose, which allows for more precise control of the recombinant protein expression, while the BL21 cells are induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) and are more prone to leaky expression, but is an all-purpose strain often used for its high level of protein expression. In order to determine the optimal conditions for maximum protein expression, different conditions were tested using 10 ml culture volumes. Two different temperatures were tested, i.e. 26°C and 37°C. Two different induction conditions were tested at each temperature. Inducer concentrations were tested based on the respective manufacturers' suggestions for optimising recombinant protein expression. For the KRX cell culture, the cultures were induced either with 1 mM IPTG and 0.05% L-rhamnose, or with 4 mM IPTG and 0.1% L-rhamnose, while the BL21 cultures were induced with IPTG at concentrations of either 1 mM or 4 mM. Even though the KRX cells are induced with rhamnose, IPTG is added because of the *lac* operator that occurs in the pGEX-4T-1 vector, as suggested by the manufacturers. It was determined that the best conditions to use were a temperature of 37°C and an IPTG concentration of 1 mM (with the 0.05% L-rhamnose for the KRX cell cultures). These conditions were then used to determine the optimal time point to harvest the cells. The culture volumes were increased to 50 ml and cell samples were taken every 2 h for 24 h. The results indicated that the optimal approach would be to use the BL21 cells for expression and to harvest the cells 4 h after induction. However, when the small test cultures were scaled up in order to achieve a larger amount of cells, and thus proteins, it was found that the conditions were not optimal for the larger volume. Even when adjusting the time to harvest to 6 h, taking into account that the larger volume means less aeration and slower growth, the proteins still appeared to be mostly insoluble. Since the literature suggested that one should either induce at a low cell density for a longer period of time, or a high cell density for a shorter period of time (Guan and Dixon, 1991; Liobikas *et al.,* 2006), a simple experiment was performed to determine the applicability of this statement to these experiments. Two flasks of 100 ml Luria-Bertani (LB) medium were inoculated with BL21 cell cultures. One culture was allowed to grow for 2 h before inducing the cells with IPTG, and then the cells were harvested after 6 h. The other culture was grown for 6 h before induction, and the cells harvested after 3 h. The expression profiles were analysed with sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), which proved that the protein was more soluble when induced at a high cell density for a shorter time period. Still, using these expression conditions, but further adhering to the basic protocol for sample preparation as described in Sigma's manual on the Glutathione-Agarose product (G4510) (as followed successfully for the Ms01 and Ms03 GST-OppA proteins), did not result in purified protein. The supernatant was too viscous to load onto the glutathione-agarose column, which can result from the nucleic acids that is released upon cell lysis. The viscosity forced batch purification instead of eluting fractions, which resulted in a protein isolate that had high levels of impurities. In addition to these complications, the visualisation of proteins on an SDS gel is an inaccurate representation of the state of proteins in the sample. Since all the proteins are denatured before loading onto the gel, it is possible that many were aggregated and inclusion body proteins were falsely thought to be soluble. Therefore, it was found that further experiments were required in order to optimise the purification of the OppA protein.

4.1.3.2. Purification of the Ms02 GST-OppA protein

Both *E. coli* BL21(DE3) and Single Step (KRX) cells were used to try to express soluble protein and isolate the protein according to the protocol that was used successfully for isolating the Ms01 and Ms03 GST-OppA proteins. This was performed according to Sigma's instructions on their glutathione-agarose column resin. While being able to isolate the GST protein as a control using the same protocol, it was confirmed after numerous attempts that the Ms02 GST-OppA protein was not soluble under these conditions. Protein denaturants such as guanidinium hydrochloride (GuHCl), SDS, urea, and β-mercapto-ethanol were tested to determine their efficacy in disrupting protein aggregation. Using GuHCl at a concentration of 6 M in the buffer increased the solubility of the protein, but the supernatant was too viscous to enable effective column loading. After considering the sample preparation for loading proteins on an SDS gel, the same conditions were scaled up and tested on a larger cell suspension. A buffer consisting of 0.125 M Tris-HCl, 4% (w/v) SDS, and 10% (v/v) β-mercapto-ethanol, was used for denaturation and the cell suspension was boiled in this solution. Unfortunately, though the results appeared promising on a small scale, the buffer cell suspension was unable to separate sufficiently during centrifugation to allow collection of the supernatant (that contained the soluble protein).

Coincidently it was discovered that by repeating the preparation steps of normal SDS sample preparation, i.e. the freeze/thaw cycles and sonication, the protein appeared to be more visible on the SDS gel. It was concluded that a possible explanation for lack of protein in the soluble fractions were due to incomplete lysis of the cells, thus not allowing release of the protein into the external buffer. To increase the release of proteins out of the cells, it was decided to employ lysozyme to digest the bacterial cell wall. To this end, lysozyme was added at a concentration of 1 µl (10 mg/ml) for every 100 µl cell suspension. After an ice incubation step, the original protocol as per Sigma was performed. This resulted in the first successful isolation of protein. Analysis with SDS-PAGE indicated the eluted protein to be impure, as the fractions contained either multiple different proteins, or damaged/truncated proteins, indicated by the multiple protein bands visible on the SDS gel. An explanation was ventured as to this result: damage to the cell walls caused by the lysozyme also caused numerous proteolytic enzymes to be released, resulting in degraded protein. To inhibit the proteases, one tube of Halt™ Protease Inhibitor Single-Use Cocktail EDTA-free (Thermo Scientific) per 10 ml cell suspension was added together with the lysozyme. The result was a narrower peak of protein in the fractions, as well as a marked reduction in the amount of additional bands, as visualised on the SDS gel. The protocol was finally repeated with separate incubation steps for the lysozyme and the protease inhibitors, and resulted in a satisfactory pure isolated protein. The concentrations achieved in the peak fractions were deemed sufficient and although the protein was not completely pure (it still contained degradation products), the results were deemed acceptable to continue.

In this study the Ms02 *oppA* gene was cloned into three mammalian DNA vaccine vectors, namely pCI-neo, VR1012 and VR1020, to test the expression of the gene in a eukaryotic cell system and to provide three ostrich Ms02 vaccines for future efficacy tests. The gene was also cloned into the prokaryotic expression vector pGEX-4T-1 in order to express the recombinant protein for subsequent purification. The purified protein was then used to generate antibodies by injecting pure protein into a rabbit and collecting its sera.

4.2. THE PRODUCTION OF THREE Ms02 *oppA* GENE CONTAINING RECOMBINANT DNA VACCINE VECTORS, PURE Ms02 OppA PROTEIN, AND ANTI-OppA ANTIBODIES

4.2.1. INTRODUCTION

Respiratory diseases in ostriches cause significant economical losses to the South African Ostrich Industry every year. Three ostrich-specific mycoplasmas are responsible for infection of the upper respiratory tract of ostriches, namely Ms01, Ms02 and Ms03 (Botes *et al*., 2005a). Mycoplasmas are small, cell wall-less bacteria known for their ability to cause pathogenesis in a wide variety of hosts (Minion, 2002; Prescott *et al*., 2002). Their small, rudimentary genome contains only the essential genes needed for independent reproduction, but they are dependent on their hosts for nutrients such as amino acids, fatty acids, vitamins, nucleic acid precursors, and other biosynthetic compounds necessary for their survival (Bradbury, 2005; Prescott *et al*., 2002). The mycoplasma's parasitic lifestyle also causes damage to the host due to the sequestering of these compounds, production of free radicals, and general damage to the host cells.

Mycoplasma infection in ostriches is characterised by reduced feed consumption, depression and tardiness of the ostriches, sneezing, coughing, and nasal exudates, ultimately causing decreased viability, downgrading of carcasses and death (Olivier, 2006). The ostrich industry suffers dramatic production losses each year. Currently, there are no effective antibiotics or registered vaccines available against the three ostrich mycoplasmas. Low or no cross-reactivity between the three ostrich mycoplasmas and poultry mycoplasmas also indicated that a vaccine will have to be specific to each ostrich mycoplasma (Pretorius, 2009). Thus an urgent need exists to develop a vaccine against each of the three ostrich mycoplasmas.

In this study the Ms02 *oppA* gene, previously identified to be a possible vaccine candidate gene, was cloned into three different eukaryotic DNA vaccine vectors. The gene was also expressed in a prokaryotic expression system, the recombinant protein isolated, and rabbit antibodies generated against the Ms02 OppA protein. This provided all the necessary components for future vaccine efficacy testing in ostriches.

4.2.2. EXPERIMENTAL PROCEDURES

4.2.2.1. Cloning of the Ms02 oppA gene into the pGEM®-T Easy vector

A pGEM®-T Easy vector (Promega) that contained the modified Ms02 o*ppA* gene (pGEM®-T Easy_OppA), was provided by Shandré Steenmans, Department of Biochemistry, Stellenbosch University. The gene was previously subjected to site-directed mutagenesis in order to change 10 TGA codons to TGG. This was done to prevent premature truncation of the protein, since mycoplasmas use TGA to code for tryptophan (a codon that is recognised by most other organisms as a stop codon). Appropriate restriction endonuclease recognition sites were incorporated into the modified o*ppA* gene for subcloning into the mammalian expression vector pCI-neo (Promega). In this study, the gene was amplified to add different restriction endonuclease recognition sites to enable subcloning into the other two DNA vaccine vectors, VR1012 and VR1020 (Vical Inc. San Diego, CA, USA), as well as into the prokaryotic expression vector pGEX-4T-1 (GE Healthcare, UK).

4.2.2.1.1. Primer design

Primers that contained the appropriate restriction endonuclease (RE) recognition sites were designed using the computer software package Primer Designer (v. 1.01, Scientific and Educational Software), and synthesised by Integrated DNA Technologies (IDT), Iowa City, USA. RE were chosen, firstly according to each expression vector, and secondly not to contain similar recognition sites as in the Ms02 *oppA* gene itself. To this end, two sets of primers were designed, with a forward and a reverse primer per set. For one set, each contained a *BamHI* recognition site, and for the other, a *SalI* recognition site (Table 4.1).

Table 4.1 Nucleotide sequences of the primers used to RE recognition sequences into the Ms02 *oppA* gene, indicated in italics. The recognition cut sites are indicated with arrows, and the initiation codon (ATG) and termination codon (TAA) of the *oppA* gene are indicated in bold.

Primer name	Primer sequence (5'→3')	Tm (℃)	Endonuclease
Ms02_OppA_SalR	*G↓TCGAC***TTA**TTTAGATGAGTTTTCTTCT	53.2	*SalI*
Ms02_OppA_SalF	*G↓TCGAC***ATG**AAATTGAAAAAAAGATTTAAT	52.6	*SalI*
Ms02_OppA_BamR	*G↓GATCC***TTA**TTTAGATGAGTTTTCTTCT	53.0	*BamHI*
Ms02_OppA_BamF	*G↓GATCC***ATG**AAATTGAAAAAAAGATTTAAT	52.4	*BamHI*

4.2.2.1.2. Amplification of the oppA gene to add RE sites

Using the pGEM®-T Easy_OppA vector construct, the gene was amplified to add the *BamHI* and *SalI* endonuclease recognition sites. Three different constructs were generated. To enable cloning into the pGEX-4T-1 vector, a *BamHI* recognition site was added to the 5' end of the Ms02 o*ppA* gene, while at the 3' end a *SalI* recognition site was added. The vector VR1012 required a reverse of these sites, thus a 5' *SalI* recognition site and a 3' *BamHI* recognition site were added. For vector VR1020, a *BamHI* recognition site was added to both the 5' and 3' ends of the *oppA* gene. In brief, each PCR contained 10 µl of 5x High Fidelity reaction buffer, 1 µl of 1 U/µl HiFi proofreading *Taq* DNA polymerase, 1.5 µl of 10 mM deoxynucleotide triphosphates (dNTPs), and 0.5 µl of 25 mM $MgCl_2$ (all from Kapa Biosystems, RSA), as well as 1.5 µl of each respective primer (20 pmol/µl), 5 ng of purified DNA, and enough Milli-Q® water to bring the final volume to 50 µl. The amplification of the gene was carried out in a 2720 Thermal Cycler (v. 2.09, Applied Biosystems Ltd., Warrington, UK), and entailed an initial denaturation step at 95℃ for 2 min, followed by 25 cycles of denaturation at 98℃ for 20 s, annealing at 50℃ for 15 s and extension at 68℃ for 90 s. A final extension step was performed at 68℃ for 5 min. Next, the PCR product was analysed on a 1% (w/v) agarose gel (Molecular Grade Agarose D1-LE, Whitehead Scientific), in 1x TAE buffer (40 mM Tris-acetate and 2 mM EDTA, pH 8.0), to establish that a single PCR product was formed. The gel was cast with 0.1 µl/ml GelRed™ Nucleic Acid Gel Stain (Biotium, Hayward, CA), to enable ultraviolet (UV) visualisation of the DNA. For each 10 µl of amplified product, 1 µl of loading buffer (50% glycerol, 1 g/l bromophenol blue, 0.5 mM EDTA pH 8.0, and 100 mM Tris-acetate, pH 8.0) was added before loading the samples. A 1 KB GeneRuler™ DNA ladder mix (Fermentas) was used to ensure the product was the correct size. The remainder of the amplified gene was then purified, using the DNA Clean & Concentrator™-5 kit (Zymo Research) according to the manufacturer's instructions. The DNA was eluted with 10 µl (60℃) Milli-Q® water. The concentration of the purified product was determined

spectrophotometrically using a Novell NanoDrop® ND-1000 (v. 3.5.1). The concentration (ng/µl) and purity were noted (A_{260}/A_{230} and A_{260}/A_{280}).

4.2.2.1.3. Polyadenylation of the purified PCR product to enable ligation into the pGEM®-T Easy vector

The purified PCR product was used as a template for polyadenylation to create overhangs for cloning, using the pGEM®-T Easy Vector System kit (Promega) as described in the manual. Each 10 µl reaction mixture contained 1 µl of 10x reaction buffer, 0.8 µl of 25 mM $MgCl_2$, 1 µl of 5 U/µl Super-Therm Taq DNA polymerase, (all from JMR-Holdings, USA) 0.4 µl of 2 mM dATP (Bioline), and 4 µl purified blunt-ended PCR product. The reaction was incubated at 70ºC for 25 min and purified, using the DNA Clean Concentrator™-5 kit according to the manufacturer's instructions.

4.2.2.1.4. Ligation of the oppA gene into the pGEM®-T Easy vector

The ligation reaction contained 2 µl of the purified polyadenylated PCR product, 5 µl of 2x ligation buffer, 50 ng of pGEM®-T Easy vector, 1 µl of 3 U/µl T4 DNA Ligase, and enough Milli-Q® water to take the final volume to 10 µl. The reactions were incubated overnight at 4°C, before proceeding to bacterial transformation. In accordance with the manufacturer's instructions, controls were included and consisted of: (a) a background reaction mixture containing no insert; and (b) a positive control reaction containing a control insert, provided with the pGEM® -T Easy Vector Kit.

4.2.2.1.5. Transformation of Escherichia coli JM-109 cells with the recombinant pGEM®-T Easy plasmid

The pGEM®-T Easy plasmid, containing the modified Ms02 *oppA* gene, was transformed into *E. coli* JM-109 competent cells (1 x 10^7 cfu/µg DNA) (Promega), using the pGEM®-T Easy Vector System kit according to the manufacturer's instructions. Two microliters of each ligation reaction were added to 50 µl of *E. coli* JM-109 competent cells. After 20 min of ice incubation, the cells were subjected to heat shock in a 42°C water bath for 45 s and returned to the ice for 2 min, before 950 µl room temperature LB-medium was added to the cells. The cells were then allowed to recover for 90 min at 37°C, while being shaken on an IKA® KS 260 Basic orbital shaker at 220 rpm. Luria-Bertani (LB) broth plates with 15 g/L agar were supplemented with 100 µg/ml ampicillin (Sigma-Aldrich) at 1 µl/ml, 0.1 M (IPTG) (Bioline) at 1.6 µl/ml, and 50 mg/ml X-Gal (Bioline) at 0.8 µl/ml. The cell culture was then plated onto the prepared LB/agar plates to enable blue/white-colony screening, and incubated for 16 h at 37°C before analyses of colony formation.

4.2.2.1.6. Diagnostic PCR to confirm insertion of the oppA gene into the pGEM®-T Easy plasmid

White colonies are indicative of plasmids with inserted DNA and were used in a diagnostic PCR for confirmation of the insert. The two primers, T7-promoter and SP6 (Table 4.2), were used in 10 µl reactions, each containing 1 µl of 10x Reaction buffer, 0.1 µl of 5 U/µl Super-Therm *Taq* DNA polymerase, and 0.6 µl of 25 mM $MgCl_2$ (all from JMR-Holdings), as well as 0.4 µl of 5 mM dNTPs (Bioline), 0.5 µl of each primer (20 pmol/µl), and 6.9 µl of Milli-Q® water. Approximately half of each of the chosen white colonies were scraped with sterile toothpicks and mixed in separate reaction tubes containing the reaction mixture, to serve as the DNA template for the PCR. The amplifications were carried out in a 2720 Thermal Cycler as follows: initial denaturation of 10 min at 94ºC, followed by 25 cycles of 94ºC for 30 s, 55ºC for 30 s, and 72ºC for 30 s, with a final extension step at 72ºC for 7 min. The resultant product was electrophoresed on a 1% (w/v) agarose gel (Section 3.2.2.4).

Table 4.2 Nucleotide sequences of the primers used with the respective cloning vectors.

Vector	Primer name	Primer sequence (5'→ 3')
pGEM®-T Easy		
	T7-promoter	5'-TAATACGACTCACTATAGGG-3'
	SP6	5'-ATTTAGGTGACACTATAGAA-3'
pGEX-4T-1		
	pGEX-F	5'-GGGCTGGCAAGCCACGTTTGGTG-3'
	pGEX-R	5'-CCGGGAGCTGCATGTGTCAGAGG-3'
pCI-neo		
	T7EEV	5'-AGGGCTAGAGTACTTAATACG-3'
	T3	5'-AATTAACCCTCACTAAAGGG-3'
VR1012		
	VR1012-F	5'-CGCGCCACCAGACATAATAG-3'
	VR1012-R	5'-AACAACAGATGGCTGGCAAC-3'
VR1020		
	VR1020-F	5'-CGTCGACAGAGCTGAGATCCTACAG-3'
	VR1020-R	5'-GACACCTACTCAGACAATGCGATGC-3'

4.2.2.1.7. Isolation of the pGEM® -T Easy vector-insert construct

The remainder of those colonies confirmed to contain the insert were scraped with a sterile toothpick and inoculated into 5 ml LB broth, supplemented with 5 µl (100 mg/ml) ampicillin, and incubated for 16 h at 37°C, while being shaken at 220 rpm on an orbital shaker. Invisorb® Spin Plasmid Mini Two (Invitrogen) was used to purify the plasmid from the culture, with slight modifications to the manufacturer's instructions. To elute the plasmid DNA, 50 µl of pre-warmed 60°C Milli-Q® water was added directly onto the filter and incubated for 10 min at room temperature. A final centrifugation step for 1 min at 10 000 x g eluted the purified plasmid. The concentration and purity was determined spectrophotometrically and visualised on an agarose gel (Section 3.2.2.2 and Section 3.2.2.4).

4.2.2.1.8. Sequencing of the plasmid inserts

The purified plasmids were used as the DNA template in a sequencing reaction to confirm successful insertion of the Ms02 o*ppA* gene, as well as to validate that no mutations (other than those introduced by site-directed mutagenesis), were introduced. Six internal primers (Table 4.3), and the two outside primers T7-promoter and SP6 (Table 4.2), were used to create overlapping sequences in order to assemble the entire 2 997 bp of the gene. The reactions were executed with the ABI BigDye® Terminator Cycle Sequencing kit (v. 3.1, Applied Biosystems). Each sequencing reaction consisted of 2 µl Terminator mix, 2 µl Half-Dye mix (Bioline), 3 µl primer (3.3 pmol/µl), and 300 ng plasmid DNA, up to a total reaction volume of 10 µl. Cycle sequencing consisted of 35 cycles of 96ºC for 10 s, 52ºC for 30 s, 60ºC for 4 min, followed by a final extension step at 60ºC for 10 min. The reactions were performed in a 2720 Thermal Cycler. The sequencing reaction products were then analysed with an ABI® 3130 Genetic Analyser (Applied Biosystems) at the Central Analytical Facility (CAF), the DNA sequencing facility of the University of Stellenbosch (RSA). Resulting sequences were manually aligned in

BioEdit® Sequence Alignment Editor (v. 7.0.5.2, Hall) and inspected for mutations, as well as for the correct restriction endonuclease recognition sites. Freeze cultures of the positive plasmids were made, by carefully mixing 250 µl of the overnight cell culture and 750 µl of an 80% sterile glycerol solution, and stored at -80°C.

Table 4.3 Nucleotide sequences of the six internal primers used to sequence the Ms02 *oppA* gene.

Primer name	Primer sequence (5'→ 3')	bp-position	T_m (°C)
Ms02_OppA_T7R	5'-CAGGTCTAATTCTGAATCTTACTTCTGAGG-3'	732-762	56
Ms02_OppA_PP2F	5'-CAGCACGAAGGTCAAACATATAGCTATTAC-3'	625-655	57
Ms02_OppA_PP2R	5'-TCCTTGTGTATATGAGTGTAGAAGTTGATC-3'	1483-1513	56
Ms02_OppA_PP3F	5'-CAGTTGAAGTACTTCTAAACAAACATCACC-3'	1364-1394	56
Ms02_OppA_PP3R	5'-AAGTTGTGGAACACCTAAGTAGTAGTTACC-3'	2242-2272	57
Ms02_OppA_SP6F	5'- TTTCAGTAAGAGATTACTTAGCAGATAGAT-3'	2084-2114	54

4.2.2.2. Subcloning of the Ms02 oppA gene into the mammalian expression vector pCI-neo

A pGEM®-T Easy plasmid that contained the modified Ms02 *oppA* gene was provided by Steenmans (2010). It had a *MluI* recognition sequence on the 5' end and an *AccI* recognition sequence on the 3' end of the gene. In order to subclone the gene into the pCI-neo vector, the gene was excised from the pGEM®-T Easy plasmid and the pCI-neo vector linearised with *AccI* and *MluI*.

4.2.2.2.1. RE excision of the pGEM®-T Easy_OppA plasmid and the pCI-neo vector with AccI

The pGEM®-T Easy plasmid containing the Ms02 o*ppA* gene and the pCI-neo vector were both first treated with the *AccI* endonuclease. Each 20 µl reaction contained 2 µl of 10x RE buffer, 0.2 µl of 10 mg/ml acetylated bovine serum albumin (BSA), 1 µg of the vector, 500 ng of the plasmid, and 1.7 µl of 6 U/µl *AccI* (Promega). The pGEM®-T Easy_OppA plasmid was incubated for 2 h at 37°C, while the pCI-neo vector was incubated for 1 h at 37°C. After incubation, the reactions were purified with the DNA Clean & Concentrator™-5 kit to remove the *AccI* enzyme and the digestion products.

4.2.2.2.2. RE excision of the purified products with MluI

The purified products were then subjected to digestion with the RE *MluI* in a 20 µl reaction, containing 8.8 µl Milli-Q® water, 2 µl 10x RE buffer, 0.2 µl (10 mg/ml) acetylated BSA, 8 µl DNA (80 ng/µl), and 1 µl of 10 U/µl *MluI* (Promega). Both reactions were incubated at 37ºC for 1 h, followed by inactivation of the enzyme for 15 min at 65ºC. To prevent reattachment of the pCI-neo vector, the digestion reaction was treated with shrimp alkaline phosphate (SAP) in a 30 µl reaction, containing 3 µl 10x SAP buffer, 16 µl DNA (80 ng/µl), and 1.3 µl of 1 U/µl SAP (Promega). Incubation for 15 min at 37°C was followed by inactivation for 15 min at 65°C. The *oppA* excised pGEM®-T Easy plasmid and the pCI-neo SAP reactions were then purified with the DNA Clean & Concentrator™-5 kit according to the manufacturer's instructions.

4.2.2.2.3. Construction of the pCI-neo vector to contain the Ms02 oppA gene

In order to construct the pCI-neo vector with the excised Ms02 *oppA* gene, the pCI-neo Mammalian Expression Vector kit (Promega) was used in a 10 µl ligation reaction, containing the purified *oppA* gene, 100 ng SAP-treated pCI-neo vector, 1 µl 10x ligation buffer, and 1 µl of 3 U/µl T4 DNA Ligase. A vector:insert ratio of 3:1 and 1:1 was calculated using the equation as seen in Figure 4.2. A negative control without DNA was also included. The ligation reactions were incubated overnight at 4°C and used directly in the transformation of *E. coli* JM-109 competent cells (Section 4.2.2.1.5). Positive clones that contained the recombinant pCI-neo vector were identified and confirmed in a diagnostic PCR. The primer pair T7EEV (Table 4.2) and Ms02_OppA_PP2R (Table 4.3) was used to verify the presence of the insert in the correct orientation. The colonies that were confirmed to be positive were cultivated and the plasmids isolated (Section 4.2.2.1.6 to Section 4.2.2.1.7).

$$\frac{vector\,(ng)\times insert\,(kb)}{vector\,(kb)}\times x(ratio) = insert\,(ng)$$

Figure 4.2 The equation used to calculate the amount of insert required (in ng), where *x* represents the chosen ratio. The ratio may be varied in order to determine the optimal ligation reaction.

4.2.2.2.4. Sequencing of the Ms02 oppA gene inserted into the pCI-neo vector

The entire 2 997 bp gene insertion of the recombinant pCI-neo plasmid was sequenced to confirm that successful cloning without any mutations was achieved. In addition to the 6 internal primers (Table 4.3), the outside primer pair T7EEV and T3 (Table 4.2) were employed in 8 separate sequencing reactions. The sequencing reactions were performed as previously described, using the ABI BigDye® Terminator Cycle Sequencing kit (Section 4.2.2.1.8). The products were analysed with an ABI® 3100 Genetic Analyser and the resulting sequences manually aligned in BioEdit® Sequence Alignment Editor.

4.2.2.3. Cloning of the Ms02 oppA gene into the eukaryotic expression vector VR1012

To enable cloning of the Ms02 *oppA* gene into the VR1012 vector, the pGEM®-T Easy_OppA plasmid with a *SalI* recognition site on the 5' end and a *BamHI* recognition site on the 3' end were used to excise the o*ppA* gene. Simultaneous digestion with the two enzymes was performed on the two vectors and the purified, digested DNA products ligated in a standard reaction.

4.2.2.3.1. RE excision of the pGEM®-T Easy_OppA plasmid and the VR1012 vector with BamHI and SalI

Both the pGEM®-T Easy_OppA plasmid and the VR1012 vector were subjected to simultaneous endonuclease excision with *BamHI* and *SalI* (both from Fermentas). The VR1012 vector reaction contained 5 µl of 10x RE buffer, 0.5 µl (10 mg/ml) acetylated BSA, 2.5 µl of each enzyme, 1.25 µg vector, and Milli-Q® water up to a total reaction volume of 50 µl. The pGEM®-T Easy_OppA reaction contained 4 µl of 10x RE buffer, 0.4 µl (10 mg/ml) acetylated BSA, 2 µl of each enzyme, 1 µg DNA (pGEM®-T Easy plasmid containing the o*ppA* gene), and Milli-Q® water up to a total reaction volume of 40 µl. Both reactions were incubated at 37°C for 30 min and inactivated for 10 min at 80°C. The products were then purified using the DNA Clean & Concentrator™-5 kit according to the manufacturer's instructions. The digested VR1012 vector was treated with SAP in a 30 µl reaction, containing 3 µl 10x SAP buffer, 8 µl (160 ng/µl) DNA, 1.3 µl of 1 U/µl SAP, and Milli-Q® water to a final

volume of 30 µl. Incubation for 15 min at 37°C was followed by 15 min inactivation at 65°C. Before proceeding, the product was purified with the DNA Clean & Concentrator™-5 kit according to the manufacturer's instructions.

4.2.2.3.2. Construction of the VR1012 vector to contain the Ms02 oppA gene

The linearised VR1012 vector and the excised Ms02 o*ppA* gene were ligated in a reaction containing 1 µl of the VR1012 vector, 3 µl of the excised *oppA* gene, 1 µl of a 10x ligation buffer, and 1 µl of 3 U/µl T4 DNA Ligase. The reactions were incubated overnight at 4°C and transformation of *E. coli* JM-109 competent cells was performed as previously described, except that 4 µl of the ligation reaction was used per transformation. Colonies were chosen to verify the presence of insert in a diagnostic PCR, with the primer pair VR1012-F (Table 4.2) and Ms02_OppA_PP2R (Table 4.3). Positive colonies were cultivated in LB-medium and the plasmids were isolated (Section 4.2.2.1.6 to Section 4.2.2.1.7).

4.2.2.3.3. Sequencing of the Ms02 oppA gene inserted into the VR1012 vector

To verify successful in-frame cloning of the Ms02 *oppA* gene into the VR1012 vector, the entire 2 997 bp gene was sequenced in 8 separate sequencing reactions. In addition to the six internal primers (Table 4.3), the outside primers VR1012-F and VR1012-R (Table 4.2) were also included. Sequencing of the product and analysis of the resultant sequences were performed as described previously. Freeze cultures were prepared of the cultures containing the correct plasmids (Section 4.2.2.1.8).

4.2.2.4. Cloning of the Ms02 oppA gene into the eukaryotic secretion expression vector VR1020

Cloning of the Ms02 o*ppA* gene into the VR1020 vector required the gene to have a *BamHI* recognition site on both ends. Due to the fact that the same RE recognition sequence occurred on both sides of the gene, the possibility of reverse insertion existed. Since an internal primer was used with an external primer in the diagnostic PCR, a clone with the gene in the correct orientation could be identified.

4.2.2.4.1. RE excision of the pGEM®-T Easy_OppA plasmid and the VR1020 vector with BamHI

The pGEM®-T Easy_OppA plasmid and the VR1020 vector were subjected to endonuclease excision with *BamHI*. The VR1020 vector reaction contained 4 µl of 10x RE buffer, 0.4 µl (10 mg/ml) acetylated BSA, 4 µl of *BamHI,* 4 µg of the vector, and Milli-Q® water up to a total reaction volume of 40 µl. The pGEM®-T Easy_OppA plasmid reaction contained 4 µl of 10x RE buffer, 0.4 µl (10 mg/ml) acetylated BSA, 4 µl of *BamHI*, 1 µg DNA (pGEM®-T Easy plasmid containing the *oppA* gene), and Milli-Q® water up to a total reaction volume of 40 µl. Both reactions were incubated at 37°C for 30 min, inactivated for 5 min at 80°C, and subsequently purified using the DNA Clean & Concentrator™-5 kit according to the manufacturer's instructions. The digested VR1020 vector was treated with SAP in a 30 µl reaction containing 3 µl 10x SAP buffer, 7 µl (160 ng/µl) DNA, 1 µl of 1 U/µl SAP, and Milli-Q® water to a final volume of 30 µl. The reaction was incubated for 15 min at 37°C, inactivated for 15 min at 65°C and then purified with the DNA Clean & Concentrator™-5 kit according to the manufacturer's instructions.

4.2.2.4.2. Construction of the VR1020 vector to contain the Ms02 oppA gene

The linearised SAP-treated VR1020 vector and the excised o*ppA* gene were then ligated in a 10 µl reaction, containing 1 µl of the VR1020 vector, 1 µl of a 10x ligase buffer and 1 µl of 3 U/µl T4 DNA Ligase. A vector:insert ratio of 3:1 and 1:1 was calculated as described before (Figure 4.2), as well as a reaction that contained the maximum possible insert. The reactions were incubated overnight at 4°C and transformation of *E. coli* JM-109 competent cells was performed. Colonies were chosen to verify the presence of the inserted gene in a diagnostic PCR, with the primer pair VR1020-F (Table 4.2) and Ms02_OppA_PP2R (Table 4.3). By using an external primer in combination with an internal primer, clones with inserts in the correct orientation were identified. Positive colonies were cultivated in LB-medium and the plasmids were isolated (Section 4.2.2.1.5 to Section 4.2.2.1.7).

4.2.2.4.3. Sequencing of the Ms02 oppA gene inserted into the VR1020 vector

To verify successful in-frame cloning of the o*ppA* gene into the VR1020 vector, the entire 2 997 bp gene was sequenced in 8 separate sequencing reactions. In addition to the six internal primers (Table 4.3), the outside primers VR1020-F and VR1020-R (Table 4.2) were also included. Sequencing of the product and analysis of the results were performed as described previously. Positive plasmids were stored as previously described (Section 4.2.2.1.8).

4.2.2.5. Subcloning of the Ms02 oppA gene into the prokaryotic expression vector pGEX-4T-1

To subclone the Ms02 *oppA* gene into the pGEX-4T-1 vector, a *BamHI* recognition site was added to the 5' end and a *SalI* recognition site added to the 3' end of the o*ppA* gene. This allowed for excision of the gene from the pGEM®-T Easy_OppA vector and linearisation of the pGEX-4T-1 vector, allowing subsequent ligation.

4.2.2.5.1. RE excision of the pGEM®-T Easy_OppA plasmid and the pGEX-4T-1 vector with BamHI and SalI

The pGEM®-T Easy_OppA plasmid and the pGEX-4T-1 vector were subjected to simultaneous endonuclease excision with *BamHI* and *SalI*. The pGEX-4T-1 reaction contained 4 µl of 10x RE buffer, 0.4 µl (10 mg/ml) acetylated BSA, 2 µl of each enzyme, 1 mg of the vector, and Milli-Q® water up to a total reaction volume of 40 µl. The pGEM®-T Easy_OppA reaction contained 4 µl of 10x RE buffer, 0.4 µl (10 mg/ml) acetylated BSA, 500 ng DNA (pGEM®-T Easy plasmid containing the *oppA* gene), and Milli-Q® water to make up the volume to 40 µl. Both reactions were incubated at 37°C; the pGEX-4T-1 reaction for 15 min and the pGEM®-T Easy reaction for 60 min. An inactivation step of 10 min at 65°C preceded purification of the product with the DNA Clean & Concentrator™-5 kit, according to the manufacturer's instructions. The digested pGEX-4T-1 vector was treated with SAP to remove a phosphate group from the 5' and 3' ends of the vector in order to prevent reattachment. The reaction consisted of 3 µl 10x SAP buffer, 8 µl (160 ng/µl) DNA, 1.3 µl of 1 U/µl SAP, and Milli-Q® water to a final volume of 30 µl. Incubation for 15 min at 37°C was followed by inactivation for 15 min at 65°C. Again, the product was purified with the DNA Clean & Concentrator™-5 kit according to the manufacturer's instructions.

4.2.2.5.2. Construction of the pGEX-4T-1 vector to contain the Ms02 oppA gene

The linearised SAP-treated pGEX-4T-1 vector and the excised Ms02 *oppA* gene were then ligated in a reaction containing 1 µl of the pGEX-4T-1 vector, 5 µl of a 2x ligation buffer and 1 µl of 3 U/µl T4 DNA Ligase. An insert:vector ratio of 5:1 and 3:1 were used to calculate the amount of the *oppA* gene to add. The reactions were incubated overnight at 4°C. Transformation of *E. coli* JM-109 competent cells was performed. To test which colonies contained the insert, a diagnostic PCR was performed, using the primer pair pGEX-F (Table 4.2) and Ms02_OppA_T7R (Table 4.3). Positive colonies were inoculated in LB-medium and the plasmids were isolated (Section 4.2.2.1.5 to Section 4.2.2.1.7).

4.2.2.5.3. Sequencing of the Ms02 oppA gene inserted into the pGEX-4T-1 vector

To verify successful in-frame cloning of the Ms02 *oppA* gene into the pGEX-4T-1 vector, 8 sequencing reactions were performed with the six internal primers and the two outside primers pGEX-F and pGEX-R (Table 4.2). The entire gene was sequenced to confirm the absence of mutations and the resultant sequences were aligned manually with BioEdit® Sequence Alignment Editor. Freeze cultures were prepared of the positive colonies (Section 4.2.2.1.8).

4.2.2.5.4. Transformation of E. coli BL21(DE3) competent cells with the pGEX-4T-1_OppA vector

The pGEX-4T-1 vector containing the Ms02 o*ppA* gene was transformed into *E. coli* BL21(DE3) competent cells (Novagen). Of the competent cells, 100 µl were mixed with 0.18 µl (14.05 M) β-mercapto-ethanol and incubated on ice for 10 min, with gentle swirls every 2 min. To this, 5 µl of the pGEX-4T-1_OppA vector was added, incubated on ice for 30 min, followed by 45 s heat-shock at 42°C. The heat-shocked cells were cooled on ice for 2 min. SOC medium was preheated to 37°C and added to the cell mixture to a final volume of 1 ml. The cells were allowed to recover for 1 h at 37°C, while being shaken on an IKA® KS 260 Basic orbital shaker at 220 rpm. The cultures were then plated on prepared LB/agar plates, supplemented with 100 µg/ml ampicillin at 1 µl/ml, and cultivated for 16 h at 37°C. After confirmation of positive transformants, using the pGEX-F and pGEX-R primer pair in a diagnostic PCR (Table 4.2), cultures were made by inoculating the positive colonies into 5 ml LB-medium supplemented with 5 µl (100 mg/ml) ampicillin. The cultures were incubated at 37°C for 16 h, while being shaken on an IKA® KS 260 Basic orbital shaker at 220 rpm.

4.2.2.5.5. Transformation of Single Step (KRX) competent cells with the pGEX-4T-1_OppA vector

The pGEX-4T-1 vector containing the *oppA* gene was transformed into Single Step (KRX) competent cells (Promega) according the manufacturer's instructions. Of the KRX competent cells, 50 µl were mixed with 5 µl of the pGEX-4T-1_OppA vector and incubated on ice for 10 min. The cells were heat-shocked for 20 s at 42°C and immediately returned to the ice for 2 min. To this, 450 µl preheated SOC media was added and the cells incubated for 1 h at 37°C, while being shaken on an IKA® KS 260 Basic orbital shaker at 220 rpm. The cultures were then plated on LB/amp/IPTG/X-gal plates prepared as described before (Section 4.2.2.1.5). A diagnostic PCR was performed using the pGEX-F and pGEX-R primer pair (Table 4.2). The positive colonies were cultured and stored for future use (Section 4.2.2.1.8).

4.2.2.6. Expression of the recombinant Ms02 OppA protein using the pGEX-4T-1 prokaryotic expression system

The Ms02 OppA protein was expressed as a GST-fusion protein with the following conditions after numerous attempts to optimise the expression of soluble OppA protein. For a complete discussion on the different attempts and associated problems, refer to Section 4.1.3.

4.2.2.6.1. Expression of the Ms02 GST-OppA protein in E. coli BL21(DE3) and Single Step (KRX) cells

The two respective cell freeze cultures, that contained the recombinant pGEX-4T-1 vector with the Ms02 *oppA* gene, were streaked onto LB/agar plates that were supplemented with 100 mg/ml ampicillin at 1 µl/ml. The plates were incubated for 16 h at 37°C before colony formation was analysed. The transformed colonies were tested in a diagnostic PCR, using the primer pair Ms02_OppA_PP2F and Ms02_OppA_PP3R (Table 4.3). To express the *oppA* gene, starter cultures were made by inoculating multiple positive colonies from each cell type in their respective flasks, containing 20 ml LB-medium supplemented with 20 µl (100 mg/ml) ampicillin and 400 µl (20%) glucose. The cultures were incubated for 14 h at 37°C, before it was diluted in 400 ml Terrific Broth (TB)-medium (360 ml TB, 40 ml (1 M) phosphate buffer, pH 7.8, 400 µl (100 mg/ml) ampicillin, and 6 ml (20%) glucose). The medium was divided in four 100 ml flasks to ensure proper aeration and incubated for 6 h at 37°C, while being shaken at 220 rpm. The cultures were then induced. The flasks that contained the BL21 culture were induced with 1 ml (0.1 M) IPTG, while the flasks with the KRX culture were induced with 1 ml (0.1 M) IPTG and 500 µl (20%) L-rhamnose. The cultures were allowed to express for 3 h, before the cells were harvested by centrifugation for 10 min at 10 000 x g and resuspended in 1x TEN50 (50 mM $NaCl_2$, 20 mM Tris-acetate, and 1 mM EDTA) buffer, supplemented with 0.2 mM DTT, 0.1% (v/v) Triton-X-100 and 10% (v/v) glycerol. Samples of the cell supension were stored for subsequent use in SDS-PAGE analysis. To prepare the protein for subsequent protein isolation, the cell suspension was subjected to three freeze/thaw cycles, cycling between 37°C and -80°C, and five cycles of sonication for 20 s followed by 20 s on ice. The cell lysate were then mixed with 1 µl of a 10 mg/µl lysozyme solution (Sigma-Aldrich) per 100 µl cell suspension, and incubated on ice for 30 min, with occasional mixing. To the lysozyme-treated cell suspension, 1 tube of Halt™ Protease Inhibitor Single-Use Cocktail EDTA-free (Thermo-Scientific) per 10 ml cell suspension was added and again incubated on ice for 30 min, with occasional mixing. The soluble fractions were then separated from the insoluble fractions by centrifugation at 10 000 x g for 20 min.

4.2.2.6.2. Visualisation of the expressed proteins using SDS-PAGE

To analyse the expression profile, an SDS-PAGE was performed. The harvested cells were lysed with 3 cycles of freezing at -80°C for 5 min, and thawing at 37°C until completely thawed. To ensure complete lysis of individual cells in order to release the intracellular proteins into the medium, the cell lysate was then forced four times through a 0.33 mm insulin needle. To prepare the cells for loading onto the SDS gel, 20 µl of cell lysis product, 12 µl of 0.1 % (m/v) bromophenol blue, and 20 µl of reducing loading buffer (0.125 M Tris-HCl, 4% (m/v) SDS, 20% (v/v) glycerol, and 10% (v/v) β-mercapto-ethanol, pH 6.8) were mixed and boiled for 2 minutes and immediately placed on ice. Twenty microliters of the samples were then loaded onto a reducing SDS gel, composed of 4.5% stacking and 12% resolving gel, and electrophoresed at a constant 20 mA. To visualise the protein bands, the gel was stained with staining buffer (0.125% (w/v) Coomassie Brilliant Blue R250, 50% (v/v)

methanol, and 10% (v/v) acetic acid) for 1 h at 37°C, while being agitated at 4 rpm. This was followed by 1 h of destaining with Destain I solution (50% (v/v) methanol and 10% (v/v) acetic acid) at 37°C, while being agitated at 4 rpm. To ensure complete destaining, the gel was left overnight in Destain II solution (5% (v/v) methanol and 7% (v/v) acetic acid) at room temperature on a rotary table (20 rpm).

4.2.2.7. Detection of the Ms02 GST-OppA protein and its purification

Western blot analysis was used to confirm that the expressed protein, as seen on the SDS gel, is in actual fact the Ms02 OppA GST-fusion protein. Confirmation was achieved before proceeding with protein isolation, however, it could only be confirmed that a GST-tagged protein of the expected size was present, because an anti-GST antibody was used as the primary antibody. Pure protein was required to generate OppA antibodies, as well as for use in possible future serological tests.

4.2.2.7.1. Western blot analysis of the expressed proteins using an anti-GST antibody

The western blot was performed with an anti-GST antibody as the primary antibody. First, to determine the optimal amount of protein needed so as not to overload the gel, different dilutions of the 24 h control GST cell sample was separated with SDS-PAGE (Section 4.2.2.6.2). After determining the optimal concentrations to use, three different dilutions were prepared, and were transferred to a 0.45 µm nitrocellulose membrane (Schleicher and Schuel, Sigma-Aldrich). The transfer occurred by electrophoresis in electrode buffer (50 mM Tris-HCl, 200 mM glycine, and 20% (v/v) methanol, pH 8.3) for 16 h at a constant of 120 mA. Casein buffer (154 mM NaCl, 0.5% (w/v) casein, 10 mM Tris-HCl, and 0.02% (w/v) thiomersal, pH 7.6) was used to block the membrane for 20 min at room temperature, while being agitated. The goat anti-GST antibody (GE Healthcare) was diluted 10 000 times with casein buffer in a volume sufficient to cover the nitrocellulose membrane. The membrane (covered with the primary antibody dilution) was then incubated for 1 h at 37°C while being agitated at 6 rpm. The nitrocellulose membrane was then washed three times with PBS-Tween (137 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4 \cdot 2H_2O$, 2 mM KH_2PO_4, and 0.1% (v/v) Tween-20, pH 7.2) for 5 min each at room temperature, while being agitated, to remove excess unbound antibody. Horse-radish peroxidase (HRP)-conjugated rabbit anti-goat antibody (Sigma-Aldrich) was diluted 5 000 times with casein buffer, added to the nitrocellulose membrane and incubated for 1 h at 37°C, while being agitated at 6 rpm. The nitrocellulose membrane was washed again, with an additional final wash of PBS without Tween-20 (15 min at room temperature) to remove excess Tween-20. The substrate, consisting of 0.05% (w/v) 4-chloro-1-naphtol, 16% (v/v) ice cold methanol, ice-cold 1x PBS, and 0.025% (v/v) cold H_2O_2, was added to the nitrocellulose membrane and allowed to develop for 30 min at room temperature, while being agitated. To prevent over-development, the nitrocellulose membrane was washed with Milli-Q® water after bands could be visualised. After the results were established the protocol was repeated using the Ms02 GST-OppA protein under the same conditions to confirm the presence of a GST-tagged protein of the expected size, before proceeding with isolation of the protein.

4.2.2.7.2. Purification of the Ms02 GST-OppA protein

Affinity chromatography was employed to purify the Ms02 GST-OppA protein from the cell suspension of the expression culture. The glutathione-agarose resin (Sigma-Aldrich) was prepared according to the manufacturer's instructions. The protein sample was prepared as described previously to obtain soluble protein

in the supernatant (Section 4.1.3.2). Purification of the protein, from the soluble sample onwards, then proceeded according to the manufacturer's instructions. The protein was eluted with elution buffer (10 mM reduced glutathione and 50 mM Tris-HCl, pH 9.4), and 20 eluted fractions were collected in 1 ml volumes. The fractions were analysed with SDS-PAGE and the concentrations were measured spectrophotometrically with the Bradford protein determination assay. In short, a standard dilution series of concentrations in the range of 0 to 175 µg/ml, made with bovine serum albumin (BSA) (Roche), was used to determine the protein concentration in each fraction (in a Greiner Bio-One 96-well microtitre plate). Of the protein fractions, 5 µl was diluted in 250 µl Bradford reagent (8.5% (w/v) phosphoric acid, 0.01% (w/v) Coomassie Brilliant Blue G-250, and 4.7% (v/v) ethanol), incubated for 5 min at room temperature, and the absorbance measured at 620 nm with a Labsystems Multiskan MS spectrophotometer. The BSA standards were used to generate a standard curve to calculate the protein concentrations of the fractional samples. An 8% resolving denaturing SDS gel was used to visualise the fractions and to determine the quality of the isolated protein.

4.2.2.8. Rabbit antibody production raised against the Ms02 GST-OppA protein

The Ms02 GST-OppA fusion protein was used to generate rabbit antibodies. The antibodies were to be used to determine if the recombinant DNA plasmid, pCI-neo, can express the OppA protein in eukaryotic cell culture. In future ostrich vaccination trials, the OppA fusion protein can be used as an antigen to determine ostrich immune responses toward the Ms02 DNA vaccines.

4.2.2.8.1. Antibodies raised against Ms02 GST-OppA purified protein in rabbits

In order to produce antibodies against the Ms02 GST-OppA protein, a rabbit was immunised with the protein, and the blood sera from the rabbit were collected. First, the protein was adsorbed onto acid-treated naked *Salmonella minnesota* bacteria (Bellstedt *et al.*, 1987) in a 1:5 ratio and dried using a rotary evaporator, the SpeedyVac Concentrator (Savant). The precipitated protein-bacteria complex was suspended in PBS to a final concentration of 0.4 mg/ml total protein. A rabbit was immunised with 200 µg (0.5 ml) of the protein-bacteria complex on days 0, 3, 6, 13, 16, 20, 27, 31 and 35. A test bleed of 2 ml was drawn on days 0 and 28 and a large bleed of 20 ml on day 42 for sera collection.

4.2.2.8.2. Enzyme-Linked Immunosorbent Assay to test rabbit antibody titres

The protein concentration as determined in Section 4.2.2.7.2 was used to coat Nunc© 96-well microtitre plates with 1, 2 or 5 µg of the purified GST-OppA protein. The protein was diluted in 100 µl carbonate buffer (50 mM $NaHCO_3$, pH 9.6). The plate was incubated overnight at 4°C and blocked with 200 µl casein buffer (154 mM NaCl, 0.5% (w/v) casein, and 10 mM Tris-HCl, pH 7.6) for 1 h at 37°C. The wells were washed three times with PBS-Tween (PBS with 0.1% (v/v) Tween-20, pH 7.2). Rabbit serum from day 42 was diluted 1:20 in casein buffer that contained 0.05% (v/v) Tween-20, and 200 µl of the diluted serum added to the first well. In each of the other wells, 100 µl casein-Tween buffer was added. By pipetting 100 µl of the first well into the next, mixing the content and pipetting 100 µl of the mixed content into the next well, and so on, a dilution series of up to 10 240 times was created. This was followed by incubation at 37°C for 1 h. The wells were washed again three times with PBS-Tween. The secondary antibody, a goat anti-rabbit antibody (Sigma-Aldrich), was diluted 1:500 in casein-Tween, 100 µl added to each well, and the plate incubated at 37°C for 1 h. This was

followed by three wash steps. The tertiary antibody, a horse-radish peroxidase (HRP)-conjugated rabbit anti-goat antibody (Sigma-Aldrich), was diluted 1:10 000 in casein-Tween and 100 µl added to each well. This was followed by a final incubation step at 37°C for 1 h, and three wash steps with PBS-Tween. Substrate solution, consisting of 0.1 M citrate buffer (pH 5) and 0.05% (w/v) 2,2'-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid) (ABTS), were prepared fresh and incubated at room temperature for 10 min. Just before use, 0.05% (v/v) cold H_2O_2 was added to the substrate solution, and 100 µl of this solution was added to each well. Absorbance was measured with a Labsystems Multiscan Original spectrophotometer, at 405 nm at five minute intervals for 30 min. In between measurements, the plate was incubated at 37°C.

4.2.2.8.3. Western blot analysis to confirm the antibody activity

To confirm the presence of Ms02 GST-OppA antibodies in the rabbit sera, and to test the cross-reactivity with the three ostrich mycoplasma GST-OppA proteins, a western blot analysis was performed as previously described, but with different antibodies (Section 4.2.2.7.1). The purified GST-OppA proteins from Ms01, Ms02 and Ms03, as well as isolated GST protein, were tested with the rabbit sera. Following the initial blocking step, three different dilutions of the rabbit sera were tested against purified Ms02 GST-OppA protein, namely 1:1 000, 1:2 500 and 1:5 000. To determine the cross-reactivity of the antibody with Ms01 and Ms03 GST-OppA proteins, a 1:2 500 rabbit antibody dilution was used against the three different purified GST-OppA proteins. A secondary antibody, goat anti-rabbit (Sigma-Aldrich), was diluted 1:2 500 in casein buffer to cover the membrane and incubated for 1 h at 37°C. The membrane was washed with PBS-Tween as described earlier. The HRP-conjugated rabbit anti-goat antibody was diluted 1:5 000 and allowed to incubate for 1 h with the membrane at 37°C. Substrate was added as described earlier (Section 4.2.2.8.3).

4.2.3. RESULTS

4.2.3.1. PCR amplification of the Ms02 oppA gene to add RE recognition sequences

The Ms02 *oppA* gene was successfully amplified with PCR from the recombinant pGEM®-T Easy plasmid, using the primers designed with recognition sequences for the restriction endonucleases *BamHI* and *SalI*. Three different amplified gene products were created: one with a *BamHI* recognition site on the 5' and 3' end of the *oppA* gene; one with a *SalI* recognition site on the 5' end and a *BamHI* recognition site on the 3' end; and one with the two sites reversed. The amplification products were of the predicted size of 3 009 bp (2 997 bp *oppA* gene + 2 x 6 bp of the RE cut sites). Figure 4.3 shows an agarose gel with the amplified gene product at just above 3 000 bp.

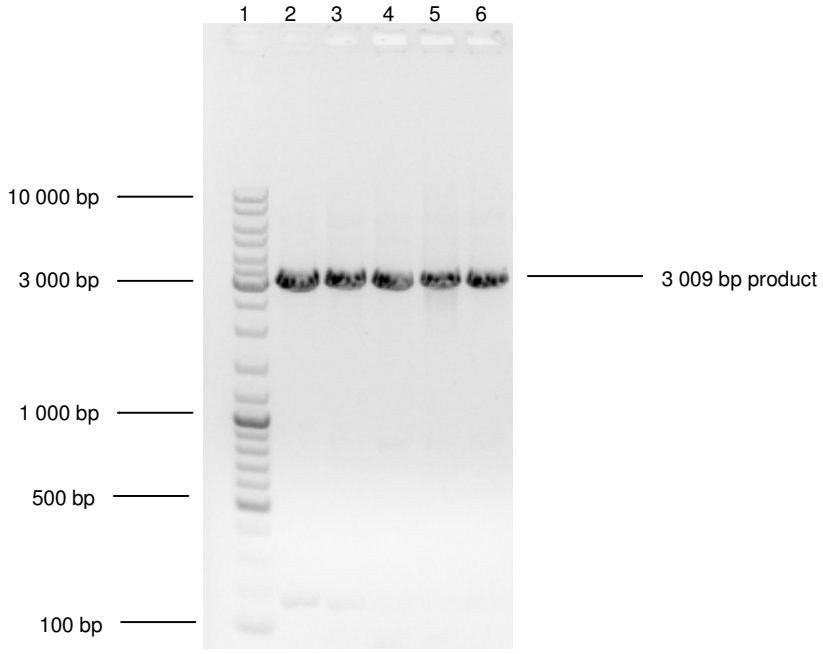

Figure 4.3 Image of the amplified *oppA* gene products with RE recognition sites added as visualised under UV light after electrophoretic separation on an agarose gel . In this image, the primers Ms02_OppA_BamF and Ms02_OppA_SalR were used to amplify the o*ppA* gene for subcloning into the pGEX-4T-1 prokaryotic expression vector. The PCR products were separated on a 1% (w/v) agarose gel with 0.1 µl/ml GelRed™ Nucleic Acid Gel Stain for UV visualisation of the DNA. Lane 1: 1 KB DNA GeneRuler™ DNA ladder mix (Fermentas). Lane 2-6: the amplified Ms02 *oppA* gene product.

4.2.3.2. Cloning of the Ms02 oppA gene into the pGEM®-T Easy plasmid

The Ms02 *oppA* gene, with the newly added RE recognition sequences, was successfully recloned into the pGEM®-T Easy plasmid and transformed into *E. coli* JM-109 competent cells, resulting in several white colonies. In order to confirm the presence of the o*ppA* gene, a diagnostic PCR with the primer pair T7-promoter and Ms02_OppA_PP2R was used to amplify a product of ~1 600 bp. In this manner, three different types of pGEM®-T Easy constructs were obtained, one with a *BamHI* recognition sequence on both ends of the gene, one with a *BamHI* on the 5' end and a *SalI* on the 3' end of the gene, and one with these two sites reversed. Figure 4.4 is an image acquired from the electophoresis of the diagnostic PCR performed on the pGEM®-T Easy plasmid containing the *oppA* gene, with restriction sites added.

The plasmids that contained the Ms02 *oppA* gene were successfully isolated from the bacteria and subjected to restriction enzyme digestion in order to ascertain the topology and size of the plasmid. The plasmids that were deemed to be of the correct size, as well as having the right digestion profile, were sequenced until a plasmid containing the Ms02 *oppA* gene, free of any mutations and flanked by the correct RE cut sites, was obtained.

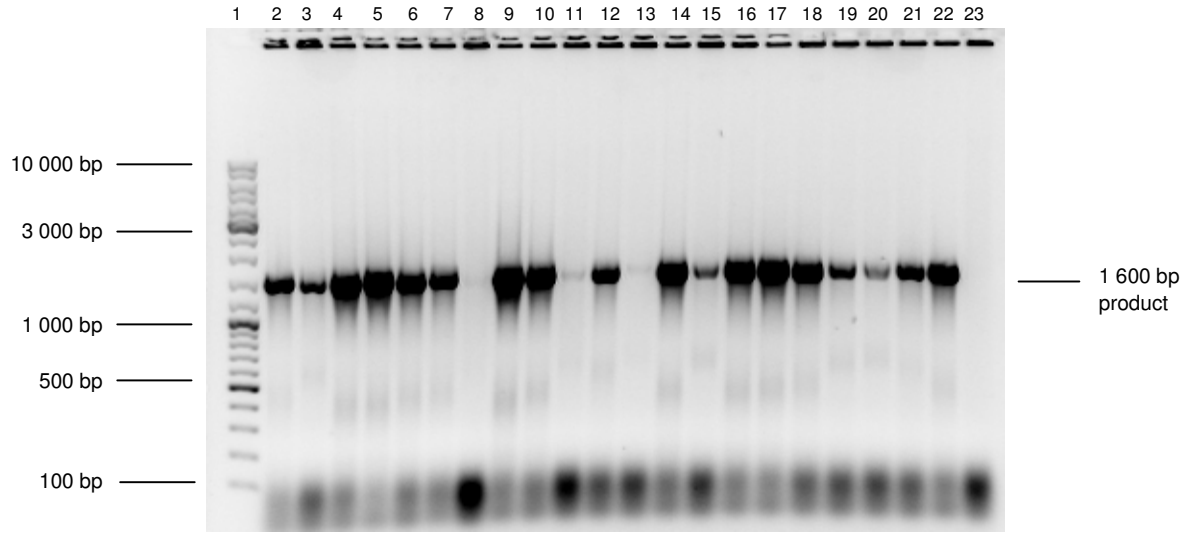

Figure 4.4 Image depicting the products of the diagnostic PCR performed on the transformed *E. coli* cells to contain the modified pGEM®-T Easy plasmid. The products were separated on a 1% (w/v) agarose gel cast with 0.1 µl/ml GelRed™ Nucleic Acid Gel Stain for UV visualisation of the DNA. Lane 1: 1 KB DNA GeneRuler™ ladder mix. Lane 2-12: amplified product from the pGEM®-T Easy plasmid containing *BamHI* recognition sites at both ends of the o*ppA* gene for subcloning into the VR1020 vector. Lane 13-23: amplified product from the pGEM®-T Easy plasmid containing a *SalI* recognition site at the 5' end and a *BamHI* recognition site at the 3' end of the o*ppA* gene for subcloning into the VR1012 vector.

4.2.3.3. Subcloning of the Ms02 oppA gene into the mammalian expression vector pCI-neo

RE digestion of the pGEM®-T Easy plasmid and the pCI-neo vector with *AccI* and *MluI* was successful and was verified by agarose gel electrophoresis. Cloning of the o*ppA* gene into the pCI-neo vector resulted in numerous colonies, but using the primer pair T7EEV and Ms02_OppA_PP2R in a diagnostic PCR indicated poor ligation efficiency. Three plasmids containing the o*ppA* gene were isolated from the bacteria and sequenced. Only two contained the full-length gene, were confirmed to be mutation-free, and stored for subsequent use.

4.2.3.4. Subcloning of the Ms02 oppA gene into the eukaryotic expression vector VR1012

The VR1012 expression vector and the pGEM®-T Easy plasmid containing a 5' *SalI* and a 3' *BamHI* RE cut site were digested with the two respective endonucleases. The ligation reactions were analysed with agarose gel electrophoresis and, due to the lack of intensity of the band corresponding with the expected size of the plasmid containing the insert, the amount of ligated product used per transformation reaction was increased. Transformation resulted in numerous colonies, of which a satisfactory amount was diagnosed to contain the *oppA* gene, using the primer pair VR1012-F and Ms02_OppA_PP2R. Sequencing confirmed the presence of the *oppA* gene and the absence of mutations, and the successful plasmid was stored for subsequent use.

4.2.3.5. Subcloning of the Ms02 oppA gene into the eukaryotic secretion expression vector VR1020

RE digestion of the relevant pGEM®-T Easy plasmid and the VR1020 expression vector was performed with the endonuclease *BamHI*. Multiple colonies were obtained with the transformation. A diagnostic PCR, using the primer pair VR1020-F and Ms02_OppA_PP2R, identified two colonies that contained the o*ppA* gene. The plasmids were isolated and subjected to sequencing. After confirming one plasmid to be free of mutations, it was stored for future use.

4.2.3.6. Subcloning of the Ms02 oppA gene into the prokaryotic expression vector pGEX-4T-1

In order to subclone the *oppA* gene into the prokaryotic expression vector pGEX-4T-1, the endonucleases *BamHI* and *SalI* were used to digest the pGEM®-T Easy plasmid containing a *BamHI* cutting site at the 5' end and a *SalI* cutting site on the 3' end of the o*ppA* gene, as well as to linearise the pGEX-4T-1 vector. Successful transformation was proven by means of a diagnostic PCR, using the primer pair pGEX-F and Ms02_OppA_T7R. Three colonies were identified, their plasmids isolated and sequenced and, when established mutation-free, stored for subsequent use in protein expression.

4.2.3.7. Transformation of the pGEX-4T-1_OppA vector into E. coli BL21(DE3) competent cells and Single Step (KRX) competent cells

The pGEX-4T-1 vector containing a mutation-free insert of the *oppA* gene was successfully cloned into *E. coli* BL21(DE3) competent cells and Single Step (KRX) competent cells. A diagnostic PCR with the primer pair Ms02_OppA_PP2F and Ms02_OppA_PP3R identified all the colonies that were tested to be positive for the o*ppA* gene. Cultures were made of selected positive colonies for expression of the recombinant OppA protein.

4.2.3.8. Expression of the Ms02 GST-OppA protein for purification

The Ms02 GST-OppA protein was successfully expressed, by growing cells to a high density and performing induction for a short time before the cells were harvested. Comparing the expression profiles of the two cell types, *E. coli* BL21(DE3) and Single Step (KRX) cells, the BL21 cells were chosen based on their faster growing capacity. Expression tests in the two cell types indicated that expression occured faster in the BL21 cells, although the KRX cells expressed a higher amount of protein over an extended period of time. BL21 cell colonies on agar plates were consistently larger, and the protein expression peaked earlier than the KRX cell cultures. Figure 4.5 indicates the first 15 h of the BL21 cell culture expression, for both the pGEX_control vector (with no insert) and the pGEX_OppA vector (with Ms02 *oppA* gene insert). The recombinant protein is predicted to be 146 kDa (120 kDa OppA protein plus the 26 kDa GST protein). A clear increase in the amount of expressed protein at the predicted size is visible in the pGEX_OppA cell samples, with the absence of a protein of the same band size in the control cell samples. In addition, the presence of the GST protein is clearly visible in the pGEX_control cell samples, and the same protein is reduced in the pGEX-OppA cell suspension. This is an indication of the expression of a GST-tagged protein.

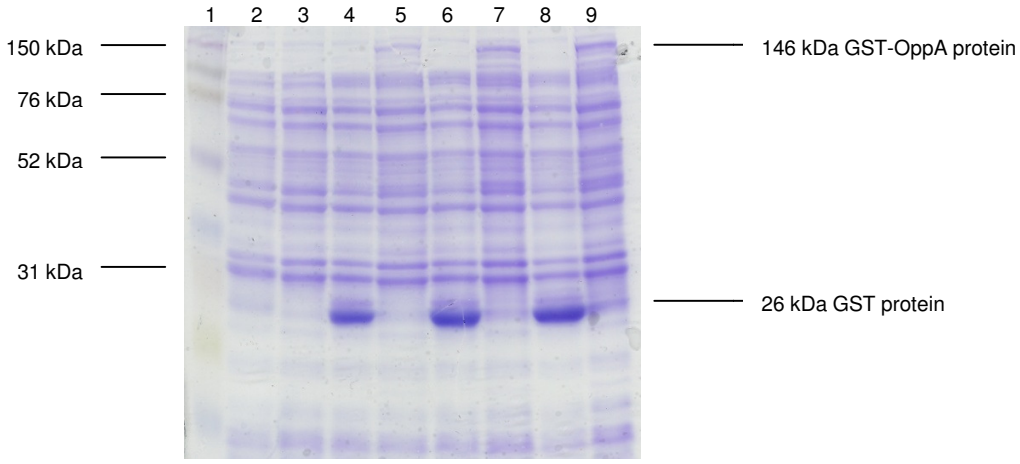

Figure 4.5 Image of a 12% denaturing SDS gel indicating the expression of *E. coli* BL21(DE3) cells over 15 h. Lane 1: PageRuler^{TM} Unstained Protein Ladder (Fermentas). Lane 2: pGEX_control (0 h). Lane 3: pGEX_OppA (0 h). Lane 4: pGEX_control (6 h). Lane 5: pGEX_OppA (6 h). Lane 6: pGEX_control (9 h). Lane 7: pGEX_OppA (9 h). Lane 8: pGEX_control (15 h). Lane 9: pGEX_OppA (15 h).

4.2.3.9. Anti-GST western blot analysis of the prokaryotic protein expression profile

A control GST protein sample was used to determine the best concentration to load onto an SDS gel in order to achieve the correct binding concentrations in the western blot analysis. Deducing from the western blot of the GST protein, a dilution of 1:5 of the 24 h cell sample was deemed to be the optimal amount of protein required (Figure 4.6a). The conditions were then successfully applied to the OppA cell samples, resulting in a western blot that confirmed the presence of a GST-tagged OppA protein (Figure 4.6b).

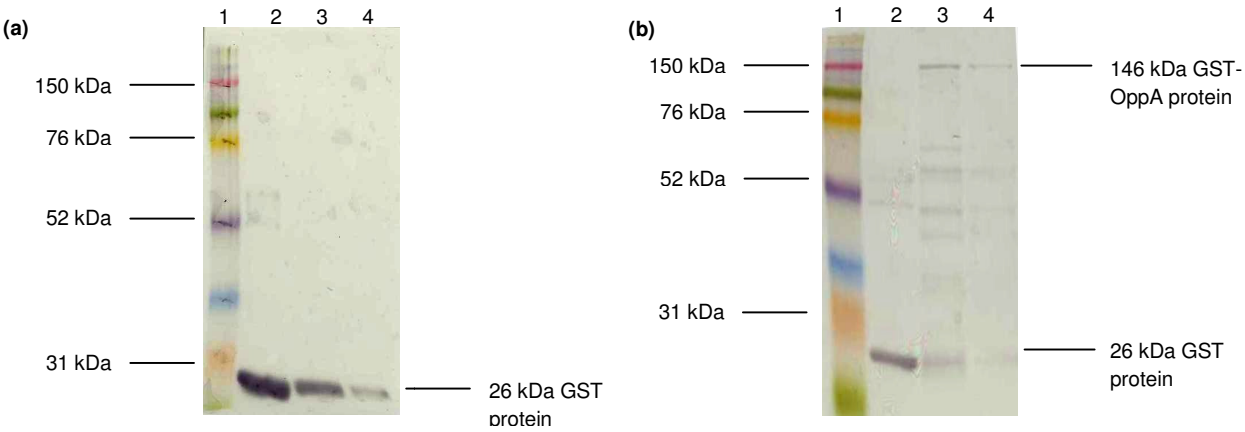

Figure 4.6(a) Image of the western blot illustrating the dilutions of the pGEX-4T-1 cells induced with 1 mM IPTG, using the primary anti-GST antibody. The proteins were separated on a 12% denaturing SDS gel before transfer to the nitrocellulose membrane. Lane 1: Full-range Rainbow marker (Amersham). Lane 2: 1:1 dilution of the GST culture. Lane 3: 1:5 dilution of the GST culture. Lane 4: 1:10 dilution of the GST culture. **Figure 4.6 (b)** Image of the western blot illustrating the detection of the GST-OppA protein, using the primary anti-GST antibody. The proteins were separated on a 12% denaturing SDS gel before transfer to the nitrocellulose membrane. Lane 1: 7 µl of the Full-range Rainbow marker. Lane 2: 1:5 dilution of the GST culture. Lane 3: 1:5 dilution of the GST-OppA culture. Lane 4: 1:10 dilution of the GST-OppA culture.

4.2.3.10. Purification of the Ms02 GST-OppA protein

The BL21 cell cultures were chosen to express the Ms02 GST-OppA protein for purification purposes. The suspended cells were treated with lysozyme to damage the cells, to allow release of the protein into the extracellular medium. To inhibit the action of released proteases, a protease inhibitor mix was added after lysozyme treatment. This proved to be the optimal conditions with which to obtain the Ms02 GST-OppA protein in a soluble form. The GST-OppA protein was successfully purified using affinity chromatography, although the protein was not as pure as would be expected (Figure 4.7a). The purity was deemed sufficient to continue, since the unwanted protein fractions were diluted in the later fractions. The concentrations of the respective fractions are indicated in Figure 4.7b. The peak occurs early in elution; after about eleven fractions all of the bound protein had eluted.

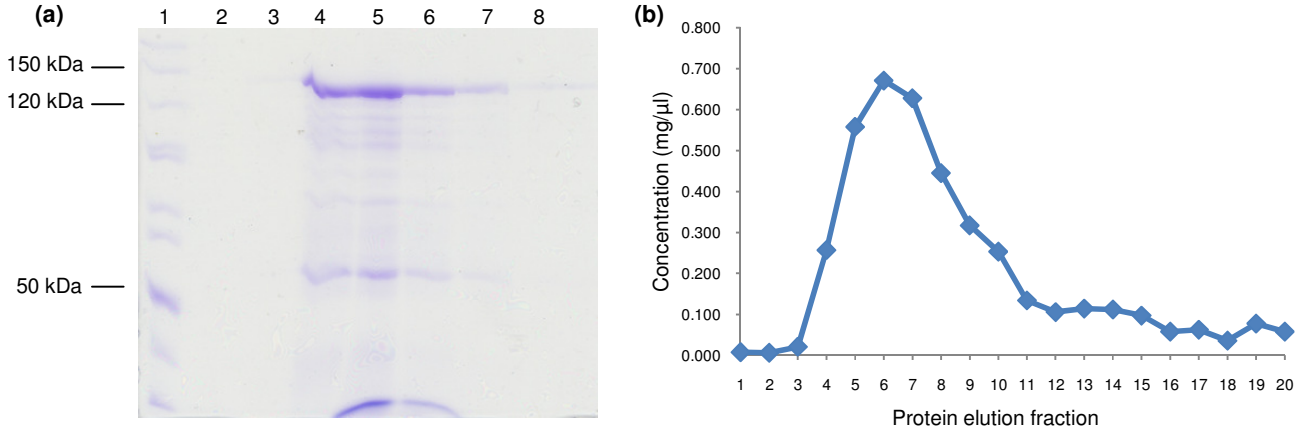

Figure 4.7 (a) Image of an 8% SDS gel depicting the quality of the eluted GST-OppA protein fractions after the introduction of reduced glutathione. The protein samples were prepared using the individual lysozyme/protease inhibitor approach. Lane 1: PageRuler^{TM} Unstained Protein Ladder. Lane 2: eluted fraction 1. Lane 3: eluted fraction 3. Lane 4: eluted fraction 5. Lane 5: eluted fraction 7. Lane 6: eluted fraction 9. Lane 7: eluted fraction 11. Lane 8: eluted fraction 16. **Figure 4.7 (b)** Graph illustrating the different concentrations (in mg/µl) of the eluted fractions of the Ms02 GST-OppA protein from the glutathione agarose column.

4.2.3.11. Western blot analysis to confirm the identity and quality of the isolated protein

To confirm that the purified protein was the GST-tagged OppA protein, a western blot was performed using the anti-GST antibody. The western blot confirmed the presence of the GST-OppA protein, and also detected the GST protein at the bottom, indicating the possibility that the Ms02 GST-OppA protein loses its GST tag either during the purification process, or soon afterwards. Figure 4.8 indicates the western blot performed with the anti-GST antibody on the eluted protein fractions acquired with the different methods of sample preparation as described in Section 4.1.3.1 to Section 4.1.3.2. The Ms03 GST-OppA protein was also tested, and therein multiple bands were detected with the GST antibody, together with the GST protein at the bottom. This might be due to degraded protein, or that the type of isolation also elutes the GST protein. Therefore, the Ms02 protein might be degraded at the linker sequence that connects the GST protein with the OppA protein, resulting in a GST-tagged OppA protein, and a GST protein that were removed from the OppA protein.

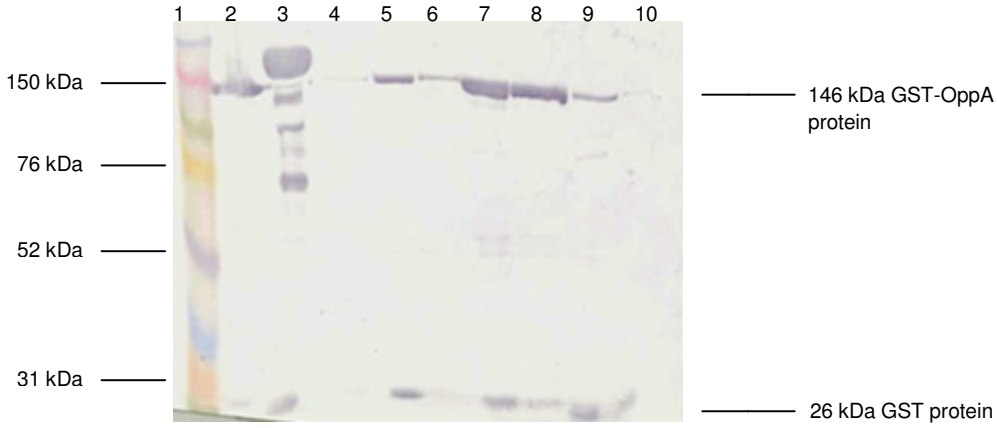

Figure 4.8 Image of a western blot depicting the GST-OppA protein fractions as detected by the anti-GST antibody. The proteins were separated on an 8% denaturing SDS gel before transfer to the nitrocellulose membrane. Lane 1: Full-range Rainbow marker. Lane 2: Ms02 GST-OppA cell culture. Lane 3: Ms03 GST-OppA purified protein. Lane 4: Ms02 GST-OppA protein from the batch purification. Lane 5 and 6: eluted fractions 8 and 11 of the simultaneous lysozyme/protease inhibitor Ms02 GST-OppA protein isolation. Lane 7 and 8: eluted fractions 12 and 15 of the individual lysozyme/protease inhibitor Ms02 GST-OppA protein isolation. Lane 9 and 10: fractions 5 and 21 of the lysozyme only Ms02 GST-OppA protein isolation.

4.2.3.12. ELISA to determine the rabbit immune response and antibody activity

An **e**nzyme-**l**inked **i**mmunosorbent **a**ssays (ELISA) was performed to test the rabbit antibody activity. For that purpose, a dilution series of the rabbit sera was prepared for day 0, day 28 and day 42 respectively (Figure 4.9). The dilutions were tested against three different protein concentrations. From the graph it is clear that no antibodies were present on day 0. A slight difference can be seen between day 28 and 42, but in order to compare the antibody titre on the different days, a graph to illustrate the rabbit antibody response was generated (Figure 4.10). All three of the protein concentrations indicate an increase in the rabbit antibody titre values over time. The 1 µg and 2 µg protein coated as antigen produced almost similar results, while the curve of the 5 µg protein is slightly lower. ELISAs depend on the balance between coating antigen and the ability of the antibody to bind to that specific antigen concentration. It is therefore necessary to test different concentrations of antigen against the sera (antibody) dilutions in order to optimise this balance. This is visible in Figure 4.9, where the last series of dilutions of the day 28 serum reacting with the 5 µg antigen coat reached negative values of absorbance. Optimising the reactivity of the antibody and the antigen will reduce any background noise and enhance the signal without compromising reliability of the results.

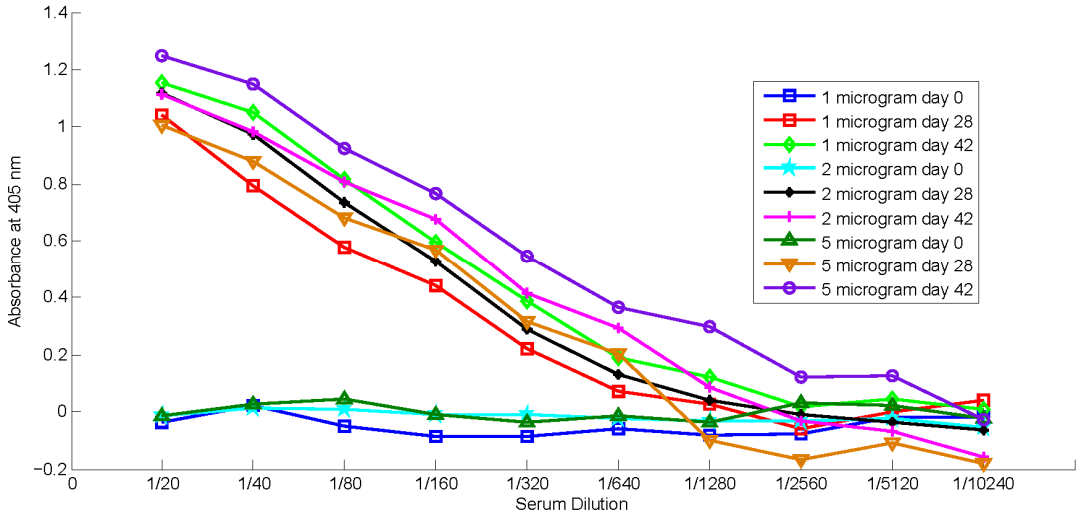

Figure 4.9 Measurement of the rabbit antibodies elicited against the Ms02 GST-OppA protein. Serum was collected on day 0, day 28 and day 42. Serum samples were diluted between 20 and 10 240 times and the absorbance recorded at 405 nm. The values were used to calculate the antibody titre values over the 42 day period for each of the different protein concentrations to illustrate the rabbit immune response against the GST-OppA protein.

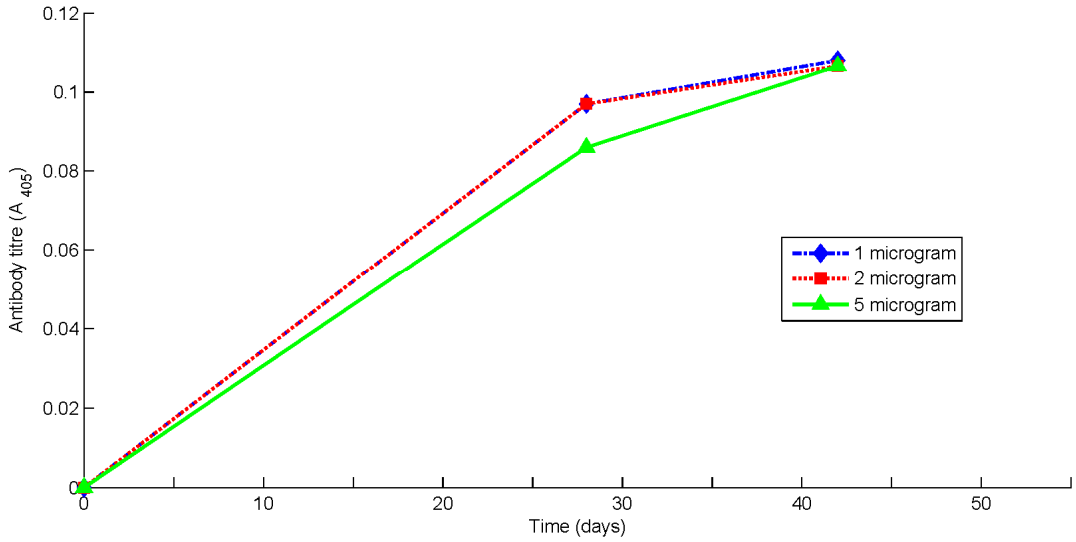

Figure 4.10 The antibody response in the rabbit against the Ms02 GST-OppA protein over a 42 day period as determined using the ELISA method. The serum was diluted and used to bind to 1, 2 or 5 µg of Ms02 GST-OppA protein utilised as the coating antigen.

4.2.3.13. Production of rabbit antibodies against the Ms02 GST-OppA protein

The rabbit sera collected on day 42 were diluted and successfully used in a western blot analysis, indicating the response of the antibodies against the Ms02 GST-OppA protein. Two bands were observed in all three dilutions (Figure 4.12a), the sizes supporting the theory that the GST protein tag is removed and that the protein exist as GST-tagged, and without the GST tag. No binding to the GST protein was observed, thus the rabbit did not form antibodies against the GST moiety itself. The possibility of truncated proteins also has to be considered, since

the codon bias of *E. coli* could cause translational problems. However, truncated proteins would cause numerous bands to be detected, both with the anti-GST antibody and the anti-GST-OppA antibody. The small size of the GST protein in relation to the Ms02 OppA protein suggested that it would not be neccessary to remove the GST tag. The antibodies were also tested against the Ms01 and Ms03 proteins using a sera dilution of 1:2 500 (Figure 4.12b). No binding could be detected on these two proteins; again the antibodies in the sera only bound to the Ms02 GST-OppA protein.

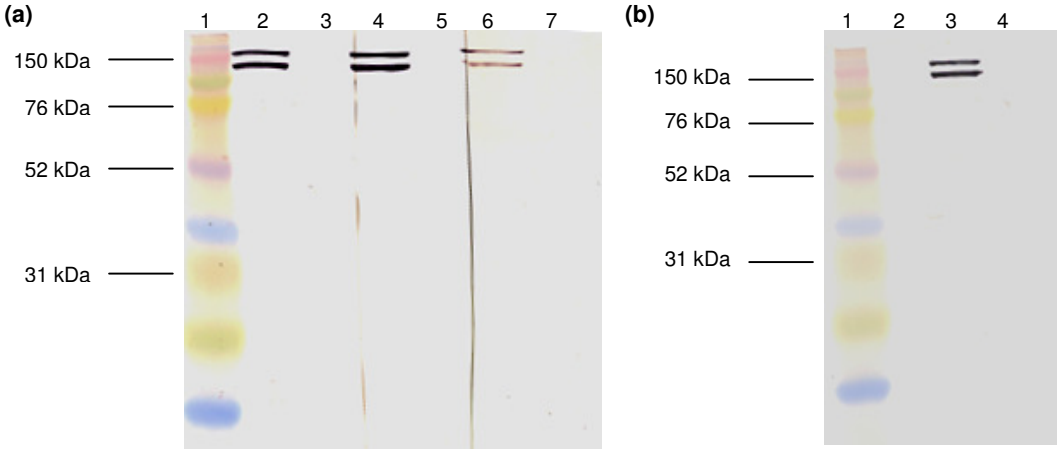

Figure 4.12(a) Image of a western blot depicting the different dilutions of the anti-OppA rabbit antibodies and their binding to the Ms02 GST-OppA protein. The proteins were separated on a 12% denaturing SDS gel before transfer to the nitrocellulose membrane. Lane 1: Full-range Rainbow marker. Lane 2: GST-OppA protein with a sera dilution of 1:2 500. Lane 3: GST protein with a sera dilution of 1:2 500. Lane 4: GST-OppA protein with a sera dilution of 1:1 000. Lane 5: GST protein with a sera dilution of 1:1 000. Lane 6: GST-OppA protein with a sera dilution of 1:5 000. Lane 7: GST protein with a sera dilution of 1:5 000. **Figure 4.12(b)** Image of a western blot demonstrating the ability of the Ms02 anti-OppA rabbit antibodies to only bind to the Ms02 GST-OppA protein and not to the Ms01 or Ms03 GST-OppA proteins. The proteins were separated on a 12% denaturing SDS gel before transfer to the nitrocellulose membrane. Lane 1: Full-range Rainbow marker. Lane 2: Ms01 GST-OppA protein. Lane 3: Ms02 GST-OppA protein. Lane 4: Ms03 GST-OppA protein.

4.2.4. DISCUSSION

The results indicated that the cloning strategy employed to insert the Ms02 *oppA* gene into the expression vector and the three DNA vaccine vectors, was successful. First the *oppA* gene was successfully amplified using the primers with the appropriate RE recognition sites, and three different pGEM®-T Easy constructs were produced. To subclone the gene into the prokaryotic expression vector, pGEX-4T-1, a *BamHI* recognition site was added to the 5' end and a *SalI* recognition site to the 3' end of the *oppA* gene. The process yielded a plasmid with the modified gene, free of any unwanted mutations, as determined by sequencing of the entire gene. In order to construct the VR1020 vector with the *oppA* gene, a pGEM®-T Easy_OppA plasmid with a *BamHI* recognition site on both ends of the *oppA* gene was generated. Even though one would expect to struggle due to the possibility of gene inversion, the VR1020_OppA vector proved to be the easiest to construct. The VR1012_OppA vector, on the other hand, was more troublesome. A successful, mutation free construct was however obtained after a few colonies had to be screened. For subcloning into the pCI-neo vector, the two restriction sites *AccI* and *MluI* were previously added to the *oppA* gene and cloned into the pGEM®-T Easy plasmid by Steenmans (2010).

The pCI-neo_OppA vector was also quite difficult to obtain: in approximately 100 colonies, only three tested positive for the gene insert, and from these three, only one was ultimately useful.

The pGEX-4T-1 vector used in this study employs the Glutathione-S-transferase (GST) gene fusion system: a multipurpose system that enables the expression, purification and detection of recombinant proteins using the *Schistosoma japonicum* GST protein tag. This GST protein is then expressed at the amino terminus of the recombinant protein. Even though this system has been used successfully for expressing a wide variety of soluble proteins in the bacterial cytoplasm, as well as several mycoplasma genes (Robino *et al*., 2007; Rosati *et al*., 2000), expressing the Ms02 GST-OppA protein proved to be quite different to the usual protocol followed. Numerous optimisation strategies were investigated, first to try and improve the expression of soluble protein, and second to improve solubilisation after expression. Tactics employing denaturing agents such as GuHCl and β-mercapto-ethanol, extreme pH conditions, and lower induction temperatures, were investigated. In the end, protein was optimally expressed just by increasing the cell density before induction. In other words, the cells were grown for a longer period of time before induction, and then cells were harvested shortly after to limit the expression time. This resulted in expression of the optimal amount of soluble protein, yielding a protein band of just below 150 kDa. The predicted size of the Ms02 OppA protein is 120 kDa, thus, adding the 26 kDa of the GST moiety, this would result in a recombinant protein of 146 kDa. Western blot analysis with an anti-GST probe indicated the presence of a GST-tagged protein in the pGEX_OppA cell samples (under 150 kDa), and a GST protein (at just below the 30 kDa marker protein band) in the control samples with no gene insert. This proved that successful expression was achieved.

Purification of the protein was achieved with affinity chromatography. Purification of the Ms02 GST-OppA protein proved to be another hurdle that took some optimisation to overcome. The optimum conditions to produce soluble protein were determined to involve a lysozyme incubation step, followed by a protease inhibitor incubation step. This allowed the cells to be thoroughly lysed and to release the protein. The protease inhibitor was added after it was noticed that the isolated protein seemed to be degraded. Multiple bands were visible on the SDS gel and it was thought that this was due to the release of proteases when the cells were lysed with lysozyme. Adding the protease inhibitor, after a 30 min incubation step with the lysozyme, resulted in purified OppA protein of a satisfying quality and quantity. Even though it is sometimes advisable to remove the fusion protein tag, it was decided that the relative size of the GST moiety with regards to the OppA protein was of lesser importance. It was decided to proceed with the rabbit immunisations using the Ms02 OppA fusion protein.

A rabbit was immunised with the purified Ms02 GST-OppA protein and its antibody activity proven with ELISA. Western blot analysis using the rabbit serum as a primary antibody indicated that the rabbit antibodies reacted only to the purified Ms02 GST-OppA protein, and not to the GST protein. In addition, the antibodies also did not bind to purified Ms01 and Ms03 GST-OppA proteins. The Ms02 and Ms01 genomes are more closely related to each other than the Ms03 genome (refer to Chapter 2, Figure 2.1) and possible cross-reactivity of the Ms02 and Ms01 OppA antibodies could be expected. Previous studies in our laboratory did detect the Ms03 protein with Ms02 antibodies, but in that instance the antibodies were tested against the whole mycoplasma organism and not just the purified protein. In addition, the Ms01 antibodies reacted only to the Ms01 organism, and the Ms03 antibodies only to the Ms03 organism. The ability of rabbit anti-Ms03 OppA antibodies to react to Ms01 GST-

OppA protein, but not the Ms02 GST-OppA protein, were illustrated in the course of an Honours study in this laboratory (Collop, 2011). It is possible in that study that the Ms02 protein sample used in the western blot analysis was not ideal, since it was not a purified sample, but a cell sample with expressed protein. At the time of the Ms03 experiment, there were still problems with the Ms02 protein expression, and a possible explanation is that in that particular cell sample, the Ms02 protein was not present, or not in a usable form. However, in this study no cross-reactivity was detected from the Ms02 OppA antibody with the other mycoplasma proteins. This requires further investigations into the cross-reactivity of the antibodies: both towards the purified proteins and the whole organisms. This serves to illustrate the necessity to develop a vaccine specific to each mycoplasma, since cross reactivity is too low, or as in this case, absent.

During the course of the study, it was discovered that the GST moiety is lost from the recombinant protein either during the protein isolation procedure, or soon afterwards. This was indicated by western blot analysis, where the anti-GST antibody would detect one large band at 150 kDa, and a small band at the GST protein size of 26 kDa, but the rabbit anti-OppA antibodies would detect two bands at the 150 kDa marker, one slightly below the other. It is possible that the Ms02 GST-OppA protein was present both with the GST moiety and without the GST moiety. The size difference supports this theory: a difference that could account for the 26 kDa GST protein. This occurence was also noted by Brandt (2012), who detected the separate GST protein and the Ms01 GST-OppA protein, both on an SDS gel as well as with a western blot performed with the anti-GST antibody. The possibility of truncated proteins exists, due to the strong codon bias of *E. coli*, but truncated proteins would have resulted in multiple detected bands, and not just the two proteins as were detected. Therefore truncation was not considered to be the cause of the double band observation. One could ascertain the identity of the two protein bands by mass spectrometry. However, this would ential the cumbersome process of excising the bands directly from the SDS gel and purifying the two proteins. The pGEX-4T-1 vector linker sequence contains a thrombin cleavage site to enable removal of the GST-tag. Natural proteases that are present in the cell could influence the stability of the recombinant protein through non-specific cleavage at the thrombin cleavage site. The GST-OppA protein would therefore its GST-tag quite rapidly, resulting in a detectable GST protein with the anti-GST antibody, and two proteins with the anti-OppA antibody.

4.2.5. CONCLUSION

This study produced three DNA vaccine vectors each containing the Ms02 *oppA* gene: pCI-neo, VR1012 and VR1020. These DNA vaccine vectors can now be used in ostrich vaccine trials to determine the efficacy thereof. Purified Ms02 GST-OppA protein and antibodies against the protein were also produced. The purified protein may be used as a booster vaccine, since studies indicated the efficacy of first vaccinating with a DNA vaccine, then following up with a protein booster vaccine. In addition to this, the purified protein can be used as antigen in future serological tests to develop ELISA assays to test for ostrich immune responses. These proteins and antibodies are also useful to test the expression efficiency of DNA vaccine vectors in eukaryotic cell systems. Testing of the expression efficiency of pCI-neo, containing the Ms02 *oppA* gene, in such a cell system, will be discussed in the next chapter.

Chapter 5

This chapter contains a broad overview on animal cell culture, transfection, the cell expression system, and its applications. This is followed by the experimental procedures and results of the preliminary study on the expression of the pCI-neo_OppA vector in a cell culture system.

5.1. A SHORT INVESTIGATION INTO CELL CULTURES AND TRANSFECTION

5.1.1. INTRODUCTION

The term cell culture refers to the complex process by which cells, mostly single-cellular eukaryotes, are grown under experimental conditions. At the end of the 19th century, Wilhelm Roux took a portion of the medullary plate from a chicken embryo and maintained it in a saline solution for a few days on a flat glass plate, demonstrating the ability of cells to survive outside of the donor organism. Thus the practice of using cell cultures arose more than a hundred years ago, but it was the need for viral vaccines in the 1950s and the onset of recombinant protein technology in the 1970s and 1980s that accelerated the design and improvement of mammalian cell culture bioprocesses (Kretzmer, 2002). In 1955, Harry Eagle mixed naturally occurring cells with a solution of synthetic nutrients and proved that pure cell lines can be preserved for an indefinite period of time. The first mass-produced product using cell culture technology was the injectable polio vaccine developed by Jonas Salk. Proteins are currently the foremost licensed biopharmaceutical product, with 70% being produced in mammalian cells (Carta and Jungbauer, 2010; Wurm, 2004). There are two types of cell lines: primary cell lines and continuous cell lines. Primary cells, such as from a mammalian embryo, cannot survive outside of the original organism indefinitely, while continuous cell lines (sometimes known as immortal cell cultures), such as those originating from tumours, can survive indefinitely outside of the organism.

5.1.2. ANIMAL CELL CULTURE

Animal cells can be used to produce therapeutic proteins, subunit vaccines, diagnostic products, and viral vaccines. Other applications include gene and cell therapy, regenerative medicine research, as well as *in vitro* research and toxicology studies (Pham *et al.*, 2006). For production of complicated eukaryotic proteins, mammalian cells are the preferred choice due to post-translational modifications needed to obtain active protein (Wurm, 2004).

Animal cells are more vulnerable than bacteria and fungi cells that are routinely used in experiments. The cells are more prone to damage and grow more slowly, which requires extremely sterile working conditions, and are dependant on several variables such as pH, temperature, osmolality and gas concentrations (Freshney *et al.*, 2005). Another characteristic of animal cells is that their metabolism in culture is different than in the organism of origin: an increase in glycolytic flux and a resulting increase in the amount of lactose formed (due to inability to metabolise glucose completely), are observed. The lactate build-up leads to high acidity levels and a rise in osmolality. Even under fully aerobic conditions, the build-up of toxic metabolites inhibits cell growth and expression of the recombinant protein (Nelson and Cox, 2004).

Certain requirements for the ideal cell culture system arise out of the need to maximise protein expression, including control of temperature, pH, gas exchange, adequate supply of nutrients in the culture medium, and steps to avoid cell contamination. Furthermore, the cell line itself is a crucial choice: it should be able to transfect easily, and be able to transcribe, translate and process the recombinant protein. An additional advantage is if the protein is secreted into the medium, because this simplifies the extraction and purification process when proteins are produced for therapeutic purposes (Pham *et al*., 2006).

5.1.2.1. Cell line cross-contamination

Studies suggest that up to 20% of the cells used in culture are either misidentified or contaminated with another cell line. Contamination has far reaching effects on the quality of research and the production of therapeutic products. Cells have to be authenticated early in passaging (splitting the cells into subcultures) to identify any possible contaminations, and contamination tests have to be repeated before freezing cell line stocks, as well as every two months during active culturing. Thus the emphasis with all cell manipulations falls on sterility, to avoid contamination with bacteria, yeast, or other cell lines. Antibiotics and antifungals can also be added as an extra measure of prevention. Mycoplasma contamination often goes undetected because mycoplasma growth is not turbid, a trait that is rather associated with bacterial and fungal growth (Rottem, 2003). Therefore, cultures need to be tested regularly for mycoplasma contamination.

5.1.2.2. Manipulation and maintenance of cells

An effective system requires a stable and consistent number of cells. This can be achieved by passaging at regular intervals. Passaging avoids the senescence associated with prolonged high cell density and the cells can be cultured for longer periods of time. Several problems arise as cells increase in the medium, such as nutrient depletion of the growth medium and accumulatation of apoptotic cells. Contact inhibition may also occur, where cell to cell contact can cause cells to stop dividing, as well as stimulate cell differentiation. Too much cell passaging may decrease the transfection efficiency, alhough this also differs between different cell types. NIH-3T3 cells demonstrated increased transfection efficiency when 56 passages were compared with 2 passages. CHO cells demonstrated stable transfection efficiency from 2 to 26 passages (Jacobsen *et al*., 2004).

5.1.3. TRANSFECTION

Transfection refers to the deliberate introduction of nucleic acids or proteins into eukaryotic cells. Genetic material such as DNA and RNA, or proteins such as antibodies, can be transfected in the cells by opening pores in the cell membrane to allow uptake of the material.

To make use of gene therapy, the introduced DNA has to enter the nuclei of the target cells. Mulligan and Berg (1980) were the first to report successful transfection of cells with recombinant DNA plasmid, using the calcium phosphate transfection method. By 1980, a liposome-mediated transfection method was developed (Fraley *et al*., 1980). Later in the same decade a polylysine (PLL)-based transfection method was developed (Wu and Wu, 1987). In 1995 Boussif *et al.* demonstrated a polyethyleneimine (pEI) based delivery system.

The ever-expanding variety of delivery systems, cell lines, and variety of protocol conditions make a direct comparison between different gene delivery studies almost impossible (van Gaal *et al*., 2011).

The different methods of transfection can be broadly classified as chemical (Orrantia and Chang, 1990), non-chemical (Escriou *et al*., 1998; Maasho *et al*., 2004), particle-based (Bertram, 2006), and viral (Fisher *et al*., 1996; Jones and Hartl, 1998; Vandenberghe *et al*., 2006).

5.1.3.1. Transfection efficiency

Transfection efficiency is incfluenced by the DNA/reagent complex formation, the incubation time of the DNA with the transfection reagent, and the dose of transfection. The complex formation of the DNA and the transfection reagent is influenced by numerous factors, including the ratio of reagent to DNA, the media pH and ionic strength, the order of mixing of the reagent, media, and DNA before the transfection (reviewed in van Gaal *et al*., 2011; Riss and Moravec, 2004). In order to develop a successful production process, the correct expression plasmid, host cell, culture medium, DNA quality, and transfection vehicle have to be selected (Graham *et al*., 1977; Pham *et al*., 2006). Optimising these factors will ensure the desired level of transfection, expression and cellular productivity.

5.1.4. CELL EXPRESSION SYSTEM

There are four basic steps in creating a mammalian cell expression system for recombinant protein production. First, the vector is constructed to include the gene(s) of interest (GOI) (refer to Basic requirements for a DNA vaccine vector in Chapter 4, Section 4.1.2.1, for more detail). Then preliminary expression is performed for evaluation purposes. Third, stable clones are developed from the cell line, and lastly, the medium should be optimised to provide the nutrients for the production clone. Therefore, there are many parameters that have to be taken into account to optimise gene delivery to the nucleus. The transfection efficiency is dependent on factors such as medium composition, cell passaging and seeding, and the protocol used in terms of media volume, incubation time, and other parameters (Promega Protocols and Applications Guide).

5.2. PRELIMINARY INVESTIGATIONS INTO THE EXPRESSION OF THE Ms02 *oppA* GENE IN THE pCI-neo DNA VACCINE VECTOR USING A EUKARYOTIC CELL SYSTEM

5.2.1. INTRODUCTION

The principle of DNA vaccination is based on the expression of the chosen vaccine candidate gene in the eukaryotic host. Good vaccine candidate genes are the genes that encode for attachment proteins involved in cytadherence, as well as proteins located on the surface of the pathogen, and proteins involved with essential cellular functions. Once an appropriate antigenic gene is chosen, the gene is cloned into a DNA vaccine vector and injected into the host. Inside the host cells, the plasmid is then expressed, resulting in the production of the recombinant antigenic protein. Since the protein is recognised as foreign by the host's immune system,

an immune response is launched against the protein. In this manner the host produces antibodies and memory cells against the antigen, enabling it to protect itself against infection by the pathogenic organism.

Different DNA vaccine vectors have different expression profiles, thus resulting in different levels of the elicited immune response. Ostrich DNA vaccine trials where the ostrich sera are tested in serological assays, such as **e**nzyme-**l**inked **i**mmunosorbent **a**ssays (ELISAs), measure the immune response against the expressed protein that was produced in the ostrich host. In cases where no immune response is detected, it does not necessarily indicate the absence of an ostrich immune response, but could indicate that the vector did not express correctly in the ostrich (or other eukaryotic host). Therefore, in order to corroborate vaccine trials, it is necessary to confirm protein expression in a eukaryotic system, so that future vaccination trials are conducted with the prior knowledge that the vaccine vector does express correctly in eukaryotic cells. One way to confirm protein expression is to investigate the expression profile of the plasmid in a cell culture system. Antibodies against the protein indicate the presence of the recombinant protein, thus confirming expression of the DNA vaccine vector.

In order to develop a DNA vaccine against the ostrich mycoplasma Ms02, the *oppA* gene was cloned into three DNA vaccine vectors, namely pCI-neo, VR1012 and VR1020 (Chapter 4). Purified OppA protein was injected into rabbits to generate OppA antibodies (Section 4.2.2.8). In this study, the pCI-neo DNA vaccine vector was used to transfect mammalian cells. The OppA antibodies were used to detect possible expression of the recombinant protein in the cell culture system.

5.2.2. EXPERIMENTAL PROCEDURES

5.2.2.1. Preparation of the pCI-neo_OppA vector to be used in transfection

In order to prepare the DNA for transfection into the mammalian cells, it was first necessary to increase the amount of vector available. To obtain a large volume of cell culture that contain the pCI-neo_OppA vector, a freeze culture of *E. coli* JM-109 cells (Promega), containing the vector, was streaked onto Luria-Bertani (LB) plates with 15 g/L agar, supplemented with 1 µl/ml (100 µg/ml) ampicillin (Sigma-Aldrich). The plates were incubated for 16 h at 37°C before colony formation was analysed.

5.2.2.1.1. Diagnostic PCR to confirm the presence of the pCI-neo_OppA vector

Before starter cultures were made for the isolation, a diagnostic PCR was performed to ensure the picked colonies did contain the pCI-neo_OppA vector. From the plates, 8 colonies were picked to serve as a DNA template in a diagnostic PCR. Approximately half of each of the chosen colonies were scraped with a sterile toothpick and mixed in respective reaction tubes containing 1 µl of 10x Reaction buffer, 0.1 µl of 5 U/µl Super-Therm *Taq* DNA polymerase, 0.6 µl of 25 mM $MgCl_2$ (all from JMR-Holdings, USA), as well as 0.4 µl of 5 mM dNTPs (Bioline), 0.5 µl of each primer (20 pmol/µl), and 6.9 µl of Milli-Q® water. The forward primer T7EEV (5'-AGGGCTAGAGTACTTAATACG-3) and the reverse primer T3 (5'-AATTAACCCTCACTAAAGGG-3') were used to amplify the gene insert. The amplifications were carried out in a 2720 Thermal Cycler Version 2.09 (Applied Biosystems Ltd., Warrington, UK) as follows: initial denaturation of 10 min at 94ºC, followed by 25 cycles of 94ºC for 30 s, 55ºC for 30 s, and 72ºC for 30 s, with a final extension step at 72ºC for 7 min. The PCR products were analysed on an 1% (w/v) agarose gel as described in Chapter 3, Section 3.2.2.4.

5.2.2.1.2. Large scale isolation of the pCI-neo_OppA vector

Four single positive colonies were then inoculated into respective Erlenmeyer flasks containing 10 ml LB-medium, supplemented with 10 µl (100 µg/ml) ampicillin and incubated on an IKA® KS 260 Basic orbital shaker for 8 h at 37°C, while being agitated at 220 rpm. After the 8 h growth, each 10 ml culture was then diluted in 100 ml fresh LB/amp-medium and allowed to grow for a further 12 h at 37°C, while being agitated at 220 rpm. The cells were harvested by centrifugation at 6 000 x g for 10 min (4°C) and the supernatant was discarded as effectively as possible. Thereafter the NucleoBond® Xtra Maxi kit (Macherey-Nagel, Germany) was used to purify the vector according to the manufacturer's instructions, save a few differences. In short, the cell pellet was resuspended in 18 ml Resuspension buffer (RES-EF) + RNaseA by pipetting up and down so that no clumps remained. To lyse the cells, 18 ml Lysis buffer (LYS-EF) was added to the suspension and the tube inverted 5 times by hand, followed by a 5 min incubation step at room temperature. During the incubation step, the NucleoBond® Xtra Column with the inserted column filter was equilibrated by applying 35 ml Equilibration buffer (EQU-EF) onto the rim of the column filter. After this step, care was taken that the column did not run dry. To neutralise the lysis buffer, 18 ml Neutralisation buffer (NEU-EF) was added and the lysate gently mixed by inverting 15 times, followed by 5 min incubation on ice. Before applying the lysate to the equilibrated column filter, the suspension was inverted 3 times, ensuring a homogenous mixture to avoid clogging the filter. The filter was washed with 15 ml Equilibration buffer (FIL-EF) to wash out the remaining lysate. The column filter was then discarded and the column washed with 90 ml Wash buffer (ENDO-EF), and allowed to empty by gravitation. This was followed by another wash step with 45 ml Wash buffer (ENDO-EF). The DNA was then eluted with 15 ml Elution buffer (ELU-EF) and collected in a 50 ml centrifuge tube. The elution was filtered through the NucleoBond® Finalizer (a DNA-binding column) and the flow-through discarded. The NucleoBond® Finalizer was washed with 70% ethanol, after which care was taken to dry the NucleoBond® Finalizer as completely as possible by plunging down a few times with the plunger with the NucleoBond® Finalizer attached. To elute the vector, 400 µl H_2O-EF was pressed through the NucleoBond® Finalizer, transferred back into the syringe and pressed through again. This was repeated twice again with 400 µl fresh H_2O-EF.

5.2.2.1.3. Determination of the quality and quantity of the pCI-neo_OppA vector

The concentration and quality of the isolated vector were determined spectrophotometrically with the Novell® Nanodrop ND-1000 (v 3.5.1) at 260 nm. Concentration (ng/µl) and purity (A_{260}/A_{230} and A_{260}/A_{280}) were measured. The vector was loaded onto an agarose gel in order to ascertain the topology and purity (Section 3.2.2.4).

5.2.2.1.4. Restriction enzyme treatment of the pCI-neo_OppA vector with MluI and AccI

The pCI-neo_OppA vector was constructed as described in Chapter 4, Section 4.2.2.2. The vector had a *MluI* recognition sequence on the 5' end and an *AccI* recognition sequence on the 3' end of the Ms02 *oppA* gene. The vector was first subjected to digestion with the restriction endonuclease (RE) *MluI* in two 20 µl reactions each containing Milli-Q® water, 2 µl 10x RE buffer, 0.2 µl (10 mg/ml) acetylated bovine serum albumin (BSA), 1 µg of the vector, and 1 µl of 10 U/µl of *MluI* (Promega). Both the reactions were incubated at 37ºC for 1 h, followed by purification with the DNA Clean & Concentrator™-5 kit (Zymo Research) according to the manufacturer's instructions. One of the reactions was then treated with the *AccI* endonuclease. The 20 µl

reaction contained 2 µl of 10x RE buffer, 0.2 µl (10 mg/ml) acetylated BSA, 8 µl of the purified DNA, and 1.7 µl of 6 U/µl *AccI* (Promega), and Milli-Q® water to make up the volume. After 1 h incubation at 37°C, the product was purified with the DNA Clean & Concentrator™-5 kit. In order to analyse the digestion products, 10 µl of each of the reactions were separated on a 1% (w/v) agarose gel, together with 2 µl of the uncut pCI-neo_OppA vector (Section 3.2.2.4).

5.2.2.2. Cell culture

The cell line MDA (human breast carcinoma cells) was kindly provided by Ms CJ Langeveldt from the Donita Africander's Reasearch Group, Department of Biochemistry, University of Stellenbosch (RSA). Cultures were maintained in 75 cm^3 culture flasks (Greiner Bio-One Int., Austria) in high glucose (4.5 g/ml) Dulbecco's Modified Eagle Medium (DMEM) (Sigma-Aldrich), supplemented with 10% heat inactivated fetal calf serum (FCS), 100 (1 U/ml) penicillin (Sigma-Aldrich) and 100 µg/ml streptomycin (Sigma-Aldrich) at 37°C in an atmosphere of 90% humidity and 5% CO_2. Cells were plated at 1.25 x 10^5 cells per well in 12-well plates (Greiner Bio-One) the day before transfection.

5.2.2.3. Transfection with the pCI-neo_OppA vector

In order to transfect the cells, the FuGENE® 6 Transfection Reagent (Promega) was used, a propriety blend of lipids and other components. The non-liposomal lipids and polymers form complexes with DNA, with the potential to form micelles. Performing transfection under aqueous conditions enables the lipophilic portion of the ampiphilic compound to form the micelle core within which the exogenous nucleic acids are ensconced. Four different vector amounts were transfected, namely 0.5 µg, 1.0 µg, 1.5 µg and 2.0 µg. Each was done in triplicate. For every µg of vector DNA, 4.5 µl FuGENE® 6 was added. The transfection reagent was added first to a serum-free medium, before it was added to the DNA. After mixing, the transfection reagent/medium/DNA was incubated for 30 min. During this time, the media on the adherent cells were aspirated and replaced with 1 ml supplemented medium. The appropriate amount of plasmid mixture was added to the wells, the plate swirled to mix, and returned to the incubator. Two 12-well plates were transfected. The cells were then harvested from one plate after 24 h and from the other plate after 48 h. The media from the cells were stored at -20°C. The cells were washed by adding 1 ml PBS (137 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4{\cdot}2H_2O$, and 2 mM KH_2PO_4, pH 7.2) to each well, with care taken not to disturb the cells on the bottom of the well. After the PBS was aspirated, 80 µl SDS lysis buffer (0.125 M Tris-HCl, 4% (m/v) sodium dodecyl sulfate (SDS), 20% (v/v) glycerol, 10% (v/v) β-mercapto-ethanol, pH 6.8), with added 0.1% (m/v) bromophenol blue) was added to each well. The buffer was mixed thoroughly inside the well with the pipette tip to ensure all the cells were loosened and suspended in the buffer, thus ensuring complete lysis. The cell samples were then incubated at 97°C for 10 min before storage at -80°C.

5.2.2.4. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis

SDS-PAGE was used to analyse the expression profiles of the cells, as well as detect whether any proteins were present in the media. For this purpose, samples were loaded onto a denaturing SDS gel composed of a 4.5% stacking and an 8% resolving gel, and electrophoresed at a constant of 20 mA. The media samples needed to

be prepared first, as the cell samples already contained the reducing treatment buffer and the loading dye, added during the cell harvesting procedure. In order to prepare the media samples for SDS-PAGE, 30 µl of medium was added to 30 µl of SDS lysis and the mixture boiled for 2 min before being placed on ice. In the same manner, 20 µl of the purified Ms02 GST-OppA protein was prepared, as well as 30 µl of a previously expressed pGEX_OppA cell culture as control samples.

5.2.2.5. Western blot analysis

For detection of the expressed Ms02 OppA protein in the cell cultures, the same antibodies and basic procedure were used as described in Chapter 4, Section 4.2.2.8.3. A few optimisations were performed in order to increase the signal-to-noise ratio, mainly regarding the blocking buffer. An SDS-PAGE was performed as described above. The separated proteins were transferred to a 0.45 µm nitrocellulose membrane (Schleicher and Schuel, Sigma-Aldrich) in an electrode buffer (50 mM Tris-HCl, 200 mM glycine and 20% (v/v) methanol, pH 8.3) for 16 h at a constant of 120 mA. After the transfer, the membrane was stained with Ponceau S (2% (w/v) Ponceau S in 30% trichloroacetic acid and 30% sulfosalicylic acid) to ensure that sufficient transfer was achieved. Staining was allowed to occur for 5 min at room temperature while being agitated, followed by destaining for 2 min with Milli-Q® water before the membrane was investigated, and a final destaining for 10 min. The membrane was then blocked for 2 h at room temperature, while being agitated, in TBST blocking buffer (20 mM Tris-acetate, 150 mM NaCl, and 0.1% (v/v) Tween-20), with 5% (w/v) fat free milk powder (Elite, available from any convenience store). The rabbit serum was diluted 1:500 in TBST blocking buffer and incubated with the membrane for 1 h at 37°C at 4 rpm. This was followed by 3 wash steps of 10 min each with TBST at room temperature, while being agitated. The secondary antibody, goat anti-rabbit antibody (Sigma-Aldrich) was diluted 1:2000 in TBST blocking buffer and the covered membrane incubated for 1 h at 37°C, while being agitated. This was followed by 3 wash steps as described above. Lastly, the horse-radish peroxidase (HRP)-conjugated rabbit anti-goat antibody (Sigma-Aldrich) was diluted 1:2 000 in TBST blocking buffer and incubated as before, followed by the 3 wash steps. An additional 3 washes were performed with TBS buffer without the Tween-20 to remove the excess Tween-20 from the membrane. Ice-cold substrate, consisting of 0.05% (w/v) 4-chloro-1-naphtol, 16% (v/v) ice cold methanol, ice cold TBS and 0.025% (v/v) cold H_2O_2, was added to the nitrocellulose membrane and allowed to develop for 30 min at room temperature, while being agitated. After band development was observed, the membrane was washed with Milli-Q® water to prevent over-development.

For the Western dot blot analysis, the membrane was dotted with approximately 25 µl media, but in separate steps. Every 15 min, 5 µl of the medium from the 4 different transfections, was added drop-wise onto the membrane. After the last 5 µl medium was added, the membrane was allowed to dry for at least one hour. Thereafter, the membrane was incubated overnight in TBST blocking buffer at 4°C. The rest of the protocol remained the same.

5.2.3. RESULTS

5.2.3.1. Diagnostic PCR to confirm the presence of the pCI-neo_OppA vector

To prepare a starter culture for the large scale isolation of the pCI-neo_OppA vector, freeze cultures of the *E. coli* JM-109 cells containing the vector was first streaked out on LB/agar plates. A diagnostic PCR was performed to identify colonies that contained the vector with the insert. From the eight colonies tested, only one did not contain the pCI-neo_OppA vector (Figure 5.1). It was decided to use colonies 1-4 for starter cultures.

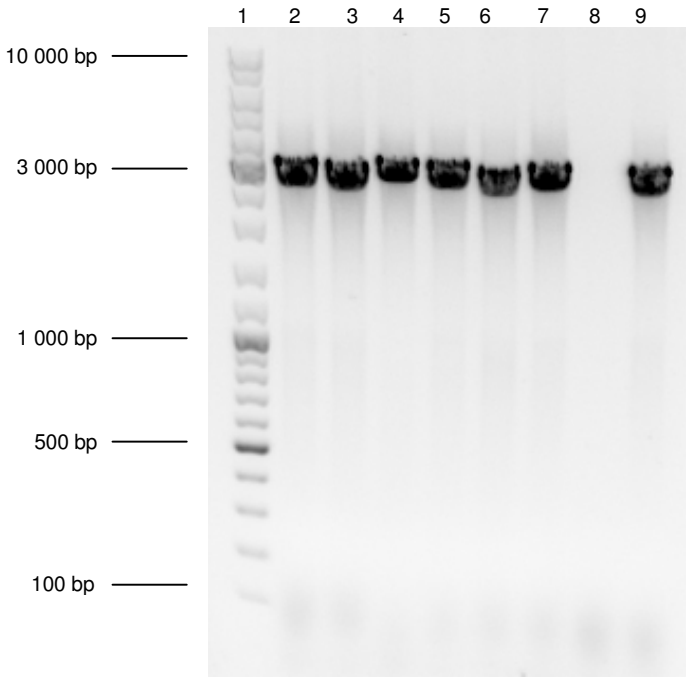

Figure 5.1 Image of the products from the diagnostic PCR on the cultures containing the pCI-neo_OppA vector. The products were separated on a 1% (w/v) agarose gel, cast with 0.1 µl/ml GelRed™ Nucleic Acid Gel Stain for UV visualisation of the DNA. Lane 1: 1 KB DNA GeneRuler™ DNA ladder mix. Lane 2-9: PCR product of colonies 1-8.

5.2.3.2. Isolation of the pCI-neo_OppA vector

The pCI-neo_OppA vector was isolated from a culture volume of 400 ml, which resulted in 1 736 ng/µl vector in an elution volume of 800 µl. The vector was subjected to restriction enzyme digestion to indicate the sizes. The conformation of plasmid DNA influences its migration during agarose separation. Five different conformations can occur. In order of slowest migration to fastest: nicked open-circular DNA (one strand cut); relaxed circular DNA (fully intact, but no supercoiled structures); linear DNA (either both strands have been cut, or the DNA exists in a linear form *in vivo*); supercoiled (DNA is fully intact with an integral twist, resulting in a compact form); and supercoiled denatured (slightly less compact than supercoiled, due to unpaired regions that may be caused by excessive alkalinity during plasmid preparation). Supercoiled DNA migrates faster than linear or open DNA due to its tight conformation. In order to indicate the size of the pCI-neo_OppA vector, it was cut with only one restriction enzyme (resulting in linear DNA), and with both restriction nucleases (resulting in the excised *oppA* gene and the linear vector without the gene), seen in Figure 5.2. The pCI-neo vector on its own is

5472 bp in length, the Ms02 *oppA* gene is 2 997 bp in length, thus accounting for an approximate added size of 8 500 bp. Thus the larger band as seen lane 5, that corresponds to the band seen in lane 4, represents the incomplete digestion (linearisation) of the vector, followed by a band that represents the linear vector without the *oppA* gene, and lastly the excised *oppA* gene itself.

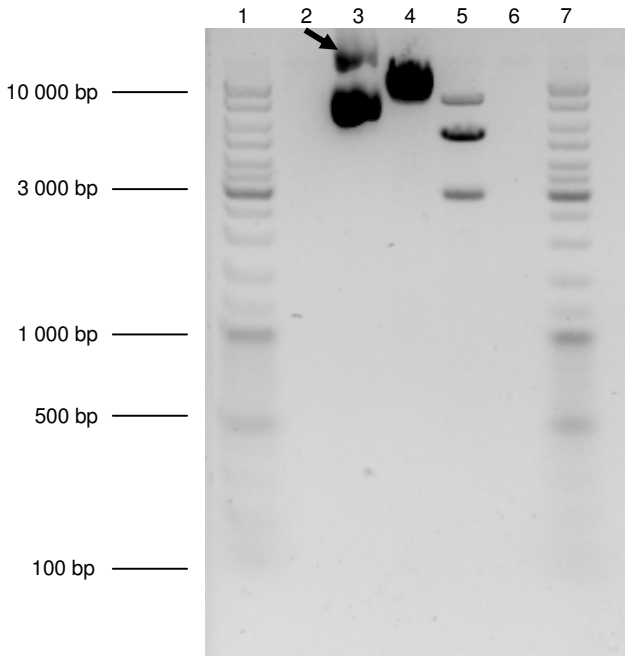

Figure 5.2 Image depicting the different conformations of the pCI-neo_OppA vector after electrophoretic separation. The nicked plasmid is indicated with an arrow. The products were separated on a 1% (w/v) agarose gel cast with 0.175 µg/ml ethidium bromide for UV visualisation of the DNA. Lane 1 and 7: 5 µl of 1 KB DNA GeneRuler^{TM} DNA ladder mix. Lanes 2 and 6 were left empty. Lane 3: uncut pCI-neo_OppA vector. Lane 4: pCI-neo_OppA vector cut with only *MluI*. Lane 5: pCI-neo_OppA vector cut with *MluI* and *AccI*.

5.2.3.3. SDS-PAGE of the transfected cells and media

An SDS-PAGE was performed to determine whether any proteins in the media or transfected cells corresponded to the size of the purified OppA protein. Figure 5.3 is an SDS gel illustrating the OppA protein sizes, both as a purified fusion protein and as part of an expressed cell suspension. The media from the transfected cells were compared and, according to the protein sizes as observed on the SDS gel, it is possible that there was OppA protein present. The OppA protein would be expressed without the GST protein, since the GST tag is a result of expression in the prokaryotic expression vector pGEX-4T-1. Thus the size of the mammalian expressed protein should correspond to the smaller size as visible in the Ms02 GST-OppA protein sample, the argument being that the smaller OppA protein is a result of the removal of the GST protein tag from the OppA protein. A light protein band in the media samples as separated in the SDS gel is visible that corresponds to the size of the smaller protein band (above 102 kDa). However, the amount of protein in the SDS sample is too low to allow definite conclusions, and the amount of sample loaded onto an SDS gel cannot be increased, due to the size of the wells.

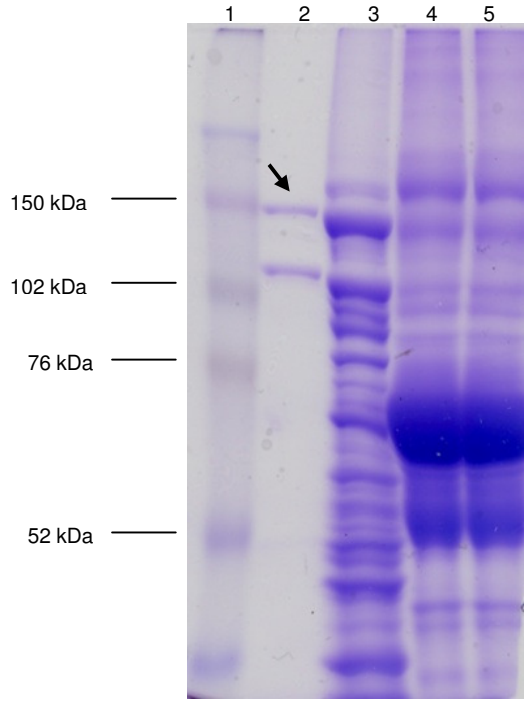

Figure 5.3 Media from the transfected eukaryotic MDA cells compared with Ms02 GST-OppA protein as a purified protein and as part of a prokaryotic cell expression sample. The proteins were separated on an 8% SDS gel. Lane 1: Full-range Rainbow marker. Lane 2: purified Ms02 GST-OppA protein. Lane 3: pGEX_OppA cell suspension sample. Lane 4: 0.5 µg transfected cell medium^{48h}. Lane 5: 2.0 µg transfected cell medium^{48h}. The Ms02 GST-OppA protein is indicated on the SDS gel with an arrow.

5.2.3.4. Western blot analysis

5.2.3.4.1. Ponceau S stain of the separated proteins

Ponceau S is a stain used to stain all of the proteins after transfer to a nitrocellulose membrane. In Figure 5.4, an image of such a stain can be seen. In this instance, both the transfected cells and the media were separated on an 8% denaturing SDS gel, together with purified GST-OppA protein as a positive control, and transferred to the nitrocellulose membrane. From the image it is clear that the SDS gel did not run uniformly, and that the proteins did not run in a straight line. During electrophoresis of the SDS gel, the observed front that is usually blue, changed colour as the proteins separated on the gel. When the media were concentrated and separated on an SDS gel, the samples were skewed and blotchy, as is also visible in the figure below. This was only observed when the media were separated: when only the cells were separated the same effect was not observed. This led to the conclusion that components of the media may possibly interfere with the normal separation on the SDS gel, and this could lead to a further influence on the western blot analysis.

The GST-OppA protein is indicated on the membrane with an arrow. The second protein band, which then represents the Ms02 OppA protein without the GST protein, is barely visible, and the same size band is also barely visible in the media samples. This is a promising result, and first a western dot blot was attempted on the media.

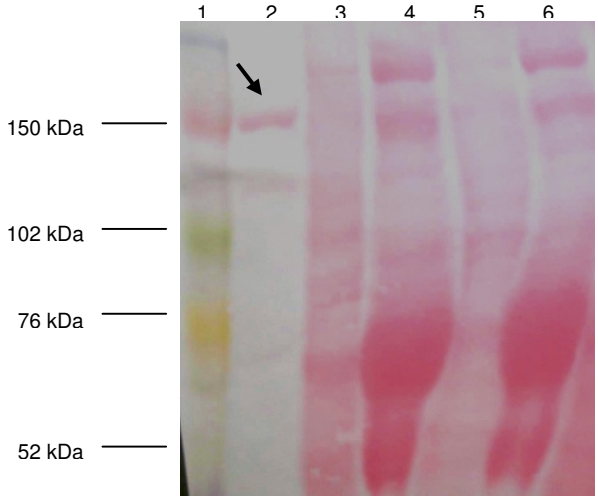

Figure 5.4 Ponceau S stain of the nitrocellulose membrane containing eukaryotic expressed proteins from the transfected MDA cells and media compared to the purified Ms02 GST-OppA protein. The proteins were seperated and transferred from an 8% SDS gel. Lane 1: Full-range Rainbow marker. Lane 2: purified GST-Ms02 OppA protein. Lane 3: 0.5 µg transfected cell^{24h}. Lane 4: 0.5 µg transfected cell medium^{24h}. Lane 5: 2.0 µg transfected cell^{24h}. Lane 6: 2.0 µg transfected cell medium^{24h}. The Ms02 GST-OppA protein is indicated on the SDS gel with an arrow.

5.2.3.4.2. Western blot with the anti-OppA antibody

An SDS gel loaded with the samples from 24 h as well as 48 h indicated that the amount of proteins (intensity of the bands on the SDS gel) were more after 48 h (results not shown). Therefore, it was decided to test only the 48 h cell harvest for protein expression.

The dot blot seemed to indicate the presence of OppA protein in the media. Direct comparisons between the intensity of the blots according to their concentrations of transfected vector cannot be made, due to the varying amount of medium that was dotted onto the membrane. The experiment was done as a preliminary and quick method of detecting the presence of proteins. No reactions were detected in the dot blot with the lysed cell samples (results not shown).

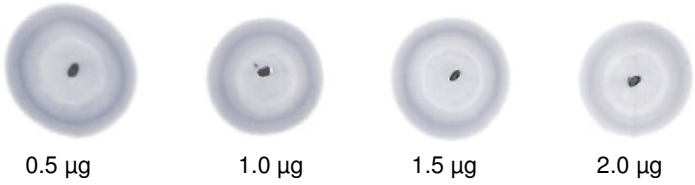

Figure 5.5 Western dot blot analysis on the media of the transfected cells harvested at 48 h. The concentrations of the pCI-neo_OppA vector that were used in the transfection of the MDA cells are indicated below its respective dot.

In Figure 5.6, the image of the western blot after SDS-PAGE separation can be seen. Only the purified Ms02 GST-OppA protein reacted with the anti-OppA rabbit antibodies (lane 2). However, bands were visible in the medium samples at a size just larger than 52 kDa, indicated by the arrows. This is thought to be due to non-specific binding, which would be explain by the western dot blot analysis. The sizes of the bands detected with the western blot corresponds to the size of the major protein present in the media samples, clearly visible in

both the SDS analysis and the Ponceau S stain of the separated proteins. One explanation could be the presence of albumin in the media, one of the major components of sera, although albumin generally occurs as a protein of an approximate 69 kDa size. Thus, the observation was that the amount of recombinant protein present in the samples may be too low to be detected with the colorimetric western blot detection method, due to the high amount of other interfering compounds.

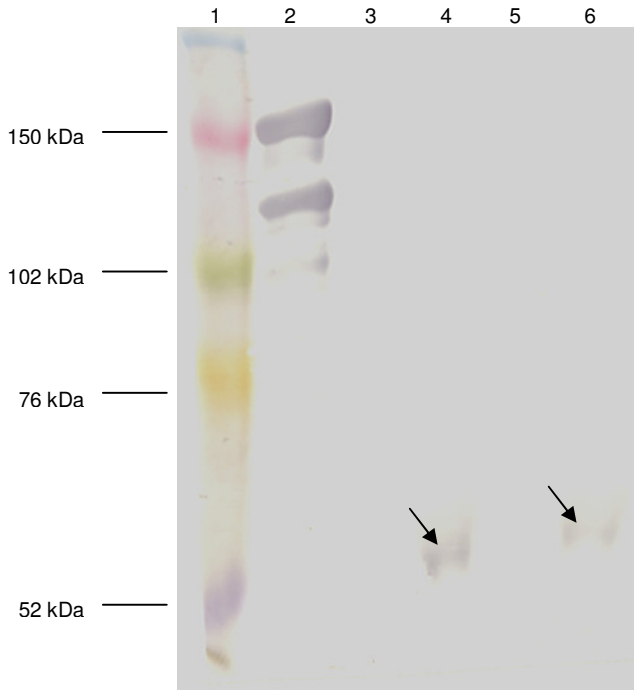

Figure 5.6 Image of the western blot showing the protein expression profile of the transfected cells and their media. Proteins were separated on an 8% denaturing SDS gel and transferred to a nitrocellulose membrane. Lane 1: Full-range Rainbow marker. Lane 2: purified Ms02 GST-OppA protein. Lane 3: 0.5 µg transfected cell^{48h}. Lane 4: 0.5 µg transfected cell medium^{48h}. Lane 5: 2.0 µg transfected cell^{48h}. Lane 6: 2.0 µg transfected cell medium^{48h}.

5.2.4. DISCUSSION

In this chapter, the expression of the Ms02 OppA protein using the pCI-neo_OppA vector was tested in an MDA cell line by transfecting the cells with the vector. First, a large amount of pCI-neo_OppA vector of the supercoiled topology was produced to use in the transfection. The supercoiled form is preferred, since literature demonstrated an inverse relationship of transfection efficiency to plasmid size, possibly due to increased mobility as a result of reduced dynamic size (Hsu and Uludag, 2008; Luke *et al.*, 2011; Mayrhofer *et al.*, 2009; Weintraub *et al.*, 1986).

The media samples separated on an SDS gel gave the impression of expressed OppA protein: a light protein band (inbetween the 150 kDa and the 102 kDa marker proteins) were visible that corresponds to a protein band of the same size in the Ms02 GST-OppA protein sample, namely the smallest of the two bands, which represents the OppA protein without the GST tag. The same could be seen on the membrane stained with

Ponceau S, although the bands were less visible. No bands of the same size were detected in the cell samples, but the cell samples contained significantly lower numbers and amount of proteins than the media samples. However, the western blot performed on the cells and media after separation in an SDS-PAGE did not result in any detectable bands, except for detection of a band of just larger than 52 kDa in size. It is thought that this was what accounted for the dot blot protein detection, and that the binding activity detected was only due to non-specific binding. From the start it was clear that the amount of protein in the lysed cell samples was too small to allow definite conclusions based on the SDS analysis. Due to the limitations of an SDS gel, a larger sample volume cannot be loaded and the recombinant protein may not be visible, and overloading the gel has negative consequences. This may be corrected for by using later harvesting times to enrich the transfection samples with OppA protein. Dialysis could also be considered, both for the protein samples and the media samples, to ensure a sample that is less contaminated with salts and small, unwanted proteins.

The western blot analysis of the proteins was performed on a colorimetric basis, a technique that is low in sensitivity. Other western blot techniques, utilizing chemiluminescence, may provide a better option for the detection of the OppA protein, due to the increased sensitivity of the technique. Preliminary optimisations with the chemiluminescent technique indicated a low signal-to-noise ratio, something that was improved when the blocking buffer was changed to contain a different form of casein, as well as a higher concentration of casein. However, there was still too high background to allow any definite conclusions. This method could be explored further, due to its numerous advantages.

The eubacterial secretion signal that is responsible for directing proteins across the cell membrane, is also present in mycoplasma proteins (Henrich *et al.*, 1999). Studies have shown that prokaryotic signal peptides are active in eukaryotic cells, directing the protein towards the medium in transfection studies (Hall *et al.*, 1990; Williamson *et al.*, 1994). The signal peptide sequence of the Ms02 *oppA* gene was not removed prior to cloning into the DNA vaccine vectors, since studies showed that genes containing a signal peptide are more immunogenic than the same genes without a signal peptide, irrespective of whether it is a eukaryotic or prokaryotic signal peptide (Baldwin *et al.*, 1999; Drew *et al.*, 2000; Haddad *et al.*, 1998; Kammath *et al.*, 1999*).* This could indicate that the expressed OppA protein in a cell culture system may be directed out of the cell and into the medium. However, the prokaryotic signal may not have been functional in these mammalian cells. The two other DNA vaccines, VR1012_OppA and VR1020_OppA also need to be tested. The VR1020 vector contains a eukaryotic signal peptide, and should eliminate doubt regarding pCI-neo's prokaryotic signal peptide efficiency. The Vical vectors are also slightly smaller than the pCI-neo vector (both approximately 500 bp), and could transfect more easily.

Different cells may also be transfected. Although no ostrich cell cultures are available yet to transfect, an acceptable choice would be to use bird cells, such as chicken cells. The KCEL cell line was established from chicken embryonic lung, which might have been a better choice (Kadoi, 2010). Due to import constraints throughout this study's time period, it was not possible to obtain chicken cells, and the transfection was therefore carried out in human cells.

Another factor to consider is stable transfection, where the antibiotic resistance genes of the vaccine vectors can be employed to ensure optimal transfection. In transient transfection, employed in this study, the efficacy

of transfection is one of the limiting factors, since transfection efficiency is dependant on the number of cells that are taken up and express the protein. Transfection efficiency should also be determined, for instance by using reporter genes.

5.2.5. CONCLUSION

In this study an attempt was made to prove expression of the OppA protein as a result of transfection with pCI-neo_OppA DNA vaccine vector in a eukaryotic cell system. The failure to detect any expressed protein could be due to a lack of expression, because of a nonfunctional signal sequence, or the levels of expression are too low for detection with the presently used detection system. However, there are still numerous optimisation strategies that could be explored: different cell lines need to be tested, later harvesting times could be necessary, and a different western blot detection technique, such as chemiluminescence, should be investigated as well.

Chapter 6

6.1. CONCLUSIONS AND FUTURE PERSPECTIVES

Most of the Ms02 genome was previously sequenced and preliminary annotated. The assembly of the generated sequences resulted in 28 large contiguous sequences, which were reduced to 14 in this study. A PCR technique developed by Liu and Whittier (1995) called **T**hermal **A**symmetric **I**nterlaced **P**olymerase **C**hain **R**eaction (TAIL-PCR) was used to accomplish the partial linkage of the large Ms02 contigs. Knowing the gene content as determined by the previous preliminary annotations, predictions were also made as to the orientation and linkages of the 14 contigs that were produced. This was done in order to attempt to locate the origin of replication and to establish if the TAIL-PCR linkages were correct, by subjecting the linked Ms02 genome sequence to open reading frame (ORF) analysis. ORFs that were located across two linked contigs, or on the additional sequences that were produced by sequencing the TAIL-PCR products, were further investigated and used in a BLASTx search to determine their gene annotations. This resulted in three predictions of linkages between contigs that were not linked using the TAIL-PCR technique. This warrants further investigation, as well as completion of the rest of the genome sequence. One observation during the assembly of the genome was the extent of repeats present, a common feature of mycoplasmas. Thus the genome was also subjected to repetitive sequence analysis, from large repeats to microsatellites to mononucleotide repeats. Further investigations are required to obtain a true representation of the repetitive nature of the Ms02 genome, but thus far it has been shown to affect the assembly of the contigs. Therefore, it is also suggested to repeat the sequencing of the Ms02 genome with paired-end technology, which will allow mapping of the resultant sequences, and so one would be able to orientate the contigs to their positions relative to one another.

This study produced three DNA vaccine vectors by cloning the Ms02 *oppA* gene into pCI-neo, VR1012, and VR1020, frequently used DNA vaccine vectors, for future vaccine efficacy trials. In order to obtain purified OppA protein, the gene was also cloned into the prokaryotic expression vector pGEX-4T-1. The protein was expressed with a GST-tag to allow for affinity purification. After purified fusion protein was obtained, it was used to immunise a rabbit in order to obtain anti-OppA antibodies. These antibodies can be used for the detection of OppA in other studies.

The pCI-neo vector was used in a preliminary study to detect expression of the OppA protein in a cell culture system. MDA cells were transfected with the pCI-neo_OppA vector and the transfected cells were harvested at 24 h and 48 h. Due to the presence of the prokaryotic secretion signal in the Ms02 *oppA* gene, the possibility existed that the protein could be exported out of the cell and would therefore reside in the medium. However, expression of the protein could not be shown. Using a colorimetric western blot on the cell samples did not result in any detectable reaction between the cell sample and the anti-OppA rabbit antibodies produced in this study. It is concluded that the transfection studies should be repeated in different cell types, as well as repeated with the other two vaccine vectors. Different detection methods of the transfected expressed protein should also be investigated, such as the use of chemiluminescent western blotting.

Future studies should also include ostrich vaccine trials, since an interesting observation by Fynan *et al.* (1993) indicates that efficiency of transfection does not necessarily determine the efficiency of vaccination. Ideally, one would like to develop a single vaccine for all three of the ostrich mycoplasmas; however, the cross-reactivity thus far is too low to use the vaccine developed for one of the ostrich mycoplasmas to protect against the other two as well. Options to overcome this include identifying other possible vaccine candidate genes and developing a multivalent (multi-component) vaccine to induce immunity simultaneously against all three mycoplasmas. Other possibilities include the use of a delivery system of the vaccine to the host. The ostrich mycoplasmas affect the upper respiratory tract of the ostriches; as such mucosal immunity might be optimal in preventing infection. Using carriers such as *Salmonella* may prove to be beneficial in developing quick, localised immunity. When successful vaccines have been developed, the concept of DNA-protein prime-boost regiments should be investigated. By vaccinating with a DNA vaccine and boosting with protein, the immune response elicited is longer and more pronounced than otherwise. The recombinant OppA protein produced in this study could therefore be of particular value for use as a booster protein, after priming with the DNA vaccines developed in this study.

Literature cited

Abdo, E.-M., Nicolet, J., Miserez, R., Gonçalves, R., Regalla, J., Griot, C., Bensaide, A., Krampe, M. and Frey, J. (1998). Humoral and bronchial immune responses in cattle experimentally infected with *Mycoplasma mycoides* subsp. *mycoides* small colony type. *Vet. Microbiol.*, **59**: 109-122.

Ada, G.L. (1994). Vaccines and immune response. In: Webster, R.G. and Granoff, A. (eds), *Encyclopedia of Virology (Vol. 3),* pp. 1503-1507. Academic Press, London.

Akbari, O., Panjwani, N., Garcia, S., Tascon, R., Lowrie, D. and Stockinger, B. (1999). DNA vaccination: Transfection and activation of dendritic cells as key events for immunity. *J. Exp. Med.*, **189**: 169-178.

Allwright, D.M., Burger, W.P,. Geyer, A. and Terblanche, A.W. (1995). Isolation of an influenza A virus from ostriches (*Struthio camelus*). *Avian Pathol.*, **22**: 59-65.

Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., Bucher, P., Cerutti, L., Corpet, F., Croning, M.D.R., Durbin, R., Falquet, L., Fleischmann, W., Gouzy, J., Hermjakob, H., Hulo, N., Jonassen, I., Kahn, D., Kanapin, A., Karavidopoulou, Y., Lopez, R., Marx, B., Mulder, N.J., Oinn, T.M., Pagni, M., Servant, F., Sigrist, J.A. and Zdobnov, E.M. (2000). InterPro – an integrated documentation resource for protein families, domains and functional sites. *Bioinformatics*, **16**: 1145-1150.

Armengaurd, J. (2009). A perfect genome annotation is within reach with the proteomics and genomics alliance. *Curr. Opin.Microbiol.,***12**: 292-300.

Babiuk, L.A. (2002). Vaccination: A management tool in veterinary medicine. *Vet. J.*, **164**: 188-201.

Babiuk, L.A., Babiuk, S.L., Loehr, B.I. and van Drunnen Littel-van den Hurk, S. (2000). Nucleic acid vaccines: Research tool or commercial reality. *Vet. Immunol. Immunopathol.*, **76**: 1–23.

Babiuk, L.A., Pontarollo, R., Babiuk, S.L., Loehr, B.I. and van Drunen Littel-van den Hurk, S. (2003). Induction of immune responses by DNA vaccines in large animals. *Vaccine,* **21**: 649-658.

Baldwin, S.L., D'Souza, C.D., Orme, I.M., Liu, M.A., Huygen, L., Denis, O., Tang, A., Zhu, L., Montgomery, D. and Ulmer, J.B. (1999). Immunogenicity and protective efficacy of DNA vaccines encoding secreted and non-secreted forms of *Mycobacterium tuberculosis* Ag85A. *Tuber. Lung Dis.*, **79**: 251-259.

Barker, J.J. (2006). Antibacterial drug discovery and structure-based design. *Drug Discov. Today*, **11**: 391-404.

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

Baseman, J.B. (1993). The cytadhesins of *Mycoplasma pneumoniae* and *M. genitalium*. *Subcell. Biochem.,* **20**: 243-259.

Baseman, J.B., Reddy, S.P. and Dallo, S.F. (1996). Interplay between mycoplasma surface proteins, airway cells, and the protean manifestations of mycoplasma-mediated human infections. *Am. J. Respir. Crit. Care Med.*, **154**: 137-144.

Baseman, J.B. and Tully, J.G. (1997). Mycoplasmas: Sophisticated, reemerging, and burdened by their notoriety. *Emerg. Infect. Dis.,* **3**: 21-32.

Bébéar, C.M., Renaudin, H., Charron, A., Bové, J.M., Bébéar, C. and Renaudin, J. (1998). Alterations in DNA topoisomerase IV and DNA gyrase in quinolone-resistant mutants of *Mycoplasma hominis* obtained *in vitro*. *Antimicrob. Agents Chemother.*, **42**: 2304-2311.

Behrens, A., Poumarat, F., Le Grand, D., Heller, M. and Rosengarten, R. (1996). A newly identified immunodominant membrane protein (pMB67) involved in *Mycoplasma bovis* surface antigenic variation. *Microbiol.*, **142**: 2463-2470.

Bellstedt, D.U., Human, P.A., Rowland, G.F. and Van der Merwe, K.J. (1987). Acid-treated naked bacteria as immune carriers for protein antigens. *J. Immunol. Methods*, **98**: 249-255.

Bencina, D. (2002). Haemaglutinins of pathogenic avian mycoplasmas. *Avian Pathol.*, **6**: 535-547.

Bennet, R.M., Gabor, G.T. and Merritt, M.M. (1985). DNA binding to human leukocytes. Evidence for a receptor-mediated association, internalization, and degradation of DNA. *J. Clin. Invest.,***76**: 2182-2190.

Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. *Nucl. Acid Res.*, **27**: 573-580.

Bergman, P.J., Camps-Palau, M.A., McKnight, J.A., Leibman, N.F., Craft, D.M., Leung, C., Liao, J., Riviere, I., Sadelain, M., Hohenhaus, A.E., Gregor, P., Houghton, A.N., Perales, M.A. and Wolchok, J.D. (2006). Development of a xenogeneic DNA vaccine program for canine malignant melanoma at the Animal Medical Center. *Vaccine,* **24**: 4852-4585.

Bergonier, D., de Simone, F., Russo, P., Solsona, M., Lambert, M. and Poumarat, F. (1996). Variable expression and geographic distribution of *Mycoplasma agalactiae* surface epitopes demonstrated with monoclonal antibodies. *FEMS Microbiol. Lett.*, **143**: 159-165.

Bernstein, P. and Ross, J. (1989). Poly(A), poly(A) binding protein and the regulation of mRNA stability. *Trends in Biochem. Sci.*, **14**: 373-377.

Bertram, J. (2006). MATra – Magnet Assisted Transfection: Combining nanotechnology and magnetic forces to improve intracellular dellivery of nucleic acids. *Curr. Pharm. Biotechnol.*, **7**: 277-285.

Binnewies, T.T., Motro, Y., Hallin, P.F., Lund, O., Dunn, D., La, T., Hampson, D.J., Bellgard, M., Wassenaar, T.M. and Ussery, D.W. (2006). Ten years of bacterial genome sequencing: Comparative-genomics-based discoveries. *Funct. Integr. Genomics*, **6**: 165-185.

Bohlin, J., Snipen, L., Hardy, S.P., Kristoffersen, A.B., Lagesen, K., Dønsvik, T., Skjerve, E. and Ussery. D.W. (2010). Analysis of intra-genomic GC content homogeneity within prokaryotes. *BMC Genomics*, **11**: 464-472.

Borovsky, Z., Tarshis, M., Zhang, P. and Rottem, S. (1998). Protein kinase C activation and vacuolation in HeLa cells invaded by *Mycoplasma penetrans*. *J. Med. Microbiol.*, **47**: 915-922.

Bos, J.D. (1997). The skin as an organ of immunity. *Clin. Exp. Immunol.*, **107**: 3-5.

Boshart, M., Weber, F., Jahn, G., Dorsch-Häsler, K., Flekcenstein, B., Schaffner, W. (1985). A very strong promoter is located upstream of an immediate early gene of Human Cytomegalovirus. *Cell*, **41**: 521-530.

Botes, A., Peyrot, B.M., Olivier, A.J., Burger, W.P. and Bellstedt, D.U. (2005a). Identification of three novel mycoplasma species from ostriches in South Africa. *Vet. Microbiol.,* **111**: 159-169.

Botes, A., Peyrot, B.M., Olivier, A.J., Burger, W.P. and Bellstedt, D.U. (2005b). Investigations into mycoplasma infections in South African ostriches. In: Proceedings of the 3rd International Ratite Science Symposium, 14-16 October, Madrid, Spain.

Bothwell, A.L., Paskind, M., Reth, M., Imanishi-Kari, T., Rajewsky, K. and Baltimore, D. (1981). Heavy chain variable region contribution to the NPb family of antibodies: Somatic mutation evident in a gamma 2a variable region. *Cell*, **24**: 625-637.

Boussif, O., Lezoualc'h, F., Zanta, M.A., Mergny, M.D., Scherman, D., Demeneix, J.P. and Behr, A. (1995). A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine. *Proc. Natl. Acad, Sci. USA*, **92**: 7297-7301.

Bové, J.M. (1993). Molecular features of Mollicutes. *Clin. Infect. Dis.,* **17**: S10-31.

Boyle, J.S., Agius, C.T., Lew, A.M. and Barr, I.G. (1998). DNA vaccines: Will they deliver? *Aus. Biotech,*. **8**: 153-159.

Bradbury, J.M. (2005). Poultry mycoplasmas: Sophisticated pathogens in simple guise. *Br. Poult. Sci.,* **46**:125-136.

Brandt, S. (2012). Preliminary investigations into the use of DNA vaccines to elicit protective immune responses against the ostrich mycoplasma Ms01. MSc dissertation, Stellenbosch University, SA.

Brenner, C., Wróblewski, H., Le Hanaff, M., Montagnier, L. and Blanchard, A. (1997). Spiralin, a mycoplasma membrane lipoprotein, induces T-cell-independent B-cell blastogenesis and secretion of proinflammatory cytokines. *Infect. Immun.*, **65**: 4322-4329.

Brewer, B.J. and Fangman, W.L. (1988). A replication fork barrier at the 3' end of yeast ribosomal RNA genes. *Cell*, **55**: 637-643.

Briggs, G.S., Smits, W.P. and Soultanas, P. (2012). Chromosomal replication initiation machinery of low G+C-content *Firmicutes. J. Bacteriol.*, **194**: 5162-5170.

Brinkmann, U., Mattes, R.E. and Buckel, P. (1989). High-level expression of recombinant genes in *Escherichia coli* is dependent on the availability of the *dnaY* gene product. *Gene,***85**: 109-114.

Canadian Ostrich Association. (2008). Cooking tips. Retrieved 17 Feb 2011.

Carson, J.L., Hu, P-C. and Collier, A.M. (1992). Cell structural and functional elements. In: Maniloff, J., McElhaney, E.D., Finch, L.R. and Baseman, J.B. (eds), *Mycoplasmas Molecular Biology and Pathogenesis,* pp. 63-72. American Society for Microbiology, Washington, DC.

Carta, G. and Jungbauer, A. (2010). Protein chromatography: Process Development and Scale-Up, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.

Casares, S., Inaba, K., Brumeanu, T.D., Steinman, R.M. and Bona, C.A. (1997). Antigen presentation by dendritic cells after immunization with DNA encoding a major histocompatibility complex class II-restricted viral epitope. *J. Exp. Med.*, **186**: 1481-1486.

Casewell, M., Friis, C., Marco, E., McMullin, P. and Phillips, I. (2003). The European ban on growth-promoting antibiotics and emerging consequences for human and animal health. *J. Antimicrob. Chemother.,***52**: 159-161.

Chambaud, I., Wróblewski, H. and Blanchard, A. (1999). Interactions between mycoplasma lipoproteins and the host immune system. *Trends Microbiol.*, **7**: 493-499.

Chaudhry, R., Nisar, N., Hora, B., Chirasani, S.R. and Malhotra, P. (2005). Expression and immunological characterization of the carboxy-terminal region of the P1 adhesin protein of *Mycoplasma pneumoniae*. *J. Clin. Microbiol.,* **43**: 321-325.

Chen, A.Y., Fry, S.R., Daggard, G.E. and Mukkur, T.K. (2008). Evaluation of immune response to recombinant potential protective antigens of *Mycoplasma hyopneumoniae* delivered as cocktail DNA and/or recombinant protein vaccines in mice. *Vaccine,* **26**: 4372-4378.

Chen, A.Y., Fry, S.R., Forbes-Faulkner, J., Daggard, G.E. and Mukkur, T.K. (2006). Comparative immunogenicity of M. hyopneumoniae NrdF encoded in different expression systems delivered orally via attenuated *S. typhimurium* aroA in mice. *Vet. Microbiol.,* **114**: 252-259.

Chen, D. and Texada, D.E. (2006). Low-usage codons and rare codons of *Escherichia coli*. *Gene Ther. Mol. Biol.,***10**: 1-12.

Chen, Y.L., Wang, S.N., Yang, W.J., Chen, Y.J., Lin, H.H. and Shiuan, D. (2003). Expression and immunogenicity of *Mycoplasma hyopneumoniae* heat shock protein antigen P42 by DNA vaccination. *Infect. Immun,.* **71**: 1155-1160.

Clifford, E.E., Martin, K.A., Dalal, P., Thomas, R. and Dubyak, G.R. (1997). Stage-specific expression of P2Y receptors, ecto-apyrase, and ecto-5'-nucleotidase in myeloid leukocytes. *Am. J. Physiol. Cell Physiol.*, **273**: 973-987.

Cole, B.C., Knudtson, K.L., Oliphant, A., Sawitzke, A.D., Pole, A., Manohar, M., Benson, L.S., Ahmen, E. and Atkin, C.L. (1996). The sequence of the *Mycoplasma arthritidis* superantigen, MAM: Identification of functional domains and comparison with microbial superantigens and plant lectin mitogens. *JEM*, **183**: 1105-1110.

Collop, N. (2011). Analysis of the cross-reactivity of antibodies produced against an OppA protein of *Mycoplasma nasistruthionis* sp. nov. (Ms03). Honnours report, Stellenbosch University, SA.

Condon, C., Watkins, S.C., Celluzi, C.M., Thompson, K. and Falo, L.D. (1996). DNA-based immunization by *in vivo* transfection of dendritic cells. *Nature Med.*, **2**: 1122-1128.

Corr, M., Lee, D.J., Carson, D.A. and Tighe, H. (1996). Gene vaccination with naked plasmid DNA: Mechanism of CTL priming. *J. Exp. Med.*, **184**: 1555-1560.

Cox, G.J.M., Zamb, T.J. and Babiuk, L.A. (1993). Bovine herpesvirus-1: Immune responses in mice and cattle infected with plasmid DNA. *J. Virol.*, **67**: 5664-5667.

DeBey, M.C. and Ross, R.F. (1994). Ciliostasis and loss of cilia induced by *Mycoplasma hyopneumoniae* in porcine tracheal organ cultures. *Infect. Immunol.*, **62**: 5312–5318.

Deitsch, K.W., Moxon, E.R. and Wellems, T.E. (1997). Shared themes of antigenic variation and virulence in bacterial, protozoal, and fungal infections. *Microbiol. Mol. Biol. Rev.*, **61**: 281-283.

Delcher, A.L., Harmon, D., Kasif, S., White, O. and Salzberg, S.L. (1999). Improved microbial gene identification with GLIMMER. *Nucl. Acids Res.*, **27**: 4636-4641.

Demina, I.A., Serebryakova, M.V., Ladygina, V.G., Galyamina, M.A., Zhukova, N.A., Alekseev, D.G., Fisunov, G.Y. and Govorun, V.M. (2011). Comparative proteomic characteristic of Mycoplasmas (Mollicutes). *Russ. J. Biochem.*, **37**: 61-70.

Denison, A.M., Clapper, B. and Dybvig, K. (2005). Avoidance of the host immune system through phase variation in *Mycoplasma pulmonis. Infect. Immun.*, **73**: 2033-2039.

Dervyn, E., Suski, C., Bruand, C., Chapuis, J., Errington, J., Jannière, L. and Ehrlich, S.D. (2001). Two essential DNA polymerases at the bacterial replication fork. *SCI*, **294**: 1716-1719.

Dhama, K., Mahendran, M., Gupta, P.K. and Rai, A. (2008). DNA vaccines: Current perspectives. *Vet. Res. Commun.*, **32**: 341-356.

Diacovich, I. and Gorvel, J.-P. (2010). Bacterial manipulation of innate immunity to promote infection. *Nature Rev. Microbiol.,***8**:117-128.

Die Burger. (6-9-2012). Volstruisbedryf steier. *Eugene Gunning*.

Dimitrov, D.S., Franzoso, G., Salman, M., Blumenthal, R., Tarshis, M., Barile, M.F. and Rottem, S. (1993). *Mycoplasma fermentans* (Incognitus Strain) cells are able to fuse with T Lymphocytes. *Clin. Infect. Dis.*, **17**: S305-S308.

Doeven, M.K., Van den Bogaart, G., Krasnikov, V. and Poolman, B. (2008). Probing receptor-translocator interactions in the oligopeptide ABC transporter by Fluorescence Correlation Spectroscopy. *Biophys. J.*, **94**: 3956–3965.

Donnelly, J.J., Ulmer, J.B., Shiver, J.W. and Liu, M.A. (1997). DNA vaccines. *Annu. Rev. Immunol.,***15**: 647-48.

Drew, D.R., Lightowlers, M. and Strugnell, R.A. (2000). Humoral immune response to DNA vaccines expressing secreted membrane bound and non-secreted forms of the *Taenia ovis* 45W antigen. *Vaccine*, **18**: 2522-2532.

Droual, R., Bickford, A.A., Charlton, B.R. and Kuney, D.R. (1990). Investigation of problems associated with intramuscular breast injection of oil-adjuvanted killed vaccines in chickens. *Avian Dis.*, **34**: 473-478.

Du, A. and Wang, S. (2005). Efficacy of a DNA vaccine delivered in attenuated *Salmonella typhimurium* against *Eimeria tenella* infection in chickens. *Int. J. Parasitol.*, **35**: 777-785.

Duensing, T.D., Wing, J.S., van Putten, J.P. (1999). Sulfated polysaccharide-directed recruitment of mammalian host proteins: a novel strategy in microbial pathogenesis. *Infect. Immunol.*, **67**: 4463–4468.

Dufour, V. (2001). DNA vaccines: New applications for veterinary medicine. *Vet. Sci. Tomorrow*, **2**:1-19.

Dybvig, K. and Voelker, L.L. (1996). Molecular biology of Mycoplasmas. *Annu. Rev. Microbiol.,* **50**: 25-57.

Dziewanowska, K., Patti, J.M., Deobald, C.F., Bayles, K.W., Trumble, W.R. and Bohach, G.A. (1999). Fibronectin binding protein and host cell tyrosine kinase are required for internalization of *Staphylococcus aureus* by epithelial cells. *Infect. Immun.*, **67**: 4873-4878.

Edward, D.G. and Fitzgerald, W.A. (1951). Cholesterol in the growth of organisms of the pleuropneumonia group. *J. Gen. Microbiol.,* **5**: 576-586.

Ellis, R.W. (1999). New technologies for making vaccines. *Vaccine*, **17**: 1596-1604.

Ellison, J.S., Olson, L.D. and Barile, M.F. (1992). Immunity and vaccine development. In: Maniloff, J., McElhaney, E.D., Finch, L.R. and Baseman, J.B. (eds), *Mycoplasmas Molecular Biology and Pathogenesis,* pp. 491-504. American Society for Microbiology, Washington, DC.

Erickson, B.A., Ross, R.F. and Bové, J.M. (1988). Isolation of *Mycoplasma salivarium* from swine. *Vet. Microbiol.*, **16**: 385-390.

Escriou, C., Ciolina, C., Helbling-Leclerc, A., Wils, P. and Scherman, D. (1998). Cationic lipid-mediated gene transfer: Analysis of cellular uptake and nuclear import of plasmid DNA. *Cell Biol. Toxicol.*, **14**: 95-104.

Evans, J.D., Leigh, S.A., Branton, S.L., Collier, S.D., Pharr, G.T. and Bearson, S.M.D. (2005). *Mycoplasma gallisepticum*: Current and developing means to control the avian pathogen. *J. Appl. Poultry Res.,* **14**: 757-763.

Fadiel, A., Lithwick, S. and Naftolin, F. (2005). The influence of environmental adaptation on bacterial genome structure. *Lett. Appl. Microbiol.,* **40**: 12-18.

Fagan, P.K., Walker, M.J., Chin, J., Eamens, G.J. and Djordjevic, S.P. (2001). Oral immunization of swine with attenuated *Salmonella typhimurium* aroA SL3261 expressing a recombinant antigen of *Mycoplasma hyopneumoniae* (NrdF) primes the immune system for a NrdF specific secretory IgA response in the lungs. *Microb. Pathog.,* **30**: 101-110.

Fairweather, N.F., Chatfield, S.N., Makoff, A.J., Strugnell, R.A., Bester, J., Maskell, D.J. and Dougan, G. (1990). Oral vaccination of mice against tetanus by use of a live attenuated Salmonella carrier. *Infect. Immun.,* **58**: 1323-1326.

Fang, G., Rocha, E. and Danchin, A. (2005). How essential are nonessential genes? *Mol. Biol. Evol.*, **22**: 2147-2456.

Ferraro, B., Morrow, M.P., Hutnick, N.A., Shin, T.H., Lucke, C.E. and Weiner, D.B. (2011). Clinical applications of DNA vaccines: Current progress. *Clin. Infect. Dis.,***53**: 296-302.

Fisher, K.J., Gao, G.P., Weitzman, M.D., DeMatteo, R., Burda, J.F. and Wilson, J.M. (1996). Transudction with recombinant adeno-associated virus for gene therapy is limited by leading-strand synthesis. *J. Virol.*, **70**: 520-532.

Flitman-Tene, R., Mudahi-Orenstein, S., Levisohn, S. and Yogev, D. (2003). Variable lipoprotein genes of *Mycoplasma agalactiae* are activated *in vivo* by promoter addition via site-specific DNA inversions. *Infect. Immun.*, **71**: 3821-3830.

Fournier, B., Aras, R. and Hooper, D.C. (2000). Expression of the multidrug resistance transporter NorA from *Staphylococcus aureus* is modified by a two-component regulatory system. *J. Bacteriol.*, **182**: 664-671.

Fraley, R., Subramani, P., Berg, P. and Papahadjopoulos, D. (1980). Introduction of liposome-encapsulatetd SV40 DNA into cells. *J. Biol. Chem.*, **255**: 10431-10435.

Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., Fritchman, R.D., Weidman, J.F., Small, K.V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T.R., Saudek, D.M., Phillips, C.A., Merrick, J.M., Tomb, J.F., Dougherty, B.A., Bott, K.F., Hu, P.C., Lucier, T.S., Peterson, S.N., Smith, H.O., Hutchison, C.A. and Venter, J.C. (1995). The minimal gene complement of *Mycoplasma genitalium*. *SCI,* **270**: 397-403.

Freshney, R.I. (2005). Culture of Animal Cells (6th Ed). Wiley-Blackwell.

Freundt, E.A. (1973). Principles of Mycoplasma classification. In: Maramorosch, K. (ed), *Mycoplasma and Mycoplasma-like agents of human, animal and plant diseases,* pp. 7-13. N.Y. Acad. Sci., New York, USA.

Fu, L.N., Ye, R.Q., Browder, L.W. and Johnston, R.N. (1991). Translational potentiation of messanger RNA with secondary structure in Xenopus. *SCI*, **251**: 807-810.

Fynan, E.F., Robinson, H.L. and Webster, R.G. (1993). Use of DNA encoding influenza hemagglutinin as an avian influenza vaccine. *DNA Cell Biol.*, **12**: 785-789.

Galvin, T.A., Muller, J. and Khan, A.S. (2000). Effect of different promoters on immune responses elicited by HIV-1 *gagl env* mulitgenic DNA vaccine in *Macaca mulatta* and *Macaca nemestrina*. *Vaccine*, **18**: 2566-2583.

Gao, F. and Zhang, C.-T. (2008). Ori-Finder: A web-based system for finding *oriC*s in unannotated bacterial genomes. *BMC Bioinform.*, **9**: 79.

Gao, J. and Chen, L.-L. (2010). Theoretical methods for identifying important functional genes in bacterial genomes. *Res. Microbiol.*, **161**: 1-8.

Garmory, H.S., Brown, K.A. and Titball, R.W. (2003). DNA vaccines: Improving expression of antigens. *Genet. Vaccines Ther.,* **1**: 1-5.

Garmory, H.S. and Titball, R.W. (2004). ATP-binding cassette transporters are targets for the development of antibacterial vaccines and therapies. *Infect. Immun.,* **72**: 6757-6763.

Gerdes, S.Y., Scholle, M.D., Campbell, J.W., Balázsi, G., Ravasz, E., Daugherty, M.D., Somera, A.L., Kyprides,N.C., Anderson, I., Gelfans, M.D., Bhattachyara, A., Kapatral, V., D'Souza, M., Baev, M.V., Grechkin, Y., Mseeh, F., Fonstein, M.Y., Overbeek, R., Barabási, A.-L., Oltvai, Z.N. and Osterman, A.L. (2003). Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J. Bacteriol.*, **185**: 5673-5684.

Giron, J.A., Lange, M. and Baseman, J.B. (1996). Adherence, fibronectin binding, and induction of cytoskeleton reorganization in cultured human cells by *Mycoplasma penetrans*. *Infect. Immun.*, **64**: 197-208.

Glass, J.I., Lefkowitz, E.J., Glass, J.S., Heiner, C.R., Ellson, Y.C. and Cassell, G.H. (2000). The complete sequence of the mucosal pathogen *Ureaplasma urealyticum*. *Nature*, **407**: 757-762.

Glew, M.D., Marenda, M., Rosengarten, R. and Citti, C. (2002). Surface diversity in *Mycoplasma agalactiae* is driven by site-specific DNA inversions within the *vpma* multigene locus. *J. Bacteriol.*, **184**: 5987-5998.

Graham, F.L., Smiley, J., Russell, W.C. and Nairn, R. (1977). Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J. Gen. Virol.*, **36**: 59-72.

Grau, O., Slizewicz, B., Tuppin, P., Launay, V., Bourgeois, E., Sagot, N,. Moynier, M., Lafeuillade, A., Bachelez, H., Clauvel, J-P., Blanchard, A., Bahraoui, E. and Montagnier, L. (1995). Association of *Mycoplasma penetrans* with Human Immunodeficiency Virus infection. *J. Infect. Dis.*, **172**: 672-681.

Groisman, E.A. and Casadesus, J. (2005). The origin and evolution of human pathogens. *Mol. Microbiol.*, **56**: 1-7.

Guan, K., and Dixon, J.E. (1991). Eukaryotic proteins expressed in *Escherichia coli*: animproved thrombin cleavage and purification procedure of fusion proteins withglutathione S-transferase. *Anal. Biochem.*, **192:** 262-267.

Guo, F.B., Ou, H.Y. and Zhang, C.-T. (2003). ZCURVE: A new system for recognizing protein coding genes in bacterial and archael genomes. *Nucl. Acids. Res.*, **31**: 1780-1789.

Guo, X. and Mrázek, J. (2008). Long simple sequence repeats in host-adapted pathogens localize near genes encoding antigens, housekeeping genes, and pseudogenes. *J. Mol. Evol.*, **67**: 497-509.

Gupta, M,. Chyi, Y.-S., Romero-Servenson, J. and Owen, J.L. (1994). Amplification of DNA markers from evolutionary diverse genomes using single primers of simple-sequence repeats. *Theor. Appl. Genet.*, **89**: 998-1006.

Gur-Arie, R., Cohen, C.J., Eitan, Y., Shelef, L., Hallerman, E.M. and Kashi, Y. (2000). Simple sequence repeats in *Escherichia coli*: Abundance, distribution, composition, and polymorphism. *Genome Res.*, **10**: 62-71.

Gurunathan, S., Klinman, D. and Seder, R.A. (2000). DNA Vaccines: Immunology, Application and Optimization. *Annu. Rev. Immunol.,* **18**: 927-974.

Gurvich, O.L., Baranov, P.V., Gesteland, R.F and Atkins, J.F. (2005). Expression levels influence ribosomal frameshifting at the tandem rare arginine codons AGG_AGG and AGA_AGA in *Escherichia coli*. *J. Bacteriol.*, **187**: 4023-4032.

Haddad, D., Liljeqvist, S., Stahl, S., Perlmann, P., Berzins, K. and Ahlborg, N. (1998). Differential induction of immunoglobin G subclasses by immunization with DNA vectors containing or lacking a signal sequence. *Immunol. Lett.*, **61**: 201-204.

Hall, J., Hazlewood, G.P., Surani, M.A., Hirst, B. and Gilbert, H.J. (1990). Eukaryotic and prokaryotic signal peptides direct secretion of a bacterial endoglucanase by mammalian cells. *J. Biol. Chemis.*, **265**: 19996-19999.

Harms, J.A. and Splitter, G.A. (1995). Interferon-γ inhibits transgene expression driven by SV40 or CMV promoters but augments expression driven by the mammalian MHC I promoter. *Human Gene Ther.*, **6**: 1291-1297.

Hempstead, P.G. (1990). An improved method for the rapid isolation of chromosomal DNA from Mycoplasma spp. *Can. J. Microbiol.*, **36**: 59-61.

Henderson, B., Poole, S. and Wilson, M. (1996). Bacterial modulins: A novel class of virulence factors which cause host tissue pathology by inducing cytokine synthesis. *Microbiol. Mol. Biol. Rev.*, **60**: 316-341.

Henrich, B., Feldman, R.-C. and Hadding, U. (1993). Cytoadhesins of *Mycoplasma hominis*. *Infect. Immun.*, **61**: 2945-2951.

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

Herrmann, R. (1992). Genome structure and organization. In: Maniloff, J., McElhaney, E.D., Finch, L.R. and Baseman, J.B. (eds), *Mycoplasmas Molecular Biology and Pathogenesis,* pp. 157-168. American Society for Microbiology, Washington, DC.

Hopfe, M., Dahlmanns, T. and Henrich, B. (2011). In *Mycoplasma hominis* the OppA-mediated cytoadhesion depends on its ATPase activity. *BMC Microbiol.*, **11**: 185-193.

Hopfe. M. and Henrich, B. (2004). OppA, the substrate-binding subunit of the oligopeptide permease, is the major ecto-ATPase of *Mycoplasma hominis*. *J. Bacteriol.,* **186**: 1021-1928.

Huang, M.T.F. and Gorman, C.M. (1990). Intervening sequences increase efficiency of RNA 3' processing and accumulation of cytoplasmic RNA. *Nucl. Acids. Res.*, **18**: 937-947.

Huchzermeyer, F.W. (1994). Transmissible diseases. In: Huchzermeyer, F.W.,*Ostrich diseases,* pp. 7-45. Agricultural Research Council, SA.

Huchzermeyer, F.W. (1998). The ostrich and other ratites. In: Huchzermeyer, F.W., *Diseases of Ostriches and other Ratites*. Agricultural Research Council, SA.

Hsu, C.Y. and Uludag, H. (2008). Effects of size and topology of DNA molecules on intracellular delivery with non-viral gene carriers. *BMC Biotechnol.*, 8: 23.

Invitrogen Corporation, (2005). Troubleshooting Transfection Experiments. Available from www.invitrogen.com.

Jacobsen, L.B., Calvin, S.A., Colvin, K.E. and Wright, M. (2004). FuGENE 6 transfection reagent: the gentle power. *Methods*, **33**: 104-112.

Jackson, R.J. and Standart, N. (1990). Do the poly(A) tail and 3' untranslated region control mRNA translation? *Cell*, **62**: 15-24.

Jakob, T., Walker, P.S., Krieg, A.M., Udey, M.C. and Vogel, J.C. (1998). Activation of cutaneous dendritic cells by CpG containing oligodeoxynucleotides: A role for dendritic cells in the augmentation of Th1 responses by immunostimulatory DNA 1. *J. Immunol.,***161**: 3042-3049.

Jan, G., Fontenelle, C., Le Hénaff, M. and Wróblewski, H. (1995). Acylation and immunological properties of *Mycoplasma gallisepticum* membrane proteins. *Res. Microbiol.*, **146**: 739-750.

Jensen, L.T., Ladefoged, S., Birkelund, S. and Christiansen, G. (1995). Selection of *Mycoplasma hominis* PG21 deletion mutants by cultivation in the presence of monoclonal antibody 552. *Infect. Immun.*, **63**: 3336-3347.

Jensen, L.T., Thorsen, P., Møsller, B., Birkelund, S. and Christiansen, G. (1998). Antigenic and genomic homogeneity of successive *Mycoplasma hominis* isolates. *J. Med. Microbiol.*, **47**: 659-666.

Jeurissen, A.H.M., Janse, E.M., Koch, G. and De Boer, G.F (1989). Postnatal development of mucosa-associated lymphoid tissues in chickens. *Cell Tissue Res.*, **258**: 119-124.

Jewell, E., Robinson, A., Savage, D., Erwin, T., Love, C.G., Lim, G.A.C., Li, X., Batley, J., Spangenberg, G.C. and Edwards, D. (2006). SSRPrimer and SSR taxonomy tree: BiomeSSR discovery. *Nucl. Acids Res.*, **34**(suppl 2): W656-W659.

Jiang, H.X., Chen, J.R., Yan, H.L., Li, X.N., Chen, Z.L. and Zeng, Z.L. (2009). Molecular viability of DR-1 and DR-2 within the *pvpA* gene in *Mycoplasma gallisepticum* isolates. *Avian Dis.*, **53**: 124-128.

Jones, E. and Hartl, D.L. (1998). Genetics: Principles and analysis. Jones and Bartlett Publishers, Boston.

Jordan, F.T.W. (1979). Avian Mycoplasmas. In: Tully, J.G. and Whitcomb, R.F. (eds), *The Mycoplasmas (Vol 2), Human and animal mycoplasmas,* pp. 1-48. Academic Press Inc., New York.

Jordan, F.T.W. (1990). Avian Mycoplasmoses. In: Jordan, F.T.W. (ed), *Poultry Diseases (3rd Ed),* pp. 74-85. Bailliere Tindall, Great Britain.

Jordan, F.T.W. (1996). Avian Mycoplasmosis. In: Jordan, F.T.W (ed), *Poultry Diseases, (3rd Ed)*, pp: 418-429. Bailliere Tindal, Great Britain.

Jordan, I.K., Rogozin, I.B., Wolf, Y.I. and Koonin, E.V. (2002). Essential genes are more evolutionary conserved that nonessential genes in bacteria. *Genome Res.*, **12**: 962-968.

Jukes, T.H. and Osawa, S. (1993). Evolutionary changes in the genetic code. *Comp. Biochem. Physiol.*, **106B**: 489-494.

Kadoi, K. (2010). Establishment of two chicken embryonic cell lines in a newly developed nutrient medium. *New Microbiol.*, **33**: 393-397.

Kane, J.F. (1995). Effects of rare codon cluster on high-level expression of heterologous proteins in *Escherichia coli*. *Curr. Opin. Biotechnol.*, **6**: 494-500.

Karlin, S., Mrázek, J. and Gentles, A.J. (2003). Genome comparisons and analysis. *Curr. Opin. Struct. Biol.*, **13**: 344-352.

Kashi, Y. and King, D.G. (2006). Simple sequence repeats as advantageous mutators in evolution. *Trends Genet.*, **22**: 253-259.

Kashi, Y., King, D.G. and Soller, M. (1997). Simple sequence repeats as a source of quantitative genetic variation. *Trends Genet.*, **13**: 74-79.

Katti, M.V., Ranjekar, P.K. and Gupta, V.S. (2001). Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Mol. Biol. Evol.*, **18**: 1161-1167.

Képès, F., Jester, B.C., Lepage, T., Rafiei, N., Rosu, B. and Junier, I. (2012). The layout of a bacterial genome. *FEBS Lett.*, **586**: 2043-2048.

Kirley, T.L. (1997). Complementary DNA cloning and sequencing of the chicken muscle ecto-ATPase. *J. Biol. Chem.*, **272**: 1076-1081.

Klainer, A.S. and Pollack, J.D. (1973). Scanning electron microscopy techniques in the study of the surface structure of mycoplasmas. *Ann. NY Acad. Sci.*, **225**: 236-245.

Kleven, S.H. (1997). *Mycoplasma synoviae* infection. In: Calnek, B.W. (ed), *Diseases of Poultry (10th Ed),* pp. 220-228. The Iowa State University Press, Iowa, USA.

Kleven, S.H. (1998). Mycoplasmas in the etiology of multifactorial respiratory disease. *Poult. Sci.*, **77**: 1146-1149.

Kleven, S.H. (2008). Control of avian mycoplasma infections in commercial poultry. *Avian Dis.,* **52**: 367-374.

Klinman, D.M. (1998). Therapeutic applications of CpG-containing oligodeoxynucleotides. *Antisense Nucl. Acid Drug Dev.*, **8**: 181-184.

Kloog, Y. and Cox, A. D. (2004). Prenyl-binding domains. Potential targets for Ras inhibitors and anti-cancer drugs. *Semin. Cancer Biol.,* **14**: 253-261.

Knowles, A.F. (1995). The rat liver ectoATPase/C-CAM cDNA detects induction of carcinoembryonic antigen but not the mercurial-insensitive ectoATPase in human hepatoma Li-7A cells treated by epidermal growth factor and cholera toxin. *Biochem. Biophys. Res. Comm.*, **207**: 529-535.

Kodihalli, S., Haynes, J.R., Robinson, H.L. and Webster, R.G. (1997). Cross-protection among lethal H5N2 influenza viruses induced by DNA vaccine to the hemagglutinin. *J. Virol.*, **71**: 3391-3396.

Koonin, E.V. (2003). Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat. R. Microbiol.*, **1**: 127-136.

Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell,* **44**: 283-292.

Kozak, M. (1989). Circumstances and mechanisms of inhibition of translation by secondary structure in eukaryotic mRNAs. *Mol. Cell Biol.*, **9**: 5134-5142.

Krass, C.J. and Gardner, M.W. (1973). Etymology of the term Mycoplasma. *Int. J. Syst. Bacteriol.,* **23**: 62-64.

Krause, D.C. (1998). *Mycoplasma pneumoniae* cytadherence: Organization and assembly of the attachment organelle. *Trends Microbiol.,* **6**: 15-18.

Krause, D.C. and Balish, M.F. (2001). Structure, function, and assembly of the terminal organelle of *Mycoplasma pneumoniae*. *FEMS Microbiol. Lett.,* **198**: 1-7.

Kretzmer, G. (2002). Industrial processes with animal cells. *Appl. Microbiol. Biotechnol.*, **59**:135-141.

Krieg, A.M., Yi, A.-K., Matson, S., Waldschmidt, T.J., Bishop, G.A., Teasdale, R., Koretzky, G.A. and Klinman, D.M. (1994). CpG motifs in bacterial DNA trigger direct B-cell activation. *Nature*, **374**: 546-549.

Kutzler, M.A. and Weiner, D.B. (2008). DNA vaccines: Ready for prime-time? *Nature Rev. Genet.,***9**: 776-788.

Langer, S. (2009). *Mycoplasma struthionis* sp. nov. and *Mycoplasma nasistruthionis* sp. nov. from Ostrich (*Struthio camelus*) with respiratory tract disease. MSc dissertation, pp. 81-91. University of Veterinary medicine. Vienna, Austria.

Le Chatelier, E., Bécherel, O.J., d'Alençon, E., Canceill, D., Erlich, S.D., Fuchs, R.P.P. and Jannière, L. (2004). Involvement of DnaE, the second replicative DNA polymerase form *Bacillus subtilis*, in DNA mutagenesis. *J. Biol. Chem.*. **279**: 1757-1767.

Lechmann, M. and Liang, T.J. (2000). Vaccine development for Hepatitis C. *Sem. Liver Dis.,* **20**: 211-226.

Ledwith, B.J., Manam, S., Troilo, P.J., Barnum, A.B,. Pauley, C.J., Griffiths, T.G,. Harper, L.B., Beare, C.M., Bagdon, W.J. and Nichols, W.W. (2000). Plasmid DNA vaccines: Investigation of integration into host cellular DNA following intramuscular injection in mice. *Intervirology*, **43**: 258-272.

Leonard, A.C. and Grimwade, J.E. (2011). Regulation of DnaA assembly and activity: Taking directions from the genome. *Ann. Rev. Microbiol.*, **65**: 19-35.

Levisohn, S. and Kleven, S.H. (2000). Avian mycoplasmosis (*Mycoplasma gallisepticum*). *Rev. Sci. Tech. Off. Int. Epiz.,* **19**: 425-442.

Lewis, P.J. and Babiuk, L.A. (1999). DNA vaccines: A review. *Adv. Virus Res.*, **54**: 129-188.

Ley, D.H. and Yoder, H.W.(1997). *Mycoplasma gallisepticum* infection. In: Calnek, B.W. (ed), *Diseases of Poultry (10th Ed),* pp. 194-205. The Iowa State University Press, Iowa, USA.

Li, Y., Korol, A.B., Fahima, T., Beiles, A. and Nevo, E. (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: A review. *Mol. Ecol.*, **11**: 2453-2465.

Lin, Y. and Zhang, R.R. (2011). Putative essential and core-essential genes in Mycoplasma genomes. *Sci. Rep.*, **53**: 1-7.

Liobikas, J., Polianskyte, Z., Speer, O., Thompson, J., Alakoskela, J.-M., Peitsaro, N., Franck, M., Whitehead, M.A., Kinnunen, P.J.K. and Eriksson, O. (2006). Expression and purification of the mitochondrial serine protease LACTB as an N-terminal GST fusion protein in *Escherichia coli*. *Protein Express. Purif.*, **45**: 335-342.

Liu, Y. and Whittier, R.F. (1995). Thermal Asymmetric Interlaced PCR: Automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. *Genomics,* **25**: 674-681.

Liu, W.G. and Yao, K.D. (2002). Chitosan and its derivatives – a promising non-viral vector for gene transfection. *J. Control. Release*, **83**: 1-11.

Lobry, J.R. (1996a). Origin of replication of *Mycoplasma genitalium*. SCI, **272**: 745-746.

Lobry, J.R. (1996b). Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol. Biol. Evol.*, **13**: 660-665.

Lockaby, S.B., Hoerr, F.J., Lauerman, L.H. and Kleven, S.H. (1998). Pathogenicity of *Mycoplasma synoviae* in broiler chickens. *Vet. Pathol.,* **35**: 178-190.

Loveless, R.W. and Feizi, T. (1989). Sialo-oligosaccharide receptor for *Mycoplasma pneumoniae* and related oligosaccharides of poly-N-acetyllactosamine series are polarized at the cilia and apical-microvillar domains of the ciliated cells in human bronchial epithelium. *Infect. Immun.*, **57**: 1285-1289.

Lucier, T.S., Hu, P.Q., Peterson, S.N., Song, X.Y., Miller, L., Heitzman, K., Bott, K.F., Hutchison, C.A. and Hu, P.C. (1994). Construction of an ordered genomic library of *Mycoplasma genitalium*. *Gene,* **150**: 27-34.

Ludwig, W., Schleifer, K.-H. and Whitman, W.B. (2009). Revised road map to the phylum *Firmicutes*. *Bergey's Manual® of Systematic Bacteriology*, pp. 1-13. Williams & Wilkins, Baltimore, USA.

Luke, J.M., Vincent, J.M., Du, S.X., Gerdemann, U., Leen, A.M., Whalen, R.G., Hodgson, C.P. and Williams, J.A. (2011). Improved antibiotic-free plasmid vector design by incorporation of transient expression enhancers. *Gene Ther.*, **18**: 334-343.

Maasho, K., Marusina, A., Reynolds, N.M., Coligan, J.E. and Borrego, F. (2004). Efficient gene transfer into the human natural killer cell line, NKL, using the amaxa nucleofection system. *J. Immunol. Methods.*, **284**: 133-140.

Mackiewicz, P., Zakrzewska-Czerwińska, J., Zawilak, A., Dudek, M.D. and Cebrat, S. (2004). Where does bacterial replication start? Rules for predicting the *oriC* region. *Nucl. Acids. Res.*, **32**: 3781-3791.

Magee, T. and Marshall, C. (1999). New insights into the interaction of Ras with the plasma membrane. *Cell,***98:** 9-12.

Makrides, S.C. (1996). Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol. Rev.*, **60**: 512-538.

Margolis, B., Bellot, F., Honegger, A.M., Ullrich, A., Schlessinger, J. and Zilberstein, A. (1990). Tyrosine kinase activity is essential for the associated of phospholipase C-gamma with the epidermal growth factor receptor. *Mol. Cell. Biol.*, **10**: 435-441.

Marin, A. and Oliver, J.L. (2003). GC-biased mutation pressure and ORF lengthening. *J. Mol. Evol.,* **56**: 371-372.

Mayrhofer, P., Schleef, M. and Jechlinger, W. (2009). Use of minicircle plasmids for gene therapy. *Methods Mol. Biol.*, **542**: 87-104.

Mazars, G.R., Moyret, C,. Jeanteur, P. and Theillet, C.G. (1991). Direct sequencing by thermal asymmetric PCR. *Nucl. Acids. Res.*, **19**: 4783.

McAuliffe, L., Ayling, R.D. and Nicholas, R.A. (2007). Identification and characterization of variable number tandem-repeat markers for molecular epidemiological analysis of *Mycoplasma mycoides* subspecies *mycoides* SC. *FEMS Microbiol. Lett.*, **276**: 181-188.

McAuliffe, L., Churchward, C.P., Lawes, J.R., Loria, G., Ayling, R.D. and Nicholas, R.A. (2008). VNTR analysis reveals an unexpected genetic diversity within *Mycoplasma agalactiae*, the main causative agent of contagious agalactia. *BMC Microbiol.*, **8**: 193-199.

McLean, M.J., Wolfe, K.H. and Devine, K.M. (1998). Base composition skews, replication orientation, and gene orientation in 12 prokaryotic genomes. *J. Molec. Evol.*, **47**: 691-696.

McTaggart, S.J. (2006). Isoprenylated proteins. *Cell. Mol. Life Sci.*, **63**: 255-267.

Médigue, C. and Moszer, I. (2007). Annotation, comparison and databases for hundreds of bacterial genomes. *Res. Microbiol.*, **158**: 724-736.

Metzgar, D., Liu, L, Hansen, C., Dybvig, K. and Willis, C. (2002). Domain-level differences in microsatellite expansion in coding DNA. *Genome Res.*, **10**: 72-80.

Minion, F.C. (2002). Molecular pathogenesis of mycoplasma animal respiratory pathogens. *Front. Biosci.*, **7**:1410-1422.

Minion, F.C., Lefkowitz, E.J., Madsen, M.L., Cleary, B.J., Swartzell, S.M. and Mahairas, G.G. (2004). The genome sequence of *Mycoplasma hyopneumoniae* strain 232 the agent of swine mycoplasmosis. *J. Bacteriol.*, **186**: 7123-7133.

Miyata, M. and Seto, S. (1999). Cell reproduction cycle of mycoplasma. *Biochimie,* **81**: 873-878.

Mora, M., Veggi, D., Santini, L., Pizza, M. and Rappuoli, R. (2003). Reverse vaccinology. *DDT*, **8**: 459-464.

Moran, N.A. (2002). Microbial minimalism: Genome reduction in bacterial pathogens. *Cell,***108**: 583-586.

Morowitz, H.J. and Wallace, D.C. (1973). Genome size and life cycle of the mycoplasma. In: Maramorosch, K. (ed), *Mycoplasma and Mycoplasma-like agents of human, animal and plant diseases,* pp. 62-73. N.Y. Acad. Sci., New York, USA.

Mrázek, J. (2006). Analysis of distribution indicates diverse functions of simple sequence repeats in Mycoplasma genomes. *Mol. Biol. Evol.*, **23**: 1370-1385.

Mrázek, J., Guo, X. and Shah, A. (2007). Simple sequence repeats in prokaryotic genomes. *Proc. Natl. Acad. Sci. USA*, **104**: 8472-8477.

Mühlradt, P.F., Kieß, M., Meyer, H., Süßmuth, R. and Jung, G. (1997). Isolation, structure elucidation, and synthesis of a macrophage stimulatory lipopeptide from *Mycoplasma fermentans* acting at picomolar concentration. *JEM*, **185**: 1951-1958.

Mühlradt, P.F., Kieß, M., Meyer, H., Süßmuth, R. and Jung, G. (1998). Structure and specific activity of macrophage-stimulating lipopeptides from *Mycoplasma hyorhinis*. *Infect. Immun.*, **66**: 4804-4810.

Musatovova, O., Kannan, T.R. and Baseman, J.B. (2012). *Mycoplasma pneumoniae* large DNA repetitive elements RepMP1 show type specific organization among strains. *PLoS ONE*, **7**: e47625.

Mushegian, A.R. and Koonin, E.V. (1996a). A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. USA*, **93**: 10268-10273.

Mushegian, A.R. and Koonin, E.V. (1996b). Gene order is not conserved in bacterial evolution. *Trends Genet.*, **12**: 289-290.

Najjar, S.M., Phillipe, N., Accili, D. and Taylor, S.I. (1993). pp120, an Endogenous substrate of the insulin receptor tyrosine kinase, is an ecto-ATPase. In*: Sereno symposia publications, (Vol 99)*, pp. 389-389. Raven Press.

Nakazaki, T., Ihara, N., Fucuta, Y. and Ikehashi, H. (2000). Abundant polymorphism in the flanking regions of two loci for basic PR-1 proteins as markers *for Indica japonica* differentiation in rice. *Breeding Sci.,***50**: 173-181.

National Institute of Allergy and Infectious Diseases. (2004). Taps Chiron to Develop Vaccine Against H9N2 Avian Influenza. Available from http://www.niaid.nih.gov/news/newsreleases-/2004/h9n2.htm. Retrieved 2010-01-31.

Necşulea, A. and Lobry, J.R. (2007). A new method for assessing the effect of replication on DNA base composition asymmetry. *Mol. Biol. Evol.*, **24**: 2169-2179.

Neimark, H. and London, J. (1982). Origins of the mycoplasmas: sterol-nonrequiring mycoplasmas evolved from streptococci. *J. Bacteriol.,* **150**: 1259-1265.

Nelson, D.L. and Cox, M.M. (2004). Principles of Biochemistry (4th Ed). W.H. Freeman and Company, New York.

Neutra, M.R., Pringault, E. and Kraehenbuhl, J.-P. (1996). Antigen sampling across epithelial barriers and induction of mucosal immune responses. *Ann. Rev. Immun.*, **14**: 275-300.

Neyrolles, O., Eliane, J.-P., Ferris, S., Ayr Florio da Cunha, R., Prevost, M.-C., Bahraoui, E. and Blanchard, A. (1999). Antigenic characterization and cytolocalization of P35, the major *Mycoplasma penetrans* antigen. *Microbiol.*, **145**: 343-355.

Nicolás, M.F., Barcellos, F.G., Hess, P.N. and Hungria, M. (2007). ABC transporters in *Mycoplasma hyopneumoniae* and *Mycoplasma synoviae*: Insights into evolution and pathogenicity. *Genet. Mol. Biol.,* **30**: 202-211.

Novelli, G. and D'Apice, M.R. (2012). Protein farnesylation and disease. *J. Inherit. Metab. Dis.*, **35**: 917-926.

Ogasawara, N., Moriya, S. and Yoshikawa, H. (1991). Initiation of chromosome replication: Structure and function of *oriC* and DnaA protein in eubacteria. *Res. Microbiol.*, **142**: 851-859.

Olivier, A.J. (2006). Respiratory diseases in ostriches: Air sac infection. Available from http://www.kleinkaroo.com/air_sac.pdf. Retrieved 2008-06-04.

Orrantia, E. and Chang, P.L. (1990). Intracellular distribution of DNA internalized through calcium phosphate precipitation. *Exp. Cell Res.*, **190**: 170-174.

Oshop, G.L., Elankumaran, S. and Heckert, R.A. (2002). DNA vaccination in the avian. *Vet. Immunol. Immunopathol.,* **89**: 1-12.

Paddenberg, R., Weber, A., Wulf, S. and Mannherz, H.G. (1998). Mycoplasma nucleases able to induce internucleosomal DNA degradation in cultured cells possess many characteristics of eukaryotic apoptotic nucleases. *Cell Death Differ.*, **5**: 517-528.

Paddenberg, R., Wulf, S., Weber, A., Heimann, P., Beck, L.A. and Mannherz H.G. (1996). Internucleosomal DNA fragmentation in cultured cells under conditions reported to induce apoptosis may be caused by mycoplasma endonucleases. *Eur. J. Cell Biol.*, **71**: 105-119.

Papazisi, L., Troy, K.E., Gorton, T.S., Liao, X. and Geary, S.J. (2000). Analysis of cytadherence-deficient, GapA-negative *Mycoplasma gallisepticum* strain R. *Infect. Immun.,* **68**: 6643-6649.

Perelman, B. (1999). Health management and veterinary procedures. In: Deeming, D.C. (ed), *The Ostrich: Biology, production and health,* pp. 321-346. CAB International, Cambridge.

Pereyre, S., Gonzalez, P., De Barbeyrac, B., Darnige, A., Renaudin, H., Charron, A., Raherison, S., Bébéar, C. and Bébéar, C.M. (2002). Mutations in 23S rRNA account for intrinsic resistance to macrolides in *Mycoplasma hominis* and *Mycoplasma fermentans* and for acquired resistance to macrolides in *M. hominis*. *Antimicrob. Agents Chemother.*,**46**: 3142-3150.

Pham, P.L., Kamen, A. and Durocher, Y. (2006). Large-scale transfection of mammalian cells for the fast production of recombinant protein. *Mol. Biotechnol.*, **34**: 225-237.

Pitcher, D.G. and Nicholas, R.A. (2005). Mycoplasma host specificity: fact or fiction? *Vet. J.,* **170**: 300-306.

Plesner, L. (1995). Ecto-ATPases: Identities and functions. In: Jeon, K.W. and Jarvik, J. (eds), *International Review of Cytology*, pp. 141-214. Academic Press, London.

Pollack, J.D. (2002). Central carbohydrate pathways: Metabolic flexibility and the extra role of some "housekeeping" enzymes. In: Razin, S. and Herrmann, R. (eds), *Molecular Biology and Pathogenicity of Mycoplasmas*, pp. 163-199. Springer, US.

Porgador, A., Irvine, K.R., Iwasaki, A., Barber, B.H., Restifo, N.P. and Germain, R.N. (1998). Predominant role for directly transfected dendritic cells in antigen presentation to CD^{8+} cells after gene gun immunization. *JEM*, **188**: 1075-1082.

Prescott, L.M., Harley, J.P. and Klein, D.A. (2002). Microbiology (5th Ed), pp 518-521. McGrawHill, NY.

Pretorius, B. (2009). The ostrich mycoplasma *Ms01*: The identification, isolation and modification of the *P100* vaccine candidate gene and immunity elicited by poultry mycoplasma vaccines. MSc dissertation. University of Stellenbosch, SA.

Price, M.N., Huang, K.H., Alm, E.J. and Arkin, A.P. (2005). A novel method for accurate operon predictions in all sequenced prokaryotes. *Nucl. Acids Res.*, **33**: 880-892.

Promega Protocols and Applications Guide: Transfection. Available from www.promega.com. Last accessed December 2012.

Rabson, A., Roitt, I.M. and Delves, P.J. (2005). Really Essential Medical Immunology (2nd Ed), Blackwell Publishing Ltd.

Rao, C.D., Pech, M., Robbins, K.C. and Aaronson, S.A. (1988). The 5' untranslated sequence of the c-sis/platelet-derived growth factor 2 transcript is a potent translational inhibitor. *Mol. Cell Biol.*, **8**: 284-292.

Rappuoli, R. (2000). Reverse vaccinology. *Curr. Opin. Microbiol.,* **3**: 445-450.

Rappuoli, R. (2001). Reverse vaccinology, a genome-based approach to vaccine development. *Vaccine*, **19**: 2688-2691.

Rawadi, G. and Roman-Roman, S. (1996). Mycoplasma membrane lipoproteins induced proinflammatory cytokines by a mechanism distinct from that of lipopolysaccharide. *Infect. Immun.*, **64**: 637-643.

Rawadi, G., Roman-Roman, S., Castedo, M., Dutilleul, V., Susin, S., Marchetti, P., Geuskens, M. and Kroemer, G. (1996). Effects of *Mycoplasma fermentans* on the myelomonocytic lineage. Different molecular entities with cytokine-inducing and cytocidal potential. *J. Immunol.*, **156**: 670-678.

Razin, S. (1978). The Mycoplasmas. *Microbiol. Rev.,* **42**: 414-470.

Razin, S. (1985). Molecular biology and genetics of mycoplasmas (Mollicutes). *Microbiol. Rev.,* **49**: 419-455.

Razin, S. (1987). Appealing attributes of mycoplasmas in cell biology research. *Isr. J. Med. Sci.,* **23**: 318-325.

Razin, S. (1992). Mycoplasma taxonomy and ecology. In: Maniloff, J., McElhaney, E.D., Finch, L.R. and Baseman, J.B. (eds), *Mycoplasmas Molecular Biology and Pathogenesis,* pp. 3-22. American Society for Microbiology, Washington, DC.

Razin, S. (1997). The minimal cellular genome of Mycoplasma. *Indian. J. Biochem. Bio.,* **34**: 124-130.

Razin, S. (1999). Adherence of pathogenic mycoplasmas to host cells. *Biosci. Rep.,* **19**: 367-372.

Razin, S. and Freundt, E.A. (1984). The Mycoplasmas. In: Krieg, N.R (ed), *Bergey's Manual® of Systematic Bacteriology (Vol 1),* pp. 740-770. Williams & Wilkins, Baltimore, USA.

Razin, S. and Jacobs, E. (1992). Mycoplasma adhesion. *J. Gen. Microbiol.,* **138**: 407-422.

Razin, S., Yogev, D. and Naot, Y. (1998). Molecular biology and pathogenicity of Mycoplasmas. *Microbiol. Mol. Biol. Rev.,* **62**: 1094-1156.

Redegeld, F., Filippini, A. and Sitovsky, M. (1991). Comparative studies of the cytotoxic T lymphocyte-mediated cytotoxicity and of extracellular ATP-induced cell lysis. Different requirements in extracellular Mg^{2+} and pH. *J. Immunol*., **147**: 3638-3645.

Reed, R.E. and Altman, S. (1983). Repeated sequences and open reading frames in the 3' flanking region of the gene for the RNA subunit of *Escherichia coli* ribonuclease P. *Proc. Natl. Acad. Sci. USA,* **80**: 5359-5363.

Reed, J.L., Famili, I., Thiele, I. and Palsson, B.O. (2006). Towards multidimensional genome annotation. *Nature Rev. Genet*., **7**: 130-141.

Ren, Q. and Paulsen, I.T. (2005). Comparative analysis of fundamental differences in membrane transport capabilities in prokaryotes and eukaryotes. *PLoS Comput. Biol*., **1**: e27.

Riss, T.L. and Moravec, R.A. (2004). Use of multiple assay endpoints to investigate the effects of incubation time, dose of toxin, and plating density in cell-based cytotoxicity assays. *ASSAY Drug. Develop. Technol*., **2**: 51-62.

Roberts, M.C. (1992). Antibiotic resistance. In: Maniloff, J., McElhaney, E.D., Finch, L.R. and Baseman, J.B. (eds), *Mycoplasmas Molecular Biology and Pathogenesis,* pp. 513-523. American Society for Microbiology, Washington, DC.

Robino, P., Alberti, A., Chessa, B., Pitau, M., Profitti, M., and Rosati, S. (2007). Molecular cloning and expression of a surface lipoprotein of *Mycoplasma capricolum* as a potential antigen for serological diagnosis. *Vet. Res. Commun*., **31**: 257–260.

Robinson, H.L. and Pertmer, T.M. (2000). DNA vaccines for viral infections: Basic studies and applications. *Adv. Virus Res*., **55**: 1-74.

Robinson, H.L. and Torres, C.A.T. (1997). DNA vaccines. *Semin. Immunol.,* **9**: 271-283.

Rocha, E.P.C. (2002). Is there a role for replication fork asymmetry in the distribution of genes in bacterial genomes? *Trends Microbiol*., **10**: 393-395.

Rocha, E.P.C. (2003). An appraisal of the potential for illegitimate recombination in bacterial genomes and its consequences: From duplications to genome reduction. *Genome Res*., **13**: 1123-1132.

Rocha, E.P.C. (2004). The replication-related organization of bacterial genomes. *Microbiol*., **150**: 1609-1627.

Rocha, E.P.C. and Blanchard, A. (2002). Genomic repeats, genome plasticity and the dynamics of *Mycoplasma* evolution. *Nucl. Acids Res.,* **30**: 2031-2042.

Rocha, E.P.C. and Danchin, A. (2002). Base composition bias might result from competition for metabolic resources. *Trends Gen*., **18**: 291-294.

Rocha, E.P.C., Guerdoux-Jamet, P., Moszer, I., Viari, A. and Danchin, A. (2000). Implication of gene distribution in the bacterial chromosome for the bacterial cell factory. *J. Biotech.*, **78**: 209-219.

Rosati, S., Robino, P., Fadda, M., Pozzi, S., Manelli, A., and Pittau, M. (2000). Expression and antigenic characterization of recombinant *Mycoplasma agalactiae* P48 major surface protein. *Vet. Microbiol.*, **71:** 201-210.

Rottem, S. (2002). Sterols and acylated proteins in Mycoplasmas. *Biochem. Biophys. Res. Commun.,* **292**: 1289-1292.

Rottem, S. (2003). Interaction of Mycoplasmas with host cells. *Physiol. Rev.,* **83**: 417-432.

Rottem, S. and Barile, M.F. (1993). Beware of mycoplasmas. *Trends Biotechnol.,* **11**: 143-151.

Rottem, S. and Naot, Y. (1998). Subversion and exploitation of host cells by mycoplasmas. *Trends Microbiol.,* **6**: 436-440.

Sachse, K., Helbig, J.H., Lysnyansky, I., Grajetzki, C, Müller, W., Jacobs, E. And Yogev, D.(2000). Epitope mapping of immunogenic and adhesive structures in repetitive domains of *Mycoplasma bovis* variable surface lipoproteins. *Infect. Immun.*, **68**: 680-687.

Sakai, H., Imamura, C., Osada, Y., Saito, R., Washio, T. and Tomita, M. (2001). Correlation between Shine-Dalgarno sequence conservation and codon usage of bacterial genes. *J. Mol. Evol.,* **52**: 164-170.

Salman, M., Deutsch, I., Tarshis, M., Naot, Y. and Rottem, S. (1994). Membrane lipids of *Mycoplasma fermentans*. *FEMS Microbiol. Lett.*, **123**: 255-260.

Salzberg, S.L., Delcher, A.L,. Kasif, S. and White, O. (1998). Microbial gene identification using interpolated Markov models. *Nucl. Acids Res.*, **26**: 544-548.

Sato, Y., Murakami, T., Funatsuki, H., Matsuba, S., Saruyama, H. and Tanida, M. (2001). Heat shock mediated APX gene expression and protection against chilling injury in rice seedlings. *J. Exp. Bot.*, **52**: 145-151.

Sato, Y., Roman, M., Tighe, H., Lee, D., Corr, M., Nguyen, M.-D., Silverman, G.J. Lotz, M., Carson, D.A. and Raz, E. (1996). Immunostimulatory DNA sequences necessary for effective intradermal gene immunization. *SCI*, **273**: 352-354.

Schmidt, E.V., Christoph, G., Zeller, R. and Leder, P. (1990). The cytomegalovirus enhancer: A pan-active control element in transgenic mice. *Mol. Cell Biol.*, **10**: 4406-4411

Schuster, S., Pfeiffer, T., Moldenhauer, F., Koch, I. and Dandekar, T. (2002). Exploring the pathway structure of metabolism: Decomposition into subnetworks and application to *Mycoplasma pneumoniae*. *Bioinform.*, **18**: 351-361.

Sedegah, M., Hedstrom, R., Hobart, P. and Hoffman, S.L. (1994). Protection against Malaria by immunization with plasmid DNA encoding circumsporozoite protein. *Proc. Natl. Acad. Sci. USA,* **91**: 9866-9870.

Senapathy, P., Shapiro, M.B. and Harris, N.L. (1990). Splice junctions, branch point sites, and exons: Sequence statistics, identification and applications to genome project. *Methods Enzymol.*, **183**: 252-278.

Shapiro, J.A. (2002). A 21st Century view of evolution. *J. Biol. Phys.*, **28**: 745-764.

Shaw, B.M., Simmons, W.L and Dybvig, K. (2012). The Vsa shield of *Mycoplasma pulmonis* is antiphagocytic. *Infect. Immun.*, **80**: 704-709.

Sheets, R.L., Stein, J., Manetz, T.S., Duffy, C., Nason, M., Andrews, C., Kong, W.-P., Nabel, G.J. and Gomez, P.L. (2006). Biodistribution of DNA plasmid vaccines against HIV-1, Ebola, Severe Acute Respiratory Syndrome, or West Nile virus is similar, without integration, despite differing plasmid backbones or gene inserts. *Toxicol. Sci.*, **91**: 610-619.

Shibata, K., Sasaki, T. and Watanabe, T. (1995). AIDS-associated mycoplasmas possess phospholipases C in the membrane. *Infect. Immun.*, **63**: 4174-4177.

Simecka, J.W., Davis, J.K., Davidson, M.K., Ross, S.E., Städtlander, C.T.K.-H. and Cassell, G.H. (1992). Mycoplasma diseases of animals. In: Maniloff, J., McElhaney, E.D., Finch, L.R. and Baseman, J.B. (eds), *Mycoplasmas Molecular Biology and Pathogenesis*,pp. 391-415. American Society for Microbiology, Washington, DC.

Simmons, W.L., Bolland, J.R., Daubenspeck, J.M. and Dybvig, K. (2007). A stochastic mechanism for biofilm formation by *Mycoplasma pulmonis. J. Bacteriol.*, **189**: 1905-1913.

Simmons, W.L. and Dybvig, K. (2007). Biofilms protect *Mycoplasma pulmonis* cells from lytic effects of complement and gramicidin. *Infect. Immun.*, **75**: 3696-3699.

Siqueira, F.M., Schrank, A. and Schrank, I.S. (2011). *Mycoplasma hyopneumoniae* transcription unit organization: Genome survey and prediction. *DNA Res.,***18**: 413-422.

Sluijter, M., Spuesens, E.B., Hartwig, N.G., van Rossum, A.M. and Vink, C. (2009). The *Mycoplasma pneumoniae* MPN490 and *Mycoplasma genitalium* MG339 genes encode reca homologs that promote homologous DNA strand exchange. *Infect Immun.*, **77**: 4905-4911.

Smooker, P.M., Steeper, K.R., Drew, D.R., Strugnell, R.A. and Spithill,T.W. (1999). Humoral responses in mice following vaccination with DNA encoding glutathione *S*-transferase of *Fasciola hepatica*: effects of mode of vaccination and the cellular compartment of antigen expression. *Parasite Immunol.*, **21**: 357-364.

Söll, D. and RajBhandary, U.L. (2006). The genetic code - Thawing the 'frozen accident'. *J. Biosci.,* **31**: 459-463.

Song, K.D., Lillehoj, H.S., Choi, K.D., Yun, C.H., Parcells, M.S., Huynh, J.T. and Han, J.Y. (2000). A DNA vaccine encoding a conserved Eimeria protein induces protective immunity against live *Eimeria acervulina* challenge. *Vaccine,***19**: 243-252.

Soultanas, P. (2011). The replication-transcription conflict. *Transcription*, **2**: 140-144.

South African Ostrich Business Chamber. (2010). Available from www.saobc.co.za.

South African Ostrich Industry. (2004). Available form http://www.ostrichsa.co.za/ostriches.php?action=-get_article&id=8.

Souza, R.C., de Almeida, D.F., Zaha, A., de Lima Morais, D.A. and de Vasconcelos, A.T.R. (2007). In search of essentiality: Mollicute-specific genes shared by twelve genomes. *Gen. and Mol. Biol.,***30**: 169-173.

Spanjaard, R.A., Chen, K. and Van Duin, J. (1990). Frameshift suppression at tandem AGA and AGG codons by cloned tRNA genes: assigning a codon to *argU* tRNA and T4 $tRNA^{Arg}$. *Nucl. Acids Res.*, **18**: 5031-5036.

Spanjaard, R.A. and Van Duin, J. (1988). Translation of the sequence AGG-AGG yields 50% ribosomal frameshift. *Proc. Natl. Acad. Sci.* USA, **85**: 7967-7971.

Steenmans, S. (2010). The identification, contiguous sequence annotation, cloning, and site-directed mutagenesis of the *P100* vaccine candidate gene of the ostrich mycoplasma Ms02. MSc dissertation, University of Stellenbosch, SA.

Stern, A.M. and Markel, H. (2005). The history of vaccines and immunization: Familiar patterns, new challenges. *Health Aff.,***24**: 611-21.

Stothard, P. and Wishart, D.S. (2006). Automated bacterial genome analysis and annotation. *Curr. Opin Microbiol.*, **9**: 505-510.

Stout, J.G., Strobel, R.S. and Kirley, T.L. (1995). Properties of and proteins associated with the extracellular ATPase of chicken gizzard smooth muscle. *J. Biol. Chem.*, **270**: 11845-11850.

Suarez, D.L. and Schultz-Cherry, S. (2000). The effect of eukaryotic expression vectors and adjuvants on DNA vaccines in chickens using an avian influenza model. *Avian Dis.*, **44**: 861-868.

Swertz, M.A. and Jansen, R.C. (2007). Beyond standardization: Dynamic software infrastructures for systems biology. *Nat. Rev. Gen.*, **8**: 253-243.

Tarshis, M., Salman, M. and Rottem, S. (1993). Cholesterol is required for the fusion of single unilamellar vesicles with *Mycoplasma capricolum*. *Biophys. J.*, **64**: 709-715.

Tavel, J.A., Martin, J.E., Kelly, G.C., Enama, M.E., Shen, J.M., Gomez, P.L, Andrews, C.A., Koup, R.A., Bailer, R.T., Stein, J.A., Roederer, M., Nabel, G.J. and Graham, B.S. (2007). Safety and immunogenicity of a Gal-Pol candidate JIC-1 DNA vaccine administered by a needle-free device in HIV-1-seronegative subjects. *J. Acquir. Immune Defic. Syndr.*, **44**: 601-605.

Theiss, P., Karpas, A. and Wise, K.S. (1996). Antigenic topology of the P29 surface lipoprotein of *Mycoplasma fermentans*: Differential display of epitopes results in high-frequency phase variation. *Infect. Immun.*, **64**: 1800-1809.

Thomas, J.M., Horspool, D., Brown, G., Tcherepanov, V. and Upton, C. (2007). GraphDNA: A Java program for graphical display of DNA composition analyses. *BMC Bioinform.,***8**: 21.

Thomson, B.J., Dewhurst, S. and Gray, D. (1984). Structure and heterogeneity of the a sequences of human herpesvirus 6 strain variants U1102 and Z29 and identification of human telomeric repeat sequences at the genomic termini. *J. Virol.*, **68**: 3007-3014.

Toth, G., Gaspari, Z. and Jurka, J. (2000). Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Res.*, **10**: 967-981.

Treangen, T.J. and Salzberg, S.L. (2012). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature Rev. Genet.*,**13**: 36-46.

Trivedi, S. (2010). Do simple sequence repeats in replication, repair and recombination genes of mycoplasmas provide genetic variability? *J. Cell Mol. Biol.*,**8**: 53-70.

Triyatni, M., Jilbert, A.R., Qiao, M., Miller, D.S. and Burrell, C.J. (1998). Protective efficacy of DNA vaccines against duck Hepatitis B Virus infection. *J. Virol.*, **72**: 84-94.

Tryon, V.V. and Baseman, J.B. (1992). Pathogenic determinants and mechanisms. *Am. Soc. Microbiol.*,pp. 457-472.Washington DC.

Tully, J.G. (1993). Current status of the Mollicute flora of humans. *Clin. Infect. Dis.*, **17**: S2-S9.

Uzzau, S., Marogna, G., Sisinnio Leori, G., Curtis, R., Schianchi, G., Stocker, B.A.D. and Rubino, S. (2005). Virulence attenuation and live vaccine potential of *aroA*, *crp cdt cya*, and plasmid-cured mutants of *Salmonella enterica* Serovar Abortusovis in mice and sheep. *Infect. Immun.*, **73**: 4302-4308.

Vandenberghe, L.H., Wilson, J.M. and Gao, G. (2006). Identification of novel adenovirus serotypes for use as vectors. In: Kaplitt, M.G. and During, M.J. (eds), *Gene Therapy of the central nervous system*, pp. 17-24. Academic Press, Amsterdam.

Van der Merwe, L. (2006). Preliminary investigations into ostrich mycoplasmas: Identification of vaccine candidate genes and immunity elicited by poultry mycoplasma vaccines. MSc dissertation, University of Stellenbosch,SA.

Van der Woude, M.W. and Baumler, A.J. (2004). Phase and antigenic variation in bacteria. *Clin. Microbiol. Rev.*, **17**: 581-611.

van Drunen Littel-van den Hurk, S., Loehr, B.I. and Babiuk, L.A. (2001). Immunization of livestock with DNA vaccines: Current studies and future prospects. *Vaccine*, **19**: 2474-2479.

Van Gaal, E.V.B.,van Eijk, R., Oosting, R.S., Jan Kok, R., Hennink, W.E., Crommelin, D.J. and Mastrobattista, E. (2011). How to screen non-viral gene delivery systems *in vitro*? *J. Control. Release*, **154**: 218-232.

Vanlaere, I. and Libert, C. (2009). Matrix metalloproteinases as drug targets in infections caused by Gram-negative bacteria and in septic shock. *Clin. Microbiol. Rev.*, **22**: 224-239.

Wang, Z., Trolio, P.J., Wang, X., Griffiths, T.G., Pacchione, S.J., Barnum, A.B., Harper, B., Pauley, C.J., Niu, Z., Denisova, L., Follmer, T.T., Rizzuto, G, Ciliberto, G., Fattori, E., Monica, N.L., Manam, S. and Ledwith, B.J. (2004). Detection of integration of plasmid DNA into host genomic DNA following intramuscular injection and electroporation. *Gene Ther.*, **11**: 711-721.

Washburn, L.R., Weaver, K.E., Weaver, E.J., Donelan, W. and Al-Sheboul, S. (1998). Molecular characterization of *Mycoplasma arthritidis* variable surface protein MAA2. *Infect. Immun.*, **66**: 2576-2586.

Wassenaar, T.M,. Bohlin, J., Binnewies, T.T and Ussery, D.W. (2009). Genome comparison of bacterial pathogens. *Genome Dyn.*, **6**: 1-20.

Webster, R.G. (1998). DNA vaccination: A potential future strategy. In: Nicholson, K.G., Webster, R.G. and Hay, A.J (eds), *Textbook of Influenza,* pp. 410-421. Blackwell Sci., Oxford.

Webster, R.G. and Robinson, H.L. (1997). DNA vaccines: A review of developments. *BioDrugs,* **8**: 273-292.

Weeratna, R., Brazolot Millan, C.L., Krieg, A.M. and Davic, H. (1998). Reduction of antigen expression from DNA vaccines by coadministrated oligodeoxynucleotides. *Antisense Nucl. Acid Drug Dev.*, **8**: 351-356.

Weintraub, H., Cheng, P.F. and Conrad, K. (1986). Expression of transfected DNA depends on DNA topology. *Cell*, **46**: 115-122.

Weisburg, W.G., Tully, J.G., Rose, D.L., Petzel, J.P., Oyaizu, H., Yang, D., Mandelco, L., Sechrest, J., Lawrence, T.G., Van Etten, J., Maniloff, J. and Woese, R. (1989). A phylogenetic analysis of Mycoplasmas: basis for their classification. *J. Bacteriol.,* **171**: 6455-6467.

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

Williamson, C.M., Bramley, A.J. and Lax, A.J. (1994). Expression of the lysostaphin gene of *Staphylococcus simulans* in a eukaryotic system. *Appl. Environ. Microbiol.*, **60**: 771-776.

Wise, K.S., Foecking, M.F., Roske, K., Lee, Y.J., Lee, Y.M., Madan, A. and Calcutt, M.J. (2006). Distinctive repertoire of contingency genes conferring mutation–based phase variation and combinatorial expression of surface lipoproteins in *Mycoplasma capricolum* subsp. *capricolum* of the *Mycoplasma mycoides* phylogenetic cluster. *J. Bacteriol.*, **188**: 4926-4941.

Woese, C.R. (1987). Bacterial evolution. *Microbiol. Rev.*, **51**: 221-271.

Woese, C.R., Maniloff, J. and Zablen, L.B. (1980). Phylogenetic analysis of the mycoplasmas. *Proc. Natl. Acad. Sci. USA,* **77**: 494-498.

Wolff, J.A., Malone, R.W., Williams, P., Chong, W., Acsadi, G., Jani, A. and Felgner, P.L. (1990). Direct gene transfer into mouse muscle *in vivo*. *SCI*, **247**: 1465-1468.

Worning, P., Jensen, L.J., Hallin, P.F., Staerfeldt, H.H. and Ussery, D.W. (2006). Origin of replication in circular prokaryotic genomes. *Environ. Microbiol.,***8**: 353-361.

Wu, G.Y. and Wu, C.H. (1987). Receptor-mediated *in vitro* gene transformation by a soluble DNA carrier system. *J. Biol. Chem.*, **262**: 4429-4432.

Wurm, F.M. (2004). Production of recombinant protein therapeutics in cultivated mammalian cells. *Nature Biotech.,***22**: 1393-1398.

Xia, X. (2012). DNA replication and strand asymmetry in prokaryotic and mitochondrial genomes. *Curr. Genomics*, **13**: 16-27.

Yang, H., Cao, S., Huang, X., Liu, J., Tang, Y. and Wen, X. (2009). Annotation confidence score for genome annotation: A genome comparison approach. *Vaccine*, **27**: 5035-5040.

Yoder, H.W.(1984). *Mycoplasma gallisepticum* infection. In: Hofstad, M.S. (ed), *Diseases of Poultry (8th Ed),* pp. 190-201. The Iowa State University Press, Iowa, USA.

Yogev, D., Rosengarten, R., Watson-McKown, R. and Wise, K.S. (1991). Molecular basis of *Mycoplasma* surface antigenic variation: a novel set of divergent genes undergo spontaneous mutation of periodic coding regions and 5' regulatory sequences. *EMBO J.*, **10**: 4069-4079.

Yotoko, K.S.C. and Bonatto, S.L. (2007). A phylogenomic appraisal of the evolutionary relationship of Mycoplasmas. *Genet. Mol. Biol.*, **30**: 270-276.

Zalacain, M., Biswas, S., Ingraham, K.A., Ambrad, J., Bryant, A., Chalker, A.F., Iordanescu, S., Fan, J., Fan, F., Lunsford, R.D., O'Dwyer, K., Palmer, L.M., So, C., Sylvester, D., Volker, C., Warren, P., McDevitt, D., Brown, J.R., Holmes, D.J. and Burnham, M.K.R. (2003). A global approach to indentify novel broad-spectrum antibacterial targets among proteins of unknown function. *J. Mol. Microbiol. Biotechnol.*, **6**: 109-126.

Zhang, Q., Young, T.F. and Ross, R.F. (1994). Microtiter plate adherence assay and receptor analogs for *Mycoplasma hyopneumoniae*. *Infect. Immunol.*, **62**: 1616-1622.

Zhang, Q., Young, T.F. and Ross, R.F. (1995). Identification and characterization of a *Mycoplasma hyopneumoniae* adhesin. *Infect. Immunol.*, **63**: 1013-1019.

Zhang, R. and Zhang, C.-T. (2004a). Isochore structures in the mouse genome. *Genomics*, **83**: 384-394.

Zhang, R. and Zhang, C.-T. (2004b). A systematic method to identify genomic islands and its applicationsin analyzing the genomes of *Corynebacterium glutamicum* and *Vibrio vulnificus* CMCP6 chromosome. *Bioinform.*, **20**: 612-622.

Zhang, R. and Zhang, C.-T. (2005). Identification of replication origins in archael genomes based on the Z curve method. *Archaea*, **1**: 335-346.

Zhang, R., Zhang, C.-T. and Ou, H.-Y. (2003). The Z curve database: A graphic representation of genome sequences. *Bioinform.*, **19**: 593-599.

Addendum A

Analysis of the distribution of essential genes and ORFs across the contigs of the Ms02 genome

Table A. Previous annotation results illustrated for each contig, with the essential gene and ORF distribution, to predict the direction of the contig. *NEI: Not enough information. Contig 16, 37 and 39 were not included, since they were not annotated previously.

			Essential genes			Total genes			
Contiguous sequence	**Size (Kbp)**	**ORFs total**	**+ strand**	**- strand**	**+ percentage**	**+ strand**	**- strand**	**+ percentage**	**Predicted direction**
contig00001	17530	17	6	2	75	14	3	82	+
contig00002	1268	1	/	/	/	1	/	100	NEI*
contig00003	1204	1	/	/	/	/	1	0	NEI*
contig00004	24502	27	1	10	9	10	17	37	-
contig00006	25664	6	/	/	/	3	3	50	NEI*
contig00007	86160	43	3	9	25	10	33	23	-
contig00008	54571	49	7	10	41	18	31	37	-
contig00009	5851	8	/	2	0	/	8	0	NEI*
contig00010	12270	14	4	4	50	6	8	43	NEI*
contig00011	1800	2	/	1	0	/	2	0	NEI*
contig00012	39046	28	1	2	33	20	8	71	NEI*
contig00013	71894	60	12	8	60	24	36	40	NEI*
contig00014	67524	37	4	14	22	6	31	16	-
contig00015	75929	61	17	6	74	40	21	66	+
contig00017	2262	2	/	/	/	2	/	100	NEI*
contig00026	9174	2	/	/	/	1	1	50	NEI*
contig00029	111665	81	8	22	27	29	52	36	-
contig00030	127294	115	12	32	27	40	75	35	-
contig00031	27764	13	/	1	0	6	7	46	NEI*
contig00032	37340	31	3	8	27	10	21	32	-
contig00033	4870	2	2	/	100	2	/	100	NEI*
contig00034	36348	42	31	/	100	38	4	90	+
contig00038	1089	2	/	/	/	/	2	0	NEI*
contig00041	43856	27	3	6	33	12	15	44	-
contig00044	5319	4	1	1	50	4	/	100	NEI*

Addendum B

Sequences of the origin of replication of the Ms02 genome

Ori-Finder

Ori sequence as detected by the Ori-Finder. The three DnaA boxes are indicated in bold and capital letters. The sequence is located 866359-866821 bp in the Ms02 genome.

taaacctatttttgtttttctttttggatgtaagaagtttcgatgtcatttagaagcctcccgatcttaacctgctcattta
ttgaaggattcattataggcattgtgtgtagtacatcttcctttattacgaatcttcttgctttagctcctgtttctctattg
aaaaaaataaatttatgtcagtaatttgataagaaaaag**TTTTCCAAA**aagtcgttggaaatatcatgcgttttaaaaacata
gtaaagtgaagaagttactcctgtaattcctgttttatttctctttattgacccgtacaagaaccc**TTTGGATAA**tga**TGTTG
AAAA**aatgaaactgttcggttcaactaatcaataacctttcattgtttcttttgaaggagtcatatcttgatcaaaaaagtct
gattgcctaacaacaccgtattttgcagtgctagttagtgtaatcgag

DoriC

Ori sequence as detected by the DoriC finder located 561223-561339 bp in the Ms02 genome. DoriC is a database containing numerous *ori*s. The following sequence has an 84% identity with *M. hyorhinis* HUB-1 chromosome.

ttttctacctttagctcttctagcagctaaaacttttctaccgttagcagttttcattcttgctctaaaaccgtgaaccttag
cgtgtttacgtttgtttggttggtatgttctttt

Addendum C

Table of the ORFs that occur over the newly linked contiguous sequences of the Ms02 genome

Table C. The first most significant hits obtained from BLASTx searches on the Ms02 genome in the NCBI database, employing the mitochondrial genetic code.

Query range	Contig	Accession number	Description	Bit score	E-value
13…1137	3	YP_003515791.1	Type I R/M system specificity subunit [*Mycoplasma agalactiae*]	270	$1e^{-25}$
2014…3204	16_2	YP_003515788.1	Type I R/M system specificity subunit [*Mycoplasma agalactiae*]	182	$3e^{-14}$
29114…30391	6_7	YP_003515788.1	Type I R/M system specificity subunit [*Mycoplasma agalactiae*]	254	$2e^{-23}$
113762…115246	7_17	ZP_08703696.1	PTS system lactose/cellobiose family IIC subunit [*Mycoplasma anatis* 1340]	1228	$7e^{-164}$
115246…116337	7_17	ZP_08703694.1	Outer surface protein [*Mycoplasma anatis* 1340]	1095	$5e^{-147}$
146983…147924	37_39	YP_003515788.1	Type I R/M system specificity subunit [*Mycoplasma agalactiae*]	139	$5e^{-09}$
202935…203603	8_AS	ZP_10029432.1	Beta-phosphoglucomutase [*Mycoplasma canis* UF31]	624	$5e^{-80}$
203853…204833	AS_15	YP_003560335.1	Hypothetical protein MCRO_0736 [*Mycoplasma crocodyli* MP145]	457	$1e^{-52}$
280435…281040	38_AS	YP_004683700.1	ISMbov4 transposase [*Mycoplasma bovis* Hubei-1]	643	$2e^{-80}$
305562…306290	41_start	ZP_14232350.1	Asparaginyl-tRNA synthetase [*Mycoplasma canis* UF33]	118	$2e^{-06}$
347720…349663	41_AS	ZP_08703784.1	Transketolase [*Mycoplasma anatis* 1340]	2623	0
361746…362387	10_AS	YP_004790383.1	Type I restriction modification DNA specificity protein [*Mycoplasma putrefaciens* KS1]	140	$8e^{-10}$
368936…370081	end_11_12	YP_003515788.1	Type I R/M system specificity subunit [*Mycoplasma agalactiae*]	180	$1e^{-16}$

Table C. The first most significant hits obtained from BLASTx searches on the Ms02 genome in the NCBI database, employing the mitochondrial genetic code.

Query range	Contig	Accession number	Description	Bit score	E-value
884164…884514	1_26	ZP_02931917.1	DNA gyrase, B subunit [*Ureaplasma urealyticum* serovar 13 str. ATCC 33698]	77	0.048
893663…894826	26_44	YP_003515791.1	Type I R/M system specificity subunit [*Mycoplasma agalactiae*]	274	$8e^{-26}$
146972…145299	AS_37	YP_002961132.1	Transposase IS1634AM [*Mycoplasma conjunctivae* HRC/581]	1441	0
203699…203821	AS	YP_004137136.1	Hypothetical protein MfeM64YM_0762 [*Mycoplasma fermentans* M64]	110	$2e^{-08}$
866822...866980	32_1	ZP_20412978.1	Type I R/M system specificity subunit domain protein, partial [*Mycoplasma* sp. G5847]	63	0.071