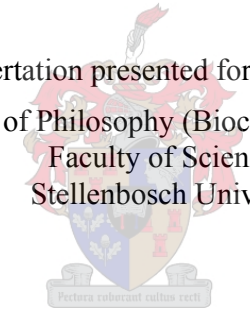
The development of a DNA vaccine against *Mycoplasma nasistruthionis* sp. nov. for use in ostriches

by

Martha Wium



Dissertation presented for the degree of
Doctor of Philosophy (Biochemistry) in the
Faculty of Science at
Stellenbosch University

Supervisor: Dr. Annelise Botes
Co-supervisor: Prof. Dirk U. Bellstedt

December 2015

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 otherwise stated), 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.

This dissertation includes one original paper published in peer-reviewed journals or books and two unpublished publications. The development and writing of the papers (published and unpublished) were the principal responsibility of myself and, for each of the cases where this is not the case, a declaration is included in the dissertation indicating the nature and extent of the contributions of co-authors.

Martha Wium

December 2015

Summary

Mycoplasma nasistruthionis sp. nov. str. Ms03 (Ms03) is one of three *Mycoplasma* species that were identified from ostriches. Mycoplasmas infections have been implicated in ostrich chick mortalities, growth retardation and downgrading of ostrich carcasses. Currently there is no vaccine available for the treatment of mycoplasmosis in ostriches. This study investigated the development of DNA vaccines against Ms03 infections in ostriches. To this end, the Ms03 genome was sequenced and annotated. The vaccine candidate gene, *oppA*, was identified within the genome sequence and characterized before DNA vaccines containing the *oppA* were developed and tested.

The genome of Ms03 was sequenced and the resulting 172 contigs were annotated. This dissertation presents the first Ms03 draft genome and annotation which contributed to the understanding of Ms03 as a miniature genetically independent organism. In Ms03, genome replication, cell division, RNA transcription, protein translation and glycolysis resemble that of the closely related *Mycoplasma synoviae* 53. Purine and pyrimidine metabolism was incomplete and *de novo* synthesis thereof was not possible. Amino acid synthesis in Ms03 was mostly absent and only the genes that convert aspartate to asparagine and glycine to serine were found. More importers than exporters were annotated owing to the lack of synthesis pathways in Ms03, which is typical for mycoplasmas that have parasitic life styles. Two oligopeptide permease (*opp*) operons were annotated within the Ms03 genome.

The potential of the *oppA* as a vaccine candidate gene was evaluated by investigating the need for a substrate-binding domain (OppA) as part of the OppBCDF transporter within *Mycoplasma* species. An *oppA* homologue could be identified for each *oppBCDF* operon in all species and therefore must play an essential role in oligopeptide transport. All mycoplasmas (except for hemoplasma) had one, two or three *opp* operons that could be divided into three types (Type A, B and C). Each type had unique InterPro and MEME domains and motifs which together with the phylogenetic analysis suggest unique roles in their survival under different conditions. Ms03 had a Type A and a Type B *opp* operon, the Type A *oppA* was used as vaccine candidate gene.

The Type A *oppA* was cloned and site-directed mutagenesis was used for codon correction before the mutated gene was sub-cloned into three DNA vaccine vectors. The three DNA vaccines (pCI-neo_*oppA*, VR1012_*oppA* and VR1020_*oppA*) were used to vaccinate ostriches and the OppA-antibody response was analysed by ELISA. The VR1020_*oppA* and pCI-neo_*oppA* constructs elicited a primary immune response in ostriches, indicating that the OppA protein was expressed *in vivo* and was immunogenic. This can therefore be viewed as the first step in the development of a DNA vaccine for the control of mycoplasma infections in ostriches.

Opsomming

Mycoplasma nasistruthionis sp. nov. str. Ms03 (Ms03) is een van drie mikoplasma spesies wat volstruise infekteer. Mikoplasma-infeksies in volstruise veroorsaak kuiken vrektes, vertraagde groei en afgradering van volstruis karkasse. Daar is tans geen geregistreerde mikoplasma entstof beskikbaar vir gebruik in volstruise nie. Hierdie studie het die ontwikkeling van DNS-entstowwe teen Ms03-infeksies in volstruise ondersoek. Vir hierdie doel was die Ms03-genoomvolgorde bepaal en geannoteer. Die entstofkandidaat-geen, *oppA*, was geïdentifiseer in die genoomvolgorde en gekarakteriseer voordat DNS-entstowwe (wat die *oppA*-geen bevat) ontwikkel en getoets is.

Die Ms03-genoomvolgorde was bepaal en die gegenereerde 172 aaneenlopende volgordes was geannoteer. Hierdie proefskrif bied die eerste voorlopige volgorde en annotering van die Ms03-genoom wat bygedra het tot die kennis van Ms03 as 'n miniatuur geneties onafhanklike organisme. Genoom-replikasie, seldeling, RNS-transkripsie, proteïen-translasie en glikolise in Ms03 stem ooreen met dié prosesse in die naverwante *Mycoplasma synoviae* 53. Die purien en pirimidien metabolisme was onvolledig en *de novo* sintese daarvan was nie moontlik in Ms03 nie. Aminosuursintese in Ms03 was meestal afwesig en net die gene wat aspartaat omskep na asparagien en glisien na serien was gevind in die annoteerde genoom. Meer invoerders as uitvoerders was geannoteer, wat dui op die gebrek aan sintesepadweë in Ms03. Dit is tipies van mikoplasmas wat ‘n parasitiese lewensstyle het. Twee oligopeptied-permeases (*opp*) operons was gevind in die Ms03-genoom.

Die potensiaal van die *oppA*-geen as 'n entstofkandidaat-geen was geëvalueer deur die behoefte van 'n substraatbindingsdomein (OppA) as deel van die OppBCDF-vervoerder binne mikoplasma spesies te ondersoek. 'n Homoloog van die *oppA*-geen kon geïdentifiseer word vir elke *oppBCDF*-operon in al die spesies en behoort daarom 'n noodsaaklike rol te speel in die vervoer van oligopeptiede. Alle mikoplasmas (behalwe vir hemoplasmas) het een, twee of drie *opp*-operons, wat verdeel kan word in drie tipes (Tipe A, B en C). Elke tipe het unieke InterPro en MEME domeine en motiewe wat saam met die filogenetiese ontleding daarop dui dat hulle unieke rolle in oorlewing onder verskillende omstandighede speel. Ms03 het 'n Tipe A en Tipe B *opp*-operon, die Tipe A *oppA* is gebruik as entstofkandidaat-geen.

Die Tipe A *oppA*-geen was gekloneer en teikengerigte-mutagenese was gebruik vir kodonregstellings voordat die gemuteerde geen in drie DNS-entstof vektore gesubkloneer was. Die drie DNS-entstowwe (pCI-neo_*oppA*, VR1012_*oppA* en VR1020_*oppA*) was gebruik om volstruise in te ent en die OppA-teenliggaamsreaksie was geanaliseer deur ELISA. Inenting met die VR1020_*oppA* en pCI-neo_*oppA* entstowwe het tot 'n primêre immuniteitsreaksie in volstruise gelei. Dit dui daarop dat die OppA proteïen *in vivo* uitgedruk en immunogenies was. Dit kan beskou word as die eerste stap in die ontwikkeling van 'n DNS-entstof vir die beheer van mikoplasma-infeksies in volstruise.

Acknowledgements

I would like to express my gratitude for the contributions the following people and institutes had made to this study:

- Dr. Annelise Botes, my supervisor for the continuous support and leadership.
- Prof. Dirk U. Bellstedt, my co-supervisor for advice and support.
- All the past and present members of the Bellstedt and Botes laboratories (2011-2015) for advice, support and friendship.
- Department of Biochemistry at the University of Stellenbosch.
- Ostrich Business Chamber for financial support.
- Technology and Human Resources for Industry Programme (THRIP) for financial support.

Last but not the least, I would like to thank my parents Manie and Cornelia Wium and my siblings Nel-mari, Howie, Eduard and Anneli as well as Stefan Opperman and all my friends for supporting me throughout writing this dissertation and my life in general.

Abbreviations

A	Adenine
A+T	Adenine and thymine
A-tailing	Adenylation of PCR products
Ab	Antibody
ABC	ATP-binding cassette
ABTS	2,2'-Azino-bis(3-ethylbenzthiazoline-6-sulphonic acid)
ADP	Adenosine diphosphate
Ag	Antigen
AI	Avian influenza
AmiD	Amino acid transport and metabolism/Inorganic ion transport and metabolism
AMP	Adenosine monophosphate
ANOVA	Analysis of variance
APCs	Antigen presenting cells
arcA	Arginine deiminase
arcB	Ornithine carbamoyltransferase
arcC	Carbamate kinase
ATP	Adenosine triphosphate
ATPase	Adenosine triphosphatase
BER	Blast Extend Repraze
BGH	Bovine growth hormone
BLAST	Basic local alignment search tool
BLOSUM62	Blocks substitution matrix 62
bp	Base pairs
BSA	Bovine serum albumin
C	Cytosine
CAF	Central Analytical Facility
CARDS	Community-acquired respiratory distress syndrome
CATH	Class architecture topology homology
CDP	Cytidine diphosphate
CMP	Cytidine monophosphate
CMV	Cytomegalovirus
CoA	Coenzyme A
COGs	Categorization of orthologs
Contigs	Contiguous sequences
CTP	Cytidine triphosphate
dADP	Deoxyadenosine diphosphate
DAFF	Department of Agriculture, Forestry and Fisheries
dAMP	Deoxyadenosine monophosphate
dATP	Deoxyadenosine triphosphate
DC	Dendritic cells
dCDP	Deoxycytidine diphosphate
dCMP	Deoxycytidine monophosphate

dCTP	Deoxycytidine triphosphate
dGDP	Deoxyguanosine diphosphate
dGMP	Deoxyguanosine monophosphate
dGTP	Deoxyguanosine triphosphate
DHAP	Dihydroxyacetone phosphate
DIVA	Differentiate infected from vaccinated animals
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
Dpp	Dipeptide permease
dTDP	Deoxythymidine diphosphate
dTMP	Deoxythymidine monophosphate
dTTP	Deoxythymidine triphosphate
EC	Enzyme commission
ECF	Energy-coupling factor
ecto-ATPase	Ecto-adenosine triphosphatase
EDTA	Ethylenediaminetetraacetic acid
EI	Enzyme I
EII	Enzyme II
ELISA	Enzyme-linked immunosorbent assay
emPCR	Emulsion-based clonal amplification
Fab	Fragment antigen-binding
Fc	Fragment crystallizable region
FIGfam	Fellowship for Interpretation of Genomes protein families
G	Guanine
G+C	Guanine and cytosine
G3P	Glycerol-3-phosphate
gDNA	Genomic DNA
GDP	Guanosine diphosphate
GMP	Guanosine monophosphate
GO	Gene ontology
GS FLX	Genome Sequencer FLX
GST	Glutathione S-transferase
GTP	Guanosine triphosphate
GUU	*Mycoplasma iowae* 695
H_2O_2	Hydrogen peroxide
HAMAP	High-quality automated and manual annotation of proteins
HGT	Horizontal gene transfer
HMM	Hidden Markov Model
HPr	Histidine-containing phospho-carrier protein
HPR	Streptavidin horseradish peroxidase
ICE	Integrated conjugative elements
id	Intradermal
IDT	Integrated DNA technologies

IE	Immediate-early
IgA	Immunoglobulin A
IgG	Immunoglobulin G
IGS	Institute for Genome Sciences
Im	Intramuscular
IMP	Inosine monophosphate
InterPro	Integrated resource of protein domains and functional sites
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IS	Insertion sequences
K	Keto nucleotides
kbp	Kilobase pairs
kDa	Kilodalton
KEGG	Kyoto Encyclopedia of Genes and Genomes
LB	Luria-Bertani
LDS	Least significant differences
M	Amino nucleotides
mA	Milliampere
MAG	*Mycoplasma agalactiae* PG2
MAGa	*Mycoplasma agalactiae*
MALL	*Mycoplasma alligatoris* A21JP2
MARTH	*Mycoplasma arthritidis* 158L3-1
Mb	Megabases
MBOVPG45	*Mycoplasma bovis* PG45
MCAP	*Mycoplasma capricolum* subsp. *capricolum* ATCC27343
MCJ	*Mycoplasma conjunctivae* HRC/581
MCRO	*Mycoplasma crocodyli* MP145
MDR	Multidrug resistance
MEME	Multiple expectation maximization for motif elicitation
MFE	*Mycoplasma fermentans* JER
MfeM64YM	*Mycoplasma fermentans* M64
MG	*Mycoplasma genitalium* G37
MGA	*Mycoplasma gallisepticum* str. R(low)
MGAH	*Mycoplasma gallisepticum* str. R(high)
MGF	*Mycoplasma gallisepticum* str. F
MHC I	Major histocompatibility complex class I
MHC II	Major histocompatibility complex class II
MHJ	*Mycoplasma hyopneumoniae* J
Mho	*Mycoplasma hominis* ATCC 23114
mhp	*Mycoplasma hyopneumoniae* 232
MHP7448	*Mycoplasma hyopneumoniae* 7448
MHR	*Mycoplasma hyorhinis* HUB-1
MLC	*Mycoplasma mycoides* subsp. capri LC str. 95010
MMB	*Mycoplasma bovis* Hubei-1

MMOB	*Mycoplasma mobile* 163K
MPN	*Mycoplasma pneumoniae* M129
MPNA	*Mycoplasma pneumoniae* 309
MPUT	*Mycoplasma putrefaciens* KS1
Ms01	*Mycoplasma struthionis* sp. nov. str. Ms01
Ms02	*Mycoplasma* sp. Ms02
Ms03	*Mycoplasma nasistruthionis* sp. nov. str. Ms03
MS53	*Mycoplasma synoviae* 53
MSB	*Mycoplasma leachii* PG50
MSC	*Mycoplasma mycoides* subsp. *mycoides* SC str. PG1
MYPE	*Mycoplasma penetrans* HF-2
MYPU	*Mycoplasma pulmonis* UAB CTIP
NADPH	Nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
NGS	Next-generation sequencing
nptII	Neomycin phosphotransferase II
nt	Nucleotides
OD	Optical density
Opp	Oligopeptide permease
OppA	Oligopeptide ABC transporter substrate-binding domain
OppB	Oligopeptide ABC transporter subunit, membrane component, B
OppC	Oligopeptide ABC transporter subunit, membrane component, C
OppD	Oligopeptide ABC transporter subunit, ATP-binding component, D
OppF	Oligopeptide ABC transporter subunit, ATP-binding component, F
ORF	Open reading frame
ori	Origin of replication
PANTHER	Protein analysis through evolutionary relationships
PAUP*	Phylogenetic analysis using parsimony (* and other methods)
PBS	Phosphate-buffered saline
PCR	Polymerase chain reactions
pDNA	Plasmid DNA
PEP	Phosphoenolpyruvate
Pfam	Protein families
PfamB	Protein families (automatically generated)
pFunc	Prokaryotic protein functional prediction
pgk	Phoshoglycerate kinase
pH	Decimal cologarithm of hydrogen ions
PIRSF	Protein information resource superfamily
PRED-LIPO	Prediction of lipoprotein signal peptides
ProDom	Protein domains
PROKAR_LIPOPROTEIN	Prokaryotic lipoprotein
PSI-BLAST	Position-specific iterative BLAST
PSORTb	Protein subcellular localization prediction of bacteria

PTS	Phosphoenolpyruvate-dependent sugar phosphotransferase transport system
R	Purine
RAST	Rapid Annotations using Subsystem Technology
RAxML-HPC2	Randomized axelerated maximum likelihood for high performance computing 2
ROS	Reactive oxygen species
rRNA	Ribosomal RNA
SAP	Shrimp alkaline phosphatase
SCOP	Structural classification of protein
SDM	Site-directed mutagenesis
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SignalP	Signal peptide
SMART	Simple modular architecture research tool
smc	Structural maintenance of chromosome proteins
sp. nov.	Species nova
SSRs	Simple sequence repeats
str.	Strain
SV40	Simian virus 40
subsp.	Subspecies
TAIL	Thermal asymmetric interlaced
tBLASTn	BLAST search in a translated nucleotide database using a protein query
tBLASTx	BLAST search in a translated nucleotide database using a translated nucleotide query
TBR	Tree bisection/reconnection
T-Coffee	Tree-based consistency objective function for alignment evaluation
T_m	Melting temperature
TMHMM	Transmembrane Protein Topology with a Hidden Markov Model
TRIS-HCl	2-Amino-2-hydroxymethyl-propane-1,3-diol hydrochloride
tRNA	Transfer ribonucleic acid
TPA	Tissue plasminogen activator
U	Uracil
UDP	Uridine diphosphate
UMP	Uridine monophosphate
UTP	Uridine triphosphate
X-Gal	5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside
XSEDE	Extreme science and engineering discovery environment
Y	Pyrimidine

Table of Contents

Chapter 1 Introduction

Struthio camelus, commonly known as the ostrich, is the largest living flightless bird in the world. Although endemic to Africa and Saudi Arabia, ostriches are farmed around the world for their meat, leather and feathers. South Africa is the world's largest producer of ostrich products and contributes to about 75% of the global ostrich market. In South Africa, ostrich farms contribute significantly to the economy, with the average gross value amounting to approximately R1.5 billion annually and provide about 20 000 jobs in rural areas. Ostrich farming requires a dry climate making the arid Klein Karoo region in the Western Cape Province of South Africa ideal. Approximately 75% of the South African ostriches are found in the Klein Karoo region (DAFF 2014). However, the South African ostrich industry is largely dependent on the export of ostrich products. This makes it particularly vulnerable to internationally notifiable diseases such as avian influenza (AI). The industry has therefore invested in research into ostrich diseases that could lead to a reduction in international embargoes and production losses.

An increased demand for ostrich meat has resulted in ostriches being reared under intensive farming conditions. Under these conditions ostriches are more exposed to diseases. In commercial ostrich farming, livestock is frequently moved since the eggs are hatched on a breeding farm or hatchery and day-old chicks (less than 72 hours) are moved to a second location where the chicks are reared for about three months (Figure 1.1). The three-month-old ostriches are moved to adult rearing farms with larger enclosures to accommodate the increase in bird size. At nine to 12 months of age, ostriches are moved again to a quarantine area for a minimum of 14 days before slaughtering. This quarantine period is required by South African regulations and was implemented in response to the repeated outbreak of AI to ensure that ostrich meat is disease free.

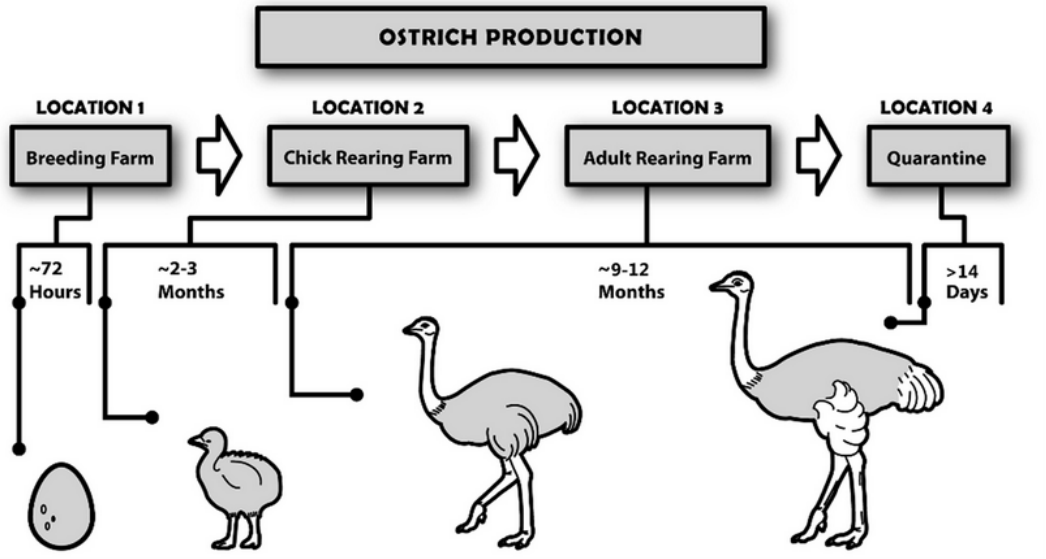


Figure 1.1 Ostrich production in South Africa (Moore et al. 2014).

In South Africa, AI is a notifiable disease and ostrich flocks are monitored bi-annually as well as before movement and slaughter as part of an official serological surveillance programme (Abolnik et al. 2013). When ostriches test positive for antibodies against AI, a stamping-out approach is followed and all ostriches on the farm are slaughtered, while all farms in a 10 km radius from the outbreak are quarantined. Export of fresh ostrich meat is immediately halted and the quarantine is only lifted once the farms in the quarantine area test negative in two rounds of testing, 28 days apart (Van Helden et al. 2012). South Africa implemented these measurements in 2013 in response to the repeated outbreaks of AI stating in 2004. The 2011 outbreak of AI had a major impact on the industry and meat production decreased from 9 000 tons in the 2009/2010 financial year to 3 000 tons in the 2012/2013 financial year (DAFF 2014) mainly as a result of the international embargo on ostrich meat.

Although AI has had a very negative short term effect on the ostrich industry, an almost more serious long term problem in the intensive rearing of ostriches has been the incidence of mycoplasma infections. Mycoplasma infections have been implicated in ostrich chick mortalities, growth retardation and downgrading of carcasses (Botes 2004). Mycoplasmas are known to be host-specific and in 2005, three unique *Mycoplasma* species were identified in South African ostriches (Botes et al. 2005a). These were provisionally named Ms01, Ms02 and Ms03. Ms01 and Ms03 have since been isolated from Namibian ostriches as well and have been described as *Mycoplasma struthionis* sp. nov. and *Mycoplasma nasistruthionis* sp. nov., respectively, but have not been formally published (Langer 2009).

Mycoplasmas infect the mucosal membranes of the ostrich respiratory tract causing eye, nose and air sac infections (Botes et al. 2005b). These infections can be treated with antibiotics, but this is not always successful, since carrier conditions exist (infected ostriches without symptoms). Currently a combination of biosecurity practice and antibiotic treatment is used to control mycoplasma infections. As yet, there are no commercial vaccines available for the prevention of mycoplasma infections in ostrich. Chicken mycoplasma vaccines have proved to be ineffective (Pretorius 2009). The development of whole organism vaccines is hampered by the fact that these mycoplasmas are slow growing and require complex media for growth (Langer 2009), making such vaccine strategies ineffective and expensive. DNA vaccines present an attractive alternative to vaccine development for numerous reasons such as, economic large scale production and that the antigen is produced within the host's own cells (Mahoney et al. 2000; Liu 2011). Recent developments in genome sequencing techniques provide a unique opportunity to characterize the genes of bacteria and *inter alia* mycoplasmas. This aids in the development of DNA vaccines by providing a repertoire of genes that can be used to select vaccine candidate genes.

The overall objective of this study was to investigate the development of DNA vaccines against *M. nasistruthionis* sp. nov. infections in South African ostriches. The genome of

M. nasistruthionis sp. nov. str. Ms03 was sequenced and annotated to allow for a better understanding of its metabolic capacity as well as for the identification of the vaccine candidate gene, *oppA*. After cloning of the gene into appropriate DNA vaccine vectors, the ability of these DNA vaccines to elicit an immune response was evaluated.

In Chapter 2 a literature overview of the genomic characteristics and essential processes in mycoplasmas is given. The second part of the literature overview focuses on the prevention of mycoplasma infections with the main focus on DNA vaccines.

The first aim of this study was to expand the knowledge of *M. nasistruthionis* sp. nov. str. Ms03 by determining the genome sequence, annotating the genes and evaluating the metabolic pathways of essential processes. The second aim was to identify and characterize the *oppA* gene as potential vaccine candidate gene. The results from these analyses are presented in Chapter 3.

The third aim was to investigate the prevalence of the *oppA* gene within all *Mycoplasma* species in order to assess its potential as a vaccine candidate gene. The results are presented in Chapter 4 as a published manuscript (Gene, 2015, Volume 558, Issue 1, pages 31-40) and are preceded by background information on certain aspects of the experimental design.

The fourth aim was to develop three DNA vaccines containing the *M. nasistruthionis* sp. nov. str. Ms03 *oppA* gene and to evaluate their potential to elicit an immune response in ostriches. The results from the DNA vaccine trial are presented as an independent manuscript in Chapter 5. This is once again preceded by background information on certain aspects of the experimental design and the negative influence an outbreak of AI had on it.

The objectives of the different parts of this study are described in the introduction of the respective chapters. The dissertation ends with an overall conclusion in Chapter 6. The references for all the chapters are given at the end of Chapter 6 and relevant supplementary data are presented in Appendix 1 to 5.

Chapter 2 The mycoplasma genome: Implication of DNA vaccine design

2.1 Introduction: Mycoplasmas in the genomic era

With genome sequencing becoming an everyday technique, the question of what we have learnt from the increasing amount of available sequencing data has become more and more important. Being the self-replicating organisms with the smallest genomes, mycoplasmas have being given considerable attention in this genomic era and a total of 91 genomes from 47 species in the class *Mollicutes* have been completely sequenced and annotated up to June 2015 (Appendix 1). The first bacterial genome, *Mycoplasma genitalium*, was sequenced in 1995 (Fraser et al. 1995) and is the smallest genome of all self-replicating organisms. Additionally, a mycoplasma genome has also served as the blueprint for the first synthetic organism (Gibson et al. 2010). Mycoplasma research aims to understand the genomic organization and composition of the mycoplasma cell. If we were to understand the smallest and "simplest" genome, it may aid in understanding more complex cell systems as well as aiding in the treatment and prevention of mycoplasma bacterial infections.

Mycoplasmas are parasitic or commensal bacteria that are of medical and agricultural importance since they are associated with a number of diseases in humans and animals. Mycoplasmas belong to the phylum *Firmicutes* and the class *Mollicutes*, 'mollis' meaning soft and 'cutes' meaning skin, reflecting the lack of cell walls in this class (Razin et al. 1998) (Table 2.1). They therefore stain negative in the Gram stain test, although they have evolved from low guanine and cytosine (G+C) containing Gram-positive bacteria through a reductive evolutionary process with their closest relatives being *Lactobacillus*, *Bacillus* and *Clostridium* (Woese et al. 1980; Weisburg et al. 1989; Wolf et al. 2004). All of the members in the order *Mycoplasmatales* require exogenous sterol for maintenance of the bacterial cell membrane

Table 2.1 Classification of the class *Mollicutes*

Phylum	*Firmicutes*		
Class	*Bacilli* *Clostridia* *Mollicutes*		
Class	**Order**	**Family**	**Genus**
Mollicutes	*Mycoplasmatales*	*Mycoplasmataceae*	*Mycoplasma* *Ureaplasma*
	Acholeplasmatales	*Acholeplasmataceae* *Incertae sedis*	*Acholeplasma* Phytoplasma *Candidatus*
	Anaeroplasmatales	*Anaeroplasmataceae*	*Anaeroplasma* *Asteroleplasma*
	Entomoplasmatales	*Spiroplasmataceae* *Entomoplasmataceae*	*Spiroplasma* *Entomoplasma* *Mesoplasma*

Reference: Razin et al. (1998); Wolf et al. (2004); Brown (2010).

which is reflected by the absence of these molecular pathways within their genomes (Brown 2010). The genus *Mycoplasma* is divided into three phylogenetic groups based on the typical bacterial 16S ribosomal RNA (rRNA) phylogeny: the hominis, pneumoniae and spiroplasma groups as described by Wolf et al. (2004) (Figure 2.1). In general, members of the genus *Mycoplasma* are known for their low G+C content and small genome size.

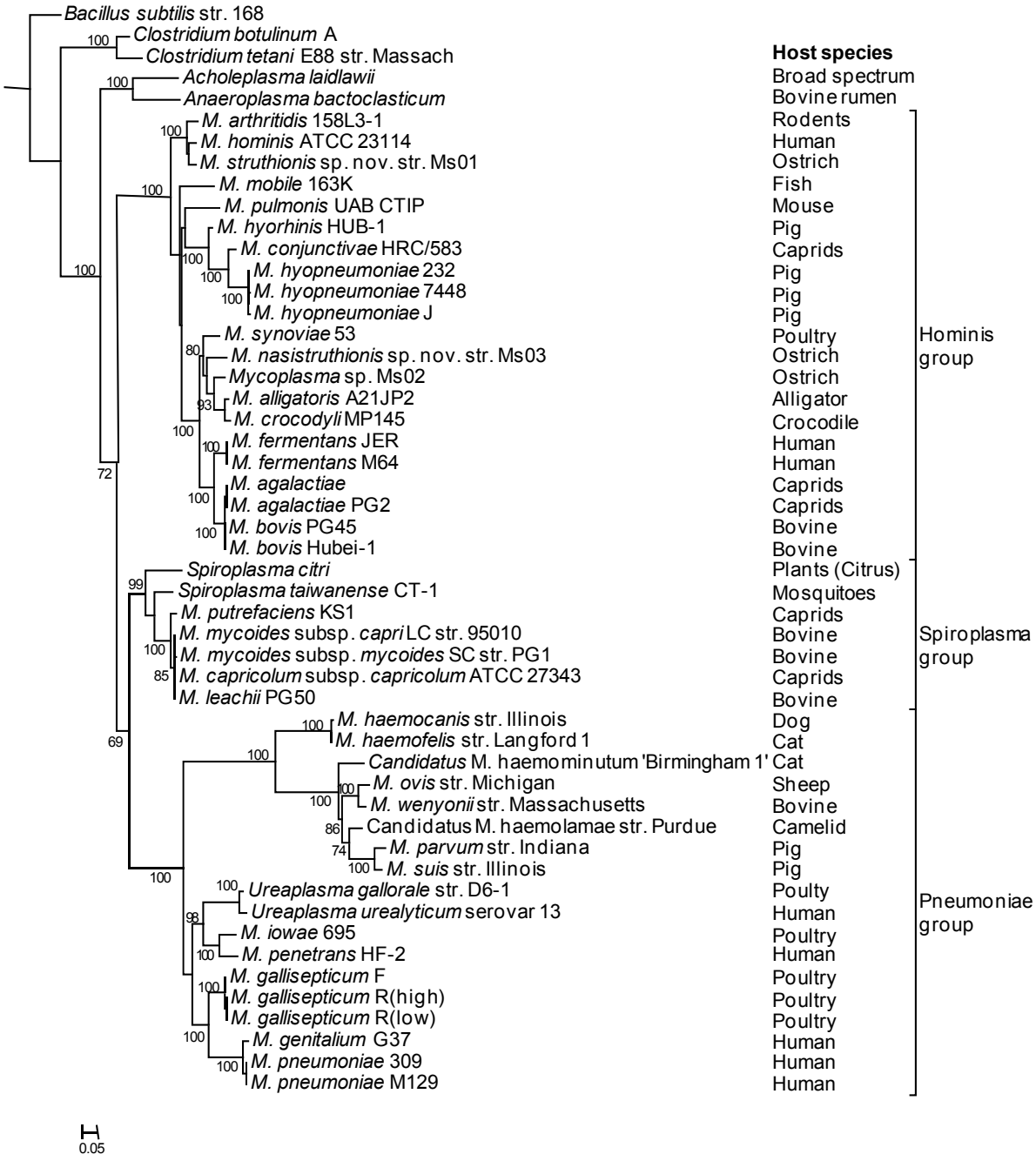


Figure 2.1 Maximum likelihood phylogeny the class *Mollicutes* based on *16S rRNA* sequences. Bootstrap values (≥70%) are indicated above or below nodes. The columns on the right-hand side indicate the hosts for each *Mycoplasma* species as well as the phylogenetic group (hominis, pneumoniae and spiroplasma).

Genome size appears to be an indication of an organism's life strategy, with small genomes associated with a parasitic life style (Moran 2002), allowing them to acquire complex metabolic building blocks from their host rather than synthesizing them. This is reflected in mycoplasma culturing studies in which a wide spectrum of substrates and factors is required for *in vitro* growth, with some species requiring undefined media components such as serum while other species are characterized as unculturable (Waites and Talkington 2004; Guimaraes et al. 2011; Flores-Medina et al. 2012; Citti and Blanchard 2013; Bueno et al. 2014).

The trivial name mycoplasma was previously used as a group term for all members in the class *Mollicutes,* however, "Bergey's Manual® of Systematic Bacteriology" reports that the more appropriate term should be mollicute(s) so that the trivial name mycoplasma can be devoted to species within the genus *Mycoplasma* (Brown 2010). In this dissertation, the trivial name mycoplasma will be used to refer to species belonging to the genus *Mycoplasma.*

The first part of this literature review investigates the genomes of mycoplasma with regard to genome characteristics, essential cellular pathways as well as pathogenicity. The second part focuses on vaccine design, in particular DNA vaccines.

2.2 Organization of mycoplasma genomes

Mycoplasmas have typical bacterial single, small circular chromosomes that range in sizes from 564 kilobase pairs (kbp) to 1 359 kbp with a G+C content between 24% and 40% (Table 2.2). Information in mycoplasma genomes is condensed with 92.54% to 99.10% (gene density) of the genome being transcribed to RNA (Table 2.2). As in other bacteria, genes in mycoplasmas frequently overlap (Fukuda et al. 1999; Fonseca et al. 2014). Most gene overlaps are short (less than 60 nucleotides), terminal unidirectional overlaps and the result of deletions resulting in the elongation of the 3' end of the genes. These overlaps are not associated with the reduction in mycoplasma genomes size (Fukuda et al. 1999; Fukuda et al. 2003).

Within bacteria, genes are organised into groups that are arranged in tandem in the genome (Mao et al. 2009; Yin et al. 2010). These groups are typically called operons and share common promoter and terminator sequences that imply that they are co-transcribed into single RNA molecules (Simons et al. 1987; Siqueira et al. 2011). Perhaps the best known example of operon gene organization in bacteria is the rRNA cluster i.e. *16S, 23S* and *5S rRNA* genes including their intergenic spacers. The largest polycistronic operon in *Mycoplasma pneumoniae* contains 20 genes but most operons only encode two or three genes (Guell et al. 2009). It is estimated that as little as 5% of all the genes in *Mycoplasma*

Table 2.2 A summary of the complete genome sequences within the class *Mollicutes* and their features

Genus *	**Number of Species**	**Number of Genome#**	**Genome Size (kbp)**	**Gene Count**	**G+C fraction**	**Gene density (%)**	**rRNA Count**	**tRNA Count**
Acholeplasma	1	1	1 497	1 422	0.32	97.05	6	36
Asteroleplasma	No complete genome sequence available							
Phytoplasma *Candidatus*	5	6	570-960	518-1 155	0.21-0.28	92.47-96.45	6	27-35
Mesoplasma	1	2	793-825	721-768	0.27	94.59-95.44	6	29
Entomoplasma	No complete genome sequence available							
Spiroplasma	8	8	945-1 175	890-1 422	0.24-0.30	96.28-97.47	3-6	29-33
Mycoplasma	30	71	564-1 359	523-1 580	0.24-0.40	92.54-99.10	3-8	28-37
Ureaplasma	2	3	752-874	642-679	0.25-0.26	94.05-95.14	6	27-30
Class: *Mollicutes*	47	91	564-1 497	518-1 580	0.21-0.40	92.47-99.10	3-8	27-37

* For the complete list of species within the class *Mollicutes* for which a complete genome sequence is available see Appendix 1. Data was downloaded from Integrated Microbial Genomes (IMG) system (http://www.jgi.doe.gov/) on 25 June 2015.

In many cases more than one strain of the same species had a complete genome sequence available.

hyopneumoniae are transcribed in monocistronic units (Siqueira et al. 2011), illustrating the abundance and importance of the operon gene organization in mycoplasma genomes.

Genes organised into an operon are usually associated with a single metabolic pathway (Yin et al. 2010). When genes of an essential metabolic pathway are organised into more than one operon, these operons will normally cluster together in the genome (Guell et al. 2009; Yin et al. 2010). In bacteria (including mycoplasma), operons may, however, also contain genes of multiple pathways, which result in related operons being found further apart (Yin et al. 2010). The operons of more frequently activated pathways are closer together than less active pathways. This arrangement was proposed to reduce effort associated with finding, transcribing and translating all of the genes of a particular pathway (Yin et al. 2010). This genomic organizational level i.e. the operons, contributes to the regulation of protein expression, resulting in similar stoichiometry among proteins of the same pathway (Rocha 2008). Depending on particular environmental conditions, a gene may be transcribed as part of more than one or different operons which is called alternative transcription. Alternative transcription is possible due to the presence of internal promoters and/or termination sequences within the transcribed sequence (Guell et al. 2009) and plays an important role in the adaption to a changing environment.

Since replication and transcription, two fundamental processes, take place on the chromosome, the organization of genes and operons will have an effect on these processes. Studies in bacteria have shown that when transcription interrupts or stalls replication, it can contribute negatively towards survival and reproduction (Srivatsan et al. 2010; Lin and Pasero 2012). Genes therefore tend to be transcribed in the same direction as the leading strand, preventing “head-on” collisions of replication and transcription processes (Price et al. 2005). Additionally frequently activated pathways and essential genes tend to be located on

the leading strand near the origin of replication to prevent the transcription of these genes to disrupt or stall replication. This allows for transcription during replication. In mycoplasmas, gene organization shows a strong bias towards the leading strand (Sirand-Pugnet et al. 2007a). About 80% of all genes in *M. genitalium* are located on the leading strand (Rocha and Blanchard 2002; Lin and Zhang 2011).

On an evolutionary time scale, mycoplasmas are located on some of the longest branches of the bacterial tree of life implying that as a group they evolve at a faster rate (Woese et al. 1984; Ciccarelli et al. 2006; Sirand-Pugnet et al. 2007a). Genome "changes" are attributed to point mutations, genome rearrangement, gene deletions and horizontal gene transfer (HGT). Mycoplasmas have some of the highest rates of base substitution mutations of all unicellular organisms (Delaney et al. 2012; Sung et al. 2012). It has also been suggested that genome rearrangements in mycoplasmas occur more frequently than previously thought (Marenda 2014). Large genome rearrangements have been observed within species by comparing the genomes of *M. hyopneumoniae* str. 232 to the J and 7448 strains (Vasconcelos et al. 2005) and also between species by comparing the genomes of *Mycoplasma bovis* Hubei-1 with *Mycoplasma agalactiae* PG2 (Li et al. 2011). Genome rearrangements contribute to the lack of gene order conservation among mycoplasmas (Sirand-Pugnet et al. 2007a).

Mycoplasma genomes are subjected to significant gene decay (deletions) and the remnants of genes (pseudogenes) are often found in mycoplasma genomes. As much as 12.99% of the *M. pneumoniae* FH genome consists of pseudogenes while 5.91% of the *M. bovis* PG45 genome consists of pseudogenes (Marenda 2014). This is in agreement with the hypothesis that mycoplasmas have evolved through a reductive evolutionary process from *Lactobacillus*, *Bacillus* and *Clostridium.* These events of gene deletions have played a significant role in the reduction of the genome size.

HGT is a mechanism whereby genes or DNA regions are transferred from one bacterium to the next (Pereyre et al. 2009; Marenda 2014). A shared environment between co-infecting species can lead to the acquisition of new genes from one another. This is viewed to be a strategy to increase the gene pool for better adaption to environmental changes. In mycoplasmas, HGT has been postulated within human (Pereyre et al. 2009), ruminant (Sirand-Pugnet et al. 2007b) and chicken (Vasconcelos et al. 2005) mycoplasmas. Furthermore, a recent study has proved the concept under laboratory conditions (Dordet Frisoni et al. 2013).

Both genome rearrangement and HGT are ascribed to the activity of mobile genetic elements. Mobile genetic elements are segments of DNA that encode for proteins that mediate the translocation of DNA regions. Mobile genetic elements in mycoplasmas include plasmids, bacteriophages, conjugative transposons and insertion sequences (IS) (Marenda 2014).

Plasmids are rarely found in mycoplasmas and have only been reported in some species, such as *Mycoplasma capricolum* subsp*. capricolum, Mycoplasma cottewii, Mycoplasma leachii, Mycoplasma mycoides* subsp*. capri* and *Mycoplasma yeatsii* (Thiaucourt et al. 2011; Breton et al. 2012; Kent et al. 2012; Marenda 2014). These circular DNA molecules range from 1 kbp to 2 kbp in size and belong to the rolling circle replication plasmid family pMV158 (Breton et al. 2012). They have similar organizations: two unidirectional open reading frames (ORFs) with the proteins of the one functioning in the replication of the plasmid and the other controlling the copy number of the plasmid in the cell (Thiaucourt et al. 2011; Breton et al. 2012). The only exception (in the genus *Mycoplasma*) reported thus far is the 3.4 kbp *M. yeatsii* plasmid, pMyBK1, that belongs to a novel rolling circle replication plasmid family (Kent et al. 2012). The pMyBK1 plasmid has two unidirectional ORFs, one may function in the replication of the plasmid but the exact functions of the two proteins are speculative (Kent et al. 2012). The retention of these plasmids within mycoplasma cells, organisms with such drastically reduced genome sizes, implies a significant role in adaptation and survival. The role of retaining plasmids remains to be identified (Breton et al. 2012).

As with plasmids, bacteriophages (prokaryotic viruses) are rarely found in mycoplasmas but have been reported in *Mycoplasma pulmonis* str. UAB 6510, *Mycoplasma arthritidis*, *Mycoplasma fermentans* and *Mycoplasma hyorhinis* (Tu et al. 2001; Röske et al. 2004; Marenda 2014). These bacteriophages have the ability to integrate into the mycoplasma genome and can contribute to genome rearrangements. The insertion of the bacteriophage into the genome may disrupt some gene functions. Interestingly in *M. arthritidis,* bacteriophages have been reported to convey virulence (Voelker et al. 1995; Marenda 2014) which is the same as in other bacteria (Penadés et al. 2015).

Conjugative transposons, or integrated conjugative elements (ICEs), are clusters of genes that can translocate within and between genomes (Dordet Frisoni et al. 2013; Marenda 2014). ICEs encode genes that assist in their excision, conjugative transfer and integration into the recipient genome (Dordet Frisoni et al. 2013). Mycoplasma ICEs are about 20 kbp to 30 kbp long, encode for about 20 genes, occur in multiple copies within genomes and have been implicated in horizontal gene transfer (Dordet Frisoni et al. 2013; Marenda 2014).

IS are short DNA elements that code for proteins with transposition activities such as transposases (Loreto et al. 2007; Lysnyansky et al. 2009; Marenda 2014). The activities of these proteins allow the insertion sequence to move within and between genomes. IS are widely found in bacteria (including mycoplasmas) and influence the genome by contributing to horizontal gene transfer and genome rearrangement that can result in acquiring new genes, deletions, insertions or gene amplification (Loreto et al. 2007; Lysnyansky et al. 2009). In the genome of *Mycoplasma mycoides* subsp. *mycoides* SC*,* the insertion sequence IS1634, which is 1 872 base pairs (bp) long, is found 60 times, illustrating the

occurrence of this element within mycoplasma genomes (Vilei et al. 1999; Westberg et al. 2004).

In addition to the above, simple sequence repeats (SSRs) are also found within mycoplasma genomes. SSRs or microsatellites are mono- to hexanucleotide tandem repeats (Trivedi 2010) and can play a role in antigenic variation (Citti et al. 2010).

Even in these small genomes, the contribution of repetitive sequences cannot be ignored since they lead to genome rearrangement and horizontal gene transfer as well as contributing to the genome size (Marenda 2014). Repetitive regions account for 21% of the *M. fermentans* M64 genome (Shu et al. 2011) and 29% of the *M. mycoides* subsp. *mycoides* SC genome (Westberg et al. 2004). The contribution of mobile and repetitive elements to the fluency of mycoplasma genomes might explain the ability of mycoplasmas to adapt to environmental changes and to new hosts.

2.3 Essential pathways in mycoplasma

The reduction in the mycoplasma genome size has lead to the loss of several metabolic pathways while preserving others, leaving these organisms with the barely essential genes to support life. Within these small genomes, the number of predicted genes ranges from 523 to 1 580 genes (Table 2.2). Through transposon mutagenesis studies, it was determined that 387 of the 482 protein-encoding genes and all 43 of the structural RNA genes in *M. genitalium* are essential (Hutchison et al. 1999; Glass et al. 2006). In similar studies, 271 of the 782 protein-encoding genes were identified as essential in *M. pulmonis* (French et al. 2008; Dybvig et al. 2010) while in *M. arthritidis*, 417 of the 635 protein-encoding genes were found to be essential (Dybvig et al. 2008). Whether or not a gene is essential will depend on its environmental or growth conditions, implying that under a particular condition, a gene may be essential whereas under another, it may be non-essential.

Some genes should be essential under all circumstances (fundamental for life) leading some authors to search for what are referred to as core-essential genes. Core-essential gene studies in mycoplasma have lead to the identification of between 153 to 196 genes that are conserved across *Mycoplasma* species' genomes (Lin and Zhang 2011; Liu et al. 2012). These genes are generally involved in critical cellular processes such as genome replication, transcription and translation (Lin and Zhang 2011). Mycoplasmas are self sustaining thus they have the ability to replicate the genome, duplicate the cell, transcribe RNA and translate protein independently of their hosts. Energy metabolism pathways are used to produce the energy necessary for cell maintenance and growth, while transporters acquire the necessary building blocks for these processes from the environment. These fundamental processes will be reviewed in the next sections.

2.3.1 Genome replication and cell division

During genome replication the circular DNA chromosome of mycoplasmas is duplicated. The process of DNA replication is complex and includes a number of steps starting with site specific initiation at the origin of replication (*ori*), followed by the unwinding of the supercoiled double stranded genomic DNA and the assembly of the replication machinery before the DNA is replicated bidirectionally and finally terminated at the termination site or when the two replication forks reach each other.

The *ori* is a well defined site where the replication forks open the parent DNA double helix. Bacterial genes in the vicinity of the *ori* tend to be conserved with regard to gene order as well as orientation. In most mycoplasmas, the *ori* is located in the vicinity of the *dnaA* and *dnaN* genes. In genome annotation, the *ori* is numbered as the first base of the circular chromosome and the ORFs would be numbered from the *dnaA* gene onwards. The *ori* has been determined experimentally for several mycoplasmas (*M. mycoides* subsp. *mycoides* LC str. Y-goat, *M. mycoides* subsp. *mycoides* SC str. PG1, *M. capricolum* subsp. *capricolum* str. California kid and *M. pulmonis* UAB CTIP and *M. hyopneumoniae*) (Cordova et al. 2002; Lartigue et al. 2003; Maglennon et al. 2013). DnaA proteins initiate replication by binding to the DnaA box sequences at the ori, which then causes the supercoiled DNA to unwind locally. DnaA boxes are 9 bp regions that are rich in adenine and thymine (A+T). The consensus sequence of mycoplasma DnaA boxes is more relaxed than in *Escherichia coli* (Cordova et al. 2002). Up to 10 putative DnaA boxes were identified in *Mycoplasma* species, however, for some mycoplasmas bioinformatic analysis could not identify any DnaA boxes (Cordova et al. 2002; Sasaki et al. 2002; Lartigue et al. 2003; Lee et al. 2008; Pereyre et al. 2009).

The replication machinery of mycoplasmas is similar to that of *Bacillus subtilis*. The leading strand is synthesized continuously while the second strand, the lagging strand, is synthesized semidiscontinuously. The leading strand requires 11 proteins to replicate while the lagging strand requires the same 11 proteins as well as DnaE and primase (Sanders et al. 2010). These different replication modes of the leading and lagging strands result in different mutation rates (Engelen et al. 2012) and as a consequence, the leading strand is rich in guanine (G) and thymine (T) while the lagging strand is rich in adenine (A) and cytosine (C) (Marín and Xia 2008). A shift from a G+T rich region to a C+A rich region is frequently observed at the *ori* in bacteria species and a GC plot can therefore be used to predict the *ori* in bacteria. This technique has been used to identify the *ori* in mycoplasma genomes with mixed success (Sirand-Pugnet et al. 2007a) most probably because mycoplasma genomes are generally A+T rich and genome rearrangement occurs frequently.

In bacteria, replication is terminated when the replication forks progressing in opposite directions meet each other or when the replication forks reach a termination site (Kono et al.

2012). Specific termination sites or the involvement of termination proteins such as Tur protein in *E. coli* and RTP protein in *B. subtilis* have not been reported in all *Firmicutes* (Kono et al. 2012). In most bacteria, replication is symmetrical implying that the termination and *ori* is 180° from each other on the circular chromosome, but it has been reported that for slow growing bacteria such as mycoplasma, replication may be asymmetrical. As with the *ori*, the termination site can be predicted with a GC plot (Zheng and Liu 2008).

Mycoplasma cells divide by binary fission and have a low number of conserved genes involved in cell division (Alarcon et al. 2007; Fisunov et al. 2011). Compared to *B. subtilis* that has 17 genes in the "cell division and cell wall synthesis" gene cluster, mycoplasma genomes only have *mraZ, mraW, ftsZ* and one gene encoding a hypothetical protein (Alarcon et al. 2007; Lluch-Senar et al. 2010). These four genes are usually but not always found (Alarcon et al. 2007; Lluch-Senar et al. 2010). The *ftsZ* gene is believed to play a central role in cell division by binary fission but has recently been shown to be non-essential in mycoplasma (Lluch-Senar et al. 2010). In the *Mycoplasma mobile* genome the *ftsZ* gene was not found (Jaffe et al. 2004) and in *M. genitalium a ftsZ* null mutation was viable (Lluch-Senar et al. 2007), which implies that an alternative mechanism of cell division can occur without the FtsZ protein. Subsequently genes involved in movement have been implicated in cell division (Hatchel and Balish 2008; Erickson and Osawa 2010; Lluch-Senar et al. 2010). It was proposed that this alternative mechanism can function on its own but Erickson and Osawa (2010) suggested that the two mechanisms work together for more effective cell division.

2.3.2 RNA transcription and protein translation

Transcription is initiated by the binding of the RNA polymerase to the promoter region. The mycoplasma promoter regions resemble those of the standard σ70 promoter region with the transcription start within 100 bp upstream of the translational start site in transcribed regions (Guell et al. 2009; Weber et al. 2012; Siqueira et al. 2014). In *M. hyopneumoniae*, promoters consist of a σ70 -10 promoter element and a -16 element, also found in the low G+C Gram-positive bacteria (Weber et al. 2012). The -35 promoter element was not identified which is typical for low G+C bacteria. Intergenic regions have also been found to transcribe and are the result of improper transcription termination as well as the initiation of transcription within these regions (Gardner and Minion 2010). The functions of these intergenic regions remain to be elucidated. Transcription is either terminated by the formation of a termination hairpin that tightly regulates the process (Guell et al. 2009) or by the gradual release of RNA polymerase from the DNA after the last gene is transcribed (Gardner and Minion 2010).

Transcription and translation in bacteria may occur simultaneously since these two processes are not compartmentalised in bacteria. Translation relies on the complex interaction of multiple proteins and RNA molecules as described by Grosjean et al. (2014).

The total number of genes involved in translation for *Mycoplasma* species ranges from 116 to 167 of which 104 are conserved in all mycoplasmas (Grosjean et al. 2014). These include genes encoding for ribosomal proteins (49), rRNA modification (4), ribosome assembly and protein maturation (8), RNA processing (4), tRNA modification (6), tRNA aminoacylation (20) and translation factors (13).

2.3.3 Energy metabolism

Glycolytic mycoplasmas use glycolysis as their main energy producing pathway while non-glycolytic mycoplasmas utilise arginine as energy source. In glycolytic mycoplasmas, the complete glycolysis pathway (Embden-Meyerhoff-Parnas pathway) that converts glucose to pyruvate is present (Halbedel et al. 2007; Shu et al. 2012). Transporters and enzymes for connecting reactions to utilize alternative substrates are also present. Mycoplasmas can use D-glucose, fructose, glycerol, mannitol and glycerol-3-phosphate that lead into this pathway as energy sources (Halbedel et al. 2007; Shu et al. 2012). Both the citric acid cycle and pentose phosphate pathways are however partly or completely absent in mycoplasmas. Pyruvate is therefore degraded further into acetate, acetyl-coenzyme A or lactate (Halbedel et al. 2007; Guimaraes et al. 2011; Shu et al. 2012).

In non-glycolytic mycoplasmas, ATP is provided by the arginine deiminase pathway encoded by the *arcA* (arginine deiminase), *arcB* (ornithine carbamoyltransferase) and *arcC* (carbamate kinase) genes as well the arginine transporter genes (Rechnitzer et al. 2011). Examples of mycoplasmas that use arginine as their major energy source are *Mycoplasma hominis* (Pereyre et al. 2009) and *M. arthritidis* (Dybvig et al. 2008). In non-glycolytic mycoplasmas some of the enzymes of glycolysis can be absent as is the case in *M. arthritidis* where hexokinase and phosphofructokinase are absent (Dybvig et al. 2008).

All mycoplasmas have a typical F_1F_0 ATP synthase that consists of eight conserved genes found within an operon (Beven et al. 2012). Unlike mitochondrial ATP synthase of which the main function is the generation of ATP, the mycoplasma ATP synthase is thought to be involved in ATP hydrolysis and maintenance of the electrochemical gradient since mycoplasma lack a cytochrome-containing electron transport chain (Rechnitzer et al. 2011). In the absence of a cell wall, mycoplasmas depend on mechanisms such as this to maintain and regulate the osmotic balance between the external environment and intracellular space.

2.3.4 Acquiring the necessary building blocks to maintain life

Mycoplasmas have limited metabolic capabilities since the massive reduction in genome size has lead to the loss of many pathways, therefore nutritional building blocks need to be imported into the cell. Pathways evicted include *de novo* synthesis of purine and pyrimidines, some amino acids, Coenzyme A, cholesterol, lipids and cell wall synthesis (Razin et al. 1998; Arraes et al. 2007; Bizarro and Schuck 2007). In addition to the above,

energy sources such as glucose, fructose, glycerol and arginine are commonly required for maintenance and growth. Mycoplasmas need to scavenge these nutrients from the environment and therefore require a number of transporters to transport these molecules into the cell. About 10% of the genes within mycoplasma genomes code for transport proteins and of these ATP-binding cassette systems (ABC) are the most prominent (Nicolás et al. 2007).

ABC transporters or traffic ATPases are conserved active membrane transport proteins that consist of four domains: two transmembrane domains that form a pore through which the substrate is transported and two cytosolic ATP-binding domains that hydrolyse ATP to provide the energy for the translocation (Berntsson et al. 2010). Additionally, ABC importers that are unique to plants, bacteria and archaea, have an extracellular substrate-binding domain (Berntsson et al. 2010; Rice et al. 2014). ABC transporters include the glycerol importer (*gtsABC*), putative glycerol-3-phosphate (*ugpACE*) importer, polyamine (putrescine/spermidine) importer (*potABCD*) and the oligopeptide permease importer (*oppABCDF*) (Vilei and Frey 2001; Pilo et al. 2005; Nicolás et al. 2007; Großhennig et al. 2013; Szczepanek et al. 2014).

Transporters other than ABC transporters include the glycerol uptake facilitator, glucose permease and the phosphoenolpyruvate-dependent sugar phosphotransferase transport system (PTS) that is responsible for import of glucose, fructose and mannitol (Halbedel et al. 2007).

2.4 Adaption towards living as a parasite

A successful pathogen possesses the ability to enter its host, reach the target tissue and adhere to it, while evading the host's immune system in order to grow, multiply and be transmitted to new hosts (Bradbury 2005; Pilo et al. 2007). With the reduced coding capacity of their small genome, mycoplasmas are remarkably well adapted to their hosts. *Mycoplasma* species have evolved to successfully live and multiply within numerous hosts. They have evolved the ability to adhere to their hosts through numerous surface proteins and are able to avoid the host's immune system through surface antigen variation and phase switching. These pathogens are transmitted through inhalation, direct and indirect contact (Frey 2002; Batista et al. 2004; Faustino et al. 2004; Fox et al. 2005). In birds, mycoplasmas can also spread from one generation to the next through eggs (vertical transmission) (Bradbury 2005) while some mammal mycoplasmas be transmitted through the milk (Fox et al. 2005). Factors such as environmental stresses (e.g. weather conditions and extreme temperatures), commercial farming practices (high stocking densities) as well as genetic factors (susceptibility of animals to infections), all contribute to the intensity and transmission rate of these infections.

2.4.1 Colonization

Mycoplasmas generally infect the mucosal membranes of the eyes, nose, respiratory- and urogenital tract as well as causing joint inflammation. Some species can, however, enter the host's cells (Fürnkranz et al. 2013; Hegde et al. 2014) and others are blood born parasites (hemoplasma) (Messick 2004; do Nascimento et al. 2012). Adhesion of a bacterium to the host cell is the first step towards colonization. Adhesion ensures close contact between the mycoplasma and the host cell for effective import of nutrients (Fürnkranz et al. 2013). A number of surface proteins have been implicated in adhesion in mycoplasmas but these are only conserved in closely related strains and species (Browning et al. 2011).

2.4.2 Avoiding the host's immune system: Antigenic, size and phase variation

Mycoplasmas face the constant challenge of recognition by the host's immune system. To effectively avoid detection, bacteria can create hypervariable surface molecules through phase and antigenic variation. The variation events cause a highly dynamic repertoire of the major immunodominant surface proteins. This creates a flexible membrane surface that allows the mycoplasma to effectively escape the attacks of the host immune system's antibodies. Mycoplasmas have a number of sophisticated systems that are associated with phase and antigenic surface variation that have been reviewed by Citti et al. (2010) and Zimmerman (2014). Mechanisms of antigenic variation include phase variation through ON/OFF expression switches, domain variation, epitope masking and locus duplication.

Phase variation is usually the result of DNA slippage or rearrangements. The expression of a gene or gene set can be turned ON/OFF when DNA slippage occurs in the promoter region due to the presence of simple sequence repeats (SSRs) (Citti et al. 2010). In mycoplasmas, SSRs in the promoter regions of the *M. hyorhinis vlp* gene and the *Mycoplasma gallisepticum vlhA* gene cause DNA slippage that results in phase variation (Yogev et al. 1991; Liu et al. 2002). DNA rearrangements can also switch expression ON/OFF. Site specific recombination (cut and paste mechanism) can result in gene and/or promoter inversion that can alter expression. Expression of the *hsd* gene in *M. pulmonis* is regulated by the inversion of the gene (Dybvig et al. 1998) while the expression of the *vpma* gene in *M. agalactiae* is regulated by the inversion of the gene and the promoter (Chopra-Dewasthaly et al. 2008). Phase variations occur at a high rate in mycoplasmas, for example, in *M. hyorhinis* phase variation occurs at 10^{-2} to 10^{-5} events/cell/generation (Rosengarten and Wise 1990). These events are reversible and inheritable which create a highly variable surface repertoire.

In domain variation, a part of the protein sequence is altered by shuffling of the domains or altering the size. Gene recombination events can structurally change the domains within genes as is the case in the *vlhA* gene of *Mycoplasma synoviae* (Noormohammadi et al.

2000). DNA slippage associated with tandems of SSRs at the C-terminal of the protein can also cause size variation in the protein (Citti et al. 2010) as is the case in the *vsa* gene of *M. pulmonis* (Simmons et al. 1996).

When constantly expressed protein is concealed by either a second protein blocking its exposure to the immune surveillance of the host or by variation of the protein size, it is called epitope masking. In *M. hominis,* the rapidly changing non-essential P120 protein conceals the constantly expressed surface protein, P56, from the host immune system by acting as a decoy (Zhang and Wise 2001).

Additionally gene or locus duplication also contributes to variation in mycoplasmas. In the genome of the field strain *M. agalactiae* str. 5632, 23 copies of the *vpma* gene were found. In these copies both gene and locus duplication contributed to the variation (Nouvel et al. 2009).

Another interesting discovery is the presence of cysteine protease in *M. gallisepticum* and *M. synoviae.* Cysteine protease is a surface exposed protease and can digest chicken IgG into Fc and Fab fragments. This digestion may disable the chicken IgG to function properly and may play a role in host invasion (Cizelj et al. 2010). Recently an antibody-binding protein, protein M, has been discovered in *M. genitalium* that can bind IgG thereby neutralising the antibody (Grover et al. 2014). These approaches may also aid in avoiding detection. All the above mechanisms contribute to the success of *Mycoplasma* species as pathogens.

2.4.3 Pathogenesis

Not all mycoplasmas are pathogenic to the host, some live in the host without causing disease symptoms. A number of diseases in human and animals are, however, caused by pathogenic or opportunistic mycoplasmas. Symptoms associated with mycoplasmosis are rarely due to toxin production and rather a result of the host response to the infection or the depletion of nutrients (Pilo et al. 2005; Browning et al. 2014).

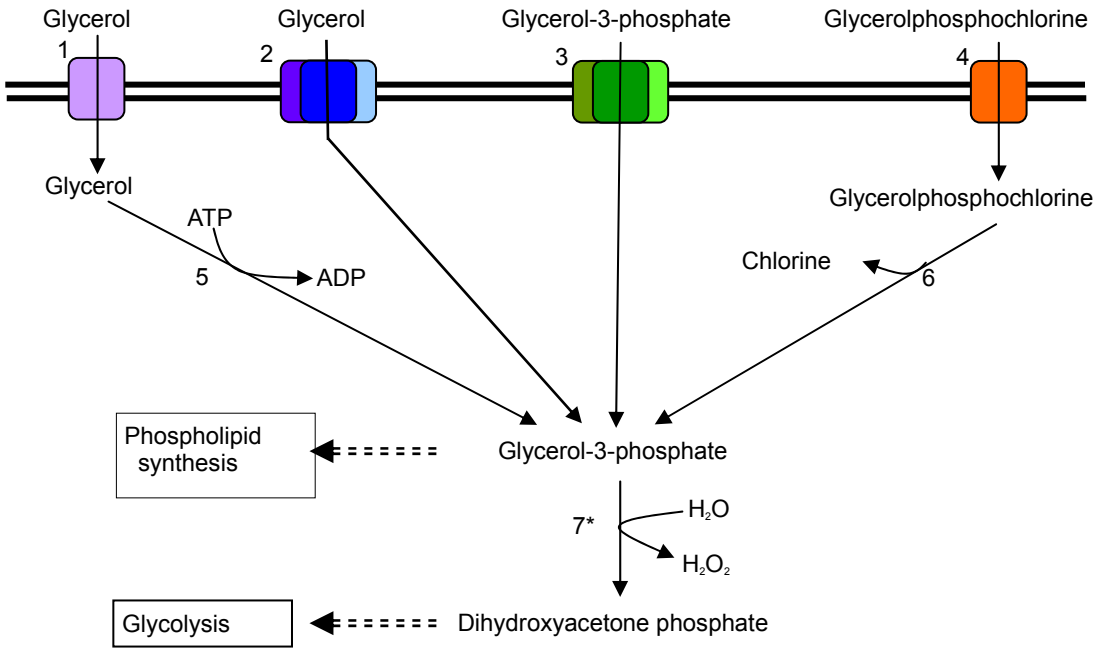
In a few cases, toxins are however reported. Community-acquired respiratory distress syndrome (CARDS) toxin is an ADP-ribosylating and vacuolating toxin produced by *M. pneumoniae*. The CARDS toxin was linked to increased mucus production (mucus metaplasia), eosinophilia, vacuolation of the bronchial and bronchiolar epithelium which is suggestive of allergic airway hyperresponsiveness. Homologues of the toxin gene, *mpn372,* were also found in *Mycoplasma penetrans* and *Mycoplasma iowae* (Medina et al. 2012; Browning et al. 2014).

The production of hydrogen peroxide (H_2O_2) in glycerol metabolism has been implicated in the virulence of *M. mycoides* subsp. *mycoides* SC, *M. gallisepticum* and *M. pneumoniae* (Figure 2.2) (Pilo et al. 2005; Großhennig et al. 2013; Szczepanek et al. 2014). The

particular enzyme implicated in this is glycerol-3-phosphate oxidase. Additionally transporters that lead to high import of glycerol, glycerol-3-phosphate and glycerolphosphochlorine also contribute to virulence of these mycoplasmas. Glycerol-3-phosphate oxidase catalyses the oxidation reaction that converts glycerol-3-phosphate and H_2O to dihydroxyacetone phosphate (DHAP) and H_2O_2. After attachment of *M. gallisepticum and M. mycoides* subsp. *mycoides* to the host cells, this enzyme acts as a transmembrane protein which releases the H_2O_2 during the conversion into the host cell. H_2O_2 and other reactive oxygen species (ROS) cause damage and inflammation in host tissue while DHAP enters the energy producing glycolytic pathway (Pilo et al. 2005). Even though the glycerol-3-phosphate oxidase is a cytosolic protein in *M. pneumoniae,* H_2O_2 production was still implicated in virulence (Schmidl et al. 2011).

OppA, the substrate-binding domain of the oligopeptide permease transporter may also play a role in pathogenicity of *M. hominis*. In addition to its role in oligopeptide transport, the OppA protein is also involved in cytoadhesion and is the major ecto-ATPase in *M. hominis* (Hopfe and Henrich 2008). The binding of OppA to the host cell induces the release of ATP from the host cell. Extracellular ATP is then hydrolysed by OppA (ecto-ATPase activity) to ADP and a phosphate group. This reduces proliferation and leads to apoptosis of the infected host cell (Hopfe and Henrich 2014).

Mycoplasmas are host-specific, therefore it is possible that the genes involved in pathogenesis are species-specific and that this may be the reason for the lack of a general mechanism of pathogenesis in the *Mycoplasma* genus. The lesions produced in different hosts by different *Mycoplasma* species are, however, similar (Browning et al. 2014). This suggests that the underlining mechanisms may be similar.



#	Enzyme	Gene	Species	
1	Glycerol uptake facilitator	*glpF*	*M. gallisepticum* *M. mycoides* subsp. *mycoides* *M. pneumoniae*	
2	ABC glycerol transporter	*gtsABC*	*M. mycoides* subsp. *mycoides*	
3	Putative glycerol-3-phosphate ABC transporter	*ugpACE*	*M. gallisepticum* *M. pneumoniae*	
4	Glycerophosphodiester transporter	*glpU*	*M. gallisepticum* *M. pneumoniae*	
5	Glycerol kinase	*glpK*	*M. gallisepticum* *M. pneumoniae*	
6	Glycerophosphodiesterase	*glpQ*	*M. pneumoniae*	
7	Glycerol-3-phosphate oxidase	*glyD*	*M. pneumoniae*	* cytosolic
		glyO	*M. gallisepticum*	* membrane
			M. mycoides subsp. m*ycoides*	* membrane

Figure 2.2 Schematic representation of H_2O_2 production of the glycerol pathway in *M. pneumoniae, M. mycoides* subsp. *mycoides* and *M. gallisepticum.* The production of H_2O_2 causes damage to host cells. Compiled from Vilei and Frey (2001); Pilo et al. (2005); Großhennig et al. (2013) and Szczepanek et al. (2014).

2.5 Mycoplasma infections in ostriches

Mycoplasma species infect mucosal membranes of the respiratory tract in ostriches and cause sinusitis, rhinitis, foaming conjunctivitis, tracheitis, coughing, laryngitis and air sac infections (Botes et al. 2005a; Botes et al. 2005b). Ostriches in intensive rearing systems are particular vulnerable to mycoplasma infections during extreme weather conditions and stress due to handling, transport or change in housing conditions. Mycoplasma infections lead to retardation of growth, downgrading of carcasses and even death in young chicks which has a major economical impact on the ostrich industry. Mycoplasma infections in ostriches are normally associated with secondary infections of other bacteria that occur in ostriches such as *E. coli*, *Pseudomonas aeruginosa*, *Pasteurella* species and *Bordetella avium* which elevate disease symptoms (Verwoerd 2000; Botes et al. 2005a).

From a phylogenetic or evolutionary perspective, the three *Mycoplasma* species that infect ostriches (*Mycoplasma struthionis* sp. nov. str. Ms01, *Mycoplasma* sp. Ms02 and *Mycoplasma nasistruthionis* sp. nov. str. Ms03) belong to the hominis group within the genus *Mycoplasma* (Figure 2.1). *M. struthionis* sp. nov. str. Ms01 is closely related to *M. hominis* ATCC 23114 and *M. arthritidis* 158L3-1 while *Mycoplasma* sp. Ms02 and *M. nasistruthionis* sp. nov. Ms03, are closely related to *M. synoviae* 53, *Mycoplasma alligatoris* A21JP2 and *Mycoplasma crocodyli* MP145.

M. nasistruthionis sp. nov. str. 2FIA^T was recently isolated from ostriches in Namibia and was characterized as a non-motile mycoplasma that ferments glucose, but cannot hydrolyse arginine or metabolise urea (Langer 2009). Serum is required for *in vitro* growth, and this mycoplasma can be cultured between 28 to 37°C (Langer 2009). The experimental part of this dissertation will focus on *M. nasistruthionis* sp. nov. str. Ms03.

2.6 Controlling mycoplasma infections

Mycoplasma species can be controlled by: 1) prevention 2) treatment and 3) vaccination. Mycoplasma infections can be prevented by biosecurity practices. Biosecurity practices are used to prevent the spread of diseases. This includes quarantine of infected animals and all-in-all-out approaches where a group of animals will be kept together throughout their life preventing close contact with other animals (same and/or other species) on the farm (Evans et al. 2005; Punyapornwithaya et al. 2012).

South Africa is the world's leader in the production of ostrich products (meat, leather and feathers) and biosecurity is of the upmost importance in order to ensure export of safe meat that meets international export standards. It is therefore enforced by law that all ostrich farms in South Africa must be registered and adhere to biosecurity regulation requirements (VPN/04/2012-01(Revision_6.0) 2012). South Africa implemented these regulations after the first outbreak avian influenza in 2004. These requirements include that poultry (excluding

ostriches) and pigs (specifically due to influenza viruses) may not be kept on registered ostrich farms, access for unauthorised persons, animals and vehicles should be restricted and contact between ostriches and wild birds should be limited by fencing off open water and constructing water troughs so that this will discourage wild water birds from perching on the sides or swimming inside them.

Mycoplasmosis can be treated with antibiotics such as tetracyclines and macrolides (targeting protein synthesis). Treatment with antibiotics that target the cell wall (penicillin) or folic acid synthesis (sulfonamides and trimethoprim) fail since the cell wall and folic acid synthesis are absent in mycoplasmas (McCormack 1993). Antibiotic treatment is, however, believed to be mostly ineffective (Frey 2002). This is also the case in ostriches where antibiotic treatment is not always successful because ostriches carrying the disease can remain asymptomatic and are therefore not treated which aids in the spread of mycoplasmas. Additionally in many countries, including South Africa, antibiotic resistance has been reported in *Mycoplasma* species (excluding ostrich-infecting mycoplasma) and are of concern (Kibeida 2011; Zhao et al. 2012; Waites et al. 2014).

Biosecurity practices and antibiotic treatment aid in limiting losses, but they do not provide optimum solutions to control mycoplasma infections. Vaccination provides the animal with the necessary defence to prevent the progression of infections and to fend off pathogens against which it has been vaccinated. This prevents or reduces the symptoms and severity of the disease. Ensuring healthy animals which can grow optimally thus increases commercial production. Vaccines not only protect the animal from the pathogen but also the human consumer from encountering the pathogen. It also decreases the associated risks of long term antibiotic use such as acquiring antibiotic resistance and of consuming pharmacological contaminants in food products (Nisha 2008; Darwish et al. 2013).

Although there are registered vaccines against a few *Mycoplasma* species there are currently no means of preventing mycoplasma infections in ostriches. The sections that follow will explore the literature on vaccines in general in order to evaluate the approaches that can be followed specifically to develop a vaccine for the treatment of mycoplasmosis in ostriches.

2.7 Live attenuated, inactivated whole organism and protein subunit vaccines

Vaccines rely on the ability of the host to gain memory of the particular antigen after exposure. This equips the host with the necessary defences to combat future exposure to the pathogenic organism. The **ideal vaccine** should be highly efficient in activating both B-cells and T-cells that lead to memory. It should be administered in a single dose and cause long term immunity. It should have no side effects in the immunized individuals, even in complicated cases such as immune-compromised, young or old individuals. The

manufacture of the ideal vaccine should also be inexpensive and quality control uncomplicated. It should be temperature stable for transport without a cold chain and have a long shelf-life. Additionally it should also be possible to differentiate infected from vaccinated animals, the so-called DIVA principle (Levine and Sztein 2004; Hoelzle et al. 2009; Uttenthal et al. 2010). Current vaccine approaches include live attenuated vaccines, inactivated whole organism vaccines, protein subunit vaccines and DNA vaccines. Most current vaccines only meet a few of the requirements for the ideal vaccine (Table 2.3).

Currently, most licensed vaccines are live attenuated and inactivated whole organism vaccines while only a few protein subunit vaccines and DNA vaccines are registered (Kutzler and Weiner 2008; Unnikrishnan et al. 2012). Live attenuated vaccines are non-pathogenic or non-virulent strains of the organism that have the ability to stimulate the immune system without causing disease symptoms. Immunity is long-lived and the vaccine can be administrated in a single dose. This approach mimics the natural route of entry of the infective organism and stimulates local immunity (Table 2.5). Immunization with live attenuated vaccines often results in both cellular and humoral immunity (Detmer and Glenting 2006). Live attenuated vaccines have, however, the potential to become reactogenic to the host, especially to immune-compromised, young and old individuals and are transmissible from one individual to the next. However, in future, a better understanding of bacterial genetics and availability of improved molecular techniques may aid in the development of safer live attenuated vaccines that may present fewer side effects (Detmer and Glenting 2006). Live attenuated vaccines are temperature sensitive and therefore require a cold chain transport and storage. Additionally they have a relative short shelf-life. Since the immunized individual is exposed to the whole pathogen it is difficult to differentiate infected from vaccinated animals (DIVA principle). Approved and commercially available live

Table 2.3 The characteristics of the ideal vaccine in comparison with current vaccine approaches

Characteristics of the ideal vaccine	**Vaccine approaches**			
	Live attenuated vaccines	**Inactivated whole organism vaccines**	**Protein subunit vaccines**	**DNA vaccines**
Humoral response (B-cell)	√	√	√	√
Cellular response (T-cell)	√	X	X	√
Single dose	√	X	X	X
No negative side effects	X	X	X	√
Inexpensive to manufacture	√	X	X	√
Uncomplicated quality controls	X	X	X	X
Temperature stable for transport without cold chain	X	X	X	√
Long shelf-life	X	X	X	√
DIVA principle	X	X	√	√

Compiled from Levine and Sztein (2004); Hoelzle et al. (2009); Uttenthal et al. (2010).

attenuated vaccines for mycoplasma control include *M. gallisepticum* str. F (FVAX-MG, Schering-Plough Animal Health), *M. gallisepticum* str. 6/85 (Mycovac-L, Intervet Inc.) and *M. gallisepticum* str. ts-11 (MG vaccine, Merial Select) for vaccination in chickens (Evans et al. 2005; Jacob et al. 2014).

Inactivated whole organism vaccines are heat or chemically inactivated pathogens. These vaccines elicit a potent humoral immune response but do not elicit a cellular response. Booster injections are almost always required. Inactivated whole organism vaccines against bacteria are composed of the entire cell content and therefore may contain toxins that can lead to local inflammation at the injection site as well as other side effects (Table 2.5). It is

Table 2.4 The advantages and disadvantages of current vaccine approaches

Live attenuated vaccines		
Advantages	**Disadvantages**	
1. Highly effective in eliciting both cellular and humoral immune response 2. Long term memory 3. Low cost 4. Can be administered orally 5. Single dose	1. Potential safety risk for immune-compromised individuals 2. Temperature sensitive, requires cold chain for transport and storage 3. May convert into virulent forms 4. Manufacturing is challenging, working with potential dangerous organisms	
Example of commercial license live attenuated vaccines:		
Product name (Supplier, attenuated strain)	**Vaccine target**	**Animal**
FVAX-MG (Schering-Plough Animal Health, strain F)	*M. gallisepticum*	Chicken
Mycovac-L (Intervet Inc., strain 6/85)	*M. gallisepticum*	Chicken
MG vaccine (Merial Select, strain ts-11)	*M. gallisepticum*	Chicken
Inactivated whole organism vaccines		
Advantages	**Disadvantages**	
1. Potent antibody response 2. Safety: Cannot revert to virulent form	1. No cellular immune response 2. Potential risk of toxins and inflammation at the immunization site 3. Temperature sensitive, requires cold chain for transport and storage 4. Manufacturing is challenging, working with potentially dangerous organisms	
Example of commercial licence inactivated whole organism vaccines:		
Product name (Supplier)	**Vaccine target**	**Animal**
Stellamune One (Pfizer Animal Health)	*M. hyopneumoniae*	Swine
Suvaxyn® MH One (Pfizer Animal Health)	*M. hyopneumoniae*	Swine
MG-Bac (Fort Dodge Animal Health)	*M. gallisepticum*	Chicken
Protein subunit vaccines		
Advantages	**Disadvantages**	
1. Potent antibody response 2. Manufacture with well established techniques 3. Safety: Eliminate working with pathogens and exposing patients or animals to live pathogens and toxins within these organisms	1. No cellular immune response 2. Folding problems, protein antigens may be in the non-native form 3. Temperature sensitive, requires cold chain for transport and storage 4. Potential risk of inflammation at the immunization site	

Table 2.5 (Continued).

DNA vaccines		
Advantages	**Disadvantages**	
1. Elicit cellular and humoral immune responses 2. Endogenous antigen production 3. Economic, large scale production 4. More temperature stable than conventional vaccines, long shelf-life and require no cold chain for transport. 5. Produce antibodies in neonates even in the presences of maternal antibodies 6. Animal acts as a bioreactor- thus eliminating problems with protein folding and post-translational modifications of protein antigens	1. Potential problems with potency 2. Would require more than one dose	
Example of commercial licence DNA vaccines:		
Product name (year, country)	**Vaccine target**	**Animal**
West Nile Innovator (2005, USA)	West Nile virus	Horse
Apex®-IHN (2005, Canada)	Infectious haematopoietic necrosis virus	Salmon
Canine melanoma vaccine (2007, USA)	Melanoma	Dogs
LifeTide-SW5 (2007, Australia)	Growth hormone release hormone for foetal loss	Swine

Compiled from: Babiuk et al. (2003); Evans et al. (2005); Detmer and Glenting (2006); Kutzler and Weiner (2008); Villarreal et al. (2011); Marchioro et al. (2013); Bueno et al. (2014); Jacob et al. (2014).

difficult to apply the DIVA principle to these vaccines since the immunized individual is exposed to the whole cell content. Production relies on the ability to culture bacteria in large quantities. Mycoplasmas are notorious for infecting cell cultures (Uphoff and Drexler 2014) creating the false impression that they are easy to culture. However for most species, the small genome size brings with it complex substrate requirements (Citti and Blanchard 2013). The implication is that some species are especially difficult to culture (Citti and Blanchard 2013; Bueno et al. 2014) and renders inactivated whole organism vaccines expensive. Examples of inactivated whole organism vaccines are Suvaxyn® MH One (Pfizer Animal Health) and Stellamune One (Pfizer Animal Health) for the prevention of *M. hyopneumoniae* infections in pigs (Villarreal et al. 2011; Marchioro et al. 2013) and MG-Bac (Fort Dodge Animal Health) for the prevention of *M. gallisepticum* infections in chicken (Bueno et al. 2014).

Protein subunit vaccines consist of a specific protein/antigen that are isolated from the bacterial cell culture or produced recombinantly. As with whole organism vaccines, protein subunit vaccines can also elicit a strong humoral response but cannot elicit a cellular response and booster injections are required (Table 2.5). Protein subunit vaccines eliminate the presence of toxins within the vaccine and therefore are less likely to elicit inflammation at the injection site as well as other side effects associated with inactivated whole organism vaccines. Recombinant proteins are commonly produced in *E. coli* after which the protein is

isolated using chemical or chromatographical methods. Isolation methods may, however, lead to the incorrect folding of the antigenic protein. Expression of mycoplasma proteins within *E. coli* is also hampered because of codon usage differences in particularly the TGA codon that encodes for tryptophan in *Mycoplasma* species but for a universal stop codon in *E. coli*. This terminates translation of mycoplasma proteins in *E. coli* prematurely and warrants the need for an additional step before the expression to change the codon to TGG. Although there are no commercial licensed protein subunit vaccines available for mycoplasmosis, their potential use, development and efficacy has been extensively studied (Galli et al. 2012; Simionatto et al. 2012; Prysliak et al. 2013).

Ostrich-infecting *Mycoplasma* species are slow growing bacteria that require serum for growth (Langer 2009). This makes traditional approaches such as inactivated whole organism vaccines for use in ostriches impractical and economically unfeasible for the relatively small vaccine market if compared to that of poultry. As an alternative DNA vaccines do not require large scale culturing and have also been reported to be cost effective (Mahoney et al. 2000).

DNA vaccines have attracted considerable attention as an alternative vaccine strategy to eliminate some of the disadvantages of other vaccine strategies (Table 2.5). This approach to vaccination was recognized in the early 1990's by Wolff et al. (1990) and will be reviewed in the following section.

2.8 DNA vaccines

A DNA vaccine is an antigen encoding double stranded circular deoxyribonucleic acid (DNA) molecule also referred to as a DNA plasmid or DNA vector. The molecules, encoding an antigenic pathogen protein, are delivered to the eukaryotic host cells. This leads to expression of the antigenic pathogen protein, which in turn is capable of eliciting a specific immune response within the immunized animal and leads to protection against the pathogen. DNA vaccines provided researches with the unique opportunity to utilize the cell's own mechanisms to produce transgenic protein (Lewis and Babiuk 1999; Findik and Çiftci 2012).

The antigen is produced *in situ* and therefore can stimulate both humoral and cellular immune responses (Kutzler and Weiner 2008) (Table 2.5). DNA vaccines circumvent the need to work with dangerous pathogens in vaccine production laboratories as is the case in the production of live attenuated or inactivated whole organism vaccines. Production also has no protein isolation step that can lead to incorrect protein folding as is the case with the production of protein subunit vaccines. Additionally DNA vaccines are relatively stable in dry form or dissolved in buffer (Quaak et al. 2010), thus they do not require traditional cold transport and storage. As with protein subunit vaccines, the DIVA principle can be applied

after vaccination since the plasmid encodes for one or a small number of pathogen proteins. Thus it would be possible to distinguish between infected and vaccinated animals by testing the immune response towards other pathogen proteins.

DNA vaccines can be produced and isolated from *E. coli* bacteria which is a relatively easy and cost effective way of producing a vaccine (Cai et al. 2009). DNA vaccines are also considered safe (Section 2.8.4). In general there are problems with the potency of DNA vaccines in large animals (Babiuk et al. 2003), but this has also been overcome in the commercial DNA vaccines available for pigs and horses.

The design of DNA vaccines is based on a two part strategy: firstly, the production of the DNA vaccine within *E. coli* and secondly, the expression of the antigen within the eukaryotic host. The two parts are interlinked and changes in one part of the plasmid may influence both processes (Hartikka et al. 1996; Williams 2013). For prokaryotic production, the plasmid contains a high copy number origin of replication and a prokaryotic marker (Figure 2.3). The prokaryotic origin of replication influences the copy number of the plasmid within the bacterial cells and therefore has a direct effect on the amount of DNA vaccine produced per unit culture. Prokaryotic markers ensure selection during the cloning process as well as stable inheritance of the plasmid DNA during bacterial growth (Oliveira and Mairhofer 2013). In general this conveys resistance to a drug, typically an antibiotic. The most used resistant marker is the neomycin phosphotransferase II (*nptII*) gene because it is tolerated by regulatory authorities (FDA 1996). The *nptII* gene encodes the enzyme aminoglycoside-3'-phosphotransferase that phosphorylates aminoglycoside antibiotics such as kanamycin and G418 thereby inactivating them. Kanamycin is not commonly use to treat human infections because of side effects (Vandermeulen et al. 2011) and it is therefore safe to use these resistance genes from the vaccine production perspective.

When DNA vaccines were first discovered all plasmids used antibiotic markers since it provided an easy way to select for bacteria containing DNA vaccines. However due to public concern about the possible horizontal transfer of these markers to other bacteria and integration into the host genome, the general perspective is to move away from using antibiotics for selection. New generations of DNA vaccine plasmids are partially or totally devoid of prokaryotic elements (Vandermeulen et al. 2011; Oliveira and Mairhofer 2013). Selection of these vectors is based either on the complementation of auxotrophic strains, toxin–antitoxin systems, operator–repressor titration, RNA markers, or on the overexpression of a growth essential gene (Vandermeulen et al. 2011; Oliveira and Mairhofer 2013). Progress had been made to remove the prokaryotic elements from DNA vaccines although there are still problems with cost effective isolation for some of the alternatives (Faurez et al. 2010).

For eukaryotic protein expression, the DNA vaccine contains a strong eukaryotic promoter, enhancer, intron and transcription terminator/polyadenylation signal in addition to the antigenic gene (Figure 2.3). The plasmid is engineered to contain has a strong and ubiquitous viral promoter that confers optimal expression of the vaccine candidate gene in eukaryotic cells (Faurez et al. 2010; Okuda et al. 2014). Additionally transcriptional transactivators, introns and other enhancer elements are included to increase transcription activity and expression (Flingai et al. 2013; Williams 2013). The polyadenylation signals ensure proper termination of transcription and RNA maturation.

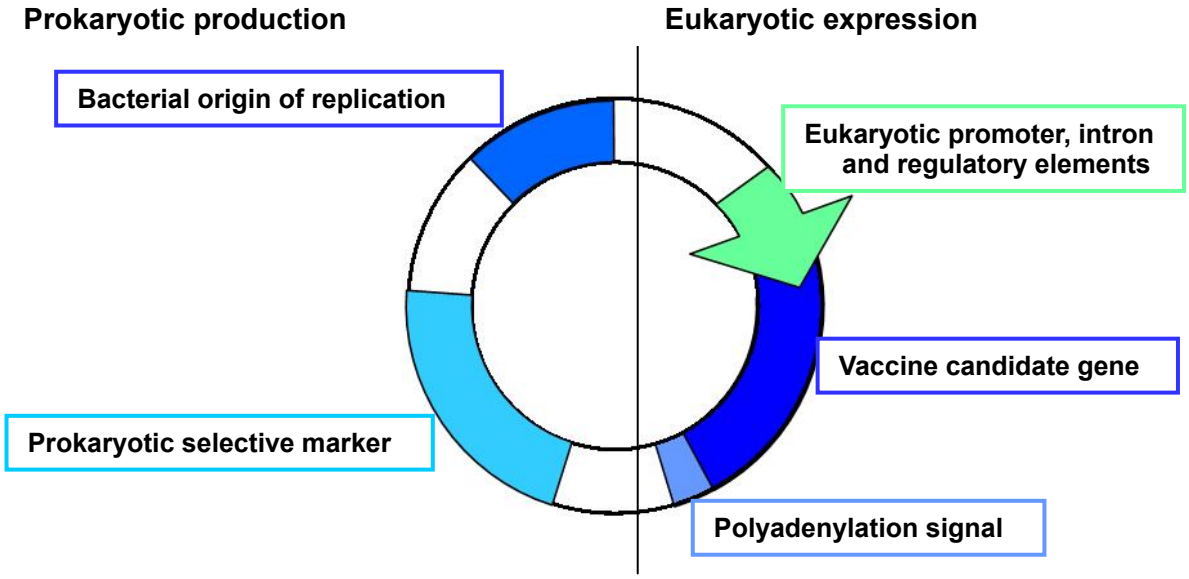


Figure 2.3 A diagram of the typical DNA vaccine vector showing the origin of replication and the selective marker which are needed for prokaryotic production, as well as the promoter, intron and polyadenylation signal, in addition to the vaccine candidate gene that is required for eukaryotic expression

2.8.1 Routes of immunization

DNA vaccines can be formulated as an aqueous solution (naked DNA vaccine), in complex with cationic liposomes and polymers, or packed within virus and bacteria vectors (Faurez et al. 2010). Naked DNA vaccines can by administrated by intramuscular or intradermal injection. Physical methods such as gene gun, electroporation and a CO_2-powered Biojector can also be used for administration of the vaccine (Escoffre et al. 2013; Manjila et al. 2013). Each method proved to have certain advantages and disadvantages, providing a range of approaches that can be applied keeping the outcome in mind (Table 2.5). For example, with bacterial carriers it is possible to administer the vaccine orally. This elicits mucosal immunity, which can be of value if the pathogen attacks/gains entrance via the airways such as in the case of mycoplasmas. These techniques are to overcome the problems associated with the low transfection rate of DNA vaccines. It is, however, important to note that because of the

effectiveness of the immune system, protein expression in the nanogram range is sufficient to elicit an effective immune response (Chastain et al. 2001).

Table 2.5 Advantages and disadvantages of DNA vaccine administration methods

Delivery method	Advantages	Disadvantages
Bacterial or viral carrier	• Easy to administer • Secretory IgA production	• Possible risk due to live bacteria
Biojector injection (id)	• Induces cellular immune response • Easy to administer	• Induce weak humoral immune response
DNA with adjuvants	• Can manipulate to get desired cellular or humoral immune response	• Unknown side effects
Electroporation (id, im)	• High level of immune response • Long duration of immune response	• Possible risk due to high voltage
Gene Gun (id)	• Small amount of DNA injected • Dominance of humoral immune response	• Induce weak cellular immune response
Intranasal immunization	• Easy to administer • Secretory IgA production • Effective for lung immunity	• Weak overall immunogenicity
Needle injection (im)	• Activate cellular immune response • pDNA spreads widely • Most commonly studied • Large amount of DNA can be injected	• Muscle pain

Abbreviations: id intradermal, im intramuscular.
Table from Bermúdez-Humarán et al. (2011); Okuda et al. (2014).

2.8.2 The journey into the cell nucleus

Understanding the mechanism of transfection into the nucleus as well as the mechanism of inducing an immune response is critical for the rational design of DNA vaccines. The proportion of DNA vaccine plasmid molecules that reach the nucleus is affected by numerous factors such as the route of administration, size and conformation of the plasmid. The tissue distribution of DNA vaccines as well as the cell type transfected will depend on the route of immunization. After intramuscular immunization, the DNA vaccine molecules will be positioned within the extracellular spaces and be transfected mainly into mycocytes. Intradermal or subcutaneous administration, on the other hand, will predominately lead to transfection of keratinocytes. Additionally a small portion will also be directly transfected into professional antigen presenting cells (APCs), dendritic cells (DC) in the muscle and the epidermal Langerhans cells or dermal dendritic cells in the skin (Kutzler and Weiner 2008; Palumbo et al. 2012).

Studies evaluating the anatomical distribution of DNA vaccines within a body have shown that after immunization most DNA vaccines either remain at the injection site, where they are taken up by cells or degraded by nuclease enzymes, or they are distributed through the blood and lymph system to various tissues (Tadokoro et al. 2001). In the blood and lymph, the plasmid will also either be degraded or taken up by APCs. Transfection occurs at a very

low frequency because the extracellular space, blood, lymph and cytoplasm are equipped with enzymes that degrade DNA.

The half-life of DNA vaccines within the blood of rats after intravenous injections ranges from 0.15 to 21 minutes depending on the topoform (linear, open circle or supercoiled) of the DNA vaccines (Houk et al. 2001). In addition to nuclease enzymes, the liver nonparenchymal cells such as Kupffer (liver resident macrophage) and liver sinusoidal endothelial cells contribute to the degradation of DNA vaccines. These cells use scavenger receptors to internalize and degraded large anionic molecules. Trace amounts of plasmids could still be detected after 28 days in rats (Tuomela et al. 2005) and pigs (Kanellos et al. 1999), 54 days in sheep (Mena et al. 2001), 70 days in goldfish (Kanellos et al. 1999) and turkeys (Kanellos et al. 1999) and 90 days in rainbow trout (Garver et al. 2005).

In order for the antigen to be expressed, the vaccine plasmid should be transfected into the nucleus. To enter the nucleus, the vaccine plasmid needs to overcome three obstacles, crossing the cell plasma membrane, travelling to the nucleus and crossing the nuclear membrane, as illustrated in Figure 2.4.

The cell membrane as well as the DNA vaccine plasmid have an overall negative charge implying that the two do not associate with each other. Despite this, DNA vaccines do cross the cell membrane and it has been proposed that it happens via endocytosis although the precise mechanism is not known (Tonheim et al. 2008). Once inside the cell, DNA vaccines

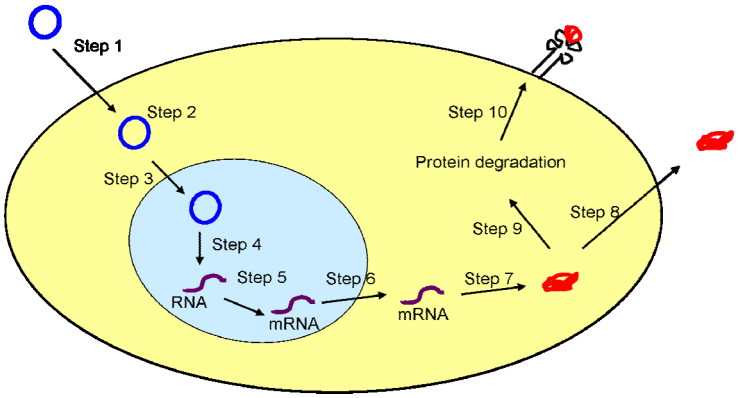


Figure 2.4 Path of a DNA vaccine into the cell nucleus and the expression of the antigenic protein. The vector is imported into the cell through endocytosis (Step 1). The DNA vector travels through the dense cytoplasm to the nucleus (Step 2) where it crosses the nucleus membrane to enter the nucleus (Step 3). Within the nucleus the antigenic gene is transcribed to RNA (Step 4) which is processed to mRNA (Step 5) and exported to the cytoplasma (Step 6) for translation of the antigen protein (Step 7). The antigenic protein can either be exported (Step 8) or it can be tagged for protein degradation (Step 9), which leads to the presentation of antigenic peptides by the major histocompatibility complex I (MHC I) on the cell membrane (Step 10).

are hardly stable, it is estimated that the half-life for DNA vaccines within the cytoplasma of HeLa and COS cells is 50 to 90 minutes (Lechardeur et al. 1999). The eukaryotic cytosol has dense and viscose characteristics with proteins, organelles and the cytoskeleton making diffusion of large molecules such as DNA vaccines slow (Lukacs et al. 2000). DNA vaccines use the microtubule network of the cytoskeleton and the molecular motor, dynein, to move to the nucleus (Vaughan and Dean 2006). A multiprotein complex is involved in this transport since the DNA vaccine does not directly bind to the dynein.

The DNA vaccine cross the nuclear membrane through a nuclear pore complex (Lechardeur and Lukacs 2006). About 0.1% of the plasmid DNA microinjected into the cytoplasma of COS-7 cells reached the nucleus and resulted in expression (Pollard et al. 1998).

2.8.3 How does the DNA vaccine result in immunity?

The antigen gene of the DNA vaccine is transcribed to RNA within the nucleus. Sequentially, the RNA is processed to remove introns (if present) and a 5' cap and 3' polyadenylate tail is added. The mature messenger RNA (mRNA) is then exported to the cytoplasma where it is translated to the antigenic protein (Figure 2.4).

Our understanding of how DNA vaccines lead to immunity is far from complete. The mechanism, as outlined in Figure 2.5 is based on the current understanding (Kutzler and Weiner 2008; Moss 2009; Liu 2011; Li et al. 2012; Pereira et al. 2014; Xu et al. 2014). After immunization, the DNA vaccine is internalized by somatic cells (myocytes or keratinocytes), as well as APC cells, leading to the expression of the protein antigen (Ag). The host-synthesized Ag molecules are subjected to immune surveillance and will be tagged by ubiquitylation for proteolysis that will degrade the antigenic protein into small fragments. The fragments enter the endoplasmic reticulum where they are loaded onto major histocompatibility complex class I (MHC I) molecules. The Ag-MHC I complex is transported via the Golgi complex to the cell surface where it facilitates binding with $CD8^+$ T-cells (Vyas et al. 2008; Wang et al. 2011). $CD8^+$ T-cells recognize the Ag-MHC I complex as non-self and elicit a cellular immune response that results in the formation of memory T-cells as well as cell death of the infected cell.

APC can either be directly transfected and produce the Ag or the Ag can be internalised by sample part of the extracellular milieu by a process called phagocytosis. This results in the display of the antigenic peptides on major histocompatibility complex class II (MHC II) molecules. This activates $CD4^+$ T-cells leading to the activation and maturation of B-cells which will produce antibodies, as well as memory B-cells, and thereby elicits humoral immunity.

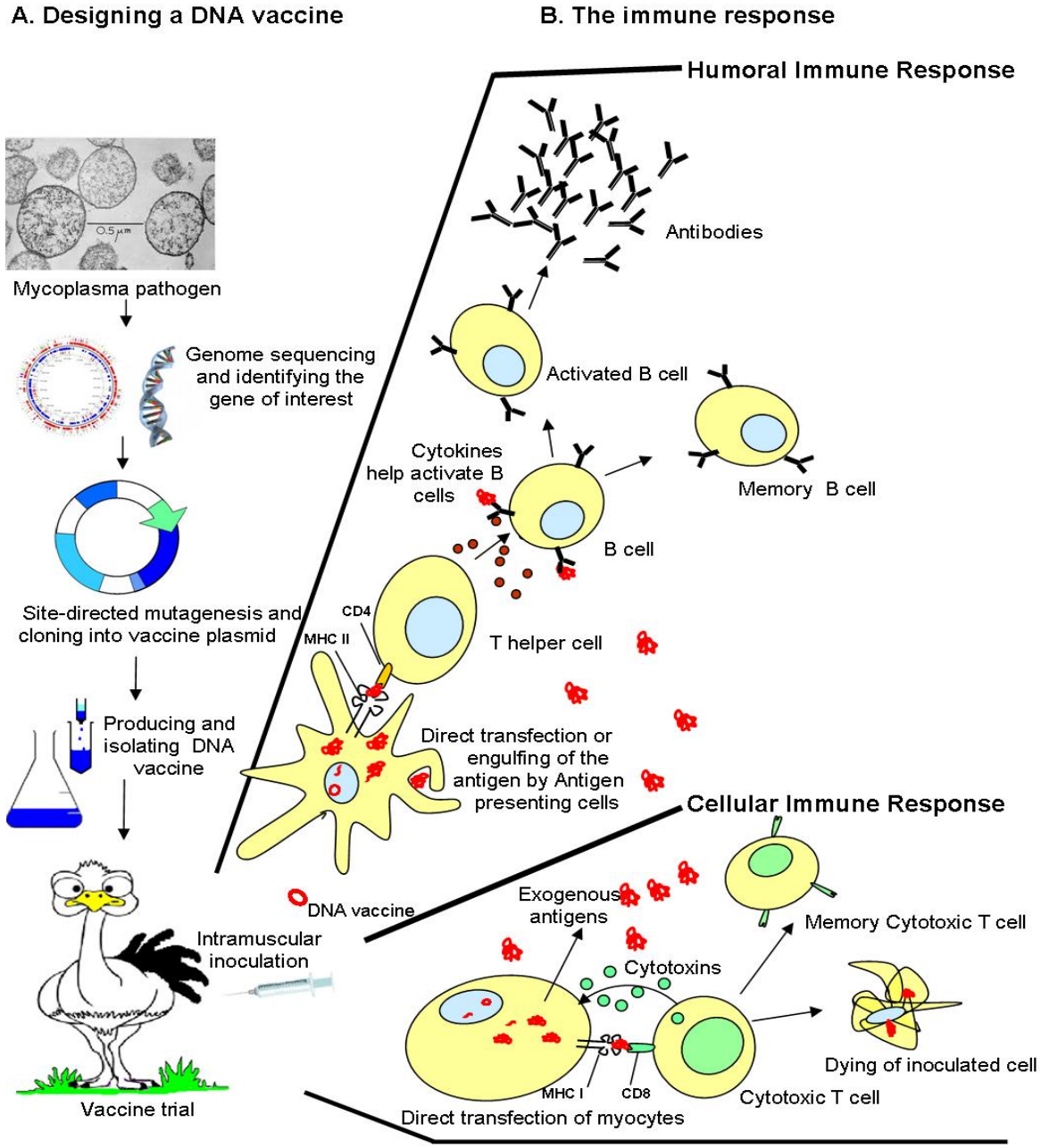


Figure 2.5 A) Design of a DNA vaccine against mycoplasmas and the immune responses elicited by it. Genomic DNA of the mycoplasma is isolated and sequenced with a next-generation sequencing platform. The vaccine candidate gene is identified from the genomic data. Site-directed mutagenesis is performed to correct for codon differences between the mycoplasma and its host, thereafter the gene is inserted into a vaccine vector for subsequent immunization trials. B) The immune response. Cellular immune response: The DNA vaccine enters myocytes or APC leading to the expression of the antigen (Ag). If a signal peptide is present the Ag may also be exported. Fragments of the antigen produced within cells are displayed by the MHC I on the cell surface for self-recognition. $CD8^+$ T-cells recognize the Ag-MHC I complex as non-self and elicits a cellular immune response that results in the formation of memory T-cells as well as cell death of the infected cell. Humoral immune response: APC cells engulf or produce Ag that leads to the presentation of Ag on MHC II molecules. This activates $CD4^+$ T-cells leading to the maturation of B-cell which will produce antibodies as well as memory B-cells. Figure complied from Kutzler and Weiner (2008); Moss (2009); Liu (2011); Li et al. (2012); Pereira et al. (2014); and Xu et al. (2014).

2.8.4 DNA vaccines and safety

Safety is an important concern with any vaccine and as such should be considered in full. It is important to understand that DNA vaccines cannot replicate within eukaryotic cells and that the transfection percentage of the cells is extremely low. This coupled with the low half-life of the vaccine within blood and tissue implies that most of the DNA vaccine degrades within the body within minutes to hours. Additionally, DNA vaccines are non-living thus cannot revert to a disease state or cause secondary infection and inflammatory responses (Ferraro et al. 2011; Xu et al. 2014). For these reasons, DNA vaccines are considered to be safe and environmentally friendly (Ferraro et al. 2011; Williams 2013; Xu et al. 2014).

Safety concerns about the use of DNA vaccines are integration into the host genome, vertical transmission, the production of anti-DNA antibodies (autoimmunity), the possible spread of selective markers leading to antibiotic resistance and the effect of consuming vaccinated food animals. Integration into the host genome may result in suppression or activation of genes, which may have devastating effects such as cancer. Preclinical and clinical studies have, however, proven the rate of integration of plasmid DNA into the genome is lower than that of spontaneous mutations (Wang et al. 2004; Sheets et al. 2006). Vertical transmission of the DNA vaccines from one generation to the next is of concern, but has not been reported yet (Langer et al. 2013; Williams 2013; Xu et al. 2014). Animal studies showed no increase in anti-DNA antibodies after DNA vaccination and there was no evidence that DNA vaccination resulted in autoimmunity (MacGregor et al. 1998; Tavel et al. 2007; Kutzler and Weiner 2008; Fioretti et al. 2014). In the production of DNA vaccines an antibiotic selective marker is generally used and concerns exist about the spread of antibiotic resistance through horizontal gene transfer. The use of antibiotic resistance genes should therefore be restricted to exclude antibiotics generally used in treatment of human disease such as kanamycin (FDA 1996) or alternatively an antibiotic free production system should be used. Humans consume DNA of plant, animal and bacterial origin daily as a part of their food. The risk of consuming plasmid DNA from vaccines should pose no greater risk than natural DNA (Schalk et al. 2006).

2.8.5 The choice of DNA vaccine candidate gene

Choosing a vaccine candidate gene for a DNA vaccine is critical as it determines the success of the vaccine. The gene candidates should conform to the following requirements:

1. It should be a unique gene with no homologous or paralogous genes within the pathogen's genome or within the host's genome (Galperin and Koonin 1999).
2. The protein product must have an essential function in the pathogen thus life without the protein should be impossible (Allan and Wren 2003). Essential functions in pathogens include nutrient uptake, DNA replication, cell division, RNA transcription,

protein translation, virulence and pathogenicity. Proteins that are involved in more than one pathway can make more potent targets since more than one pathway will be affected.

3. The protein product must not be hypermutable or involved in antigen variation thus serving as decoys in preventing an effective immune response (Liu 2011).

4. The protein product must be accessible. Critical epitopes must not be masked thereby preventing an effective immune response (Liu 2011). Surface proteins generically have good accessibility since these proteins are exposed to the immune system of the host. An added advantage is that some have the essential role of host attachment, the first step in pathogenicity. Surface proteins are also involved in transport of essential nutrients, which in the case of mycoplasma, makes life possible.

In addition to the above, *in silico* prediction programs also use the availability of 3D structures and protein size as parameters to select suitable vaccine candidate genes (Agüero et al. 2008; Caffrey et al. 2009). This however does not mean that larger proteins without resolved 3D structures cannot make good and effective vaccines, it simply means that the experimental procedure may be more complicated.

2.9 DNA vaccine strategies for ostrich-infecting mycoplamas

The process of designing and developing a DNA vaccine for mycoplasmas is outlined in Figure 2.5 A. The steps include the identification of a vaccine candidate gene from a sequenced and annotated genome of the pathogen, the cloning of the vaccine candidate gene into a vaccine vector and the production of the DNA vaccine in *E. coli* before the DNA vaccine can be evaluated. Additionally for mycoplasmas, codon optimization is also required.

2.9.1 Codon usage in *Mycoplasma* species

Tryptophan is coded as UGG in the universal genetic code whilst UGA is one of the three universal stop codons. In the class *Mollicutes*, however, the universal stop codon UGA, encodes for the amino acid tryptophan (Inamine et al. 1990). The use of UGA as an additional tryptophan codon is attributed to the reduced guanine and cytosine (G+C) content in mycoplasma genomes (Halbedel and Stulke 2007). Consequently the expression of mycoplasma genes in *E. coli*, avian or mammalian cells is jeopardized by the occurrence of the TGA codon within genes (UGA in RNA sequences) that result in premature termination of protein expression (Halbedel and Stulke 2007). To circumvent this problem a number of approaches can be used. Firstly, an expression system that employs the same codon can be used. The expression of mycoplasma proteins in *Spiroplasma citri* has been reported but

this bacterium is difficult to cultivate (Stamburski et al. 1991). Secondly, an *E. coli* suppressor strain that expresses UGA suppressor tRNA codons can be used to express mycoplasma proteins but they fail if the gene contains multiple UGA codons (Smiley and Minion 1993). Thirdly, genes can be synthesized *in vitro* with appropriate codons but this is expensive for long genes. Lastly, the codons can be changed by site-directed mutagenesis (SDM). SDM is a polymerase chain reaction (PCR) based technique in which primers that contain the mutated sites, are used to amplify the gene as well as the vector. After PCR amplification, the original template is removed with DpnI endonuclease enzyme digestion. DpnI endonuclease digests methylated and hemimethylated DNA i.e. the DNA of bacterial origin. The product is then transformed into *E. coli* and the mutation sites confirmed by sequencing (Ishii et al. 1998).

The influence of a low G+C content can further be observed in codon preferences in mycoplasma. The occurrence of rare codons may influence the expression of recombinant protein in *E. coli* as well as in the ostrich (targeted host) after DNA vaccination. Each animal and bacterium has its own codon preferences that result in optimal expression within the organism (Babiuk et al. 2003). No codon preference table is, however, currently available for the optimization of expression in ostrich. Although codon optimization can lead to increased protein expression, synonymous codon changes may also affect the protein conformation and stability, change sites of post-translational modifications and even alter protein function (Tsai et al. 2008; Zhang et al. 2010; Spencer et al. 2012; Zhou et al. 2013; Mauro and Chappell 2014). The immune system is, however, very effective and protein expression in the nanogram range can elicit an effective immune response (Chastain et al. 2001).

For the above reasons the only codon optimization done in the current study was the mutation of TGA codons to TGG by using SDM within the vaccine candidate gene.

2.9.2 DNA vaccine vectors for use in ostrich

The three vaccine vectors chosen for this study were pCI-neo, VR1012 and VR1020 (Figure 2.6). These were selected based on DNA vaccine studies in other birds (Lee et al. 2003; McCutchan et al. 2004; Klotz et al. 2007) and on special characteristics such as a tissue plasminogen activator (TPA) signal peptide in VR1020. The TPA signal should result in the export of the translated protein from the cell in which it is produced and thus better activation of the immune system. The prokaryotic elements in these vectors include an origin of replication and a prokaryotic selection marker. These elements are required for the production of the DNA vaccines within *E. coli*. The prokaryotic selection marker in pCI-neo is ampicillin resistance while in both VR1012 and VR1020 the prokaryotic selection marker is kanamycin resistance. The use of kanamycin resistance is preferable because kanamycin not commonly use to treat human infections due to side effects (Vandermeulen et al. 2011).

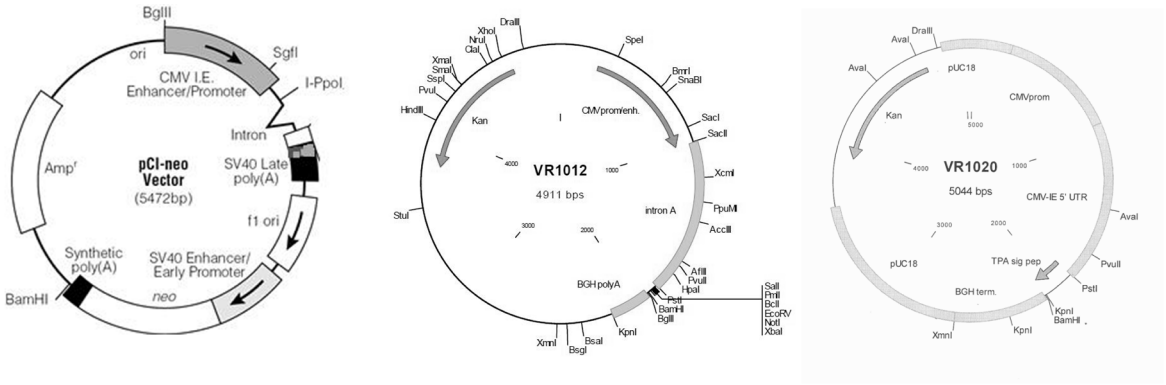


Figure 2.6 The DNA vaccine vector map of pCI-neo, VR1012 and VR1020.

The eukaryotic elements in all of these vectors include a promoter region, intron and polyadenylation signal. These regions are required for optimal expression in eukaryotic cells. All three of these vectors use a cytomegalovirus (CMV) immediate-early (IE) enhancer/promoter region. The CMV promoter is a virally-derived promoter that leads to a high-level of transgenic expression in a wide range of eukaryotic cells making it an ideal choice for DNA vaccines (Garmory et al. 2003; Kutzler and Weiner 2008). Furthermore pCI-neo has a chimeric intron and a simian virus 40 (SV40) late polyadenylation signal while both VR1012 and VR1020 have the intron A from CMV and the bovine growth hormone (BGH) terminator site. Introns have been reported to increase expression (Sridhar et al. 2008) while the polyadenylation signal or terminator site is required for proper termination of transcription and export of the mRNA from the nucleus (Kutzler and Weiner 2008). Polyadenylation sequences such as BGH or SV40 also increase the stability of mRNA transcripts (Garmory et al. 2003).

2.9.3 DNA vaccine candidate gene, *oppA*

Numerous potential target genes for DNA vaccines are described in the literature and amongst these are the ABC transporters. ABC importers are involved in the active import of nutrients and are located on the surface of the cell membrane of mycoplasma.

One such importer is the oligopeptide permease (Opp) transporter that is responsible for the active import of oligopeptides due to the absence of *de novo* amino acid synthesis. The Opp transporter contains a substrate-binding domain referred to as OppA. The OppA proteins of *Brachyspira pilosicoli* (Movahedi and Hampson 2010), *Moraxella catarrhalis* (Yang et al. 2011) and *Yersinia pestis* (Tanabe et al. 2006) have been evaluated as candidate vaccine antigens and could similarly be a potential vaccine antigen in Ms03.

The *oppA* gene and its corresponding protein product have several characteristics that make it a suitable vaccine candidate gene:

- The substrate-binding domain is unique to plants, bacteria and archaea (Berntsson et al. 2010; Rice et al. 2014) which implies that the *oppA* gene should have no homologous genes in the ostrich genome.
- Oligopeptide transport is an essential function in mycoplasmas and the substrate-binding domain is a prominent part of the transporter. Furthermore, transposon mutagenesis studies in *M. genitalium* and *M. pulmonis* found the *oppA* gene to be essential for survival (Glass et al. 2006; French et al. 2008).
- The OppA is not involved in antigen variation due to its essential function in oligopeptide binding.
- The substrate-binding domain OppA is surface located and therefore exposed to the immune system of the host.

Additionally the OppA of *M. hominis* has been reported to function in cytoadhesion (Henrich et al. 1999; Hopfe and Henrich 2004). If this is also the case in *M. nasistruthionis* sp. nov. str. Ms03, then antibodies against the OppA would not only limit or prevent oligopeptide transport but also cytoadhesion and therefore infections. In order to identify an *oppA* gene in *M. nasistruthionis* sp. nov. str. Ms03, the genome was sequenced and annotated. The sequencing and annotation results are presented in the following chapter together with the identification and characterization of the *opp* operon.

Chapter 3 The sequencing and annotation of the *Mycoplasma nasistruthionis* sp. nov. str. Ms03 genome and the identification and characterization of the proposed vaccine candidate gene, *oppA*

3.1 Introduction

Recent developments in genome sequencing techniques provide a unique opportunity to characterize individual genes of bacteria and thereby obtain an overview of their metabolic capacities. Knowledge of the genome also allows for a reverse vaccinology approach to be taken where bioinformatic tools are used to identify genes that conform to the requirements for a vaccine candidate gene. Having an annotated genome further enables the quick and effective identification of a vaccine candidate gene.

The first aim of this study was therefore to expand the knowledge of *Mycoplasma nasistruthionis* sp. nov. str. Ms03 (Ms03, this abbreviation is used in this chapter for brevity) by determining the Ms03 genome sequence, annotating the genes and evaluating the metabolic pathways of essential processes. This would also allow future identification of vaccine candidate genes. The objectives were therefore set as follows:

- To sequence and assemble the Ms03 genome.
- To annotate the Ms03 genome using two annotation programs.
- To generate a metabolic overview of Ms03.

A number of next-generation sequencing (NGS) platforms are available for genome sequencing. These include the Roche 454 GS FLX Titanium, Illumina Solexa GAII, Life APG SOLiD 3 and Pacific Biosciences PacBio RS; these are compared in Table 3.1. Pyrosequencing (such as the Roche 454 platform) is a sequence-by-synthesis technique that relies on the detection of pyrophosphate release during nucleotide incorporation. The released inorganic pyrophosphate initiates a chain of enzyme reactions that ultimately produces light. The sequential nucleotide specific light signals are then used to determine the nucleotide sequence (Harrington et al. 2013). The Roche 454 GS FLX Titanium platform was used in this study because the longer read length is beneficial for *de novo* sequencing of genomes in which repeats frequently occur as in the case of mycoplasmas. *De novo* sequencing was used to determine the Ms03 genome sequence since no reference genome was available for Ms03. Pyrosequencing produces reads that can be assembled into contiguous sequences (contigs) and these contigs can then be used as a draft genome.

Table 3.1 Comparison of next-generation sequencing techniques

Platform	**Library (Template preparation)**	**NGS chemistry (Read length)**	**Advantages and disadvantages**
Roche 454 GS FLX Titanium	Fragment, paired-end (emPCR)	Pyrosequencing (400 bases)	**Advantage**: Longer reads improve mapping in repetitive regions; fast run times **Disadvantage**: High reagent cost; high error rates in homopolymer repeats
Illumina Solexa GA_{II}	Fragment, mate-pair (emPCR)	Sequencing by synthesis with reversible terminators (75 or 100 bases)	**Advantage**: Low cost per Mb of data generated **Disadvantage**: Low multiplexing capability of samples
Life APG SOLiD 3	Fragment, mate-pair (solid-phase)	Cleavable probe sequencing by ligation (50 bases)	**Advantage**: Two-base encoding provides inherent error correction **Disadvantage**: Long run times
Pacific Biosciences PacBio RS	Fragment (single molecule)	Single DNA polymerase detection (mean 2246 bases, maximum 23000 bases)	**Advantage**: Longer reads improve mapping in repetitive regions and *de novo* sequencing **Disadvantage**: Highest error rates compared with other NGS chemistries

Abbreviation emPCR: emulsion-based clonal amplification.
Adapted from Metzker (2009); Glenn (2011); van Dijk et al. (2014).

For genome annotation, two automatic prokaryotic servers were used, namely the Institute for Genome Sciences (IGS) prokaryotic annotation pipeline and the Rapid Annotations using Subsystem Technology (RAST) annotation pipeline. IGS and RAST annotation servers are both web-base analyses for the annotation of bacterial and archaeal genomes that allow the user to manually curate data (Table 3.2). The two servers use different approaches to annotation as outlined in Table 3.3 and Table 3.4. Automated annotation results from different servers are often reported to provide different and sometimes conflicting predictions (Ederveen et al. 2013). The two annotation pipelines were therefore compared in attempt to generate a more complete/accurate annotation of the Ms03 genome. Additionally RAST allows for a metabolic reconstruction of the annotated data using KEGG pathways. An overview of the metabolic capacity of Ms03 could therefore be generated using the gene annotation of both annotation servers as basis.

Traditional whole organism vaccine development against ostrich-infecting mycoplasmas is impractical and expensive because Ms03 is a slow growing bacterium that requires undefined medium components such as serum for *in* vivo growth. Knowledge of the Ms03 genome makes a DNA vaccine approach against Ms03 infections in ostriches possible. DNA vaccine development requires knowledge of the mycoplasma genome in order to identify a vaccine candidate gene that can be used in the DNA vaccine.

The o*ppA* gene (substrate-binding domain of the oligopeptide permease (Opp) transporter) was chosen as vaccine candidate gene in previous *Mycoplasma struthionis sp.* nov. str. Ms01 and *Mycoplasma* sp. Ms02 studies within our research group (Pretorius 2009; Steenmans 2010).

The second aim of this study was therefore to identify an *oppA* gene within the Ms03 genome for use as a DNA vaccine candidate gene. The objectives were therefore:

- To identify an *opp* operon and its associated *oppA* gene within the Ms03 draft genome sequence.
- To confirm the identity of the *opp* operon and therefore the *oppA* gene by using bioinformatic analysis.

The *opp* operon in Ms03 was identified using tBLASTn and Glimmer searches. Furthermore, the identity of the *opp* genes was confirmed by identifying the functional protein motifs within their respective protein products.

Table 3.2 Comparison of the IGS and RAST annotation servers

	IGS annotation server	**RAST annotation server**
Developed by	Institute for Genome Sciences, University of Maryland School of Medicine	The National Microbial Pathogen Data Resource
Website	http://ae.igs.umaryland.edu/cgi/index	http://rast.nmpdr.org/
Annotation of	Bacteria and archaea	Bacteria and archaea
Prediction of non-coding RNA and coding protein	Search for non-coding RNA and protein coding genes simultaneously	Search for non-coding RNA first then protein-encoding genes Do not allow protein-encoding genes to overlap with non-coding RNA
Programs used for rRNA prediction	RNAmmer	search-for-rrna BLASTn against RNA database for endpoint adjustments
Programs used for tRNA prediction	tRNAscanSE	tRNAscanSE
Prediction of protein-encoding genes	Self-training method with Glimmer3*	Self-training method with Glimmer3*
Metabolic reconstruction	No	Yes

* Note the self-training methods are different between the two programs, see Table 3.3 and Table 3.4.
Compiled from Aziz et al. (2008); Galens et al. (2011); Overbeek et al. (2013).

Table 3.3 The IGS annotation pipeline

The IGS annotation pipeline can be used to annotate both draft and finished genomes. The pipeline includes a number of steps for gene finding, protein searching and functional assignment (Galens et al. 2011). These steps are:

1. Parallel prediction of non-coding RNA and protein-encoding genes:
   - Non-coding RNA genes are identified by using tRNAscanSE for transfer RNA (tRNA) gene identification and RNAmmer for ribosomal RNA (rRNA) gene identification.
   - Protein-encoding genes are predicted with an *ab initio* Glimmer3 search. A set of long non-overlapping genes are generated. From this set the relative frequencies of start sites are calculated and an upstream position weight matrix is created to aid in the identification ribosomal binding sites. These along with the set of long non-overlapping genes are used as input for the second Glimmer3 iteration. The predicted genes from this second round of Glimmer3 are used for downstream annotation.
2. Similarity searches (Round 1) using the predicted genes (Round 2 of Glimmer3) as queries:
   - An initial BLASTx search is performed against the UniRef100 database.
   - The resulting pairwise alignments are used as input into BER (Blast Extend Repraze). BER translates the extended nucleotide query (300 nucleotides (nt) up- and downstream from the predicted gene) to protein and aligns it to each protein match found in the BLASTx search. This may result in up to 150 alignments for each predicted gene. By aligning the extended nucleotide, it is possible to adjust for frameshifts or in-frame stop codons due to sequencing errors. The BER tool aligns the sequence through the error and allows for better curation. One round of extension is performed; further manual curation is needed if the predicted gene needs to be extended more than 300 nt.
   - HMMER package is used to search the predicted polypeptide sequences against the TIGRFAM and Pfam databases.
3. Refinement of the gene predictions based on automated evidence-directed structural curation:
   - The start and stop codons are adjusted based on the BER analysis and the presence of a ribosomal binding site upstream of the start codon.
   - All overlapping genes (non-coding RNA and protein-encoding) are identified. The BER analysis is evaluated for overlaps larger than 60 nt. If no evidence from a BER or HMM (Hidden Markov Model) search are found for one of the overlapping genes, it suggests a false positive and the gene is removed from the gene set. The same principle is applied to RNA genes overlapping protein-encoding genes. All other overlaps greater than 60 nt are flagged for manual reviewing.
   - Regions between genes are explored with BLASTx searches against the UniRef100 database to identify possible false negatives.
4. Similarity searches (Round 2) using the newly identified or curated genes as queries:
   - Similarity searches include an initial BLASTx search as well as BER and HMM searches (against the TIGRFAM and Pfam databases). In the second round of similarity searches the genes are not extended with 300 nt as is the case within the first round of similarity searches.
5. Motif searches on each predicted polypeptide:
   - Protein motif searches include SignalP (for prediction of signal peptide cleavage sites), LipoP (for prediction of lipoprotein signal peptides), TMHMM (for prediction of transmembrane helices), PROSITE using ScanProsite (for prediction of binding sites, active sites etc.), NCBI COGs (for assessment of orthologous groups) and Priam (to assign enzyme commission (EC) numbers) analyses are performed on each predicted polypeptide.
6. Assignment of a common name, gene symbol, Gene Ontology (GO) terms, EC number and TIGRFAM roles to each polypeptide is based on the results of the similarity and motif searches using the program pFunc (prokaryotic protein functional prediction).
7. All evidences are presented for online manual curation in Manatee.

Table 3.4 The RAST annotation pipeline

The RAST annotation pipeline can be used to annotate both draft and finished prokaryotic genomes. The pipeline uses the FIGFam and subsystem approach to annotation (Aziz et al. 2008; Overbeek et al. 2013). The steps within the RAST annotation are as follows:

1. Special case genes (such as the genes for selenoproteins and pyrrolysoproteins) are identified first. These genes require domain-specific knowledge for identification.
2. Non-coding RNAs are identified using tRNAscanSE for tRNA gene identification and "search-for-rrna" for rRNA gene identification. BLASTn searches against an rRNA database are used for endpoint adjustments of the rRNA genes.
3. An *ab initio* Glimmer3 search is used to identify gene-candidates. The resulting gene-candidates are compared to a set of universal, plus up to 200 "unduplicated", proteins in order to determine the 30 closest phylogenetic neighbours.
4. All gene-candidates are searched against proteins in the subsystems of the 30 closest phylogenetic neighbours. Subsystems are based on the FIGfam database. Genes are retained if they are similar to proteins in a subsystem and do not overlap significantly with a gene previously called. These genes are assigned a functional role base on these similarities.
5. Glimmer is retrained based on the retained genes. The gene-candidates from this iteration are searched against proteins in the subsystems as in step 4. This is repeated until no new gene-candidates are found to be similar to subsystem proteins.
6. The remaining gene-candidates that do not match subsystem proteins and do not overlap with existing called genes are retained if they are similar to any protein in the 30 closest phylogenetic neighbours.
7. The rest of the gene-candidates are now retained if they do not overlap with the genes called in step 4 to 6. This reduces the false negative gene calling.
8. Gene fragments that may contain frameshifts due to low quality sequencing are detected by comparing it to the genes from the 30 closest phylogenetic neighbours. Upon user request, the fragments are joined to a single gene.
9. Stretches of DNA longer than 1 500 base pairs (bp) that do not contain any genes are compared to the proteins of the 30 closest phylogenetic neighbours using BLASTx in order to identify missing genes.
10. Functions for the genes that did not have subsystem matches are assigned base on BLASTp similarities.
11. If a gene-candidate lacks subsystem assignment but the genes flanking it have subsystem assignments, then it is compared to the 30 closest phylogenetic neighbours. The gene may then be assigned into a subsystem provided that the three genes have bidirectional-best-hits in a neighbouring genome.
12. Additionally gaps are re-examined if they are flanked by genes with bidirectional-best-hits to genes that are in a subsystem in one of the neighbouring genomes. This aids in the identification of missing genes.
13. Gene-candidates are removed from the annotation if 1) they do not have subsystem or BLAST support, 2) are embedded in another gene, 3) significantly overlap with another gene or 4) are extremely short (>90 nt).
14. Subsystem analysis and initial metabolic reconstruction (using the KEGG pathways) are performed.
15. Pairs of close bidirectional-best-hits are computed against the genomes in PubSEED to support estimates of functional coupling based on conserved contiguity.
16. Results are presented online in SEED Viewer that provides the opportunity for viewing and manual curation of the annotated genome.

3.2 Material and methods

3.2.1 DNA isolation and confirming the identity of Ms03 genomic DNA

Cultures of Ms03 were obtained from Mr J.J. Gouws (Faculty of Veterinary Science, Onderstepoort, University of Pretoria). Genomic DNA (gDNA) was isolated from these cultures using a method described by Hempstead (1990) with the following modifications: The cells were collected from 25 ml of culture by centrifugation and directly resuspended in 1 ml concentrated TE buffer (50 mM Tris-HCl, pH 7.5, 10 mM EDTA). Following the sodium acetate/ethanol precipitation step, samples were incubated overnight at -20°C before collecting the gDNA by centrifugation at 16 000 × g for 1 h at 4°C. The pellet was washed twice by adding 1.5 ml 70% ethanol and incubation for 20 min at room temperature before centrifugation at 16 000 × g for 10 min at 4°C. The pellet was air dried at room temperature and dissolved in 50 µl Milli-Q® water. The gDNA concentration was determined spectrophotometrically (NanoDrop spectrophotometer, ND-1000), while the integrity of the gDNA samples was assessed on a 1% (w/v) agarose gel (Lonza, Switzerland). Electrophoresis was performed in 1 × TAE buffer (40 mM Tris-acetate, 1 mM EDTA, pH 8, 20 mM glacial acetic acid) at 100 V for 5 min followed by 12 h at 12 V. After electrophoresis the agarose gel was stained for 10 min in 0.5 µg/ml ethidium bromide solution and visualised under UV light.

Polymerase chain reactions (PCR) were performed to confirm the identity of Ms03 gDNA and the absence of possible contamination from the other ostrich mycoplasmas, *M. struthionis* sp. nov. str. Ms01 and *Mycoplasma* sp. Ms02. The primer sets used to identify the specific ostrich mycoplasmas was based on *16S rRNA* gene sequences (Botes et al. 2005a) (Table 3.5). Each PCR reaction contained 1 × reaction buffer, 0.2 mM of each deoxynucleotide (dATP, dCTP, dGTP and dTTP, Kapa Biosystems, South Africa) 2 mM $MgCl_2$, 1 pmol/ml of each primer (Table 3.5), 0.2 units of Super-Therm Taq DNA polymerase (JMR Holdings, USA) and 2 µl gDNA (diluted 1:100) in a total volume of 10 µl. Positive controls containing *M. struthionis* sp. nov. str. Ms01, *Mycoplasma* sp. Ms02 and Ms03 gDNA as well as a negative control without template were included. PCR fragments were amplified with 25 cycles of denaturation at 94°C for 30 sec, annealing at 57°C for 15 sec and elongation at 72°C for 1 min followed by a final extension step at 72°C for 6 min in a Veriti 96 well Thermal Cycler (Applied Biosystems, USA). PCR products were separated using 2% (w/v) agarose gel containing 0.5 µg/ml ethidium bromide. Electrophoresis was carried out at 100 V for 40 min in 1 × TAE buffer and products were visualised under UV light. The 100 bp DNA ladder (Promega, USA) was used for size determination.

Table 3.5 Primer sequences, based on *16S rRNA* sequences, for the identification of ostrich *Mycoplasma* species

	Primer	Sequence	Fragment size (bp)
M. struthionis sp. nov. str. Ms01 specific primers	MS012 (F)	5'-AACATTAGTTAATGCCGGATACGC-3'	499
	MS01D (R)	5'-GCCAGTATCCAAAGCGAGCC-3'	
Mycoplasma sp. Ms02 specific primers	MS02H (F)	5'-AATATAAAAGGAGCGTTTGC-3'	247
	MS02A (R)	5'-AAGGCAATAGCATTTCCTCTACT-3'	
Ms03 specific primers	MS03A (F)	5'-AGTGCTAATGCCGGATACTTATAC-3'	521
	MS03C (R)	5'-CGTTAACCTCTATACAATTCTAGCG-3'	

Reference: Botes et al. (2005a).

3.2.2 Ms03 genome sequencing and assembly

The gDNA was sequenced using the Roche 454 GS FLX Titanium system by inqaba biotec (South Africa). Due to low gDNA concentrations, whole genome amplification was performed using the GenomiPhi kit (GE Healthcare, UK). In attempt to obtain a complete assembly, a second run, using the same library, as well as a third and a fourth run with newly isolated gDNA were done. Reads were assembled into contigs using the GS *de novo* assembler in Newbler (Roche) with default parameters. The genome parameters were also calculated with this program to assess the quality of the assembly. The quality parameters of the Ms03 draft genome were compared to that of *M. struthionis* sp. nov. str. Ms01 (Pretorius 2009) and *Mycoplasma* sp. Ms02 (Steenmans 2010).

3.2.3 IGS and RAST annotations and comparison of the annotations for the Ms03 genome

The IGS annotation engine was used for structural and functional annotation of the sequences (http://ae.igs.umaryland.edu/cgi/index.cgi, Galens et al. (2011)) and Manatee was used to view annotations (http://manatee.sourceforge.net/). Upon submission of the contigs to the IGS annotation engine, contigs were arranged from longest to shortest and assembled together into a pseudomolecule. The contigs were separated by a spacer (NNNNNCACACACTTAATTAATTAAGTGTGTGNNNNN) that introduces stop and start codons in all six reading frames. All the contigs (≥ 100 bp) in the final assembly was used to generated the pseudomolecule.

Annotation was also performed with a second automatic annotation platform: RAST prokaryotic genome annotation server (http://rast.nmpdr.org, Aziz et al. (2008)). For comparison purposes, the same pseudomolecule generated by assembling all contigs was used as input for the RAST annotation.

Both the IGS and RAST are online annotation servers which allow easy access to an annotation pipeline. All analyses and results are provided online which means that extensive

computer capacity is not required by the user. In the case of IGS, annotation results can be viewed online in Manatee that has a results page for each gene showing the evidence from the BER-searches and the protein bioinformatic analyses (SignalP, LipoP, TMHMM, PROSITE, NCBI COGs and Priam) on which the annotation was based. This information is presented in a clear systematic manner with links to the BER-alignments or to websites with the necessary motif information. The alignments are also provided for manual curation of the start and stop codons (Galens et al. 2011). Annotation results from RAST can be viewed online in The SEED Viewer. The viewer has options to browse the genome, to compare the genome based on function, sequence or KEGG pathways as well as the option for manual curation. The evidence that the annotation was based on, is provided on the gene's feature page (Aziz et al. 2008; Overbeek et al. 2013).

A metabolic pathway reconstruction (using KEGG reference pathways) was automatically done within the Seed Viewer of RAST using the RAST-annotated genes. The IGS-annotated genes were added to the constructed metabolic pathways in order to use as basis for comparing the IGS and RAST annotation servers. The reconstructed pathways based on the gene annotations were further compared to the KEGG pathways of the closes related species, *Mycoplasma synoviae* 53 (NC_007294), to determine the extent of missing genes due to incomplete sequencing or annotation of the Ms03 genome.

3.2.4 Identification of the Ms03 origin of replication

The origin of replication (*ori*) as well as the DnaA boxes were predicted using the web-based tool, Ori-Finder (Gao and Zhang 2008). DnaA boxes were identified using the consensus sequence TT(A/T/C)TCCACA and allowing 1 or 2 mismatches.

3.2.5 Identification and bioinformatic characterization of the Ms03 *opp* operon

Genes within the oligopeptide permease (*opp*) transporter operon (*oppABCDF*) were identified by using the operon of *Mycoplasma hominis* (YP_003302688.1) as query sequence in tBLASTn searches against the initial assembly of the Ms03 draft genome within CLC-bio. Glimmer v3.02 (Delcher et al. 1999; Delcher et al. 2007) was used to identify open reading frames (ORFs) in the contigs that contained the *opp* genes. Results were compared to that obtained after annotation of the draft Ms03 genome. The detailed analysis of functional motifs in the Opp proteins of *M. hominis* (Henrich et al. 1999) was used together with InterPro (Hunter et al. 2012) to identity the functional motifs in the Ms03 *oppABCDF* genes. Additionally, the similarity and identity between the Opp proteins of Ms03 and *M. hominis* was calculated using the EMBOSS WATER pairwise alignment algorithm (http://www.ebi.ac.uk/Tools/psa/emboss_water/). The program was set to use the BLOSUM62 matrix with a gap penalty of 10 and an extension penalty of 0.5.

3.3 Results and discussion

3.3.1 DNA isolation and confirming the identity of Ms03 genomic DNA

gDNA was isolated from Ms03 cultures to use as template in genome sequencing. Nanodrop analysis indicated adequate purity, but low concentration while the agarose gel electrophoresis confirmed the presence and integrity of the gDNA. No contamination with *M. struthionis* sp. nov. str. Ms01 and *Mycoplasma* sp. Ms02 gDNA was detected within the Ms03 gDNA samples by PCR. Multiple isolations were needed due to low yields. The purity and concentration of the gDNA was also assessed by inqaba biotec before genome sequencing. Due to low yield, whole genome amplification (GenomiPhi kit) was performed by inqaba biotec prior to genome sequencing.

3.3.2 Ms03 genome sequencing and assembly

The genome of Ms03 was sequenced with the GS FLX Titanium system by inqaba biotec. The first three Titanium runs produced 21.6 Mb of data that consist of 104 488 reads (Table 3.6). Using the GS *de novo* assembler, the reads could be assembled into 427 contigs larger or equal to 100 bp of which 212 contigs were equal or larger than 500 bp. In an attempt to decrease the number of large contigs, a fourth Titanium run was performed which produced 3.6 Mb of data that consist of 13 117 reads. In total, 25.2 Mb of data that consisted of 117 605 reads was generated (Table 3.6). This was assembled together with the initial data into 314 contigs larger or equal to 100 bp. The larger contigs (≥ 500 bp) were only reduced from 212 to 172 contigs. All contigs from the final assembly are listed in Appendix 2 Supplementary Table 2.1.

Table 3.6 The 454 sequencing data of the four 454 Titanium sequencing runs and the *de novo* assembly of Ms03 genome

454 Titanium sequencing run	**Data per run (Mb)**	**Reads**	
1st Run	15.4	71 383	
2nd Run	5.1	28 994	
3rd Run	1.1	4 111	
4th Run	3.6	13 117	
Newbler Assembly	**Total Mb Data**	**Total reads**	**Number of contigs ≥500**
Data from 1st run	15.4	71 383	221
1st and 2nd	20.5	100 377	220
1st, 2nd and 3rd	21.6	104 488	212
All data	25.2	117 605	172

The 172 large contigs (≥ 500 bp) had a total length of 845 856 bp with an average size of 4 917 bp (Table 3.7). The longest contig was 39 354 bp in length. The N50 contig size was 9 874 bp. This implies that more than half of the assembled sequence length (845 856 bp / 2 = 422 928 bp) is in contigs longer or equal to 9 874 bp. Q40 plus indicated that 99.3% (839 779 bp) of the bases had Phred-like consensus quality scores higher than 40. Therefore the probability of these bases being incorrect was less than 1 in 10 000. The Q39 minus was 0.72% indicating that 6 077 bp of the 845 856 bp had Phred-like consensus quality scores less than 40 thus had a higher probability of being incorrect (>1 in 10 000).

During *in vitro* culture, the Ms03 mycoplasma grows slowly and does not reach a high cell density. As a result it was difficult to isolate intact gDNA at high concentrations. This may have contributed to the quantity and quality of the reads generated. Whole genome amplification using the GenomiPhi kit was required to obtain an adequate concentration of gDNA for sequencing. This amplification step may have introduced a small amount of bias into the data (Pinard et al. 2006), that may have caused areas of the genome not to be sequenced. Problems with the generation of the sequencing library cannot be excluded since coverage of the individual contigs was irregular (uneven) as visualized in Newbler.

Table 3.7 Comparison of the *de novo* assemble genomes of Ms03, *Mycoplasma* sp. Ms02 and *M. struthionis* sp. nov. str. Ms01

***De novo* genome assemble parameters**	**Ostrich-infecting mycoplasma genome assembles**		
	Ms03	*Mycoplasma* sp. Ms02	*M. struthionis* sp. nov. str. Ms01
454 instrument used to sequence genome (read length)	GS FLX Titanium (400 bases)	GS FLX (200 bases)	GS20 (100 bases)
Amount sequencing data generated	25.2 Mb	15.8 Mb	30 Mb
Number of reads	117 605 reads	70 115 reads	345 863 reads
Number of contigs ≥500 bp	172 contigs	28 contigs	65 contigs
Number of bases (≥500 bp)	845 856 bp	895 119 bp	693 513 bp
Average contig size	4 917 bp	31 968 bp	10 669 bp
N50 contig size	9 874 bp	71 894 bp	44 778 bp
Largest contig size	39 354 bp	127 294 bp	86 684 bp
Q40 plus bases	839 779 bp, 99.3%	893 151 bp, 99.8%	691 611 bp, 99.7%
Q39 minus bases	6 077 bp, 0.72%	1 968 bp, 0.22%	1 902 bp, 0.3%
Number of contigs ≥100 bp	314 contigs	-	-
Number of bases (≥100 bp)	881 370 bp	-	-
Reference	This study	Steenmans (2010)	Pretorius (2009)

The genomes of ostrich-infecting mycoplasmas, *M. struthionis* sp. nov. str. Ms01 and *Mycoplasma* sp. Ms02 were previously sequenced by Pretorius (2009) and Steenmans (2010), respectively. In spite of the improvements in 454 sequencing chemistry (read length increased from 100 bases to 400 bases), the Ms03 genome assembly had more contigs and a lower N50 contig length compared to the assembly of *M. struthionis* sp. nov. str. Ms01 or *Mycoplasma* sp. Ms02 (Table 3.7). Results were also not significantly improved by generating more data. This may indicate that the full assembly of the Ms03 genome was not only influenced by the read length and chemistry (454 short gun) used, but also the sequence characteristics of the genome.

A+T rich genomes with numerous runs and repeats (as found in mycoplasmas) cause anomalies that result in incomplete genome drafts. A limitation of 454 sequencing technology is the inability to accurately detect runs of identical bases (mononucleotide repeats or homopolymer runs). During 454 sequencing, nucleotide bases that are incorporated into the newly synthesised DNA, release inorganic pyrophosphate. This release initiates a chain of enzymatic reactions leading to the production of a light signal. There is however no termination that would prevent multiple consecutive incorporation of the same base; therefore the technology relies on the intensity of the light signal to detect multiple incorporations of the same base (Shendure and Ji 2008; Kircher and Kelso 2010). Multiple incorporation of long homopolymer runs (>10 nt) are prone to error and may have contributed to the state of the A+T rich Ms03 genome assembly.

Additionally, repeats longer than the read length creates gaps since they are erroneously collapsed on top of one another during the assembly of reads into contigs (Treangen and Salzberg 2011). Attempts to complete the *Mycoplasma* sp. Ms02 genome by using thermal asymmetric interlaced (TAIL) PCR failed due to numerous long repeats within the contigs (Strydom 2013). If the incomplete assembly of the Ms03 draft genome was the result of repetitive regions, generating paired-end or longer (>1 000 bases) reads may solve the genome assembly problems. Paired-end reads are generated from a single DNA fragment of a fixed size, from which both ends are sequenced. An assembler can then use the expected distance and the orientation of the reads to reconstruct the genome (Treangen and Salzberg 2011). Additionally, sequencing data from platforms such as PacBio that generate longer read lengths (up to 23 000 bp) can be combined with the 454 reads. Combining data from PacBio with 454 data have been reported to overcome the assembly problems due to repeats and sequence bias (Koren et al. 2012).

The absence of a reference genome, that would make a comparative approach to the Ms03 genome assembly possible, further contributes to the incomplete state of the assembly. A number of complete mycoplasma genome sequences (Appendix 1) are available, but the lack of gene order conservation amongst *Mycoplasma* species, due to genome

rearrangements (Sirand-Pugnet et al. 2007a), implies that closely related species cannot serve as suitable reference genomes for the assembly of the Ms03 genome.

3.3.3 IGS and RAST annotations of the Ms03 draft genome

For annotation, a pseudomolecule was generated by joining all the contigs larger or equal to 100 bp together with spacers between them. The spacers introduced stop and start codons, in all six reading frames, between each contig. This aids in identifying possible gene fragments on the ends of contigs. The generated pseudomolecule was 852 084 bp in length and had a G+C content of 28.7% with base frequencies of 35.8% adenine (A), 14.4% cytosine (C), 14.3% guanine (G) and 35.5% thymine (T) (Table 3.8). This is similar to other members of the genus *Mycoplasma* which typically have genome sizes ranging from 564 kb to 1 359 kb and a G+C content of 24% to 40% (Chapter 2, Table 2.2).

Table 3.8 A summary of the Ms03 draft genome annotation as provided by the IGS and RAST annotation servers*

Pseudomolecule generated by the IGS annotation server	
Pseudomolecule length	852 084 bp
G+C content	28.7%
Base frequencies	(A) 35.8% (C) 14.4% (G) 14.3% (T) 35.5%
IGS annotation server	
Protein-encoding sequences	763 genes
Conserved hypothetical genes	46 genes (5.9%)
Hypothetical genes	227 genes (29.0%)
Total hypothetical genes	273 genes (34.8%)
tRNA	19
rRNA	3
Average gene length	998 nt
Percent coding	91.9%
Percent coding, tRNA, rRNA, or repeat	92.9%
Frequency of the translational start sites (excluding hypothetical genes)	ATG: 79.3% GTG: 18.5% TTG: 2.2%
RAST annotation server	
Protein-encoding sequences	635 genes
Total RNAs	24 (5 rRNAs and 19 tRNAs)
Number of subsystems	115
Hypothetical genes	244 genes (38.4%)

* The parameters for each annotations server differ, however, this is an output of the summary provided by the annotation server itself.

The IGS server annotated 763 genes in the Ms03 draft genome of which 273 were annotated as hypothetical genes (Table 3.8). The RAST server annotated 635 genes in the Ms03 draft genome of which 244 genes were annotated as hypothetical genes (Table 3.8). The number of predicted genes in the complete mycoplasma genomes ranges from 525 to 1 580 annotated genes, which is similar to what was annotated by IGS and RAST.

The most common start codon within the IGS annotation was ATG, the average gene length was 998 nt and 92.9% of the genome was predicted to be transcribed to RNA (Table 3.8). This information was not provided in the summary generated by the RAST analysis.

All the genes annotated with the IGS and RAST annotation pipeline are listed in a comparative manner (based on the location on the pseudomolecule) in Appendix 2 Supplementary Table 2.2. Annotated protein-encoding and non-coding RNA genes were named mnas (abbreviation for Ms03 chosen by the IGS server) followed by a number in the IGS annotation. In the RAST annotation protein-encoding and non-coding RNA genes were named peg (abbreviation for protein-encoding gene) and rna, respectively, followed by a number. Genes in IGS were numbered according to the position of the ORF on the pseudomolecule (i.e. the first gene was mnas_1, the second mnas_2, the third mnas_3), while in RAST this was not the case.

3.3.4 Comparison of IGS and RAST annotation

The IGS platform provided more information regarding the overall characteristics of the Ms03 draft genome than RAST (Table 3.8). The information was also easily available and included the number of conserved hypothetical genes, gene density (percent coding) and the frequency of the translational starting sites that were not provided by RAST. The only genome characteristics provided by RAST were number of protein-encoding sequences, number of RNAs, number of subsystems and total number of hypothetical genes (Table 3.8). It is possible to calculate all the additional parameters for the RAST annotation, except for number of conserved hypothetical genes (conserved hypothetical genes are defined as genes found in several phylogenetic lineages but have not been assigned function).

The actual Ms03 annotation results of the IGS and RAST servers differ with regards to:

- Number of protein-encoding and rRNA genes predicted
- Start and stop positions of some genes
- Names assigned to an annotated genes
- IGS assigns gene abbreviations to the annotated genes, where possible.

3.3.4.1 Genes annotated

The IGS annotation found 128 genes more than the RAST annotation while the hypothetical genes were 273 in the IGS annotation compared to 244 in the RAST annotation (Table 3.8). This implies that many of the genes annotated in IGS were discarded in the RAST annotation. Both IGS and RAST use the self-training method, Glimmer3, to identify genes. The results of the initial Glimmer3 searches were, however, implemented in a different manner as shown in Table 3.3 and Table 3.4. In the IGS pipeline, an initial Glimmer3 ORF search was performed to identify 1) non-overlapping long ORFs, 2) the frequency of the translational start sites and 3) an upstream position weight matrix for ribosomal binding sites. These results were used as input for the second Glimmer3 iteration. All ORFs predicted in the second run were used for downstream analysis (Galens et al. 2011). In the RAST pipeline, the ORF results from the initial Glimmer3 search were used to identify the 30 closest phylogenetic neighbours. All gene-candidates were then vetted to identify genes that were similar to proteins in the subsystems of the 30 closest phylogenetic neighbours. In RAST, a subsystem is a set of related functional roles to which genes were assigned. These functional roles frequently make up a metabolic pathway, a complex (like a ribosome) or a class of proteins. The ORFs with subsystem matches were retained and used to retrain Glimmer3. This process was repeated until no new gene-candidates were found to belong to a subsystem. At this point new gene-candidates were retained if they matched genes called in the 30 closest phylogenetic neighbours, but did not overlap significantly with previously called genes (Overbeek et al. 2013). Although both methods (IGS and RAST) had additional steps to limit false positive and negative gene calling, most genes were called from these Glimmer3 iterations. RAST retained most genes based on matches with subsystem genes (provided that they do not overlap) whereas IGS predicted a gene irrespective of functional assignment or overlap with each other which leads to a larger number of genes identified within IGS.

The IGS annotation found all three rRNAs (*5S rRNA, 16S rRNA* and *23S rRNA*). Within the RAST annotation, however, five rRNAs were found, three (rna.5, rna.8 and rna.12) corresponding to the *5S rRNA, 16S rRNA* and *23S rRNA* (Appendix 2 Supplementary Table 2.3) while the additional two were annotated as “Large subunit ribosomal RNA” genes. IGS predicted rRNA genes with the program RNAmmer that uses HMM datasets to locate rRNA genes. RAST, on the other hand, uses “search-for-rrna” followed by endpoint adjustments using a BLASTn search against a RNA database. The number of predicted rRNA genes was also similar to that annotated in complete mycoplasma genomes where the rRNA genes range from 3 to 8 (Chapter 2, Table 2.2).

Both annotation servers use the same program to identify tRNA and both annotated 19 tRNAs in the Ms03 draft genome (Appendix 2 Supplementary Table 2.4) which was lower

than in complete mycoplasma genomes where the tRNA genes range from 28 to 37 (Chapter 2, Table 2.2). This was probably a result of the incomplete Ms03 genome sequence.

3.3.4.2 Functional categories

Of the 763 genes predicted in the IGS annotation, 54% were assigned to functional categories (Figure 3.1). Of the genes not assigned to functional categories (46%), most (273 genes, 35%) were annotated as hypothetical genes while the remainder had a functional annotation but was not assigned to a functional category.

Of the 635 genes predicted in the RAST annotation, 42% (265 genes) were assigned to functional categories (Figure 3.2), this included five genes annotated as hypothetical. The remaining 58% (370 genes) was not assigned into a functional category. Of these, most (238 genes, 37%) were annotated as hypothetical while 132 genes (21%) had a functional annotation even though they were not assigned into functional categories.

IGS uses a more elaborate system to confirm and therefore assign function to a predicted gene. In IGS, functional categories are based on TIGRFAM roles and genes are assigned to a category using the program pFunc (uses BER and HMMER searches against TIGRFAM and Pfam databases as well as SignalP, LipoP, TMHMM and PROSITE searches), while in RAST functional categories are based on subsystems as defined by the FIGfam database and genes are assigned based on similarity searches. As a result more genes were assigned to functional categories using the IGS analysis.

The IGS category for protein synthesis (95 genes, 12.1%) and the category transport and binding proteins (86 genes, 11.0%) account for 23% of all genes within the Ms03 draft genome (Figure 3.1). Other major categories were cellular processes (46 genes, 5.9%), DNA metabolism (44 genes, 5.6%), cell envelope (39 genes, 5.0%), protein fate (36 genes, 4.6%) and energy metabolism (29 genes, 3.7%). The main categories in the RAST annotation were protein metabolism (125 genes, 19.7%), DNA metabolism (58 genes, 9.1%), RNA metabolism (40 genes, 6.3%) and carbohydrate metabolism (21 genes, 3.3%) (Figure 3.2). In spite of the absence of a cell wall in mycoplasmas, two genes were present in the subcategory cell wall and capsule within the RAST annotation. These genes, *tsaB* (peg.484) and *tsaE* (peg.483), are found in many mycoplasmas. According to the RAST annotation website, placement of these genes into another category is being considered. In IGS these genes were annotated as “tRNA threonylcarbamoyl adenosine modification protein YeaZ” and “tRNA threonylcarbamoyl adenosine modification protein YjeE” and placed into the category “Transport and binding proteins” and “Unknown: conserved”, respectively.

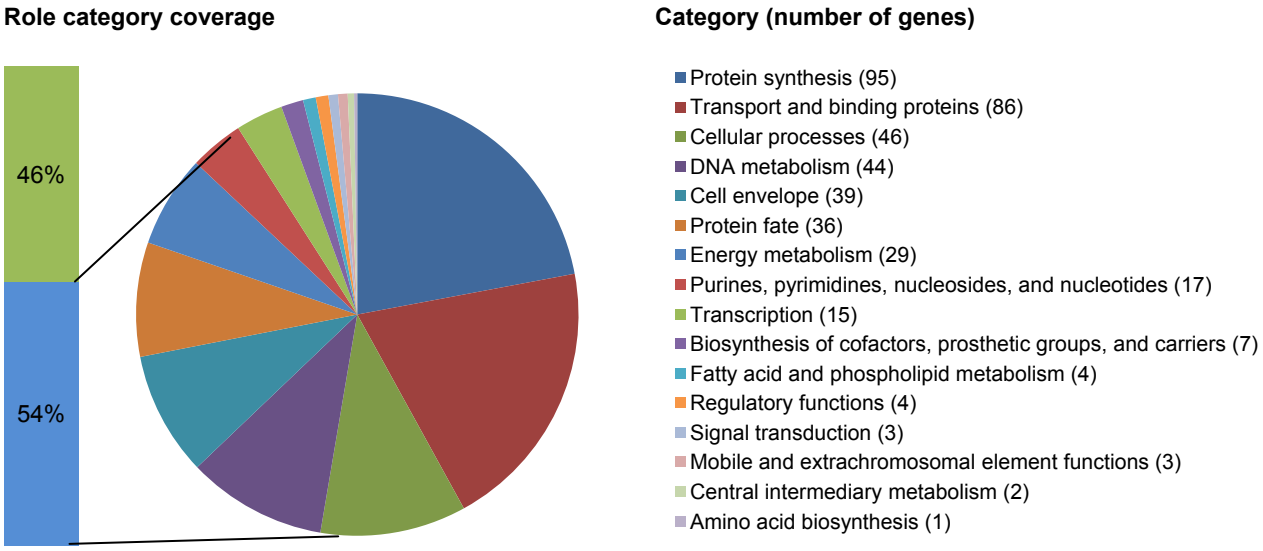


Figure 3.1 The IGS annotation role category breakdown of the Ms03 draft genome. The role categories are defined by TIGRfam. The bar-graph indicates the percentage genes assigned into a functional category in blue and the percentage genes not assigned into a functional category in green. The pie-chart displays the portions of each category. The categories are indicated on the right hand side with the number of genes in the category in brackets. The chart was generated in Microsoft Excel from the IGS data. Genes can be assigned into more than one functional category.

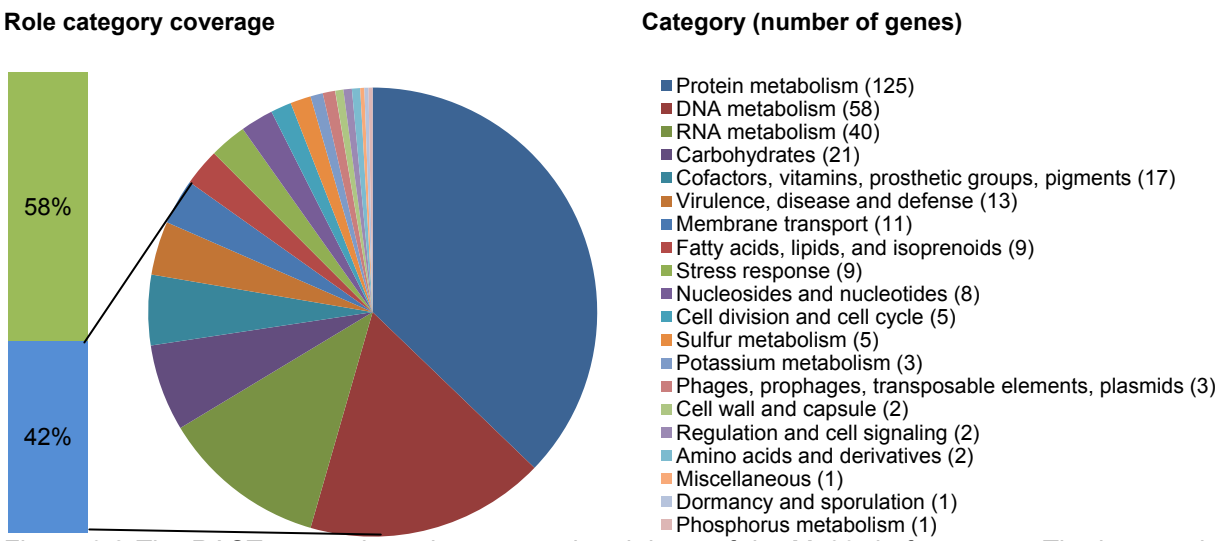


Figure 3.2 The RAST annotation role category breakdown of the Ms03 draft genome. The bar-graph indicates the percentage genes assigned into a functional subsystem category in blue and the percentage genes not assigned into a functional category in green. The pie-chart displays the portions of each category. The categories are indicated on the right hand side with the number of genes in the category in brackets. The chart was generated in Microsoft Excel from the RAST data. Genes can be assigned to more than one subsystem category.

It should be noted that role categories in the two annotation programs differ and that this complicates direct comparisons. For example the largest category in the IGS annotation was “protein synthesis” (95 genes) while the largest category for the RAST annotation was “protein metabolism” (125 genes). But the RAST category “protein metabolism” contained most of the genes within the IGS categories “protein synthesis” as well as “protein fate” (36 genes). For this reason the results from the two annotations were further compared using metabolic pathways.

3.3.5 Metabolic overview of Ms03

RAST allowed the construction of metabolic pathways for the annotated Ms03 draft genome using KEGG reference pathways. Construction of the metabolic pathways not only gave an overview of the metabolic capacity of Ms03 but also allowed a comparison of the annotation ability of IGS and RAST servers. The IGS annotation data was therefore added to each of the metabolic pathway reconstructions. The pathways of essential processes namely genome replication, cell division, RNA transcription, protein translation, energy metabolism and transport were used as basis for the comparison of the two annotation servers.

From the gene annotation it could be seen that genes were missing in both annotations. To determine the extent to which genes were possibly missing and therefore the extent to which the genome sequence was incomplete, genes annotated within the *M. synoviae* 53 genome were added to each metabolic pathway. The *M. synoviae* 53 genome is a closely related species to Ms03 (Chapter 2, Figure 2.1) for which the complete genome sequence as well as its annotation is available.

3.3.5.1 Genome replication and cell division in Ms03

Genome replication is initiated by the binding of DnaA proteins to conserved regions (called DnaA boxes) at the *ori* (Brown 2002). The *ori* is a well defined site at which the replication forks open and is usually located in the vicinity of the *dnaA* and *dnaN* genes. In Ms03, the *dnaA* gene (mnas_99 or peg.9), encoding for the chromosomal replication initiator protein, was located from 142 051 to 143 460 nt on the DNA pseudomolecule as annotated by both IGS and RAST (Appendix 2 Supplementary Table 2.5). The *dnaN* gene (mnas_100 or peg.10) that encodes for the beta subunit of DNA polymerase III (EC 2.7.7.7) was located adjacent to the *dnaA* gene. This arrangement is conserved across *Mycoplasma* species (Figure 3.3). The Ms03 *ori* was predicted with Ori-Finder to be located between 141 769 and 142 049 nt (length 281 nt) (Figure 3.4, Table 3.9). This region was between the mnas_98 (peg.8, RecD-like DNA helicase YrrC) and *dnaA* gene (indicated with an orange arrow in Figure 3.3) and had an A+T content of 85.05%. A high A+T content for the *ori* is typical since it reduces the energy associated with opening the double stranded DNA molecule for replication. This confirms the position of the predicted *ori*.

DnaA boxes are A+T rich 9 bp regions that are usually found in the *ori*. Ori-Finder identified two DnaA boxes in this region when allowing one mismatch to the consensus sequence (TT(A/T/C)TCCACA), however when relaxing the conditions to allow two mismatches, nine DnaA boxes could be identified (Figure 3.4, Table 3.9). This was in range with other *Mycoplasma* species in which up to 10 DnaA boxes were identified (Cordova et al. 2002; Lartigue et al. 2003; Lee et al. 2008; Pereyre et al. 2009).

A change in gene direction in the Ms03 draft genome annotations (both the IGS and RAST annotations) was observed at the *ori* (Figure 3.3). This is typically observed at the *ori* since it would prevent the head-on collision of replication and translation. Head-on collisions could stall replication and lead to a decrease in fitness (Price et al. 2005; Srivatsan et al. 2010; Lin and Pasero 2012). In the annotation of complete genomes, genes are numbered from the *dnaA* gene. The Ms03 annotation was, however, only a draft and the contig order was unknown therefore the gene numbers were not re-assigned based on the *dnaA* gene's position. The linear genome sequences of *Mycoplasma agalatiae* PG2, *Mycoplasma pulmonis* UAB CTIP, *Mycoplasma arthritidis* 158L3-1, *Mycoplasma hyopneumoniae* 232 and *M. synoviae* 53 were used to generate Figure 3.3 therefore the *dnaA* gene was the first gene in the genome sequence. As such the figure shows no genes upstream of the *dnaA* gene for these genomes, but bacterial genomes are obviously circular and there will be genes upstream. It should be mentioned that the *dnaA*, *dnaN* genes, the predicted *ori* as well as all the genes shown for Ms03 in Figure 3.3 were located on one contig (contig00350) therefore the arrangement and orientation of genes in this regions is not due to the linking of contigs into the pseudomolecule.

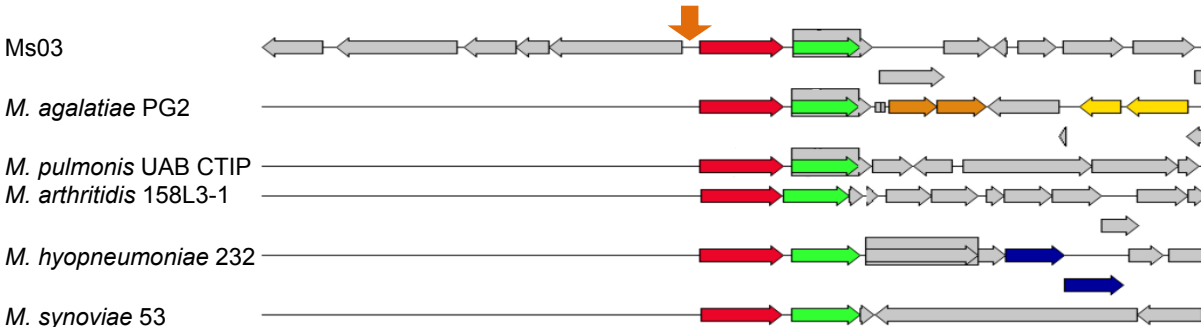


Figure 3.3 The annotated genes at the *ori*. The coloured genes indicate the same function with the red genes being *dnaA* and the green genes being *dnaN*. The grey boxes around the genes indicate similar function to the *dnaA* gene. The orange arrow indicates the *ori* as predicted with Ori-Finder. The linear genome sequences with the first gene being the *dnaA* gene of *M. agalatiae* PG2, *M. pulmonis* UAB CTIP, *M. arthritidis* 158L3-1, *M. hyopneumoniae* 232 and *M. synoviae* 53 were used to generate the figure. The figure was generated within RAST, results for the IGS annotation are not shown but the gene order, orientation and strand for all genes in both analyses are listed in Appendix 2, Supplementary Table 2.2.

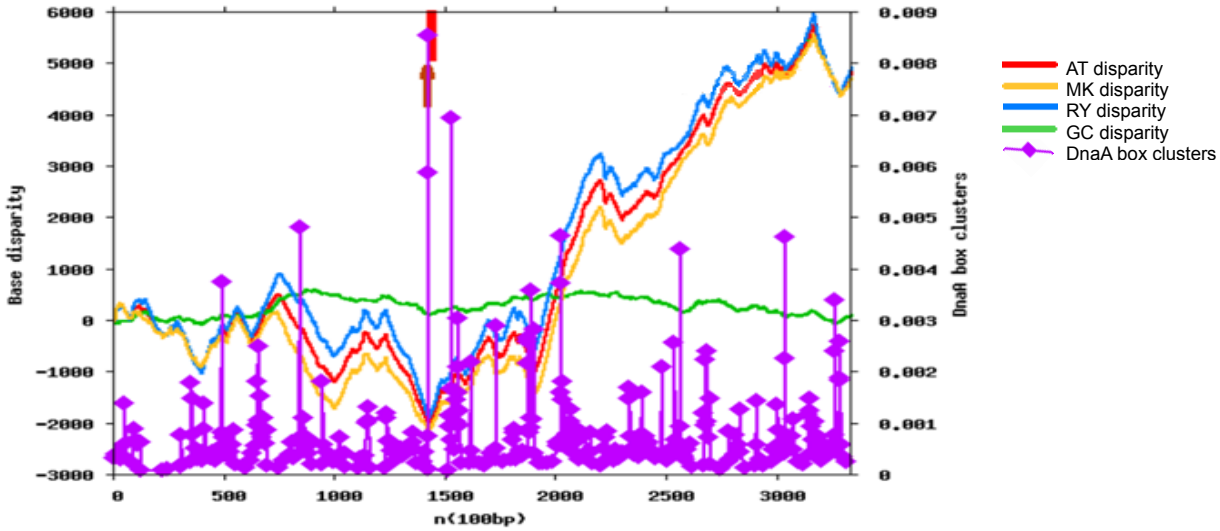


Figure 3.4 The Ms03 *ori* as predicted with Ori-Finder. The graph shows the Z-curves (AT, MK, RY and GC disparity curves) for the first 330000 nt of the Ms03 pseudomolecule sequence. The short vertical red line indicates the location of the indicator genes (*dnaA, dnaN*) and the short vertical dark orange arrow indicates the location of the predicted *ori*. The purple peaks with the diamonds indicate DnaA box clusters (allowing 1 mismatch from the consensus). The extremes for the AT disparity was 142 029 nt (minimum) and 316 379 nt (maximum), for MK disparity 142 053 nt (minimum) and 316 382 nt (maximum), for RY disparity 142 029 nt (minimum) and 316 379 nt (maximum) and for GC disparity 39 426 nt (minimum) and 86 718 nt (maximum). Abbreviations of nucleotides: A, adenine; T, thymine; G, guanine; C, cytosine; M, amino (A and C); K, keto (G and T); R, purine (A and G) and Y, pyrimidine (C and T).

Table 3.9 Characteristics of the Ms03 *ori* as predicted in Ori-Finder

Ori-Finder parameter	
***Ori* length**	281 nt
A+T content of *ori*	85.05%
The location of *ori* region	141 769..142 049 nt
The location of dnaA gene	142 051..143 460 nt
The number of DnaA boxes allowing one mismatch*	2
Location of DnaA boxes (sequence)	141 841 (TTATTCACA) 141 911 (TTATTCACA)
The number of DnaA boxes allowing two mismatches*	9
Location of DnaA boxes (sequence)	141 828 (TTTTGCACC) 141 841 (TTATTCACA) 141 853 (TTATTAACA) 141 888 (TTATTAACA) 141 911 (TTATTCACA) 141 922 (TTATTAACA) 141 964 (TAATTCACA) 142 033 (TTATACACT) 142 035 (TTTTATACA)

* Consensus sequence TT(A/T/C)TCCACA

Following site specific initiation of genome replication at the *ori*, the supercoiled DNA unwinds and the replication machinery is assembled. This proposed process in Ms03 is outlined in Figure 3.5. In total 14 DNA replication proteins were annotated in the Ms03 draft genome with both the IGS and RAST annotations (Figure 3.5 and Appendix 2 Supplementary Table 2.5). These include, DnaA, two helicase (DnaB), DNA primase (DnaG), single-stranded DNA-binding protein (SSB), ribonuclease HII (RNase HII), DNA ligase (LigA), DNA polymerase I (DpoI) and six genes for the DNA polymerase III holoenzyme. The six DNA polymerase III holoenzyme subunits annotated were beta (*dnaN* gene), delta (*holA* gene), delta' (*holB* gene), gamma/tau (*dnaX* gene) and two alpha (*dnaE* and *polC* genes) subunits (Figure 3.5). Binding of DnaA proteins to the *ori* initiates replication by causing melting or opening of the double stranded DNA chromosome. Helicase separates the double stranded DNA and the binding of the SSB proteins stabilize the single stranded parental DNA. Subsequently, the DNA polymerase holoenzyme synthesizes the new strand complementary to the single strand parent DNA. The two DNA polymerase III alpha subunits, PolC and DnaE are responsible for the synthesis of the leading and lagging strand, respectively. This would likely represent all the genes for replication within the Ms03 genome since the same DNA replication proteins were annotated in the *M. synoviae* 53 genome (Figure 3.5) as well as *Mycoplasma bovis* str. Hubei-1 (Li et al. 2011).

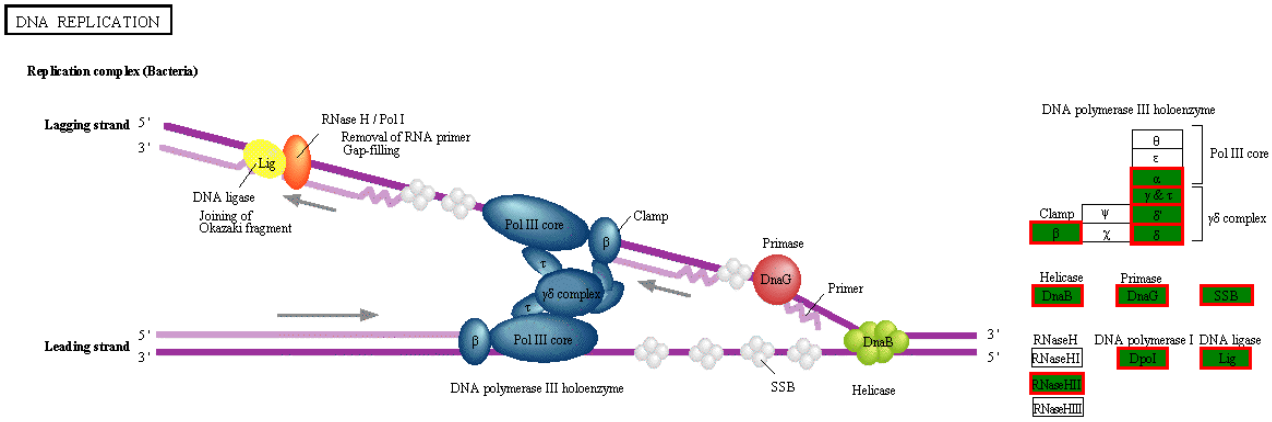


Figure 3.5 Proposed DNA replication in Ms03. The proteins (rectangular blocks on the right hand side) coloured in green were found in both the IGS and RAST annotations while proteins coloured in white were not found in the Ms03 draft genome annotations. For comparison, the annotated proteins of the complete *M. synoviae* 53 genome are indicated with red boxes. The schematic was downloaded from the Ms03 metabolic analysis in RAST and adapted to include the Ms03 IGS annotated proteins as well as the KEGG pathway of the *M. synoviae* 53 genome (msy03030). This schematic represents the KEGG reference pathway 03030.

During replication the circular chromosome of the mycoplasma is duplicated through DNA synthesis for which purine and pyrimidine bases are required. These bases can either be generated by *de novo* synthesis or recovered by the salvage pathways. Purine and pyrimidine metabolism in Ms03 is shown in Figure 3.6 and Figure 3.7 respectively using the KEGG reference pathways as basis with the purine and pyrimidine *de novo* synthesis pathways indicated with pink arrows.

In the *de novo* synthesis of purine bases, inosine monophosphate (IMP) is synthesised from glycine, aspartate, glutamine, formate, HCO_3^- and ribose-5-phosphate. AMP and GMP are subsequently synthesised from IMP. In the *de novo* synthesis of pyrimidine bases, UTP and CTP are synthesised from glutamine, HCO_3^- and aspartate. Like most mycoplasmas (Bizarro and Schuck 2007), Ms03 is unable to synthesised purine (Figure 3.6, Appendix 2 Supplementary Table 2.6) and pyrimidine bases (Figure 3.7, Appendix 2 Supplementary Table 2.7) *de novo* and have to rely on the respective salvage pathways.

For purine and pyrimidine metabolism in Ms03, the IGS and RAST analysis annotated the same genes (Figure 3.6, Figure 3.7, Appendix 2 Supplementary Table 2.6 and Supplementary Table 2.7). These genes encode for enzymes that form part of the purine and pyrimidine salvage pathways. The enzyme, EC 3.1.3.5, that catalyses several reversible reactions was annotated by RAST (peg.299) as a “5'-nucleotidase”. In IGS, however, this gene (mnas_205) was only annotated as a “calcineurin-like phosphoesterase family protein”. The nucleotidase enzyme belongs to the “calcineurin-like phosphoesterase family protein” and therefore was indicated as annotated in both the IGS and RAST analysis. This enzyme, 5'-nucleotidase, was however not annotated in the *M. synoviae* 53 genome although it is found within all other members of the hominis group (Vasconcelos et al. 2005).

The purine metabolic pathway (Figure 3.6) of *M. synoviae* 53 had two enzymes that were not annotated in the Ms03 draft genome, ribonucleoside-diphosphate reductase (EC 1.17.4.1) and purine-nucleoside phosphorylase (EC 2.4.2.1). In addition to these two enzymes, the pyrimidine metabolic pathway in *M. synoviae* 53 also had the enzymes thymidylate synthase (EC 2.1.1.45) and dCMP deaminase (EC 3.5.4.12) annotated that were absent in the Ms03 draft genome annotations (Figure 3.7). The ribonucleoside-diphosphate reductase enzyme and thymidylate synthase were also absent in *M. hominis* (KEGG pathways mho00230 and mho00240). Given the absence of the ribonucleoside-diphosphate reductase enzyme and thymidylate synthase in *M. hominis,* it is possible that only the absence of purine-nucleoside phosphorylase and dCMP deaminase in the Ms03 draft genome was due to incomplete genome sequence or the incomplete annotation (genes annotated as hypothetical).

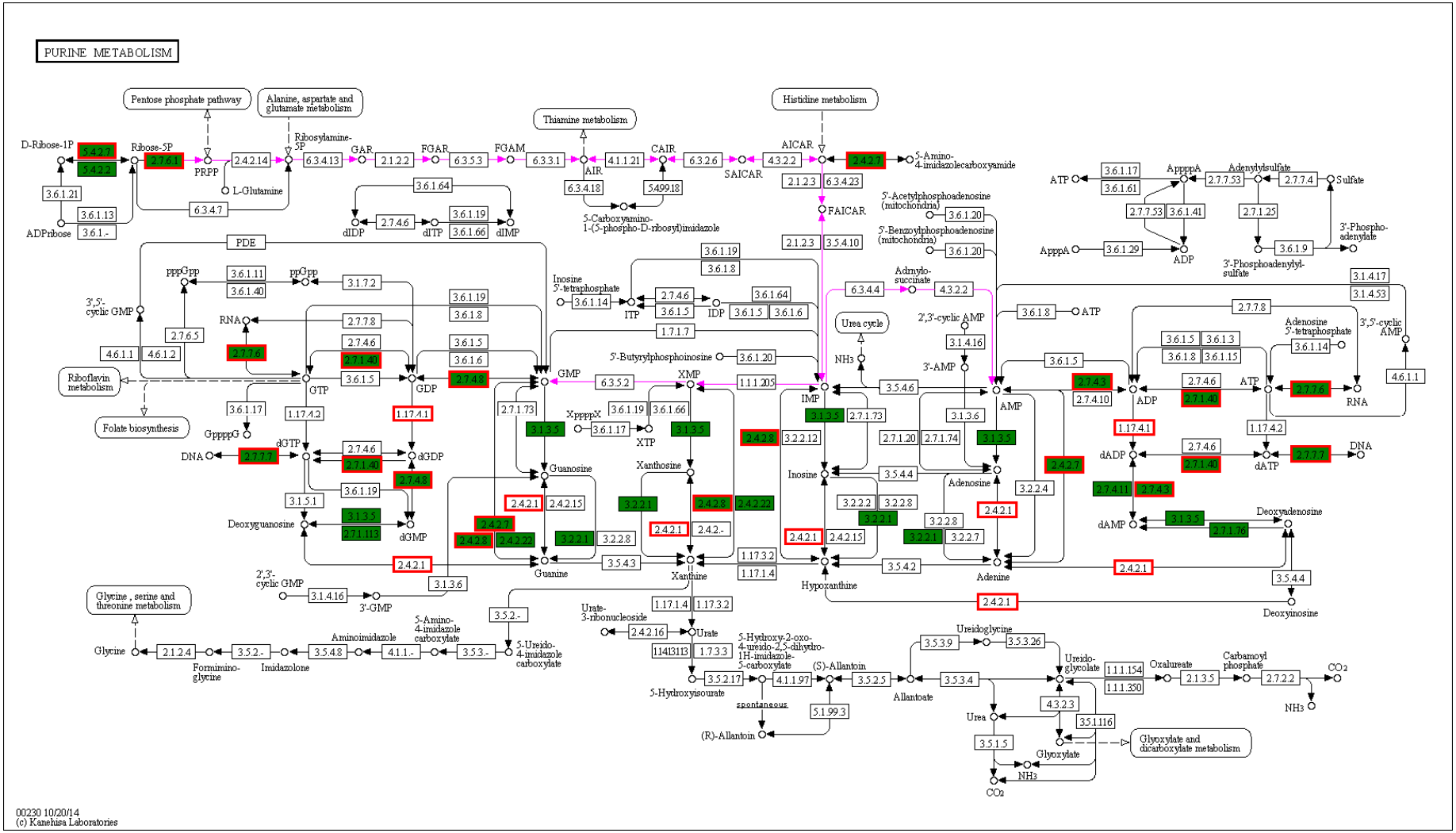


Figure 3.6 Proposed purine metabolism in Ms03. The enzymes coloured in green were found in both the IGS and RAST annotations while the enzymes in white were not found in the Ms03 draft genome annotations. For comparison, annotated enzymes of the complete *M. synoviae* 53 genome are indicated with red boxes. The purine *de novo* synthesis pathway is indicated with pink arrows. The schematic was downloaded from the Ms03 metabolic analysis in RAST and adapted to include the Ms03 IGS annotated proteins as well as the KEGG pathway of the *M. synoviae* 53 genome (msy00230). This schematic represents the KEGG reference pathway 00230.

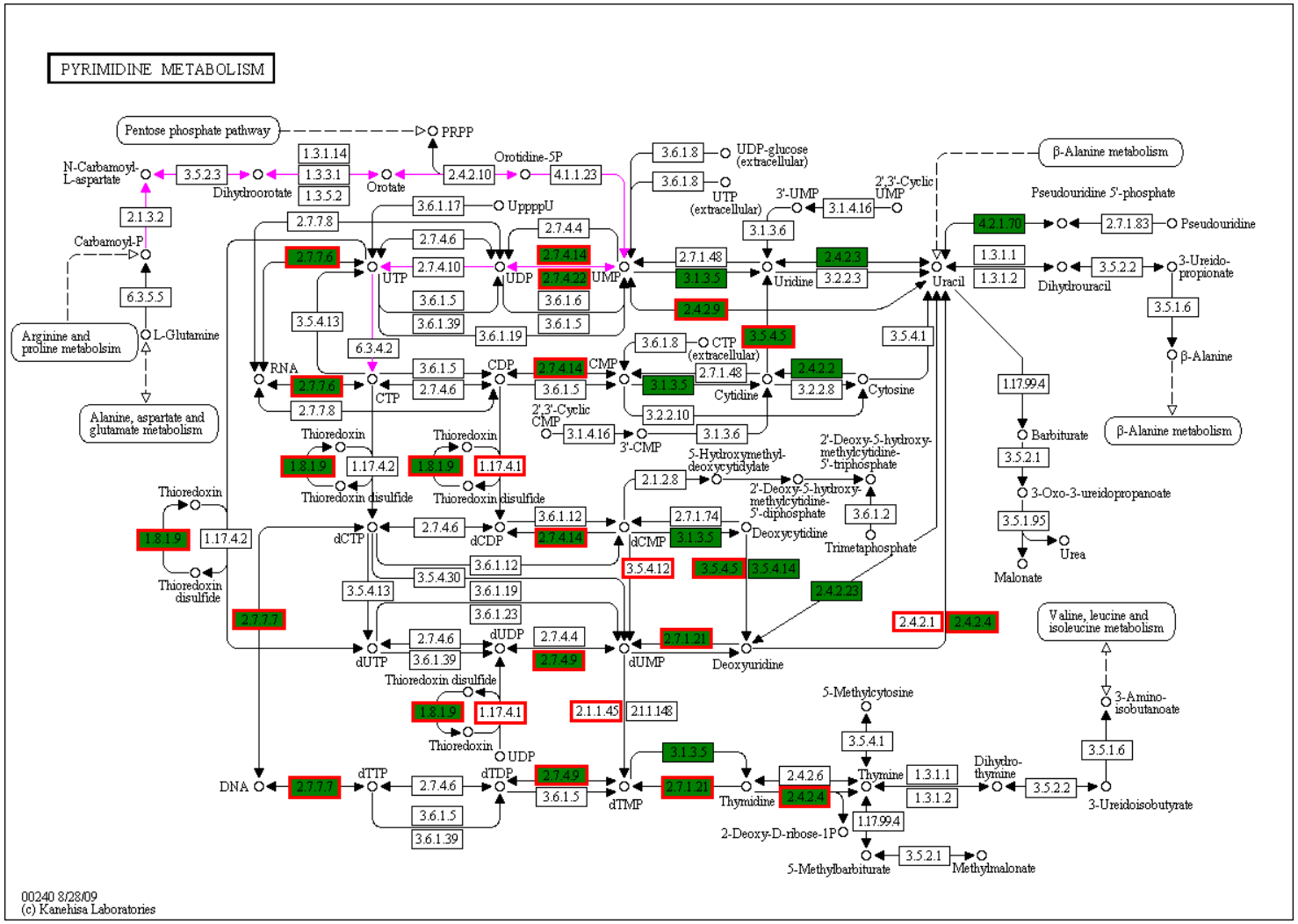


Figure 3.7 Proposed pyrimidine metabolism in Ms03. The enzymes coloured in green were found in both the IGS and RAST annotations while the enzymes in white were not found in the Ms03 draft genome annotations. For comparison, annotated enzymes of the complete *M. synoviae* 53 genome are indicated with red boxes. The pyrimidine *de novo* synthesis pathway is indicated with pink arrows. The schematic was downloaded from the Ms03 metabolic analysis in RAST and adapted to include the Ms03 IGS annotated proteins as well as the KEGG pathway of the *M. synoviae* 53 genome (msy00240). This schematic represents the KEGG reference pathway 00240.

The purine salvage pathways in Ms03 include the stepwise conversion of deoxyadenosine to dAMP, to dADP, to dATP as well as the stepwise conversion of deoxyguanosine to dGMP, to dGDP, to dGTP (Figure 3.6). The pyrimidine salvage pathways in Ms03 include the stepwise conversion of thymine to thymidine, to dTMP, to dTDP as well as the stepwise conversion of uracil to deoxyuridine, to deoxycytidine, to dCMP, to dCDP (Figure 3.7). The enzyme nucleoside diphosphate kinase (EC 2.7.4.6) that is needed to convert dTDP to dTTP and dCDP to dCTP was, however, not annotated. This enzyme is also absent in the *M. synoviae* 53 genome annotation as is the case in all *Mollicutes* (Bizarro and Schuck 2007). It has been reported that this function can be replaced by the glycolytic kinases, 6-phosphofructokinase (EC 2.7.1.11), phosphoglycerate kinase (EC 2.7.2.3) and pyruvate kinase (EC 2.7.1.40) although these enzymes have a lower substrate-specificity for nucleosides other than ATP (Pollack et al. 2002; Bizarro and Schuck 2007; Pachkov et al. 2007). The bases dATP, dGTP, dCTP and dTTP are incorporated into the newly synthesised DNA by the DNA polymerase III holoenzyme (EC 2.7.7.7).

The limited purine and pyrimidine metabolic capacity suggests that Ms03 would be dependent on environment-derived nucleotides and precursors. It also implies the need to import these bases for DNA synthesis.

Mycoplasmas have a low level of conservation of proteins related to cell division (Alarcon et al. 2007; Fisunov et al. 2011). Seven genes were annotated that encode for proteins involved in cell division with both the IGS and RAST annotations (Table 3.10). The genes are *ftsH*, *era*, *scpA, scpB* and *ftsY* that were annotated in both the IGS and RAST annotations. The FtsH protein plays a role in the quality control of integral membrane proteins and also influences cell division (Ito and Akiyama 2005), the GTP-binding protein, Era, may contribute to cell cycle regulation (Gollop and March 1991), the segregation and condensation protein A (ScpA) and B (ScpB) play a role in chromosomal segregation during cell division (Soppa et al. 2002), while the signal recognition particle-docking protein, FtsY, plays a role in protein targeting and also influences the cell cycle (Macao et al. 1997; Alarcon et al. 2007).

In addition to these five genes, the IGS annotation had two additional genes (mnas_318 and mnas_803) that were both annotated as “recF/RecN/SMC N-terminal domain proteins” (Table 3.10). The mnas_318 protein was annotated as “chromosome partition protein smc” (peg.446) in the RAST annotation (Table 3.10) while the ORF for the mnas_803 gene was not predicted in RAST. An additional protein, peg.192, was also annotated as “chromosome partition protein smc” in the RAST annotation, but was only annotated as a conserved hypothetical protein in the IGS annotation. The *ftsH* and *smc* genes are found in all mycoplasma genomes (Alarcon et al. 2007).

Table 3.10 Cell division genes annotated with the Ms03 draft genome

IGS				**RAST**	
Gene_id	**Gene name**	**Gene symbol**	**EC**	**Gene_id**	**Gene name**
mnas_95	ATP-dependent zinc metalloprotease FtsH domain protein	*ftsH*	3.4.24.-	peg.5	Cell division protein FtsH
mnas_167	Conserved hypothetical protein			peg.192	Chromosome partition protein smc
mnas_291	GTP-binding protein Era	*era*		peg.404	GTP-binding protein Era
mnas_318	RecF/RecN/SMC N terminal domain protein			peg.446	Chromosome partition protein smc
mnas_329	Segregation and condensation protein B	*scpB*		peg.455	Segregation and condensation protein B
mnas_330	Segregation and condensation protein A	*scpA*		peg.456	Segregation and condensation protein A
mnas_803	RecF/RecN/SMC N terminal domain protein				
mnas_606	Signal recognition particle-docking protein FtsY	*ftsY*		peg.115	Signal recognition particle receptor protein FtsY (alpha subunit) (TC 3.A.5.1.1)

Mycoplasmas divide by binary fission and the cell division protein. FtsZ is believed to play a central role in cell division by binary fission (Lluch-Senar et al. 2010). In Ms03, the *ftsZ* gene was, however, not found in either the IGS or RAST annotations. The *ftsZ* gene was also not found in the genome of *Mycoplasma mobile* (Jaffe et al. 2004) and in *Mycoplasma genitalium* a *ftsZ* null mutation was viable (Lluch-Senar et al. 2007). The *ftsZ* gene is, however, found in almost all mycoplasmas, including the close related species, *M. synoviae* 53. In the absences of the *fstZ* gene, proteins involved in cell movement were found to be implicated in cell division (Hatchel and Balish 2008; Erickson and Osawa 2010; Lluch-Senar et al. 2010). However, Ms03 has been characterised as a non-motile *Mycoplasma* species (Langer 2009) which questions the absence of the *ftsZ* gene. Furthermore, the cell division genes *mraZ* and *mraW* were also not annotated in the Ms03 draft genome. The *mraZ* and *mraW* genes are conserved in *Mycoplasma* species (Alarcon et al. 2007). The absence of these genes (*fstZ, mraZ* and *mraW*) may be a result of the incomplete genome sequence or these genes may be annotated as hypothetical genes. The possibility that one or more of these gene are not found in Ms03 can however not be excluded.

3.3.5.2 RNA transcription and protein translation in Ms03

RNA polymerase is essential for life and is found in all organisms including bacteria. This enzyme is responsible for the DNA-directed synthesis of RNA, i.e. transcription. In mycoplasma, the core subunits of RNA polymerase resemble that of other bacteria and are encoded by the *rpoA*, *rpoB* and *rpoC* genes. Genes for the alpha (*rpoA*), beta (*rpoB*) and beta' (*rpoC*) subunits of RNA polymerase, were annotated within both the IGS and RAST annotations of Ms03 (Figure 3.8, Appendix 2 Supplementary Table 2.8). In addition, the transcription initiation sigma factor (*rpoD* gene) that assists binding of the RNA polymerase to the promoter sequence was also annotated (Appendix 2 Supplementary Table 2.8). Only one sigma factor is common in mycoplasmas (Dorman 2011). The annotated genes correspond to that annotated in *M. synoviae* 53 (Figure 3.8). The omega (ω) subunit of RNA polymerase is absent in mycoplasmas while the delta (δ) is found in some but not all mycoplasma genomes (Madeira and Gabriel 2007). The delta subunit that is required for the expression of several stress-responses is found in *M. pneumoniae, M. genitalium, Mycoplasma gallisepticum* and *Ureaplasma urealyticum* (Fraser et al. 1995; Himmelreich et al. 1996; Madeira and Gabriel 2007).

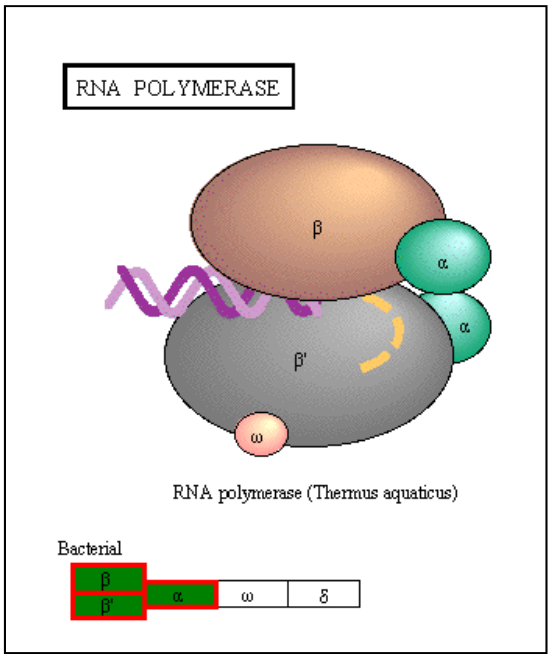


Figure 3.8 Proposed RNA polymerase in Ms03. The proteins coloured in green were found in both IGS and RAST annotations while the enzymes in white were not found in the Ms03 draft genome annotations. For comparison, annotated enzymes of the complete *M. synoviae* 53 genome are indicated with red boxes. The schematic was downloaded from the Ms03 metabolic analysis in RAST and adapted to include the Ms03 IGS annotated proteins as well as the KEGG pathway of the *M. synoviae* 53 genome (msy03020). This schematic represents the KEGG reference pathway 03020.

Transcription relies on the availability of nucleotides to be incorporated into the newly synthesised RNA molecule. As with DNA synthesis these nucleotides need to be salvaged from the environment because *de novo* synthesis of purines and pyrimidines is absent (Figure 3.6 and Figure 3.7). The purine salvage pathways for RNA bases in Ms03 include the stepwise conversion of adenine to adenosine, to AMP, to ADP, to ATP as well as the stepwise conversion of guanine to guanosine, to GMP, to GDP, to GTP (Figure 3.6). The pyrimidine salvage pathways for RNA bases in Ms03 include the stepwise conversion of uracil to uridine, to cytidine, to CMP, to CDP as well as the stepwise conversion of uridine to UMP, to UDP (Figure 3.7). The enzyme nucleoside diphosphate kinase (EC 2.7.4.6) that converts CDP to CTP and UDP to UTP was not annotated in the Ms03 draft genome and was also not present in the *M. synoviae* 53 genome annotation, although the function may be performed by other enzymes (Section 3.3.5.1). The bases ATP, GTP, CTP and UTP can be incorporated into the newly synthesised RNA by the RNA polymerase (EC 2.7.7.6). This limited metabolic capacity of Ms03 suggests that it depends on the import of ribonucleotide bases for RNA synthesis.

Ribosomes are responsible for the RNA-directed synthesis of proteins, i.e. translation. A ribosome is a RNA-protein complex that uses tRNA molecules as adaptors to translate RNA to protein. The IGS annotation as well as the RAST annotation had 50 ribosomal proteins annotated (Figure 3.9, Appendix 2 Supplementary Table 2.9). This is equivalent to the ribosomal proteins annotated in the *M. synoviae* 53 genome (Figure 3.9). Additionally the rRNA genes *5S rRNA, 16S rRNA* and *23S rRNA* were annotated (Figure 3.9, Appendix 2 Supplementary Table 2.3) as described above (Section 3.3.4.1).

The two annotation programs, IGS and RAST both use the same tRNA prediction tool, tRNAscan-SE (Aziz et al. 2008; Galens et al. 2011) and annotated the same tRNA genes in the Ms03 draft genome. In Ms03, 19 tRNA genes were found that cover 14 of the 20 standard amino acids as well as tRNA-SeC for the translation of the amino acid selenocysteine (Figure 3.10, Appendix 2 Supplementary Table 2.4). Selenocysteine is the 21st amino acid and is incorporated into selenoproteins (Commans and Böck 1999). The tRNA for selenocysteine is reported in *Mycoplasma alligatoris, Mycoplasma capricolum, Mycoplasma crocodyli, M. gallisepticum, M. genitalium, Mycoplasma hyorhinis, Mycoplasma suis* and *M. synoviae* 53 as well as many other bacteria (Commans and Böck 1999; Guimaraes et al. 2011).

The tRNA genes for alanine, aspartate, isoleucine, methionine, phenylalanine and proline were not annotated (Figure 3.10, Appendix 2 Supplementary Table 2.4). In *M. synoviae* 53, however, tRNA genes for all 20 amino acids as well as the tRNA-Sec are found (Figure 3.10).

In Ms03 aminoacyl-tRNA synthetase genes were annotated for 19 of the 20 amino acids (Figure 3.10, Appendix 2 Supplementary Table 2.10). Only 18 aminoacyl-tRNA synthetase genes were, however, annotated in the RAST annotation. The enzyme for serine tRNA synthesis (EC 6.1.1.11) was annotated in IGS however the ORF was not found in the RAST analysis. The RAST enzyme for glutamine tRNA synthesis (EC 6.1.1.24) was annotated in IGS as “tRNA synthetases class I (E and Q), catalytic domain protein“ and was therefore added to the IGS pathway in Figure 3.10. The enzyme for asparagine tRNA synthesis (EC 6.1.1.23 or EC 6.1.1.22) was not annotated in either annotation. In the aminoacyl-tRNA biosynthesis pathway of *M. synoviae* 53 (msy00970) the enzyme EC 6.1.1.22 was however annotated.

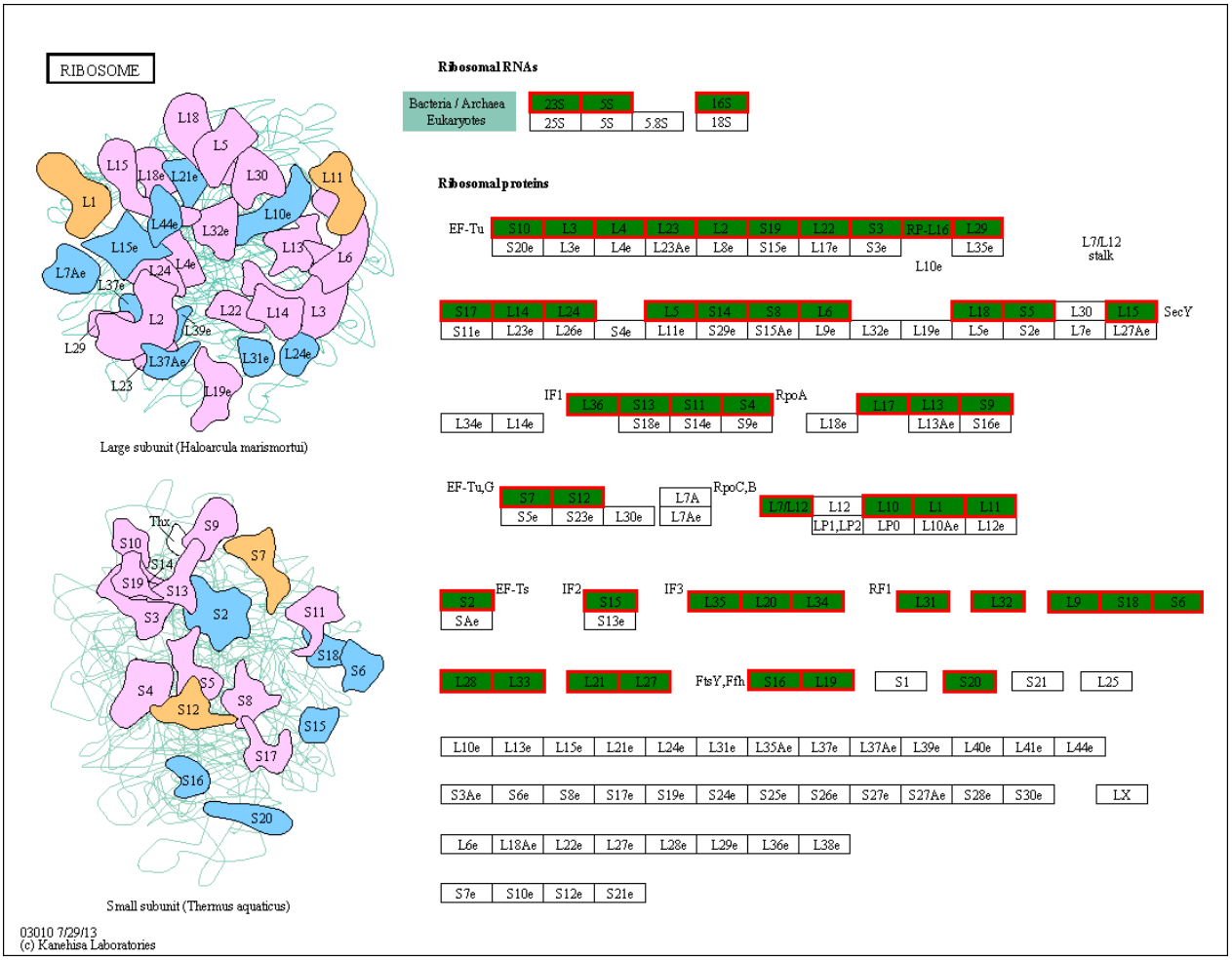


Figure 3.9 The proposed ribosomal genes in Ms03. The proteins coloured in green were found in both IGS and RAST annotations while the proteins in white were not found in the Ms03 draft genome annotations. For comparison, annotated proteins of the complete *M. synoviae* 53 genome are indicated with red boxes. The schematic was downloaded from the Ms03 metabolic analysis in RAST and adapted to include the Ms03 IGS annotated proteins as well as the KEGG pathway of the *M. synoviae* 53 genome (msy03010). This schematic represents the KEGG reference pathway 03010.

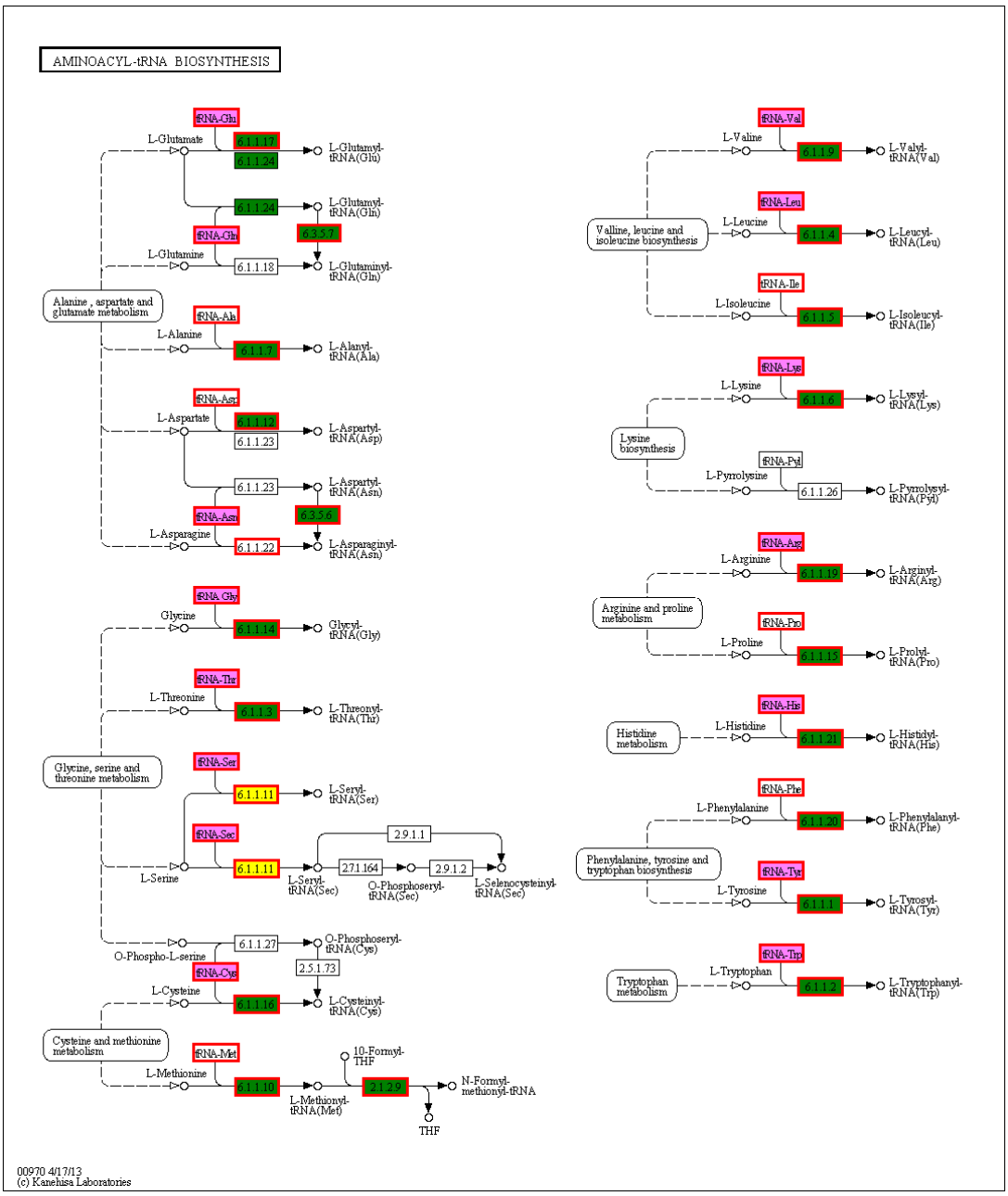


Figure 3.10 The proposed aminoacyl-tRNA biosynthesis in Ms03. The tRNA genes found with both IGS and RAST annotations are coloured in pink. The aminoacyl-tRNA synthetase enzymes coloured in yellow were only found within the IGS annotation, the enzymes coloured in green were annotated in both IGS and RAST annotations while the enzymes coloured in white were not found in the Ms03 draft genome annotations. For comparison, annotated tRNAs and enzymes of the complete *M. synoviae* 53 genome are indicated with red boxes. The schematic was downloaded from the Ms03 metabolic analysis in RAST and adapted to include the Ms03 IGS annotated proteins as well as the KEGG pathway of the *M. synoviae* 53 genome (msy00970). This schematic represents the KEGG reference pathway 00970.

The aminoacyl-tRNA biosynthesis of all standard 20 amino acids are essential to most organisms (Sheppard et al. 2008). The tRNA genes and aminoacyl tRNA synthetase enzymes that were not found in the Ms03 draft annotation do not overlap with each other (Figure 3.10). This indicates that it is likely that all the tRNA and synthetase enzymes for all 20 amino acids will be present in a completely assembled and annotated genome of Ms03. In general, tRNA genes range from 71 to 91 nt in size. The absences of these small genes are likely an effect of the incomplete assembly of the genome. Additionally, detailed manual curation of the hypothetical and annotated genes may lead to the assignment or re-assignment of function to some genes which could include some of the “missing” functions.

Translation, the synthesis of protein, relies on the availability of amino acids to incorporate/ build into the protein molecule. Annotated Ms03 genes for *de novo* amino acid synthesis paint a rather incomplete picture. Table 3.11 provides an overview of the amino acid metabolism pathways, with only six genes annotated in the Ms03 draft genome.

Table 3.11 Overview of the number of genes annotated in amino acid metabolic pathways in the Ms03 draft genome

KEGG pathway maps for amino acid metabolism (Reference pathway number)	**Number of genes annotated in the genome**	
	Ms03	***M. synoviae* 53**
Alanine, aspartate and glutamate metabolism (00250)	1	0
Glycine, serine and threonine metabolism (00260)	2	3
Cysteine and methionine metabolism (00270)	2	2
Valine, leucine and isoleucine degradation (00280)	0	1
Valine, leucine and isoleucine biosynthesis (00290)	0	0
Lysine biosynthesis (00300)	0	0
Lysine degradation (00310)	0	0
Arginine and proline metabolism (00330)	1	0
Histidine metabolism (00340)	0	0
Tyrosine metabolism (00350)	0	0
Phenylalanine metabolism (00360)	0	0
Tryptophan metabolism (00380)	0	0
Phenylalanine, tyrosine and tryptophan biosynthesis (00400)	0	0
Total number of genes	6	6

The enzyme L-aspartate:ammonia ligase (EC 6.3.1.1) that catalyses the conversion of aspartate to asparagines was the only enzyme annotated in the alanine, aspartate and glutamate metabolic pathway in both the IGS and RAST annotations of Ms03 (Table 3.12, Appendix 2 Supplementary Figure 2.1). In the genomes of *M. synoviae* 53 (Appendix 2 Supplementary Figure 2.1) and *M. hominis* no enzymes were annotated for this pathway.

The enzyme serine hydroxymethyltransferase (EC 2.1.2.1) was annotated in IGS but not in the RAST annotation (Table 3.12, Appendix 2 Supplementary Figure 2.2). In RAST the ORF for this gene was not predicted. This enzyme forms part of the glycine, serine and threonine metabolic pathway and catalyses the conversion of glycine to serine. In addition, phosphoglyceromutase (EC 5.4.2.12) was annotated within both the IGS and RAST annotations. Phosphoglyceromutase also takes part in glycolysis (Section 3.3.5.3). Both these enzymes were present in the genome of *M. synoviae* 53 (Appendix 2 Supplementary Figure 2.2).

Table 3.12 Genes annotated in the amino acid metabolic pathways of the Ms03 draft genome

Reaction (KEGG reaction number)	**IGS**			**RAST**	
	Gene_id	**Description (EC number)**	**Gene symbol**	**Gene_id**	**Description (EC number)**
Alanine, aspartate and glutamate metabolism (KEGG reference pathway 00250)					
ATP + NH3 + L-Aspartate <=> PPi + AMP + L-Asparagine (R00483)	mnas_499	Aspartate-ammonia ligase (EC 6.3.1.1)	*asnA*	peg.628	L-Aspartate: ammonia ligase (AMP-forming) (EC 6.3.1.1)
Glycine, serine and threonine metabolism (KEGG reference pathway 00260)					
5,10-Methylenetetrahydrofolate + Glycine + H2O <=> Tetrahydrofolate + L-Serine (R00945)	mnas_278	Serine hydroxymethyltransferase (EC 2.1.2.1)	*glyA*		
Cysteine and methionine metabolism (KEGG reference pathway 00270)					
H2O + ATP + L-Methionine <=> Phosphate + PPi + S-Adenosyl-L-methionine (R00177)	mnas_482	Methionine adenosyltransferase (EC 2.5.1.6)	*metK*	peg.617	ATP:L-methione S-adenosyltransferase (EC 2.5.1.6)
S-Adenosyl-L-methionine + DNA <=> S-Adenosyl-L-homocysteine + DNA 5-methylcytosine (R00380)	mnas_497	Modification methylase BanI (EC 2.1.1.37)	*banIM*	peg.626	DNA-cytosine methyltransferase (EC 2.1.1.37)
Arginine and proline metabolism (KEGG reference pathway 00330)					
Peptide with N-terminal proline + H2O <=> L-Proline + Peptide (R00135)	mnas_42	prolyl aminopeptidase (EC 3.4.11.5)	*pip*	peg.422	Proline iminopeptidase (EC 3.4.11.5)

The enzyme methionine adenosyltransferase (EC 2.5.1.6) that catalyses the conversion of methionine to adenosyl-L-methionine and the enzyme DNA-cytosine methyltransferase (EC 2.1.1.37) that catalyses the methylation of the cytosine base in DNA were the only two enzymes annotated in the cysteine and methionine metabolism pathway of the Ms03 draft genome (Table 3.12, Appendix 2 Supplementary Figure 2.3). These two enzymes were also found in the *M. synoviae* 53 genome (Appendix 2 Supplementary Figure 2.3).

In the arginine and proline metabolism pathway, the enzyme prolyl aminopeptidase (EC 3.4.11.5) was the only enzyme annotated in the Ms03 draft genome (Table 3.12, Appendix 2 Supplementary Figure 2.4). This enzyme catalyses the release of an N-terminal proline from a peptide and is also found in the *M. synoviae* 53 genome (Appendix 2 Supplementary Figure 2.4).

To conclude this section, annotation of the genes in the amino acid metabolism pathways indicated that Ms03 can convert aspartate to asparagine and glycine to serine and imply that most amino acids must be imported for protein synthesis. Ms03, like other mycoplasmas, is unable to synthesise *de novo* amino acids.

3.3.5.3 Energy metabolism in Ms03

From experimental evidence it is known that Ms03 uses glucose as its main energy source (Langer 2009). Ten glycolysis enzymes were annotated in the IGS annotation and nine in the RAST annotation (Figure 3.11, Appendix 2 Supplementary Table 2.11). The enzyme phosphoglycerate kinase (EC 2.7.2.3) was annotated in the IGS annotation, however in the RAST analysis, the ORF of this gene was not predicted. Two glucose-6-phosphate isomerase (EC 5.3.1.9) genes were annotated in both analyses although in IGS one of these was only annotated as a “phosphoglucose isomerase family protein”. Enzyme II of the phosphoenolpyruvate-dependent sugar phosphotransferase transport system (PTS) (EC 2.7.1.69) was included in the glycolysis pathway. The PTS system phosphorylates extracellular glucose and other sugars upon transport into the cell. This circumvents the need for glucokinase (EC 2.7.1.2) that was not annotated in Ms03 as is the case in *M. synoviae* 53 (Figure 3.11).

Both the IGS and RAST annotations lack the gene for phosphofructokinase (EC 2.7.1.11) that converts β-D-fructose-6-phosphate to β-D-fructose-1,6-bisphosphate (Figure 3.11). In the *M. synoviae* 53 genome, this enzyme was annotated, however in the genome of *M. hominis* phosphofructokinase is also absent (KEGG mho0010) (Pereyre et al. 2009). *M. hominis* is a non-glycolytic mycoplasma that uses arginine as its main energy source but can also utilize glucose. In *M. hominis,* the steps to circumvent the need for phosphofructokinase form part of the pentose phosphate pathway (KEGG mho00030). The

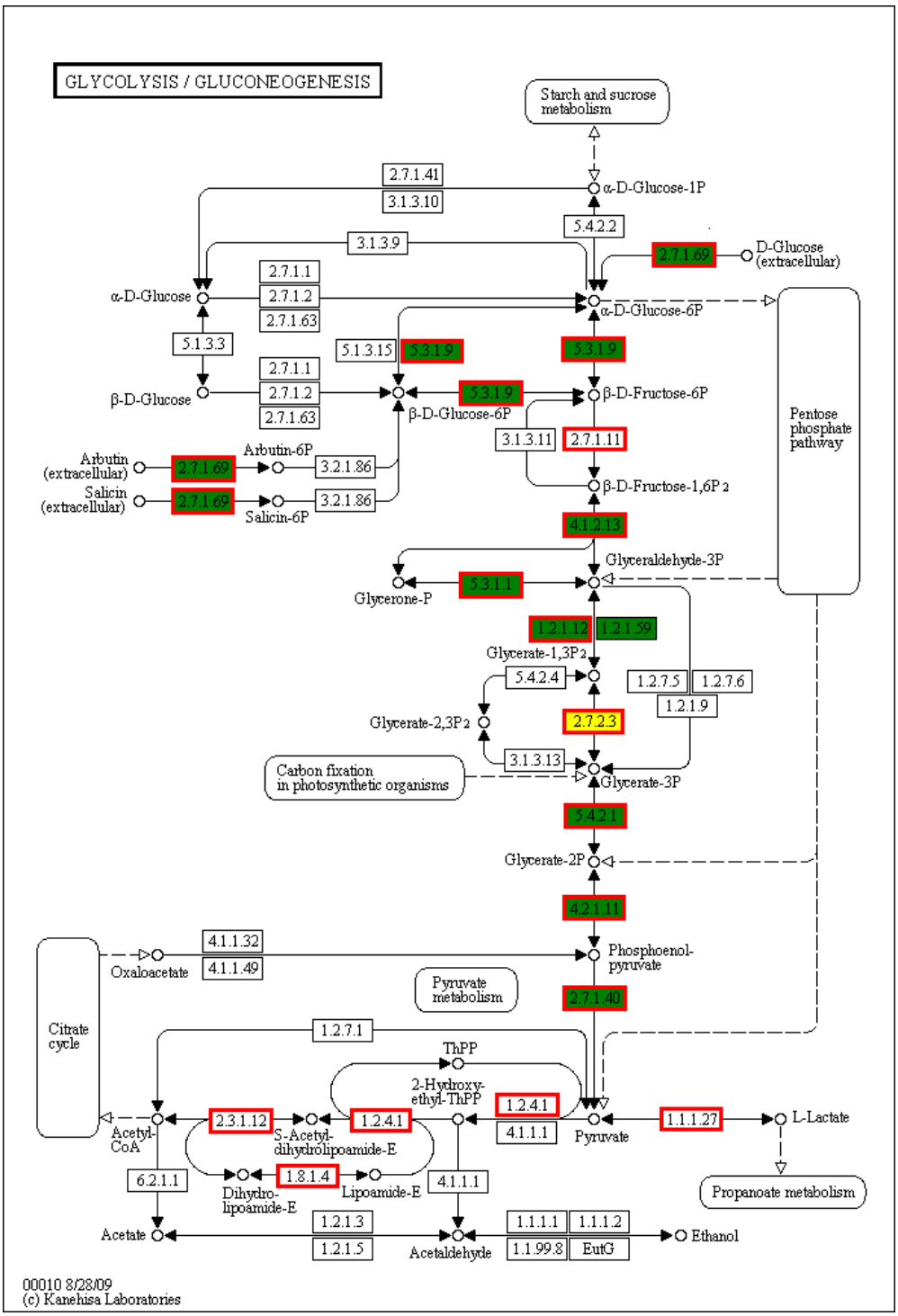


Figure 3.11 Glycolysis / gluconeogenesis pathway in Ms03. The enzymes coloured in yellow were only found within the IGS annotation, enzymes coloured in green were found in both IGS and RAST annotations while the enzymes in white were not found in the Ms03 draft genome annotations. For comparison, annotated enzymes of the complete *M. synoviae* 53 genome are indicated with red boxes. The schematic was downloaded from the Ms03 metabolic analysis in RAST and adapted to include the Ms03 IGS annotated proteins as well as the KEGG pathway of the *M. synoviae* 53 genome (msy00010). This schematic represents the KEGG reference pathway 00010.

pentose phosphate pathway for MS03 is presented in Figure 3.12 (Appendix 2 Supplementary Table 2.12). In this pathway, β-D-fructose-6-phosphate from glycolysis is converted to xylulose-5-phosphate (transketolase EC 2.2.1.1). Xylulose-5-phosphate can follow one of two paths to continue with glycolysis. Firstly xylulose-5-phosphate can be converted directly to glyceraldehyde-3-phosphate by phosphoketolase (EC 4.1.2.9). Secondly xylulose-5-phosphate can be converted to ribulose-5-phosphate by ribulose-5-phosphate 3-epimerase (EC 5.1.3.1), ribulose-5-phosphate to ribose-5-phosphate by ribose-5-phosphate isomerase B (EC 5.3.1.6) and ribose-5-phosphate to glyceraldehyde-3-phosphate by transketolase (EC 2.2.1.1). Glyceraldehyde-3-phosphate can then enter glycolysis. Although most of the enzymes to circumvent the need for phosphofructokinase were annotated in Ms03, transketolase (EC 2.2.1.1) was not annotated. In order to catalyse the degradation of glucose to pyruvate, either phosphofructokinase or transketolase would be needed. Although these “missing” genes could likely be due to the incomplete assembly that complicates accurate ORF calling or the incomplete annotation with genes annotated as hypothetical, the possibility that one of these genes is absent in the Ms03 genome cannot be excluded.

Pyruvate is further metabolised to D-lactate (Figure 3.13, Appendix 2 Supplementary Table 2.13). In contrast to *M. synoviae* 53*,* enzymes to convert pyruvate to acetyl-CoA were not annotated in Ms03 (Figure 3.11 and Figure 3.13). These enzymes were however also not annotated in *M. hominis* (KEGG mho0010)*.* Additionally no genes were annotated with roles in the citric acid cycle pathway, which suggest that the citric acid cycle is absent in Ms03. This is expected as the citric acid cycle is absent in most mycoplasmas (Razin et al. 1998; Jaffe et al. 2004; Chen et al. 2012).

The pentose phosphate pathway that regenerates NADPH from $NADP^+$ was incomplete (Figure 3.12). The enzymes glucose-6-phosphate dehydrogenase (EC 1.1.1.49) and 6-phosphogluconate dehydrogenase (EC 1.1.1.44) that produce NADPH were not annotated in the Ms03 genome as in the case of *M. synoviae* 53. The pentose phosphate pathway is, however, also incomplete in many other mycoplasmas (Himmelreich et al. 1996; Razin et al. 1998; Halbedel et al. 2007; Pereyre et al. 2009). The absence of these genes is therefore not necessarily related to the incomplete genome sequence of Ms03.

As in the *M. synoviae* 53 genome, no genes were annotated within the Ms03 draft genome for the arginine deiminase pathway and the urea cycle. This was consistent with the finding that Ms03 is unable to utilize arginine and urea as energy source during *in vitro* culture (Langer 2009).

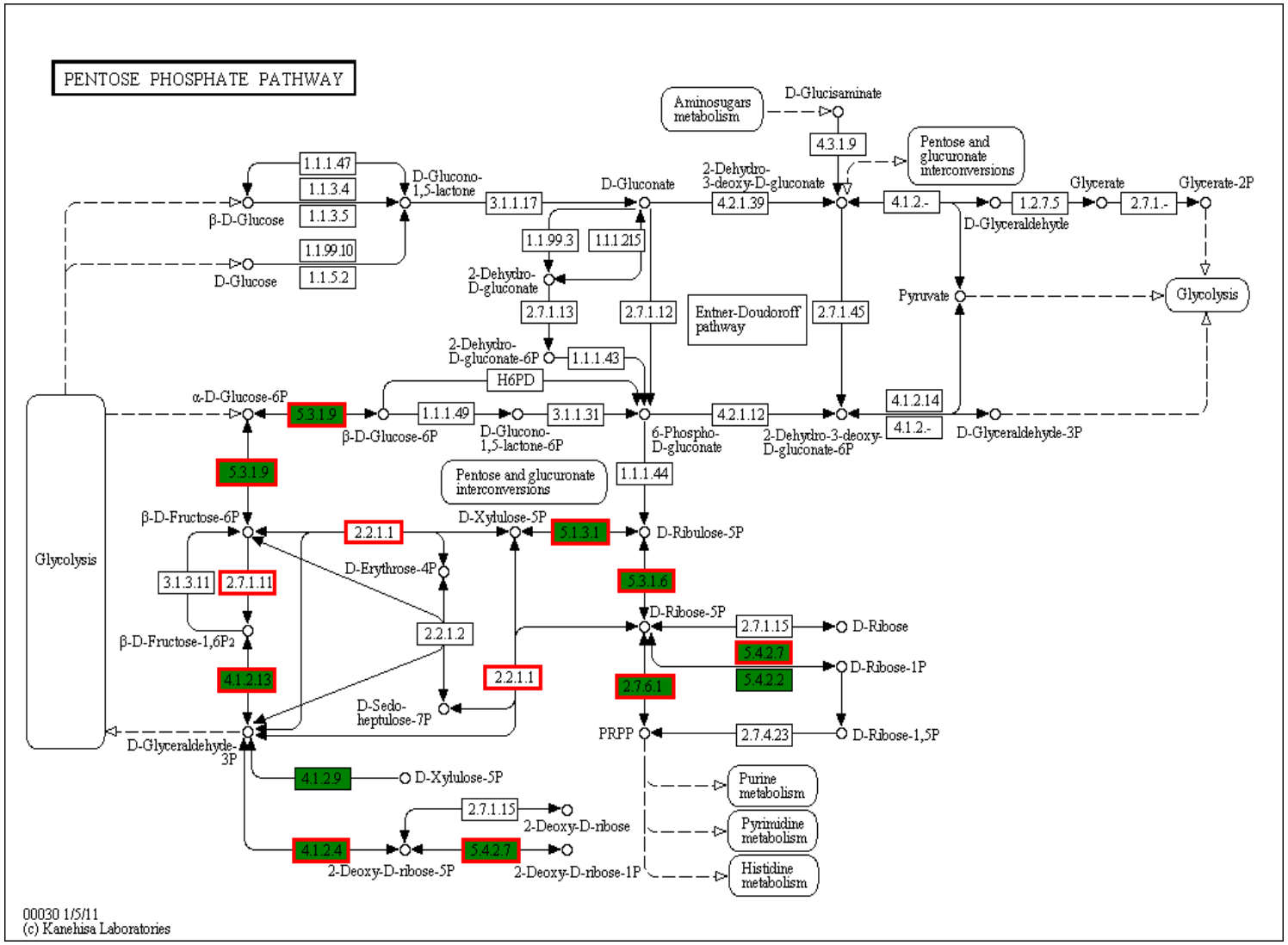


Figure 3.12 The proposed pentose phosphate pathway in Ms03. The enzymes coloured in green were found in both the IGS and RAST annotations while the enzymes in white were not found in the Ms03 draft genome annotations. For comparison, annotated enzymes of the complete *M. synoviae* 53 genome are indicated with red boxes. The schematic was downloaded from the Ms03 metabolic analysis in RAST and adapted to include the Ms03 IGS annotated proteins as well as the KEGG pathway of the *M. synoviae* 53 genome (msy00030). This schematic represents the KEGG reference pathway 00030.

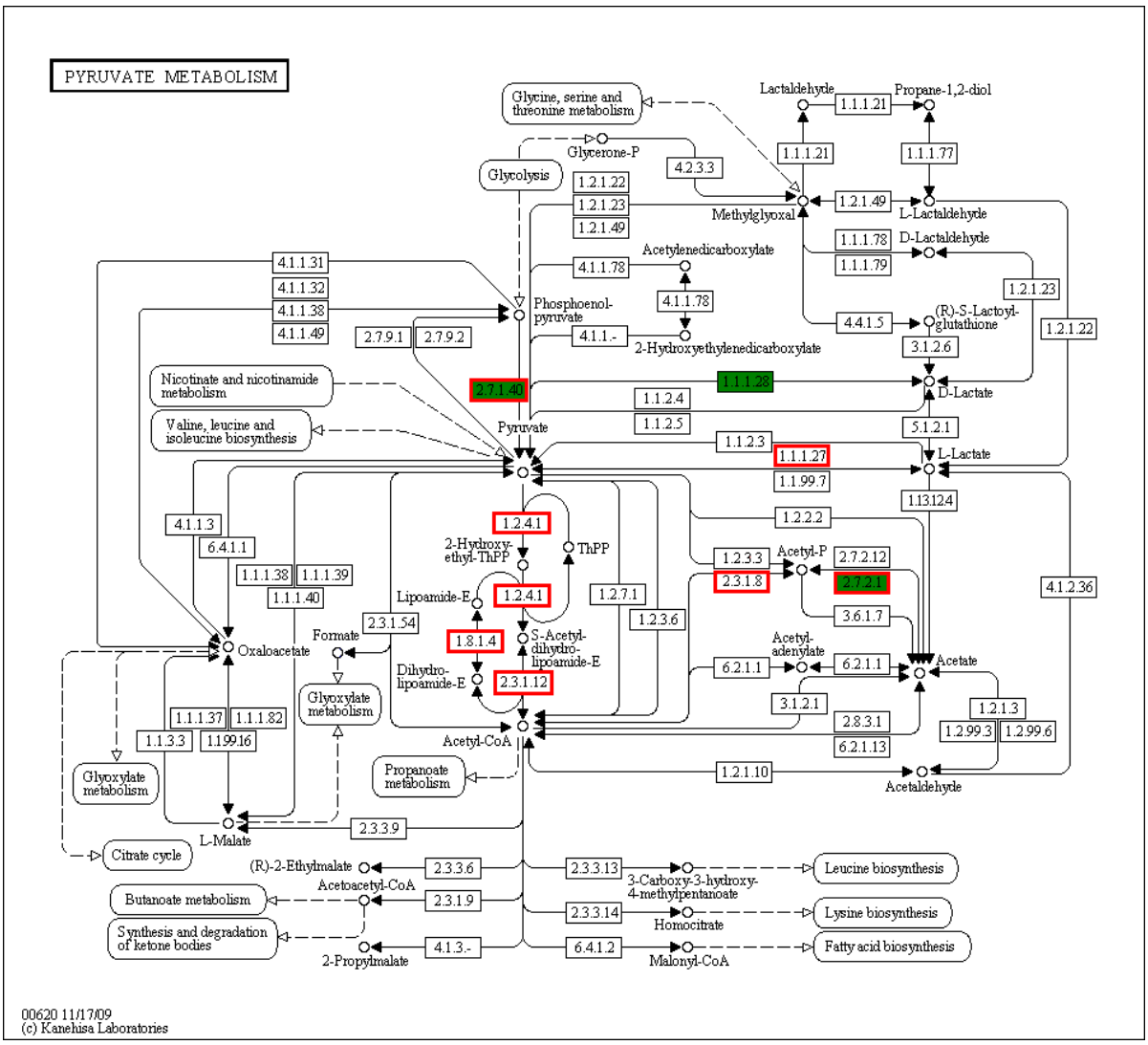


Figure 3.13 The proposed pyruvate metabolism in Ms03. The enzymes coloured in green were found in both IGS and RAST annotations while the enzymes in white were not found in the Ms03 draft genome annotations. For comparison, annotated enzymes of the complete *M. synoviae* 53 genome are indicated with red boxes. The schematic was downloaded from the Ms03 metabolic analysis in RAST and adapted to include the Ms03 IGS annotated proteins as well as the KEGG pathway of the *M. synoviae* 53 genome (msy00620). This schematic represents the KEGG reference pathway 00620.

In mycoplasma, ATP synthase is thought to be involved in ATP hydrolysis and the maintenance of the electrochemical gradient (Rechnitzer et al. 2011). IGS annotated 15 genes that encode for ATP synthase subunits, while RAST annotated 10 genes (Figure 3.14 and Table 3.13). All mycoplasmas have a typical F_1F_0 ATP synthase that consists of eight conserved genes found within an operon (Beven et al. 2012). In the IGS annotation, six of the ATP synthase subunit genes were found adjacent to each other (*atpB, atpE, atpF, atpH, atpA* and *atpG;* mnas_84 to mnas_89). The *atpD* and *atpC* (mnas_767 and mnas_766) were found next to each other but in a separate location (Table 3.13). The first six genes are located on contig00408 while the *atpD* and *atpC* genes were on contig00433. In a complete assembly of the Ms03 genome, these genes would likely be located next to each other in an

operon system as in other *Mycoplasma* species. The ORFs for the IGS genes, *atpG* (mnas_89), *atpD* (mnas_765), *atpD* (mnas_767), mnas_630 and mnas_643 were not predicted in the RAST annotation. Extra copies of the ATP synthase subunits (*atpA* and *atpD*) have been reported in many mycoplasmas although it is not the case in most bacteria (Beven et al. 2012).

3.3.5.4 Acquiring the necessary building blocks to maintain life: Transport

The transporters annotated in the Ms03 draft genome are illustrated in Figure 3.14 and the proteins involved are listed in Appendix 2 Supplementary Table 2.14. The F_1F_0 ATP synthase (Section 3.3.5.3, Table 3.13) is a transmembrane protein complex (Figure 3.14 A) and has been reported to play a role in ATP hydrolysis and maintaining the electrochemical gradient (Rechnitzer et al. 2011).

Putative potassium, magnesium and cation (possibly for copper, lead, cadmium, zinc and mercury ions) importers were annotated in both the Ms03 IGS and RAST annotations (Figure 3.14 B-D, Appendix 2 Supplementary Table 2.14). These ions play an important role as cofactors in enzymatic reactions as well as in maintaining an electrochemical gradient for osmotic balance.

Table 3.13 The annotated genes for ATP synthase (EC 3.6.3.14) in the Ms03 draft genome

IGS annotation			**RAST annotation**	
Gene_id	**Gene name**	**Gene symbol**	**Gene_id**	**Gene name**
mnas_84	ATP synthase A chain	*atpB*	peg.590	ATP synthase F_0 sector subunit a
mnas_85	ATP synthase F_0, C subunit	*atpE*	peg.591	ATP synthase F_0 sector subunit c
mnas_86	ATP synthase F_0, B subunit	*atpF*	peg.592	ATP synthase F_0 sector subunit b
mnas_87	ATP synthase F_1, delta subunit	*atpH*	peg.593	ATP synthase delta chain
mnas_88	ATP synthase F_1, alpha subunit	*atpA*	peg.594	ATP synthase alpha chain
mnas_89	ATP synthase family protein	*atpG*		
mnas_293	ATP synthase alpha/beta family, nucleotide-binding domain protein		peg.405	ATP synthase alpha chain
mnas_294	ATP synthase F_1, beta subunit	*atpD*	peg.406	ATP synthase beta chain
mnas_614	ATP synthase alpha/beta family, nucleotide-binding domain protein	*atpA*	peg.141	ATP synthase alpha chain
mnas_615	ATP synthase F_1, beta subunit	*atpD*	peg.142	ATP synthase beta chain
mnas_765	ATP synthase F_1, beta subunit	*atpD*		
mnas_766	ATP synthase F_1, epsilon subunit	*atpC*	peg.347	ATP synthase epsilon chain
mnas_767	ATP synthase, subunit beta	*atpD*		
mnas_630	Putative ATP synthase alpha chain domain protein			
mnas_643	Putative ATP synthase alpha chain domain protein			

The phosphoenolpyruvate-dependent sugar phosphotransferase transport system (PTS) imports and phoshorylates sugars (Clore and Venditti 2013). This transport involves a series of phosphorylation events of enzyme I (EI), histidine-containing phospho-carrier protein (HPr) and enzyme II (EII). EII consists out of three components, IIA, IIB and IIC. A high-energy phosphate moiety from phosphoenolpyruvate (PEP) is transferred to EI in the cytoplasma, subsequently to HPr and to IIA (cytoplasma) then to IIB which is associated with the membrane embedded IIC component. Finally the transmembrane domain IIC catalyzes the coupled translocation and phosphorylation (phosphate group from IIB) of the incoming sugar (Clore and Venditti 2013). In both the IGS and RAST annotations of the Ms03 draft genome, two HPr, glucose/glucosamine/beta-glucoside-specific IIA, IIB and IIC components (EC 2.7.1.69) and an additional putative PTS system glucose-specific enzyme IIB component were annotated (Figure 3.14 E, Appendix 2 Supplementary Table 2.14). Furthermore, an enzyme was annotated as ”HPr(Ser) kinase/phosphatise” (EC 2.7.1.-) which could possibly be the Ms03 EI. The glycolytic mycoplasma *M. genitalium* had two complete glucose and fructose PTS systems while in the non-glycolytic mycoplasmas, *M. hominis* and *Ureaplasma parvum,* EI is absent (Glass et al. 2000; Pereyre et al. 2009). *In vitro* culturing has shown that Ms03 mainly uses glucose as energy source (Langer 2009) and with the absence of glucokinase (EC 2.7.1.2) in glycolysis (Figure 3.11), the phosphorylation of extracellular glucose could be required by Ms03.

ATP-binding cassette (ABC) transporters are present in all organisms and play an important role in the active transport of substrate across the cell membrane (Rice et al. 2014). All ABC transporters have two transmembrane domains that form a pore and two intracellular domains that bind and hydrolyze ATP thereby supplying energy for substrate transport. Additionally, ABC importers in bacteria have an extracellular substrate-binding domain (Rice et al. 2014).

Energy-coupling factor (ECF) transporters are a type of ABC transporter that mediate the uptake of micronutrients. ECF transporters are present in approximately 50% of prokaryotic species but are particularly abundant in the phylum *Firmicutes* of Gram-positive bacteria (Slotboom 2014). The transporter consists of an integral membrane protein, EcfT (the T component) and integral membrane substrate-binding domain (the S component) as well as two similar or identical cytosolic ATP synthases, EcfA and EcfA′ (the A components). A substrate-binding domain for different substrates can interact with the same T and A components. In the Ms03 draft genome, an ECF transporter was annotated (Figure 3.14 F). Two EcfA proteins were found in both the IGS and RAST annotations. An EcfT protein was annotated within IGS, however in the RAST annotation, this ORF was not predicted. Only one S component was annotated with both programs. The annotated S component was

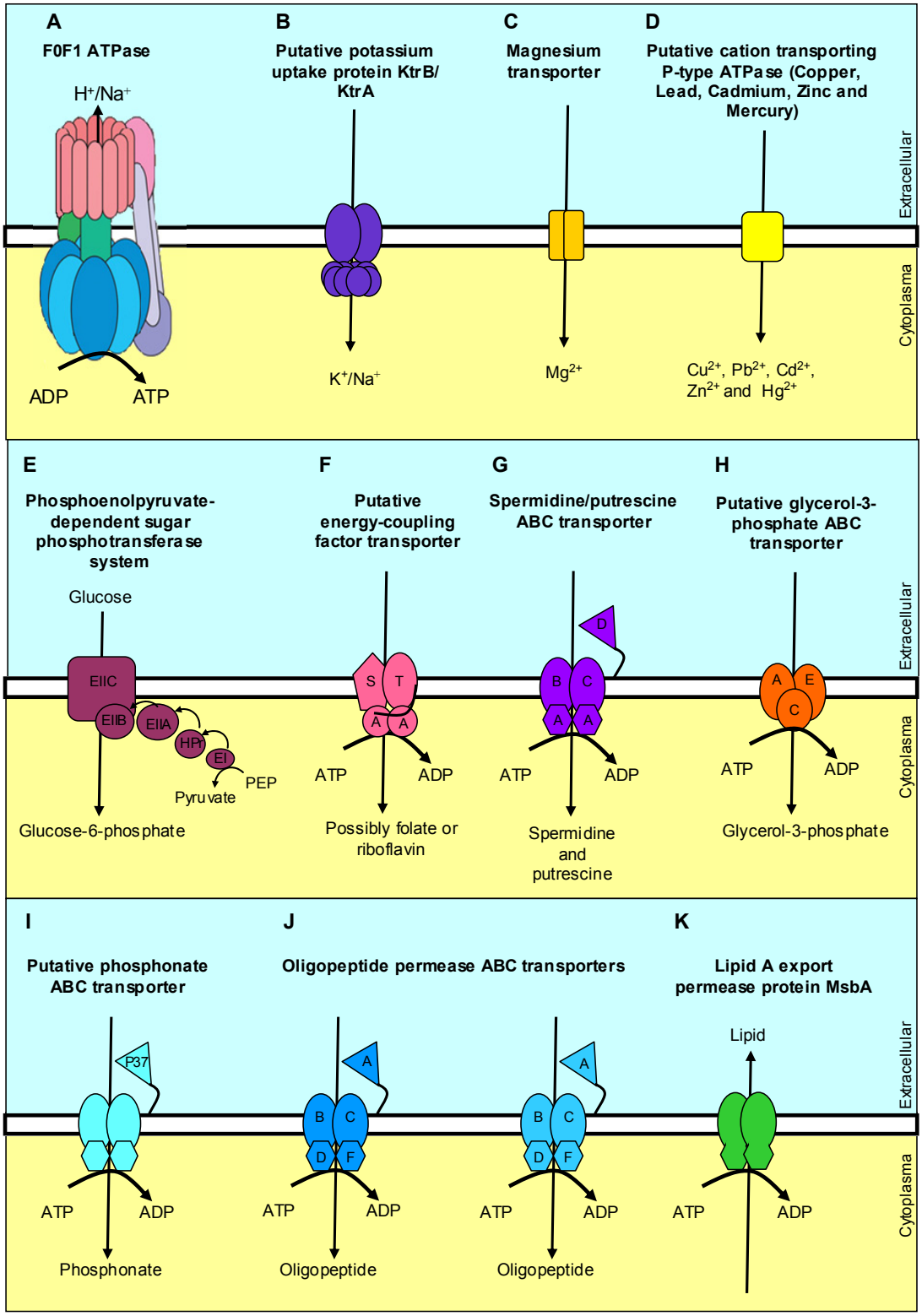


Figure 3.14 Transporters annotated in the Ms03 draft genome. A) The F_1F_0 ATP synthase, B) Putative potassium uptake protein (KtrB/KtrA), C) Magnesium transporter, D) Putative cation transporting P-type ATP synthase (copper, lead, cadmium, zinc and mercury), E) Phosphoenolpyruvate-dependent sugar phosphotransferase system, F) Putative energy-coupling factor transporter, G) Putative spermidine/putrescine ABC transporter, H) Putative glycerol-3-phosphate ABC transporter, I) Putative

phosphonate ABC transporter, J) Oligopeptide permease ABC transporters and K) Lipid A export permease protein MsbA. Abbreviation: PEP phosphoenolpyruvate.

predicted to have the substrate specificity for riboflavin with the IGS server but folate with the RAST server. In general, the S components share low sequence similarity (10 to 20% identity) (Slotboom 2014) which may explain the difference in the annotations.

Within the Ms03 draft genome, ABC importers for spermidine/putrescine, glycerol-3-phosphate, phosphonate and two Opp transporters were annotated (Figure 3.14 G-J, Appendix 2 Supplementary Table 2.14). A complete spermidine/putrescine (Pot) transporter consists of two transmembrane domains (PotB and PotC), two homodimeric intracellular nucleotide binding domains (PotA) as well as the substrate-binding domain (PotD) (Figure 3.14 G). In the IGS annotation, eight genes were denoted as components of a spermidine/putrescine (PotABCD) transporter while the RAST annotation had only three genes (Appendix 2 Supplementary Table 2.14). The ORFs for the additional genes annotated in IGS were not predicted in the RAST annotation. In the IGS annotation, a complete gene set, *potABCD* was annotated with two additional *potA* genes, one *potC* and *potD* gene.

In both the IGS and RAST annotations, the putative glycerol-3-phosphate ABC transporter consisted of UgpA, UgpC and UgpE (Figure 3.14 H). The IGS annotation had annotated an additional *ugpC* gene. The substrate-binding domain, UgpB was not annotated. In *M. gallisepticum* and *M. pneumoniae,* the UgpACE transporter has been implicated in virulence (Szczepanek et al. 2014).

P37 is an extracytoplasmic thiamine-binding lipoprotein that forms part of the ABC transport system and has been proposed to be the putative phosphonate substrate-binding protein (Sippel et al. 2009; Sippel et al. 2011). It was found to be part of the operon that included a transmembrane domain and an intracellular ATPase domain. In both the IGS and RAST annotations all three of the components were annotated (Figure 3.14 I, Appendix 2 Supplementary Table 2.14).

In the absences of *de novo* synthesis of amino acids (Section 3.3.5.2), import of oligopeptides will be required. The oligopeptide permease (Opp) transport system consists of a substrate-binding domain (OppA), two transmembrane domains (OppB and OppC) and two ATP-binding domains (OppD and OppF). Two complete Opp transporters were annotated with the IGS annotation server (Figure 3.14 J, Appendix 2 Supplementary Table 2.14). The RAST annotation, however, did not predict an ORF for the second *oppB* gene. Additionally, the RAST annotation had a gene annotated as “Oligopeptide ABC transporter ATP-binding protein” (peg.424) but the IGS annotated the same region as a hypothetical protein (mnas_50). The IGS annotation also had a “putative oligopeptide ABC transporter, ATP binding protein OppF C-terminal” (mnas_806) protein annotated; this region did not

have an ORF prediction in RAST. Both these proteins (mnas_806 and mnas_50) were truncated and were located on different contigs.

In the RAST annotation two proteins were annotated as “lipid A export ATP-binding/permease protein MsbA” Figure 3.14 K, Appendix 2 Supplementary Table 2.14) These were only annotated as “ABC transporter family protein” in the IGS annotation. In *E. coli,* MsbA is a homodimer ABC exporter that plays a role in translocation of lipopolysaccharide and had also been characterized as a multidrug resistance (MDR) efflux system. Putative ABC MDR genes were also identified in the genomes of *M. genitalium*, *M. pneumoniae* and *M. hominis* (Van Veen and Konings 1998; Paulsen et al. 2000; Raherison et al. 2002; Raherison et al. 2005).

Both IGS and RAST annotations include four membrane proteins predicted to have β-galactosidase activity (EC 3.2.1.23) (Appendix 2 Supplementary Table 2.14). In mycoplasmas, a very small minority of species have β-galactosidase activity and annotated genes are only found in *M. alligatoris* (Brown et al. 2011; May and Brown 2014) a close related species to Ms03. β-Galactosidase is a large membrane bound protein that hydrolysis terminal β-D-galactosides attached to glycoproteins and glycosaminoglycans. It also hydrolyses free lactose to galactose and glucose (May and Brown 2014; Singh et al. 2014). In *Streptococcus pneumoniae* β-galactosidase is postulated to play a role in adhesion (Singh et al. 2014).

As illustrated in Figure 3.14, a larger number of importers were annotated compared to exporters. Mycoplasmas, with such a reduced genome, have limited biosynthetic capacity and therefore rely on the import of nutrients, cofactors and other building blocks.

In addition to the above annotated transporters, both annotation analyses identified a number of substrate-binding, transmembrane and ATP-binding domains of putative ABC transporters (Appendix 2, Supplementary Table 2.14). These were components of transporters and/or the functions were unknown and as such are not discussed.

Furthermore, no transporter for nucleoside or nucleotide bases (Section 3.3.5.1) was found in both the IGS and RAST annotations. This may be an effect of the incomplete assembly of the Ms03 genome or the genes involved may be annotated as hypothetical in the current annotation. Transporters for these bases were, however, also not found in the genomes of *M. genitalium* and *M. pneumoniae* although they are essential components of mycoplasma growth medium (Himmelreich et al. 1996; Mushegian and Koonin 1996; Castellanos et al. 2004). Additionally, the loss of specificity and broadened substrate selectivity so that a single carrier can transport several metabolites across the cell membrane could explain the lack of some transport systems in mycoplasmas (Saurin and Dassa 1996).

3.3.6 Identification and bioinformatic characterization of the Ms03 *opp* operon

The Ms03 *opp* operon of which the *oppA* gene forms part was identified. Contigs that contain *opp* genes were identified by using the *opp* operon of *M. hominis* as query in a tBLASTn search with the initial assembly of Ms03 draft genome as database. Two contigs were identified to contain *opp* genes and the ORFs within these two contigs were predicted. The identities of the genes were confirmed with BLAST searches in NCBI. Two *opp* operons were subsequently identified (Figure 3.15) and submitted to NCBI with the following accession numbers, KM410300.1 and KM410301.1. Genes within the *opp1* operon were only named *oppA, oppB, oppC, oppD* and *oppF* while the genes in the *opp2* operon were numbered *oppA2, oppB2, oppC2 oppD2* and *oppF2.*

The first draft assembly (15.4 Mb, Table 3.6) of the Ms03 genome was used to identify the *opp* genes, however for the annotation, the pseudomolecule was generated from the contigs in the final assembly (25.2 Mb, Table 3.6). In the first draft assembly each set of *opp* genes (*oppA, oppB, oppC, oppD, oppF*) was located adjacent to each other within an operon. This was not the case in the automated annotation of the final assembly. The parameters in the assembly program, Newbler, that was used to assemble the 454 reads into contigs, are conserved. If the program is unsure whether or not two regions are adjacent to each other, it would rather split the regions than link them erroneously. In an attempt to decrease the number of contigs, more data was generated using the 454 platform. This did decrease the total number of contigs, however in the region of one of the two *opp* operons (*opp1* operon) more data may have introduced uncertainty which led to the genes of this *opp* operon to be distributed over more than one contig.

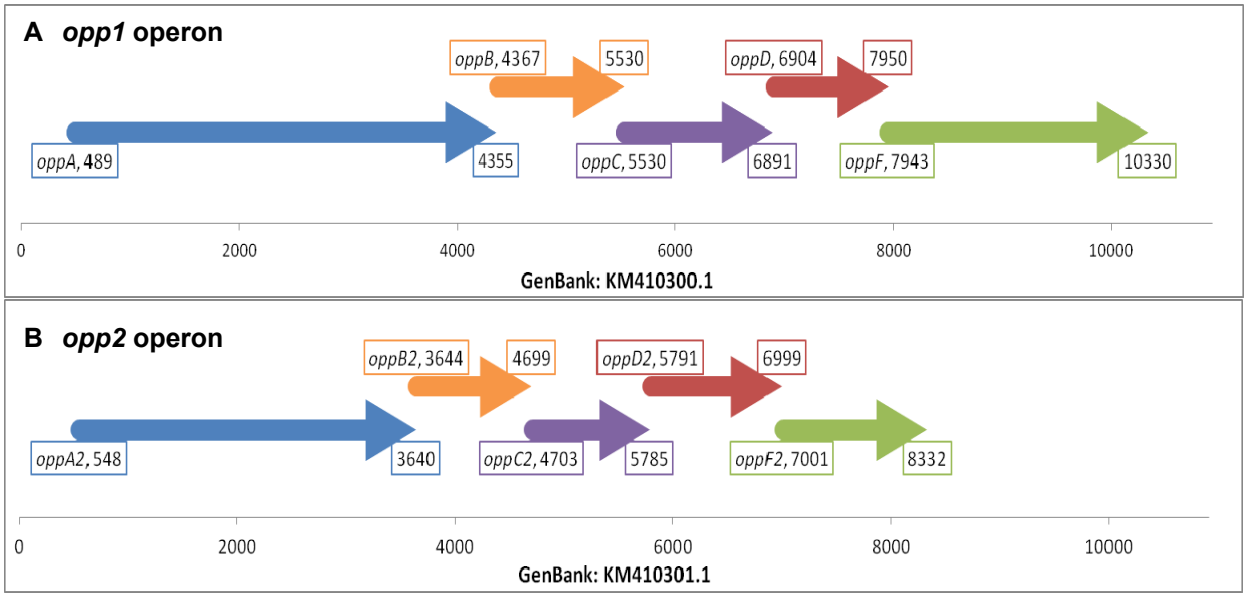


Figure 3.15 The *opp* operons in Ms03. A) The organization of the *opp1* operon (KM410300.1) B) The organization of the *opp2* operon (KM410301.1).

Table 3.14 A sequence comparison of the Ms03 *opp1* and *opp2* operon proteins with that of *M. hominis*

Gene (Gene size, Protein size)	**Ms03_OppA**	**Ms03_OppA2**
Mho_OppA (2 886 bp, 961 aa)	Protein identity: 27.1% Protein similarity: 45.7% Gaps: 14.5%	Protein identity: 19.0% Protein similarity: 32.6% Gaps: 34.6%
Ms03_OppA (3 867 bp, 1 288 aa)		Protein identity: 19.7% Protein similarity: 33.1% Gaps: 35.0%
Ms03_OppA2 (3 093 bp, 1 030 aa)		
	Ms03_OppB	**Ms03_OppB2**
Mho_OppB (1 146 bp, 381 aa)	Protein identity: 47.4% Protein similarity: 64.7% Gaps: 12.0%	Protein identity: 28.1% Protein similarity: 50.9% Gaps: 14.6%
Ms03_OppB (1 164 bp, 387 aa)		Protein identity: 26.3% Protein similarity: 42.2% Gaps: 18.2%
Ms03_OppB2 (1 056 bp, 351 aa)		
	Ms03_OppC	**Ms03_OppC2**
Mho_OppC (1 275 bp,424 aa)	Protein identity: 38.1% Protein similarity: 57.9% Gaps: 16.5%	Protein identity: 26.1% Protein similarity: 46.6% Gaps: 5.0%
Ms03_OppC (1 362 bp, 453 aa)		Protein identity: 25.3% Protein similarity: 44.8% Gaps: 12.5%
Ms03_OppC2 (1 083 bp, 360 aa)		
	Ms03_OppD	**Ms03_OppD2**
Mho_OppD (1 167 bp, 388 aa)	Protein identity: 71.2% Protein similarity: 84.3% Gaps: 0.0%	Protein identity: 36.4% Protein similarity: 60.2% Gaps: 7.8%
Ms03_OppD (1 047 bp, 348 aa)		Protein identity: 42.6% Protein similarity: 63.6% Gaps: 8.0%
Ms03_OppD2 (1 209 bp, 402 aa)		
	Ms03_OppF	**Ms03_OppF2**
Mho_OppF (2 529 bp, 842 aa)	Protein identity: 45.5% Protein similarity: 62.5% Gaps: 8.6%	Protein identity: 19.4% Protein similarity: 30.8% Gaps: 53.8%
Ms03_OppF (2 388 bp, 795 aa)		Protein identity: 19.8% Protein similarity: 31.5% Gaps: 52.1%
Ms03_OppF2 (1 332 bp, 443 aa)		

Similarity and Identity calculated with EMBOSS WATER Pairwise alignment (http://www.ebi.ac.uk/Tools/psa/emboss_water/) using the BLOSUM62 matrix with a gap penalty of 10 and extension penalty of 0.5

Compared to *M. hominis*, the gene organization of the two Ms03 *opp* operons was found to be the same (ABCDF) and the sizes of the genes and their protein products similar, except for the Ms03 *oppA* that was 981 bp larger than the *M. hominis oppA* and the Ms03 *oppF2* that was 1 217 bp smaller than the *M. hominis oppF* gene (Figure 3.15 and Table 3.14). The percentage protein identity and similarity between the *M. hominis* Opp proteins and that of the Ms03 Opp1 were higher than that between *M. hominis* and Ms03 Opp2 proteins or between Ms03 Opp1 and Ms03 Opp2 proteins (Table 3.14).

Vaccine candidate genes should be unique to the pathogen, with no homologous or paralogous genes within the genome of the host. The genome of the ostrich was sequenced and published recently and it was confirmed with tBLASTx searches that the two *opp* operons of Ms03 had no protein homologues or paralogues within the ostrich genome.

To confirm the identity of the predicted *opp* genes, functional motifs and domains were identified within the protein products using the online program, InterPro. InterPro combines protein signatures from 11 databases into one resource (McDowall and Hunter 2011). A summary of the InterPro analyses of the Opp proteins are provided in Table 3.15 while a detailed report of the individual motifs and their locations is provided in Appendix 2 Supplementary Table 2.15. Additionally the Opp proteins were manually examined for the same protein motifs as found in the Opp proteins of *M. hominis.*

Table 3.15 Bioinformatic analyses of the Opp proteins

	Ms03 *opp1* operon	**Ms03 *opp2* operon**
	OppA (1 288 aa)	**OppA2 (1 030 aa)**
Protein family membership	None predicted	ABC-type oligopeptide transport system, solute-binding component, Mycoplasmataceae, predicted (IPR016880)
InterPro domains	None predicted	Solute-binding protein family 5 domain (IPR000914)
GO term prediction	None predicted	None predicted
	OppB (387 aa)	**OppB2 (351 aa)**
Protein family membership	None predicted	None predicted
InterPro domains	ABC transporter type 1, transmembrane domain MetI-like (IPR000515)	ABC transporter type 1, transmembrane domain MetI-like (IPR000515)
GO term **Biological process**	GO:0006810 transport	GO:0006810 transport
Molecular function	GO:0005215 transporter activity	None predicted
Cellular component	GO:0016020 membrane	GO:0016020 membrane

Table 3.15 (Continued).

	Ms03 *opp1* operon	**Ms03 *opp2* operon**
	OppC (453 aa)	**OppC2 (360 aa)**
Protein family membership	None predicted	None predicted
InterPro domains	Oligopeptide transport permease C-like, N- terminal domain (IPR025966) ABC transporter type 1, transmembrane domain MetI-like (IPR000515)	Oligopeptide transport permease C-like, N- terminal domain (IPR025966) ABC transporter type 1, transmembrane domain MetI-like (IPR000515)
GO term prediction **Biological process**	GO:0006810 transport	GO:0006810 transport
Molecular function	None predicted	None predicted
Cellular component	GO:0016020 membrane	GO:0016020 membrane
	OppD (348 aa)	**OppD2 (402 aa)**
Protein family membership:	None predicted.	None predicted.
InterPro domains	P-loop containing nucleoside triphosphate hydrolase (IPR027417) AAA+ ATPase domain (IPR003593) ABC transporter-like (IPR003439) Oligopeptide/dipeptide ABC transporter, C- terminal (IPR013563) Oligopeptide/dipeptide ABC transporter, ATP-binding protein, C-terminal (IPR010066)	P-loop containing nucleoside triphosphate hydrolase (IPR027417) AAA+ ATPase domain (IPR003593) ABC transporter-like (IPR003439) Oligopeptide/dipeptide ABC transporter, C- terminal (IPR013563) Oligopeptide/dipeptide ABC transporter, ATP-binding protein, C-terminal (IPR010066)
GO term prediction **Biological process**	GO:0015833 peptide transport	GO:0015833 peptide transport
Molecular function	GO:0000166 nucleotide binding GO:0005524 ATP binding GO:0015197 peptide transporter activity GO:0016887 ATPase activity	GO:0000166 nucleotide binding GO:0005524 ATP binding GO:0015197 peptide transporter activity GO:0016887 ATPase activity
Cellular component	GO:0016020 membrane	GO:0016020 membrane
	OppF (795 aa)	**OppF2 (443 aa)**
Protein family membership	None predicted	None predicted
InterPro domains	P-loop containing nucleoside triphosphate hydrolase (IPR027417) AAA+ ATPase domain (IPR003593) ABC transporter-like (IPR003439) Oligopeptide/dipeptide ABC transporter, C- terminal (IPR013563)	P-loop containing nucleoside triphosphate hydrolase (IPR027417) AAA+ ATPase domain (IPR003593) ABC transporter-like (IPR003439) Oligopeptide/dipeptide ABC transporter, C- terminal (IPR013563)
GO term prediction **Biological process**	GO:0015833 peptide transport	GO:0015833 peptide transport
Molecular function	GO:0000166 nucleotide binding GO:0005524 ATP binding GO:0016887 ATPase activity GO:0017111 nucleoside-triphosphatase activity	GO:0000166 nucleotide binding GO:0005524 ATP binding GO:0016887 ATPase activity
Cellular component	None predicted	None predicted

OppA

No protein family, domains or GO terms were predicted with InterPro for the OppA protein of the *opp1* operon. InterPro predicted the OppA2 protein of the *opp2* operon to belong to the protein family: ABC-type oligopeptide transport system, solute-binding component of *Mycoplasmataceae* (IPR016880), pertaining to the solute-binding protein family 5 domain (IPR000914) (Table 3.15). GO terms were however not predicted for the OppA2 protein. For both OppA proteins the PROSITE prokaryotic lipoprotein (PS51257) and the PHOBIUS signal peptide motifs were found within the first 23 (OppA) or 27 (OppA2) amino acids (Appendix 2 Supplementary Table 2.15).

In the OppA protein of *M. hominis* four functional sites were identified: a signal peptide, an oligopeptide binding site and Walker B and A motifs (Henrich et al. 1999). Through manual alignment of the Ms03 OppA proteins with that of *M. hominis*, a putative signal peptide could be identified in both OppA (first 23 aa) and OppA2 (first 27 aa). The liposignal peptide motif with a conserved cysteine residue at position 23 of OppA and 27 of OppA2 should result in the proteins being anchored to the cell via a lipo anchor. These signal peptides were also predicted with InterPro (Appendix 2 Supplementary Table 2.15). Manual alignment also allowed the identification of a putative oligopeptide binding site (552-FRVRPGHFW-560) as well as putative Walker B (1 074-YWTGTSPFSLAGWGYD-1 089) and putative Walker A (1 094-GSGIDGYS-1 101) motifs in the Ms03 OppA (*opp1* operon). These motifs could not be found in the OppA2 protein.

In bacterial OppA proteins, the oligopeptide-binding site determines substrate specificity (Doeven et al. 2005) and is in general less conserved which may explain the wide range of oligopeptides that can be transported. The Walker B and A motifs are responsible for the ecto-ATPase activity of *M. hominis* OppA. The consensus sequence for the Walker A motif is -GXXXXG**K**[S/T]- where Lys (K) is essential for nucleotide-binding. A mutation of the Lys to an Arg (both large and positively charged) resulted in a loss of ATPase activity (Hopfe and Henrich 2004). In Ms03 OppA the amino acid at this position is Tyr (large and possessing polar hydroxyl group) therefore although the motif is present in the Ms03 OppA sequence, it is likely not functional.

Based on the identified functional motifs and domains, the identity of the OppA (*opp1* operon) and the OppA2 (*opp2* operon) proteins could be confirmed as the proposed substrate-binding domains of the Opp transporters in Ms03.

OppB and OppC

InterPro could not predict the protein family for either Ms03 OppB or OppB2 however the InterPro domain “ABC transporter type 1, transmembrane domain MetI-like” (IPR000515) was found in both proteins (Table 3.15). The GO terms predicted for these two proteins were transport (GO:0006810) for biological process and membrane (GO:0016020) for cellular

component. Additionally the GO term transporter activity (GO:0005215) for molecular function was only predicted for the Ms03 OppB protein.

In Ms03 OppC and OppC2, InterPro detected two domains: oligopeptide transport permease C-like, N- terminal domain (IPR025966) and ABC transporter type 1, transmembrane domain MetI-like (IPR000515) (Table 3.15) The GO term for biological process was transport (GO:0006810) and for cellular component was membrane (GO:0016020).

In the Opp transporter, OppB and OppC domains form the pore through which the substrate is imported into the cell. Six transmembrane-spanning segments were predicted with InterPro in OppB, OppB2, OppC and OppC2 of Ms03 (Appendix 2 Supplementary Table 2.15) which should result in a membrane pore with 12 membrane spanning segments (six from the OppB domain and six from OppC domain) which is typical for ABC transporters (Rees et al. 2009).

In addition manual alignment revealed the OppB conserved motif (RTAK-KGLXXXXI/VZXXHZLR, with Z representing a hydrophobic residue and X any residue) in OppB (256-IAKSKGLSRKEIFFKYVLR-274) and OppB2 (250-FAYLKGVSKNRFVWKHALK-268) and the OppC conserved motif (XAAXXZGAXXXRXIFXHILP) in the OppC (346-TASKSVGASKARLIYKHALP-365) and OppC2 (257-VSASKILGTPTWKILKNYVP-276). These motifs were found between the fourth and fifth membrane-spanning segments as predicted in the OppB and OppC proteins and correspond to the proposed interaction sites with OppD and OppF domains on the cytoplasmic side of the membrane (Mourez et al. 1997; Henrich et al. 1999). These motifs and domains confirmed the identity of OppB and OppC as well as the OppB2 and OppC2 proteins as the proposed transmembrane pores of the Opp transporters in Ms03.

OppD and OppF

In each of the OppD, OppD2, OppF and OppF2 proteins, the following InterPro domains were identified: P-loop containing nucleoside triphosphate hydrolase (IPR027417); AAA+ ATPase domain (IPR003593); ABC transporter-like (IPR003439) and oligopeptide/dipeptide ABC transporter, C-terminal (IPR013563). Additionally the oligopeptide/dipeptide ABC transporter and ATP-binding protein, C-terminal (IPR010066) domain was also found in the OppD and OppD2 (Table 3.15). For these four proteins the GO term for biological process was peptide transport (GO:0015833) and for molecular function was nucleotide binding (GO:0000166), ATP binding (GO:0005524), and ATPase activity (GO:0016887). Additionally for molecular function, the GO term, nucleoside-triphosphatase activity (GO:0017111) was only found in OppF and the GO term, peptide transporter activity (GO:0015197) was only predicted for OppD and OppD2. The GO term membrane (GO:0016020) for cellular component was found in OppD and OppD2 but not in OppF and OppF2.

In the Opp transporter, the OppD and OppF domains hydrolyse ATP to provide the energy needed for the translocation of the oligopeptide across the membrane. This activity is attributed to the presence of protein motifs such as the Walker A and Walker B motifs. The Walker A motif (GXXGXGK[T/S]) was present in OppD (43-GESGSGKS-50), OppD2 (95-GESGSGKS-102), OppF (42-GESGSGKT-49) and OppF2 (79-GESGSGKS-86). The Walker B motif (RXXXGXXXLZZZZD) was also found in OppD (180-LIIADEPTTALD-191), OppD2 (234-ILVMDEPTTALD-243), OppF (303-VIVADEPIASLD-336) and OppF2 (305-LLIADEPISALD-316). Furthermore, the ABC signature motif (LSGGQ) was also present in OppD (160-LSGGM-164), OppD2 (212-MSGGM-216), OppF (632-FSGGQ-636) and OppF2 (285-FSGGQ-289). The presence of these motifs was reflected in the InterPro motifs identified, that correspond to ATPase activity and family domains for ABC transporters. These motifs and domains confirmed the identity of the OppD and OppF as well as the OppD2 and OppF2 proteins as the proposed cytosolic ATP-binding domains of the Opp transporters in Ms03.

3.4 Conclusion

The genome of Ms03 was sequenced and assembled into 172 large contigs. It was however not possible to generate the complete genome sequence of Ms03. Two annotation servers, IGS and RAST, were used to annotate the Ms03 draft genome and predicted 763 and 635 protein-encoding genes, respectively. In spite of the differences, the two annotations augment to each other to generate a more complete annotation of the Ms03 draft genome. The IGS annotation yielded more annotated genes which resulted in a more complete representation of the Ms03 metabolic pathways. The RAST annotation provided the option for metabolic analysis which benefits the interpretation of the annotation result. Each annotation pipeline (IGS and RAST) has strengths and weaknesses. The combinational approach described here not only provided the opportunity to compare the pipelines, but also increased the confidence in the gene annotation of Ms03 and resulted in a more complete overview of the Ms03 metabolic capacity.

Although some of the genes within the Ms03 pathways analysed are potentially “missing”, most annotated genes within the Ms03 draft genome correspond with the gene functions annotated in the complete genome of *M. synoviae* 53. This indicates that the annotation of the Ms03 genome is mostly complete. The generation of a complete genome sequence may however be advantageous when considering the number of genes that are potentially truncated because the Ms03 draft genome consists of 172 contigs.

Genome replication, cell division, RNA transcription and protein translation in Ms03 is similar to that of *M. synoviae* 53. From the annotations it is likely that Ms03 is a glycolytic mycoplasma that cannot utilize arginine and urea which is in agreement with the experimental results of Langer (2009). Like other mycoplasmas, Ms03 cannot synthesise

purines and pyrimidines or amino acids using *de novo* pathways. Most of the annotated transporters were importers which is a reflection of the reduced metabolic capacity of the mycoplasma and the need to obtain nutrients from the environment.

The current annotation of the Ms03 genes together with the metabolic analysis can assist in the identification of future vaccine gene targets. Two Opp transporters were found in the Ms03 genome. The genes for each transporter are located in an operon that consists of five adjacent genes with the order ABCDF. Protein motifs within the respective genes confirmed the functional annotations of the respective genes as *oppA*, *oppB, oppC*, *oppD* and *oppF*. These proteins were characterized in an attempt to confirm the identity of the *opp* operon and also the chosen vaccine candidate gene, *oppA*. Two *opp* operons (two *oppA* genes) were however identified within the Ms03 genome. Additionally, *oppA* genes in mycoplasma genomes are not always annotated, which questions the need for the substrate-binding domain in Opp transporter (Staats et al. 2007). The above warranted a more detailed analysis of the *opp* operons with special reference to the *oppA* gene within *Mycoplasma* species.

Acknowledgements:

We would like to thank the Institute for Genome Sciences Annotation Engine service at the University of Maryland, School of Medicine for providing structural and functional annotation of the sequences.

Chapter 4 The identification of *oppA* gene homologues as part of the oligopeptide transport system in mycoplasmas

4.1 Introduction

The *oppA* gene was chosen as a vaccine candidate gene in this dissertation, however it would only be a suitable target if it plays a prominent part in oligopeptide transport in mycoplasma. The oligopeptide permease (Opp) transport system is responsible for the import of oligopeptides into the bacterial cell. This transporter belongs to the ATP-binding cassette (ABC) family and consists of five components; a substrate-binding domain, OppA, two transmembrane domains, OppB and OppC that form a pore for transport and two cytoplasmic domains, OppD and OppF that hydrolyse ATP to provide the energy necessary for the translocation of the oligopeptide. In Gram-positive bacteria, OppA is anchored to the membrane through a lipid anchor, while in Gram-negative bacteria it is located in the periplasmic space. The oligopeptide-binding site (and therefore OppA) determines the specificity of the transporter. This site is in general less conserved and has a wide size range of peptides that are transported in different bacteria. The OppA of *Escherichia coli* has a preference for oligopeptides that contain positively charged amino acids and are 3 to 4 residues in length (Klepsch et al. 2011) while the OppA of *Lactoccocus lactis* can transport oligopeptides between 4 and 35 residues in size and has a preference for proline-rich peptides containing at least one isoleucine (Berntsson et al. 2009). This illustrates the diversity in substrates that can be transported by bacterial OppA proteins. As a result low sequence similarity is, in general, reported between *oppA* genes and their protein products (Doeven et al. 2005).

In *Mycoplasma* species, the need for an OppA in oligopeptide transport was, however, questioned because of the absence of an annotated *oppA* gene as part of some *oppBCDF* operons in mycoplasma genomes (Staats et al. 2007). The progress in genome sequencing has led significantly increase in the number of available and annotated mycoplasma genomes. Within these genomes, *oppA* genes are not always found as part of the *oppBCDF* operons.

In order to assess whether the *oppA* gene of Ms03 has the potential to serve as a vaccine candidate gene the oligopeptide transporter systems in *Mycoplasma* species was assessed. The aim was therefore to investigate the prevalence of the *oppA* gene as part of the *oppBCDF* operon in *Mycoplasma* species. To this end, the following objectives were set:

- To identify *opp* genes in the annotated genomes of *Mycoplasma* species available on NCBI.
- To identify an *oppA* gene for each *oppBCDF* operon.

- To determine the phylogenetic relationships amongst the *Mycoplasma* species as well as their respective OppABCDF operons.
- To identify conserved and functional motifs in the annotated and newly postulated OppA proteins, in order to confirm that the postulated *oppA* genes annotated as "hypothetical protein" are indeed *oppA* genes.

The results obtained are presented in an independent manuscript following this introduction with the supplementary data presented in Appendix 3 at the end of this dissertation. The following sections will highlight certain aspects of the experimental design.

4.2 Phylogenetic analysis

Phylogenies are used to address various questions such as the relationships among genes or species, epidemiological dynamics of pathogens and the evolutionary histories of populations (Hartfield et al. 2014). The relationships of species can be determined by using a marker gene or protein. The selection of this gene is critical; it should be present in all the species in the assessment, have the same function within the species and be under the same selective pressure.

The *16S rRNA* gene is highly conserved and universally distributed in the different species of bacteria and archaea (Brocchieri 2001). Ribosomes are present in all self-replicating cells and function in protein synthesis (Schuwirth et al. 2005; Kitahara et al. 2012). The *16S rRNA* genes have the same structural role in the ribosomes irrespective of species and are under approximately the same selective evolutionary pressure. Although other phylogenetic markers such as the RNA polymerase subunit B (*rpoB*) gene, 16S–23S intergenic transcribed spacers and phosphoglycerate kinase amino acid sequence have been evaluated, the *16S rRNA* gene sequences are currently recommended for description of *Mollicutes* species (Wolf et al. 2004; Brown 2010; Volokhov et al. 2012). For the above reasons, the *16S rRNA* gene is an ideal marker to analyse the relationships among *Mycoplasma* species and was used in the current study.

A phylogeny was also constructed from the OppABCDF proteins to analyse the functional relationship amongst different copies of the *oppABCDF* operon within and between species. These genes are not necessarily under the same selective pressure but are subjected to the selection constraints of protein function and protein structure. The amino acid sequences rather than the nucleotide sequences were aligned because at nucleotide level the sequences were heterogeneous and could not by aligned with confidence.

The *16S rRNA* gene and the OppABCDF protein sequences were first aligned with the online alignment program T-Coffee (tree-based consistency objective function for alignment evaluation, Notredame et al. (2000)) followed by further manual refinement. T-Coffee is a multiple sequence alignment program that uses a progressive approach to combine local

and global pair-wise alignments creating a reliable alignment of the sequences (Notredame et al. 2000). This allows for the alignment of genes or proteins that differ in length as was the case with the OppA and OppF proteins in the current study.

The construction of phylogenies is either distance or tree-searching based. In distance based methods, such as neighbour-joining, distances are calculated between pairs of sequences (within the larger alignment) resulting in a distance matrix. The distance matrix is then used to construct the phylogenetic tree. Although distance analyses are computer efficient, they perform poorly for very divergent sequences and are sensitive to gaps in the alignment (Yang and Rannala 2012). Tree-searching methods consider all characters of the entire alignment one site (column) at a time and calculate a score for each possible tree. The tree with the best score is retained. Examples of tree-searching methods are parsimony, maximum likelihood and Bayesian inference. In parsimony “tree-scores” are minimum number of changes, in maximum likelihood, log-likelihood values and in Bayesian inference, posterior probability (Wiley and Lieberman 2011b; Wiley and Lieberman 2011a). Each of these analyses uses their own formulas, algorithms and models that contribute to the outcome.

The major weakness of a parsimony analysis such as the software package PAUP* (Phylogenetic Analysis Using Parsimony* and other methods, Swofford (2002)) is its lack of explicit assumptions which makes it nearly impossible to incorporate any knowledge of sequence evolution into the tree reconstruction (Yang and Rannala 2012). Maximum likelihood is an exhaustive analysis that searches for the best tree. Maximum likelihood is a powerful framework for estimating parameters and testing hypotheses and can accommodate variable amino acid substitution rates among sites and other evolutionary models. Exhaustive searches require a large amount of computational power, however online platforms such as the Cipres Web portal (Miller et al. 2010) are available for these analyses. RAxML (Randomized Axelerated Maximum Likelihood) is a program for maximum likelihood phylogenetic analysis of large datasets (Stamatakis 2014) that is available from the Cipres Web portal.

OppA proteins in mycoplasma have low sequence identity and gaps were introduced to allow alignment. Since distance based methods perform poor with very divergent sequences and gaps, the phylogenies were generated using both RAxML (maximum likelihood analysis) and PAUP (parsimony analysis). The generated phylogenies had the same topology and only the RAxML phylogenies are shown.

4.3 Bioinformatic analysis of annotated and newly postulated *oppA* homologues genes

In order to confirm that newly postulated *oppA* genes annotated as “hypothetical protein” were indeed *oppA* genes, a number of bioinformatic programs were used to evaluate the annotated and newly postulated OppA protein sequences. OppA is located on the surface of the mycoplasma cell and the location of the postulated and annotated OppA proteins were predicted with PSORTb version 3.0.2 (Yu et al. 2010). This program is specific for bacterial proteins and has five subcellular categories, namely cytoplasmic, cytoplasmic membrane, cell wall, extracellular or unknown. Additionally the presences of signal peptides were predicted using SignalP (Petersen et al. 2011) and PRED-LIPO (Bagos et al. 2008). SignalP predicts whether or not a protein has a signal peptide while PRED-LIPO specifically predicts signal peptidase I (secretory protein) and signal peptidase II (lipoproteins). A signal peptidase II was found in the OppA of *M. hominis* (the best described OppA in the *Mycoplasma* genus).

InterPro (Hunter et al. 2012) was used to predict the functional domains of the postulated and annotated OppA proteins. The InterPro consortium includes the following databases: PROSITE, HAMAP (high-quality automated and manual annotation of proteins), Pfam (protein families), PRINTS, ProDom (protein domain families), SMART (simple modular architecture research tool), TIGRFAMs, PIRSF (protein information resource superfamily), SUPERFAMILY, CATH-Gene3D (class architecture topology homology) and PANTHER (protein analysis through evolutionary relationships). Using InterPro the information of these databases can be searched simultaneously to predict the function, domains and important sites within postulated and annotated OppA proteins.

It was however not always possible to predicted function, domains or important sites within the postulated and annotated OppA proteins. Therefore the motif discovery program, MEME (Multiple expectation maximization for motif elicitation) (Bailey et al. 2009) was used to locate conserved motifs within the protein sequences. MEME locates repeated patterns within a group of related proteins and does not require functional or database information to identify these motifs.

4.4 The identification of *oppA* gene homologues as part of the oligopeptide transport system in mycoplasmas

4.4.1 Contributions of co-authors

The following manuscript contains the original work of the author of this dissertation. The contributions of each of the co-authors were as follows:

Dr. Annelise Botes is a lecturer of Biochemistry at the University of Stellenbosch and the supervisor of this dissertation. In this capacity, she was involved in the conceptual development and execution of all aspects of this study.

Prof. Dirk U. Bellstedt is a professor of Biochemistry at the University of Stellenbosch and the co-supervisor of this dissertation. In this capacity, he was involved in the conceptual development and execution of this study.

Gene 558 (2015) 31–40

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

The identification of *oppA* gene homologues as part of the oligopeptide transport system in mycoplasmas

Martha Wium, Annelise Botes *, Dirk U. Bellstedt

Department of Biochemistry, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa

ARTICLE INFO

Article history:
Received 31 October 2014
Received in revised form 3 December 2014
Accepted 16 December 2014
Available online 17 December 2014

Keywords:
Oligopeptide permease transport
Mycoplasma
OppA
opp operon
ABC transporters

ABSTRACT

The lack of an annotated *oppA* gene as part of many oligopeptide permease (*opp*) operons has questioned the necessity of the oligopeptide-binding domain (OppA) as a part of the Opp transport system in mycoplasmas. This study investigated the occurrence of an *oppA* gene as part of the *oppBCDF* operon in 42 mycoplasma genomes. Except for hemoplasma, all mycoplasmas were found to possess one or more copies of the *oppBCDF* operon and with the help of similarity searches their *oppA* genes could be identified. Phylogenetic analysis of the combined OppABCDF amino acid sequences allowed them to be grouped into three types. Each type has a unique set of conserved motifs, which are likely to reflect substrate preference and adaption strategies. Our approach allowed the identification of *oppA* gene homologues for all mycoplasma *opp* operons and thereby provides a method for re-evaluating the current annotation of *oppA* genes in mycoplasma genomes.

1. Introduction

Mycoplasmas are the smallest self-replicating prokaryotes that have evolved through a degenerative process from low G + C content Gram-positive bacteria (Razin et al., 1998; Brown, 2010). The evolution of mycoplasmas was accompanied by a substantial loss of their genomes resulting in the loss of the cell wall as well as many of their biosynthetic pathways (Yus et al., 2009). Mycoplasmas lack the capacity to synthesize molecules such as cholesterol, fatty acids, some amino acids, purines and pyrimidines, and therefore need to acquire these and other nutrients from their host (Razin et al., 1998). Acquisition of these molecules is mediated through membrane transport proteins of which the ATP-binding cassette (ABC) transporters are the most prominent (Nicolás et al., 2007). The role of ABC transporters in the growth and survival of bacteria makes them an attractive target for the development of vaccines and antibacterial therapies (Garmory and Titball, 2004; Nicolás et al., 2007). To this end, the OppA proteins of *Brachyspira pilosicoli* (Movahedi and Hampson, 2010), *Moraxella catarrhalis* (Yang et al., 2011) and *Yersinia pestis* (Tanabe et al., 2006) have been evaluated as candidate vaccine antigens and could similarly be a potential vaccine antigen in mycoplasmas.

OppA is a lipoprotein that forms part of a bacterial oligopeptide permease (Opp) transport system that is responsible for the import of oligopeptides into the cell. The Opp transport system has five structural domains; OppA, the extracellular substrate-binding domain, OppB and OppC, the two transmembrane domains and OppD and OppF, the two intracellular domains that bind and hydrolyze ATP thereby supplying energy for peptide transport. A separate gene encodes for each domain and the five genes are located within a polycistronic operon in most instances. The Opp system is well characterized in bacteria and the function of OppA as substrate-binding protein is established (Levdikov et al., 2005; Berntsson et al., 2009; Klepsch et al., 2011). In Gram-negative bacteria the OppA is found within the periplasmic space, while in Gram-positive bacteria OppA is anchored to the membrane as is the case in mycoplasmas. The necessity of this protein in the Opp transport system in mycoplasmas was, however, questioned due to the fact that only one out of eleven genomes available at the time contained an annotated *oppA* gene as part of the *opp* operon (Staats et al., 2007).

Within the 39 mycoplasma genomes chosen to be included in this study, there were 55 annotated *oppBCDF* operons, but only 16 of these had an annotated *oppA* gene as part of the *opp* operon. In addition to this, some genomes had two copies of the *oppBCDF* operon, but only one contained an annotated *oppA* gene. The need for an *opp* operon, and specifically the *oppA* gene, for the survival of mycoplasmas was studied using transposon mutagenesis. In both *Mycoplasma genitalium* and *Mycoplasma pulmonis* the *opp* operons and specifically the *oppA*

Abbreviations: ABC, ATP-binding cassette; BLAST, basic local alignment search tool; CATH, class architecture topology homologous superfamily; Dpp, dipeptide permease; GO, gene ontology; InterPro, protein sequence analysis and classification; MEME, multiple em for motif elicitation; NCBI, national center for biotechnology information; Opp, oligopeptide permease; ORF, open reading frame; Pfam, protein families; PSORTb, subcellular localization prediction tool; PRED-LIPO, prediction of lipoprotein signal peptides; PSI-BLAST, position specific iterated blast; rRNA, ribosomal RNA; SCOP, structural classification of proteins; sp. nov, novel species; str, strain; tBLASTx, translated nucleotide databases using a translated nucleotide query.

* Corresponding author.
E-mail address: annelise@sun.ac.za (A. Botes).

http://dx.doi.org/10.1016/j.gene.2014.12.036
0378-1119/

gene were found to be essential for survival (Glass et al., 2006; French et al., 2008). Detailed studies done on the *opp* operon within *Mycoplasma hominis* found that the OppA protein not only acts as a substrate-binding protein but also possesses a cytoadherence function and has ecto-ATPase activity (Henrich et al., 1993; Hopfe and Henrich, 2004). It is therefore possible that OppA proteins are essential for the survival of all mycoplasmas and that the apparent lack of an *oppA* gene within an operon may not be because of its absence, but rather incorrect/incomplete annotation. Annotation of genome data generated through next generation sequencing projects is usually achieved through automatic servers that rely on sequence similarity. Compared to the OppBCDF protein, there is a low sequence similarity between the OppA proteins of different mycoplasmas. Genes or proteins that show low sequence similarity are often "missed" by these automated analyses and as such lack either an appropriate annotation or functional assignment.

The purpose of this study was to determine the number of *opp* operons in mycoplasma genomes and the extent to which *oppA* genes form part of these operons. Due to our interest in OppA as a possible vaccine antigen, genomes of ostrich-infecting mycoplasmas were included in the current study. Three unique *Mycoplasma* species were found to infect ostriches and were provisionally named Ms01, Ms02, and Ms03 (Botes et al., 2005). Ms01 has since been described as *Mycoplasma struthionis* sp. nov. and Ms03 as *Mycoplasma nasistruthionis* sp. nov., but have not been formally published (Langer, 2009). The phylogenetic relationships of all the mycoplasmas included in this study were determined using *16S rRNA* sequences. In this study an approach was developed to identify *oppA* genes in mycoplasma genomes. Annotation of *oppA* genes was re-evaluated using a combination of bioinformatic analyses including sequence similarity, prediction of cellular location and identification of protein sequence motifs unique to OppA proteins. The relationships of the *opp* operons were determined using phylogenetic analysis of an amino acid alignment of the OppABCDF proteins.

2. Materials and methods

2.1. Multiple sequence alignment and phylogenetic analysis of 16S rRNA gene sequences

The *16S rRNA* gene sequences (available from NCBI) of the 42 *Mycoplasma* species listed in Supplementary Table 1 used in this study were aligned with the objective of identifying the evolutionary relationships between them. The *16S rRNA* sequences of related *Mollicutes* were included namely: *Anaeroplasma bactoclasticum* (NR_044675.1), *Acholeplasma laidlawii* (M23932.1), *Spiroplasma citri* (M23942.1), *Spiroplasma taiwanense* CT-1 (HM037992.1), *Ureaplasma urealyticum* serovar 13 (AF073455.1) and *Ureaplasma gallorale* str. D6-1 (NR_026027.1). The *16S rRNA* sequence of *Clostridium botulinum* A (NR_074124.1) was used as outgroup.

Initial alignments were done with T-Coffee (Notredame et al., 2000; Di Tommaso et al., 2011) and further refined through manual editing in BioEdit v7.0.5.2 (Hall, 1999). Maximum likelihood trees were constructed using RAxML-HPC2 on XSEDE (8.024) (Stamatakis, 2014) using the Cipres Web portal (Miller et al., 2010). Clade support was evaluated using Bootstrap analysis with an automatic stop when the majority rule criterion was reached, as recommended by the program. Bootstrap values $\geq 75\%$ were considered well supported while values below 75% considered moderately or weakly supported. As a result only values above 75% are indicated on the phylogenetic tree.

2.2. Identification of opp operons in published mycoplasma genomes

Thirty-nine mycoplasma genomes available on NCBI were selected to include representative strains of all species as well as members of the evolutionary groups (pneumoniae, hominis and spiroplasma) listed in Supplementary Table 1. In some genomes more than one copy of the *opp* operon was annotated in which case all were used in the analysis. Where an annotated *opp* operon was not found, a genome search was performed using the gene function as criterium namely "oligopeptide" and finally tBLASTx searches were performed to confirm the absence of the operon. In such instances the annotated *oppBCDF* genes of the nearest relative, based on the 16S rRNA phylogeny, were used as query.

2.3. Identification of opp operons in mycoplasmas infecting ostriches

The three genomes of the ostrich-infecting mycoplasmas isolated from ostriches in South Africa were sequenced using the Roche 454 system (Pretorius, 2009; Steenmans, 2010; Wium, unpublished results) and reads assembled to contiguous sequence level. Although the full genome sequence of each still needs to be completed, the contiguous sequences were sufficiently complete to allow the identification of their relevant *opp* operons.

The *opp* operons in the three ostrich-infecting mycoplasma genomes, *M. struthionis* sp. nov. str. Ms01, *Mycoplasma* sp. Ms02 and *M. nasistruthionis* sp. nov. str. Ms03, were identified using a local tBLASTx search in CLCbio with the genome assembly data as database and the *opp* operon of *M. hominis* (X99740.1) as query sequence. Open reading frames (ORFs) were identified using Glimmer v3.02 (Delcher et al., 2007) and gene sizes were confirmed by comparison with other mycoplasma *opp* operons. The *oppA* gene sequences were confirmed by PCR amplification from genomic DNA with appropriate primers and proofreading Taq polymerase followed by Sanger sequencing to ensure error free sequences.

2.4. Identification of oppA gene homologues in mycoplasma genomes

For the *opp* operons in which the *oppA* gene was not annotated, the four neighboring genes on either side of the *oppBCDF* operon were evaluated in an attempt to identity their *oppA* genes. This was done by comparing the gene size of the neighboring genes with other annotated *oppA* genes, followed by a tBLASTx search to determine sequence similarity with other known *oppA* genes of mycoplasmas. Additionally a PSI-BLAST was conducted if the above approach was not successful. A typical gene size of 2499-3868 nucleotides was chosen and sequence similarity was chosen as 30% amino acid identity over 70% of the sequence length. If an *oppA* gene could not be identified as part of an *opp* operon, the whole genome was searched using the NCBI tBLASTx and PSI-BLAST algorithms. In such instances the annotated *oppA* gene of the nearest relative, based on the 16S rRNA phylogeny, was used as query. Additionally, the position of the *oppA* gene relative to the *oppBCDF* genes was determined.

2.5. Amino acid sequence alignment and phylogenetic analysis of OppABCDF

All previously annotated OppA amino acid sequences as well as those identified using the approach outlined above were subsequently combined with their OppBCDF amino acid sequences to investigate the relationships between these transporters. The reason why this was done was because these are the complete functional protein components of each of the OppABCDF oligopeptide transporters. The analysis included the OppABCDF amino acid sequences from the 31 *Mycoplasma* species available on NCBI, the three mycoplasmas isolated from ostriches and *C. botulinum* A (CAL84708.1-CAL84712.1). If a genome contained more than one copy of the *oppABCDF* operon, all were included. One of the operons of *Mycoplasma penetrans* HF-2 was identified to have five OppA's and therefore the OppBCDF sequence was represented five times in the alignment, each with a different OppA. A total of 65 sequences were therefore included in the final alignment.

Amino acid sequence alignment was done using T-Coffee (Notredame et al., 2000) with further manual refinement in BioEdit v7.0.5.2 (Hall, 1999). Compared to the OppBCDF protein sequences, the OppA proteins had very low sequence similarity and therefore only the areas containing conserved sequence regions for OppA were

retained. The total alignment consisted of 7038 characters (Supplementary File 1) with the alignment of OppA amino acid sequences spanning from 1 to 2851, OppB from 2855 to 3461, OppC from 3465 to 4018, OppD from 4022 to 5343 and OppF from 5347 to 6776. In an attempt to align sequences, gaps were inserted in the alignments of the different gene segments. The lengths of the different alignments therefore do not represent the actual length of the different proteins, but also account of the inserted gaps. The following regions within the total alignment were used for the construction of the phylogeny: OppA regions: 10–71, 91–113, 154–168, 505–524, 550–577, 610–619, 689–698, 749–768, 1039–1073, 1081–1134, 1387–1445, 1499–1509, 1529–1543, 1548–1566, 1572–1577, 1591–1629, 1635–1686, 1706–1725, 1729–1733, 1760–1796, 1894–1916, 2043–2088, 2104–2117, 2158–2168, 2194–2220, 2211–2224, 2229–2258, 2280–2289, 2643–2651, 2677–2686, 2711–2717, 2740–2750, 2761–2768, and 2779–2783 (766 amino acids); OppB regions: 2950–2992, 3002–3018, 3043–3065, and 3097–3353 (340 amino acids); OppC regions: 3528–3602, and 3760–4009 (325 amino acids); OppD regions: 4531–4613, 4838–4890, and 5015–5236 (358 amino acids); and OppF regions: 5395–5476, 5744–5912, 6561–6718 and 6737–6776 (442 amino acids).

Maximum likelihood trees were constructed using RAxML-HPC2 on XSEDE (8.024) (Stamatakis, 2014) using the Cipres Web portal (Miller et al., 2010). For the analysis of the OppABCDF alignment the protein GAMMA model with the protein substitution matrix BLOSUM62 was used. Clade support was evaluated using Bootstrap analysis.

2.6. Bioinformatic analysis of annotated and newly postulated oppA gene homologues

In order to confirm that identified gene sequences annotated as "hypothetical proteins" were indeed *oppA* genes, a bioinformatic analysis was conducted on the protein sequences of both annotated and newly identified *oppA* genes. The mycoplasma OppA is an extracellular lipoprotein attached to the outer surface of the plasma membrane (Henrich et al., 1999). The cellular localization of the annotated and newly identified OppA proteins was therefore predicted using PSORTb version 3.0.2 (Yu et al., 2010) with the "Advanced Gram stain options" set to "negative without outer membrane", as is recommended for mycoplasma. The presence of a signal peptide was predicted using SignalP (Petersen et al., 2011) and PRED-LIPO. PRED-LIPO (Bagos et al., 2008) specifically predicts signal peptidase I and signal peptidase II associated with secretory protein and lipoproteins, respectively.

InterPro (Hunter et al., 2012) and MEME (Bailey et al., 2009) were used to identify family relationships, conserved domains and protein motifs in OppA proteins. InterPro uses signatures provided by a combination of several member databases. MEME is a web based motif discovery program that does not rely on existing information in databases to find motifs. MEME identified motifs simply indicate similarities amongst group members and do not necessarily coincide with a specific function. MEME was set to search for motifs between 6 and 100 amino acids in length with a maximum of 6 different motifs per sequence. The identified MEME motifs were compared to functional motifs as described for *M. hominis* OppA (Henrich et al., 1999). Similar motifs to those of *M. hominis* were also identified in *M. struthionis* sp. nov. str. Ms01, *Mycoplasma* sp. Ms02 and *M. nasistruthionis* sp. nov. str. Ms03 (Botes et al., 2010). These were also compared to the identified MEME motifs.

3. Results

3.1. Multiple sequence alignment and phylogenetic analysis of 16S rRNA sequences

The phylogeny resulting from the phylogenetic analysis of the *16S rRNA* sequences allows the distinction of three groups of mycoplasmas i.e. hominis, pneumoniae and spiroplasma (Fig. 1). Bootstrap values were high for all three groups retrieved in the analysis. Furthermore, there were clades within these groups which were retrieved with high bootstrap support. Many of the clades of *Mycoplasma* species shared hosts or infected closely related hosts.

3.2. Identification of opp operons in published mycoplasma genomes

A total of 55 *opp* operons were found in 31 of the NCBI mycoplasma genomes used in this study (Supplementary Table 1), but only 41 of these operons had genes annotated as *oppB*, *oppC*, *oppD* and *oppF*. Eight of the genomes have a single *opp* operon, 22 have two *opp* operons and one, *Mycoplasma alligatoris* A21JP2, has three *opp* operons. Of the single operon genomes, three belong to the hominis group, four to the pneumoniae group and one to the spiroplasma group. In eight of the genomes an *opp* operon could not be identified. These included *Candidatus* Mycoplasma haemolamae str. Purdue, *Candidatus* Mycoplasma haemominutum 'Birmingham 1', *Mycoplasma haemocanis* str. Illinois, *Mycoplasma haemofelis* str. Langford 1, *Mycoplasma ovis* str. Michigan, *Mycoplasma parvum* str. Indiana, *Mycoplasma suis* str. Illinois and *Mycoplasma wenyonii* str. Massachusetts, which are all hemotropic mycoplasmas or hemoplasmas.

Except for *Mycoplasma gallisepticum*, where the two operons were found in tandem, operons were found far apart within genomes and the order of genes within operons often differed between operons within the same species and between species. Different copies in a genome were usually distinguished by a different numbering annotation of their *opp* genes. This numbering annotation was inconsistent between genomes.

A completely annotated *opp* operon consisting of an *oppA*, *oppB*, *oppC*, *oppD* and *oppF* gene in a single operon could only be found in 13 of the annotated genomes available on NCBI (Supplementary Table 1). Of these, only *Mycoplasma bovis* Hubei-1, *Mycoplasma mycoides* subsp. *capri* LC str. 95010 and str. PG1 had a complete annotation for both of their *opp* operons. Seventeen of the species had only the *oppBCDF* genes annotated for one or both of their *opp* operons with 10 of the species having one or more of the *oppBCDF* genes annotated as a hypothetical protein, *dpp*, *pgk* or *amiD*. In most instances the products of the hypothetical proteins were defined as ABC-type proteins or oligopeptide ABC transporter permeases. Dpp and Opp transporters are thought to differ only with regard to their substrate preference, i.e. dipeptides vs. oligopeptides (Dassa and Bouige, 2001). *OppC* is also annotated as *amiD* (Staats et al., 2007) which is found in *Mycoplasma pneumoniae* M129. In *Mycoplasma mobile* 163 K, *pgk* (phosphoglycerate kinase) is the annotation given for *oppD*. This gene contains a N-terminal phosphoglycerate kinase domain and a C-terminal oligopeptide ABC transporter ATPase domain. Both the PGK enzyme and the OppD protein contain a nucleotide-binding site and have an ATPase function.

In addition to the above, three *oppB*, four *oppBC* and five *oppF* genes were found as single genes or incomplete operons (Supplementary Table 2) and were therefore excluded from further analyses.

3.3. Identification of opp operons in mycoplasmas infecting ostriches

In the ostrich-infecting mycoplasma genomes, an *opp* operon was previously identified in both *M. struthionis* sp. nov. str. Ms01 and *Mycoplasma* sp. Ms02 (Pretorius, 2009; Steenmans, 2010). A second *opp* operon could be identified in the genome of *Mycoplasma* sp. Ms02 and two *opp* operons in *M. nasistruthionis* sp. nov. str. Ms03. The gene sequences of the five *opp* operons were submitted to NCBI and their accession numbers are listed in Supplementary Table 1.

3.4. Identification of oppA gene homologues in mycoplasma genomes

The positions in which the *oppA* gene relative to the rest of the operon were found, are shown in Fig. 2. Of all the NCBI genomes listed in Supplementary Table 1, only 13 species (16 operons) contained an annotated *oppA* gene as part of an *opp* operon. In operons in which an

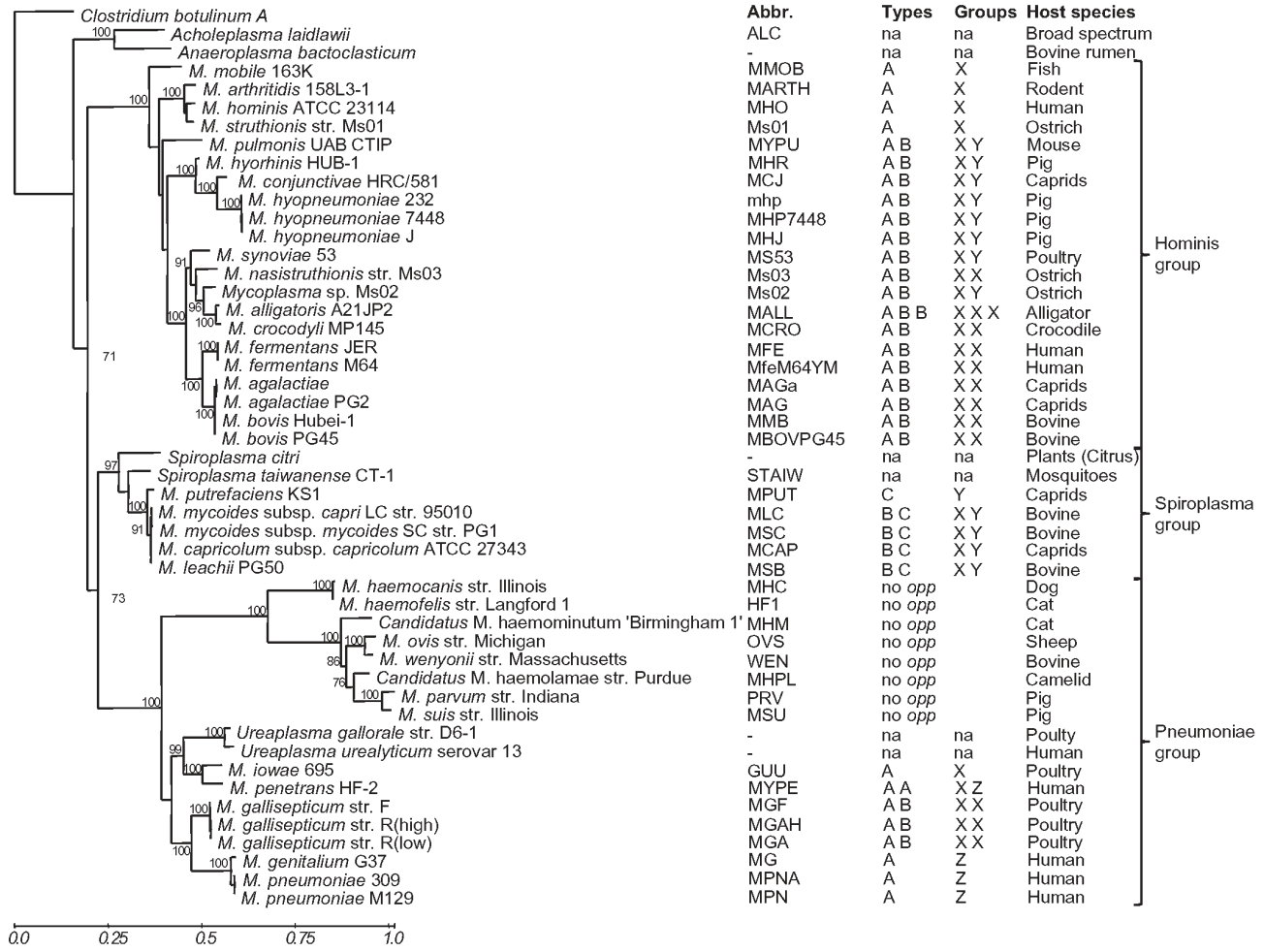


Fig. 1. The maximum likelihood phylogeny of *16S rRNA* sequences. Bootstrap values (≥75%) are indicated above or below the lines. The table indicates the NCBI abbreviation of genes within the respective genomes, the OppABCDF type (A, B, or C) and the gene arrangement group (X, Y, or Z) of the *opp* operons in the genome and host species. Abbreviations: na, not analyzed; no *opp*, no *opp* operons identified within the genome.

oppA was not annotated, genes on either side of the *oppBCDF* operon were evaluated using gene size and tBLASTx. Sixteen operons contained a gene directly before the *oppB* or directly after the *oppF* gene that had the coding title description of lipoprotein, oligopeptide ABC-transporter permease or substrate-binding protein. In six instances a gene with such a description could be found downstream of the *oppBCDF* operon separated by either one or two genes. Gene size and sequence similarity analysis identified most of these genes as *oppA*, but for one of the operons of *Mycoplasma gallisepticum* str. R(low) (MGA_0226) and *Mycoplasma gallisepticum* str. R(high) (MGAH_0226) a PSI-BLAST was needed to confirm their identity.

Eleven of the operons had genes with the coding title "hypothetical protein" adjacent to the *oppBCDF* genes. Of these, nine could be identified as *oppA* based on gene size and tBLASTx analysis, while two (*M. gallisepticum* str. F (MGF_2297) and *M. penetrans* HF-2 (MYPE5560)) could only be confirmed using PSI-BLAST analysis. In four of the species (*M. genitalium* G37, *M. penetrans* HF-2 (MYPE7570-MYPE7620), *M. pneumoniae* 309 and M129) an *oppA* gene was identified several thousand base pairs downstream of the *oppBCDF* operon. Except for *M. pneumoniae* M129, they all had a coding title indicating a lipoprotein as gene product, but sequence similarity to other known *oppA* genes could only be confirmed using PSI-BLAST.

In *M. penetrans* HF-2 five copies of the *oppA* gene were identified downstream of the *opp* operon and adjacent to each other (MYPE7570-MYPE7620) (Supplementary Fig. 1). The first gene copy was found to be separated into two open reading frames (MYPE7570 and MYPE7580). The five copies are paralogous and probably evolved by recent duplication events within this species since BLAST analysis revealed them to be more closely related to each other (75–85% amino acid similarity) than to any other mycoplasma OppA protein.

In all three ostrich-infecting mycoplasma genomes, an *oppA* gene was found as part of an *oppBCDF* operon and was situated directly before the *oppB* gene except for one of the *Mycoplasma* sp. Ms02 operons where the *oppA* gene was separated by three genes on the *oppF* side. In all instances the *oppA* gene could be identified based on gene size and sequence similarity.

Additionally, the order of the genes in the different genomes was analyzed and the gene arrangements were grouped into three types i.e. X, Y and Z as indicated in Fig. 2. These groupings relative to the distribution of species within the 16S rRNA and OppABCDF phylogeny are shown in Figs. 1 and 3.

3.5. Amino acid sequence alignment and phylogenetic analysis of OppABCDF

The *OppBCDF* part of the operon was conserved across *Mycoplasma* species but the annotated and newly identified OppA proteins differed in sequence length and had poor sequence similarity. The entire coding

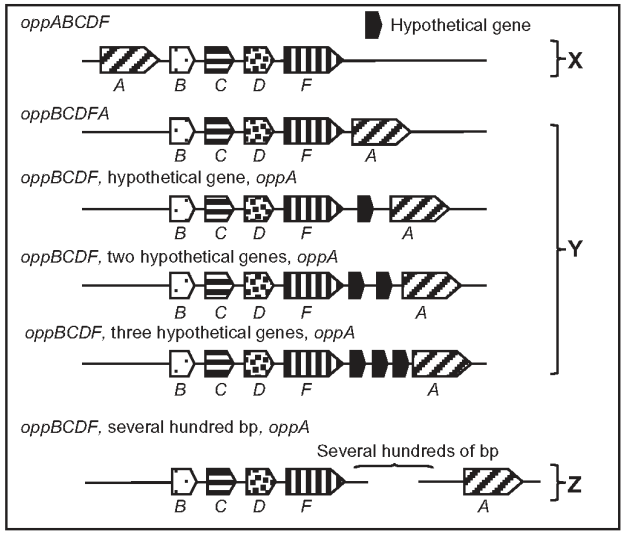


Fig. 2. The gene arrangement within *opp* operons in mycoplasma genomes. The genes in mycoplasma *opp* operons were found to be arranged in the order *oppABCDF*, *oppBCDF* hypothetical gene *oppA*, *oppBCDF* two hypothetical genes *oppA*, *oppBCDF* three hypothetical genes *oppA* or *oppBCDF* and *oppA* located several hundreds of base pairs downstream. The gene arrangement were grouped into X, Y and Z as indicated.

region of OppA could not be aligned with confidence and therefore only the conserved regions of the alignment were used (766 amino acids), in combination with the OppBCDF amino acid sequences (OppB 340 amino acids; OppC 325 amino acids; OppD 358 amino acids; OppF 442 amino acids), to construct a resolved and supported phylogenetic tree.

Based on the OppABCDF phylogeny (Fig. 3) the *opp* operons could be divided into three groups: Types A, B and C. The type did not correspond to the position of the operons relative to one another in the genome. When a phylogenetic analysis of each of the individual *opp* genes was performed three groups were also consistently retrieved but with lower bootstrap support and less resolution within the three clades. The function of the different genes in a single operon is therefore closely associated causing them to evolve as a unit despite the *oppA* gene not always being clustered together with the *oppBCDF* genes.

All mycoplasmas with a single *opp* operon clustered in Type A except for *Mycoplasma putrefaciens* KS1 (Type C). With the exception of *M. penetrans* HF-2 and *M. alligatoris* A21JP2, *opp* operon copies within the same species grouped into different types in the phylogeny. In the case of *M. penetrans* HF-2, where both operons were Type A they were distinguished as Types A1 and A2 and in *M. alligatoris* A21JP2 where two operons belong to Type B as Type B1 and B2. *Opp* operons of the hominis and the pneumoniae groups were always Type A or Types A and B while *opp* operons of the spiroplasma group were always Type C or Types B and C (16S rRNA phylogeny, Fig. 1). Most of the *oppA* genes that were not annotated as *oppA* were within the Type A operons and had been annotated as hypothetical proteins or lipoproteins. Half of the *oppA* genes in Types B and C were also not annotated before with most of them annotated as lipoproteins or ABC transporter permeases.

3.6. Bioinformatic analysis of annotated and newly postulated oppA gene homologues

PSORTb was only able to predict the cellular localization as being extracellular of the cytoplasmic membrane in 32% of the proteins and this included only 4 of the 16 OppA proteins already annotated (Supplementary Table 3). This was due the inability of the software to detect lipoprotein motifs (Yu et al., 2010). With the exception of three proteins, SignalP was able to find a signal peptide sequence in the N-terminal of all OppA proteins. PRED-LIPO on the other hand was able to identify a signal sequence in all OppA proteins except one (*Mycoplasma conjunctivae* HRC/581 Type A) but was not always able to predict the signal sequence as being that of a lipoprotein. This included proteins that had already been annotated as OppA and might be due to the fact that PRED-LIPO is optimized for Gram-positive bacteria and not mycoplasmas that lack a cell wall. The Prosite motif, PROKAR_LIPOPROTEIN (PS51257, signal peptide containing the lipoprotein attachment site) was found by InterPro within the first 35 amino acids of most of the OppA proteins.

InterPro could predict the functional classification of the OppA proteins in Type B. These proteins belong to the InterPro family IPR016880, which is dedicated to the solute-binding component of ABC-type oligopeptide transport systems in *Mycoplasmataceae* (Supplementary Table 3). Domains and their relative position found within Type B include the (i) bacterial extracellular solute-binding protein, family 5 (IPR000914, Pfam PF00496) within the region 146–804, (ii) dipeptide-binding protein, domain 3 (CATH superfamily 3.10.105.10) within the region 482–992 and (iii) periplasmic binding protein-like II domain (SCOP superfamily 53850) within the region 144–947. The gene ontology terms predicted for biological process was transport (GO: 0006810) and molecular function was transporter activity (GO: 005215).

InterPro could not predict a family for the OppA proteins in Type C (Supplementary Table 3) although similar domains as for Type B were found and included (i) bacterial extracellular solute-binding protein, family 5 (IPR000914, Pfam PF00496) within the region 85–550, (ii) dipeptide-binding protein, domain 1 (CATH superfamily 3.90.76.10) within the regions 74–161 and 243–287, (iii) periplasmic binding protein-like II domain (CATH superfamily 3.40.190.10) within the region 271–443, (iv) dipeptide-binding protein, domain 3 (CATH superfamily 3.10.105.10) within the region 757–903 and (v) periplasmic binding protein-like II domain (SCOP superfamily 53850) within the regions 42–158 and 236–446. The gene ontology terms predicted were the same as for Type B; transport (GO: 0006810) and transporter activity (GO: 005215). Additionally, a conserved site (IPR023765, PS01040) containing the bacterial extracellular solute-binding protein family 5 signature was predicted for *M. putrefaciens* KS1.

InterPro could not predict a family for most of the Type A OppA proteins, although some were found to belong to a family of uncharacterized conserved lipoproteins (IPR017012, UCP032899). In a few of the Type A OppA proteins, domains similar to those of Types B and C were identified and included the (i) dipeptide-binding protein, domain 1 (CATH superfamily 3.90.76.10) within the region 162–373, (ii) periplasmic binding protein-like II domain (SCOP superfamily 53850) within the region 165–658 and (iii) periplasmic binding protein-like II domain (CATH superfamily 3.40.190.10) within the region 317–480. In addition to these domains, an unintegrated signature (PD024071) was found within the regions 129–158 and 494–1008 amongst members of the pneumoniae group.

The motif discovery program MEME was used to identify conserved sequence motifs within the OppA proteins of each of the types (A, B and C) (Fig. 4, Supplementary Tables 4 and 5). The MEME suite could identify six conserved motifs within OppA Types A, B and C (motifs 1A–6A, motifs 1B–6B and motifs 1C–6C respectively) and all motifs overlapped with one or more domains or motifs identified using InterPro. Within the OppA proteins of Types B and C the position of each motif was in the same region in each sequence and the same motifs were found in all members of a type.

In OppA Type A, the positions of all motifs varied much more. Only motifs 1A and 2A were found in all group members with the most common motif positions being 620–820 and 160–280, respectively (Fig. 4, Supplementary Table 4). The exceptions were *M. gallisepticum* strains, *M. penetrans* HF-2 (MYPE5560) and *M. nasistruthionis* sp. nov. str. Ms03 in which the position of motifs 1A and 2A differed considerably from the rest of the members. With a few exceptions, motif 3A and motif 5A were only found in species within the pneumoniae group,

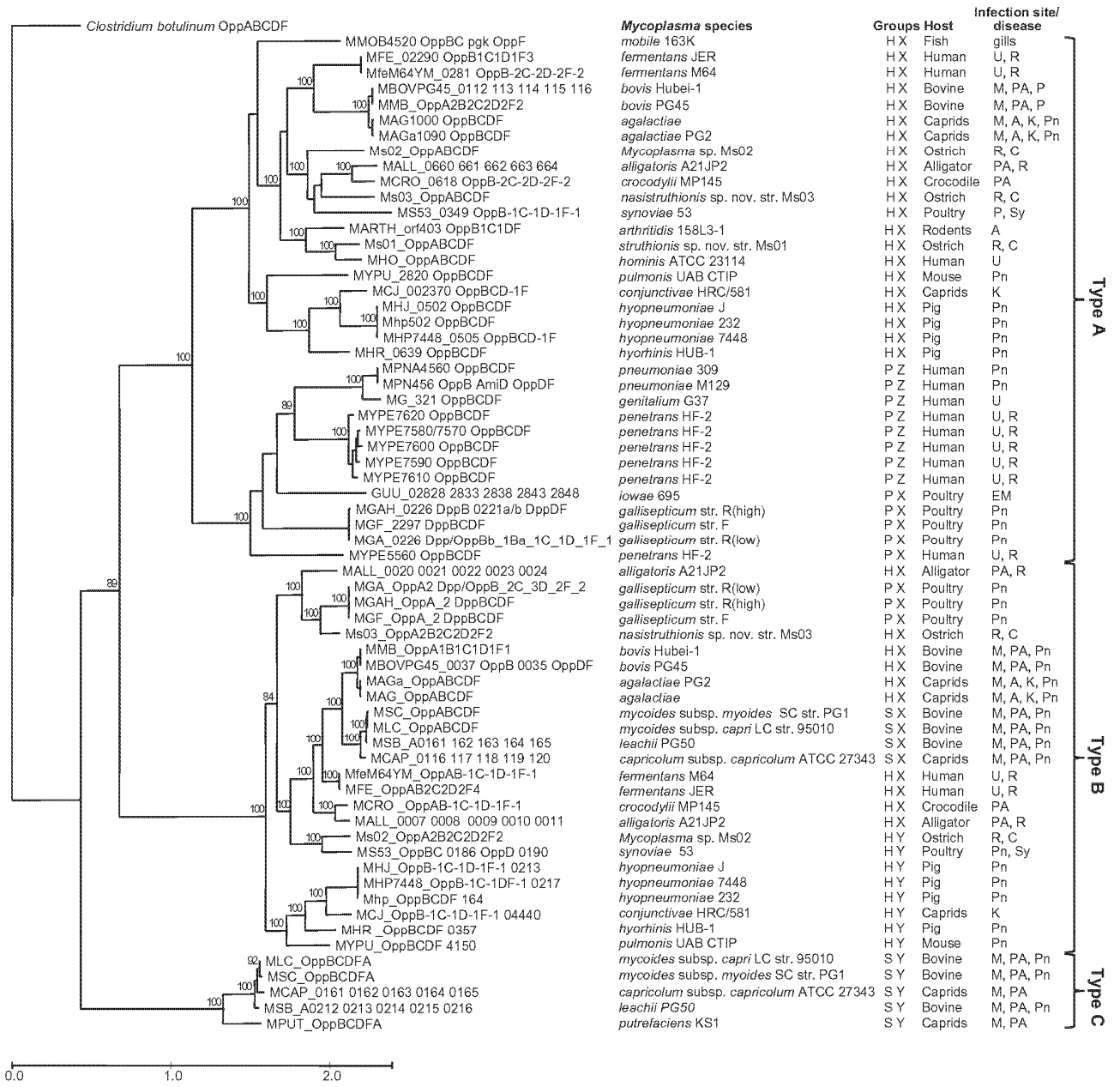


Fig. 3. The maximum likelihood phylogeny of OppABCDF amino acid sequences. Bootstrap values (≥75%) are indicated above the lines. Each operon is indicated with the species abbreviation followed by the annotation of each gene in the order that they occur within the operon (ABCDF or BCDF A). The table indicates *Mycoplasma* species, the 16S rRNA phylogenetic group (H, P or S) and the gene arrangement group (X, Y or Z), host species and infection site/disease. Abbreviations: A, arthritis; C, conjunctivitis; Em, embryonic mortality; H, hominis 16S rRNA phylogenetic group; K, keratoconjunctivitis; M, mastitis; P, pneumoniae 16S rRNA phylogenetic group; Pn, pneumonia; PA, polyarthritis; R, respiratory infection; S, spiroplasma 16S rRNA phylogenetic group; Sy, synovitis; U, urogenital infection.

motif 4A was only found in species within the hominis group, while motif 6A was found in species of both the pneumoniae and the hominis groups.

The positioning of the known functional motifs, as described for *M. hominis* OppA was compared to the MEME motifs identified in the Type A group. Amongst the Type A members the oligopeptide-binding site and Walker B motifs were consistently identified as MEME motifs and therefore the MEME analysis could guide the identification of these functional regions. MEME motifs 1A and 2A contained the Walker B and oligopeptide-binding motif of *M. hominis* respectively, but the Walker A and signal peptidase II site was not identified as a MEME motif. Motifs similar to the functional motifs described for *M. hominis* were identified in the three ostrich-infecting mycoplasmas (Supplementary Table 6). Similar to *M. hominis*, the Walker B and oligopeptide-binding motifs of *M. struthionis* sp. nov. str. Ms01, *Mycoplasma* sp. Ms02 and *M. nasistruthionis* sp. nov. str. Ms03 were identified as motifs 1A and 2A. As in *M. hominis*, MEME did not identify the signal peptidase II site, but different to *M. hominis*, all the putative Walker A motifs were also identified as part of MEME motif 1A. The postulated Walker A motif of each of the ostrich-infecting mycoplasmas have a 75% identity to the Walker A consensus sequence and only 37.5% to the Walker A of *M. hominis* (Supplementary Table 6). In

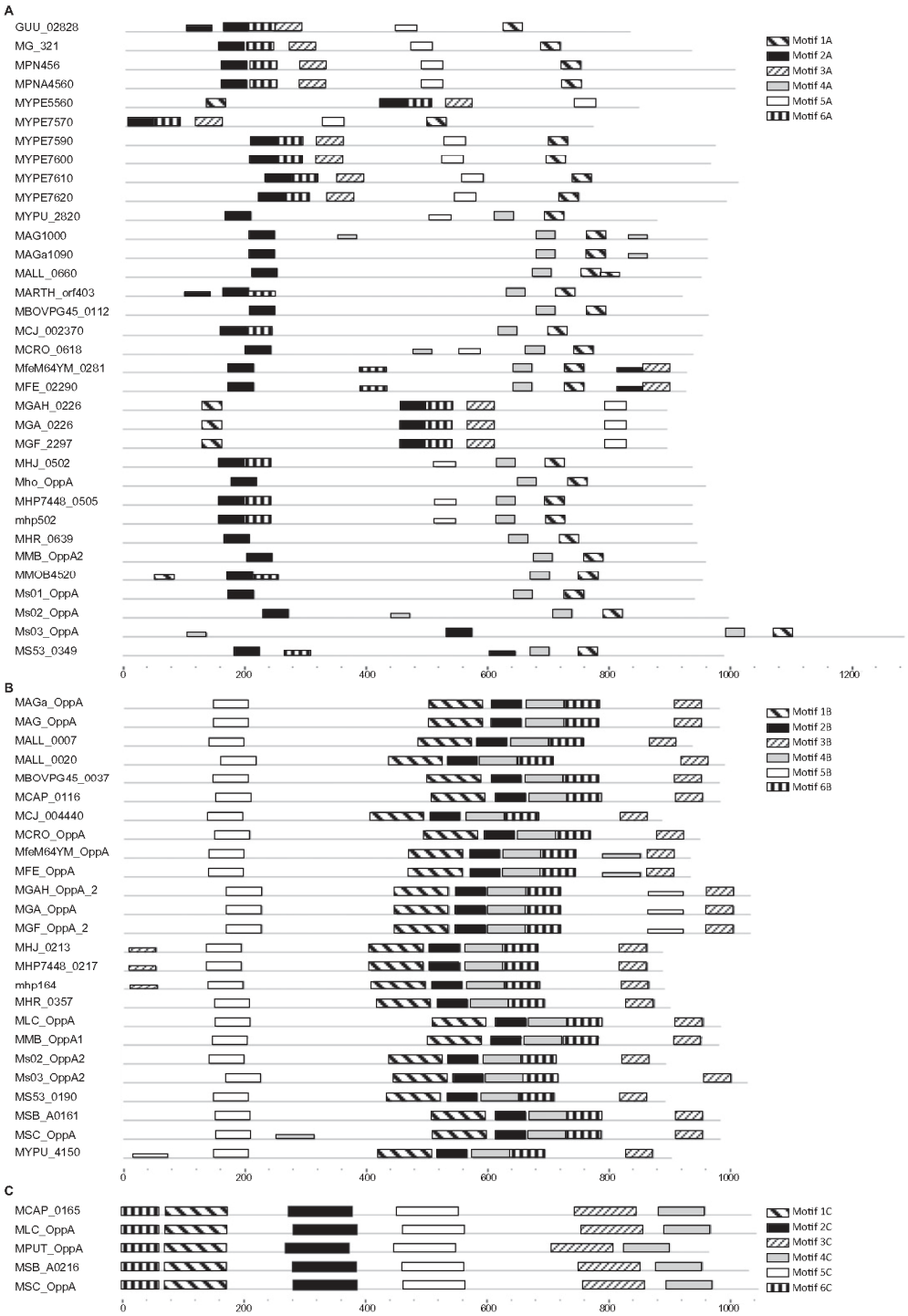


Fig. 4. Locations of the MEME motifs with in Types A, B and C *opp* operons. Motifs are non-overlapping sites with a p-value better than 0.0001. The height of the motif "block" is proportional to −log(p-value) and truncated at the height for a motif with a p-value of 1e-10.

general, Walker A motifs involved in ATPase activity contain a highly conserved lysine residue in position 7 (Ter Beek et al., 2014) as is the case with *M. hominis* (Hopfe and Henrich, 2004). The Walker A motif of *M. struthionis* sp. nov. str. Ms01 contains a leucine and *Mycoplasma* sp. Ms02 and *M. nasistruthionis* sp. nov. str. Ms03 a tyrosine in position 7. They are therefore not likely to have ATPase activity.

4. Discussion

Phylogenetic analysis of *16S rRNA* sequences gave a phylogeny (Fig. 1) which is in agreement with similar analyses of *16S rRNA* sequences of *Mycoplasma* species (Woese et al., 1980; Maniloff et al., 1985; Weisburg et al., 1989) and reflects their phylogenetic

relationships including those of the ostrich-infecting mycoplasmas. Similar to Brown (2010), the hemoplasma species grouped within the pneumoniae group.

All hemoplasma genomes were found not to contain any *opp* operon. Hemoplasmas are unique amongst mycoplasmas as they are blood-borne pathogens that infect the erythrocyte and are transmitted through arthropod vectors, such as lice and ticks (Messick, 2004; Guimaraes et al., 2014). The import of additional peptides through an Opp system might therefore not be required in hemoplasmas which therefore appears to be a further step in their degenerative evolution. The remaining mycoplasma genomes were all found to encode at least one *opp* operon with most mycoplasmas having two and one three. Furthermore based on phylogenetic analysis of the combined OppABCDF amino acid sequences the *opp* operons could be grouped into Types A, B and C (Fig. 3).

More than one copy of the *opp* operon is common in Gram-positive bacteria. Most lactic acid bacteria, such as *Lactococcus lactis* have two copies of the *opp* operon (Doeven et al., 2005; Eitinger et al., 2011) while *in silico* analysis of the *Staphylococcus aureus* genome has revealed four *opp* operons (Hiron et al., 2007). The expression of all four *opp* operons was, however, found to vary under different experimental conditions such as nutritional availability and the presence of antibiotics (Hiron et al., 2007; Date et al., 2014). These authors (Hiron et al., 2007; Date et al., 2014) concluded that multiple copies of the same transport systems and their differential expression may be an adaptive strategy towards survival under different conditions.

Studies in *Mycoplasma agalactiae* PG2 and *Mycoplasma hyopneumoniae* 7448 found expression of both of their *opp* operon copies (Pinto et al., 2009; Cacciotto et al., 2010). In *M. gallisepticum*, however, the functionality might be influenced by the presence of a stop codon in the Type A *oppB* gene of strain R(high) and the Type A *oppC* gene of the R(low) strain. Browning et al. (2011) also indicated the Type B *oppF* gene of R(low) to be truncated. We found most of the OppF proteins in Type B to be shorter (381–698 aa) than that of Type A (758–883 aa) without affecting the protein motifs (Walker A, C-motif and Walker B). It is therefore unlikely that the shorter amino acid sequence has an effect on the functionality of the OppF protein.

An *oppA* gene could be identified as part of all *opp* operons investigated in this study. The position of the *oppA*, however, differed between operons within the same species and between species (Fig. 2). Genome rearrangement is a common phenomenon within mycoplasma genomes and may have contributed to the location of these genes. In the Type A2 operon of *M. penetrans* HF-2, five copies of the *oppA* gene were identified. Oligopeptide transporters with more than one substrate domain have been reported in many bacteria. *Borrelia burgdorferi* has a single OppBCDF transporter that can interact with five different OppA domains, three of which are located within the same operon as *oppBCDF* (Wang et al., 2004). These five *oppA* genes are individually regulated, have different substrate affinities and expression varies with environmental conditions. This, together with a higher level of expression as a result of multiple copies, may contribute to the bacterium's ability to adapt to the environment as well as the host (Medrano et al., 2007; Raju et al., 2011). This may also apply to *Mycoplasma* species.

Different bioinformatic analyses were performed in an attempt to identify domains or motifs that can be associated with OppA proteins and therefore used to confirm the identity of hypothetical or lipoproteins during annotation. The *M. hominis* genome contains a single *opp* operon (Type A) and its OppA is probably the best characterized amongst *Mycoplasma* species. It is an extracellular protein that was found to contain a signal peptidase II recognition site, an oligopeptide-binding site as well as Walker A and Walker B motifs (Henrich et al., 1999). Using a combination of SignalP, PRED-LIPO and InterPro a signal peptide (containing the lipoprotein attachment site) was found in the first 35 amino acids of the majority of identified OppA proteins as is the case for *M. hominis*. With InterPro an oligopeptide binding domain could be identified in all OppA proteins in the Types B and C group, but only for a few in the Type A group. No Walker A or B motifs were identified in any of the sequences using InterPro not even in OppA of *M. hominis* with experimentally confirmed Walker A and B motifs.

MEME identified a consistent pattern of motifs in the OppA protein within Types B and C, which could assist in the identification of new OppA proteins (Fig. 4). Although not as consistent as in Types B and C, the pattern of motifs in Type A still allowed the identification of a gene as *oppA*. However the MEME motifs could also be used as a guide to identify functional motifs and thereby provide additional confirmation of the identity of an *oppA* gene. A comparison of the motifs identified in *M. hominis*, *M. struthionis* sp. nov. str. Ms01, *Mycoplasma* sp. Ms02 and *M. nasistruthionis* sp. nov. str. Ms03 using MEME found only the oligopeptide-binding site and Walker B. The Walker A motif is postulated to be highly conserved compared to that of the Walker B. One would therefore expect a MEME motif to be found in the region of the *M. hominis* Walker A given that the less conserved Walker B motif was identified. However, the regions upstream and downstream of the Walker A motif may not be conserved amongst species resulting in an inability of MEME to identify a region containing this motif. According to Matte and Delbaere (2010), a greater variation in Walker A sequences is being recognized as more ATP-binding proteins are characterized.

C. botulinum A was used to root both the 16S rRNA and OppABCDF phylogeny. The retrieval of the Type C operons in a basal position in the OppABCDF phylogeny (Fig. 3) appears to be in conflict with the 16S rRNA phylogeny as the spiroplasma group appears in a proximal position in that phylogeny (Fig. 1). However, when search tools were applied to search for the OppABCDF operons the most similar operons found were the Type C operons. This indicates that the basal position of Type C operons may rather be a product of the search tool than its true position in the phylogeny. More searches for OppABCDF operons in outgroups will have to be performed to discover which bacterial *opp* operons are most similar to the three types of mycoplasma *opp* operons identified in this study and which of these are basal in the phylogeny. However the OppABCDF phylogeny does indicate the three types of Opp proteins and their relationships. The phylogeny indicates that the distribution of the different *opp* operons may reflect the adaptation of a mycoplasma to a particular environment, i.e. host and infection site. The grouping of *opp* operons of the same species into separate types implies that across multiple species, the evolution of operon types were not the result of recent duplication events. Yu et al. (2014) was able to classify *Staphylococcus* isolates into different classes based on the arrangement of genes within the different *opp* operons. Although there are groupings (X, Y and Z) with similar gene order amongst the mycoplasma *opp* operons, this grouping is not the same as that based on sequence similarity and phylogeny.

Although a rare event, the distribution of species in the OppABCDF phylogeny might also for some species be ascribed to horizontal gene transfer (HGT) where a shared environment between co-infecting species can lead to the acquisition of new genes. This is viewed as a strategy to increase the gene pool for better adaption to environmental changes. In mycoplasma HGT has been postulated within human (Pereyre et al., 2009), ruminant (Sirand-Pugnet et al., 2007) and chicken (Vasconcelos et al., 2005) mycoplasmas. Although the spiroplasma ruminant species (*M. putrefaciens* KS1, *Mycoplasma capricolum* subsp. *capricolum* ATCC27343, *Mycoplasma leachii* PG50, *M. mycoides* subsp. *capri* LC str. 95010, *M. mycoides* subsp. *mycoides* SC str. PG1) are not closely related to the hominis ruminant species (*M. agalactiae*, *M. agalactiae* PG2, *M. bovis* Hubei-1 and *M. bovis* PG45) (see 16S rRNA phylogeny, Fig. 1) their Type B OppABCDF proteins are closely related. This indicates a possible HGT of the Type B operon genes while the Type C operon is orthologous to the spiroplasma and Type A to the hominis groups. Sirand-Pugnet et al. (2007) proposed that *M. agalactiae* acquired a number of genes, including those coding for the *opp* genes, from members of the *M. mycoides* cluster. This argument probably also holds for the *M. bovis* strains. Furthermore Browning et al. (2011) has proposed HGT of the *opp* operon from *M. synoviae* to *M. gallisepticum*. *M. gallisepticum* strains and *M. synoviae* 53 are not

closely related to one another, as they are members of the pneumoniae or hominis group respectively yet their Type B OppABCDF proteins are closely related. This may therefore again be the result of HGT of the Type B operon genes while the Type A operon is orthologous for both the pneumoniae and hominis groups.

5. Concluding remarks

All mycoplasmas (except hemoplasmas) were found to possess one or more *opp* operons in their genomes. Although the size of the *oppA* gene was fairly consistent across operons and species, low sequence similarity between species and between operons within a single species resulted in tBLASTx not always being able to identify distant homologues. By combining tBLASTx with a consensus sequence search in the form of PSI-BLAST, an *oppA* could be identified in all mycoplasma *opp* operons suggesting that it plays an essential role as part of the Opp transporter in *Mycoplasma* species. The parasitic lifestyle of mycoplasmas had allowed them to reduce their genome size. Therefore the retention of more than one copy of the *oppABCDF* transporter operon appears to indicate their essential role for mycoplasma survival. A signal peptide as well as a putitive oligopeptide binding site was identified in most of OppA protein sequences, which proves the extracellular nature as well as the substrate binding function of these proteins. Phylogenetic analysis revealed three types of *opp* operons that together with the unique InterPro and MEME domains and motifs found in each type suggest that each type may have a unique role and may be an adaptive strategy towards survival under different conditions. Our approach allowed the identification of *oppA* gene homologues for all mycoplasma *opp* operons and thereby provides a method for re-evaluating the current annotation of *oppA* genes of mycoplasma genomes.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gene.2014.12.036.

Acknowledgments

We would like to thank the Ostrich Business Chamber and The Technology and Human Resources for Industry Programme (THRIP) (Project reference TP50090719) for their financial support.

References

Bagos, P.G., Tsirigos, K.D., Liakopoulos, T.D., Hamodrakas, S.J., 2008. Prediction of lipoprotein signal peptides in Gram-positive bacteria with a hidden Markov model. J. Proteome Res. 7 (12), 5082–5093.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., Noble, W.S., 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37 (Web Server issue), W202–W208.

Berntsson, R.P., Doeven, M.K., Fusetti, F., Duurkens, R.H., Sengupta, D., Marrink, S.J., Thunnissen, A.M., Poolman, B., Slotboom, D.J., 2009. The structural basis for peptide selection by the transport receptor OppA. EMBO J. 28 (9), 1332–1340.

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

Botes, A., Pretorius, B., Steenmans, S., Bellstedt, D.U., 2010. The identification of an oligopeptide permease (Opp) transport system in ostrich mycoplasmas. *18th Meeting of the International Organization for Mycoplasmology*, Chianciano Terme, Siena, Italy.

Brown DR. 2010. Phylum XVI. Tenericutes Murray 1984a, 356^{VP} (Effective publication: Murray 1984b, 33.). In *Bergey's Manual® of Systematic Bacteriology*, (ed. NR Krieg, JT Staley, DR Brown, BP Hedlund, BJ Paster, NL Ward, W Ludwig, WB Whitman), pp. 567-723. Springer New York.

Browning, G.F., Marenda, M.S., Noormohammadi, A.H., Markham, P.F., 2011. The central role of lipoproteins in the pathogenesis of mycoplasmoses. Vet. Microbiol. 153, 44–50.

Cacciotto, C., Addis, M.F., Pagnozzi, D., Chessa, B., Coradduzza, E., Carcangiu, L., Uzzau, S., Alberti, A., Pittau, M., 2010. The liposoluble proteome of *Mycoplasma agalactiae*: an insight into the minimal protein complement of a bacterial membrane. BMC Microbiol. 10, 225–236.

Dassa, E., Bouige, P., 2001. The ABC of ABCs: a phylogenetic and functional classification of ABC systems in living organisms. Res. Microbiol. 152, 211–229.

Date, S.V., Modrusan, Z., Lawrence, M., Morisaki, J.H., Toy, K., Shah, I.M., Kim, J., Park, S., Xu, M., Basuino, L., et al., 2014. Global gene expression of methicillin-resistant *Staphylococcus aureus* USA300 during human and mouse infection. J. Infect. Dis. 209 (10), 1542–1550.

Delcher, A.L., Bratke, K.A., Powers, E.C., Salzberg, S.L., 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 23 (6), 673–679.

Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J.M., Taly, J.F., Notredame, C., 2011. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids Res. 39 (Web Server issue), W13–W17.

Doeven, M.K., Kok, J., Poolman, B., 2005. Specificity and selectivity determinants of peptide transport in *Lactococcus lactis* and other microorganisms. Mol. Microbiol. 57 (3), 640–649.

Eitinger, T., Rodionov, D.A., Grote, M., Schneider, E., 2011. Canonical and ECF-type ATP-binding cassette importers in prokaryotes: diversity in modular organization and cellular functions. FEMS Microbiol. Rev. 35 (1), 3–67.

French, C.T., Lao, P., Loraine, A.E., Matthews, B.T., Yu, H., Dybvig, K., 2008. Large-scale transposon mutagenesis of *Mycoplasma pulmonis*. Mol. Microbiol. 69 (1), 67–76.

Garmory, H.S., Titball, R.W., 2004. ATP-binding cassette transporters are targets for the development of antibacterial vaccines and therapies. Infect. Immun. 72 (12), 6757–6763.

Glass, J.I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M.R., Maruf, M., Hutchison, C.A., Smith, H.O., Venter, J.C., 2006. Essential genes of a minimal bacterium. Proc. Natl. Acad. Sci. U. S. A. 103 (2), 425–430.

Guimaraes, A.M.S., Santos, A.P., do Nascimento, N.C., Timenetsky, J., Messick, J.B., 2014. Comparative genomics and phylogenomics of hemotrophic mycoplasmas. PLoS One 9 (3), e91445.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.

Henrich, B., Feldmann, R.C., Hadding, U., 1993. Cytoadhesins of *Mycoplasma hominis*. Infect. Immun. 61 (7), 2945–2951.

Henrich, B., Hopfe, M., Kitzerow, A., 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 (16), 4873–4878.

Hiron, A., Borezée-Durant, E., Piard, J.C., Juillard, V., 2007. Only one of four oligopeptide transport systems mediates nitrogen nutrition in *Staphylococcus aureus*. J. Bacteriol. 189 (14), 5119–5129.

Hopfe, M., Henrich, B., 2004. OppA, the substrate-binding subunit of the oligopeptide permease, is the major ecto-ATPase of *Mycoplasma hominis*. J. Bacteriol. 186 (4), 1021–1028.

Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T.K., Bateman, A., Bernard, T., Binns, D., Bork, P., Burge, S., et al., 2012. InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res. 40 (Database issue), D306–D312.

Klepsch, M.M., Kovermann, M., Löw, C., Balbach, J., Permentier, H.P., Fusetti, F., de Gier, J.W., Slotboom, D.J., Berntsson, R.P., 2011. *Escherichia coli* peptide binding protein OppA has a preference for positively charged peptides. J. Mol. Biol. 414 (1), 75–85.

Langer, S., 2009. Proposal for molecular tools for the epidemiology of contagious bovine pleuro pneumonia and classification of unknown *Mycoplasma* sp. isolated from *Struthio camelus*. Vol Mag. rer.nat. Universität Wien, Wien.

Levdikov, V.M., Blagova, E.V., Brannigan, J.A., Wright, L., Vagin, A.A., Wilkinson, A.J., 2005. The structure of the oligopeptide-binding protein, AppA, from *Bacillus subtilis* in complex with a nonapeptide. J. Mol. Biol. 345 (4), 879–892.

Maniloff, J., Rogers, M.J., Simmons, J., Walker, R.T., Weisburg, W.G., Woese, C.R., 1985. Analysis of *Mycoplasma* 5 s ribosomal-RNA sequences — evolution of the smallest cells. Biophys. J. 47 (2), A315.

Matte, A., Delbaere, L.T.J., 2010. ATP-binding motifs. Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd, Chichester.

Medrano, M.S., Ding, Y., Wang, X.G., Lu, P., Coburn, J., Hu, L.T., 2007. Regulators of expression of the oligopeptide permease A proteins of *Borrelia burgdorferi*. J. Bacteriol. 189 (7), 2653–2659.

Messick, J.B., 2004. Hemotrophic mycoplasmas (hemoplasmas): a review and new insight into pathogenic potential. Vet. Clin. Pathol. 33 (1), 2–13.

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Gateway Computing Environments Workshop (GCE), pp. 1–8.

Movahedi, A., Hampson, D.J., 2010. Evaluation of recombinant *Brachyspira pilosicoli* oligopeptide-binding proteins as vaccine candidates in a mouse model of intestinal spirochaetosis. J. Med. Microbiol. 59, 353–359.

Nicolás, M.F., Barcellos, F.G., Hess, P.N., Hungria, M., 2007. ABC transporters in *Mycoplasma hyopneumoniae* and *Mycoplasma synoviae*: insights into evolution and pathogenicity. Genet. Mol. Biol. 30 (1), 202–211.

Notredame, C., Higgins, D.G., Heringa, J., 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. 302 (1), 205–217.

Pereyre, S., Sirand-Pugnet, P., Beven, L., Charron, A., Renaudin, H., Barre, A., Avenaud, P., Jacob, D., Couloux, A., Barbe, V., et al., 2009. Life on arginine for *Mycoplasma hominis*: clues from its minimal genome and comparison with other human urogenital mycoplasmas. PLoS Genet. 5 (10), e1000677.

Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8 (10), 785–786.

Pinto, P.M., Klein, C.S., Zaha, A., Ferreira, H.B., 2009. Comparative proteomic analysis of pathogenic and non-pathogenic strains from the swine pathogen *Mycoplasma hyopneumoniae*. Proteome Sci. 7, 45.

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. *Department of Biochemistry*, Vol MSc (Biochemistry). University of Stellenbosch, Stellenbosch, South Africa.

Raju, B.V., Esteve-Gassent, M.D., Karna, S.L., Miller, C.L., Van Laar, T.A., Seshu, J., 2011. Oligopeptide permease A5 modulates vertebrate host-specific adaptation of *Borrelia burgdorferi*. Infect. Immun. 79 (8), 3407–3420.

Razin, S., Yogev, D., Naot, Y., 1998. Molecular biology and pathogenicity of mycoplasmas. Microbiol. Mol. Biol. Rev. 62 (4), 1094–1156.

Sirand-Pugnet, P., Lartigue, C., Marenda, M., Jacob, D., Barre, A., Barbe, V., Schenowitz, C., Mangenot, S., Couloux, A., Segurens, B., et al., 2007. Being pathogenic, plastic, and sexual while living with a nearly minimal bacterial genome. PLoS Genet. 3 (5), 744–758.

Staats, C.C., Boldo, J., Broetto, L., Vainstein, M., Schrank, A., 2007. Comparative genome analysis of proteases, oligopeptide uptake and secretion systems in *Mycoplasma* spp. Genet. Mol. Biol. 30 (1), 225–229.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30 (9), 1312–1313.

Steenmans, S., 2010. The identification, contiguous sequence annotation, cloning and site-directed mutagenesis of the P100 vaccine candidate gene of the ostrich mycoplasma *Ms02*. *Department of Biochemistry*, Vol MSc (Biochemistry). University of Stellenbosch, Stellenbosch, South Africa.

Tanabe, M., Atkins, H.S., Harland, D.N., Elvin, S.J., Stagg, A.J., Mirza, O., Titball, R.W., Byrne, B., Brown, K.A., 2006. The ABC transporter protein OppA provides protection against experimental *Yersinia pestis* infection. Infect. Immun. 74 (6), 3687–3691.

Ter Beek, J., Guskov, A., Slotboom, D.J., 2014. Structural diversity of ABC transporters. J. Gen. Physiol. 143 (4), 419–435.

Vasconcelos, A.T., Ferreira, H.B., Bizarro, C.V., Bonatto, S.L., Carvalho, M.O., Pinto, P.M., Almeida, D.F., Almeida, L.G., Almeida, R., Alves-Filho, L., et al., 2005. Swine and poultry pathogens: the complete genome sequences of two strains of *Mycoplasma hyopneumoniae* and a strain of *Mycoplasma synoviae*. J. Bacteriol. 187 (16), 5568–5577.

Wang, X.G., Kidder, J.M., Scagliotti, J.P., Klempner, M.S., Noring, R., Hu, L.T., 2004. Analysis of differences in the functional properties of the substrate binding proteins of the *Borrelia burgdorferi* oligopeptide permease (*opp*) operon. J. Bacteriol. 186 (1), 51–60.

Weisburg, W.G., Tully, J.G., Rose, D.L., Petzel, J.P., Oyaizu, H., Yang, D., Mandelco, L., Sechrest, J., Lawrence, T.G., Van Etten, J., et al., 1989. A phylogenetic analysis of the mycoplasmas: basis for their classification. J. Bacteriol. 171 (12), 6455–6467.

Wium M. unpublished results. The development of a DNA vaccine against *Mycoplasma nasistruthionis* sp. nov. for the use in ostrich In *Department of Biochemistry*, Vol PhD. University of Stellenbosch, Stellenbosch, South Africa.

Woese, C.R., Maniloff, J., Zablen, L.B., 1980. Phylogenetic analysis of the mycoplasmas. Proc. Natl. Acad. Sci. U. S. A. 77 (1), 494–498.

Yang, M., Johnson, A., Murphy, T.F., 2011. Characterization and evaluation of the *Moraxella catarrhalis* oligopeptide permease A as a mucosal vaccine antigen. Infect. Immun. 79 (2), 846–857.

Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S.C., Ester, M., Foster, L.J., et al., 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26 (13), 1608–1615.

Yu, D., Pi, B., Yu, M., Wang, Y., Ruan, Z., Feng, Y., Yu, Y., 2014. Diversity and evolution of oligopeptide permease systems in staphylococcal species. Genomics 104 (1), 8–13.

Yus, E., Maier, T., Michalodimitrakis, K., van Noort, V., Yamada, T., Chen, W.H., Wodke, J.A., Guell, M., Martinez, S., Bourgeois, R., et al., 2009. Impact of genome reduction on bacterial metabolism and its regulation. Science 326 (5957), 1263–1268.

Chapter 5 Evaluation of the *Mycoplasma nasistruthionis* sp. nov. str. Ms03 oligopeptide permease substrate-binding domain (OppA) as a DNA vaccine candidate in ostriches

5.1 Introduction

DNA vaccines provide researchers with the unique opportunity to utilise the cell's own mechanisms to produce transgenic protein antigens. DNA vaccines are DNA vectors that are constructed to have a combination of prokaryotic and eukaryotic elements. This allows for the production of the DNA vaccine in a prokaryotic organism such as *Escherichia coli* as well as the expression of antigenic protein within eukaryotic cells. The expressed antigenic protein, originating from the gene encoding it in the pathogen, can then in turn elicit a specific immune response that can lead to protection against the pathogen.

The aim of this part of the study was to evaluate the potential of using a DNA vaccine containing the *Mycoplasma nasistruthionis* sp. nov. str. Ms03 *oppA* gene to elicit an immune response in ostriches. The objectives were as follows:

- To clone the *oppA* gene of *M. nasistruthionis* sp. nov. str. Ms03 into the pGEM®-T Easy vector and to use site-directed mutagenesis (SDM) to change all the mycoplasma tryptophan encoding TGA codons to universal tryptophan encoding TGG codons in the Ms03 *oppA* gene.
- To sub-clone the mutated *oppA* gene into the pCI-neo, VR1012 and VR1020 vectors for use as DNA vaccines.
- To sub-clone the mutated *oppA* gene into pGEX-4T-1 in order to express and isolate the recombinant OppA protein for use as coating antigen in an anti-OppA ELISA.
- To evaluate the potential of the three DNA vaccines (pCI-neo_*oppA*, VR1012_*oppA* and VR1020_*oppA*) to elicit an anti-OppA humoral immune response in ostriches during a vaccination trial.

The results obtained are presented in an independent manuscript in this chapter with the supplementary data in Appendix 4 and 5 at the end of this dissertation. The following sections will give additional background to certain aspects of the experimental design described in the manuscript and the negative influence an outbreak of AI had on it.

5.2 Primer design

Primers were designed for SDM, *oppA* gene amplification and sequencing for the DNA vaccine development described in this chapter. Primers are polynucleotide sequences that are usually 18 to 30 nucleotides (nt) long. They are designed for the specific amplification of a unique region. For optimal amplification, the G+C content of primers should be higher than 40% and the G+C content of the forward and reverse primers should not differ by more than 5%. Binding of the primer to the template may be promoted with a G or C in the 3' end position (GC-clamp). The annealing temperature should be between 60°C and 75°C and the annealing temperature of the forward and reverse primers should be within 5°C of each other. Secondary structures and mononucleotide repeats (runs) within a primer should be avoided. Furthermore, interactions within (self-dimerizing) and between primers should be avoided as they cause primer dimer formation during PCR amplification. The primer sequence should be unique, thus homologue sequences should not be present within the template. This will ensure that the binding of the primer is specific to the region it was designed to bind to. For cloning, restriction enzyme sites can be added to the 5' end of the primers for cloning.

When designing primers for SDM, the mutagenesis site should be more or less in the middle of the primer sequence and the annealing temperature of the portion (18 to 24 nt) on both sides of the mutagenesis site should be similar (within 5°C).

5.3 Aspects of the vaccine trial

In this study, the ability of DNA vaccines to elicit an immune response in ostriches was assessed in a vaccination trial. In this trial, ostriches were vaccinated at three months of age before they were introduced into feedlot systems where mycoplasma exposure increases. The vaccine trial was conducted on a commercial ostrich farm near Oudtshoorn (the Klein Karoo region, Western Cape, South Africa). This implies that the exposure of the ostriches to pathogens and weather conditions could not be controlled. These are however the conditions under which a commercial vaccine would be used.

The vaccine trial consisted of three experimental groups that were vaccinated with pCI-neo_*oppA,* VR1012_*oppA* and VR1012_*oppA* constructs respectively, and a control group that did not receive any vaccine. The initial aim was to collect samples every three weeks up to 12 weeks after the first vaccination at day 0 with a second booster immunization at week 6. The objective of the first vaccination was to stimulate a primary immune response against the OppA protein. The objective of the second booster

injection was to stimulate a larger and more rapid secondary immune response implying that memory cells were formed due to the first vaccination.

An outbreak of highly pathogenic avian influenza (AI) was reported in the Oudtshoorn region in March 2011 (Van Helden et al. 2012) which coincided with the vaccination trial. AI is a notifiable disease and in ostrich symptoms include conjunctivitis, ocular discharge, nasal discharge, tracheal foam, pharyngitis, coughing and green urine (Toffan et al. 2010). High mortality rates (>80%) are reported in ostrich chicks under one month while the mortality rate in young birds (up to eight months) is between 15% and 60% whilst adults seldom show clinical symptoms of AI (Allwright et al. 1993). Ostriches (non-trial birds) were tested positive for AI on the farm where the vaccine trial was conducted and in accordance with South African legislation all ostriches on the farm were slaughtered. This included all the ostriches that were part of the trial. No further sampling was therefore possible. It was therefore only possible to collect blood and saliva samples at week 0 and week 3. At the start of the trial there were 30 ostriches in each group. Due to deaths amongst the trial birds in all groups, only 25 ostriches within the pCI-neo_*oppA* group, 21 ostriches within the VR1012_*oppA* group, 21 ostriches within the VR1020_*oppA* group and 23 ostriches within the control group survived which could be sampled. None of the trial birds were confirmed positive for AI and no symptoms were observed, although AI as a cause for the high mortality in the test birds cannot be excluded.

However, mortality of ostriches is affected by many factors such as stress, nutrition, temperature fluctuations and other diseases. Temperature fluctuations may have been a contributing factor since the trial was conducted in autumn (March to April) when there is a sudden drop in temperatures in the region. High mortality is reported in ostrich chicks (up to three months) with the mortality ranging between 10% to 50% while in three- to six-month-old ostriches the mortality ranges between 10% to 30% (Black and Glatz 2011; Wang 2012). The mortality within the trial groups of the current study ranged from 16% to 30% which is therefore not higher than normal. It should be noted that although the reasons for the deaths were unknown, the mortality in the control group was also high (23%). It can therefore be concluded that mortalities during the trial cannot be ascribed to the vaccination that the birds received.

For these reasons, the manuscript that follows contains only data from week 0 and week 3 and it was therefore only possible to test whether the DNA vaccine elicited a primary immune response against OppA.

5.4 Evaluation of the *Mycoplasma nasistruthionis* sp. nov. str. Ms03 oligopeptide permease substrate-binding domain (OppA) as a DNA vaccine candidate in ostriches

5.4.1 Contributions of co-authors

The following manuscript contains the original work of the author of this dissertation. The contributions of each of the co-authors were as follows:

Dr. Annelise Botes is a lecturer of Biochemistry at the University of Stellenbosch and the supervisor of this dissertation. In this capacity, she was involved in the conceptual development and execution of all aspects of this study.

Prof. Dirk U. Bellstedt is a professor of Biochemistry at the University of Stellenbosch and the co-supervisor of this dissertation. In this capacity, he was involved in the conceptual development and execution of this study.

Evaluation of the *Mycoplasma nasistruthionis* sp. nov. str. Ms03 oligopeptide permease substrate-binding domain (OppA) as a DNA vaccine candidate in ostriches

Martha Wium, Annelise Botes and Dirk U. Bellstedt

Abstract

DNA vaccines containing the *oppA* gene of *Mycoplasma nasistruthionis* sp. nov. str. Ms03, were developed and the potential of these DNA vaccine to elicit an immune response in ostriches was evaluated. The *oppA* gene was cloned and site-directed mutagenesis was used for codon correction. The mutated *oppA* gene was then sub-cloned into the DNA vaccine vectors: pCI-neo, VR1012 and VR1020. Three-month-old ostriches were vaccinated with these vaccines and the ability to elicit an anti-OppA immune response was evaluated by enzyme-linked immunosorbent assay (ELISA) in which recombinant OppA protein was used as coating antigen. The VR1020_*oppA* and pCI-neo_*oppA* constructs elicited a statistically significant immune response indicating their ability to induce a primary immune response in ostriches. This can therefore be viewed as the first step in the development of a DNA vaccine for the control of mycoplasma infections in ostriches.

Introduction

Mycoplasmas are a unique group of bacteria that can be distinguished from other bacteria by their lack of a cell wall and small A+T (adenine and thymine) rich genome (Razin et al. 1998). Mycoplasmas are host-specific (Pitcher and Nicholas 2005) and can infect the mucosal membranes of the eyes, nose, respiratory- and urogenital tract as well as causing joint inflammation. These infections cause numerous diseases in humans, animals and birds (Razin et al. 1998). Three ostrich-infecting *Mycoplasma* species have been identified, Ms01, Ms02 and Ms03 (Botes et al. 2005a) of which two Ms01 and Ms03 have been provisionally described and given the species names, *Mycoplasma struthionis* sp. nov. and *Mycoplasma nasistruthionis* sp. nov., respectively (Langer 2009). These mycoplasmas cause upper respiratory tract infections in ostriches. Mycoplasmas are specifically problematic amongst feedlot ostriches where infections can cause retardation of growth, mortalities and carcass downgrading, which have a financial impact on rearing of slaughter birds. The majority of ostriches destined for slaughter are reared in feedlots from three months of age (Hoffman and Lambrechts 2011).

Infections can be treated with antibiotics, but due to carrier conditions long-term treatment may be required. Long-term antibiotic treatment can lead to unwanted

accumulation of antibiotic residues in meat with concomitant risks for consumers (Nisha 2008; Darwish et al. 2013). An alternative control measure would be the use of vaccines. Live attenuated and inactivated whole organism mycoplasma vaccines are available for commercial use in chickens (Kleven 2008; Jacob et al. 2014), but chicken vaccines were found to be ineffective in treating mycoplasmosis in ostriches (Pretorius 2009). Traditional approaches for mycoplasma vaccine development for use in ostriches is hampered by the fact that ostrich-infecting *Mycoplasma* species are difficult to culture in large volumes and require complex medium for growth. This makes whole organism vaccines impractical and economically unfeasible for the relatively small ostrich vaccine market if compared to that of poultry.

As an alternative to whole organism vaccines, DNA vaccines are an attractive option. Over the past two decades, four DNA vaccines were registered for veterinary use in salmon, swine, horse and dogs (Kutzler and Weiner 2008; Giese 2012). Compared to inactivated whole organism vaccines and protein subunit vaccines, DNA vaccines have the advantage of being able to stimulate both humoral and cellular immune responses (Liu 2011). DNA vaccines also have an excellent safety profile (Schalk et al. 2006; Giese 2012) and can be produced at relatively low cost (Mahoney et al. 2000). Production of DNA vaccines also does not rely on large scale culturing of the pathogen which is problematic in some mycoplasmas including Ms03 (Waites and Talkington 2004; Flores-Medina et al. 2012; Citti and Blanchard 2013; Bueno et al. 2014). Furthermore, they are more temperature stable than other vaccines (Giese 1998) making them a particular attractive option for veterinary use under extensive farming conditions. Ostriches are usually farmed in semi-desert and desert regions where high temperatures are frequently found making cold storage of vaccines problematic.

The *Mycoplasma* genus is known for its small genome size ranging from 564 kbp in *Mycoplasma parvum* (do Nascimento et al. 2013) to 1 359 kbp in *Mycoplasma penetrans* (Sasaki et al. 2002). With it comes a parasitic lifestyle and dependence on the host for nutrients. Mycoplasmas therefore have a wide range of transporters that are essential for survival. Transport proteins have been indicated as possible targets in vaccine development (Garmory and Titball 2004; Grandi 2010). Among these are ATP binding cassette (ABC) transporters that play a crucial role in the energy dependent import of nutrients. The oligopeptide permease transport (Opp) system is an ABC transporter that is responsible for the import of oligopeptides. Similar to other ABC transporters (Berntsson et al. 2010) this transporter has five domains: OppA is the substrate-binding domain, OppB and OppC form a pore within the cell membrane and OppD and OppF hydrolyse ATP to provide energy for translocation of the oligopeptide. These domains are

each encoded by an open reading frame (ORF), which is usually found in tandem within an operon structure in the genome.

In bacteria, OppA has also been reported to play a role in spore formation (Rudner et al. 1991), cytadherence (Henrich et al. 1999), antibiotic resistance (Mistry et al. 2013), biofilm formation (Lee et al. 2004) and ecto-ATPase activity (Hopfe and Henrich 2004). OppA is an attractive target for vaccine development because of its essential function as well as its location on the external bacterial membrane surface, which exposes the protein to the immune system of the host. To this end OppA has been evaluated as a target in subunit vaccines for *Brachyspira pilosicoli* (Movahedi and Hampson 2010), *Moraxella catarrhalis* (Yang et al. 2011) and *Yersinia pestis* (Tanabe et al. 2006).

The genome of *M. nasistruthionis* sp. nov. str. Ms03 (Ms03, this abbreviation is used for brevity) was found to have two complete *opp* operons and therefore two possible *oppA* genes that could be used as vaccine candidates. Phylogenetic analysis of Opp proteins in *Mycoplasma* species allowed *opp* operons to be grouped into three types; A, B and C (See Chapter 3 of this dissertation, Wium et al. (2015)). Only Type A is found in all *Mycoplasma* species of the pneumoniae and hominis groups indicating an essential biological role for this operon. Furthermore, the immunogenicity and essential nature of the Type A OppA protein in *M. hominis* was experimentally confirmed (Henrich et al. 1993; Henrich et al. 1999). The *oppA* gene that forms part of the Type A *opp* operon of Ms03 was therefore selected as vaccine candidate gene (KM410300).

This study investigated the potential of using the Ms03 Type A *oppA* gene in the development of a DNA vaccine for the control of Ms03 infections in ostriches. For this purpose, the Ms03 *oppA* gene was cloned into the pGEM®-T Easy vector followed by codon correction using site-directed mutagenesis (SDM). The mutated *oppA* gene was sub-cloned into an expression vector and the gene product was expressed as a GST-fusion protein that was used to develop an anti-OppA enzyme-linked immunosorbent assay (ELISA). The mutated *oppA* gene was also sub-cloned into three different DNA vaccine vectors that were assessed for their ability to elicit an anti-OppA humoral immune response in ostriches using the aforementioned ELISA.

Materials and methods

Mycoplasma cultivation and DNA isolation

Cultures of Ms03 were obtained from Mr J.J. Gouws (Faculty of Veterinary Science, Onderstepoort, University of Pretoria). Genomic DNA (gDNA) was isolated from these cultures using a method described by Hempstead (1990). In short: The cells were harvested from 25 ml of Chanock's medium by centrifugation at 27 200 × g for 60 min at

4°C (Beckman Model J2-21 centrifuge, JA-20 rotor). The harvested cells were washed by resuspension in 1 ml concentrated TE buffer (50 mM Tris-HCl, pH 7.5, 10 mM EDTA) followed by centrifugation, 10 000 × g for 30 min at 4°C (Biofuge Fresco, Heraeus). The pellet was resuspended in 100 µl concentrated TE buffer and incubated at -20°C for 60 min or stored overnight. The frozen cells were lysed by a rapid thaw step at 55°C for 2 to 5 min. The contaminating proteins within the lysed cell sample was digested by adding 1 ml of SDS TE buffer (50 mM Tris-HCl, 10 mM EDTA, 1% (w/v) SDS, pH 7.5) and proteinase K (Roche, Switzerland) to a final concentration of 20 mg/ml. This was followed by an incubation step at 55°C for 30 min to allow the cell suspense to clarify. Proteins are denatured by the combination of high detergent concentration (SDS) and proteinase K (Hilz et al. 1975). RNA was digested by adding 6 µl RNase A (20 mg/ml, PureLinkTM RNase A, Invitrogen, USA) and 100 µl concentrated TE buffer followed by incubation for 60 min at 37°C. The SDS was precipitated from the solution by adding 100 µl of potassium acetate (5 M) and incubating it on ice for 30 min. The SDS precipitation was removed by centrifugation, 16 000 × g for 10 min at 4°C. Contaminating proteins were removed from the supernatant by three equal volume extractions with chloroform: isopropanol (24:1, v/v). In each extraction, the solutions were mixed by inverting them for 10 min followed by centrifugation, 16 000 × g for 2 min at room temperature. gDNA was precipitated by adding 1/10 of a volume of sodium acetate solution (3 M) and 2 volumes of 95% ethanol followed by an overnight incubation at -20°C. The precipitated gDNA was collected by centrifugation, 16 000 × g for 60 min at 4°C, and was washed by adding 1.5 ml 70% ethanol followed by incubation for 20 min at room temperature. gDNA was collected by centrifugation, 16 000 × g for 10 min at 4°C and air dried at room temperature before dissolving it in 50 µl Milli-Q® water by incubation at 4°C overnight.

The concentrations of gDNA samples were determined spectrophotometrically (NanoDrop spectrophotometer, ND-1000), while the integrity of the gDNA samples were assessed on a 1% (w/v) agarose gel.

Cloning and site-directed mutagenesis of the *oppA* gene

The *oppA* gene was amplified from Ms03 gDNA using proof-reading Kapa HiFi DNA polymerase (Kapa Biosystems, South Africa) according to the manufacturer's instructions. Restriction enzyme sites of MluI and AccI were added to the Ms03_P100F and Ms03_P100R primers respectively (Table 5.1). All primers used within this study were synthesized and purified by Integrated DNA Technologies (IDT, USA). PCR products were separated and visualised using a 1% (w/v) agarose gel containing 0.5 µg/ml ethidium bromide under UV light. PCR products were purified using the GFX™

PCR DNA and gel band purification kit (GE Healthcare Life Science, UK) according to the manufacturer's instructions.

For SDM the purified PCR product was cloned into the pGEM®-T Easy vector (Promega, USA). To this end, 3' terminal deoxyadenosine residues were added to the blunt ends of the purified PCR product using the A-tailing procedure as described in the pGEM®-T Easy system I manual. A DNA clean and concentrator™-5 kit (ZYMO Research, USA) was used to clean and concentrate the A-tailing product followed by overnight ligation (vector: insert ration 1:1 and 1:3) at 4°C into the pGEM®–T Easy vector, according to the manufacturer's instructions. The ligation products (pGEM_*oppA*) were transformed into *E. coli* JM109 cells and grown on Luria-Bertani (LB) agar plates (15 g/L) containing 100 µg/ml ampicillin (Sigma-Aldrich, USA), 0.16 mM isopropyl β-D-1-thiogalactopyranoside (IPTG, Bioline, UK) and 0.04 mg/ml X-Gal (Bioline, UK) for 16-18 h at 37°C.

Table 5.1 Primers used in this study

Primer name	**Sequence***	**Anneal. Temp**	**Comment**
Ms03_P100F	5'-ACGCGT**ATG**AAAAAATGATGATTATT-3'	55	For cloning into pCI-neo (5' MluI restriction enzyme site)
Ms03_P100R	5'-GTCGAC**CTA**TTTAGGTCTTACACCGT-3'		For cloning into pCI-neo (5' AccI restriction enzyme site)
Ms03_P100SalI_F	5'-GTCGAC**ATG**AAAAAATGGTGGTTAT-3'	55	For cloning into VR1012 (5' SalI restriction enzyme site)
Ms03_P100NotI_R2	5'-ATGCGGCCGC**CTA**TTTAGGTCTTACAC-3'		For cloning into VR1012 and pGEX-4T-1 (5' NotI restriction enzyme site)
Ms03_P100BamHI_F	5'-GGATCC**ATG**AAAAAATGGTGGTTAT-3'	60	For cloning into VR1020 and pGEX-4T-1 (5' BamIH restriction enzyme site)
Ms03_P100BamHI_R	5'-GGATCC**CTA**TTTAGGTCTTACACCG-3'		For cloning into VR1020 (5' BamIH restriction enzyme site)
T7_R	5'-GCTGTAATTTGGGCATTTTCTTG-3'	74	Internal *oppA* primers, position 592-615
P100_1F	5'-CATTAATTTAGCTTTATTAAAAGAT-3'	64	Internal *oppA* primers, position 549-573
P100_1R	5'-GAAACAAATGAAATTGAAACAGTAT-3'	68	Internal *oppA* primers, position 1 220-1 244
P100_2F	5'-GACAACACTGTAAGTTTTGGAAATC-3'	70	Internal *oppA* primers, position 1 177-1 201
P100_2R	5'-ATTTTTCTGGTTTTAATAAGTCATC-3'	67	Internal *oppA* primers, position 1 906-1 930
P100_3F	5'-CCCTAGAGCAAACTATGGAAATAAA-3'	72	Internal *oppA* primers, position 1 851-1 875
P100_3R	5'-CAGAAGAAGGTATTTACTAATGTGT-3'	65	Internal *oppA* primers, position 2 525-2 549
P100_4F	5'-AGAGATGCTGTAAATAAAGATCCTG-3'	69	Internal *oppA* primers, position 2 461-2 485
P100_4R	5'-TGTTCAGTAGTTGGTTTAACAAAGT-3'	68	Internal *oppA* primers, position 3 185-3 209
SP6_F	5'-GCTTTAGATTTAGTAATTGCTGCTT-3'	70	Internal *oppA* primers, position 3 130-3 154
pGEM-T Easy_T7	5'-TAATACGACTCACTATAGGG-3'		pGEM®-T Easy vector
pGEM-T Easy_SP6	5'-ATTTAGGTGACACTATAGAA-3'		
pGEX_F	5'-GGGCTGGCAAGCCACGTTTGGTG-3'		pGEX-4T-1 vector
pGEX_R	5'-CCGGGAGCTGCATGTGTCAGAGG-3'		
pCI-neo_T7EEV	5'-AAGGCTAGAGTACTTAATACGA-3'		pCI-neo vector
pCI-neo_T3	5'-AATTAACCCTCACTAAAGGG-3'		

Table 5.1 (Continued).

Primer name	Sequence	Anneal. Temp	Comment
VR1012_F	5'-CGCGCCACCAGACATAATAG-3'		VR1012 vector
VR1012_R	5'-AACAACAGATGGCTGGCAAC-3'		
VR1020_F	5'-CGTCGACAGAGCTGAGATCCTACAG-3'		VR1020 vector
VR1020_R	5'-GACACCTACTCAGACAATGCGATGC-3'		
Site1&2_F	5'-ACGCGT**ATG**AAAAAATG**G**TG**G**TTATTACCAGTAGCTAGTACA-3'	66	SDM reaction 1, mutate positions 12 &15
Site1&2_R	5'-TGTACTAGCTACTGGTAATAACC**A**CC**A**TTTTTT**CAT**ACGCGT-3'		
Site10_F	5'-TTTTAACAGCAGCTATTGACTG**G**AACTCAATCGCTTCAATT-3'		SDM reaction 1, mutate position 2 727
Site10_R	5'-AATTGAAGCGATTGAGTTCC**A**GTCAATAGCTGCTGTTAAAA-3'		
Site9_F	5'-TGATGAATATGCTTACACAATGTG**G**GGAATGTCAGCAGCAGA-3'	68	SDM reaction 2, mutate position 2 626
Site9_R	5'-TCTGCTGCTGACATTCCCC**A**CATTGTGTAAGCATATTCATCA-3'		
Site13&14_F	5'-AACCAACTACTGAACAATG**G**AGAAATTACTG**G**ACAGGAACTTC TCCATT-3'		SDM reaction 2, mutate positions 3 213 & 3 225
Site13&14_R	5'-AATGGAGAAGTTCCTGTCC**A**GTAATTTCTCC**A**TTGTTCAGTAG TTGGTT-3'		
Site3_F	5'-ACTTAAAGCATCTGATAAGTG**G**GAATTAAATGAAAACG-3'	61	SDM reaction 3, mutate position 1 290
Site3_R	5'-CGTTTTCATTTAATTCCC**A**CTTATCAGATGCTTTAAGT-3'		
Site7_F	5'-ACAGTAATTAATAAGAATTACTG**G**GACACAGAATATGTTAAT-3'		SDM reaction 3, mutate position 2 301
Site7_R	5'-ATTAACATATTCTGTGTCCC**A**GTAATTCTTATTAATTACTGT-3'		
Site4_F	5'-GAGTTAGACCAGGTCATTTCTG**G**ACCGATGCTAAAGGAAA-3'	65	SDM reaction 4, mutate position 1 680
Site4_R	5'-TTTCCTTTAGCATCGGTCC**A**GAAATGACCTGGTCTAACTC-3'		
Site12_F	5'-TTAGATTTAGTAATTGCTGCTTG**G**AACGGATTAGATCCAAGA-3'		SDM reaction 4, mutate position 3 156
Site12_R	5'-TCTTGGATCTAATCCGTTCC**A**AGCAGCAATTACTAAATCTAA-3'		
Site5_F	5'-GATTATTAAGAACACAAATGTG**G**GACACACCTTATAGGCTA-3'	64	SDM reaction 5, mutate position 1 762
Site5_R	5'-TAGCCTATAAGGTGTGTCCC**A**CATTTGTGTTCTTAATAATC-3'		
Site8_F	5'-CACATTAGTAAATACCTTCTTCTG**G**TCAATTCTGCCAAAAGA-3'		SDM reaction 5, mutate position 2 550
Site8_R	5'-TCTTTTGGCAGAATTGACC**A**GAAGAAGGTATTTACTAATGTG-3'		
Site6_F	5'-CTAAATTGAGTGGTATTTACTG**G**TATGGACTTTCAGTTGAT-3'	62	SDM reaction 6, mutate position 2 196
Site6_R	5'-ATCAACTGAAAGTCCATACC**A**GTAAATACCACTCAATTTAG-3'		
Site16_F	□□-TGTATTTAGTTCAAGATTCTG**G**CTAAACTATACAACATCAC-3'		SDM reaction 6, mutate position 3 609
Site16_R	5'-GTGATGTTGTATAGTTTAGCC**A**GAATCTTGAACTAAATACA-3'		
Site11_F	5'-CCTCAACCGGTTAAACCTTG**G**ATTACAGGATTGTCTCCTGAC-3'	68	SDM reaction 7, mutate position 2 775
Site11_R	5'-GTCAGGAGACAATCCTGTAATCC**A**AGGTTTAACCGGTTGAGG-3'		
Site15_F	5'-TCCATTTTCACTAGCAGGTTG**G**GGTTATGACTACGATGGTAT-3'		SDM reaction 7, mutate position 3 258
Site15_R	5'-ATACCATCGTAGTCATAACCCC**A**ACCTGCTAGTGAAAATGGA-3'		

*Restriction enzyme sites, included for cloning purposes, are underlined, the start and termination codon are bold and the nucleotides change by site-directed mutagenesis are bold and underlined.

White colonies were screened with colony PCR to confirm the presence of the insert using the pGEM–T Easy_T7 and P100_2R primers (Table 5.1). Each colony PCR reaction contained 1 × reaction buffer, 0.2 mM of each dNTP (Kapa Biosystems, South Africa), 2 mM $MgCl_2$, 1 pmol/ml of each primer and 0.2 units of Super-Therm Taq DNA polymerase (JMR Holdings, USA) in a final volume of 10 µl. Bacteria from a single colony were transferred with a sterilized toothpick into the PCR mixture. PCR conditions were as follows: 25 cycles of 94°C for 30 sec, 55°C for 15 sec and 72°C for 1 min, followed by a final step at 72°C for 6 min in a Veriti 96 well Thermal Cycler (Applied Biosystems, USA). Positive colonies were inoculated into 5 ml LB medium containing 100 µg/ml ampicillin and grown at 37°C for 16 h, shaking at 250 rpm. Plasmid DNA (pDNA) was isolated from

2 ml culture with the Invisorb® spin plasmid mini two kit (Invitek GmbH, Germany), according to the manufacturer's instruction. All pDNA samples were stored at 4°C. Freezer stocks were prepared from left over culture by diluting 1:1 with 80% glycerol and stored at -80°C.

Accurate insertion of the gene into the vector was confirmed by sequencing. Each sequencing reaction contained 1 µl BigDye® Terminator mix (v3.1, Applied Biosystems, USA), 3 µl Halfdye (Bioline, UK), 300 ng pDNA, 1 pmol primer (Table 5.1) and Milli-Q® water to a final volume of 10 µl. PCR conditions were as follows: 35 cycles of 94°C for 10 sec, 55°C for 30 sec and 60°C for 4 min, followed by a final step at 60°C for 10 min. The Central Analytical Facility (CAF) of the University of Stellenbosch, South Africa analysed the products with an ABI® 3100 Genetic Analyser (Applied Biosystems, USA). The resulting sequences were edited and aligned against the vector and the gene sequence of Ms03 *oppA* using BioEdit v7.0.5.2 (Hall 1999).

Seven consecutive SDM steps were carried out to change all 16 of TGA codons to the universal codon for tryptophan i.e. TGG. The positions of the nucleotides mutated within the *oppA* gene were 12, 15, 1 290, 1 680, 1 762, 2 196, 2 301, 2 550, 2 626, 2 727, 2 775, 3 156, 3 213, 3 225, 3 258 and 3 609. A strategy of using two primer pairs that bound to the *oppA* gene at distant positions and possessed very similar annealing temperatures was followed, to allow combinations of primers to be used in order to reduce the number of SDM steps to seven in total. In each step, the pGEM_*oppA* vector was amplified with two sets of SDM primers (Table 5.1) with Kapa HiFi DNA polymerase according to the manufacturer's instructions using 2.25 mM $MgCl_2$. The SDM PCR products were treated with DpnI (Promega, USA) according to the manufacturer's instructions and the enzymes removed with the DNA clean and concentrator™-5 kit (ZYMO Research, USA) before transforming the plasmids into *E. coli* JM109 cells. Colony PCR was used to select positive colonies and the plasmid was subsequently isolated and the insert sequenced as before to confirm the success of each SDM step.

Expression and purification of recombinant OppA protein

Recombinant OppA protein was produced for use as coating antigen in an enzyme-linked immunosorbent assay (ELISA). For expression, the SDM *oppA* gene was sub-cloned from SDM pGEM_*oppA* vector into the pGEX-4T-1 vector (GE Healthcare Life Science, UK). The mutated *oppA* gene was first amplified by PCR from the pGEM_*oppA* vector using the Ms03_P100BamHI_F and Ms03_P100NotI_R2 primers with BamHI and NotI restriction sites (Table 5.1). The PCR product was next sub-cloned into a pGEM®-T Easy vector and the pGEM plasmid containing *oppA* and the BamHI and NotI restriction sites (pGEM_$oppA_{BamNot}$) isolated as describe above. Double digests of pGEM_$oppA_{BamNot}$ and

the pGEX-4T-1 vector were performed with BamHI and NotI (FastDigest, Thermo Scientific, USA) and the digested products purified with the Zymo-DNA clean and concentrator™-5 kit (Zymo Research, USA). The BamHI/NotI digested linear pGEX-4T-1 product was further treated with shrimp alkaline phosphatase (SAP, Promega, USA) according to the manufacturer's instructions and purified as before. Ligation was performed overnight at 4°C with a vector to insert ration of 1:1 and 1:3 using T4 DNA ligase (Promega, USA) before transforming the plasmids into *E. coli* JM109 cells. A control was included which contained digested pGEX-4T-1 without insert. Positive colonies were identified using colony PCR with the pGEX_F and P100_2R primers (Table 5.1) as described above followed by plasmid sequencing as before to confirm accurate insertion.

The pGEX-4T-1_*oppA* plasmid was transformed into *E. coli* BL21(DE3)pLysS cells. Expression of the glutathione S-transferase (GST) OppA fusion protein was induced at an OD_{600} of 0.3 by the addition of 0.4 mM IPTG and cells harvested after 6 h. Terrific broth with 100 µg/ml ampicillin, 1% (w/v) glucose and 34 µg/ml chloramphenicol was used for expression experiments. The GST-OppA protein was isolated using glutathione-agarose chromatography (Sigma-Aldrich, USA) within a gravity flow column system at 4°C according to the manufacturer's instructions. An extra 50 ml phosphate-buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4 and 1.5 mM KH_2PO_4, pH 7.2) wash step was added before bound GST-OppA protein was eluted with 10 mM reduced L-glutathione in Tris-HCl buffer (50 mM , pH 9.5). Fractions (1 ml) were collected and their protein concentration determined using a Bradford assay. In brief, 5 µl the protein fractions or BSA standards (concentrations in the range of 0 to 1.75 mg/ml) were 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) and incubated for 5 min at room temperature within 96-well microtitre plate. Absorbance was measured at 620 nm. A standard curve using BSA was generated in order to calculate the protein concentrations of the fractions. Expression and isolation products were analysed using SDS-PAGE and Coomassie blue staining.

For western blot analysis the proteins were transferred to a 0.45 µm nitrocellulose membrane (Schleicher and Schuel, Sigma-Aldrich) by electrophoresis (20 mA for 16 h), followed by three 1 h incubation steps at 37°C: blocking with casein buffer (10 mM Tris-HCl, 150 mM NaCl, 0.5% casein and 0.02% Thiomersal), 1:12 000 dilution of primary antibody (goat anti-GST, GE Healthcare Life Science, UK) and a 1:10 000 dilution of secondary antibody (rabbit anti-goat IgG (whole molecule) peroxidase conjugate, Sigma-Aldrich, USA). Antibodies were diluted with casein-Tween buffer (0.5% casein buffer containing 0.1% (v/v) Tween 20). After each binding step the membrane was washed

three times with PBS-Tween 20 (0.1%). The membrane was developed for 20 min in a solution containing 0.05% (w/v) 4-chloro-1-naphthol, 16.6% (v/v) methanol, 114 mM NaCl, 2.25 mM KCl, 8.3 mM Na_2HPO_4 and 1.25 mM KH_2PO_4 and 0.025 µl/ml H_2O_2 (30%). The colour reaction was stopped by washing the membrane in Milli-Q® water.

Preparation of DNA vaccines

The SDM *oppA* gene was sub-cloned into pCI-neo (Promega, USA), VR1012 and VR1020 (Vical Inc., USA). For sub-cloning into pCI-neo single consecutive digests of the pCI-neo and the SDM pGEM_*oppA* vectors were performed with AccI and MluI (FastDigest, Thermo Scientific, USA). AccI and MluI sites were previously added to amplification primers (Table 5.1) Digested products were purified and pCI-neo further treated with SAP followed by ligation and transformation as described above. A control was included which contained pCI-neo without insert. Positive colonies were identified using T7EEV and P100_2R primers (Table 5.1) by colony PCR and the insert sequenced as described above to confirm accurate insertion.

For sub-cloning into VR1012 and VR1020, the *oppA* gene was PCR amplified from the SDM pGEM_*oppA* vector. For cloning into VR1012, SalI and NotI restriction enzyme sites were included in the forward and reverse primers, respectively and for cloning into VR1020, BamIH restriction enzyme sites were included in both the forward and reverse primers (Table 5.1). The PCR product with appropriate restriction sites was cloned into a pGEM®-T Easy vector as describe above. Plasmid DNA was isolated (as above) and a double digest was performed with SalI and NotI for VR1012 and a single digest with BamHI for VR1020. Purification, SAP treatment, ligation and transformation were performed as described above except that a different antibiotic was used (kanamycin, 50 µg/ml). Positive colonies were identified in VR1012 using VR1012_F and P100_2R primers and in VR1020 using VR1020_F and P100_2R primers (Table 5.1) followed by sequencing of the insert as described above to confirm accurate insertion.

For large scale production of the pCI-neo_*oppA* vaccine, five large scale cultures were used, while four large scale cultures were used for both the VR1012_*oppA* and VR1020_*oppA* vaccines. For each large scale culture, 450 ml of LB medium was inoculated with *E. coli* JM109 cells that contain the respective vaccine plasmids and cultivated for 16 h (A_{600} between 0.650 and 1.150). The DNA vaccine plasmids (pCI-neo_*oppA*, VR1012_*oppA* and VR1020_*oppA*) were purified with an Endotoxin-free plasmid DNA purification kit (NucleoBond® Xtra Maxi plus EF, Macherey-Nagel, Germany). The yields were determined using a Nanodrop (Thermo Scientific, USA) while the integrity of the plasmid DNA was confirmed with electrophoresis on a 1% agarose

gel. The DNA vaccines were diluted to final concentrations of 100 µg/ml in sterile PBS buffer for use as vaccines.

DNA vaccine trial

A vaccine trial on ostriches in the Oudtshoorn district (Western Cape) in South Africa in which of all three vaccines were to be tested was planned. Ethical clearance to perform this trial was obtained from the University of Stellenbosch Animal Ethics Committee (Ref: 10NB_BOT01) prior to the start of the trial. Three-month-old ostriches used for the trial were from, and housed on, a commercial ostrich farm near Oudtshoorn in the Klein Karoo region. Trial ostriches were not kept in isolation, but were housed and treated in the same manner as non-trial ostriches on the farm, receiving food and water *ad libitum*. Each ostrich was tagged with a unique number according to standard ostrich farming practices. The trial was therefore conducted under similar conditions to which a commercial vaccine would be administered.

The vaccine trial consisted of three experimental groups. Group 1 (25 ostriches) was vaccinated with pCI-neo_*oppA*, group 2 (21 ostriches) with VR1012_*oppA*, group 3 (21 ostriches) with VR1020_*oppA* and the control group (23 ostriches) received no vaccination. Ostriches were vaccinated intramuscularly in the upper thigh with a single dose (100 µg in 1 ml PBS) at week 0. Blood (4 ml) was drawn from the jugular vein in Vacuette® Z serum sep clot activator tubes with 18G x 1” needles (Vacuette, UK). Serum was separated by centrifugation at low speed for 20 min and transferred to a 1.5 ml microcentrifuge tube before storage at -20°C. Blood samples were collected and the weight of each ostrich recorded before vaccination at week 0 and after vaccination at week 3 (The reader is referred to page 101 where the implications of an outbreak of avian influenza during this trial are outlined).

Monitoring of existing mycoplasma infections

Saliva samples were used to monitor existing mycoplasma infections in the trial ostriches in order to interpret immune responses after vaccination, as existing infections may have influenced the results. Saliva samples were collected before vaccination at week 0 and at week 3 after vaccination. Saliva samples were collected from all ostriches by swabbing the trachea with a sterile transport swab (plain rayon tipped with plastic applicator, Copan, Italy) after which the swab was again placed in its sterile polypropylene tube. In the laboratory, each swab was rinsed in a microcentrifuge tube containing 200 µl PBS buffer followed by PCR testing with species-specific primers for the three ostrich-infecting mycoplasmas, *M. struthionis* sp. nov. str. Ms01 (Ms01), *Mycoplasma* sp. Ms02 (Ms02)

and Ms03 using Super-Therm Taq DNA polymerase (JMR Holdings, USA) as described by Botes et al. (2005a).

Evaluation of the immune response

Microtiter plates (Maxisorp, Nunc) were coated overnight at 4°C with 100 µl of recombinant GST-OppA protein diluted to 10 µg/ml in carbonate buffer (50 mM, pH 9.6). Non-specific binding was prevented by blocking with 300 µl/well of 0.5% casein buffer for 2 h at 37°C after which the casein buffer was decanted. Serum samples were diluted 1:100 with casein-Tween buffer (0.5% casein buffer containing 0.1% (v/v) Tween 20) and 100 µl/well loaded in triplicate before incubation for 1 h at 37°C. The plate was decanted and washed eight times with PBS-Tween (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4 and 1.5 mM KH_2PO_4, pH 7.2, 0.1% (v/v) Tween 20), before 100 µl/well of biotinylated rabbit anti-ostrich IgG antibodies (Blignaut et al. 2000) (diluted 1:100 with casein-Tween buffer) was added followed by incubation for 1 h at 37°C. After decantation and washing (as above), a 100 µl/well streptavidin horseradish peroxidase (HRP) conjugate mixture (2 ml Streptavidin (Invitrogen), 38 ml 0.5% casein buffer and 40 ml (50%) glycerol) diluted 1:100 with casein-Tween buffer was added and incubated for 1 h at 37°C. After decantation and washing, 100 µl of substrate solution (0.5 mg/ml ABTS, 0.5 µl/ml H_2O_2, 0.1 M citrate buffer, pH 5) was added. Absorbances were read at 405 nm after 30 min incubation at 37°C.

Serum collected from ostrich 7736 (VR1020_*oppA* group) at week 0 and 3 was used as internal control and column blanks (containing all components except ostrich serum) were used to standardize the results.

Statistical analysis of data

Statistical analysis of the data was performed with the Agrobase Generation II® (Agronomix Software Inc.) software. ANOVA and least significant differences (LDS) between the groups were calculated.

Results

Mycoplasma cultivation and DNA isolation

Ms03 cultures were successfully cultivated by Mr J.J. Gouws (Faculty of Veterinary Science, Onderstepoort, University of Pretoria). In medium, Ms03 grew slowly and did not reach high cell densities. Genomic DNA was successfully isolated from the cultured Ms03. The concentration of the gDNA samples range from 755 to 1022 ng/µl and the 260/280 ratio range from 2.00 to 2.05. The higher 260/280 ratios (above 2) are not surprising since absorbance is depend on the nucleotide composition of the DNA sample.

Mycoplasma genomes are A+T rich and the nucleotide, adenine has a high 260/280 ratio.

Cloning, site-directed mutagenesis, expression and purification of OppA

Subsequently, the gDNA was used as template for the amplification of the *oppA* gene. The *oppA* gene was successfully cloned into the pGEM®-T Easy vector which was confirmed by sequencing (results shown in Appendix 4). The 16 TGA codons were mutated successfully to TGG in seven consecutive steps of site-directed mutagenesis as confirmed by sequencing. The results of each consecutive step are shown in Appendix 4. This therefore gave an *oppA* gene that was suitable for transcription according to the universal code, cloned into the pGEM®-T Easy vector.

Restriction enzyme sites (BamHI and NotI) were added to mutated *oppA* gene by PCR amplification before it was again cloned into the pGEM®-T Easy vector. The success of the resulting pGEM_$oppA_{BamNot}$ vector was confirmed by sequencing (Appendix 4). Restriction digestion and ligation was used to sub-clone the $oppA_{BamNot}$ gene into the prokaryotic pGEX-T4-1 expression vector as confirmed by sequencing (results shown in Appendix 4).

SDS-PAGE analysis showed that the recombinant OppA protein was expressed successfully as an N-terminal GST-fusion protein of the predicted size of about 170 kDa (26 kDa due to the GST tag), Figure 5.1 A, lane 2. The isolation of the GST-OppA fusion protein was achieved by affinity purification as shown in Figure 5.1 A, lane 3. This fraction was shown to contain free GST and GST-OppA fusion protein (Figure 5.1 A, lane 3). This could have resulted from enzymatic cleavage during isolation and storage (Braun et al. 2002). Success of the isolation was confirmed by western blot analysis using anti-GST antibodies (Figure 5.1 B). The concentration of the isolated recombinant OppA protein was determined by Bradford analysis using BSA as a standard. An additional PBS wash step was introduced during isolation due to interference of Triton-X in the first wash step with the Bradford analysis. Protein was eluted between fraction 7 and 12 with the highest concentration obtained in the 9th fraction (Figure 5.2).

Preparation of DNA vaccines

The mutated *oppA* gene was successfully sub-cloned into pCI-neo with the use of the restriction enzyme sites (MluI and AccI) previously added. The results were confirmed by sequencing as shown in Appendix 4.

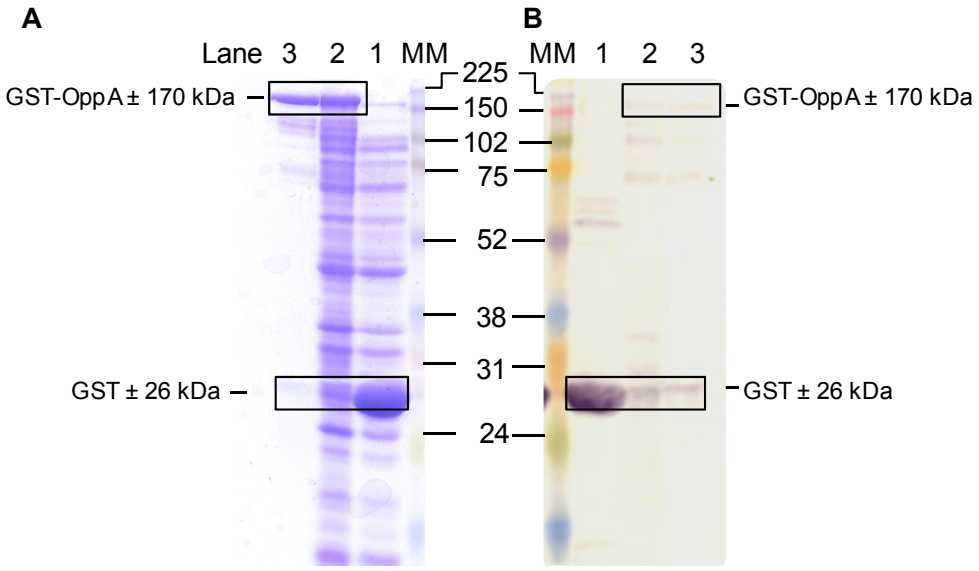


Figure 5.1 Expression of recombinant GST-OppA protein. SDS-PAGE (A) and western blot (B) analysis of the expressed OppA protein. Lane 1 *E. coli* BL21(DE3)pLysS cells expressing GST-control, lane 2 *E. coli* BL21(DE3)pLysS cells expressing GST-OppA protein and lane 3 isolated recombinant GST-OppA protein. MM is the GE Healthcare full-range rainbow molecular weight marker, molecular sizes indicated in kDa.

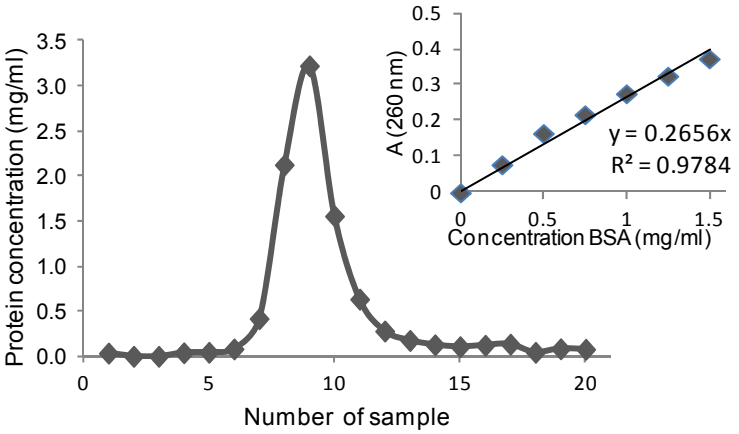


Figure 5.2 Protein isolation profile of the recombinant GST-OppA. Inserted in top right-hand corner is the BSA standard curve used to calculate protein concentrations.

For sub-cloning into VR1012, restriction enzyme sites (SalI and NotI) were added to the mutated *oppA* gene with PCR amplification before it was successfully cloned into the pGEM®-T Easy vector. The success of the resulting pGEM_*oppA*$_{SalNot}$ vector was confirmed by sequencing (Appendix 4). The mutated *oppA*$_{SalNot}$ was sub-cloned into VR1012 with success as confirmed by sequencing (Appendix 4).

For sub-cloning into VR1020, PCR amplification was used to add restriction enzyme sites to the mutated *oppA* gene before it was cloned into the pGEM®-T Easy vector. The success of the resulting pGEM_*oppA*$_{BamBam}$ vector was confirmed by sequencing

(Appendix 4). The mutated *oppA*$_{BamBam}$ gene was sub-cloned into VR1020 as confirmed by sequencing (Appendix 4).

Large scale production of the DNA vaccines was achieved. In total 3522 µg of pCI-neo_*oppA* plasmid DNA was isolated from 2 250 ml of medium, 6303 µg of VR1012_*oppA* plasmid DNA was isolated from 1 800 ml of medium and 4572 µg of VR1020_*oppA* plasmid DNA was isolated from 1 800 ml of medium. The 260/280 ratio of the isolated plasmids ranged from 1.89 to 1.95. These were used to dilute the three vaccines to the required concentration of 100 µg/ml for vaccination.

DNA vaccine trial

Ostriches were successfully vaccinated with the prepared DNA vaccines. No adverse reactions were observed at the injection sites. Blood and saliva samples were successfully collected. It was observed that within minutes after vaccination and sample collection (saliva and blood), the ostriches resumed normal behaviour such as eating, walking around and exploring, which indicates that the procedure did not have an immediate effect on the birds' behaviour. Later adverse behaviour and inflammation at the vaccination sites was also not observed.

Monitoring of existing mycoplasma infections

The results of the mycoplasma infections that were monitored with PCR during the trial are presented in Table 5.2. At week 0, 43 of the 90 birds (48%) had mycoplasma infections (Table 5.2). Of the ostriches, 34 (38%) were infected with Ms03 and 33 (36%) with Ms02. This includes 24 ostriches that were not only infected with a single species, but had dual infections of Ms02 and Ms03. No Ms01 infections could be detected. These infections were spread over the different vaccine and control groups. At week 3, the total number of infections were 38 (42%) with 21 of the 90 ostriches (23%) being infected with Ms03, 25 (28%) with Ms02 and 3 ostriches infected with Ms01. Once again multiple infections were present in some birds, with seven ostriches being infected with Ms02 and Ms03 infections and two ostriches were infected with Ms01, Ms02 and Ms03. Infections at week 3 were spread over all groups, but the three Ms01 infections were only found in the VR1012_*oppA* group (Appendix 5, Supplementary Table 5.1).

Overall the number of infections decreased from week 0 to week 3 within the control and VR1012_*oppA* groups but increased for pCI-neo_*oppA* and VR1020_*oppA* groups. These changes in infection status could not be related to the vaccines since the VR1012_*oppA* and control group received a non-scheduled treatment with Terramycin four days before the week 3 sampling point. This was prompted by lack of weight increase of ostriches in the control group and a weight decrease in the VR1012_*oppA*

group (Table 5.3) with an overall poor condition in both of these groups that could have been due to an underlying infection other than just mycoplasma. Terramycin is commonly used under field conditions to treat birds with physical signs of infection and thereby limit possible disease progression. It is a long-acting tetracycline with a broad-spectrum antibiotic activity, which also has some activity against mycoplasma.

Table 5.2 Summary of the mycoplasma infections during the vaccine trial as determined by PCR

	Ms01		**Ms02**		**Ms03**		**Total number of infected birds***		**Number of birds with infections in week 0 and/or week 3***
Week	0	3	0	3	0	3	0	3	
Control	0/23	0/23	15/23	5/23	19/23	1/23	21/23$^{\Delta}$	6/23	21/23
pCI-neo_*oppA*	0/25	0/25	1/25	7/25	1/25	13/25	2/25	16/25$^{\#}$	16/25
VR1012_*oppA*	0/21	3/21	16/21	9/21	13/21	4/21	18/21$^{\blacklozenge}$	10/21$^{\Omega}$	18/21
VR1020_*oppA*	0/21	0/21	1/21	4/21	1/21	3/21	2/21	6/21$^{¥}$	7/21
Total	0/90	3/90	33/90	25/90	34/90	21/90	43/90	38/90	62/90

*Since ostriches were infected with more than one *Mycoplasma* species, total numbers of infected birds are not the sum of the individual infections.
$^{\Delta}$ 13 dual infections of Ms02 and Ms03 in the control group in week 0
$^{\#}$ 4 dual infections of Ms02 and Ms03 in the pCI-neo_*oppA* group in week 3
$^{\blacklozenge}$ 11 dual infections of Ms02 and Ms03 in the VR1012_*oppA* group in week 0
$^{\Omega}$ 2 dual infections of Ms02 and Ms03 as well as 2 infected with Ms01, Ms02 and Ms03 in the VR1012_*oppA* group in week 3
$^{¥}$ 1 dual infections of Ms02 and Ms03 in the VR1020_*oppA* group in week 3

Evaluation of the immune response

The ELISA was successfully optimized with regard to the coating concentration of the recombinant OppA protein as well as the serum dilution. It was found that when serum samples of ostrich 7736 were used they gave consistent absorbance values of about 0.750 (week 0) and 1.750 (week 3) repeatedly. Consequently these serum samples were included on each plate as internal controls to monitor plate-to-plate variation. Plate-to-plate variation was within 20% of the mean. The number of freeze-thaw cycles of the 7736 serum sample may have increased the variation. The negative controls gave consistent low absorbance values.

The results of the vaccination trial are shown in Figure 5.3. Mean titre values are shown for week 0 (before vaccination) and week 3 (after vaccination). At week 0, the mean ELISA titre values (405 nm) for the control and vaccinated groups ranged between 0.938 and 1.142, but these differences were not statistically significant. The titre value for the control group was maintained at week 3 with a mean of 0.958. High absorbance values for all of the groups at the start of the trial and the control group at week 3 may be a result of mycoplasma infections occurring before and during the trial. This may be due to possible existing antibodies against OppA in the serum or due to increased background

values of the ELISA. The ANOVA analysis showed a significant treatment x time interaction (P=0.0428) between the vaccinated groups. However using the calculated least significant difference (LSD, P≤ 0.05) as basis for determining statistically significant differences, only the pCI-neo_*oppA* and VR1020_*oppA* treatment groups differed significantly from that of the control group. This implied that the ostriches were responding to vaccination. The OppA protein was therefore expressed *in vivo* and was sufficiently immunogenic to induce a primary antibody response.

The weight of the ostriches was monitored during the vaccine trial as an indication of the birds' health. An ANOVA analysis was performed for the recorded ostrich weights at week 0 and week 3. The average weight per group remained constant or increased slightly from week 0 to week 3 (except for VR1012_*oppA* which showed a slight decrease). No statistically significant difference over time or treatment x time interaction was found. A significant difference was found between treatments (P=0.0000) with a LSD value of 2.713, but this difference was not significant over time as it was already present at the onset of the trial due to the variation in weight between groups (Appendix 5, Supplementary Table 5.2). At the start of the trial the average weight per group ranged from 28.4 to 38.9 kg (Table 5.3) with the VR1012_*oppA* and VR1020_*oppA* groups not differing significantly. All weights were, however, in the expected range for the age group used since the weight of three-month-old ostriches can range from 4 to 44 kg (Bunter and Cloete 2004). The lack of an increase in weight between groups during the trial could therefore not be ascribed to the DNA vaccine administered as the unvaccinated control group also did not gain weight.

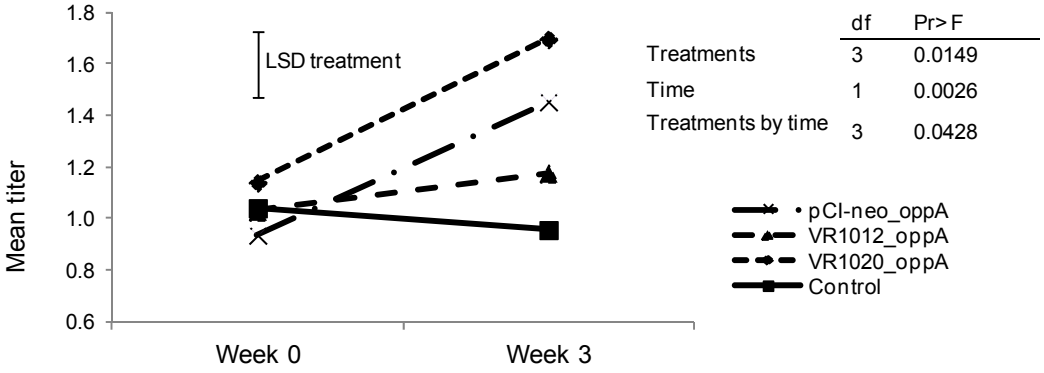


Figure 5.3 Anti-OppA immune response of ostriches vaccinated with three different DNA vaccine vectors containing the mutated *oppA* gene of *M. nasistruthionis* sp. nov. str. Ms03. Vaccinated ostriches received a single DNA vaccine dose (100 µg/ml) at week 0 and the control group did not receive any vaccine. Statistical parameters are indicated in the top right-hand corner. Additional ANOVA statistical data are available in Appendix 5, Supplementary Table 5.3.

Table 5.3 Summary of the mean weight recorded per vaccine group during the vaccine trial

	Mean weight	
	Week 0	**Week 3**
Control	28.4	28.6
pCI-neo_*oppA*	38.9	39.9
VR1012_*oppA*	35.1	34.9
VR1020_*oppA*	33.6	36.8

Discussion

In this study, DNA vaccines were developed using the Ms03 *oppA* gene as vaccine candidate gene. The codon optimization was required due to the use of the universal stop codon TGA as a tryptophan codon in mycoplasmas. The Ms03 *oppA* gene contains 16 TGA codons. A strategy of combining two primer pairs to amplify the vector and *oppA* gene during SDM PCR was successfully employed to reduce the number of consecutive steps from 14 to 7 and to thereby correct all 16 TGA codons within *oppA* to TGG codons.

Three different eukaryotic expression vectors (pCI-neo, VR1012 and VR1020) were used to develop the DNA vaccines within this study. They were selected based on DNA vaccine studies in other birds (Lee et al. 2003; McCutchan et al. 2004; Klotz et al. 2007) and on special characteristics such as a TPA-signal peptide in VR1020 that should result in the export of the translated protein from the cell and thus better activation of the immune system. The *oppA* gene was cloned into each vector using different restriction enzymes to ensure that the *oppA* gene is in-frame for *in vivo* protein expression. Large scale production of the constructs (pCI-neo_*oppA*, VR1012_*oppA* and VR1020_*oppA*) was successfully achieved, indicating that scaling up for commercial production should be possible.

The humoral immune response elicited by the ostriches against these DNA vaccines was evaluated with an ELISA that was developed to using recombinant OppA protein as coating antigen. ELISA titre values were already raised at the start of the vaccination trial. The same observation was made by Yang et al. (2011) when using recombinant OppA protein as subunit vaccine against *M. catarrhalis* in mice. They suggested that the high titres were due to background and not recombinant OppA reacting to native OppA antibodies. Negative controls in our ELISA (serum replaced with casein buffer containing Tween 20) were used to test for nonspecific background and gave consistent low absorbance values (0.078-0.150). The raised titre values can therefore not be ascribed to non-specific binding.

Yang et al. (2011) also postulated the possible presence of existing systemic antibodies to OppA as a result of existing infections. In this study, a large proportion of the ostriches

in all groups had existing mycoplasma infections at the start of the trial, which could account for the presence of anti-OppA antibodies. There were, however, ostriches that tested negative for mycoplasma with high titre values.

Possible existing antibodies did, however, not mask the detection of an immune response to the OppA during vaccination. Vaccination resulted in a statistical significant increase in ELISA titre values for both the VR1020_*oppA* and pCI-neo_*oppA* vaccines with the VR1020_*oppA* construct resulting in the largest increase from week 0 to week 3. No significant increase was observed for VR1012_*oppA*. The DNA vaccine vectors (VR1020_*oppA* and pCI-neo_*oppA*) were therefore able to express the OppA protein *in vivo*, which in turn was sufficiently immunogenic to induce an anti-OppA immune response.

Despite the fact that all three the vaccine vectors contain a CMV promoter, which is known to enhance the expression of genes, this may not have occurred in all three vaccines as an increase in anti-OppA antibodies was not observed in the VR1012_*oppA* vaccinated birds. The CMV promoter of VR1012 is optimized for use in mice and could therefore possibly not function in ostriches (Suarez and Schultz-Cherry 2000). However the main difference between the VR1012 and VR1020 vectors is the presence of the TPA-signal peptide within VR1020 vector. The export of the expressed OppA protein may therefore explain the significant antibody response following vaccination with the VR1020_*oppA* vaccine compared to the VR1012_*oppA* vaccine.

An increase or decrease of PCR-detected infections did not correlate with immune responses. Ostriches typically produce a maximum primary immune response upon vaccination after 21 days, irrespective of the vaccine used (Blignaut et al. 2000; Bonato et al. 2009). The PCR was used to detect the presence or absents of mycoplasmas in the trial ostriches, but the technique does not quantify the bacterial load within the ostriches. Although a decrease in mycoplasma infections was not observed within the duration of this vaccine trial, it does not exclude the possibility that the bacterial load within the infected ostriches might have decreased. Thus the DNA vaccines may prove to be effective in eliminating existing mycoplasma infections given a longer duration, optimum dose or a second booster immunization. In future studies, the use of a quantitative technique such as real-time PCR to determine the bacterial load during vaccination trials may be advantageous.

Although not statistically significant, the weight of the pCI-neo_*oppA* and VR1020_*oppA* vaccine groups increased while that of VR1012_*oppA* group decreased and the control group showed no change. Growth rate of birds is affected by many factors such as stress, nutrition, temperature fluctuations and diseases. Before the age of three months,

chicks were kept in small camps where chick numbers were low and they receive stimuli to encourage eating. At three months of age they were moved in to larger camps where there were a larger number of birds. In the new camps food ration as well as food and water bowls were different. The changes in social dynamic along with their housing environment result in birds being stressed and not eating sufficiently which can impact on growth. Ostriches are known to be severely affected by stress (Hoffman and Lambrechts 2011) and during this adaption period retardation in weight gain and even weight loss in some case is therefore normal among farm ostriches. Furthermore stress is known to depress immune responses. This indicates that, in future trials, serious attention should be given to reducing stress. Ostriches have only been farmed intensively since about 1860 and this may indicated that stress may play a major role in reducing immune responses in ostriches in general.

Due to an outbreak of avian influenza, it was not possible to evaluate a secondary immune response during this study (the reader is referred to page 101 where the implications of an outbreak of avian influenza during this trial are outlined). This would have indicated whether or not the vaccinations had lead to immune memory. Further trials are now required to establish the optimal doses for immune memory and whether or not these vaccines will lead to protection against mycoplasma infections.

In conclusion, this study developed three DNA vaccines using the *oppA* gene as vaccine candidate gene. These DNA vaccines were produced on large scale and used to immunize ostriches. The pCI-neo_*oppA* and VR1020_*oppA* vaccines elicited an immune response in the vaccinated ostriches which implies that the OppA protein was expressed *in vivo.* As antigen, the OppA protein was therefore sufficiently immunogenic to elicit a primary immune response. This is the first study that shown a DNA vaccine was capable of eliciting an immune response in ostriches. This study can therefore be viewed as the first step in the development of a DNA vaccine for the control of mycoplasma infections in ostriches.

Acknowledgements

We would like to thank Vical Inc., USA for supplying the VR1012 and VR1020 vectors for research purposes and Mr JJ Gouws (Faculty of Veterinary Science, Onderstepoort, University of Pretoria) for the culturing of the mycoplasma bacteria. We would like to acknowledge the funding bodies, The Technology and Human Resources for Industry Programme (THRIP) (Project reference TP50090719) and South African Ostrich Business Chamber.

Chapter 6 Conclusion

Vaccines are one of the great achievements of modern medicine. The discovery of DNA vaccines, more than two decades ago, has opened a new economical and safe possibility for the treatment of diseases.

This study set out to explore the genome of *Mycoplasma nasistruthionis* sp. nov. str. Ms03 in terms of its metabolic capacity (Chapter 3) and found that i) genome replication, cell division, RNA transcription, protein translation and glycolysis in Ms03 resembles that of the closely related *Mycoplasma synoviae* 53, ii) purine and pyrimidine metabolism is incomplete and *de novo* synthesis is not possible, iii) amino acid synthesis is mostly absent, with Ms03 only able to convert aspartate to asparagine and glycine to serine and iv) more importers than exporters were annotated owing to the lack of synthesis pathways in Ms03. This is typical for mycoplasmas that have parasitic life styles.

This study was unable to generate a complete genome sequence for Ms03. Considering that 25.2 Mb of 454 sequencing data was used to assemble the Ms03 genome and that the genome size of the genus *Mycoplasma* ranges from 564 to 1 359 kbp, the coverage of the Ms03 genome would be about 25 times which should have been sufficient to generate a complete or near complete genome. The characteristics of the genome (A+T content as well as the occurrence of repeats and runs) and the technology used to sequence the genome (454 pyrosequencing in this case) influenced the assembly outcome. Two approaches to overcome this may be to generate data with longer read lengths (like PacBio) or to generate paired-end data.

The KEGG pathways generated from the Ms03 genome annotation were comparable to that of *M. synoviae* 53 and the annotation mostly completed. The absence of some genes within pathways is, however, questionable since it would render the pathways incomplete. The incomplete genome sequence of the Ms03 genome could have resulted in some ORFs not being predicted by the annotation programs. Additionally ORFs may also have been truncated due to the incomplete genome sequence. Annotation of these truncated proteins would be complicated as some functional motifs may be located on the “missing” part of the sequence. A large number of genes were annotated as “hypothetical protein” implying that it was not possible to predict the function of these proteins due to a lack of evidence. A more comprehensive annotation may assign function to some of these hypothetical proteins and lead to the identification of some of the genes which absence is questionable.

This study sought to identify and characterize the *opp* operon and associated *oppA* gene within the Ms03 genome and found that the Ms03 genome had two *opp* operons,

therefore two *oppA* genes and that each of the identified *opp* genes (A, B, C, D and F) had typical motifs associated with the respective function of the protein products in Opp transporters.

Furthermore the prevalence of *oppA* genes as part of the *oppBCDF* operons within *Mycoplasma* species was evaluated (Chapter 4) and it was confirmed that the substrate-binding domain is present in all species and therefore must play an essential part in oligopeptide transport. All mycoplasmas (except for hemoplasma) had one to three *opp* operons that could be divided into three types (Type A, B and C). Each type had unique InterPro and MEME domains and motifs which together with the phylogenetic analysis suggest a unique role that may be a strategy towards survival under different conditions. Ms03 had a Type A and a Type B *opp* operon, the Type A *oppA* was used as vaccine candidate gene.

The Ms03 Type A *oppA* is a suitable vaccine candidate choice because:

- The importers of the ATP-binding cassette (ABC) family are unique to plants, bacteria and archaea (Berntsson et al. 2010; Rice et al. 2014). This implies that no homologue of the sequence should be found in the ostrich. At the time this study was initiated the genome sequence of the ostrich was not available. It was however recently published and it was confirmed with a tBLASTx search that no homologues of the *oppA* gene were present in the ostrich genome (3).
- OppA is located on the outer surface of the mycoplasma cell membrane (Chapter 4).
- Oligopeptide import is an essential process in Ms03 as deducted from the lack of genes for amino acid synthesis in the Ms03 draft genome (Chapter 3). Additionally transposon mutagenesis studies in *Mycoplasma genitalium* and *Mycoplasma pulmonis* had found the *oppA* gene to be essential (Glass et al. 2006; French et al. 2008). Furthermore the *Mycoplasma hominis* OppA protein not only acts as a substrate-binding protein but also possesses a cytoadherence function and has ecto-ATPase activity (Henrich et al. 1993; Hopfe and Henrich 2004). The *M. hominis opp* operon belongs to Type A.
- The two OppA proteins of Ms03 only share 19.7% amino acid identity (Chapter 3). This along with the phylogenetic and other bioinformatic analyses suggest that each type of OppABCDF transporter should have a unique role (Chapter 4).

Lastly the study aimed to develop and evaluate DNA vaccines containing the Ms03 Type A *oppA* gene for the treatment of Ms03 infections in ostriches and found that DNA vaccines were able to elicit a primary immune response in the ostriches. Due to an

outbreak of avian influenza only two sample points were included in the DNA vaccine trial which was unfortunate. As a result, a second booster vaccination was not possible which could have contributed positively to the vaccine study. However this study shown that, the use of DNA vaccines may be a viable approach for the control of mycoplasma infections in ostriches in future.

In this dissertation, the first Ms03 draft genome and annotation was presented which contributed to our understanding of Ms03 as a miniature genetically independent bacterium. The analysis of the *opp* operons provides insight into the organization and relationships amongst oligopeptide transporters in *Mycoplasma* species. Three DNA vaccines were developed using the *oppA* gene as vaccine candidate gene. This is the first study to show that DNA vaccines are capable of eliciting an immune response in ostriches.

Reference list

Abolnik C, Fehrsen J, Olivier A, van Wyngaardt W, Fosgate G, Ellis C. 2013. Serological investigation of highly pathogenic avian influenza H5N2 in ostriches (*Struthio camelus*). *Avian Pathology* **42**(3): 206.

Agüero F, Al-Lazikani B, Aslett M, Berriman M, Buckner FS, Campbell RK, Carmona S, Carruthers IM, Chan AWE, Chen F. 2008. Genomic-scale prioritization of drug targets: The TDR targets database. *Nature Reviews Drug Discovery* **7**(11): 900.

Alarcon F, de Vasconcelos ATR, Yim L, Zaha A. 2007. Genes involved in cell division in mycoplasmas. *Genetics and Molecular Biology* **30**(1): 174.

Allan E, Wren BW. 2003. Genes to genetic immunization: Identification of bacterial vaccine candidates. *Methods* **31**(3): 193.

Allwright DM, Burger WP, Geyer A, Terblanche AW. 1993. Isolation of an influenza A virus from ostriches (*Struthio camelus*). *Avian Pathology* **22**(1): 59.

Arraes FBM, de Carvalho MJA, Maranhao AQ, Brigido MM, Pedrosa FO, Felipe MSS. 2007. Differential metabolism of *Mycoplasma* species as revealed by their genomes. *Genetics and Molecular Biology* **30**(1): 182.

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M. 2008. The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics* **9**(1): 75.

Babiuk LA, Pontarollo R, Babiuk S, Loehr B, van Drunen Littel-van den Hurk S. 2003. Induction of immune responses by DNA vaccines in large animals. *Vaccine* **21**: 649.

Bagos PG, Tsirigos KD, Liakopoulos TD, Hamodrakas SJ. 2008. Prediction of lipoprotein signal peptides in Gram-positive bacteria with a Hidden Markov Model. *Journal of Proteome Research* **7**(12): 5082.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Research* **37**(Web Server issue): W202.

Batista L, Pijoan C, Ruiz A, Utrera V, Dee S. 2004. Assessment of transmission of *Mycoplasma hyopneumoniae* by personnel. *Journal of Swine Health and Production* **12**(2): 75.

Bermúdez-Humarán LG, Kharrat P, Chatel JM, Langella P. 2011. Lactococci and lactobacilli as mucosal delivery vectors for therapeutic proteins and DNA vaccines. *Microbial Cell Factories* **10**(Supplement 1): S4.

Berntsson RP, Doeven MK, Fusetti F, Duurkens RH, Sengupta D, Marrink SJ, Thunnissen AM, Poolman B, Slotboom DJ. 2009. The structural basis for peptide selection by the transport receptor OppA. *The EMBO Journal* **28**(9): 1332.

Berntsson RP, Smits SH, Schmitt L, Slotboom DJ, Poolman B. 2010. A structural classification of substrate-binding proteins. *FEBS Letters* **584**(12): 2606.

Beven L, Charenton C, Dautant A, Bouyssou G, Labroussaa F, Skollermo A, Persson A, Blanchard A, Sirand-Pugnet P. 2012. Specific evolution of F1-like ATPases in mycoplasmas. *PLoS One* **7**(6): e38793.

Bizarro CV, Schuck DC. 2007. Purine and pyrimidine nucleotide metabolism in *Mollicutes*. *Genetics and Molecular Biology* **30**(1): 190.

Black D, Glatz PC. 2011. Ratite health: Welfare implications. In *The Welfare of Farmed Ratites* (ed. PC Glatz, C Lunam, I Malecki), pp. 165-194. Springer, Berlin Heidelberg.

Blignaut A, Burger WP, Morley AJ, Bellstedt DU. 2000. Antibody responses to La Sota strain vaccines of Newcastle disease virus in ostriches (*Struthio camelus*) as detected by enzyme-linked immunosorbent assay. *Avian Diseases* **44**(2): 390.

Bonato M, Evans MR, Hasselquist D, Cloete SWP, Cherry MI. 2009. Growth rate and hatching date in ostrich chicks reflect humoral but not cell-mediated immune function. *Behavioral Ecology and Sociobiology* **64**(2): 183.

Botes A. 2004. Immunological and epidemiological investigations in South African ostriches and penguins. PhD Dissertation, Stellenbosch University, South Africa.

Botes A, Peyrot BM, Olivier AJ, Burger WP, Bellstedt DU. 2005a. Identification of three novel *Mycoplasma* species from ostriches in South Africa. *Veterinary Microbiology* **111**: 159.

Botes A, Peyrot BM, Olivier AJ, Burger WP, Bellstedt DU. 2005b. Investigations into mycoplasma infections in South African ostriches. In *Proceedings of the 3rd International Ratite Science Symposium of the World's Poultry Science Association (WPSA) and 12th World Ostrich Congress*, pp. 211-216, Madrid, Spain.

Bradbury JM. 2005. Gordon memorial lecture. Poultry mycoplasmas: Sophisticated pathogens in simple guise. *British Poultry Science* **46**(2): 125.

Braun P, Hu Y, Shen B, Halleck A, Koundinya M, Harlow E, LaBaer J. 2002. Proteome-scale purification of human proteins from bacteria. *Proceedings of the National Academy of Sciences* **99**(5): 2654.

Breton M, Tardy F, Dordet-Frisoni E, Sagne E, Mick V, Renaudin J, Sirand-Pugnet P, Citti C, Blanchard A. 2012. Distribution and diversity of mycoplasma plasmids: Lessons from cryptic genetic elements. *BMC Microbiology* **12**(1): 257.

Brocchieri L. 2001. Phylogenetic inferences from molecular sequences: Review and critique. *Theoretical Population Biology* **59**(1): 27.

Brown DR. 2010. Phylum XVI. Tenericutes Murray 1984a, 356^{VP} (Effective publication: Murray 1984b, 33.). In *Bergey's Manual® of Systematic Bacteriology* (ed. NR Krieg, JT Staley, DR Brown, BP Hedlund, BJ Paster, NL Ward, W Ludwig, WB Whitman), pp. 567-723. Springer New York.

Brown DR, Farmerie WG, May M, Benders GA, Durkin AS, Hlavinka K, Hostetler J, Jackson J, Johnson J, Miller RH *et al.* 2011. Genome sequences of *Mycoplasma alligatoris* A21JP2T and *Mycoplasma crocodyli* MP145T. *Journal of Bacteriology* **193**(11): 2892.

Brown TA. 2002. Chapter 13, Genome Replication. In *Genomes 2nd edition*. Oxford: Wiley-Liss Available from: http://www.ncbi.nlm.nih.gov/books/NBK21113/.

Browning GF, Marenda MS, Noormohammadi AH, Markham PF. 2011. The central role of lipoproteins in the pathogenesis of mycoplasmoses. *Veterinary Microbiology* **153**: 44.

Browning GF, Noormohammadi AH, Markham PF. 2014. Identification and characterization of virulence genes in mycoplasmas. In *Mollicutes: Molecular Biology and Pathogenesis* (ed. GF Browning, C Citti), pp. 77-90. Caister Academic Press, Norfolk, UK.

Bueno TP, Miranda LS, Ruiz-Apodaca IR, Bravo TD, Edreira MR, Marrero SM. 2014. Influence of culture medium components on *Mycoplasma gallisepticum* growth. *Revista de Salud Animal* **36**(2): 130.

Bunter KL, Cloete SWP. 2004. Genetic parameters for egg-, chick- and live-weight traits recorded in farmed ostriches (*Struthio camelus*). *Livestock Production Science* **91**(1): 9.

Caffrey CR, Rohwer A, Oellien F, Marhöfer RJ, Braschi S, Oliveira G, McKerrow JH, Selzer PM. 2009. A comparative chemogenomics strategy to predict potential drug targets in the metazoan pathogen, *Schistosoma mansoni*. *PLoS One* **4**(2): e4413.

Cai Y, Rodriguez S, Hebel H. 2009. DNA vaccine manufacture: Scale and quality. *Expert Review of Vaccines* **8**(9): 1277.

Castellanos M, Wilson DB, Shuler ML. 2004. A modular minimal cell model: Purine and pyrimidine transport and metabolism. *Proceedings of the National Academy of Sciences of the United States of America* **101**(17): 6681.

Chastain M, Simon AJ, Soper KA, Holder DJ, Montgomery DL, Sagar SL, Casimiro DR, Middaugh CR. 2001. Antigen levels and antibody titers after DNA vaccination. *Journal of Pharmaceutical Sciences* **90**(4): 474.

Chen LL, Chung WC, Lin CP, Kuo CH. 2012. Comparative analysis of gene content evolution in phytoplasmas and mycoplasmas. *PLoS One* **7**(3): e34407.

Chopra-Dewasthaly R, Citti C, Glew MD, Zimmermann M, Rosengarten R, Jechlinger W. 2008. Phase-locked mutants of *Mycoplasma agalactiae*: Defining the molecular switch of high-frequency Vpma antigenic variation. *Molecular Microbiology* **67**(6): 1196.

Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P. 2006. Toward automatic reconstruction of a highly resolved tree of life. *Science* **311**(5765): 1283.

Citti C, Blanchard A. 2013. Mycoplasmas and their host: Emerging and re-emerging minimal pathogens. *Trends in Microbiology* **21**(4): 196.

Citti C, Nouvel LX, Baranowski E. 2010. Phase and antigenic variation in mycoplasmas. *Future Microbiology* **5**(7): 1073.

Cizelj I, Bercic RL, Dusanic D, Narat M, Kos J, Dovc P, Bencina D. 2010. My*coplasma gallisepticum* and *Mycoplasma synoviae* express a cysteine protease, CysP, which can cleave chicken IgG into Fab and Fc. *Microbiology* **157**(2): 362.

Clore GM, Venditti V. 2013. Structure, dynamics and biophysics of the cytoplasmic protein–protein complexes of the bacterial phosphoenolpyruvate: sugar phosphotransferase system. *Trends in Biochemical Sciences* **38**(10): 515.

Commans S, Böck A. 1999. Selenocysteine inserting tRNAs: An overview. *FEMS Microbiology Reviews* **23**(3): 335.

Cordova CMM, Lartigue C, Sirand-Pugnet P, Renaudin J, Cunha RAF, Blanchard A. 2002. Identification of the origin of replication of the *Mycoplasma pulmonis* chromosome and its use in *oriC* replicative plasmids. *Journal of Bacteriology* **184**(19): 5426.

DAFF. 2014. A profile of the South African ostrich market value chain. Department of Agriculture, Forestry and Fisheries, Pretoria, South Africa.

Darwish WS, Eldaly EA, El-Abbasy MT, Ikenaka Y, Nakayama S, Ishizuka M. 2013. Antibiotic residues in food: The African scenario. *Japanese Journal of Veterinary Research* **61**(Supplement): S13.

Delaney NF, Balenger S, Bonneaud C, Marx CJ, Hill GE, Ferguson-Noel N, Tsai P, Rodrigo A, Edwards SV. 2012. Ultrafast evolution and loss of CRISPRs following a host shift in a novel wildlife pathogen, *Mycoplasma gallisepticum*. *PLoS Genetics* **8**(2): e1002511.

Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**(6): 673.

Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Research* **27**(23): 4636.

Detmer A, Glenting J. 2006. Live bacterial vaccines- A review and identification of potential hazards. *Microbial Cell Factories* **5**: 23.

do Nascimento NC, Dos Santos AP, Chu Y, Guimaraes AM, Pagliaro A, Messick JB. 2013. Genome sequence of *Mycoplasma parvum* (formerly *Eperythrozoon parvum*), a diminutive hemoplasma of the pig. *Genome Announcements* **1**(6): e00986.

do Nascimento NC, Santos AP, Guimaraes AM, Sanmiguel PJ, Messick JB. 2012. *Mycoplasma haemocanis* - The canine hemoplasma and its feline counterpart in the genomic era. *Veterinary Research* **43**(1): 66.

Doeven MK, Kok J, Poolman B. 2005. Specificity and selectivity determinants of peptide transport in *Lactococcus lactis* and other microorganisms. *Molecular Microbiology* **57**(3): 640.

Dordet Frisoni E, Marenda MS, Sagné E, Nouvel LX, Guérillot R, Glaser P, Blanchard A, Tardy F, Sirand-Pugnet P, Baranowski E. 2013. ICEA of *Mycoplasma agalactiae*: A new family of self-transmissible integrative elements that confers conjugative properties to the recipient strain. *Molecular Microbiology* **89**(6): 1226.

Dorman CJ. 2011. Regulation of transcription by DNA supercoiling in *Mycoplasma genitalium*: Global control in the smallest known self-replicating genome. *Molecular Microbiology* **81**(2): 302.

Dybvig K, Cao ZH, Lao P, Jordan DS, French CT, Tu AHT, Loraine AE. 2008. Genome of *Mycoplasma arthritidis*. *Infection and Immunity* **76**(9): 4000.

Dybvig K, Lao P, Jordan DS, Simmons WL. 2010. Fewer essential genes in mycoplasmas than previous studies suggest. *FEMS Microbiology Letters* **311**(1): 51.

Dybvig K, Sitaraman R, French CT. 1998. A family of phase-variable restriction enzymes with differing specificities generated by high-frequency gene rearrangements. *Proceedings of the National Academy of Sciences* **95**(23): 13923.

Ederveen THA, Overmars L, van Hijum SAFT. 2013. Reduce manual curation by combining gene predictions from multiple annotation engines, a case study of start codon prediction. *PLoS One* **8**(5): e63523.

Engelen S, Vallenet D, Médigue C, Danchin A. 2012. Distinct co-evolution patterns of genes associated to DNA polymerase III DnaE and PolC. *BMC Genomics* **13**(1): 69.

Erickson HP, Osawa M. 2010. Cell division without FtsZ - A variety of redundant mechanisms. *Molecular Microbiology* **78**(2): 267.

Escoffre JM, Zeghimi A, Novell A, Bouakaz A. 2013. *In-vivo* gene delivery by sonoporation: Recent progress and prospects. *Current Gene Therapy* **13**(1): 2.

Evans JD, Leigh SA, Branton SL, Collier SD, Pharr GT, Bearson SMD. 2005. *Mycoplasma gallisepticum*: Current and developing means to control the avian pathogen. *Journal of Applied Poultry Research* **14**(4): 757.

Faurez F, Dory D, Le Moigne V, Gravier R, Jestin A. 2010. Biosafety of DNA vaccines: New generation of DNA vectors and current knowledge on the fate of plasmids after injection. *Vaccine* **28**(23): 3888.

Faustino CR, Jennelle CS, Connolly V, Davis AK, Swarthout EC, Dhondt AA, Cooch EG. 2004. *Mycoplasma gallisepticum* infection dynamics in a house finch population: Seasonal variation in survival, encounter and transmission rate. *Journal of Animal Ecology* **73**(4): 651.

FDA. 1996. Points to consider on plasmid DNA vaccines for preventive infectious disease indications. Docket No. 96-N-0400. *Food and Drug Administration, Washington, DC*.

Ferraro B, Morrow MP, Hutnick NA, Shin TH, Lucke CE, Weiner DB. 2011. Clinical applications of DNA vaccines: Current progress. *Clinical Infectious Diseases* **53**(3): 296.

Findik A, Çiftci A. 2012. Bacterial DNA vaccines in veterinary medicine: A review. *Journal of Veterinary Advances* **2**(4): 139.

Fioretti D, Iurescia S, Rinaldi M. 2014. Recent advances in design of immunogenic and effective naked DNA vaccines against cancer. *Recent Patents on Anti-cancer Drug Discovery* **9**(1): 66.

Fisunov GY, Alexeev DG, Bazaleev NA, Ladygina VG, Galyamina MA, Kondratov IG, Zhukova NA, Serebryakova MV, Demina IA, Govorun VM. 2011. Core proteome of the minimal cell: Comparative proteomics of three mollicute species. *PLoS One* **6**(7): e21964.

Flingai S, Czerwonko M, Goodman J, Kudchodkar SB, Muthumani K, Weiner DB. 2013. Synthetic DNA vaccines: Improved vaccine potency by electroporation and co-delivered genetic adjuvants. *Frontiers in Immunology* **4**: 354.

Flores-Medina S, Soriano-Becerril DM, Diaz-Garcia FJ. 2012. Molecular diagnostics of mycoplasmas: Perspectives from the microbiology standpoint. In *Polymerase Chain Reaction* (ed. P Hernandez-Rodriguez, APR Gomez), pp. 119-142. INTECH Open Access Publisher.

Fonseca MM, Harris DJ, Posada D. 2014. Origin and length distribution of unidirectional prokaryotic overlapping genes. *G3: Genes| Genomes| Genetics* **4**(1): 19.

Fox LK, Kirk JH, Britten A. 2005. Mycoplasma mastitis: A review of transmission and control. *Journal of Veterinary Medicine Series B-Infectious Diseases and Veterinary Public Health* **52**(4): 153.

Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM *et al.* 1995. The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**(5235): 397.

French CT, Lao P, Loraine AE, Matthews BT, Yu H, Dybvig K. 2008. Large-scale transposon mutagenesis of *Mycoplasma pulmonis*. *Molecular Microbiology* **69**(1): 67.

Frey J. 2002. Mycoplasmas of animals. In *Molecular biology and pathogenicity of mycoplasmas* (ed. S Razin, R Herrmann), pp. 73-90. Springer US.

Fukuda Y, Nakayama Y, Tomita M. 2003. On dynamics of overlapping genes in bacterial genomes. *Gene* **323**: 181.

Fukuda Y, Washio T, Tomita M. 1999. Comparative study of overlapping genes in the genomes of *Mycoplasma genitalium* and *Mycoplasma pneumoniae*. *Nucleic Acids Research* **27**(8): 1847.

Fürnkranz U, Siebert-Gulle K, Rosengarten R, Szostak MP. 2013. Factors influencing the cell adhesion and invasion capacity of *Mycoplasma gallisepticum*. *Acta Veterinaria Scandinavica* **55**(1): 63.

Galens K, Orvis J, Daugherty S, Creasy HH, Angiuoli S, White O, Wortman J, Mahurkar A, Giglio MG. 2011. The IGS standard operating procedure for automated prokaryotic annotation. *Standards in Genomic Sciences* **4**(2): 244.

Galli V, Simionatto S, Marchioro SB, Fisch A, Gomes CK, Conceição FR, Dellagostin OA. 2012. Immunisation of mice with *Mycoplasma hyopneumoniae* antigens P37, P42, P46 and P95 delivered as recombinant subunit or DNA vaccines. *Vaccine* **31**(1): 135.

Galperin MY, Koonin EV. 1999. Searching for drug targets in microbial genomes. *Current Opinion in Biotechnology* **10**(6): 571.

Gao F, Zhang CT. 2008. Ori-Finder: A web-based system for finding oriCs in unannotated bacterial genomes. *BMC Bioinformatics* **9**(1): 79.

Gardner SW, Minion FC. 2010. Detection and quantification of intergenic transcription in *Mycoplasma hyopneumoniae*. *Microbiology* **156**(8): 2305.

Garmory HS, Brown KA, Titball RW. 2003. DNA vaccines: Improving expression of antigens. *Genetic Vaccines and Therapy* **1**(1): 2.

Garmory HS, Titball RW. 2004. ATP-binding cassette transporters are targets for the development of antibacterial vaccines and therapies. *Infection and Immunity* **72**(12): 6757.

Garver KA, Conway CM, Elliott DG, Kurath G. 2005. Analysis of DNA-vaccinated fish reveals viral antigen in muscle, kidney and thymus, and transient histopathologic changes. *Marine Biotechnology* **7**(5): 540.

Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM *et al.* 2010. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**(5987): 52.

Giese M. 1998. DNA-antiviral vaccines: New developments and approaches- A review. *Virus Genes* **17**(3): 219.

Giese M. 2012. Approved veterinary vaccines: Paving the way to products for human health. In *Gene Vaccines* (ed. J Thalhamer, R Weiss, S Scheiblhofer), pp. 265-283. Springer, Vienna.

Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, Hutchison CA, Smith HO, Venter JC. 2006. Essential genes of a minimal bacterium. *Proceedings of the National Academy of Sciences* **103**(2): 425.

Glass JI, Lefkowitz EJ, Glass JS, Heiner CR, Chen EY, Cassell GH. 2000. The complete sequence of the mucosal pathogen *Ureaplasma urealyticum*. *Nature* **407**(6805): 757.

Glenn TC. 2011. Field guide to next-generation DNA sequencers. *Molecular Ecology Resources* **11**(5): 759.

Gollop N, March PE. 1991. A GTP-binding protein (Era) has an essential role in growth rate and cell cycle control in *Escherichia coli*. *Journal of Bacteriology* **173**(7): 2265.

Grandi G. 2010. Bacterial surface proteins and vaccines. *F1000 Biology Reports* **2**: 36.

Grosjean H, Breton M, Sirand-Pugnet P, Tardy F, Thiaucourt F, Citti C, Barré A, Yoshizawa S, Fourmy D, de Crécy-Lagard V. 2014. Predicting the minimal translation apparatus: Lessons from the reductive evolution of *Mollicutes*. *PLoS Genetics* **10**(5): e1004363.

Großhennig S, Schmidl SR, Schmeisky G, Busse J, Stülke J. 2013. Implication of glycerol and phospholipid transporters in *Mycoplasma pneumoniae* growth and virulence. *Infection and Immunity* **81**(3): 896.

Grover RK, Zhu X, Nieusma T, Jones T, Boero I, MacLeod AS, Mark A, Niessen S, Kim HJ, Kong L. 2014. A structurally distinct human mycoplasma protein that generically blocks antigen-antibody union. *Science* **343**(6171): 656.

Guell M, van Noort V, Yus E, Chen WH, Leigh-Bell J, Michalodimitrakis K, Yamada T, Arumugam M, Doerks T, Kuhner S *et al.* 2009. Transcriptome complexity in a genome-reduced bacterium. *Science* **326**(5957): 1268.

Guimaraes AM, Santos AP, SanMiguel P, Walter T, Timenetsky J, Messick JB. 2011. Complete genome sequence of *Mycoplasma suis* and insights into its biology and adaption to an erythrocyte niche. *PLoS One* **6**(5): e19574.

Halbedel S, Hames C, Stulke J. 2007. Regulation of carbon metabolism in the mollicutes and its relation to virulence. *Journal of Molecular Microbiology and Biotechnology* **12**(1): 147.

Halbedel S, Stulke J. 2007. Tools for the genetic analysis of *Mycoplasma*. *International Journal of Medical Microbiology* **297**(1): 37.

Hall TA. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**: 95.

Harrington CT, Lin EI, Olson MT, Eshleman JR. 2013. Fundamentals of pyrosequencing. *Archives of Pathology and Laboratory Medicine* **137**(9): 1296.

Hartfield M, Murall CL, Alizon S. 2014. Clinical applications of pathogen phylogenies. *Trends in Molecular Medicine* **21**(4): 196.

Hartikka J, Sawdey M, Cornefert-Jensen F, Margalith M, Barnhart K, Nolasco M, Vahlsing HL, Meek J, Marquet M, Hobart P. 1996. An improved plasmid DNA expression vector for direct injection into skeletal muscle. *Human Gene Therapy* **7**(10): 1205.

Hatchel JM, Balish MF. 2008. Attachment organelle ultrastructure correlates with phylogeny, not gliding motility properties, in *Mycoplasma pneumoniae* relatives. *Microbiology* **154**(1): 286.

Hegde S, Hegde S, Spergser J, Brunthaler R, Rosengarten R, Chopra-Dewasthaly R. 2014. *In vitro* and *in vivo* cell invasion and systemic spreading of *Mycoplasma agalactiae* in the sheep infection model. *International Journal of Medical Microbiology* **304**(8): 1024.

Hempstead PG. 1990. An improved method for the rapid isolation of chromosomal DNA from *Mycoplasma* spp. *Canadian Journal of Microbiology* **36**(1): 59.

Henrich B, Feldmann RC, Hadding U. 1993. Cytoadhesins of *Mycoplasma hominis*. *Infection and Immunity* **61**(7): 2945.

Henrich B, Hopfe M, Kitzerow A, 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*. *Journal of Bacteriology* **181**(16): 4873.

Hilz H, Wiegers U, Adamietz P. 1975. Stimulation of proteinase K action by denaturing agents: Application to the isolation of nucleic acids and the degradation of 'masked' proteins. *European Journal of Biochemistry* **56**(1): 103.

Himmelreich R, Hilbert H, Plagens H, Pirkl E, Li BC, Herrmann R. 1996. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Research* **24**(22): 4420.

Hoelzle K, Doser S, Ritzmann M, Heinritzi K, Palzer A, Elicker S, Kramer M, Felder KM, Hoelzle LE. 2009. Vaccination with the *Mycoplasma suis* recombinant adhesion protein MSG1 elicits a strong immune response but fails to induce protection in pigs. *Vaccine* **27**(39): 5376.

Hoffman L, Lambrechts H. 2011. Bird handling, transportation, lairage, and slaughter: Implications for bird welfare and meat quality. In *The Welfare of Farmed Ratites* (ed. P Glatz, C Lunam, I Malecki), pp. 195-235. Springer, Berlin Heidelberg.

Hopfe M, Henrich B. 2004. OppA, the substrate-binding subunit of the oligopeptide permease, is the major ecto-ATPase of *Mycoplasma hominis*. *Journal of Bacteriology* **186**(4): 1021.

Hopfe M, Henrich B. 2008. OppA, the ecto-ATPase of *Mycoplasma hominis* induces ATP release and cell death in HeLa cells. *BMC Microbiology* **8**(1): 55.

Hopfe M, Henrich B. 2014. Multifunctional cytoadherence factors. In *Mollicutes: Molecular Biology and Pathogenesis* (ed. GF Browning, C Citti), pp. 107-129. Caister Academic Press, Norfolk, UK.

Houk BE, Martin R, Hochhaus G, Hughes JA. 2001. Pharmacokinetics of plasmid DNA in the rat. *Pharmaceutical Research* **18**(1): 67.

Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, Bernard T, Binns D, Bork P, Burge S *et al.* 2012. InterPro in 2011: New developments in the family and domain prediction database. *Nucleic Acids Research* **40**(Database issue): D306.

Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, Smith HO, Venter JC. 1999. Global transposon mutagenesis and a minimal mycoplasma genome. *Science* **286**(5447): 2165.

Inamine JM, Ho KC, Loechel S, Hu PC. 1990. Evidence that UGA is read as a tryptophan codon rather than as a stop codon by *Mycoplasma pneumoniae*, *Mycoplasma genitalium*, and *Mycoplasma gallisepticum*. *Journal of Bacteriology* **172**(1): 504.

Ishii TM, Zerr P, Xia XM, Bond CT, Maylie J, Adelman JP. 1998. Site-directed mutagenesis. *Methods in Enzymology* **293**: 53.

Ito K, Akiyama Y. 2005. Cellular functions, mechanism of action, and regulation of FtsH protease. *Annual Review of Microbiology* **59**: 211.

Jacob R, Branton SL, Evans JD, Leigh SA, Peebles ED. 2014. Effects of live and killed vaccines against *Mycoplasma gallisepticum* on the performance characteristics of commercial layer chickens. *Poultry Science* **93**(6): 1403.

Jaffe JD, Stange-Thomann N, Smith C, DeCaprio D, Fisher S, Butler J, Calvo S, Elkins T, FitzGerald MG, Hafez N *et al.* 2004. The complete genome and proteome of *Mycoplasma mobile*. *Genome Research* **14**(8): 1447.

Kanellos T, Sylvester ID, Ambali AG, Howard CR, Russell PH. 1999. The safety and longevity of DNA vaccines for fish. *Immunology* **96**: 307.

Kent BN, Foecking MF, Calcutt MJ. 2012. Development of a novel plasmid as a shuttle vector for heterologous gene expression in *Mycoplasma yeatsii*. *Journal of Microbiological Methods* **91**(1): 121.

Kibeida OAI. 2011. A comparison of methods used to measure the *in vitro* antimicrobial susceptibilities of *Mycoplasma* species of animal origin. MSc Thesis, University of Pretoria South Africa.

Kircher M, Kelso J. 2010. High-throughput DNA sequencing - Concepts and limitations. *Bioessays* **32**(6): 524.

Kitahara K, Yasutake Y, Miyazaki K. 2012. Mutational robustness of 16S ribosomal RNA, shown by experimental horizontal gene transfer in *Escherichia coli*. *Proceedings of the National Academy of Sciences* **109**(47): 19220.

Klepsch MM, Kovermann M, Löw C, Balbach J, Permentier HP, Fusetti F, de Gier JW, Slotboom DJ, Berntsson RP. 2011. *Escherichia coli* peptide binding protein OppA has a preference for positively charged peptides. *Journal of Molecular Biology* **414**(1): 75.

Kleven SH. 2008. Control of avian mycoplasma infections in commercial poultry. *Avian Diseases* **52**(3): 367.

Klotz C, Gehre F, Lucius R, Pogonka T. 2007. Identification of *Eimeria tenella* genes encoding for secretory proteins and evaluation of candidates by DNA immunisation studies in chickens. *Vaccine* **25**(36): 6625.

Kono N, Arakawa K, Tomita M. 2012. Validation of bacterial replication termination models using simulation of genomic mutations. *PLoS One* **7**(4): e34526.

Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, Wang Z, Rasko DA, McCombie WR, Jarvis ED. 2012. Hybrid error correction and *de novo* assembly of single-molecule sequencing reads. *Nature Biotechnology* **30**(7): 693.

Kutzler MA, Weiner DB. 2008. DNA vaccines: ready for prime time? *Nature Reviews Genetics* **9**(10): 776.

Langer B, Renner M, Scherer J, Schüle S, Cichutek K. 2013. Safety assessment of biolistic DNA vaccination. In *Biolistic DNA delivery*, pp. 371-388. Springer.

Langer S. 2009. Proposal for molecular tools for the epidemiology of contagious bovine pleuro pneumonia and classification of unknown *Mycoplasma* sp. isolated from *Struthio camelus*. Mag. rer.nat., Universität Wien, Republic of Austria.

Lartigue C, Blanchard A, Renaudin J, Thiaucourt F, Sirand-Pugnet P. 2003. Host specificity of *Mollicutes oriC* plasmids: Functional analysis of replication origin. *Nucleic Acids Research* **31**(22): 6610.

Lechardeur D, Lukacs GL. 2006. Nucleocytoplasmic transport of plasmid DNA: A perilous journey from the cytoplasm to the nucleus. *Human Gene Therapy* **17**(9): 882.

Lechardeur D, Sohn KJ, Haardt M, Joshi PB, Monck M, Graham RW, Beatty B, Squire J, O'Brodovich H, Lukacs GL. 1999. Metabolic instability of plasmid DNA in the cytosol: A potential barrier to gene transfer. *Gene Therapy* **6**(4): 482.

Lee CW, Senne DA, Suarez DL. 2003. Development of hemagglutinin subtype-specific reference antisera by DNA vaccination of chickens. *Avian Diseases* **47**(S3): 1051.

Lee EM, Ahn SH, Park JH, Lee JH, Ahn SC, Kong IS. 2004. Identification of oligopeptide permease (*opp*) gene cluster in *Vibrio fluvialis* and characterization of biofilm production by *oppA* knockout mutation. *FEMS Microbiology Letters* **240**(1): 21.

Lee SW, Browning GF, Markham PF. 2008. Development of a replicable *oriC* plasmid for *Mycoplasma gallisepticum* and *Mycoplasma imitans*, and gene disruption through homologous recombination in *M. gallisepticum*. *Microbiology* **154**(9): 2571.

Levine MM, Sztein MB. 2004. Vaccine development strategies for improving immunization: The role of modern immunology. *Nature Immunology* **5**(5): 460.

Lewis PJ, Babiuk LA. 1999. DNA vaccines: A review. *Advances in Virus Research* **54**: 129.

Li L, Saade F, Petrovsky N. 2012. The future of human DNA vaccines. *Journal of Biotechnology* **162**(2): 171.

Li Y, Zheng H, Liu Y, Jiang Y, Xin J, Chen W, Song Z. 2011. The complete genome sequence of *Mycoplasma bovis* strain Hubei-1. *PLoS One* **6**(6): e20999.

Lin Y, Zhang RR. 2011. Putative essential and core-essential genes in mycoplasma genomes. *Scientific reports* **1**: 53.

Lin YL, Pasero P. 2012. Interference between DNA replication and transcription as a cause of genomic instability. *Current Genomics* **13**(1): 65.

Liu L, Panangala VS, Dybvig K. 2002. Trinucleotide GAA repeats dictate pMGA gene expression in *Mycoplasma gallisepticum* by affecting spacing between flanking regions. *Journal of Bacteriology* **184**(5): 1335.

Liu MA. 2011. DNA vaccines: An historical perspective and view to the future. *Immunological Reviews* **239**(1): 62.

Liu W, Fang L, Li M, Li S, Guo S, Luo R, Feng Z, Li B, Zhou Z, Shao G *et al.* 2012. Comparative genomics of mycoplasma: Analysis of conserved essential genes and diversity of the pan-genome. *PLoS One* **7**(4): e35698.

Lluch-Senar M, Querol E, Pinol J. 2010. Cell division in a minimal bacterium in the absence of *ftsZ*. *Molecular Microbiology* **78**(2): 278.

Lluch-Senar M, Vallmitjana M, Querol E, Pinol J. 2007. A new promoterless reporter vector reveals antisense transcription in *Mycoplasma genitalium*. *Microbiology* **153**(8): 2743.

Loreto ELS, Ortiz MF, Port JIR. 2007. Insertion sequences as variability generators in the *Mycoplasma hyopneumoniae* and *M. synoviae* genomes. *Genetics and Molecular Biology* **30**(1): 283.

Lukacs GL, Haggie P, Seksek O, Lechardeur D, Freedman N, Verkman AS. 2000. Size-dependent DNA mobility in cytoplasm and nucleus. *Journal of Biological Chemistry* **275**(3): 1625.

Lysnyansky I, Calcutt MJ, Ben-Barak I, Ron Y, Levisohn S, Methe BA, Yogev D. 2009. Molecular characterization of newly identified IS3, IS4 and IS30 insertion sequence-like elements in *Mycoplasma bovis* and their possible roles in genome plasticity. *FEMS Microbiology Letters* **294**(2): 172.

Macao B, Luirink J, Samuelsson T. 1997. Ffh and FtsY in a *Mycoplasma mycoides* signal-recognition particle pathway: SRP RNA and M domain of Ffh are not required for stimulation of GTPase activity *in vitro*. *Molecular Microbiology* **24**(3): 523.

MacGregor RR, Boyer JD, Ugen KE, Lacy KE, Gluckman SJ, Bagarazzi ML, Chattergoon MA, Baine Y, Higgins TJ, Ciccarelli RB *et al.* 1998. First human trial of a DNA-based vaccine for treatment of human immunodeficiency virus type 1 infection: Safety and host response. *Journal of Infectious Diseases* **178**(1): 92.

Madeira HMF, Gabriel JE. 2007. Regulation of gene expression in mycoplasmas: Contribution from *Mycoplasma hyopneumoniae* and *Mycoplasma synoviae* genome sequences. *Genetics and Molecular Biology* **30**(1): 277.

Maglennon GA, Cook BS, Matthews D, Deeney AS, Bossé JT, Langford PR, Maskell DJ, Tucker AW, Wren BW, Rycroft AN. 2013. Development of a self-replicating plasmid system for *Mycoplasma hyopneumoniae*. *Veterinary Research* **44**(1): 63.

Mahoney RT, Wen YM, Wilde H, Xu ZY. 2000. The introduction of new DNA vaccines into developing countries. In *Madame Curie Bioscience Database*. Landes Biosciences, Georgetown. Available from: http://www.ncbi.nlm.nih.gov/books/NBK5994/.

Manjila SB, Baby JN, Bijin EN, Constantine I, Pramod K, Valsalakumari J. 2013. Novel gene delivery systems. *International Journal of Pharmaceutical Investigation* **3**(1): 1.

Mao F, Dam P, Chou J, Olman V, Xu Y. 2009. DOOR: A database for prokaryotic operons. *Nucleic Acids Research* **37**(Database issue): D459.

Marchioro SB, Maes D, Flahou B, Pasmans F, Del Pozo Sacristán R, Vranckx K, Melkebeek V, Cox E, Wuyts N, Haesebrouck F. 2013. Local and systemic immune responses in pigs intramuscularly injected with an inactivated *Mycoplasma hyopneumoniae* vaccine. *Vaccine* **31**(9): 1305.

Marenda MS. 2014. Genomic mosaics. In *Mollicutes: Molecular Biology and Pathogenesis* (ed. GF Browning, C Citti), pp. 15-54. Caister Academic Press, Norfolk, UK.

Marín A, Xia X. 2008. GC skew in protein-coding genes between the leading and lagging strands in bacterial genomes: New substitution models incorporating strand bias. *Journal of Theoretical Biology* **253**(3): 508.

Mauro VP, Chappell SA. 2014. A critical analysis of codon optimization in human therapeutics. *Trends in Molecular Medicine* **20**(11): 604.

May M, Brown DR. 2014. Glycosidase activity in *Mollicutes*. In *Mollicutes: Molecular Biology and Pathogenesis* (ed. GF Browning, C Citti), p. 149. Caister Academic Press, Norfolk, UK.

McCormack WM. 1993. Susceptibility of mycoplasmas to antimicrobial agents: Clinical implications. *Clinical Infectious Diseases* **17**(Supplement 1): S200.

McCutchan TF, Grim KC, Li J, Weiss W, Rathore D, Sullivan M, Graczyk TK, Kumar S, Cranfield MR. 2004. Measuring the effects of an ever-changing environment on malaria control. *Infection and Immunity* **72**(4): 2248.

McDowall J, Hunter S. 2011. InterPro protein classification. *Methods in Molecular Biology* **694**: 37.

Medina JL, Coalson JJ, Brooks EG, Winter VT, Chaparro A, Principe MF, Kannan TR, Baseman JB, Dube PH. 2012. *Mycoplasma pneumoniae* CARDS toxin induces pulmonary eosinophilic and lymphocytic inflammation. *American Journal of Respiratory Cell and Molecular Biology* **46**(6): 815.

Mena A, Andrew ME, Coupar BEH. 2001. Rapid dissemination of intramuscularly inoculated DNA vaccines. *Immunology and Cell Biology* **79**(1): 87.

Messick JB. 2004. Hemotrophic mycoplasmas (hemoplasmas): A review and new insights into pathogenic potential. *Veterinary Clinical Pathology* **33**(1): 2.

Metzker ML. 2009. Sequencing technologies- The next generation. *Nature Reviews Genetics* **11**(1): 31.

Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *Gateway Computing Environments Workshop (GCE)*, pp. 1-8.

Mistry A, Warren MS, Cusick JK, Karkhoff-Schweizer RR, Lomovskaya O, Schweizer HP. 2013. High-level pacidamycin resistance in *Pseudomonas aeruginosa* is mediated by an *opp* oligopeptide permease encoded by the *opp-fabI* operon. *Antimicrobial Agents and Chemotherapy* **57**(11): 5565.

Moore C, Cumming GS, Slingsby J, Grewar J. 2014. Tracking socioeconomic vulnerability using network analysis: Insights from an avian influenza outbreak in an ostrich production network. *PLoS One* **9**(1): e86973.

Moran NA. 2002. Microbial minimalism: Genome reduction in bacterial pathogens. *Cell* **108**(5): 583.

Moss RB. 2009. Prospects for control of emerging infectious diseases with plasmid DNA vaccines. *Journal of Immune Based Therapies and Vaccines* **7**(1): 3.

Mourez M, Hofnung M, Dassa E. 1997. Subunit interactions in ABC transporters: A conserved sequence in hydrophobic membrane proteins of periplasmic permeases defines an important site of interaction with the ATPase subunits. *The EMBO journal* **16**(11): 3066.

Movahedi A, Hampson DJ. 2010. Evaluation of recombinant *Brachyspira pilosicoli* oligopeptide-binding proteins as vaccine candidates in a mouse model of intestinal spirochaetosis. *Journal of Medical Microbiology* **59**: 353.

Mushegian AR, Koonin EV. 1996. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proceedings of the National Academy of Sciences of the United States of America* **93**(19): 10268.

Nicolás MF, Barcellos FG, Hess PN, Hungria M. 2007. ABC transporters in *Mycoplasma hyopneumoniae* and *Mycoplasma synoviae*: Insights into evolution and pathogenicity. *Genetics and Molecular Biology* **30**(1): 202.

Nisha AR. 2008. Antibiotic residues- A global health hazard. *Veterinary World* **1**(12): 375.

Noormohammadi AH, Markham PF, Kanci A, Whithear KG, Browning GF. 2000. A novel mechanism for control of antigenic variation in the haemagglutinin gene family of *Mycoplasma synoviae*. *Molecular Microbiology* **35**(4): 911.

Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology* **302**(1): 205.

Nouvel LX, Marenda M, Sirand-Pugnet P, Sagne E, Glew M, Mangenot S, Barbe V, Barre A, Claverol S, Citti C. 2009. Occurrence, plasticity, and evolution of the *vpma* gene family, a genetic system devoted to high-frequency surface variation in *Mycoplasma agalactiae*. *Journal of Bacteriology* **191**(13): 4111.

Okuda K, Wada Y, Shimada M. 2014. Recent developments in preclinical DNA vaccination. *Vaccines* **2**(1): 89.

Oliveira PH, Mairhofer J. 2013. Marker-free plasmids for biotechnological applications - Implications and perspectives. *Trends in Biotechnology* **31**(9): 539.

Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M. 2013. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research* **42**(D1): D206.

Pachkov M, Dandekar T, Korbel J, Bork P, Schuster S. 2007. Use of pathway analysis and genome context methods for functional genomics of *Mycoplasma pneumoniae* nucleotide metabolism. *Gene* **396**(2): 215.

Palumbo RN, Zhong X, Panus D, Han W, Ji W, Wang C. 2012. Transgene expression and local tissue distribution of naked and polymer-condensed plasmid DNA after intradermal administration in mice. *Journal of Controlled Release* **159**(2): 232.

Paulsen IT, Nguyen L, Sliwinski MK, Rabus R, Saier MH. 2000. Microbial genome analyses: Comparative transport capabilities in eighteen prokaryotes. *Journal of Molecular Biology* **301**(1): 75.

Penadés JR, Chen J, Quiles-Puchalt N, Carpena N, Novick RP. 2015. Bacteriophage-mediated spread of bacterial virulence genes. *Current Opinion in Microbiology* **23**: 171.

Pereira VB, Zurita-Turk M, Saraiva TDL, De Castro CP, Souza BM, Agresti PM, Lima FA, Pfeiffer VN, Azevedo MSP, Rocha CS. 2014. DNA vaccines approach: From concepts to applications. *World Journal of Vaccines* **4**(2): 50.

Pereyre S, Sirand-Pugnet P, Beven L, Charron A, Renaudin H, Barre A, Avenaud P, Jacob D, Couloux A, Barbe V *et al.* 2009. Life on arginine for *Mycoplasma hominis*: Clues from its minimal genome and comparison with other human urogenital mycoplasmas. *PLoS Genetics* **5**(10): e1000677.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature Methods* **8**(10): 785.

Pilo P, Frey J, Vilei EM. 2007. Molecular mechanisms of pathogenicity of *Mycoplasma mycoides* subsp mycoides SC. *The Veterinary Journal* **174**(3): 513.

Pilo P, Vilei EM, Peterhans E, Bonvin-Klotz L, Stoffel MH, Dobbelaere D, Frey J. 2005. A metabolic enzyme as a primary virulence factor of *Mycoplasma mycoides* subsp *mycoides* small colony. *Journal of Bacteriology* **187**(19): 6824.

Pinard R, de Winter A, Sarkis GJ, Gerstein MB, Tartaro KR, Plant RN, Egholm M, Rothberg JM, Leamon JH. 2006. Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. *BMC Genomics* **7**(1): 216.

Pitcher DG, Nicholas RAJ. 2005. Mycoplasma host specificity: Fact or fiction? *The Veterinary Journal* **170**(3): 300.

Pollack JD, Myers MA, Dandekar T, Herrmann R. 2002. Suspected utility of enzymes with multiple activities in the small genome *Mycoplasma* species: The replacement of the missing "household" nucleoside diphosphate kinase gene and activity by glycolytic kinases. *OMICS* **6**(3): 247.

Pollard H, Remy JS, Loussouarn G, Demolombe S, Behr JP, Escande D. 1998. Polyethylenimine but not cationic lipids promotes transgene delivery to the nucleus in mammalian cells. *Journal of Biological Chemistry* **273**(13): 7507.

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 Thesis, Stellenbosch University, South Africa.

Price MN, Alm EJ, Arkin AP. 2005. Interruptions in gene expression drive highly expressed operons to the leading strand of DNA replication. *Nucleic Acids Research* **33**(10): 3224.

Prysliak T, van der Merwe J, Perez-Casal J. 2013. Vaccination with recombinant *Mycoplasma bovis* GAPDH results in a strong humoral immune response but does not protect feedlot cattle from an experimental challenge with *M. bovis*. *Microbial Pathogenesis* **55**: 1.

Punyapornwithaya V, Fox LK, Hancock DD, Gay JM, Alldredge JR. 2012. Time to clearance of mycoplasma mastitis: The effect of management factors including milking time hygiene and preferential culling. *The Canadian Veterinary Journal* **53**(10): 1119.

Quaak SG, Haanen JB, Beijnen JH, Nuijen B. 2010. Naked plasmid DNA formulation: Effect of different disaccharides on stability after lyophilisation. *AAPS PharmSciTech* **11**(1): 344.

Raherison S, Gonzalez P, Renaudin H, Charron A, Bébéar C, Bébéar C. 2005. Increased expression of two multidrug transporter-like genes is associated with ethidium bromide and ciprofloxacin resistance in *Mycoplasma hominis*. *Antimicrobial Agents and Chemotherapy* **49**(1): 421.

Raherison S, Gonzalez P, Renaudin H, Charron A, Bebear C, Bebear CM. 2002. Evidence of active efflux in resistance to ciprofloxacin and to ethidium bromide by *Mycoplasma hominis*. *Antimicrobial Agents and Chemotherapy* **46**(3): 672.

Razin S, Yogev D, Naot Y. 1998. Molecular biology and pathogenicity of mycoplasmas. *Microbiology and Molecular Biology Reviews* **62**(4): 1094.

Rechnitzer H, Brzuszkiewicz E, Strittmatter A, Liesegang H, Lysnyansky I, Daniel R, Gottschalk G, Rottem S. 2011. Genomic features and insights into the biology of *Mycoplasma fermentans*. *Microbiology* **157**(3): 760.

Rees DC, Johnson E, Lewinson O. 2009. ABC transporters: The power to change. *Nature Reviews Molecular Cell Biology* **10**(3): 218.

Rice AJ, Park A, Pinkett HW. 2014. Diversity in ABC transporters: Type I, II and III importers. *Critical Reviews in Biochemistry and Molecular Biology* **49**(5): 426.

Rocha EPC. 2008. The organization of the bacterial genome. *Annual Review of Genetics* **42**: 211.

Rocha EPC, Blanchard A. 2002. Genomic repeats, genome plasticity and the dynamics of *Mycoplasma* evolution. *Nucleic Acids Research* **30**(9): 2031.

Rosengarten R, Wise KS. 1990. Phenotypic switching in mycoplasmas: Phase variation of diverse surface lipoproteins. *Science* **247**(4940): 315.

Röske K, Calcutt MJ, Wise KS. 2004. The *Mycoplasma fermentans* prophage φMFV1: Genome organization, mobility and variable expression of an encoded surface protein. *Molecular Microbiology* **52**(6): 1703.

Rudner DZ, Ledeaux JR, Ireton K, Grossman AD. 1991. The *spook* locus of *Bacillus subtilis* is homologous to the oligopeptide permease locus and is required for sporulation and competence. *Journal of Bacteriology* **173**(4): 1388.

Sanders GM, Dallmann HG, McHenry CS. 2010. Reconstitution of the *B. subtilis* replisome with 13 proteins including two distinct replicases. *Molecular Cell* **37**(2): 273.

Sasaki Y, Ishikawa J, Yamashita A, Oshima K, Kenri T, Furuya K, Yoshino C, Horino A, Shiba T, Sasaki T *et al.* 2002. The complete genomic sequence of *Mycoplasma penetrans*, an intracellular bacterial pathogen in humans. *Nucleic Acids Research* **30**(23): 5293.

Saurin W, Dassa E. 1996. In the search of *Mycoplasma genitalium* lost substrate-binding proteins: Sequence divergence could be the result of a broader substrate specificity. *Molecular Microbiology* **22**(2): 389.

Schalk JAC, Mooi FR, Berbers GAM, Van Aerts LAGJM, Ovelgonne H, Kimman TG. 2006. Preclinical and clinical safety studies on DNA vaccines. *Human Vaccines* **2**(2): 45.

Schmidl SR, Otto A, Lluch-Senar M, Piñol J, Busse J, Becher D, Stülke J. 2011. A trigger enzyme in *Mycoplasma pneumoniae*: Impact of the glycerophosphodiesterase GlpQ on virulence and gene expression. *PLoS Pathogens* **7**(9): e1002263.

Schuwirth BS, Borovinskaya MA, Hau CW, Zhang W, Vila-Sanjurjo A, Holton JM, Cate JHD. 2005. Structures of the bacterial ribosome at 3.5 Å resolution. *Science* **310**(5749): 827.

Sheets RL, Stein J, Manetz TS, Duffy C, Nason M, Andrews C, Kong WP, Nabel GJ, Gomez PL. 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. *Toxicological Sciences* **91**(2): 610.

Shendure J, Ji H. 2008. Next-generation DNA sequencing. *Nature Biotechnology* **26**(10): 1135.

Sheppard K, Yuan J, Hohn MJ, Jester B, Devine KM, Söll D. 2008. From one amino acid to another: tRNA-dependent amino acid biosynthesis. *Nucleic Acids Research* **36**(6): 1813.

Shu HW, Liu TT, Chan HI, Liu YM, Wu KM, Shu HY, Tsai SF, Hsiao KJ, Hu WS, Ng WV. 2011. Genome sequence of the repetitive-sequence-rich *Mycoplasma fermentans* strain M64. *Journal of Bacteriology* **193**(16): 4302.

Shu HW, Liu TT, Chan HI, Liu YM, Wu KM, Shu HY, Tsai SF, Hsiao KJ, Hu WS, Ng WV. 2012. Complexity of the *Mycoplasma fermentans* M64 genome and metabolic essentiality and diversity among mycoplasmas. *PLoS One* **7**(4): e32940.

Simionatto S, Marchioro SB, Galli V, Brum CB, Klein CS, Rebelatto R, Silva EF, Borsuk S, Conceição FR, Dellagostin OA. 2012. Immunological characterization of *Mycoplasma hyopneumoniae* recombinant proteins. *Comparative Immunology, Microbiology and Infectious Diseases* **35**(2): 209.

Simmons WL, Cao ZH, Glass JI, Simecka JW, Cassell GH, Watson HL. 1996. Sequence analysis of the chromosomal region around and within the V-1-encoding gene of *Mycoplasma pulmonis*: Evidence for DNA inversion as a mechanism for V-1 variation. *Infection and Immunity* **64**(2): 472.

Simons R, Houman F, Kleckner N. 1987. Improved single and multicopy lac-based cloning vectors for protein and operon fusions. *Gene* **53**(1): 85.

Singh AK, Pluvinage B, Higgins MA, Dalia AB, Woodiga SA, Flynn M, Lloyd AR, Weiser JN, Stubbs KA, Boraston AB. 2014. Unravelling the multiple functions of the architecturally intricate *Streptococcus pneumoniae* β-galactosidase, BgaA. *PLoS Pathogens* **10**(9): e1004364.

Sippel KH, Robbins AH, Reutzel R, Boehlein SK, Namiki K, Goodison S, Agbandje-McKenna M, Rosser CJ, McKenna R. 2009. Structural insights into the extracytoplasmic thiamine-binding lipoprotein p37 of *Mycoplasma hyorhinis*. *Journal of Bacteriology* **191**(8): 2585.

Sippel KH, Venkatakrishnan B, Boehlein SK, Sankaran B, Quirit JG, Govindasamy L, Agbandje-McKenna M, Goodison S, Rosser CJ, McKenna R. 2011. Insights into *Mycoplasma genitalium* metabolism revealed by the structure of MG289, an extracytoplasmic thiamine binding lipoprotein. *Proteins* **79**(2): 528.

Siqueira FM, de Souto Weber S, Cattani AM, Schrank IS. 2014. Genome organization in *Mycoplasma hyopneumoniae*: Identification of promoter-like sequences. *Molecular Biology Reports* **41**(8): 5395.

Siqueira FM, Schrank A, Schrank IS. 2011. *Mycoplasma hyopneumoniae* transcription unit organization: Genome survey and prediction. *DNA Research* **18**(6): 413.

Sirand-Pugnet P, Citti C, Barre A, Blanchard A. 2007a. Evolution of *Mollicutes*: Down a bumpy road with twists and turns. *Research in Microbiology* **158**(10): 754.

Sirand-Pugnet P, Lartigue C, Marenda M, Jacob D, Barre A, Barbe V, Schenowitz C, Mangenot S, Couloux A, Segurens B *et al.* 2007b. Being pathogenic, plastic, and sexual while living with a nearly minimal bacterial genome. *PLoS Genetics* **3**(5): 744.

Slotboom DJ. 2014. Structural and mechanistic insights into prokaryotic energy-coupling factor transporters. *Nature Reviews Microbiology* **12**(2): 79.

Smiley BK, Minion FC. 1993. Enhanced readthrough of opal (UGA) stop codons and production of *Mycoplasma pneumoniae* P1 epitopes in *Escherichia coli*. *Gene* **134**(1): 33.

Soppa J, Kobayashi K, Noirot-Gros MF, Oesterhelt D, Ehrlich SD, Dervyn E, Ogasawara N, Moriya S. 2002. Discovery of two novel families of proteins that are proposed to interact with prokaryotic SMC proteins, and characterization of the *Bacillus subtilis* family members ScpA and ScpB. *Molecular Microbiology* **45**(1): 59.

Spencer PS, Siller E, Anderson JF, Barral JM. 2012. Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. *Journal of Molecular Biology* **422**(3): 328.

Sridhar S, Reyes-Sandoval A, Draper S, Moore A, Gilbert S, Gao G, Wilson J, Hill A. 2008. Single-dose protection against *Plasmodium berghei* by a simian adenovirus vector using a human cytomegalovirus promoter containing intron A. *Journal of Virology* **82**(8): 3822.

Srivatsan A, Tehranchi A, MacAlpine DM, Wang JD. 2010. Co-orientation of replication and transcription preserves genome integrity. *PLoS Genetics* **6**(1): e1000810.

Staats CC, Boldo J, Broetto L, Vainstein M, Schrank A. 2007. Comparative genome analysis of proteases, oligopeptide uptake and secretion systems in *Mycoplasma* spp. *Genetics and Molecular Biology* **30**(1): 225.

Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**(9): 1312.

Stamburski C, Renaudin J, Bové JM. 1991. First step toward a virus-derived vector for gene cloning and expression in spiroplasmas, organisms which read UGA as a tryptophan codon: Synthesis of chloramphenicol acetyltransferase in *Spiroplasma citri*. *Journal of Bacteriology* **173**(7): 2225.

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 Thesis, Stellenbosch University, South Africa.

Strydom M. 2013. The ostrich mycoplasma Ms02 partial genome assembly, bioinformatic analysis and the development of three DNA vaccines. MSc Thesis, Stellenbosch University, South Africa.

Suarez DL, Schultz-Cherry S. 2000. The effect of eukaryotic expression vectors and adjuvants on DNA vaccines in chickens using an avian influenza model. *Avian Diseases* **44**(4): 861.

Sung W, Ackerman MS, Miller SF, Doak TG, Lynch M. 2012. Drift-barrier hypothesis and mutation-rate evolution. *Proceedings of the National Academy of Sciences* **109**(45): 18488.

Swofford DL. 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods) Version 4.0b10 Sinauer Associates, Sunderland, Massachusetts.

Szczepanek SM, Boccaccio M, Pflaum K, Liao X, Geary SJ. 2014. Hydrogen peroxide production from glycerol metabolism is dispensable for the virulence of *Mycoplasma gallisepticum* in the tracheas of chickens. *Infection and Immunity* **82**(12): 4915.

Tadokoro K, Koizumi Y, Miyagi Y, Kojima Y, Kawamoto S, Hamajima K, Okuda K, Tanaka S, Onari K, Wahren B. 2001. Rapid and wide-reaching delivery of HIV-1 env DNA vaccine by intranasal administration. *Viral Immunology* **14**(2): 159.

Tanabe M, Atkins HS, Harland DN, Elvin SJ, Stagg AJ, Mirza O, Titball RW, Byrne B, Brown KA. 2006. The ABC transporter protein OppA provides protection against experimental *Yersinia pestis* infection. *Infection and Immunity* **74**(6): 3687.

Tavel JA, Martin JE, Kelly GG, Enama ME, Shen JM, Gomez PL, Andrews CA, Koup RA, Bailer RT, Stein JA *et al.* 2007. Safety and immunogenicity of a Gag-Pol candidate HIV-1 DNA vaccine administered by a needle-free device in HIV-1-seronegative subjects. *Journal of Acquired Immune Deficiency Syndromes* **44**(5): 601.

Thiaucourt F, Manso-Silvan L, Salah W, Barbe V, Vacherie B, Jacob D, Breton M, Dupuy V, Lomenech AM, Blanchard A *et al.* 2011. *Mycoplasma mycoides*, from "mycoides small colony" to "capri". A microevolutionary perspective. *BMC Genomics* **12**: 114.

Toffan A, Olivier A, Mancin M, Tuttoilmondo V, Facco D, Capua I, Terregino C. 2010. Evaluation of different serological tests for the detection of antibodies against highly pathogenic avian influenza in experimentally infected ostriches (*Struthio camelus*). *Avian Pathology* **39**(1): 11.

Tonheim TC, Bøgwald J, Dalmo RA. 2008. What happens to the DNA vaccine in fish? A review of current knowledge. *Fish & Shellfish Immunology* **25**(1): 1.

Treangen TJ, Salzberg SL. 2011. Repetitive DNA and next-generation sequencing: Computational challenges and solutions. *Nature Reviews Genetics* **13**(1): 36.

Trivedi S. 2010. Do simple sequence repeats in replication, repair and recombination genes of mycoplasmas providing genetic variability? *Journal of Cell and Molecular Biology* **7**(2): 53.

Tsai CJ, Sauna ZE, Kimchi-Sarfaty C, Ambudkar SV, Gottesman MM, Nussinov R. 2008. Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. *Journal of Molecular Biology* **383**(2): 281.

Tu AH, Voelker LL, Shen X, Dybvig K. 2001. Complete nucleotide sequence of the mycoplasma virus P1 genome. *Plasmid* **45**(2): 122.

Tuomela M, Malm M, Wallen M, Stanescu I, Krohn K, Peterson P. 2005. Biodistribution and general safety of a naked DNA plasmid, GTU®-MultiHIV, in a rat, using a quantitative PCR method. *Vaccine* **23**(7): 890.

Unnikrishnan M, Rappuoli R, Serruto D. 2012. Recombinant bacterial vaccines. *Current Opinion in Immunology* **24**(3): 337.

Uphoff CC, Drexler HG. 2014. Detection of mycoplasma contamination in cell cultures. *Current Protocols in Molecular Biology* **106**: 28.4.

Uttenthal A, Parida S, Rasmussen TB, Paton DJ, Haas B, Dundon WG. 2010. Strategies for differentiating infection in vaccinated animals (DIVA) for foot-and-mouth disease, classical swine fever and avian influenza. *Expert Review of Vaccines* **9**(1): 73.

van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. 2014. Ten years of next-generation sequencing technology. *Trends in Genetics* **30**(9): 418.

Van Helden LS, Grewar JD, Visser D, Dyason E, Koen P. 2012. An outbreak of highly pathogenic avian influenza in domestic ostriches: The current situation in South Africa. In *Proceedings of the 10th annual congress of the Southern African society for veterinary epidemiology and preventive medicine*, pp. 11-12.

Van Veen HW, Konings WN. 1998. The ABC family of multidrug transporters in microorganisms. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **1365**(1): 31.

Vandermeulen G, Marie C, Scherman D, Préat V. 2011. New generation of plasmid backbones devoid of antibiotic resistance marker for gene therapy trials. *Molecular Therapy* **19**(11): 1942.

Vasconcelos AT, Ferreira HB, Bizarro CV, Bonatto SL, Carvalho MO, Pinto PM, Almeida DF, Almeida LG, Almeida R, Alves-Filho L *et al.* 2005. Swine and poultry pathogens: The complete genome sequences of two strains of *Mycoplasma hyopneumoniae* and a strain of *Mycoplasma synoviae*. *Journal of Bacteriology* **187**(16): 5568.

Vaughan EE, Dean DA. 2006. Intracellular trafficking of plasmids during transfection is mediated by microtubules. *Molecular Therapy* **13**(2): 422.

Verwoerd DJ. 2000. Ostrich diseases. *Revue scientifique et technique (International Office of Epizootics)* **19**(2): 638.

Vilei EM, Frey J. 2001. Genetic and biochemical characterization of glycerol uptake in *Mycoplasma mycoides* subsp *mycoides* SC: Its impact on H_2O_2 production and virulence. *Clinical and Diagnostic Laboratory Immunology* **8**(1): 85.

Vilei EM, Nicolet J, Frey J. 1999. IS1634, a novel insertion element creating long, variable-length direct repeats which is specific for *Mycoplasma mycoides* subsp. *mycoides* small-colony type. *Journal of Bacteriology* **181**(4): 1319.

Villarreal I, Maes D, Vranckx K, Calus D, Pasmans F, Haesebrouck F. 2011. Effect of vaccination of pigs against experimental infection with high and low virulence *Mycoplasma hyopneumoniae* strains. *Vaccine* **29**(9): 1731.

Voelker LL, Weaver KE, Ehle LJ, Washburn LR. 1995. Association of lysogenic bacteriophage MAV1 with virulence of *Mycoplasma arthritidis*. *Infection and Immunity* **63**(10): 4016.

Volokhov DV, Simonyan V, Davidson MK, Chizhikov VE. 2012. RNA polymerase beta subunit (*rpoB*) gene and the 16S–23S rRNA intergenic transcribed spacer region (ITS) as complementary molecular markers in addition to the *16S rRNA* gene for phylogenetic analysis and identification of the species of the family *Mycoplasmataceae*. *Molecular Phylogenetics and Evolution* **62**(1): 515.

VPN/04/2012-01(Revision_6.0). 2012. Standard for the requirements, registration, maintenance of registration and official control of ostrich compartments in South Africa. Department of Agriculture, Forestry and Fisheries, South Africa.

Vyas JM, Van der Veen AG, Ploegh HL. 2008. The known unknowns of antigen processing and presentation. *Nature Reviews Immunology* **8**(8): 607.

Waites KB, Lysnyansky I, Bebear C. 2014. Emerging antimicrobial resistance in mycoplasmas of humans andanimals. In *Mollicutes: Molecular Biology and Pathogenesis* (ed. GF Browning, C Citti), pp. 289-322. Caister Academic Press, Norfolk, UK.

Waites KB, Talkington DF. 2004. *Mycoplasma pneumoniae* and its role as a human pathogen. *Clinical Microbiology Reviews* **17**(4): 697.

Wang G, Pan L, Zhang Y. 2011. Approaches to improved targeting of DNA vaccines. *Human Vaccines* **7**(12): 1271.

Wang MD. 2012. Ostrich (*Struthio camelus*) chick survival to 6 months post hatch: Estimation of environmental and genetic parameters and the effect of imprinting, foster parenting and deliberate care. MSc Thesis, Stellenbosch University, South Africa.

Wang Z, Troilo PJ, Wang X, Griffiths TG, Pacchione SJ, Barnum AB, Harper LB, Pauley CJ, Niu Z, Denisova L. 2004. Detection of integration of plasmid DNA into host genomic DNA following intramuscular injection and electroporation. *Gene Therapy* **11**(8): 711.

Weber SS, Sant'Anna FH, Schrank IS. 2012. Unveiling *Mycoplasma hyopneumoniae* promoters: Sequence definition and genomic distribution. *DNA Research* **19**(2): 103.

Weisburg WG, Tully JG, Rose DL, Petzel JP, Oyaizu H, Yang D, Mandelco L, Sechrest J, Lawrence TG, Van Etten J *et al.* 1989. A phylogenetic analysis of the mycoplasmas: Basis for their classification. *Journal of Bacteriology* **171**(12): 6455.

Westberg J, Persson A, Holmberg A, Goesmann A, Lundeberg J, Johansson KE, Pettersson B, Uhlen M. 2004. The genome sequence of *Mycoplasma mycoides* subsp. *mycoides* SC type strain PG1(T), the causative agent of contagious bovine pleuropneumonia (CBPP). *Genome Research* **14**(2): 221.

Wiley EO, Lieberman BS. 2011a. Parametric phylogenetics. In *Phylogenetics: Theory and Practice of Phylogenetic Systematics, Second Edition*, pp. 203-228. John Wiley & Sons, USA.

Wiley EO, Lieberman BS. 2011b. Parsimony and parsimony analysis. In *Phylogenetics: Theory and Practice of Phylogenetic Systematics, Second Edition*, pp. 152-202. John Wiley & Sons, USA.

Williams JA. 2013. Vector design for improved DNA vaccine efficacy, safety and production. *Vaccines* **1**(3): 225.

Wium M, Botes A, Bellstedt DU. 2015. The identification of *oppA* gene homologues as part of the oligopeptide transport system in mycoplasmas. *Gene* **558**(1): 31.

Woese CR, Maniloff J, Zablen LB. 1980. Phylogenetic analysis of the mycoplasmas. *Proceedings of the National Academy of Sciences* **77**(1): 494.

Woese CR, Stackebrandt E, Ludwig W. 1984. What are mycoplasmas: The relationship of tempo and mode in bacterial evolution. *Journal of Molecular Evolution* **21**(4): 305.

Wolf M, Müller T, Dandekar T, Pollack JD. 2004. Phylogeny of *Firmicutes* with special reference to *Mycoplasma* (*Mollicutes*) as inferred from phosphoglycerate kinase amino acid sequence data. *International Journal of Systematic and Evolutionary Microbiology* **54**: 871.

Wolff JA, Malone RW, Williams P, Chong W, Acsadi G, Jani A, Felgner PL. 1990. Direct gene-transfer into mouse muscle *in vivo*. *Science* **247**(4949): 1465.

Xu Y, Yuen PW, Lam JK. 2014. Intranasal DNA vaccine for protection against respiratory infectious diseases: The delivery perspectives. *Pharmaceutics* **6**(3): 378.

Yang M, Johnson A, Murphy TF. 2011. Characterization and evaluation of the *Moraxella catarrhalis* oligopeptide permease A as a mucosal vaccine antigen. *Infection and Immunity* **79**(2): 846.

Yang Z, Rannala B. 2012. Molecular phylogenetics: Principles and practice. *Nature Reviews Genetics* **13**(5): 303.

Yin Y, Zhang H, Olman V, Xu Y. 2010. Genomic arrangement of bacterial operons is constrained by biological pathways encoded in the genome. *Proceedings of the National Academy of Sciences* **107**(14): 6310.

Yogev D, Rosengarten R, Watson-McKown R, Wise K. 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. *The EMBO Journal* **10**(13): 4069.

Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ *et al.* 2010. PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26**(13): 1608.

Zhang F, Saha S, Shabalina SA, Kashina A. 2010. Differential arginylation of actin isoforms is regulated by coding sequence–dependent degradation. *Science* **329**(5998): 1534.

Zhang Q, Wise KS. 2001. Coupled phase-variable expression and epitope masking of selective surface lipoproteins increase surface phenotypic diversity in *Mycoplasma hominis*. *Infection and Immunity* **69**(8): 5177.

Zhao F, Lv M, Tao X, Huang H, Zhang B, Zhang Z, Zhang J. 2012. Antibiotic sensitivity of 40 *Mycoplasma pneumoniae* isolates and molecular analysis of macrolide-resistant isolates from Beijing, China. *Antimicrobial Agents and Chemotherapy* **56**(2): 1108.

Zheng X, Liu S. 2008. The relaxing ori-ter balance of *Mycoplasma* genomes. *Science in China Series C: Life Sciences* **51**(2): 182.

Zhou JH, You YN, Chen HT, Zhang J, Ma LN, Ding YZ, Pejsak Z, Liu YS. 2013. The effects of the synonymous codon usage and tRNA abundance on protein folding of the 3C protease of foot-and-mouth disease virus. *Infection, Genetics and Evolution* **16**: 270.

Zimmerman CU. 2014. Current insights into phase and antigenic variation in mycoplasmas. In *Mollicutes: Molecular Biology and Pathogenesis* (ed. GF Browning, C Citti), p. 165. Caister Academic Press, Norfolk, UK.

Appendix 1 A inventory of *Mollicutes* species with an available complete genome sequence and their genome characteristics

Species	Strain	IMG Genome ID	Release Date	Host Name	Genome Size	Gene Count	GC	CDS Count	CDS %	rRNA Count	tRNA Count	w/o function prediction %
Order: *Acholeplasmatales*	**Family: *Acholeplasmataceae***		**Genus: *Acholeplasma***									
Acholeplasma laidlawii	PG-8A	641522601	01/08/2008		1 496 992	1 422	0.32	1 380	97.05	6	36	25.74
Order: *Acholeplasmatales*	**Family: *Acholeplasmataceae***		**Genus: *Candidatus* Phytoplasma**									
Aster yellows witches'-broom phytoplasma	AY-WB	2606217303	10/03/2015	Plant: Lettuce	723 970	706	0.27	665	94.19	6	32	32.58
Candidatus Phytoplasma australiense		642555117	01/12/2008		879 959	727	0.27	684	94.09	6	35	37.55
	NZSb11	2554235337	01/01/2014		959 779	1 155	0.27	1 114	96.45	6	35	45.97
Candidatus Phytoplasma mali	AT	642555118	01/12/2008		601 943	518	0.21	479	92.47	6	32	29.54
Candidatus Phytoplasma solani	284/09	2558309089	26/02/2014		570 238	547	0.28	520	95.06	-	27	23.58
Onion yellows phytoplasma	OY-M	2606217364	10/03/2015		853 092	977	0.28	935	95.70	6	32	39.00
Order: *Entomoplasmatales*	**Family: *Entomoplasmataceae***		**Genus: *Mesoplasma***									
Mesoplasma florum	L1	2606217385	10/03/2015	Plant: Solidago	793 224	721	0.27	682	94.59	6	29	19.00
	W37	2558309067	26/02/2014	Plant: Solidago	825 824	768	0.27	733	95.44	6	29	20.05
Order: *Entomoplasmatales*	**Family: *Spiroplasmataceae***		**Genus: *Spiroplasma***									
Spiroplasma apis	B31	2563366575	14/04/2014		1 160 554	1 030	0.28	998	96.89	3	29	30.39
Spiroplasma chrysopicola	DF-1	2554235371	01/01/2014		1 123 322	1 051	0.29	1 015	96.57	3	33	30.54
Spiroplasma culicicola	AES-1	2558860239	07/03/2014	Mosquito	1 175 131	1 103	0.26	1 071	97.10	3	29	32.64
Spiroplasma diminutum	CUAS-1	2554235383	01/01/2014	Mosquito	945 296	890	0.25	858	96.40	3	29	27.75
Spiroplasma mirum	SMCA	2558860251	07/03/2014	Rabbit tick	1 132 608	1 422	0.29	1 386	97.47	3	33	51.76
Spiroplasma sabaudiense	Ar-1343	2558860238	07/03/2014	Mosquito	1 075 953	967	0.30	931	96.28	6	30	30.40
Spiroplasma syrphidicola	EA-1	2554235381	01/01/2014	Hoverfly	1 107 344	1 044	0.29	1 009	96.65	3	32	31.99
Spiroplasma taiwanense	CT-1	2561511192	23/03/2014	Mosquito	1 086 278	1 093	0.24	1 061	97.07	3	29	39.98
Order: *Mycoplasmatales*	**Family: *Mycoplasmataceae***		**Genus: *Mycoplasma***									
Candidatus Mycoplasma haemolamae	Purdue	2521172720	01/02/2013	Alpaca	756 845	961	0.39	925	96.25	3	33	65.87
Mycoplasma agalactiae	5632	2540341079	19/09/2013		1 006 702	866	0.30	825	95.27	6	34	20.67
	PG2	646564553	01/08/2010		1 006 702	854	0.30	813	95.20	6	34	43.09
Mycoplasma arthritidis	158L3-1	642555141	01/12/2008	Human	820 453	666	0.31	631	94.74	3	32	34.38
Mycoplasma bovis	CQ-W70	2576861438		Cattle	948 516	816	0.29	778	95.34	4	34	16.42
	Donetta PG45	649633071	01/07/2011	Cattle	1 003 404	806	0.29	765	94.91	6	34	14.27
	HB0801	2521172704	01/02/2013	Cattle	991 702	849	0.29	809	95.29	6	34	13.90
	Hubei-1	650716062	01/12/2011	Cattle	948 121	839	0.29	801	95.47	4	34	35.40
Mycoplasma bovoculi	M165/69	2558860179	07/03/2014	Cattle	760 240	625	0.28	591	94.56	3	30	18.40
Mycoplasma californicum	ST-6	2576861462		Cattle	793 841	672	0.31	635	94.49	6	31	19.94
Mycoplasma capricolum	ATCC 27343	2606217281	10/03/2015	Goat	1 010 023	875	0.24	835	95.43	6	30	22.74
Mycoplasma conjunctivae	HRC/581	644736392	01/12/2009	Sheep	846 214	725	0.29	691	95.31	3	29	42.62
Mycoplasma crocodyli	MP145	646564554	01/08/2010	Crocodile	934 379	731	0.27	689	94.25	6	34	25.58
Mycoplasma cynos	C142	2540341156	19/09/2013	Dog	998 123	891	0.26	883	99.10	8	-	26.15
Mycoplasma fermentans	JER	648028044	01/01/2011	Human	977 524	838	0.27	797	95.11	5	36	35.68
	M64	649633072	01/07/2011	Human	1 118 751	1 091	0.27	1 050	96.24	5	36	43.54

Species	Strain	IMG Genome ID	Release Date	Host Name	Genome Size	Gene Count	GC	CDS Count	CDS %	rRNA Count	tRNA Count	w/o function prediction %
Mycoplasma gallisepticum		2517093027	06/09/2012		986 257	809	0.32	771	95.30	6	32	20.02
	CA06_2006.052-5-2P	2517093024	06/09/2012		976 412	801	0.32	763	95.26	6	32	19.85
	F	646862333	01/08/2010		977 612	795	0.31	756	95.09	7	32	25.53
	NC06_2006.080-5-2P	2518645544	04/12/2012	House finch	938 869	782	0.32	744	95.14	6	32	19.95
	NC08_2008.031-4-3P	2517093025	06/09/2012		926 650	777	0.32	739	95.11	6	32	19.95
	NC95_13295-2-2P	2518645562	04/12/2012		953 989	792	0.32	754	95.20	6	32	20.33
	NY01_2001.047-5-1P	2518645563	04/12/2012		965 525	798	0.32	760	95.24	6	32	20.18
	R(high)	646862334	01/08/2010		1 012 027	805	0.31	766	95.16	7	32	24.97
	R(low)	2606217644	10/03/2015		1 012 800	826	0.31	788	95.40	4	32	25.30
	S6	2531839059	23/08/2013	Chicken	929 411	757	0.31	720	95.11	4	33	20.74
	VA94_7994-1-7P	2518645561	04/12/2012		964 110	805	0.32	767	95.28	6	32	20.50
	WI01_2001.043-13-2P	2518645557	04/12/2012		939 844	784	0.32	746	95.15	6	32	20.28
Mycoplasma genitalium	G37	2606217665	10/03/2015	Human	580 076	563	0.32	521	92.54	3	36	17.76
	M2288	2540341178	19/09/2013	Human	579 558	545	0.32	506	92.84	3	36	14.31
	M2321	2512564088	31/01/2014	Human	579 977	588	0.32	547	93.03	3	36	20.41
	M6282	2540341080	19/09/2013	Human	579 504	523	0.32	484	92.54	3	36	14.72
	M6320	2540341081	19/09/2013	Human	579 796	548	0.32	509	92.88	3	36	14.23
Mycoplasma haemocanis	Illinois	2511231066	28/02/2012	Dog	919 992	1 190	0.35	1 156	97.14	3	31	71.60
Mycoplasma haemofelis	Langford 1	649633073	01/07/2011	Cat	1 147 259	1 580	0.39	1 545	97.78	3	31	79.49
	Ohio2	651053045	01/12/2011		1 155 937	1 561	0.39	1 527	97.82	3	31	78.99
Mycoplasma hominis	PG21, ATCC 23114	646311946	01/04/2010	Human	665 445	563	0.27	523	92.90	7	33	33.04
Mycoplasma hyopneumoniae	168	650377960	01/07/2011	Pig	925 576	728	0.28	695	95.47	3	30	15.25
	168-L	2554235345	01/01/2014	Pig	921 093	726	0.28	693	95.45	3	30	27.55
	232	2606217661	10/03/2015		892 758	713	0.29	677	94.95	3	30	30.43
	7422	2554235315	01/01/2014	Pig	898 495	711	0.29	677	95.22	3	30	27.29
	7448	2606217470	10/03/2015	Pig	920 079	722	0.28	686	95.01	3	30	29.92
	J	2606217353	10/03/2015		897 405	728	0.29	690	94.78	3	30	29.95
Mycoplasma hyorhinis	DBS 1050	2554235454	01/01/2014		837 447	782	0.26	749	95.78	3	30	21.61
	GDL-1	2511231163	28/02/2012		837 480	740	0.26	707	95.54	3	30	28.38
	HUB-1	648028045	01/01/2011	Pig	839 615	687	0.26	654	95.20	3	30	23.44
	MCLD	2512047042	23/03/2012		829 709	811	0.26	778	95.93	3	30	27.37
	SK76	2519103106			836 897	786	0.26	753	95.80	3	30	21.88
Mycoplasma leachii	99/014/6	651053046	01/12/2011	Cattle	1 017 232	942	0.24	905	96.07	6	31	6.48
	PG50	649633074	01/07/2011	Cattle	1 008 951	922	0.24	882	95.66	6	30	25.27
Mycoplasma mobile	163K	637000180	01/12/2006		777 079	669	0.25	633	94.62	3	28	9.27
	163K	2606217309	10/03/2015		777 079	688	0.25	655	95.20	3	28	21.80
Mycoplasma mycoides	95010	650716063	01/12/2011		1 155 838	959	0.24	922	96.14	6	30	27.84
	GM12	646862335	01/08/2010		1 084 586	868	0.24	830	95.62	6	30	18.89
	GM12	646862336	01/08/2010		1 089 202	870	0.24	832	95.63	6	30	18.97
	Gladysdale	648231714	01/01/2011		1 193 808	1 134	0.24	1 095	96.56	6	30	25.93
	PG1	2606217549	10/03/2015		1 211 703	1 178	0.24	1 138	96.60	6	30	29.37
Mycoplasma ovis	Michigan	2558309058	26/02/2014		702 511	886	0.32	851	96.05	4	31	62.19
Mycoplasma parvum	Indiana	2554235465	01/01/2014		564 395	616	0.27	582	94.48	3	31	49.84
Mycoplasma penetrans	HF-2	2606217263	10/03/2015	Human	1 358 633	1 075	0.26	1 038	96.56	3	30	27.53

Species	Strain	IMG Genome ID	Release Date	Host Name	Genome Size	Gene Count	GC	CDS Count	CDS %	rRNA Count	tRNA Count	w/o function prediction %
Mycoplasma pneumoniae	309	2511231209	28/02/2012	Human	817 176	749	0.40	707	94.39	3	36	17.89
	FH	648231715	01/01/2011	Human	811 088	670	0.40	629	93.88	3	36	31.94
	M129-B7	2606217423	10/03/2015	Human	816 394	796	0.40	753	94.60	3	37	27.39
	M29	2597489969	08/12/2014	Human	857 799	837	0.40	794	94.86	3	37	27.24
Mycoplasma pulmonis	UAB CTIP	2606217467	10/03/2015		963 879	789	0.27	753	95.44	4	29	24.84
Mycoplasma putrefaciens	KS1	2511231058	28/02/2012	Goat	832 603	724	0.27	686	94.75	6	30	21.69
	Mput9231	2540341179	19/09/2013		859 996	746	0.27	709	95.04	6	30	27.21
Mycoplasma suis	Illinois	650716064	01/12/2011	Pig	742 431	879	0.31	844	96.02	3	32	62.80
	KI_3806	2511231073	28/02/2012	Pig	709 270	844	0.31	809	95.85	3	32	62.56
Mycoplasma synoviae	53	2606217384	10/03/2015		799 476	736	0.28	695	94.43	4	34	27.72
Mycoplasma wenyonii	Massachusetts	2517093037	06/09/2012	Cattle	650 228	687	0.34	652	94.91	3	32	50.22
Order: *Mycoplasmatales*	**Family: *Mycoplasmataceae***		**Genus: *Ureaplasma***									
Ureaplasma parvum		641522658	01/08/2008	Human	751 679	642	0.25	609	94.86	6	27	30.69
	ATCC 700970	2606217444	10/03/2015	Human	751 719	656	0.25	617	94.05	6	30	27.59
Ureaplasma urealyticum	Western	643348586	01/04/2009	Human	874 478	679	0.26	646	95.14	6	27	31.08

Data was downloaded from Integrated Microbial Genomes (IMG) system (http://www.jgi.doe.gov/) on 25 June 2015

Summary of the above genomes

Order	Family	Genus	Number of Species	Number of Genome*	Genome Size	Gene Count	GC	CDS count	CDS %	rRNA Count	tRNA Count	w/o function prediction %
Acholeplasmatales	***Acholeplasmataceae***	***Acholeplasma***	1	1	1 497	1 422	0.32	1 380	97.05	6	36	25.74
		Phytoplasma *Candidatus*	5	6	570-960	518-1 155	0.21-0.28	479-1 114	92.47-96.45	6	27-35	23.58-45.97
Entomoplasmatales	***Entomoplasmataceae***	***Mesoplasma***	1	2	793-825	721-768	0.27	682-733	94.59-95.44	6	29	19.00-20.05
	Spiroplasmataceae	***Spiroplasma***	8	8	945-1 175	890-1 422	0.24-0.30	858-1 386	96.28-97.47	3-6	29-33	27.75-51.76
Mycoplasmatales	***Mycoplasmataceae***	***Mycoplasma***	30	71	564-1 359	523-1 580	0.24-0.40	506-1 545	92.54-99.10	3-8	28-37	6.48-79.49
		Ureaplasma	2	3	752-874	642-679	0.25-0.26	609-646	94.05-95.14	6	27-30	27.59-31.08
Class: *Mollicutes*			**47**	**91**	**564-1 497**	**518-1 580**	**0.21-0.40**	**479-1 545**	**92.47-99.10**	**3-8**	**27-37**	**6.48-79.49**
*** In many cases more than one strain of the same species had a complete genome sequence available**												

Appendix 2 Supplementary tables and figures for Chapter 3

Supplementary Table 2.1 Contigs generated by the GS *de novo* assembler arrange from longest to shortest*

Contig	Length (bp)	Number of reads	Contig	Length (bp)	Number of reads
contig00239	39354	4190	contig00083	4229	558
contig00095	31157	2304	contig00206	4137	341
contig00059	26529	2537	contig00144	4125	216
contig00408	25616	1634	contig00343	4099	233
contig00350	24392	2473	contig00427	4059	251
contig00132	21804	1480	contig00198	3922	325
contig00135	21012	1729	contig00204	3897	256
contig00160	20872	2102	contig00088	3768	202
contig00061	17129	2026	contig00255	3753	170
contig00166	15961	819	contig00020	3618	211
contig00051	15263	1508	contig00267	3155	229
contig00316	14165	808	contig00433	3153	159
contig00311	14034	1003	contig00155	3089	190
contig00116	13910	923	contig00222	2966	110
contig00353	13542	915	contig00067	2955	177
contig00034	13149	631	contig00118	2906	155
contig00161	13130	1297	contig00465	2859	231
contig00180	12250	1026	contig00200	2856	214
contig00098	12239	1315	contig00077	2780	298
contig00013	11983	922	contig00097	2750	124
contig00163	10933	816	contig00091	2742	180
contig00142	10406	1062	contig00019	2724	119
contig00177	10160	1298	contig00371	2712	310
contig00070	10116	861	contig00085	2663	168
contig00285	9874	463	contig00004	2638	385
contig00148	9719	637	contig00189	2614	370
contig00389	9507	909	contig00196	2344	45
contig00191	9482	1026	contig00092	2248	327
contig00075	9466	531	contig00048	2201	69
contig00009	8637	1062	contig00322	2191	176
contig00058	8088	508	contig00209	2171	67
contig00068	7993	702	contig00257	2092	81
contig00159	7930	691	contig00169	2078	107
contig00011	7663	540	contig00175	2060	150
contig00046	7436	298	contig00219	1999	53
contig00130	7434	708	contig00474	1981	121
contig00450	7396	353	contig00165	1978	1054
contig00056	6932	222	contig00221	1944	79
contig00284	6856	475	contig00007	1938	59
contig00151	6529	740	contig00245	1892	181
contig00052	6515	329	contig00038	1884	67
contig00050	6462	388	contig00187	1859	60
contig00014	6425	745	contig00253	1845	138
contig00117	6219	348	contig00317	1811	460
contig00212	6173	251	contig00185	1796	111
contig00152	6137	571	contig00030	1731	56
contig00006	6028	268	contig00018	1719	42
contig00031	5819	479	contig00054	1677	221
contig00262	5683	327	contig00251	1658	49
contig00274	5663	641	contig00261	1601	70
contig00258	5545	157	contig00381	1576	276
contig00218	5293	298	contig00352	1545	75
contig00467	5147	528	contig00190	1532	235
contig00002	4807	161	contig00210	1514	52
contig00364	4675	241	contig00233	1476	103
contig00199	4521	253	contig00005	1419	64
contig00237	4423	277	contig00432	1349	294
contig00178	4323	250	contig00125	1317	65
contig00133	4272	353	contig00357	1312	63
contig00220	4231	235	contig00032	1276	56

Contig	Length (bp)	Number of reads	Contig	Length (bp)	Number of reads
contig00081	1219	313	contig00482	444	68
contig00354	1182	85	contig00057	440	7
contig00036	1176	97	contig00229	439	12
contig00359	1176	276	contig00250	437	6
contig00269	1153	58	contig00044	426	6
contig00129	1133	181	contig00137	424	137
contig00039	1124	39	contig00410	422	39
contig00053	1122	14	contig00265	421	40
contig00291	1089	74	contig00275	419	23
contig00468	1055	54	contig00028	410	8
contig00043	1050	20	contig00485	409	24
contig00033	1004	26	contig00197	405	148
contig00292	977	141	contig00490	405	18
contig00186	926	88	contig00402	404	127
contig00248	910	17	contig00386	398	233
contig00042	897	17	contig00012	392	15
contig00379	881	38	contig00158	392	98
contig00146	846	17	contig00066	390	212
contig00045	806	12	contig00141	387	145
contig00041	805	24	contig00451	373	16
contig00035	797	4	contig00367	370	174
contig00241	796	23	contig00404	369	286
contig00214	786	22	contig00243	367	18
contig00182	778	18	contig00242	361	11
contig00349	769	137	contig00103	358	13
contig00208	768	26	contig00211	356	130
contig00195	750	42	contig00266	353	10
contig00368	733	132	contig00476	344	78
contig00089	727	46	contig00104	342	15
contig00079	726	345	contig00029	331	3
contig00240	719	28	contig00150	331	32
contig00156	705	171	contig00145	327	12
contig00188	696	49	contig00127	325	53
contig00207	687	22	contig00162	324	9
contig00121	657	50	contig00399	320	4
contig00510	643	84	contig00437	315	137
contig00064	621	262	contig00430	308	113
contig00037	617	11	contig00139	305	132
contig00082	614	143	contig00015	297	15
contig00179	607	136	contig00181	297	10
contig00026	589	67	contig00331	295	36
contig00078	582	255	contig00010	290	51
contig00306	570	19	contig00469	290	33
contig00049	561	25	contig00710	281	103
contig00140	556	71	contig00143	278	12
contig00264	547	56	contig00297	276	12
contig00157	546	157	contig00080	275	175
contig00047	540	5	contig00246	270	7
contig00124	527	139	contig00496	266	64
contig00084	515	84	contig00409	265	471
contig00023	504	30	contig00477	263	42
contig00205	500	60	contig00380	256	7
contig00183	492	19	contig00238	252	3
contig00016	488	18	contig00471	251	102
contig00120	488	24	contig00351	242	112
contig00055	474	11	contig00090	240	51
contig00217	469	44	contig00119	235	12
contig00027	467	9	contig00147	231	12
contig00236	465	47	contig00279	229	30
contig00001	462	34	contig00230	222	2
contig00390	459	116	contig00472	221	18
contig00405	453	204	contig00065	215	182
contig00247	444	12	contig00268	214	13

Contig	Length (bp)	Number of reads	Contig	Length (bp)	Number of reads
contig00231	198	2	contig00716	136	74
contig00176	198	54	contig00303	134	7
contig00069	195	202	contig00022	132	11
contig00101	190	26	contig00428	132	51
contig00366	189	151	contig00076	132	39
contig00460	188	6	contig00107	129	2
contig00416	186	649	contig00192	140	280
contig00270	183	172	contig00713	140	17
contig00213	182	27	contig00168	127	29
contig00415	181	6	contig00003	126	10
contig00363	176	9	contig00707	126	7
contig00324	173	65	contig00310	122	16
contig00108	166	1432	contig00733	120	115
contig00374	166	39	contig00714	120	20
contig00711	165	203	contig00479	119	124
contig00588	164	814	contig00730	118	1142
contig00466	164	99	contig00260	116	30
contig00344	162	3	contig00294	116	10
contig00223	159	7	contig00216	113	13
contig00226	154	2	contig00300	112	10
contig00228	151	14	contig00376	110	29
contig00328	151	44	contig00319	110	9
contig00184	151	77	contig00330	108	1723
contig00346	150	25	contig00272	108	32
contig00295	149	160	contig00340	108	61
contig00735	148	370	contig00273	106	5
contig00128	146	94	contig00280	105	16
contig00301	146	11	contig00715	103	258
contig00480	146	97	contig00024	102	25
contig00734	144	19	contig00481	102	58
contig00335	142	7	contig00123	101	19
contig00235	141	110	contig00731	101	103
contig00400	141	6	contig00017	100	5
contig00040	140	3	contig00448	100	452

* Contigs from the final assemble with 25.2 Mb of 454 data

Supplementary Table 2.2 List of the IGS and RAST annotate genes gene within the Ms03 draft genome

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_1	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	28	157	+	Unknown: General						
mnas_2	endonuclease/Exonuclease/phosphatase family protein			270	2244	+	Unknown: Enzymes of unknown specificity	**peg.214**	Membrane nuclease, lipoprotein	- none -	271	2244	+
mnas_3	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	2191	2392	+	Unknown: General						
								peg.215	hypothetical protein	- none -	2406	2287	-
mnas_4	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	2457	3255	+	Unknown: General	**peg.216**	hypothetical protein	- none -	2458	3255	+
mnas_5	putative lipoprotein		GO:0008150, GO:0003674, GO:0016020	3284	6431	-	Cell envelope: Other	**peg.217**	hypothetical protein	- none -	6431	3285	-
mnas_6	metallo-beta-lactamase superfamily protein			6674	8564	+	Unknown: General Hypothetical	**peg.218**	Ribonuclease J1 (endonuclease and 5' exonuclease)	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: Ribonucleases in Bacillus	6675	8564	+
mnas_7	M42 glutamyl aminopeptidase family protein			8572	9652	+	Protein fate: Degradation of proteins, peptides, and glycopeptides	**peg.219**	Endo-1,4-beta-glucanase	- none -	8573	9652	+
mnas_8	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	9672	9909	+	Unknown: Conserved	**peg.220**	*M. genitalium* predicted coding region MG335.1	- none -	9673	9909	+
mnas_9	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	10188	10656	+	Unknown: General						
mnas_10	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	10657	10990	+	Unknown: General	**peg.221**	hypothetical protein	- none -	10658	10990	+
mnas_11	Putative potassium uptake protein KtrB	*ktrB*		11069	12809	+	Transport and binding proteins: Unknown substrate	**peg.222**	Potassium uptake protein, integral membrane component, KtrB	- none -	11124	12809	+
mnas_12	trkA-C domain protein	*ktrA*		12815	13490	+	Unknown: General Hypothetical	**peg.223**	Trk system potassium uptake protein TrkA	**Category**: Potassium metabolism **Subcategory**: Potassium metabolism **Subsystem**: Potassium homeostasis	12816	13490	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_13	DNA gyrase, A subunit (EC 5.99.1.3)	*gyrA*	GO:0003916, GO:0003918, GO:0006265, GO:0009330	13588	16267	-	DNA metabolism: DNA replication, recombination, and repair	**peg.224**	DNA gyrase subunit A (EC 5.99.1.3)	**Category**: DNA Metabolism **Subcategory**: DNA replication **Subsystem**: DNA topoisomerases, Type II, ATP-dependent **Category**: Virulence, Disease and Defense **Subcategory**: Resistance to antibiotics and toxic compounds **Subsystem**: Resistance to fluoroquinolones **Subsystem**: Cell Division Subsystem including YidCD	16267	13589	-
mnas_14	ribosomal large subunit pseudouridine synthase B (EC 5.4.99.22)	*rluB*	GO:0003723, GO:0006364, GO:0001522, GO:0009982	16366	17083	-	Cellular processes: Other, Transport and binding proteins: Unknown substrate, Unknown: Enzymes of unknown specificity	**peg.225**	Ribosomal large subunit pseudouridine synthase B (EC 4.2.1.70)	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: RNA pseudouridine syntheses	17083	16367	-
mnas_15	nicotinate (nicotinamide) nucleotide adenylyltransferase (EC 2.7.7.18)	*nadD*	GO:0000309, GO:0004515, GO:0005737, GO:0009435	17082	18165	-	Biosynthesis of cofactors, prosthetic groups, and carriers: Pyridine nucleotides	**peg.226**	Nicotinate-nucleotide adenylyltransferase (EC 2.7.7.18) / Hydrolase (HAD superfamily), YqeK	**Category**: Cofactors, Vitamins, Prosthetic Groups, Pigments **Subcategory**: NAD and NADP **Subsystem**: NAD and NADP cofactor biosynthesis global	18165	17083	-
mnas_16	Damage-repair DNA polymerase IV (EC 2.7.7.7)	*dinB*		18170	19400	-	DNA metabolism: DNA replication, recombination, and repair	**peg.227**	DNA polymerase IV (EC 2.7.7.7)	**Category**: DNA Metabolism **Subcategory**: DNA repair **Subsystem**: DNA repair, bacterial	19400	18171	-
mnas_17	nifU-like N terminal domain protein			19399	19810	-	Unknown: General Hypothetical	**peg.228**	Putative iron-sulfur cluster assembly scaffold protein for SUF system, SufE2	- none -	19810	19400	-
mnas_18	aminotransferase class-V family protein			19799	20963	-	Unknown: Enzymes of unknown specificity	**peg.229**	Cysteine desulfurase (EC 2.8.1.7), SufS subfamily	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: mnm5U34 biosynthesis bacteria **Category**: Amino Acids and Derivatives **Subcategory**: Alanine, serine, and glycine **Subsystem**: Alanine biosynthesis	20963	19800	-
mnas_19	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	20969	21491	-	Unknown: General	**peg.230**	hypothetical protein	- none -	21368	20970	-
mnas_20	nusB family protein			21471	21954	-	Unclassified: Role category not yet assigned	**peg.231**	Transcription termination protein NusB	**Category**: RNA Metabolism **Subcategory**: Transcription **Subsystem**: Transcription factors bacterial	21954	21472	-

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_21	putative membrane protein		GO:0008150, GO:0003674, GO:0016020	22078	23686	-	Cell envelope: Other	**peg.232**	hypothetical protein	- none -	23686	22079	-
mnas_22	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	23764	24154	-	Unknown: General	**peg.233**	hypothetical protein	- none -	24154	23765	-
mnas_23	translation elongation factor P	*efp*	GO:0003746, GO:0005737, GO:0006414	24304	24868	+	Protein synthesis: Translation factors	**peg.234**	Translation elongation factor P	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Translation elongation factors bacterial	24305	24868	+
mnas_24	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	24867	25179	+	Unknown: General	**peg.235**	hypothetical protein	- none -	24868	25179	+
mnas_25	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	25178	25508	+	Unknown: General	**peg.236**	Expressed protein	- none -	25521	26450	+
mnas_26	putative membrane protein		GO:0008150, GO:0003674, GO:0016020	25520	26450	+	Cell envelope: Other						
mnas_27	putative nicotinate phosphoribosyltransferase		GO:0008152, GO:0003824, GO:0004516, GO:0016757, GO:0019363, GO:0004514, GO:0016740, GO:0009435	26439	27447	+	Cellular processes: Other, Unknown: Enzymes of unknown specificity	**peg.237**	Nicotinate phosphoribosyltransferase (EC 2.4.2.11)	**Category**: Cofactors, Vitamins, Prosthetic Groups, Pigments **Subcategory**: NAD and NADP **Subsystem**: NAD and NADP cofactor biosynthesis global	26440	27447	+
mnas_28	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	27517	28297	+	Unknown: General	**peg.238**	Mobile element protein	- none -	27518	28297	+
mnas_29	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	28310	29933	-	Unknown: General	**peg.239**	hypothetical protein	- none -	29933	28311	-
mnas_30	putative lipoprotein		GO:0008150, GO:0003674, GO:0016020	29898	33624	-	Cell envelope: Other	**peg.240**	hypothetical protein	- none -	33624	29899	-
mnas_31	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	33705	35799	-	Unknown: General	**peg.241**	hypothetical protein	- none -	35799	33706	-
mnas_32	peptidase M60-like family protein			35900	37121	-	Protein fate: Degradation of proteins, peptides, and glycopeptides	**peg.242**	Integral membrane protein (Rhomboid family)	- none -	37118	35901	-
mnas_33	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	37086	38961	-	Unknown: General	**peg.243**	hypothetical protein	- none -	38961	37087	-
mnas_34	DNA methylase family protein			39414	40404	+	DNA metabolism: Restriction/modification	**peg.414**	Type III restriction-modification system methylation subunit (EC 2.1.1.72)	**Category**: DNA Metabolism **Subcategory**: no subcategory **Subsystem**: Restriction-Modification System	39496	40404	+

IGS								RAST					
Gene_id	Common_name	Symbol[1]	GO terms[2]	fmin[3]	fmax[3]	S[4]	TIGR_roles	Gene_id	Function	Subsystem	Start[5]	Stop[5]	S[4]
mnas_35	putative type III restricction-modification system: methylase		GO:0003677, GO:0032775, GO:0008170, GO:0016740, GO:0032259, GO:0006306, GO:0009007	40506	40806	+	Cellular processes: Other, Unknown: Enzymes of unknown specificity	**peg.415**	Type III restriction-modification system methylation subunit (EC 2.1.1.72)	**Category**: DNA Metabolism **Subcategory**: no subcategory **Subsystem**: Restriction-Modification System	40507	40806	+
mnas_36	DEAD/DEAH box helicase family protein			40807	43450	+	Unknown: Enzymes of unknown specificity	**peg.416**	Type III restriction-modification system DNA endonuclease res (EC 3.1.21.5)	**Category**: DNA Metabolism **Subcategory**: no subcategory **Subsystem**: Restriction-Modification System	40808	43450	+
mnas_37	helix-turn-helix family protein			43698	44694	+	Regulatory functions: DNA interactions	**peg.417**	Sucrose operon repressor ScrR, LacI family	- none -	43720	44694	+
mnas_38	Putative potassium channel protein			44758	45874	+	Transport and binding proteins: Cations and iron carrying compounds	**peg.418**	putative potassium channel protein	- none -	44759	45874	+
mnas_39	putative endonuclease 4		GO:0003677, GO:0006281, GO:0008270, GO:0005622, GO:0008833	45866	46694	+	Cellular processes: Other, Unknown: Enzymes of unknown specificity	**peg.419**	Endonuclease IV (EC 3.1.21.2)	**Category**: DNA Metabolism **Subcategory**: DNA repair **Subsystem**: DNA repair, bacterial	45867	46694	+
mnas_40	ribosomal protein L33	*rpmG*	GO:0000315, GO:0003735, GO:0006412, GO:0022625	46776	46929	+	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.420**	LSU ribosomal protein L33p @ LSU ribosomal protein L33p, zinc-dependent	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial	46777	46929	+
mnas_41	metallopeptidase M24 family protein			46989	48051	+	Protein fate: Degradation of proteins, peptides, and glycopeptides	**peg.421**	Aminopeptidase YpdF (MP-, MA-, MS-, AP-, NP- specific)	**Category**: Protein Metabolism **Subcategory**: Protein degradation **Subsystem**: Protein degradation	46990	48051	+
mnas_42	prolyl aminopeptidase (EC 3.4.11.5)	*pip*	GO:0016804, GO:0030163	48034	48985	+	Protein fate: Degradation of proteins, peptides, and glycopeptides	**peg.422**	Proline iminopeptidase (EC 3.4.11.5)	- none -	48035	48985	+
mnas_43	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	49099	49483	-	Unknown: Conserved						
mnas_44	tRNA-Asn			49170	49245	+		**rna.15**	tRNA-Asn-GTT	- none -	49171	49242	+
mnas_45	tRNA-Glu			49251	49327	+		rna.16	tRNA-Glu-TTC	- none -	49252	49324	+
mnas_46	tRNA-Val			49332	49408	+		**rna.17**	tRNA-Val-TAC	- none -	49333	49405	+
mnas_47	tRNA-Thr			49409	49485	+		**rna.18**	tRNA-Thr-TGT	- none -	49410	49482	+
mnas_48	tRNA-Leu			49520	49604	+		**rna.19**	tRNA-Leu-TAG	- none -	49521	49601	+
mnas_49	ECF-type riboflavin transporter, S component family protein			49839	50841	+	Transport and binding proteins: Unknown substrate	**peg.423**	Substrate-specific component FolT of folate ECF transporter	**Category**: Cofactors, Vitamins, Prosthetic Groups, Pigments **Subcategory**: Folate and pterines **Subsystem**: Folate biosynthesis	49840	50841	+
mnas_50	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	50851	53320	+	Unknown: General	**peg.424**	oligopeptide ABC transporter ATP-binding protein	- none -	50852	53320	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_51	high affinity transport system p37 family protein			53466	54666	+	Transport and binding proteins: Unknown substrate	**peg.425**	High affinity transport system protein p37 precursor	- none -	53467	54666	+
mnas_52	Phosphate/phosphonate ABC transporter, ATP-binding protein (EC 3.6.3.28)	*PhnC*		54667	55396	+	Transport and binding proteins: Unknown substrate	**peg.426**	ABC transporter ATP-binding protein	- none -	54668	55396	+
mnas_53	binding--dependent transport system inner membrane component family protein			55490	57062	+	Transport and binding proteins: Unknown substrate	**peg.427**	Transport system permease protein p69	- none -	55389	57062	+
mnas_54	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	57342	57462	-	Unknown: General	**peg.428**	hypothetical protein	- none -	57462	57343	-
mnas_55	subtilase family protein			57637	60025	-	Protein fate: Degradation of proteins, peptides, and glycopeptides	**peg.429**	hypothetical protein	- none -	60025	57638	-
mnas_56	putative lipoprotein		GO:0008150, GO:0003674, GO:0016020	60033	61671	-	Cell envelope: Other	**peg.430**	hypothetical protein	- none -	61671	60034	-
mnas_57	RNA polymerase beta subunit			62087	64340	+	Transcription: DNA-dependent RNA polymerase	**peg.431**	DNA-directed RNA polymerase beta subunit (EC 2.7.7.6)	**Category**: RNA Metabolism **Subcategory**: Transcription **Subsystem**: RNA polymerase bacterial **Category**: Virulence, Disease and Defense **Subcategory**: Invasion and intracellular resistance **Subsystem**: Mycobacterium virulence operon involved in DNA transcription	62088	65722	+
mnas_58	DNA-directed RNA polymerase subunit beta (EC 2.7.7.6)	*rpoB*		64459	65722	+	Transcription: DNA-dependent RNA polymerase						
mnas_59	DNA-directed RNA polymerase, beta' subunit (EC 2.7.7.6)	*rpoC*	GO:0000345, GO:0003899, GO:0006350	65714	70202	+	Transcription: DNA-dependent RNA polymerase	**peg.432**	DNA-directed RNA polymerase beta'; subunit (EC 2.7.7.6)	**Category**: RNA Metabolism **Subcategory**: Transcription **Subsystem**: RNA polymerase bacterial **Category**: Virulence, Disease and Defense **Subcategory**: Invasion and intracellular resistance **Subsystem**: Mycobacterium virulence operon involved in DNA transcription	65715	70202	+
mnas_60	dUTPase family protein			70296	70599	+	Unclassified: Role category not yet assigned						

IGS								RAST					
Gene_id	Common_name	Symbol[1]	GO terms[2]	fmin[3]	fmax[3]	S[4]	TIGR_roles	Gene_id	Function	Subsystem	Start[5]	Stop[5]	S[4]
mnas_61	excinuclease ABC subunit A (EC 3.1.25.-)	*uvrA*	GO:0006289, GO:0009380, GO:0009381	70611	73044	+	DNA metabolism: DNA replication, recombination, and repair	**peg.504**	Excinuclease ABC subunit A	**Category**: DNA Metabolism **Subcategory**: DNA repair **Subsystem**: DNA repair, UvrABC system	70873	73044	+
mnas_62	HPr(Ser) kinase/phosphatase (EC 2.7.1.-)	*hprK*	GO:0004674, GO:0009401, GO:0016791	73096	74038	+	Regulatory functions: Protein interactions, Signal transduction: PTS	**peg.505**	HPr kinase/phosphorylase (EC 2.7.1.-) (EC 2.7.4.-)	**Category**: Regulation and Cell signaling **Subcategory**: no subcategory **Subsystem**: HPr catabolite repression system	73097	74038	+
mnas_63	prolipoprotein diacylglyceryl transferase (EC 2.4.99.-)	*lgt*	GO:0008961, GO:0009249, GO:0016021	74039	75065	+	Protein fate: Protein modification and repair	**peg.506**	Prolipoprotein diacylglyceryl transferase (EC 2.4.99.-)	**Category**: Protein Metabolism **Subcategory**: Protein processing and modification **Subsystem**: Lipoprotein Biosynthesis	74040	75065	+
mnas_64	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	75106	79135	-	Unknown: General	**peg.507**	DNA helicase	- none -	79135	75107	-
mnas_65	AAA domain protein			79229	83915	-	Unknown: General Hypothetical	**peg.508**	DNA helicase	- none -	83915	79230	-
mnas_66	thioredoxin reductase (EC 1.8.1.9)	*trxB*	GO:0019430, GO:0050660, GO:0005737, GO:0004791	84164	85091	+	Cellular processes: Other, Unknown: Enzymes of unknown specificity	**peg.509**	Thioredoxin reductase (EC 1.8.1.9)	**Category**: Sulfur Metabolism **Subcategory**: no subcategory **Subsystem**: Thioredoxin-disulfide reductase	84165	85091	+
mnas_67	putative lipoprotein		GO:0008150, GO:0003674, GO:0016020	85115	86762	-	Cell envelope: Other	**peg.510**	hypothetical protein	- none -	86363	85116	-
mnas_68	subtilase family protein			86745	89082	-	Protein fate: Degradation of proteins, peptides, and glycopeptides	**peg.511**	hypothetical protein	- none -	89082	86746	-
mnas_69	putative lipoprotein		GO:0008150, GO:0003674, GO:0016020	89097	90441	-	Cell envelope: Other	**peg.512**	hypothetical protein	- none -	90441	89098	-
mnas_70	Transposase, IS4 family			90710	92354	-	Mobile and extrachromosomal element functions: Transposon functions	**peg.513**	Mobile element protein	- none -	92354	90711	-
mnas_71	DNA polymerase III, alpha subunit, Gram-positive type (EC 2.7.7.7)	*polC*	GO:0003887, GO:0006260, GO:0009360	92549	96941	-	DNA metabolism: DNA replication, recombination, and repair	**peg.514**	DNA polymerase III alpha subunit (EC 2.7.7.7)	**Category**: DNA Metabolism **Subcategory**: DNA replication **Subsystem**: DNA-replication	96941	92550	-
mnas_72	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	97267	98905	-	Unknown: General	**peg.579**	hypothetical protein	- none -	98905	97268	-
mnas_73	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	99049	104383	+	Unknown: Conserved	**peg.580**	FIG00834693: hypothetical protein	- none -	99086	104383	+
mnas_74	bacterial extracellular solute-binding protein	*oppA*		104397	107490	+	Transport and binding proteins: Unknown substrate	**peg.581**	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA (TC 3.A.1.5.1)	**Category**: Membrane Transport **Subcategory**: ABC transporters **Subsystem**: ABC transporter oligopeptide (TC 3.A.1.5.1)	104398	107490	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_75	Oligopeptide ABC transporter, permease protein (OppB)	*oppB*		107493	108549	+	Transport and binding proteins: Unknown substrate	**peg.582**	Oligopeptide transport system permease protein OppB (TC 3.A.1.5.1)	**Category**: Membrane Transport **Subcategory**: ABC transporters **Subsystem**: ABC transporter oligopeptide (TC 3.A.1.5.1)	107494	108549	+
mnas_76	Oligopeptide ABC transporter, permease protein (OppC)	*oppC*		108552	109635	+	Transport and binding proteins: Unknown substrate	**peg.583**	Oligopeptide transport system permease protein OppC (TC 3.A.1.5.1)	**Category**: Membrane Transport **Subcategory**: ABC transporters **Subsystem**: ABC transporter oligopeptide (TC 3.A.1.5.1)	108553	109635	+
mnas_77	oligopeptide/dipeptide ATP-binding protein	*oppD*	GO:0005524, GO:0009898, GO:0015440, GO:0015833, GO:0043190	109640	110849	+	Transport and binding proteins: Amino acids, peptides and amines	**peg.584**	Oligopeptide transport ATP-binding protein OppD (TC 3.A.1.5.1)	**Category**: Membrane Transport **Subcategory**: ABC transporters **Subsystem**: ABC transporter oligopeptide (TC 3.A.1.5.1)	109641	110849	+
mnas_78	Oligopeptide ABC transporter, ATP-binding protein OppF	*OppF*		110850	112182	+	Transport and binding proteins: Unknown substrate	**peg.585**	Oligopeptide transport ATP-binding protein OppF (TC 3.A.1.5.1)	**Category**: Membrane Transport **Subcategory**: ABC transporters **Subsystem**: ABC transporter oligopeptide (TC 3.A.1.5.1)	110851	112182	+
mnas_79	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	112196	112457	+	Unknown: General	**peg.586**	hypothetical protein	- none -	112197	112457	+
mnas_80	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	113305	114154	-	Transport and binding proteins: Unknown substrate, Unknown: Enzymes of unknown specificity	**peg.587**	TsaC protein (YrdC domain) required for threonylcarbamoyladenosine t(6)A37 modification in tRNA	- none -	114154	113306	-
mnas_81	archaeal ATPase family protein			114282	114936	-	Unknown: General Hypothetical	**peg.588**	hypothetical protein	- none -	114936	114283	-
mnas_82	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	115097	117698	-	Unknown: Conserved	**peg.589**	hypothetical protein	- none -	117659	115098	-
mnas_83	putative membrane protein		GO:0008150, GO:0003674, GO:0016020	117790	118237	+	Cell envelope: Other						
mnas_84	ATP synthase A chain family protein	*atpB*		118238	119045	+	Energy metabolism: ATP-proton motive force interconversion	**peg.590**	ATP synthase F0 sector subunit a	- none -	118239	119045	+
mnas_85	ATP synthase F0, C subunit (EC 3.6.3.14)	*atpE*	GO:0000276, GO:0015986, GO:0045263, GO:0045264, GO:0046933	119062	119380	+	Energy metabolism: ATP-proton motive force interconversion	**peg.591**	ATP synthase F0 sector subunit c	- none -	119063	119380	+
mnas_86	ATP synthase F0, B subunit (EC 3.6.3.14)	*atpF*	GO:0000276, GO:0015986, GO:0045263, GO:0045264, GO:0046933	119389	119947	+	Energy metabolism: ATP-proton motive force interconversion	**peg.592**	ATP synthase F0 sector subunit b (EC 3.6.3.14)	- none -	119405	119947	+
mnas_87	ATP synthase F1, delta subunit (EC 3.6.3.14)	*atpH*	GO:0000275, GO:0015986, GO:0045261, GO:0045262, GO:0046933	119962	120511	+	Energy metabolism: ATP-proton motive force interconversion	**peg.593**	ATP synthase delta chain (EC 3.6.3.14)	- none -	119963	120511	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_88	ATP synthase F1, alpha subunit (EC 3.6.3.14)	*atpA*	GO:0000275, GO:0015986, GO:0045261, GO:0045262, GO:0046933	120511	122092	+	Energy metabolism: ATP-proton motive force interconversion	**peg.594**	ATP synthase alpha chain (EC 3.6.3.14)	- none -	120512	122092	+
mnas_89	ATP synthase family protein	*atpG*		122094	122820	+	Energy metabolism: ATP-proton motive force interconversion						
mnas_90	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	122820	124143	-	Unknown: General						
mnas_91	protein yebR	*yebR*		124152	124599	-	Unclassified: Role category not yet assigned	**peg.1**	Free methionine-(R)-sulfoxide reductase, contains GAF domain	- none -	124599	124153	-
mnas_92	lysine--tRNA ligase (EC 6.1.1.6)	*lysS*	GO:0004824, GO:0005737, GO:0006430	124659	126129	-	Protein synthesis: tRNA aminoacylation	**peg.2**	Lysyl-tRNA synthetase (class II) (EC 6.1.1.6)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: tRNA aminoacylation, Lys	126129	124660	-
mnas_93	ABC transporter family protein			126266	127307	-	Transport and binding proteins: Unknown substrate	**peg.3**	ABC transporter ATP-binding protein	- none -	127307	126267	-
mnas_94	Putative ABC transporter permease protein			127308	135783	-	Transport and binding proteins: Unknown substrate	**peg.4**	ABC transporter permease protein	- none -	135783	127309	-
mnas_95	ATP-dependent zinc metalloprotease FtsH domain protein (EC 3.4.24.-)	*ftsH*	GO:0004222, GO:0005524, GO:0043934, GO:0006950, GO:0030428, GO:0030163, GO:0006508, GO:0051301, GO:0016021, GO:0017111, GO:0008270, GO:0005886, GO:0007049	136009	138019	-	Cellular processes: Other, Transport and binding proteins: Unknown substrate, Cellular processes: Adaptations to atypical conditions, Protein fate: Other, Unknown: Enzymes of unknown specificity	**peg.5**	Cell division protein FtsH (EC 3.4.24.-)	**Category**: Cofactors, Vitamins, Prosthetic Groups, Pigments **Subcategory**: Folate and pterines **Subsystem**: Folate biosynthesis cluster **Subsystem**: Cell division-ribosomal stress proteins cluster	138019	136010	-
mnas_96	tRNA(Ile)-lysidine synthetase (EC 6.3.4.-)	*tilS*	GO:0006400, GO:0016879	138132	138987	-	Protein synthesis: tRNA and rRNA base modification	**peg.6**	tRNA(Ile)-lysidine synthetase (EC 6.3.4.19)	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: tRNA processing **Category**: Cofactors, Vitamins, Prosthetic Groups, Pigments **Subcategory**: Folate and pterines **Subsystem**: Folate biosynthesis cluster **Subsystem**: Cell division-ribosomal stress proteins cluster	138987	138133	-

IGS								RAST					
Gene_id	**Common_name**	**Symbol[1]**	**GO terms[2]**	**fmin[3]**	**fmax[3]**	**S[4]**	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start[5]**	**Stop[5]**	**S[4]**
mnas_97	peptidyl-tRNA hydrolase (EC 3.1.1.29)	*pth*	GO:0004045, GO:0006412	138986	139544	-	Protein synthesis: Other	**peg.7**	Peptidyl-tRNA hydrolase (EC 3.1.1.29)	**Category**: Dormancy and Sporulation **Subcategory**: no subcategory **Subsystem**: Sporulation-associated proteins with broader functions **Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Translation termination factors bacterial **Subsystem**: Cell division-ribosomal stress proteins cluster	139544	138987	-
mnas_98	AAA domain protein			139551	141768	-	Unknown: General Hypothetical	**peg.8**	RecD-like DNA helicase YrrC	**Category**: DNA Metabolism **Subcategory**: DNA repair **Subsystem**: DNA repair, bacterial RecBCD pathway	141768	139552	-
mnas_99	chromosomal replication initiator protein DnaA	*dnaA*	GO:0003677, GO:0003688, GO:0005524, GO:0006270, GO:0006275	142050	143460	+	DNA metabolism: DNA replication, recombination, and repair	**peg.9**	Chromosomal replication initiator protein DnaA	**Category**: DNA Metabolism **Subcategory**: DNA replication **Subsystem**: DNA-replication **Subsystem**: Cell Division Subsystem including YidCD	142051	143460	+
mnas_100	DNA polymerase III beta subunit, central domain protein (EC 2.7.7.7)	*dnaN*		143607	144738	+	DNA metabolism: DNA replication, recombination, and repair	**peg.10**	DNA polymerase III beta subunit (EC 2.7.7.7)	**Category**: DNA Metabolism **Subcategory**: DNA replication **Subsystem**: DNA-replication **Subsystem**: Cell Division Subsystem including YidCD	143608	144738	+
mnas_101	S4 domain protein			144739	144946	+	Unknown: General Hypothetical	**peg.11**	FIG002958: hypothetical protein	**Subsystem**: DNA replication cluster 1	144740	144946	+
mnas_102	DNA methylase family protein			145044	146124	+	DNA metabolism: Restriction/modification	**peg.12**	Adenine specific DNA methyltransferase (HINFIM)	- none -	145045	146124	+
mnas_103	putative type-2 restriction enzyme HinfI		GO:0016787, GO:0009036, GO:0009307, GO:0004519, GO:0004518	146116	146911	+	Cellular processes: Other, Unknown: Enzymes of unknown specificity	**peg.13**	Type II restriction enzyme HinfI (EC 3.1.21.4) (Endonuclease HinfI) (R.HinfI)	- none -	146117	146911	+
mnas_104	putative lipoprotein		GO:0008150, GO:0003674, GO:0016020	146990	147200	+	Cell envelope: Other	**peg.14**	hypothetical protein	- none -	147187	146957	-
mnas_105	2,3-bisphosphoglycerate-independent phosphoglycerate mutase (EC 5.4.2.1)			147256	148009	+	Unknown: Enzymes of unknown specificity	**peg.61**	2,3-bisphosphoglycerate-independent phosphoglycerate mutase (EC 5.4.2.1)	**Category**: Carbohydrates **Subcategory**: Central carbohydrate metabolism **Subsystem**: Glycolysis and Gluconeogenesis **Category**: Miscellaneous **Subcategory**: -no subcategory **Subsystem**: Phosphoglycerate mutase protein family	147371	148009	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_106	integral membrane family protein			148120	149113	+	Cell envelope: Other	**peg.62**	hypothetical protein	- none -	148121	149113	+
mnas_107	D-lactate dehydrogenase (EC 1.1.1.28)	*ldhA*	GO:0008720, GO:0006950, GO:0009236, GO:0051287, GO:0008939	149269	150307	+	Cellular processes: Other, Transport and binding proteins: Unknown substrate, Cellular processes: Adaptations to atypical conditions, Unknown: Enzymes of unknown specificity	**peg.63**	D-lactate dehydrogenase (EC 1.1.1.28)	**Category**: Carbohydrates **Subcategory**: Fermentation **Subsystem**: Fermentations: Lactate	149270	150307	+
mnas_108	major intrinsic family protein			150299	151364	+	Unclassified: Role category not yet assigned	**peg.64**	Aquaporin Z	**Category**: Stress Response **Subcategory**: Osmotic stress **Subsystem**: Osmoregulation	150300	151364	+
mnas_109	S1 RNA binding domain protein			151475	153614	+	Unknown: General Hypothetical	**peg.65**	Transcription accessory protein (S1 RNA-binding domain)	**Category**: RNA Metabolism **Subcategory**: Transcription **Subsystem**: Transcription factors bacterial **Subsystem**: Cell division-ribosomal stress proteins cluster	151476	153614	+
mnas_110	tRNA-Gln			153820	153745	-		**rna.3**	tRNA-Gln-TTG	- none -	153820	153749	-
mnas_111	tRNA-Tyr			153909	153825	-		**rna.4**	tRNA-Tyr-GTA	- none -	153909	153829	-
mnas_112	putative lipoprotein		GO:0008150, GO:0003674, GO:0016020	154014	154989	+	Cell envelope: Other	**peg.66**	hypothetical protein	- none -	154015	154989	+
mnas_113	ribulose-phosphate 3 epimerase family protein			155012	155678	-	Unknown: Enzymes of unknown specificity	**peg.67**	Ribulose-phosphate 3-epimerase (EC 5.1.3.1)	**Category**: Carbohydrates **Subcategory**: Central carbohydrate metabolism **Subsystem**: Pentose phosphate pathway **Subsystem**: Conserved gene cluster associated with Met-tRNA formyltransferase	155678	155013	-
mnas_114	ribosome small subunit-dependent GTPase A (EC 3.6.-.-)	*rsgA*	GO:0005525, GO:0006412, GO:0043022	155677	156526	-	Protein synthesis: Translation factors	**peg.68**	Ribosome small subunit-stimulated GTPase EngC	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Universal GTPases	156526	155678	-
mnas_115	phosphotransferase enzyme family protein			156525	157521	-	Unknown: Enzymes of unknown specificity	**peg.69**	Serine/threonine protein kinase PrkC, regulator of stationary phase	**Subsystem**: Conserved gene cluster associated with Met-tRNA formyltransferase	157521	156526	-
mnas_116	phosphatase 2C family protein			157507	158287	-	Unknown: Enzymes of unknown specificity	**peg.70**	Protein serine/threonine phosphatase PrpC, regulation of stationary phase	**Subsystem**: Conserved gene cluster associated with Met-tRNA formyltransferase	158287	157508	-
mnas_117	guanylate kinase (EC 2.7.4.8)	*gmk*	GO:0004385, GO:0015949	158276	158891	-	Purines, pyrimidines, nucleosides, and nucleotides: Nucleotide and nucleoside interconversions	**peg.71**	Guanylate kinase (EC 2.7.4.8)	**Category**: Nucleosides and Nucleotides **Subcategory**: Purines **Subsystem**: Purine conversions	158891	158277	-

IGS								RAST					
Gene_id	**Common_name**	**Symbol[1]**	**GO terms[2]**	**fmin[3]**	**fmax[3]**	**S[4]**	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start[5]**	**Stop[5]**	**S[4]**
mnas_118	tRNA pseudouridine(55) synthase (EC 5.4.99.25)	*truB*	GO:0004730, GO:0006400, GO:0016870	159005	159848	+	Protein synthesis: tRNA and rRNA base modification	**peg.72**	tRNA pseudouridine synthase B (EC 4.2.1.70)	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: RNA pseudouridine syntheses **Subsystem**: tRNA processing **Category**: Cofactors, Vitamins, Prosthetic Groups, Pigments **Subcategory**: Riboflavin, FMN, FAD **Subsystem**: Riboflavin, FMN and FAD metabolism in plants	159006	159848	+
mnas_119	HAD hydrolase, IIB family protein		GO:0008152, GO:0016787	159847	160666	+	Unknown: Enzymes of unknown specificity	**peg.73**	FIG00834275: hypothetical protein	- none -	159848	160666	+
mnas_120	FAD synthetase family protein			160665	161508	+	Unknown: Enzymes of unknown specificity	**peg.74**	Riboflavin kinase (EC 2.7.1.26) / FMN adenylyltransferase (EC 2.7.7.2)	**Category**: Cofactors, Vitamins, Prosthetic Groups, Pigments **Subcategory**: Riboflavin, FMN, FAD **Subsystem**: Riboflavin, FMN and FAD metabolism in plants **Subsystem**: Riboflavin, FMN and FAD metabolism in plants **Subsystem**: Riboflavin to FAD **Subsystem**: Riboflavin to FAD **Subsystem**: Riboflavin, FMN and FAD metabolism **Subsystem**: Riboflavin, FMN and FAD metabolism	160666	161508	+
mnas_121	putative lipoprotein		GO:0008150, GO:0003674, GO:0016020	161659	164827	+	Cell envelope: Other	**peg.75**	no similarity found Pfscan: pos. 16-26 PS00013 | PROKAR_LIPOPROTEIN Prokaryotic membrane lipoprotein lipid attachment site	- none -	161660	164827	+
mnas_122	ribosomal protein S15	*rpsO*	GO:0000314, GO:0003735, GO:0006412, GO:0022627	164971	165238	+	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.76**	SSU ribosomal protein S15p (S13e)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome SSU bacterial	164972	165238	+
mnas_123	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	165357	165678	+	Unknown: General	**peg.77**	hypothetical protein	- none -	165358	165678	+
mnas_124	DAK2 domain fusion YloV family protein			165670	167311	+	Unknown: General Hypothetical	**peg.78**	Dihydroxyacetone kinase family protein	**Category**: Fatty Acids, Lipids, and Isoprenoids **Subcategory**: Phospholipids **Subsystem**: Glycerolipid and Glycerophospholipid Metabolism in Bacteria	165671	167311	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_125	fatty acid/phospholipid synthesis protein PlsX	*plsX*	GO:0003824, GO:0008610	167310	168321	+	Fatty acid and phospholipid metabolism: Biosynthesis	**peg.79**	Phosphate:acyl-ACP acyltransferase PlsX	**Category**: Fatty Acids, Lipids, and Isoprenoids **Subcategory**: Phospholipids **Subsystem**: Glycerolipid and Glycerophospholipid Metabolism in Bacteria	167311	168321	+
mnas_126	ribonuclease III (EC 3.1.26.3)	*rnc*	GO:0004525, GO:0006396	168313	169006	+	Transcription: RNA processing	**peg.80**	Ribonuclease III (EC 3.1.26.3)	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: RNA processing and degradation, bacterial	168314	169006	+
mnas_127	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	169088	169754	-	Unknown: General						
mnas_128	tRNA sulfurtransferase ThiI (EC 2.8.1.4)	*thiI*	GO:0003824, GO:0006400, GO:0009228	169746	170895	-	Biosynthesis of cofactors, prosthetic groups, and carriers: Thiamine, Protein synthesis: tRNA and rRNA base modification	**peg.118**	tRNA S(4)U 4-thiouridine synthase (former ThiI)	**Category**: Cofactors, Vitamins, Prosthetic Groups, Pigments **Subcategory**: Cofactors, Vitamins, Prosthetic Groups, Pigments - no subcategory **Subsystem**: Thiamin biosynthesis	170895	169747	-
mnas_129	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	170884	172480	-	Unknown: General	**peg.119**	DNA double-strand break repair Rad50 ATPase	- none -	172480	170885	-
mnas_130	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	172691	173186	+	Unknown: General	**peg.120**	hypothetical protein	- none -	172584	173186	+
mnas_131	HAD hydrolase, IIB family protein		GO:0008152, GO:0016787	173197	174184	-	Unknown: Enzymes of unknown specificity	**peg.121**	Hydrolase (HAD superfamily)	- none -	174184	173198	-
mnas_132	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	174240	174702	+	Unknown: Conserved	**peg.122**	predicted coding region	- none -	174241	174702	+
mnas_133	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	174701	175199	+	Unknown: Conserved	**peg.123**	hypothetical protein	- none -	174994	174794	-
								peg.124	hypothetical protein	- none -	175032	175199	+
mnas_134	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	175182	175566	+	Unknown: Conserved	**peg.125**	Expressed protein	- none -	175183	175566	+
mnas_135	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	176098	176542	+	Unknown: General	**peg.126**	hypothetical protein	- none -	176099	176542	+
mnas_136	nusA-like KH domain protein			176549	178202	+	Unknown: General Hypothetical	**peg.127**	Transcription termination protein NusA	**Category**: RNA Metabolism **Subcategory**: Transcription **Subsystem**: Transcription factors bacterial **Subsystem**: NusA-TFII Cluster	176550	178202	+
mnas_137	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	178185	178473	+	Unknown: Conserved	**peg.128**	hypothetical protein	- none -	178186	178473	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_138	translation initiation factor IF-2	*infB*	GO:0003743, GO:0005737, GO:0006413	178456	179602	+	Protein synthesis: Translation factors	**peg.129**	Translation initiation factor 2	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Universal GTPases **Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Translation initiation factors bacterial **Subsystem**: NusA-TFII Cluster	178457	180291	+
mnas_139	translation-initiation factor 2 family protein			179643	180291	+	Protein synthesis: Translation factors, Disrupted reading frame: NULL						
mnas_140	adenine phosphoribosyltransferase (EC 2.4.2.7)	*apt*	GO:0003999, GO:0005737, GO:0006166	180290	180797	+	Purines, pyrimidines, nucleosides, and nucleotides: Salvage of nucleosides and nucleotides	**peg.130**	Adenine phosphoribosyltransferase (EC 2.4.2.7)	**Category**: Nucleosides and Nucleotides **Subcategory**: Purines **Subsystem**: Purine conversions	180291	180797	+
mnas_141	ribosomal protein L1	*rplA*	GO:0000311, GO:0003735, GO:0006412, GO:0022625	180893	181589	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.131**	LSU ribosomal protein L1p (L10Ae)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial	181589	180894	-
mnas_142	ribosomal protein L11	*rplK*	GO:0000315, GO:0003735, GO:0006412, GO:0022625	181588	182179	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.132**	LSU ribosomal protein L11p (L12e)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial **Subsystem**: LSU ribosomal proteins cluster	182179	181589	-
mnas_143	LMP repeated region family protein			182519	183395	+	Unclassified: Role category not yet assigned	**peg.133**	VlhA.4.04	- none -	182520	183395	+
mnas_144	Putative Beta-galactosidase			183431	186473	-	Unknown: General Hypothetical	**peg.134**	Beta-galactosidase (EC 3.2.1.23)	**Category**: Carbohydrates **Subcategory**: Di- and oligosaccharides **Subsystem**: Lactose utilization **Category**: Sulfur Metabolism **Subcategory**: no subcategory **Subsystem**: Galactosylceramide and Sulfatide metabolism	186473	183432	-

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_145	beta-galactosidase domain protein (EC 3.2.1.23	*bgaA 2*	GO:0016020, GO:0008152, GO:0003824, GO:0016798, GO:0009986, GO:0043169, GO:0004553, GO:0016787, GO:0004565, GO:0005618, GO:0005975	186593	186770	-	Energy metabolism: Sugars, Transport and binding proteins: Unknown substrate, Disrupted reading frame: NULL, Unknown: Enzymes of unknown specificity	**peg.135**	hypothetical protein	- none -	186770	186594	-
mnas_146	beta-galactosidase domain protein (EC 3.2.1.23)	*bgaA*	GO:0016020, GO:0008152, GO:0003824, GO:0016798, GO:0009986, GO:0043169, GO:0004553, GO:0016787, GO:0004565, GO:0005618, GO:0005975	186883	187102	-	Energy metabolism: Sugars, Transport and binding proteins: Unknown substrate, Unknown: Enzymes of unknown specificity	**peg.136**	Beta-galactosidase (EC 3.2.1.23)	**Category**: Carbohydrates **Subcategory**: Di- and oligosaccharides **Subsystem**: Lactose utilization **Category**: Sulfur Metabolism **Subcategory**: no subcategory **Subsystem**: Galactosylceramide and Sulfatide metabolism	187102	186884	-
mnas_147	glycosyl hydrolases family 2, sugar binding domain protein			187112	189515	-	Unknown: Enzymes of unknown specificity	**peg.137**	Beta-galactosidase (EC 3.2.1.23)	**Category**: Carbohydrates **Subcategory**: Di- and oligosaccharides **Subsystem**: Lactose utilization **Category**: Sulfur Metabolism **Subcategory**: no subcategory **Subsystem**: Galactosylceramide and Sulfatide metabolism	189515	187113	-
mnas_148	ABC transporter family protein			189740	190124	-	Transport and binding proteins: Unknown substrate	**peg.138**	Lipid A export ATP-binding/permease protein MsbA	- none -	190100	189741	-
mnas_149	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	190132	190285	-	Unknown: General						
mnas_150	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	190495	200602	+	Unknown: General	**peg.157**	Putative peptidoglycan bound protein (LPXTG motif) Lmo1799 homolog	- none -	190496	200602	+
mnas_151	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	200893	204526	+	Unknown: General	**peg.158**	Exonuclease SbcC	**Category**: DNA Metabolism **Subcategory**: DNA repair **Subsystem**: DNA repair, bacterial	200894	204526	+
mnas_152	type I restriction-modification system, M subunit (EC 2.1.1.72)	*hsdM*	GO:0006306, GO:0009007, GO:0009307, GO:0015666	204706	205987	+	DNA metabolism: Restriction/modification	**peg.159**	Type I restriction-modification system, DNA-methyltransferase subunit M (EC 2.1.1.72)	**Category**: DNA Metabolism **Subcategory**: no subcategory **Subsystem**: Type I Restriction-Modification **Subsystem**: Restriction-Modification System	204707	206292	+
mnas_153	N-6 DNA Methylase family protein			205944	206292	+	Unknown: Enzymes of unknown specificity						

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_154	type I restriction modification DNA specificity domain protein			206314	207580	+	DNA metabolism: Restriction/modification	**peg.160**	Type I restriction-modification system, specificity subunit S (EC 3.1.21.3)	**Category**: DNA Metabolism **Subcategory**: no subcategory **Subsystem**: Type I Restriction-Modification **Subsystem**: Restriction-Modification System	206315	207580	+
mnas_155	type I restriction modification DNA specificity domain protein			207533	208235	+	DNA metabolism: Restriction/modification	**peg.161**	Type I restriction-modification system, specificity subunit S (EC 3.1.21.3)	**Category**: DNA Metabolism **Subcategory**: no subcategory **Subsystem**: Type I Restriction-Modification **Subsystem**: Restriction-Modification System	207534	208235	+
mnas_156	type I site-specific deoxyribonuclease, HsdR family protein (EC 3.1.21.3)	*hsdR*	GO:0009035, GO:0009307, GO:0019812	208238	211040	+	DNA metabolism: Restriction/modification						
mnas_157	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	211050	212709	+	Unknown: General	**peg.182**	massive surface protein MspF	- none -	211366	212709	+
mnas_158	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	212867	212987	+	Unknown: General	**peg.183**	hypothetical protein	- none -	212868	212987	+
mnas_159	DNA polymerase III, subunit gamma and tau (EC 2.7.7.7)	*dnaX*	GO:0003887, GO:0006260, GO:0009360	213184	215344	+	DNA metabolism: DNA replication, recombination, and repair	**peg.184**	DNA polymerase III subunits gamma and tau (EC 2.7.7.7)	**Category**: DNA Metabolism **Subcategory**: DNA replication **Subsystem**: DNA-replication **Subcategory**: DNA uptake, competence **Subsystem**: DNA processing cluster	213185	215344	+
mnas_160	DNA-binding protein, YbaB/EbfC family		GO:0003677, GO:0005737, GO:0008150	215353	215638	+	Unknown: Conserved	**peg.185**	FIG000557: hypothetical protein co-occurring with RecR	**Category**: DNA Metabolism **Subcategory**: DNA uptake, competence **Subsystem**: DNA processing cluster	215354	215638	+
mnas_161	recR family protein			215637	216222	+	DNA metabolism: DNA replication, recombination, and repair	**peg.186**	Recombination protein RecR	**Category**: DNA Metabolism **Subcategory**: DNA replication **Subsystem**: DNA-replication **Subcategory**: DNA uptake, competence **Subsystem**: DNA processing cluster	215638	216222	+
mnas_162	thymidylate kinase (EC 2.7.4.9)	*tmk*	GO:0004798, GO:0015949	216239	216896	+	Purines, pyrimidines, nucleosides, and nucleotides: Nucleotide and nucleoside interconversions	**peg.187**	Thymidylate kinase (EC 2.7.4.9)	- none -	216240	216896	+
mnas_163	Putative DNA polymerase III, delta' subunit (EC 2.7.7.7)	*holB*		216882	217782	+	Unknown: General Hypothetical	**peg.188**	DNA polymerase III delta prime subunit (EC 2.7.7.7)	**Category**: DNA Metabolism **Subcategory**: DNA replication **Subsystem**: DNA-replication	216883	217782	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_164	tetrapyrrole (Corrin/Porphyrin) Methylases family protein			217774	218506	+	Unknown: Enzymes of unknown specificity	**peg.189**	rRNA small subunit methyltransferase I	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: 16S rRNA modification within P site of ribosome **Category**: Stress Response **Subcategory**: Heat shock **Subsystem**: Heat shock dnaK gene cluster extended	217775	218506	+
mnas_165	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	218495	219779	+	Unknown: General	**peg.190**	hypothetical protein	- none -	218496	219779	+
mnas_166	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	219800	220277	-	Unknown: Conserved	**peg.191**	Protein yjgK	- none -	220277	219801	-
mnas_167	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	220343	222005	-	Unknown: Conserved	**peg.192**	Chromosome partition protein smc	**Category**: DNA Metabolism **Subcategory**: no subcategory **Subsystem**: DNA structural proteins, bacterial	222005	220344	-
mnas_168	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	222372	222588	+	Unknown: General	**peg.193**	hypothetical protein	- none -	222373	222588	+
mnas_169	cytidylyltransferase family protein			222687	223755	-	Fatty acid and phospholipid metabolism: Biosynthesis	**peg.194**	Phosphatidate cytidylyltransferase (EC 2.7.7.41)	**Category**: Fatty Acids, Lipids, and Isoprenoids **Subcategory**: Phospholipids **Subsystem**: Glycerolipid and Glycerophospholipid Metabolism in Bacteria	223755	222688	-
mnas_170	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	224286	225714	-	Unknown: General	**peg.195**	hypothetical protein	- none -	225714	224287	-
mnas_171	subtilase family protein			225911	227657	-	Protein fate: Degradation of proteins, peptides, and glycopeptides	**peg.196**	hypothetical protein	- none -	227513	225912	-
mnas_172	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	227599	228199	-	Unknown: General	**peg.197**	hypothetical protein	- none -	228175	227600	-
mnas_173	DNA topoisomerase family protein			228205	229624	-	DNA metabolism: DNA replication, recombination, and repair						
mnas_174	ribosomal protein S6	*rpsF*	GO:0000314, GO:0003735, GO:0006412, GO:0022627	229809	230271	+	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.244**	SSU ribosomal protein S6p	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome SSU bacterial	229867	230271	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_175	single-stranded DNA-binding family protein	*ssb*	GO:0003697, GO:0006260, GO:0006281, GO:0006310, GO:0006350	230304	230943	+	DNA metabolism: DNA replication, recombination, and repair	**peg.245**	Single-stranded DNA-binding protein	**Category**: DNA Metabolism **Subcategory**: DNA repair **Subsystem**: DNA repair, bacterial	230305	230943	+
mnas_176	ribosomal protein S18	*rpsR*	GO:0000314, GO:0003735, GO:0006412, GO:0022627	231029	231233	+	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.246**	SSU ribosomal protein S18p @ SSU ribosomal protein S18p, zinc-independent	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome SSU bacterial **Subsystem**: Ribosome SSU bacterial	231030	231233	+
mnas_177	tRNA (guanine-N(7)-)-methyltransferase (EC 2.1.1.33)	*trmB*	GO:0006400, GO:0008176	231282	231891	-	Protein synthesis: tRNA and rRNA base modification	**peg.247**	tRNA (guanine46-N7-)-methyltransferase (EC 2.1.1.33)	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: RNA methylation	231891	231283	-
mnas_178	RNA methyltransferase, RsmD family (EC 2.1.1.-)		GO:0008168, GO:0031167	231951	232497	+	Unknown: Enzymes of unknown specificity	**peg.248**	16S rRNA (guanine(966)-N(2))-methyltransferase (EC 2.1.1.171) ## SSU rRNA m(2)G966	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: RNA methylation	231952	232497	+
mnas_179	23S rRNA (uracil-5-)-methyltransferase RumA (EC 2.1.1.-)	*rumA*	GO:0000154, GO:0008649	232484	233780	+	Protein synthesis: tRNA and rRNA base modification	**peg.249**	RNA methyltransferase, TrmA family	- none -	232485	233780	+
mnas_180	glycosyl transferase 2 family protein			234006	235026	+	Cell envelope: Biosynthesis and degradation of surface polysaccharides and lipopolysaccharides	**peg.250**	Glycosyltransferase	- none -	234007	235026	+
mnas_181	RDD family protein			235025	235646	+		**peg.251**	predicted coding region	- none -	235026	235646	+
mnas_182	uvrD/REP helicase N-terminal domain protein			235648	237832	+	DNA metabolism: DNA replication, recombination, and repair	**peg.252**	ATP-dependent DNA helicase UvrD/PcrA	**Category**: DNA Metabolism **Subcategory**: DNA repair **Subsystem**: DNA repair, bacterial UvrD and related helicases	235649	237832	+
mnas_183	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	237834	239493	+	Unknown: General	**peg.253**	predicted coding region	- none -	237835	239493	+
mnas_184	EDD, DegV family domain protein		GO:0005575, GO:0008150, GO:0016740	239605	240448	+	Unknown: General Hypothetical	**peg.254**	FIG00836191: hypothetical protein	- none -	239456	240448	+
mnas_185	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	240447	240690	+	Unknown: General	**peg.255**	hypothetical protein	- none -	240448	240690	+
mnas_186	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase (EC 2.7.8.5)	*pgsA*	GO:0008444, GO:0008654	240741	241395	-	Fatty acid and phospholipid metabolism: Biosynthesis	**peg.256**	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase (EC 2.7.8.5)	**Category**: Fatty Acids, Lipids, and Isoprenoids **Subcategory**: Phospholipids **Subsystem**: Glycerolipid and Glycerophospholipid Metabolism in Bacteria	241395	240742	-
mnas_187	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	241378	241942	-	Unknown: General	**peg.257**	hypothetical protein	- none -	241942	241379	-

IGS								RAST					
Gene_id	Common_name	Symbol[1]	GO terms[2]	fmin[3]	fmax[3]	S[4]	TIGR_roles	Gene_id	Function	Subsystem	Start[5]	Stop[5]	S[4]
mnas_188	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	241950	242274	-	Unknown: Conserved	**peg.258**	hypothetical protein	- none -	242274	241951	-
mnas_189	magnesium transporter	*mgtE*	GO:0015095, GO:0015693	242282	243617	-	Transport and binding proteins: Cations and iron carrying compounds	**peg.259**	Mg/Co/Ni transporter MgtE / CBS domain	**Category**: Membrane Transport **Subcategory**: Cation transporters **Subsystem**: Magnesium transport	243617	242283	-
mnas_190	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	243694	243784	-	Unknown: General						
mnas_191	disA bacterial checkpoint controller nucleotide-binding family protein			243785	244196	-	Unclassified: Role category not yet assigned						
mnas_192	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	244214	244547	+	Unknown: General	**peg.277**	hypothetical protein	- none -	244296	244547	+
mnas_193	tRNA-Arg			244829	244752	-		**rna.11**	tRNA-Arg-TCT	- none -	244829	244756	-
mnas_194	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	245280	248232	+	Unknown: General	**peg.278**	Predicted cell-wall-anchored protein SasA (LPXTG motif)	- none -	245281	248232	+
mnas_195	GDSL-like Lipase/Acylhydrolase family protein			248502	258885	+	Unknown: Enzymes of unknown specificity	**peg.279**	hypothetical protein	- none -	248524	258885	+
mnas_196	cation transporting ATPase, family protein			258959	259481	-	Transport and binding proteins: Unknown substrate	**peg.280**	Lead, cadmium, zinc and mercury transporting ATPase (EC 3.6.3.3) (EC 3.6.3.5); Copper-translocating P-type ATPase (EC 3.6.3.4)	**Category**: Membrane Transport **Subcategory**: Cation transporters **Subsystem**: Copper Transport System **Category**: Virulence, Disease and Defense **Subcategory**: Resistance to antibiotics and toxic compounds **Subsystem**: Copper homeostasis	259481	258960	-
mnas_197	recombination O N terminal family protein			259497	259776	-	Unclassified: Role category not yet assigned						
mnas_198	chaperone protein DnaJ	*dnaJ*	GO:0005515, GO:0006457, GO:0009408	259848	260949	+	Protein fate: Protein folding and stabilization	**peg.291**	Chaperone protein DnaJ	**Category**: Protein Metabolism **Subcategory**: Protein folding **Subsystem**: Protein chaperones **Category**: Stress Response **Subcategory**: Heat shock **Subsystem**: Heat shock dnaK gene cluster extended	259849	260949	+
mnas_199	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	260949	261324	+	Unknown: Conserved	**peg.292**	FIG000605: protein co-occurring with transport systems (COG1739)	- none -	260950	261324	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_200	cytosol aminopeptidase family, catalytic domain protein			261323	262694	+	Protein fate: Degradation of proteins, peptides, and glycopeptides	**peg.293**	Cytosol aminopeptidase PepA (EC 3.4.11.1)	**Category**: Protein Metabolism **Subcategory**: Protein degradation **Subsystem**: Aminopeptidases (EC 3.4.11.-)	261324	262694	+
mnas_201	cytosol aminopeptidase family, catalytic domain protein			262702	264076	+	Protein fate: Degradation of proteins, peptides, and glycopeptides	**peg.294**	Cytosol aminopeptidase PepA (EC 3.4.11.1)	**Category**: Protein Metabolism **Subcategory**: Protein degradation **Subsystem**: Aminopeptidases (EC 3.4.11.-)	262703	264076	+
mnas_202	peptidase S41 family protein			264139	266107	+	Protein fate: Degradation of proteins, peptides, and glycopeptides	**peg.295**	hypothetical protein	- none -	264140	266107	+
mnas_203	endonuclease I family protein			266153	267725	-	Unknown: Enzymes of unknown specificity	**peg.296**	Endonuclease I	- none -	267725	266154	-
mnas_204	HAD hydrolase, IIB family protein		GO:0008152, GO:0016787	267833	268661	-	Unknown: Enzymes of unknown specificity	**peg.297**	hypothetical protein	- none -	268661	267834	-
mnas_205	calcineurin-like phosphoesterase family protein			268775	271316	+	Unknown: Enzymes of unknown specificity	**peg.299**	5'-nucleotidase (EC 3.1.3.5)	**Category**: Nucleosides and Nucleotides **Subcategory**: Purines **Subsystem**: Purine conversions	268959	271316	+
								peg.298	hypothetical protein	- none -	268999	268877	-
mnas_206	ribosomal protein S12	*rpsL*	GO:0000314, GO:0003735, GO:0006412, GO:0022627	271479	271890	+	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.300**	SSU ribosomal protein S12p (S23e)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome SSU bacterial **Category**: Virulence, Disease and Defense **Subcategory**: Invasion and intracellular resistance **Subsystem**: Mycobacterium virulence operon involved in protein synthesis (SSU ribosomal proteins)	271480	271890	+
mnas_207	ribosomal protein S7	*rpsG*	GO:0000314, GO:0003735, GO:0006412, GO:0022627	271950	272421	+	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.301**	SSU ribosomal protein S7p (S5e)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome SSU bacterial **Category**: Virulence, Disease and Defense **Subcategory**: Invasion and intracellular resistance **Subsystem**: Mycobacterium virulence operon involved in protein synthesis (SSU ribosomal proteins)	271951	272421	+
mnas_208	translation elongation factor G	*fusA*	GO:0003746, GO:0006414	272433	273702	+	Protein synthesis: Translation factors						

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_209	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	273712	274435	+	Unknown: General	**peg.308**	hypothetical protein	- none -	273743	274435	+
								peg.309	hypothetical protein	- none -	274609	274761	+
mnas_210	preprotein translocase, SecE subunit	*secE*	GO:0005887, GO:0015450, GO:0043952	274760	275003	+	Protein fate:Protein and peptide secretion and trafficking	**peg.310**	Preprotein translocase subunit SecE (TC 3.A.5.1.1)	**Subsystem**: LSU ribosomal proteins cluster	274761	275003	+
mnas_211	KOW motif family protein			275027	275633	+	Unclassified: Role category not yet assigned	**peg.311**	Transcription antitermination protein NusG	**Category**: RNA Metabolism **Subcategory**: Transcription **Subsystem**: Transcription factors bacterial **Subsystem**: LSU ribosomal proteins cluster	275028	275633	+
mnas_212	ABC-2 type transporter family protein			275685	276825	-	Transport and binding proteins: Unknown substrate	**peg.312**	ABC transporter, permease protein	- none -	276825	275686	-
mnas_213	ABC transporter family protein			276805	277750	-	Transport and binding proteins: Unknown substrate	**peg.313**	Methionine ABC transporter ATP-binding protein	- none -	277750	276806	-
mnas_214	tRNA-Trp			277866	277792	-		**rna.13**	tRNA-Trp-CCA	- none -	277866	277796	-
mnas_215	putative membrane protein		GO:0008150, GO:0003674, GO:0016020	278060	278345	+	Cell envelope: Other	**peg.314**	hypothetical protein	- none -	278061	278345	+
mnas_216	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	278507	278636	+	Unknown: General	**peg.315**	hypothetical protein	- none -	278508	278636	+
mnas_217	50S ribosome-binding GTPase family protein			278781	279882	-	Unknown: General Hypothetical	**peg.316**	GTP-binding and nucleic acid-binding protein YchF	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Universal GTPases	279882	278782	-
mnas_218	HNH endonuclease family protein			279918	281205	-	Unclassified: Role category not yet assigned	**peg.317**	hypothetical protein	- none -	281205	279919	-
mnas_219	adenine-specific methyltransferase EcoRI family protein			281191	281860	-	Unknown: Enzymes of unknown specificity	**peg.318**	Phage protein	- none -	281860	281192	-
mnas_220	DNA (cytosine-5-)-methyltransferase family protein (EC 2.1.1.37)	*dcm*	GO:0003886, GO:0006304	282059	283031	+	DNA metabolism: DNA replication, recombination, and repair	**peg.319**	DNA-cytosine methyltransferase (EC 2.1.1.37)	**Category**: DNA Metabolism **Subcategory**: DNA repair **Subsystem**: DNA repair, bacterial	282060	283031	+
mnas_221	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	283064	283388	+	Unknown: General	**peg.320**	hypothetical protein	- none -	283065	283388	+
mnas_222	fic/DOC family protein			283395	284010	+	Unclassified: Role category not yet assigned	**peg.321**	hypothetical protein	- none -	283396	284010	+
mnas_223	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	284169	285819	+	Unknown: General	**peg.322**	predicted coding region	- none -	284170	285819	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_224	histidine--tRNA ligase (EC 6.1.1.21)	*hisS*	GO:0004821, GO:0005737, GO:0006427	285808	287191	+	Protein synthesis: tRNA aminoacylation	**peg.323**	Histidyl-tRNA synthetase (EC 6.1.1.21)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: tRNA aminoacylation, His	285809	287191	+
mnas_225	OB-fold nucleic acid binding domain protein			287201	287780	+	Unknown: General Hypothetical						
mnas_226	hemolysin C domain protein	*hlyC*	GO:0050660, GO:0003824, GO:0055114	287782	288106	+	Transport and binding proteins: Unknown substrate, Unknown: Enzymes of unknown specificity	**peg.333**	hypothetical protein	- none -	287891	288106	+
mnas_227	Probable spermidine/putrescine/ ABC transporter substrate	*potD*	GO:0008150, GO:0003674, GO:0005575	288456	290397	+	Unknown: General	**peg.334**	Probable spermidine/putrescine substrate binding protein in Mollicutes	- none -	288508	290397	+
mnas_228	ribosomal protein S20	*rpsT*	GO:0000312, GO:0003735, GO:0006412, GO:0022627	290447	290708	+	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.335**	SSU ribosomal protein S20p	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome SSU bacterial	290448	290708	+
mnas_229	thymidine kinase (EC 2.7.1.21)	*tdk*	GO:0046872, GO:0005524, GO:0006260, GO:0004797, GO:0005737	290785	291346	-	Cellular processes: Other, Transport and binding proteins: Unknown substrate, Unknown: Enzymes of unknown specificity	**peg.336**	Thymidine kinase (EC 2.7.1.21)	- none -	291346	290786	-
mnas_230	sporulation Regulator WhiA C terminal domain protein			291364	292240	-	Unknown: Conserved	**peg.337**	FIG001886: Cytoplasmic hypothetical protein	- none -	292231	291365	-
mnas_231	translation initiation factor IF-3	*infC*	GO:0003743, GO:0006413	292734	293247	+	Protein synthesis: Translation factors	**peg.338**	Translation initiation factor 3	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Translation initiation factors bacterial **Category**: Virulence, Disease and Defense **Subcategory**: Invasion and intracellular resistance **Subsystem**: Mycobacterium virulence operon involved in protein synthesis (LSU ribosomal proteins)	292735	293247	+

IGS								RAST					
Gene_id	Common_name	Symbol[1]	GO terms[2]	fmin[3]	fmax[3]	S[4]	TIGR_roles	Gene_id	Function	Subsystem	Start[5]	Stop[5]	S[4]
mnas_232	ribosomal protein L35	*rpmI*	GO:0000315, GO:0003735, GO:0006412, GO:0022625	293227	293416	+	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.339**	LSU ribosomal protein L35p	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial **Category**: Virulence, Disease and Defense **Subcategory**: Invasion and intracellular resistance **Subsystem**: Mycobacterium virulence operon involved in protein synthesis (LSU ribosomal proteins)	293228	293416	+
mnas_233	ribosomal protein L20	*rplT*	GO:0000315, GO:0003735, GO:0006412, GO:0022625	293473	293827	+	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.340**	LSU ribosomal protein L20p	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial **Category**: Virulence, Disease and Defense **Subcategory**: Invasion and intracellular resistance **Subsystem**: Mycobacterium virulence operon involved in protein synthesis (LSU ribosomal proteins)	293474	293827	+
mnas_234	DHHA1 domain protein			294264	295239	-	Unknown: General Hypothetical	**peg.341**	FIG146085: 3'-to-5' oligoribonuclease A, Bacillus type	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: RNA processing and degradation, bacterial	295239	294265	-
mnas_235	DHHA1 domain protein			295238	296201	-	Unknown: General Hypothetical	**peg.342**	FIG146085: 3'-to-5' oligoribonuclease A, Bacillus type	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: RNA processing and degradation, bacterial	296201	295239	-
mnas_236	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	296202	297330	-	Unknown: General	**peg.343**	hypothetical protein	- none -	297330	296203	-
mnas_237	HAD super, subIIIB family protein			297495	298860	+	Unknown: Enzymes of unknown specificity	**peg.344**	Multiple banded antigen	- none -	297496	298860	+
mnas_238	branched-chain amino acid transport system / permease component family protein			298946	299903	-	Transport and binding proteins: Other	**peg.345**	Unspecified monosaccharide ABC transport system, permease component 2	- none -	299903	298947	-
mnas_239	Putative sugar ABC transporter permease protein			299902	301714	-	Transport and binding proteins: Other	**peg.346**	Sugar ABC transporter, permease protein	- none -	301636	299903	-

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_240	spermidine/putrescine ABC transporter permease PotB domain protein	*potB*	GO:0006810, GO:0016021, GO:0005886, GO:0005215	301718	302042	-	Transport and binding proteins: Unknown substrate						
mnas_241	putative permidine/putrescine transport system ATP-binding protein (EC 3.6.3.31)	*potA*		302025	302700	-	Transport and binding proteins: Unknown substrate	**peg.353**	Putrescine transport ATP-binding protein PotA (TC 3.A.1.11.1)	- none -	303408	302026	-
mnas_242	putative spermidine/putrescine import ATP-binding protein PotA		GO:0006810, GO:0005524, GO:0016820, GO:0016787, GO:0000166, GO:0006200, GO:0043190, GO:0005886, GO:0015417, GO:0016887, GO:0005215	302678	303149	-	Cellular processes: Other, Transport and binding proteins: Unknown substrate, Disrupted reading frame: NULL, Unknown: Enzymes of unknown specificity						
mnas_243	spermidine/putrescine import ATP-binding PotA domain protein (EC 3.6.3.31)	*potA*	GO:0006810, GO:0005524, GO:0016820, GO:0016787, GO:0000166, GO:0006200, GO:0017111, GO:0043190, GO:0005886, GO:0015417, GO:0016887, GO:0005215	303195	303408	-	Cellular processes: Other, Transport and binding proteins: Unknown substrate, Disrupted reading frame: NULL, Unknown: Enzymes of unknown specificity						
mnas_244	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	303450	303588	-	Unknown: General	**peg.354**	hypothetical protein	- none -	303588	303451	-
mnas_245	DNA polymerase III, alpha subunit (EC 2.7.7.7)	*dnaE*	GO:0003887, GO:0006260, GO:0009360	303755	306668	+	DNA metabolism: DNA replication, recombination, and repair	**peg.355**	DNA polymerase III alpha subunit (EC 2.7.7.7)	**Category**: DNA Metabolism **Subcategory**: DNA replication **Subsystem**: DNA-replication	303756	306668	+
mnas_246	Putative DNA polymerase I (EC 2.7.7.7)			306660	307536	+	DNA metabolism: Degradation of DNA	**peg.356**	DNA polymerase I (EC 2.7.7.7)	**Category**: DNA Metabolism **Subcategory**: DNA repair **Subsystem**: DNA Repair Base Excision **Subcategory**: DNA replication **Subsystem**: DNA-replication	306661	307536	+
mnas_247	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	307811	308810	+	Unknown: General	**peg.357**	Helicase loader DnaB	- none -	307812	308810	+
mnas_248	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	308877	309813	+	Unknown: General	**peg.358**	Putative DNA helicase	- none -	308878	309813	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_249	methionyl-tRNA formyltransferase (EC 2.1.2.9)	*fmt*	GO:0004479, GO:0005737, GO:0006431	309803	310640	+	Protein synthesis: tRNA aminoacylation	**peg.359**	Methionyl-tRNA formyltransferase (EC 2.1.2.9)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Translation initiation factors bacterial **Subsystem**: Conserved gene cluster associated with Met-tRNA formyltransferase	309804	310640	+
mnas_250	glyceraldehyde-3-phosphate dehydrogenase, type I (EC 1.2.1.-)	*gap*	GO:0006094, GO:0006096, GO:0008943, GO:0019682	310772	311777	+	Energy metabolism: Glycolysis/gluconeogenesis	**peg.360**	NAD-dependent glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12)	**Category**: Carbohydrates **Subcategory**: Central carbohydrate metabolism **Subsystem**: Glycolysis and Gluconeogenesis	310773	311777	+
mnas_251	tRNA uridine 5-carboxymethylaminomethyl modification enzyme GidA	*gidA*	GO:0006400, GO:0016740	311865	313701	+	Protein synthesis: tRNA and rRNA base modification	**peg.361**	tRNA uridine 5-carboxymethylaminomethyl modification enzyme GidA	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: mnm5U34 biosynthesis bacteria **Subsystem**: Cell Division Subsystem including YidCD **Subsystem**: RNA modification and chromosome partitioning cluster	311866	313701	+
mnas_252	putative type II DNA modification enzyme		GO:0003677, GO:0008168, GO:0016740, GO:0032259	313720	314827	+	Cellular processes: Other, Transport and binding proteins: Unknown substrate, Unknown: Enzymes of unknown specificity	**peg.362**	putative type II DNA modification enzyme (methyltransferase)	- none -	313721	314827	+
mnas_253	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	314816	315296	+	Unknown: General						
mnas_254	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	315304	315889	+	Unknown: General	**peg.367**	hypothetical protein	- none -	315338	315889	+
mnas_255	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	315891	316371	+	Unknown: General	**peg.368**	Phage excisionase	- none -	3158141	316371	+
mnas_256	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	316382	316541	-	Unknown: General	**peg.369**	hypothetical protein	- none -	316541	316383	-
mnas_257	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	316542	316881	-	Unknown: General	**peg.370**	hypothetical protein	- none -	316881	316543	-
mnas_258	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	316897	318844	-	Unknown: General	**peg.371**	massive surface protein MspJ';	- none -	318727	316898	-
mnas_259	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	318845	320141	-	Unknown: General	**peg.372**	Phage protein	- none -	320141	318846	-

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_260	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	320145	321108	-	Unknown: General	**peg.373**	Phage protein	- none -	321108	320146	-
mnas_261	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	321122	321608	-	Unknown: General	**peg.374**	Phage protein	- none -	321608	321123	-
mnas_262	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	321649	323716	-	Unknown: General	**peg.375**	Phage protein	- none -	323716	321650	-
mnas_263	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	323690	324881	-	Unknown: General	**peg.376**	Phage protein	- none -	324881	323691	-
mnas_264	putative marRP			324867	326547	-	Unclassified: Role category not yet assigned	**peg.377**	MarRP	- none -	326547	324868	-
mnas_265	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	326551	327529	-	Unknown: General	**peg.378**	hypothetical protein	- none -	327529	326552	-
mnas_266	dnaB-like helicase C terminal domain protein			327515	328475	-	Unknown: General Hypothetical	**peg.379**	Replicative DNA helicase (DnaB) (EC 3.6.4.12)	**Category**: DNA Metabolism **Subcategory**: DNA replication **Subsystem**: DNA-replication	328106	327516	-
mnas_267	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	328489	329512	+	Unknown: General	**peg.382**	hypothetical protein	- none -	328703	329512	+
mnas_268	glycine--tRNA ligase (EC 6.1.1.14)	*glyS*	GO:0004820, GO:0006426, GO:0009345	329639	330980	+	Protein synthesis: tRNA aminoacylation	**peg.383**	Glycyl-tRNA synthetase (EC 6.1.1.14)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: tRNA aminoacylation, Gly	329616	330980	+
mnas_269	DNA primase (EC 2.7.7.-)	*dnaG*	GO:0003896, GO:0006269	331007	332948	+	DNA metabolism: DNA replication, recombination, and repair	**peg.384**	DNA primase (EC 2.7.7.-)	**Category**: Cell Division and Cell Cycle **Subcategory**: no subcategory **Subsystem**: Macromolecular synthesis operon **Category**: DNA Metabolism **Subcategory**: DNA replication **Subsystem**: DNA-replication	331008	332948	+
mnas_270	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	332950	333211	+	Unknown: General	**peg.385**	hypothetical protein	- none -	332951	333211	+
mnas_271	RNA polymerase sigma factor, sigma-70 family protein	*rpoD*		333300	334452	+	Transcription: DNA-dependent RNA polymerase	**peg.386**	RNA polymerase sigma factor RpoD	**Category**: RNA Metabolism **Subcategory**: Transcription **Subsystem**: Transcription initiation, bacterial sigma factors **Category**: Cell Division and Cell Cycle **Subcategory**: no subcategory **Subsystem**: Macromolecular synthesis operon	333301	334452	+
mnas_272	NIF3 family protein			334459	335239	+	Unclassified: Role category not yet assigned	**peg.387**	UPF0135 protein Bsu YqfO	- none -	334460	335239	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_273	AAA ATPase, central region		GO:0005524, GO:0017111	335501	336722	+	Transport and binding proteins: Unknown substrate, Unknown: Enzymes of unknown specificity	**peg.388**	ATPase, AAA family	- none -	335502	336722	+
mnas_274	phenylalanine--tRNA ligase, alpha subunit (EC 6.1.1.20)	*pheS*	GO:0004826, GO:0006432, GO:0009328	336721	337675	+	Protein synthesis: tRNA aminoacylation	**peg.389**	Phenylalanyl-tRNA synthetase alpha chain (EC 6.1.1.20)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: tRNA aminoacylation, Phe	336722	337675	+
mnas_275	uracil DNA glycosylase superfamily protein			337664	337904	+	DNA metabolism: DNA replication, recombination, and repair	**peg.390**	Uracil-DNA glycosylase, family 1	**Category**: DNA Metabolism **Subcategory**: DNA repair **Subsystem**: Uracil-DNA glycosylase **Subsystem**: DNA Repair Base Excision	337665	338322	+
mnas_276	uracil DNA glycosylase superfamily protein			337980	338322	+	DNA metabolism: DNA replication, recombination, and repair						
mnas_277	tRNA synthetase B5 domain protein		GO:0004826, GO:0006432	338330	340490	+	Protein synthesis: tRNA aminoacylation	**peg.391**	Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: tRNA aminoacylation, Phe	338331	340490	+
mnas_278	serine hydroxymethyltransferase (EC 2.1.2.1)	*glyA*		340489	341647	+	Surface structures: Serine family, Biosynthesis of cofactors, prosthetic groups, and carriers: Folic acid, Purines, pyrimidines, nucleosides, and nucleotides: Nucleotide and nucleoside interconversions						
mnas_279	hsp70 family protein			341651	342440	-	Protein fate: Protein folding and stabilization						
mnas_280	grpE family protein			342528	343383	-	Protein fate: Protein folding and stabilization	**peg.393**	Heat shock protein GrpE	**Category**: Protein Metabolism **Subcategory**: Protein folding **Subsystem**: Protein chaperones **Category**: Stress Response **Subcategory**: Heat shock **Subsystem**: Heat shock dnaK gene cluster extended	343383	342529	-
mnas_281	hrcA C terminal domain protein			343397	344435	-	Unknown: General Hypothetical	**peg.394**	Heat-inducible transcription repressor HrcA	**Category**: Stress Response **Subcategory**: Heat shock **Subsystem**: Heat shock dnaK gene cluster extended	344435	343398	-
mnas_282	putative lipoprotein		GO:0008150, GO:0003674, GO:0016020	344704	346750	+	Cell envelope: Other	**peg.395**	massive surface protein MspC	- none -	344705	346750	+

IGS								RAST					
Gene_id	Common_name	Symbol[1]	GO terms[2]	fmin[3]	fmax[3]	S[4]	TIGR_roles	Gene_id	Function	Subsystem	Start[5]	Stop[5]	S[4]
mnas_283	phosphocarrier protein HPr (EC 2.7.11.-)	*ptsH*	GO:0006355, GO:0016301, GO:0006351, GO:0005351, GO:0009401, GO:0005737	346838	347102	-	Cellular processes: Other, Transport and binding proteins: Unknown substrate, Signal transduction: PTS, Unknown: Enzymes of unknown specificity	**peg.396**	Phosphotransferase system, phosphocarrier protein HPr	**Category**: Regulation and Cell signaling **Subcategory**: no subcategory **Subsystem**: HPr catabolite repression system	347102	346839	-
mnas_284	phospholipase D family protein			347174	348692	-	Fatty acid and phospholipid metabolism: Degradation	**peg.397**	Cardiolipin synthetase (EC 2.7.8.-)	**Category**: Fatty Acids, Lipids, and Isoprenoids **Subcategory**: Phospholipids **Subsystem**: Cardiolipin synthesis **Subsystem**: Glycerolipid and Glycerophospholipid Metabolism in Bacteria	348692	347175	-
mnas_285	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	348732	348978	-	Unknown: Conserved	**peg.398**	hypothetical protein	- none -	348978	348733	-
mnas_286	Putative ribose-5-phosphate isomerase (EC 5.3.1.6)			348980	349421	-	Unknown: Enzymes of unknown specificity	**peg.399**	Ribose 5-phosphate isomerase B (EC 5.3.1.6)	**Category**: Carbohydrates **Subcategory**: Central carbohydrate metabolism **Subsystem**: Pentose phosphate pathway	349421	348981	-
mnas_287	ABC transporter family protein			349525	351553	+	Transport and binding proteins: Unknown substrate	**peg.400**	ABC transporter atp-binding and permease protein (MDR homolog)	- none -	349526	351553	+
mnas_288	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	351584	351914	+	Unknown: General	**peg.401**	hypothetical protein	- none -	351600	351914	+
mnas_289	putative rRNA maturation factor YbeY		GO:0008150, GO:0046872	351954	352437	+	Unknown: Conserved	**peg.402**	Metal-dependent hydrolase YbeY, involved in rRNA and/or ribosome maturation and assembly	- none -	351955	352437	+
mnas_290	cytidine deaminase (EC 3.5.4.5)	*cdd*	GO:0004126, GO:0005737, GO:0008655	352423	352822	+	Purines, pyrimidines, nucleosides, and nucleotides: Salvage of nucleosides and nucleotides	**peg.403**	Cytidine deaminase (EC 3.5.4.5)	- none -	352424	352822	+
mnas_291	GTP-binding protein Era	*era*	GO:0003924, GO:0005525, GO:0019843, GO:0042254, GO:0043022	352823	353726	+	Cellular processes: Cell division, Regulatory functions: RNA interactions	**peg.404**	GTP-binding protein Era	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Universal GTPases	352824	353726	+
mnas_292	putative SPOUT methyltransferase family protein			353767	353929	-	Unknown: Enzymes of unknown specificity						
mnas_293	ATP synthase alpha/beta family, nucleotide-binding domain protein			353957	354818	+	Unknown: Enzymes of unknown specificity	**peg.405**	ATP synthase alpha chain (EC 3.6.3.14)	- none -	353958	354818	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_294	ATP synthase F1, beta subunit (EC 3.6.3.14)	*atpD*	GO:0000275, GO:0015986, GO:0045261, GO:0045262, GO:0046933	354817	356179	+	Energy metabolism: ATP-proton motive force interconversion	**peg.406**	ATP synthase beta chain (EC 3.6.3.14)	- none -	354818	356179	+
mnas_295	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	356190	358470	+	Unknown: Conserved	**peg.407**	Prolipoprotein	- none -	356191	358470	+
mnas_296	putative lipoprotein		GO:0008150, GO:0003674, GO:0016020	358513	360538	-	Cell envelope: Other	**peg.408**	massive surface protein MspG	- none -	360538	358514	-
mnas_297	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	360673	362746	+	Unknown: Conserved	**peg.409**	FIG00833989: hypothetical protein	- none -	360674	362746	+
mnas_298	phosphotransferase enzyme family protein			362821	363574	+	Unknown: Enzymes of unknown specificity	**peg.410**	Choline kinase family	- none -	362822	363574	+
mnas_299	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	363701	363899	+	Unknown: General	**peg.411**	hypothetical protein	- none -	363717	363899	+
mnas_300	putative membrane protein		GO:0008150, GO:0003674, GO:0016020	363888	364953	+	Cell envelope: Other	**peg.412**	hypothetical protein	- none -	363889	364953	+
mnas_301	putative membrane protein		GO:0008150, GO:0003674, GO:0016020	365046	366111	+	Cell envelope: Other	**peg.413**	hypothetical protein	- none -	365047	366111	+
mnas_302	putative lipoprotein		GO:0008150, GO:0003674, GO:0016020	366212	366557	-	Cell envelope: Other						
mnas_303	putative lipoprotein		GO:0008150, GO:0003674, GO:0016020	366761	368090	-	Cell envelope: Other	**peg.433**	hypothetical protein	- none -	368090	366762	-
mnas_304	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	368166	377586	+	Unknown: General	**peg.434**	Putative peptidoglycan bound protein (LPXTG motif) Lmo1799 homolog	- none -	368242	377586	+
mnas_305	ribonuclease HII family protein			378227	378746	-	Transcription: Degradation of RNA						
mnas_306	phosphorylase superfamily protein			377758	378235	+	Unclassified: Role category not yet assigned	**peg.435**	HMP-PP hydrolase (pyridoxal phosphatase) Cof, detected in genetic screen for thiamin metabolic genes (PMID:15292217)	- none -	379658	378750	-
mnas_307	HAD hydrolase, IIB family protein		GO:0008152, GO:0016787	378749	379658	-	Unknown: Enzymes of unknown specificity	**peg.436**	Dihydrofolate reductase (EC 1.5.1.3)	**Category**: Cofactors, Vitamins, Prosthetic Groups, Pigments **Subcategory**: Folate and pterines **Subsystem**: Folate biosynthesis	380268	379741	-
mnas_308	dihydrofolate reductase family protein			379740	380268	-	Biosynthesis of cofactors, prosthetic groups, and carriers: Folic acid						

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_309	UMP kinase (EC 2.7.4.22)	*pyrH*	GO:0015949, GO:0033862	380436	381156	+	Purines, pyrimidines, nucleosides, and nucleotides: Nucleotide and nucleoside interconversions	**peg.437**	Uridine monophosphate kinase (EC 2.7.4.22)	**Subsystem**: Ribosome recycling related cluster	380437	381156	+
mnas_310	ribosome recycling factor	*frr*	GO:0005737, GO:0006412, GO:0008135	381155	381707	+	Protein synthesis: Translation factors	**peg.438**	Ribosome recycling factor	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Translation termination factors bacterial **Subsystem**: Ribosome recycling related cluster,	381156	381707	+
mnas_311	tRNA modification GTPase TrmE (EC 3.6.-.-)	*trmE*	GO:0003924, GO:0006400	381748	383146	+	Protein synthesis: tRNA and rRNA base modification	**peg.439**	GTPase and tRNA-U34 5-formylation enzyme TrmE	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: mnm5U34 biosynthesis bacteria **Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Universal GTPases **Subsystem**: Cell Division Subsystem including YidCD **Subsystem**: RNA modification and chromosome partitioning cluster **Subsystem**: RNA modification cluster	381749	383146	+
mnas_312	tatD related DNase family protein			383147	383942	+	Unknown: Enzymes of unknown specificity	**peg.440**	Putative deoxyribonuclease YcfH	**Category**: DNA Metabolism **Subcategory**: no subcategory **Subsystem**: YcfH	383148	383942	+
mnas_313	dimethyladenosine transferase (EC 2.1.1.-)	*ksgA*	GO:0000154, GO:0000179	383941	384724	+	Protein synthesis: tRNA and rRNA base modification	**peg.441**	SSU rRNA (adenine(1518)-N(6)/adenine(1519)-N(6))-dimethyltransferase (EC 2.1.1.182)	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: RNA methylation **Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome biogenesis bacterial	383942	384724	+
mnas_314	HIT domain protein			384750	385098	-	Unknown: Enzymes of unknown specificity	**peg.442**	HIT-family hydrolase protein	- none -	385098	384751	-
mnas_315	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	385097	386327	-	Unknown: General	**peg.443**	P60-like lipoprotein	- none -	386327	385098	-
mnas_316	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	386326	388534	-	Unknown: General	**peg.444**	*M. hominis* p80-related protein	- none -	388534	386327	-
mnas_317	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	388650	389151	+	Unknown: General	**peg.445**	hypothetical protein	- none -	388615	389151	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_318	recF/RecN/SMC N terminal domain protein			389204	391247	+	Cellular processes: Cell division	**peg.446**	Chromosome partition protein smc	**Category**: DNA Metabolism **Subcategory**: no subcategory **Subsystem**: DNA structural proteins, bacterial	389790	391247	+
mnas_319	GTP-binding protein LepA (EC 3.6.5.-)	*lepA*	GO:0005525	391496	393299	+	Unknown: General Hypothetical	**peg.447**	Translation elongation factor LepA	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Universal GTPases **Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Translation elongation factors bacterial **Category**: Stress Response **Subcategory**: Heat shock **Subsystem**: Heat shock dnaK gene cluster extended	391497	393299	+
mnas_320	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	393377	395834	-	Unknown: Conserved	**peg.448**	hypothetical protein	- none -	395783	393378	-
mnas_321	peptidase M13 family protein			396052	397960	+	Protein fate: Degradation of proteins, peptides, and glycopeptides	**peg.449**	Neutral endopeptidase O (EC 3.4.24.-)	- none -	396053	397960	+
mnas_322	OB-fold nucleic acid binding domain protein			399642	400500	-	Unknown: General Hypothetical						
mnas_323	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	398311	399646	+	Unknown: General						
mnas_324	virulence factor BrkB family protein			400523	401621	-	Unclassified: Role category not yet assigned	**peg.450**	FIG00835188: hypothetical protein	- none -	401621	400524	-
mnas_325	aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase, B subunit (EC 6.3.5.-)	*gatB*	GO:0006424, GO:0030956, GO:0050567	401604	403026	-	Protein synthesis: tRNA aminoacylation	**peg.451**	Aspartyl-tRNA(Asn) amidotransferase subunit B (EC 6.3.5.6) @ Glutamyl-tRNA(Gln) amidotransferase subunit B (EC 6.3.5.7)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: tRNA aminoacylation, Asp and Asn; **Subsystem**: tRNA aminoacylation, Glu and Gln	403026	401605	-
mnas_326	amidase family protein			403027	404350	-	Unknown: Enzymes of unknown specificity	**peg.452**	Aspartyl-tRNA(Asn) amidotransferase subunit A (EC 6.3.5.6) @ Glutamyl-tRNA(Gln) amidotransferase subunit A (EC 6.3.5.7)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: tRNA aminoacylation, Asp and Asn **Subsystem**: tRNA aminoacylation, Glu and Gln	404350	403028	-

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_327	glu-tRNAGln amidotransferase C subunit			404349	404652	-	Protein synthesis: tRNA aminoacylation	**peg.453**	Aspartyl-tRNA(Asn) amidotransferase subunit C (EC 6.3.5.6) @ Glutamyl-tRNA(Gln) amidotransferase subunit C (EC 6.3.5.7)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: tRNA aminoacylation, Asp and Asn **Subsystem**: tRNA aminoacylation, Glu and Gln	404652	404350	-
mnas_328	RNA pseudouridylate synthase family protein			404658	405534	-	Protein synthesis: tRNA and rRNA base modification	**peg.454**	Ribosomal large subunit pseudouridine synthase C (EC 4.2.1.70)	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: RNA pseudouridine syntheses **Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome biogenesis bacterial	405534	404659	-
mnas_329	segregation and condensation protein B	*scpB*	GO:0005515, GO:0007059, GO:0030261	405576	406152	-	Cellular processes: Cell division, DNA metabolism: Chromosome-associated proteins	**peg.455**	Segregation and condensation protein B	CBSS-314276.3.peg.1499	406152	405577	-
mnas_330	segregation and condensation protein A	*scpA*		406141	406906	-	Cellular processes: Cell division, DNA metabolism: Chromosome-associated proteins	**peg.456**	Segregation and condensation protein A	CBSS-314276.3.peg.1499	406906	406142	-
mnas_331	Putative 1-acyl-sn-glycerol-3-phosphate acyltransferase (EC 2.3.1.51)	*plsC*		406895	407666	-	Unknown: Enzymes of unknown specificity	**peg.457**	1-acyl-sn-glycerol-3-phosphate acyltransferase (EC 2.3.1.51)	**Category**: Fatty Acids, Lipids, and Isoprenoids **Subcategory**: Phospholipids **Subsystem**: Glycerolipid and Glycerophospholipid Metabolism in Bacteria	407666	406896	-
mnas_332	phosphopantetheine--transferase domain protein			407658	408033	-	Protein fate: Protein modification and repair	**peg.458**	Holo-[acyl-carrier protein] synthase (EC 2.7.8.7)	- none -	407985	407659	-
mnas_333	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	408072	409734	-	Unknown: General	**peg.459**	hypothetical protein	- none -	409734	408073	-
mnas_334	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	410384	419783	-	Unknown: General	**peg.460**	glycosyl hydrolase, family 31/fibronectin type III domain protein	- none -	419759	410385	-
mnas_335	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	419994	420525	+	Unknown: General	**peg.461**	unknown; predicted coding region	- none -	420022	420525	+
mnas_336	transcription elongation factor GreA domain protein	*greA*	GO:0003711, GO:0006354	420668	421184	+	Transcription: Transcription factors	**peg.462**	Transcription elongation factor GreA	**Category**: RNA Metabolism **Subcategory**: Transcription **Subsystem**: Transcription factors bacterial	420669	421184	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_337	deoxynucleoside kinase family protein			421303	421960	+	Purines, pyrimidines, nucleosides, and nucleotides: Nucleotide and nucleoside interconversions	**peg.463**	Deoxyadenosine kinase (EC 2.7.1.76) / Deoxyguanosine kinase (EC 2.7.1.113)	**Category**: Nucleosides and Nucleotides **Subcategory**: Purines **Subsystem**: Purine conversions **Subsystem**: Purine conversions	421304	421960	+
mnas_338	hypoxanthine phosphoribosyltransferase (EC 2.4.2.8)	*hpt*	GO:0004422, GO:0006166	421965	422574	+	Purines, pyrimidines, nucleosides, and nucleotides: Salvage of nucleosides and nucleotides	**peg.464**	Hypoxanthine-guanine phosphoribosyltransferase (EC 2.4.2.8)	**Category**: Cofactors, Vitamins, Prosthetic Groups, Pigments **Subcategory**: Folate and pterines **Subsystem**: Folate biosynthesis cluster **Category**: Nucleosides and Nucleotides **Subcategory**: Purines **Subsystem**: Purine conversions **Subsystem**: Cell division-ribosomal stress proteins cluster	421966	422574	+
mnas_339	preprotein translocase, SecA subunit	*secA*	GO:0009898, GO:0015450, GO:0016887, GO:0043952	422797	425623	+	Protein fate:Protein and peptide secretion and trafficking	**peg.465**	Protein export cytoplasm protein SecA ATPase RNA helicase (TC 3.A.5.1.1)	- none -	422798	425623	+
mnas_340	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	425653	426412	+	Unknown: General	**peg.466**	hypothetical protein	- none -	425654	426412	+
mnas_341	ABC transporter family protein			426586	428206	+	Transport and binding proteins: Unknown substrate	**peg.467**	ABC transporter ATP-binding protein uup	- none -	426587	428206	+
mnas_342	RNA methyltransferase, RsmE family protein (EC 2.1.1.-)		GO:0016436, GO:0031167	428205	428886	+	Unknown: Conserved	**peg.468**	Ribosomal RNA small subunit methyltransferase E (EC 2.1.1.-)	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: RNA methylation **Category**: Stress Response **Subcategory**: Heat shock **Subsystem**: Heat shock dnaK gene cluster extended	428206	428886	+
mnas_343	ribosomal protein L13	*rplM*	GO:0000315, GO:0003735, GO:0006412, GO:0022625	429035	429470	+	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.469**	LSU ribosomal protein L13p (L13Ae)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial	429036	429470	+
mnas_344	30S ribosomal protein S9	*rpsI*	GO:0005840, GO:0006412, GO:0003735	429469	429862	+	Cellular processes: Other	**peg.470**	SSU ribosomal protein S9p (S16e)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome SSU bacterial	429470	429862	+
mnas_345	mycoplasma MG185/MG260 family protein			429892	431416	-	Unclassified: Role category not yet assigned	**peg.471**	hypothetical protein	- none -	430488	430892	+
								peg.472	hypothetical protein	- none -	431106	431438	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_346	antibiotic biosynthesis monooxygenase family protein			431616	431901	-	Cellular processes: Toxin production and resistance	**peg.473**	hypothetical protein	- none -	431901	431617	-
mnas_347	methionine--tRNA ligase domain protein (EC 6.1.1.10)	*metG*	GO:0005524, GO:0000166, GO:0004825, GO:0006431, GO:0005737, GO:0016874	431917	432442	-	Cellular processes: Other, Transport and binding proteins: Unknown substrate, Unknown: Enzymes of unknown specificity	**peg.474**	Methionyl-tRNA synthetase (EC 6.1.1.10)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: tRNA aminoacylation, Met	433472	431918	-
mnas_348	methionine--tRNA ligase (EC 6.1.1.10)	*metG*	GO:0004825, GO:0005737, GO:0006431	432401	433472	-	Protein synthesis: tRNA aminoacylation						
mnas_349	conserved hypothetical 95 family protein			433474	434278	-	Unclassified: Role category not yet assigned	**peg.475**	tRNA (adenine37-N(6))-methyltransferase TrmN6 (EC 2.1.1.223)	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: RNA methylation	434278	433475	-
mnas_350	ribonuclease R (EC 3.1.13.1)	*rnr*	GO:0006401, GO:0008997	434243	436445	-	Transcription: Degradation of RNA	**peg.476**	3'-to-5' exoribonuclease RNase R	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: RNA processing and degradation, bacterial	436445	434244	-
mnas_351	preprotein translocase, SecG subunit	*secG*	GO:0005887, GO:0015450, GO:0043952	436455	436695	-	Protein fate:Protein and peptide secretion and trafficking	**peg.477**	hypothetical protein	- none -	436695	436456	-
mnas_352	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	437093	439496	+	Unknown: General	**peg.478**	hypothetical protein	- none -	437094	439496	+
mnas_353	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	439555	439645	-	Unknown: General						
mnas_354	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	439651	441286	-	Unknown: General	**peg.479**	hypothetical protein	- none -	441313	441984	+
mnas_355	putative membrane protein		GO:0008150, GO:0003674, GO:0016020	441312	441984	+	Cell envelope: Other						
mnas_356	tyrosine--tRNA ligase (EC 6.1.1.1)	*tyrS*	GO:0004831, GO:0005737, GO:0006437	442022	443264	-	Protein synthesis: tRNA aminoacylation	**peg.480**	Tyrosyl-tRNA synthetase (EC 6.1.1.1)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: tRNA aminoacylation, Tyr	443264	442023	-
mnas_357	putative tRNA binding domain protein			443266	443875	-	Unknown: General Hypothetical	**peg.481**	Phenylalanyl-tRNA synthetase domain protein (Bsu YtpR)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: tRNA aminoacylation, Phe	443875	443267	-
mnas_358	EDD, DegV family domain protein		GO:0005575, GO:0008150, GO:0016740	443900	444767	+	Unknown: General Hypothetical	**peg.482**	hypothetical protein	- none -	443997	444767	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_359	tRNA threonylcarbamoyl adenosine modification protein YjeE		GO:0005737, GO:0006400, GO:0016887	444760	445147	+	Unknown: Conserved	**peg.483**	TsaE protein, required for threonylcarbamoyladenosine t(6)A37 formation in tRNA	**Category**: Cell Wall and Capsule **Subcategory**: no subcategory **Subsystem**: YjeE	444761	445147	+
mnas_360	tRNA threonylcarbamoyl adenosine modification protein YeaZ	*yeaZ*		445156	445717	+	Transport and binding proteins: Cations and iron carrying compounds	**peg.484**	TsaB protein, required for threonylcarbamoyladenosine (t(6)A) formation in tRNA	**Category**: Cell Division and Cell Cycle **Subcategory**: no subcategory **Subsystem**: YgjD and YeaZ **Category**: Cell Wall and Capsule **Subcategory**: no subcategory **Subsystem**: YjeE **Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome biogenesis bacterial **Subsystem**: Bacterial RNA-metabolizing Zn-dependent hydrolases **Subsystem**: Conserved gene cluster associated with Met-tRNA formyltransferase	445157	445717	+
mnas_361	tRNA threonylcarbamoyl adenosine modification protein YgjD			445716	446646	+	Transport and binding proteins: Cations and iron carrying compounds	**peg.485**	TsaD/Kae1/Qri7 protein, required for threonylcarbamoyladenosine t(6)A37 formation in tRNA	**Category**: Cell Division and Cell Cycle **Subcategory**: no subcategory **Subsystem**: YgjD and YeaZ **Subsystem**: Macromolecular synthesis operon **Subsystem**: Bacterial RNA-metabolizing Zn-dependent hydrolases **Subsystem**: Conserved gene cluster associated with Met-tRNA formyltransferase	445717	446646	+
mnas_362	type I restriction modification DNA specificity domain protein			446816	448133	+	DNA metabolism: Restriction/modification	**peg.486**	Type I restriction-modification system, specificity subunit S (EC 3.1.21.3)	**Category**: DNA Metabolism **Subcategory**: no subcategory **Subsystem**: Type I Restriction-Modification **Subsystem**: Restriction-Modification System	446817	448133	+
mnas_363	type I restriction modification DNA specificity domain protein			448080	448788	+	DNA metabolism: Restriction/modification	**peg.487**	Type I restriction-modification system, specificity subunit S (EC 3.1.21.3)	**Category**: DNA Metabolism **Subcategory**: no subcategory **Subsystem**: Type I Restriction-Modification **Subsystem**: Restriction-Modification System	448372	448788	+
mnas_364	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	449190	449838	-	Unknown: General	**peg.488**	hypothetical protein	- none -	449375	449872	+

IGS								RAST					
Gene_id	Common_name	Symbol[1]	GO terms[2]	fmin[3]	fmax[3]	S[4]	TIGR_roles	Gene_id	Function	Subsystem	Start[5]	Stop[5]	S[4]
mnas_365	type III restriction-modification system methylation subunit domain protein (EC 2.1.1.72)	*sthIM*	GO:0003677, GO:0032775, GO:0008170, GO:0016740, GO:0032259, GO:0006306, GO:0009007	448982	449198	+	Cellular processes: Other, Transport and binding proteins: Unknown substrate, Unknown: Enzymes of unknown specificity						
mnas_366	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	450101	452267	-	Unknown: General	**peg.489**	hypothetical protein	- none -	452267	450102	-
mnas_367	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	452547	454830	-	Unknown: General	**peg.490**	VlhA.5.03	- none -	454830	452548	-
								peg.491	hypothetical protein	- none -	454832	454975	+
mnas_368	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	455138	457310	-	Unknown: General	**peg.492**	VlhA.4.08	- none -	457010	455139	-
mnas_369	ribosomal protein L19	*rplS*	GO:0000311, GO:0003735, GO:0006412, GO:0022625	457336	457693	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.493**	LSU ribosomal protein L19p	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial	457693	457337	-
mnas_370	tRNA (guanine(37)-N(1))-methyltransferase	*trmD*	GO:0006400, GO:0009019	457682	458390	-	Protein synthesis: tRNA and rRNA base modification	**peg.494**	tRNA (Guanine37-N1) -methyltransferase (EC 2.1.1.31)	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: RNA methylation **Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome biogenesis bacterial	458390	457683	-
mnas_371	ribosomal protein S16	*rpsP*	GO:0000314, GO:0003735, GO:0006412, GO:0022627	458389	458638	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.495**	SSU ribosomal protein S16p	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome SSU bacterial	458638	458390	-
mnas_372	RNA 2'-O ribose methyltransferase substrate binding family protein			458750	459512	-	Protein synthesis: tRNA and rRNA base modification	**peg.496**	23S rRNA (guanosine-2';-O-)-methyltransferase rlmB (EC 2.1.1.-)	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: RNA methylation	459512	458751	-
mnas_373	cysteine--tRNA ligase (EC 6.1.1.16)	*cysS*	GO:0004817, GO:0006423	459540	460776	-	Protein synthesis: tRNA aminoacylation	**peg.497**	Cysteinyl-tRNA synthetase (EC 6.1.1.16)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: tRNA aminoacylation, Cys	460692	459541	-
mnas_374	arginine--tRNA ligase (EC 6.1.1.19)	*argS*	GO:0004814, GO:0005737, GO:0006420	460892	462548	+	Protein synthesis: tRNA aminoacylation	**peg.498**	Arginyl-tRNA synthetase (EC 6.1.1.19)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: tRNA aminoacylation, Arg	460893	462548	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_375	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	462561	462825	+	Cellular processes: Other, Transport and binding proteins: Unknown substrate	**peg.499**	hypothetical protein	- none -	462562	462825	+
mnas_376	fructose-1,6-bisphosphate aldolase, class II (EC 4.1.2.13)	*fba*	GO:0004332, GO:0006096	462826	463696	+	Energy metabolism: Glycolysis/gluconeogenesis	**peg.500**	Fructose-bisphosphate aldolase class II (EC 4.1.2.13)	**Category**: Carbohydrates **Subcategory**: Central carbohydrate metabolism **Subsystem**: Glycolysis and Gluconeogenesis	462827	463696	+
mnas_377	eco57I restriction-modification methylase family protein			463734	464700	+	DNA metabolism: Restriction/modification	**peg.501**	Transcriptional regulator	- none -	463735	464700	+
mnas_378	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	464822	465893	+	Unknown: General Hypothetical	**peg.502**	Thermonuclease family protein	- none -	464823	465893	+
mnas_379	DNA gyrase subunit B (EC 5.99.1.3)	*gyrB*		465983	467939	+	DNA metabolism: Other	**peg.503**	DNA gyrase subunit B (EC 5.99.1.3)	**Category**: DNA Metabolism **Subcategory**: DNA replication **Subsystem**: DNA topoisomerases, Type II, ATP-dependent **Category**: Virulence, Disease and Defense **Subcategory**: Resistance to antibiotics and toxic compounds **Subsystem**: Resistance to fluoroquinolones **Subsystem**: Cell Division Subsystem including YidCD	465984	467931	+
mnas_380	tetrahydrofolate dehydrogenase/cyclohydrolase, NAD(P)-binding domain protein			468061	468208	-	Biosynthesis of cofactors, prosthetic groups, and carriers: Folic acid						
mnas_381	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	468224	476891	+	Unknown: General	**peg.515**	hypothetical protein	- none -	473356	472508	-
								peg.516	hypothetical protein	- none -	474529	473522	-
								peg.517	hypothetical protein	- none -	475507	474623	-
								peg.518	hypothetical protein	- none -	476803	476126	-
mnas_382	ribosomal protein L15	*rplO*	GO:0000315, GO:0003735, GO:0006412, GO:0022625	477052	477487	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.519**	LSU ribosomal protein L15p (L27Ae)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial	477487	477053	-
mnas_383	ribosomal protein S5	*rpsE*	GO:0000312, GO:0003735, GO:0006412, GO:0022627	477491	478136	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.520**	SSU ribosomal protein S5p (S2e)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome SSU bacterial	478136	477492	-

IGS								RAST					
Gene_id	Common_name	Symbol[1]	GO terms[2]	fmin[3]	fmax[3]	S[4]	TIGR_roles	Gene_id	Function	Subsystem	Start[5]	Stop[5]	S[4]
mnas_384	ribosomal protein L18	*rplR*	GO:0000315, GO:0003735, GO:0006412, GO:0022625	478135	478486	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.521**	LSU ribosomal protein L18p (L5e)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial	478486	478136	-
mnas_385	ribosomal protein L6	*rplF*	GO:0003735, GO:0006412, GO:0019843, GO:0022625	478510	479050	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.522**	LSU ribosomal protein L6p (L9e)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial	479050	478511	-
mnas_386	ribosomal S8 family protein			479060	479456	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.523**	SSU ribosomal protein S8p (S15Ae)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome SSU bacterial	479456	479061	-
mnas_387	30S ribosomal protein S14 type Z	*rpsZ*	GO:0005840, GO:0046872, GO:0006412, GO:0003735, GO:0019843	479455	479641	-	Cellular processes: Other, Transport and binding proteins: Unknown substrate	**peg.524**	SSU ribosomal protein S14p (S29e) @ SSU ribosomal protein S14p (S29e), zinc-dependent	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome SSU bacterial **Subsystem**: Ribosome SSU bacterial	479641	479456	-
mnas_388	50S ribosomal protein L5	*rplE*	GO:0005840, GO:0006412, GO:0000049, GO:0003735, GO:0019843	479640	480186	-	Cellular processes: Other	**peg.525**	LSU ribosomal protein L5p (L11e)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial	480186	479641	-
mnas_389	ribosomal protein L24	*rplX*	GO:0000315, GO:0003735, GO:0006412, GO:0022625	480187	480520	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.526**	LSU ribosomal protein L24p (L26e)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial	480520	480188	-
mnas_390	ribosomal protein L14	*rplN*	GO:0000315, GO:0003735, GO:0006412, GO:0022625	480538	480904	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.527**	LSU ribosomal protein L14p (L23e)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial	480904	480539	-
mnas_391	30S ribosomal protein S17	*rpsQ*	GO:0003735, GO:0006412, GO:0022627	480903	481173	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.528**	SSU ribosomal protein S17p (S11e)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome SSU bacterial	481173	480904	-
mnas_392	ribosomal protein L29	*rpmC*	GO:0000315, GO:0003735, GO:0006412, GO:0022625	481172	481364	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.529**	LSU ribosomal protein L29p (L35e)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial	481364	481146	-
mnas_393	ribosomal protein L16	*rplP*	GO:0000315, GO:0003735, GO:0006412, GO:0022625	481363	481789	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.530**	LSU ribosomal protein L16p (L10e)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial	481789	481364	-

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_394	ribosomal protein S3	*rpsC*	GO:0000312, GO:0003735, GO:0006412, GO:0022627	481788	482442	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.531**	SSU ribosomal protein S3p (S3e)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome SSU bacterial	482442	481789	-
mnas_395	ribosomal protein L22	*rplV*	GO:0000311, GO:0003735, GO:0006412, GO:0022625	482443	482779	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.532**	LSU ribosomal protein L22p (L17e)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial	482779	482444	-
mnas_396	ribosomal protein S19	*rpsS*	GO:0000314, GO:0003735, GO:0006412, GO:0022627	482778	483057	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.533**	SSU ribosomal protein S19p (S15e)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome SSU bacterial	483057	482779	-
mnas_397	ribosomal protein L2	*rplB*	GO:0000315, GO:0003735, GO:0006412, GO:0022625	483056	483902	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.534**	LSU ribosomal protein L2p (L8e)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial	483902	483057	-
mnas_398	50S ribosomal L23 domain protein	*rplW*	GO:0005840, GO:0000166, GO:0006412, GO:0003735, GO:0019843	484016	484490	-	Cellular processes: Other	**peg.535**	LSU ribosomal protein L23p (L23Ae)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial	484490	484017	-
mnas_399	ribosomal L4/L1 family protein			484489	484780	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.536**	LSU ribosomal protein L4p (L1e)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial	484780	484490	-
mnas_400	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	485007	486828	-	Unknown: General						
mnas_401	DNA topoisomerase 4 subunit A (EC 5.99.1.-)	*parC*	GO:0005524, GO:0003918, GO:0005694, GO:0006265, GO:0003700	487275	489867	+	Cellular processes: Other, Transport and binding proteins: Unknown substrate, Transcription: Transcription factors, Unknown: Enzymes of unknown specificity	**peg.537**	Topoisomerase IV subunit A (EC 5.99.1.-)	**Category**: DNA Metabolism **Subcategory**: DNA replication **Subsystem**: DNA topoisomerases, Type II, ATP-dependent **Category**: Virulence, Disease and Defense **Subcategory**: Resistance to antibiotics and toxic compounds **Subsystem**: Resistance to fluoroquinolones	487276	489867	+
mnas_402	DNA (cytosine-5-)-methyltransferase family protein (EC 2.1.1.37)	*dcm*	GO:0003886, GO:0006304	490036	490495	+	DNA metabolism: DNA replication, recombination, and repair, Disrupted reading frame: NULL	**peg.538**	DNA-cytosine methyltransferase (EC 2.1.1.37)	**Category**: DNA Metabolism **Subcategory**: DNA repair **Subsystem**: DNA repair, bacterial	490037	490495	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol[1]**	**GO terms[2]**	**fmin[3]**	**fmax[3]**	**S[4]**	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start[5]**	**Stop[5]**	**S[4]**
mnas_403	putative cytosine-specific methyltransferase		GO:0003677, GO:0009307, GO:0008168, GO:0006306, GO:0003886, GO:0016740, GO:0032259	490518	490869	+	Cellular processes: Other, Transport and binding proteins: Unknown substrate, Unknown: Enzymes of unknown specificity	**peg.539**	DNA-cytosine methyltransferase	- none -	490519	490869	+
mnas_404	type-2 restriction enzyme DpnII (EC 3.1.21.4)	*dpnB*	GO:0003677, GO:0009307, GO:0009036	491431	492319	-	Cellular processes: Other, Transport and binding proteins: Unknown substrate, Unknown: Enzymes of unknown specificity	**peg.540**	Type II restriction enzyme MjaIII (EC 3.1.21.4)	- none -	492319	491432	-
mnas_405	modification methylase DpnIIA (EC 2.1.1.72)	*dpnM*	GO:0009307, GO:0009007, GO:0003676	492319	493036	-	Cellular processes: Other, Transport and binding proteins: Unknown substrate, Unknown: Enzymes of unknown specificity	**peg.541**	DNA adenine methylase (EC 2.1.1.72)	- none -	492877	492320	-
mnas_406	ABC transporter transmembrane region family protein			493036	493804	-	Transport and binding proteins: Unknown substrate	**peg.542**	Lipid A export ATP-binding/permease protein MsbA	- none -	495533	493788	-
mnas_407	ABC transporter family protein			493787	495533	-	Transport and binding proteins: Unknown substrate						
mnas_408	bacterial regulatory , arsR family protein			495532	495898	-	Regulatory functions: DNA interactions	**peg.543**	arsenical resistance operon repressor family protein	- none -	495898	495533	-
mnas_409	AAA domain protein			495979	500236	-	Unknown: General Hypothetical	**peg.544**	Apolipoprotein N-acyltransferase (EC 2.3.1.-) in lipid-linked oligosaccharide synthesis cluster	- none -	500236	495980	-
mnas_410	Putative PTS system glucose-specific enzyme IIB component		GO:0008150, GO:0003674, GO:0005575	500247	500607	-	Transport and binding proteins: Unknown substrate, Signal transduction: PTS, Unknown: Enzymes of unknown specificity	**peg.545**	PTS system glucose-specific enzyme IIB component	- none -	500607	500248	-
mnas_411	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	500832	501000	-	Unknown: General						
mnas_412	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	501010	501196	-	Unknown: General						
mnas_413	oligoendopeptidase F (EC 3.4.24.-)	*pepF*	GO:0004222, GO:0006508	502226	504056	+	Protein fate: Degradation of proteins, peptides, and glycopeptides	**peg.546**	Oligoendopeptidase F (EC 3.4.24.-)	- none -	502227	504056	+
mnas_414	oligopeptide ABC transporter substrate-binding protein	*oppA*	GO:0008150, GO:0003674, GO:0005575	504068	507935	+	Transport and binding proteins: Unknown substrate	**peg.547**	Lipoprotein	- none -	504069	507935	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_415	Oligopeptide ABC transporter, permease protein OppB	*oppB*		507946	508705	+	Transport and binding proteins: Unknown substrate						
mnas_416	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	508709	509093	-	Unknown: General						
mnas_417	terminase-like family protein			509193	511155	-	Unclassified: Role category not yet assigned	**peg.548**	Phage terminase, large subunit	**Category**: Phage packaging machinPhages, Prophages, Transposable elements, Plasmids **Subcategory**: Phages, Prophages **Subsystem**: Phage packaging machineryery	511155	509194	-
mnas_418	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	511154	511388	-	Unknown: General	**peg.549**	hypothetical protein	- none -	511388	511155	-
mnas_419	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	511389	511728	-	Unknown: General	**peg.550**	hypothetical protein	- none -	511728	511390	-
mnas_420	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	511714	511939	-	Unknown: General	**peg.551**	hypothetical protein	- none -	511939	511715	-
mnas_421	phage Mu F like family protein			511931	512765	-	Mobile and extrachromosomal element functions: Prophage functions	**peg.552**	Phage protein	- none -	512765	511932	-
mnas_422	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	512764	514513	-	Unknown: General	**peg.553**	DNA polymerase B region	- none -	514513	512765	-
mnas_423	thymidylate kinase (EC 2.7.4.9)	*tmk*	GO:0004798, GO:0015949	514514	515099	-	Purines, pyrimidines, nucleosides, and nucleotides: Nucleotide and nucleoside interconversions	**peg.554**	Thymidylate kinase (EC 2.7.4.9)	- none -	515099	514515	-
mnas_424	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	515100	515457	-	Unknown: General	**peg.555**	hypothetical protein	- none -	515457	515101	-
mnas_425	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	515485	515980	-	Unknown: General						
mnas_426	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	515981	516137	-	Unknown: General	**peg.556**	hypothetical protein	- none -	516125	515982	-
mnas_427	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	516185	520046	+	Unknown: General	**peg.557**	hypothetical protein	- none -	516249	520046	+
mnas_428	putative membrane protein		GO:0008150, GO:0003674, GO:0016020	520183	521065	-	Cell envelope: Other	**peg.558**	hypothetical protein	- none -	521065	520184	-

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_429	excinuclease ABC subunit B (EC 3.1.25.-)	*uvrB*	GO:0006289, GO:0009380, GO:0009381	521216	523220	+	DNA metabolism: DNA replication, recombination, and repair	**peg.559**	Excinuclease ABC subunit B	**Category**: DNA Metabolism **Subcategory**: DNA repair **Subsystem**: DNA repair, UvrABC system	521217	523220	+
mnas_430	excinuclease ABC subunit A domain protein	*uvrA*	GO:0003677, GO:0005524, GO:0009432, GO:0005737, GO:0006289, GO:0009380, GO:0046872, GO:0009381, GO:0016887	523209	523647	+	Cellular processes: Other, Transport and binding proteins: Unknown substrate, Unknown: Enzymes of unknown specificity						
mnas_431	ribosome biogenesis GTP-binding protein YlqF	*ylqF*	GO:0003924, GO:0005525, GO:0005737, GO:0042254	523647	524412	-	Unclassified: Role category not yet assigned						
mnas_432	divIVA domain protein			524401	524668	-	Unknown: General Hypothetical	**peg.560**	hypothetical protein	- none -	524668	524402	-
mnas_433	RNA 2'-O ribose methyltransferase substrate binding family protein			524670	525402	-	Protein synthesis: tRNA and rRNA base modification	**peg.561**	FIG011178: rRNA methylase	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: RNA methylation	525402	524671	-
mnas_434	spoU rRNA Methylase family protein			525385	525922	-	Protein synthesis: tRNA and rRNA base modification	**peg.562**	tRNA (cytidine(34)-2';-O)-methyltransferase (EC 2.1.1.207)	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: RNA methylation	525922	525386	-
mnas_435	translation elongation factor Tu	*tuf*	GO:0003746, GO:0006414	525991	527179	-	Protein synthesis: Translation factors	**peg.563**	Translation elongation factor Tu	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Universal GTPases **Subsystem**: Translation elongation factors bacterial **Category**: Virulence, Disease and Defense **Subcategory**: Invasion and intracellular resistance **Subsystem**: Mycobacterium virulence operon involved in protein synthesis (SSU ribosomal proteins)	527179	525992	-

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_436	xylulose-5-phosphate phosphoketolase (EC 4.1.2.9)	*xpkA*	GO:0005975, GO:0050193	527686	530062	+	Energy metabolism: Sugars, Unknown: Enzymes of unknown specificity	**peg.564**	Xylulose-5-phosphate phosphoketolase (EC 4.1.2.9) @ Fructose-6-phosphate phosphoketolase (EC 4.1.2.22)	**Category**: Carbohydrates **Subcategory**: Central carbohydrate metabolism **Subsystem**: Pentose phosphate pathway **Subsystem**: Pentose phosphate pathway **Subcategory**: Fermentation **Subsystem**: Fermentations: Lactate **Subsystem**: Fermentations: Lactate	527687	530062	+
mnas_437	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	531075	531387	-	Unknown: Conserved						
mnas_438	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	530223	531087	+	Unknown: General						
mnas_439	AAA domain protein			531412	532342	-	Unknown: General Hypothetical	**peg.565**	hypothetical protein	- none -	532342	531413	-
mnas_440	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	532325	532718	-	Unknown: Conserved	**peg.566**	hypothetical protein	- none -	532718	532326	-
mnas_441	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	532727	533117	-	Unknown: General	**peg.567**	hypothetical protein	- none -	533117	532728	-
mnas_442	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	533203	533617	-	Unknown: General	**peg.568**	hypothetical protein	- none -	533581	533204	-
mnas_443	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	533558	533714	-	Unknown: General	**peg.569**	hypothetical protein	- none -	533714	533559	-
mnas_444	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	533985	534894	-	Unknown: General	**peg.570**	hypothetical protein	- none -	534894	533986	-
mnas_445	inosine-uridine preferring nucleoside hydrolase family protein			534886	535915	-	Purines, pyrimidines, nucleosides, and nucleotides: Salvage of nucleosides and nucleotides	**peg.571**	preQ1-regulated inosine-uridine nucleoside hydrolase (EC 3.2.2.1)	**Category**: Nucleosides and Nucleotides **Subcategory**: Purines **Subsystem**: Purine conversions	535915	534887	-
mnas_446	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	535916	536267	-	Unknown: General	**peg.572**	hypothetical protein	- none -	536267	535917	-
mnas_447	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	536259	536577	-	Unknown: General						
mnas_448	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	536730	536910	-	Unknown: General	**peg.573**	hypothetical protein	- none -	536910	536731	-
mnas_449	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	537104	538043	-	Unknown: Conserved	**peg.574**	hypothetical protein	- none -	537986	537105	-

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_450	trigger factor (EC 5.2.1.8)	*tig*	GO:0003755, GO:0005854, GO:0051082, GO:0051083	538059	539262	+	Protein fate: Protein folding and stabilization						
mnas_451	papain cysteine protease family protein			539360	540677	+	Protein fate: Degradation of proteins, peptides, and glycopeptides	**peg.575**	Aminopeptidase C (EC 3.4.22.40)	**Category**: Protein Metabolism **Subcategory**: Protein degradation **Subsystem**: Protein degradation	539361	540677	+
mnas_452	tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase (EC 2.1.1.61)	*trmU*	GO:0004808, GO:0005737, GO:0006396	540686	541802	+	Protein synthesis: tRNA and rRNA base modification	**peg.576**	tRNA-specific 2-thiouridylase MnmA	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: RNA methylation	540687	541802	+
mnas_453	alanine--tRNA ligase (EC 6.1.1.7)	*alaS*	GO:0004813, GO:0005737, GO:0006419	541809	544434	+	Protein synthesis: tRNA aminoacylation	**peg.577**	Alanyl-tRNA synthetase (EC 6.1.1.7)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: tRNA aminoacylation, Ala	541810	544434	+
mnas_454	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	544421	544847	+	Unknown: Conserved	**peg.578**	Putative Holliday junction resolvase YqgF	- none -	544422	544847	+
mnas_455	phosphatidylinositol-specific phospholipase C, X domain protein			544947	546894	+	Unknown: General Hypothetical	**peg.595**	Phosphatidylinositol-specific phospholipase C (EC 4.6.1.13)	**Category**: Phage packaging machinPhages, Prophages, Transposable elements, Plasmids **Subcategory**: Pathogenicity islands **Subsystem**: Listeria Pathogenicity Island LIPI-1 extended	545065	546894	+
mnas_456	lipoassociated domain protein			546869	550376	+	Unknown: General Hypothetical	**peg.596**	hypothetical protein	- none -	546870	550376	+
mnas_457	Aspartate--tRNA ligase (EC 6.1.1.12)			550450	551500	-	Unknown: Enzymes of unknown specificity	**peg.597**	Aspartyl-tRNA synthetase (EC 6.1.1.12)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: tRNA aminoacylation, Asp and Asn	551428	550451	-
mnas_458	Glutamyl-tRNA synthetase (EC 6.1.1.17)		GO:0004818	551508	552183	-	Unknown: Enzymes of unknown specificity						
mnas_459	tRNA-Lys			552267	552343	+		**rna.20**	tRNA-Lys-CTT	- none -	552268	552340	+
mnas_460	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	552533	553844	+	Unknown: Conserved	**peg.598**	FIG00834982: hypothetical protein	- none -	552534	553844	+
mnas_461	tRNA-Thr			554066	553992	-		**rna.21**	tRNA-Thr-CGT	- none -	554066	553996	-

IGS								RAST					
Gene_id	Common_name	Symbol[1]	GO terms[2]	fmin[3]	fmax[3]	S[4]	TIGR_roles	Gene_id	Function	Subsystem	Start[5]	Stop[5]	S[4]
mnas_462	ssrA-binding protein	*smpB*	GO:0003723, GO:0006450	554439	554877	-	Protein synthesis: Other	**peg.599**	tmRNA-binding protein SmpB	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Translation termination factors bacterial **Category**: Stress Response **Subcategory**: Heat shock **Subsystem**: Heat shock dnaK gene cluster extended	554877	554440	-
mnas_463	valine--tRNA ligase (EC 6.1.1.9)	*valS*	GO:0004832, GO:0005737, GO:0006438	555014	557501	+	Protein synthesis: tRNA aminoacylation	**peg.600**	Valyl-tRNA synthetase (EC 6.1.1.9)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: tRNA aminoacylation, Val	555015	557501	+
mnas_464	tetrahydrofolate dehydrogenase/cyclohydrolase, catalytic domain protein			557485	558055	+	Biosynthesis of cofactors, prosthetic groups, and carriers: Folic acid						
mnas_465	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	558059	564545	-	Unknown: General	**peg.601**	hypothetical protein	- none -	558124	559089	+
								peg.602	hypothetical protein	- none -	559465	560025	+
mnas_466	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	564553	565924	-	Unknown: General						
mnas_467	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	565946	567257	-	Unknown: General	**peg.603**	hypothetical protein	- none -	567257	565947	-
mnas_468	DJ-1/Pfpl family protein			567420	567987	-	Unclassified: Role category not yet assigned	**peg.604**	hypothetical protein	- none -	567987	567421	-
mnas_469	putative membrane protein		GO:0008150, GO:0003674, GO:0016020	567986	569000	-	Cell envelope: Other	**peg.605**	hypothetical protein	- none -	569000	567987	-
mnas_470	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	569002	570289	-	Unknown: General	**peg.606**	hypothetical protein	- none -	570289	569003	-
mnas_471	peptide deformylase (EC 3.5.1.88)	*def*	GO:0006464, GO:0042586	570303	570861	-	Protein fate: Protein modification and repair	**peg.607**	Peptide deformylase (EC 3.5.1.88)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Translation termination factors bacterial **Subsystem**: Bacterial RNA-metabolizing Zn-dependent hydrolases **Subsystem**: Conserved gene cluster associated with Met-tRNA formyltransferase	570798	570304	-
mnas_472	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	571010	571736	-	Unknown: Conserved						

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_473	putative Replicative DNA helicase (EC 3.6.1.-)	*dnaB*		571777	573238	-	Unknown: General Hypothetical	**peg.608**	Replicative DNA helicase (EC 3.6.1.-)	- none -	573238	571778	-
mnas_474	ribosomal protein L9	*rplI*	GO:0000311, GO:0003735, GO:0006412, GO:0022625	573224	573686	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.609**	LSU ribosomal protein L9p	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial	573686	573225	-
mnas_475	DHHA1 domain protein			573687	575673	-	Unknown: General Hypothetical	**peg.610**	Phosphoesterase, DHH family protein	- none -	575673	573688	-
mnas_476	ZIP Zinc transporter family protein			575782	576790	-	Transport and binding proteins: Cations and iron carrying compounds	**peg.611**	hypothetical protein	- none -	576790	575783	-
mnas_477	transcriptional regulator family protein			576789	577263	-	Transcription: Transcription factors	**peg.612**	FIG000859: hypothetical protein YebC	**Category**: Cofactors, Vitamins, Prosthetic Groups, Pigments **Subcategory**: Riboflavin, FMN, FAD **Subsystem**: Riboflavin, FMN and FAD metabolism in plants **Category**: DNA Metabolism **Subcategory**: DNA recombination **Subsystem**: RuvABC plus a hypothetical	577215	576790	-
mnas_478	Spermidine/putrescine/ ABC transporter substrate binding protein	*potD*	GO:0016020	577279	579238	+	Transport and binding proteins: Unknown substrate	**peg.613**	Probable spermidine/putrescine substrate binding protein in Mollicutes	- none -	577451	579238	+
mnas_479	ribosome-binding factor A	*rbfA*	GO:0006396, GO:0019843	579306	579648	+	Transcription: RNA processing	**peg.614**	Ribosome-binding factor A	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Translation initiation factors bacterial **Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: RNA processing and degradation, bacterial **Subsystem**: NusA-TFII Cluster	579307	579648	+
mnas_480	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	579843	580533	+	Unknown: General	**peg.615**	hypothetical protein	- none -	579844	580533	+
mnas_481	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	580729	581755	+	Unknown: General	**peg.616**	hypothetical protein	- none -	580730	581755	+
mnas_482	methionine adenosyltransferase (EC 2.5.1.6)	*metK*	GO:0004478, GO:0006556	581907	583050	+	Central intermediary metabolism: Other	**peg.617**	S-adenosylmethionine synthetase (EC 2.5.1.6)	- none -	581908	583050	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_483	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	583101	583230	+	Unknown: General	**peg.618**	hypothetical protein	- none -	583102	583230	+
mnas_484	ATP-dependent protease La (EC 3.4.21.53)	*lon*	GO:0004252, GO:0006508, GO:0008846	583478	586061	-	Protein fate: Degradation of proteins, peptides, and glycopeptides						
mnas_485	leucine--tRNA ligase (EC 6.1.1.4)	*leuS*	GO:0004823, GO:0005737, GO:0006429	586071	588402	-	Protein synthesis: tRNA aminoacylation, Disrupted reading frame: NULL	**peg.619**	Leucyl-tRNA synthetase (EC 6.1.1.4)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: tRNA aminoacylation, Leu	588402	586072	-
mnas_486	NAD^+ synthetase (EC 6.3.1.5)	*nadE*	GO:0005737, GO:0008795, GO:0009435	588691	589474	+	Biosynthesis of cofactors, prosthetic groups, and carriers: Pyridine nucleotides	**peg.620**	NAD synthetase (EC 6.3.1.5)	**Category**: Cofactors, Vitamins, Prosthetic Groups, Pigments **Subcategory**: NAD and NADP **Subsystem**: NAD and NADP cofactor biosynthesis global	588692	589474	+
mnas_487	transcriptional regulator family protein			589488	589659	+	Transcription: Transcription factors						
mnas_488	putative sn-glycerol-3-phosphate ABC transporter, ATP-binding protein	*ugpC*	GO:0008150, GO:0003674, GO:0005575	589659	591540	+	Transport and binding proteins: Unknown substrate, Unknown: Enzymes of unknown specificity	**peg.621**	Multiple sugar ABC transporter, ATP-binding protein	- none -	589912	591540	+
mnas_489	putative sn-glycerol-3-phosphate transport system permease protein	*ugpA*	GO:0008150, GO:0003674, GO:0016020	591529	592576	+	Cell envelope: Other	**peg.622**	N-Acetyl-D-glucosamine ABC transport system, permease protein	- none -	591530	592576	+
mnas_490	ABC transporter permease protein ugpE	*ugpE*		592520	593543	+	Transport and binding proteins: Unknown substrate	**peg.623**	ABC transporter permease protein	- none -	592554	593543	+
mnas_491	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	593545	593821	+	Unknown: General	**peg.624**	hypothetical protein	- none -	593546	593821	+
mnas_492	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	593823	595719	+	Unknown: General						
mnas_493	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	595725	596658	+	Unknown: General	**peg.625**	hypothetical protein	- none -	595786	596658	+
mnas_494	tRNA-Gly			597015	596941	-		**rna.22**	tRNA-Gly-TCC	- none -	597015	596945	-
mnas_495	tRNA-Arg			597174	597097	-		**rna.23**	tRNA-Arg-ACG	- none -	597174	597101	-
mnas_496	tRNA-Cys			597257	597182	-		**rna.24**	tRNA-Cys-GCA	- none -	597257	597186	-
mnas_497	modification methylase BanI (EC 2.1.1.37)	*banIM*	GO:0003677, GO:0009307, GO:0003886	597393	598746	+	Cellular processes: Other, Transport and binding proteins: Unknown substrate, Unknown: Enzymes of unknown specificity	**peg.626**	DNA-cytosine methyltransferase (EC 2.1.1.37)	**Category**: DNA Metabolism **Subcategory**: DNA repair **Subsystem**: DNA repair, bacterial	597394	598746	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_498	endonuclease/Exonuclease/phosphatase family protein			598815	599928	-	Unknown: Enzymes of unknown specificity	**peg.627**	membrane nuclease	- none -	599928	598816	-
mnas_499	aspartate-ammonia ligase (EC 6.3.1.1)	*asnA*		599974	600955	-	Unclassified: Role category not yet assigned	**peg.628**	Aspartate--ammonia ligase (EC 6.3.1.1)	**Category**: Amino Acids and Derivatives **Subcategory**: Glutamine, glutamate, aspartate, asparagine; ammonia assimilation **Subsystem**: Glutamine, Glutamate, Aspartate and Asparagine Biosynthesis	600955	599975	-
mnas_500	smr domain protein			600962	601280	-	Unknown: General Hypothetical	**peg.629**	unknown; predicted coding region	- none -	601280	600963	-
mnas_501	isochorismatase family protein			601267	601564	-	Unknown: Enzymes of unknown specificity	**peg.630**	hypothetical protein	- none -	601450	601268	-
mnas_502	putative hydrolase		GO:0016787	601570	602803	-	Unknown: Enzymes of unknown specificity						
mnas_503	putative hydrolase domain protein		GO:0016787	602847	603048	-	Unknown: Enzymes of unknown specificity	**peg.631**	Ribonuclease J2 (endoribonuclease in RNA processing)	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: Ribonucleases in Bacillus **Subsystem**: Bacterial RNA-metabolizing Zn-dependent hydrolases	603048	602848	-
mnas_504	pyruvate kinase (EC 2.7.1.40)	*pyk*	GO:0004743, GO:0005737, GO:0006096	603175	604609	-	Energy metabolism: Glycolysis/gluconeogenesis	**peg.632**	Pyruvate kinase (EC 2.7.1.40)	**Category**: Carbohydrates **Subcategory**: Central carbohydrate metabolism **Subsystem**: Pyruvate metabolism I: anaplerotic reactions, PEP **Subsystem**: Glycolysis and Gluconeogenesis;	604609	603176	-
mnas_505	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	604675	605578	-	Unknown: General	**peg.633**	hypothetical protein	- none -	605578	604676	-
mnas_506	ribosome biogenesis GTP-binding protein YsxC	*ysxC*	GO:0003924, GO:0005525, GO:0005737, GO:0042254, GO:0043022	605567	606158	-	Unclassified: Role category not yet assigned	**peg.634**	GTP-binding protein EngB	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Universal GTPases	606158	605568	-
mnas_507	acetate kinase (EC 2.7.2.1)	*ackA*	GO:0005737, GO:0006083, GO:0006113, GO:0008776	606159	607281	-	Energy metabolism: Fermentation	**peg.635**	Acetate kinase (EC 2.7.2.1)	**Category**: Carbohydrates **Subcategory**: Fermentation **Subsystem**: Fermentations: Lactate	607185	606160	-

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_508	hsp70 family protein			607297	607717	+	Protein fate: Protein folding and stabilization	**peg.15**	Chaperone protein DnaK	**Category**: Protein Metabolism **Subcategory**: Protein folding **Subsystem**: Protein chaperones **Category**: Stress Response **Subcategory**: Heat shock **Subsystem**: Heat shock dnaK gene cluster extended	607388	607717	+
mnas_509	phosphoglucose isomerase family protein			607836	609141	+	Energy metabolism: Pentose phosphate pathway	**peg.16**	Glucose-6-phosphate isomerase (EC 5.3.1.9)	**Category**: Carbohydrates **Subcategory**: Central carbohydrate metabolism **Subsystem**: Glycolysis and Gluconeogenesis	607837	609141	+
mnas_510	RNA pseudouridylate synthase family protein			609140	609644	+	Protein synthesis: tRNA and rRNA base modification	**peg.17**	Ribosomal small subunit pseudouridine synthase A (EC 4.2.1.70)	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: RNA pseudouridine syntheses	609141	609865	+
mnas_511	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	609709	609865	+	Unknown: General						
mnas_512	neisseria meningitidis TspB family protein			609854	611543	+	Unclassified: Role category not yet assigned	**peg.18**	hypothetical lipoprotein	- none -	609855	611543	+
mnas_513	Phosphopentomutase (EC 5.4.2.7)	*deoB*	GO:0006139, GO:0008973	611597	612788	-	Purines, pyrimidines, nucleosides, and nucleotides: Other	**peg.19**	Phosphopentomutase (EC 5.4.2.7)	- none -	612788	611598	-
mnas_514	dUTPase family protein			613000	613219	+	Unclassified: Role category not yet assigned	**peg.20**	hypothetical protein	- none -	613070	613219	+
mnas_515	tRNA-Leu			613243	613327	+		**rna.1**	tRNA-Leu-CAA	- none -	613244	613324	+
mnas_516	ahpC/TSA family protein			613365	613863	+	Cellular processes: Detoxification, Cellular processes: Adaptations to atypical conditions	**peg.21**	Thiol peroxidase, Tpx-type (EC 1.11.1.15)	**Category**: Sulfur Metabolism **Subcategory**: no subcategory **Subsystem**: Thioredoxin-disulfide reductase	613366	613863	+
mnas_517	aminopeptidase domain protein (EC 3.4.11.-)	*celM 2*	GO:0004177, GO:0016787	613864	614053	+	Unknown: Enzymes of unknown specificity	**peg.22**	hypothetical protein	- none -	613865	614053	+
mnas_518	M42 glutamyl aminopeptidase family protein			614052	614949	+	Protein fate: Degradation of proteins, peptides, and glycopeptides	**peg.23**	Endo-1,4-beta-glucanase	- none -	614068	614949	+
mnas_519	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	615080	615446	+	Unknown: Conserved	**peg.24**	hypothetical protein	- none -	615210	615446	+
mnas_520	NUDIX domain protein			615484	615934	-	Unknown: Enzymes of unknown specificity	**peg.25**	MutT/nudix family protein	- none -	615934	615485	-

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_521	ribosomal protein S2	*rpsB*	GO:0000314, GO:0003735, GO:0006412, GO:0022627	616187	617054	+	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.26**	SSU ribosomal protein S2p (SAe)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome SSU bacterial **Subsystem**: Ribosome recycling related cluster	616188	617054	+
mnas_522	translation elongation factor Ts	*tsf*	GO:0003746, GO:0006414	617095	617974	+	Protein synthesis: Translation factors	**peg.27**	Translation elongation factor Ts	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Translation elongation factors bacterial **Subsystem**: Ribosome recycling related cluster	617096	617974	+
mnas_523	putative lipoprotein		GO:0008150, GO:0003674, GO:0016020	618098	61General Unknown9	+	Cell envelope: Other						
mnas_524	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	618594	622437	+	Unknown: General	**peg.28**	Siderophore-mediated iron transport protein	- none -	618595	622437	+
mnas_525	putative membrane protein		GO:0008150, GO:0003674, GO:0016020	622451	623906	+	Cell envelope: Other						
mnas_526	chromate transporter family protein			623910	624492	+	Transport and binding proteins: Anions	**peg.29**	Chromate transport protein	- none -	623947	624492	+
mnas_527	chromate transporter family protein			624491	625139	+	Transport and binding proteins: Anions	**peg.30**	chromate ion transporter (CHR) family, putative	- none -	624492	625139	+
mnas_528	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	625754	626051	+	Unknown: General	**peg.31**	hypothetical protein	- none -	625863	626051	+
								peg.32	putative	- none -	626843	627547	+
mnas_529	eco57I restriction-modification methylase family protein			627536	628559	-	DNA metabolism: Restriction/modification	**peg.33**	N6 adenine-specific DNA methyltransferase, N12 class	- none -	628559	627537	-
mnas_530	putative restriction endonuclease Hpy8I		GO:0004519	626842	627547	+	Unknown: Enzymes of unknown specificity						
mnas_531	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	628740	629073	-	Unknown: General	**peg.34**	hypothetical protein	- none -	628992	628741	-
mnas_532	anticodon binding domain protein			629093	629480	+	Unknown: General Hypothetical	**peg.35**	Threonyl-tRNA synthetase (EC 6.1.1.3)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: tRNA aminoacylation, Thr	629127	629480	+
mnas_533	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	629472	630051	+	Unknown: General	**peg.36**	hypothetical protein	- none -	629473	630051	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_534	PTS system, glucose/glucosamine/beta-glucoside-specific (EC 2.7.1.69)			630093	630882	-	Unknown: Enzymes of unknown specificity	**peg.37**	PTS system, N-acetylglucosamine-specific IIA component (EC 2.7.1.69) / PTS system, N-acetylglucosamine-specific IIB component (EC 2.7.1.69) / PTS system, N-acetylglucosamine-specific IIC component (EC 2.7.1.69)	- none -	630882	630094	-
mnas_535	ribosomal protein L7/L12	*rplL*	GO:0000315, GO:0003735, GO:0006412, GO:0022625	631037	631409	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.38**	LSU ribosomal protein L7/L12 (P1/P2)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial **Subsystem**: LSU ribosomal proteins cluster	631409	631038	-
mnas_536	50S ribosomal protein L10	*rplJ*	GO:0005840, GO:0042254, GO:0006412, GO:0006950, GO:0003735	631440	631938	-	Cellular processes: Other, Cellular processes: Adaptations to atypical conditions	**peg.39**	LSU ribosomal protein L10p (P0)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial **Subsystem**: LSU ribosomal proteins cluster	631938	631441	-
mnas_537	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	632180	632645	-	Unknown: General	**peg.40**	hypothetical protein	- none -	632609	632181	-
mnas_538	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	632629	633916	-	Unknown: General	**peg.41**	Sialidase NanB (EC 3.2.1.18)	- none -	633445	632630	-
mnas_539	ruvA, C-terminal domain protein			633980	634598	+	Unknown: General Hypothetical	**peg.42**	Holliday junction DNA helicase RuvA	**Category**: DNA Metabolism **Subcategory**: DNA replication **Subsystem**: DNA-replication **Subcategory**: DNA recombination **Subsystem**: RuvABC plus a hypothetical	633981	634598	+
mnas_540	holliday junction DNA helicase RuvB (EC 3.6.4.12)	*ruvB*	GO:0006310, GO:0009378, GO:0009379	634572	635526	+	DNA metabolism: DNA replication, recombination, and repair	**peg.43**	Holliday junction DNA helicase RuvB	**Category**: DNA Metabolism **Subcategory**: DNA replication **Subsystem**: DNA-replication **Subcategory**: DNA recombination **Subsystem**: RuvABC plus a hypothetical	634573	635526	+
mnas_541	export membrane protein SecF	*secF*	GO:0015450, GO:0031522, GO:0043952	635590	638176	+	Protein fate:Protein and peptide secretion and trafficking	**peg.44**	Protein-export membrane protein SecD (TC 3.A.5.1.1) / Protein-export membrane protein SecF (TC 3.A.5.1.1)	- none -	635591	638176	+
mnas_542	GIY-YIG catalytic domain protein			638639	639758	-	DNA metabolism: Restriction/modification						

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_543	recombination U family protein			638276	638639	+	DNA metabolism: DNA replication, recombination, and repair						
mnas_544	NADPH-dependent FMN reductase family protein			639818	640409	-	Energy metabolism: Electron transport	**peg.45**	FMN-dependent NADH-azoreductase	- none -	640409	639819	-
mnas_545	NADPH-dependent FMN reductase family protein			640414	641005	-	Energy metabolism: Electron transport	**peg.46**	FMN-dependent NADH-azoreductase	- none -	641005	640415	-
mnas_546	tRNA-Ser			641317	641408	+		**rna.2**	tRNA-Pseudo-GCT	- none -	641318	641405	+
mnas_547	triose-phosphate isomerase (EC 5.3.1.1)	*tpiA*	GO:0004807, GO:0006096	641493	642237	+	Energy metabolism: Glycolysis/gluconeogenesis	**peg.47**	Triosephosphate isomerase (EC 5.3.1.1)	**Category**: Carbohydrates **Subcategory**: Central carbohydrate metabolism **Subsystem**: Glycolysis and Gluconeogenesis	641494	642237	+
mnas_548	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	643188	644022	-	Unknown: General						
mnas_549	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	642223	643192	+	Unknown: General						
mnas_550	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	644182	644398	+	Unknown: Conserved	**peg.48**	FIG00837018: hypothetical protein	- none -	644183	644398	+
mnas_551	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	644504	645458	-	Unknown: General	**peg.49**	predicted coding region	- none -	645458	644505	-
mnas_552	pseudouridine synthase, RluA family protein (EC 5.4.99.-)		GO:0001522, GO:0009451, GO:0009982	645458	646388	-	Protein synthesis: tRNA and rRNA base modification	**peg.50**	Ribosomal large subunit pseudouridine synthase D (EC 4.2.1.70)	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: RNA pseudouridine syntheses **Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome biogenesis bacterial	646388	645459	-
mnas_553	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	646482	646980	-	Unknown: Conserved	**peg.51**	hypothetical protein	- none -	646980	646483	-
mnas_554	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	646955	647645	-	Unknown: Conserved	**peg.52**	hypothetical protein	- none -	647264	646956	-
mnas_555	zinc finger found in FPG and IleRS family protein			647663	648242	+	Unclassified: Role category not yet assigned	**peg.53**	Isoleucyl-tRNA synthetase (EC 6.1.1.5)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: tRNA aminoacylation, Ile	647673	648242	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_556	signal peptidase (SPase) II family protein			648250	648964	+	Protein fate:Protein and peptide secretion and trafficking	**peg.54**	Lipoprotein signal peptidase (EC 3.4.23.36)	**Category**: Protein Metabolism **Subcategory**: Protein processing and modification **Subsystem**: Signal peptidase **Subsystem**: Lipoprotein Biosynthesis	648251	648964	+
mnas_557	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	649069	650386	+	Unknown: Conserved	**peg.55**	Conserved expressed protein	- none -	649070	650386	+
mnas_558	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	650471	650714	+	Unknown: General	**peg.56**	hypothetical protein	- none -	650472	650714	+
mnas_559	ribosomal protein L21	*rplU*	GO:0000311, GO:0003735, GO:0006412, GO:0022625	650820	651120	+	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.57**	LSU ribosomal protein L21p	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial	650821	651120	+
mnas_560	ribosomal protein L27	*rpmA*	GO:0000315, GO:0003735, GO:0006412, GO:0022625	651127	651388	+	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.58**	LSU ribosomal protein L27p	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial	651128	651388	+
mnas_561	DNA recombination-mediator A family protein			651448	652006	-	Unclassified: Role category not yet assigned	**peg.59**	SMF family protein, DNA processing chain A (dprA)	- none -	651943	651449	-
mnas_562	integrase core domain protein			652014	652500	-	Unknown: General Hypothetical						
mnas_563	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	652676	652895	-	Unknown: General						
mnas_564	ABC transporter family protein			653065	654874	+	Transport and binding proteins: Unknown substrate	**peg.60**	FIG00836097: hypothetical protein	- none -	653066	654874	+
mnas_565	ABC transporter transmembrane region family protein			654882	656322	+	Transport and binding proteins: Unknown substrate						
mnas_566	large-conductance mechanosensitive channel, MscL family protein			656550	656955	-	Transport and binding proteins: Unknown substrate	**peg.81**	Large-conductance mechanosensitive channel	**Category**: Potassium metabolism **Subcategory**: Potassium metabolism **Subsystem**: Potassium homeostasis	656955	656551	-
mnas_567	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	657185	658805	+	Unknown: General	**peg.82**	hypothetical protein	- none -	657186	658805	+
mnas_568	ribosomal protein S4	*rpsD*	GO:0000314, GO:0003735, GO:0006412, GO:0022627	659041	659641	+	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.83**	SSU ribosomal protein S4p (S9e)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome SSU bacterial	659042	659641	+

IGS								RAST					
Gene_id	Common_name	Symbol[1]	GO terms[2]	fmin[3]	fmax[3]	S[4]	TIGR_roles	Gene_id	Function	Subsystem	Start[5]	Stop[5]	S[4]
mnas_569	ribosomal protein L31	*rpmE*	GO:0000315, GO:0003735, GO:0006412, GO:0022625	659658	659868	+	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.84**	LSU ribosomal protein L31p	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial	659659	659868	+
mnas_570	uvrC Helix-hairpin-helix N-terminal family protein			660055	660577	-	Unclassified: Role category not yet assigned	**peg.85**	Excinuclease ABC subunit C	**Category**: DNA Metabolism **Subcategory**: DNA repair **Subsystem**: DNA repair, UvrABC system	660394	660056	-
mnas_571	DNA polymerase III, delta subunit	*holA*		660635	661577	-	DNA metabolism: DNA replication, recombination, and repair	**peg.86**	DNA polymerase III delta subunit (EC 2.7.7.7)	**Category**: DNA Metabolism **Subcategory**: DNA replication **Subsystem**: DNA-replication	661577	660636	-
mnas_572	comEC/Rec2-related domain protein			661589	662885	-	Unknown: General Hypothetical	**peg.87**	hypothetical protein	- none -	662885	661590	-
mnas_573	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	662905	663325	-	Unknown: General	**peg.88**	hypothetical protein	- none -	663364	662906	-
mnas_574	uracil phosphoribosyltransferase (EC 2.4.2.9)	*upp*	GO:0004845, GO:0008655	664082	664712	+	Purines, pyrimidines, nucleosides, and nucleotides: Salvage of nucleosides and nucleotides	**peg.89**	Uracil phosphoribosyltransferase (EC 2.4.2.9)	- none -	664083	664712	+
mnas_575	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	664860	665001	+	Unknown: General						
mnas_576	Oligopeptide ABC transporter system, permeaseprotein (OppC)	*OppC*		665000	666362	+	Transport and binding proteins: Unknown substrate	**peg.90**	Oligopeptide transport system permease protein OppC (TC 3.A.1.5.1)	**Category**: Membrane Transport **Subcategory**: ABC transporters **Subsystem**: ABC transporter oligopeptide (TC 3.A.1.5.1)	665001	666362	+
mnas_577	Oligopeptide ABC transporter, ATP-binding protein OppD	*oppD*	GO:0005524, GO:0009898, GO:0015440, GO:0015833, GO:0043190	666374	667421	+	Transport and binding proteins: Amino acids, peptides and amines	**peg.91**	Oligopeptide transport ATP-binding protein OppD (TC 3.A.1.5.1)	**Category**: Membrane Transport **Subcategory**: ABC transporters **Subsystem**: ABC transporter oligopeptide (TC 3.A.1.5.1)	666375	667421	+
mnas_578	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	669019	669571	-	Unknown: Conserved						
mnas_579	ABC superfamily ATP binding cassette transporter, ABC domain protein (EC 3.6.3.24)	*nikE*	GO:0005524, GO:0016787, GO:0000166, GO:0017111, GO:0006200, GO:0015833, GO:0015413, GO:0016887	667413	669027	+	Cellular processes: Other, Transport and binding proteins: Unknown substrate, Unknown: Enzymes of unknown specificity						
mnas_580	putative membrane protein		GO:0008150, GO:0003674, GO:0016020	669549	670995	-	Cell envelope: Other	**peg.92**	FIG065159: hypothetical 2 cooccurring with ATP synthase chains	- none -	670995	669550	-
mnas_581	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	671013	673176	-	Unknown: General	**peg.93**	Siderophore-mediated iron transport protein	- none -	673047	671014	-

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_582	5s_rRNA			673329	673226	-		**rna.5**	5S RNA	- none -	673333	673227	-
mnas_583	60Kd inner membrane family protein			673412	675431	-	Cell envelope: Other	**peg.94**	Inner membrane protein translocase component YidC, long form	**Subsystem**: Cell Division Subsystem including YidCD **Subsystem**: RNA modification cluster	675395	673413	-
mnas_584	ribonuclease P protein component (EC 3.1.26.5)	*rnpA*	GO:0004526, GO:0006396	675381	675720	-	Transcription: RNA processing	**peg.95**	Ribonuclease P protein component (EC 3.1.26.5)	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: tRNA processing **Subsystem**: Cell Division Subsystem including YidCD **Subsystem**: RNA modification cluster	675720	675382	-
mnas_585	ribosomal protein L34	*rpmH*	GO:0000315, GO:0003735, GO:0006412, GO:0022625	675734	675881	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.96**	LSU ribosomal protein L34p	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial **Subsystem**: Cell Division Subsystem including YidCD **Subsystem**: RNA modification cluster	675881	675735	-
mnas_586	serine--tRNA ligase (EC 6.1.1.11)	*serS*	GO:0004828, GO:0005737, GO:0006434	676092	677319	+	Protein synthesis: tRNA aminoacylation						
mnas_587	preprotein translocase, SecY subunit	*secY*	GO:0005887, GO:0015450, GO:0043952	677514	678822	+	Protein fate:Protein and peptide secretion and trafficking	**peg.97**	Preprotein translocase secY subunit (TC 3.A.5.1.1)	- none -	677533	678822	+
mnas_588	adenylate kinase (EC 2.7.4.3)	*adk*		678815	679481	+	Purines, pyrimidines, nucleosides, and nucleotides: Nucleotide and nucleoside interconversions	**peg.98**	Adenylate kinase (EC 2.7.4.3)	**Category**: Nucleosides and Nucleotides **Subcategory**: Purines **Subsystem**: Purine conversions	678816	679481	+
mnas_589	methionine aminopeptidase, type I (EC 3.4.11.18)	*map*	GO:0004239, GO:0006464	679496	680267	+	Protein fate: Protein modification and repair	**peg.99**	Methionine aminopeptidase (EC 3.4.11.18)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Translation termination factors bacterial	679515	680267	+
mnas_590	translation initiation factor IF-1	*infA*	GO:0003743, GO:0006413	680267	680483	+	Protein synthesis: Translation factors	**peg.100**	Translation initiation factor 1	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Translation initiation factors bacterial	680268	680483	+
mnas_591	ribosomal protein L36	*rpmJ*	GO:0000315, GO:0003735, GO:0006412, GO:0022625	680497	680611	+	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.101**	LSU ribosomal protein L36p	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial	680498	680611	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_592	30S ribosomal protein S13	*rpsM*	GO:0003735, GO:0006412, GO:0022627	680645	681011	+	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.102**	SSU ribosomal protein S13p (S18e)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome SSU bacterial	680646	681011	+
mnas_593	ribosomal S11 family protein			681117	681414	+	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.103**	SSU ribosomal protein S11p (S14e)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome SSU bacterial	681018	681224	+
mnas_594	PQ loop repeat family protein			681803	682541	+	Unclassified: Role category not yet assigned	**peg.104**	predicted coding region	- none -	681804	682541	+
mnas_595	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	683097	684696	-	Unknown: General	**peg.105**	hypothetical protein	- none -	684771	683098	-
mnas_596	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	684727	685366	-	Unknown: General	**peg.106**	hypothetical protein	- none -	685339	684728	-
mnas_597	phosphoglycerate kinase (EC 2.7.2.3)		GO:0004618	685368	686469	-	Energy metabolism: Glycolysis/gluconeogenesis						
mnas_598	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	686587	687697	-	Unknown: General	**peg.107**	hypothetical protein	- none -	687697	686588	-
mnas_599	lemA family protein			687732	688386	-	Unclassified: Role category not yet assigned	**peg.108**	LemA PROTEIN	- none -	688386	687733	-
mnas_600	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	688508	689003	-	Unknown: General	**peg.109**	hypothetical protein	- none -	689003	688509	-
mnas_601	Oligopeptide ABC transporter, ATP-binding protein OppF	*oppF*		688995	689301	-	Transport and binding proteins: Unknown substrate	**peg.110**	Oligopeptide ABC transporter ATP-binding protein	- none -	689271	688996	-
mnas_602	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	689313	690792	+	Unknown: General	**peg.111**	predicted coding region	- none -	689443	690792	+
mnas_603	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	690809	691469	+	Unknown: Conserved	**peg.112**	Cysteinyl-tRNA synthetase related protein	CBSS-261594.1.peg.788	690810	691469	+
mnas_604	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	691558	692149	-	Unknown: General	**peg.113**	hypothetical protein	- none -	692149	691559	-
mnas_605	sigma-70, region 4 family protein			692240	692465	-	Unclassified: Role category not yet assigned	**peg.114**	hypothetical protein	- none -	692450	692241	-

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_606	signal recognition particle-docking protein FtsY	*ftsY*	GO:0003924, GO:0005047, GO:0009306	692457	693084	-	Protein fate:Protein and peptide secretion and trafficking	**peg.115**	Signal recognition particle receptor protein FtsY (=alpha subunit) (TC 3.A.5.1.1)	**Category**: Membrane Transport **Subcategory**: Protein translocation across cytoplasmic membrane **Subsystem**: Bacterial signal recognition particle (SRP) **Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Universal GTPases	693084	692458	-
mnas_607	DNA ligase, NAD-dependent (EC 6.5.1.2)	*ligA*	GO:0003911, GO:0006260, GO:0006281, GO:0006310	693519	695565	-	DNA metabolism: DNA replication, recombination, and repair	**peg.116**	DNA ligase (EC 6.5.1.2)	**Category**: DNA Metabolism **Subcategory**: DNA repair **Subsystem**: DNA Repair Base Excision	695565	693520	-
mnas_608	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	695711	696842	+	Unknown: General	**peg.117**	hypothetical protein	- none -	695712	696842	+
mnas_609	phage integrase family protein			696894	697806	-	DNA metabolism: DNA replication, recombination, and repair, Mobile and extrachromosomal element functions: Prophage functions						
mnas_610	collagen triple helix repeat family protein			698876	699716	+	Unclassified: Role category not yet assigned	**peg.139**	Phage tail fiber protein	**Category**: Phage packaging machinPhages, Prophages, Transposable elements, Plasmids **Subcategory**: Phages, Prophages **Subsystem**: Phage tail fiber proteins	698877	699716	+
mnas_611	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	699772	700552	+	Unknown: General						
mnas_612	peptidase M60-like family protein			700552	701815	-	Protein fate: Degradation of proteins, peptides, and glycopeptides						
mnas_613	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	701851	703735	-	Unknown: General	**peg.140**	hypothetical protein	- none -	703507	701852	-
mnas_614	ATP synthase alpha/beta family, nucleotide-binding domain protein (EC 3.6.3.14)	*atpA*		703767	704628	+	Energy metabolism: ATP-proton motive force interconversion	**peg.141**	ATP synthase alpha chain (EC 3.6.3.14)	- none -	703768	704628	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_615	ATP synthase F1, beta subunit (EC 3.6.3.14)	*atpD*	GO:0000275, GO:0015986, GO:0045261, GO:0045262, GO:0046933	704627	705989	+	Energy metabolism: ATP-proton motive force interconversion	**peg.142**	ATP synthase beta chain (EC 3.6.3.14)	- none -	704628	705989	+
mnas_616	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	706000	706936	+	Unknown: Conserved						
mnas_617	metalloenzyme superfamily protein			706936	707713	-	Unclassified: Role category not yet assigned						
mnas_618	protein-(glutamine-N5) methyltransferase, release factor-specific (EC 2.1.1.-)	*prmC*	GO:0006412, GO:0008757, GO:0018364	707723	708449	-	Unknown: Enzymes of unknown specificity	**peg.143**	Protein-N(5)-glutamine methyltransferase PrmC, methylates polypeptide chain release factors RF1 and RF2	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Translation termination factors bacterial	708449	707724	-
mnas_619	peptide chain release factor 1	*prfA*	GO:0003747, GO:0006415	708448	709513	-	Protein synthesis: Translation factors	**peg.144**	Peptide chain release factor 1	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Translation termination factors bacterial	709513	708449	-
mnas_620	putative membrane protein		GO:0008150, GO:0003674, GO:0016020	709608	710049	-	Cell envelope: Other	**peg.145**	Amino acid permease	- none -	709983	709609	-
mnas_621	putative membrane protein		GO:0008150, GO:0003674, GO:0016020	710069	710993	+	Cell envelope: Other	**peg.146**	hypothetical protein	- none -	710163	710993	+
mnas_622	signal recognition particle protein	*ffh*	GO:0003924, GO:0005048, GO:0009306, GO:0048501	711337	712693	+	Protein fate:Protein and peptide secretion and trafficking	**peg.147**	Signal recognition particle, subunit Ffh SRP54 (TC 3.A.5.1.1)	**Category**: Membrane Transport **Subcategory**: Protein translocation across cytoplasmic membrane **Subsystem**: Bacterial signal recognition particle (SRP) **Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Universal GTPases	711383	712693	+
mnas_623	tRNA-Thr			712768	712843	+		**rna.6**	tRNA-Thr-GGT	- none -	712769	712840	+
mnas_624	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	713229	716046	-	Unknown: General	**peg.148**	contains gram positive anchor domain	- none -	715617	713230	-
mnas_625	mraW methylase family protein			716062	716491	+	Unknown: Enzymes of unknown specificity	**peg.149**	rRNA small subunit methyltransferase H	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: 16S rRNA modification within P site of ribosome	716285	716491	+
mnas_626	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	716507	717770	+	Unknown: General	**peg.150**	hypothetical protein	- none -	716676	717770	+

IGS								RAST					
Gene_id	Common_name	Symbol[1]	GO terms[2]	fmin[3]	fmax[3]	S[4]	TIGR_roles	Gene_id	Function	Subsystem	Start[5]	Stop[5]	S[4]
mnas_627	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	717769	718930	+	Unknown: General	**peg.151**	hypothetical protein	- none -	717770	718930	+
mnas_628	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	719004	719304	+	Unknown: General						
mnas_629	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	719294	721619	+	Unknown: General	**peg.152**	FIG057547: hypothetical cooccurring with ATP synthase chains	- none -	719295	721619	+
mnas_630	Putative ATP synthase alpha chain domain protein		GO:0045261, GO:0015986, GO:0016787, GO:0005524, GO:0046933, GO:0046961	721618	721891	+	Cellular processes: Other, Transport and binding proteins: Unknown substrate, Unknown: Enzymes of unknown specificity						
mnas_631	obg family GTPase CgtA	*cgtA*	GO:0005525, GO:0042254	722034	723309	-	Protein synthesis: Other	**peg.153**	GTP-binding protein Obg	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Universal GTPases	723309	722035	-
mnas_632	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	723422	724283	+	Unknown: General	**peg.154**	hypothetical protein	- none -	723423	724283	+
mnas_633	bacterial DNA-binding family protein			724369	724666	+	DNA metabolism: Chromosome-associated proteins	**peg.155**	Integration host factor alpha/beta	**Category**: DNA Metabolism **Subcategory**: no subcategory **Subsystem**: DNA structural proteins, bacterial	724370	724666	+
mnas_634	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	724789	725398	+	Unknown: General	**peg.156**	hypothetical protein	- none -	724829	725398	+
mnas_635	tRNA-His			725837	725913	+		**rna.7**	tRNA-His-GTG	- none -	725838	725910	+
mnas_636	putative membrane protein		GO:0008150, GO:0003674, GO:0016020	726121	727603	+	Cell envelope: Other						
mnas_637	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	727607	728645	+	Unknown: General	**peg.162**	hypothetical protein	- none -	727860	728645	+
mnas_638	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	728684	730373	-	Unknown: General	**peg.163**	hypothetical protein	- none -	730286	728685	-
mnas_639	putative membrane protein		GO:0008150, GO:0003674, GO:0016020	730389	730800	+	Cell envelope: Other	**peg.164**	hypothetical protein	- none -	730546	730800	+
mnas_640	proline--tRNA ligase (EC 6.1.1.15)	*proS*	GO:0004827, GO:0006433	730877	732368	+	Protein synthesis: tRNA aminoacylation	**peg.165**	Prolyl-tRNA synthetase (EC 6.1.1.15), archaeal/eukaryal type	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: tRNA aminoacylation, Pro	730911	732368	+
mnas_641	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	732586	733084	+	Unknown: General	**peg.166**	hypothetical protein	- none -	732587	733084	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_642	16s_rRNA			733261	734756	+		**rna.8**	Small Subunit Ribosomal RNA; ssuRNA; SSU rRNA	- none -	733282	734818	+
								rna.9	Large Subunit Ribosomal RNA; lsuRNA; LSU rRNA	- none -	735059	735903	+
mnas_643	Putative ATP synthase alpha chain domain protein		GO:0045261, GO:0015986, GO:0016787, GO:0005524, GO:0046933, GO:0046961	735919	736192	-	Cellular processes: Other, Transport and binding proteins: Unknown substrate, Unknown: Enzymes of unknown specificity						
mnas_644	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	736191	738369	-	Unknown: General	**peg.167**	FIG057547: hypothetical cooccurring with ATP synthase chains	- none -	738369	736192	-
mnas_645	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	738359	738659	-	Unknown: General						
mnas_646	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	738677	738941	+	Unknown: General	**peg.168**	hypothetical protein	- none -	738747	738941	+
mnas_647	cytidylate kinase (EC 2.7.4.14)	*cmk*	GO:0004127, GO:0015949	738992	739673	+	Purines, pyrimidines, nucleosides, and nucleotides: Nucleotide and nucleoside interconversions	**peg.169**	Cytidylate kinase (EC 2.7.4.25)	- none -	738993	739673	+
mnas_648	ribosome-associated GTPase EngA	*engA*	GO:0003924, GO:0005525, GO:0005737, GO:0042254, GO:0043022	739665	740979	+	Unclassified: Role category not yet assigned	**peg.170**	GTP-binding protein EngA	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Universal GTPases	739666	740979	+
mnas_649	ribosomal protein L28	*rpmB*	GO:0000311, GO:0003735, GO:0006412, GO:0022625	741085	741277	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.171**	LSU ribosomal protein L28p @ LSU ribosomal protein L28p, zinc-independent	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial	741277	741086	-
mnas_650	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	741374	744044	+	Unknown: General						
mnas_651	energy-coupling factor transporter ATP-binding EcfA 1 domain protein (EC 3.6.3.-)	*ecfA1*	GO:0006810, GO:0016787, GO:0005524, GO:0000166, GO:0006200, GO:0005886, GO:0016887	744048	744384	+	Cellular processes: Other, Transport and binding proteins: Unknown substrate, Unknown: Enzymes of unknown specificity	**peg.172**	ATPase component of general energizing module of ECF transporters	- none -	744082	744384	+
mnas_652	energy-coupling factor transporter ATP-binding protein EcfA2 (EC 3.6.3.-)	*ecfA2*	GO:0006824, GO:0016820, GO:0005524, GO:0005886, GO:0016887	744374	745718	+	Transport and binding proteins: Unknown substrate, Unknown: Enzymes of unknown specificity	**peg.173**	ATPase component of general energizing module of ECF transporters	- none -	744375	745718	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_653	energy-coupling factor transporter transmembrane protein EcfT	*ecfT*	GO:0015087, GO:0016021, GO:0009236, GO:0005886	745717	746698	+	Cellular processes: Other, Transport and binding proteins: Unknown substrate						
								peg.175	hypothetical protein	- none -	748105	748272	+
mnas_654	putative membrane protein		GO:0008150, GO:0003674, GO:0016020	746762	748139	-	Cell envelope: Other	**peg.174**	Unspecified monosaccharide ABC transport system, ATP-binding protein	- none -	748139	746763	-
mnas_655	Putative Sugar ABC transporter ATP-binding protein	*mglA*		748233	749064	-	Transport and binding proteins: Unknown substrate	**peg.176**	Unspecified monosaccharide ABC transport system, ATP-binding protein	- none -	748917	748234	-
mnas_656	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	749080	750265	+	Unknown: General	**peg.177**	hypothetical protein	- none -	749207	750265	+
mnas_657	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	750251	750698	+	Unknown: General	**peg.178**	hypothetical protein	- none -	750252	750698	+
mnas_658	phosphopantetheine attachment site family protein			750705	750924	+	Unclassified: Role category not yet assigned	**peg.179**	hypothetical protein	- none -	750706	750924	+
mnas_659	tRNA-SeC			750949	751024	+		**rna.10**	tRNA-SeC-TCA	- none -	750950	751021	+
mnas_660	basic membrane family protein			751354	752332	-	Cell envelope: Other						
mnas_661	NAD-dependent glycerol-3-phosphate dehydrogenase (EC 1.1.1.94)			751067	751358	+	Energy metabolism: Other						
mnas_662	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	752542	752701	-	Unknown: General	**peg.180**	hypothetical protein	- none -	752701	752543	-
mnas_663	inorganic pyrophosphatase family protein			752851	753406	-	Central intermediary metabolism: Phosphorus compounds	**peg.181**	Inorganic pyrophosphatase (EC 3.6.1.1)	**Category**: Phosphorus Metabolism **Subcategory**: no subcategory **Subsystem**: Phosphate metabolism	753406	752852	-
mnas_664	ribosomal protein L32	*rpmF*	GO:0000315, GO:0003735, GO:0006412, GO:0022625	753660	753861	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.198**	LSU ribosomal protein L32p	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial	753861	753661	-
								peg.199	FIG007079: UPF0348 protein family	- none -	754166	755134	+
mnas_665	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	755123	755810	-	Unknown: General	**peg.200**	hypothetical protein	- none -	755651	755124	-
mnas_666	cytidyltransferase-like domain protein			754165	755134	+	Unknown: Enzymes of unknown specificity						
mnas_667	PD-(D/E)XK nuclease superfamily protein			755814	756633	-	Unclassified: Role category not yet assigned						

IGS								RAST					
Gene_id	**Common_name**	**Symbol[1]**	**GO terms[2]**	**fmin[3]**	**fmax[3]**	**S[4]**	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start[5]**	**Stop[5]**	**S[4]**
mnas_668	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	756656	757046	-	Unknown: General	**peg.201**	hypothetical protein	- none -	757046	756657	-
mnas_669	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	757164	757398	-	Unknown: General	**peg.202**	hypothetical protein	- none -	757398	757165	-
mnas_670	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	757502	758021	-	Unknown: General	**peg.203**	hypothetical protein	- none -	757916	757503	-
mnas_671	pyrimidine-nucleoside phosphorylase family protein (EC 2.4.2.2)	*pdp*	GO:0006213, GO:0016154	758052	759222	+	Purines, pyrimidines, nucleosides, and nucleotides: Other	**peg.204**	Pyrimidine-nucleoside phosphorylase (EC 2.4.2.2)	- none -	758053	759222	+
mnas_672	deoxyribose-phosphate aldolase (EC 4.1.2.4)	*deoC*	GO:0004139, GO:0009264	759247	759577	+	Purines, pyrimidines, nucleosides, and nucleotides: Other, Energy metabolism: Other	**peg.205**	Deoxyribose-phosphate aldolase (EC 4.1.2.4)	- none -	759248	759911	+
mnas_673	deoC/LacD aldolase family protein			759674	759911	+	Unknown: Enzymes of unknown specificity						
mnas_674	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	760153	760660	-	Unknown: General						
mnas_675	cof-like hydrolase family protein		GO:0008152, GO:0016787	760876	761701	-	Unknown: Enzymes of unknown specificity	**peg.206**	Hydrolase (HAD superfamily)	- none -	761701	760877	-
mnas_676	TM2 domain protein			761891	762110	+	Unknown: General Hypothetical	**peg.207**	hypothetical protein	- none -	761892	762110	+
mnas_677	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	762151	762259	-	Unknown: General						
mnas_678	isoleucine--tRNA ligase (EC 6.1.1.5)	*ileS*	GO:0004822, GO:0005737, GO:0006428	762331	764371	+	Protein synthesis: tRNA aminoacylation						
mnas_679	50S ribosomal protein L3	*rplC*		764427	765093	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.208**	LSU ribosomal protein L3p (L3e)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial	765226	764428	-
mnas_680	putative lSU ribosomal protein L3P		GO:0005840, GO:0006412, GO:0005622, GO:0003735	765061	765226	-	Cellular processes: Other						
mnas_681	ribosomal protein S10	*rpsJ*	GO:0000314, GO:0003735, GO:0006412, GO:0022627	765261	765570	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.209**	SSU ribosomal protein S10p (S20e)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome SSU bacterial	765570	765262	-
mnas_682	thioredoxin family protein			765767	766109	-	Energy metabolism: Electron transport	**peg.210**	Possible periplasmic thiredoxin	- none -	766109	765768	-
mnas_683	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	766186	766390	-	Unknown: Conserved						

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_684	threonyl and Alanyl tRNA synthetase second additional domain protein			766394	767297	-	Protein synthesis: tRNA aminoacylation						
mnas_685	tryptophan--tRNA ligase (EC 6.1.1.2)	*trpS*	GO:0004830, GO:0005737, GO:0006436	767304	768297	-	Protein synthesis: tRNA aminoacylation	**peg.211**	Tryptophanyl-tRNA synthetase (EC 6.1.1.2)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: tRNA aminoacylation, Trp	768297	767305	-
mnas_686	DNA gyrase B subunit, carboxyl terminus family protein (EC 5.99.1.-)	*gyrB1*		76General Unknown 9	770429	+	DNA metabolism: DNA replication, recombination, and repair						
mnas_687	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	770437	772243	+	Unknown: General	**peg.212**	VlhA.4.04	- none -	770516	772243	+
mnas_688	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	772296	772401	-	Unknown: General						
mnas_689	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	772405	772546	-	Unknown: General	**peg.213**	hypothetical protein	- none -	774263	772632	-
mnas_690	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	772487	772619	-	Unknown: General						
mnas_691	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	772631	774263	-	Unknown: General						
mnas_692	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	774391	775252	+	Unknown: General	**peg.260**	hypothetical protein	- none -	774809	775252	+
mnas_693	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	775500	775953	+	Unknown: General	**peg.261**	hypothetical protein	- none -	775501	775953	+
mnas_694	glycerol-3-phosphate acyltransferase family protein			776321	776816	+	Unknown: Enzymes of unknown specificity	**peg.262**	Acyl-phosphate:glycerol-3-phosphate O-acyltransferase PlsY	**Category**: Fatty Acids, Lipids, and Isoprenoids **Subcategory**: Phospholipids **Subsystem**: Glycerolipid and Glycerophospholipid Metabolism in Bacteria	776343	776816	+
mnas_695	phosphopyruvate hydratase (EC 4.2.1.11)	*eno*	GO:0004634, GO:0006096	776General Unknown	778227	-	Energy metabolism: Glycolysis/gluconeogenesis	**peg.263**	Enolase (EC 4.2.1.11)	**Category**: Carbohydrates **Subcategory**: Central carbohydrate metabolism **Subsystem**: Glycolysis and Gluconeogenesis	778014	776857	-
mnas_696	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	778276	779878	-	Unknown: General	**peg.264**	hypothetical protein	- none -	779878	778277	-
mnas_697	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	779858	780122	-	Unknown: General	**peg.265**	hypothetical protein	- none -	780104	779859	-

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_698	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	780243	782001	-	Unknown: General	**peg.266**	hypothetical protein	- none -	781524	780244	-
mnas_699	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	782015	782402	+	Unknown: General	**peg.267**	hypothetical protein	- none -	782067	782402	+
mnas_700	Putative glucose-6-phosphate isomerase (EC 5.3.1.9)		GO:0004347	782379	783636	+	Energy metabolism: Pentose phosphate pathway	**peg.268**	Glucose-6-phosphate isomerase (EC 5.3.1.9)	**Category**: Carbohydrates **Subcategory**: Central carbohydrate metabolism **Subsystem**: Glycolysis and Gluconeogenesis	782380	783636	+
mnas_701	multidrug resistance ABC superfamily ATP binding cassette transporter, membrane domain protein (EC 3.6.3.42)	*ndvA*	GO:0006810, GO:0055085, GO:0015441, GO:0005524, GO:0016787, GO:0000166, GO:0006200, GO:0017111, GO:0016021, GO:0016887	783659	783848	-	Cellular processes: Other, Transport and binding proteins: Unknown substrate, Unknown: Enzymes of unknown specificity						
mnas_702	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	783858	784389	-	Unknown: General						
mnas_703	DNA (cytosine-5-)-methyltransferase family protein (EC 2.1.1.37)	*dcm*	GO:0003886, GO:0006304	784390	785557	-	DNA metabolism: DNA replication, recombination, and repair	**peg.269**	CPG DNA methylase	- none -	785557	784391	-
mnas_704	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	785842	786892	+	Unknown: General	**peg.270**	hypothetical protein	- none -	785843	786892	+
mnas_705	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	786891	787461	+	Unknown: General						
mnas_706	protein RecA	*recA*	GO:0003677, GO:0005737, GO:0006281, GO:0006310, GO:0008094, GO:0009432	787461	788181	+	DNA metabolism: DNA replication, recombination, and repair	**peg.271**	RecA protein	**Category**: DNA Metabolism **Subcategory**: DNA repair **Subsystem**: DNA repair, bacterial **Subcategory**: DNA replication **Subsystem**: DNA-replication	787684	788181	+
mnas_707	ymdB-like family protein			788238	789069	+	Unclassified: Role category not yet assigned	**peg.272**	FIG006542: Phosphoesterase	- none -	788239	789069	+
mnas_708	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	789218	789437	+	Unknown: General	**peg.273**	hypothetical protein	- none -	789437	789775	+
mnas_709	phage holin family protein			789436	789775	+	Unclassified: Role category not yet assigned						
mnas_710	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	789764	790073	+	Unknown: General	**peg.274**	hypothetical protein	- none -	789765	790073	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_711	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	790023	790494	+	Unknown: General	**peg.275**	hypothetical protein	- none -	790024	790494	+
mnas_712	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	790502	790925	+	Unknown: General						
								peg.276	hypothetical protein	- none -	790997	791128	+
mnas_713	HAD ATPase, P-type, IC family protein (EC 3.6.3.-)			791145	792615	+	Unclassified: Role category not yet assigned						
mnas_714	ribosomal protein L17	*rplQ*	GO:0003735, GO:0005762, GO:0006412, GO:0022625	792787	793150	-	Protein synthesis: Ribosomal proteins: synthesis and modification	**peg.281**	LSU ribosomal protein L17p	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Ribosome LSU bacterial	793150	792788	-
mnas_715	DNA-directed RNA polymerase, alpha subunit (EC 2.7.7.6)	*rpoA*	GO:0000345, GO:0003899, GO:0006350	793151	794198	-	Transcription: DNA-dependent RNA polymerase	**peg.282**	DNA-directed RNA polymerase alpha subunit (EC 2.7.7.6)	**Category**: RNA Metabolism **Subcategory**: Transcription **Subsystem**: RNA polymerase bacterial	794198	793152	-
mnas_716	DNA methylase family protein			794260	795649	+	DNA metabolism: Restriction/modification	**peg.283**	Type III restriction-modification system methylation subunit (EC 2.1.1.72)	**Category**: DNA Metabolism **Subcategory**: no subcategory **Subsystem**: Restriction-Modification System	794339	795649	+
mnas_717	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	795658	795868	+	Unknown: General						
mnas_718	peptidase S41 family protein			795892	797401	-	Protein fate: Degradation of proteins, peptides, and glycopeptides	**peg.284**	hypothetical protein	- none -	797401	795893	-
mnas_719	ribonuclease HII family protein			797530	798136	-	Transcription: Degradation of RNA	**peg.285**	Ribonuclease HII (EC 3.1.26.4)	**Category**: RNA Metabolism **Subcategory**: RNA processing and modification **Subsystem**: Ribonucleases in Bacillus **Subsystem**: Ribonuclease H; **Subsystem**: Conserved gene cluster associated with Met-tRNA formyltransferase	798136	797531	-
mnas_720	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	798711	799005	-	Unknown: General	**peg.286**	hypothetical protein	- none -	798188	798715	+
mnas_721	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	798187	798715	+	Unknown: General						
mnas_722	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	799009	800440	-	Unknown: Conserved	**peg.287**	hypothetical protein	- none -	800082	800477	+

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_723	16S rRNA (guanine(527)-N(7))-methyltransferase GidB (EC 2.1.1.170)	*gidB*	GO:0005575, GO:0008168, GO:0031167, GO:0046118	800563	801184	-	Unknown: Enzymes of unknown specificity						
mnas_724	ribose-phosphate diphosphokinase family protein (EC 2.7.6.1)	*prs*	GO:0004749, GO:0009152	801185	802067	-	Purines, pyrimidines, nucleosides, and nucleotides: Purine ribonucleotide biosynthesis						
mnas_725	putative lipoprotein		GO:0008150, GO:0003674, GO:0016020	802075	802666	-	Cell envelope: Other	**peg.288**	Ribose-phosphate pyrophosphokinase (EC 2.7.6.1)	**Category**: Carbohydrates **Subcategory**: Central carbohydrate metabolism **Subsystem**: Pentose phosphate pathway	802034	801186	-
mnas_726	mycoplasma lipo, C-terminal region family protein			802737	803526	-	Unclassified: Role category not yet assigned	**peg.289**	hypothetical protein	- none -	803019	802738	-
mnas_727	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	803538	804915	+	Unknown: General						
mnas_728	elongation factor G, domain IV family protein			804927	805773	+	Unknown: General Hypothetical	**peg.290**	Translation elongation factor G	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Universal GTPases **Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: Translation elongation factors bacterial **Subsystem**: Translation elongation factor G family **Category**: Virulence, Disease and Defense **Subcategory**: Invasion and intracellular resistance **Subsystem**: Mycobacterium virulence operon involved in protein synthesis (SSU ribosomal proteins)	805033	805773	+
mnas_729	recA bacterial DNA recombination family protein			805892	806276	+	DNA metabolism: DNA replication, recombination, and repair						
mnas_730	23s_rRNA			811845	806381	-		**rna.12**	Large Subunit Ribosomal RNA; lsuRNA; LSU rRNA	- none -	807600	806382	-
mnas_731	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	806553	806925	-	Unknown: Conserved						
mnas_732	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	807620	808544	-	Unknown: General						

IGS								RAST					
Gene_id	Common_name	Symbol[1]	GO terms[2]	fmin[3]	fmax[3]	S[4]	TIGR_roles	Gene_id	Function	Subsystem	Start[5]	Stop[5]	S[4]
mnas_733	aromatic cluster surface family protein			808932	809841	-	Cell envelope: Surface structures						
mnas_734	aromatic cluster surface family protein			809840	810179	-	Cell envelope: Surface structures						
mnas_735	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	810195	811104	+	Unknown: Conserved	**peg.302**	Prolipoprotein	- none -	810415	811104	+
mnas_736	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	811876	812617	+	Unknown: Conserved						
mnas_737	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	812621	812828	+	Unknown: General	**peg.303**	hypothetical protein	- none -	812649	812828	+
mnas_738	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	812839	813313	+	Unknown: General	**peg.304**	hypothetical protein	- none -	812840	813313	+
mnas_739	putative lipoprotein		GO:0008150, GO:0003674, GO:0016020	813351	813825	+	Cell envelope: Other						
mnas_740	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	813829	814135	-	Unknown: General						
mnas_741	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	814137	814308	-	Unknown: General						
mnas_742	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	814352	814664	-	Unknown: General	**peg.305**	hypothetical protein	- none -	814664	814353	-
mnas_743	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	814670	815006	-	Unknown: General						
mnas_744	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	816183	817335	-	Unknown: General						
mnas_745	GA module family protein			815024	816191	+	Unclassified: Role category not yet assigned						
mnas_746	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	817339	817936	-	Unknown: Conserved						
mnas_747	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	818071	818410	-	Unknown: General	**peg.306**	hypothetical protein	- none -	818410	818072	-
mnas_748	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	818497	818941	-	Unknown: General						

IGS								RAST					
Gene_id	**Common_name**	**Symbol**[1]	**GO terms**[2]	**fmin**[3]	**fmax**[3]	**S**[4]	**TIGR_roles**	**Gene_id**	**Function**	**Subsystem**	**Start**[5]	**Stop**[5]	**S**[4]
mnas_749	tRNA synthetases class I (E and Q), catalytic domain protein			818933	819605	-	Unknown: Enzymes of unknown specificity	**peg.307**	Glutamyl-tRNA synthetase (EC 6.1.1.17) @ Glutamyl-tRNA(Gln) synthetase (EC 6.1.1.24)	**Category**: Protein Metabolism **Subcategory**: Protein biosynthesis **Subsystem**: tRNA aminoacylation, Glu and Gln **Subsystem**: tRNA aminoacylation, Glu and Gln	819605	818934	-
mnas_750	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	819636	820374	+	Unknown: General Hypothetical	**peg.324**	hypothetical protein	- none -	820240	820374	+
mnas_751	sn-glycerol-3-phosphate ABC transporter, ATP-binding UgpC domain protein (EC 3.6.3.20)	*ugpC*	GO:0006810, GO:0005524, GO:0016820, GO:0016787, GO:0000166, GO:0006200, GO:0017111, GO:0015430, GO:0043190, GO:0016887, GO:0005215	820488	820725	+	Cellular processes: Other, Transport and binding proteins: Unknown substrate, Unknown: Enzymes of unknown specificity						
mnas_752	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	820725	821136	+	Unknown: General	**peg.325**	hypothetical protein	- none -	820783	821136	+
mnas_753	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	821150	821807	+	Unknown: Conserved						
mnas_754	AAA domain protein			822073	822688	-	Unknown: General Hypothetical						
mnas_755	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	821906	822095	+	Unknown: General	**peg.326**	hypothetical protein	- none -	821907	822095	+
								peg.327	hypothetical protein	- none -	822709	822074	-
mnas_756	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	822839	823850	-	Unknown: General						
mnas_757	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	823852	824059	-	Unknown: General						
mnas_758	HNH endonuclease family protein			824147	824810	-	Unclassified: Role category not yet assigned	**peg.328**	hypothetical protein	- none -	824612	824148	-
mnas_759	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	824814	825372	-	Unknown: General						
mnas_760	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	825376	825760	-	Unknown: General	**peg.329**	hypothetical protein	- none -	825673	825377	-
mnas_761	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	825764	825875	-	Unknown: General	**peg.330**	hypothetical protein	- none -	826290	825877	-

IGS								RAST					
Gene_id	Common_name	Symbol[1]	GO terms[2]	fmin[3]	fmax[3]	S[4]	TIGR_roles	Gene_id	Function	Subsystem	Start[5]	Stop[5]	S[4]
mnas_762	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	825876	826290	-	Unknown: General						
mnas_763	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	826289	826691	-	Unknown: Conserved	**peg.331**	hypothetical protein	- none -	826541	826290	-
mnas_764	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	826724	827606	-	Unknown: General	**peg.332**	hypothetical protein	- none -	827492	826725	-
mnas_765	ATP synthase F1, beta subunit (EC 3.6.3.14)	*atpD*	GO:0000275, GO:0015986, GO:0045261, GO:0045262, GO:0046933	827622	828504	+	Energy metabolism: ATP-proton motive force interconversion						
mnas_766	ATP synthase F1, epsilon subunit (EC 3.6.3.14)	*atpC*	GO:0000275, GO:0015986, GO:0045261, GO:0045262, GO:0046933	829415	829835	-	Energy metabolism: ATP-proton motive force interconversion	**peg.347**	ATP synthase epsilon chain (EC 3.6.3.14)	- none -	829835	829416	-
mnas_767	ATP synthase subunit beta (EC 3.6.3.14)	*atpD*		829837	830173	-	Energy metabolism: ATP-proton motive force interconversion						
mnas_768	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	831020	831707	+	Unknown: General	**peg.348**	hypothetical protein	- none -	831378	831707	+
mnas_769	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	831706	831844	+	Unknown: General						
mnas_770	Spermidine/putrescine ABC transporter permease protein PotB	*PotB*		831867	832221	+	Transport and binding proteins: Unknown substrate	**peg.349**	Spermidine Putrescine ABC transporter permease component PotB (TC 3.A.1.11.1)	- none -	831868	832221	+
mnas_771	spermidine/putrescine transport system permease PotC domain protein	*potC*	GO:0016020, GO:0006810, GO:0005215	832213	832666	+	Transport and binding proteins: Unknown substrate						
mnas_772	telomere recombination family protein			832846	833314	-	Unclassified: Role category not yet assigned	**peg.350**	TsaC protein (YrdC domain) required for threonylcarbamoyladenosine t(6)A37 modification in tRNA	- none -	833314	832847	-
mnas_773	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	833387	833480	+	Unknown: General						
mnas_774	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	833705	834293	+	Unknown: General						
mnas_775	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	834607	834718	-	Unknown: General						
mnas_776	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	835093	835480	-	Unknown: General						

IGS								RAST					
Gene_id	Common_name	Symbol[1]	GO terms[2]	fmin[3]	fmax[3]	S[4]	TIGR_roles	Gene_id	Function	Subsystem	Start[5]	Stop[5]	S[4]
mnas_777	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	834821	835097	+	Unknown: General						
mnas_778	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	835442	835871	-	Unknown: General	**peg.351**	hypothetical protein	- none -	835706	835443	-
mnas_779	NAD-dependent glycerol-3-phosphate dehydrogenase family protein			835906	836638	-	Energy metabolism: Other	**peg.352**	Glycerol-3-phosphate dehydrogenase [NAD(P)$^+$] (EC 1.1.1.94)	**Category**: Fatty Acids, Lipids, and Isoprenoids **Subcategory**: Phospholipids **Subsystem**: Glycerolipid and Glycerophospholipid Metabolism in Bacteria	836626	835907	-
mnas_780	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	836648	837320	-	Unknown: General						
mnas_781	ABC transporter family protein			837407	838070	-	Transport and binding proteins: Unknown substrate						
mnas_782	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	838173	838920	-	Unknown: General						
mnas_783	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	838920	839268	-	Unknown: Conserved						
mnas_784	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	839749	840229	-	Unknown: General	**peg.363**	hypothetical protein	- none -	839887	839750	-
mnas_785	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	840222	840393	-	Unknown: General	**peg.364**	hypothetical protein	- none -	840375	840223	-
mnas_786	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	840405	840531	+	Unknown: General						
mnas_787	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	840530	840986	+	Unknown: General	**peg.365**	hypothetical protein	- none -	840531	840986	+
mnas_788	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	841116	841809	-	Unknown: General						
mnas_789	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	840960	841128	+	Unknown: General						
mnas_790	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	841821	842451	+	Unknown: General	**peg.366**	hypothetical protein	- none -	841870	842451	+
mnas_791	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	843153	843459	-	Unknown: General						
mnas_792	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	842500	843157	+	Unknown: General						

IGS								RAST					
Gene_id	Common_name	Symbol[1]	GO terms[2]	fmin[3]	fmax[3]	S[4]	TIGR_roles	Gene_id	Function	Subsystem	Start[5]	Stop[5]	S[4]
mnas_793	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	843442	843796	-	Unknown: General						
mnas_794	putative lipoprotein		GO:0008150, GO:0003674, GO:0016020	843802	844339	-	Cell envelope: Other						
mnas_795	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	844342	844444	-	Unknown: General						
mnas_796	hsp70 family protein			844464	845103	+	Protein fate: Protein folding and stabilization						
mnas_797	amino acid permease family protein			845105	845720	+	Transport and binding proteins: Unknown substrate	**peg.380**	hypothetical protein	- none -	845404	845204	-
mnas_798	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	845724	846330	-	Unknown: General						
mnas_799	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	846342	846549	-	Unknown: General						
mnas_800	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	847090	847447	+	Unknown: Conserved	**peg.381**	hypothetical protein	- none -	847091	847447	+
mnas_801	conserved hypothetical protein		GO:0008150, GO:0003674, GO:0005575	847595	848141	+	Unknown: Conserved						
mnas_802	ABC transporter family protein			848720	849290	-	Transport and binding proteins: Unknown substrate						
mnas_803	recF/RecN/SMC N terminal domain protein			848141	848720	+	Cellular processes: Cell division						
mnas_804	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	849882	850437	+	Unknown: General						
mnas_805	hypothetical protein		GO:0008150, GO:0003674, GO:0005575	850671	850980	-	Unknown: General	**peg.392**	hypothetical protein	- none -	850869	850672	-
								rna.14	Large Subunit Ribosomal RNA; lsuRNA; LSU rRNA	- none -	851512	851009	-
mnas_806	Putative oligopeptide ABC transporter, ATP-binding protein OppF C-terminal			851524	852058	-	Transport and binding proteins: Unknown substrate						

1. Gene symbol
2. Gene Ontology terms
3. The start and stop condon positions (location include stop codon)
4. DNA strand (+ or -) on which the predicted gene is located
5. The start and stop codon postions (location include stop codon)

Supplementary Table 2.3 Ribosomal RNAs found within the IGS and RAST annotations of the Ms03 draft genome

	IGS			RAST			
	RNA_ID	**Location**	**Size (nt)**	**RNA_ID**	**Description**	**Location**	**Size (nt)**
5s_rRNA	mnas_582	673329-673226	103	rna.5	5S RNA	673333-673227	107
16s_rRNA	mnas_642	733261-734756	1495	rna.8	Small Subunit Ribosomal RNA; ssuRNA; SSU rRNA	733282-734818	1537
				rna.9	Large Subunit Ribosomal RNA; lsuRNA; LSU rRNA	735059-735903	845
23s_rRNA	mnas_730	811845-806381	5464	rna.12	Large Subunit Ribosomal RNA; lsuRNA; LSU rRNA	807600-806382	1219
				rna.14	Large Subunit Ribosomal RNA; lsuRNA; LSU rRNA	851512-851009	504

Supplementary Table 2.4 The transfer RNAs found within the IGS and RAST annotations of the Ms03 draft genome

aa	**IGS**				**RAST**			
	Gene_ID	**RNA**	**Location**	**Size (nt)**	**RNA_ID**	**RNA**	**Location**	**Size (nt)**
Ala	No annotation tRNA							
Arg	mnas_193	tRNA-Arg	244829-244752	77	rna.11	tRNA-Arg-TCT	244829-244756	74
	mnas_495	tRNA-Arg	597174-597097	77	rna.23	tRNA-Arg-ACG	597174-597101	74
Asn	mnas_44	tRNA-Asn	49170-49245	75	rna.15	tRNA-Asn-GTT	49171-49242	72
Asp	No annotation tRNA							
Cys	mnas_496	tRNA-Cys	597257-597182	75	rna.24	tRNA-Cys-GCA	597257-597186	72
Gln	mnas_110	tRNA-Gln	153820-153745	75	rna.3	tRNA-Gln-TTG	153820-153749	72
Glu	mnas_45	tRNA-Glu	49251-49327	76	rna.16	tRNA-Glu-TTC	49252-49324	73
Gly	mnas_494	tRNA-Gly	597015-596941	74	rna.22	tRNA-Gly-TCC	597015-596945	71
His	mnas_635	tRNA-His	725837-725913	76	rna.7	tRNA-His-GTG	725838-725910	73
Ile	No annotation tRNA							
Leu	mnas_48	tRNA-Leu	49520-49604	84	rna.19	tRNA-Leu-TAG	49521-49601	81
	mnas_515	tRNA-Leu	613243-613327	84	rna.1	tRNA-Leu-CAA	613244-613324	81
Lys	mnas_459	tRNA-Lys	552267-552343	76	rna.20	tRNA-Lys-CTT	552268-552340	73
Met	No annotation tRNA							
Phe	No annotation tRNA							
Pro	No annotation tRNA							
Ser	mnas_546	tRNA-Ser	641317-641408	91	rna.2	tRNA-Pseudo-GCT	641318-641405	88
	mnas_659	tRNA-SeC	750949-751024	75	rna.10	tRNA-SeC-TCA	750950-751021	72
Thr	mnas_47	tRNA-Thr	49409-49485	76	rna.18	tRNA-Thr-TGT	49410-49482	73
	mnas_461	tRNA-Thr	554066-553992	74	rna.21	tRNA-Thr-CGT	554066-553996	71
	mnas_623	tRNA-Thr	712768-712843	75	rna.6	tRNA-Thr-GGT	712769-712840	72
Trp	mnas_214	tRNA-Trp	77866-77792	74	rna.13	tRNA-Trp-CCA	277866-277796	71
Tyr	mnas_111	tRNA-Tyr	153909-153825	84	rna.4	tRNA-Tyr-GTA	153909-153829	81
Val	mnas_46	tRNA-Val	49332-49408	76	rna.17	tRNA-Val-TAC	49333-49405	73

Supplementary Table 2.5 DNA replication genes found within the IGS and RAST annotations of the Ms03 draft genome

EC	IGS			RAST	
	Gene_id	Gene name	Gene symbol	Gene_id	Gene name
2.7.7.-	mnas_269	DNA primase	*dnaG*	peg.384	DNA primase
2.7.7.7	mnas_71	DNA polymerase III, alpha subunit, Gram-positive type	*polC*	peg.514	DNA polymerase III alpha subunit
2.7.7.7	mnas_100	DNA polymerase III beta subunit, central domain protein	*dnaN*	peg.10	DNA polymerase III beta subunit
2.7.7.7	mnas_159	DNA polymerase III, subunit gamma and tau	*dnaX*	peg.184	DNA polymerase III subunits gamma and tau
2.7.7.7	mnas_163	Putative DNA polymerase III, delta' subunit	*holB*	peg.188	DNA polymerase III delta prime subunit
2.7.7.7	mnas_245	DNA polymerase III, alpha subunit	*dnaE*	peg.355	DNA polymerase III alpha subunit
2.7.7.7	mnas_571	DNA polymerase III, delta subunit	*holA*	peg.86	DNA polymerase III delta subunit
2.7.7.7	mnas_246*	Putative DNA polymerase I		peg.356*	DNA polymerase I
3.1.26.4	mnas_719#	Ribonuclease HII family protein		peg.285#	Ribonuclease HII
3.6.1.-	mnas_473	Putative replicative DNA helicase	*dnaB*	peg.608	Replicative DNA helicase
3.6.4.12	mnas_266	DnaB-like helicase C terminal domain protein		peg.379	Replicative DNA helicase (DnaB)
6.5.1.2	mnas_607$	DNA ligase, NAD-dependent	*ligA*	peg.116$	DNA ligase
-	mnas_175	Single-stranded DNA-binding family protein	*ssb*	peg.245	Single-stranded DNA-binding protein
-	mnas_99	Chromosomal replication initiator protein DnaA	*dnaA*	peg.9	Chromosomal replication initiator protein DnaA

* Correspond to the DpoI of DNA polymerase I in Figure 3.5
# Correspond to the RNaseHII in Figure 3.5
$ Correspond to the Lig of DNA ligase in Figure 3.5

Supplementary Table 2.6 Purine metabolism genes found within the IGS and RAST annotations of the Ms03 draft genome

EC	IGS			RAST	
	Gene_id	Gene name	Gene symbol	Gene_id	Gene name
2.4.2.7	mnas_140	Adenine phosphoribosyltransferase	*apt*	peg.130	Adenine phosphoribosyltransferase
2.4.2.8	mnas_338*	Hypoxanthine phosphoribosyltransferase	*hpt*	peg.464*	Hypoxanthine-guanine phosphoribosyltransferase
2.7.1.40#	mnas_504	Pyruvate kinase	*pyk*	peg.632	Pyruvate kinase
2.7.1.113 2.7.1.76	mnas_337	Deoxynucleoside kinase family protein		peg.463	Deoxyadenosine kinase (EC 2.7.1.76) / Deoxyguanosine kinase (EC 2.7.1.113)
2.7.4.3	mnas_588$	Adenylate kinase family protein	*adk*	peg.98$	Adenylate kinase
2.7.4.8	mnas_117	Guanylate kinase	*gmk*	peg.71	Guanylate kinase
2.7.6.1	mnas_724Δ	Ribose-phosphate diphosphokinase family protein	*prs*	peg.288Δ	Ribose-phosphate pyrophosphokinase
3.1.3.5	mnas_205	Calcineurin-like phosphoesterase family protein		peg.299	5'-nucleotidase
3.2.2.1	mnas_445	Inosine-uridine preferring nucleoside hydrolase family protein		peg.571	preQ1-regulated inosine-uridine nucleoside hydrolase
5.4.2.7	mnas_513£Δ	Phosphopentomutase	*deoB*	peg.19£Δ	Phosphopentomutase

* Enzyme number for reverse reaction is EC 2.4.2.22
# Part of glycolysis
$ Enzyme number for reverse reaction is EC 2.7.4.11
Δ Part of Pentose phosphate pathway
£ Enzyme number for reverse reaction is EC 5.4.2.2

Supplementary Table 2.7 Pyrimidine metabolism genes found within the IGS and RAST annotations of the Ms03 draft genome

EC	IGS			RAST	
	Gene_id	**Gene name**	**Gene symbol**	**Gene_id**	**Gene name**
1.8.1.9	mnas_66	Yhioredoxin reductase	*trxB*	peg.509	Thioredoxin reductase
2.4.2.2 [△]	mnas_671[θ]	Pyrimidine-nucleoside phosphorylase family protein	*pdp*	peg.204[θ]	Pyrimidine-nucleoside phosphorylase
2.4.2.9	mnas_574	Uracil phosphoribosyltransferase	*upp*	peg.89	Uracil phosphoribosyltransferase
2.7.1.21	mnas_229	Thymidine kinase	*tdk*	peg.336	Thymidine kinase
2.7.4.9	mnas_423	Thymidylate kinase	*tmk*	peg.554	Thymidylate kinase
2.7.4.9	mnas_162	Thymidylate kinase	*tmk*	peg.187	Thymidylate kinase
2.7.4.22	mnas_309	UMP kinase	*pyrH*	peg.437	Uridine monophosphate kinase
2.7.4.14[Ω]	mnas_647	Cytidylate kinase	*cmk*	peg.169	Cytidylate kinase (EC 2.7.4.25[Ω])
4.2.1.70	mnas_14	Ribosomal large subunit pseudouridine synthase B (EC 5.4.99.22)	*rluB*	peg.225	Ribosomal large subunit pseudouridine synthase B
4.2.1.70	mnas_118	tRNA pseudouridine(55) synthase (EC 5.4.99.25)	*truB*	peg.72	tRNA pseudouridine synthase B (EC 4.2.1.70)
4.2.1.70	mnas_328	RNA pseudouridylate synthase family protein		peg.454	Ribosomal large subunit pseudouridine synthase C (EC 4.2.1.70)
4.2.1.70	mnas_552	Pseudouridine synthase, RluA family protein (EC 5.4.99.-)		peg.50	Ribosomal large subunit pseudouridine synthase D (EC 4.2.1.70)
4.2.1.70	mnas_510	RNA pseudouridylate synthase family protein		peg.17	Ribosomal small subunit pseudouridine synthase A (EC 4.2.1.70)
3.1.3.5	mnas_205	Calcineurin-like phosphoesterase family protein		peg.299[¥]	5'-nucleotidase
3.5.4.5[♦]	mnas_290	Cytidine deaminase	*cdd*	peg.403	Cytidine deaminase

* Part of glycolysis
[$]Enzyme number for reverse reaction is EC 2.7.4.11
[△] Part of Pentose phosphate pathway
[¥]Part of Purine metabolise
[θ] Enzyme number for reverse reaction is EC 2.4.2.3
[Ω] The eukaryotic enzyme EC 2.7.4.14 is a bifunctional enzyme that catalyses the phosphorylation of both CMP and UMP with similar efficiency (dCMP can also act as acceptor). Different from the monofunctional prokaryotic enzymes EC 2.7.4.25, (d)CMP kinase and EC 2.7.4.22, UMP kinase.
[⌂]This enzyme can accept both the ribonucleoside uridine (EC 2.4.2.3) and the 2'-deoxyribonucleosides 2'-deoxyuridine (EC 2.4.2.3) and thymidine (EC 2.4.2.4). Formally known as EC 2.4.2.23.
[♦] Enzyme number for reverse reaction is EC 3.5.4.14

Supplementary Table 2.8 RNA polymerase genes found within the IGS and RAST annotations of the Ms03 draft genome

EC	IGS			RAST	
	Gene_id	**Gene name**	**Gene symbol**	**Gene_id**	**Gene name**
-	mnas_271	RNA polymerase sigma factor, sigma-70 family protein	*rpoD*	peg.386	RNA polymerase sigma factor RpoD
2.7.7.6	mnas_715	DNA-directed RNA polymerase, alpha subunit	*rpoA*	peg.282	DNA-directed RNA polymerase alpha subunit
2.7.7.6	mnas_59	DNA-directed RNA polymerase, beta' subunit	*rpoC*	peg.432	DNA-directed RNA polymerase beta' subunit
2.7.7.6	mnas_57* mnas_58*	RNA polymerase beta subunit rpoB	*rpoB*	peg.431*	DNA-directed RNA polymerase beta subunit

* The ORF of mnas_57 and mnas_58 from 64459 to 65722 and 65714 to 70202, respectively. The ORF of peg.431 are from 62088 to 65722 and therefore include both mnas_57 and mnas_58

Supplementary Table 2.9 Ribosomal genes found within the IGS and RAST annotations of the Ms03 draft genome

IGS			RAST	
Gene_id	**Gene name**	**Gene symbol**	**Gene_id**	**Gene name**
mnas_40	ribosomal protein L33	*rpmG*	peg.420	LSU ribosomal protein L33p
mnas_122	ribosomal protein S15	*rpsO*	peg.76	SSU ribosomal protein S15p (S13e)
mnas_141	ribosomal protein L1	*rplA*	peg.131	LSU ribosomal protein L1p (L10Ae)
mnas_142	ribosomal protein L11	*rplK*	peg.132	LSU ribosomal protein L11p (L12e)
mnas_174	ribosomal protein S6	*rpsF*	peg.244	SSU ribosomal protein S6p
mnas_176	ribosomal protein S18	*rpsR*	peg.246	SSU ribosomal protein S18p @ SSU ribosomal protein S18p, zinc-independent
mnas_206	ribosomal protein S12	*rpsL*	peg.300	SSU ribosomal protein S12p (S23e)
mnas_207	ribosomal protein S7	*rpsG*	peg.301	SSU ribosomal protein S7p (S5e)
mnas_228	ribosomal protein S20	*rpsT*	peg.335	SSU ribosomal protein S20p
mnas_232	ribosomal protein L35	*rpmI*	peg.339	LSU ribosomal protein L35p
mnas_233	ribosomal protein L20	*rplT*	peg.340	LSU ribosomal protein L20p
mnas_343	ribosomal protein L13	*rplM*	peg.469	LSU ribosomal protein L13p (L13Ae)
mnas_344	30S ribosomal protein S9	*rpsI*	peg.470	SSU ribosomal protein S9p (S16e)
mnas_369	ribosomal protein L19	*rplS*	peg.493	LSU ribosomal protein L19p
mnas_371	ribosomal protein S16	*rpsP*	peg.495	SSU ribosomal protein S16p
mnas_382	ribosomal protein L15	*rplO*	peg.519	LSU ribosomal protein L15p (L27Ae)
mnas_383	ribosomal protein S5	*rpsE*	peg.520	SSU ribosomal protein S5p (S2e)
mnas_384	ribosomal protein L18	*rplR*	peg.521	LSU ribosomal protein L18p (L5e)
mnas_385	ribosomal protein L6	*rplF*	peg.522	LSU ribosomal protein L6p (L9e)
mnas_386	ribosomal S8 family protein		peg.523	SSU ribosomal protein S8p (S15Ae)
mnas_387	30S ribosomal protein S14 type Z	*rpsZ*	peg.524	SSU ribosomal protein S14p (S29e) @ SSU ribosomal protein S14p (S29e), zinc-dependent
mnas_388	50S ribosomal protein L5	*rplE*	peg.525	LSU ribosomal protein L5p (L11e)
mnas_389	ribosomal protein L24	*rplX*	peg.526	LSU ribosomal protein L24p (L26e)
mnas_390	ribosomal protein L14	*rplN*	peg.527	LSU ribosomal protein L14p (L23e)
mnas_391	30S ribosomal protein S17	*rpsQ*	peg.528	SSU ribosomal protein S17p (S11e)
mnas_392	ribosomal protein L29	*rpmC*	peg.529	LSU ribosomal protein L29p (L35e)
mnas_393	ribosomal protein L16	*rplP*	peg.530	LSU ribosomal protein L16p (L10e)
mnas_394	ribosomal protein S3	*rpsC*	peg.531	SSU ribosomal protein S3p (S3e)
mnas_395	ribosomal protein L22	*rplV*	peg.532	LSU ribosomal protein L22p (L17e)
mnas_396	ribosomal protein S19	*rpsS*	peg.533	SSU ribosomal protein S19p (S15e)
mnas_397	ribosomal protein L2	*rplB*	peg.534	LSU ribosomal protein L2p (L8e)
mnas_398	50S ribosomal L23 domain protein	*rplW*	peg.535	LSU ribosomal protein L23p (L23Ae)
mnas_399	ribosomal L4/L1 family protein		peg.536	LSU ribosomal protein L4p (L1e)
mnas_474	ribosomal protein L9	*rplI*	peg.609	LSU ribosomal protein L9p
mnas_521	ribosomal protein S2	*rpsB*	peg.26	SSU ribosomal protein S2p (SAe)
mnas_535	ribosomal protein L7/L12	*rplL*	peg.38	LSU ribosomal protein L7/L12 (P1/P2)
mnas_536	50S ribosomal protein L10	*rplJ*	peg.39	LSU ribosomal protein L10p (P0)
mnas_559	ribosomal protein L21	*rplU*	peg.57	LSU ribosomal protein L21p
mnas_560	ribosomal protein L27	*rpmA*	peg.58	LSU ribosomal protein L27p
mnas_568	ribosomal protein S4	*rpsD*	peg.83	SSU ribosomal protein S4p (S9e)
mnas_569	ribosomal protein L31	*rpmE*	peg.84	LSU ribosomal protein L31p
mnas_585	ribosomal protein L34	*rpmH*	peg.96	LSU ribosomal protein L34p
mnas_591	ribosomal protein L36	*rpmJ*	peg.101	LSU ribosomal protein L36p
mnas_592	30S ribosomal protein S13	*rpsM*	peg.102	SSU ribosomal protein S13p (S18e)
mnas_593	ribosomal S11 family protein		peg.103	SSU ribosomal protein S11p (S14e)
mnas_649	ribosomal protein L28	*rpmB*	peg.171	LSU ribosomal protein L28p
mnas_664	ribosomal protein L32	*rpmF*	peg.198	LSU ribosomal protein L32p
mnas_679	50S ribosomal protein L3	*rplC*	peg.208	LSU ribosomal protein L3p (L3e)
mnas_681	ribosomal protein S10	*rpsJ*	peg.209	SSU ribosomal protein S10p (S20e)
mnas_714	ribosomal protein L17	*rplQ*	peg.281	LSU ribosomal protein L17p

Supplementary Table 2.10 The aminoacyl-tRNA biosynthesis genes found within the IGS and RAST annotations of the Ms03 draft genome

EC	IGS		RAST	
	Gene name	**Gene id**	**Gene name**	**Gene id**
2.1.2.9	methionyl-tRNA formyltransferase	mnas_249	Methionyl-tRNA formyltransferase	peg.359
6.1.1.1	tyrosine--tRNA ligase	mnas_356	Tyrosyl-tRNA synthetase	peg.480
6.1.1.2	tryptophan--tRNA ligase	mnas_685	Tryptophanyl-tRNA synthetase	peg.211
6.1.1.3	threonyl and Alanyl tRNA synthetase second additional domain protein	mnas_684	Threonyl-tRNA synthetase	peg.35
6.1.1.4	leucine--tRNA ligase	mnas_485	Leucyl-tRNA synthetase	peg.619
6.1.1.5	isoleucine--tRNA ligase	mnas_678	Isoleucyl-tRNA synthetase	peg.53
6.1.1.6	lysine--tRNA ligase	mnas_92	Lysyl-tRNA synthetase (class II)	peg.2
6.1.1.7	alanine--tRNA ligase	mnas_453	Alanyl-tRNA synthetase	peg.577
6.1.1.9	valine--tRNA ligase	mnas_463	Valyl-tRNA synthetase	peg.600
6.1.1.10	methionine--tRNA ligase	mnas_348	Methionyl-tRNA synthetase	peg.474
6.1.1.11	serine--tRNA ligase	mnas_586		
6.1.1.12	aspartate--tRNA ligase	mnas_457	Aspartyl-tRNA synthetase	peg.597
6.1.1.14	glycine--tRNA ligase	mnas_268	Glycyl-tRNA synthetase	peg.383
6.1.1.15	proline--tRNA ligase	mnas_640	Prolyl-tRNA synthetase	peg.165
6.1.1.16	cysteine--tRNA ligase	mnas_373	Cysteinyl-tRNA synthetase	peg.497
6.1.1.17	glutamyl-tRNA synthetase	mnas_458	Glutamyl-tRNA synthetase	peg.307
6.1.1.19	arginine--tRNA ligase	mnas_374	Arginyl-tRNA synthetase	peg.498
6.1.1.20	phenylalanine--tRNA ligase, alpha subunit	mnas_274	Phenylalanyl-tRNA synthetase alpha chain	peg.389
	tRNA synthetase B5 domain protein	mnas_277	Phenylalanyl-tRNA synthetase domain protein (Bsu YtpR)	peg.391
			Phenylalanyl-tRNA synthetase beta chain	peg.481
6.1.1.21	histidine--tRNA ligase	mnas_224	Histidyl-tRNA synthetase	peg.323
6.1.1.24	tRNA synthetases class I (E and Q), catalytic domain protein	mnas_749	Glutamyl-tRNA(Gln) synthetase	peg.307
6.3.5.6	glu-tRNAGln amidotransferase C subunit	mnas_327	Aspartyl-tRNA(Asn) amidotransferase subunit A	peg.452
			Aspartyl-tRNA(Asn) amidotransferase subunit B	peg.451
			Aspartyl-tRNA(Asn) amidotransferase subunit C	peg.453
6.3.5.7	aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase, B subunit	mnas_325	Glutamyl-tRNA(Gln) amidotransferase subunit A	peg.452
			Glutamyl-tRNA(Gln) amidotransferase subunit B	peg.451
			Glutamyl-tRNA(Gln) amidotransferase subunit C	peg.453

Supplementary Table 2.11 Glycolysis / gluconeogenesis genes found within the IGS and RAST annotations of the Ms03 draft genome

IGS				**RAST**	
EC	**Gene_id**	**Gene name**	**Gene symbol**	**Gene_id**	**Gene name**
1.2.1.12	mnas_250[#]	Glyceraldehyde-3-phosphate dehydrogenase, type I	*gap*	peg.360[#]	NAD-dependent glyceraldehyde-3-phosphate dehydrogenase
2.7.1.40	mnas_504	Pyruvate kinase	*pyk*	peg.632	Pyruvate kinase
2.7.1.69	mnas_534	PTS system, glucose/glucosamine/beta-glucoside-specific	-	peg.37	PTS system, N-acetylglucosamine-specific IIA component / PTS system, N-acetylglucosamine-specific IIB component / PTS system, N-acetylglucosamine-specific IIC component
2.7.2.3	mnas_597	Phosphoglycerate kinase	-		
4.1.2.13	mnas_376	Fructose-1,6-bisphosphate aldolase, class II	*fba*	peg.500	Fructose-bisphosphate aldolase class II
4.2.1.11	mnas_695	Phosphopyruvate hydratase	*eno*	peg.263	Enolase
5.3.1.1	mnas_547	Triose-phosphate isomerase	*tpiA*	peg.47	Triosephosphate isomerise
5.3.1.9	mnas_509	Phosphoglucose isomerase family protein		peg.16	Glucose-6-phosphate isomerise
5.3.1.9	mnas_700	Putative glucose-6-phosphate isomerase	-	peg.268	Glucose-6-phosphate isomerise
5.4.2.1	mnas_105	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	-	peg.61	2,3-bisphosphoglycerate-independent phosphoglycerate mutase

[#]Enzyme number for reverse reaction is EC 1.2.1.59

Supplementary Table 2.12 Pentose phosphate pathway

EC	IGS			RAST	
	Gene_id	**Gene name**	**Gene symbol**	**Gene_id**	**Gene name**
2.7.6.1	mnas_724	ribose-phosphate diphosphokinase family protein	*prs*	peg.288	Ribose-phosphate pyrophosphokinase
4.1.2.4	mnas_672	deoxyribose-phosphate aldolase	*deoC*	peg.205	Deoxyribose-phosphate aldolase
4.1.2.9	mnas_436	xylulose-5-phosphate phosphoketolase	*xpkA*	peg.564	Xylulose-5-phosphate phosphoketolase / Fructose-6-phosphate phosphoketolase
4.1.2.13*	mnas_376	Fructose-1,6-bisphosphate aldolase, class II	*fba*	peg.500	Fructose-bisphosphate aldolase class II
5.1.3.1	mnas_113	Ribulose-phosphate 3 epimerase family protein	-	peg.67	Ribulose-phosphate 3-epimerase
5.4.2.7	mnas_513[#]	Phosphopentomutase	*deoB*	peg.19[#]	Phosphopentomutase
5.3.1.6	mnas_286	Putative ribose-5-phosphate isomerase	-	peg.399	Ribose 5-phosphate isomerase B
5.3.1.9*	mnas_509	Phosphoglucose isomerase family protein		peg.16	Glucose-6-phosphate isomerise
5.3.1.9*	mnas_700	Putative glucose-6-phosphate isomerase	-	peg.268	Glucose-6-phosphate isomerase

*Part of glycolysis
[#] Enzyme number for reverse reaction is 5.4.2.2

Supplementary Table 2.13 Pyruvate metabolism

EC	IGS			RAST	
	Gene_id	**Gene name**	**Gene symbol**	**Gene_id**	**Gene name**
1.1.1.28	mnas_107	D-lactate dehydrogenase	*ldhA*	peg.63	D-lactate dehydrogenase
2.7.2.1	mnas_507	Acetate kinase	*ackA*	peg.635	Acetate kinase
2.7.1.40*	mnas_504	Pyruvate kinase	*pyk*	peg.632	Pyruvate kinase

*Part of glycolysis

Supplementary Table 2.14 Membrane associated, substrate-binding and transport proteins in the Ms03 draft genome.

IGS			RAST	
Gene_id	**Gene name**	**Gene symbol**	**Gene_id**	**Gene name**
Putative potassium uptake (ktrB/ktrA)				
mnas_11	Putative potassium uptake protein KtrB	*ktrB*	peg.222	Potassium uptake protein, integral membrane component, KtrB
mnas_12	TrkA-C domain protein	*ktrA*	peg.223	Trk system potassium uptake protein TrkA
Magnesium (channel-type mechanism)				
mnas_189	Magnesium transporter	*mgtE*	peg.259	Mg/Co/Ni transporter MgtE / CBS domain
Putative cation transporting P-type ATPase (Copper, Lead, cadmium, zinc and mercury)				
mnas_196	Cation transporting ATPase, family protein		peg.280	Lead, cadmium, zinc and mercury transporting ATPase (EC 3.6.3.3) (EC 3.6.3.5); Copper-translocating P-type ATPase (EC 3.6.3.4) 259481-258960
Phosphoenolpyruvate-dependent sugar phosphotransferase system				
mnas_283	phosphocarrier protein HPr (EC 7.11.-)	*ptsH*	peg.396	Phosphotransferase system, phosphocarrier protein HPr
mnas_410	Putative PTS system glucose-specific enzyme IIB component		peg.545	PTS system glucose-specific enzyme IIB component
mnas_62	HPr(Ser) kinase/phosphatase (EC 2.7.1.-)	*hprK*	peg.505	HPr kinase/phosphorylase (EC 2.7.1.-) (EC 2.7.4.-)
mnas_534	PTS system, glucose/glucosamine/beta-glucoside-specific (EC 2.7.1.69)		peg.37	PTS system, N-acetylglucosamine-specific IIA component (EC 2.7.1.69) / PTS system, N-acetylglucosamine-specific IIB component (EC 2.7.1.69) / PTS system, N-acetylglucosamine-specific IIC component (EC 2.7.1.69)
Putative energy-coupling factor transporter				
mnas_651	Energy-coupling factor transporter ATP-binding EcfA1 domain protein (EC 3.6.3.-)	*ecfA1*	peg.172	ATPase component of general energizing module of ECF transporters
mnas_652	Energy-coupling factor transporter ATP-binding protein EcfA2 (EC 3.6.3.-)	*ecfA2*	peg.173	ATPase component of general energizing module of ECF transporters
mnas_653	energy-coupling factor transporter transmembrane protein EcfT	*ecfT*		
mnas_49	ECF-type riboflavin transporter, S component family protein		peg.423	Substrate-specific component FolT of folate ECF transporter
Putative spermidine/putrescine ABC transporter permease				
mnas_227	Probable spermidine/putrescine/ABC transporter substrate	*potD*	peg.334	Probable spermidine/putrescine substrate binding protein in Mollicutes
mnas_240	Spermidine/putrescine ABC transporter permease PotB domain protein	*potB*		
mnas_241	Putative spermidine/putrescine transport system ATP-binding protein (EC 3.6.3.31)	*potA*		
mnas_242	Putative spermidine/putrescine import ATP-binding protein	*potA*		
mnas_243	Spermidine/putrescine import ATP-binding PotA domain protein (EC 3.6.3.31)	*potA*		
mnas_478	Spermidine/putrescine/ABC transporter substrate binding protein	*potD*	peg.613	Probable spermidine/putrescine substrate binding protein in Mollicutes
mnas_770	Spermidine/putrescine ABC transporter permease protein PotB	*potB*	peg.349	Spermidine Putrescine ABC transporter permease component PotB (TC 3.A.1.11.1)
mnas_771	Spermidine/putrescine transport system permease PotC domain protein	*potC*		

IGS			RAST	
Gene_id	**Gene name**	**Gene symbol**	**Gene_id**	**Gene name**
Putative glycerol-3-phosphate ABC transporter				
mnas_488	Putative sn-glycerol-3-phosphate ABC transporter, ATP-binding protein	*ugpC*	peg.621	Multiple sugar ABC transporter, ATP-binding protein
mnas_489	Putative sn-glycerol-3-phosphate transport system permease protein	*ugpA*	peg.622	N-Acetyl-D-glucosamine ABC transport system, permease protein
mnas_490	ABC transporter permease protein ugpE	*ugpE*	peg.623	ABC transporter permease protein
mnas_751	sn-glycerol-3-phosphate ABC transporter, ATP-binding UgpC domain protein	*ugpC*		
p37 transport system				
mnas_51	High affinity transport system p37 family protein		peg.425	High affinity transport system protein p37 precursor
mnas_52	Phosphate/phosphonate ABC transporter, ATP-binding protein (EC 3.6.3.28)	*PhnC*	peg.426	ABC transporter ATP-binding protein
mnas_53	Binding-dependent transport system inner membrane component family protein		peg.427	Transport system permease protein p69
Oligopeptide permease ABC transporter				
mnas_50	Hypothetical protein		peg.424	Oligopeptide ABC transporter ATP-binding protein
mnas_74	Bacterial extracellular solute-binding protein	*oppA*	peg.581	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA (TC 3.A.1.5.1)
mnas_75	Oligopeptide ABC transporter, permease protein (OppB)	*oppB*	peg.582	Oligopeptide transport system permease protein OppB (TC 3.A.1.5.1)
mnas_76	Oligopeptide ABC transporter, permease protein (OppC)	*oppC*	peg.583	Oligopeptide transport system permease protein OppC (TC 3.A.1.5.1)
mnas_77	Oligopeptide/dipeptide ATP-binding protein	*oppD*	peg.584	Oligopeptide transport ATP-binding protein OppD (TC 3.A.1.5.1)
mnas_78	Oligopeptide ABC transporter, ATP-binding protein OppF	*oppF*	peg.585	Oligopeptide transport ATP-binding protein OppF (TC 3.A.1.5.1)
mnas_414	Oligopeptide ABC transporter substrate-binding protein	*oppA*	peg.547	Lipoprotein
mnas_415	Oligopeptide ABC transporter, permease protein OppB	*oppB*		
mnas_576	Oligopeptide ABC transporter system, permease protein OppC	*oppC*	peg.90	Oligopeptide transport system permease protein OppC (TC 3.A.1.5.1)
mnas_577	Oligopeptide ABC transporter, ATP-binding protein OppD	*oppD*	peg.91	Oligopeptide transport ATP-binding protein OppD
mnas_601	Oligopeptide ABC transporter, ATP-binding protein OppF	*oppF*	peg.110	Oligopeptide ABC transporter ATP-binding protein
mnas_806	Putative oligopeptide ABC transporter, ATP-binding protein OppF C-terminal			
Putative Lipid A export				
mnas_148	ABC transporter family protein		peg.138	Lipid A export ATP-binding/permease protein MsbA
mnas_407	ABC transporter family protein		peg.542	Lipid A export ATP-binding/permease protein MsbA
Membrane protein with beta galactosidase activity				
mnas_144	Putative Beta-galactosidase		peg.134	Beta-galactosidase (EC 3.2.1.23)
mnas_145	Beta-galactosidase domain protein (EC 3.2.1.23)	*bgaA2*	peg.135	hypothetical protein
mnas_146	Beta-galactosidase domain protein (EC 3.2.1.23)	*bgaA*	peg.136	Beta-galactosidase (EC 3.2.1.23)
mnas_147	Glycosyl hydrolases family 2, sugar binding domain protein		peg.137	Beta-galactosidase (EC 3.2.1.23)

IGS			RAST	
Gene_id	**Gene name**	**Gene symbol**	**Gene_id**	**Gene name**
Hypothetical membrane, peripheral, transport and putative ABC transporters				
mnas_476	ZIP Zinc transporter family protein		peg.611	hypothetical protein
mnas_199	conserved hypothetical protein		peg.292	FIG000605: protein co-occurring with transport systems (COG1739)
mnas_226	hemolysin C domain protein	*hlyC*	peg.333	Hypothetical protein
mnas_797	amino acid permease family protein			
mnas_107	D-lactate dehydrogenase (EC 1.1.1.28)	*ldhA*	mnas_107	D-lactate dehydrogenase
mnas_401	DNA topoisomerase 4 subunit A (EC 5.99.1.-)	*parC*	peg.537	Topoisomerase IV subunit A (EC 5.99.1.-)
mnas_239	Putative sugar ABC transporter permease protein		peg.345	Unspecified monosaccharide ABC transport system, permease component 2
			peg.346	Sugar ABC transporter, permease protein
mnas_93	ABC transporter family protein		peg.3	ABC transporter ATP-binding protein
mnas_94	Putative ABC transporter permease protein		peg.4	ABC transporter permease protein
mnas_701	Multidrug resistance ABC superfamily ATP binding cassette transporter, membrane domain protein (EC 3.6.3.42)	*ndvA*		
mnas_273	AAA ATPase, central region		peg.388	ATPase, AAA family
mnas_212	ABC-2 type transporter family protein		peg.312	ABC transporter, permease protein
mnas_213	ABC transporter family protein		peg.313	Methionine ABC transporter ATP-binding protein
mnas_287	ABC transporter family protein		peg.400	ABC transporter ATP-binding and permease protein (MDR homolog)
mnas_341	ABC transporter family protein		peg.467	ABC transporter ATP-binding protein uup
mnas_406	ABC transporter transmembrane region family protein			
mnas_565	ABC transporter transmembrane region family protein			
mnas_564	ABC transporter family protein		peg.60	FIG00836097: hypothetical protein
mnas_781	ABC transporter family protein			
mnas_802	ABC transporter family protein			
mnas_654	putative membrane protein		peg.174	Unspecified monosaccharide ABC transport system, ATP-binding protein 748139-746763
mnas_579	ABC superfamily ATP binding cassette transporter, ABC domain protein (EC 3.6.3.24)	*nikE*		
mnas_655	Putative Sugar ABC transporter ATP-binding protein	*mglA*	peg.176	Unspecified monosaccharide ABC transport system, ATP-binding protein

Supplementary Table 2.15 The InterPro results of the Ms03 Opp proteins.

OppA (1288)	**Protein family membership: None predicted** **InterPro Domains:** None predicted **GO term prediction** **Biological Process:** None predicted **Molecular Function:** None predicted **Cellular Component:** None predicted **Signature matches**		
	Program	**Motif/domain**	**Regions**
	PROSITE	PS51257 (PROKAR_LIPOPROTEIN)	1-23
	COILS	Coil	68-103, 158-224, 249-284
	PHOBIUS	SIGNAL_PEPTIDE	1-20
		SIGNAL_PEPTIDE_C_REGION	17-20
		SIGNAL_PEPTIDE_H_REGION	5-16
		SIGNAL_PEPTIDE_N_REGION	1-4
		Non_Cytoplasmic_Domains	21-1288
OppA2 (1030)	**Protein family membership:** ABC-type oligopeptide transport system, solute-binding component, Mycoplasmataceae, predicted (IPR016880) **InterPro Domains:** Solute-binding protein family 5 domain (IPR000914) **GO term prediction** **Biological Process:** None predicted **Molecular Function:** None predicted **Cellular Component:** None predicted **Signature matches**		
	Program	**Motif/domain**	**Regions**
	PIR	PIRSF028335 (ABC_oligopep_OppA_prd)	2-1030
	Pfam	PF00496 (SBP_bac_5)	191-710
	GENE3D	CATH Superfamily: G3DSA:3.10.105.10	617-749
	PROSITE	PS51257 (PROKAR_LIPOPROTEIN)	1-27
	PHOBIUS	SIGNAL_PEPTIDE	1-27
		SIGNAL_PEPTIDE_C_REGION	20-27
		SIGNAL_PEPTIDE_H_REGION	8-19
		SIGNAL_PEPTIDE_N_REGION	1-7
	SUPERFAMILY	SSF53850:Periplasmic binding protein-like II	391-589, 632-776
	COILS	Coil	784-805

OppB (387)

Protein family membership: None predicted
InterPro Domains: ABC transporter type 1, transmembrane domain MetI-like (IPR000515)
GO term prediction
Biological Process: GO:0006810 transport
Molecular Function: GO:0005215 transporter activity
Cellular Component: GO:0016020 membrane
Signature matches

Program	Motif/domain	Regions
Pfam	PF00528: BPD_transp_1	150-353
PROSITE	PS50928: ABC_TM1	132-341
GENE3D	CATH Superfamily: G3DSA:1.10.3720.10	126-346
SUPERFAMILY	SSF161098: MetI-like	126-336
PANTHER	PTHR30465 : Peptide transport system permease protein	1-359
	PTHR30465:SF0: Oligopeptide Transport System Permease Protein OppB	1-359
PHOBIUS	Cytoplasmic_Domains	1-8, 160-170, 235-279, 349-387
	Non_Cytoplasmic_Domains	31-133, 196-214, 303-321
	Transmembrane_Domains	9-30, 134-159, 171-195, 215-234, 280-302, 322-348
TMHMM	TMhelix	9-30, 136-158, 171-193, 218-240, 280-302, 322-344

OppB2 (351)

Protein family membership: None predicted
InterPro Domains: ABC transporter type 1, transmembrane domain MetI-like (IPR000515)
GO term prediction
Biological Process: GO:0006810 transport
Molecular Function: None predicted
Cellular Component: GO:0016020 membrane
Signature matches

Program	Motif/domain	Regions
Pfam	PF00528: BPD_transp_1	157-348
PROSITE	PS50928: ABC_TM1	139-335
GENE3D	CATH Superfamily: G3DSA:1.10.3720.10	134-344
SUPERFAMILY	SSF161098: MetI-like	133-348
PANTHER	PTHR30465 : Peptide transport system permease protein	27-350
	PTHR30465:SF0: Oligopeptide Transport System Permease Protein OppB	27-350
PHOBIUS	Cytoplasmic_Domains	1-52, 166-176, 233-269, 399-351
	Non_Cytoplasmic_Domains	73-142, 202-212, 297-315
	Transmembrane_Domains	53-71, 143-165, 177-201, 213-232, 270-296, 316-338
TMHMM	TMhelix	53-75, 142-164, 176-198, 213-235, 270-292, 317-339

OppC (453)

Protein family membership: None predicted
InterPro Domains: Oligopeptide transport permease C-like, N- terminal domain (IPR025966)
ABC transporter type 1, transmembrane domain MetI-like (IPR000515)
GO term prediction
Biological Process: GO:0006810 transport
Molecular Function: None predicted
Cellular Component: GO:0016020 membrane
Signature matches:

Program	Motif/domain	Regions
Pfam	PF12911: OppC_N	40-82
	PF00528: BPD_transp_1	267-450
PROSITE	PS50928: ABC_TM1	253-438
SUPERFAMILY	SSF161098: MetI-like family	244-431
PANTHER	PTHR30465: Peptide Transport System Permease Protein	25-448
	PTHR30465:SF2 : Dipeptide transport system permease protein dppc-related	25-448
PHOBIUS	Cytoplasmic_Domain	1-49, 277-287, 337-374, 442-453
	Non_Cytoplasmic_Domain	74-251, 311-375, 396-417
	Transmembrane	50-73, 252-276, 288-310, 316-336, 375-395, 418-441
TMHMM	TMhelix	50-72, 255-277, 289-311, 315-337, 371-393, 419-441

Opp2C (360)

Protein family membership: None Predicted
InterPro Domains: Oligopeptide transport permease C-like, N- terminal domain (IPR025966)
ABC transporter type 1, transmembrane domain MetI-like (IPR000515)
GO term prediction
Biological Process: GO:0006810 transport
Molecular Function: None predicted
Cellular Component: GO:0016020 membrane
Signature matches:

Program	Motif/domain	Regions
Pfam	PF12911: OppC_N	**84-132**
	PF00528: BPD_transp_1	181-354
PROSITE	PS50928: ABC_TM1	163-349
SUPERFAMILY	SSF161098: MetI-like family	156-342
GENE3D	G3DSA:1.10.3720.10	154-357
PANTHER	PTHR30465: Peptide Transport System Permease Protein	68-356
	PTHR30465:SF2 : Dipeptide transport system permease protein dppc-related	68-356
PHOBIUS	Cytoplasmic_Domain	1-95, 189-199, 245-276, 353-360
	Non_Cytoplasmic_Domain	118-164, 220-224, 308-326
	Transmembrane	96-117, 165-188, 200-219, 225-244, 277-307, 327-352
TMHMM	TMhelix	96-118, 165-187. 199-221, 225-244, 273-295, 330-352

OppD (348)	Protein family membership: None predicted. InterPro Domains: P-loop containing nucleoside triphosphate hydrolase (IPR027417) AAA+ ATPase domain (IPR003593) ABC transporter-like (IPR003439) Oligopeptide/dipeptide ABC transporter, C- terminal (IPR013563) Oligopeptide/dipeptide ABC transporter, ATP-binding protein, C-terminal (IPR010066) GO term prediction Biological Process: GO:0015833 peptide transport Molecular Function: GO:0000166 nucleotide binding GO:0005524 ATP binding GO:0015197 peptide transporter activity GO:0016887 ATPase activity Cellular Component: GO:0016020 membrane Signature matches:		
	Program	Motif/domain	Regions
	Pfam	PF00005: ABC_tran	27-187
		PF08352: oligo_HPY	239-300
	PROSITE	PS50893: ABC_TRANSPORTER_2	8-260
	GENE3D	G3DSA: 3.40.50.300	2-263
	SUPERFAMILY	SSF52540: P-loop containing nucleoside triphosphate hydrolases	34-310
	PANTHER	PTHR24220: FAMILY NOT NAMED	1-321
		PTHR24220:SF177: D,D-DIPEPTIDE TRANSPORT ATP-BINDING PROTEIN DDPD-RELATED	1-321
	TIGRFAMs	TIGR01727	237-320
	SMART	SM00382 (AAA)	35-237
OppD2 (402)	Protein family membership: None predicted. InterPro Domains: P-loop containing nucleoside triphosphate hydrolase (IPR027417) AAA+ ATPase domain (IPR003593) ABC transporter-like (IPR003439) Oligopeptide/dipeptide ABC transporter, C- terminal (IPR013563) Oligopeptide/dipeptide ABC transporter, ATP-binding protein, C-terminal (IPR010066) GO term prediction Biological Process: GO:0015833 peptide transport Molecular Function: GO:0000166 nucleotide binding GO:0005524 ATP binding GO:0015197 peptide transporter activity GO:0016887 ATPase activity Cellular Component: GO:0016020 membrane Signature matches:		
	Program	Motif/domain	Regions
	Pfam	PF00005: ABC_tran	80-239
		PF08352: oligo_HPY	291-352
	PROSITE	PS00211: ABC_TRANSPORTER_1	212-226
		PS50893: ABC_TRANSPORTER_2	58-312
	GENE3D	G3DSA: 3.40.50.300	58-312
	SUPERFAMILY	SSF52540: P-loop containing nucleoside triphosphate hydrolases	62-310
	PANTHER	PTHR24220: FAMILY NOT NAMED	57-388
		PTHR24220:SF177: D,D-dipeptide transport ATP-binding protein ddpD-related	57-388
	TIGRFAMs	TIGR01727: oligo_HPY	289-371
	SMART	SM00382 (AAA)	87-297

OppF (795)

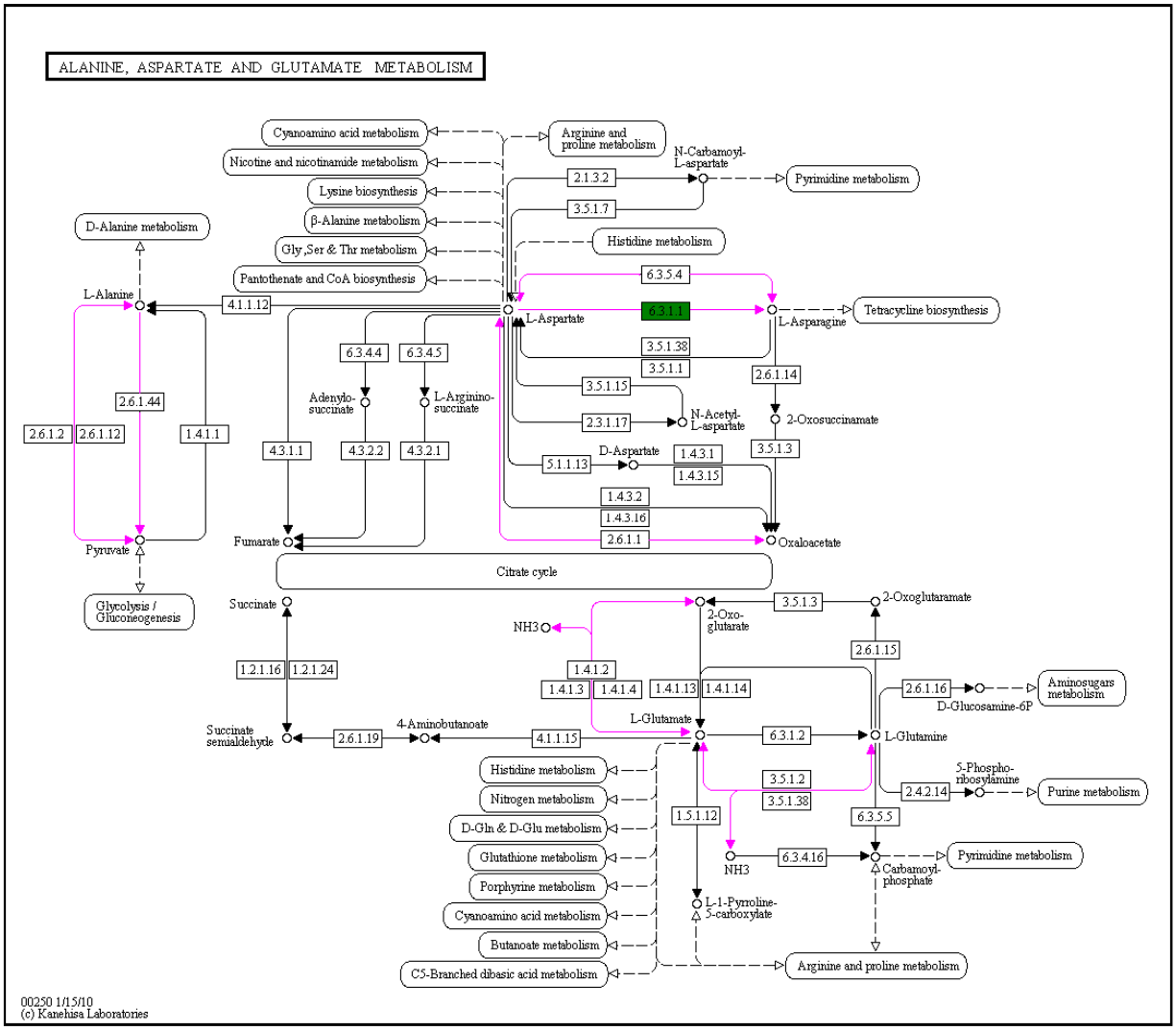
Protein family membership: None predicted.
InterPro Domains: P-loop containing nucleoside triphosphate hydrolase (IPR027417)
AAA+ ATPase domain (IPR003593)
ABC transporter-like (IPR003439)
Oligopeptide/dipeptide ABC transporter, C- terminal (IPR013563)
GO term prediction
Biological Process: GO:0015833 peptide transport
Molecular Function: GO:0000166 nucleotide binding
GO:0005524 ATP binding
GO:0016887 ATPase activity
GO:0017111 nucleoside-triphosphatase activity
Cellular Component: None predicted.
Signature matches

Program	Motif/domain	Regions
Pfam	PF00005: ABC_tran	26-127, 594-660
	PF08352: oligo_HPY	711-743
PROSITE	PS50893: ABC_TRANSPORTER_2	6-732
	PS00211: ABC_TRANSPORTER_1	632-646
GENE3D	G3DSA: 3.40.50.300	2-127, 610-735
SUPERFAMILY	SSF52540: P-loop containing nucleoside triphosphate hydrolases	4-126, 596-723
PANTHER	PTHR24220: FAMILY NOT NAMED	1-744
	PTHR24220:SF214: D,D-DIPEPTIDE TRANSPORT ATP-BINDING PROTEIN DDPF-RELATED	1-744
SMART	SM00382: AAA	34-709
COILS	Coil	394-419,441-462, 470-491

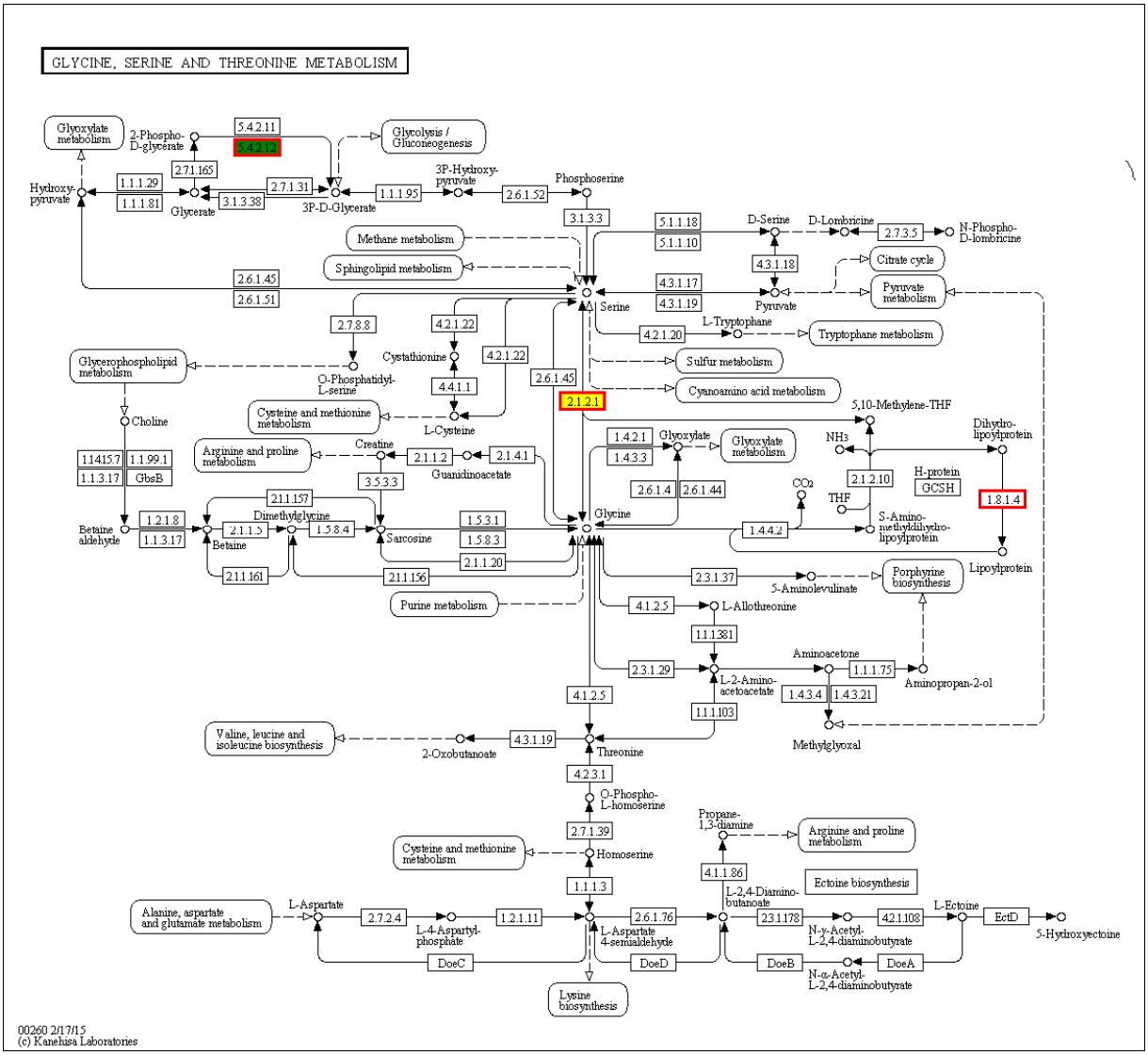
OppF2 (443)

Protein family membership: None predicted.
InterPro Domains: P-loop containing nucleoside triphosphate hydrolase (IPR027417)
AAA+ ATPase domain (IPR003593)
ABC transporter-like (IPR003439)
Oligopeptide/dipeptide ABC transporter, C- terminal (IPR013563)
GO term prediction
Biological Process: GO:0015833 peptide transport
Molecular Function: GO:0000166 nucleotide binding
GO:0005524 ATP binding
GO:0016887 ATPase activity
Cellular Component: None predicted.
Signature matches

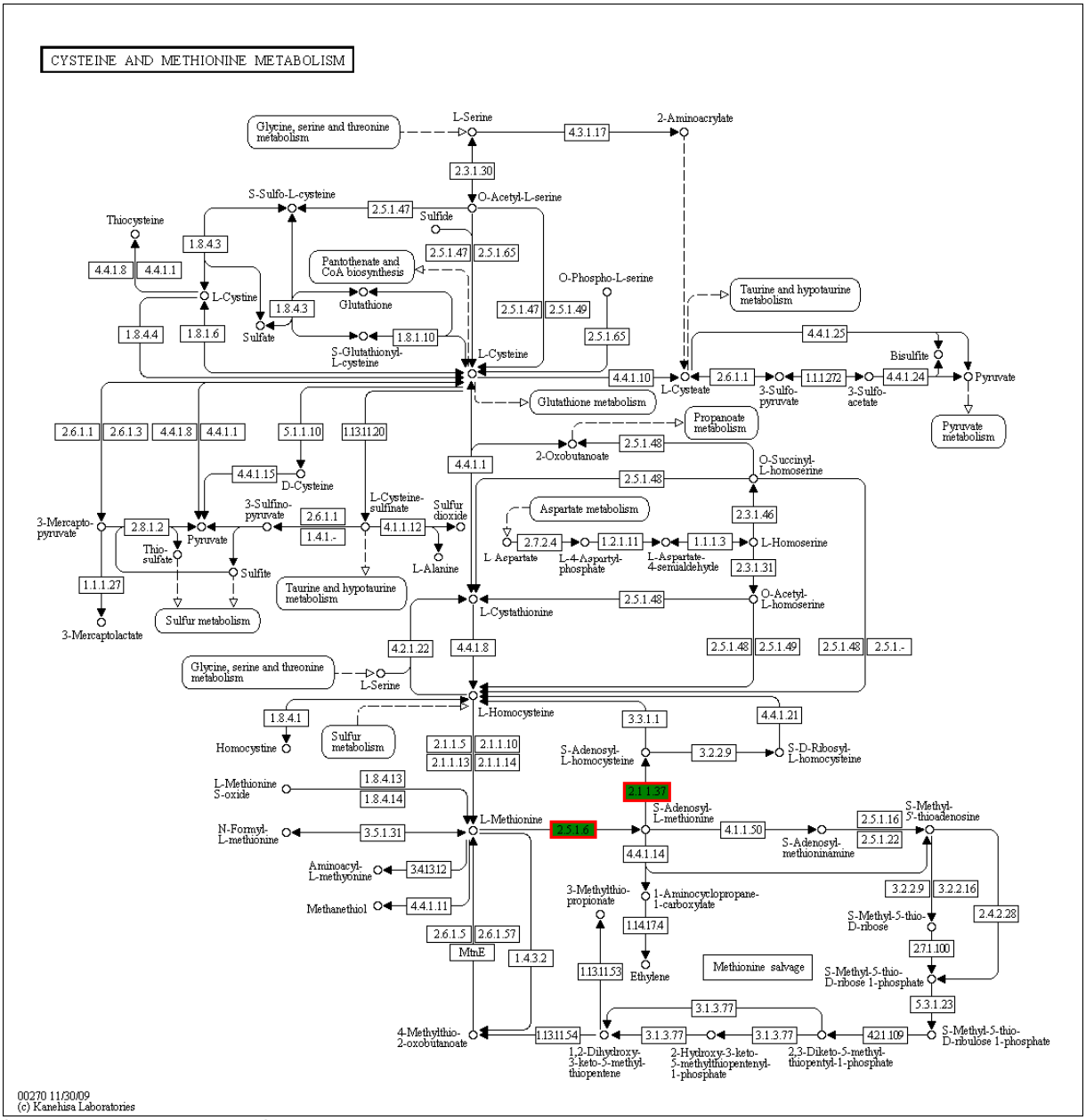
Program	Motif/domain	Regions
Pfam	PF00005: ABC_tran	64-312
	PF08352: oligo_HPY	364-396
PROSITE	PS50893: ABC_TRANSPORTER_2	45-385
	PS00211: ABC_TRANSPORTER_1	285-299
GENE3D	G3DSA: 3.40.50.300	44-162, 249-395
SUPERFAMILY	SSF52540: P-loop containing nucleoside triphosphate hydrolases	46-367
PANTHER	PTHR24220: FAMILY NOT NAMED	44-394
	PTHR24220:SF336: D,D-DIPEPTIDE TRANSPORT ATP-BINDING PROTEIN DDPF-RELATED	44-394
SMART	SM00382: AAA	71-362
COILS	Coil	234-258



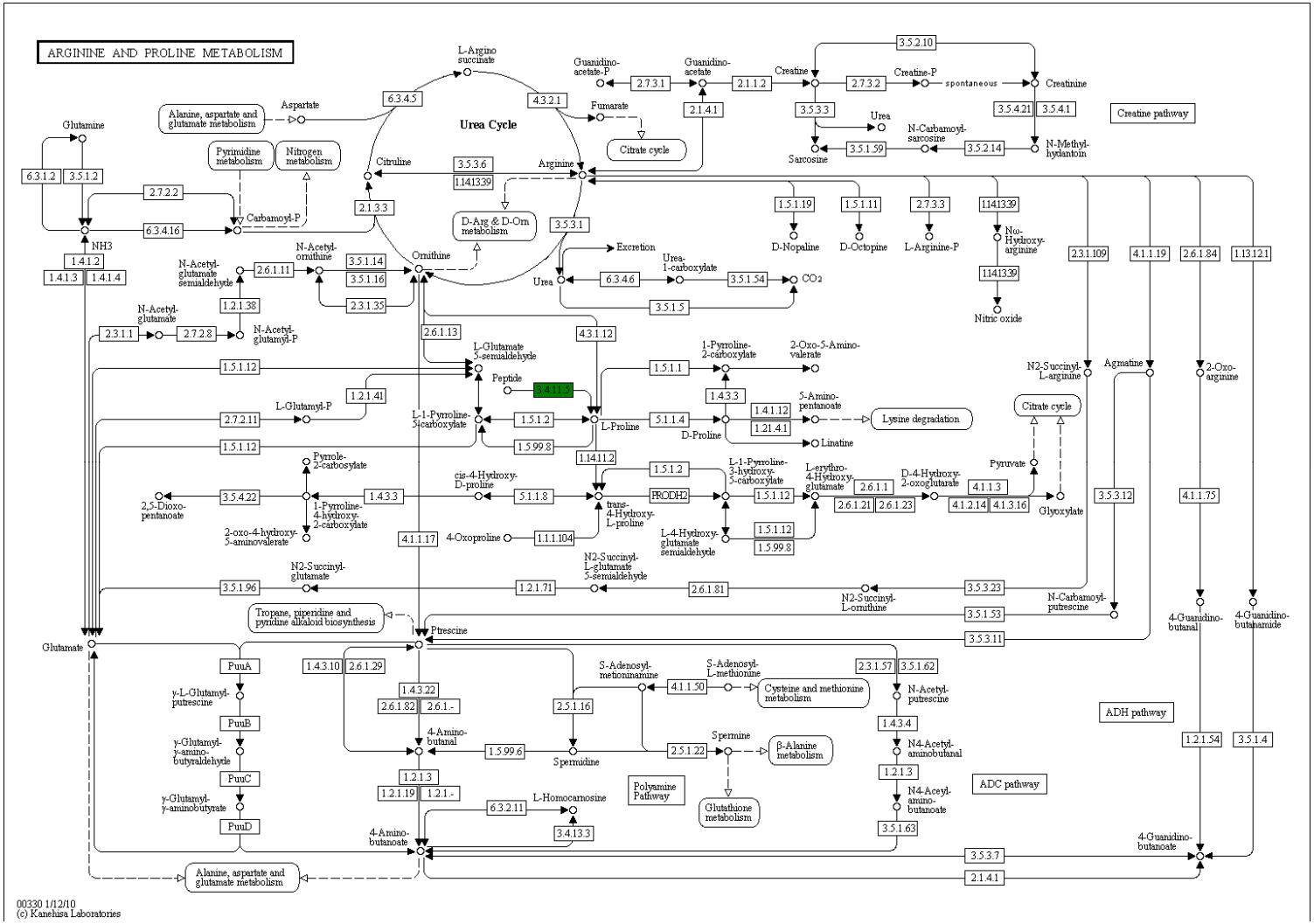
Supplementary Figure 2.1 Alanine, aspartate and glutamate metabolism in Ms03. The enzymes coloured in green were annotated in both RAST and IGS analyses while the enzymes in white were not annotated in Ms03 draft genome annotations. In the *M. synoviae 53* genome no enzymes were annotated for this pathway. The schematic was downloaded from metabolic analysis in RAST and adapted to include the Ms03 IGS annotated proteins as well as the KEGG pathway of the *M. synoviae 53*. This schematic represents the KEGG reference pathway 00250.



Supplementary Figure 2.2 Glycine, serine and theonine metabolism in Ms03. The enzymes coloured in green were annotated in both RAST and IGS analyses, enzymes in yellow were only annotated in IGS analyses while the enzymes in white were not annotated in Ms03 draft genome annotations. Annotated enzymes of the complete *M. synoviae 53* genome are indicated with red boxes. The schematic was downloaded from metabolic analysis in RAST and adapted to include the Ms03 IGS annotated proteins as well as the KEGG pathway of the *M. synoviae 53* (msy00270). This schematic represents the KEGG reference pathway 00270.



Supplementary Figure 2.3 Cysteine and methionine metabolism in Ms03. The enzymes coloured in green were annotated in both RAST and IGS analyses while the enzymes in white were not annotated in Ms03 draft genome annotations. Annotated enzymes of the complete *M. synoviae 53* genome are indicated with red boxes. The schematic was downloaded from metabolic analysis in RAST and adapted to include the Ms03 IGS annotated proteins as well as the KEGG pathway of the *M. synoviae 53* (msy00270). This schematic represents the KEGG reference pathway 00270.



Supplementary Figure 2.4 Arginine and proline metabolism in Ms03. The enzymes coloured in green were annotated in both RAST and IGS analyses while the enzymes in white were not annotated in Ms03 draft genome annotations. In the *M. synoviae 53* genome no enzymes were annotated for this pathway. The schematic was downloaded from metabolic analysis in RAST and adapted to include the Ms03 IGS annotated proteins. This schematic represents the KEGG reference pathway 00330.

Appendix 3 Supplementary tables and figures for Chapter 4

Supplementary Table 1 Inventory of the *opp* operons within the mycoplasma genomes used for analyses

Species	Abbr	NCBI nr	Annotation *opp* operons (genomic location[1])					
Candidatus Mycoplasma haemolamae str. Purdue	MHPL	NC_018219.1	ND					
Candidatus Mycoplasma haemominutum 'Birmingham 1'	MHM	NC_021007.1	ND					
M. agalactiae	MAGa	NC_013948.1	*oppA*[2]*BCDF* (39649-34950)	B	*MAGa1090*[3]*, oppBCDF* (115992-124849)	A		
M. agalactiae PG2	MAG	NC_009497.1	*oppA*[2]*BCDF* (40944-32960)	B	*MAG1000*[3]*, oppBCDF* (112247-121103)	A		
M. alligatoris A21JP2	MALL	NZ_ADNC00000000.1	*MALL_0660*[3]*-0664* (contig: ADNC01000007.1, 22108-31018)	A	*MALL_0007*[2]*-0011* (contig: ADNC01000018.1, 2432-10724)	B1	*MALL_0020*[2]*-0024* (contig: ADNC01000020.1, 7011-14606)	B2
M. arthritidis 158L3-1	MARTH	NC_011025.1	*MARTH_orf403*[2]*, oppB1C1DF* (366708-357886)	A				
M. bovis Hubei-1	MMB	NC_015725.1	*oppA1*[2]*B1C1D1F1* (40995-32981)	B	*oppA2*[3]*B2C2D2F2* (121202-127178)	A		
M. bovis PG45	MBOVPG45	NC_014760.1	*MBOVPG45_0037*[2]*, oppB,* MBOVPG45_0035, oppDF (39666-31652)	B	*MBOVPG45_0112*[3]*-0116* (119020-127881)	A		
M. capricolum subsp. *capricolum* ATCC 27343	MCAP	NC_007633.1	*MCAP_0116*[2]*-0120* (138537-146357)	B	*MCAP_0161-0164, MCAP_0165*[2] (195862-204849)	C		
M. conjunctivae HRC/581	MCJ	NC_012806.1	*MCJ_002370*[3]*, oppBCD-1F* (220854-229281)	A	*oppB-1C-1D-1F-1,* 2 genes[4], *MCJ_004440*[2] (490827-506717)	B		
M. crocodyli MP145	MCRO	NC_014014.1	*oppA*[2]*B-1C-1D-1F-1* (389928-399020)	B	*MCRO_0618*[3]*, oppB-2C-2D-2F-2* (726278-717361)	A		
M. fermentans JER	MFE	NC_014552.1	*MFE_02290*[3]*, oppB1C1D1F3* (245804-254675)	A	*oppA*[2]*B2C2D2F4* (197036-205524)	B		
M. fermentans M64	MfeM64YM	NC_014921.1	*oppA*[2]*B-1C-1D-1F-1* (213463-221843)	B	*MfeM64YM_0281*[3]*, oppB-2C-2D-2F-2* (282323-291194)	A		
M. gallisepticum str. R(low)	MGA	NC_004829.2	*MGA_0226*[3, 5], dpp/oppBb_1Ba_1C_1D_1F_1 (721805-716646)	A	*oppA*[2]*, dpp/oppB_2C_3D_2F_2* (724918-717083)	B		
M. gallisepticum str. R(high)	MGAH	CP001872.1	*MGAH_0226*[3, 5]*, dppB, MGAH_0221a, 0221b, dppDF* (716475-707492)	A	*oppA_2*[2]*, dppBCDF*[11] (724747-716912)	B		
M. gallisepticum str. F	MGF	CP001873.1	*MGF_2297*[3, 5]*, dppBCDF* (389898-398879)	A	*oppA_2*[2]*, dppBCDF* (381626-389461)	B		
M. genitalium G37	MG	NC_000908.2	*OppBCDF* (99383-109206) *MG_321*[3,5] (400119-402923)	A				
M. haemocanis str. Illinois	MHC	NC_016638.1	ND					
M. haemofelis str. Langford 1	HF1	NC_014970.1	ND					
M. hominis ATCC 23114	MHO	NC_013511.1	*oppA*[3]*BCDF* (178627-187638)	A				

Species	Abbr	NCBI nr	Annotation *opp* operons (genomic location[1])					
M. hyopneumoniae 232	mhp	NC_006360.1	*oppBCDF, gene*[6]*, mhp164*[2] (192435-206054)	B	*mhp502*[3]*, oppBCDF* (633310-624928)	A		
M. hyopneumoniae 7448	MHP7448	NC_007332.1	*oppB-1C-1DF-1*, gene[6], *MHP7448_0217*[2] (243270-256889)	B	*MHP7448_0505*[3]*, oppBCD-1F* (670413-662031)	A		
M. hyopneumoniae J	MHJ	NC_007295.1	*oppB-1C-1D-1F-1*, gene[6], *MHJ_0213*[2] (235856-249475)	B	*MHJ_0502*[3]*, oppBCDF* (647232- 638850)	A		
M. hyorhinis HUB-1	MHR	NC_014448.1	*oppBCDF*, gene[6], *MHR_0357*[2] (467543-451680)	B	*MHR_0639*[3]*, oppBCDF* (786329-777920)	A		
M. iowae 695	GUU	NZ_AGFP00000000.1	*GUU_02828*[2]*, 02833, 02838, 02843, 02848* (contig: AGFP01000029.1, 7327-16111)	A				
M. leachii PG50 [17]	MSB	NC_014751.1	*MSB_A0161*[2]*,0162-165* (180561-188377)	B	*MSB_A0212-0215, MSB_A0216*[2] (244151- 253117)	C		
M. mobile 163K	MMOB	NC_006908.1	*MMOB4520*[3]*, oppBC, pgk, oppF* (564402-554792)	A				
M. mycoides subsp. *capri* LC str. 95010	MLC	NC_015431.1	*oppBCDFA*[2] (205424-214429)	C	*oppA*[2]*BCDF* (1029320-1021513)	B		
M. mycoides subsp. *mycoides* SC str. PG1	MSC	BX293980.2	*oppBCDFA*[2] (210699-219713)	C	*oppA*[2]*BCDF* (1097481-1105303)	B		
M. ovis str. Michigan	OVS	NC_023062.1	ND					
M. parvum str. Indiana	PRV	NC_022575.1	ND					
M. penetrans HF-2	MYPE	NC_004432.1	*MYPE5560*[3, 5]*, oppBCDF* (709149-700280)	A1	*oppBCDF* (1183321-1176983) *MYPE7570-MYPE7620*[3, 5] (993492-1008762)	A2		
M. pneumoniae 309	MPNA	NC_016807.1	*oppBCDF*, (264311-270394) *MPNA4560*[3, 5] (553754-556771)	A				
M. pneumoniae M129	MPN	NC_000912.1	*oppB, amiD, oppDF* (265910-271993) *MPN456*[3, 5] (555398-558415)	A				
M. pulmonis UAB CTIP	MYPU	NC_002771.1	*MYPU_2820*[2]*, oppBCDF* (327933-336205)	A	*oppBCDF, gene*[6]*, MYPU_4150*[2] (485792-500096)	B		
M. putrefaciens KS1	MPUT	NC_015946.1	*oppBCDFA*[3] (745676-736919)	C				
M. suis str. Illinois	MSU	NC_015155.1	ND					
M. synoviae 53	MS53	NC_007294.1	*oppBC, MS53_0186, oppD*[7], 2 genes[4], *MS53_0190*[3] (199080-213247)	B	*MS53_0349*[3]*, oppB-1C-1D-1F-1* (401341-392438)	A		
M. nasistruthionis sp. nov str. Ms03	Ms03	KM410300 KM410301	*oppA*[3]*BCDF*	A	*oppA2*[2]*B2C2D2F2*	B		
Mycoplasma sp. Ms02	Ms02	KM410302 KM410304	*oppA*[3]*BCDF*	A	*B2C2F2D2*, 3 genes, *oppA2*[2]	B		

Species	Abbr	NCBI nr	Annotation *opp* operons (genomic location[1])					
M. struthionis sp. nov. str. Ms01	Ms01	KM410303	*oppA*[3]*BCDF*	A				
M. wenyonii str. Massachusetts	WEN	NC_018149.1	ND					

[1] From the start codon of the first gene to the stop codon of the last gene
[2] Conserve domain found in NCBI BLAST search
[3] No putative conserve domain found in NCBI BLAST
[4] Two genes are located between *oppBCDF* and the putative *oppA*, both located on the same strand, transcribe unidirectional and code for hypothetical proteins
[5] *oppA* gene identified with two PSI-BLAST iterations
[6] Gene is located between *oppBCDF* and the putative *oppA*, located on the same strand, transcribe unidirectional and codes for a hypothetical protein
[7] Note the order of genes BCFD
[8] Three genes are located between *oppBCDF* and the putative *oppA*, both located on the same strand, transcribe unidirectional and code for hypothetical proteins
ND No *opp* operon was identified with in the genome

Supplementary Table 2 Annotated *opp* genes that are located within single or incomplete *opp* operon

Species	**Abbr**	**NCBI nr**	***opp* genes**	**Location**
M. agalactiae	MAGa	NC_013948.1	*oppB*	598047-598982
M. agalactiae PG2	MAG	NC_009497.1	*oppB*	548505-549440
M. arthritidis 158L3-1	MARTH	NC_011025.1	*oppB2C2*	656656-654525
M. bovis Hubei-1	MMB	NC_015725.1	*oppB3*	613224-614159
M. fermentans JER	MFE	NC_014552.1	*oppF pseudo* *oppB3C3*	341222-343764 722489-720386
M. fermentans M64	MfeM64YM	NC_014921.1	*oppF3* *oppB-3C2*	352166-354709 799513-797410
M. hominis ATCC 23114	MHO	NC_013511.1	*oppBC*	210656-212795
M. mobile 163K	MMOB	NC_006908.1	*oppF*	212201-214753
M. mycoides subsp. *mycoides* SC str. PG1	MSC	BX293980.2	*oppF* *oppF*	1112073-1113035 1121742-1122701
Total		*OppB*	3	
		OppBC	4	
		OppF	5	
			16	

Supplementary Table 3 PSORTb, PRED-LIPO, SignalP and InterPro results for mycoplasma OppA proteins

OppA Type A protein	Size	PSORT	PRED-LIPO	SignalP 4.1	InterPro						
					Prosite	Family	Domains	Region	Unintegrated signatures	Region	GO term
GUU_02828	832	Unknown	LIPO	Y	-				G3DSA:3.40.190.10	317-480	
									G3DSA:3.90.76.10	173-288	
									PD024071	609-648	
									SSF53850	172-270, 319-506, 555-658	
MAG1000	959	Unknown	LIPO	Y	-						
MAGa1090	959	Extracellular	LIPO	Y	-						
MALL_0660	949	Unknown	LIPO	Y	PS51257 (1-27)						
MARTH_orf403	918	Unknown	LIPO	Y	PS51257 (1-31)				G3DSA:3.90.76.10	174-224	
MBOVPG45_0112	961	Unknown	LIPO	Y	-						
MCJ_002370	955	Extracellular	Membrane	Y	PS51257 (1-25)						
MCRO_0618	940	Unknown	SIGNAL	Y	PS51257 (1-28) PS00387 (172-178)				G3DSA:3.90.76.10	213-269	
MfeM64YM_0281	929	Unknown	LIPO	Y	PS51257 (1-29)						
MFE_02290	928	Unknown	LIPO	Y	PS51257 (1-29)						
MGAH_0226	897	Cytoplasmic Membrane	LIPO	Y	PS51257 (1-28)						
MGA_0226	897	Cytoplasmic Membrane	LIPO	Y	PS51257 (1-28)						
MGF_2297	897	Cytoplasmic Membrane	LIPO	Y	PS51257 (1-28)						
MG_321	934	Extracellular	LIPO	Y	PS51257 (1-25)				G3DSA:3.90.76.10	162-356	
									PD024071	664-934	
									SSF53850	165-302, 335-373	
MHJ_0502	926	Unknown	SIGNAL	Y	-	IPR017012	PIRSF032899	1-926			
MHO_OppA	961	Unknown	SIGNAL	Y	PS51257 (1-28)	IPR017012	PIRSF032899	1-960	G3DSA:3.90.76.10	192-237	
MHP7448_0505	938	Cytoplasmic Membrane	LIPO	Y	-	IPR017012	PIRSF032899	1-926			
mhp502	938	Unknown	SIGNAL	Y	-	IPR017012	PIRSF032899	1-926			
MHR_0639	946	Unknown	LIPO	Y	PS51257 (1-24)						
MMB_OppA2	961	Unknown	LIPO	Y	-						
MMOB4520	955	Unknown	SIGNAL	Y	-	IPR017012	PIRSF032899	1-955			
MPN456	1005	Unknown	LIPO	Y	PS51257 (1-25)				G3DSA:3.90.76.10	167-224, 274-373	
									PD024071	698-1003	
									SSF53850	167-228, 268-326, 356-409	
MPNA4560	1005	Unknown	LIPO	Y	PS51257 (1-25)				G3DSA:3.90.76.10	167-224, 274-373	
									PD024071	698-1003	
									SSF53850	167-233, 279-319, 354-414	
Ms01_OppA	942	Unknown	LIPO	Y	PS51257 (1-27)						
Ms02_OppA	998	Unknown	LIPO	Y	PS51257 (1-28)				G3DSA:3.90.76.10	192-213, 245-292	
Ms03_OppA	1288	Unknown	LIPO	Y	PS51257 (1-23)						
MS53_0349	991	Unknown	LIPO	Y	PS51257 (1-26)						
MYPE5560	847	Unknown	SIGNAL	Y	-				PD024071	129-158	
MYPE7580/MYPE7570*	982	Extracellular	SIGNAL	Y	PS51257 (1-26)				PD024071	494-769	
MYPE7590	972	Extracellular	LIPO	Y	PS51257 (1-26)				PD024071	694-969	
MYPE7600	965	Extracellular	SIGNAL	Y	PS51257 (1-26)				PD024071	690-963	
MYPE7610	1010	Extracellular	LIPO	Y	PS51257 (1-35)				PD024071	720-1008	
MYPE7620	990	Extracellular	LIPO	Y	PS51257 (1-26)				PD024071	697-988	
MYPU_2820	876	Unknown	LIPO	Y	-	IPR017012	PIRSF032899	2-876			

OppA Type B Protein	Size	PSORT	PRED-LIPO	SignalP 4.1	InterPro						
					Prosite	Family	Domains	Region	Unintegrated signatures	Region	GO terms
MAGa_OppA	983	Unknown	LIPO	Y	-	IPR016880	PIRSF028335	1-982	G3DSA:3.10.105.10	576-640, 683-813, 894-938	GO:0006810 GO:0005215
							IPR000914	437-798	SSF53850	437-639, 681-804, 892-946	
MAG_OppA	982	Unknown	LIPO	Y	-	IPR016880	PIRSF028335	1-982	G3DSA:3.10.105.10	576-640, 683-813, 894-938	GO:0006810 GO:0005215
							IPR000914	437-798	SSF53850	437-651, 698-796	
MALL_0007	938	Unknown	LIPO	Y	PS51257 (1-25)	IPR016880	PIRSF028335	1-937	G3DSA:3.10.105.10	559-616, 657-787, 852-896	GO:0006810 GO:0005215
							IPR000914	464-772	G3DSA:3.90.76.10	136-204	
									SSF53850	443-615, 655-778, 854-906	
MALL_0020	991	Unknown	LIPO	Y	PS51257 (1-24)	IPR016880	PIRSF028335	1-991	G3DSA:3.10.105.10	510-567, 606-737, 905-949	GO:0006810 GO:0005215
							IPR000914	360-700	SSF53850	348-578, 621-729	
MBOVPG45_0037	983	Unknown	LIPO	Y	-	IPR016880	PIRSF028335	1-982	G3DSA:3.10.105.10	576-640, 683-813, 894-938	GO:0006810 GO:0005215
							IPR000914	437-798	SSF53850	437-639, 678-804, 892-947	
MCAP_0116	984	Unknown	SIGNAL	Y	-	IPR016880	PIRSF028335	1-984	G3DSA:3.10.105.10	583-647, 688-818, 896-940	GO:0006810 GO:0005215
							IPR000914	444-803	SSF53850	444-646, 683-809, 898-949	
MCJ_004440	888	Unknown	LIPO	Y	PS51257 (1-21)	IPR016880	PIRSF028335	1-888	G3DSA:3.10.105.10	482-539, 580-716, 580-716	GO:0006810 GO:000521
							IPR000914	146-677	G3DSA:3.90.76.10	147-235	
									SSF53850	144-197, 337-538, 582-727	
MCRO_OppA	950	Unknown	LIPO	Y	PS51257 (1-26)	IPR016880	PIRSF028335	1-949	G3DSA:3.10.105.10	571-628, 669-799, 864-908	GO:0006810 GO:0005215
							IPR000914	476-784	SSF53850	456-627, 667-829	
MfeM64YM_OppA	935	Unknown	LIPO	Y	PS51257 (1-26)	IPR016880	PIRSF028335	1-935	G3DSA:3.10.105.10	547-604, 644-774, 853-893	GO:0006810 GO:0005215
							IPR000914	426-757	SSF53850	405-603, 639-774	
MFE_OppA	935	Unknown	LIPO	Y	PS51257 (1-26)	IPR016880	PIRSF028335	1-935	G3DSA:3.10.105.10	547-604, 644-774, 853-893	GO:0006810 GO:0005215
							IPR000914	426-757	SSF53850	405-603, 639-774	
MGAH_OppA_2	1034	Cytoplasmic Membrane	LIPO	Y	PS51257 (1-26)	IPR016880	PIRSF028335	1-1034	G3DSA:3.10.105.10	524-581, 620-751, 948-992	GO:0006810 GO:0005215
							IPR000914	382-716	SSF53850	382-592, 635-779	
MGA_OppA	1034	Cytoplasmic Membrane	LIPO	Y	PS51257 (1-26)	IPRO16880	PIRSF028335	1-1034	G3DSA:3.10.105.10	524-581, 620-751, 948-992	GO:0006810 GO:0005215
							IPR000914	382-716	SSF53850	382-592, 635-779	
MGF_OppA_2	1034	Cytoplasmic Membrane	LIPO	Y	PS51257 (1-26)	IPR016880	PIRSF028335	1-1034	G3DSA:3.10.105.10	524-581, 620-751, 948-992	GO:0006810 GO:0005215
							IPR000914	382-716	SSF53850	382-592, 635-807	
MHJ_0213	889	Cytoplasmic Membrane	LIPO	Y	-	IPR016880	PIRSF028335	1-889	G3DSA:3.10.105.10	482-539, 580-715, 805-849	GO:0006810 GO:0005215
							IPR000914	159-678	SSF53850	361-538, 582-740	
MHP7448_0217	889	Unknown	LIPO	Y	-	IPR016880	PIRSF028335	1-889	G3DSA:3.10.105.10	482-539, 580-715, 805-849	GO:0006810 GO:0005215
							IPR000914	159-679	SSF53850	361-538, 582-738	
mhp164	892	Unknown	LIPO	Y	PS51257 (1-24)	IPR016880	PIRSF028335	1-892	G3DSA:3.10.105.10	485-542, 583-718, 808-852	GO:0006810 GO:0005215
							IPR000914	162-682	SSF53850	364-541, 585-740	
MHR_0357	901	Unknown	LIPO	Y	PS51257 (1-25)	IPR016880	PIRSF028335	1-901	G3DSA:3.10.105.10	494-551, 595-724, 816-860	GO:0006810 GO:0005215
							IPR000914	348-710	G3DSA:3.90.76.10	159-214, 340-370	
									SSF53850	154-207, 349-550, 593-735	
MLC_OppA	985	Unknown	SIGNAL	N	-	IPR016880	PIRSF028335	1-985	G3DSA:3.10.105.10	584-648, 689-819, 896-941	GO:0006810 GO:0005215
							IPR000914	446-804	SSF53850	445-647, 684-810	
MMB_OppA1	982	Unknown	LIPO	Y	-	IPR016880	PIRSF028335	1-982	G3DSA:3.10.105.10	576-640, 683-813, 894-938	GO:0006810 GO:0005215
							IPR000914	437-798	SSF53850	437-639, 678-804, 892-947	
Ms02_OppA2	895	Unknown	LIPO	Y	PS51257 (1-28)	IPR016880	PIRSF028335	1-895	G3DSA:3.10.105.10	512-569, 614-744, 809-853	GO:0006810 GO:0005215
							IPR000914	374-729	SSF53850	374-568, 612-735, 794-863	
Ms03_OppA2	1030	Unknown	LIPO	Y	PS51257 (1-27)	IPR016880	PIRSF028335	1-1030	G3DSA:3.10.105.10	521-578, 617-749, 783-794	GO:0006810 GO:0005215
							IPR000914	191-710	SSF53850	391-589, 632-776	
MS53_0190	894	Unknown	LIPO	Y	-		IPR000914	367-722	G3DSA:3.10.105.10	511-568, 611-741, 805-849	GO:0006810 GO:0005215
									G3DSA:3.90.76.10	158-212	
									SSF53850	371-567, 607-732, 805-858	

OppA Type B protein	Size	PSORT	PRED-LIPO	SignalP 4.1	InterPro						
					Prosite	Family	Domains	Region	Unintegrated signatures	Region	GO terms
MSB_A0161	985	Unknown	SIGNAL	N	-	IPR016880	PIRSF028335	1-985	G3DSA:3.10.105.10	584-648, 689-819, 896-941	GO:0006810 GO:0005215
							IPR000914	445-804	SSF53850	445-647, 684-810	
MSC_OppA	985	Unknown	SIGNAL	N	-	IPR016880	PIRSF028335	1-985	G3DSA:3.10.105.10	584-648, 689-819, 894-941	GO:0006810 GO:0005215
							IPR000914	445-804	SSF53850	445-647, 684-810	
MYPU_4150	904	Cytoplasmic Membrane	LIPO	Y	PS51257 (1-23)	IPR016880	PIRSF028335	1-903	G3DSA:3.10.105.10	494-551, 596-725, 811-859	GO:0006810 GO:0005215
							IPR000914	346-711	G3DSA:3.90.76.10	163-214, 339-371	
									SSF53850	171-207, 171-207, 611-751	

OppA Type C protein	Size	PSORT	PRED-LIPO	SignalP 4.1	InterPro						
					Prosite	Family	Domains	Region	Unintegrated signatures	Region	GO terms
MCAP_0165	1035	Unknown	LIPO	Y	PS51257 (1-24)		IPR000914	85-550	G3DSA:3.10.105.10	791-853	GO:0006810 GO:0005215
									G3DSA:3.40.190.10	278-430	
									G3DSA:3.90.76.10	76-161, 247-277	
									SSF53850	43-157, 243-430	
MLC_OppA	1044	Extracellular	LIPO	Y	PS51257 (1-24)		IPR000914	330-540	G3DSA:3.10.105.10	802-903	GO:0006810 GO:0005215
									Code: 3.40.190.10	288-440	
									G3DSA:3.90.76.10	76-158, 253-287	
									SSF53850	43-158, 252-438	
MPUT_OppA	966	Unknown	LIPO	Y	PS51257 (1-24) PS01040 (90-112)		IPR000914	85-435	G3DSA:3.10.105.10	757-876	GO:0006810 GO:0005215
									G3DSA:3.40.190.10	271-435	
									G3DSA:3.90.76.10	74-161, 243-270	
									SSF53850	47-156, 236-435	
MSB_A0216	1031	Extracellular	LIPO	Y	PS51257 (1-24)		IPR000914	325-497	G3DSA:3.10.105.10	798-888	GO:0006810 GO:0005215
									G3DSA:3.40.190.10	287-443	
									G3DSA:3.90.76.10	76-161, 252-286	
									SSF53850	42-148, 245-446	
MSC_OppA	1047	Extracellular	LIPO	Y	PS51257 (1-24)		IPR000914	330-540	G3DSA:3.10.105.10	805-859	GO:0006810 GO:0005215
									G3DSA:3.40.190.10	288-440	
									G3DSA:3.90.76.10	76-161, 252-287	
									SSF53850	43-158, 252-438	

* Open reading frame MYPE7570 and MYPE7580 represent the complete copy of one OppA protein

G3DSA:3.10.105.10: CATH Classification, Class 3: Alpha Beta, Architecture 3.10: Roll, Topology 3.10.105: Dipeptide-binding Protein; domain 3, Homologous Superfamily 3.10.105.10: Dipeptide-binding Protein; domain 3

G3DSA:3.40.190.10: CATH Classification, Class 3: Alpha Beta, Architecture 3.40: 3-Layer(aba) Sandwich, Topology 3.40.190: D-Maltodextrin-Binding Protein; domain 2, Homologous Superfamily 3.40.190.10: Periplasmic binding protein-like II

G3DSA:3.90.76.10: CATH Classification, Class 3: Alpha Beta, Architecture 3.90: Alpha-Beta Complex, Topology 3.90.76: Dipeptide-binding Protein; domain 1, Homologous Superfamily 3.90.76.10: Dipeptide-binding Protein; domain 1

GO:0006810 transport: Biological Process

GO:0005215 transporter activity: Molecular Function

IPR000914: Pfam PF00496, Bacterial extracellular solute-binding proteins, family 5

IPR016880: InterPro, ABC-type oligopeptide transport system, solute-binding component, *Mycoplasmataceae*, predicted

IPR017012: InterPro family, Uncharacterised conserved protein UCP032899 (PIRSF032899), lipoprotein

PD024071: ProDom family, lipoprotein related to MG321

PIRSF028335: Protein information Resource, ABC-type oligopeptide transport system, solute-binding component, *Mycoplasmataceae* type

PIRSF032899: Protein Information Resource, uncharacterized conserved lipoprotein

PS00387: Prosite, PPASE, Inorganic pyrophosphatase signature

PS01040: Prosite, SBP_BACTERIAL_5 Bacterial extracellular solute-binding proteins, family 5 signature

PS51257: Prosite, PROKAR_LIPOPROTEIN, Prokaryotic membrane lipoprotein lipid attachment site

SSF53850: SCOP Classification, Class: Alpha and beta proteins (a/b) [51349], Fold: Periplasmic binding protein-like II [53849], Superfamily: Periplasmic binding protein-like II [53850]

Supplementary Table 4 MEME motifs in Type A, B and C OppA proteins

Motif	**Motif 1A**	**Motif 2A**	**Motif 3A**	**Motif 4A**	**Motif 5A**	**Motif 6A**
Size of motif	29	41	41	29	32	41
E-value	1.0e-280	3.3e-339	8.2e-157	2.0e-155	5.3e-151	1.2e-193
GUU_02828	624-652	162-202	247-287	-	447-478	203-243
MAG1000	762-790	205-245	-	678-706	-	-
MAGa1090	762-790	205-245	-	678-706	-	-
MALL_0660	754-782	209-249	-	672-700	-	-
MARTH_orf403	711-739	161-201	-	628-656	-	203-243
MBOVPG45_0112	762-790	205-245	-	678-706	-	-
MCJ_002370	702-730	162-202	-	620-648	-	204-244
MCRO_0618	746-774	203-243	-	665-693	555-586	-
MfeM64YM_0281	730-758	174-214	859-899	645-673	-	392-432
MFE_02290	730-758	174-214	859-899	645-673	-	392-432
MGAH_0226	134-162	458-498	570-610	-	796-827	499-539
MGA_0226	134-162	458-498	570-610	-	796-827	499-539
MGF_2297	134-162	458-498	570-610	-	796-827	499-539
MG_321	687-715	154-194	271-311	-	473-504	201-241
MHJ_0502	698-726	159-199	-	616-644	516-547	201-241
MHO_OppA	736-764	179-219	-	651-679	-	-
MHP7448_0505	698-726	159-199	-	616-644	516-547	201-241
mhp502	698-726	159-199	-	616-644	516-547	201-241
MHR_0639	720-748	167-207	-	637-665	-	-
MMB_OppA2	762-790	205-245	-	678-706	-	-
MMOB4520	754-782	173-213	-	673-701	-	215-255
MPN456	721-749	159-199	288-328	-	490-521	207-247
MPNA4560	721-749	159-199	288-328	-	490-521	207-247
Ms01_OppA	730-758	174-214	-	645-673	-	-
Ms02_OppA	794-822	232-272	-	711-739	-	-
Ms03_OppA	1073-1101	534-574	-	994-1022	-	-
MS53_0349	752-780	184-224	-	673-701	-	267-307
MYPE5560	134-162	419-459	529-569	-	741-772	460-500
MYPE7570/MYPE7580*	499-527	6-46	118-158	-	327-358	47-87
MYPE7590	699-727	208-248	317-357	-	527-588	329-289
MYPE7600	695-723	207-247	315-355	-	524-555	248-288
MYPE7610	738-766	232-272	349-389	-	557-588	273-313
MYPE7620	717-745	221-261	333-373	-	543-574	262-302
MYPU_2820	692-720	165-205	-	609-637	502-533	-
Motif	**Motif 1B**	**Motif 2B**	**Motif 3B**	**Motif 4B**	**Motif 5B**	**Motif 6B**
Size	85	49	43	62	55	56
E-value	2.2e-1119	2.1e-710	3.0e-640	5.3e-723	1.2e-613	1.2e-623
MAGa_OppA	503-587	607-655	909-951	664-725	148-202	727-782
MAG_OppA	503-587	607-655	909-951	664-725	148-202	727-782
MALL_0007	486-570	583-631	867-909	638-699	141-195	701-756
MALL_0020	437-521	534-582	920-962	587-648	163-217	650-705
MBOVPG45_0037	503-587	607-655	909-951	664-725	148-202	727-782
MCAP_0116	510-594	614-662	911-953	669-730	152-206	732-787
MCJ_004440	409-493	506-554	820-862	565-626	140-194	627-682
MCRO_OppA	498-582	595-643	879-921	650-711	153-207	713-768
MfeM64YM_OppA	474-558	571-619	864-906	625-686	141-195	688-743
MFE_OppA	474-558	571-619	864-906	625-686	141-195	688-743
MGAH_OppA_2	451-535	548-596	963-1005	601-662	172-226	664-719
MGA_OppA	451-535	548-596	963-1005	601-662	172-226	664-719
MGF_OppA_2	451-535	548-596	963-1005	601-662	172-226	664-719
MHJ_0213	409-493	506-554	820-862	565-626	139-193	627-682

Motif	**Motif 1B**	**Motif 2B**	**Motif 3B**	**Motif 4B**	**Motif 5B**	**Motif 6B**
MHP7448_0217	409-493	506-554	820-862	565-626	139-193	627-682
mhp164	412-496	509-557	823-865	568-629	142-196	630-685
MHR_0357	421-505	518-566	834-874	576-637	151-205	638-693
MLC_OppA	511-595	615-663	912-954	670-731	153-207	735-788
MMB_OppA1	503-587	607-655	909-951	664-725	148-202	727-782
Ms02_OppA2	439-523	536-584	824-866	595-656	144-198	658-713
Ms03_OppA2	448-532	545-693	959-1001	592-653	171-225	661-716
MS53_0190	438-522	535-583	820-862	592-653	149-203	655-710
MSB_A0161	511-599	615-663	912-954	670-731	153-207	733-788
MSC_OppA	511-595	615-663	912-954	670-731	153-207	733-788
MYPU_4150	421-505	518-566	830-872	577-638	151-205	639-694
Motif	**Motif 1C**	**Motif 2C**	**Motif 3C**	**Motif 4C**	**Motif 5C**	**Motif 6C**
Size	100	100	100	74	100	57
E-value	1.0e-184	2.8e-181	8.0e-160	5.7e-111	4.8e-119	4.7e-092
MCAP_0165	72-171	275-374	745-844	884-957	453-552	1-57
MLC_OppA	72-171	284-383	756-855	894-967	463-562	1-57
MPUT_OppA	72-171	271-370	707-806	826-899	449-548	1-57
MSB_A0216	72-171	283-382	751-850	880-953	461-560	1-57
MSC_OppA	72-171	271-370	707-806	826-899	449-548	1-57
* Open reading frame MYPE7570 and MYPE7580 represent the complete copy of one OppA protein						

Supplementary Table 5 Regular expressions of the MEME motifs within Type A, B and C OppA protein

Type A OppA	
Motif	**Regular expression**
Motif 1A	[YW][FIY]xxG[TS]SP[LA]GFS[GS]W[SG][YP]DY[ND]G[IV]G[TS][GW][LIY][DE][GA]L[TIS]
Motif 2A	SINS[EK]xF[KQ][EKD]A[LI]KKA[KT]K[LYV]QFx[IV][DRK][KT][GDN][IV]KWV[DN][SN][KN]G[NEK]PTKYE[VS][VG][APK]
Motif 3A	F[TNR][FILV]Y[LIV][TK][KQS][PA][YF][PA][FY][TAL][FLM][GQS][MT][LFM][STK][KI][ES][YF][FL][AFT][AP][MLI]P[HY][TW][NDH][PQT][KER][VI][KR][AN][IL][HKR][LI][GQS][SGK][GDQ][STAG][PD]
Motif 4A	[KQR]FK[SA][AP][KN]FKE[LI][KQ][AE]E[MV]KK[IL]LD[KE][FY]Y[QAK]DN[PKN]LG[PA]
Motif 5A	[MET][AR][KN][FI][LIF][YR][NQ][WS][NE]S[DEK][DE][SA][YRL][TI]IRA[GA]I[ANV][GN]LINWY[NQR]L[AS][LIQ][IV]
Motif 6A	D[FY]ER[GA][FIL]E[SI]Y[WY]L[AS][AS]SLGYNRNGYF[LI]DL[IL]GLD[FL]EKT[VA][GN]Y[TD]P[KS][SN]
Type B OppA	
Motif	**Regular expression**
Motif 1B	LGPQG[YS][LMI][LI]LSKN[PL][SN]Y[YF]S[AS][DE][KN]TI[SP][NE][KR]I[KR]I[FY]F[SA][SQ][DN]P[NE][ILT][NL][AS][AL][LM][FY][DE]D[GK]YI[AS][AQ]T[RK]IPAIQQ[LN][AKR][YF]W[TAS][ND][KP]E[YT][RK][KQ][YF][ML][KN]K[SN][SQT]G[FY]GTI[AG][LF][AQG][FL]NLD
Motif 2B	DLRNAIYY[AG]INR[DE][DE][ML]L[NK][IL]VG[WL][ND][SF]S[FY]PV[IT]TWTAFGQG[SK][ST]S[FDR]GD[ANP][LI]E[LIT][FG]FD
Motif 3B	[KAE][EQ][GN]WT[EQ][QN][SK][VA]F[AG][IF]I[AG][AG][LF]EKI[IV]R[DE]AAP[VI][VI]PLMEVDT[YN]WEI[SNT]R[VI][GNR]G
Motif 4B	[KD][TVD][GDK][KN][KN][KT][EAP][IF]P[VLI][QL]NY[SDN][HYF][IV][DNV]HL[SA]KS[YF][NK]FE[HAK][VT]DR[TK]D[KL][GA][FY][DN][LP][KE][IVT]A[NKR][FKQ]Y[LM]D[RL]FK[AK]K[HY]P[ND][LV]K[KQS][VI]TL[KN][YF][IL]
Motif 5B	[ILN][QP][NT][NS][NGS]N[KN][IFV][QLMV][SA][ML][TN][IG]LLNDG[KA][SN][RK]WSN[GN]DEV[VT]A[DQ]D[YF]ID[AY][IL][HL]YILD[LI][NS]T[GA]SQ[KR][LQ][TDV][NT][IL]LQ
Motif 6B	[NY]STDE[QH]Q[NK][AV][GA][IT][AG]L[QK]D[FA][ML][RS]KA[FT][NG]G[FY][ILV]NI[ED]IK[SG]LPE[NG][VI]YE[DS][FAR][RI][TE]KG[EQ][FY]D[LI][IL]Y[RQK]NFD[AT]F
Type C OppA	
Motif	**Regular expression**
Motif 1C	LQD[VIT]L[LI][TA]V[ND][RKN]HD[HN][YF]EGALA[EL][YK]W[DK][HA][ND][KST][DN][KFS][DK][YHT]W[KS]F[RK][LI]RKN[AI][YK]WT[KAR]IENGK[QA]V[KD][GK][DGP][LAQ]IT[GA][LKQ]D[IFLV]FNTFRYV[LF]NKNN[LR]ALT[TL][ED][HI][FW][LS][TS][NAK][FL][KA][HN][VA][PHN][QEK]L[MI][DEN]F[IL][ND]KLSDP
Motif 2C	YFE[ST][VI]ISYL[AS]FAP[IM]P[ED]I[AS][LV][FQL]Y[AV][NKQS]D[KS][DG]Q[VEGK][SY][NS][IK][YF]AGT[NAL]Y[GA]KP[LS][GAE][KS][KQ]SGY[NE][GT][LM]WYSG[PA]Y[VI][IV][EDQ][DE]Y[FV][PS]G[SR]NLNL[TK][KR]NE[FH]Y[YF]NK[EDQ][NK]V[YH]I[EH]K[IM][NL][YF]SY[VT]NK[AG]D[AP][AS][TS][RS]R
Motif 3C	[TI][IN]P[FG][LN][LS][NDS][PT][TG]G[AS][DS][DG][FY]K[IN][KY][IVL][AIQ][QR][FAL]F[GK][SAT]FN[YF]LVR[KN][HKF][DGN][NSG][NGS]D[IM][DN]SP[FI]V[FI]D[IP][DY][KE]P[IG][DT][FAQ][SK][NAE]Y[LG][KE][EDLQ][LIRV]R[ADE][GS]K[YFS][GA][LI][AGM][AE][FI]GW[SA]PDY[DA]DPTNYL[AY]T[LV][KL]Y[GD]G[VA][YF][ED][HY]I[QM][SGN][WM][TDK]K[VL]F
Motif 4C	[AH][YF][KAES][DEQ]LK[NDK][IA][LIT][QE][HF][FL][TS][NK]E[LV][TE][YEN][IV]D[EK]N[EK][ASTV][DE][IK][YK][KDE]RYT[EQR]LA[KQ]LEN[YF][YL][TN]L[SA]SA[IL]IIPT[HY][TV][HR][LQE][AS][DE][TF]LP[SIT][IV]S[YF][VL]D[EQ][FL][ST][KI][PS][TR][WF]P[TIS]
Motif 5C	R[TPV]RT[PA][EQ]EDSILNRA[LI]ALKS[LV]RI[LM][AT]R[YF][AV]LNRSLYAKF[YFH]SEA[RK]DG[NV][DN][RH]P[TV]S[TS]QLRNTFTS[KQ]Y[VI]ST[FY][EN]D[DK][EKQS][HK][QKR]V[LIV]D[KE][DNS][SL][KGT][QDE][KT]VADYADFLAK[DN]YYDI[RT]KYDD[ND]
Motif 6C	MKK[VI]LG[ML]T[LT][LS][GT][SL][IL]IA[TS][AS][AS][AM][ST][AI]VSCS[VL][GR][IV][SGN][LP][DY]K[IL]L[NS]R[KR][NI]S[ND]T[KT][VI][LY][RK][ED]L[TI][NT][YQ][SP][LI][AV][NT][LW]NS

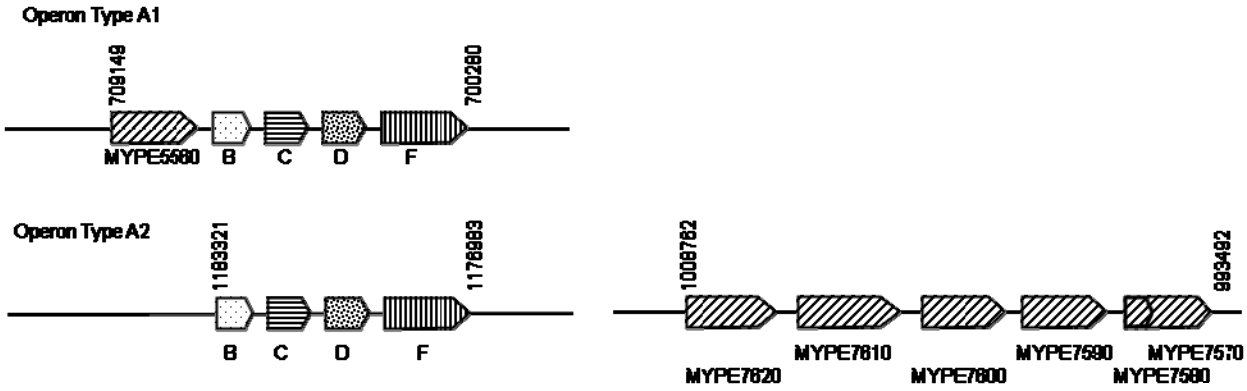
Supplementary Table 6 A comparison of the conserved functional motifs in *M. hominis* with the postulated motifs in *M. struthionis* sp. nov. str. Ms01, *Mycoplasma* sp. Ms02 and *M. nasistruthionis* sp. nov. str. Ms03

Sequence of the conserved motifs				
	Signal peptidase II recognition site	**Oligopeptide binding site**	**Walker B**	**Walker A**
Consensus motif[1, 2]	(+4)-VAASC-(-1)	(F/Y)X(L/I)RKGVK(W/F)	RXXXGXXXLZZZZD	GXXXXGKS/T
Mho_OppA	-LVAASC-28	197-FVIRKGVKW-205	737-RFDGVTGENLLAWSAD-752	869-GKDSSGKS-876
Ms01_OppA	-LVAAAC-28	192-YVLKDNLKW-200	732-INQGANGTRNVGWSYD-747	752-GSGYDGLS-759
Ms02_OppA	-SVAVSC-27	250-FRIRPEQVW-258	795-YWYSVSPARRGRWNYD-810	815-GTGRDGYS-822
Ms03_OppA	-ALALSC-28	552-FRVRPGHFW-560	1074-YWTGTSPFSLAGWGYD-1089	1094-GSGIDGYS-1101
Percentage identity between the postulated and consensus motifs				
	Signal peptidase II recognition site	**Oligopeptide binding site**	**Walker B**	**Walker A**
Mho_OppA	100%	100%	75%	100%
Ms01_OppA	80%	50%	50%	75%
Ms02_OppA	80%	50%	37.5%	75%
Ms03_OppA	60%	50%	50%	75%
Percentage identity and similarity of the Ms01, Ms02 and Ms03 motifs to that of *M. hominis*[3]				
	Signal peptidase II recognition site	**Oligopeptide binding site**	**Walker B**	**Walker A**
Ms01_OppA	Identity: 5/6 (83.3%) Similarity: 6/6 (100.0%)	Identity: 3/9 (33.3%) Similarity: 7/9 (77.8%)	Identity: 5/16 (31.2%) Similarity: 6/16 (37.5%)	Identity: 3/8 (37.5%) Similarity: 3/8 (37.5%)
Ms02_OppA	Identity: 4/6 (66.7%) Similarity: 4/6 (66.7%)	Identity: 4/9 (44.4%) Similarity: 4/9 (44.4%)	Identity: 3/16 (18.8%) Similarity: 6/16 (37.5%)	Identity: 3/8 (37.5%) Similarity: 3/8 (37.5%)
Ms03_OppA	Identity: 3/6 (50.0%) Similarity: 4/6 (66.7%)	Identity: 4/9 (44.4%) Similarity: 5/9 (55.6%)	Identity: 4/16 (25.0%) Similarity: 7/16 (43.8%)	Identity: 3/8 (37.5%) Similarity: 3/8 (37.5%)

[1] X represents any amino acids

[2] Z represents all hydrophobic amino acids

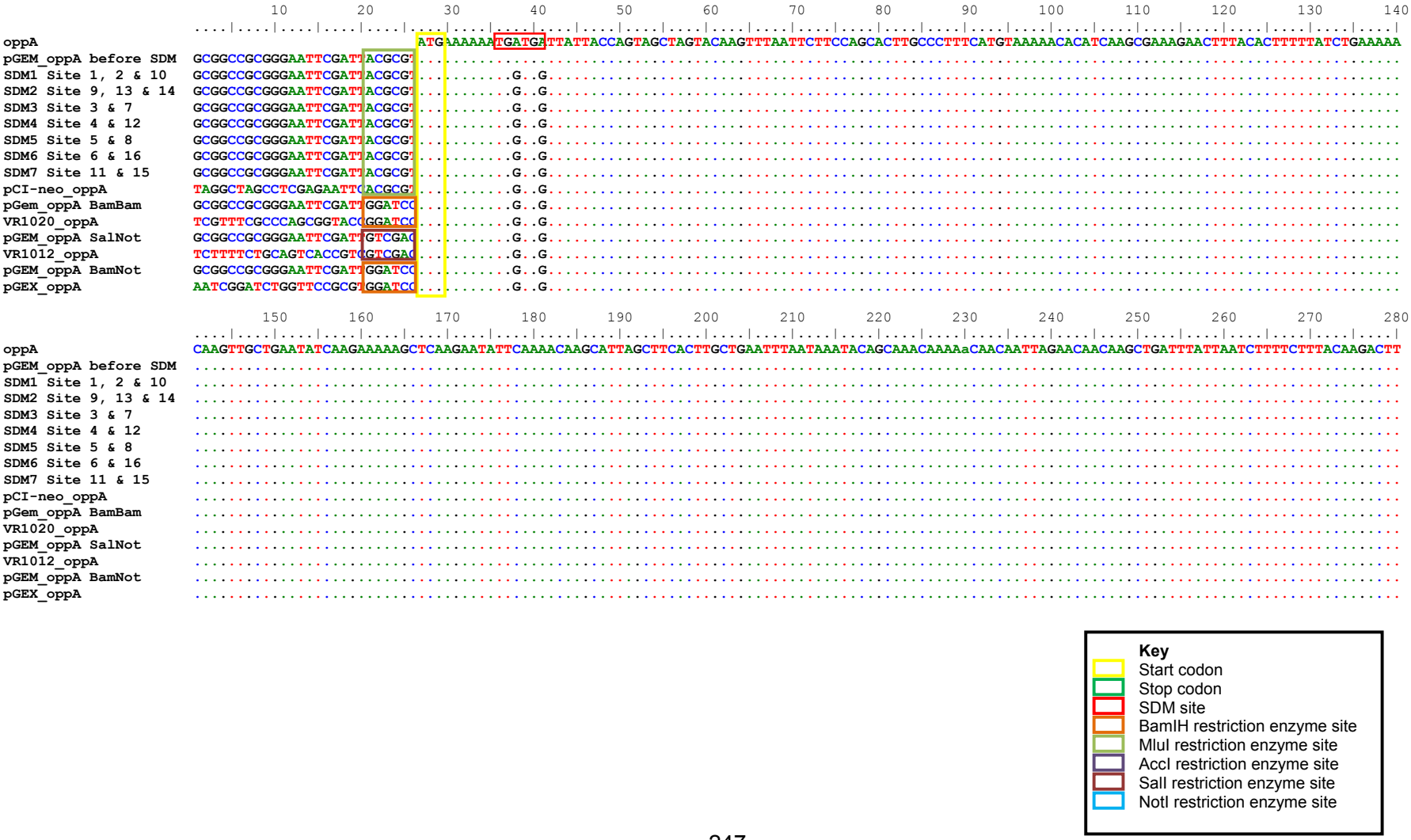
[3] Similarity and identity calculated with EMBOSS Stretcher Pairwise alignment (http://www.ebi.ac.uk/Tools/psa/emboss_stretcher/) using the BLOSUM62 matrix with a gap penalty of 10 and extension penalty of 0.5

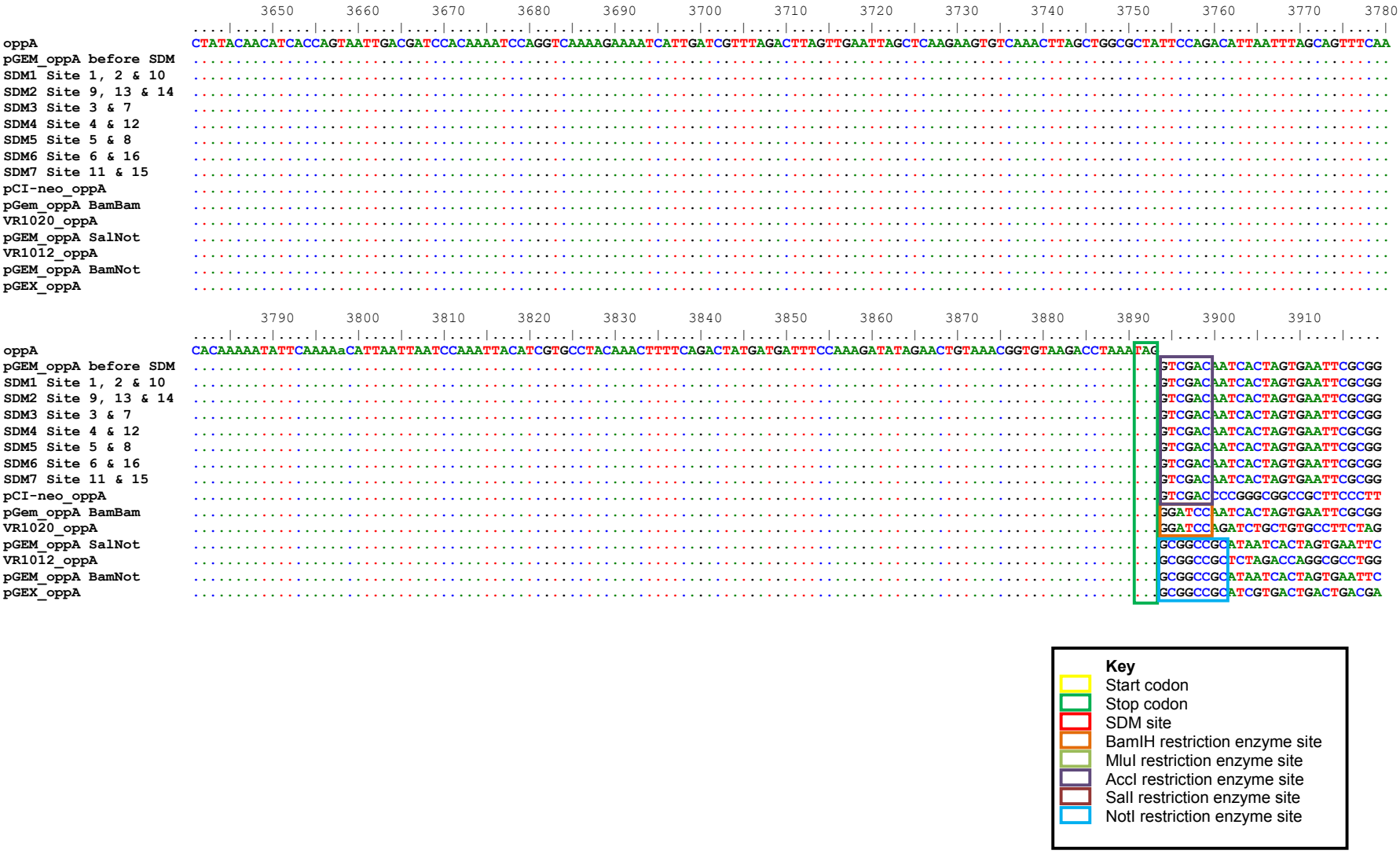


Supplementary Figure 1 The opp operons of *M. penetrans* HF-2. Type A1 has the oppA gene next to the OppBCDF genes in a polycistronic unit while the Type A2 opp operon has 5 copies of oppA encoded by six open reading frames, MYPE7570-MYPE7620 that is located several hundreds of bases from the opp operon. MYPE7570 and MYPE7580 represent a single copy of oppA.

Appendix 4 Sequencing data of Chapter 5

Alignments of the *Mycoplasma nasistruthionis* sp. nov. str. Ms03 Type A *oppA* gene was done in BioEdit, as obtained by sequencing after each round of SDM modification and after each clone experiment.





Appendix 5 Supplementary tables and figures for Chapter 5

Supplementary Table 5.1 The weight, mycoplasma infections and ELISA data of each ostrich in the vaccine trial

Vaccine group	Ostrich number	Weight		Ostrich mycoplasma infections						Average A_{260}	
				Week 0			Week 3			Week 0	Week 3
		28/3/11	18/4/11	Ms01	Ms02	Ms03	Ms01	Ms02	Ms03	Mean	Mean
pCI-neo_*oppA*	**7689**	42.4	42.4							0.556	0.436
pCI-neo_*oppA*	**7690**	34.8	37.4							0.707	1.617
pCI-neo_*oppA*	**7692**	37.8	40.4							0.511	0.950
pCI-neo_*oppA*	**7695**	42.4	42.6							0.938	1.298
pCI-neo_*oppA*	**7696**	48.6	50.0					+		0.632	2.697
pCI-neo_*oppA*	**7697**	41.2	47.0							0.928	0.865
pCI-neo_*oppA*	**7698**	47.6	38.4			+			+	0.842	0.916
pCI-neo_*oppA*	**7699**	37.4	39.2					+		0.687	1.999
pCI-neo_*oppA*	**7701**	36.6	33.8						+	1.028	1.092
pCI-neo_*oppA*	**7702**	39.4	42.2						+	0.706	0.462
pCI-neo_*oppA*	**7703**	38.8	42.0					+		1.574	1.646
pCI-neo_*oppA*	**7704**	36.0	38.6					+	+	1.372	1.689
pCI-neo_*oppA*	**7705**	34.4	38.0						+	1.441	2.639
pCI-neo_*oppA*	**7706**	31.6	29.2					+	+	1.493	1.120
pCI-neo_*oppA*	**7707**	42.4	45.8							0.557	0.518
pCI-neo_*oppA*	**7708**	34.0	40.4						+	0.616	1.706
pCI-neo_*oppA*	**7709**	47.6	52.0						+	0.441	0.728
pCI-neo_*oppA*	**7710**	41.6	42.4		+				+	0.886	0.960
pCI-neo_*oppA*	**7711**	38.0	-					+	+	0.870	1.876
pCI-neo_*oppA*	**7712**	38.0	39.8							1.215	1.888
pCI-neo_*oppA*	**7713**	39.8	25.2						+	1.858	2.942
pCI-neo_*oppA*	**7714**	33.0	38.4						+	1.295	1.426
pCI-neo_*oppA*	**7715**	38.0	39.8							1.231	2.003
pCI-neo_*oppA*	**7717**	35.8	36.4					+	+	0.288	1.643
pCI-neo_*oppA*	**7718**	36.0	35.4							0.779	1.258
Mean weight		**38.9**	**39.9**							**0.938**	**1.455**
Total Ms01/02/03 Infected ostriches				**0**	**1**	**1**	**0**	**7**	**13**		
Total infected ostriches per week					**2**			**16**			
Total infected ostriches						**16**					
VR1020_*oppA*	**7720**	29.4	32.4							0.848	1.476
VR1020_*oppA*	**7721**	38.0	39.6							0.895	1.908
VR1020_*oppA*	**7722**	33.8	35.8					+		0.675	2.779
VR1020_*oppA*	**7724**	34.6	39.4							0.677	0.975
VR1020_*oppA*	**7725**	30.6	37.2							1.441	0.977
VR1020_*oppA*	**7727**	38.2	41.4							0.504	0.793
VR1020_*oppA*	**7728**	33.0	32.6						+	0.881	0.857
VR1020_*oppA*	**7729**	32.0	29.0						+	0.668	0.492
VR1020_*oppA*	**7730**	40.2	42.8					+		1.481	2.696
VR1020_*oppA*	**7731**	33.2	36.8							1.936	2.622
VR1020_*oppA*	**7732**	32.6	37.8							1.722	1.369
VR1020_*oppA*	**7733**	32.4	27.0							1.228	0.285
VR1020_*oppA*	**7736**	31.8	38.6							0.834	2.312
VR1020_*oppA*	**7739**	32.4	38.4					+	+	0.446	0.934
VR1020_*oppA*	**7740**	31.4	39.2							1.222	2.260
VR1020_*oppA*	**7741**	35.2	37.8			+				1.841	3.029
VR1020_*oppA*	**7742**	36.6	42.0							1.164	1.870
VR1020_*oppA*	**7744**	33.4	34.6							1.397	2.726
VR1020_*oppA*	**7745**	29.2	36.4		+			+		1.797	2.775
VR1020_*oppA*	**7746**	30.8	35.8							0.808	1.016
VR1020_*oppA*	**7748**	36.4	37.8							1.515	1.503
Mean weight		**33.6**	**36.8**							**1.142**	**1.698**
Total Ms01/02/03 Infected ostriches				**0**	**1**	**1**	**0**	**4**	**3**		
Total infected ostriches per week					**2**			**6**			
Total infected ostriches						**7**					

Vaccine group	Ostrich number	Weight		Ostrich mycoplasma infections						Average A_{260}	
		Week 0	Week 3	Week 0			Week 3			Week 0	Week 3
		28/3/11	18/4/11	Ms01	Ms02	Ms03	Ms01	Ms02	Ms03	28/3/11	18/4/11
VR1012_*oppA*	**7750**	33.6	32.6		+	+				1.173	2.375
VR1012_*oppA*	**7752**	38.0	30.8		+	+				1.031	0.729
VR1012_*oppA*	**7753**	36.8	44.4			+		+		0.954	1.724
VR1012_*oppA*	**7754**	39.6	39.0		+					2.702	3.003
VR1012_*oppA*	**7755**	41.6	46.8		+	+		+		1.524	1.129
VR1012_*oppA*	**7756**	45.6	49.6		+		+			2.923	1.784
VR1012_*oppA*	**7757**	44.4	45.6							0.762	0.642
VR1012_*oppA*	**7759**	42.8	43.0		+					0.824	0.720
VR1012_*oppA*	**7760**	43.4	31.6		+			+		0.406	0.586
VR1012_*oppA*	**7761**	44.4	29.6		+	+	+	+	+	0.807	0.807
VR1012_*oppA*	**7762**	36.2	35.4		+	+		+	+	0.679	1.044
VR1012_*oppA*	**7763**	32.0	33.8		+	+	+	+	+	0.707	0.986
VR1012_*oppA*	**7765**	31.0	35.8		+	+				0.526	0.842
VR1012_*oppA*	**7767**	27.8	30.6		+	+				0.689	1.959
VR1012_*oppA*	**7768**	26.8	27.6							1.576	1.242
VR1012_*oppA*	**7769**	31.6	44.4		+	+		+	+	0.497	0.682
VR1012_*oppA*	**7771**	30.0	25.4			+				1.556	1.342
VR1012_*oppA*	**7773**	27.5	24.4		+	+		+		0.001	1.096
VR1012_*oppA*	**7775**	28.6	22.2		+	+				0.950	0.767
VR1012_*oppA*	**7776**	28.5	32.4		+			+		0.867	0.328
VR1012_*oppA*	**7777**	27.6	28.8							0.475	0.937
Mean weight		**35.1**	**34.9**							**1.030**	**1.177**
Total Ms01/02/03 Infected ostriches				**0**	**16**	**13**	**3**	**9**	**4**		
Total infected ostriches per week					**18**			**10**			
Total infected ostriches						**18**					
Control	**7779**	28.0	34.6			+			+	0.816	1.374
Control	**7780**	31.6	32.6		+	+				1.010	0.852
Control	**7781**	27.5	30.0		+	+		+		0.410	1.102
Control	**7782**	30.5	26.8							0.251	2.044
Control	**7783**	29.6	32.6			+				0.510	0.291
Control	**7784**	32.6	34.8		+	+				0.806	0.611
Control	**7785**	28.6	29.0		+	+		+		0.643	0.787
Control	**7787**	31.2	27.8		+	+		+		0.311	0.511
Control	**7788**	34.2	36.8		+	+		+		0.342	0.618
Control	**7789**	21.4	22.8		+	+				0.854	0.632
Control	**7790**	30.2	18.4		+	+				0.215	0.536
Control	**7791**	27.4	26.2			+				1.414	1.466
Control	**7792**	29.2	31.0							2.276	0.638
Control	**7795**	25.8	24.6			+		+		1.231	1.797
Control	**7797**	26.0	27.4			+				1.965	1.967
Control	**7798**	27.2	30.0		+					0.912	0.637
Control	**7799**	30.4	34.4		+	+				1.759	1.662
Control	**7800**	23.6	24.2		+	+				1.750	0.684
Control	**7802**	27.6	26.0			+				1.434	1.018
Control	**7803**	25.6	26.0		+	+				1.646	0.965
Control	**7805**	28.6	29.6		+	+				1.057	0.999
Control	**7806**	-	24.0		+	+				1.464	0.461
Control	**7807**	-	28.2		+					0.908	0.384
Mean weight		**28.4**	**28.6**							**1.043**	**0.958**
Total Ms01/02/03 Infected ostriches				**0**	**15**	**19**	**0**	**5**	**1**		
Total infected ostriches per week					**21**			**6**			
Total infected ostriches						**21**					

Supplementary Table 5.2 ANOVA analysis of the weight data

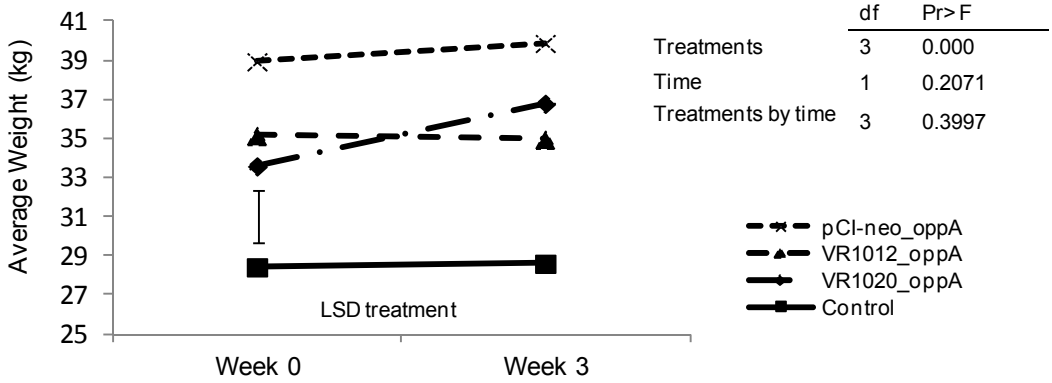
Source	df	SS	MS	F-value	Pr> F
Treatment	3	3241.832	1080.611	24.72	0.0000
Time	1	70.104	70.104	1.60	0.2071
Treatment by time	3	129.612	43.204	0.99	0.3997
Residual	172	7519.021	43.715		
Total	179	10960.569			

Df Degrees of freedom
SS Sum of squares
MS Mean of SS

Supplementary Table 5.3 ANOVA analysis of the ELISA data

Source	df	SS	MS	F-value	Pr> F
Treatments	3	4.166	1.389	3.59	0.0149
Time	1	3.611	3.611	9.33	0.0026
Treatments by time	3	3.224	1.075	2.78	0.0428
Residual	172	66.545	0.387		
Total	179	77.547			

Df Degrees of freedom
SS Sum of squares
MS Mean of SS



Supplementary Figure 5.1 Average weight of the ostriches vaccinated with three different DNA vaccines containing the mutated oppA gene of *M. nasistruthionis* sp. nov. strain Ms03. Vaccinated ostriches received a single DNA vaccine dose (100 µg/ml) at week 0 and the control group did not receive any vaccine. Statistical parameters are indicated.