

***Ehrlichia ruminantium*: Genome assembly and
analysis with the identification and testing of
vaccine candidate genes**

by

JUNITA LIEBENBERG

Submitted in partial fulfilment of the requirements for the degree Doctor of Philosophy

in the Department of Veterinary Tropical Diseases, Faculty of Veterinary Science,

University of Pretoria

September 2010

ACKNOWLEDGEMENTS

I would like to express my gratitude to my promoters Prof. Basil Allsopp and Dr. Nicola Collins for the opportunity to obtain a Ph.D. at the University of Pretoria. Special thanks to Prof. Allsopp for his mentorship and for the time he spent in correcting my writing. My thanks also go to Dr. Collins for her valuable advice, scientific contribution and motivation.

I would also like to extend my appreciation to my colleagues at the ARC-OVI for their contributions in the animal trials and immunological assays. A special word of thanks Dr. Mirinda van Kleef and Dr. Alri Pretorius for their assistance with the lymphocyte proliferation and ELISpot assays, and their expertise in cellular immunology. I would also like to thank Dr. Erich Zwegarth and Antoinette Josemans for providing the *E. ruminantium* cell cultures and Helena Steyn for her assistance with the immunisation of animals.

The work presented in this thesis was supported by the Department of Science and Technology of South Africa (LEAD 37/2001 (87)), the European Union (FP6-003713), the National Research Foundation of South Africa (FA2004042200063) and the Agricultural Research Council of South Africa.

Finally special thanks to Frans for his understanding and support.

TABLE OF CONTENTS

LIST OF FIGURES	v
LIST OF TABLES	vii
LIST OF ABBREVIATIONS	viii
SUMMARY	xi
CHAPTER 1: LITERATURE REVIEW	1
1.1. Heartwater	1
1.1.1. History of heartwater research	1
1.1.2. The organism	2
1.1.3. The vector	3
1.1.4. The disease	4
1.1.5. Immune responses to <i>E. ruminantium</i> infection	5
1.1.6. Heartwater vaccine development	6
1.1.6.1. Attenuated heartwater vaccines	6
1.1.6.2. Inactivated heartwater vaccines	7
1.1.6.3. DNA vaccines	7
1.2. Genome sequencing	8
1.2.1. DNA sequencing	8
1.2.1.1. Novel sequencing technologies	9
1.2.1.1.1. <i>In vitro</i> amplification sequencing technologies	10
1.2.1.1.2. Single-molecule sequencing technologies	11
1.2.1.1.3. Limitations and advantages of the latest technologies	12
1.2.2. Identification of novel vaccine candidate genes from whole genome sequence data	13
1.2.2.1. Reverse vaccinology	14
1.2.2.1.1. <i>Neisseria meningitidis</i> vaccine candidates	14
1.2.2.1.2. <i>Streptococcus pneumoniae</i> vaccine candidates	14
1.2.2.1.3. <i>Chlamydia pneumoniae</i> vaccine candidates	15
1.2.2.1.4. <i>Porphyromonas gingivalis</i> vaccine candidates	15
1.2.2.1.5. <i>Bacillus anthracis</i> vaccine candidates	16
1.2.2.2. Comparative genomics	17
1.2.2.3. Expression profiling	17
1.3. Aims of this study	18

CHAPTER 2: THE COMPLETION AND ANNOTATION OF THE GENOME SEQUENCE OF <i>EHRlichia RUMINANTIUM</i> (WELGEVONDEN)	20
2.1. Introduction	20
2.2. Materials and Methods	23
2.2.1. Genome sequencing and assembly	23
2.2.1.1. DNA extraction	23
2.2.1.2. Construction of small insert libraries	23
2.2.1.3. Template preparation for DNA sequencing	24
2.2.1.4. DNA sequencing	24
2.2.1.5. Sequencing data analysis and assembly	25
2.2.1.6. Gap closure and quality assessment	25
2.2.2. Annotation and analysis	26
2.2.2.1. Selection of a gene set	26
2.2.2.2. Similarity searches and domain identification	27
2.2.2.3. Subcellular localisation prediction of ORFs	28
2.3. Results and Discussion	29
2.3.1. Sequence determination of the entire genome	29
2.3.1.1. Library construction	29
2.3.1.2. Genome assembly	30
2.3.2. Annotation of the <i>E. ruminantium</i> genome sequence	33
2.3.2.1. Assignment of potential coding regions	34
2.3.2.2. Functional assignment of protein-encoding genes	34
2.3.2.3. General features of the genome	34
2.3.2.4. Subcellular localisation of ORFs	45
2.3.2.5. Paralogous gene families of membrane proteins	48
2.3.2.6. Pathogenicity-associated genes	49
2.4. Conclusions	50
CHAPTER 3: METABOLIC RECONSTRUCTION AND COMPARATIVE GENOMIC ANALYSIS OF SPECIES WITHIN THE ORDER RICKETTSIALES	62
3.1. Introduction	62
3.2. Materials and Methods	67
3.2.1. Metabolic reconstruction	67
3.2.2. <i>In silico</i> genome comparisons	67
3.3. Results and Discussion	69
3.3.1. Pathway analysis	69
3.3.1.1. Central metabolic pathways	69



3.3.1.1.1. Carbohydrate metabolism	69
3.3.1.1.2. Nucleoside biosynthesis	70
3.3.1.1.3. Amino acid biosynthesis	72
3.3.1.1.4. Cofactor biosynthesis	72
3.3.1.1.5. Lipid metabolism and cell wall components	75
3.3.1.2. Energy metabolism	78
3.3.1.3. Replication, repair and recombination	79
3.3.1.4. Transcription and translation	79
3.3.2. Transporters	80
3.3.3. Synteny analysis	81
3.3.4. Shared and genus-specific genes	82
3.4. Conclusions	86
CHAPTER 4: REPETITIVE DNA IN THE COMPLETE GENOME SEQUENCE OF <i>EHRlichia RUMINANTIUM</i> (WELGEVONDEN)	98
4.1. Introduction	98
4.2. Materials and Methods	100
4.2.1. Analysis of genomic repeat sequences	100
4.2.2. Amplification and cloning of variable repeat regions	100
4.2.3. Amplification of the regions around the <i>rho</i> and <i>tuf</i> genes	101
4.3. Results and Discussion	102
4.3.1. Repeat sequences in the <i>E. ruminantium</i> genome sequence	102
4.3.2. Simple sequence repeats (SSRs)	104
4.3.3. Longer tandem repeats (LTRs)	104
4.3.3.1. Tandem repeats in coding regions	108
4.3.3.2. Repeat regions with variable number of repeat units	111
4.3.4. Interspersed repetitive DNA	113
4.3.4.1. Homologous recombination between repetitive sequences	113
4.3.4.2. Duplications appear to generate new genes	116
4.3.5. Ankyrin repeats	120
4.4. Conclusions	120
CHAPTER 5: SELECTION OF POSSIBLE VACCINE CANDIDATES	127
5.1. Introduction	127
5.2. Materials and Methods	129
5.2.1. <i>In silico</i> selection strategy	129
5.2.2. Expression of recombinant proteins	129

5.2.2.1. Directional cloning into the pET vector	129
5.2.2.2. Expression and purification of recombinant proteins	130
5.2.2.3. Western blot analysis	131
5.2.3. Immunological assays	131
5.2.3.1. Lymphocyte proliferation assays	131
5.2.3.2. IFN- γ ELISpot assays	132
5.2.4. Vaccine trials in sheep	133
5.2.4.1. Challenge material	133
5.2.4.2. DNA immunisation	133
5.2.4.2.1. Cloning of ORFs into pCMViUBs	133
5.2.4.2.2. Large scale DNA preparation	134
5.2.4.2.3. DNA immunisation of sheep	134
5.2.4.3. DNA prime–recombinant protein boost immunisation strategy	135
5.2.4.3.1. Large scale preparation of recombinant proteins	135
5.2.4.3.2. Immunisation of sheep	135
5.3. Results and Discussion	137
5.3.1. <i>In silico</i> selection of possible vaccine candidates	137
5.3.2. Expression of recombinant proteins	140
5.3.3. Physical characteristics of recombinant proteins	140
5.3.4. Recombinant proteins inducing specific Th1 cellular immune responses	141
5.3.5. Vaccine trials in sheep	146
5.4. Conclusions	151
CHAPTER 6: CONCLUDING DISCUSSION	158
APPENDIX A: REFERENCES	165
APPENDIX B: MATERIALS, BUFFERS, MEDIA AND SOLUTIONS	201
APPENDIX C: PRIMERS	204
APPENDIX D: PROTEIN CLASSIFICATION SCHEME	219
APPENDIX E: <i>E. RUMINANTIUM</i> GENE LIST	221
APPENDIX F: WEB BASED TOOLS	246
APPENDIX G: PUBLICATIONS AND ETHICS	247

LIST OF FIGURES

Figure 2.1. A. The physical map of De Villiers <i>et al.</i> (2000). B. A computer-generated restriction map of the completed <i>E. ruminantium</i> genome sequence, showing the cutting sites of the endonucleases <i>KspI</i> , <i>RsrII</i> and <i>SmaI</i> .	32
Figure 2.2. Circular representation of the genome of <i>E. ruminantium</i> (Welgevonden isolate).	35
Figure 2.3. Linear representation of the <i>E. ruminantium</i> (Welgevonden isolate) genome.	44
Figure 2.4. Predicted compartmentalisation of putative proteins by pSORTb and CELLO.	47
Figure 3.1. Neighbour-joining tree based on 16S rRNA sequences showing the phylogenetic relationships of <i>E. ruminantium</i> with other Rickettsiales for which complete genome sequences had been published at the time of this study.	66
Figure 3.2. Schematic overview of metabolic pathways and substrate transport in <i>E. ruminantium</i> .	71
Figure 3.3. <i>E. ruminantium</i> genes coding for the enzymes involved in the TCA cycle, heme biosynthesis and amino acid biosynthesis.	73
Figure 3.4. <i>E. ruminantium</i> genes involved in the pentose phosphate and gluconeogenesis pathways.	74
Figure 3.5. <i>E. ruminantium</i> genes involved in nucleotide metabolism.	76
Figure 3.6. <i>E. ruminantium</i> genes involved in cofactor biosynthesis.	77
Figure 3.7. Global comparison between <i>E. ruminantium</i> (middle), <i>E. chaffeensis</i> (top) and <i>E. canis</i> (bottom) displayed using ACT.	83
Figure 3.8. Comparison of chromosomal synteny between <i>E. ruminantium</i> (middle), <i>A. marginale</i> (top) and <i>A. phagocytophilum</i> (bottom).	83
Figure 3.9. Genomic location of the homologous genes in <i>E. ruminantium</i> (middle) and the two <i>Wolbachia</i> species.	84
Figure 3.10. <i>E. ruminantium</i> gene order compared to <i>N. sennetsu</i> (top) and <i>P. ubiquus</i> (bottom).	84
Figure 3.11. A. Comparison of relative positions of conserved genes between <i>E. ruminantium</i> , <i>R. bellii</i> (top) and <i>R. conorii</i> (bottom). B. <i>E. ruminantium</i> gene order compared to <i>R. felis</i> (top) and <i>R. prowazekii</i> (bottom).	85

Figure 4.1. Amplification and cloning of variable repeat regions from <i>E. ruminantium</i> Welgevonden genomic DNA.	112
Figure 4.2. PCR amplification across the <i>rho</i> repeat regions in <i>E. ruminantium</i> isolates.	115
Figure 4.3. Schematic representation of <i>E. ruminantium</i> genes that may have arisen through duplication events.	118
Figure 4.4. Screen capture from ACT of the area around Erum2490, Erum2500 and Erum2510 in <i>E. ruminantium</i> (middle), compared to <i>E. chaffeensis</i> (top) and <i>E. canis</i> (bottom).	119
Figure 5.1. ELISpot and lymphocyte proliferation assays (LPA) of PBMCs stimulated with recombinant proteins (plate 1).	143
Figure 5.2. ELISpot and lymphocyte proliferation assays (LPA) of PBMCs stimulated with recombinant proteins (plate 2).	144
Figure 5.3. Anti-His ₆ Western blot of the seven selected ORFs expressed in <i>E. coli</i> .	147
Figure 5.4. Reaction index of sheep.	153
Figure 5.5. Daily post-challenge temperatures of the challenge control group (A) and the infected and treated group (B).	154
Figure 5.6. Daily post-challenge temperatures of the negative control groups. A: Sheep inoculated 3x with empty pCMViUBs vector. B: Sheep inoculated twice with empty pCMViUBs vector, followed by a recombinant β -galactosidase protein boost.	155
Figure 5.7. Daily post-challenge temperatures of sheep inoculated 3x with ORF cocktail 1 (A) or ORF cocktail 2 (B) DNA.	156
Figure 5.8. Daily post-challenge temperatures of the prime–boost vaccinated groups. A: Sheep immunised twice with ORF cocktail 1 DNA followed by an ORF cocktail 1 recombinant protein boost. B: Sheep immunised twice with ORF cocktail 2 DNA followed by an ORF cocktail 2 recombinant protein boost.	157

LIST OF TABLES

Table 2.1. General features of the genome of the Welgevonden strain of <i>E. ruminantium</i> .	33
Table 2.2. Functional classification of <i>Ehrlichia ruminantium</i> protein-coding genes.	52
Table 3.1. Characteristics of the Rickettsiales for which genome sequences were available at the time this study commenced.	65
Table 3.2. <i>E. ruminantium</i> genes shared by other Rickettsiales.	87
Table 4.1. Genome properties of the sequenced genomes in the order Rickettsiales.	103
Table 4.2. Tandem repeats in the <i>E. ruminantium</i> genome.	105
Table 4.3. CDSs containing LTRs.	110
Table 4.4. Dispersed repeats in the <i>E. ruminantium</i> genome.	121
Table 5.1. The immunisation strategy for the animal trial.	136
Table 5.2. Number of ORFs identified as possible vaccine candidates grouped according to their putative function, during several rounds of selection and elimination.	139
Table 5.3. Lymphocyte proliferation assays using PBMCs from a naïve and an infected and treated sheep stimulated with recombinant proteins.	142
Table 5.4. Characteristics of the seven ORFs that elicited significant PBMC proliferation and IFN- γ production <i>in vitro</i> .	145
Table 5.5. Predicted sizes of the seven possible vaccine candidates.	147

LIST OF ABBREVIATIONS

A	adenine
aa	amino acids
ABC	ATP-binding cassette
ACT	Artemis Comparison Tool
ADP	adenosine diphosphate
ATP	adenosine triphosphate
BAC	bacterial artificial chromosome
BCG	bacillus Calmette Guérin
bp	base pairs
BSA	bovine serum albumin
C	cytosine
CD	cluster of differentiation
CDS	coding sequence
cfu	colony forming units
CoA	coenzyme A
ConA	Concanavalin A
cpm	counts per minute
CTL	cytotoxic T-lymphocytes
DHF	dihydrofolate
dNTP	deoxynucleotide tri-phosphate
DNA	deoxyribonucleic acid
EC	Enzyme Commission
EDTA	ethylene diamine tetraacetic acid
ELISA	enzyme-linked immunosorbent assay
ELISpot	enzyme-linked immunosorbent spot
FACS	fluorescent-activated cell sorting
FAD	flavin adenine dinucleotide
FCS	foetal calf serum
G	guanine
Gb	gigabase(s)
His	histidine
HRP	horseradish peroxidase
IFA	indirect fluorescent antibody
IFN- γ	interferon-gamma



IgG	immunoglobulin G
IHF	integration host factor
IL	interleukin
IPTG	isopropyl- β -D-thiogalactoside
kb	kilobase(s)
kDa	kilodalton
kPa	kilopascal
LB	Luria-Bertani
LD ₅₀	lethal dose, 50%
LPA	lymphocyte proliferation assay(s)
LTRs	longer tandem repeats
μ Ci	microcurie
M	molar
MAP	major antigenic protein
Mb	megabases
MMR	measles, mumps and rubella
mRNA	messenger ribonucleic acid
MW	molecular weight
NAD ⁺	nicotinamide adenine dinucleotide
NADH	nicotinamide adenine dinucleotide - hydrogen (reduced)
NK	natural killer
N-terminal	amino terminal
OMP	outer membrane protein
ORF	open reading frame
PBMC	peripheral blood mononuclear cells
PBS	phosphate buffered saline
PBS-T	phosphate buffered saline-Tween
PCR	polymerase chain reaction
PT	pertussis toxin
PVDF	polyvinylidene fluoride
r	recombinant
RBS	ribosomal binding site
RI	reaction index
RNA	ribonucleic acid
RNA-Seq	ribonucleic acid sequencing
rpt	repeat
rRNA	ribosomal ribonucleic acid



ru	repeat unit
SDS	sodium dodecyl sulphate
SFC	spot forming cells
SI	stimulation index
SPG	sucrose potassium glutamate
ssDNA	single-stranded DNA
SSRs	simple sequence repeats
T	thymine
TCA	tricarboxylic acid
th	transmembrane helix
Th1	T-helper 1
tmRNA	transfer-messenger ribonucleic acid
TNF- β	tumour necrosis factor-beta
tRNA	transfer ribonucleic acid
U	enzyme unit(s)
Vlp	variable surface lipoprotein

SUMMARY

***Ehrlichia ruminantium*: Genome assembly and analysis, with the identification and testing of vaccine candidate genes**

by

JUNITA LIEBENBERG

PROMOTOR: Prof. B. A. Allsopp

CO-PROMOTOR: Dr. N. E. Collins

DEPARTMENT: Veterinary Tropical Diseases

DEGREE: Ph.D.

A shotgun genome sequencing project was undertaken in the expectation that access to the entire protein coding potential of *E. ruminantium* (Welgevonden) will facilitate the identification of vaccine candidate genes against heartwater. The 1,516,355 bp sequence is predicted to encode 888 proteins and 41 stable RNA species. The most prominent feature is the large number of tandemly repeated and duplicated sequences, some of continuously variable copy number. These repeats have mediated numerous translocation and inversion events and seem to be responsible for the generation of both new full and partial protein coding sequences. There are 32 predicted pseudogenes, most of which are truncated fragments of genes associated with repeats. Of the 13 members of the order Rickettsiales compared in this study, *E. ruminantium* has the lowest coding capacity (62%), lowest GC content (27.5%), but the highest proportion of repetitive sequences, which comprise 8.5% of the genome. Metabolic reconstruction of *E. ruminantium* revealed the metabolic and biosynthetic capabilities typical of an obligate intracellular organism. We identified a number of genes unique to *E. ruminantium*, most of which are not functionally

characterised in any organism, and those shared with 12 other members of the Rickettsiales. Bioinformatic tools were used to identify possible vaccine candidates from the annotated genome sequence. The protective properties of seven open reading frames (ORFs), which induced cellular immune responses *in vitro*, were tested *in vivo*. Only 20% survival was obtained in sheep immunised with a DNA formulation consisting of three ORFs. We found that the levels of peripheral blood mononuclear cell proliferation and interferon-gamma (IFN- γ) production did not correlate with each other, nor with the levels of protection, suggesting that the current assays are just not reliable and that IFN- γ expression alone is not an indicator of protection. Therefore more cytokines and different assays will have to be investigated to define in detail what constitutes a protective immune response against *E. ruminantium* infection. However, the data generated from the genome sequence will continue to facilitate novel approaches to study the organism and to develop an efficacious vaccine against heartwater.