UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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 Zweygarth 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

LIST OF TABLES

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 peroxide
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
µCi	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.