UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

University of Pretoria

Faculty of Veterinary Science
Department of Veterinary Tropical Diseases

MSc Dissertation

Of

Mr Teboho Tsotetsi

Supervisor : Dr Kgomotso Sibeko-Matjila

Co-supervisors : Dr Nicola Collins

Prof Marinda Oosthuizen

Validation of expression profiles of differentially expressed transcripts identified in cattle-derived and buffalo-derived *Theileria parva* isolates by RNA sequencing

A dissertation submitted in partial fulfilment of the requirements for the degree of

MAGISTER SCIENTIAE (VETERINARY SCIENCE)

In the

Department of Veterinary Tropical Diseases

Faculty of Veterinary Science University of Pretoria

South Africa

April 2017

DECLARATION

I declare that this dissertation, which I hereby submit for the Master of Science degree in the Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, is my original work and has not been submitted by me for a degree to any other university.

Signed:_____ Date:_____

Teboho Tsotetsi

ACKNOWLEDGEMENTS

This research project was supported by the Genomics Research Institute of the University of Pretoria and the postgraduate bursary from the National Research Foundation (NRF), in collaboration with the Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria.

This dissertation emanates from project V040/14, approved by the Research Committee of the Faculty of Veterinary Science and the Animal Use and Care Committee of the University of Pretoria.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DISSERTATION SUMMARY

The intracellular apicomplexan protozoan parasite *Theileria parva* is a causative agent of cattle theileriosis which manifests in two disease syndromes, namely East Coast fever (ECF) and Corridor disease. Although ECF was eradicated from South Africa, cattle theileriosis still persists in the form of Corridor disease. Moreover, it is not known if the *T. parva* parasites present in buffalo in South Africa could cause ECF if they were to become established in cattle. This has made it essential to identify genetic differences that would allow successful discrimination of cattle-derived (causative agents of ECF) and buffalo-derived (causative agents of Corridor disease) *T. parva* parasites. Consequently, Next Generation Sequencing (NGS) was utilized to analyse *T. parva* transcriptomes from two isolates representing cattle-derived and buffalo-derived parasites, in order to identify gene expression profiles that may characterize cattle-derived and buffalo-derived *T. parva* isolates. However, RNA-sequencing (RNA-seq) experiments can be influenced by variability caused by technical effects including multiple template preparation stages, diverse sequencing chemistries and complex data processing of NGS experiments; it is thus crucial that data from these experiments is validated using other technologies. Thus, the aim of this study was to use quantitative real-time polymerase chain reaction (qPCR) for validation of differentially expressed genes (DEGs) identified from the RNA-seq study using NGS. Three groups of genes representing different expression profiles, including: 1. constitutively expressed genes; 2. up- and down-regulated genes and 3. genes exclusively expressed in one isolate or the other, were selected for validation.

Prior to validation of expression profiles for the selected set of genes using qPCR, endogenous control genes had to be selected in order to normalize qPCR gene expression data. Since there is no information available on the evaluation of the expression stability of these genes in *T. parva* isolates, the expression stability of five candidate reference genes, β-actin, glyceraldehyde-3-phosphate dehydrogenase, 28S rRNA, cytochrome b and fructose bisphosphatase aldolase (F6P), was evaluated for identification of reliable reference genes. The outcome of the stability rankings for each gene varied according to the program showing that the criteria for stability ranking differ from program to program. It is for this reason that the RefFinder tool, used in this study, integrates the different programs and gives a recommended comprehensive ranking. Therefore, based on this comprehensive analysis between the two *T. parva* isolates investigated, 28S rRNA and β-actin genes were selected as

the most suitable reference genes for this study. Intra- and inter-assay variation analysis of the selected reference genes showed that there was no significant variation in the expression of these genes between the two *T. parva* isolates with the p values being less than 0.05 and the coefficient of variation percentage being low (<2) for all the genes tested. Thus, we propose that genes coding for 28S rRNA and β-actin proteins be employed as endogenous control genes in studies that involve gene expression analysis of *T. parva*.

Validation of expression profiles from RNA-seq data obtained using NGS was performed using qPCR. In this study, the comparative C_T method for qPCR data analysis was employed to analyse the expression profiles of selected genes. The use of this method requires initial validation by ensuring that the target genes have approximately the same amplification efficiency as the endogenous control genes. Therefore, the amplification efficiencies of target genes and endogenous control genes were evaluated by constructing validation plots from standard curves generated from selected constitutively expressed and differentially expressed genes, in comparison to the standard curves of the two endogenous control genes. Initially, cDNA was prepared from total RNA isolated from bovine and buffalo lymphoblastoid cell cultures infected with *Theileria parva* (Muguga) and *Theileria parva* (7014), respectively, previously used for RNA-seq by NGS. The quantity of the parasite cDNA from the two isolates was interpolated from the standard curve and standardized to a concentration of 36.03 ng/µl, to eliminate concentration bias in downstream gene expression analysis. This study passed the comparative C_T method validation experiment since the absolute slopes of ΔC_T vs. Log input cDNA for the selected target genes were all less than 0.1 as required.

Twenty DEGs, constituting up- and down-regulated genes and genes exclusively expressed in one isolate or the other, and 10 stably expressed (constitutive) genes were selected for validation of expression profiles from RNA-seq data obtained using NGS. Discrepancies between RNA-seq and qPCR analyses were observed from all three groups of target genes but mostly in the constitutively expressed group of genes; in this group only 40% of the qPCR results corroborated with RNA-seq findings while 60% demonstrated variations in expression with four genes down-regulated and one up-regulated in *T. parva* 7014 relative to *T. parva* Muguga. Since most of the disagreements in the two datasets were down-regulated expression, this finding suggests that RNA-seq was more sensitive in detecting low abundant RNA transcripts.

From the genes shown to be exclusively expressed in either *T. parva* 7014 or *T. parva* Muguga by RNA-seq, 80% of the genes selected for validation produced the same outcome using qPCR. According to qPCR analysis, one gene coding for 40S ribosomal protein was detected in both *T. parva* Muguga and 7014 while it was only detected in *T. parva* Muguga according to RNA-seq. This finding suggests that this gene transcript may occur in low abundance in *T. parva* 7014, demonstrating that when using NGS approaches for gene expression profiling; genes with low expression can go undetected if the sequence depth is not high enough to allow detection of transcripts occurring at low levels.

For the DEGs selected for validation of up- and down-regulated genes, qPCR results were consistent with NGS data for 60% of the evaluated genes. Notable were two genes which were up-regulated in *T. parva* Muguga according to RNA-seq were only detected in *T. parva* Muguga using qPCR, while another gene, expected to be up-regulated in *T. parva* 7014 from RNA-seq analysis, was only detected in *T. parva* 7014 using qPCR. This result suggests that, although RNA-seq may be the most sensitive technique for detecting differential gene expression at low expression levels, the sensitivity of the NGS platform may be highly dependent on the abundance of the specific transcript.

Nevertheless, from the corroboration of NGS data and qPCR data in 70% of tested DEGs, it is evident that there are differences in the expression levels of some genes between a cattle-derived *T. parva* isolate (Muguga) and a buffalo-derived *T. parva* isolate (7014). Thus, it is proposed that the differential expression of some genes between the two *T. parva* isolates is the reason why the two isolates manifest different disease syndromes in cattle. It should be noted, however, that the cDNA in this study was isolated from lymphoblastoid cell cultures, and it remains to be seen whether the two isolates show the same expression profiles *in vivo*.

Chapter 1 : INTRODUCTION

1.1 General introduction

High mortality has been observed in exotic and crossbred cattle due to bovine theileriosis. The disease (Perry and Young, 1995) also adversely affects indigenous calves and adult cattle in endemically unstable areas. In affected African countries, bovine theileriosis is caused by the tick-borne protozoan parasite, *Theileria parva,* whose natural host is the African buffalo (*Syncerus caffer*). The sporozoite stage of the parasite infects T and B lymphocytes of cattle where it develops into a schizont stage, which induces clonal proliferation of the target cell. The parasite divides in synchrony with the host cells leading to the spread of parasitized cells in various tissues throughout the body, resulting in classical bovine theileriosis (Geysen *et al*., 1999). The effects of the disease result in major constrains in cattle production and the expansion of the meat and dairy industry in affected countries. Bovine theileriosis is a disease of major economic importance; the cost of bovine theileriosis, particularly for one form of the disease known as East Coast fever (ECF), was estimated at $186 million in the 1980s in all affected countries in the African region (Mukhebi *et al*., 1992).

In addition to livestock, theileriosis has great impact on the game industry. African buffalo (*Syncerus caffer*) are natural reservoirs of the *T. parva* parasite and other diseases of veterinary importance such as foot-and-mouth disease, bovine tuberculosis and brucellosis. Together with theileriosis, these diseases have devastating effects to livestock farming. From the ecotourism and economic perspective, the business of breeding disease-free buffalo is very lucrative. The average price of a disease-free buffalo is ZAR150 000, with annual price increases estimated to be at 29% (Laubscher and Hoffman, 2012). In 2012, on an exceptional event, a staggering ZAR26 million (ZAR1 ~ US$0.12) was paid for a buffalo bull. The market value of buffalo depreciates significantly, in fact up to ten times less, for infected animals. The financial implication extends to the loss of revenue if the game properties lose their attraction for tourists or hunters due to the absence of buffalo. As a member of the ‘big five game family’, the African buffalo is a sought-after trophy attracting over $10, 000 to hunt one.

Cattle-derived and buffalo-derived *T. parva* infections cause two different disease syndromes in cattle, ECF and Corridor disease respectively, and both are fatal to susceptible cattle. East

Coast fever was introduced into South Africa, presumably from East Africa in 1902 and was eradicated by 1956 through an extensive quarantine, systemic dipping and slaughter campaign (Theiler, 1904; Neitz, 1957). Corridor disease still occurs in South Africa wherever infected buffalo and cattle graze on the same pasture in the presence of the tick vector *Rhipicephalus appendiculatus*. The reason behind the difference in disease syndromes resulting from *T. parva* infection is not clearly understood. From molecular evidence, it is known that cattle-derived *T. parva* parasites are genetically homogeneous while there is a great deal of genetic heterogeneity between buffalo-derived *T. parva* isolates (Geysen *et al*., 1999), but no distinctive differences have been identified to date to distinguish between the two groups of *T. parva* parasites. It is therefore not known if the *T. parva* parasites present in buffalo in South Africa could cause ECF if they were to become established in cattle. These challenges make it essential to identify genetic differences that would allow successful discrimination of cattle-derived (causative agents of ECF) and buffalo-derived (causative agents of Corridor disease) *T. parva* parasites.

Consequently, next generation sequencing (NGS) was utilized to analyse *T. parva* transcriptomes from two isolates representing cattle-derived and buffalo-derived parasites, in order to identify gene expression profiles that may characterize cattle-derived and buffalo-derived *T. parva* isolates (KP Sibeko, Department of Veterinary Tropical Diseases, University of Pretoria, personal communication). Thus this study sought to validate differentially expressed genes (DEGs) identified from these *T. parva* isolates, using quantitative real-time polymerase chain reaction (qPCR). Successful validation of these DEGs could allow identification of genetic markers that will allow differentiation of cattle-derived and buffalo-derived *T. parva* parasites. Understanding the genetic diversity between cattle-derived and buffalo-derived *T. parva* parasites may give insight into the evolution and diversification of the parasite for improvement of disease control measures and to further understand the epidemiology of bovine theileriosis.

1.2 Literature review

1.2.1 Introduction

Cattle and wildlife in Africa are immensely vulnerable to a wide range of tick-borne pathogens, one of economic importance being *Theileria parva*. *Theileria parva* is a haemoprotozoan parasite which is principally transmitted by the ticks *Rhipicephalus appendiculatus* and *Rhipicephalus zambeziensis* with its natural reservoir host being the African buffalo (*Syncerus caffer). Theileria parva* infections in cattle are associated with three disease syndromes which are East Coast fever (ECF), January disease and Corridor disease (Uilenburg, 1976).

In most East African countries bovine theileriosis is controlled by immunization using the infection and treatment method. The wide-spread application of this method is limited by the existence of multiple strains of *T. parva* since immunization with single strain vaccines does not protect against challenge with all heterologous stocks, especially infections resulting from buffalo-derived *T. parva* (Radley *et al*., 1975; Irvin *et al*., 1983). Although the development of the cocktail vaccine has partially overcome this problem, the presence of a reservoir of the parasite in the African buffalo complicates the epidemiology of the disease. The carrier host is believed to increase the polymorphic traits of *T. parva* isolates (Radley *et al*., 1975; Musisi, 1990). The basis of variation in immunity between *T. parva* strains remains unclear and underpins the need for markers that will distinguish parasite strains, thus providing more insight into the epidemiology of bovine theileriosis.

Various molecular techniques based on the parasite polymorphic antigen-encoding genes such as p104, p67 and polymorphic immunodominant molecule (PIM), and the serology-based assay using monoclonal antibodies against a *T. parva* surface protein, PIM, have been employed to evaluate genetic variations between different *T. parva* strains (Geysen *et al*., 1999; Bishop *et al*., 2001; Minami *et al*., 1983). However, the resolution of genetic differentiation in these studies is limited because of the relatively low marker density. Recently, Hayashida *et al.* (2013) performed a genome sequence comparative study of nine *T. parva* strains from Eastern, Central and southern Africa using Next Generation Sequencing (NGS). Genome-wide comparison of strains revealed genetic polymorphisms on a fine scale and was used to infer phylogenetic relationships among the parasites investigated, showing

that new tools such as NGS have overcome the limitations seen with using serological methods and RFLP for characterization of *T. parva* stocks.

1.2.2 *Theileria parva*: The parasite

Theileria parasites are currently classified in the class Sporozoa together with human pathogens, including *Plasmodium* and *Toxoplasma* (Young *et al*., 1986). *Theileria* species fall under a sub-phylum called Apicomplexa, as these parasites possess an apical complex, which contains secretory organelles involved in invasion, or establishment, in the cells of their mammalian and invertebrate hosts. The evolutionary and functional equivalence of the apical complex between different genera of the Apicomplexa is, however, still vague (Bishop *et al*., 2004).

Theileria parva is a tick-borne protozoan parasite that causes uncontrolled proliferation of bovine lymphocytes. It is recognized as a complex of intracellular protozoan parasites distinguishable on the basis of the clinical and epidemiological features of the infections they cause. Three subspecies were previously recognized: *T. parva parva*, *T. parva lawrencei* and *T. parva bovis* associated with ECF, Corridor disease and January disease, respectively (Lawrence, 1979; Uilenberg, 1976). However, this taxonomy was later discarded due to lack of biological evidence to distinguish between the three subspecies. *Theileria parva* parasites that are transmitted between cattle (formerly *T. p. bovis* and *T. p. parva*) are now classified as cattle-derived *T. parva*, while *T. parva* parasites that are transmitted from buffalo to cattle (formerly *T. p. lawrencei*) are classified as buffalo-derived *T. parva* (Perry and Young, 1995).

The geographical distributed of *T. parva* often coincide with that of its tick vectors. The principal tick vector of *T. parva* is *Rhipicephalus appendiculatus,* which finds its host primarily in cattle, buffalo and large antelope, but it can occur on small ruminant species including sheep and goats (Morzaria, 1988). The distribution of *R. appendiculatus* occurs in eastern, central and south-eastern Africa, it is limited to suitable environments with appropriate hosts. In the subtropical central and southern regions of Africa, *R. appendiculatus* completes one life cycle annually, and the occurrence of adults, nymphs or larvae is seasonal, while in tropical areas, more than one life cycle can be completed each year and all stages occur at one time (Arthur, 1961; Bishop *et al*., 2004).

The different species of *Theileria* may differ enormously in their virulence depending on the strain of the parasite, degree of host susceptibility and the dose of parasite and *T. parva*

parasites are no exception. However, fundamental questions concerning the population genetics of *T. parva* have not yet been addressed due to the limited number of molecular markers available. These questions include the population genetic structure of *T. parva* in cattle and African buffalo (*Syncerus caffer*), the degree to which parasite populations are modified by vaccination and the extent of sexual recombination in populations of the parasite in the field (Oura *et al*., 2003). All of these factors combined contribute to the genetic diversity of the *T. parva* strains and the mystery around different disease syndromes caused by these parasites in infected cattle.

1.2.3 Life cycle

Theileria parva has a complex life cycle (Fig. 1.1) involving several distinct intracellular stages in both the arthropod vector and the mammalian host (Bishop *et al*., 2004). The sporozoite stage of the parasite is transmitted during feeding by the tick vector (Stagg *et al*., 1981). The initial interaction between the sporozoites the host cell occurs by chance due to the immobility of the *Theileria* sporozoites and it is temperature independent. This initial contact results in a relatively strong binding of the sporozoite to the host cell surface and, as far as can be seen, is not reversible. The sporozoites bind to and enter host cells in any orientation, and invasion does not require re-orientation of the parasite to bring the apical end into close contact with the host cell membrane. This differs from other apicomplexans which require re-orientation for the apical-end to access the host first and thus expedite invasion (Shaw, 2003).

In the animal host the sporozoites penetrate the lymphocytes and differentiate into schizonts inducing a lympho-proliferative disorder. The schizonts later differentiate in the lymphocytes into merozoites that, once released from the lymphocytes, invade the erythrocytes. Once in the erythrocytes, merozoites develop into piroplasms which are infective to ticks (Norval *et al*., 1992).

The sexual stage of development of *T. parva* occurs in the gut of the tick where gametogenesis and fertilization take place resulting in the production of a zygote (Melhorn and Schein, 1984). The zygote invades the gut cell and remains there throughout the moulting cycle and develops into a single motile kinete. These kinetes escape the gut cells and invade the salivary glands. The parasites remain in the salivary gland until transmitted to another mammalian host when the resulting post-moult nymph or adult feeds. Tick feeding initiates

rapid sporozoite development and infective sporozoites are released during the later stages of feeding (Norval *et al*., 1992).

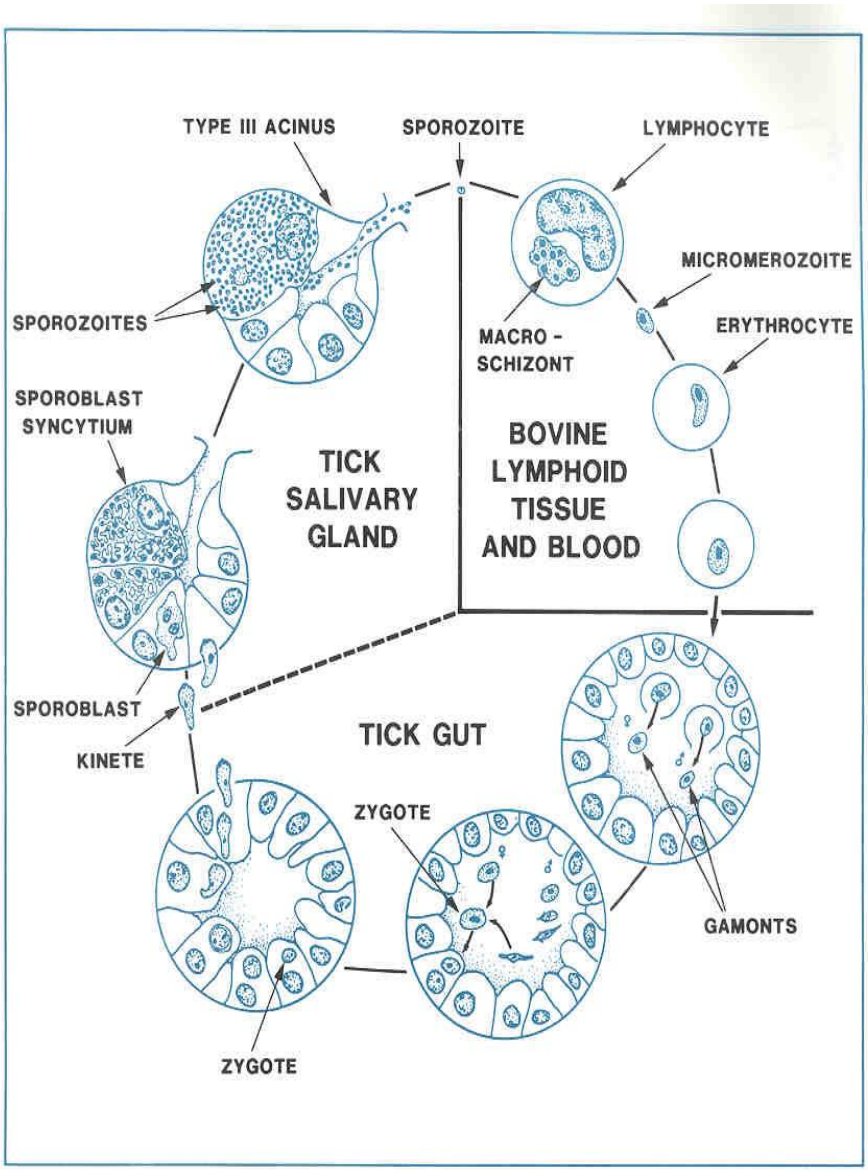


Figure 1.1: **Representation of the life cycle of *Theileria parva***

(From the International Laboratory for Research on Animal Diseases, 1983, https://www.ilri.org/InfoServ/Webpub/fulldocs/Ilrad82/Theiler.htm).

1.2.4 Bovine theileriosis

Buffalo are considered to be natural hosts of *T. parva* in which the parasite appears not to manifest any clinical disease (Grootenhuis *et al*., 1987). It is believed that co-evolution of African buffalo with *T. parva* must have occurred prior to infection of cattle and that ticks infected with *T. parva* from buffalo first came into contact with *Bos indicus* cattle in sub-Saharan Africa approximately 4500 years ago (Epstein, 1971; Oura *et al*., 2011).

The infection of cattle with *T. parva* parasites causes three disease syndromes, namely East Cost fever, January disease and Corridor disease.

1.2.4.1. East Coast fever

The causative agent of East Coast fever (ECF) was first observed by Koch in 1898 in infected cattle blood smears but was later extensively studied by Theiler (1904). East Coast fever arises as a result of transmission of the cattle-derived *T. parva* parasite from infected cattle to susceptible cattle through tick transmission. The disease is characterized by the proliferation of lymphoblasts infected with theilerial schizonts throughout the body, particularly in the lymph nodes, lymphoid aggregates, spleen, kidneys, liver and lungs (Musoke *et al*., 2004).

East Coast Fever is characterized by the swelling of the draining lymph node in cattle 7 to 15 days post attachment of infected ticks. This is followed by a generalized lymphadenopathy in which superficial subcutaneous lymph nodes such as the parotid, prescapular and prefemoral lymph nodes can easily be seen and palpated (Elsheikha and Khan, 2011). Fever supervenes and continues throughout the period of infection with the increase of the body temperature being rapid and usually in excess of 39.5°C but may reach 42°C. Death usually results 18 to 30 days post-infection of susceptible cattle by infected ticks. The most striking post-mortem lesions are lymph node enlargement and massive pulmonary oedema and hyperaemia. Haemorrhages are common on the serosal and mucosal surfaces of many organs, sometimes together with obvious areas of necrosis in the lymph nodes and thymus (Elsheikha and Khan, 2011). The severity and time course of the disease is dependent on, among other factors, the degree of the infected tick challenge and the strain of the parasite (Irvin and Mwamachi, 1983).

1.2.4.2. Corridor disease

Corridor disease, also referred to as buffalo-derived theileriosis, is a deadly disease of cattle caused by buffalo-derived strains of *T. parva.* This disease is characterised by low schizont

parasitosis and piroplasm parasitaemia. Clinical characteristics of Corridor disease are similar to those of ECF except that the course of the disease is usually shorter, death occurring only 3-4 days after the onset of first clinical signs (Lawrence *et al*., 1994). Outbreaks of Corridor disease are generally encountered when there is contact between infected buffalo and susceptible cattle in the presence of the tick vector or when susceptible cattle graze pastures where carrier buffaloes have been.

Corridor disease was first diagnosed in 1953 in a corridor of land between the then separate Hluhluwe and Imfolozi game reserves in South Africa, hence the name Corridor disease. The disease has generally been regarded as self-limiting as cattle usually die in the acute stage before the parasite develops into the erythrocytic piroplasm stage infective to the tick vector (Norval *et al*., 1992). However, an *in vitro* study has demonstrated that carrier state can develop in animals infected with buffalo-derived *T. parva* (Potgieter *et al*., 1988). Recently, a study by Mbizeni *et al*., (2013) corroborated these reports, clearly demonstrating that Corridor disease occurs in localities at the cattle/game interface in KwaZulu-Natal, and showing that *T. parva* positive cattle could be detected by *T. parva*-specific qPCR in these areas. Mbizeni *et al.* (2013) could not demonstrate transmission of the parasite from these cattle as the infection was not maintained for longer than a few months.

1.2.4.3. January disease

Upon the elimination of ECF in Zimbabwe, another type of theileriosis known as January disease emerged. This disease is caused by the cattle-derived *T. parva* parasite formerly known as *T. parva bovis.* The name January disease is derived from the strict seasonality of the disease occurrence which is between December and March, coinciding with the seasonal activity of the tick vector *R. appendiculatus* in Zimbabwe (Matson, 1967).

The disease shows the same pathogenesis and clinical characteristics as ECF but it has a lower mortality rate compared to ECF and it regularly occurs in Zimbabwe (Lawrence *et al*., 1994).

1.2.5 The epidemiology of cattle theileriosis

Bovine theileriosis causes high mortality in cattle of introduced breeds and in indigenous cattle outside areas where the disease is endemic. The causative agents of classical ECF and Corridor disease were formerly classified as subspecies of *T. parva,* known as *T. parva parva* and *T. parva lawrencei*, respectively. Barnett and Brocklesby (1966) reported transformation

of *Theileria parva lawrencei* to *T. parva parva* by passage through cattle; however transformation could not be successfully demonstrated in South Africa (Potgieter *et al*., 1988).

Bovine theileriosis is predominant across the eastern, central, and southern parts of Africa, and has been reported in 11 countries in the region: Kenya, Uganda, Tanzania, Burundi, Rwanda, Malawi, Mozambique, southern Sudan, Democratic Republic of Congo (DRC), Zambia and Zimbabwe (Lawrence *et al*., 1992). East Coast fever was also reported in Comoros between 2003 and 2004 for the first time. The latter incident was suggested to result from importation of immunized cattle from Tanzania, which were fed upon by naïve ticks that subsequently transmitted the infection to a vulnerable local cattle population (De Deken *et al*., 2007).

In South Africa the introduction of ECF occurred during 1901–1903 through cattle imported from Kenya and Tanzania for restocking after the ravages of the rinderpest epidemic and the Anglo-Boer War. About 1.25 million out of 4 million cattle in the affected area had died of ECF by 1914. The disease was exterminated from southern Africa between 1946 and 1955 during a 50-year campaign comprising movement control, tick control, destocking of infected pastures and slaughtering of herds which were infected (Lawrence, 1992). When ECF was eradicated from South Africa, the vector tick was not eradicated allowing the emergence of another form of theileriosis, Corridor disease. The original buffalo-derived *T. parva*, the causative agent of Corridor disease, remains endemic in South African and Zimbabwean buffalo populations (Lawrence *et al*., 2004). This form of cattle theileriosis is a controlled disease in South Africa meaning that all incidents of the disease must be reported to the authorities and infected animals should be destroyed in order to prevent the establishment of the parasite in cattle. There is evidence that cattle infected with buffalo-derived *T. parva* (the causative agent of Corridor disease) can recover after receiving low doses of sporozoite stabilate and become carriers of the parasite, hence buffalo must be certified pathogen-free before they can be translocated to Corridor disease-free areas (Potgieter *et al*., 1988).

Since the eradication of ECF from South Africa, there have been no reports of ECF infections in cattle (Musoke *et al*., 2004). However, there is a concern that the disease might re-emerge following the identification of p67 alleles associated with *T. parva* isolates (*T. parva* Muguga) responsible for causing ECF in Kenya, East Africa, from *T. parva* isolates from cattle from a farm in KwaZulu-Natal (Sibeko *et al*., 2010). However, although the cattle on

this farm had theileriosis, the origin of the parasites could not be confirmed, and cattle-to-cattle transmission of the parasite could not be confirmed (Thompson *et al*., 2008).

1.2.6 Control and treatment of theileriosis

Various methods are used to control theileriosis including tick control by acaricide application, immunization and chemotherapy (Mutugi *et al*., 1987). In South Africa tick control by acaricide application is applied together with strict physical separation of buffalo and cattle; methods that involve immunization and chemotherapy are prohibited due to the threat of development of carrier state. Ticks infecting livestock in tropical and sub-tropical countries are prone to developing acaricide resistance. Thus, an optimal integrated TBD control program is likely to include several approaches including use of resistant breeds and vaccination, thereby allowing more strategic application of acaricides. The development of combined vaccines conferring protection against pathogens transmitted and the respective vector may be a prerequisite for the success of vaccine development against protozoan pathogens, including *Theileria parva* although such a vaccine could not be applied in South Africa (Dolan, 1999; Radley, 1981; Neitz, 1953)

1.2.6.1 Immunization

Cattle can be immunized against *T. parva* by simultaneous inoculation of live sporozoites and long-acting formulations of oxytetracycline, a process known as the infection and treatment method (Radley *et al*., 1975). Protective immunity with this method is achieved on the principle that an animal gets a mild or asymptomatic body reaction and in the same process its body elicits immune responses that will ensure protection upon infection by field strains of the parasite. There is solid evidence that the subsequent protection is mediated by parasite-specific major histocompatibility complex (MHC) class 1-restricted cytotoxic T lymphocytes (CTLs), which eliminate schizont-infected lymphoblasts (Mckeever *et al*., 1994). The response is highly specific and therefore vulnerable to breakthrough by heterologous strains. This has been a major challenge with the development of infection and treatment vaccination as different *T. parva* stocks have different immunogenic properties and may not cross-protect (Young *et al.,* 1973) especially against infections resulting from buffalo-derived *T. parva*. This problem has been partially overcome by the use of a cocktail of stabilates, such as the Muguga cocktail, as the combination has shown to present broad protection even in the field (Musisi, 1990). However, a concern with using a stabilate cocktail is possible introduction of new *T. parva* stocks in areas where they never existed before immunization.

Another problem with the ITM is that it uses oxytetracyclines, which have been shown to be less effective particularly where high doses of *T. parva* stocks are used, and the infection is not always contained. Parvaquone, although efficacious in controlling such infections, is too expensive for use in field immunization (Dolan, 1986).

In view of the above, a lot of effort has gone into research for development of recombinant vaccines. A study by Musoke *et al*. (1991) showed that immunization using a recombinant vaccine emulsified in 3% saponin induced sporozoite neutralizing antibodies in cattle and provided protection in six of nine animals on homologous challenge with *T. parva* sporozoites.

A 67 kDa glycoprotein (p67) from the surface of the *T. parva* sporozoite has been isolated and used in a variety of immunization protocols, with little success reported so far in the development of practical levels of immune mediated disease resistance (Nene *et al*., 1996**)**. However, cattle recovering from a single infection with *T. parva* sporozoites resist infection upon homologous challenge; such animals have weak antibody and T cell responses to p67. There is still a need therefore to identify *T. parva* antigens that can induce antigen-specific class 1 MHC-restricted CD8 cytotoxic T lymphocytes (CTLs). Graham *et al.* (2005) already illustrated an approach that provides a basis for developing a CTL-targeted anti-ECF subunit vaccine and have identified the CTL-target antigens recognized by T cells from immune cattle to include ε̆-TCP1, e1F-1A, Hsp90, cysteine proteinase and IFN-γ.

1.2.6.2 Tick control

Bovine theileriosis has mainly been controlled by means of tick control. Tick control has been achieved through various combinations of pasture spelling, control of cattle movement and acaricide application (Dolan, 1999). Pasture spelling has been abandoned due to the long periods it involved. The rapidly rising costs of acaricides and problems with their application has led to the development of other control methods such as vaccines against ticks, slow-release acaricide devices and more efficient means of topical application of acaricide (Norval *et al*., 1992). Olds *et al*. (2012) established the effect of anti-tick vaccines by employing insect cell-expressed recombinant versions of the *R. appendiculatus* homologs of Bm86, named Ra86, to vaccinate cattle. The Ra86 vaccination of cattle significantly decreased the moulting success of nymphal ticks to the adult stage; this suggests that repeated vaccinations using Ra86 could reduce tick populations over successive generations. Vaccination with

Ra86 could thus form a component of integrated control strategies for *R. appendiculatus* leading to a reduction in use of environmentally damaging acaricides.

1.2.6.3. Chemotherapy

Cattle theileriosis is best treated by the intramuscular injection of the antitheilerial hydroxynaphthoquinone compounds, parvaquone or buparvaquone (Muraguri *et al*., 1999). However, successful application of chemotherapy requires early diagnosis so that treatment can be given at the early stages of clinical disease.

According to Lizundia *et al*. (2009), *T. parva* parasites possess a relic plastid (apicoplast), whose metabolic pathways include several promising drug targets. Inhibition assays have indicated that several putative inhibitors of apicoplast function inhibit *Theileria*-induced proliferation of lymphocytes but their modes of action are unclear.

1.2.7 Characterization of *Theileria parva*

Identification and characterization of different stocks of *T. parva* is intricate due to the fact that these parasites cannot be differentiated from each other on the basis of vector or host specificity, morphological appearance or bovine serological responses which can be detected by standard immunological methods (Conrad *et al*., 1987).

Due to the limited number of molecular markers available, questions about the population genetic structure of *T. parva* in cattle and African buffalo, the degree to which parasite populations are modified by vaccination and the extent of sexual recombination in populations of parasites in the field remain unresolved (Oura *et al*., 2003).

A number of serological and molecular assays have been developed in attempts to characterize buffalo-derived and cattle-derived *T. parva* isolates.

1.2.7.1 Monoclonal antibody screening

Theileria parva antischizont monoclonal antibodies (MAb) have been synthesized and utilized in an indirect fluorescent antibody (IFA) test against schizont-infected cells derived from *in vitro* cultures to demonstrate stock-specific diversity. The presence or absence of binding to MAbs 2 and 3 and to 15 and 16 was a convenient way of dividing *T. p. parva* stocks according to the three former groups which are *T. p. parva, T. p. lawrencei* and *T. p. bovis* (Minami *et al*., 1983).

1.2.7.2 Molecular characterization

1.2.7.2.1 Characterization based on genes encoding antigenic proteins:

In the pursuit to characterize *T. parva* stocks, Geysen *et al.* (1999) conducted a study to develop new molecular tools for characterization of *T. parva* stocks using 12 parasites for analysis of parasites in the field. The methods involved southern blotting on DNA subjected to digestion with restriction enzymes (RFLP-DNA); four radiolabeled probes, Tpr probes (Sohanpal *et al*. 1995), the telomere probe (Allsopp *et al*., 1993), the LA6 probe (Bishop *et al*.,1998) and the minisatellite probe were used on genomic DNA from tissue culture parasites digested with *Eco*R1. Restriction fragment length polymorphism assays of gene-specific PCR products (RFLP-PCR) were developed for three polymorphic antigenic loci for use on field samples. The PIM locus was characterized by Toye *et al*. (1991), p104 locus by Iams *et al*. (1990) and the p150 locus was characterized in 1998 by Skilton *et al*. Southern blot on RFLP-DNA illustrated minor variations among various cloned cultures of the same parent stock when using the Tpr and telomere probes. Analysis of the profiles of the different probes revealed that the Zambian isolate RFLP profiles were analogous to, suggesting a relative homogeneity among stocks from Zambia. The p150 and p104 antigen loci exhibited moderate polymorphism with four alleles identified among the isolates studied. The PIM locus was the most polymorphic with 10 alleles besides marked size polymorphism in both DNA and protein length. Clear differentiation of the isolates coming from two geographical areas in Zambia and within each Province was possible. The PIM polymorphisms alone could characterize subdivisions in the *T. parva* population in the field.

In a study by Sibeko *et al*. (2010) the variable region of the p67 gene was amplified and the gene sequences were analyzed to characterize South African *T. parva* parasites in cattle from farms where Corridor disease outbreaks were diagnosed and in experimentally infected cattle. Four p67 alleles were identified, including alleles 1 and 2 previously detected in East African cattle and buffalo, respectively, as well as two novel alleles, one with a different 174 bp deletion (allele 3), the other with a similar sequence to allele 3 but with no deletion (allele 4). The findings of this study showed that the p67 genetic profile was more diverse than previously thought and could not be used to distinguish between cattle- and buffalo-derived *T. parva* isolates in South Africa. Nonetheless, p67 allele 1, identical to that of *T. parva* which causes classical ECF, was identified in the cattle, raising fears that *T. parva* parasites with characteristics of cattle-derived *T. parva* might be present in cattle in South Africa. The

p67 gene encodes an antigenic protein which has been identified as a vaccine candidate, as a result this gene and its product has been investigated extensively. Thus the implications of identification of extensive diversity in this gene are that immunisation using a p67-based vaccine could induce immunity to homologous *T. parva* strains common in cattle, but heterologous strains could still prove lethal in a given area.

1.2.7.2.2 Characterization based on micro- and mini-satellite markers:

Micro and mini-satellite PCR amplification of *T. parva*-specific sequences have confirmed that there is genetic diversity amongst *T. parva* stocks derived from both cattle and buffalo (Oura *et al*., 2003). Micro and mini-satellite markers developed for characterizing *T. parva* stocks enable detection of higher levels of polymorphism than PCR-RFLP methods which are based on individual loci. Microsatellite markers have been used to characterize field isolates by defining multilocus genotypes for each sample based on the sizes of the predominant alleles at each locus (Oura *et al*., 2004).

Molecular studies findings suggest that only a limited subset of the total *T. parva* gene pool present within buffalo has become established in cattle (Oura *et al*., 2011). It is unclear whether the transfer of buffalo-derived strains to cattle resulting in ECF was a single event or whether there is a constant dribble of new strains transferring from buffalo to cattle. This has very important implications since it has been demonstrated that there is incomplete cross-protection between animals immunised with cattle-derived *T. parva* on challenge with parasites from buffalo (Young *et al*., 1973), which implies that any live vaccination programme based on homogenous stocks, as it is common with cattle-derived *T. parva* stocks, may not protect against Corridor disease.

1.2.8. *Theileria parva* genome and transcriptome analysis

The genome of *T. parva* has been sequenced in order to facilitate research on the biology of the parasite and assist in the identification of schizont antigens for vaccine development. The genome of *T. parva* is approximately 8.5 Mbp in size, which is rather small for a eukaryote with such a complex life cycle, and it contains about 4035 protein-encoding genes (Gardner *et al*., 2005).

The availability of the genome of *T. parva* allowed studies that can elucidate biological and molecular processes in this parasite, as well as determinants of species variation within the *T.*

parva parasite group. The separation of species has been studied using gene conversion, based on the low-resolution VNTR markers. This analysis revealed that the crossover rate in *T. parva* is relatively high and varies across different regions of the genome. Non-crossovers and crossover-associated gene conversions have not yet been characterized due to the lack of informative markers. However, analysis of the genome using NGS, provided the necessary high marker resolution resulting in the detection of all recombination events, from sequencing of the haploid genomes of two parental strains of *T. parva*, and two recombinant clones derived from ticks fed on cattle that had been simultaneously co-infected with two different *T. parva* isolates (Henson *et al*., 2012).

A study by Hayashida *et al.* (2013) elucidated the phylogeny of *T. parva* strains based on analysis of genome-wide SNPs with prediction of possible past recombination events, providing insight into the migration, diversification, and evolution of this parasite species in the African continent. However, to investigate the virulence and evolution of bovine theileriosis after their diversification from buffalo will require further phylogenetic analysis in combination with phenotypic data.

In 2005, Bishop *et al.* described the transcriptome of *T. parva* using massively parallel signature sequencing (MPSS), which revealed that the majority of genes are transcriptionally active in the schizont stage of the parasite. The transcriptome can be analysed to identify differentially expressed genes (DEGs) that may aid in revealing the mystery behind the differences in the disease syndromes caused by cattle-derived and buffalo-derived *T. parva* isolates.

1.2.9 Validation of gene expression data

Next generation sequencing is an empirical tool that can provide robust transcriptomics information (RNA-sequencing or RNA-seq). However, RNA-sequencing experiments can be influenced by the variability coming from technical effects. One of these technical effects comes from the generation of libraries of cDNA fragments, which involves various ligations of adaptors and PCR amplifications. It is thus essential for RNA-sequencing data to be validated (Wang *et al*., 2009). The tanscriptome dataset arising from NGS is usually validated employing techniques such as qPCR, DNA microarray or even the quantitative proteomic approach. Quantitative real-time PCR distinguishes itself from other methods available for gene expression analysis in terms of accuracy, sensitivity and fast results. As a consequence,

the technology has established itself as the golden standard for medium throughput gene expression analysis (Derveaux *et al*., 2010). Using qPCR as a quantification method of choice depends on the target sequence, the degree of accuracy required and whether quantification needs to be relative or absolute. qPCR generally offers two quantification methods. (i) A relative quantification based on the relative expression of a target gene versus a reference gene. To investigate the physiological changes in gene expression, the relative expression ratio (fold change) is adequate for most purposes. (ii) An absolute quantification based either on an internal or external calibration curve (Pfaffi, 2001).

Despite being an exceptionally powerful technique, qPCR suffers from certain pitfalls, most importantly being the need to normalize the data with an endogenous control or housekeeping gene (Bustin and Nola, 2004). The expression of endogenous control genes used for normalization in qPCR analysis should remain constant between cells of different tissues and under different experimental conditions; otherwise, it can lead to erroneous results (Pfaffi *et al*., 2004). Thus prior to relative gene expression analysis certain validation steps are required.

High-throughput RNA-sequencing (RNA-Seq) technology revolutionized the approach for transcriptome analysis and has been applied to many organisms. Illumina/HiSeq™ 2000 RNA-seq technology has become more popular over the last decade because of higher throughput and relatively low cost, and has been used with several apicomplexan parasite species (Hyashida *et al*., 2013), including *T. parva* (KP Sibeko, Department of Veterinary Tropical Diseases, University of Pretoria, unpublished data). RNA-sequence analysis consists of four fundamental analysis steps, providing that an already sequenced reference genome or transcriptome is available for the reviewed organism. Firstly, raw image data have to be converted into short read sequences, which are subsequently aligned to the reference transcriptome. The amount of mapped reads is counted and the gene expression level is calculated by peak calling algorithms. Finally, statistical tests are used to determine differential gene expression (Carstens *et al*., 2012). In a recent study the RNA of a cattle-derived *T. parva* (Muguga) isolate and a buffalo-derived *T. parva* (7014) isolate were sequenced using NGS. Unlike other sequencing techniques such as Sanger sequencing, NGS requires extensive processing before analysis, for reasons that are largely related to the scale on which the data are collected. With sequencing capacities up to ~3.0×10^9 reads (e.g., Illumina's HiSeq platform), large numbers of loci and individual samples can be included on

a single run. Post-run processing includes the de-indexing of individual samples, quality control, alignment, and calling of single-nucleotide polymorphisms (Carstens *et al*., 2012). A total of 3954 transcripts were successfully mapped to the *T. parva* genome sequence and of these 1048 (26.5%) were differentially expressed between the two isolates.

1.3 Study rationale

There is limited understanding as to why cattle-derived and buffalo-derived *T. parva* infections cause different disease syndromes in cattle, i.e. ECF and Corridor disease, respectively. Also of more concern to South Africa is the possibility that ECF might re-emerge since its eradication in the 1950s; a *T. parva* p67 allele that is associated with ECF-causing *T. parva* stocks (*T. parva* Muguga) has been identified from cattle samples collected from a farm near Ladysmith in KwaZulu-Natal (Sibeko *et al*., 2010). The recurrence of ECF in South Africa would have devastating consequences to an industry consisting of a vulnerable cattle population. Therefore, there is a need to identify genetic markers that would allow successful discrimination of the causative agents of ECF and Corridor disease; moreover, to understand the epidemiology of bovine theileriosis.

The use of molecular characterization tools has revealed a high level of genetic diversity among buffalo-derived *T. parva* stocks compared to cattle-derived *T. parva* stocks (Bishop *et al*., 1994; Collins and Allospp, 1999; Geysen *et al*., 2004; Oura *et al*., 2004). Although these tools have succeeded in the broad differentiation of cattle- and buffalo-derived *T. parva* isolates, they have failed to distinguish between *T. parva* stocks according to the different disease syndromes they cause, which requires a superior resolution of genetic differentiation. The use of Next Generation Sequencing (NGS) in genome sequence analysis of different *T. parva* stocks in a recent study has allowed analysis of genetic polymorphisms on a fine scale (Hayashida *et al*., 2013). Next Generation Sequencing has also enabled identification of transcripts that are differentially expressed between cattle-derived and buffalo-derived *T. parva* (KP Sibeko, Department of Veterinary Tropical Diseases, University of Pretoria, personal communication). It is thought that these transcripts could hold key answers as to why infections by cattle-derived and buffalo-derived *T. parva* isolates cause different disease syndromes in cattle.

The differentially expressed genes identified using NGS are often validated using qPCR. It is worth pointing out that validation using qPCR on the same cDNA samples assayed in the

RNA-seq analysis only validates the technology. It does not validate the conclusion about the treatments/conditions. It is the validation using different biological replicates from the same populations that can further validate the biological conclusions from RNA-seq experiments (Allison *et al*., 2006).

1.4 Aims and objectives

This study forms part of a larger *T. parva* transcriptomics project in which transcriptomes of two *T. parva* isolates, *T. parva* (Muguga) and *T. parva* (7014), respectively representing cattle-derived parasites (causative agents of ECF) and buffalo-derived parasites (causative agents of Corridor disease), were investigated for variations in gene expression and other significant transcriptome sequence differences using Next Generation Sequencing (NGS). Transcriptome data analysis revealed differential expression in a significant number of genes between the cattle-derived and buffalo-derived *T. parva* isolates investigated.

Consequently, the aim of this MSc project was to use quantitative real-time PCR (qPCR) to provide validated data for differentially expressed genes (DEGs) identified from the RNA-seq dataset generated using NGS and this aim is embedded in the following objectives:

i. Identification of DEGs from the *T. parva* (Muguga) and *T. parva* (7014) NGS transcriptome dataset for validation using qPCR.
ii. Identification and evaluation of internal control genes for qPCR.
iii. Primer design, development and evaluation of a custom qPCR array for validation of DEGs.
iv. Validation of gene expression profiles from selected DEGs using the qPCR array.

Chapter 2 : MATERIALS AND METHODS

2.1 RNA isolation and cDNA synthesis

The RNA used in this study was prepared previously by Dr KP Sibeko (Department of Veterinary Tropical Diseases, University of Pretoria). Total RNA was isolated from bovine and buffalo lymphoblastoid cell cultures infected with *Theileria parva* (Muguga) and *Theileria parva* (7014), respectively, previously used for RNA-seq by Next Generation sequencing (NGS). RNA extractions were performed using the RNeasy plus mini kit (Qiagen), according to the manufacturer's instructions. Two RNA extractions were performed for each *T. parva* stock, each from a replicate cell culture plate of a specific stock. A maximum of $1x10^7$ cells was used for a single extraction. Total RNA was eluted in 50 µl of Rnase-free water.

The quality of the RNA was determined using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies). To minimize adverse effects of protein contamination, only the RNA samples with 260/280 ratio between 1.9 and 2.1 and 260/230 ratio greater than 2.0 were used for the analysis.

For RNA-seq, poly(A) mRNA was isolated from total RNA using beads with oligo(dT). cDNA was synthesized from short mRNA fragments (200-700 nucleotides) produced using an in house fragmentation buffer, using Random hexamer-primers. Subsequently, the second-strand cDNA was synthesized which was used in the preparation of the cDNA libraries by purification of short fragments using QiaQuick PCR extraction kit (Qiagen) and subsequent adding of poly(A) and ligation of sequencing adaptors. Suitable fragments, with respect to the result of agarose gel electrophoresis, were selected for amplification with PCR. Finally, the libraries were subjected to sequencing using Illumina HiSeq™ 2000.

Sequence mapping: Since the RNA material used for sequencing was prepared from bovine and buffalo lymphoblast cell cultures infected with *T. parva* Muguga and *T. parva* 7014 schizonts, respectively, all reads were mapped against the *Bos taurus* genome sequence (accession number: DAAA00000000) to exclude host sequence reads from the analysis. For mapping of parasite reads, *T. parva* Muguga genome sequence (accession number: AAGK00000000) was used as a reference. Thus, using TopHat (v2.0.4), SOAPsplice and SOAP softwares from the ENSEMBL database, clean reads were mapped to reference

genome and gene sequences of *T. parva and Bos taurus*. The total number of reads aligned to reference genes was counted and located in relative positions in reference genes.

Gene expression analysis: The Reads Per kb per Million reads method (RPKM) (Mortazavi et al. 2008) was applied to calculate the Unigene expression. The calculated gene expression was directly used for comparing the differences in gene expression between *T. parva* 7014 and *T. parva* Muguga samples, using the latter isolate as a reference. FDR (False Discovery Rate) control is a statistical method used in multiple hypothesis testing to correct for p-value. Thus, transcripts with the FDR≤0.001 and Fold Change >2 were regarded as significantly differentially expressed.

For validation of differentially expressed transcripts, cDNA was synthesized from total RNA prepared for RNA-Seq, using the iScript Advanced cDNA Synthesis kit (BIO-RAD) according to the manufacturer's instructions.

2.2 Primer design and synthesis

Twenty differentially expressed genes (DEGs) and 10 stably expressed genes were selected for validation of expression profiles from RNA-seq data obtained using NGS. DEGs were selected from genes with the false discovery rate (FDR) ≤0.001 and Fold Change >2. Genes with RPKM (Reads per kilo base per million) value >10 but <100 were selected to represent moderately expressed genes while transcripts with RPKM value >100 were selected to represent highly expressed genes. Five house-keeping genes (HKGs) were also analyzed for expression stability. Oligonucleotide primers for amplification and quantification of DEGs and HKGs were designed using Primer-Blast online software (NCBI, USA) and synthesized by Life technologies (USA). Primers of lengths between 18–24 bp were designed to generate amplicons of sizes ranging between 80 and 125 bp, with melting temperatures between 57-60°C (Table 2.1). To determine if the primers will amplify the specific target genes, the primer sequences were screened for hairpins, dimer formation, and target specificity using BLASTN (NCBI, USA). Wherever possible, primers were designed spanning an intron to detect any genomic DNA contamination. Primer sequences and parameters of the target genes involved in this study are depicted in Table 2.1; these include primers for HKGs identified from existing literature. Primers for constitutively expressed genes (n=10), genes expressed exclusively in one isolate (n=5 for each isolate), up-regulated (n=5) and down-

regulated (n=5) genes, identified from the NGS dataset. The gene annotations for the selected genes were determined using Blast2GO (Conesa *et al*., 2005).

To further confirm the specificity of the primers for amplification, qPCR was performed as described in section 2.3 followed by analysis of amplicons employing agarose gel electrophoresis using 2 % agarose gels prepared and run in 1% Tris Borate EDTA (TBE) buffer for 30 minutes. Ethidium bromide with a final concentration of 0.3 µg/ml was added to the agarose gel mix to allow visualization of amplicons under UV light.

2.3 Quantitative real-time PCR reaction and amplification conditions

cDNA prepared from total RNA was used for qPCR experiments which were performed employing the ABI StepOnePlus™ system and software (Applied Biosystems, USA). The qPCR reaction mixture contained 2.5 µl of template cDNA (~36.0ng/µl), 10 µl of 2× Platinum SYBR Green SuperMix-UDG[1] (Life technologies, SA), and 0.4 µl of 10 µM stock concentration of each gene-specific primer in a final volume of 20 µl. All qPCR reactions were performed under the following conditions: UDG activation for 2 minutes at 50 °C, denaturation for 2 minutes at 95 °C, and 40 amplification cycles of denaturation for 10 seconds at 95 °C, primer annealing for 10 seconds at 58 °C and final extension for 1 minute at 72 °C. The specificity of the qPCR reaction for each amplified product was verified by melting curve analysis which was carried out as follows: 15 seconds at 95° C, 1 minute at 60 °C, (with 20°C/s transition rate, and then ramping to 95°C at 0.2°C/s transition rate) and 15 seconds at 95 °C. Two biological replicates for each target gene were used for qPCR analysis and three technical replicates were analyzed for each biological replicate in two independent runs. In each qPCR run, a no template control as well as an uninfected bovine cDNA control were included.

Since samples used for this study were contaminated with host RNA both from bovine and buffalo it was not possible to calculate the copy numbers or concentration of the parasite RNA material. Thus, to quantify the parasite cDNA employing a standard curve generated using qPCR, the blue tongue viral (BTV) cDNA was used. Prior to quantification, the efficiency of the qPCR reaction was evaluated using *T. parva* cDNA and blue tongue viral

[1] Platinum® SYBR® Green qPCR SuperMix-UDG is a ready-to-use cocktail containing all components required for qPCR, except primers. It is supplied at a 2x concentration and contains Platinum® Taq DNA polymerase, SYBR® Green I dye, Tris-HCl, KCl, 6 mM MgCl2, 400 µM dGTP, 400 µM dATP, 400 µM dCTP, 800 µM dUTP, uracil DNA glycosylase (UDG), and stabilizers.

(BTV) cDNA in order to be able to quantify *T. parva* cDNA using a standard curve generated from BTV cDNA. The BTV gene which was amplified was the gene that encodes the variable outer shell protein using the same qPCR conditions as described above. The following forward and reverse primers were used to amplify the BTV gene; F 5'CGG ACC GCA TTA TGG TAT AAC C 3' and R 5' ACT CTT GTG TCT CGT ACT TTC AAC 3' (Hoffman *et al*., 2009). The concentration for the dilutions ranged from 1000 ng/µl to 0.1 ng/µl and triplicate samples of each dilution were subjected to qPCR. All qPCR reactions were carried out using the same conditions as described above.

Table 2.1: Primer sequences used for evaluation of housekeeping genes and validation of constitutively and differentially expressed genes.

Accession No.	Gene description	Primer sequence 5'-3'	Amplicon length (bp)	Primer melting temperature (°C)
Housekeeping Genes				
XM_760535.1	Fructose bisphosphate aldolase (F6P)	F- TATGCGCTGGTGTGTCAGTC R- CACTCCTCTGCTGAATGGCT	86	58
AB499089.1	Cytochrome b	F- GGTAAGACCCTGTGCACCTT R- CACCTCCATGTCGTCTCACC	84	58
L28998.1	28S rRNA	F- CGACTGTCCGTGAAAAGGGA R- AACCTTGGAGACCTGATGCG	80	58
XM_760378.1	β-Actin	F- ATGTTGCAATCCAGGCCGTA R- GTGGGTGACACCATCTCCAG	86	58
XM_760333.1	Glyceraldehyde-3-phosphate dehydrogenase	F- CCCCTCAATGACGCCAGAAT R- TTCCTCGAGCAGCAATGTGT	88	58
Constitutively Expressed Genes				
XM_758090.1	ABC transporter	F- AATCCTGCCACATACACCAC R- CTCGTCTGTGTGCTTGTCTT	120	58
XM_759942.1	DNA polymerase delta small subunit	F- TCCTACTCGGAAAGGTTCGT R- CATGGATTGGAGCTTCGACA	120	58
XM_757859.1	Transcription factor	F- ATCGTCTCAGTGGACAGGAA R- ACGTTATTTGCCCTGGAGTC	125	58
XM_758301.1	Membrane transporter	F- TGACTGCACACCACTTCTTC R- CAGTTAAACCTGACACCGCT	120	58
XM_758360.1	ATP-dependent RNA helicase ddx56-like	F- AAAGCGGTTCATCCCTTACC R- TTATCGTTTGCGCTTGAGGA	120	58
XM_758834.1	26S proteasome non-ATPase regulatory subunit 13	F- AGGGCAAATCACACACTCTG R- AGGTTTTCTGAACGCCAAGT	120	58
XM_759535.1	Spermatogenesis-associated protein 5	F- GGGAGCAGAAGTCGTCAACA	115	60

		R- TCCAAGCATTTCTGTGGGCT		
XM_760594.1	MPP 10 protein	F- TGGAGAGGTTCGCAGATGAG R- ACATGCTGGTGTTGAGTCGT	125	60
XM_760631.1	GTP-binding protein	F- CTCCGGGAAGAACAAGGCAT R- GCACATCCATATCCGGGGAG	120	60
XM_759300.1	DCP 1-like decapping family protein	F- CAAACGCCTTTCGTTACTGGA R- CCCTTCAATTCCCATCCTAGACC	120	60
Genes Exclusively Expressed in *T. parva* Muguga				
XM_759484.1	Hypothetical protein	F- TAACGCCGATGTGGAGTGTT R- TGAGTTAAGTAAATTGCTCTCGGAT	90	59
XM_759485.1	Hypothetical protein	F- CCAAGGGACCAGAAGGGATTG R- ATCACTTGGTTCGGATGGCTG	80	60
XM_767539.1	Hypothetical protein	F- GGCCACTGAAGTAGTAGACGC R- GCTGCTTTGAGGGTTGCTATG	99	60
XM_760433.1	Sub-telomeric SVSP protein family	F- CCAACAAGGGTCTGGGCTTA R- CTCCCTCGGTAGTTTCAGTCA	81	60
XM_758501.1	40Sribosomal protein	F- AAGTTCATGGATCGCTCGCT R- TTGTCTCTTCTTCGCTCGCC	107	60
Genes Exclusively Expressed in *T. parva* 7014				
XM_757579.1	Ribosomal protein L2	F- AGCAAAAAGCTGGTACTGAAAGA R- TGAGGATGGTCACAAGCGTT	82	60
XM_757605.1	DNA-directed RNA polymerase	F- ACCTATTTTTGTTGAAGGTGAGTCT R- TTGGTCTCCATCAAAATCTGCG	82	59
XM_758244.1	Hypothetical protein	F- ACCAACAGCAGCCGATAAGA R- GGCGGTGCCATCGAATAAAA	94	59
XM_757584.1	Ribosomal protein L14	F- AAATCCGATATGGTTCGTGTGT R- TTGTGCCCTTATTAACAACAAAATC	70	58
XM_757615.1	Hypothetical protein	F- TGGTATTTAGAGGAGTTTTGGGC R- CACCCATTCCAATCTCAATAGTTCA	120	59
Genes up-regulated in *T. parva* Muguga compared to T. *parva*7014				
XM_759970.1	Transcription factor with ap2 domain	F- GCTAGGAACGGGTTTGGATT R- ACTGCATAACAGGGACCACT	80	58
XM_757780.1	Hypothetical protein	F- GTGAGAAGGGAGTCAGATTCG R- CTTGATTGACTCAAATACATGCGA	80	58

XM_760940.1	Myosin light chain	F- GCAAATTGGCCTTGTGTCAT R- GCAAGGCCCACTGTAACTAT	96	58
XM_761074.1	ABC transporter	F- CCAGCCATAAGCAGAAGTGT R- CCTCATCCTCCAATGTCACG	88	58
XM_761046.1	Tash 1-like protein	F- AAGGGGCATCCTGGTGAATA R- AATTTCGAGAGAGGCGAGTC	84	58
Genes up-regulated in *T. parva*7014 compared to T. *parva* Muguga				
XM_757708.1	Regulator of chromosome condensation	F- TAGATTCCGAGCCAGCAGTA R- AGGAGAGGAACATGGTGTGA	84	58
XM_757582.1	50S ribosomal protein	F- AACACCGAACCAAATCGAAG R- ATGAGTAGGACGTACTTTTATCCA	91	57
XM_757773.1	Hypothetical protein	F- GAGCAACTTGCCCTTGGATT R- ATCTGACTTCTCGGAAAGCG	111	58
XM_758397.1	Apical membrane antigen-1	F- TAAATGACAGCGCTCAGGAC R- GACGAGTAGTACGTGAAGCC	88	58
XM_757777.1	DNA-directed polymerase iii subunit rpc 2	F- ACCTGGAAATCAGCCCAATG R- ATGTGTTTCTTGGGCTCTGG	87	58

F= Forward primer; R= Reverse primer

2.4 Evaluation of housekeeping genes expression stability

For identification of reliable reference genes for normalization of qPCR data, five housekeeping genes (HKGs) expressed in *Theileria parva* including genes encoding β-actin, glyceraldehyde-3-phosphate dehydrogenase, 28S rRNA, cytochrome b and fructose bisphosphatase aldolase (F6P), were identified from literature and evaluated. The amplification of HKGs was performed in cDNA prepared from both *T. parva* Muguga and *T. parva* 7014 isolates, using the qPCR protocol presented in section 2.3. To determine which of the five HKGs show stable expression in the two *T. parva* isolates, the cycle of quantification (Cq) values obtained from qPCR were used. The qPCR data was analyzed employing RefFinder (https://www.reffinder.net/), a web-based tool that integrates the currently available major computational programs (geNorm, Normfinder, BestKeeper, and the comparative $\Delta\Delta C_t$ method) to compare and rank the tested candidate reference genes. The most stable candidate genes within and between the test groups are those with the lowest variation values. The use of the two reference genes in combination is generally more accurate than just using the most stable gene (Gimeno *et al*., 2014). Therefore, the two most stable HKGs were ultimately used as internal control genes in the validation of gene expression profiles of DEGs selected from NGS dataset.

2.5 Evaluation of qPCR assays for precision and reproducibility

Quantitative real-time PCR assay precision was determined using a randomly selected gene, that codes for a *T. parva* transmembrane protein (accession number XM_758301.1). Amplification was performed across a 10-fold dilution series (36.0 ng/µl to 0.0036 ng/µl) prepared from *T. parva* Muguga and 7014 cDNA samples and the Cq values for each isolate were separately plotted against the cDNA concentrations. The linearity of amplification for the selected genes was determined by the R^2 values of each dilution series. Intra-assay (across the plate) and inter-assay (between plates) variability was determined according to MIQE (Minimal information for Publication of Quantitative Real-time PCR Experiments) guidelines (Bustin *et al*., 2009). The mean of Cq values, standard deviation (S.D) and coefficient of variation (CV) of Cq values were calculated separately for amplification of representative target genes from *T. parva* Muguga and *T. parva* 7014 cDNA. The intra-and inter-assay variability was assessed using the CV value, which was determined through dividing the S.D by the mean Cq value; the resulting value was multiplied by 100 to express CV as a percentage.

2.6 Quantification of gene expression

2.6.1 Validation experiment for relative quantitative analysis using comparative C_t (or $\Delta\Delta C_t$) analysis

To analyze the expression profiles of the selected target genes from the NGS dataset, the $\Delta\Delta C_t$ method using the StepOnePlus™ software was employed. However, for a valid $\Delta\Delta C_t$ analysis, the gene target and endogenous control amplification assays are required to perform at approximately the same efficiency (97-110%). Thus, the efficiencies of the assays of the two isolates were evaluated by generating standard curves for target genes and two endogenous control genes (β-Actin and 28S rRNA) using the qPCR conditions described in section 2.3. Serial dilutions of concentrations ranging from 116.0 ng/µl to 0.0116 ng/µl were prepared from *T. parva* Muguga and *T. parva* 7014 cDNA. For qPCR reactions, 2.5 µl of each dilution was used as a template and all dilutions were run in triplicate. The mean C_t values generated from equivalent standard curve mass points (target gene vs endogenous control gene) were then used in the ΔC_t calculation ($C_{T\ \text{target gene}}$- $C_{T\ \text{endogenous control gene}}$). These ΔC_t values were plotted against log input amount of cDNA to create a semi-log regression line. The slope of the semi-log regression line was used as a general criterion for passing the validation experiment. The absolute value of the slope of ΔC_t versus log input has to be <0.1 for the experiment to be considered valid (Pfaffl, 2001).

2.6.2 Gene expression analysis for differentially expressed and constitutively expressed genes

In all of the following qPCR validation experiments, *T. parva* Muguga was used as the reference sample while *T. parva* 7014 was used as the test sample.

The expression profiles of constitutively expressed genes and DEGs, including genes exclusively expressed, and up- and down-regulated genes, were validated using qPCR. The selected genes were amplified from both *T. parva* Muguga and *T. parva* 7014 using qPCR (in the same run) using the conditions described in section 2.3. Each gene target and two HKGs were run in triplicate and the mean Cq values were used for analysis. Using the $\Delta\Delta C_T$ method, the data was presented as the fold change in gene expression normalized to the two endogenous control genes and relative to the reference sample (*T. parva* Muguga). Each target gene in each sample was normalized by subtracting the mean C_T value of the two

endogenous control genes (β-actin and 28S rRNA) from the mean C_T value of the target gene (C_{Tmean} target gene -C_{Tmean} endogenous control genes), thereafter the difference of each normalized target gene was obtained (ΔC_T 7014 - ΔC_T Muguga) and the fold difference was calculated using the equation $2^{-\Delta\Delta C_T}$. The up-regulated and down-regulated genes amplification data was analyzed with relative quantification min/max confidence of 95%.

Chapter 3 : RESULTS

3.1 Identification of differentially expressed genes from *T. parva* Muguga and *T. parva* 7014 Next Generation Sequencing dataset

RNA from *T. parva* Muguga (cattle-derived) and *T. parva* 7014 (buffalo-derived) isolates was sequenced using Next Generation Sequencing (NGS) and the dataset was analyzed to identify genes that are differentially expressed between the two isolates. Gene expression variation analysis was performed against the *T. parva* Muguga transcriptome as a reference. The differentially expressed genes (DEGs) were selected from genes with the false discovery rate (FDR) ≤0.001 and Fold Change >2. From the NGS dataset analysis 1048 DEGs were detected and 2906 genes were constitutively expressed between *T. parva* Muguga and *T. parva* 7014 (Fig. 3.1). Noteworthy was the 742 (70.8%) genes significantly up-regulated in *T. parva* Muguga and 309 (29.5%) in *T. parva* 7014. Among the DEGs were genes that were exclusively expressed in one of the isolates; 69 genes were detected only in *T. parva* Muguga while 52 genes were detected only in *T. parva* 7014.

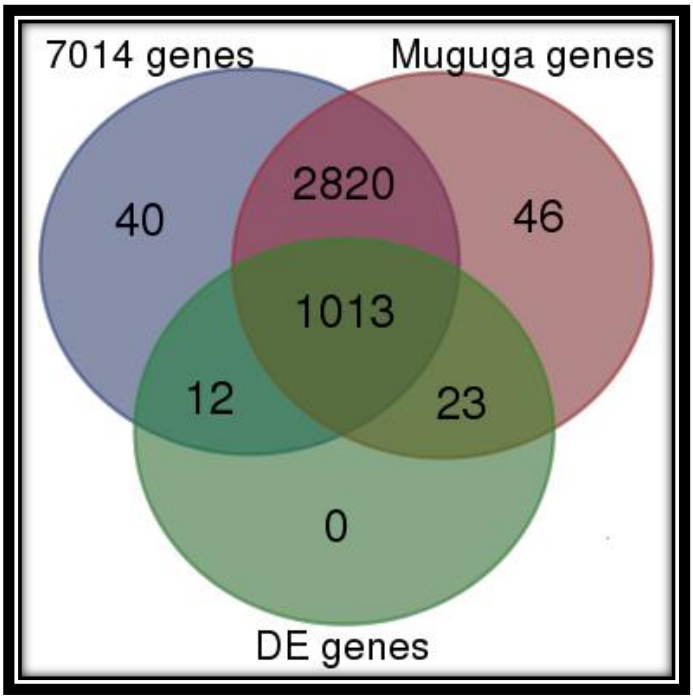


Figure 3.1: The number of expressed genes identified in *T. parva* (Muguga) and *T. parva* (7014) transcriptomes. A total of 3954 transcripts were successfully mapped to the *T. parva* genome sequence and of these 1048 (26.5%) were differentially expressed (DE) [(FDR) ≤0.001 and Fold Change >2] with 69 exclusively detected in *T. parva* Muguga and 52 in *T. parva* 7014 (KP Sibeko, Department of Veterinary Tropical Diseases, University of Pretoria, unpublished data).

3.2 Evaluation of housekeeping genes for use as internal control genes for normalization of qPCR data

3.2.1 Quantitative real-time PCR amplification efficiency and *T. parva* cDNA quantification

A standard curve was generated by qPCR amplification of a gene encoding beta-actin from blue tongue viral cDNA of known concentration to allow quantification of *T. parva* Muguga and *T. parva* 7014 cDNA since the RNA samples used for cDNA preparation were contaminated with bovine and buffalo host nucleic acid material, thus making it difficult to quantify parasite cDNA exclusively using standard methods. The target gene for *T. parva* was the 28S rRNA gene. Before quantifying the parasite cDNA, the working efficiency of the standard curve was evaluated and determined to be 97%, with a correlation coefficient (R^2) value of 0.995 and a slope of -3.387 (Fig. 3.2).Comparison of BTV cDNA and *T. parva* cDNA amplification show that the two had more or less an equivalent amplification efficiency. The parasite cDNA quantity interpolated from the standard curve was 36.03 ng/µl for *T. parva* Muguga and 127.77 ng/µl for *T. parva* 7014. *Theileria parva* 7014 cDNA samples were then diluted to match the 36.03 ng/µl concentration of *T. parva* Muguga in order to eliminate concentration bias in downstream gene expression analysis.

3.2.2 Confirmation of primer specificity

Oligonucleotide primers were designed for evaluation of five housekeeping genes (HKGs). The primer specificity to amplify the targeted gene regions using qPCR was determined by melting curve analysis. *T. parva* (Muguga and 7014) cDNA was used as template. Single product specific melting peaks were detected at different melting temperatures for each gene product: fructose bisphosphate aldolase (F6P), 86°C; cytochrome b, 80°C; 28S rRNA, 85°C; β-actin, 81°C; and glyceraldehyde phosphate dehydrogenase (GAPDH), 79°C (Fig. 3.3 A). In addition, PCR products were analyzed with agarose gel electrophoresis to determine the amplicon size and single products of desired length, i.e. F6P = 86 bp; cytochrome b = 84 bp; 28S rRNA = 80 bp; β-actin = 86 bp; and GAPDH = 88 bp, were detected (Fig. 3.3 B). No primer-dimers were generated during the applied 40-cycle qPCR assays.

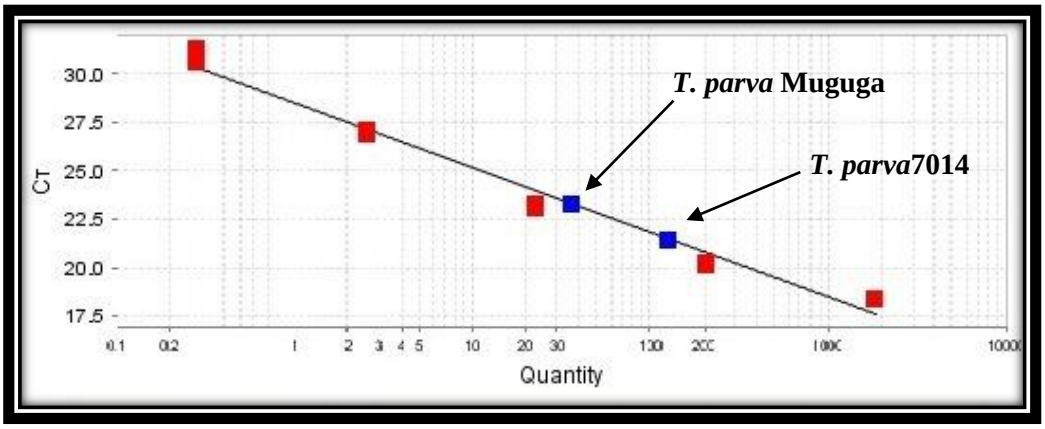
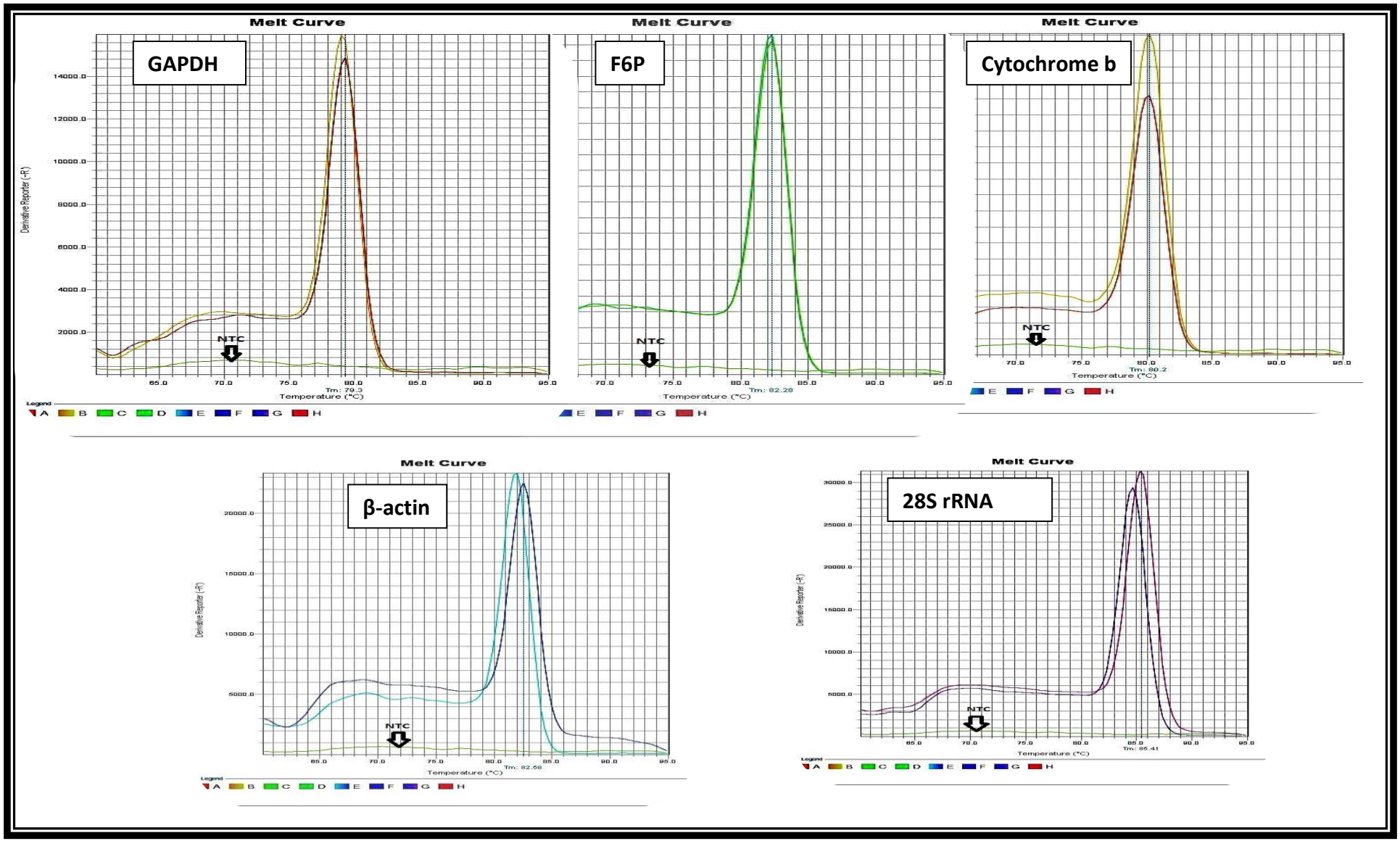


Figure 3.2: A standard curve generated using 10-fold dilutions of blue tongue viral cDNA of known concentrations (red blocks) for quantification of the amount of *T. parva* cDNA present in the cDNA samples known to be contaminated with bovine and buffalo host nucleic acid material (blue blocks).

A



B

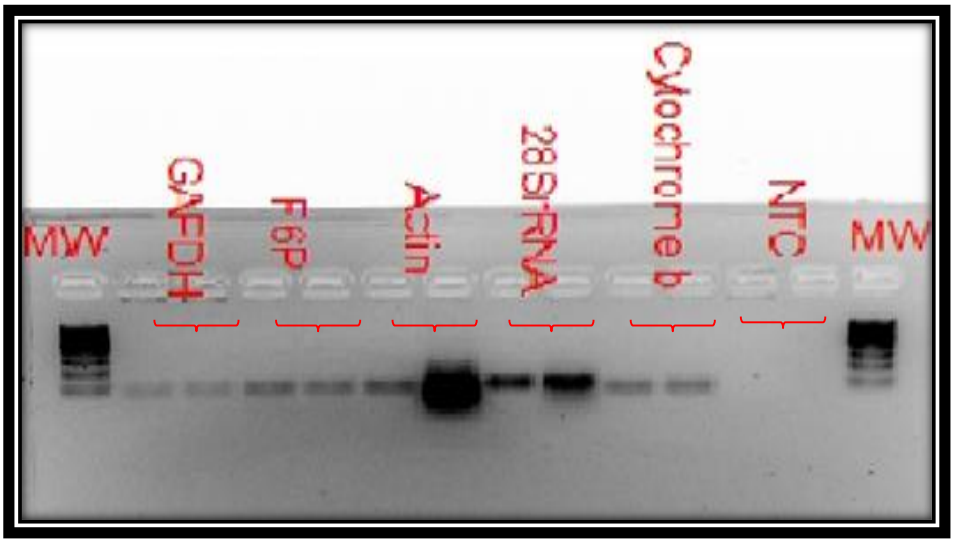


Figure 3.3: **Specificity of qPCR amplification.** A) **Melting curves of the five housekeeping genes showing a single melting peak for each product (each including two biological replicates and a no template control [indicated as NTC]) obtained from one of the *T. parva* isolates.** B) **Agarose gel (2%) showing amplification of a specific PCR product of the expected size for each HKG tested in the study using a 100 bp molecular weight marker.**

3.2.3 Analysis of HKGs expression stability

The web-based software RefFind was employed to analyze the expression stability of the HKGs and rank them accordingly. ReFfinder integrates major computational programs, such as geNorm, Normfinder, BestKeeper, and the comparative delta C_T method to compare and rank the tested candidate endogenous control genes. Based on the rankings from each program, it assigns an appropriate weight to an individual gene and calculates the geometric mean of their weights for the overall final ranking, then each gene is assigned an expression stability value (*M*). Genes with the lowest *M* value have the most stable expression.

Two separate qPCR runs were performed for analysis of expression stability of the HKGs. In the first run, the expression stability values of 28S rRNA (*M*=0.171) and β-actin (*M*=0.171) were the lowest (Fig. 3.4). In the second run, the same genes, 28S rRNA (*M*=0.141) and β-actin (*M*=0.213), had the lowest expression stability values (Fig. 3.4).

Table 3.1 illustrates the overall expression stability of the HKGs in both runs according to each computational program. Consequently, 28S rRNA and β-actin were selected for use as internal control genes for normalization of qPCR data in DEG expression profile validation experiments.

3.2.4 Intra- and inter-assay variation analysis

In order to determine the precision of the resulting data, intra- and inter-assay variation analyses were performed for expression of HKGs between *T. parva* Muguga and *T. parva* 7014, using the Student's *t*-test (Table 3.2; Fig. 3.5 A-B). The Student's *t*-test analysis for both inter- and intra-assay variation illustrated that there was no significant variation in expression of the housekeeping genes between the two *T. parva* isolates under investigation. Inter-assay (run to run) variation analysis illustrated low coefficient of variation percentage in all five HKGs (Table 3.2).

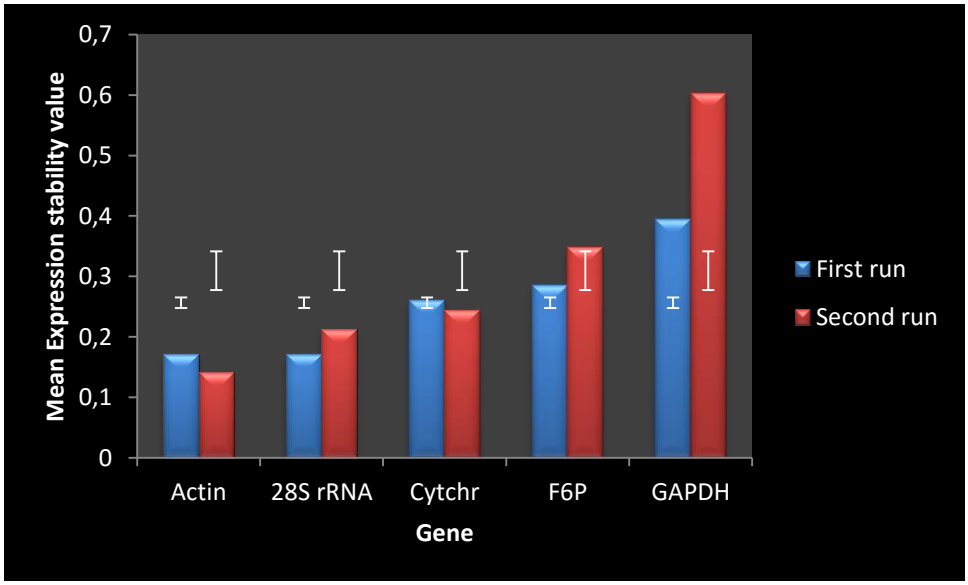


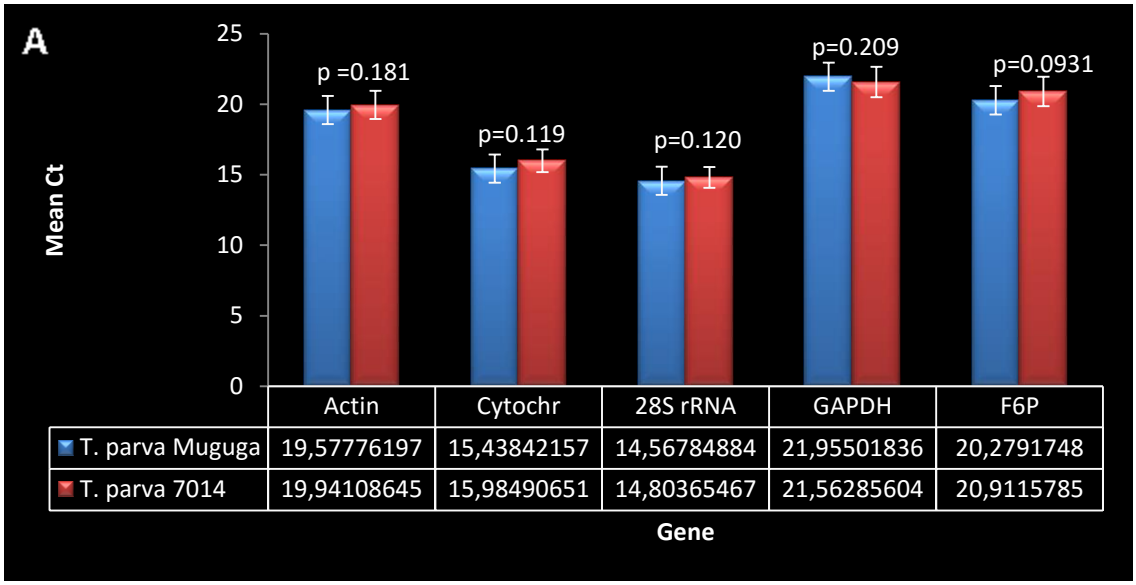
Figure 3.4: **Expression stability rankings of the five selected candidate reference genes obtained from the two qPCR runs for HKGs evaluation. Actin = β-actin; Cytochr = cytochrome b; F6P = fructose bisphosphatase aldolase; GAPDH = glyceraldehyde phosphate dehydrogenase.**

Table 3.1: Gene expression stability of selected HKGs as assessed by RefFinder.

Ranking Order (Better--Good--Average)					
Method	**1**	**2**	**3**	**4**	**5**
Delta C_T	β-actin	Cytochrome b	28S rRNA	F6P	GAPDH
Bestkeeper	28S rRNA	β-actin	Cytochrome b	F6P	GAPDH
Normfinder	β-actin	Cytochrome b	28S rRNA	F6P	GAPDH
geNorm	Cytochrome b/28S rRNA		β-actin	F6P	GAPDH
Recommended comprehensive ranking	**β-actin**	**28S rRNA**	**Cytochrome b**	**F6P**	**GAPDH**

Table 3.2: **Analysis of inter-assay variations in gene expression of the HKGs for *T. parva* Muguga and *T. parva* 7014 as determined by the Student's *t*-test.**

Gene	Run 1 mean Ct	Run 2 mean Ct	Standard deviation	Coefficient of variation (%)
28S rRNA	14.54018	14.68575	0.072788	0.49
Cytochrome b	15.76403	15.71166	0.026	0.17
β-actin	20.05952	19.75942	0.150047	0.75
GAPDH	23.63574	21.75894	0.9384	4.1
F6P	21.03138	20.59538	0.218001	1.1



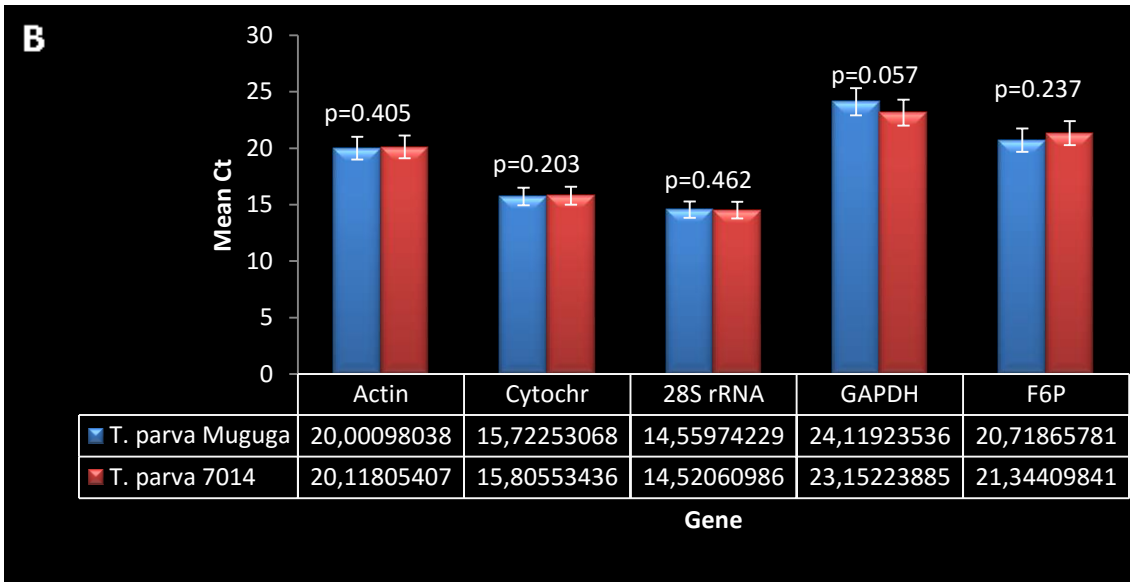


Figure 3.5: **Student's *t*-test results for analysis of intra-assay variations in the expression of HKGs between *T. parva* Muguga and 7014. No significant differences were observed (p>0.05). The results from the first run are presented in** panel A **and results from the second run in** panel B**. Actin = β-actin; Cytochr = cytochrome b; F6P = fructose-2.6-bisphosphatase; GAPDH = glyceraldehyde phosphate dehydrogenase.**

3.3 Validation of NGS data by relative quantification of expression using comparative CT (or ΔΔCT) analysis

The expression profiles of constitutively expressed genes and DEGs were analyzed by qPCR using the comparative C_T method. The following sections describe all the steps involved in achieving the final fold change analysis.

3.3.1 Evaluation of the amplification efficiency of target genes (DEGs and constitutively expressed genes) and endogenous control genes (28S rRNA and β-actin).

The comparative C_T method requires that the assays of the target and endogenous control genes have similar amplification efficiency. Therefore, the amplification efficiencies of target genes and endogenous control genes were evaluated by constructing validation plots from standard curves generated from selected constitutively expressed (n=1) and differentially expressed genes (n=1), in comparison to the standard curves of the two endogenous control genes (28SrRNA and β-actin). The validation plots were constructed for each parasite isolate (Fig. 3.6 A-D) and the slopes of the curves for all the validation plots were below 0.1 as required, ranging from 0.08 to -0.1.

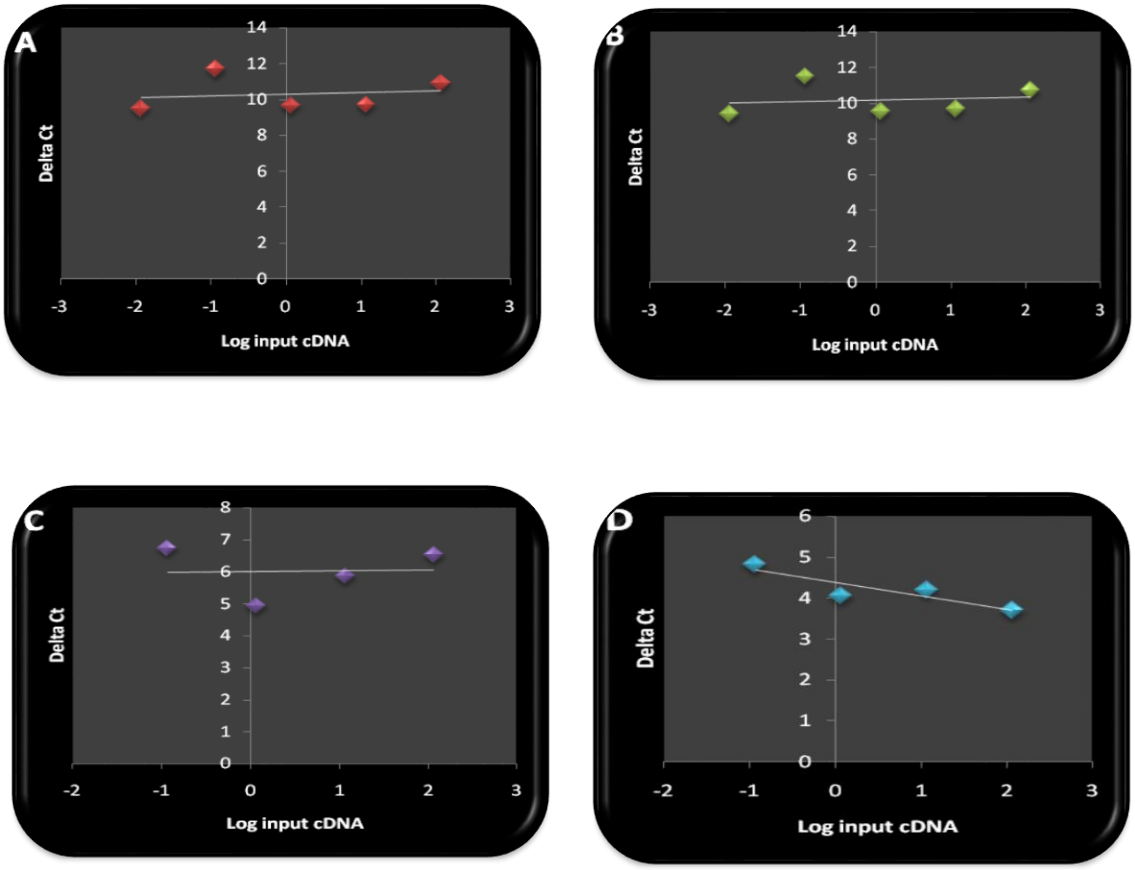


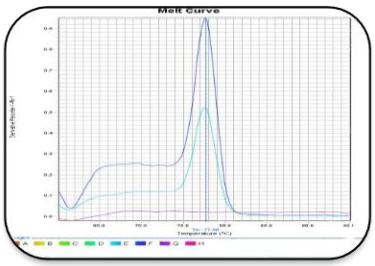
Figure 3.6: Validation plots of ΔC_T vs. log input cDNA amount. Panel A represents amplification of a constitutively expressed target gene (gene encoding a transcription factor; accession # XM_757859.1), amplified from a dilution series prepared from *T. parva* Muguga cDNA. Panel B represents amplification of the same constitutively expressed target gene amplified from a dilution series prepared from *T. parva* 7014 cDNA. Panel C represents amplification of a differentially expressed target gene (gene encoding tash1-like protein; accession # XM_761046.1), amplified from a dilution series prepared from *T. parva* Muguga cDNA. Panel D represents amplification of the same differentially expressed target gene amplified from a dilution series prepared from *T. parva* 7014 cDNA. The slopes of the plots in panel A-D were 0.083889; 0.080895; 0.028094; and -0.32435, respectively.

3.3.2 Quantitative real-time PCR validation analysis of constitutively expressed genes

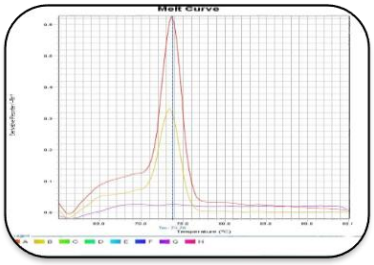
Ten genes identified from NGS data as constitutively expressed in *T. parva* Muguga and *T. parva* 7014 were selected for validation using qPCR. Prior to validation of the expression profiles of the selected genes, the specificity of the primers to amplify the targeted regions using qPCR was evaluated by melting curve analysis and agarose gel electrophoresis. *Theileria parva* (Muguga and 7014) cDNA was used as template. Single product specific melting peaks were detected at different melting temperatures for each gene product (Fig. 3.7 A) and agarose gel electrophoresis analysis indicated that single amplicons of the desired lengths were obtained (Fig. 3.7 B).

The expression profiles of the ten constitutively expressed genes were then analyzed using the comparative C_T method in which each constitutively expressed gene was first normalized using two endogenous control genes (β-actin and 28S rRNA) (Fig. 3.8) and then the fold difference in expression of each gene between the two *T. parva* isolates was calculated (Table 3.3). The qPCR results indicated that only four genes showed insignificant variation in expression ($0.5 \leq$ fold change ≥ 2) between the two *T. parva* isolates (Table 3.3.2). Contrary to the NGS data, four of the ten selected genes were up-regulated in *T. parva* Muguga (fold change >2) while one gene was up-regulated in *T. parva* 7014 (fold change <0.5).

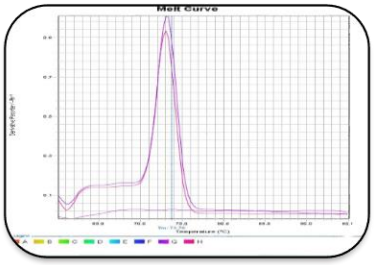
A



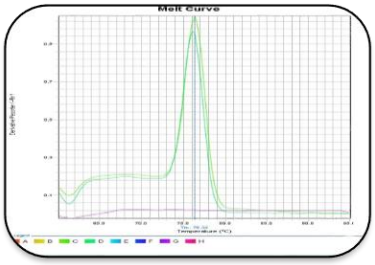
XM_758090.1



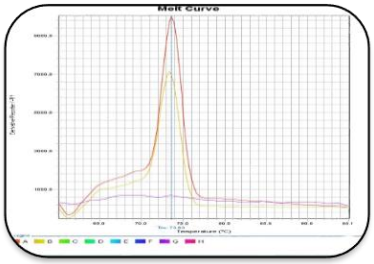
XM_757859.1



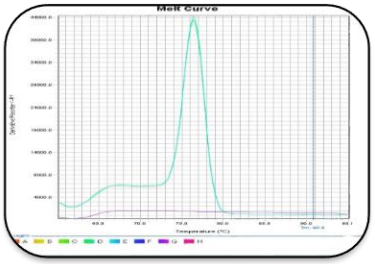
XM_758360.1



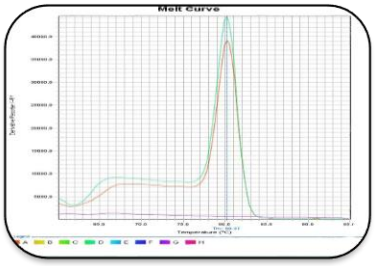
XM_758301.1



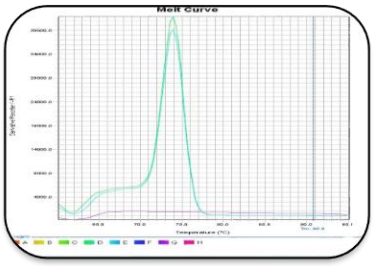
XM_758834.1



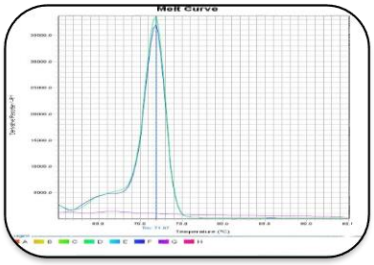
XM_759300.1



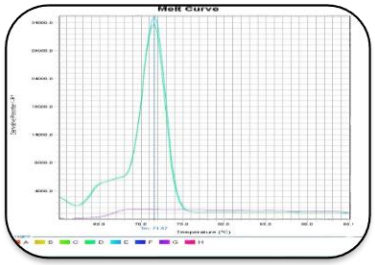
XM_759535.1



XM_75994.1



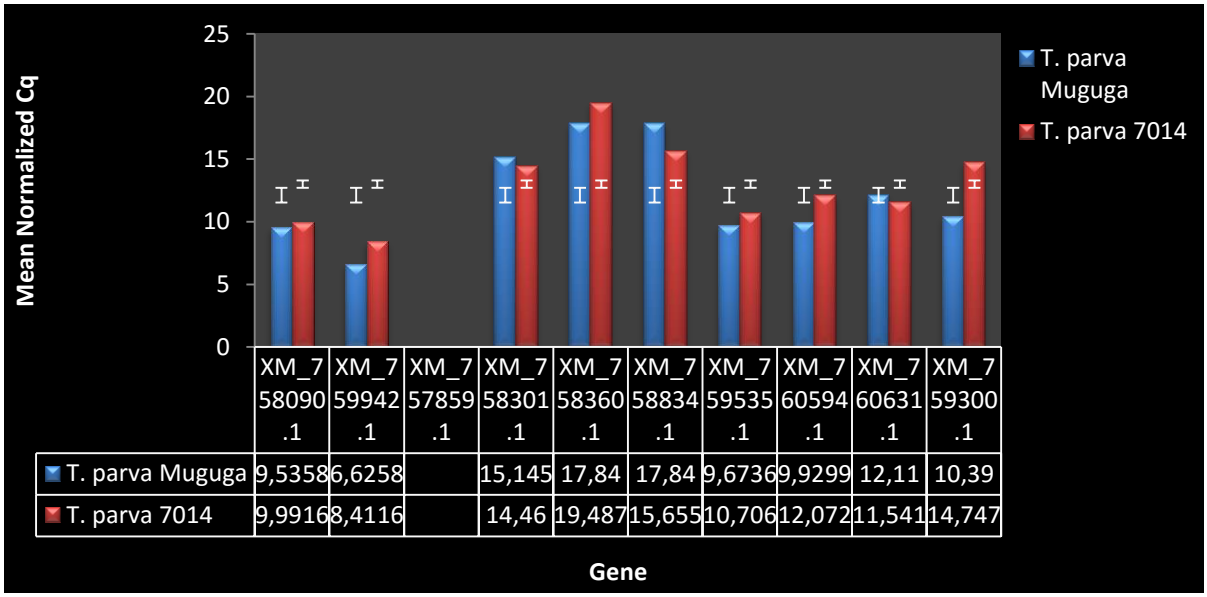
XM_760594.1



XM_760631.1

B

Figure 3.7: **Specificity of qPCR amplification.** A**) Melting curves of the ten genes selected for constitutive expression validation showing a single melting peak for each product (each including two biological replicates and a no template control).** B**) Agarose gel (2%) showing qPCR amplicons generated by specific amplification of the 10 selected stably expressed genes. Lanes 1-10 represent genes with the accession number: XM_758090.1; XM_757859.1; XM_758360.1; XM_758301.1; XM_758834.1; XM_759300.1; XM_759535.1:XM_759942.1; XM_760594.1; XM_760631.1 respectively.**



	XM_758090.1	XM_759942.1	XM_757859.1	XM_758301.1	XM_758360.1	XM_758834.1	XM_759535.1	XM_760594.1	XM_760631.1	XM_759300.1
T. parva Muguga	9,5358	6,6258		15,145	17,84	17,84	9,6736	9,9299	12,11	10,39
T. parva 7014	9,9916	8,4116		14,46	19,487	15,655	10,706	12,072	11,541	14,747

Figure 3.8: **Normalization of targeted constitutively expressed genes (n=10) against endogenous control genes (β-actin and 28S rRNA) in *T. parva* Muguga and *T. parva* 7014 isolates.**

Table 3.3: **Fold change in expression of ten selected genes in *T. parva* 7014 relative to *T. parva* Muguga.**

Transcript	$\Delta\Delta C_t$ (ΔC_t 7014-ΔC_tMuguga)	$2^{-\Delta\Delta Ct}$
XM_758090.1	0.455	1.37*
XM_759942.1	1.786	3.45**
XM_757859.1	undetermined[#]	undetermined[#]
XM_758301.1	-0.685	0.62*
XM_758360.1	1.647	3.13**
XM_758834.1	-2.185	0.20***
XM_759535.1	1.032	2.05*
XM_760594.1	2.143	4.42**
XM_760631.1	-0.569	0.67*
XM_759300.1	4.357	20.49**

*0.5≤ fold change ≥2=invariant expression; ** fold change >2=up-regulated in Muguga;*** fold change <0.5=up-regulated in 7014; [#]=no amplification of target gene.

3.3.3 Quantitative real-time PCR validation analysis of genes exclusively expressed in *T. parva* Muguga and *T. parva* 7014

Genes identified as exclusively expressed in each isolate by NGS (n=5 from each isolate) were selected for validation using qPCR. Prior to validation of the expression profiles of the selected genes, the specificity of the primers to amplify the targeted regions using qPCR was evaluated by melting curve analysis and agarose gel electrophoresis. *Theileria parva* (Muguga and 7014) cDNA was used as template. Single product specific melting peaks were detected at different melting temperatures for each gene product (Fig. 3.9 A-B) and from the agarose gel electrophoresis analysis single amplicons of desired lengths were obtained (Fig. 3.9 C).

A

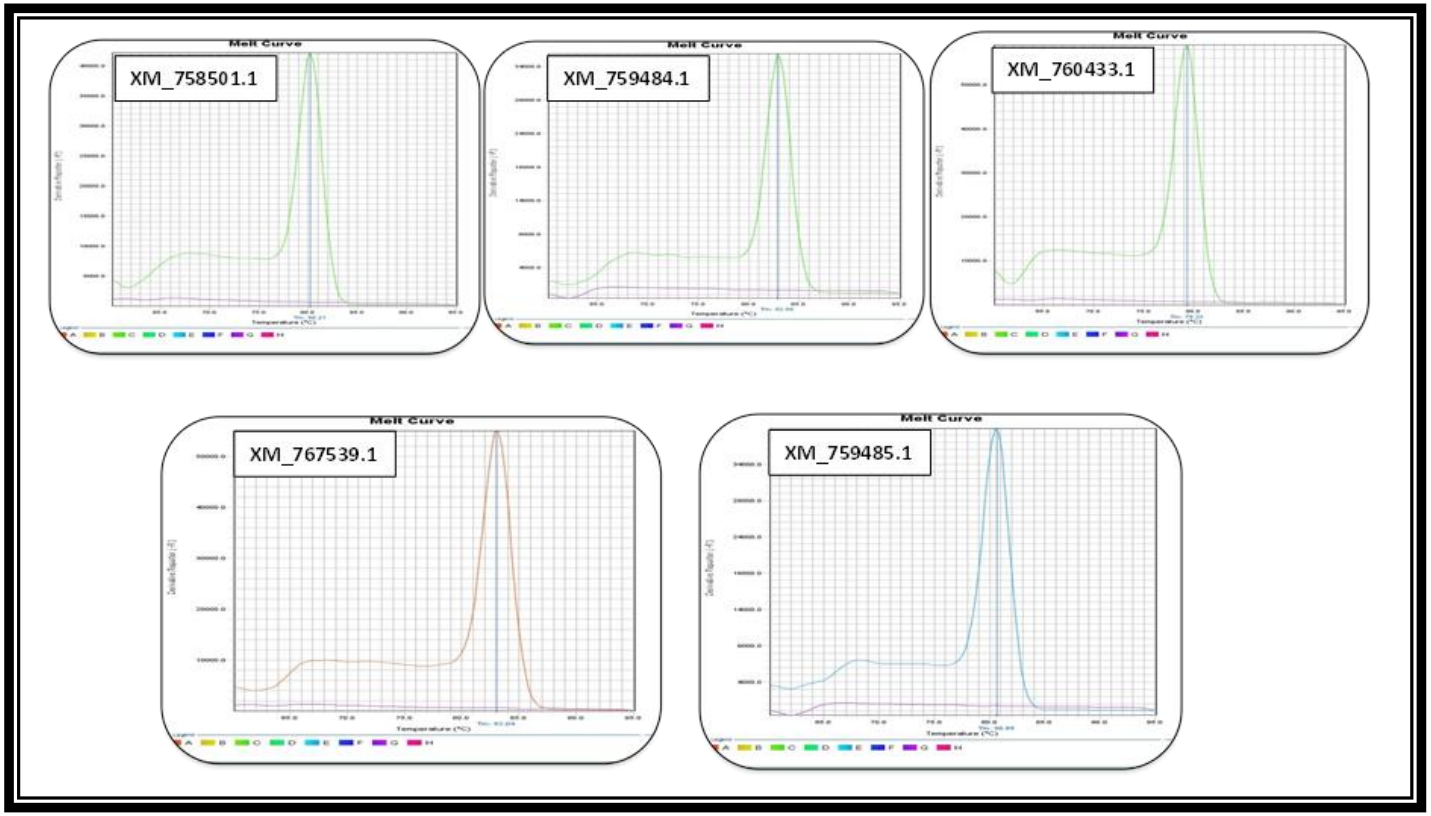


Figure 3.9: Panel A **illustrates the melting curve analyses (each including two biological replicates and a no template control) of the five genes exclusively expressed in *T. parva* Muguga selected for validation while** panel B **shows the melt curves of the five genes exclusively expressed in *T. parva* 7014.** Panel C **illustrates the agarose gel (2%) showing amplicons from specific amplification of the five genes exclusively expressed in *T. parva* Muguga (lane 1-5: accession number: XM_759484.1; XM_759485.1; XM_767539.1; XM_760433.1; XM_758501.1 respectively) and the five genes exclusively expressed in *T. parva* 7014 (lane 6-10: accession number: XM_757579.1; XM_757605.1; XM_758244.1; XM_757584.1; XM_757615.1). The figure continues in the next page.**

B

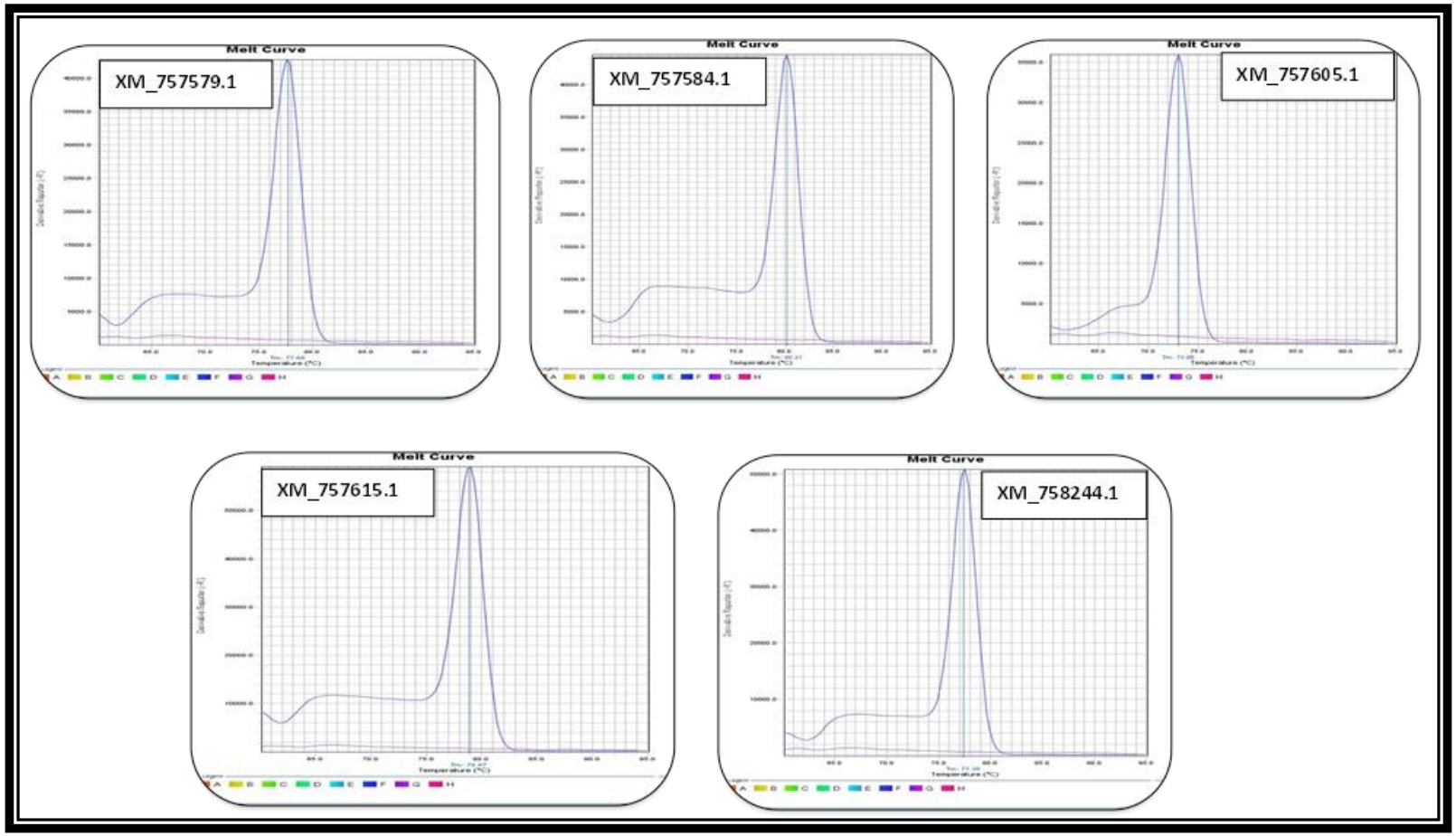


Figure 3.9: Continues from page 45

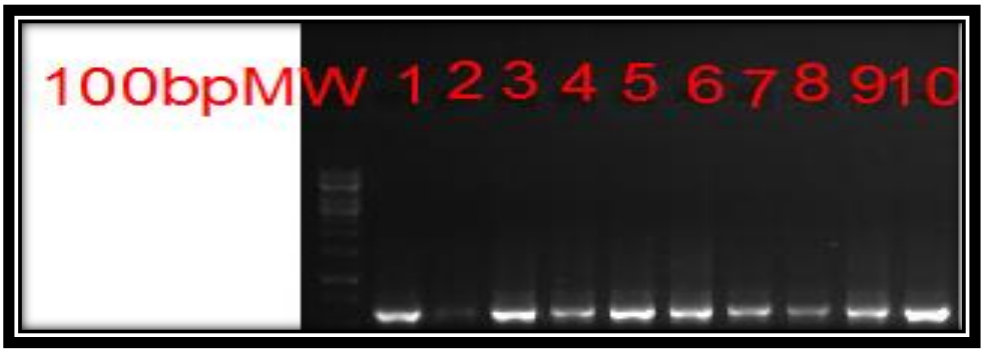


Figure 3.9: Continues from pages 44 and 45.

The expression profiles of the selected exclusively expressed genes were then analyzed using the comparative C_T method in which each gene was first normalized using two endogenous control genes (β-actin and 28S rRNA) and then the fold difference in expression of each gene between the two *T. parva* isolates was calculated (Table 3.4). The qPCR results illustrated that out of the five genes selected for validation from 69 genes exclusively expressed in *T. parva* Muguga according to NGS analysis, four (XM_759484.1; XM_759485.1; XM_767539.1; XM_760433.1) were confirmed to be expressed in *T. parva* Muguga and were not detected in *T. parva* 7014 cDNA (Fig. 3.10A). The one remaining gene (XM_758501.1) was detected in both isolates but it was shown to be highly up-regulated in *T. parva* Muguga, with a fold change of 10942.8. Out of the five genes selected for validation from 52 genes exclusively expressed in *T. parva* 7014 according to NGS analysis, qPCR analysis confirmed that four (XM_757579.1; XM_757605.1; XM_757584.1; XM_757615.1) were expressed only in *T. parva* 7014 and not in *T. parva* Muguga, while one gene (XM_758244.1) could only be amplified from *T. parva* Muguga (Fig. 3.10B).

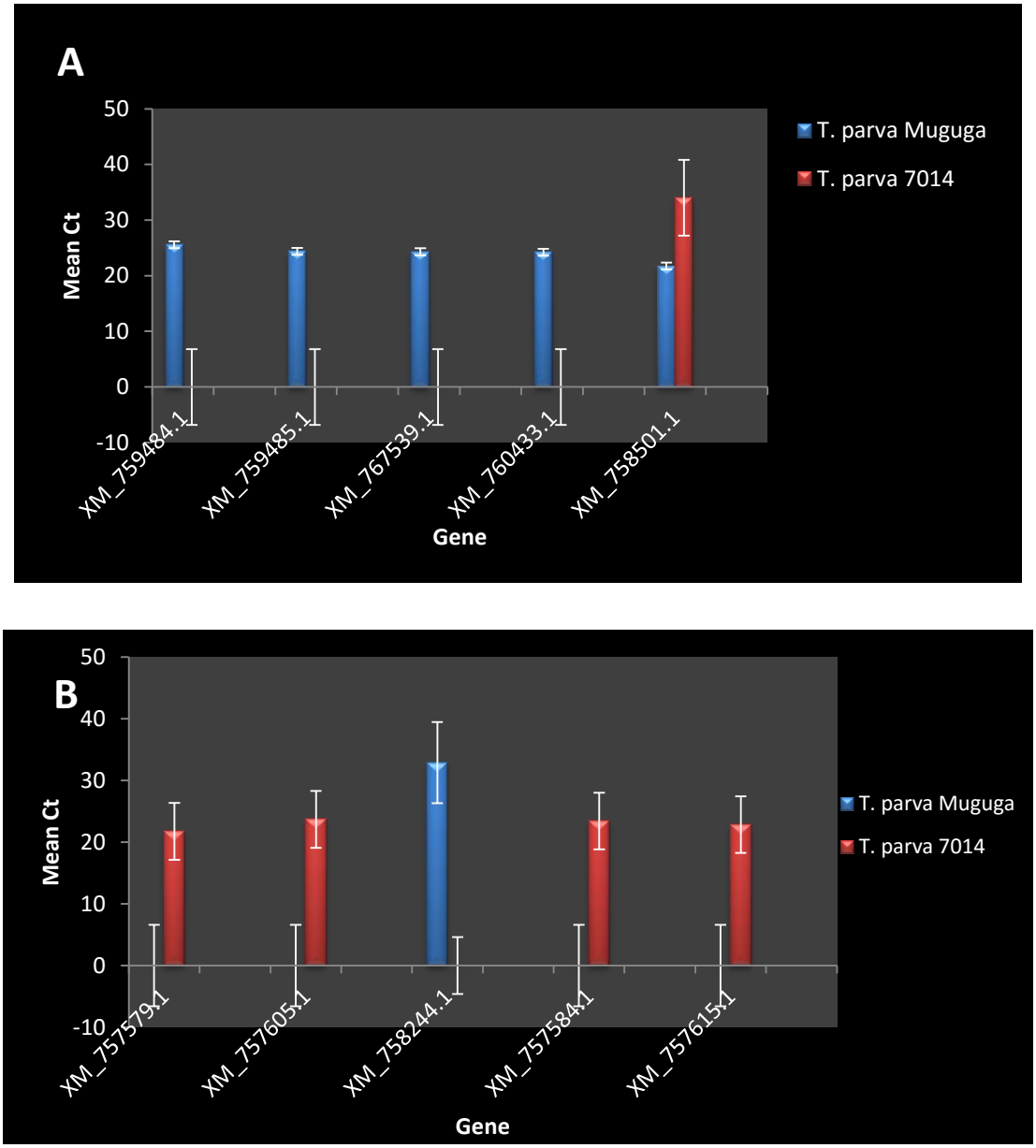


Figure 3.10: **Panel A- Comparison of expression profiles of the five selected genes in *T. parva* Muguga and *T. parva* 7014 using mean Ct values obtained from qPCR analysis. Four genes were detected only in *T. parva* Muguga (XM_759484.1; XM_759485.1; XM_767539.1; XM_760433.1) confirming that they are exclusively expressed in *T. parva* Muguga. The fifth gene (XM_758501.1) was detected in both *T. parva* Muguga and *T. parva* 7014. Panel B- Comparison of expression of the five selected genes in *T. parva* Muguga and *T. parva* 7014 using mean Ct values obtained from qPCR analysis. Four genes were detected only in *T. parva* 7014 (XM_759484.1; XM_759485.1; XM_767539.1; XM_760433.1) confirming that they are exclusively expressed in *T. parva* 7014. The fifth gene (XM_758244.1) was detected only in *T. parva* Muguga.**

3.3.4 Analysis of genes up and down-regulated in *T. parva* 7014 relative to *T. parva* Muguga

From the 742 and 309 genes identified as significantly up-regulated in *T. parva* Muguga and *T. parva* 7014, respectively according to NGS analysis, five genes for each isolate were selected for validation using qPCR. Prior to validation of the expression profiles of the selected DEGs the specificity of the primers to amplify the targeted regions using qPCR was evaluated by melting curve analysis and agarose gel electrophoresis. Single product specific melting peaks were detected at different melting temperatures for each gene product (Fig. 3.11 A-B) and from the agarose gel electrophoresis analysis single amplicons of desired lengths were obtained (Fig. 3.11 C).

A

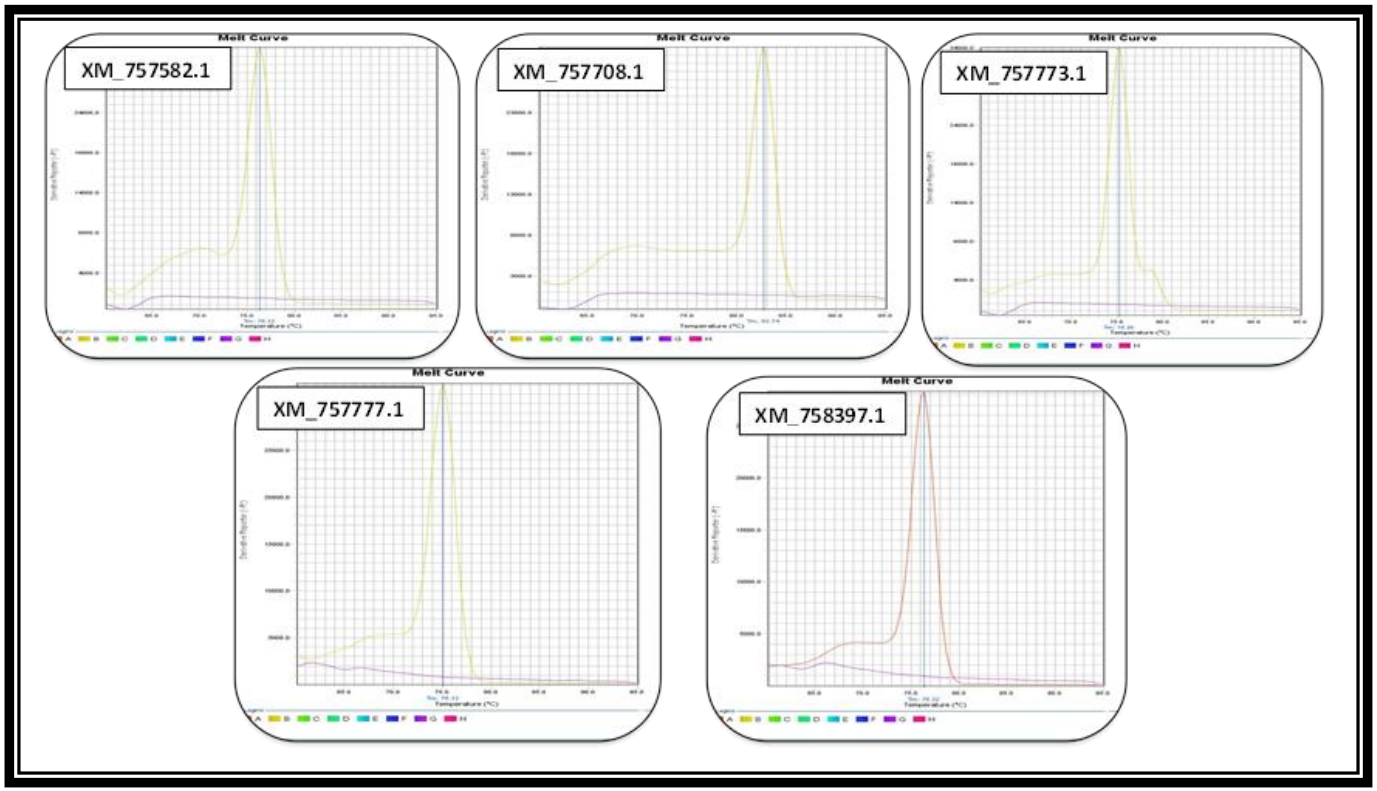


Figure 3.11: Panel A **illustrates the melting curve analysis (each including two biological replicates and a no template control) of the five genes up-regulated in *T. parva* 7014 relative to *T. parva* Muguga selected for validation while** panel B **shows the melt curves of the five genes down-regulated in *T. parva* 7014 relative to *T. parva* Muguga.** Panel C **illustrates the agarose gel (2%) showing amplicons from specific amplification of the five genes up-regulated in *T. parva* 7014 relative to *T. parva* Muguga (lane 1-5: accession number: XM_757708.1; XM_757582.1; XM_758397.1; XM_747773.1; XM_757777.1 respectively) and five genes down-regulated in *T. parva* 7014 relative to *T. parva* Muguga (lane 6-10: accession number: XM_759970.1; XM_757780.1; XM_761074.1; XM_760940.1; XM_761046.1). The figure continues in the next page.**

B

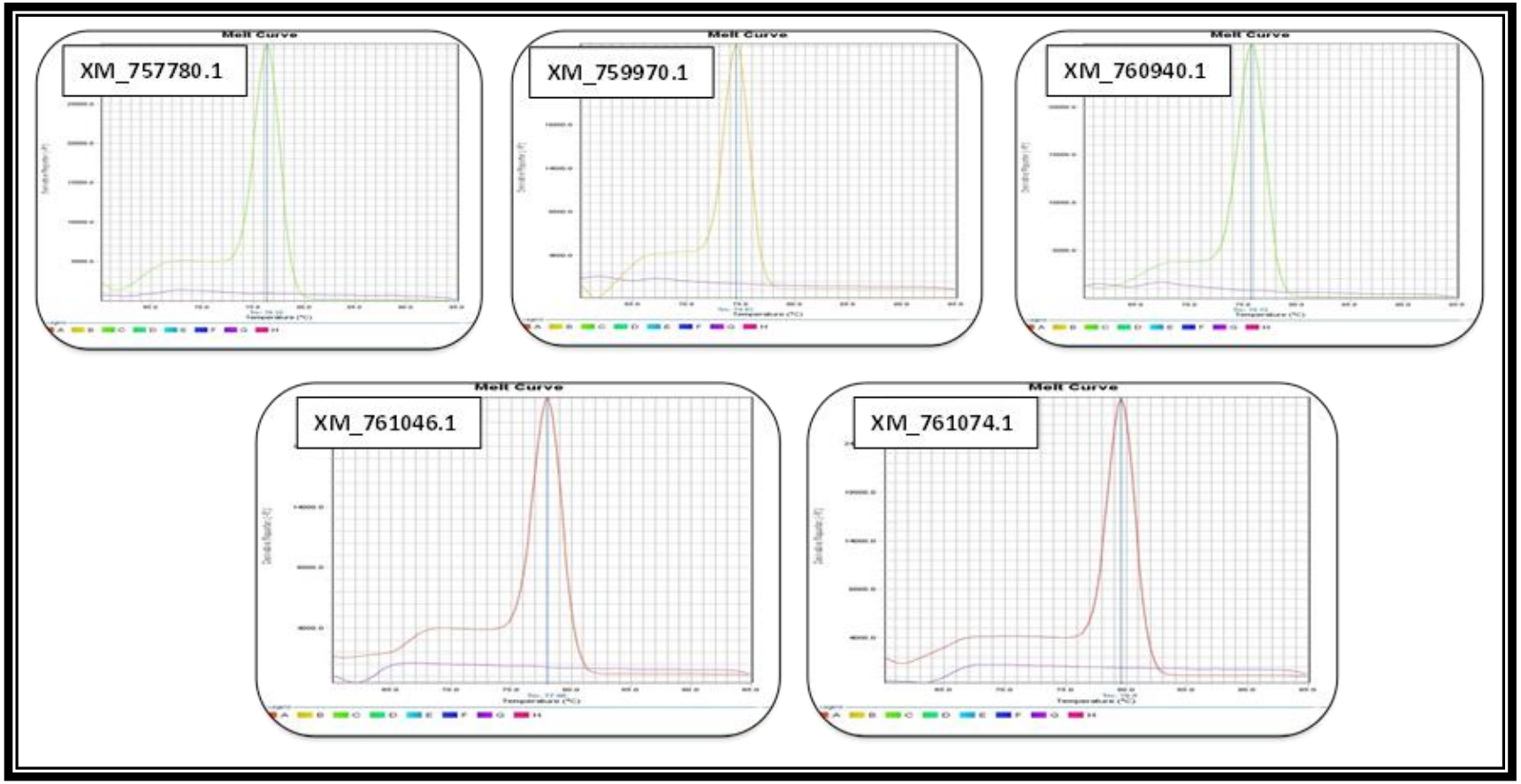


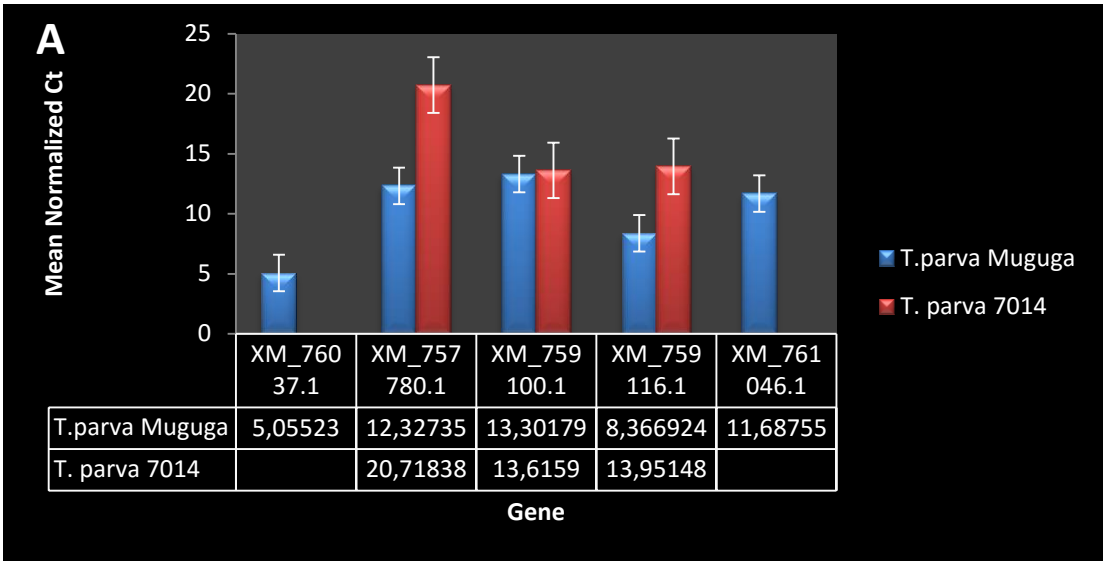
Figure 3.11: Continues from page 50.

C

Figure 3.3.4.1: Continues from the previous pages 50 and 51.

The expression profiles of the five genes up-regulated in *T. parva* Muguga and five genes up-regulated in *T. parva* 7014 were then analyzed using the comparative C_T method in which each of the DEGs was first normalized (Fig. 3.12 A-B) using two endogenous control genes (β-actin and 28S rRNA) and then the fold difference in expression of each gene between the two *T. parva* isolates was calculated (Table 3.4).

Consistent with NGS data, from the ten DEGs selected for validation of up- and down-regulated genes, three genes with the following accession numbers: XM_757708.1; XM_757582.1; XM_757777.1 were up-regulated in *T. parva* 7014 and three genes (XM_757780.1; XM_761074.1; XM_760940.1) were down regulated in the same isolate (Table 3.4). Notable were two genes (XM_761046.1 and XM_758397.1) which only amplified in *T. parva* Muguga and one (XM_757582.1) that was only detected in *T. parva* 7014 cDNA. One gene (XM_757773.1) showed to be constitutively expressed between the two isolates.



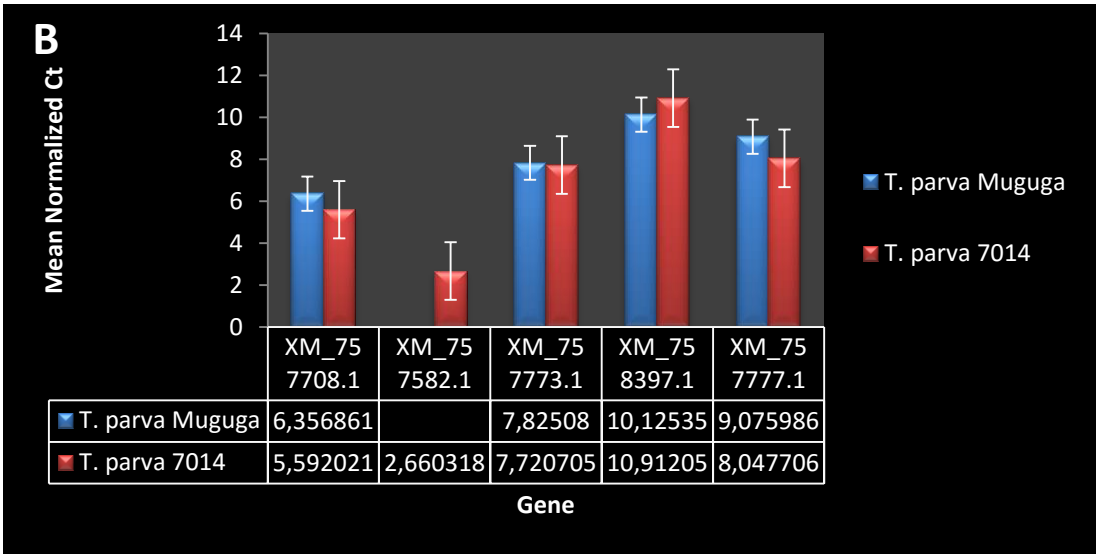


Figure 3.12: **Normalization of the ten DEGs selected for validation.** Panel A **represents five genes shown by NGS analysis to be down-regulated in *T. parva* 7014 relative to *T. parva* Muguga.** Panel B **represents five genes shown by NGS analysis to be up-regulated in *T. parva* 7014 relative to *T. parva* Muguga.**

Table 3.4: **Illustration of the fold difference in expression of the selected DEGs in *T. parva* 7014 relative to *T. parva* Muguga and comparison to NGS analysis.**

Gene ID	$2^{-\Delta\Delta C_T}$	Expression profile in *T. parva* 7014 according to NGS	Expression profile in *T. parva* 7014 according to qPCR (relative to *T. parva* Muguga)	RPKM value
XM_759970.1	Undetermined **(Only amplified in *T. parva* Muguga/7014)**	Down-regulated	Down-regulated	0.089487
XM_757780.1	335.7	Down-regulated	Down-regulated**	0.865323
XM_760940.1	1.2	Down -regulated	Down-regulated*	1.976318
XM_761074.1	48.0	Down -regulated	Down-regulated**	1.34485
XM_761046.1	Undetermined **(Only amplified in *T. parva* Muguga)**	Down -regulated	Down-regulated	0.156637
XM_757708.1	0.6	Up-regulated	Up-regulated*	62.89662
XM_757582.1	Undetermined (**Only amplified in 7014**)	Up-regulated	Up-regulated	731.2102
XM_757773.1	0.9	**Up-regulated**	**Constitutively expressed**	35.587
XM_758397.1	1.7	**Up-regulated**	**Down-regulated***	39.35173
XM_757777.1	0.4	Up-regulated	Up-regulated**	23.92806

** Significantly up or down-regulated (0.5≤ fold change ≥2); * insignificantly up or down-regulated (0.5> fold change <2) (Dalman *et al*., 2011).

Overall, the comparison of NGS and qPCR data demonstrated more discrepancies (70%) for genes identified to be constitutively expressed by NGS RNA-seq analysis (Table 3.5). On the

contrary, data from 70% of the DEGs used for validation by qPCR corroborated with NGS data.

Table 3.5: **Summary of results showing data agreement or disagreement between transcriptome analysis using NGS and expression analysis using qPCR**

	AGREE	DISAGREE
	Constitutively expressed (n=10)	
Muguga/7014	3	7 (5 up-regulated in Muguga, 1 up-regulated in 7014, 1 no amplification)
	Exclusively expressed (n=5 for each isolate)	
Muguga	4	1 (up-regulated in Muguga)
7014	4	1 (detected in Muguga)
	Up-regulated (n=5 for each isolate)	
Muguga	3 (2 detected only in Muguga)	2
7014	3	2
TOTAL (DEGs)	**14 (70%)**	**6 (30%)**

Chapter 4 : DISCUSSION

Transcriptomics refers to the study of the complete set of transcripts in a specific cell, tissue, or organism for a given developmental stage or physiological condition (Wang *et al*., 2009). This complete set of transcripts is known as a transcriptome, including protein-coding messenger RNA (mRNA) and non-coding RNA [ncRNA: ribosomal RNA (rRNA), transfer RNA (tRNA), and other ncRNAs] (McCarthy *et al*., 2012). Unlike the relatively stable genome, the transcriptome varies with developmental stage, physiological condition, and external environment. Transcriptome analysis is a powerful tool for dissecting the relationship between genotype and phenotype, leading to a better understanding of the underlying pathways and mechanisms controlling cell fate, development, and disease progression. The aims of transcriptomics are not limited to the quantification of change in expression level for each gene among different transcriptome samples but include also the mapping and annotation of the transcriptome and the determination of the functional structure of each gene in the genome (Wang *et al*., 2009).

Next Generation Sequencing technologies present the possibility of hypothetical discovery of novel transcripts and isoforms in a shorter fraction of time. However, multiple template preparation stages, diverse sequencing chemistries and complex data processing of NGS experiments may impact on the verification of authentic RNA biomarkers (Robles *et al*., 2012); it is thus imperative to validate gene expression data by employing different techniques in order to eliminate any biasness and discrepancies presented by the gene expression analysis techniques. Quantitative real-time PCR is central to biomarker validation where potential markers need to be measured with greater accuracy and precision. Consequently, the study at hand focused on validating the NGS transcriptomics data by employing qPCR.

Before validation of differential expression of a set of genes using qPCR, endogenous control genes had to be selected in order to normalize qPCR gene expression data (Robles *et al*., 2012). The prerequisite of a suitable reference gene is that it should be adequately expressed in the tissue of interest, but most importantly, it should show minimal variability in expression between samples and under the experimental conditions used (Dheda *et al*., 2004). In this study five candidate reference genes were selected which encoded GAPDH, F6P, cytochrome b, 28S rRNA and β-actin. All of these genes have previously been described as

housekeeping genes of protozoan parasites (Gomez *et al*., 2010; Ersfeld, 2003). However to our knowledge, there are no reports on the evaluation of the expression stability of these genes in *T. parva* isolates; hence the expression stability of the selected candidate reference genes was evaluated employing the web-based software RefFinder. The RefFinder tool integrates the currently available major computational programs (geNorm, Normfinder, BestKeeper, and the comparative ΔΔCt method) to compare and rank the tested candidate reference genes. The most stable candidate genes within and between the test groups are those with the lowest variation values. The use of two reference genes in combination is generally more accurate than just using the most stable gene (Robles *et al*., 2012). It is evident from Table 3.2 that the criteria for stability ranking differ from program to program; hence the outcome of the stability rankings varied according to the program. It is for this reason that the RefFinder tool integrates the different programs and gives a recommended comprehensive ranking. Therefore, based on this comprehensive analysis between the two *T. parva* isolates, the use of 28S rRNA and β-actin as endogenous control genes for studies that involve gene expression analysis of *T. parva* is proposed.

A number of parameters is elementary to the success and adaptability of a qPCR assay for gene expression analysis. Firstly, it is essential that the qPCR system has the capability to recognize multiple strains of the *T. parva* parasite with a high level of specificity for each target gene. In addition, with multiple gene targets on one plate, optimized performance with a single thermal cycling protocol and reaction conditions are necessary for utilizing the system as a qPCR array. Lastly, it is imperative for the qPCR system to quantify RNA levels accurately across a wide linear range of template concentrations with minimal intra- and inter-assay variability (Bullard *et al*., 2010). Intra- and inter-assay variation analysis of the selected HKGs showed that there was no significant variation in the expression of these genes between the two *T. parva* isolates with the p values being greater than 0.05 and the coefficient of variation percentage being low (<2) for all the genes tested. The absence of significant intra-or inter-assay variation enables plate-to-plate comparisons between results obtained and provides statistical significance when examining replicate datasets.

Quantitative real-time PCR is a highly sensitive technique that allows quantification of rare transcripts and small changes in gene expression. In this study the comparative C_T method of qPCR was employed to analyse the expression profiles of genes selected from the NGS dataset between two *T. parva* isolates. The comparative C_T method utilizes arithmetic

formulas to achieve the result for relative quantification which is expressed as a fold change in genes expression ratio. However, the prerequisite for using this method is that it should be validated by ensuring that the target genes have approximately the same amplification efficiency as the endogenous control genes. This study passed the comparative C_T method validation experiment since the absolute slopes of ΔC_T vs. Log input cDNA for the selected target genes were all less than 0.1.

Using the NGS RNA-seq data, differentially expressed genes were selected from genes with the false discovery rate (FDR) ≤0.001 and Fold Change >2. Transcripts with RPKM (Reads per kilo base per million) value >10 but <100 were selected to represent moderately expressed genes while transcripts with RPKM value >100 were selected to represent highly expressed transcripts. Three groups selected for validation included: 1. constitutively expressed genes; 2. up- and down-regulated genes and 3. genes exclusively expressed in one isolate or the other. Discrepancies between RNA-seq and qPCR analysis were observed from all three groups but mostly in the constitutively expressed group of genes; in this group only 40% of the qPCR results corroborated with RNA-seq findings while 60% demonstrated variations in expression with four genes down-regulated and one up-regulated in *T. parva* 7014 relative to *T. parva* Muguga. Since most of the disagreements in the two datasets were down-regulated expression, this finding suggests that RNA-seq was more sensitive in detecting low abundant RNA transcripts. Comparing absolute quantitation of expression levels across platforms can result in less than ideal correlation, especially when normalization is handled differently for the approaches being compared. This could be a problem even when comparing two NGS platforms using normalized counts because each platform has its own associated biases. The solution would be to correlate relative expression across platforms. Studies in micro RNA (miRNA) analysis using Illumina (formerly Solexa sequencing) revealed that when all miRNAs were considered in the analysis, there was no significant correlation between Illumina reads and qPCR data. However, a significant positive correlation was observed when miRNAs with more than 100 Illumina reads were considered. These results suggest that NGS data with less than 100 reads can only roughly represent the relative abundance of miRNAs (Cristino *et al*., 2011). RNA-seq and qPCR are both valuable technologies for gene expression analysis, each with their own strengths and limitations. The main benefits of RNA-seq are the broad scope of genes being interrogated, its compatibility with allele and transcript-specific RNA quantification, and the possibility to discover hitherto unknown transcripts. Quantitative real-time PCR on the other hand provides

excellence for sensitive RNA quantification of a targeted set of genes. It is thus important to understand the relationship between sensitivity and read depth. Results from this study clearly indicate that the detection and quantification sensitivity of RNA-seq is very much dependent on the read depth. RNA-seq data showed to have a high coverage compared to qPCR, however, at an affordable 100M coverage, RNA-seq begins to suffer from reduced quantification and detection sensitivity (compared to RT-qPCR). In our study, the expression profiles determined by RNA-seq and qPCR correlated for four genes; these were genes coding for an ABC transporter protein, a membrane transporter protein, a spermatogenesis associated protein 5 and a GTP-binding protein. Transporters of the ATP-Binding Cassette (ABC) family are known to provide the basis of multidrug resistance, thus it is essential to the parasite that the genes coding for these proteins be conserved in different strains of the parasite (Higgins, 1993). It may be inferred from this validation finding that these four constitutively expressed genes may be used as possible candidate reference genes for *T. parva* gene expression studies; however, further validation using other *T. parva* isolates is needed.

From the genes shown to be exclusively expressed in either *T. parva* 7014 or *T. parva* Muguga by RNA-seq, 80% (4/5 for *T. parva* 7014 and 4/5 for *T. parva* Muguga) of the genes selected for validation produced the same outcome on qPCR. According to qPCR analysis, one gene coding for 40S ribosomal protein was up-regulated in both *T. parva* Muguga and 7014 while it was only detected in *T. parva* Muguga according to NGS. This finding suggests that this gene transcript may occur in low abundance in *T. parva* 7014, demonstrating that when using NGS approaches for gene expression profiling, genes with low expression can go undetected if the sequence depth is not high enough to allow detection of transcript occurring at low levels, corroborating with previous reports which have shown that RNA-seq is less precise for weakly expressed genes (McIntyre *et al.*, 2011; Mooney *et al.*, 2013). This was not the case though with the second gene which was expected to be detected only in *T. parva* 7014 but was only detected in *T. parva* Muguga. It has been reported that discrepancies in gene expression results, obtained using different platforms, are most likely when measuring genes expressed at low levels (Wang *et al.*, 2009). Therefore, it is possible that our gene occurred in low abundance in both isolates and therefore is not consistently detected by different platforms used for gene expression studies, in this case RNA-seq and qPCR.

A similar scenario was also observed in the qPCR analysis of DEGs up or down-regulated in *T. parva* 7014 relative to *T. parva* Muguga. Two genes which were up-regulated in *T. parva* Muguga according to RNA-seq were only detected in *T. parva* Muguga using qPCR, while another gene, expected to be up-regulated in *T. parva* 7014 from RNA-seq analysis, was only detected in *T. parva* 7014 using qPCR. Generally, RNA-seq is the most sensitive technique for detecting differential gene expression at low expression levels (Rapaport *et al*., 2013); however, the sensitivity of the NGS platform seems to be highly dependent on the abundance of the specific transcript. Wang *et al*. (2009) reported a high cross-platform correlation, mainly in highly expressed genes, demonstrating that the above-median expressed genes have a good transferability between different platforms. Discrepancies can also be explained by analysis of alternatively spliced genes; in this case the primers only hybridize on one spliced variant and its abundance could therefore be different to what has been observed in RNA-seq. Nevertheless, the corroboration of NGS data and qPCR data for the three genes up-regulated in *T. parva* 7014 relative to *T. parva* Muguga and three genes down-regulated in *T. parva* 7014 relative to *T. parva* Muguga, may indicate a possible role of these genes in the different disease syndromes caused by these two *T. parva* isolates in cattle.

Most of the DEGs selected for validation in this study are involved in immunity, host invasion and transformation of host cells. This suggests that the two *T. parva* isolates will engender different immune responses depending on whether the genes involved are up or down regulated in each of the isolates. Among the differentially expressed genes selected for validation were genes coding for CD8 T-cell antigens. These play a vital role in host cell invasion and pathogenesis (Pelle *et al.,* 2011). Two genes encoding CD8 T-cell antigens selected for validation were up-regulated in *T. parva* Muguga relative to *T. parva* 7014. The DEGs selected for validation also included genes that code for antigenic proteins p150 microsphere and apical membrane antigen-1. A gene encoding TashHN was also among the genes selected for validation. TashHN has been associated with severe attenuation of the potential to differentiate in *T. annulata*-infected cell lines (Shiels *et al*., 2006), suggesting that regulation of this gene may play a role in the resultant virulence differences caused by the parasite. The other genes selected for validation included genes encoding transcription factors with an ap2 domain which are vital in gene regulation for survival and successful propagation of the parasite.

4.1 CONCLUSION

Transcriptome expression results are generally fair estimates for genes which are abundantly expressed in a particular tissue. In our study we have observed that for rare transcripts the expression estimates given by RNA-seq did not correlate with qPCR. Second, designing of primers is important for expression validation. If the primers are designed from junctions which participate in alternative splicing then there would not be good correlation. One may have to ensure that primer sequences are designed from the reads only and not from the reference annotation.

It is evident from the findings of this study that there are differences in the expression levels of some genes between a cattle-derived *T. parva* isolate (Muguga) and a buffalo-derived *T. parva* isolate (7014). It is proposed that the differential expression of some genes between the two *T. parva* isolates is the reason why the two isolates manifest different disease syndromes in cattle. It should be noted, however, that the cDNA in this study was isolated from lymphoblastoid cell cultures, and it remains to be seen whether the two isolates show the same expression profiles *in vivo*. This study highlighted that most of the differentially expressed genes between *T. parva* Muguga and *T. parva* 7014 are genes that code for hypothetical proteins. It is therefore essential that further work is carried out to determine the functional annotation of these genes.

Chapter 5 : REFERENCES

Allison, D.B., Cui, X. and Sabripour, M. (2006). Microarray data analysis: from disarray to consolidation and consensus. Nature Review Genetics, **406**, 55-65

Allsopp, B.A., Baylis, H.A., Allsopp, M.T.E.P., Cavalier-Smith, T., Bishop, R.P., Carrington, D.M., Sohanpal, B. And Spooner, P. (1993). Discrimination of six species of *Theileria* using oligonucleotide probes which detect small subunit ribosomal RNA sequence. Parasitology, **107**, 157-165.

Arthur, D.R. (1961). Diagnosis of *Rhipicephalus appendiculatus.* Ticks and disease. New York: Pergamon Press, p. 70-73.

Barnett, S.F. and Brockleysby, D. W. (1966). The susceptibility of the African buffalo (*Syncerus caffer*) to infection with *Theileria parva* (Theiler 1904). British Veterinary Journal, **122**, 379-86.

Bishop, R., Shah, T., Pelle, R., Hoyle, D., Pearson, T., Haines, L., Brass, A., Hulme, H., Graham, S., Taracha, E.L.N., Kanga, S., Lu, C., Hass, B., Wortman, J., White, O., Gardner, M.J., Nene, V. And de Villiers, E.P. (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 Acid Research, **33**, 5503-5511.

Bishop, R., Musoke, A., Morzaria, S., Gardner, M. and Nene, V. (2004). *Theileria*: intracellular protozoan parasites of wild and domestic ruminants transmitted by ixodid ticks. Parasitology, **129**, S271–S283.

Bishop, R., Geysen, D., Spooner, P., Skilton, R., Nene, V., Dolan, T. and Morzoria, S. (2001). Molecular and Immunological characterization of *Theileria parva* stocks which are components of 'Muguga cocktail' used for vaccination against East Coast fever in cattle. Veterinary Parasitology**, 94**, 227-37.

Bishop, R., Morzaria, S. and Gobright, E. (1998). Linkage of two distinct AT-rich minisatellites at multiple loci in the genome of *Theileria parva*. Gene, **216**, 245-54.

Bishop, R.P., Spooner P.R., Kanhai, G.K., Kiarie, J., Latif, A.A., Hove, T., Masaka, S. and Dolan, T.T. (1994). Molecular characterization of *Theileria* parasites: application to the epidemiology of theileriosis in Zimbabwe. Parasitology, **109**, 573-581.

Bullard, J.H., Purdom, E., Hansen, K.D. and Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics, **11**,94.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J. and Wittwer, C.T. (2009). The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. Clinical Chemistry **55(4),** 611–622.

Bustin, S.A. and Nolan, T. (2004). Pitfalls of quantitative reverse transcription polymerase chain reaction. Journal of Biomolecular Techniques, **15**, 155-166.

Carstens, B., Lemmon, A.R. and emmon E.M. (2012). The Promises and pitfalls of Next Generation Squencing Data in Phylogeography. Systems Biology, **50**, 713-715.

Cristino, A.S., Tanaka, E.D., Rubio, M., Piulachs, M.D. and Belles, X. (2011) Deep Sequencing of Organ- and Stage-Specific microRNAs in the Evolutionarily Basal Insect *Blattella germanica* (L.) (Dictyoptera, Blattellidae). PLoS ONE, **6(4):** e19350.

Collins, N.E. and Allsopp, B.A. (1999). *Theileria parva* ribosomal internal transcribed spacer sequences exhibit extensive polymorphism and mosaic evolution: application to the characterization of parasites from cattle and buffalo. Parasitology, **118**, 541-551.

Conesa, A., Götz, S., Garcia-Gomez, J. M., Terol, J., Talon, M. and Robles M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics, **21**, 3674-3676.

Conrad, P.A., Iams, K., Brown, W.c., Sohanpal, B. And Ole-moiyoi, O. (1987). DNA probes detect genetic diversity in *Theileria parva* stocks. Molecular and Biochemical Parasitology, **25**, 213-26.

De Deken, R., Martin, V., Saido, A., Madder, M., Brandt, J., Geysen, D. (2007). An outbreak of east coast fever on the Comoros: a consequence of the import of immunised cattle from Tanzania? Veterinary Parasitology, **143**, 245-253.

Derveaux, S., vandesomple, J. and Hellenemans, J. (2010) How to do successful gene expression analysis using real-time PCR. Methods, **50**, 227-230.

Dolan, T.T. (1999). Dogmas and misunderstanding in East Coast fever. Tropical Medicine and International Health, **4**, A3-11.

Dolan, T.T. (1986). Dogmas and misunderstanding in East Coats Fever. Tropical Medicine and International Health, **4**, A3-11.

Elsheikha, H.M. and Khan, N.A. (2011). Essentials of Veterinary Parasitology. Parasites and Vectors, **4,** 67.

Ersfeld, H. (2003). Genomes and Genome Projects of Protozoan Parasites. Curr. Issues Molecular Biololgy, **5**, 61-74.

Epstein, H. (1971). On the classification of cattle. The origin of domestic animals of Africa. African Publishing Corporation, New York, p 187.

Gachohi, J., Skilton, R., Hansen, F., Ngumi, P. and Kitala, P. (2012). Epidemiology of East Coast fever (*Theileria parva* infection) in Kenya: past, present and the future. Parasites and Vectors, **5**, 1-13.

Gardner, M.J., Bishop, R., Shah, T., de Villiers, E.P., Carlton, J.M., Hall, N., Ren, Q., Paulsen, I.T., Pain, A., Berriman, M., Wilson, R.J.M., Sato, S., Ralph, S.A., Mann, D.J., Xiong, Z., Shallom, S.J., Weidman, J., Jiang, L., Lynn, J., Weaver, B., Shoaibi, A., Domingo, A.R., Wasawo, D., Crabtree, J., Wortman, J.R., Haas, B., Angiuoli, S.V., Creasy, T.H., Lu, C., Suh, B., Silva, J.C., Utterback, T.R., Feldblyum, T.V., Pertea, M., Allen, J., Nierman, W.C., Taracha, E.L.N., Salzberg, S.L., White, O.R., Fitzhugh, H.A., Morzaria, S., Venter, J.C., Fraser, C.M. and Nene, V. (2005). Genome Sequence of *Theileria parva*, a Bovine Pathogen That Transforms Lymphocytes. Science, **309**, 1-5

Geysen, D., Bishop, R., Skilton, R., Dolan, T.T and Morzaria, S. (1999). Molecular epidemiology of *Theileria parva* in the field. Tropical Medicine and International Health, **4**, A21-27.

Geysen, D., Bazarusongs, T., Brand, T and Dolan, T.T. (2004). An unusual mosaic structure of the PIM gene of *Theileria parva* and its relationship to allelic diversity. Molecular and Biochemical Parasitology, **133**, 163-74.

Grootenhuis, J.G., Leitch, B.L., Stagg, D.A., Dolan, T.T. and Young, A.S. (1987). Experimental induction of *Theileria parva lawrencei* carrier state in an African buffalo (*Syncerus caffer*). Parasitology, **94**, 425-431.

Gomez, C., Ramirez, M.E., Calixto-Galvez, M., Medel, O. An Rodr´ıguez, M.A. (2010). Regulation of Gene Expression in Protozoa Parasites. Journal of Biomedicine and Biotechnology,10, 1-24.

Hayashida, K., Abe, T., Weir, W., Nakao, R., Ito, K., Kajino, K., Suzuki, Y., Jongejan, F., Geysen, D. and Sugimoto, C. (2013). Whole-Genome Sequencing of *Theileria parva* Strains Provides Insight into Parasite Migration and Diversification in the African Continent. DNA research, **1093**, 1-12.

Henson, S., Bishop, R.P., Morzaria, S., Spooner, P.R., Pelle, R., Poveda, L., Ebeling, M., Küng, E., Certa, U., Daubenberger, C.A. and Qi, W. (2012). High-resolution genotyping and mapping of recombination and gene conversion in the protozoan Theileria parva using whole genome sequencing. BMC Genomics, **13**, 1-15.

Iams, K.P., Hall, R., Webster, P. and Musoke, A.J. (1990). Identification of ʎgt11 clones encoding the major antigenic determinants expressed by *Theileria parva* sporozoites. Infection and immunity, **58**, 1828-34.

Irvin, A.D. and Mwamachi, D.M. (1983). Clinical and diagnostic features of East Coast fever (*Theileria parva* infection of cattle). Veterinary Record, **113**,192–198.

Irvin, A.D., Dobbelaere, D.A.E., Mwamachi, D.M., Minami, T., Spooner, P.R. and Ocama, J.G.R. (1983). Immunization against East Coast fever: correlation between monoclonal antibody profiles of *Theileria parva* stocks and cross-immunity *in vivo. Research in* Veterinary Science, **35**,341-346.

Laubscher, L. and Hoffman, L. (2012). An Overview of Disease-Free Buffalo Breeding Projects with Reference to the Different systems used in South Africa. Sustainability, **4**, 3124-3140.

Lawrence, J.A., Perry, B.D. and Williamson, S.M. (2004). East Coast fever. In Coetzer J A W, Tustin R C (eds) *Infectious diseases of livestock* (2nd edn). Oxford University Press Southern Africa, Cape Town: 448–463.

Lawrence, J.A., De vos, A.J and Irvin, A.D. (1994). East Coast Fever: Infectious disease of Livestock, **1**. J.A.W Coetzer, G.R Thompson and R.C Justin (Eds), Oxford University press, London, pp 309-25.

Lawrence, J.A., Perry, B.D. and Williamson, S.M. (1992). East coast fever. In Infectious Diseases of Livestock. 2nd edition, Coetzer JAW, Tustin RC (Eds), Oxford University Press; **1**,448–467.

Lawrence, J.A. (1992). History of bovine theileriosis in southern Africa. In Norval RAI, Perry BD,YoungAS (eds) The epidemiology of theileriosis in Africa. Academic Press, London: 1–31.

Lawrence, J.A. (1979). The differential diagnosis of the bovine theilerias of Southern Africa. Journal of the South African Veterinary Association, **50**, 311-313.

Lindberg, J. and Lundeberg, J. (2010). The plasticity of mammalian transcriptome. Genomics, **95**, 1-6.

Lizundia, R., Werling, Dirk., Langskey, G. And Ralph, S.A. (2009). *Theileria* Apicoplast as a Target for Chemotherapy. American Society for Microbiology, **53**, 1213-1217.

McCarthy, D.J., Chen, Y. and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res, **40**,4288-97.

Matson, B.A. (1967) Theileriosis in Rhodesia: 1. A study of diagnostic specimens of two seasons. Journal of South African Medical Association, **38**, 93-101.

Mbizeni, S., Potgiter, F.T., Troskie, C., Mans, B.J., Penzhorn, B.I. and Latif, A.A. (2013) Field and laboratory studies on Corridor disease (*Theileria parva* infection) in cattle

population at the livestock/game interface of uPongolo-Mkuze area, South Africa. Ticks and Tick borne Diseases, 3, 227-34.

Mckeever, D.J., Taracha, E.L.N., Innes, E.A., Machugh, N.D., Awino, E., Goddeeris, B.M. and Morrison, W.I. (1994). Adoptive transfer of immunity to Theileria parva in the CD8 fraction of responding efferent lymph. Proceedings of national Academy of science, **91**.

Mclntyre, M. (2011) RNA-seq: technical variability and sampling. BMC Genomics, **12**, 293.

Melhorn, H. and Schein, E. (1984). The piroplasms: Life cycle and sexual stages. Advances in Parasitology, **23**, 37-103.

Minami, T., P.R., Irvin, A.D., Ocama, J.G.R., Dobbelaere, D.A.E. and Fujinaga, T. (1983) Characterization of stocks of *Theileria parva* by monoclonal antibody profiles. *Research in Veterinary Science* **35**,334-340.

Mooney, M. (2013) Comparative RNA-seq and microassay analysis of gene expression changes in B-cell lymphomas of *Canis familiaris* . PLOS ONE, **8**, e61088.

Morrison, W.I. and Mckeever, D.J. (1995). Theileriosis: Progress towards vaccine Development through understanding immune responses to the parasite. Veterinary Parasitology, 57, 177-187.

Morzaria, S.P. (1988). Characterization of *T. parva* stocks using contour clamped homogeneous electric fields and field inversion gel electrophoresis. In: ILRAD 1987 Annual Scientific Report. Nairobi. International Laboratory for Research on Animal Diseases, pp 8-9.

Mukhebi, A.W., Perry, B.D. and Kruska, R. (1992). Estimated economics of theileriosis control in Africa. Preventive Veterinary Medicine, **12**, 73-85.

Muraguri, G.R., Kiara, H.K. and McHardy, N. (1999). Treatment of East Coast fever: a comparison of parvaquone and buparvaquone. Veterinary Parasitology, **87**, 25–37.

Musisi, F.L. (1990). Methods currently used for control of East Coast fever: Their validity and proposals for future control strategies. Parassitologia, **30**, 73-86.

Musoke, R.A., Tweyongyere., R., Bizimenyera, E., Waiswa, C., Mugisha, A., Biryomumaisho, S. and McHardy, N. (2004). Treatment of East Coast Fever of Cattle with

a Combination of Parvaquone and Frusemide. Tropical Animal Health and Production, **36**, 233-245.

Musoke, A., Morzaria, S., Nkonge, C., Jones, E. and Nene, V. (1991). A recombinant sporozoite surface antigen of Theileria parva induces

protection in cattle. Proc. Nati. Acad. Sci. **89**, 514-518.

Mutugi, J.J., Young, A.S., Martim, A.C., Ndungu, S.G., Stagg, D.A., Grootenhuis, J.G. and Leitch, B.L. (1987). Immunization of Cattle against Theileriosis using varying doses of *Theileria parva lawrencei* and *Theileria parva parva* sporozoites and oxytetracycline treatments. International Journal for Parasitology, **18**, 453-461.

Nene, V., Musoke, A., Gobright, E. and Morzaria S. (1996). Conservation of the Sporozoite p67 Vaccine Antigen in Cattle-derived *Theileria parva* stocks with different cross-immunity profiles. Infection and immunity, **64**, 2056-2061.

Neitz, W.O. (1957). Theileriosis, gonderiosis and cytauxzoonoses: a review. Onderstepoort Journal of Veterinary Research, **27**, 275-430.

Norval, R.A.L., Perry, B.D. and Young, A.S. (1992). The Epidemiology of Theileriosis in Africa. Academic Press, London.

Olds, C., Mwauraa, S., Crowderd, D., Odongoa, D., van Oerse, M., Owend, J., Bishopa, R. and Daubenberger, C. (2012). Immunization of cattle with Ra86 impedes *Rhipicephalus appendiculatus* nymphal-to-adult molting. Ticks and Tick-borne Diseases, **3**, 170-178.

Oura, C.A., Odongo, D.O., Lubega, G.W., Spooner, P.R., Tait, A. and Bishop, R.P. (2003). A panel of microsatellite and minisatellite markers for the characterization of field isolates of *Theileria parva*. International Journal of Parasitology, **33**, 1641-1653.

Oura, C.A., Bishop, R., Wampande, E.M., Lubega, G.W. and Tait, A. (2004). The persistence of component *Theileria parva* stocks in cattle immunized with the 'Muguga cocktail' live vaccine against East Coast Fever in Uganda. Parasitology, **129**, 27-42.

Oura, C.A.L., Tait, A., Asiimwe, B., Lubega, G.W. and Weir, W. (2011). Haemoparasite prevalence and *Theileria parva* strain diversity in Cape buffalo (*Syncerus caffer*) in Uganda. Veterinary Parasitology, **175**, 212-219.

Pelle, R., Graham, S.P., Njahira, M.N., Osaso, J., Saya, R.M., Odongo, D.O., Toye, P.G., Spooner, P.R., Musoke, A.J., Mwangi1, D.M., Taracha, L.M., Morrison, W.I., Weir, W., Silva, J.C. and Bishop, R.P. (2011). Two Theileria parva CD8 T Cell Antigen Genes Are More Variable in Buffalo than Cattle Parasites, but Differ in Pattern of Sequence Diversity. Gordon Langsley, Institut national de la sante´ et de la recherche me´dicale - Institut Cochin, France, **6**, 1-14.

Perry, B.D. and Young, A.S. (1995). The past and future roles of epidemiology and economics in the control of tick-borne diseases of livestock in Africa: the case of theileriosis.Veterinary Medicine, **25**, 107-120.

Pfaffi, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acid Research, **29**,1-6.

Pfaffi, M.W., Tichopad, A., Prgomet, C. And Neuvians, T.P. (2004) Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: Bestkeeper-Excel- based tool using pair- wise correlations. Biotechnology, **6**, 509-15.

Potgieter, F. T., Stoltsz, W. H., Blouin, E. F. and Roos, J. A. (1988). Corridor disease in South Africa: a review of the current status. Journal of the South African Veterinary Association, **59**, 155–160.

Radley, D.E. (1981). Infection and treatment immunization against theileriosis. In: Advances in the Control of Theileriosis. Irvin A.D., Cunningham M.P. and Young A.S., eds. Martinus Nijhoff Publishers, The Netherlands, 227-237.

Radley, D.E., Brown, C.G.D., Burridge, M.J., Cunningham, M.P., Kirimi, I.M., Purnell, R.E. and Young, A.S. (1975). East Coast Fever 1. Chemoprophylactic immunization of cattle against *Theileria parva* (Muguga) and five theilerial strains. Veterinary Parasitology, **1**, 35-41.

Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., Mason, C.E., Socci, N.D. and Betel, D. (2013). Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. Genome Biology, **14**, 1-13.

Robles, J.A., Qureshi, S.E., Stephen, S.J., Wilson S.R., Burden C.J. and Taylor, J.M. (2012): Efficient experimental design and analysis strategies for the detection of differential expression using RNA-Sequencing. BMC Genomics, **13**:484.

Shaw, M.K. (2003). Cell invasion by Theileria sporozoites. Trends in parasitology, **19**, 2-6.

Shiels, B., Langsley, G., Weir, W., Pain, A., Mckellar, S. And Dobbelaere, D. (2006). Alteration of host cell phenotype by *Theileria annulata* and Theileria parva: mining for manipulators in the parasite genomes. International Journal for Parasitology, **36,** 9-21.

Sibeko K.P., Geysen, D., Oosthuizen, M.C.,Matthee, C.A., Troskie, M., Potgieter, F.T., Coetzer, J.A.W. and Collins, N.E. (2010). Four p67 alleles identified in South African *Theileria parva* field samples. Veterinary Parasitology, **167**, 244–254.

Skilton, R.A., Bishop, R.P., Wells, C.W., Spooner, P.R., Gobright, E., Nkonge, C., Musoke, A.J., Macklin, M. and Iams, K.P. (1998). Cloning and characterization of a 150 kDa microsphere antigen of *Theileria parva* that is immunologically cross- reactive with the polymorphic immunodominant molecule (PIM). Parasitology, **117**, 321-30.

Sohanpal, B.K., Morzaria, S.P., Gobright, E.I. and Bishop, R.P. (1995). Characterisation of the telomeres at opposite ends of a 3Mb *Theileria parva* chromosome. Nucleic acids Reserved, **11**, 1942 7.

Stagg, D.A., Dolan, T.T., Leitch, B.L. and Young, A.S. (1981). The initial stages of infection of cattle cells with *Theileria parva* sporozoites *in vitro*. Parasitology, **83**, 191-97.

Theiler, A. (1904) Rhodesian fever. Transvaal Agricultural Journal, **2**, 421-438.

Thompson, B.E., Latif, A.A., Oosthuizen, M.C., Troskie, M. And Penzhorn, B.L. (2008). Occurrence of *Theileria parva* infection in cattle on a farm in the Ladysmith district, KwaZulu-Natal, South Africa. Journal of South African Veterinary Association, **79**, 31-35.

Toye, P.G., Goddeeris, B.M., Iams, K., Musoke, A.J. and Morrison, W.I. (1991). Characterization of a polymorphic immunodominant molecular in sporozoites and schizonts of Theileria parva. Parasite Immunol. **13**, 49e52.

Uilenberg, G. (1976). Tick-borne livestock diseases and their vectors. Epizootiology of tick-borne diseases. World animal review, **17**, 8-15.

Wang, Z., Gerstein, M. and Snyder, M. (2009). RNA-Seq: revolutionary too for transcriptomice. Nature Reviews Genetics, **10**, 57-63.

Young, A.S., Brown, C.G., Burridge, M.J., Cunningham, M.P., Kirimi, I.M. and Irvin, A.D. (1973). Observations on the cross-immunity between *Theileria lawrencei* (Serengeti) and *Theileria parva* (Muguga) in cattle. International Journal of Parasitology, **3**, 723-728.

Young, A. S., Leitch, B. L., Newson, R. M. and Cunningham, M. P. (1986). Maintenance of *Theileria parva* infection in an endemic area of Kenya. Parasitology **93**, 9–16.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Animal Ethics Committee

PROJECT TITLE	**DEVELOPMENT AND EVALUATION OF A REAL-TIME PCR ASSAY FOR VALIDATION OF *THEILERIA PARVA* DIFFERENTIALLY EXPRESSED TRANSCRIPTS**
PROJECT NUMBER	**V040-14**
RESEARCHER/PRINCIPAL INVESTIGATOR	**T Tsotetsi**

STUDENT NUMBER (where applicable)	**UP_12290859**
DISSERTATION/THESIS SUBMITTED FOR	**MSc**

ANIMAL SPECIES	**n/a**	
NUMBER OF ANIMALS	**n/a**	
Approval period to use animals for research/testing purposes		**1 April 2014 - 1 April 2015**
SUPERVISOR	**Dr. K Sibeko-Matjila**	

KINDLY NOTE:

Should there be a change in the species or number of animal/s required, or the experimental procedure/s - please submit an amendment form to the UP Animal Ethics Committee for approval before commencing with the experiment

APPROVED	Date	April 201
CHAIRMAN: UP Animal Ethics Committee	Signature	

S***4285-15***