

**ANALYSIS OF *BABESIA ROSSI* TRANSCRIPTOME IN DOGS DIAGNOSED WITH
CANINE BABESIOSIS**

By

Keneilwe Pelloakgosi-Shikwambani

Submitted in accordance with the requirements for the degree

Master of Science

in the subject

Life Science

at the

UNIVERSITY OF SOUTH AFRICA

Supervisor

Prof P.T. Matjila

Co-supervisor

Dr K. P. Sibeko

April 2018

Declaration

I, Keneilwe Peloakgosi-Shikwambani, hereby declare that this dissertation, which I hereby submit for the degree of Master of Science (Life Sciences) at the University of South Africa, is my own work and has not previously been submitted by me for a degree at this or any other university.

I declare that the dissertation does not contain any written work presented by other persons whether written, pictures, graphs or data or any other information without acknowledging the source.

I declare that where words from a written source have been used, the words have been paraphrased and referenced and where exact words from a source have been used the words have been placed inside quotation marks and referenced.

I declare that I have not copied and pasted any information from the Internet, without specifically acknowledging the source and have inserted appropriate references to these sources in the reference section of the dissertation or thesis.

I declare that during my study I adhered to the Research Ethics Policy of the University of South Africa, received ethics approval for the duration of my study prior to the commencement of data gathering, and have not acted outside the approval conditions.

I declare that the content of my dissertation has been submitted through an electronic plagiarism detection program before the final submission for examination.

Student signature:



Peloakgosi

Date: 23 April 2018

Acknowledgements

I take this opportunity to gratefully acknowledge my supervisor, Prof Tshepo Matjila and co-supervisor, Dr Kgomotso Sibeko for their time, guidance, support and contribution in this study. Your wisdom and knowledge to the highest standard inspired me.

A sincere thank you to Prof Andrew Leisewitz for collection of the blood samples used in this study.

Deserving a special mention are my fellow postgraduate students, Dr Agatha Kolo, for her patience and assistance in the laboratory and Dr Petunia Malatji for her assistance during the transcriptome data analysis.

A warm felt appreciation to my parents and husband, for lovingly and unselfishly caring for me and believing in my endeavors.

A special thanks to family members and friends who have always supported and encouraged me throughout this study.

A huge thank you to the following organizations: Unisa Postgraduate Bursary, National Research Foundation and University of Pretoria (Genomic Research Institute) for their generous funding of this study, Department of Veterinary and Tropical Diseases at the University of Pretoria for giving us permission to use their facilities to perform the laboratory work and Agricultural Research Council, Biotechnology Platform for allowing us to use their Trinity software for transcriptome data analysis.

Saving the best for last, is my ever-faithful God Almighty, for the strength, patience determination and guidance He provided to me to complete this study, Amen.

Table of contents

Acknowledgements	ii
Table of contents.....	iii
List of figures.....	vii
List of tables	ix
List of abbreviations and symbols	x
ABSTRACT	xiii
CHAPTER 1	1
1. INTRODUCTION	1
CHAPTER 2	3
2. LITERATURE REVIEW.....	3
2.1. <i>Babesia rossi</i>	3
2.2. Life cycle of <i>Babesia</i> species.....	4
2.2.1. Development in the tick	4
2.2.2. Development in the vertebrate host.....	4
2.3. Occurrence of <i>Babesia rossi</i> infection in South Africa	5
2.4. Canine babesiosis	7
2.4.1. Epidemiology of <i>Babesia</i> species causing canine babesiosis	7
2.4.2. Canine babesiosis induced by <i>Babesia rossi</i>	8
2.4.3. Pathogenesis of canine babesiosis	8
2.5. Diagnosis of canine babesiosis.....	10
2.5.1. Microscopic identification.....	10
2.5.2. Serology-based methods.....	11
2.5.3. Molecular-based methods	11
2.6. Control of <i>Babesia rossi</i> infections.....	13
2.7. <i>Babesia rossi</i> genotypes.....	13
2.8. Transcriptome.....	15

2.8.1. Transcriptome sequencing	15
2.8.2. De novo assembly	17
2.9. Functional annotation	18
2.10. Problem identification	18
2.11. Study aim.....	19
2.12. Research objectives	19
CHAPTER 3	20
3. METHODOLOGY.....	20
3.1. Sample collection.....	20
3.2. DNA extraction and quantification.....	20
3.3. Screening of DNA samples for hemoparasite infection using the Reverse Line Blot Hybridization Assay	21
3.4. Genotyping by qualitative Polymerase Chain Reaction (PCR)	25
3.5. DNA sequencing and analysis	26
3.6. RNA extraction and quality assessment	26
3.7. Transcriptome sequencing.....	28
3.7.1. Library construction and validation	28
3.7.2. Pre-processing of sequence data and data delivery	29
3.7.3. <i>De novo</i> transcriptome assembly	30
3.8. Functional annotation of the <i>Babesia rossi</i> transcriptome	30
3.9. Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway analysis ...	31
3.10. Cluster of Orthologous Groups (COG) Classification analysis.....	31
CHAPTER 4	32
4. RESULTS	32
4.1. DNA quality assessment.....	32
4.2. Detection of <i>Babesia rossi</i> DNA by Reverse Line Blot hybridization assay .	32
4.3. Qualitative PCR analysis for detection of <i>Babesia rossi</i> DNA from dog.....	32

blood samples.....	32
4.4. DNA sequence analysis.....	32
4.5. RNA quantity and quality	33
4.6. <i>De novo</i> transcriptome assembly statistics of <i>Babesia rossi</i> genotypes 19 and 29.....	35
4.7. Pairwise analysis of <i>Babesia rossi</i> genotypes 19 and 29 transcriptomes. ..	37
4.8. Length distribution of the unigenes detected in <i>Babesia rossi</i>	38
transcriptome.....	38
4.9. Annotation statistics of the combined unigenes of <i>Babesia rossi</i>	39
transcriptome.....	39
4.10. Sequence homology analysis of <i>Babesia rossi</i> transcriptome	39
4.10.1 Similarity distribution.....	39
4.10.2. E-value distribution	40
4.10.3. Top hit species distribution	41
4.11. Functional annotation of the unigenes in <i>Babesia rossi</i> transcriptome	41
4.11.1. Comparative analysis of gene ontology outputs for unigenes detected in transcriptomes of <i>Babesia rossi</i> genotypes 19 and 29.	42
4.11.2. Gene ontology analysis of the combined unigenes in <i>Babesia rossi</i> ..	44
transcriptome.....	44
4.12. Pathway associations of combined unigenes in <i>Babesia rossi</i>	46
transcriptome	46
4.13. Cluster of Orthologous Groups (COG) classification	47
CHAPTER 5	48
5.1. DISCUSSION	48
5.2. CONCLUSION.....	55
5.3. RECOMMENDATION STUDIES	55
5.4. LIMITATIONS	56

LIST OF REFERENCES	57
APPENDIX	74

List of figures

Figure 1: Morphological appearance of Giemsa-stained red blood cells infected with <i>Babesia</i> (arrows).....	3
Figure 2: Life cycle of <i>Babesia</i> in a vertebrate host.....	5
Figure 3: The number of blood samples collected from domestic dogs in each of five South African provinces.....	6
Figure 4: Tree construction showing the phylogenetic relationship between various genotypes based on <i>BrEMA1</i> gene sequences. <i>BrAK</i> (adenosine kinase gene from <i>B. rossi</i> , GenBank accession number AJ223322. Numbers at the notes indicates bootstrap values.....	15
Figure 5: Representation of RLB hybridization analysis for detection of <i>B. rossi</i> DNA in blood samples collected from sick domestic dogs. Lane 1: positive controls, lane 2: negative control, lane 3-12: PCR products.....	75
Figure 6A: Melting curves of <i>BrEMA1</i> gene in positive <i>B. rossi</i> DNA samples (A and B small arrows) collected from sick domestic dogs.....	77
Figure 6B: Melting peaks of <i>BrEMA 1</i> gene in positive DNA samples (A, B) collected from sick domestic dogs.....	77
Figure 7: Identification of four <i>B. rossi</i> genotypes in the positive <i>B. rossi</i> blood samples collected from sick domestic dogs.....	33
Figure 8: Venn diagram showing the distribution of the common and exclusively detected unigenes between the <i>B. rossi</i> genotypes 19 and 29. Numbers inside the coloured circles indicate the number of unigenes. <i>Br</i> geno 19= <i>B. rossi</i> genotype 19; <i>Br</i> geno 29= <i>B. rossi</i> genotype 29.....	37
Figure 9: The length distribution of the unigenes in <i>Babesia rossi</i> transcriptome.....	38

Figure 10: Data distribution of the unigenes in <i>Babesia rossi</i> transcriptome. Numbers indicate the number of unigenes.....	39
Figure 11: Sequence similarity distribution of the blast hits of each unigene in <i>B. rossi</i> transcriptome.....	40
Figure 12: E-value distribution of the blast hits for unigenes from <i>B. rossi</i> transcriptome, using a cut -off value of 10E-3.....	41
Figure 13: Species distribution for homologous sequences of <i>B. rossi</i> transcriptome unigenes.....	42
Figure 14: Top 10 most enriched GO terms, for specific ontology categories, detected from analysis of the unigenes of <i>Babesia rossi</i> genotypes 19 and 29. The x-axis indicates the gene ontology terms and the bar shows the number of unigenes.....	43
Figure 15: Top 10 GO terms of the combined unigenes in <i>Babesia rossi</i> transcriptome. The x-axis indicated the gene ontology terms and the bar and numbers show the number of unigenes.....	45
Figure 16: Top 10 most enriched pathways detected from the combined unigenes of <i>Babesia rossi</i> . The x-axis indicates the pathway and the bar shows the number of unigenes.....	46
Figure 17: Cluster of Orthologous Groups classification of unigenes in <i>Babesia rossi</i> transcriptome.....	47

List of tables

Table 1: Geographical distribution of <i>Babesia</i> species causing canine babesiosis across the world.....	7
Table 2: List of probes and their corresponding sequences used to identify the pathogen DNA on the RLB.....	24
Table 3: DNA concentration and 260/280 ratio values from 20 domestic dog samples.....	74
Table 4: RLB and qualitative PCR results of 20 DNA samples.....	75
Table 5: Samples and their melting temperatures on real-time PCR.....	77
Table 6: RNA concentration and 260/280 ratio values from 20 domestic dog samples.....	79
Table 7: RNA quality and quantity of the samples infected with different <i>B. rossi</i> genotypes using BioAnalyzer at the University of Pretoria, South Africa.....	35
Table 8: RNA testing report using the Agilent 2100 and Nanodrop at BGI, Hong Kong.....	36
Table 9: Summary statistics of the transcriptome assemblies of <i>B. rossi</i> genotypes 19 and 29.....	37

List of abbreviations and symbols

A	adenine
ARDS	acute respiratory distress syndrome
bp	base pairs
BLAST	basic local alignment search tool
<i>Br</i> EMA1	<i>Babesia rossi</i> erythrocyte antigen membrane 1
C	cytosine
cDNA	complementary DNA
COG	cluster of orthologous groups
DNA	deoxyribonucleic acid
G	guanine
ECL	enhanced chemiluminescence
EDTA	ethylenediamine tetra-acetic acid
ELISA	enzyme immunosorbent assay
GO	gene ontology
ICT	immune chromatographic test
IFA	indirect fluorescent antibody
KEGG	Kyoto Encyclopaedia of Genes and Genomes
kg	kilogram
M	molar
min	minute/s
mg	milligram
ml	millilitre

MODS	multiple organ dysfunction syndrome
mM	millimolar
NaHCO ₃	sodium bicarbonate
NaOH	sodium hydroxide
NCBI	national centre for biotechnology information
ng	nanogram
NGS	next generation sequencing
nm	nanometre
OVAH	Onderstepoort Veterinary Academic Hospital
PCR	polymerase chain reaction
qPCR	quantitative PCR
R	rand
RIN	integrity number
RLB	reverse line blot
rpm	revolutions per minute
RNA	ribonucleic acid
RNA-Seq	RNA-sequencing
sec	seconds
SDS	sodium dodecyl sulphate
SPA	soluble parasite antigens
SSPE	sodium chloride/sodium phosphate/EDTA
T	thymine

T _m	melting temperature
μM	micromolar
μl	microlitre
μg	microgram
°C	degrees Celsius
%	percentage
≥	greater than

ABSTRACT

Background: Canine babesiosis is a tick-borne disease causing detrimental health effects on the domestic dogs with huge economic impact on the owners. The most complicated form of canine babesiosis is caused by a pathogenic *Babesia rossi* parasite. Canine babesiosis induced by *B. rossi* still remains the cause of mortality and morbidity in South African dogs, yet, the transcriptomic and genomic information of this parasite species is still not available. The transcriptomic and genomic information is essential in the disease development and processes for the design of effective disease control strategies. Consequently, our understanding of the mechanisms underlying the pathogenesis of the different genotypes of *B. rossi* remains limited. A previous study suggested a relationship between the parasite genotype and the disease phenotype. To date, thirteen *B. rossi* genotypes have been identified and associated with diverse clinical signs in their hosts. Hence the aim of this study was to sequence RNA from samples representing *B. rossi* genotypes, 19, 29 and 31, in order to have insight on the overall transcriptome of this parasite and to establish if there would be significant differences among the genotypes.

Methodology: To screen for *B. rossi* positive samples, total DNA was extracted from 20 blood samples collected from sick domestic dogs presented at the Onderstepoort Veterinary Academic Hospital (OVAH). *Babesia rossi* infections were confirmed using the PCR-Reverse Line Blot (RLB) hybridization assay. Further confirmation of infection status was done by amplification of the *B. rossi* Erythrocyte Membrane Antigen 1 (*BrEMA1*) gene in all the DNA samples using qualitative PCR (qPCR), followed by sequencing of PCR products. Subsequently, total RNA was extracted from the 20 *B. rossi*-infected blood samples collected from the same dogs in which DNA was extracted. Three samples representing *B. rossi* genotypes 19, 29 and 31 were selected for transcriptome analysis. RNA sequencing was performed using the Illumina HiSeq 2000 to allow transcriptome analysis. *De novo* assembly was performed independently for all three transcriptomes using the Trinity software. The unigenes generated from specific transcriptome assemblies were subjected to global functional annotation using Blast2GO version 2.8.0 software, followed by KEGG

database for annotation of biological pathways, and DAVID version 6.7, for COG classification to predict and classify their functions.

Results: The sample representing *B. rossi* genotype 31 was excluded in the transcriptome analysis due to low RNA mass, which usually compromises the quality of the library used in RNA sequencing. Thus, a total of 26 747 238 and 25 709 627 paired-end reads were obtained from *B. rossi* genotypes 19 and 29, respectively. *De novo* transcriptome assembly produced a total of 3019 unigenes, with an average length of 419 bp and N50 of 362 bp in *B. rossi* genotype 19, and 2727 unigenes with an average length of 441 bp and N50 of 362 in *B. rossi* genotype 29. A total of 1193 unigenes were common between *B. rossi* genotype 19 and 29, while 1828 unigenes were exclusively detected in *B. rossi* genotype 19; and 1534 were specific to *B. rossi* genotype 29. Between the two *B. rossi* genotypes, a total of 4553 unigenes were obtained, representing the overall *B. rossi* transcriptome. From the overall transcriptome, 12.3% (n=558) of the unigenes could be annotated with 53 different gene ontology (GO) functional categories. About 34% (n=1550) of the unigenes represented in the overall transcriptome mapped to 237 KEGG pathways and only 2.5% (114) could be annotated in the COG database.

Conclusion: Although, there were no striking differences in the transcriptomes of *B. rossi* genotypes 19 and 29, this study presents the first transcriptomic resource for *B. rossi*, which will highly contribute to our genetic understanding of *B. rossi* and provide a platform for future gene expression studies. Hypothetical proteins identified in this study will require further characterization as they may have a critical role in the biology and pathogenicity of *B. rossi* parasite.

Keywords: *B. rossi* genotypes, *de novo* analysis, transcriptome, canine babesiosis, RNA-sequencing.

CHAPTER 1

1. INTRODUCTION

Babesia rossi is the most virulent tick-transmitted apicomplexan parasite responsible for causing canine babesiosis in dogs (Schetters *et al.*, 2009). Canine babesiosis is the common cause of complicated clinical syndromes in dogs in South Africa (Last *et al.*, 2007). The disease presentation ranges widely from acute to chronic or even subclinical (Schoeman, 2009). The clinical manifestations of canine babesiosis include fever, severe anaemia, shock, splenogamy, multiple organ failure, such as acute renal failure (ARF), acute respiratory distress syndrome (ARDS), and hepatic dysfunction, and in some cases *B. rossi* infection result in death of the infected dogs (Jacobson, 2006).

Twelve percent (1253 out of 10710) of the sick dogs were diagnosed with canine babesiosis in a study conducted at Onderstepoort Veterinary Academic Hospital (OVAH) in Gauteng from the records of outpatients from 1988 to 1993 (Shakespeare, 1995). The death rate in *B. rossi* infections was around 12% and only 1% for the least pathogenic *Babesia vogeli* (Lobetti, 2006), resulting in an average of 104 canine babesiosis cases per month or just over 3 dogs per day with the disease. Hence canine babesiosis is a major disease of domestic dogs in Southern Africa, and causes considerable veterinary expenditure and a lot of suffering and a loss of life of dogs 1993 (Shakespeare, 1995).

The babesial infections are usually diagnosed based on the size and morphological appearance of intra-erythrocytic forms in peripheral blood smears (Schoeman, 2009). Other studies have reported that diagnosis based on serological methods is not specific due to cross reactivity between species and even between genera (Yamane *et al.*, 1993; Ano *et al.*, 2001). Hence, advanced molecular techniques such as polymerase chain reaction (PCR)-based tests and sequencing technologies have been established and are widely used to detect canine piroplasm nucleotides due to their high sensitivity and specificity (Holman *et al.*, 2002).

Currently, chemical drugs and vaccination are used to control the diseases caused by the apicomplexan parasites (Jakalski *et al.*, 2014). These drugs and vaccines do

not have a direct effect on the parasite itself but only lessen the severity of the clinical symptoms in the host (Aboge *et al.*, 2008). However, the continual development of parasites that are resistant to treatment is a serious concern, hence a valuable resource that could provide a platform to understand the underlying interaction mechanisms of apicomplexan parasite and their host are needed (Jakalski *et al.*, 2014).

Over the years, the approach to determine genome sequences of an organism has greatly improved the efficiency and speed of gene discovery (Ansorge, 2009). Other recent advances in sequencing technologies, such as transcriptome (also known as RNA sequence) analysis, have been developed (Pareek *et al.*, 2011). Transcriptome reflects cellular activity within a tissue at a given point in time inside the organism (Twine *et al.*, 2011). Genome and transcriptome sequencing of various apicomplexan parasites, such as *Plasmodium falciparum*, *Theileria parva*, *Babesia bigemina* and *Babesia bovis* have been completed (Bishop *et al.*, 2005; Bozdech *et al.*, 2003; Pedroni *et al.*, 2013). The transcriptome analysis of the aforementioned organisms improved the understanding of gene expression in connection with parasite-host interaction (Jongejan *et al.*, 1987), whereas this still needs to be established for *B. rossi*.

Apicomplexans, such as *Babesia*, are intracellular parasites that have a close relationship with their hosts and therefore have established pathways that help in their transmission (Boyle *et al.*, 2006). Transcriptome analysis has allowed gene expression studies to identify the genes whose transcripts are up or down regulated during the parasite's life cycle in order to improve the understanding of the parasite's development and differentiation in the host (Boyle *et al.*, 2006). A previous report linked *B. rossi* with varying clinical signs and life-threatening disease in the dog host; however, the underlying mechanisms of such association are still unknown (Matjila *et al.*, 2009). These mechanisms could be determined from genomic or transcriptome information of *B. rossi*. Thus, this study presents the first ever report of the analysis of the whole transcriptome of *B. rossi* parasite.

CHAPTER 2

2. LITERATURE REVIEW

2.1. *Babesia rossi*

Babesia rossi was initially discovered in a side-striped jackal (*Canis adustus*) in East Africa (Nuttall, 1910) and was classified as a large *Babesia* species (3.0-5.0 μm) (Carret *et al.*, 1999). This infective agent is transmitted by the *Haemaphysallis elliptica*, previously misidentified as *Haemaphysallis leachi* (Apanaskevich *et al.*, 2007). *Babesia rossi* occurs in Sub Saharan Africa (Lewis *et al.*, 1996) and is a common infection in South African dogs (Matjila *et al.*, 2004).

Babesia infection is recognized in dogs based on the morphological appearance of the parasite in the erythrocyte (Solano-Gallego and Baneth, 2011). A typical intraerythrocytic piroplasm is pear shaped and often occurs in pairs (Homer *et al.*, 2000) as shown in Figure 1. In acute infection, a red blood cell may contain four to sixteen parasites whereas in chronic infection, an amoeboid and ring form can be seen (Agnieszka *et al.*, 2005).

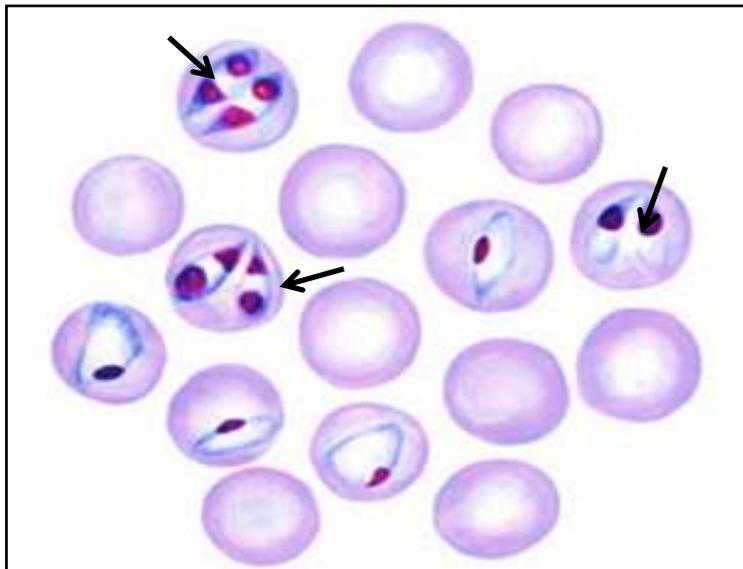


Figure 1: Morphological appearance of Giemsa-stained red blood cells infected with *Babesia* (arrows). The Merck Veterinary Manual, courtesy of State of Queensland, Department of Agriculture, Fisheries and Forestry.

2.2. Life cycle of *Babesia* species

2.2.1. Development in the tick

Once the tick feeds on the red blood cell infected with the *Babesia*, most of the parasites become less functional and are destroyed (Chauvin *et al.*, 2009). However, some stages of the parasite survive and develop into gametocytes (Mehlhorn and Schein, 1984). The gametes then fuse inside the digestive tract of the tick to form the zygotes. The zygotes penetrate the midgut cells and develop into ookinetes (Mehlhorn and Schein, 1984). The ookinetes can then move freely and appear to be haploid. In the midgut, meiosis occurs, indicating the beginning of sporogony (Hunfeld *et al.*, 2008). The ookinetes subsequently break free from the midgut and enters the body tissues of the tick. In the female tick, the ookinetes invade the ovary and the eggs become infected with the *Babesia*. Kinetes then enter the salivary glands of the tick, where the sporozoites are produced (Chauvin *et al.*, 2009). The sporozoites are the infectious stage of the parasite and they introduce the protozoa into the mammalian or vertebrate host (Chauvin *et al.*, 2009; Hunfeld *et al.*, 2008).

2.2.2. Development in the vertebrate host

The sporozoites are transferred into the vertebrate host during the tick bite. Sporozoites invade the red blood cells of the host and thereafter are called trophozoites (Chauvin *et al.*, 2009). Binary fission takes place and the parasite produces two merozoites. After the division, the red blood cells burst to discharge the merozoites, which invade new erythrocytes (Hunfeld *et al.*, 2008). Multiplication of the parasite and infected cells progresses until either the host dies or more usually until the immune system of the host intervenes to stop multiplication (Chauvin *et al.*, 2009; Hunfeld *et al.*, 2008). The general life cycle of *Babesia* species in a host is demonstrated in Figure 2.

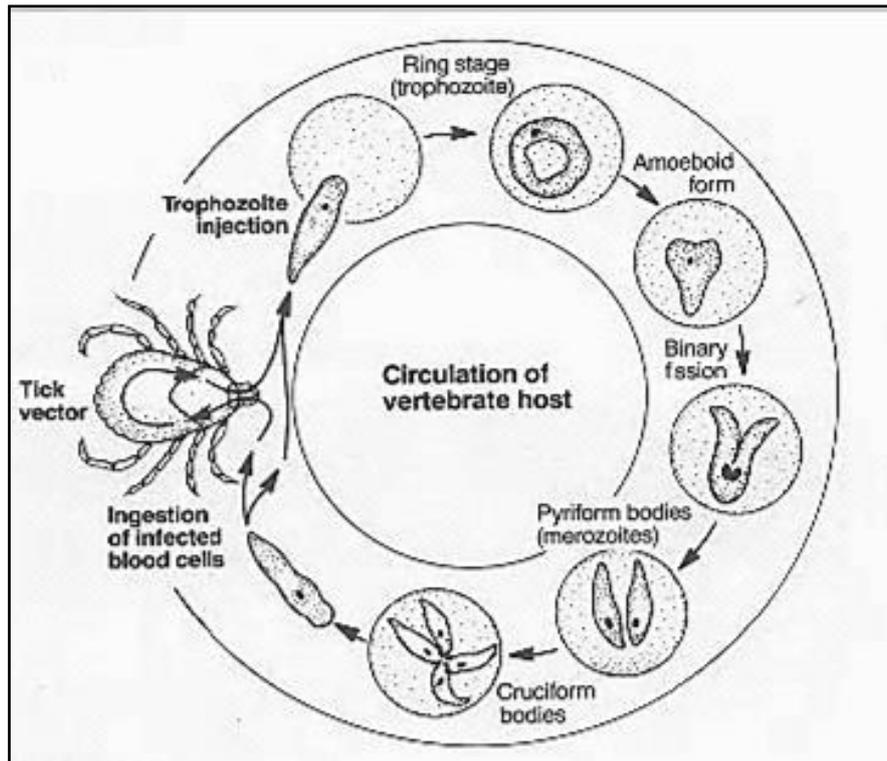


Figure 2: Life cycle of *Babesia* in a vertebrate host (Oppredoes. 1998).

2.3. Occurrence of *Babesia rossi* infection in South Africa

Babesia rossi infection is widely spread in South Africa (Matjila *et al.*, 2004). In the study conducted by Matjila *et al* (2004), blood samples from domestic dogs were collected from different provinces in South Africa as indicated in Figure 3. The proportion of sampled domestic dogs that tested positive for *B. rossi* for the different provinces were as follows: (345/527; 65%) from Gauteng (OVAH), (2/90; 2%) from Southern Gauteng, (36/38; 95%) from Mpumalanga, (7/129; 5%) from Free State, (16/253; 6%) from Kwa-Zulu Natal, (12/54; 22%) from Eastern Cape and (2/47; 4%) from Western Cape.

In Southern Gauteng, Mpumalanga, Free State, Kwa-Zulu Natal, Eastern Cape and Western Cape, co-infection of *B. rossi* and *Ehrlichia canis* was observed, while triple infection of *B. rossi*, *E. canis* and *Babesia vogeli* was detected from samples obtained from the OVAH (Matjila *et al.*, 2004). Most *Babesia*-infected dogs were infested with *Haemaphysallis elliptica*, the tick responsible for transmitting *B. rossi*. The high prevalence of *B. rossi* infections in dogs presented at the OVAH corresponded with the high number of ticks collected from these dogs. These findings suggested that many tick vectors found on the dogs causes the recurring

transmission of tick-borne diseases (Matjila *et al.*, 2004). Therefore, there is a demand for efficient control measures to avoid the transmission of tick-borne pathogens between domestic dogs in South Africa. Certainly, this study highlighted the alarming number of tick-borne pathogens in the dog population of South Africa (Matjila *et al.*, 2004).

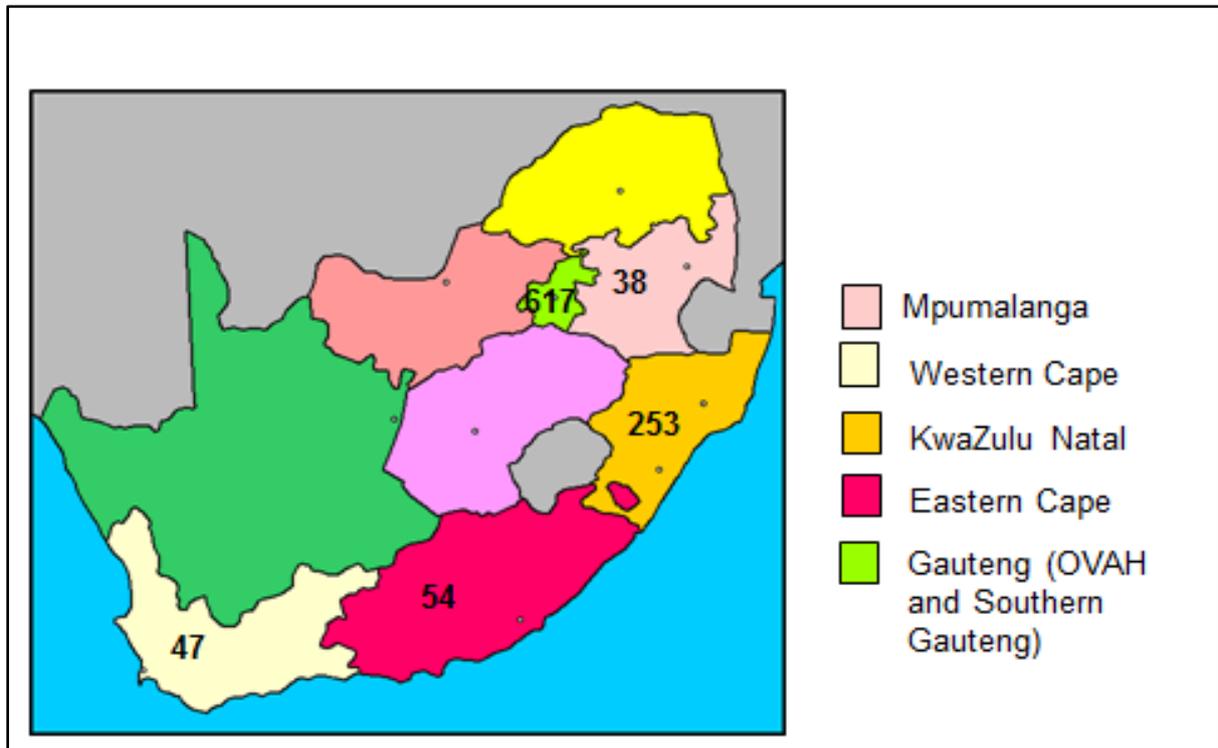


Figure 3: The number of blood samples collected from domestic dogs in each five South African provinces (Matjila *et al.*, 2004). Source: www.exploresouthafrica.net/images/provmap.jpg.

As noted previously, *B. rossi* was first reported in a side-striped jackal (*Canus adustus*) in Kenya (Nutall, 1910). Thereafter, it was also reported in blood and organ smear from a jackal pup (Nutall, 1912). Two black-backed jackals were reported to be infected with *B. rossi* and without any apparent clinical signs (Van Heerden, 1980). *Babesia rossi* has also been observed in wild dogs (*Lycaon pictus*) (17/227, 7%) at a breeding centre (Matjila *et al.*, 2008). Since clinical cases of babesiosis had never been recorded in the wild dogs, it appeared that *B. rossi* occurs naturally in wild dogs. This enables the wild dogs to cope with the *B. rossi* infections and become the carriers of the parasite (Penzhorn, 2011).

2.4. Canine babesiosis

2.4.1. Epidemiology of *Babesia* species causing canine babesiosis

The occurrence of the three *Babesia* species, namely; *B. rossi*, *B. canis* and *B. vogeli*, causing canine babesiosis in a certain region is influenced by the presence of the specific tick vector in that area (Matjila *et al.*, 2004). The ticks responsible for transmitting *B. rossi*, *B. canis* and *B. vogeli* are *Haemaphysallis elliptica*, *Dermacentor reticularis* and *Rhipicephalus sanguineus* respectively (Caccio *et al.*, 2002; Matjila *et al.*, 2004). *Babesia rossi* is common in South Africa (Matjila *et al.*, 2004). Currently, *B. rossi* has been recorded in other African regions such as Nigeria (Sasaki *et al.*, 2007) and Sudan (Oyamada *et al.*, 2005), where its tick vector is endemic. *Babesia canis* (senso-stricto) occurs in Europe (Uilenberg *et al.*, 1989), while *Babesia vogeli* occurs in North America and Brazil (Caccio *et al.*, 2002) with only a few cases reported in South Africa (Matjila *et al.*, 2004) (Table 1).

Table 1: Geographical distribution of *Babesia* species causing canine babesiosis across the world (Schoeman, 2009).

<i>Babesia</i> species	Tick vector	Country
<i>B. rossi</i>	<i>Haemophysallis Elliptica</i>	South Africa Nigeria, Sudan.
<i>B. canis</i>	<i>Dermacentor reticulatus</i>	France , Spain Hungary, Switzerland Germany, Belgium Netherlands.
<i>B. vogeli</i>	<i>Rhipicephalus sanguineus</i>	United States of America Okinawa, Japan France, Australia South Africa.

2.4.2. Canine babesiosis induced by *Babesia rossi*

Canine babesiosis is a tick-borne disease caused by large and small piroplasms inside the red blood cells (Taboada and Merchant, 1991). Canine babesiosis caused by *B. rossi* in dogs may show clinical signs which have been classified as uncomplicated or complicated (Böhm *et al.*, 2006). The disease is characterised as an uncomplicated case if there is only a mild anaemia, with no indication of an organ dysfunction or failure. Examples of the clinical signs in the uncomplicated cases of *B. rossi* infection include fever, lethargy and splenomegaly (Jacobson and Lobetti, 1996). Dogs in this category are treated with antibabesial drugs and also blood transfusion if necessary. Moreover, these dogs have been reported to have a full recovery and high survival rate (Jacobson, 2006).

In contrast, complicated cases of *B. rossi* infection in dogs are evidenced by severe anaemia, organ failure and haemoconcentration (Jacobson and Lobetti, 1996). Examples of clinical signs in complicated cases are acute renal failure, acute respiratory distress syndrome, hypotension, haemolytic anaemia and shock (Jacobson and Lobetti, 1996). Dogs with complicated cases need to receive an intensive treatment at a veterinary medical emergency care facility. Their mortality rate is usually averaged at 15%, regardless of treatment (Shakespeare, 1995; Collet, 2000).

Canine babesiosis in South Africa is caused by the virulent *B. rossi*, although the less pathogenic species, *B. vogeli* also occurs albeit in fewer cases (Matjila *et al.*, 2004). This common disease was detected in 12% of dogs admitted at the OVAH and 31% of the affected dogs were admitted in hospital in serious condition (Shakespeare, 1995). Matijatko *et al.* (2009) reported the mortality rate of 13.9% in Zagreb, which is parallel to the 10-15% mortality rate recorded in studies of canine babesiosis caused by *B. rossi* in South Africa.

2.4.3. Pathogenesis of canine babesiosis

The most common clinical sign of acute canine babesiosis in South Africa was reported to be hemolytic anemia (Jacobson, 2006). The parasite directly disrupts the red blood cells and haemolytic anaemia occurs (Shinuo, 2013). Reyers *et al.* (1998) reported that dogs with canine babesiosis had severe anemia, neutrophilia and monocytosis from their two hematological studies. This supports the fact that *B.*

rossi; the most prevalent African strain causes more severe hematological changes than the less virulent *B. canis* found in Europe (Scheepers *et al.*, 2011).

Shock often exists in dogs with canine babesiosis. One of the proposed mechanisms leading to the development of shock in canine babesiosis is related to the parasite parasitemia (Matijatko *et al.*, 2009). The link between a higher level of parasitemia and the shock is not yet understood (Matijatko *et al.*, 2009). However, severe parasitemia is unlikely the only trigger of the shock in canine babesiosis, especially since some dogs with low levels of parasitemia is also known to develop shock (Böhm *et al.*, 2006).

Hypotension is also common in canine babesiosis due to the release of vasoactive amines and cytokines, which produce vasodilation (Shinuo, 2013). The reduction in vascular volume caused by increased vascular permeability and decreased myocardial function also play a role in hypotension (Jacobson and Lobetti, 1996). Hypotension is one of the main causes of poor tissue perfusion, which is especially harmful to renal tissue (Bone *et al.*, 1992). Collapse and other neurological signs often indicated in cerebral babesiosis could be caused by hypoglycaemia, which has been reported in dogs with canine babesiosis caused by *B. rossi* (Keller *et al.*, 2004).

The development of multiple organ dysfunction syndromes (MODS) has been recorded in canine babesiosis induced by the *B. rossi* (Welzl *et al.*, 2001). Multiple organ dysfunction syndromes occur due to the dysregulation of the pro-inflammatory and anti-inflammatory mechanisms and results in generalized auto destructive inflammation (Goris *et al.*, 1985). Welzl *et al* (2001) reported that 34% of the dogs with complicated babesiosis caused by *B. rossi* showed high serum creatinine levels. In the same study, the most frequently affected organ was the liver, followed by kidneys, muscles, lungs and central nervous system (Welzl *et al.*, 2001). The above mentioned clinical signs demonstrated in dogs diagnosed with canine babesiosis caused by *B. rossi* are severe and life-threatening, highlighting the need to develop rapid diagnostic techniques for early detection of infection and eventually the vaccines for control of the disease.

2.5. Diagnosis of canine babesiosis

There is a challenge in diagnosing babesiosis in the chronically infected dogs because of very low, often periodic parasitemia. Hematological and molecular tests are recommended since diagnosis and signs of babesiosis are nearly nonspecific (Vanier and Krause, 2009). The laboratory methods that are used to diagnose babesiosis are discussed below.

2.5.1. Microscopic identification

Microscopic examinations continue to be the simplest diagnostic test to detect intraerythrocytic parasites during acute infections (Irwin, 2009). Microscopy detects parasite inside the red blood cells using Giemsa or Wright's stained blood smears (Böhm *et al.*, 2006). Parasitized red blood cells found in the infected host vary over time. Sometimes the parasitemia in the host is quickly detectable by light microscopy even though the host does not show any symptoms of the disease (Calder *et al.*, 1996). Blood smears are used to express the quantity of parasitemia in *Babesia* research (Lewis *et al.*, 1995). The method is rapid and low cost with high specificity but lacks sensitivity to detect low level parasitemia (Dantas-Torres and Figueredo, 2006). It also fails to differentiate among parasites and is not applicable for large scale epidemiology, prevalence and distribution studies (Silva *et al.*, 2010).

In South Africa, the diagnosis of canine babesiosis is comprised of a capillary smear taken from the ear pinna (Keller *et al.*, 2004). Chances to detect the organism are increased by using smears prepared from the capillary blood such as ear tip or nail bed. This is due to the fact that infected red blood cells tend to sludge in the capillaries (Böhm *et al.*, 2006). Level of parasitemia is usually higher in the capillary blood than in venous blood as often seen in dogs with *B. rossi* infections. Therefore, capillary samples are the most desired diagnostic samples (Böhm *et al.*, 2006). In spite of the extensive blood smears applications, there is still limited information about the relationship between parasitemia and the outcome in *B. rossi* infection. Failure to detect parasites in animals with haemolytic anaemia or thrombocytopenia has led to an incorrect diagnosis in reported cases (Birkenheuer *et al.*, 2005; Yeagley *et al.*, 2009).

2.5.2. Serology-based methods

Serological testing is often used but limited because of uniform tests, cross-reactivity, and false-negative results in very young animals or early infection (Jia *et al.*, 2006). An indirect fluorescent antibody (IFA) is the most frequently used serological test for babesiosis (Levy *et al.*, 1987). The technique involves a reaction between antibodies from the serum of the infected individual with the corresponding fluorescent-labelled antigen (Skotarcvazak, 2008). Furthermore, the method lacks sensitivity because of cross-reactions between *Babesia* species and with other apicomplexan parasites and it is not sufficient for large scale screening (Aboge *et al.*, 2007). Immunochromatographic test (ICT) was also developed for quicker diagnosis of canine *Babesia* species infection; however, it needed a great amount of antigen and is still being improved (Jia *et al.*, 2007).

Enzyme immunosorbent assay (ELISA) is another serological test with fair sensitivity, but low specificity. This assay is suitable to test large number of samples but the poor quality of antigens has limited its use (Reiter and Weiland, 1989). Previously, ELISA had been shown to detect antibodies of the canine *Babesia* species using natural antigens (Boonchit *et al.*, 2006). However, studies have shown that the recombinant antigens are better to be used than the native antigens since recombinant antigens are usually available in pure forms and give higher specific reaction (Boonchit *et al.*, 2006; Jia *et al.*, 2006; Aboge *et al.*, 2007). This prevents the problem of false positive results seen with the native antigen; therefore, the ELISA with the pure form of recombinant antigens could be the next promising serological diagnostic technique (Boonchit *et al.*, 2006).

2.5.3. Molecular-based methods

Molecular techniques are currently used as alternative tools to identify the genes of all types of pathogens (Gasser, 2006). Diagnostic techniques such as, Polymerase Chain Reaction (PCR) and PCR-reverse line blot (RLB), detect DNA from infective agent and differentiate the infecting organisms in a given animal at the same time (Schnittger *et al.*, 2004). Polymerase Chain Reaction has been used to diagnose canine babesiosis over other techniques because it is highly sensitive and specific to detect the target pathogen in peripheral blood (Inokuma *et al.*, 2004). The technique is able to detect parasitemia of 0.0001% and it is used when the parasites are

unidentified on the blood smears but the clinical signs suggest babesiosis (Vannier and Krause, 2009).

There are conventional, nested and semi-nested and real-time PCR machines that are currently commercially available. There is also development of advanced single-step tests with the ability to differentiate the subspecies (Birkenheuer *et al.*, 2003). However, qualitative PCR, such as real-time PCR can be a better option to detect the infection than the conventional, semi- and nested PCR and other sequencing methods because of its superior sensitivity and accurate measurements of specific DNA that is present in the sample (Oyamada *et al.*, 2005). Real-time PCR has advantages over the conventional PCR since it allows high throughput analysis, does not need handling after the amplification, can be used for quantitation and to differentiate amplicons of varying sequences by melting curve analysis (Muller *et al.*, 2002).

Despite the advantages of the above mentioned assays, it would be useful to have a universal test which detect and differentiate all organisms at the same time; and the RLB assay meets these criteria (Gubbels *et al.*, 1999). Originally, this assay was developed to diagnose sickle cell anaemia (Subeki *et al.*, 2007). The Reverse Line Blot (RLB) assay involves hybridization of PCR products to specific probes immobilized on a membrane in order to identify the differences in the amplified sequences (Gubbels *et al.*, 1999). In a single "line" on the membrane, many samples can be assessed against multiple probes to allow detection at the same time (Gubbels *et al.*, 1999). Previous work demonstrated that the RLB is a useful technique to simultaneously identify the presence of *Babesia* and *Theileria* species in a DNA sample (García-Sanmartín *et al.*, 2006; Altay *et al.*, 2007). The RLB assay is extremely practical as one PCR reaction is required to discriminate between species that are difficult to differentiate by microscopic examination of blood smears. In brief, molecular based tests are fast, more economical and clinically efficient to be used in the diagnostic parasitology laboratory (Morgan and Thompson, 1998).

2.6. Control of *Babesia rossi* infections

Soluble parasite antigens (SPA) obtained from various *Babesia* species have been used to immunise animals against infection with these parasites (Schetters *et al.*, 2001). However, immunization does not significantly influence the parasite infection, but diminishes the manifestations of the clinical disease during infection (Finizio *et al.*, 2011). *Babesia canis*, another causative agent of canine babesiosis and *B. rossi*, are two genetically distinct species that cause different clinical signs (Schetters *et al.*, 2001). It has been reported that adding the *B. rossi* SPA to the *B. canis* SPA allows protective immunity against the heterologous *B. canis* infection, suggesting that some antigens that might be important in the development of immunity against canine babesiosis are shared by *B. canis* and *B. rossi* (Schetters *et al.*, 2001).

Another study done by Adaszek *et al.* (2012) reported that the *B. canis* SPA could be safe and tolerated by the dogs. However, efforts to produce a vaccine against *B. rossi* infection using SPA from *B. rossi* culture supernatants did not have favourable outcome (Schetters *et al.*, 2001). Nevertheless, research into developing vaccines for all the *Babesia* species that cause canine babesiosis still continues (Schoeman, 2009).

Dipping or using tick-collars are also one of the effective ways for preventing canine babesiosis in most parts of the world (Schoeman, 2009). Ticks must be attached on the host for three days for transmission to occur; therefore, pet owners are advised to check if their pets have ticks every day. Since blood transfusion is used during treatment when needed, blood donors in areas where the disease is common must be tested all the time to prevent the spread of the disease (Schoeman, 2009).

2.7. *Babesia rossi* genotypes

Canine babesiosis induced by *B. rossi* in South Africa has multiple disease manifestations (Boozer and Macintire, 2003). There is preliminary data that suggests a relationship between the parasite genotype and the disease phenotype (Matjila *et al.*, 2009). *Babesia rossi* synthesizes *B. rossi* erythrocyte membrane antigen 1 (BrEMA1) gene, which was discovered on the cytoplasmic surface of the *B. rossi*-infected red blood cells (B. Carcy, unpublished data). Analysis of the BrEMA1 genes of different strains of *B. rossi* indicated that these genes code for the polymorphic

proteins that contain the various numbers of repetitive hexapeptide motifs (Matjila *et al.*, 2009).

Babesia rossi erythrocyte membrane antigen 1 gene was believed to be present only in this parasite as it was not detected in other *Babesia* spp., *B. vogeli* and *B. canis*, infecting dogs (Matjila *et al.*, 2009). The exact function of *BrEMA1* gene has not been discovered but it was suspected that the gene may have a role in virulence since *B. rossi* is the most pathogenic species of the large *Babesia* of the dogs (Matjila *et al.*, 2009).

In a study by Matjila *et al.* (2009), the *BrEMA1* gene was used as a genetic marker to characterize the *B. rossi* isolates collected in South Africa and to investigate the link between distinct genotypes and the severity of the clinical manifestations of canine babesiosis caused by *B. rossi*. To date, thirteen *B.rossi* genotypes have been identified using *BrEMA1* primers (Figure 4). *Babesia rossi* genotype 19 was reported to be the most virulent and seen in most of the dogs diagnosed with canine babesiosis caused by *B. rossi*. *Babesia rossi* genotype 19 was also the most common genotype identified amongst fatal cases (Matjila *et al.*, 2009).

In another study, *B. rossi* genotype 29 was the most encountered genotype and associated with the severe clinical signs (Malatji, 2011). *Babesia rossi* genotypes 19, 28, 29 and 11 were the most encountered amongst the 13 identified genotypes and *B. rossi* genotypes 19 and 29 were identified in the sick domestic dogs with the most severe clinical cases (Matjila *et al.*, 2009; Malatji, 2011). Matjila *et al.* (2009) reported *B. rossi* 19 as the most virulent due to severe clinical signs observed in the sick dogs, *B. rossi* genotype 29 presented mild clinical signs, whereas *B. rossi* genotype 31 was reported as the least virulent because of the less severe clinical signs observed in the dogs. Hence, the transcriptome of these three *B. rossi* genotypes were sequenced and assembled in this study.

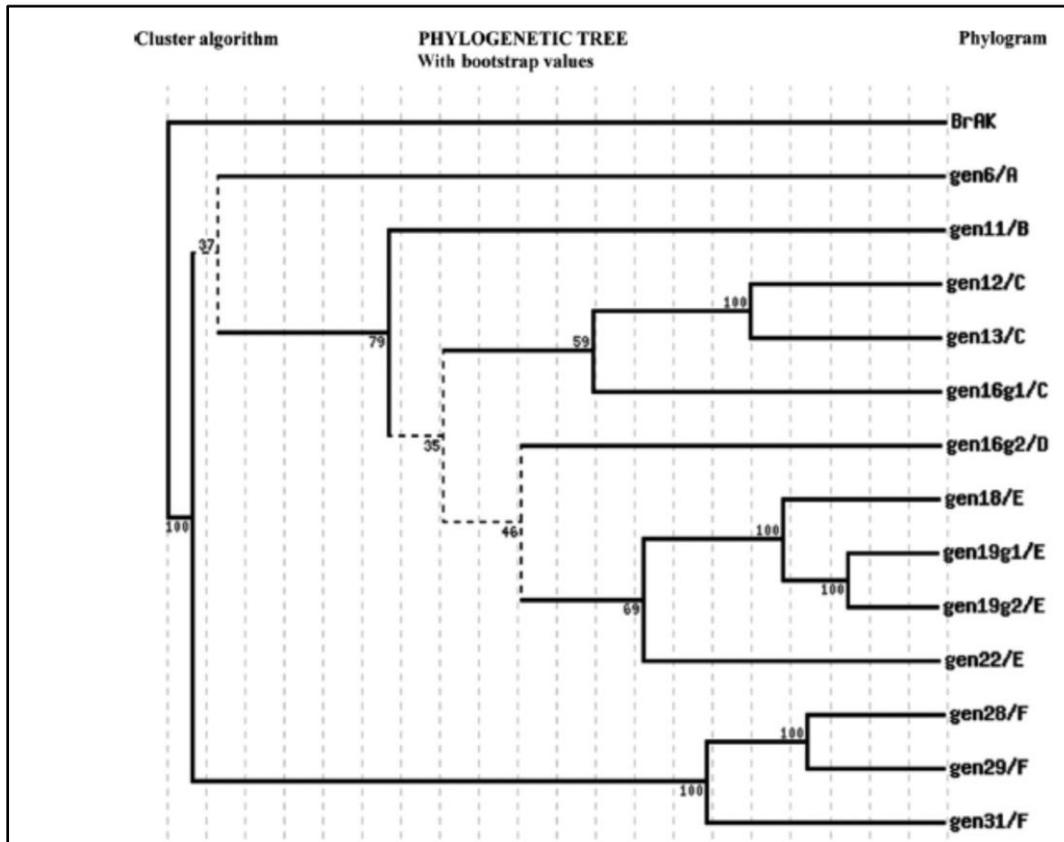


Figure 4. Tree construction showing the phylogenetic relationship between various genotypes based on *BrEMA1* gene sequences. *BrAK* (adenosine kinase gene from *B. rossi*, GenBank accession number AJ223322). Numbers at the nodes indicate bootstrap values (Matjila *et al.*, 2009).

2.8. Transcriptome

2.8.1. Transcriptome sequencing

Transcriptome is described as the full set of transcripts formed by a cell in an organism (Morozova *et al.*, 2009). It is difficult to understand how cells with the same genetic make-up can produce different specific function in the multicellular organism (Saha *et al.*, 2002). This phenotypic variety has been linked to the fact that different cell types within the organism express different sets of genes that lead to different cell functions (Morozova and Marr, 2008). Therefore, it is very important to understand the transcriptomes as they reveal the molecular components of cells and tissues and complex biological processes (Gallardo-Escarate *et al.*, 2014).

Furthermore, it is very crucial to understand the mechanisms by which the blood parasites are transmitted and cause the disease in order to develop effective strategies for control, including the development of vaccines (Brayton *et al.*, 2007). Thus, the sequencing and analysis of transcriptome of apicomplexan parasites, such as *Plasmodium falciparum*, improved the understanding of gene expression during which the host and parasite interact during the infection (Bozdech *et al.*, 2003). Canine babesiosis caused by *B. rossi* and the clinical signs of the human *P. falciparum* malaria have been noted to be similar (Allred and Al-Khedery, 2004), therefore malaria studies can be used as a model for research in canine babesiosis (Malherbe and Parkin, 1951).

There are different sequencing technologies that are used to characterize transcriptomes, including serial analysis of gene expression (SAGE), massively parallel signature sequencing (MPSS) and next generation sequencing (NGS) platforms (Wang *et al.*, 2009). The NGS has revolutionized the transcriptome field by providing the possibility to investigate the transcriptomes in high coverage in a short period of time at a low cost (Morozova *et al.*, 2009). The information generated using this technology provide important details regarding the physiology of different species (Qian *et al.*, 2014). Although, there are improved sequencing methods, the transcriptome of *B. rossi* is still not available.

The benefit of NGS is that it can be used for gene expression profiling of organisms which lack a reference genome, by *de novo* assembly of sequenced reads (Hoffman, 2011). However, the relatively short reads produced by some of the NGS platforms such as Illumina, posed a challenge for effective *de novo* assembly. Nonetheless, advancements in the bioinformatics pipelines have overcome this limitation (Rosenkranz *et al.*, 2008). Consequently, a large number of transcriptomic and genomic sequences have become available in both model and non-model organisms (Eleftherianos *et al.*, 2006). Since *B. rossi* is a non-model organism, NGS-based approaches will be ideal to investigate transcriptome of this pathogen.

2.8.2. *De novo* assembly

RNA sequencing (RNA-Seq) has quickly become a great tool to quantify the transcriptome for different organisms (Martin and Wang, 2011). The transcriptome can be assembled by aligning the RNA sequence reads to the reference genome, provided the research is carried out on a model species which has the reference genome sequence available (Clarke *et al.*, 2013). However, non-model species do not have the genome sequences available; therefore, the transcriptome is constructed and quantified by performing *de novo* assembly (Chen *et al.*, 2011).

Read alignment by *de novo* transcriptome assembly is the initial step of analysis for the organisms without reference genome sequences (Chen *et al.*, 2011). As indicated earlier, the major challenge with *de novo* assembly is the assembly of short reads; hence most *de novo* transcriptome sequencing projects using NGS have previously preferred the 454/Roche system which provides considerably longer read lengths compared to other NGS methods. Furthermore, the speed and accuracy of discovery of new genes have also been a challenge (Gallardo-Escárate *et al.*, 2014). Nevertheless, the *de novo* transcriptome assembly is becoming less challenging because the length of the NGS reads and short read assembly algorithms have since improved to allow efficient assembly of generation of accurate and informative contigs (Zhang *et al.*, 2012).

These improvements have led to development of open source bioinformatics pipelines, such as Trinity and commercial softwares, such as CLC Genomics Workbench (CLC bio), to perform the transcriptome assemblies of the organisms without the reference genome (Robertson *et al.*, 2010). Besides being a useful tool for read alignment for non-model organisms, the *de novo* transcriptome assembly also further enables researchers to investigate genes that are missing from incomplete reference genomes (Chen *et al.*, 2011).

2.9. Functional annotation

Functional annotation is very important in the investigation of the transcriptomic data, particularly for non-model species (Kumar *et al.*, 2014). It assigns a specific or general function on the basis of sequence comparison to proteins in protein sequence databases such as the GenBank, SWISS-PROT, and other similar databases (Ouzounis and Karp, 2002). Gene ontology, a bioinformatics process that has unified the representation of gene and gene product attributes across all species, allows functional characterization of unannotated genes and gene products (Gotz *et al.*, 2008). The ontology covers three domains, namely: biological processes, molecular functions and cellular components. It is very challenging to predict accurate annotations of the transcriptome in species without the available reference genome (Kumar *et al.*, 2014). Hence, functional annotation based on gene ontology terms is usually performed; using different bioinformatics programs or softwares such as Blast2GO (Prasath *et al.*, 2014).

2.10. Problem identification

Canine babesiosis induced by *B. rossi* still remains the cause of high numbers of mortality and morbidity in dogs in South Africa (Collet, 2000). The harmful effect of *B. rossi* is not only restricted to the health issues of the dog but has a negative impact on the economy as well. For example, canine babesiosis caused by *B. rossi* costs the dog owners R20 million per year in diagnosis and treatment in South Africa (Collet, 2000). Hence, understanding the mechanisms leading to the range of complicated clinical syndromes development is very important for controlling parasite transmission and the disease. Unfortunately, critical this may be, the research at molecular level is limited due to the absence of *B. rossi* genome and transcriptome sequence data.

At present, *B. rossi* transcriptome data is practically non-existent in public databases. Therefore, there is no reliable database, which can be used to retrieve information that can be useful in studying the molecular mechanisms behind canine babesiosis caused by *B. rossi* and the complicated clinical symptoms presented in their hosts. Given the negative impact infections by this parasite has on the host, there is an urgent need to generate *B. rossi* transcriptome data.

2.11. Study aim

The aim of this study is to sequence, assemble and annotate *B. rossi* transcriptome in dogs diagnosed with canine babesiosis.

2.12. Research objectives

The objectives of this study are to:

- (i) Sequence the whole transcriptome of virulent *B. rossi* genotypes 19, 29 and 31, obtained from dogs naturally infected with canine babesiosis.
- (ii) Perform *de novo* assembly of transcriptomes genotypes 19, 29 and 31 sequence reads.
- (iii) Perform functional annotation (gene ontology) on unigenes from the transcriptome assembly of genotypes 19, 29 and 31.

CHAPTER 3

3. METHODOLOGY

3.1. Sample collection

Ethical clearance (reference number: 2014/CAES/042) was granted by the UNISA Ethics Committee prior to resuming with the research activities of the study. Ethical clearance had also been obtained from the University of Pretoria Animal Ethics Committee through Prof Andrew Leisewitz for his parallel study on canine babesiosis, and he provided blood samples from sick domestic dogs for the current study. Forty clinical blood samples diagnosed with canine babesiosis, based on the microscopic examination of blood smears at the Onderstepoort Veterinary Academic Hospital (OVAH), Faculty of Veterinary Science, University of Pretoria, were made available. Blood was collected from the cephalic vein into the EDTA tubes and immediately stored at 4°C refrigerator prior to DNA and RNA extractions. Two tubes of clinical blood samples were collected from each dog for DNA and RNA extractions.

3.2. DNA extraction and quantification

DNA was extracted from each of the 20 clinical blood samples collected from the sick dogs using the QIAmp blood and tissue extraction kit (QIAGEN), according to the manufacturer's protocol. Approximately 200 µl of each blood sample was added to 20 µl QIAGEN Proteinase into a 1.5 ml microcentrifuge tube. Subsequent to this, the solution was mixed properly by vortexing, followed by the addition of 200 µl Buffer AL. The mixture was then mixed by vortexing for 15 sec and incubated at 56°C for 10 min to allow sample lysis. The tube contents were then briefly centrifuged to remove drops from the inside of the lid. Two hundred (200) µl of ethanol (96-100%) was added to the sample mixture and mixed by vortexing for 15 sec, followed by brief centrifugation to remove drops from the inside of the lid. Subsequently, the sample mixture was carefully transferred to the QIAamp mini spin column, in a 2 ml collection tube, without wetting the rim. The assembly was then centrifuged at 6 000 xg for 1 min. The QIAamp mini spin column was placed in another clean 2 ml collection tube and the collection tube containing the filtrate was discarded. The mini spin column was carefully opened and 500 µl of Buffer AWL1 was added without

wetting the rim. The tube was centrifuged again at 6 000 xg for 1 min. The QIAamp mini spin column was placed in another clean 2 ml collection tube. The collection tube containing the filtrate was discarded. The QIAamp mini spin column was opened with care and 500 µl of Buffer AW2 was added, making sure the rims are not wet. The tube was centrifuged at 20 000 xg for 3 min. The QIAamp mini spin column was placed in a new 2 ml collection tube and centrifuged at full speed for further 1 min. This step was performed to eliminate the chance of possible AW2 carryover. The flow-through in the collection tube was discarded. For DNA elution, the QIAamp mini spin column was placed in a 1.5 ml microcentrifuge tube and the collection tube containing the flow-through was discarded. Two hundred (200) microlitres of Buffer AE was added to the QIAamp mini spin column, incubated at room temperature (15-25°C) for 1 min and then centrifuged at 6 000 xg for 1 min. DNA contamination between the samples was minimized by following strict precautions during the extractions. Eluted DNA samples were collected and stored in the collection tubes.

One µL of each extracted DNA samples was used to determine the quality of the DNA in each sample and to check the success of the extraction procedure. The DNA concentration was evaluated by measuring the absorbance of light at 260nm using the Nanodrop® ND-1000 Spectrophotometer (Thermo Scientific). The remaining DNA samples were stored at -20°C until further use.

3.3. Screening of DNA samples for hemoparasite infection using the Reverse Line Blot Hybridization Assay

The detection of nucleic acids of specific species of *Babesia*, *Theileria*, *Anaplasma* and *Ehrlichia* on all the DNA samples was performed using Polymerase Chain Reaction (PCR) technique, followed by Reverse Line Blotting (RLB) assay, conducted at the Department of Veterinary and Tropical Diseases, University of Pretoria. *Babesia rossi* positive DNA sample was included in all the RLB reactions to check whether the PCR products were specific and hybridized to the specific DNA probes on the RLB membrane (Matjila *et al.*, 2005). The RLB assay was performed as described by Gubbels *et al* (1999).

Polymerase Chain Reaction (PCR): PCR was performed as described below: A forward primer, RLB-F2 (5'- GAC ACA GGG TAG TGA G-3') and the reverse primer,

RLB-R2 (biotin-5'- CTA AGA ATT TCA CCT CTG ACA GT -3') were used for amplification of a 460 to 520 bp of the 18S SSU rRNA spanning the V4 hypervariable region. These set of primers was also hybridized with regions conserved for *Theileria* and *Babesia* (Matjila *et al.* 2004). The PCR for *Ehrlichia/Anaplasma* was conducted with the forward primer Ehr-F (5'-GGA ATT CAG AGT TGG ATC MTG GYT CAG-3') and Ehr-R (5'-biotin-CGG GAT CCC GAG TTT GCC GGG ACT TYT TCT-3') to amplify a 460-520 bp fragment from the V1 hypervariable region of the 16S SSU rRNA gene. The PCR reaction consisted of the following: 12.5 µl of the PCR Supermix-UDG (Invitrogen, The Scientific Group), 20 pmol (0.25 µl) of each primer, and 2.5µl of DNA to a total volume of 25 µl. The Gene Amp® PCR System 9700 and 2720 thermal cyclers (Applied Biosystems) were used for amplification. The PCR conditions were set as follows: Initial step of 3 min at 37°C, 10 min at 94°C, and 10 cycles at 94°C for 20 sec, 67°C for 30 sec and 72°C for 30 sec, with lowering of the annealing step by 2°C after every second cycle (Touchdown PCR), followed by 40 cycles of denaturation at 94°C for 30 sec and extension at 72°C for 30 sec.

Membrane preparation: All the oligonucleotide probes listed in Table 2 were diluted with 150 µl of 0.5 M NaHCO₃ (pH 8.4). The membrane was incubated for 10 min in 16% EDTA at room temperature, washed for two minutes with distilled water and placed in a miniblottedter. The probes were then loaded into the lanes of the miniblottedter so that they can be covalently linked to the membrane, followed by one minute of incubation. The probes were aspirated and the membrane inactivated by incubating it in 100 ml of 100 mM NaOH solution for 10 minutes at room temperature. The membrane was washed in the SSPE-0.1% sodium dodecyl sulphate (SDS) solution by shaking for 5 min at 60°C.

Reverse line blot hybridization: Prior to use, the membrane was washed for 5 min at 42°C with 125 ml 2X SSPE-0.1% SDS, and then placed in the miniblottedter with the slots perpendicular on the previously applied probes. A volume of 25 µl of PCR product was diluted to an end volume of 150 µl 2X SSPE-0.1% SDS. The PCR products were heated for 10 min at 100°C in a thermocycler, put on ice immediately to cool, and centrifuged. The denatured PCR products were applied into the slots and incubated for 60 min at 42°C. The PCR products were also aspirated and the blots were washed twice in 125 ml of 2X SSPE-0.1% SDS for 10 min at 42°C by

shaking. The membrane was incubated for 60 min at 42°C for hybridization to take place. Samples were discarded by aspiration; thereafter the membrane was removed from the blotter. The membrane was washed twice in preheated 2 X SSPE/0.5% SDS for 10 min at 50°C in a water bath with gentle shaking. The membrane was then incubated with 10 ml 2 X SSPE/0.5% SDS + 2.5 µl streptavidin-POD (peroxidase-labeled) conjugate (Roche Diagnostics, South Africa) (1.25 U) for 30 min at 42°C. The membrane was washed twice again in preheated 2 X SSPE/0.5% SDS for 10 min at 42°C in water bath with shaking. The membrane was washed again twice with 2 X SSPE for 5 min at room temperature with shaking. The membrane was then incubated for one minute in 10 ml of ECL detection fluid (DNA Thunder™, Perkin Elmer, Separation Scientific). The membrane was covered with overhead sheet and air bubbles were removed by rolling a tube over the sheet. It was placed between two clean overhead sheets and placed in the exposure cassette. The X-ray film (X-OMAT™ Blue XB-1, Kodak, Separation Scientific) was exposed for 5-20 min and then developed for detection of hybridized PCR products, which were visualized by chemiluminescence. The film was placed in a grid and each sample lane correlated with the DNA probes. After use, the PCR products were stripped from the membrane by two washes for 30 min each time at 80°C in 1% SDS solution. The membrane was rinsed in 20 mM EDTA (pH 8.0) for 15 min and then stored in fresh 20 mM EDTA solution at 4°C for reuse.

Table 2: List of probes and their corresponding sequences used to identify the pathogen DNA on the RLB.

Probes	Probe sequence (5'- to 3'-end)
<i>Ehrlichia/Anaplasm</i> genus-specific	GGG GGA AAG ATT TAT CGC TA
<i>Anaplasma centrale</i>	TCG AAC GGA CCA TAC GC
<i>Anaplasma marginale</i>	GAC CGT ATA CGC AGC TTG
<i>Anaplasma phagocytophilum</i>	TTG CTA TAA AGA ATA ATT AGT GG
<i>Anaplasma bovis</i>	GTA GCT TGC TAT GRG AAC A
<i>Ehrlichia ruminantium</i>	AGT ATC TGTBTAG TGG CAG
<i>Ehrlichia chaffeensis</i>	ACC TTT TGG TTA TAA ATA ATT GTT
<i>Ehrlichia.sp. omatjenne</i>	CGG ATT TTT ATC ATA GCT TGC
<i>Ehrlichia canis</i>	TCT GGC TATBAGG AAA TTG TTA
<i>Theileria/Babesia</i> genus specific	TAA TGG TTA ATA GGA RCR GTT G
<i>Theileriagenus</i> specific	ATT AGA GTG CTC AAA GCA GGC
<i>Babesia</i> genus specific 1	ATT AGA GTG TTT CAA GCA GAC
<i>Babesia</i> genus specific 2	ACT AGA GTG TTT CAA ACA GGC
<i>Babesia felis</i>	TTA TGC GTT TTC CGA CTG GC
<i>Babesia divergens</i>	ACT RAT GTC GAG ATT GCA C
<i>Babesia microti</i>	GRC TTG CGA TCW TCT GGA
<i>Babesia bigemina</i>	CGT TTT TTC CCT TTT GTT GG
<i>Babesia bovis</i>	CAG GTT TCG CCT GTA TAA TTG AG
<i>Babesia rossi</i>	CGG TTT GTT GCC TTT GTG
<i>Babesia canis</i>	TGC GTT GAC CGT TTG AC
<i>Babesia vogeli</i>	AGC GTG TTC GAG TTT GCC
<i>Babesia major</i>	TCC GAC TTT GGT TGG TGT
<i>Babesia bicornis</i>	TTG GTA AAT CGC CTT GGT C
<i>Babesia caballi</i>	GTG TTT ATC GCA GAC TTT TGT
<i>Babesia gibsoni</i>	CAT CCC TCT GGT TAA TTT G
<i>Babesia leo</i>	ATC TTG TTG CCT TGC AGC T
<i>Babesia sp (sable)</i>	GCT GCA TTG CCT TTT CTC C

<i>Theileria</i> sp (kudu)	CTG CAT TGT TTC TTT CCT TTG
<i>Theileria</i> sp (sable)	GCT GCA TTG CCT TTT CTC C
<i>Theileria bicornis</i>	GCG TTG TGG CTT TTT TCT G
<i>Theileria annulata</i>	CCT CTG GGG TCT GTG CA
<i>Theileria buffeli</i>	GGC TTA TTT CGG WTT GAT TTT
<i>Theileria</i> sp. (buffalo)	CAG ACG GAG TTT ACT TTG T
<i>Theileria mutans</i>	CTT GCG TCT CCG AAT GTT
<i>Theileria parva</i>	GGA CGG AGT TCG CTT TG
<i>Theileria taurotragi</i>	TCT TGG CAC GTG GCT TTT
<i>Theileria velifera</i>	CCT ATT CTC CTT TAC GAG T
<i>Theileria equi</i>	TTC GTT GAC TGC GYT TGG
<i>Theileria lestoquardi</i>	CTT GTG TCC CTC CGG G
<i>Theileria ovis</i>	TTG CTT TTG CTC CTT TAC CAG
<i>Theileria annae</i>	CCG AAC GTA ATT TTA TTG ATT TG
<i>Theileria separata</i>	GGT CGT GGT TTT CCT CGT

3.4. Genotyping by qualitative Polymerase Chain Reaction (PCR)

Amplification of the *BrEMA1* genotype in the positive *B. rossi* DNA samples was done using the Light cycler 2.0 (Roche, South Africa). The primers Frep*BrEMA1* (5'-CCA ACA TTG ATG ATG ACA A-3') and Rrep*BrEMA1* (5'-CTG CAT GTC AGC TTC ATC A-3') are specifically designed to target the 18-nucleotide repetitive sequence region of the *BrEMA1* gene (accession number AJ416994). This region has previously been used to define the *B. rossi* genotypes based on the number of 18-nucleotide repetitive sequences, which ranges from 16 to 31, resulting in PCR product fragments ranging from 375 to 645 bp in size. The template of 2.5 µL of each DNA sample was used in a total volume of 20 µL. The extracted DNA infected with *BrEMA1* genotype was used as positive control and nuclease free H₂O as negative control. The PCR reaction mixture contained the following: dH₂O (11.5 µL), one µL of the forward primer (10 µM), one µL of reverse primer (10 µM), Universal SYBGreen Supermix (BioRad) (4 µL). The cycling conditions were as follows: pre-incubation cycle at 95°C for 10 min, amplification cycle at 95°C for 10 sec, 55°C for 10 sec and

72°C for 15 sec, melting curves at 95°C for 0 sec, 65°C for 15 sec and 95°C for 0 sec and cooling cycle at 40°C for 30 sec. Fluorescence values were measured at 530nm.

3.5. DNA sequencing and analysis

Twenty PCR products obtained from the amplification of *BrEMA 1* gene were sent to Inqaba-Biotec, Pretoria, South Africa for DNA sequencing. The PCR amplicons were sequenced directly in both the forward and reverse directions with the Frep*BrEMA1* (forward) and Rrep*BrEMA1* (reverse) primers using the ABI Big dye termination V3.1 Kit, on the ABI 3500XL genetic analyzer (Applied Biosystems), according to manufacturer's instructions. The obtained sequences were assembled and edited using preGap and Gap (Genome Assembly Programme) 4 of Standen Package (standen.sourceforge.net). The sequences were then compared with other related genera using BLAST (Basic Local Alignment Search Tool) (www.ncbi.nlm.nih.gov/BLAST/).

3.6. RNA extraction and quality assessment

Twenty (20) blood samples were collected in RNAprotect® Animal Blood Tubes, from the same 20 sick domestic dogs investigated by RLB hybridization assay and DNA sequence analysis. Total RNA was extracted from the clinical blood samples of the sick dogs, using the RNeasy® Protect Animal Blood Kit (QIAGEN) according to the manufacturer's instructions as follows: The RNAprotect® Animal Blood Tubes were centrifuged for 3 min at 5 000 xg. The supernatant was removed by decanting or pipetting, making sure that the pellet was not disturbed. One millilitre of RNase-free water was added to the pellet and the tube was closed. The tube was mixed by vortexing until the pellet dissolved and centrifugation followed for 3 min at 5 000 xg. The entire supernatant was removed by decanting and discarded. Two-hundred and forty (240) µl of Buffer RSB was added to the pellet. The pellet was mixed using vortex mixer until it was dissolved. The sample was pipetted into a 1.5 ml collection tube. Two-hundred (200) µl of Buffer RBT and 20 µl of Proteinase K were added. The mixture was mixed by vortexing for 5 sec and incubated for 10 min at 55°C in a shaker-incubator. The sample was pipetted into a QIAshredder (lilac) spin column in a 2 ml collection tube. After pipetting, the sample was centrifuged for 3 min at full speed (20 000 xg). The entire flow-through was transferred to a new 1.5 ml tube. Two hundred and forty microliters of ethanol (96-100%) was added to the flow-

through and mixed by vortexing. The sample was transferred into the RNeasy MinElute spin column (pink) and placed in a 2 ml collection tube. The sample was centrifuged for one minute at $\geq 8\ 000\ \text{xg}$. The filtrate was discarded and the RNeasy MinElute spin column was placed in a 2 ml collection tube. Three hundred and fifty (350) μl of Buffer RW1 was added to the RNeasy MinElute spin column, followed by centrifugation for 15 sec at $\geq 8000\ \text{xg}$. Ten microlitres of DNase1 stock was added to 70 μl of Buffer RDD in a 1.5 ml microcentrifuge. The solution was mixed by gently inverting the tube and centrifuged briefly to collect the residual liquid from the sides of the tube. Eighty microliters of the DNase I mixture was added to the RNeasy MinElute spin column, followed by incubation for 15 min on the bench top. After incubation, 350 μl of Buffer RW1 was added to the RNeasy MinElute spin column, followed by centrifugation for 15 min at $\geq 8\ 000\ \text{xg}$. Then 350 μl of Buffer RPE was added to the RNeasy MinElute spin column, followed by centrifugation for 15 sec at $\geq 8\ 000\ \text{xg}$. The flow-through was discarded and the RNeasy MinElute spin column was placed in a 2 ml collection tube. Five-hundred (500) μl of ethanol (80%) was added to the RNeasy MinElute spin column, followed by centrifugation for two minutes at $\geq 8000\ \text{xg}$. After centrifugation, the RNeasy MinElute spin column was placed in a new 2 ml collection tube. The lid of the spin column was opened and the spin column was centrifuged at full speed for five minutes. The filtrate in the collection tube was discarded. The RNeasy MinElute spin column was placed in a 1.5 ml collection tube and 14-30 μl of the Buffer REB was pipetted directly onto the spin column membrane. Then, centrifugation for 1 min at $\geq 8000\ \text{xg}$ was done for RNA elution. The RNA eluate was incubated for five minutes at 65°C . After incubation, the RNA eluate was put on ice. All total RNA samples were stored at -4°C until further use.

The quality of the RNA samples was determined prior to RNA sequencing. Firstly, one microliter of each RNA samples was used to evaluate the quality by using the Nanodrop® ND-1000 Spectrophotometer (Thermo Scientific) at an absorbance of 260/280 nm at the Department of Veterinary and Tropical Diseases, University of Pretoria. Six total RNA samples (dog 5 and 17 representing *B. rossi* genotype 19; dog 8 and 13 representing *B. rossi* genotype 28; dog 20 representing *B. rossi* genotype 29 and dog 16 representing *B. rossi* genotype 31) (Table 7) were selected based on the severity of the clinical signs observed on the sick dogs, and four

microliters of each six total RNA samples was sent to ACGT Microarray Facility and Ion Torrent Sequencing Facility, at the University of Pretoria for the second assessment on the concentration, integrity number (RIN) and the 28s/18s ratio of the RNA samples using the Agilent BioAnalyzer (Agilent Technologies). The rest of each total RNA samples were stored at -80°C for RNA Seq.

3.7. Transcriptome sequencing

Samples for transcriptome sequencing were selected based on the diversity of the clinical signs observed in the host, represented by three *B. rossi* genotypes which were established through DNA sequencing of the *BrEMA1* gene. *Babesia rossi* genotypes 19 caused the most severe clinical signs in the hosts; *B. rossi* genotype 29 presented the moderately severe clinical manifestations in the sick domestic dogs while *B. rossi* genotype 31, showed the less severe clinical syndromes in the hosts in our study. Therefore, the three total RNA samples infected with *B. rossi* genotypes 19 (dog 17), 29 (dog 20) and 31 (dog 16) (Table 8) were prepared from three dogs showing clinical signs associated with each of the three genotypes, to fulfil the standard sample requirements for transcriptome sequencing at the Beijing Genomics Institute (BGI) in Hong Kong, namely: mass \geq 2 μ g, concentration \geq 20 ng, 28S/18S ratio \geq 1 and RIN \geq 6.5. The RNA samples were submitted to Beijing Genomics Institution, Hong Kong for library construction and transcriptome sequencing. Prior to library construction, the final assessment on the total RNA quality (concentration, RIN, 28S/18S) was performed by using the Agilent 2100 Bioanalyzer (Agilent RNA 6000 Nano Kit); the purity of the total RNA samples was also confirmed by NanoDrop™.

3.7.1. Library construction and validation

Briefly, the three libraries from *B. rossi* genotypes 19, 29 and 31 were prepared using the Illumina TruSeq RNA Sample Preparation kit v2 (Illumina, USA) at BGI, Hong Kong as follows:

mRNA isolation and fragmentation: Two hundred (200) ng of total RNA was purified by oligo-dT beads, then poly (A)-containing mRNA was fragmented into small pieces with Elute/ prime/ fragment mix.

cDNA synthesis: First-strand cDNA was generated from fragmented mRNA using the first strand master mix and superscript II reverse transcriptase (Invitrogen) at 25°C for 10 min; 42°C for 50 min and 70°C for 15 min. Subsequently, the second strand master mix was added to synthesize the second strand cDNA at 16°C for one hour.

End repair, A-tailing and adaptor ligation: The purified fragmented cDNA was combined with the end repair mix and incubated at 30°C for 30 min. The end-repaired DNA was purified with Ampure XP Beads (Agencourt). The A-tailing mix was then added and the solution was mixed by pipetting before incubation at 37°C for 30 min.

Adapter ligation: In this step, the adenylate 3' ends DNA, RNA Index adapter and ligation mix were combined and mixed well by pipetting. The ligation reaction was incubated at 30°C for 10 min and the end-paired DNA was purified with Ampure XP Beads (Agencourt).

Polymerase Chain Reaction (PCR): To enrich the cDNA fragments, PCR was performed using PCR primer cocktail and PCR master mix. The PCR products were then purified with Ampure XP Beads (Agencourt).

Library quantitation and sequencing: The average molecule length was determined using the Agilent 2100 Bioanalyzer instrument (Agilent DNA 1000 Reagents); and the library was quantified by real-time quantitative PCR (qPCR) (TaqMan Probe). The qualified libraries were amplified on cBot to generate the cluster on the flow cell (TruSeq PE Cluster Kit V3-cBot-HS, Illumina) and sequenced paired-end reads on the HiSeq® 2000 system (TruSeq SBS KIT-HS V3, Illumina).

3.7.2. Pre-processing of sequence data and data delivery

Post transcriptome sequence analysis, raw reads which contained adaptors, unknown nucleotides larger than 5% and low quality reads were removed from the transcriptome data using the filterFQ software. The clean raw reads were subsequently used for downstream bioinformatics analysis.

3.7.3. *De novo* transcriptome assembly

Transcriptome *de novo* assembly of the three *B. rossi* genotypes was performed with short reads assembling program, the Trinity software package (release 2013-02-25) as described in Grabherr *et al.*, 2011. Trinity is a program which incorporates three separate software modules, namely; Inchworm, Chrysalis and Butterfly.

In brief: In the first step, the RNA-Seq data is assembled into unique sequences of transcripts using Inchworm, to produce full length-transcripts for a dominant isoform, but reports only the unique part of the alternatively spliced transcripts. In the second step, the Chrysalis clusters the Inchworm contigs into clusters and creates complete de Bruijn graph for each cluster. Each cluster represents the transcriptional complexity for a given gene. In the final step, the Butterfly processes the separate graphs in parallel, tracing the path that reads and pairs of reads take within the graph, eventually reporting full length transcripts for alternatively spliced isoforms, and teasing apart transcripts that corresponds to paralogous genes.

The longest transcript among similar transcripts, representing the same gene, was considered as the unigene. Since three total RNA samples of different *B. rossi* genotypes were sequenced in this study, the unigenes from each genotype were further subjected to sequence splicing. Subsequently, the redundancies were discarded using sequence clustering software embedded in Trinity to obtain non-redundant unigenes.

3.8. Functional annotation of the *Babesia rossi* transcriptome

The unigenes obtained from *B. rossi* genotype 19, 29 and 31 were used as queries in BLASTx against the NCBI (full) non-redundant protein database (nr) with expectation (E) value of 1.0E-3 and 20 maximum number of hits applied to acquire desired output (Altschul *et al.*, 1990). The gene identifier (GI) numbers of the gene sequences matching specific unigenes from each genotype, according to the set parameters, were retrieved and uploaded into the Batch Entrez program (<http://www.ncbi.nlm.nih.gov/sites/batchentrez>) to remove the possible duplicates. Subsequently, unigenes from *B. rossi* genotypes 19, 29 and 31, were uploaded on Venn diagram (<http://bioinformatics.psb.ugent.be/webtools/Venn/>) to identify the

unigenes unique to each *B. rossi* genotype and those common between the three genotypes.

Functional annotation of unigenes from each *B. rossi* genotype was performed on the Blast2GO program, version 2.8.0 software. In Blast2GO, the fasta sequences of the unigenes were used as queries in BLASTx against the NCBI (full) non-redundant protein database (nr), with expectation (E) value of 1.0E-3 and the maximum number of hits set at 20n (Altschul *et al.*, 1990), followed by mapping, annotation and Interproscan analysis.

During mapping, the Gene Ontology (GO) terms were retrieved for each blast hit. Gene Ontology terms described the biological processes, molecular functions and cellular components. The annotation step assigned the annotation to the query sequence, thus providing a better understanding on the distribution of gene functions. Level 2 of gene ontology was employed to obtain reasonable coverage and useful term specificity.

3.9. Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway analysis

The KEGG (Kyoto Encyclopedia of Genes and Genomes) database (<http://www.genome.jp/kegg/mapper.html>) was used to annotate the biological pathways of all unigenes from the three *B. rossi* genotypes. The fasta sequences of the unigenes were used as query sequences. The BlastKOALA server was used to assign K numbers to a given sequence data set for consecutive analysis using the Reconstruct Pathway tool.

3.10. Cluster of Orthologous Groups (COG) Classification analysis

The unigenes from the three *B. rossi* genotypes were further analysed into gene families using DAVID (the Database for Annotation, Visualization and Intergrated Discovery version 6.7 (<https://david.ncifcrf.gov>), to predict and classify their functions. The unigene GI number list was converted into DAVID identifier using the Gene ID Conversion Tool within DAVID. The converted unigene list was resubmitted into the DAVID for Cluster of Orthologous Group (COG) classification.

CHAPTER 4

4. RESULTS

4.1. DNA quality assessment

The DNA samples had a concentration ranging from -0, 22 and 898, 57ng/μl (Table 3 in the Appendix). The DNA concentration greatly varied among the samples. The DNA samples had the 260/280 values ranging between -20 and 2.07 (Table 3 in the Appendix). The DNA samples with the negative concentration values and 260/280 ratios were included in the downstream DNA analysis.

4.2. Detection of *Babesia rossi* DNA by Reverse Line Blot hybridization assay

The RLB assay results confirmed the presence of *B. rossi* parasite in 18 (90%) of the 20 DNA samples (Table 4 and Figure 5 in the Appendix). There was no nucleic acid detected in the other two (10%) DNA samples, which could be due to the presence of very low *B. rossi* parasite DNA in those DNA samples. There was no co-infection of *B. rossi* with *Anaplasma* and *Ehrlichia* in all the positive RLB samples. The negative control did not hybridize to any probe included in the blot.

4.3. Qualitative PCR analysis for detection of *Babesia rossi* DNA from dog blood samples

The Frep*BrEMA1* and Rrep*BrEMA1* primers successfully amplified the *BrEMA1* gene in all 20 extracted DNA samples, including the two DNA samples in which the nucleic acid could not be detected in the RLB assay, and the DNA samples with negative concentration values. All the positive *B. rossi* DNA samples melted at a temperature between 72°C and 83°C, with most of the samples melting at around 80°C (Table 5 and Figure 6A, 6B in the Appendix). There was no amplification observed in the negative sample.

4.4. DNA sequence analysis

Eighteen out of 20 samples passed the quality control test for DNA sequencing. The *BrEMA1* gene sequences retrieved from those samples matched the 13 corresponding sequences already available at the NCBI database. Four different *BrEMA1* genotypes, namely, *B. rossi* genotype 19 (FM164400), *B. rossi* genotype 28

(FM164403), *B. rossi* genotype 29 (FM164404) and *B. rossi* genotype 31 (FM164405), with 96-100% identity to the query sequences, were identified. *Babesia rossi* genotype 28 was the most encountered, followed by *B. rossi* genotype 19, then *B. rossi* genotypes 29 and 31 (Figure 7).

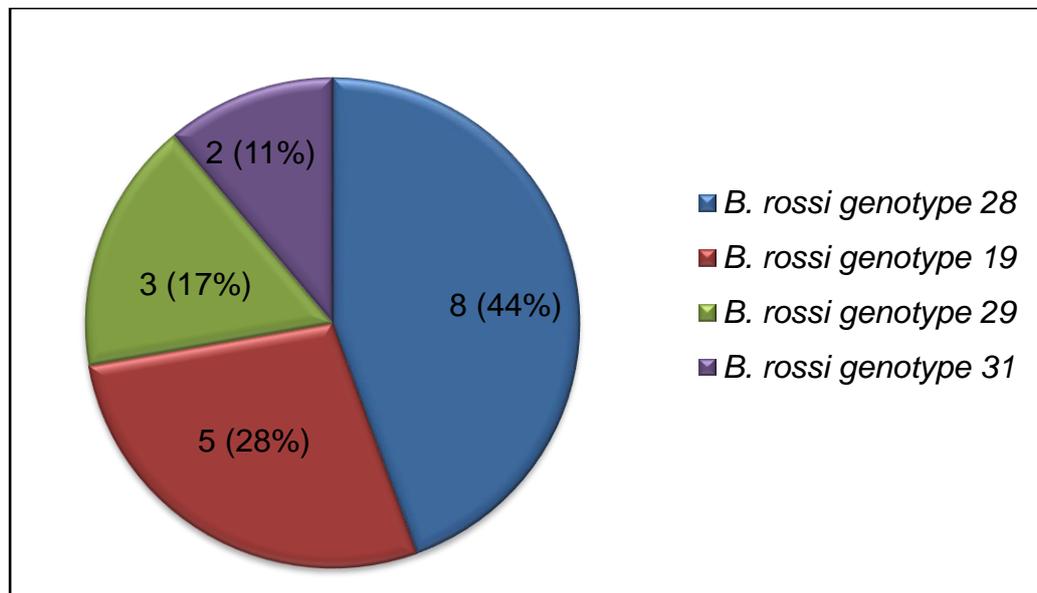


Figure 7: Identification of four *B. rossi* genotypes in the positive *B. rossi* blood samples collected from sick domestic dogs.

4.5. RNA quantity and quality

Total RNA yields obtained from the 20 blood samples infected with *B. rossi* genotypes 19, 28, 29 and 31 ranged between -5,414 and 284,982 ng/ μ l and the 260/280 ratio values ranged from 1.43 to 18.66. (Table 6 in the Appendix). Six total RNA samples were selected based on their concentration values and the 260/280 ratio (Table 6 in the Appendix) for reassessment of RNA quality and quantity. The concentration values differed significantly between the six different RNA samples, ranging from 44.66ng/ μ l to 409.48ng/ μ l. The 28S/18S ratio of 2 is acceptable and indicates intact RNA (www.thermofischer.com). However, the 28S/18S ratio in the six selected RNA samples ranged from 1.05 to 1.28 (Table 7). The scope of the RNA integrity number (RIN) value ranges from 1 to 10, with 1 being bad quality and 10 being the good quality (Fleige and Pfaffl, 2006). Thus, the RIN of the different RNA samples was between 7 and 9 (Table 7), indicating that they were of good quality to be used in transcriptome sequencing.

Table 7: RNA quality and quantity of the samples infected with different *B. rossi* genotypes using BioAnalyzer at the University of Pretoria, South Africa.

Sample number	<i>B. rossi</i> genotype detected	RNA concentration (ng/μL)	RIN	Ratio (28S/18S)
1 (dog 5)	19	86.450	9.1	1.09
2 (dog 8)	28	409.48	7.9	1.22
3 (dog 28)	28	44.660	9.1	1.28
4 (dog 16)	31	132.73	8.9	1.20
5 (dog 17)	19	235.63	9.1	1.16
6 (dog 20)	29	227.43	9.3	1.05

The quality of the three selected RNA samples, representing each genotype, was also re-assessed at BGI prior to transcriptome sequencing. Sample number 4 (representing *B. rossi* genotype 31) had a lower RNA concentration and total RNA mass compared to its counterparts. However the same sample had higher RIN and 28S/18S ratio than sample number 5 and 6 (representing *B. rossi* genotypes 19 and 29, respectively). Below is the report of the quality and quantity of the RNA quality check analysis.

Table 8: RNA testing report using the Agilent 2100 and Nanodrop at BGI, Hong Kong.

Sample number	<i>B. rossi</i> genotype detected	RNA concentration (ng/μL)	Total Mass (μg)	RIN	Ratio (28S/18S)
4 (dog 16)	31	50	0.7	9.7	1.9
5 (dog 17)	19	170	2.38	8.8	1.3
6 (dog 20)	29	172	2.41	9.3	1.8

4.6. *De novo* transcriptome assembly statistics of *Babesia rossi* genotypes 19 and 29

The transcriptome of the three genotypes was sequenced, but the sequencing was only successful for *B. rossi* genotypes 19 and 29. From the quality check results at BGI, RNA samples number 5 (*B. rossi* genotype 19) and 6 (*B. rossi* genotype 29) had a good total RNA mass compared to RNA samples number 4 (*B. rossi* genotype 31), which had a poor RNA total mass. Because of this, the quality of the library for RNA samples number 4 (*B. rossi* genotype 31) was likely compromised (Johnson *et al.*, 2012). Consequently; the transcriptome data from this genotype was excluded from the analysis.

Babesia rossi genotype 19 generated a high number of clean paired-end reads, with a slightly high GC content and Q20 compared to genotype 29. However, *de novo* assembly analysis of genotype 19 generated less numbers of contigs with a short mean length and N50 compared to its counterpart. Following the clustering of contigs into unigenes and exclusion of duplicate unigenes, the number of unigenes generated in genotype 19 was increased significantly, but with a shorter mean length

than genotype 29. Both genotypes had the same unigene N50 value subsequent to this analysis (Table 9).

Table 9: Summary statistics of transcriptome assemblies of *B. rossi* genotypes 19 and 29.

	<i>B. rossi</i> genotype 19	<i>B. rossi</i> genotype 29
Total number of clean paired-end reads	26 747 238	25 709 672
GC content	62.63%	61.18%
Q20	94.69%	94.49%
Total number of contigs	50 839	62 308
Mean contig length(bp)	225	230
Contig N50	250	258
Total number of unigenes	3 019	2727
Unigene mean length (bp)	419	441
Unigene N50	362	362

4.7. Pairwise analysis of *Babesia rossi* genotypes 19 and 29 transcriptomes.

A comparative analysis of the unigenes from *B. rossi* genotypes 19 and 29 revealed that a total of 1193 unigenes were common in the two genotypes (Figure 9), which consisted of 39.5% of unigenes detected from genotype 19 and 43.7% of unigenes in genotype 29. A total of 1826 (60.5 %) and 1534 (56.3%) unigenes were exclusively detected in *B. rossi* genotype 19 and 29, respectively. Between the two genotypes, a total of 4553 unigenes were obtained and will hereon be referred to as combined unigenes (total number of unigenes detected for the *B. rossi* transcriptome) for the purpose of this dissertation.

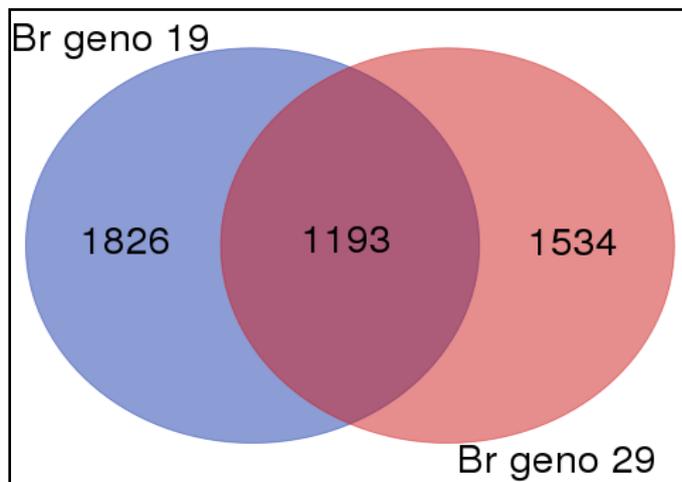


Figure 8: Venn diagram showing the distribution of the common and exclusively detected unigenes between the *B. rossi* genotypes 19 and 29. Numbers inside the coloured circles indicate the number of unigenes. Br geno 19= *B. rossi* genotype 19; Br geno 29= *B. rossi* genotype 29.

4.8. Length distribution of the unigenes detected in *Babesia rossi*

transcriptome.

Overall, the number of unigenes decreased with an increasing length (101-1000bp) (Figure 10). Only 4.8% (n=220) of detected unigenes had a length less than 100 bp. A high portion of unigenes (n=969; 21.3%) had a length between 101-200 bp, followed by the unigenes with a length ranging from 201-300 bp (n=891; 19.6%), and unigenes (n=704; 15.5%) with a length between 301-400 bp, and 519 (11.4%) unigenes with a length ranging from 401-500 bp. Furthermore, a total of 372 (8.2%), 205 (4.5%) and 176 (3.9%) unigenes had a length between 501-600 bp, 601-700 bp, respectively. Only 2.2% (n=99) of unigenes had a length between 801-900 bp and 2.4% (n=111) of unigenes had a length ranging between 900-1000 bp. A total of 287 unigenes (6.3%) had a length more than 1 000 bp (Figure 9).

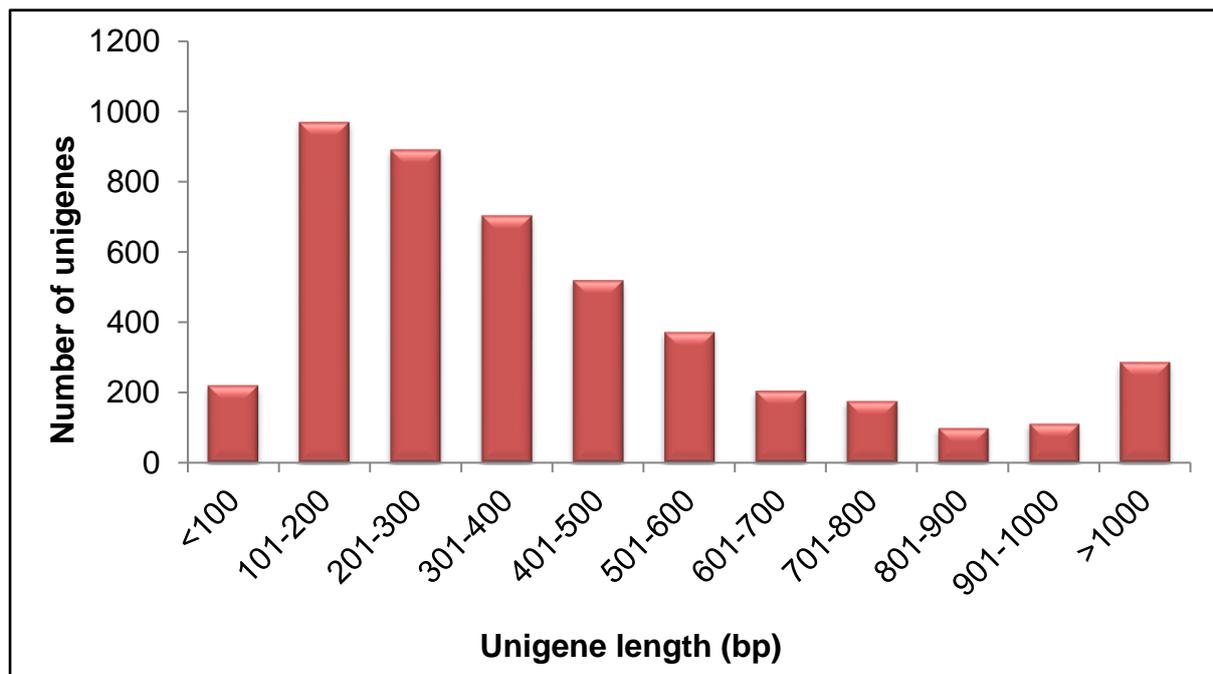


Figure 9: The length distribution of the combined unigenes in *Babesia rossi* transcriptome

4.9. Annotation statistics of the combined unigenes of *Babesia rossi* transcriptome.

Only 16.71% unigenes of the *B. rossi* transcriptome retrieved hits following analysis by BLAST and 100% had interproscan output (Figure 10). From these, 12.25% of the unigenes could be annotated based on the GO categories, and 14.01% were successfully mapped. Functional annotation also revealed that 133 out of 4553 unigenes were identified as encoding hypothetical proteins.

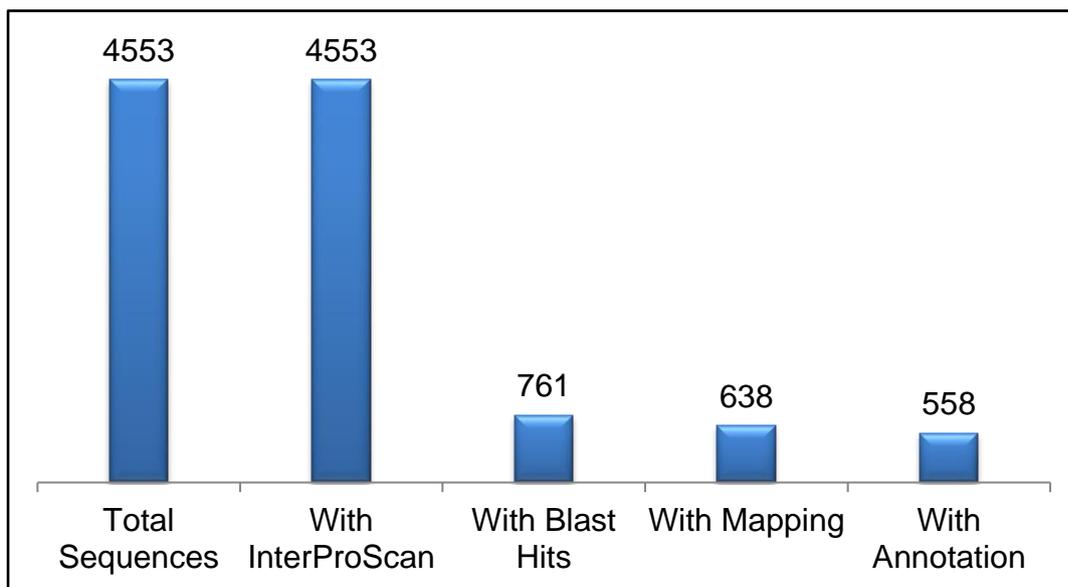


Figure 10: Data distribution of the combined unigenes detected from *Babesia rossi* transcriptome. Numbers indicate the number of unigenes.

4.10. Sequence homology analysis of *Babesia rossi* transcriptome

4.10.1 Similarity distribution

Babesia rossi transcriptome blast sequence similarity analysis revealed the highest percentage of unigenes (41%) with products which have sequence similarity between 60-80% to known proteins; while 35% and 23% of unigenes had sequence similarity ranging from 40-60% and 80-100%, respectively. Only 1% of the unigenes had sequence similarity between 20%-40% to known proteins (Figure 11).

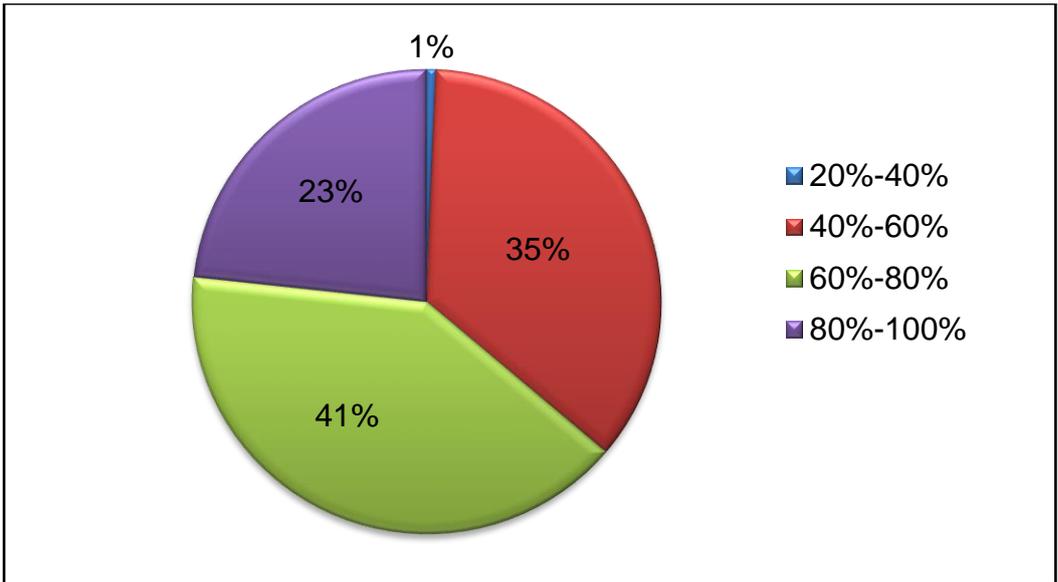


Figure 11: Sequence similarity of the blast hits of each unigene in *B. rossi* transcriptome.

4.10.2. E-value distribution

The E-value (expectation value) describes the expected number of random hits when searching a database; thus, the lower the E-value, the more significant a match to a query (Itai *et al.*, 2005). E-value distribution of the blast hits obtained from analysis of combined unigenes revealed that 26% had the E-value ranging between $1E-150$ to $1E-180$, followed by 23% of unigenes with the E-value between $1E-0$ to $1E-30$ and 20% of unigenes had E-value of $1E-30$ to $1E-60$. Between 7% and 14% of unigenes were distributed in $1E-60$ to $1E150$ clusters (Figure 12).

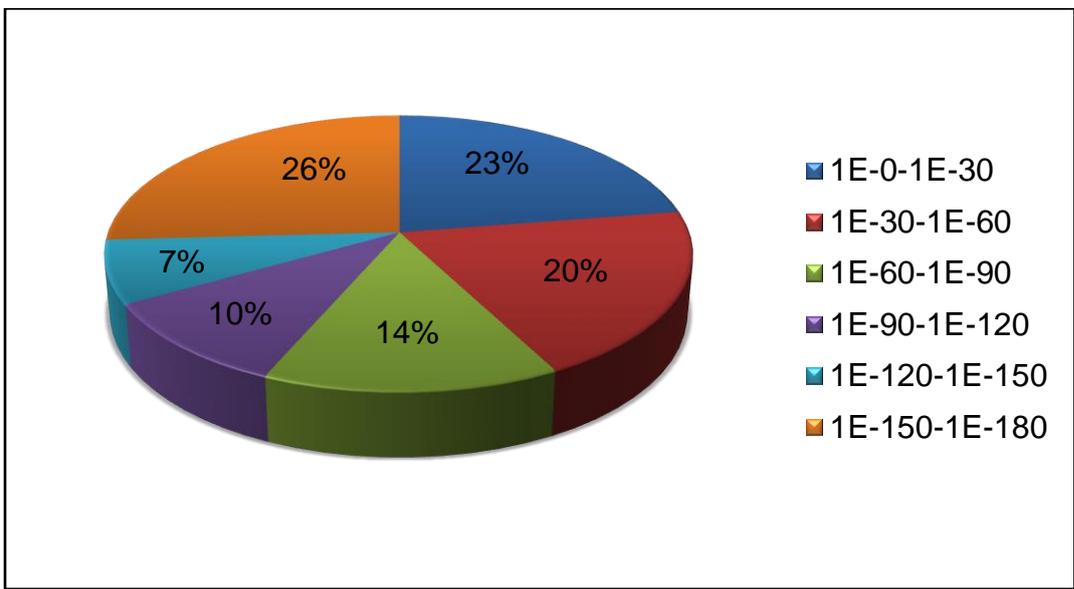


Figure 12: E-value distribution of the blast hits for unigenes from *B. rossi* transcriptome, using a cut -off value of 10E-3.

4.10.3. Top hit species distribution

Species distribution analysis showed that 41% of combined unigenes had homologs with *B. bigemina*, followed by *B. bovis* (26%), then *B. bovis* T2Bo (14%) and *Theileria equi* (9%). Minority of unigenes had sequence homologs with parasite species such as *B. equi*, *Theileria annulata*, *Theileria parva* and *B. gibsoni*. “Others” species which had homologs to some of the *B. rossi* unigenes included *B. canis*, *B. divergens*, *B. orientalis*, *B. microti* and *P. berghei*, amongst others (Figure 13).

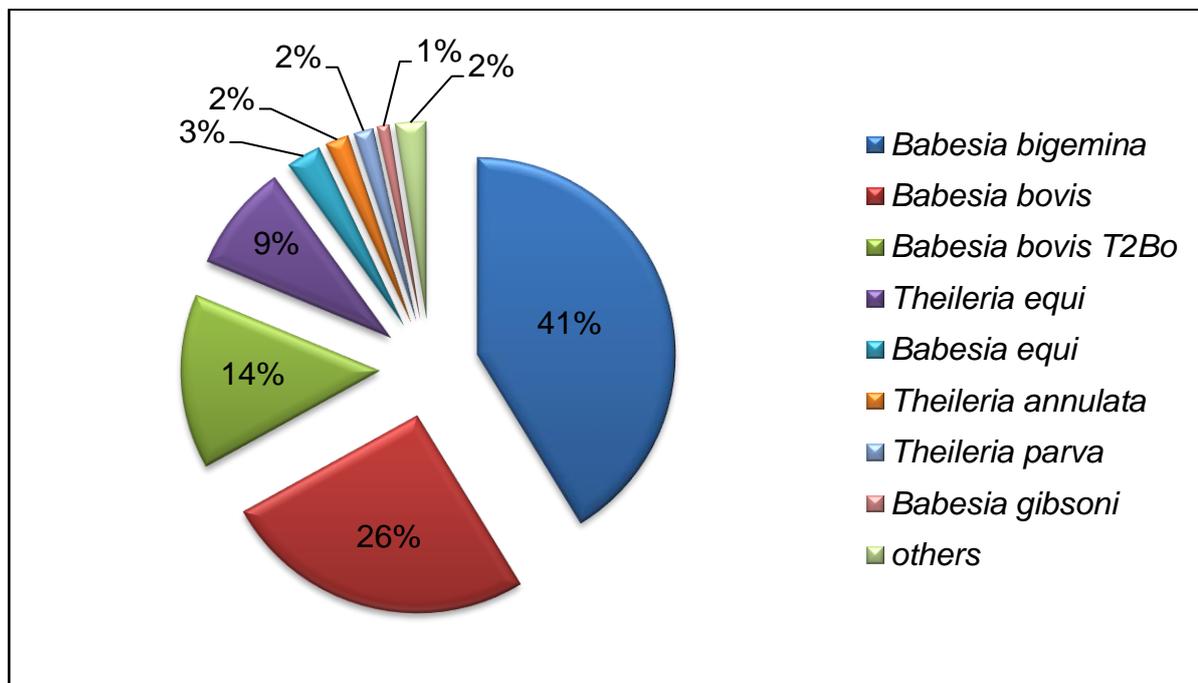


Figure 13: Species distribution for homologous sequences of *B. rossi* transcriptome unigenes.

4.11. Functional annotation of the unigenes in *Babesia rossi* transcriptome

Functional annotation is very important in understanding the transcriptomic data of species without the reference genomes (Kumar *et al.*, 2014). Gene ontology helps the functional characterization of genes and proteins of many organisms in terms of

biological processes, molecular functions and cellular components (Gotz *et al.*, 2008).

4.11.1. Comparative analysis of gene ontology outputs for unigenes detected in transcriptomes of *Babesia rossi* genotypes 19 and 29.

Based on the gene ontology analysis of the unigenes in *B. rossi* genotypes 19 and 29, the overall GO terms detected were similar in both genotypes (Figure 14). Only GO terms such as 'developmental process'; 'cell junction'; 'extracellular region part' and 'symplast' were presented only in *B. rossi* genotype 19 while 'multicellular organism process'; 'metallochaperone activity' and 'other organism part' were unique to *B. rossi* genotype 29.

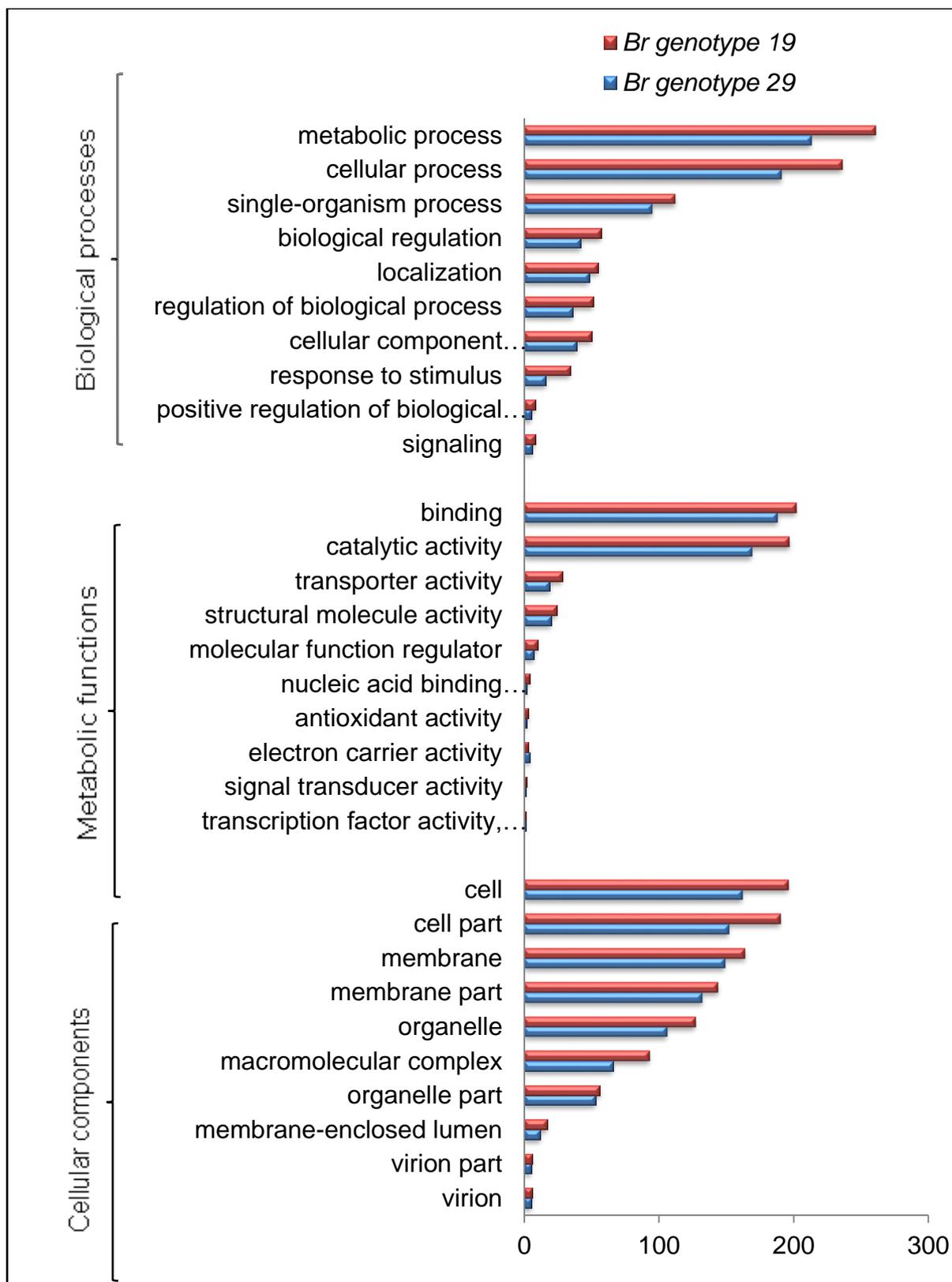


Figure 14: Top 10 most enriched GO terms, for specific ontology categories, detected from analysis of the unigenes of *Babesia rossi* genotypes 19 and 29. The x-axis indicates the gene ontology terms and the bar shows the number of unigenes.

4.11.2. Gene ontology analysis of the combined unigenes in *Babesia rossi* transcriptome

From GO analysis, 12.3% of the combined unigenes could be associated with three ontology categories, the biological process (37.3%), molecular function (20.1%) and cellular component (42.7%). The cellular component category was the most enriched GO category consisting of 17 GO terms. The subgroups that were most abundant in this category were the 'cell', 'cell part', 'membrane', 'membrane part' and 'organelle' (Figure 15).

The biological process category was the second most abundant, with 16 GO terms; the most enriched GO terms were the 'metabolic process', followed by the 'cellular process', 'single-organism process', 'localization' and 'biological' (Figure 15). The molecular function category had the least unigenes, with 8 GO terms. The most enriched GO terms in this category were mainly 'binding', 'catalytic activity', 'transporter activity', structural molecule activity and 'molecular function regulator' (Figure 15).

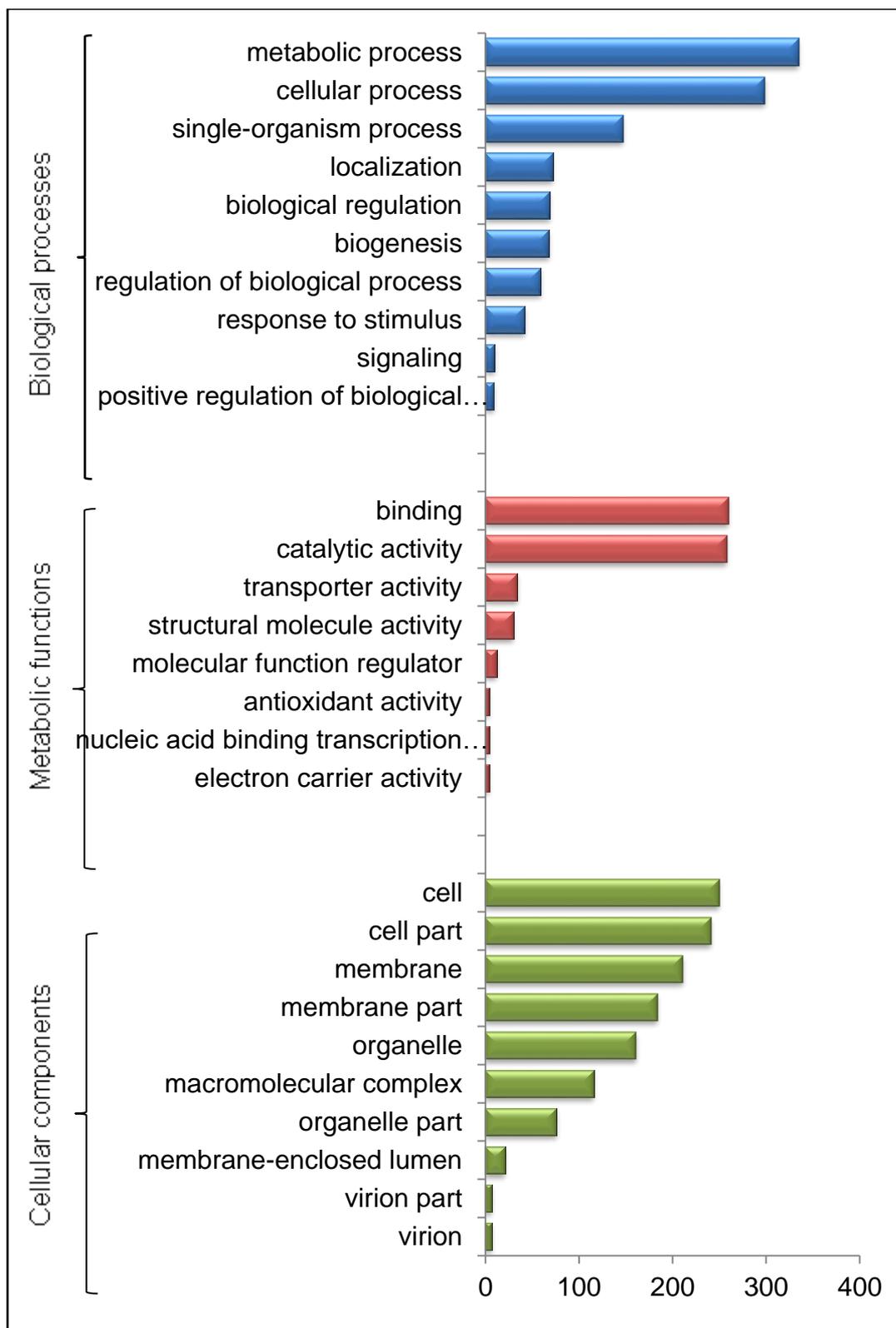


Figure 15: Top 10 most enriched GO terms, for specific ontology categories, detected from analysis of the combined unigenes of *Babesia rossi*. The x-axis indicates the gene ontology terms and the bar shows the number of unigenes.

4.12. Pathway associations of combined unigenes in *Babesia rossi* transcriptome

A total of 1550 unigenes in the *B. rossi* transcriptome could be assigned to 263 KEGG pathways. Among these, a large number of unigenes were assigned to metabolic pathways, followed by ribosomes, then biosynthetic of secondary metabolites, biosynthesis of antibiotics, splisosome, microbial metabolism in diverse environments, carbon metabolism, purine metabolism, RNA transport and pyrimidine metabolism (Figure 16).

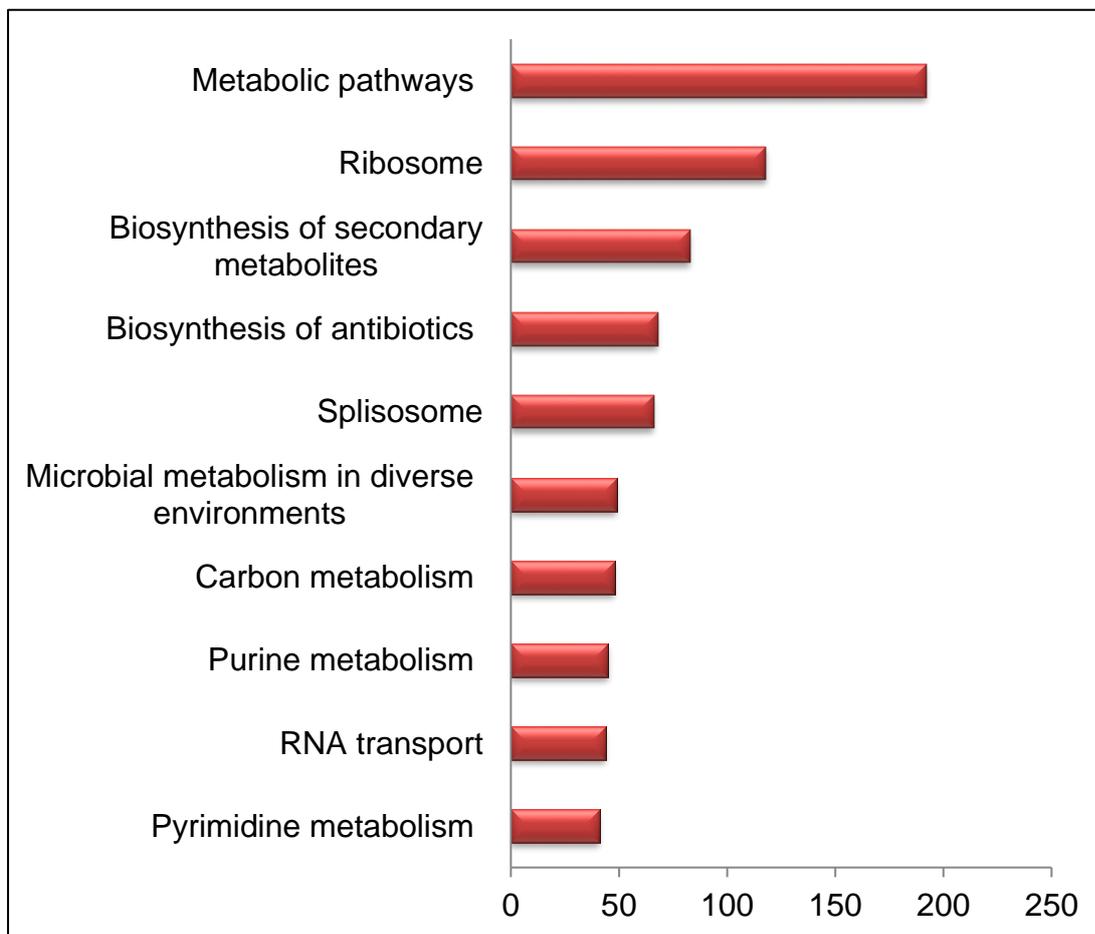


Figure 16: Top 10 most enriched pathways detected from the combined unigenes of *Babesia rossi*. The x-axis indicates the pathway and the bar shows the number of unigenes.

4.13. Cluster of Orthologous Groups (COG) classification

Out of 4553 combined unigenes, only 114 (2.5%) unigenes were clustered into 8 categories of COG. The largest category was 'nucleotide transport and metabolism' (n=60; 52.6%), the second was the 'general function prediction only' (n=17; 14.9%) followed by 'function unknown' (n=16; 14%) (Figure 17). Only one unigene was represented under 'carbohydrate transport and metabolism', 'DNA replication, recombination and repair' and 'energy production' each (Figure 17).

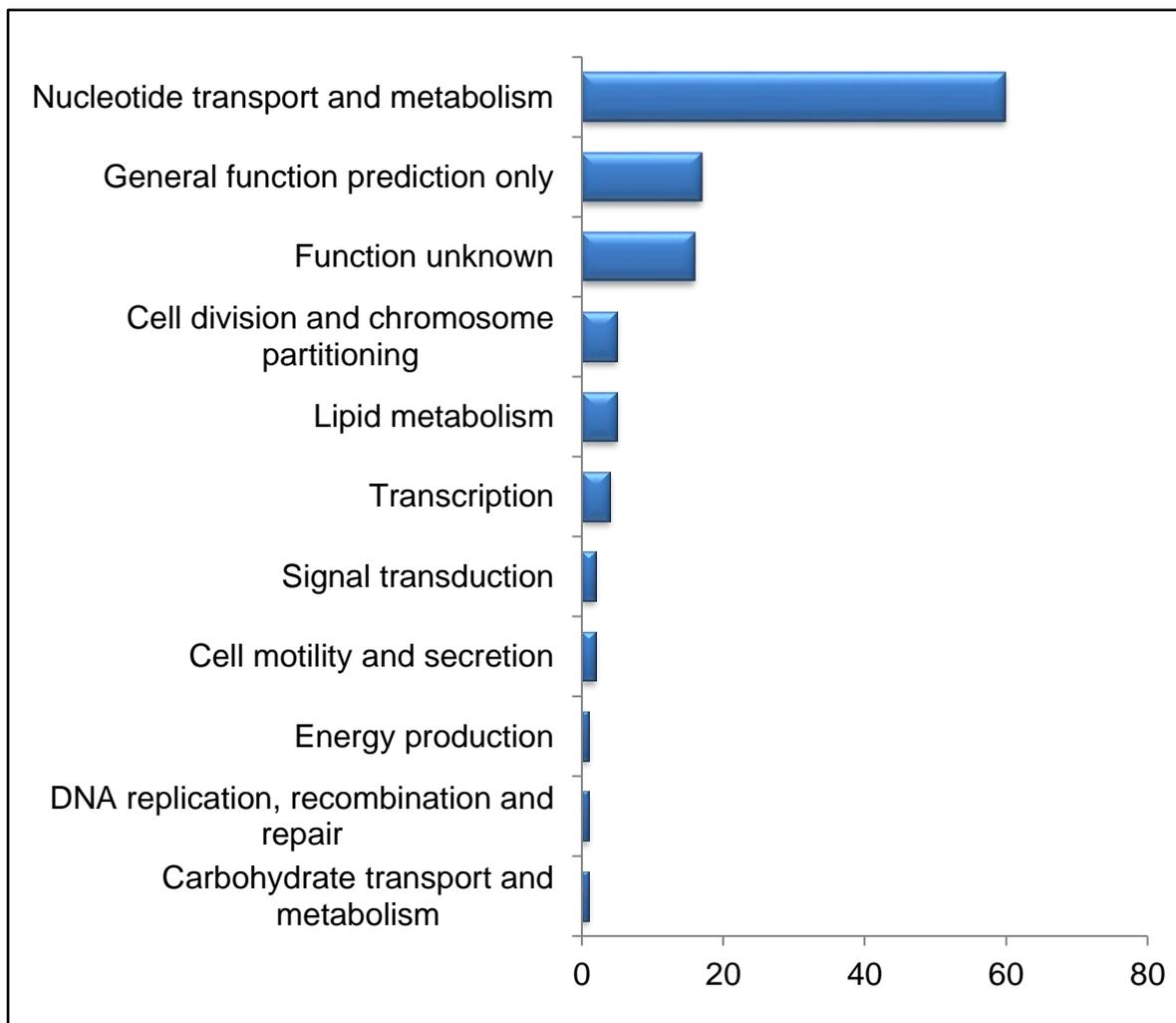


Figure 17: Cluster of Orthologous Groups (COG) classification of the combined unigenes of *Babesia rossi* transcriptome.

CHAPTER 5

5.1. DISCUSSION

The transcriptome profiles contribute an important insight to help understand developmental and disease processes by revealing the molecular components of cells (Wang *et al.*, 2011). Hence, the main aim of this study was to perform RNA-Seq and assemble the transcriptome of *B. rossi* from pathogens representing genotypes 19 and 29, each sampled from a random dog presenting varied clinical syndromes of canine babesiosis.

Two samples which tested negative for *B. rossi* parasite in the RLB hybridization assay produced negative DNA concentration values when measured using Nanodrop® ND-1000 Spectrophotometer. It was noted that the corresponding blood samples were clotted and this could have resulted in unsuccessful DNA extractions, as suggested in previous reports (Dell'anno *et al.*, 1998).

Using RLB hybridization assay, *B. rossi* nucleic acid material was detected in all tested samples except for two samples. However, targeting the *BrEMA1* gene using qPCR, *B. rossi* DNA was detected from all the samples, including the two samples that tested negative using the RLB hybridization assay. Thus, the lack of positive detection of *B. rossi* DNA from some of the samples by RLB hybridization analysis could have been due to low levels of *B. rossi* parasite DNA present in these samples. This confirms that RLB hybridization assay has a limited detection capability at certain concentrations or levels of parasitemia compared to qPCR, which has generally been proven to have high sensitivity (Vannier and Krause, 2009).

Notably, the melting temperatures of the *BrEMA 1* gene in the positive *B. rossi* DNA samples were comparable to the results in the previous study (Matjila *et al.*, 2009), confirming the specific amplification of the target gene region. Consequently, four *B. rossi* parasite genotypes, namely; 19, 28, 29 and 31 were identified from samples positive for *B. rossi* DNA, based on previously defined specific melting temperatures of the *BrEMA 1* gene and DNA sequence analysis (Matjila *et al.*, 2009).

To investigate the transcriptome of *B. rossi*, RNA was extracted from 20 blood samples from sick domestic dogs. Similar to results obtained from DNA extractions, negative RNA concentration values were detected from two samples, which could also be attributed to clotting of the blood samples used. About 65% of the RNA samples prepared from 20 blood samples investigated were deemed pure and of good quality to be used in transcriptome by RNA sequencing, based on the absorbance ratio, which was approximately 2 (Schoeder *et al.*, 2006) and RIN value, which ranged between 7 and 9 (Fleige and Pfaffl, 2006).

Furthermore, the 28S/18S ratio, which is normally used to indicate the integrity of RNA (Skrypina *et al.*, 2003), was also considered during the sample quality check, since the RNA molecule is very unstable and can be easily digested by RNase enzymes (Auer *et al.*, 2003), leading to the presence of the shorter fragments of RNA in the sample (Imbeaud *et al.*, 2005). Although the ratio of the 28S/18S bands was below the acceptable value (<2) in all the RNA samples selected for sequencing, the RNA samples number 4 (representing *B. rossi* genotype 31), 5 (representing *B. rossi* genotype 19) and 6 (representing *B. rossi* genotype 29) fulfilled other requirements for RNA sequencing by Illumina (TrueSeq). Finally, RNA samples were selected for sequencing to represent the three different *B. rossi* genotypes which have been associated with different clinical signs in the infected host.

Babesia rossi genotypes 19, 29 and 31 were included for transcriptome analysis; however, data for genotype 31 had to be excluded from the analysis due to low RNA mass in the selected samples. It has been reported that low RNA mass may lead to library construction failure or too low library construction to sequence or insufficient sequencing data (McCormic *et al.*, 2011). Moreover, the high number of unigenes produced in the transcriptome of *B. rossi* genotype 31 (5041), could be due to high proportion of duplication, poor randomness of sequencing data (Kirscher *et al.*, 2011).

According to Pedroni *et al.* (2013), the genomic comparison between three virulent and attenuated *B. bovis* strains showed no changes among the protein coding genes shared between the strains. This information led to the conclusion that the exclusion of *B. rossi* genotype 31 may not have a negative impact on the whole *B. rossi*

transcriptome. However, there was a reduction in the genome of all the attenuated strains (Pedroni *et al.*, 2013). Similarly; the less virulent *B. rossi* genotype 29 had reduced number of unigenes than the most severe *B. rossi* genotype 19. Therefore, this data proposed that the difference between the virulent and attenuated strains may lie in the non-coding regions which have an effect on the transcriptomic variability, or the loss of genome content during attenuation process (Pedroni *et al.*, 2013).

The assembly of reads from genotypes 19 and 29 revealed that *B. rossi* genotype 19 (n=26 747 238) had a high number of clean paired-end reads compared to its counterpart (genotype 29; n=25 709 672). The length of the sequences and the N50 are often used to measure the quality of reads assemblies (Schliesky *et al.*, 2012). In *B. rossi* transcriptomes from the different genotypes, similar N50 values were obtained. The slight varying average length and same N50 indicated that there was no biasness in the quality of the two assemblies.

The GC content plays an important role in the gene and genome regulation and assists to determine the physical properties of the genome (Carels *et al.*, 1998). The GC content was comparable at 62.63% in *B. rossi* genotype 19 and 61.18% in *B. rossi* genotype 29. This is higher than the GC content of other apicomplexan parasites, such as *B. bovis* (41, 8%), *Plasmodium falciparum* (19%) and *Theileria parva* (34, 1%) (Hayashida *et al.*, 2012). This finding could indicate the stability of the *B. rossi* genome.

The formation of unigenes from assembled transcriptome of *B. rossi* genotypes 19 and 29 resulted in 4553 unigenes detected between the two genotypes; notwithstanding that each genotype contributed a varying number of unigenes. Nonetheless, the number of combined unigenes detected from *B. rossi* transcriptome were comparable to that of the apicomplexan parasite, *T. parva* (n=4,035). There are no significant differences in known metabolic pathways between *Theileria* and *Babesia* species which have been indicated, eventhough there is a known preference to reproduce in different host cell types (Hayashida *et al.*, 2012). For instance, *Babesia* species proliferate in the red blood cells of the vertebrate host and are transovarially transmitted in the vector host, whereas *Theileria* species first reproduce in the lymphocytes and then red blood cells and is not transovarially

transmitted inside the tick (Chauvin *et al.*, 2009). Other apicomplexam parasites with sequenced genome include *P. falciparum* (5438 unigenes) and *B. bovis* (3641 unigenes) (Allred and Al-Khedery, 2004).

The considerable portion of the assembled unigenes had top blast hits with other apicomplexan parasites, mainly *Babesia spp.*, *Babesia bigemina*, followed by *B. bovis*. Notably, only a small proportion of sequences returned hits against other *Babesia species*, such as, *B. vogeli*, *B. canis* and *B. gibsoni*, other *Babesia species* causing canine babesiosis in dogs (Homer *et al.*, 2000). This result could be due to the unavailability of genome sequences for these species. As a general observation from this study and others which used *de novo* assembly, there could be biasness in analysis towards the species with annotated genome because of the more accurate gene predictions and high number of sequences available in public sequence databases compared to non-model organisms (Wang *et al.*, 2010), .

The ultimate goal in many transcriptome studies is to annotate the sequences by connecting them to the biological information (Ashrafi *et al.*, 2012). Low percentage of the combined unigenes (16.7%) had blast hits in the assembled *B. rossi* transcriptome, which might be due to the absence of *B. rossi* genome information in public databases. This finding shows the importance of generating the large collection of *B. rossi* sequences. Previous *de novo* transcriptome studies of other species reported the annotation rate ranging between 20%-40% (Ye *et al.*, 2006). In this study, less than 20% of unigenes in the *B. rossi* transcriptome could match known proteins, which is lower than the previously reported *de novo* transcriptome studies. This could be attributed to many unigenes with short length resulting in annotation of few unigenes. Alternatively, the unmatched unigenes could be unique to *B. rossi*. Regardless, all the unigenes in the *B. rossi* transcriptome were successfully annotated by InterProScan, which is an indication of coding sequences (Pallavicini *et al.*, 2013) and this shows that *de novo* assembled unigenes in *B. rossi* are coding for proteins characterized by known Interproscan domains.

Functional annotation based on GO terms, according to the biological process, cellular component and molecular function categories, was also performed. A broad range of GO terms representing the three main GO categories was detected from a considerable number of unigenes. This indicates that the transcriptome data

expressed diverse range of unigenes in the *B. rossi* involved in different pathways. In many *de novo* transcriptome studies (Ashrafi *et al.*, 2012; Annadurai *et al.*, 2013; Gallardo-Eccarate *et al.*, 2014), many unigenes are represented in the biological process GO group, followed by metabolic function and cellular component GO categories (Ashrafi *et al.*, 2012; Annadurai *et al.*, 2013; Gallardo-Eccarate *et al.*, 2014). However, in this study, the cellular component was the most abundant GO category, followed by the biological process and molecular function GO categories.

The large portion of unigenes in *B. rossi* transcriptome was represented in the 'cell'; 'cell part' and membrane within the cellular component category, implying that the *B. rossi* gene products are most active in the cell. The *Plasmodium falciparum* and *B. bovis* are able to attach to the capillary endothelium through sequestration (Allred and Al-Khedery, 2004), in which their proteins are expressed on the infected erythrocytes, which seem to transform the adhesive properties of the infected cells (Cooke *et al.*, 2001). Cytoadherence is mediated by *P. falciparum* Erythrocyte Membrane Protein 1 (PfEMP1) and Variant Erythrocyte Surface Antigen 1 (VESA1) in *P. falciparum* (Sherman *et al.*, 2003) and *B. bovis* (Chauvin *et al.*, 2009), respectively.

Although, it has not been proven, the expression of the parasite-derived antigens, with adhesive properties on the extracellular surface of cells infected with *B. rossi* can be accountable for their ability to sequester *in vivo* and to aggregate *in vitro* (Schetters and Eling, 1999). It has been established that the adhesive properties of *P. falciparum* and *B. bovis*-infected erythrocytes play an important role in their ability to sequester parasite, and therefore in the virulence of these parasites (Hutchings *et al.*, 2007). If virulence is directly linked to cytoadhesion, further research should be conducted to establish the relationship between *BrEMA1* genotypes and cytoadhesion.

Most host cell invasion stages by Apicomplexan parasites rely on the gliding motility to propel themselves into the host cell (Kemp *et al.*, 2012), which is powered by actinmyosin structure found beneath the plasma membrane of the parasite (Frènal *et al.*, 2010). Apicomplexan parasites are characterized by apical complex made up of unique organelles, namely: rhoptries, micromeres, and dense granules (Carruthers and Sibley, 1997). The contents released from these organelles are important for

invasion and survival of the parasite within the host cell (Carruthers and Sibley, 1997).

The 'metabolic process' and 'cellular process' were the most abundant in the biological process category, which indicates the cellular and metabolic processes in the developmental stage of the parasite. As intracellular parasites, Apicomplexans obtain nutrients and perform their energy requirements by manipulating the host metabolism (Xu *et al.*, 2010), causing changes in host cell signalling, nutrient delivery to the parasite, protein synthesis, apoptosis and evasion of host immunity (Edwards and Palsson, 2000). Therefore, this could explain the survival of *B. rossi* parasite in its host.

Lastly, 'binding' and 'catalytic activity' were dominant within the molecular function category, suggesting high gene regulation and enzymatic activity processes in *B. rossi* parasite. Regarding 'binding', parasite proteins, such as *Babesia* RAP-1, are involved with host receptors during invasion and are commonly found in the apical organelles (Rodriguez *et al.*, 2014). Members of the RAP-1 family in *Babesia*, such as *B. divergens*, are involved in the invasion of host cells through direct binding with receptors on the host cell surface (Yokoyama *et al.*, 2006). Moreover, parasite ligands and their complementary red blood cell receptors in invasion stages are very important in understanding the invasion biology of *Babesia* parasites and provides information for the development of effective and prevention strategies (Rodriguez *et al.*, 2014).

Recently, *B. canis* merozoite surface antigen 1 (*BcMSA1*) has been identified (Zhou *et al.*, 2016). In general, merozoite surface proteins in *Babesia* species are involved in merozoite invasion of host erythrocytes and are targets for host immune responses (Goo *et al.*, 2012). However, there is a genetic diversity in merozoite surface antigens of *Babesia* species (Zhou *et al.*, 2016), therefore, merozoite surface antigens in *B. rossi* need to be identified since it might be a potential candidate for diagnostic antigen and vaccines for diagnosis and prevention of *B. rossi* infection.

Regarding the 'catalytic activity', two molecular targets, namely dihydrofolate reductase (DHFR) enzyme and to a lesser extent, thymidylate synthase (TS) enzyme have been developed as a chemotherapeutics in *Babesia* and *Plasmodium*

parasites (Jiang *et al.*, 2000). The DHFR and TS in protozoans occur as bifunctional enzyme, in which they are expressed in single polypeptide (Ivanetich and Santi, 1990). The DHFR is very important for the rapid growth of cells due to its function in biosynthesis and cell replication (Aboge *et al.*, 2008). There is evidence that the *Babesia* DHFR-TS enzyme activity is crucial for the parasite survival, possibly by playing a role in the synthesis of thymidine nucleotides required for DNA synthesis and cell replication (Aboge *et al.*, 2008). Therefore, proteins with catalytic activity, depending on the pathways they are utilized, can be targets for control of *B. rossi* infections.

Metabolism was the largest category in KEGG pathways, indicating that active metabolic processes take place in *B. rossi* infection. Intracellular parasites, such as apicomplexan, retrieve nutrients and achieve their energy requirements by manipulating the host (Xu *et al.*, 2010). Cellular and biochemical approaches have been developed to investigate molecular functions in metabolic pathways (Tellez *et al.*, 2003). For example, most of the host-parasite common enzymes were down-regulated in fatty acid metabolic pathway in *Plasmodium*, while host specific host enzymes were up-regulated (Xu *et al.*, 2010). Furthermore, a recent study showed that *Plasmodium* rely on the *de novo* fatty acid synthesis only for liver-stage development, indicating that *Plasmodium* scavenges fatty acids mainly from its host blood cell (Tarun *et al.*, 2009).

'Ribosomes-associated pathways' were the second category identified in the top 10 KEGG pathways detected from analysis of *B. rossi* transcriptome. Ribosomes are important cellular components and play a role in protein synthesis (Van Spaendonk *et al.*, 2001). In apicomplexan parasites, ribosomes are produced with structurally distinct rRNA molecules encoded by single copy gene units (Bishop *et al.*, 2000). This distinction was demonstrated in *Plasmodium* through the stage specific expression of different rRNA genes (Waters *et al.*, 1997). It has been suggested that the maintenance of functionally different ribosome types provide *Plasmodium* with a unique post-transcriptional control mechanism to regulate gene expression (Thompson *et al.*, 1999).

Different ribosome types could be a step in commitment of the parasite to a change in life cycle stage by influencing the population of messenger rRNA (mRNAs) that

are translated (Van Spaendonk *et al.*, 2001). On the other hand, the expression of functionally different ribosomes might be able to optimize protein synthesis during growth in the different environments of the host and vector (Van Spaendonk *et al.*, 2001). *Babesia* and *Plasmodium* are closely related and share common clinical signs (Vial and Gorenflot, 2006), therefore; similar metabolic pathways could be involved in *B. rossi* and *P. falciparum* infections.

'Nucleotide and transport metabolism' category was the most enriched COG in *B. rossi* transcriptome. For example, parasites, such as *P. falciparum* and *Cryptosporidium*, depend on nucleotide uptake from the host, and are incapable of *de novo* amino acid synthesis and therefore have reliance upon salvage from the host (Garner *et al.*, 2002). The same mechanism could be taking place in *B. rossi* parasite. Notably, 11 out of 25 COG categories were identified; this could be due to the unavailability of the *B. rossi* unigenes in the DAVID software used. However, these data will provide information which would help to explore major genes for molecular mechanisms involved in *B. rossi* parasite infection.

5.2. CONCLUSION

In conclusion, the transcriptome of the two *B. rossi* genotypes representing different clinical manifestations on the host infected with *B. rossi* have been sequenced, assembled and functionally characterized. To our knowledge, this is the first study reporting on the transcriptome of *B. rossi*. Therefore, including samples from strains causing different clinical diseases was crucial in order to provide a maximally integrative transcriptome. Although, there were no significant differences between the *B. rossi* genotypes 19 and 29, the unigenes that were assigned to GO terms were mostly involved in biological processes, metabolic functions and cellular components, giving researchers a glimpse into the molecular nature of *B. rossi*. The generated transcriptomic data will also provide a valuable public resource of *B. rossi* gene sequences for future genetic and genomic studies of *B. rossi* parasite.

5.3. RECOMMENDATION STUDIES

Future studies could include gene expression analysis, which will allow determination of differentially expressed genes between these two *B. rossi* genotypes; this will assist in explaining the diverse clinical presentations often

observed in dogs diagnosed with canine babesiosis. Furthermore, the functional roles of the hypothetical proteins identified in the *B. rossi* parasite need to be characterized and investigated for possible disease modulators, novel drug targets or vaccine candidates against the *B. rossi* infection.

5.4. LIMITATIONS

Transcriptome sequencing is usually expensive (Merino *et al.*, 2016) and due to budget constraints, a small sample size could be investigated, and replicates could not be included in this study. Since studies with small sample size can only provide preliminary data, thus, future studies will have to examine a large sample size to confirm the findings of this study and allow deduction of significant conclusions.

LIST OF REFERENCES

- Aboge, G.O., Jia, H., Kuriki, K., Zhou, J., Nishikawa, Y., Igarashi, K., Fugisaki, K., Suzuki, H. and Xuan, X. 2007. Molecular characterization of a novel 32-kDa merozoite antigen of *Babesia gibsoni* with a better diagnostic performance by enzyme-linked immunosorbent assay. *Parasitology*. **134**: 1185-1194.
- Aboge, G.O., Jia, H., Terkawi, M.A., Goo, Y.K., Nishikawa, Y., Sunaga, F., Namikawa, K., Tsuji, N., Igarashi, I., Suzuki, H., *et al.* 2008. Cloning, expression, and characterization of *Babesia gibsoni* dihydrofolate reductase-thymidylate synthase: inhibitory effect of antifolates on its catalytic activity and parasite proliferation. *Antimicrob. Agents. Chemother.* **52** (11): 4072-4080.
- Adaszek, L., Wernicka-Furmaga, R. and Winiarczyk, S. 2012. Preliminary study on the safety of a new vaccine against canine babesiosis containing soluble parasitic antigen. *Bull. Vet. Inst. Pulawy*. **56**: 145-148.
- Agnieszka, S.S., Grzegorz, K., Pawel, G. and Halina, W. 2005. Usefulness of touch-down PCR assay for the diagnosis of typical cases of *Babesia canis canis* infections in dogs. *Bull. Vet. Inst. Pulawy*. **49**: 407-410
- Allred, D.R. and Al-Khedery, B. 2004. Antigenic variation and cytoadhesion in *Babesia bovis* and *Plasmodium falciparum*: different logics achieve the same goal. *Mol. Biochem. Parasitol.* **134**: 27-35.
- Altay, K., Dumanli, N. and Aktas, M. 2007. Molecular identification, genetic diversity and distribution of *Theileria* and *Babesia* species infecting small ruminants. *Vet. Parasitol.* **147**: 161-165.
- Altschul, S.F., Gish, W., Miller, W., Miller, E.W. and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**: 403-410.
- Annadurai, R.S., Jayakumar, V., Muggasimangalam, R.C., Katta, A.V., Anand S., *et al.* 2013. *De novo* transcriptome assembly (NGS) of *Curcuma longa* L. rhizome reveals novel transcripts related to anti-cancer and anti-malarial terpenoids. *PLoS One*. **8** (2): e56217.

Ano, H., Makiruma, S. and Harasawa, R. 2001. Detection of *Babesia* species from infected dog blood by polymerase chain reaction. *J. Vet. Med. Sci.* **63**: 111-113.

Ansorge, W.J. 2009. Next-generation DNA sequencing techniques. *Nature Biotechnol.* **4**: 195-203.

Apanaskevich, D.A., Horak, I.G. and Camicas, J.L. 2007. Redescription of *Haemaphysalis (Rhipistoma) elliptica* (Koch, 1844), an old taxon of the *Haemaphysalis (Rhipistoma) leachi* group from East and southern Africa, and of *Haemaphysalis (Rhipistoma) leachi* (Audouin, 1826) (*Ixodida, Ixodidae*). *Onderstepoort J. Vet. Res.* **74** (3): 181-208.

Ashrafi, H., Hill, T., Stoffel, K., Kozik, A., Yao, J., Chin-Wo, S.R. and Van Deynze, A. 2012. *De novo* assembly of the pepper transcriptome (*Capsium annuum*): a benchmark for the *silico* discovery of the SNPs, SSRs and candidate genes. *BMC.* **13** (751): 1-15.

Auer, H., Lyianarachchi, S., Newsom, D., Klisovic, M.I., Marcucci, U and Kornacker, K. 2003. Chipping aways at the chip bias: RNA degradation in microarray analysis. *Nat. Genet.* **35**: 292-293.

Birkenheuer, A. J., Levi, M. G. and Breitschwerdt, E. B. Development and evaluation of a seminested PCR for detection and differentiation of *Babesia gibsoni* (Asian genotype) and *B. canis* DNA in canine blood samples. *J. Clin. Microbiol.* **41** (9): 4172-4177.

Birkenheuer, A.J., Correa, M.T., Levy, M.G. and Breitschwerdt, E.B. 2005. Geographic distribution of babesiosis among dogs in the United States and association with dog bites: 150 cases (2000-2003). *J Am. Vet. Med. Assoc.* **227**: 942-947.

Bishop, R., Gobrigh, E., Spooner, P., Allsopp, B., Sohanpal, B. and Collins, N. 2000. Microsequence heterogeneity and expression of the LSU rRNA genes within two single copy ribosomal transcription units of *Theileria parva*. *Gene.* **257**: 299-305.

Bishop, R., Shah, T., Pelle, R., Hoyle, D., Pearson, T., Haines, L., Brass, A., Hulme, H., Graham, S.P., Taracha, E.L.N., *et. al.* 2005. Analysis of the transcriptome of the

protozoan *Theileria parva* using MPSS reveals that the majority of genes are transcriptionally active in the schizont stage. *Nucleic Acids Res.* **33** (17): 5503-5511.

Böhm, M., Leisewitz, A.L., Thompson, P.N. and Schoeman, J.P. 2006. Capillary and venous *Babesia canis rossi* parasitaemias and their association with outcome of infection and circulatory compromise. *Vet. Parasitol.* **141** (2): 18-29.

Bone, R.C., Balk, R.A., Cerra, F.B., Dellinger, R.P., Fein, A.M., Knaus, W.A., Schein, R.M. and Sibbald, W.J. 1992. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee, American College of Chest Physicians/Society of Critical Care Medicine, Chest. **101**: 1644-1655.

Boonchit, S., Alhassan, A., Chan, B., Xuan, X., Yokoyama, N., Ooshiro, M., Goff, W.L., Waghela, S.D., Wagner, G. and Igarashi, I. 2006. Expression of C-terminal truncated and full-length *Babesia bigemina* rhoptry-associated protein 1 and their potential use in enzyme-linked immunosorbent assay. *Vet. Parasitol.* **137**: 28-35.

Boozer, A.L. and Macintire, D.K. 2003. Canine babesiosis. *Vet. Clin. Nor. Am: Small Anim. Pract.* **33** (4): 885-904.

Boyle, J.P., Saeij, J.P.P., Cleary, M.D. and Boothroyd, J.C. 2006. Analysis of gene expression during development: lessons from the Apicomplexan. *Microb. Infect.* **8**: 1623-1630.

Bozdech, Z., Zhu, J., Joachimiak, M.P., Cohen, F.E., Pulliam, B., *et al.* 2003. Expression profiling of the schizont and trophozoite stages of *Plasmodium falciparum* with a long-oligonucleotide microarray. *Genome Biol.* **4**: 44-49.

Brayton, K.A., Lau, A.O., Herndon, D.R., Hannick, L., Kappmeyer, L.S., Berens, S.J., Bidwell, S.L., Brown, W.C., Crabtree, J., Fadrosch, D., Feldblum, T., Forberger, H.A., *et al.* 2007. Genome Sequence of *Babesia bovis* and comparative analysis of apicomplexan hemoprotozoa". *PLoS Pathogens.* **3** (10): e148. doi:10.1371/journal.ppat.0030148.

Caccio, S.M., Antunovic, B., Moretti, A., Mangili, V., Marinculic, A., Baric- Rafaj, R., Slemend, S.B. and Pieniazek, N.J. 2002. Molecular characterisation of *Babesia*

canis canis and *Babesia canis vogeli* from naturally infected European dogs. *Vet. Parasitol.* **106**: 285-292.

Calder, J.A.M., Reddy,G.R., Chieves,L., Littell,C.H., Livengood,J.R., Norval,R.A.I., Smith,C.and Dame, J.B. 1996. Monitoring *Babesia bovis* infections in cattle by using PCR-based tests. *J. Clin. Microbiol.* **34**: 2748-2755.

Carels, N., Hatey, P., Jabbari, K. and Bernardi, G. 1998. Compositional distribution of homologous coding sequences from plants. *J. Mol. Evol.* **46**: 45-43.

Carret, C., Walas, F., Carcy, B., Grande, N., Precigout, E., Moubri, K., Schetters, T.P. and Gorenflot, A. 1999. *Babesia canis canis*, *Babesia canis vogeli*, *Babesia canis rossi*: differentiation of the three subspecies by a restriction fragment length polymorphism analysis on amplified small subunit ribosomal RNA genes. *J. Euk. Microbiol.* **46**: 298-303.

Carruthers, V.B. and Sibley, L.D. 1997. Sequential protein secretion from three distinct organelles of *Toxoplasma gondii* accompanies invasion of human fibroblasts. *Eur. J. Cell Biol.* **73**: 114-123.

Chauvin, A., Moreau, E., Bonnet, S., Plantard, O. and Malandrin, L. 2009. *Babesia* and its hosts: adaptation to long-lasting interactions as a way to achieve efficient transmission. *Vet. Res.* **40** (37): 1-18.

Chen, G., Yin, K.P., Wang, C. and Shi, T. 2011. *De novo* transcriptome assembly of RNA-Seq reads with different strategies. *Sci. China Life Sci.* **54** (12): 1129-1133.

Clarke, K., Yang, Y., Mash, R., Xie, L.and Zhang, K.K. 2013. Comparative analysis of *de novo* transcriptome assembly. *Sci. China Life Sci.* **56**: 156-162.

Cooke, B. M., Mohandas, N. and Coppel, R.L. 2001. The malaria-infected red blood cell: structural and functional changes. *Adv. Parasitol.* **50**: 1-86.

Collett, M.G. 2000. Survey of canine babesiosis in South Africa. *J. S. Afr. Vet. Assoc.* **71** (3): 180-186.

Dantas-Torres, F. and Figueredo, L.A. 2006. Canine babesiosis: a Brazilian Perspective. *Vet. Parasitol.* **141**: 197-203.

- Dell'anno, A., Fabiano, M., Duineveld, G.C.A., Kok, A. and Danovaro, R. 1998. Nucleic Acid (DNA, RNA) Quantification and RNA/DNA Ratio Determination in marine sediments: comparison of spectrophotometric, fluorometric, and highperformance liquid chromatography methods and estimation of detrital DNA. *Appl. Environ. Microbiol.* **64** (9): 3238-3245.
- Edwards, J.S. and Palsson, B.O. 2000. Robustness analysis of the *Escherichia coli* metabolic network. *Biotechnol. Prog.* **16** (6):927-939.
- Eleftherianos, I., Marokhazi, J., Millichap, P.J., Hodgkinson, A.J., Sriboonlert, A., *et al.* 2006. Prior infection of *Manduca sexta* with non-pathogenic *Photorhabdus luminiscens*: roles of immunorelated proteins shown by RNA interference. *Insect Biochem. Mol. Biol.* **36**: 517-525.
- Finizio, A.L., Kleuskens, J.A., Van De Crommert, J., Gorenflot, A., Carcy, B. and Schetters, T.P. 2011. Soluble parasite antigens from *Babesia canis* do not directly activate the kallikrein system in dogs diagnosed with *Babesia canis*. *Vet. Parasitol.* **176** (3): 132-138.
- Fleige, S. and Pfaffl, M.W. 2006. RNA integrity and the effect on the real-time qRT-PCR performance. *Mol. Aspect. Med.* **27**: 126-139.
- Frenal, K., Polonais, V., Marq. J.B., Stratmann, R., Limenitakis, J. and Soldati-Favre, D. 2010. Functional dissection of the apicomplexan glideosome molecular architecture. *Cell Host Microb.* **8**: 343-357.
- Gallardo-escárate, C., Valenzuela-muñoz, V. and Nuñez-acuña, G. 2014. A RNA-Seq analysis using *de novo* transcriptome assembly as a reference for the *Salmon Louse Caligus rogercressey*. *PLoS ONE.* **9** (4): e92239.
- García-Sanmartín, J., Nagore, D., García-pérez, A.L., Juste, R.A. and Hurtado, A. 2006. Molecular diagnosis of *Theileria* and *Babesia* species infecting cattle in Northern Spain using reverse line blot microarrays. *Vet. Res.* **2**: 16-22.
- Garner, M. J., Bishop, R., Shah, T., De Villiers, E.P., Carlton, J.M., Hall, N., Ren, Q., Paulsen, I.T., Pain, A., Berriman, M., *et al.* 2002. Genome sequence of *Theileria parva*, a bovine pathogen that transforms lymphocytes. *Science.* **309**: 134-137.

Gasser, R.B. 2006. Molecular tools - advances, opportunities and prospects. *Vet. Parasitol.* **136**: 69-89.

Goo, Y.K., Aboge, G.O., Terkawi, M.A., Jia, H., Yamagishi, J., Sunaga, F., Namikawa, K., *et al.* 2012. Four promising antigens, *BgP32*, *BgP45*, *BgP47*, and *BgP50*, for serodiagnosis of *Babesia gibsoni* infection were classified as *B. gibsoni* merozoite surface protein family. *Parasitol. Int.* **61** (2): 364-368.

Goris, R.J., Bockhorst, T.P., Nuytinek, J.K. and Gimbrere, J.S. 1985. Multiple organ failure. Generalized auto destructive inflammation. *Arch. Surg.* **120**: 1109-1115.

Gotz, S., Garcia-Gomez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H.*et al.* 2008. High-throughput functional annotation and data mining with the BLAST2GO suite. *Nucleic acids Res.* **36**: 3420-3435.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I, *et al.* 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29** (7): 644-652.

Gubbels, J.M., De Vos, A.P., Van Der Weide, M., Viseras, J., Schouls, E. De Vries, L.M.and Jongejan, F. 1999. Simultaneous detection of bovine *Theileria* and *Babesia* species by reverse line blot hybridization. *J. Clin. Microbiol.* **37**: 1782-1789.

Hackshaw, A. 2008. Small studies: strengths and limitations. *Eur. Respir. J.* **32** (5): 1141-1143.

Hayashida, K., Hara, Y., Abe, T., Yamasaki, C., *et al.* 2012. Comparative genome analysis of three eukaryotic parasites with differing abilities to transform leukocytes reveals key mediator of *Theileria*-induced leukocyte transformation. *mBio.* **3** (5):e00204-12. doi:10.1128/mBio.00204-12.

Hoffman, J.I. 2011. Gene discovery in the Antarctic fur seal (*Arctocephalus gazella*) skin transcriptome. *Mol. Ecol. Res.* **11**: 703-710.

Holman, P.J., Swift, P.K., Frey, R.E., Bennett, J., Cruz, D. and Wagner, G.G. (2002). Genotypically unique *Babesia* spp. isolated from reindeer (*Rangifer tarandus*) in the United States. *Parasitol. Res.* **88**: 405-409.

Homer, M.J., Bruinsma, E.S., Lodes, M.J., Moro, M.H. 3rd Telford, S.R., Krause, P.J., Reynolds, L.D., Mohamath, R., Benson, D.R., Houghton, R.L., *et al.* 2000. A polymorphic multigene family encoding an immunodominant protein from *Babesia microti*. *J. Clin. Microbiol.* **38**: 362-368.

Hunfeld, K.P., Hildebrandt, A. and Gray, J.S. 2008. Babesiosis: Recent insights into ancient diseases. *Int. J. Parasitol.* **38**: 1219-1237.

Hutchings, C. L., Li, A., Fernandez, K.M., Fletcher, T., *et al.* 2007. New insights into the altered adhesive and mechanical properties of red blood cells parasitized by *Babesia bovis*. *Mol. Microbiol.* **65**: 1092-1105.

Imbeaud, S., Graudens, E., Boulanger, V., Barlet, X., Zaborski, P., Eveno, E., *et al.* 2005. Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces. *Nucleic Acids Res.* **33** (6): e56.<http://doi.org/10.1093/nar/gni054>.

Ivanetich, K.M. and Santi, D.V. 1990. Thymidylate synthase-dihydrofolate reductase in protozoa. *Exp. Parasitol.* **70** (3), 367-371. Inokuma, H., Yoshizaki, Y., Matsumoto, K., Okuda, M., Onishi, T., Nakagome, K., Kosugi, R. and Hirakawa, M. 2004. Molecular survey of *Babesia* infection in dogs in Okinawa, Japan. *Vet. Parasitol.* **121**: 341-346.

Inokuma, H., Yoshizaki, Y., Matsumoto, K., Okuda, M., Onishi, T., Nakagome, K., Kosugi, R. and Hirakawa, M. Molecular survey of *Babesia* infection in dogs in Okinawa, Japan. 2004. *Vet Parasitol.* **121**: 341-346.

Irwin, P.J. 2009. Canine babesiosis: from molecular taxonomy to control. *Parasit. Vectors.* **2** (Suppl 1) S4.[doi: 10.1186/1756-3305-2-S1-S4](https://doi.org/10.1186/1756-3305-2-S1-S4).

Itai, S., Birkland, A., Chang, K., El-Yaniv, R. and Yona, G. 2005. Correcting BLAST e-Values for low-complexity segments. *J. Comput. Biol.* **12** (7): 978-1001.

Jacobson, L.S. 2006. The South African form of severe and complicated canine babesiosis: clinical advances 1994-2004. *Vet. Parasitol.* **138**: 126-139.

Jacobson, L.S. and Lobetti, R.G. 1996. Rhabdomyolysis as a complication of canine babesiosis. *JSAP*. **37**: 286-291.

Jakalski, T.G.M., Wakaguri, H., Kischka, T.G., Nishikawa, Y., Kawazu, S., Matsubayashi, M., Kawahara, F., Tsuji, N., Cao, S. and *et al.* 2014. Database issue D631-D636 DB-AT: a 2015 update to the Full-parasites database brings a multitude of new transcriptomic data for apicomplexan parasites. *Nucleic Acids Res.* **43**: 1-6.

Jia, H., Zhou, J., Ikadai, H., Matsuu, A., Suzuki, H., Igarashi, H., Fujisaki, K. and Xuan, X. 2006. Identification of a novel gene encoding a secreted antigen 1 of *Babesia gibsoni* and evaluation of its use in serodiagnosis. *Am. J. Trop. Med. Hyg.* **75**: 843-850.

Jia, H., Liao, M., Lee, E., Nishikawa, Y., Inokuma, H., Matsuu, A., Garashi, I. I. and Xuan, X. 2007. Development of an immunochromatographic test with recombinant BgSA1 for the diagnosis of *Babesia gibsoni* infection in dogs. *Parasitol. Res.* **100**: 1381-1384.

Jiang, L., Lee, P.C., White, J. and Rathod, P.K. 2000. Potent and selective activity of a combination of thymidine and 1843U89, a folate-based thymidylate synthase inhibitor, against *Plasmodium falciparum*. *Antimicrob. Agents Chemother.* **44**: 1047-1050.

Johnson, M.T.J., Carpenter, E.J., Tian, Z., Bruskiwich, R., Burris, J.N., *et al.* 2012. Evaluating Methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. *PLoS ONE*. **7** (11): e50226. doi:10.1371/journal.pone.0050226.

Jongejan, F., Zivkovic, D., Pegram, R.G., Tatchell, R.J., Fison, T., Latif, A.A. and Paine, G. 1987. Ticks (*Acari: Ixodidae*) of the Blue and White Nile ecosystems in the Sudan with particular reference to the *Rhipicephalus sanguineus* group. *Exp Appl Acarol.* **3**:331-346.

Keller, N., Jacobson, L.S., Nel, M., De Clerq, M., Thompson, P.N. and Schoeman, J.P. 2004. Prevalence and risk factors of hypoglycemia in virulent canine babesiosis. *J. Vet. Intern. Med.* **18**: 265-270.

- Kemp, L.E., Yamamoto, M. and Soldati-Favre, D. 2012. Subversion of host cellular functions by the apicomplexan parasites. *Microbiol. Rev.* **37**: 607-631.
- Kircher, M., Heyn, P. and Keiso, J. 2011. Addressing challenges in the production and analysis of Illumina sequencing data. *BMC Genomics.* **12** (382): 1-14.
- Kumar, A., Congui, L., Lindström, L., Piironen, S., Vidotto, M. and Grapputo, A. 2014. Sequencing, *de novo* assembly and annotation of the Colorado potato beetle, *Leptinotarsa decemlineata*, transcriptome. *PLoS ONE.* **9** (1): e86012.
- Last, R.D., Hilla, J.M., Matjila, P.T. and Rème, C.A. 2007. A field trial evaluation of the prophylactic efficacy of amitraz-impregnated collars against canine babesiosis (*Babesia canis rossi*) in South Africa. *J. S. Afr. Vet. Assoc.* **78** (2): 63-65.
- Levy, M.G., Breitschwerdt, E.B. and Moncol, D.J. 1987. Antibody activity to *Babesia canis* in dogs in North Carolina. *Am. J. Vet. Res.* **48**: 339-341.
- Lewis, B.D., Penzhorn, B.L. and Lopez-Rebollar, L.M. 1995. Immune responses to South African *Babesia canis* and the development of a preliminary vaccine. *J. S. Afr. Vet. Assoc.* **66**: 61-65.
- Lewis, B.D., Penzhorn, B.L., Lopez-Rebollar, L.M. and De Waal, D.T. 1996. Isolation of a South African vector specific strain of *Babesia canis*. *Vet. Parasitol.* **63**: 9-16.
- Lobetti, R.G. 2006. Babesiosis, in *Infectious diseases of the dog and cat*, 3rd edition, C.E. Greene (Eds). Philadelphia: W.B. Saunders: 61-68.
- Malatji, D.P. 2011. Detection of *Babesia rossi* genotypes using real-time PCR. MSc dissertation report. Pretoria: University of Pretoria, Department of Veterinary Tropical Diseases, Faculty of Veterinary Science.
- Malherbe, W.D. and Parkin, B.S. 1951. Atypical symptomatology in *Babesia canis* infection. *J. S. Afr. Vet. Assoc.* **22**: 25-61.
- Martin, J.A. and Wang, Z. 2011. Next-generation transcriptome assembly. *Nat. Rev. Genet.* **12** (10): 671-682.

- Matjila, P.T., Penzhorn, B.L., Bekker, C.P., Nijhof, A.M. and Jongejan, F. 2004. Confirmation of occurrence of *Babesia canis vogeli* in domestic dogs in South Africa. *Vet. Parasitol.* **122** (2):119-125.
- Matjila, T.P., Nijhof, A.M., Taoufik, A., Houwers, D., Teske, E., Penzhorn, B.L., De Lange, T.D., Jongejan, F. 2005. Autochthonous canine babesiosis in The Netherlands. *Vet. Parasitol.* **131**: 23-29.
- Matjila, P.T., Leisewitz, A.L., Jongejan, F. and Penzhorn, B.L. 2008. Molecular detection of tick borne protozoal and *ehrlichial* infections in domestic dogs in South Africa. *Vet. Parasitol.* **155** (1-2): 152-157.
- Matjila, P.T., Carcy, B., Leisewitz, A.L., Schettters, T., Jongejan, F., Gorenflot, A. and Penzhorn, B.L. 2009. Preliminary evaluation of the *BrEMA1* gene as a tool for associating *babesia rossi* genotypes and clinical manifestation of canine babesiosis. *J. Clin. Microbiol.* **47** (11): 3586-3592.
- Matijatko, V., Kis, I., Torti, M., Brkljacic, M., Kucer, N., Rafaj, R.B., Grden, D., Zivicnjak, T. and Mrljak, V. 2009. Septic shock in canine babesiosis. *Vet. Parasitol.* **162** (3-4): 263-270.
- McCormic, K.P., Willman, M.R. and Meyers, B.C. 2011. Experimental design, preprocessing, normalization and differential expression analysis of small RNA sequencing experiment. *Silence.* **2** (2): 1-19.
- Mehlhorn, H. and Schein, E. 1984. The piroplasms: life cycle and sexual stages. *Adv. Parasitol.* **23**: 37-103.
- Merino, A.M., Fresno, C., Netto, F., Netto, E.M, Pratto, L and Fernandez E.A. 2016. The impact of quality control in RNA-seq experiments. *J. Phys. Conf. Ser.* **705**. 1-10.
- Morgan, U.M. and Thompson, R.C.A. 1998. Molecular detection of parasitic protozoa. *Parasitology.* **117**: 73-85.
- Morozova, O., Hirst, M. and Marra, M.A. 2008. Applications of next-generation sequencing technologies in functional genomics. *Genomics.* **92** (5): 255-264.

Morozova, O., Hirst, M. and Marra M.A. 2009. Applications of new sequencing technologies for transcriptome analysis. *Annu Rev Genomics Hum Genet* .10: 135-151.

Muller, P.Y., Janovjak, H., Miserez, A.R. and Dobbie, Z. 2002. Processing of gene expression data generated by quantitative real-time RT-PCR. *BioTechniques*. **32** (6): 1372-1379.

Nutall, G.F.H. 1910. On Haematozoa occurring in wild animals in Africa. 1. *Piroplasma rossi* N. Sp. And *Haemogregarina canis adust* N. Sp found in jackal. *Parasitology*. **3**: 108-116.

Nutall, G.H.F. 1912. Note on *Rosella rossi* (Nuttall, 1910) occurring in the jackal in British East Africa. *Parasitology*. **3**: 108-116.

Opperdoes, R.F. 1998. Research unit for tropical diseases"christian de duve" institute of cellular pathology and laboratory of biochemistry, Université Catholique de Louvain, Avenue Hippocrate, B-1200, Brussels, Belgium: 74-75.

Ouzounis, C.A. and Karp, P.D. 2002. The past, present and future of genome-wide re-annotation. *Genome Biol*. **3** (2): 1-6.

Oyamada, M., Davoust, B., Boni, M., Dereure, J., Bucheton, B., Hammad, A., Itamoto, K., Okuda, M.and Inokuma, H. 2005. Detection of *Babesia canis rossi*, *B. canis vogeli*, and *Hepatozoon canis* in dogs in a village of Eastern Sudan by using a screening PCR and sequencing methodologies. *Clin. Diagn. Lab. Immunol*. **12** (11): 1343-1346.

Pallavicini, A., Canapa, A., Barucca, M., Alfoldi, J., *et al.* 2013. Analysis of the transcriptome of the Indonesian *coelacanth Latimeria menadoensis*. *BMC Genomics*. **14** (538): 1-15.

Pareek, C.S., Smoczynski, R. and Tretyn, A. 2011. Sequencing technologies and genome sequencing. *J. Appl. Genet*. **52** (4): 413-435.

Pedroni, M.J., Sondgeroth, K.S., Gallego-Lopez, G.M., Echaide, I. and Lau, A.O.T. 2013. Comparative transcriptome analysis of geographically distinct virulent and

attenuated *Babesia bovis* strains reveals similar gene expression changes through attenuation. *BMC Genomics*.**14** (763): 1-13.

Penzhorn, B.L. 2011. Why is Southern African canine babesiosis so virulent? An evolutionary perspective. *Parasit. Vectors*. **4** (51):1-6.

Prasath D., Karthika, R., Habeeba, N.T., Suraby, E.J., Rosana, O.B., Shaji, A., *et al.* 2014. Comparison of the transcriptomes of ginger (*Zingiber officinale* Rosc.) and mango ginger (*Curcuma amada* Roxb.) in response to the bacterial wilt infection. *PLoS ONE*. **9**(6): e99731. <https://doi.org/10.1371/journal.pone>.

Qian, X., Ba, Y., Zhuang, Q. and Zhong, G. 2014. RNA-Seq technology and its application in fish transcriptomics. *OMICs*. **18** (2): 98-110.

Reiter, I. and Weiland, G. 1989. Recently developed methods for the detection of babesial infections. *Trans. R. Soc. Trop. Med. Hyg.* **83**: 21-23.

Reyers F., Leisewitz, A.L., Lobetti, R.G., Milner, R.J. and Jacobson, L.S. 1998. Canine babesiosis in South Africa: more than one disease. Does this serve as a model for *Falciparum* malaria? *Ann. Trop. Med. Parasitol.* **92** (4): 503-511.

Robertson, G., Schein, J., Chiu, R., Field, M., Jackman, S.D., Mungall, K., Lee, S., Okada, H.M., *et al.* 2010. *De novo* assembly and analysis of RNA-seq data. *Nature Methods*. **7** (11): 909-912.

Rodriguez, M., Alhassan, A., Ord, R.L., Cursino-Santos, J.R., Singh, M., *et al.* 2014. Identification and characterization of the RouenBd 1987 *Babesia divergens* Rhopty-associated protein 1. *PLoS ONE*. **9** (9): e107727. doi:10.1371/journal.pone.0107727.

Rosenkranz, R., Borodina, T., Lehrach, H. and Himmelbauer, H. 2008. Characterizing the mouse ES cell transcriptome with Illumina sequencing. *Genomics*. **92**: 187-194.

Saha, S., Sparks, A.B., Rago, C., Akmaev, V., Wang, C.J., Vogelstein, B., Kinzler, K.W., and Velculescu, V.E. 2002. Using the transcriptome to annotate the genome. *Nature Biotechnol.* **20**: 508-512.

Sasaki, M., Omobowale, O., Tozuka, M., Ohta, K., Matsuu, A., Nottidge, H.O., Hirata, H., Ikadai, H. and Oyamada, T. 2007. Molecular survey of *Babesia canis* in dogs in Nigeria. *J. Vet. Med. Sci.* **69**: 1191-1193.

Scheepers, E., Leisewitz, A.L., Thompson, P.N. and Christopher, M.M. 2011. Serial haematology results in transfused and non-transfused dogs naturally infected with *Babesia rossi*. *J. S. Afr. Vet. Assoc.* **82** (3): 136-143.

Schettters, T.P.M. and Eling, W.M.C. 1999. Can *Babesia* infections be used as a model for cerebral malaria? *Parasitol. Today.* **15**: 492-497.

Schettters, T.P.M., Kleuskens, J.A.G.M., Scholtes, N.C., Gorenflot, A., Moubri, K. and Vermeulen, A.N. 2001. Vaccination of dogs against heterologous *Babesia canis* infection using antigens from culture supernatants. *Vet. Parasitol.* **100**: 75-86.

Schettters, T.P.M., Kleuskens, J.A.G.M., Crommert, J., Leeuw, P.W.J., Finizio, A.L. and Gorenflot, A. 2009. Systemic inflammatory responses in dogs experimentally infected with *Babesia canis*; a haematological study. *Vet. Parasitol.* **162**: 7-15.

Schliesky, S., Gowik, U., Beber, A.P.M and Brautigam, A. 2012. RNA-Seq assembly- are we there yet? *Front. Plant Sci.* **3** (220): doi: 10.3389/fpls.2012.00220.

Schnittger, L., Yin, H., Qi, B., Gubbels, M.J., Beber, D., Niemann, S., Jongejan, F. and Ahmed, J.S. 2004. Simultaneous detection and differentiation of *Theileria* and *Babesia* parasites infecting small ruminants by reverse line blotting. *Parasitol. Res.* **92**: 189-196.

Schoeman, J.P. 2009. Canine babesiosis. *Onderstepoort J. Vet. Res.* **76**: 59-66.

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M. and Ragg, T. 2006. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol.* **7** (3): 1-14.

Shakespeare, A.S. 1995. The incidence of canine babesiosis amongst sick dogs presented to the Onderstepoort Veterinary Academic Hospital. *J. S. Afr. Vet. Assoc.* **66**: 247-250.

- Sherman, I.W., Eda, S. and Winograd, E. 2003. Cytoadherence and sequestration in *Plasmodium falciparum*: defining the ties that bind. *Microbes. Infect.* **5** (10): 897-909.
- Shinuo, C.A.O. 2013. Studies on development of serological diagnostic methods and molecular targeted drugs against canine babesiosis, Doctor's Course of Animal and Food Hygiene, Graduate School of Obihiro University of Agriculture and Veterinary Medicine.
- Silva, M.G., Ueti, M.W., Norimine, J., Florin-Christensen, M., Bastos, R.G. Goff, W.L., et al. 2010. *Babesia bovis* expresses a neutralization-sensitive antigen that contains a microneme adhesive repeat (MAR) domain. *Parasitol. Int.* **59**: 294-297.
- Skotarczak, B. 2008. Babesiosis as a disease of people and dogs. Molecular diagnostics: a review. *Vet. Med-Czech.* **53** (5): 229-235.
- Skrypina, N.A., Timofeeva, A.V., Khaspekov, G.L., Savochkina, L.P and Beabealashvili, R. 2003. Total RNA suitable for molecular biology analysis. *J Biotechnol.* **105** (1-2): 1-9.
- Solano-Gallego, L. and Baneth, G. 2011. Babesiosis in dogs and cats - expanding parasitological and clinical spectra. *Vet. Parasitol.* **18**: 48-60.
- Subeki, M.H., Takahashi, K., Nabeta, K., Yamasaki, M., Maede, Y. and Katakura, K. 2007. Screening of Indonesian medicinal plant extracts for antibabesial activity and isolation of new quassinoids from *Brucea javanica*. *J. Nat. Prod.* **70** (10): 1654-1657.
- Taboada, J. and Merchant, S.R. 1991. Babesiosis of companion animals and man. *Vet. Clin. North. Am. Small. Anim. Pract.* **21**: 103-123.
- Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A. and Conesa, A. 2011. Differential expression in RNA-seq: a matter of depth. *Genome Res.* **21**: 2213-2223.
- Tarun, A.S., Vaughan, A.M. and Kappe, S.H. 2009. Redefining the role of *de novo* fatty acid synthesis in *Plasmodium* parasites. *Trends Parasitol.* **25** (12):545-550.
- Tellez, M., Matesanz, F. and Alcina, A. 2003. The C-terminal domain of the *Plasmodium falciparum* acyl-CoA synthetases PfACS1 and PfACS3 functions as ligand for ankyrin. *Mol. Biochem. Parasitol.* **129** (2):191-198.

The Merck Veterinary Manual for Veterinary Professionals, Circulatory system Babesiosis, 2010-2013. Merck Sharp and Dohme Corp, A subsidiary of Merck & Co, Inc, White House Station, N.J, U.S.A. http://www.merckmanuals.com/vet/circulatory_system/blood_parasites/babesiosis.html.

Thompson, J., van Spaendonk, R. M. L., Choudhuri, R., Sinden, R. E., Janse, C. J., and Waters, A. P. 1999. *Mol. Microbiol.* **31**: 253-260.

Twine, N.A., Janitz, K., Wilkins, M.R., and Janitz, M. 2011. Whole Transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. *PLoS ONE.* **6** (1): e16266. doi:10.1371/journal.pone.0016266.

Uilenberg, G., Franssen, F.F., Perie, N.M. and Spanjer, A.A. 1989. Three groups of *Babesia canis* distinguished and a proposal for nomenclature. *Vet. Quart.* **11**: 33-40.

Vanier, E. and Krause, P.J. 2009. Update on babesiosis. *Interdiscip. Perspec. Infects. Dis.* ID 984568, doi:10.1155/2009/984568.

Van Heerden, J. 1980. The transmission of *Babesia canis* to the wild dog *Lycaon pictus* (Temminck) and black-backed jackal (*Canis mesomelas*) Schreber, *J. S. Afr. Vet. Assoc.* **51**: 119-120.

Van Spaendonk, R.M.L., Ramaesar, J., Van Wig coherent, A., *et al.* 2001. Functional equivalence of structurally distinct ribosomes in the malaria parasite. *Biol. Chem.* **276**: 22638-22647.

Vial, H.J. and Gorenflot, A. 2006. Chemotherapy against babesiosis. *Vet. Parasitol.* **138** (1-2): 147-160.

Wang, Z., Gerstein, M. and Snyder, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Gen.* **10**: 57-63.

Wang, Z., Fang, B., Chen, J., Zhang, X., and *et al.* 2010. *De novo* assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweetpotato (*Ipomoea batatas*). *BMC Genomics.* **11** (726): 1-14.

- Wang, Y., Ghaffari, N., Johnson, C.D., Braga-Neto, U.M., Wang, H., Chen, R. and Zhou, H. 2011. Evaluation of the coverage and depth of transcriptome by RNA-Seq in chickens. *BMC Bioinformatics*. **12**: 1-7.
- Waters, A. P., van Spaendonk, R. M. L., Ramesar, J., Vervenne, H. A. W., Dirks, R. W., Thompson, J. and Janse, C. J. 1997. Species-specific regulation and switching of transcription between stage-specific ribosomal RNA genes in *Plasmodium berghei*. *J. Biol. Chem.* **272**, 3583-3589.
- Welzl, C., Leisewitz, A.L., Jacobson, L.S., Vaughanscott, T. and Myburgh, E. 2001. Systemic inflammatory response syndrome and multiple-organ damage/dysfunction in complicated canine babesiosis. *J. S. Afr. Vet. Assoc.* **72** (3): 158-162.
- Xu, T., Ping, J., Yu, Y., Yu, F., Yu, Y., Hao, P., *et al.* 2010. Revealing parasite influence in metabolic pathways in Apicomplexa infected patients. *BMC Bioinformatics*. 11: 1-9.
- Yamane, I., Thomford, J.W., Gardner, I.A., Dubey, J.P., Levy, M. and Conrad, P.A. 1993. Evaluation of the indirect fluorescent antibody test for diagnosis of *Babesia gibsoni* infections in dogs. *Am. J. Vet. Res.* **54**: 1579-1584.
- Ye, J., Fang, L., Zhang, Y., Chen, J., *et al.* 2006. WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* **34**: 293-297.
- Yeagley, T.J., Reichard, M.V., Hempstead, J.E., Allen, K.E., Parsons, L.M., White, M.A., Little, S.E. and Meinkoth, J.H. 2009. Detection of *Babesia gibsoni* and the canine small *Babesia* 'Spanish isolate' in blood samples obtained from dogs confiscated from dog fighting operations. *J. Am. Vet. Assoc.* **235** (5): 535-539.
- Yokoyama, N., Okamura, M. and Igarashi, I. 2006. Erythrocyte invasion by *Babesia* parasites: current advances in the elucidation of the molecular interactions between the protozoan ligands and host receptors in the invasion stage. *Vet. Parasitol.* **138** (1-2): 22-32.
- Zhang, J., Liang, S., Duan, J., Chen, S., Cheng, Z., Zhang, Q., Liang, X. and Li, Y. 2012. *De novo* assembly and characterisation of the transcriptome during seed

development, and generation of genic-SSR markers in Peanut (*Arachis hypogaea* L.). *BMC Genomics*. **13** (90): 1-6.

Zhou, M., Cao, S., Luo, Y., Liu, M., Wang, G., Moumouni, P.F.A., et al. 2016. Molecular identification and antigenic characterization of a merozoite surface antigen and a secreted antigen of *Babesia canis* (*BcMSA1* and *BcSA1*). *Parasit. Vectors*. **9** (257): 1-11.

APPENDIX

Table 3: DNA concentration and 260/280 ratio values from the 20 domestic dog samples.

Sample name	DNA concentration (ng/ μ L)	260/280 ratio
Dog 1	-0.22	0.40
Dog 2	4.22	3.25
Dog 3	4.29	-20
Dog 4	0.76	0.33
Dog 5	633.39	1.79
Dog 6	898.57	1.83
Dog 7	307.34	1.61
Dog 8	56.18	2.00
Dog 9	-3.22	1.67
Dog 10	-3.13	1.21
Dog 11	33.58	1.75
Dog 12	97.88	2.01
Dog 13	39.22	2.12
Dog 14	37.03	2.18
Dog 15	15.43	1.66
Dog 16	160.25	1.93
Dog 17	105.37	1.97
Dog 18	30.65	2.07
Dog 19	60.03	1.93
Dog 20	161.05	1.95

Table 4: RLB and qualitative PCR results of 20 DNA samples.

Sample name	Pathogen DNA detected by RLB	Pathogen DNA detected by Real-time PCR
Dog 1	No nucleic acid detected	<i>B. rossi</i>
Dog 2	<i>B. rossi</i>	<i>B. rossi</i>
Dog 3	<i>B. rossi</i>	<i>B. rossi</i>
Dog 4	<i>B. rossi</i>	<i>B. rossi</i>
Dog 5	<i>B. rossi</i>	<i>B. rossi</i>
Dog 6	<i>B. rossi</i>	<i>B. rossi</i>
Dog 7	<i>B. rossi</i>	<i>B. rossi</i>
Dog 8	<i>B. rossi</i>	<i>B. rossi</i>
Dog 9	<i>B. rossi</i>	<i>B. rossi</i>
Dog 10	No nucleic acid detected	<i>B. rossi</i>
Dog 11	<i>B. rossi</i>	<i>B. rossi</i>
Dog 12	<i>B. rossi</i>	<i>B. rossi</i>
Dog 13	<i>B. rossi</i>	<i>B. rossi</i>
Dog 14	<i>B. rossi</i>	<i>B. rossi</i>
Dog 15	<i>B. rossi</i>	<i>B. rossi</i>
Dog 16	<i>B. rossi</i>	<i>B. rossi</i>
Dog 17	<i>B. rossi</i>	<i>B. rossi</i>
Dog 18	<i>B. rossi</i>	<i>B. rossi</i>
Dog 19	<i>B. rossi</i>	<i>B. rossi</i>
Dog 20	<i>B. rossi</i>	<i>B. rossi</i>

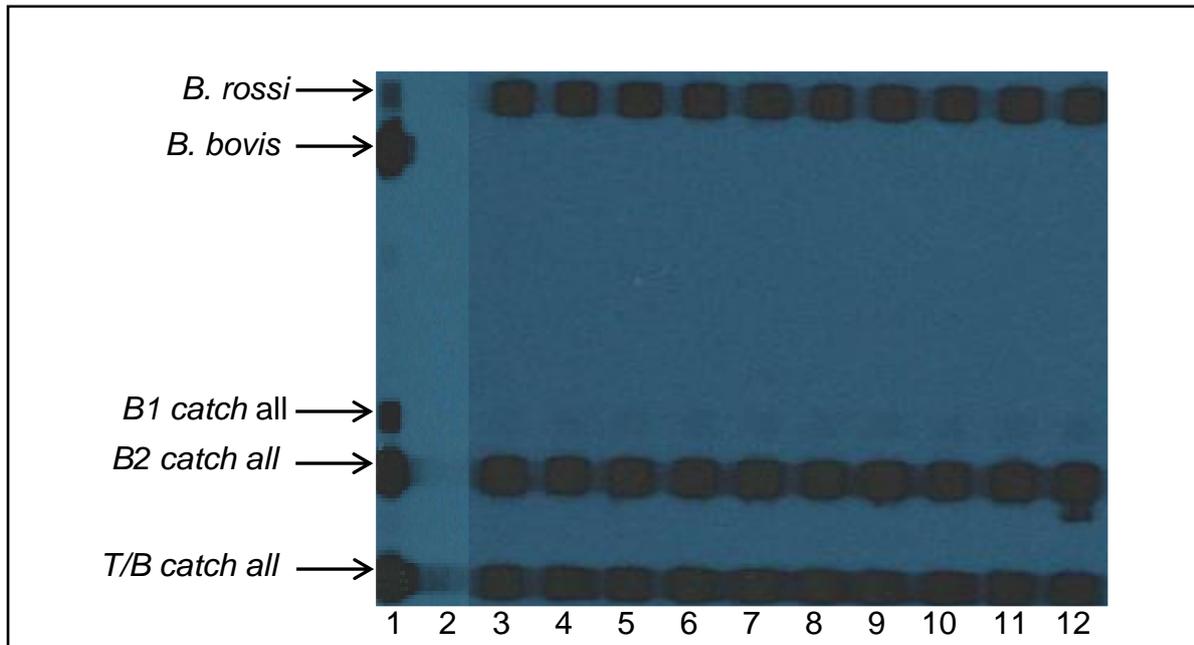


Figure 5: Representation of RLB hybridization analysis for detection of *B. rossi* DNA in blood samples collected from sick domestic dogs. Lane 1: positive controls, lane 2: negative control, lane 3-12: PCR products.

Table 5: Samples and their melting temperatures on the qualitative PCR.

Sample 1D	Melting temperature (Tm)(°C)
Dog1	72.26
Dog2	72.19
Dog3	72.35
Dog4	82.57
Dog5	82.77
Dog6	82.82
Dog7	82.68
Dog8	82.67
Dog9	72.51
Dog10	82.27
Dog11	82.52
Dog12	82.97
Dog13	82.57
Dog14	82.75
Dog15	77.01
Dog16	82.34
Dog17	82.35
Dog18	82.49
Dog19	82.40
Dog20	82.68

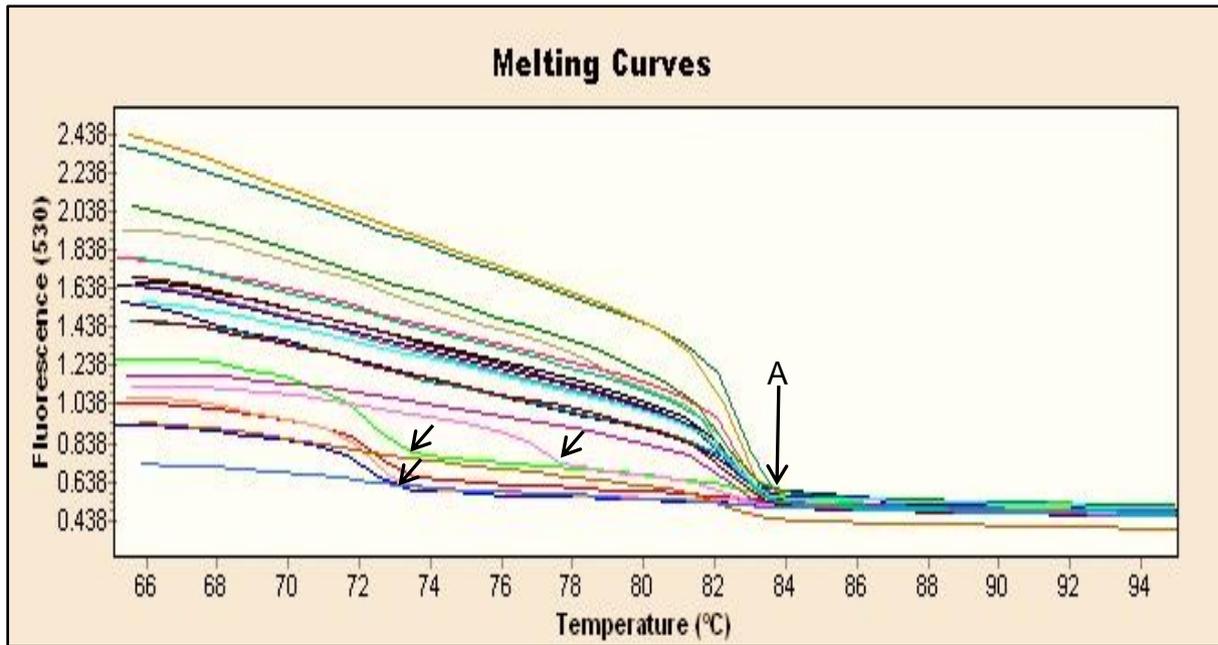


Figure 6A: Melting curves of *BrEMA1* gene in positive *B. rossii* DNA samples (A and small arrows) collected from sick domestic dogs.

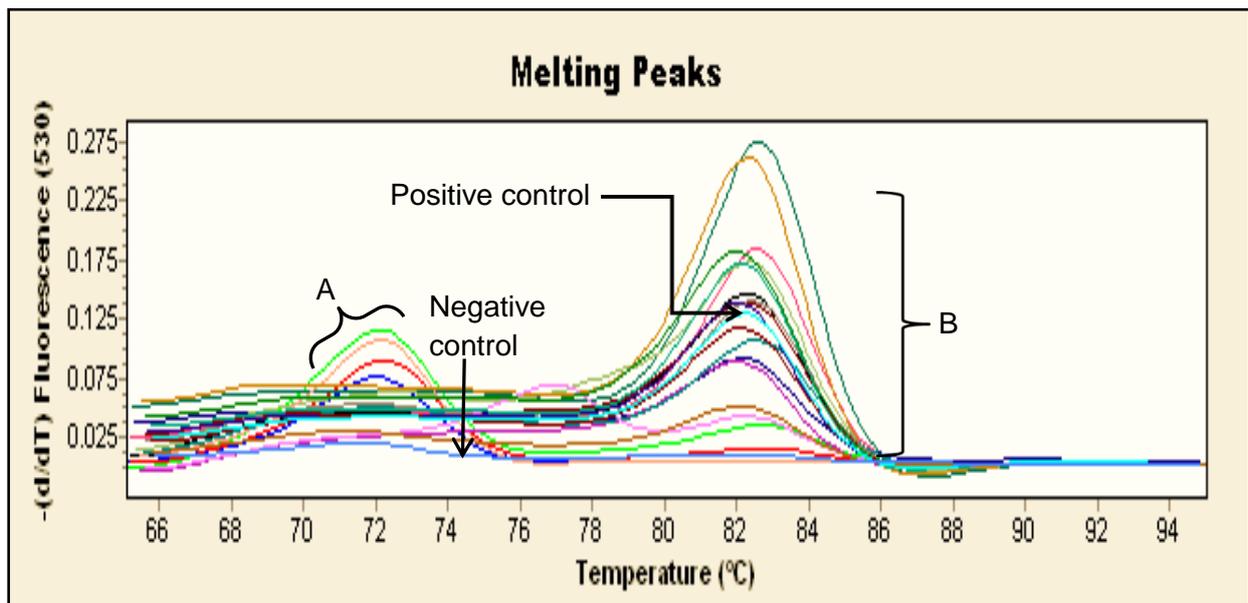


Figure 6B: Melting peaks of *BrEMA 1* gene in positive DNA samples (A, B) collected from sick domestic dogs.

Table 6: RNA concentration and 260/280 ratio values from 20 domestic dog samples.

Sample name	RNA concentration (ng/μL)	260/280 ratio
Dog 1	12.78	2.44
Dog 2	10.90	2.25
Dog 3	20.31	2.01
Dog 4	5.98	1.97
Dog 5	73.82	1.96
Dog 6	9.65	2.09
Dog 7	67.78	1.88
Dog 8	284.98	1.99
Dog 9	16.64	1.79
Dog 10	133.85	1.93
Dog 11	11.59	2.92
Dog 12	6.06	18.66
Dog 13	49.40	2.04
Dog 14	-5.41	1.43
Dog 15	7.44	2.16
Dog 16	110.79	2.14
Dog 17	198.33	2.03
Dog 18	-1.52	2.33
Dog 19	16.29	2.24
Dog 20	190.66	2.06

Analysis of *Babesia rossi* transcriptome in dogs diagnosed with canine babesiosis

Keneilwe Peloakgosi-Shikwambani ¹KP Sibeko ² PT Matjila ²

¹ Department of Life and Consumer Sciences, University of South Africa, South Africa

² Department of Veterinary Tropical Diseases, Faculty of Veterinary Science,
University of Pretoria, South Africa

*Corresponding author

E-mail: pkeneilwe@yahoo.com (K)

Abstract

Background: Canine babesiosis induced by *B. rossi* still remains the cause of mortality and morbidity in South African dogs, yet, the transcriptomic information of this parasite species is still not available. A previous study suggested a relationship between the parasite genotype and the disease phenotype. Therefore, the aim of this study was to sequence RNA and assemble the transcriptomes from samples representing *B. rossi* genotypes, 19, 29 and 31, in order to have insight on the overall transcriptome of this parasite.

Results: The sample representing *B. rossi* genotype 31 was excluded in the transcriptome analysis due to low RNA mass. Thus, a total of 26 747 238 and 25 709 627 paired-end reads were obtained from *B. rossi* genotypes 19 and 29, respectively. *De novo* transcriptome assembly produced a total of 3019 unigenes, with an average length of 419 bp and N50 of 362 bp in *B. rossi* genotype 19, and 2727 unigenes with an average length of 441 bp and N50 of 362 in *B. rossi* genotype 29. A total of 1193 unigenes were common between *B. rossi* genotype 19 and 29, while 1828 unigenes were exclusively detected in *B. rossi* genotype 19; and 1534 were specific to *B. rossi* genotype 29. Between the two *B. rossi* genotypes, a total of 4553 unigenes were obtained, representing the overall *B. rossi* transcriptome. From the overall transcriptome, 12.3% (n=558) of the unigenes could be annotated with 53 different gene ontology (GO) functional categories. About 34% (n=1550) of the unigenes represented in the overall transcriptome mapped to 237 KEGG pathways and only 2.5% (114) could be annotated in the COG database.

Conclusion: There were no striking differences in the transcriptomes of *B. rossi* genotypes 19 and 29, however, this study presents the first transcriptomic resource

for *B. rossi*, which will highly contribute to our genetic understanding of *B. rossi* and provide a platform for future gene expression studies

Keywords: *B. rossi* genotypes, *de novo* analysis, transcriptome, canine babesiosis, RNA-sequencing.

Author's summary

Tick-transmitted blood parasites cause some of the serious and poorly controlled diseases, including the most complicated form of canine babesiosis caused by *Babesia rossi* in domestic dogs. Therefore, there is a need to understand the mechanisms leading to the development of range of complicated clinical syndromes for controlling parasite transmission and the disease. This study reports the first transcriptomic data of *Babesia rossi*, which will assist researchers in understanding the complicated clinical syndromes often seen in dogs diagnosed with canine babesiosis induced by *B. rossi*.

Introduction

Babesia rossi is the most virulent apicomplexan tick-transmitted parasite responsible for causing canine babesiosis in dogs [1]. Canine babesiosis is the common cause of complicated clinical syndromes in dogs in South Africa [2]. The harmful effect of *B. rossi* is not only restricted to the health issues of the dog but has a negative impact on the economy as well, costing the dog owners R20 million per year in diagnosis and treatment of canine babesiosis in South Africa [3]. Hence, understanding the mechanisms leading to the range of complicated clinical syndromes development is very important for controlling parasite transmission and the disease. The clinical manifestations of canine babesiosis include fever, severe anaemia, shock, splenogamy, multiple organ failure, such as acute renal failure (ARF), acute respiratory distress syndrome (ARDS), and hepatic dysfunction, and in some cases *B. rossi* infection result in death of the infected dogs [4].

Over the years, the approach to determine genome sequences of an organism has greatly improved the efficiency and speed of gene discovery [5]. Other recent advances in sequencing technologies, such as transcriptome (also known as RNA sequence) analysis, have been developed [6]. Transcriptome reflects cellular activity within a tissue at a given point in time inside the organism [7]. Genome and transcriptome sequencing of various apicomplexan parasites, such as *Plasmodium falciparum*, *Theileria parva*, *Babesia bigemina* and *Babesia bovis* have been completed [8, 9, 10]. The transcriptome analysis of the aforementioned organisms improved the understanding of gene expression in connection with parasite-host interaction [11].

Apicomplexans, such as *Babesia*, are intracellular parasites that have a close relationship with their hosts and therefore have established pathways that help in their transmission [12]. Transcriptome analysis has allowed gene expression studies to identify the genes whose transcripts are up or down regulated during the parasite's life cycle in order to improve the understanding of the parasite's development and differentiation in the host [12]. A previous report linked *B. rossi* with varying clinical signs and life-threatening disease in the dog host; however, the underlying mechanisms of such association are still unknown [13]. These mechanisms could be determined from genomic or transcriptome information of *B. rossi*. Thus, this study presents the first ever report of the transcriptome sequence and analysis of *B. rossi* parasite.

Results and discussion

Sequence analysis and *de novo* assembly

Three cDNA samples representing *Babesia rossi* genotypes 19, 29 and 31 were prepared and sequenced using HiSeq® 2000 system. Post transcriptome sequencing, raw reads containing adaptors, unknown nucleotides larger than 5% and low quality reads were removed from the transcriptome data using the filterFQ software. After quality assessment and data filtering, the sample representing *B. rossi* genotype 31 was excluded in the transcriptome analysis due to low RNA mass, which usually compromises the quality of the library used in RNA sequencing [14]. According to [10], the genomic comparison between three virulent and attenuated *B. bovis* strains showed no changes among the protein coding genes shared between the strains. This information led to the conclusion that the exclusion of *B. rossi* genotype 31 may not have a negative impact on the whole *B. rossi* transcriptome.

Babesia rossi genotype 19 generated a high number of clean paired-end reads, with a slightly high GC content and Q20 compared to genotype 29 (Table 1). The GC content plays an important role in the gene and genome regulation and assists to determine the physical properties of the genome [15]. The GC content was comparable at 62.63% in *B. rossi* genotype 19 and 61.18% in *B. rossi* genotype 29. This is higher than the GC content of other apicomplexan parasites, such as *B. bovis* (41, 8%), *Plasmodium falciparum* (19%) and *Theileria parva* (34, 1%) [16]. This finding could indicate the stability of the *B. rossi* genome.

The assembly of reads from genotypes 19 and 29 revealed that *B. rossi* genotype 19 had a high number of clean paired-end reads compared genotype 29. Using Trinity program, *de novo* assembly generated less numbers of contigs with a short mean

length and N50 in genotype 19 compared to its counterpart (Table 1). Following the clustering of contigs into unigenes and exclusion of duplicate unigenes, the number of unigenes generated in genotype 19 was increased significantly, but with a shorter mean length than genotype 29. The less virulent *B. rossi* genotype 29 had reduced number of unigenes than the most severe *B. rossi* genotype 19. Therefore, this data proposed that the difference between the virulent and attenuated strains may lie in the non-coding regions which have an effect on the transcriptomic variability, or the loss of genome content during attenuation process [10]

Both genotypes had the same unigene N50 value subsequent to this analysis (Table 1). The length of the sequences and the N50 are often used to measure the quality of reads assemblies [17]. In *B. rossi* transcriptomes from the different genotypes, similar N50 values were obtained. The slight varying average length and same N50 indicated that there was no biasness in the quality of the two assemblies.

Table 1: Summary statistics of transcriptome assemblies of *B. rossi* genotypes 19 and 29.

	<i>B. rossi</i> genotype 19	<i>B. rossi</i> genotype 29
Total number of clean paired-end reads	26 747 238	25 709 672
GC content	62.63%	61.18%
Q20	94.69%	94.49%
Total number of contigs	50 839	62 308
Mean contig length(bp)	225	230
Contig N50	250	258
Total number of unigenes	3 019	2727
Unigene mean length (bp)	419	441
Unigene N50	362	362

Sequence annotation

The unigenes from *B. rossi* genotypes 19 and 29 were uploaded on Venn diagram (<http://bioinformatics.psb.ugent.be/webtools/Venn/>) to identify the unigenes unique to each *B. rossi* genotype and those common between the two genotypes. The combined unigenes from assembled transcriptome of *B. rossi* genotypes 19 and 29 resulted in 4553 unigenes detected between the two genotypes; notwithstanding that each genotype contributed a varying number of unigenes, and the two genotypes shared a total of 1193 unigenes (Figure 1). A total of 4553 will hereon be referred to as combined unigenes (total number of unigenes detected for the *B. rossi* transcriptome)

for the purpose of this manuscript Nonetheless, the number of combined unigenes detected from *B. rossi* transcriptome were comparable to that of the apicomplexan parasite, *T. parva* (n=4,035). Other apicomplexam parasites with sequenced genome include *P. falciparum* (5438 unigenes) and *B. bovis* (3641 unigenes) [18].

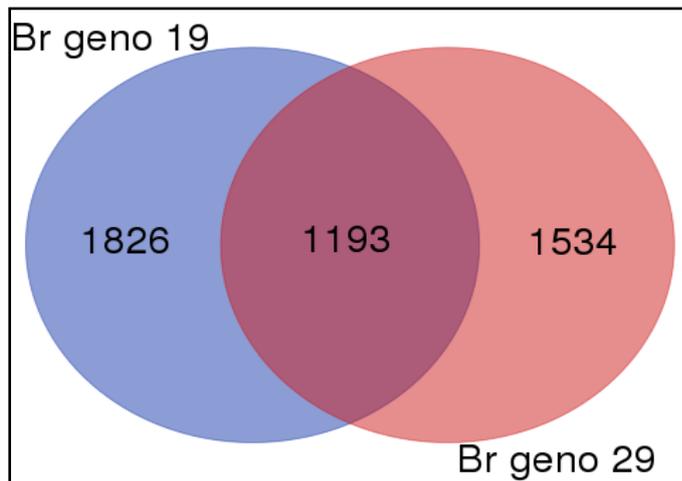


Figure 1: Venn diagram showing the distribution of the common and exclusively detected unigenes between the *B. rossi* genotypes 19 and 29. Numbers inside the coloured circles indicate the number of unigenes. Br geno 19= *B. rossi* genotype 19; Br geno 29= *B. rossi* genotype 29.

The combined unigenes from *Babesia rossi* transcriptome were compared with NCBI (full) non-redundant protein database (nr) with expectation (E) value of 1.0E-3 and 20 maximum number of hits to acquire desired output [19] (Altschul *et al.*, 1990). Overall, the number of unigenes decreased with an increasing length (101-1000bp). Only 4.8% (n=220) of detected unigenes had a length less than 100 bp and a total of 287 unigenes (6.3%) had a length more than 1 000 bp (Figure 2).

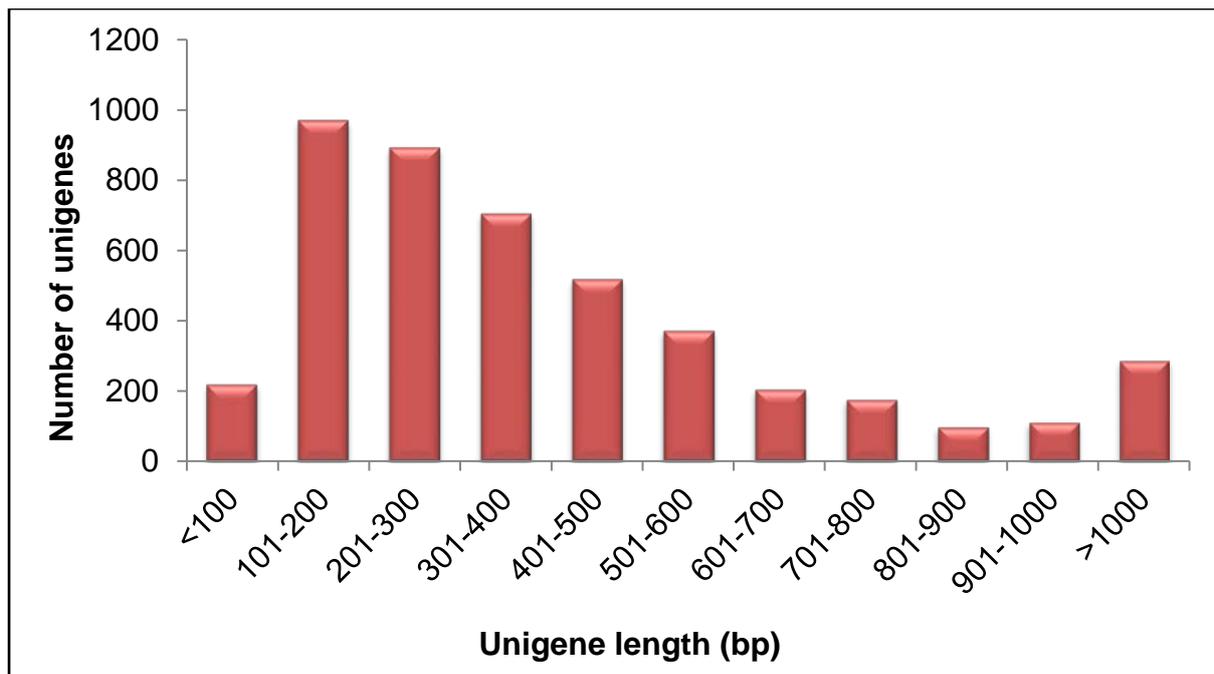


Figure 2: The length distribution of the combined unigenes in *Babesia rossi* transcriptome

Only 16.71% unigenes of the *B. rossi* transcriptome retrieved hits following analysis by BLAST and 100% had interproscan output. From these, 12.25% of the unigenes could be annotated based on the GO categories, and 14.01% were successfully mapped (Table 2). Regardless, all the unigenes in the *B. rossi* transcriptome were successfully annotated by InterProscan, which is an indication of coding sequences [20] and this shows that *de novo* assembled unigenes in *B. rossi* are coding for proteins characterized by known Interproscan domains. Functional annotation also revealed that 133 out of 4553 unigenes were identified as encoding hypothetical proteins.

Table 2: Data distribution of the combined unigenes detected from *Babesia rossi* transcriptome.

Category	Number of unigenes
Total sequences	4553
With interproscan	4553
With blast hits	761
With mapping	638
With annotation	558

Babesia rossi transcriptome blast sequence similarity analysis revealed the highest percentage of unigenes (41%) with products which have sequence similarity between 60-80% to known proteins; while 35% and 23% of unigenes had sequence similarity ranging from 40-60% and 80-100%, respectively. Only 1% of the unigenes had sequence similarity between 20%-40% to known proteins (Figure 3).

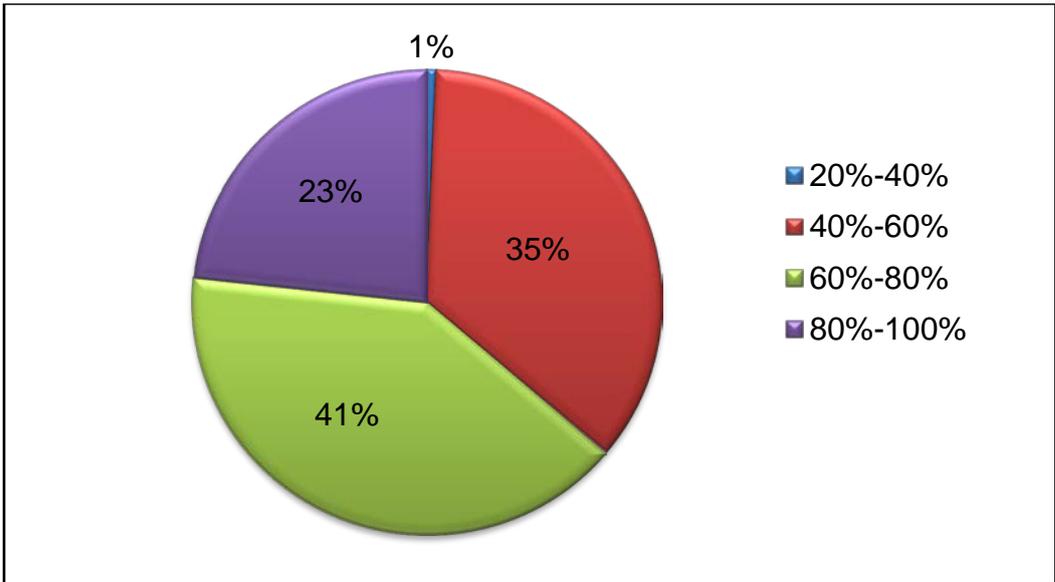


Figure 3: Sequence similarity of the blast hits of each unigene in *B. rossi* transcriptome.

The E-value (expectation value) describes the expected number of random hits when searching a database; thus, the lower the E-value, the more significant a match to a query [21]. A large percentage of combined unigenes had the E-value ranging between 1E-150 to 1E-180, followed by unigenes with the E-value between 1E-0 to 1E-30 and then unigenes with the E-value of 1E-30 to 1E-60. A smaller percentage of combined unigenes were distributed in 1E-60 to 1E150 clusters (Figure 4).

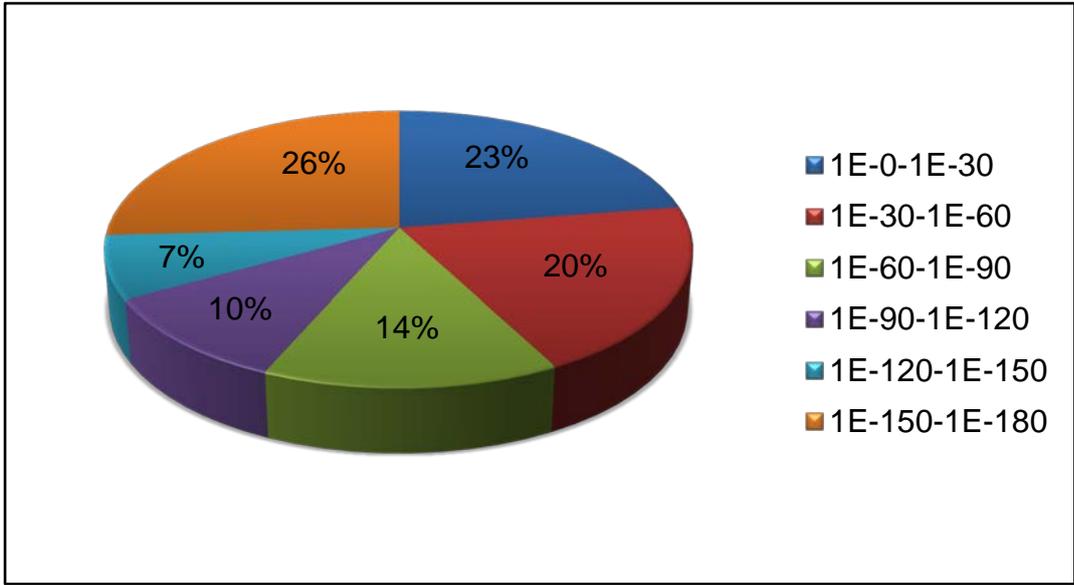


Figure 4: E-value distribution of the blast hits for unigenes from *B. rossi* transcriptome, using a cut-off value of 10E-3.

The considerable portion of the assembled unigenes had top blast hits with other apicomplexan parasites, mainly *Babesia spp.*, *Babesia bigemina* (41%), followed by *B. bovis* (26%) (Figure 5). Notably, only a small proportion of sequences returned hits against other *Babesia species*, such as, *B. vogeli*, *B. canis* and *B. gibsoni*, other *Babesia species* causing canine babesiosis in dogs [2]. This result could be due to the unavailability of genome sequences for these species. As a general observation from this study and others which used *de novo* assembly, there could be biasness in analysis towards the species with annotated genome because of the more accurate gene predictions and high number of sequences available in public sequence databases compared to non-model organisms [23].

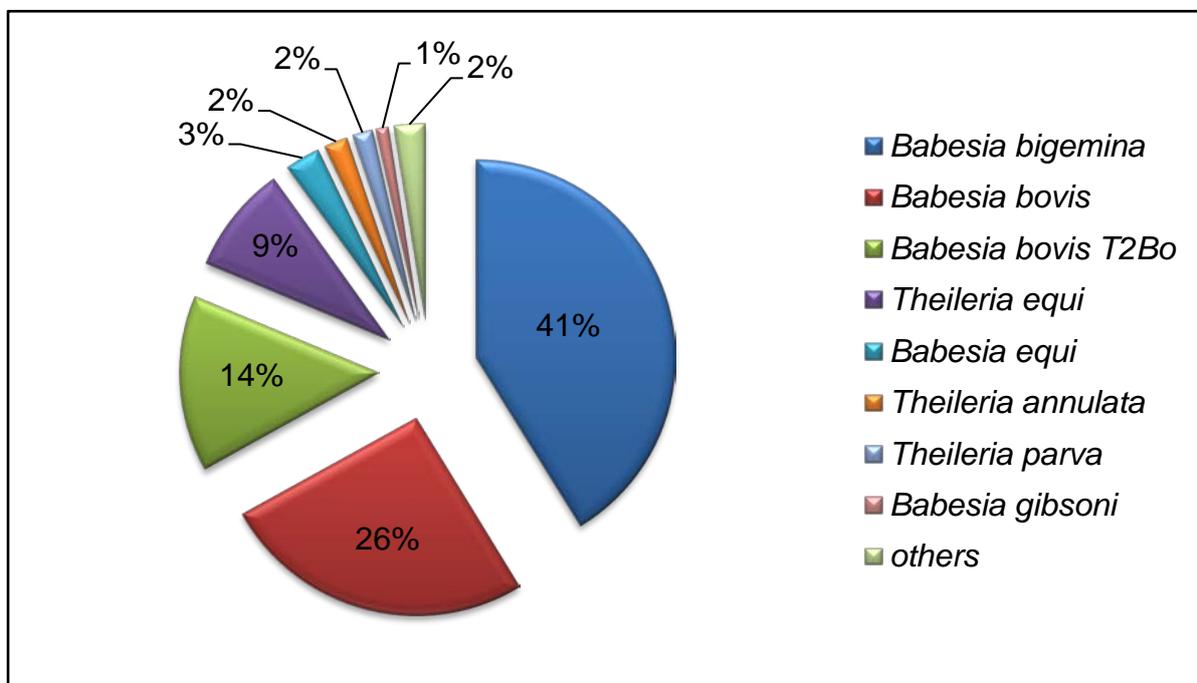


Figure 5: Species distribution for homologous sequences of *B. rossi* transcriptome unigenes.

Gene ontology (GO) annotation was also performed for the combined unigenes in the *B. rossi* transcriptome, in terms of biological process, molecular function and cellular component. Twelve percent of the combined unigenes could be associated with three ontology categories, the biological process (37.3%), molecular function (20.1%) and cellular component (42.7%). The cellular component category was the most enriched GO category consisting of 17 GO terms, with the most abundant subgroups such as 'cell', 'cell part', 'membrane', 'membrane part' and 'organelle' (Figure 6), implying that the *B. rossi* gene products are most active in the cell. The *Plasmodium falciparum* and *B. bovis* are able to attach to the capillary endothelium through sequestration [18], in which their proteins are expressed on the infected erythrocytes, which seem to transform the adhesive properties of the infected cells [24]. Cytoadherence is mediated by *P. falciparum* Erythrocyte Membrane Protein 1 (PfEMP1) and Variant Erythrocyte Surface Antigen 1 (VESA1) in *P. falciparum* [25] and *B. bovis* [17], respectively. Although, it has not been proven, the expression of the parasite-derived antigens, with adhesive properties on the extracellular surface of cells infected with *B. rossi* can be accountable for their ability to sequester *in vivo* and to aggregate *in vitro* [27].

The biological process category was the second most abundant, with 16 GO terms; the most enriched GO terms were the 'metabolic process', followed by the 'cellular process', 'single-organism process', 'localization' and 'biological' (Figure 6), which indicates the cellular and metabolic processes in the developmental stage of the parasite. As intracellular parasites, apicomplexans obtain nutrients and perform their energy requirements by manipulating the host metabolism [28], causing changes in host cell signalling, nutrient delivery to the parasite, protein synthesis, apoptosis and

evasion of host immunity [29]. Therefore, this could explain the survival of *B. rossi* parasite in its host.

The molecular function category had the least unigenes, with 8 GO terms. The most enriched GO terms in this category were mainly 'binding', 'catalytic activity', 'transporter activity', structural molecule activity and 'molecular function regulator' (Figure 6), suggesting high gene regulation and enzymatic activity processes in *B. rossi* parasite. Regarding 'binding', parasite proteins, such as *Babesia* RAP-1, are involved with host receptors during invasion and are commonly found in the apical organelles [30]. Members of the RAP-1 family in *Babesia*, such as *B. divergens*, are involved in the invasion of host cells through direct binding with receptors on the host cell surface [31]. Moreover, parasite ligands and their complementary red blood cell receptors in invasion stages are very important in understanding the invasion biology of *Babesia* parasites and provide information for the development of effective and prevention strategies [30].

.A broad range of GO terms representing the three main GO categories was detected from a considerable number of unigenes. This indicates that the transcriptome data expressed diverse range of unigenes in the *B. rossi* involved in different pathways. In many *de novo* transcriptome studies [32, 33, 34], many unigenes are represented in the biological process GO group, followed by metabolic function and cellular component GO categories [32, 33, 34]. However, in this study, the cellular component was the most abundant GO category, followed by the biological process and molecular function GO categories.

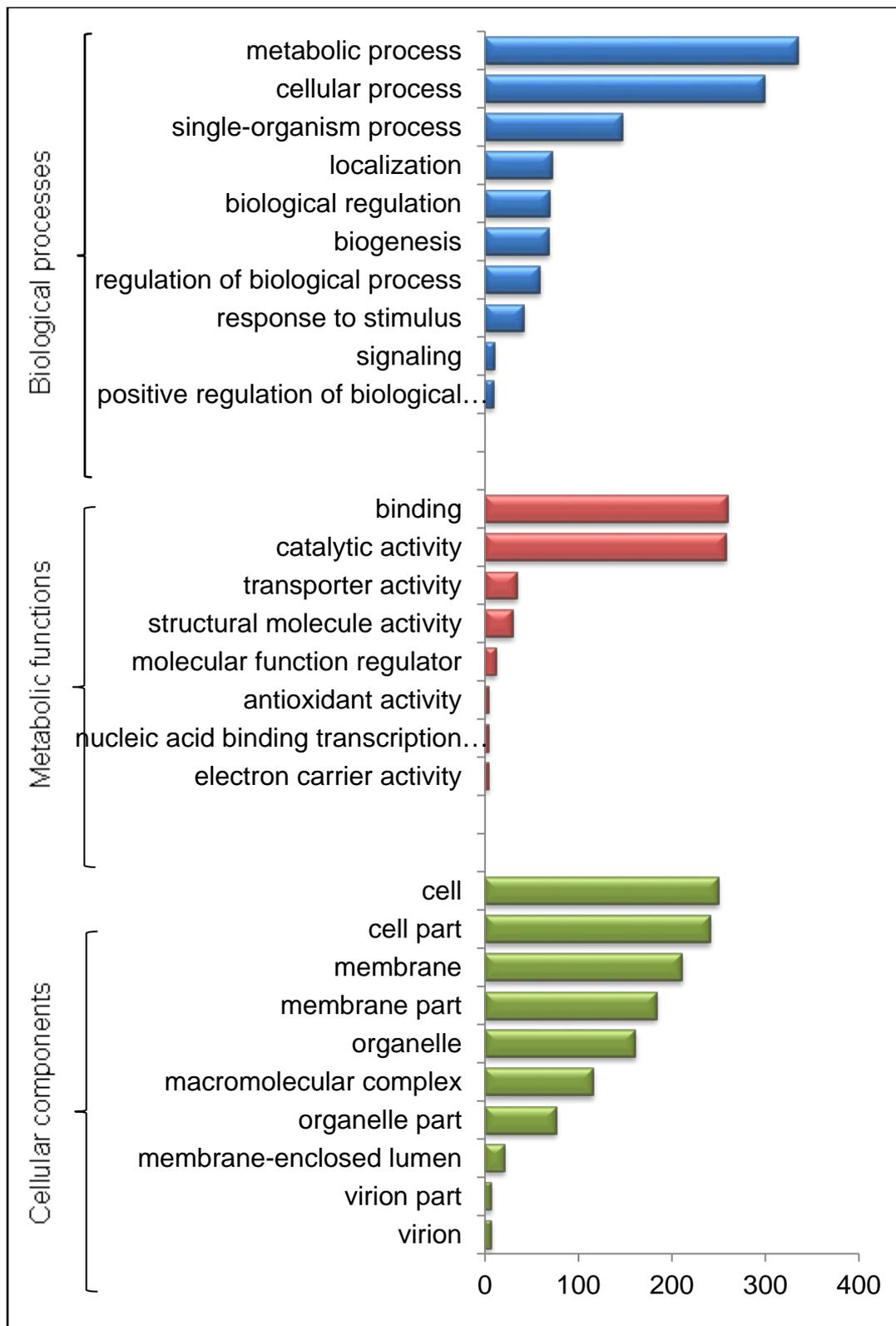


Figure 6: Top 10 most enriched GO terms, for specific ontology categories, detected from analysis of the combined unigenes of *Babesia rossi*. The x-axis indicates the gene ontology terms and the bar shows the number of unigenes.

The ultimate goal in many transcriptome studies is to annotate the sequences by connecting them to the biological information [32]. Low percentage of the combined unigenes (16.7%) had blast hits in the assembled *B. rossi* transcriptome, which might be due to the absence of *B. rossi* genome information in public databases. This finding shows the importance of generating the large collection of *B. rossi* sequences. Previous *de novo* transcriptome studies of other species reported the annotation rate ranging between 20%-40% [35] (Ye *et al.*, 2006). In this study, less than 20% of unigenes in the *B. rossi* transcriptome could match known proteins, which is lower than the previously reported *de novo* transcriptome studies. This could be attributed to many unigenes with short length resulting in annotation of few unigenes. Alternatively, the unmatched unigenes could be unique to *B. rossi*.

Besides GO annotation, KEGG pathway analysis was also carried out for the combined unigenes in the *B. rossi* transcriptome, which is an alternative approach to categorize genes functions with the emphasis on biochemical pathways. A total of 1550 unigenes in the *B. rossi* transcriptome could be assigned to 263 KEGG pathways. Among these, a large number of unigenes were assigned to metabolic pathways, followed by ribosomes, then biosynthetic of secondary metabolites, biosynthesis of antibiotics, splicing, microbial metabolism in diverse environments, carbon metabolism, purine metabolism, RNA transport and pyrimidine metabolism (Figure 7). Metabolism was the largest category in KEGG pathways, indicating that active metabolic processes take place in *B. rossi* infection.

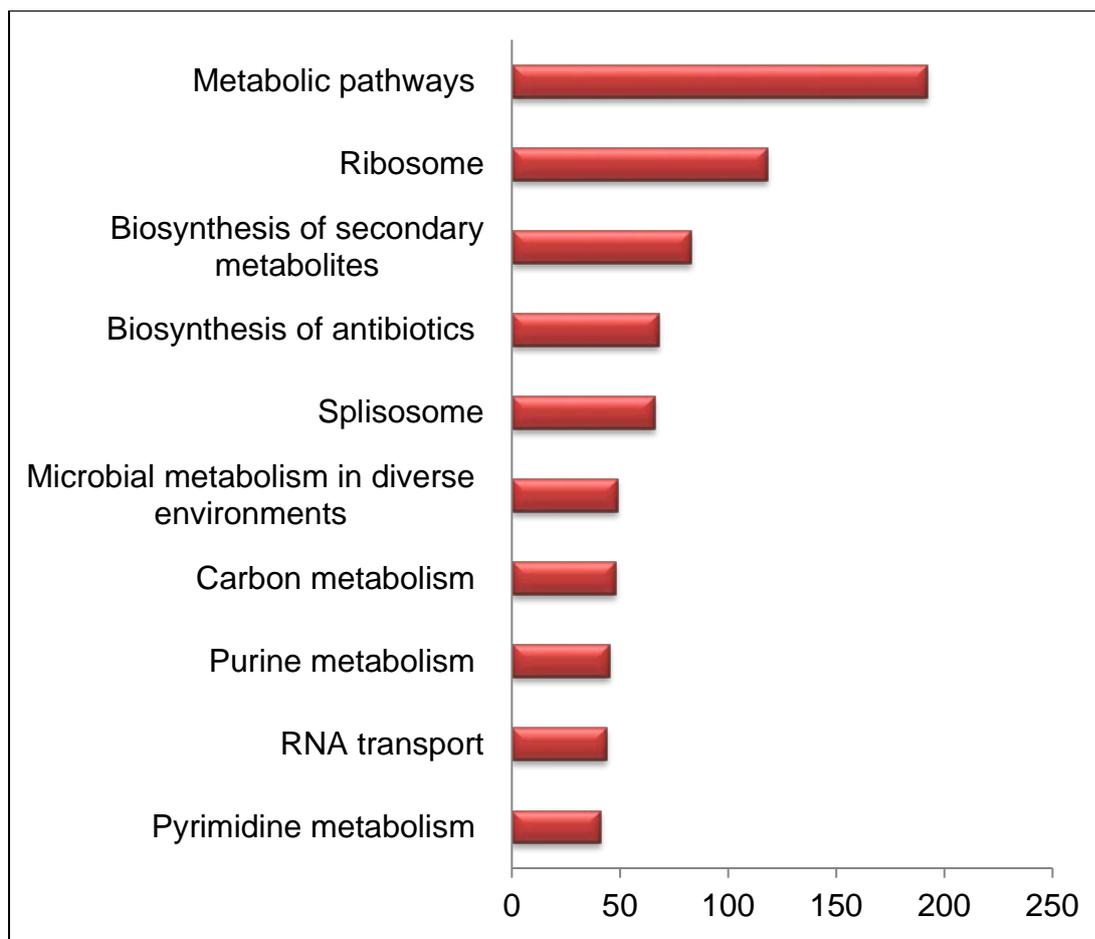


Figure 7: Top 10 most enriched pathways detected from the combined unigenes of *Babesia rossi*. The x-axis indicates the pathway and the bar shows the number of unigenes.

In addition, the combined unigenes in the *B. rossi* transcriptome were subjected to a search against the COG database for functional prediction and classification. Out of 4553 combined unigenes, only 114 (2.5%) unigenes were clustered into 8 categories of COG. The largest category was 'nucleotide transport and metabolism (n=60; 52.6%), the second was the 'general function prediction only (n=17; 14.9%) followed by 'function unknown' (n=16; 14%) (Figure 8). Only one unigene was represented under 'carbohydrate transport and metabolism, 'DNA replication, recombination and repair' and 'energy production' each (Figure 8). Nucleotide and transport metabolism' category was the most enriched COG in *B. rossi* transcriptome. For example,

parasites, such as *P. falciparum* and *Cryptosporidium*, depend on nucleotide uptake from the host, and are incapable of *de novo* amino acid synthesis and therefore have reliance upon salvage from the host [36]. The same mechanism could be taking place in *B. rossi* parasite. Notably, 11 out of 25 COG categories were identified; this could be due to the unavailability of the *B. rossi* unigenes in the DAVID software used. However, these data will provide information which would help to explore major genes for molecular mechanisms involved in *B. rossi* parasite infection

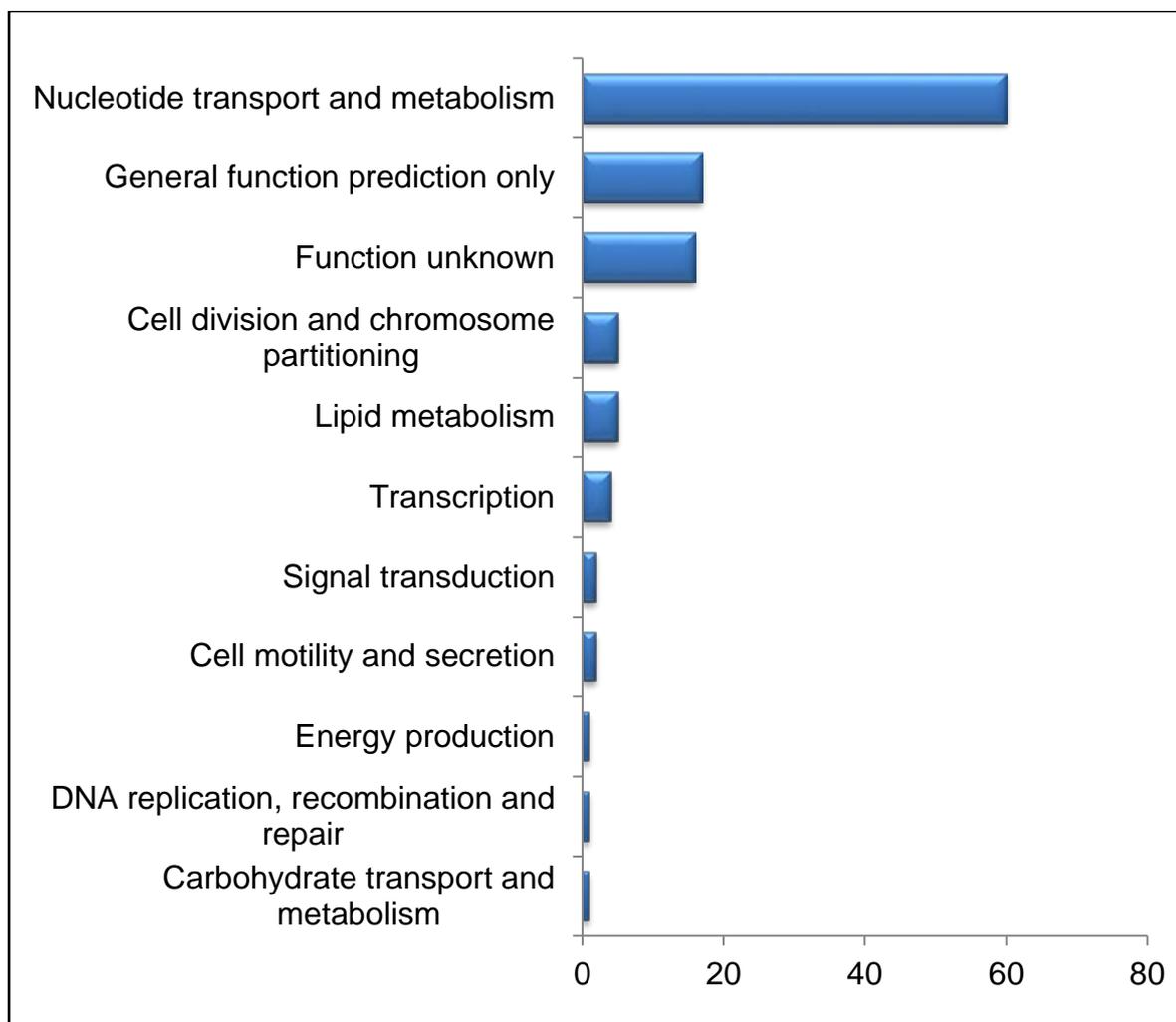


Figure 8: Cluster of Orthologous Groups (COG) classification of the combined unigenes of *Babesia rossi* transcriptome.

Materials and methods

Ethics statement.

Ethical clearance was granted by the UNISA Ethics Committee prior to resuming with the research activities of the study. Another ethical clearance had also been obtained from the University of Pretoria Animal Ethics Committee through Prof Andrew Leisewitz for his parallel study on canine babesiosis, and he provided blood samples from sick domestic dogs for the current study.

Sample collection

Forty clinical blood samples were withdrawn from the cephalic vein of the 20 sick domestic dogs, i.e. two clinical blood samples from each dog for DNA and RNA extractions. Based on the microscopic examination of the blood smears at the Onderstepoort Veterinary Academic Hospital (OVAH), Faculty of Veterinary Science, University of Pretoria, the clinical blood samples were diagnosed with canine babesiosis.

RNA extractions

The total RNA was extracted from the 20 *B. rossi*-infected blood samples collected from the same dogs using the RNeasy® Protect Animal Blood Kit (QIAGEN) according to the manufacturer's instructions. The quality and quantity of the RNA samples was determined using the Nanodrop® ND-1000 Spectrophotometer (Thermo Scientific) at an absorbance of 260/280 nm at the Department of Veterinary and Tropical Diseases, University of Pretoria, prior to RNA sequencing. Three total RNA samples infected with *B. rossi* genotypes 19, 29 and 31 were prepared from three dogs showing clinical signs associated with each of the three genotypes. These three total RNA samples fulfilled the standard sample requirements for transcriptome sequencing at the Beijing

Genomics Institute (BGI) in Hong Kong, namely: mass \geq 2 μ g, concentration \geq 20 ng, 28S/18S ratio \geq 1 and RIN \geq 6.5. The RNA samples were submitted to Beijing Genomics Institution, Hong Kong for library construction and transcriptome sequencing.

mRNA-seq Library Construction for Illumina Sequencing

Briefly, the three libraries from *B. rossi* genotypes 19, 29 and 31 were prepared using the Illumina TruSeq RNA Sample Preparation kit v2 (Illumina, USA) at BGI, Hong Kong as follows:

mRNA isolation and fragmentation: Two hundred (200) ng of total RNA was purified by oligo-dT beads, then poly (A)-containing mRNA was fragmented into small pieces with Elute/ prime/ fragment mix.

cDNA synthesis: First-strand cDNA was generated from fragmented mRNA using the first strand master mix and superscript II reverse transcriptase (Invitrogen) at 25°C for 10 min; 42°C for 50 min and 70°C for 15 min. Subsequently, the second strand master mix was added to synthesize the second strand cDNA at 16°C for one hour.

End repair, A-tailing and adaptor ligation: The purified fragmented cDNA was combined with the end repair mix and incubated at 30°C for 30 min. The end-repaired DNA was purified with Ampure XP Beads (Agencourt). The A-tailing mix was then added and the solution was mixed by pipetting before incubation at 37°C for 30 min.

Adapter ligation: In this step, the adenylate 3' ends DNA, RNA Index adapter and ligation mix were combined and mixed well by pipetting. The ligation reaction was incubated at 30°C for 10 min and the end-paired DNA was purified with Ampure XP Beads (Agencourt).

Polymerase Chain Reaction (PCR): To enrich the cDNA fragments, PCR was performed using PCR primer cocktail and PCR master mix. The PCR products were then purified with Ampure XP Beads (Agencourt).

Library quantitation and sequencing: The average molecule length was determined using the Agilent 2100 Bioanalyzer instrument (Agilent DNA 1000 Reagents); and the library was quantified by real-time quantitative PCR (qPCR) (TaqMan Probe). The qualified libraries were amplified on cBot to generate the cluster on the flow cell (TruSeq PE Cluster Kit V3-cBot-HS, Illumina) and sequenced paired-end reads on the HiSeq[®] 2000 system (TruSeq SBS KIT-HS V3, Illumina).

Sequence Data Analysis and Assembly

Post transcriptome sequence analysis, raw reads which contained adaptors, unknown nucleotides larger than 5% and low quality reads were removed from the transcriptome data using the filterFQ software. Transcriptome *de novo* assembly of the three *B. rossi* genotypes was performed with short reads assembling program, the Trinity software package (release 2013-02-25) as described in [37].

Sequence Annotation

The unigenes obtained from *B. rossi* genotype 19, 29 and 31 were used as queries in BLASTx against the NCBI (full) non-redundant protein database (nr) with expectation (E) value of 1.0E-3 and 20 maximum number of hits applied to acquire desired output (Altschul *et al.*, 1990). Subsequently, unigenes from *B. rossi* genotypes 19, 29 and 31, were uploaded on Venn diagram (<http://bioinformatics.psb.ugent.be/webtools/Venn/>) to identify the unigenes unique to each *B. rossi* genotype and those common between the three genotypes. Functional annotation of unigenes from each *B. rossi* genotype was performed on the Blast2GO program, version 2.8.0 software. The KEGG (

Encyclopedia of Genes and Genomes) database (<http://www.genome.jp/kegg/mapper.html>) was used to annotate the biological pathways of all unigenes from the three *B. rossi* genotypes. The unigenes from the three *B. rossi* genotypes were further analysed into gene families using DAVID (the Database for Annotation, Visualization and Integrated Discovery version 6.7 (<https://david.ncifcrf.gov>), to predict and classify their functions.

Conclusions

In conclusion, the transcriptome of the two *B. rossi* genotypes representing different clinical manifestations on the host infected with *B. rossi* have been sequenced, assembled and functionally characterized. To our knowledge, this is the first study reporting on the transcriptome of *B. rossi*. Therefore, including samples from strains

causing different clinical diseases was crucial in order to provide a maximally integrative transcriptome. Although, there were no significant differences between the *B. rossi* genotypes 19 and 29, the unigenes that were assigned to GO terms were mostly involved in biological processes, metabolic functions and cellular components, giving researchers a glimpse into the molecular nature of *B. rossi*. The generated transcriptomic data will also provide a valuable public resource of *B. rossi* gene sequences for future genetic and genomic studies of *B. rossi* parasite. Future studies could include gene expression analysis, which will allow determination of differentially expressed genes between these two *B. rossi* genotypes; this will assist in explaining the diverse clinical presentations often observed in dogs diagnosed with canine babesiosis. Furthermore, the functional roles of the hypothetical proteins identified in the *B. rossi* parasite need to be characterized and investigated for possible disease modulators, novel drug targets or vaccine candidates against the *B. rossi* infection.

Acknowledgements

We sincerely thank the following people and institutes: Prof Leisewitz for providing clinical samples, Dr Agatha Kolo for laboratory assistance, Dr Petunia Malatji for transcriptome analysis, Department of Veterinary and Tropical Diseases at the University of Pretoria for permission to use their facilities to perform the laboratory

work and Agricultural Research Council, Biotechnology Platform for allowing us to use their Trinity software for transcriptome data analysis.

Funding: This research project was supported by Unisa Postgraduate Bursary, National Research Foundation and University of Pretoria (Genomic Research Institute),

Author's contributions: Matjila T.P conceived and designed the experiments. Peloakgosi-Shikwambani K. performed the experiments. Peloakgosi-Shikwambani analyzed the data. Sibeko K. contributed analysis tools. Peloakgosi-Shikwambani K. wrote the paper.

References

1. Schetters TPM, Kleuskens JAGM, Crommert J, Leeuw PWJ, Finizio, AL, Gorenflot A (2009) Systemic inflammatory responses in dogs experimentally infected with *Babesia canis*; a haematological study. *Vet Parasitol* 162: 7-15.
2. Last RD, Hilla JM, Matjila PT. and Rème, CA (2007) A field trial evaluation of the prophylactic efficacy of amitraz-impregnated collars against canine babesiosis (*Babesia canis rossii*) in South Africa. *J S Afr Vet Assoc* 78 (2): 63-65.

3. Collett MG (2000) Survey of canine babesiosis in South Africa. *J S Afr Vet Assoc* 71 (3): 180-186.
4. Jacobson LS (2006) The South African form of severe and complicated canine babesiosis: clinical advances 1994-2004. *Vet Parasitol* 138: 126-139.
5. Ansorge WJ (2009) Next-generation DNA sequencing techniques. *Nature Biotechnology* 4: 195-203.
6. Pareek CS, Smoczynski R, Tretyn A (2011) Sequencing technologies and genome sequencing. *J Appl Genet* 52 (4): 413-435.
7. Twine NA, Janitz K, Wilkins MR, Janitz M (2011) Whole Transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. *PLoS ONE*. 6 (1): e16266. doi:10.1371/journal.pone.0016266.
8. Bishop R, Shah T, Pelle R, Hoyle D, Pearson T, Haines L, Brass A, Hulme H, Graham SP, Taracha ELN, *et. al.* 2005. Analysis of the transcriptome of the protozoan *Theileria parva* using MPSS reveals that the majority of genes are transcriptionally active in the schizont stage. *Nucleic Acids Res* 33 (17): 5503-5511.
9. Bozdech Z, Zhu J, Joachimiak MP, Cohen FE, Pulliam, B, *et al* (2003) Expression profiling of the schizont and trophozoite stages of *Plasmodium falciparum* with a long-oligonucleotide microarray. *Genome Biol* 4: 44-49.
10. Pedroni MJ, Sondgeroth KS, Gallego-Lopez GM, Echaide I. Lau AOT (2013) Comparative transcriptome analysis of geographically distinct virulent and attenuated *Babesia bovis* strains reveals similar gene expression changes through attenuation. *BMC Genomics* 14 (763): 1-13.

11. Jongejan F, Zivkovic D, Pegram RG, Tatchell RJ, Fison T, Latif AA, Paine G (1987) Ticks (*Acari:Ixodidae*) of the Blue and White Nile ecosystems in the Sudan with particular reference to the *Rhipicephalus sanguineus* group. *Exp Appl Acarol* 3:331-346.
12. Boyle JP, Saeij JPP, Cleary MD, Boothroyd JC (2006) Analysis of gene expression during development: lessons from the Apicomplexan. *Microb. Infect* 8: 1623-1630.
13. Matjila PT, Carcy B, Leisewitz AL, Schetters T, Jongejan F, Gorenflot A, Penzhorn BL (2009) Preliminary evaluation of the *BrEMA1* gene as a tool for associating *Babesia rossi* genotypes and clinical manifestation of canine babesiosis. *J Clin Microbiol* 47 (11): 3586-3592.
14. McCormic KP, Willman MR, Meyers BC (2011) Experimental design, preprocessing, normalization and differential expression analysis of small RNA sequencing experiment. *Silence* 2 (2): 1-19.
15. Carels N, Hatey P, Jabbari K, Bernardi G (1998) Compositional distribution of homologous coding sequences from plants. *J Mol Evol* 46: 45-43.
16. Hayashida K, Hara Y, Abe T, Yamasaki C, *et al* (2012) Comparative genome analysis of three eukaryotic parasites with differing abilities to transform leukocytes reveals key mediator of *Theileria*-induced leukocyte transformation. *mBio* 3 (5):e00204-12. doi:10.1128/mBio.00204-12.
17. Chauvin A, Moreau E, Bonnet S, Plantard O, Malandrin L (2009) *Babesia* and its hosts: adaptation to long-lasting interactions as a way to achieve efficient transmission. *Vet Res* 40 (37): 1-18.

18. Allred DR, Al-Khedery B (2004) Antigenic variation and cytoadhesion in *Babesia bovis* and *Plasmodium falciparum*: different logics achieve the same goal. *Mol. Biochem Parasitol* 134: 27-35.
19. Altschul SF, Gish W, Miller W, Miller EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403-410.
20. Pallavicini A, Canapa A, Barucca M, Alfoldi J, *et al* (2013) Analysis of the transcriptome of the Indonesian *coelacanth* *Latimeria menadoensis*. *BMC Genomics* 14 (538): 1-15.
21. Itai S, Birkland A, Chang K, El-Yaniv R, Yona G (2005) Correcting BLAST e-Values for low-complexity segments. *J Comput Biol* 12 (7): 978-1001.
22. Homer MJ, Bruinsma ES, Lodes MJ, Moro MH, 3rd Telford SR, Krause PJ, *et al* (2000). A polymorphic multigene family encoding an immunodominant protein from *Babesia microti*. *J Clin Microbiol* 38: 362-368.
23. Wang Z, Fang B, Chen J, Zhang X, *et al* (2010) *De novo* assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweetpotato (*Ipomoea batatas*). *BMC Genomics* 11 (726): 1-14.
24. Cooke B M, Mohandas N, Coppel RL (2001) the malaria-infected red blood cell: structural and functional changes. *Adv Parasitol* 50: 1-86.
25. Sherman IW, Eda S, Winograd E (2003). Cytoadherence and sequestration in *Plasmodium falciparum*: defining the ties that bind. *Microbes Infect* 5 (10): 897-909.
26. Chauvin A, Moreau E, Bonnet S, Plantard O, Malandrin L (2009) *Babesia* and its hosts: adaptation to long-lasting interactions as a way to achieve efficient transmission. *Vet Res* 40 (37): 1-18.

27. Schettlers TPM, Eling WMC (1999) Can *Babesia* infections be used as a model for cerebral malaria? *Parasitol Today* 15: 492-497.
28. Xu T, Ping J, Yu Y, Yu F, Yu Y, Hao P, *et al* (2010) Revealing parasite influence in metabolic pathways in Apicomplexa infected patients. *BMC Bioinformatics* 11: 1-9.
29. Edwards JS, Palsson BO (2000). Robustness analysis of the *Escherichia coli* metabolic network. *Biotechnol Prog* 16 (6):927-939.
30. Rodriguez M, Alhassan A, Ord RL, Cursino-Santos JR, Singh M, *et al* (2014) Identification and characterization of the RouenBd 1987 *Babesia divergens* Rhopty-associated protein 1. *PLoS ONE* 9 (9): e107727. doi:10.1371/journal.pone.0107727.
31. Yokoyama N, Okamura M, Igarashi I (2006) Erythrocyte invasion by *Babesia* parasites: current advances in the elucidation of the molecular interactions between the protozoan ligands and host receptors in the invasion stage. *Vet Parasitol* 138 (1-2): 22-32.
32. Ashrafi H, Hill T, Stoffel K, Kozik A, Yao J, Chin-Wo SR, Van Deynze A (2012) *De novo* assembly of the pepper transcriptome (*Capsium annuum*): a benchmark for the *silico* discovery of the SNPs, SSRs and candidate genes. *BMC Genomics* 13 (751): 1-15.
33. Annadurai RS, Jayakumar V, Muggasimangalam RC, Katta AV, Anand S, *et al* (2013) *De novo* transcriptome assembly (NGS) of *Curcuma longa L.* rhizome reveals novel transcripts related to anti-cancer and anti-malarial terpenoids. *PLoS One* 8 (2): e56217.

34. Gallardo-escárate C, Valenzuela-muñoz V, Nuñez-acuña G (2014) A RNA-Seq analysis using *de novo* transcriptome assembly as a reference for the *Salmon Louse Caligus rogercresseyi*. PLoS ONE 9 (4): e92239.
35. Ye J, Fang L, Zhang Y, Chen J, *et al* (2006) WEGO: a web tool for plotting GO annotations. Nucleic Acids Res 34: 293-297.
36. Garner MJ, Bishop R, Shah T, De Villiers EP, Carlton JM, Hall, N, *et al* (2002) Genome sequence of *Theileria parva*, a bovine pathogen that transforms lymphocytes. Science 309: 134-137.
37. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29 (7): 644-652.