Transmembrane proteins – mining the cattle tick transcriptome

Sabine A. Richards\textsuperscript{a}, Christian Stutzer\textsuperscript{a}, Anna-Mari Bosman\textsuperscript{b}, Christine Maritz-Olivier\textsuperscript{a}\textsuperscript{*}

\textsuperscript{a} Department of Genetics, Faculty of Natural and Agricultural Sciences, University of Pretoria, South Africa

\textsuperscript{b} Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, South Africa

* Corresponding author:

Christine Maritz-Olivier, Tel: +27 (0)12 420 3945, Fax: +27 (0)12 362 5327, E-mail address: christine.maritz@up.ac.za.
Abstract

Managing the spread and load of pathogen-transmitting ticks is an important task worldwide. The cattle tick, *Rhipicephalus microplus*, not only impacts the economy through losses in dairy and meat production, but also raises concerns for human health in regards to the potential of certain transmitted pathogens becoming zoonotic. However, novel strategies to control *R. microplus* are hindered by lack of understanding tick biology and the discovery of suitable vaccine or acaricide targets. The importance of transmembrane proteins as vaccine targets are well known, as is the case in tick vaccines with Bm86 as antigen. In this study, we describe the localization and functional annotation of 878 putative transmembrane proteins. Thirty proteins could be confirmed in the *R. microplus* gut using LC-MS/MS analysis and their roles in tick biology are discussed. To the best of our knowledge, 19 targets have not been reported before in any proteomics study in various tick species and the possibility of using the identified proteins as targets for tick control are discussed. Although tissue expression of identified putative proteins through expansive proteomics is necessary, this study demonstrates the possibility of using bioinformatics for the identification of targets for further evaluation in tick control strategies.
Introduction

The cattle tick, *Rhipicephalus microplus*, is an obligate hematophagous ectoparasite of cattle occurring in the tropical and subtropical regions of the world and is a competent vector of disease causing pathogens including *Babesia bovis*, *Babesia bigemina* and *Anaplasma marginale* (de Castro 1997, de Vos et al. 2001). It is also this vector competence of *R. microplus* for various bacterial and protozoan pathogens that emphasizes this species' potential role in zoonosis, although parasitism on human hosts is thought to be a very rare occurrence (Andreotti et al. 2011). Globally, ticks are considered to be second to mosquitoes with regards to the number and incidences of pathogens transmitted to humans and they surpass all arthropods with regards to pathogen transmission to wild and domestic animals (Needham 1985). *Rhipicephalus microplus* occurs together with *R. decoloratus* in many regions of southern Africa resulting in the displacement of the latter less pathogenic species and increased incidence of *Babesia bovis* infection (Tønnesen et al. 2004).

Conventional control of *R. microplus* is based mainly on the use of chemical acaricides (de Castro 1997). However, the frequent use of these chemicals results in the spread of resistance with reduced or no efficacy being reported against major classes of available acaricides in several developing countries including Latin America, India and South Africa (Castro-Janer et al. 2010). The steady increase of resistance to current chemical control stresses the need for identification of new or altered acaricides in combination with other improved control strategies such as immunological control via vaccination (Willadsen 2006, de la Fuente et al. 2007).

Plasma membrane-associated proteins have long been of interest in vaccine design since their antigenic regions are easily accessible to antibodies produced by an immunized host (Rappuoli and Bagnoli 2011). In this regard, more than 50% of all drug targets under development are
membrane proteins, highlighting their important pharmacological and biological roles (Drews 2000, Terstappen and Reggiani 2001). These proteins account for some 25-30% of all open reading frames identified in sequenced genomes of both prokaryotic and eukaryotic organisms (Martin-Galiano and Frishman 2006). Furthermore, membrane-associated proteins are involved in a vast array of pivotal cellular functions including cell-cell interaction, transport, metabolism, regulation, signal transduction and recognition (Hubert et al. 2010). Several vaccines and vaccine candidates against prokaryotes target membrane proteins with extracellular domains, such as filamentous hemagglutinin and pertactin that have been used successfully in vaccines against Bordetella pertussis (Poolman 1999, Sheu et al. 2001). Efforts are also being made for the use of cell surface proteins in vaccine development for many other bacteria that include Neisseria gonorrhoeae, N. meningitides, Moraxella catarrhalis and Klebsiella pneumonia (Poolman 1999, Kurupati et al. 2011).

Immunological control of R. microplus was first suggested by Allen and Humphreys (1979) and the feasibility of targeting membrane proteins was demonstrated by the 87% protection obtained in R. microplus challenged cattle following immunization with membrane fractions from partially fed adult tick midguts (Opdebeeck et al. 1988). Subsequent research by Willadsen and colleagues led to the development of two R. microplus vaccines (TickGARD™ and Gavac™), using the 89 kDa GPI-linked midgut glycoprotein (named Bm86) of unknown function as antigen (Rand et al. 1989, Willadsen et al. 1989). Although the Bm86-based vaccines were inexpensive and nontoxic relative to acaricides, varying efficacy was reported ranging from 0% in Argentina to 91% in Cuba (de la Fuente et al. 2000). As a result of commercialization considerations, TickGARD™ (Hoechst Animal Health; Australia) and TickGARDPLUS™ (Intervet Australia (Pty) Ltd., Australia) are no longer available (Guerrero et al. 2012). In contrast, Gavac™ and GAVAC Plus™ are still being produced and Gavac™ was recently used in a country-wide two year
vaccination program (including 18 states) in the Republic of Venezuela, South America (Suarez et al., unpublished). In this program over 1.9 million head of cattle were immunized with Gavac™ that resulted in an 83.7% reduction in acaricide use, as well as an estimated 81.5% cost saving on traditional chemical control.

To date, numerous tick vaccine candidates have been identified, but targeting most of these proteins have not proven effective enough in pilot trials to result in a commercial vaccine (Table 1). Moreover, transmembrane proteins (as a protein class) have not been greatly exploited in vaccination studies, though their potential has recently been demonstrated in pilot cattle trials (i.e. Aquaporin) (Table 1) (Guerrero et al. 2014). Such proteins have successfully been tested against several other organisms and viruses including feline leukemia virus, Edwardsiella tarda, Pseudomonas aeruginosa and several Vibrio species (Baumann et al. 2004, Langhammer et al. 2011, Lun et al. 2014, Maiti et al. 2011). Therefore, the identification of novel protective antigens still remains the biggest obstacle in tick vaccine development.
Table 1: Summary of *Rhipicephalus* antigens evaluated in vaccination trials. Indicated are various antigens with their protein identity tested for various *Rhipicephalus* tick species in vaccination trials. Topology was derived from published and/or UniProt data available for the original full-length mature protein. Also indicated are the source tissue(s), vaccine antigen type (native/recombinant/synthetic), experimental animal and vaccine efficacy (N/i: not indicated). Vaccine efficacies indicated in percentage were determined, otherwise individual mortality, reduction in infestation, reduction in eggs and egg viability are indicated. Denoted are no significant effect (NSE) or the efficacy achieved (%).

<table>
<thead>
<tr>
<th>Protein identity</th>
<th>Antigen name</th>
<th>Species</th>
<th>Source</th>
<th>Protein type(s)</th>
<th>Experimental host(s)</th>
<th>Vaccine efficacy</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-nucleotidase</td>
<td>5'-nucleotidase (4F8)</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Malpighian tubules</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>NSE</td>
<td>Hope et al., 2010</td>
</tr>
<tr>
<td>Akirin</td>
<td>rAKR</td>
<td>Various</td>
<td>Whole mosquito</td>
<td>Recombinant</td>
<td>Red deer</td>
<td>25-33% reduction in tick infestation</td>
<td>Carréòn et al. 2012</td>
</tr>
<tr>
<td>Angiotensin-converting enzyme</td>
<td>Bm91</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Salivary glands</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>6 and 8% reproductive efficiency and egg viability</td>
<td>Lambertz et al., 2012</td>
</tr>
<tr>
<td>Angiotensin-converting enzyme</td>
<td>Bm91</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Salivary glands</td>
<td>Native</td>
<td>Cattle</td>
<td>~37% reduction in eggs</td>
<td>Riding et al., 1994</td>
</tr>
<tr>
<td>Aquaporin</td>
<td>Aquaporin</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Midgut</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>75% and 68%</td>
<td>Guerrero et al. 2014</td>
</tr>
<tr>
<td>Aspartic proteinase</td>
<td>BYC</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Eggs</td>
<td>Native</td>
<td>Cattle</td>
<td>14% and 36%</td>
<td>da Silva Vaz et al., 1998; Seixas et al., 2012</td>
</tr>
<tr>
<td>Aspartic proteinase</td>
<td>BYC</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Eggs</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>25%</td>
<td>Leal et al., 2006; Seixas et al., 2012</td>
</tr>
<tr>
<td>Cement protein</td>
<td>64TRP</td>
<td><em>Rhipicephalus appendiculatus</em></td>
<td>Salivary glands</td>
<td>Recombinant</td>
<td>Guinea pigs</td>
<td>~62% mortality</td>
<td>Trimnell et al., 2002</td>
</tr>
<tr>
<td>Cement protein</td>
<td>64TRP</td>
<td><em>Rhipicephalus sanguineus</em></td>
<td>Salivary glands</td>
<td>Recombinant</td>
<td>Guinea pigs</td>
<td>~47% mortality</td>
<td>Trimnell et al., 2005</td>
</tr>
<tr>
<td>Cement protein</td>
<td>TrP64</td>
<td><em>Rhipicephalus appendiculatus</em></td>
<td>N/i</td>
<td>Recombinant</td>
<td>Rabbits</td>
<td>43.9%</td>
<td>Saimo et al. 2011</td>
</tr>
<tr>
<td>Component of vitellin</td>
<td>GP80</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Various</td>
<td>Native</td>
<td>Sheep</td>
<td>68%</td>
<td>Tellam et al., 2002</td>
</tr>
<tr>
<td>Component of vitellogenin, vitellin</td>
<td>VIT87</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Eggs</td>
<td>Native</td>
<td>Sheep</td>
<td>68%</td>
<td>Tellam et al., 2002</td>
</tr>
<tr>
<td>Elongation factor 1 alpha</td>
<td>Ef1a</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Various</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>31%</td>
<td>Almazán et al., 2012</td>
</tr>
<tr>
<td>Protein identity</td>
<td>Antigen name</td>
<td>Species</td>
<td>Source</td>
<td>Protein type(s)</td>
<td>Experimental host(s)</td>
<td>Vaccine efficacy</td>
<td>Reference</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>-----------------</td>
<td>-----------------------------</td>
<td>-----------------------------</td>
<td>-----------------</td>
<td>----------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Extracellular matrix protein, Glycine-rich</td>
<td>RH50</td>
<td>Rhipicephalus haemaphysalooides</td>
<td>Salivary glands</td>
<td>Recombinant</td>
<td>Rabbits</td>
<td>30.5% mortality</td>
<td>Zhou et al., 2006</td>
</tr>
<tr>
<td>protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ferritin 2</td>
<td>RaFER2</td>
<td>Rhipicephalus annulatus</td>
<td>Midgut</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>72%</td>
<td>Hajdusek et al., 2010</td>
</tr>
<tr>
<td>Ferritin 2</td>
<td>RmFER2</td>
<td>Rhipicephalus microplus</td>
<td>Midgut</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>64%</td>
<td>Hajdusek et al., 2010</td>
</tr>
<tr>
<td>Glutathione S-transferase</td>
<td>GST-HI</td>
<td>Rhipicephalus microplus</td>
<td>Various</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>57%</td>
<td>Parizi et al., 2011</td>
</tr>
<tr>
<td>Glutathione S-transferase, Vitellin-degrading</td>
<td>GST-HI, VTDCE</td>
<td>Rhipicephalus microplus</td>
<td>Various</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>35.3 to 61.6% protection against infestation</td>
<td>Parizi et al., 2012</td>
</tr>
<tr>
<td>enzyme and Aspartic proteinase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mating factor, voraxina</td>
<td>Voraxinc</td>
<td>Rhipicephalus appendiculatus</td>
<td>Testis</td>
<td>Recombinant</td>
<td>Rabbits</td>
<td>~50% reduction in eggs</td>
<td>Yamada et al., 2009</td>
</tr>
<tr>
<td>Metalloprotease</td>
<td>BrRm-MP4</td>
<td>Rhipicephalus microplus</td>
<td>Larvae</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>60%</td>
<td>Ali et al. 2015</td>
</tr>
<tr>
<td>Mucin</td>
<td>BMA7</td>
<td>Rhipicephalus microplus</td>
<td>Whole ticks, membrane fractions</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>~21% reduction in eggs</td>
<td>McKenna et al., 1998</td>
</tr>
<tr>
<td>Ribosomal protein P0</td>
<td>pP0-KHL</td>
<td>Rhipicephalus sanguineus</td>
<td>Various</td>
<td>Synthetic peptide</td>
<td>Rabbits</td>
<td>90.25%</td>
<td>Rodríguez-Mallon et al., 2012</td>
</tr>
<tr>
<td>Serpin-3 (RAS), Serpin-4 (RAS-4) and 36kDa</td>
<td>RAS-3, RAS-4 and</td>
<td>Rhipicephalus appendiculatus</td>
<td>Salivary glands</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>~27% mortality</td>
<td>Imamura et al., 2008</td>
</tr>
<tr>
<td>immuno-dominant protein (RIM36)</td>
<td>RIM36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain variant of Bm86</td>
<td>Bm95</td>
<td>Rhipicephalus microplus</td>
<td>Midgut</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>58-89%</td>
<td>Sugumaran et al., 2011</td>
</tr>
<tr>
<td>Strain variant of Bm86</td>
<td>Bm95</td>
<td>Rhipicephalus haemaphysalooides</td>
<td>Midgut</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>78.9-84.6%</td>
<td>Garcia-Garcia et al., 2000</td>
</tr>
<tr>
<td>Subolesin/Akirin</td>
<td>Subolesin (4D8)</td>
<td>Various</td>
<td>N/i</td>
<td>Recombinant</td>
<td>Cattle, Sheep</td>
<td>0% (cattle), 60-63% (sheep) reduction of tick infestation</td>
<td>Torina et al. 2014</td>
</tr>
<tr>
<td>Subolesin</td>
<td>Subolesin (4D8)</td>
<td>Rhipicephalus microplus</td>
<td>N/i</td>
<td>Synthetic</td>
<td>White-tailed deer</td>
<td>83%</td>
<td>Carreón et al. 2012</td>
</tr>
<tr>
<td>Subolesin</td>
<td>rBmSu</td>
<td>Rhipicephalus microplus</td>
<td>Whole ticks</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>44% and 37.2% for subsequent challenges</td>
<td>Shakya et al. 2014</td>
</tr>
<tr>
<td>Subolesin/Akirin</td>
<td>Subolesin (4D8)</td>
<td>Rhipicephalus annulatus</td>
<td>Various</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>60%</td>
<td>Almazán et al., 2010, Almazán et al., 2012</td>
</tr>
<tr>
<td>Subolesin/Akirin</td>
<td>Subolesin (4D8)</td>
<td>Rhipicephalus microplus</td>
<td>Various</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>51% and 81% depending on construct</td>
<td>Almazán et al., 2010, Almazán et al., 2012</td>
</tr>
<tr>
<td>Protein identity</td>
<td>Antigen name</td>
<td>Species</td>
<td>Source</td>
<td>Protein type(s)</td>
<td>Experimental host(s)</td>
<td>Vaccine efficacy</td>
<td>Reference</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------</td>
<td>------------------------</td>
<td>--------------</td>
<td>-----------------</td>
<td>----------------------</td>
<td>-------------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Trypsin inhibitor</td>
<td>BmLTI</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Various</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>32%</td>
<td>Andreotti et al., 2012</td>
</tr>
<tr>
<td>Trypsin inhibitor</td>
<td>BmTl</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Salivary glands</td>
<td>Native</td>
<td>Cattle</td>
<td>72.8%</td>
<td>Andreotti et al., 2002</td>
</tr>
<tr>
<td>Trypsin inhibitor</td>
<td>BmTl-A</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Salivary glands</td>
<td>Synthetic</td>
<td>Cattle</td>
<td>~18.4%</td>
<td>Andreotti et al., 2007</td>
</tr>
<tr>
<td>Ubiquitin</td>
<td>UBE</td>
<td><em>Rhipicephalus annulatus</em></td>
<td>Various</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>15% and 22% depending construct preparation</td>
<td>Almazán et al., 2010, Almazán et al., 2012</td>
</tr>
<tr>
<td>Ubiquitin</td>
<td>UBE</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Various</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>55%</td>
<td>Almazán et al., 2010</td>
</tr>
<tr>
<td>Unknown</td>
<td>ARS antigen 1</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Midgut</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>73-76%</td>
<td>Rachinsky et al., 2008; Guerrero et al., 2012</td>
</tr>
<tr>
<td>Unknown</td>
<td>ARS antigen 1 and 2</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Midgut</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>71%</td>
<td>Rachinsky et al., 2008; Guerrero et al., 2012</td>
</tr>
<tr>
<td>Unknown</td>
<td>ARS antigen 2</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Midgut</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>63%</td>
<td>Rachinsky et al., 2008; Guerrero et al., 2012</td>
</tr>
<tr>
<td>Unknown</td>
<td>Bm86</td>
<td><em>Rhipicephalus annulatus</em></td>
<td>Midgut</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>~100%</td>
<td>de Vos et al., 2001, Odongo et al., 2007</td>
</tr>
<tr>
<td>Unknown</td>
<td>Bm86</td>
<td><em>Rhipicephalus appendiculatus</em></td>
<td>Midgut</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>NSE</td>
<td>de Vos et al., 2001, Odongo et al., 2007</td>
</tr>
<tr>
<td>Unknown</td>
<td>Bm86</td>
<td><em>Rhipicephalus decoloratus</em></td>
<td>Midgut</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>70% reduction in eggs</td>
<td>de Vos et al., 2001, Odongo et al., 2007</td>
</tr>
<tr>
<td>Unknown</td>
<td>Bm86</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Midgut</td>
<td>Native</td>
<td>Cattle</td>
<td>92%</td>
<td>Willadsen et al., 1989</td>
</tr>
<tr>
<td>Unknown</td>
<td>Bm86</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Midgut</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>51-91%, depending on tick strain</td>
<td>Willadsen et al., 1995; Rodriguez et al., 1995; Garcia-Garcia et al., 1998; Jonsson et al., 2000; Garcia-Garcia et al., 2000; Canales et al., 2009, de la Fuente et al., 2000</td>
</tr>
<tr>
<td>Unknown</td>
<td>Bm86</td>
<td><em>Rhipicephalus sanguineus</em></td>
<td>Midgut</td>
<td>Recombinant</td>
<td>Dogs</td>
<td>~31% reduction in infestation</td>
<td>Perez-Perez et al., 2010</td>
</tr>
<tr>
<td>Unknown</td>
<td>Bm86</td>
<td><em>Rhipicephalus microplus</em></td>
<td>N/i</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>31%</td>
<td>Cunha et al. 2012</td>
</tr>
<tr>
<td>Unknown</td>
<td>Bm86</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Whole ticks</td>
<td>Recombinant</td>
<td>White-tailed deer</td>
<td>76%</td>
<td>Carreón et al. 2012</td>
</tr>
<tr>
<td>Protein identity</td>
<td>Antigen name</td>
<td>Species</td>
<td>Source</td>
<td>Protein type(s)</td>
<td>Experimental host(s)</td>
<td>Vaccine efficacy</td>
<td>Reference</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------</td>
<td>---------</td>
<td>--------</td>
<td>-----------------</td>
<td>---------------------</td>
<td>-----------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Unknown and Angiotensin-converting enzyme</td>
<td>Bm86 &amp; Bm91</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Midgut &amp; Salivary glands</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>~100%</td>
<td>Willadsen et al., 1996</td>
</tr>
<tr>
<td>Unknown and mucin</td>
<td>Bm86 &amp; BMA7</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Whole ticks, membrane fractions</td>
<td>Recombinant (Bm86) and Native (BMA7)</td>
<td>Cattle</td>
<td>~90% reduction in eggs</td>
<td>McKenna et al., 1998</td>
</tr>
<tr>
<td>Unknown and nucleotidase</td>
<td>Bm86 &amp; 4F8</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Midgut</td>
<td>Recombinant</td>
<td>Sheep</td>
<td>85%</td>
<td>Hope et al., 2010</td>
</tr>
<tr>
<td>Unknown, Bm86 peptide derivative</td>
<td>SBm19733</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Various</td>
<td>Synthetic peptide</td>
<td>Cattle</td>
<td>35.87%</td>
<td>Pataroyo et al., 2002</td>
</tr>
<tr>
<td>Unknown, Bm86 peptide derivative</td>
<td>SBm4912</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Various</td>
<td>Synthetic peptide</td>
<td>Cattle</td>
<td>72.4%</td>
<td>Pataroyo et al., 2002</td>
</tr>
<tr>
<td>Unknown, Bm86 peptide derivative</td>
<td>SBm7462</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Various</td>
<td>Synthetic peptide</td>
<td>Cattle</td>
<td>81.05%</td>
<td>Pataroyo et al., 2002</td>
</tr>
<tr>
<td>Unknown, homologue of Bm86</td>
<td>Haa86</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Midgut</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>~36.5%</td>
<td>Kumar et al., 2012</td>
</tr>
<tr>
<td>Unknown, homologue of Bm86</td>
<td>Ba86</td>
<td><em>Rhipicephalus annulatus</em></td>
<td>Midgut</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>83%</td>
<td>Canales et al., 2009</td>
</tr>
<tr>
<td>Unknown, homologue of Bm86</td>
<td>Ba86</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Midgut</td>
<td>Recombinant</td>
<td>Cattle</td>
<td>~71.5%</td>
<td>Canales et al., 2009</td>
</tr>
<tr>
<td>Unknown, homologue of Bm86</td>
<td>Ra86</td>
<td><em>Rhipicephalus appendiculatus</em></td>
<td>N/I</td>
<td>Recombinant</td>
<td>Rabbits</td>
<td>69.7%</td>
<td>Saimo et al. 2011</td>
</tr>
<tr>
<td>Vitellin-degrading enzyme</td>
<td>VTDCE</td>
<td><em>Rhipicephalus microplus</em></td>
<td>Eggs</td>
<td>Native</td>
<td>Cattle</td>
<td>21%</td>
<td>Seixas et al., 2008; Seixas et al., 2012</td>
</tr>
</tbody>
</table>
Prior to the increased availability of genomic and transcriptomic data from high-throughput analyses, membrane proteins were identified mainly through the use of costly, laborious and time consuming experiments involving the isolation of membrane fractions via differential subcellular fractionation and amino acid sequencing that at best provided partial sequence information (Carroll et al. 2007). More daunting is the classic determination of membrane protein topology, as high-resolution microscopy or crystallographic techniques are required that are hindered by transmembrane proteins being notoriously difficult to crystallize (von Heijne 2006). However, the expansion of nucleotide sequence databases in the post-genomic era has allowed for fast and inexpensive data mining (including functional annotation, localization and topology) to predict encoded proteins with desirable traits using in silico tools. In this regard, assembled genome sequence data is currently available mainly for *Ixodes scapularis* (Pagel Van Zee et al. 2007) in addition to preliminary sequences for *R. microplus* (Bellgard et al. 2012) allowing data mining for transcripts of interest. Also, the number of RNA-seq projects to mine transcriptomes and aid in genome assembly is rapidly expanding (i.e. *Amblyomma triste*, *A. parvum* and *A. cajennense*) (Garcia et al. 2014).

At this stage, the identification of transmembrane proteins using predictive tools is reliable due to validated signals and the physicochemical properties of membrane spanning regions, in comparison to prediction of some peripheral membrane proteins (Bhardwaj et al. 2006, Chen and Li 2013). This study will therefore focus on transmembrane proteins as a novel group for the identification of potential candidates for future evaluation in tick control. This gene expression and in silico analysis study, is combined with a pilot proteomics analysis of the membrane fraction of *R. microplus* midgut tissues and aims to: identify putative transmembrane proteins that are expressed in various *R. microplus* life stages and tissues, confirm expression of proteins using proteomics techniques in midgut tissues and describe the putative functional
role of these single- and multi-spanning proteins in the biology of the cattle tick *R. microplus* and their relevance for future tick control.

Materials and Methods

**In silico identification of transmembrane proteins**

Gene expression data of *R. microplus* immature life stages (larvae and nymph), as well as tissues (ovary, salivary gland and midgut) were obtained from previous microarray studies (Maritz-Olivier et al. 2012; Stutzer et al., 2013). From this data, the normalized intensity values for Cy5-labeled test groups (i.e. larvae, nymphs, ovaries, midgut, and salivary glands) were used for selection and a minimum signal intensity threshold of 1,000 (M-values > 0) was chosen for further evaluation of expressed transcripts to identify transmembrane proteins. The putative open reading frames of these transcripts were determined using the Prot4EST (http://www.compsysbio.org/lab/?q=prot4EST) software (Wasmuth and Blaxter, 2004). Consequently, the number of predicted transmembrane helices and putative signal peptides that would indicate protein secretion were determined from the encoded protein sequences using the TMHMM (v.2.0, http://www.cbs.dtu.dk/services/TMHMM/) and SignalP (v.3.0, http://www.cbs.dtu.dk/services/SignalP-3.0/) servers, respectively.

The outputs of both TMHMM and SignalP were used to obtain a list of predicted transmembrane proteins. Briefly, at least one identified transmembrane region must have been predicted with TMHMM if no signal peptide was identified with SignalP, while at least two transmembrane regions were needed in case a signal peptide was present in the first 60 amino acids. The obtained list was further analysed to differentiate between predicted single- and multi-spanning
proteins. Therefore, transcripts with one predicted transmembrane region (and no predicted signal peptide), as well as two predicted regions (containing a predicted signal peptide) were considered as single-spanning. Transcripts with two predicted transmembrane regions lacking a predicted signal peptide or containing more than two predicted transmembrane helices were defined as multi-spanning.

**Functional annotation and immuno-informatics**

For functional annotation, sequence alignment searches –BLASTs- were performed against non-redundant -NR-, eukaryotic gene ontology -KOG-, gene ontology -GO-, Protein family -Pfam-, simple modular architecture research tool -SMART- and mitochondrial plastid –MitPla-databases (http://exon.niaid.nih.gov), using the dCAS (v.1.4.3) desktop cDNA annotation software (Guo et al. 2009). The transcript list of predicted transmembrane proteins was then manually curated and functional annotation was based on a minimum expected value -E-value- of \(10^{-10}\) for at least two of the databases used (NR, KOG and GO) with similar protein hits. Reviewed entries of orthologous proteins in the UniProtKB database (http://www.uniprot.org/help/uniprotkb) were used for comparison of localization, topology and function (accessed November 2014). The immuno-informatic web-based tool VaxiJen (http://www.ddg-pharmfac.net/vaxijen/vaxijen/vaxijen.html) was employed with a threshold of 0.5 (parasite target organism database) to determine the likelihood of the predicted transmembrane proteins to be protective antigens (Doytchinova and Flower 2007). The only commercialized tick vaccine antigen, Bm86, obtained a score of 0.7698 and serves as an example of a good vaccine candidate.

The final set of transmembrane proteins confirmed with LC-MS/MS analysis (Table S1) were also screened for their presence in the midgut transcriptome of *R. decoloratus* and were
considered present if it showed a log₂ fold change of ≤1.7 -with P value < 0.05- (van Zyl et al. 2015). Matching refseq and GenBank numbers from *I. scapularis* and consensus sequences as described by Prudencio et al. (2011) were also identified.

**LC-MS/MS analysis of the midgut proteome**

A pilot proteomic study was performed in collaboration with Stellenbosch University (Proteomics Laboratory, Central Analytical Facility, South Africa) to develop a procedure for analysis of *R. microplus* proteins. Engorged adult female *R. microplus* ticks (n = 40), from a laboratory bred strain (South African strain, ClinVet Inc.), were collected at days 18 to 21 post infestation from Hereford cattle kept in a controlled environment at the Onderstepoort Veterinary Institute (Pretoria, South Africa). Ethical clearance was obtained from the South African Dept. of Agriculture, Forestry and Fisheries as well as the University of Pretoria’s Animal Use and Care Committee (Project approval number EC022-10).

Tick midgut tissues were dissected, pooled and submerged in sterile cold phosphate-buffered saline (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 2 mM KH₂PO₄, pH 7.4) containing protease inhibitor (cOmplete ULTRS Tablets; Roche Diagnostics, South Africa). Dissected midgut material was washed 10x in 1 ml PBS with protease inhibitor, by inverting the suspension followed by centrifugation (10,000 g for 10 min). The washed midgut material was subjected to protein extraction using the ReadyPrep™ Protein Extraction kit (Membrane) as described by the manufacturer (Bio-Rad; USA). Three fractions were collected during the extraction procedure: a pellet, hydrophilic phase and hydrophobic phase. The hydrophilic and hydrophobic phases were subjected to protein concentration determination using using the 2-D Quant kit (Amersham Biosciences, USA) and electrophoretic separation on polyacrylamide gel (10%).
Some 8 µg/ul total protein was submitted for mass spectrometry (Proteomics laboratory, Central Analytical Facility, Stellenbosch University, South Africa). Samples were cut from the polyacrylamide gel and processed using filter-aided sample preparation (FASP) similar to the method described by Wiśniewski et al. (2009). The final trypsinized proteins were dried in vacuo and the resultant peptides were re-dissolved in 5% acetonitrile containing 0.1% formic acid. Coupled high performance liquid chromatography and electrospray tandem mass spectrometry - LC-MS/MS- was performed using an EASY-nLC II system connected to a linear trap quadropole -LTQ- Orbitrap Velos mass spectrometer (Thermo Scientific, Bremen, Germany). Chromatographic separation of peptides was achieved on a XBridge BEH130 C18 NanoEase column (Waters Corporation, USA) fitted with a pre-column using a step-wise chemical gradient and a flow rate of 300 nl/min. The step gradient consisted of an increase of solvent B (0.1% v/v formic acid in 100% acetonitrile) from solvent A (0.1% v/v formic acid in HPLC grade water) by: 5-17% in 5 min, 17-25% in 90 min, 25-60% in 10 min, 60-80% in 5 min and then for 10 min at 80% solvent B.

For automatic acquisition of mass spectra, the LTQ-Orbitrap system was operated in data-dependent mode to switch between Orbitrap-MS and LTQ-MS/MS analyses. Dynamic exclusion was performed (60 sec) with an ion selection threshold of 500 counts for MS/MS (activation Q-value of 0.25 for 10 ms). The raw MS/MS spectra were acquired using the Xcalibur™ software package (Thermo Scientific, Bremen, Germany). The resolution was set at 60,000 and precursor ion scan MS spectra (m/z 400 – 2000) were acquired. The 20 most intense ions were finally isolated for fragmentation using collision induced dissociation. The lock mass option was enabled and the polydimethylcyclosiloxane ions generated during electrospray were used for internal recalibration (m/z = 445.120025) for accurate mass measurement in both the MS and MS/MS modes (Schlosser and Volkmer-Engert, 2003; Olsen et al., 2005).
Data was analysed using Thermo Proteome Discoverer (v.1.3) (Thermo Scientific, Bremen, Germany) to identify proteins via automated database searching (Mascot, Matrix Science, London, UK) of all tandem mass spectra against an in-house database (Maritz-Olivier et al., 2012). A fixed modification for carbamidomethyl cysteine and variable modifications for oxidized methionine, N-acetylation and deamidation –NO- was set, as well as a precursor mass tolerance of 10 parts per million –ppm- with a fragment mass tolerance of 0.8 Da. Two missed tryptic cleavages were allowed. Proteins were considered positively identified when a Mascot score with a probability value of less than 0.05 (p<0.05) was determined by Percolator in the Proteome Discoverer suite. Validation of search results was further performed in Percolator where a decoy database was searched with a strict false discovery rate -FDR- of 0.02 and a relaxed FDR of 0.05 with final validation based on the q-value.

Results

Transcriptional profile of putative *R. microplus* transmembrane proteins

Microarray analysis of transcripts expressed throughout the life cycle and various tissues of *R. microplus* identified 13,456 transcripts for *R. microplus* adult female tissues (salivary glands, midguts and ovaries), as well as immature life stages (larvae and nymph) (Maritz-Olivier et al. 2012). Final topology and signal peptide prediction for all transcripts were performed and a list of 878 putative transmembrane proteins was identified of which 215 contained a predicted signal peptide. These transcripts were subsequently clustered according to the different tick tissues and life stages where they were expressed (Figure 1A). Ten genes (1.1%) were found to be transcribed in all immature life stages, as well as adult female tissues. Furthermore, 23 transcripts (2.6%) were found to be present only in the immature life stages and midgut tissues,
along with 46 transcripts (5.2%) present only in all the adult female tissues (salivary gland, ovary and midgut) tested.

**Figure 1: Distribution of identified transcripts containing predicted transmembrane domains in *R. microplus* life stages and tissues.** A) The Venn diagram represents the number of genes which are expressed above a threshold uniquely in or shared between tick midgut, larvae, nymph, salivary gland, and ovary. B) A table summarizing the shared genes expressed in and shared between life stages and tissues as percentages based on all identified 878 putative transmembrane proteins (G: midgut; N: nymph; L: larvae; SG: salivary gland; O: ovary).

Numerous putative transmembrane transcripts are ovary- and larvae specific, making up 12% and 11% of the 878 transmembrane transcripts, respectively (Figure 1B). Generally, the least number of transcripts were found to be shared between four or all five of the life stages and tissues investigated with 0.5% to a maximum of 1.6% of all transcripts. One exception is transcripts shared between all immature life stages and tissues (excluding ovary) that make up
almost 7% of the 878 total sequences investigated. Additionally, transcripts shared between ovary, salivary gland and larvae, as well as ovary, larvae and nymph only represented 1% of the total sequences investigated (Figure 1B).

**Single-spanning predicted transmembrane proteins**

Analysis of all 878 transmembrane proteins according to their number of transmembrane regions identified 510 single-spanning transcripts. Following manual curation, 79 transcripts could be annotated and their subcellular localization predicted. The majority of transcripts were localized to the endoplasmatic reticulum (22.8%), mitochondria (16.5%) and plasma membrane (13.9%) (Figure 2A). Functional annotation showed that most transcripts are involved in intracellular trafficking, secretion and vesicular transport (26.6%) followed by unknown functions (19.0%) and posttranslational modification, protein turnover, chaperones (13.9%) (Figure 2B).
Figure 2: Localization and functional prediction of identified transmembrane proteins. Numbers are given as percentages of all identified annotatable single- or multi-spanning transcripts. Regulated transcripts in all tissues are classified according to their eukaryotic orthologous functional groups (KOGs) based on Tatusov et al. (2003). A) Percentages of single- and C) multi-spanning transmembrane proteins localized to a specific subcellular compartment. Functional prediction of single-spanning (B) and multi-spanning (D) transcripts. Abbreviations: ER, endoplasmatic reticulum; GA, Golgi apparatus; Mit, mitochondrion; PM, plasma membrane; Per, peroxisome; Lys, lysosome; Ves, vesicle; End, endosome; Nuc, nucleus; Unk, unknown. Functional groups are assigned as follows: (A) RNA processing and modification; (B) chromatin structure and dynamics; (C) energy production and conversion; (D) cell cycle control, cell division, chromosome partitioning; (E) amino acid transport and metabolism; (F) nucleotide transport and metabolism; (G) carbohydrate transport and metabolism; (H) coenzyme transport and metabolism; (I) lipid transport and metabolism; (J) translation, ribosomal structure and biogenesis; (K) transcription; (L) replication, recombination and repair; (M) cell wall/membrane/envelope biogenesis; (N) cell motility; (O) posttranslational modification, protein turnover, chaperones; (P) inorganic ion transport
Multi-spanning predicted transmembrane proteins

Prediction of the number of transmembrane regions identified 368 putative multi-spanning transcripts of which 188 could be annotated. Distribution of the transcripts identified most transcripts to be localized to the endoplasmatic reticulum (28.2%), the plasma membrane (25.5%) and the mitochondria (16.0%) (Figure 2C). No subcellular localization could be assigned to 10.1% of the identified transcripts, stressing the need for proteome analysis of *R. microplus*. Following functional annotation, most annotatable transcripts were found to be involved in ion transport and metabolism (13.8%), lipid transport and metabolism (11.7%), as well as energy production and conversion (11.7%) (Figure 2D). However, the largest proportion of transcripts (15.4%) had unknown functions.

Proteomics analysis of transmembrane proteins in *R. microplus* midgut tissue

A total number of 292 putative proteins were identified using LC-MS/MS analysis, of which 30 transmembrane proteins could be mapped to annotatable open reading frames obtained from the analysed transcriptomic data (Table 2). Eleven of these were also identified as being present in *R. decoloratus* (van Zyl et al., 2015). Homologous sequences for all of these 30 proteins were identified in *I. scapularis*. Six protein sequences contained consensus sequences representing potential *R. microplus* vaccine targets as identified by Prudencio et al. (2011). To the best of our knowledge 19 proteins have not been identified using proteomics approaches in
ticks, and only 11 have been previously reported (Table 2). For the prediction of protective antigens and subunit vaccines the program VaxiJen was used. Nine confirmed proteins were found to have an antigenicity probability score below a threshold of 0.5. However, most proteins had a score from 0.5 to 0.7 (15) and 6 additional proteins showed predicted scores above 0.7 (Table 2).
Table 2: List of annotatable single- and multi-spanning transmembrane proteins. Proteins were identified through manual curation. Abbreviations: mit – mitochondrion; er - endoplasmatic reticulum, plasma mem – plasma membrane. A protein was marked as confirmed if LC-MS/MS analysis resulted in a Mascot score of above 30. Presence of a transcript is indicated in the midgut of *R. decoloratus* and genome of *Ixodes scapularis* is indicated. Published proteins identified in any tick species by mass spectrometry are also shown. Consensus sequences identified according to Prudencio et al. (2011) are indicated in the motif column. Contiguous sequence numbers are indicated with additional short numbers referred to in Figure 3.

<p>| General Class(^a) | Contig | Description(^b) | VaxiJen (score) | Motifs according to Prudencio et al. 2011 | Localization inferred via UniProt and/or literature(^a) | LC-MS-MS of midgut tissues Mascot score (nr. of unique peptides) | Presence in tissue and life stage transcriptome of <em>R. microplus</em>(^a) | Presence in midgut transcriptome of <em>R. decoloratus</em>(^a) | Transcripts form <em>I. scapularis</em> Similar transcripts identified from published proteomics data in ticks |
|----------------------|--------|------------------|-----------------|---------------------------------------------|-------------------------------------------------|-------------------------------------------------|-------------------------------------------------|---------------------------------------------|-------------------------------------------------|-------------------------------------------------|
| <strong>Single- spanning</strong> |
| Energy production and conversion | Contig5414 | Putative cytochrome c1 (CYC1), heme protein, mitochondrial Cytochrome c family | 0.54 | None | mit | 30.98 (1) | N,SG,MG | Yes | ref|XP_002411993.1| gb|EEC14489.1| Kongsuwan 2010 |
| | Contig3319 | Putative cytochrome c oxidase subunit IV, mitochondrial (COX41) Cytochrome c oxidase IV family | 0.57 | None | mit | 74.22 (1) | SG,MG | No | ref|XP_002400930.1| gb|EEC05750.1| Diaz-Martin 2013, Francischetti 2011 |
| Function unknown | Contig4076 | Putative lysosome-associated membrane glycoprotein 1 (LAMP1) LAMP family | 0.72 | None | lysosome | 33.73 (1) | MG,OV | No | ref|XP_002401516.1| gb|EEC00677.1| n.i. |
| | Contig278A | Putative antigen B membrane protein PAN-1 domain | 0.44 | None | unknown | 607.76 (14) | MG | Yes | ref|XP_002404323.1| gb|EEC01169.1| Kongsuwan 2010 |
| Intracellular trafficking, secretion, and vesicular transport | Contig2515 | Putative vesicle-trafficking protein SEC22b (SEC22B) Synaptobrevin family | 0.48 | None | vesicle | 41.77 (1) | N | No | ref|XP_00240323.1| gb|EEC01169.1| n.i. |
| | Contig3777 | Putative vesicular integral-membrane protein VIP36 (LMAN2) L-type lectin-like domain | 0.74 | None | vesicle | 53.52 (2) | SG,MG | No | ref|XP_00240938.1| gb|EEC17711.1| n.i. |
| | Contig5295 | Putative translocon-associated protein subunit delta (SSR4) TRAP-delta family | 0.58 | None | er | 89.14 (2) | SG,MG | Yes | ref|XP_002405540.1| gb|EEC15348.1| n.i. |
| Posttranslational modification, protein turnover, chaperones | Contig5725 | Putative dolichyl-diphosphooligosaccharide−protein glycosyltransferase subunit 1 (RPN1) OST1 family | 0.49 | None | er | 70.58 (5) | SG,MG | Yes | ref|XP_002409944.1| gb|EEC02477.1| Kongsuwan 2010 |
| | Contig1372 | Putative calnexin (Canx) Calreticulin family | 0.91 | PXFF | er | 261.89 (1) | N,SG,MG,OV | Yes | ref|XP_002401232.1| gb|EEC05774.1| Kongsuwan 2010 |
| | Contig1356 | Putative dolichyl-diphosphooligosaccharide−protein glycosyltransferase 48 kDa subunit (DDOST) DDOST 48 kDa subunit family | 0.54 | PxxKxH | er | 607.81 (8) | N,SG,MG | No | ref|XP_002403874.1| gb|EEC08860.1| Kongsuwan 2010 |</p>
<table>
<thead>
<tr>
<th>General Class*</th>
<th>Contig</th>
<th>Descriptiona</th>
<th>VaxiJen (score)</th>
<th>Motifs according to Pudeficio et al. 2011</th>
<th>Localization inferred via UniProt and/or literatureb</th>
<th>LC-MS-MS of midgut tissues Mascot score (nr. of unique peptides)</th>
<th>Presence in tissue and life stage transcriptome of R. microplus*</th>
<th>Presence in midgut transcriptome of R. decoloratus*</th>
<th>Transcripts form I. scapularis</th>
<th>Similar transcripts identified from published proteomics data in ticks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secondary metabolites biosynthesi s, transport and catabolism</td>
<td>CV438075 (11)</td>
<td>Putative thioredoxin-related transmembrane protein 1 (TMX1)</td>
<td>0.67</td>
<td>None</td>
<td>er</td>
<td>42.41 (1)</td>
<td>N,M,G,OV</td>
<td>Yes</td>
<td>ref[XP_002407565.1]</td>
<td>gb</td>
</tr>
<tr>
<td></td>
<td>Contig7373 (12)</td>
<td>Putative aldehyde dehydrogenase, dimeric NADP-prefering (Aldh2a1)</td>
<td>0.53</td>
<td>None</td>
<td>unknown</td>
<td>59.83 (3)</td>
<td>Ov</td>
<td>No</td>
<td>ref[XP_002414335.1]</td>
<td>gb</td>
</tr>
<tr>
<td></td>
<td>Contig1118 (13)</td>
<td>Putative dimethylamine monooxygenase [N-oxide-forming] 5 (FMOS)</td>
<td>0.56</td>
<td>None</td>
<td>er</td>
<td>105.90 (5)</td>
<td>N,M,G</td>
<td>No</td>
<td>ref[XP_002414517.1]</td>
<td>gb</td>
</tr>
<tr>
<td>Multi-spanning</td>
<td>Contig492 (14)</td>
<td>Putative aflavin 2 (ATL2)</td>
<td>0.53</td>
<td>None</td>
<td>er</td>
<td>41.06 (1)</td>
<td>L,N,SG,MG</td>
<td>No</td>
<td>ref[XP_002408177.1]</td>
<td>gb</td>
</tr>
<tr>
<td></td>
<td>Contig7051 (15)</td>
<td>Putative V-type ATPase subunit a</td>
<td>0.47</td>
<td>None</td>
<td>vesicle</td>
<td>69.47 (2)</td>
<td>SG,M,G,OV</td>
<td>Yes</td>
<td>ref[XP_002414796.1]</td>
<td>gb</td>
</tr>
<tr>
<td></td>
<td>Contig4119 (16)</td>
<td>Putative cytochrome c oxidase subunit 2 (COII)</td>
<td>0.40</td>
<td>None</td>
<td>mit</td>
<td>72.59 (1)</td>
<td>N</td>
<td>No</td>
<td>ref[NP_008512.1]</td>
<td>gb</td>
</tr>
<tr>
<td></td>
<td>Contig1009 (17)</td>
<td>Putative NAD(P) transhydrogenase (NNT), mitochondrial</td>
<td>0.50</td>
<td>None</td>
<td>mit</td>
<td>77.75 (1)</td>
<td>L</td>
<td>No</td>
<td>ref[XP_002408762.1]</td>
<td>gb</td>
</tr>
<tr>
<td></td>
<td>Contig4681 (18)</td>
<td>Putative ADP/ATP carrier protein (sesB) Mitochondrial carrier (TC 2.A.29) family</td>
<td>0.31</td>
<td>None</td>
<td>mit</td>
<td>1300.84 (11)</td>
<td>SG,M,G,OV</td>
<td>No</td>
<td>ref[XP_002414592.1]</td>
<td>gb</td>
</tr>
<tr>
<td>Inorganic ion transport and metabolism</td>
<td>Contig862 (19)</td>
<td>Putative innexin 2 (inx2) Pannexin family</td>
<td>0.71</td>
<td>None</td>
<td>plasma mem</td>
<td>138.45 (1)</td>
<td>L,N,SG,OV</td>
<td>Yes</td>
<td>ref[XP_002400049.1]</td>
<td>gb</td>
</tr>
<tr>
<td>Intracellular trafficking, secretion, and vesicular transport</td>
<td>Contig2157 (20)</td>
<td>Putative signal peptidase complex subunit 2 (SPCS2)</td>
<td>0.46</td>
<td>None</td>
<td>er</td>
<td>50.24 (1)</td>
<td>SG,M,G</td>
<td>No</td>
<td>ref[XP_002401970.1]</td>
<td>gb</td>
</tr>
<tr>
<td></td>
<td>Contig4349 (21)</td>
<td>Putative translocon-associated protein subunit gamma (SSR3)</td>
<td>0.48</td>
<td>None</td>
<td>er</td>
<td>53.48 (1)</td>
<td>L,N,SG,MG</td>
<td>No</td>
<td>ref[XP_002411611.1]</td>
<td>gb</td>
</tr>
<tr>
<td></td>
<td>Contig8566 (22)</td>
<td>Putative reticulon-1 (Rtn1) reticulon domain</td>
<td>0.46</td>
<td>None</td>
<td>er</td>
<td>59.46 (3)</td>
<td>L,N,SG,MG</td>
<td>No</td>
<td>ref[XP_002407248.1]</td>
<td>gb</td>
</tr>
<tr>
<td>Lipid transport and metabolism</td>
<td>Contig181A (23)</td>
<td>Putative STAR-related lipid transfer protein 3 (STARD3) MENTAL &amp; START domain</td>
<td>0.62</td>
<td>None</td>
<td>er</td>
<td>40.20 (1)</td>
<td>MG,OV</td>
<td>Yes</td>
<td>ref[XP_002413856.1]</td>
<td>gb</td>
</tr>
<tr>
<td></td>
<td>Contig1926 (24)</td>
<td>Putative very-long-chain enoyl-CoA reductase (TECR) Steroid 5-alpha reductase family</td>
<td>0.52</td>
<td>None</td>
<td>er</td>
<td>51.56 (1)</td>
<td>OV</td>
<td>No</td>
<td>ref[XP_002401986.1]</td>
<td>gb</td>
</tr>
<tr>
<td>General Class</td>
<td>Contig</td>
<td>Description</td>
<td>VaxiJen (score)</td>
<td>Motifs according to Pudencio et al. 2011</td>
<td>Localization inferred via UniProt and/or literature</td>
<td>LC-MS-MS of midgut tissues Mascot score (nr. of unique peptides)</td>
<td>Presence in tissue and life stage transcriptome of R. microplus</td>
<td>Presence in midgut transcriptome of R. decoloratus</td>
<td>Transcripts form I. scapularis</td>
<td>Similar transcripts identified from published proteomics data in ticks</td>
</tr>
<tr>
<td>---------------</td>
<td>--------</td>
<td>-------------</td>
<td>----------------</td>
<td>------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>----------------------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Contig4425 (25)</td>
<td>Putative Niemann-Pick C1 protein (NPC1) Patched family</td>
<td>0.54</td>
<td>None</td>
<td>vesicle</td>
<td>106.04 (2)</td>
<td>MG,OV</td>
<td>Yes</td>
<td>ref</td>
<td>XP_002411148.1</td>
</tr>
<tr>
<td></td>
<td>Contig8380 (26)</td>
<td>Putative sterol O-acyltransferase 1 (Soat1) Membrane-bound acyltransferase family, Sterol o-acyltransferase subfamily</td>
<td>0.73</td>
<td>LHXXL 2x</td>
<td>er</td>
<td>177.54 (2)</td>
<td>MG,OV</td>
<td>No</td>
<td>ref</td>
<td>XP_002401663.1</td>
</tr>
<tr>
<td>Posttranslational modification, protein turnover, chaperones</td>
<td>Contig4332 (27)</td>
<td>Putative dolichyl-diphosphooligosaccharide–protein glycosyltransferase subunit STT3A (STT3A) STT3 family</td>
<td>0.56</td>
<td>None</td>
<td>er</td>
<td>44.79 (2)</td>
<td>N,SG,MG</td>
<td>Yes</td>
<td>ref</td>
<td>NP_001001202.1</td>
</tr>
<tr>
<td></td>
<td>Contig1719 (28)</td>
<td>Putative dolichyl-diphosphooligosaccharide–protein glycosyltransferase subunit 2 (Rpn2) SWP1 family</td>
<td>0.59</td>
<td>None</td>
<td>er</td>
<td>151.68 (1)</td>
<td>SG,MG</td>
<td>No</td>
<td>ref</td>
<td>XP_002399904.1</td>
</tr>
<tr>
<td>Signal transduction mechanisms</td>
<td>Contig1297 (29)</td>
<td>Putative tetraspanin, CD63 antigen (CD63) Tetrascarin (TM4SF) family</td>
<td>0.71</td>
<td>None</td>
<td>vesicle</td>
<td>40.29 (1)</td>
<td>MG</td>
<td>No</td>
<td>ref</td>
<td>XP_002433632.1</td>
</tr>
<tr>
<td></td>
<td>CK181827 (30)</td>
<td>Putative tetraspanin, CD9 antigen (CD9) Tetrascarin (TM4SF) family</td>
<td>0.61</td>
<td>None</td>
<td>plasma mem</td>
<td>87.13 (1)</td>
<td>L,OV</td>
<td>No</td>
<td>ref</td>
<td>XP_002435757.1</td>
</tr>
</tbody>
</table>

a. Classification of transcripts according to eukaryotic orthologous group terms for gene ontology.
b. Functional annotation of genes based on comparison among BLAST output from NR, GO and KOG databases. All transcript descriptions are based on consensus with database entries with UniProt.
c. Indicated are the putative cellular localizations of each transcript obtained from literature and UniProt outputs for homologues proteins. Vesicle denotes localization of proteins in several subcellular compartments such as the golgi, endosome, lysosome or peroxisome.
d. Based on data from Maritz-Oliver et al. (2012)
e. Based on data from van Zyl et al. (2015)

Abbreviations: er, endoplasmic reticulum; mit, mitochondrion; n.i., not identified; nr., number; plasma mem, plasma membrane.
Discussion

Only 6.5% of all 13,456 transcripts previously analysed by Maritz-Olivier et al. (2012) could be identified as putative transmembrane proteins. This is less than the generally predicted percentage for all pro- and eukaryotes that ranges between 20 – 30% (Jones 1998, Krogh et al. 2001, Martin-Galiano and Frishman 2006). These results may likely be due to false negative predictions as a result of very stringent selection criteria or due to limitations in the prediction software used, rather than distinctive differences in the biology of ticks.

Distribution of predicted transmembrane proteins

Due to the fact that little is known about tick proteomes and their networks, the description and annotation of putative proteins play a crucial role in order to improve our knowledge on basic tick biology. The largest proportions of *R. microplus* transmembrane transcripts identified from transcriptome data in this study were specific to larvae (11.2%) and adult ovaries (11.7%) (Figure 1), highlighting the fact that a unique set of membrane proteins is required for larval and ovarian development and tissue-specific function. Similar results were obtained in the previous study conducted by Maritz-Olivier et al. (2012), where a global comparison of transcription indicated that ovary-specific transcripts represented the highest number (larvae second) of uniquely transcribed genes present in *R. microplus*.

In contrast to the small number of predicted transmembrane proteins shared between all life stages and tissues (1.1%) from this study, for all transcripts included in the analysis by Maritz-Olivier et al. (2012), 23% were identified to be expressed in all tick life stages and tissues above an arbitrary intensity threshold of 1,000. These transcripts are of interest especially in regards to tick control, where target sets can be chosen that will affect the immature life stages, as well as...
vital pathways involved in feeding (salivary glands), nutrient acquisition (midguts) and reproduction (ovaries) within adult female ticks.

**Single- versus multi-spanning predicted proteins**

The differentiation between single- and multi-spanning proteins are of interest in regards to development of new vaccine candidates, since proteins with multiple transmembrane regions are more difficult to express than with a single transmembrane domain and will require costly fractionation steps to provide pure antigens for vaccine formulation making it economically less viable to reach market (Pizza et al. 2000). Of all 878 identified putative transmembrane proteins, 24% contain a predicted signal peptide. However, the possibility of truncated sequences (lacking the N-terminal signal peptide) cannot be excluded and the topology prediction performed cannot distinguish between a signal peptide and a transmembrane region, consequently resulting in false predictions (Krogh et al. 2001). To overcome this limitation, all predicted transmembrane regions that were also identified as signal peptide during SignalP analyses are not included in this study.

As found for membrane proteins of other analysed genomes (Hubert et al. 2010), more than half of all the predicted tick transmembrane proteins were single-spanning (58%). However, 51% of multi-spanning proteins were annotatable compared to only 15.5% of single-spanning ones. Previous studies have shown that more than half of tick transcripts cannot be annotated (Wang et al. 2007), thus subcellular localization and functional class predictions are hampered especially if predictable leader sequences are lacking.
Subcellular localization and function prediction

Most of the annotatable single- and multi-spanning transcripts localized to the endoplasmatic reticulum are associated with a function in posttranslational modification, protein turnover, chaperones (protein folding) as well as signal transduction mechanisms. This highlights the organelle’s important and multi-layered functions in protein synthesis and transport (Figure 3).

Figure 3: An overview of the identified transmembrane proteins and associated biological processes for *R. microplus* ticks. Indicated is a simplified illustration of the cell and cellular compartments with key processes (in shaded blocks), as well as the proteins involved in each process or associated with a particular compartment. All associated proteins are identified in closed brackets with numbers ranging from 1 to 30 that correlate to their entries described in Table 2. Abbreviation: ER, endoplasmic reticulum.
Interestingly, a difference in the general class of intracellular trafficking, secretion and vesicular transport between single- and multi-spanning transcripts was observed. While this function is mostly associated with single-spanning predicted proteins (26.6%), only 10.6% of multi-spanning transcripts were predicted to have this function. Furthermore, several functions were found to be more enriched for multi-spanning compared to single-spanning transcripts including transport and metabolism of inorganic ions, carbohydrates and lipids (Figure 2B and 2D).

Genes encoding transporters, as well as signal transduction mechanisms of the complete gene content of sequenced organisms range between 5 to 15% (Lee et al. 2008). A functional difference between proteins depending on their number of transmembrane regions is apparent, as most transporters for example have multiple (> 5) membrane spanning regions, compared to proteins which span the membrane only a few times (von Heijne 2006). Therefore, differentiating between single- and multi-spanning proteins could also give insights into the function of a protein. In this regard, the involvement of especially multi-spanning putative proteins as transporters can be illustrated in this study. For the 79 annotatable predicted single-spanning transcripts, 11.4% were predicted to be involved in metabolic transport of coenzymes, carbohydrates, lipids, nucleotides, amino acids, inorganic ions and secondary metabolites. In contrast, of the 188 annotatable predicted multi-spanning transcripts, 37.7% were involved in these functional classes (Figure 2B and 2D). The functions and pathways that specific proteins (Table 2) are involved in will be discussed in following sections.

Annotation and Vaxijen analysis - Enrichment of our understanding of tick biology and novel tick antigens

Out of all 878 transmembrane proteins 510 were identified to be single-spanning. However, only 79 were found to be annotatable after manual curation. Similarly, only 188 multi-spanning
proteins were annotatable of the 368 predicted proteins. Therefore, out of the annotatable membrane spanning proteins obtained from the transcriptomic data, 30 (consisting of 13 single-spanning and 17 multi-spanning), could be confirmed using mass spectrometry analysis (Table 2). However, as this analysis was a pilot study and only used *R. microplus* midgut fractions from partially (day 18) to fully (day 21) engorged female ticks, expression of proteins present in other tick tissues from transcriptomic analysis remains to be confirmed. Furthermore, although *R. microplus* and *R. decoloratus* are closely related species (Murrell et al. 2001), only 11 of the identified proteins that were detected via proteomics analysis in *R. microplus* could be identified in *R. decoloratus* via transcriptomic analysis. This, however, could most likely be attributed to the fact that van Zyl et al. (2015) used an *R. microplus* microarray platform for the hybridization of *R. decoloratus* transcripts. For all the transcripts identified, orthologous sequences could be identified in *I. scapularis* highlighting the possibility of further development of candidate sequences for possible cross-species tick vaccines. Six proteins contained consensus sequences identified by Prudencio et al. (2011), representing potential targets for further evaluation in *R. microplus*. To the best of our knowledge 19 proteins have not been identified using proteomics approaches before in ticks while 11 have been previously reported (Table 2).

For the prediction of protective antigens and subunit vaccines the program VaxiJen was used in this study. Currently, identification of cattle tick vaccine candidates is hindered by the fact that the major histocompatibility alleles used for epitope screening is mostly based on human or murine alleles (de Groot et al. 2006, Nene et al. 2012). Furthermore, tick proteins are most likely a product of divergent or convergent evolution, since they lack obvious sequence similarity to most available sequences found in databases. Considering that most algorithms still rely on sequence alignment to identify sequence similarities or motifs characteristic of antigens (Flower et al. 2010), an alignment-free approach such as VaxiJen (Doytchinova and Flower 2007a; Doytchinova and Flower 2007b), is more robust to identify putative tick protective antigens.
Although never applied to ectoparasites, VaxiJen has shown impressive prediction accuracy of up to 70-89% for bacterial, viral and tumor antigens and 78-97% accuracy for endoparasitic and fungal antigens (Doytchinova and Flower, 2007a; Flower et al. 2010). The highest accuracy of antigen prediction was found at a threshold of 0.5 (Doytchinova and Flower 2007a). In this regard, 9 confirmed proteins were found to have a score below this threshold, while 15 proteins had a score ranging between 0.5 and 0.7 with 6 proteins that showed a score above 0.7 (Table 2). Interestingly, a putative calnexin (Contig 1372) identified in this study showed a higher VaxiJen score (0.91) than that predicted for Bm86 (Contig8501) (VaxiJen score: 0.77).

Proteins involved in protein translocation and modification

Nine proteins were found to be involved in the protein translocation and modification taking place in the ER - endoplasmatic reticulum - (Table 2, Figure 3). Several proteins associated with the translocon complex in the ER could be identified. This complex is responsible for protein translocation and integration into the ER membrane prior to protein sorting via the Golgi apparatus. In eukaryotes this complex consists of several transmembrane proteins that are part of the Sec family and TRAM protein family in mammals (Johnson and van Waes 1999). Proteins identified in this study include two translocon-associated protein (TRAP) subunits (Contig 5295 and 4349) which have not been reported before in ticks. This complex contains four subunits and has been found to be involved in the initiation of protein translocation (Fons et al. 2003), as well as the accelerated degradation of misfolded proteins (Nagasawa et al. 2007). A transmembrane signal peptidase (SP) and oligosaccharyltransferase (OST) complex is involved in cleaving off the signal peptide of proteins and to glycosylate the polypeptide, respectively (Mohorko et al. 2011, Haase Gilbert et al. 2013). One of the subunits of the SP complex has been newly identified in this study (Contig2157), as well as four subunits of the OST complex (Contig5725, 1356, 4332 and 1719) (Table 2) which were, however, all previously identified in
other publications in *R. microplus* in addition to *R. sanguineus* and *Ornithodoros moubata* for Contigs4332 and 1719 (Diaz-Martin et al. 2013, Oliveira et al. 2013). The latter is a hetero-oligomeric membrane complex which facilitates N-glycosylation, affects protein folding, aids in protein sorting as well as enables interactions of the cell with its environment (Mohorko et al. 2011). In eukaryotes eight polypeptides have been identified as being part of the OST complex and similar subunit compositions have been found between yeast and vertebrates (Kelleher and Gilmore 2006). Two proteins important for quality control and correct folding in the ER were also identified and include a putative thioredoxin-related transmembrane protein (CV438075) (Table 2). This protein is an oxidoreductase and a role in disulfide-linked protein folding in the ER is suggested (Matsuo et al. 2004). The second protein, calnexin (Contig1372), is a chaperone involved in the quality control of proteins ensuring correct folding and assembly of many glycoproteins in the ER. It is also suggested to be involved in stabilization of incorrectly folded proteins and retaining them until they are correctly folded or targeting them for degradation (Williams 2006). Calnexin has previously been reported in *R. microplus* while CV438075 has not been identified up to date to the best of our knowledge. Five of the proteins involved in protein translocation and modification were also identified in transcriptome data of *R. decoloratus*.

**Proteins involved in trafficking and metabolism**

An endoplasmic reticulum resident vesicle-trafficking protein, Sec22b (Contig2515), was identified as highly up-regulated in the nymphal stages of *R. microplus* via transcriptome data (Table 2, Figure 3). This SNARE protein is integral to the ER membrane and has been implicated in ER-mediated phagocytosis, vesicular transport between the cis-Golgi and ER, as well as transport between the trans-Golgi network to the lysosomes (Hatsuzawa et al. 2006, Okumura et al. 2006, Aoki et al. 2008, Siddiqi et al. 2010). An additional trafficking protein
(Contig377), the vesicular integral membrane protein of 36 kDa (VIP36), was also identified as highly expressed in adult *R. microplus* tissues that were supported by midgut proteomics analysis (Table 2, Figure 3). This type I transmembrane protein, is an L-type lectin involved in protein trafficking between the ER-Golgi intermediate compartment and the cis-Golgi, as well as retrograde transport to the ER of miss folded/modified glycoproteins for refolding or degradation (Kamiya et al. 2005, Kroeger et al. 2012). This process is calcium-dependent and optimal binding is achieved in the more acidic conditions (pH 6.5) prevailing in the cis- to trans-Golgi compartments (Weisz 2003, Kamiya et al. 2005).

Luminal acidification is achieved by intraluminal proton pumps, such as the V-type ATPases (Weisz 2003). In this regard, transcription of a V-type ATPase related subunit (Contig7051) was identified within adult female tissues of *R. microplus* of which the protein expression was shown in midgut tissues, as well as transcription within *R. decoloratus* midgut tissues (Table 2, Figure 3). This subunit forms part of the V0 membrane-imbedded domain of the V-type ATPase proton transport complex (Marshansky et al. 2014). The complex is required for luminal acidification of the endosomal compartments involved in various cellular processes including protein degradation, endosomal trafficking and protein sorting (Hurtado-Lorenzo et al. 2006, Hinton et al. 2009).

Lastly, two additional transcripts related to trafficking and organelle morphogenesis were identified that include putative reticulon (Contig8566) and atlastin (Contig492) proteins (Table 2, Figure 3). Reticulons have been identified in many eukaryotes, however, not in archaea or bacteria (Yang and Strittmatter 2007). These proteins play a role in shaping the membrane of the ER, trafficking between the ER and Golgi apparatus, as well as in apoptosis (Yang and Strittmatter 2007). The family of atlastins function as GTPases and atlastin-2 was found to be involved in ER and Golgi apparatus morphogenesis (Rismanchi et al. 2008). Both transcripts are novel proteins identified from transcriptomic data of the immature life stages and adult
female tissues of *R. microplus*, along with evidence of protein expression in midgut tissues (Table 2, Figure 3).

**Proteins involved in lipid transport and metabolism**

Recently, lipid transport and metabolism was identified as a key feature in the metabolism of both *R. microplus* and *R. decoloratus* (van Zyl et al. 2015). In this study several proteins related to lipid metabolism were identified (Table 2, Figure 3). Firstly, a transmembrane (Contig4425) protein sharing sequence similarity to the Niemann-Pick C family proteins (NPC1) was identified from *R. microplus* midgut transcriptome and proteome, as well as *R. decoloratus* midgut transcriptome data (Table 2, Figure 3). These proteins have been proposed to be involved in intracellular cholesterol cycling and have been identified from midgut transcripts of other tick species, such as *D. variabilis* and *I. ricinus* (Vanier and Millat 2004, Anderson et al. 2008, Horácková et al. 2010, Vance and Peake 2011). The Niemann-Pick C1 protein has also been shown to occur in the late-endosomal/lysosomal compartments, along with the V0 domain of the V-ATPase complex and the lysosome-associated membrane glycoprotein 1 (LAMP1) during cholesterol homeostasis and degradation in *Drosophila* cells (Swetha et al. 2011). A similar LAMP1 protein (Contig4076) was identified that was up-regulated in *R. microplus* tissues (Table 2, Figure 3).

Three additional proteins with putative functions in lipid transport and metabolism were identified including: a StAR-related lipid transfer protein (Contig181A), a putative very-long-chain enoyl-CoA reductase (Contig1926) and sterol O-acyltransferase 1 (Soat1) (Contig8380). These proteins are involved in processes such as absorption and metabolism of cholesterol, as well as lipoprotein assembly (Chang et al. 2006, Chang et al. 2009, Alpy and Tomasetto 2014).
three proteins have high VaxiJen scores and contain antigenic motifs previously identified by Prudencio et al. (2011) (Table 2, Figure 3).

**Proteins involved in energy transport and metabolism**

Five mitochondrial-related transcripts involved in energy transport and metabolism were identified including: three putative cytochrome-related transcripts (Contig3319, Contig4119 and Contig5414), a putative NAD(P) transhydrogenase (Contig1009) and a putative ADP/ATP carrier protein (Contig4681) (Table 2, Figure 3). Two subunits of cytochrome c oxidase (Contig3319 and Contig4119) were identified from *R. microplus* transcriptomic data in adult and nymphal tissues, respectively. Cytochrome c oxidase is a component of the mitochondrial respiratory chain involved in the transfer of electrons from cytochrome c to oxygen and similar proteins have been identified from the mialomes and sialomes of both hard and soft tick species (Anderson et al. 2008, Francischetti et al. 2010, Kongsuwan et al. 2010, Sarewicz and Osyczka 2015). A related cytochrome c hemeprotein (Contig5414) was identified, which acts as an electron carrier protein within the electron transport chain, as well as acting as an intermediary effector in programmed cell death (Liu et al. 1996, Sarewicz and Osyczka 2015). The mitochondrial enzyme, NAD(P) transhydrogenase, is involved in proton translocation across the inner mitochondrial membrane via the transfer between NAD(H) and NADP(+) to produce high concentrations of NADPH (Hatefi and Yamaguchi 1996). The produced NADPH is consequently utilized in biosynthetic pathways, as well as in free radical detoxification (Jezek and Hlavatá 2005). Lastly, the ADP/ATP carrier protein is an adenine nucleotide antiporter that catalyzes the exchange of cytosolic ADP and matrix-produced ATP across the inner membrane of the mitochondrion and was identified from transcriptomic and proteomics data in adult *R. microplus* tissues (Table 2, Figure 3) (Liu and Chen 2013, Monné and Palmieri 2014). The latter protein
Additional identified membrane proteins

Numerous additional transcripts were identified that are transcribed within the life stages and tissues of *R. microplus*. Two putative tetraspanin proteins, namely the CD63 antigen (Contig1297) and the CD9 antigen (CK181827), were identified in this study (Table 2, Figure 3). Tetraspanins encompass a large protein family present in cell membranes, the exocytotic pathway and the endosome-lysosome system (Pols and Klumperman 2009). These proteins are involved in cell migration, fusion and signalling events (Hemler 2005). CD63 in particular is one of the few tetraspanins that are predominantly found in the endosome-lysosome system and might function in intracellular transport of other proteins (Pols and Klumperman 2009).

Two transcripts involved in secondary metabolite metabolism and detoxification were identified that are expressed in adult *R. microplus* midgut tissues (Table 2, Figure 3). A putative aldehyde dehydrogenase (Aldh3a1) (Contig7373) was identified that is hypothesized to be involved in aldehyde detoxification resulting from UV-induced lipid peroxidation in humans (Pappa et al. 2003). Expression of this protein type has been previously reported in related proteomics studies (Rachinsky et al. 2007, Francischetti et al. 2011, Popara et al. 2013). Flavin-containing monooxygenase (Contig1118) is another transcript involved in secondary metabolism, identified from both transcriptomic and proteomics data (Table 2, Figure 3). This ER-resident enzyme functions in catabolism of foreign chemicals, including drugs and pesticides (Hodgson et al. 1995, Phillips and Shephard 2008). Lastly, an antigen B protein previously identified by Willadsen (Willadsen 2001) from the Australian *Rhipicephalus* species was also identified in this work in both transcriptomic and proteomics midgut datasets (Contig278A, Table 2). Though this
transcript appears to be shared with the *R. decoloratus* transcriptome (Table 2, Figure 3), the biological role of this protein remains to be elucidated.

**Implication for tick control**

Due in part to understanding and knowledge of tick biology to date, the study of membrane proteins is further complicated by the difficulty in isolation of functional protein in comparison to soluble counterparts. Therefore, bioinformatics is an invaluable tool which, although predictive, is a faster the limited and sometimes only method of gaining information about possible vaccine targets.

The importance of transmembrane proteins in basically all the metabolic functions of an organism makes them ideal targets for tick control. Particularly for tick vaccines, the presence of an antigen in the tick midgut is a promising target, as was seen with the Bm86-based vaccine (Willadsen et al. 1989, Willadsen 2001). In this study, two proteins located to the plasma membrane were identified (Table 2). Both have a VaxiJen score close to 0.7 and above the threshold of 0.5, indicating their potential as novel antigens. Tetraspanin proteins have been investigated as vaccine target against other organisms such as the parasitic nematode *Brugia malayi* (Dakshinamoorthy et al. 2013). For the latter organism, this targeted tetraspanin is present on the outside of the nematode and the vaccine resulted in an almost 64% protection (Dakshinamoorthy et al. 2013). Members of this protein family were also tested against the helminth *Schistosomes mansoni* with a reduction of 57% and 64% for mean adult worm burdens and liver egg burdens for tetraspanin-1, respectively, and 34% and 52% for tetraspanin-2 (Tran et al. 2006). Both tested tetraspanins are expressed on the outside of the organism and are exposed to the immune system of the host (Tran et al. 2006). Calnexin (Table 2) has also recently been identified as a promising vaccine candidate against multiple fungal pathogens.
(Wüthrich et al. 2015). Although the signal peptidase complex subunit 2 protein (Table 2) has not been tested in vaccine trials to date, signal peptidase type I vaccination of BALB/c mice resulted in a 81% reduction of *Leishmania major* loads (Rafati et al. 2006). Although the possibility of targeting an intracellular antigen in ticks has also been demonstrated (de la Fuente et al. 2011), the actual intracellular binding of antibodies to antigen has not been shown. Evidence is also lacking on the possibility that cross-reaction with other antigen(s) is the basis of protection. For the identification of further putative vaccine targets, proteins with a VaxiJen score above 0.5 should be expressed and validated in cattle vaccine trials, especially if a consensus motif according to Table 2 is present as well. Interestingly, all six proteins that were found to contain a consensus sequence representing potential *R. microplus* targets (Prudencio et al. 2011) also had a VaxiJen score above 0.5. All proteins were also identified in *I. scapularis* and several in *R. decoloratus*. The presence of putative antigens in several other tick species highlights the possibility of a “universal” vaccine targeting many species of economic importance at the same time. For the future development of next generation vaccine targets, immunoinformatics as a tool enables the identification of epitope regions or overall antigenicity of proteins of interest prior to costly wet lab experiments. This allows for a systematic selection of a finite set of targets that can be evaluated using in vivo and in vitro methods.

The most common tick control strategy is the use of acaricides. These can work in two ways, through ingestion of the chemical while the tick is feeding or through absorption through the tick cuticle (Kroeber and Guerin 2006). Therefore, knowledge of the localization and abundance of a target for chemical control in any given tissue (i.e. midgut) could also be used to maximise acaricide efficacy. Novel acaricides could be identified by looking at previously untargeted pathways or proteins, and their subsequent evaluation using computational biology, combinatorial chemistry and high-throughput screening (Ghosh et al. 2007). The most common
acaricides used for chemical control target mostly components of the nervous system and contractile/muscle tissues and include chemical classes such as carbamates, organophosphates, organochlorines, pyrethrins/pyrethroids, macrocyclic lactones (i.e. avermectins and milbemycins) and formamidines (George et al. 2004, Abbas et al. 2014). Other targets of acaricides includes proteins involved in the respiratory chain and chemicals developed to date that target this process include Diafenthiuron, Chlorfenapyr, Acequinocyl, Fenazaquin and Cyenopyrafen (Ruder and Kayser 1993, Nauen and Bretschneider 2002, Raghavendra et al. 2011, Van Leeuwen et al. 2014). A prime example is the vacuolar ATPase that has shown promise as an insecticidal target for inhibitory compounds such as the pea albumin 1b (PA1b) protein, with activity in several pest species including the cereal weevil and mosquitoes (Gressent et al. 2007, Huss and Wieczorek 2009, Gressent et al. 2011, Muench et al. 2014). Finally, some acaricides are involved in regulation of growth and development of an organism and include examples such as tetronic acid derivatives (such as Spirodiclofen) which inhibit lipid biosynthesis (Nauen and Bretschneider 2002). From Table 2 several proteins were identified that could be developed into novel acaricide targets. This includes proteins that are involved in energy, as well as lipid transport and metabolism.

Limitations and conclusions

This study on the transmembrane proteins in different life stages and tissues of R. microplus (and limited work on R. decoloratus) contributes to our understanding of the tick transcriptome and biology as a whole, paving the way for further studies in order to understand and consequently control this organism. However, a bioinformatics-based approach has several inherent limitations. Firstly, most alignment algorithms used to predict transmembrane proteins treat a protein as a single unit, not taking into consideration the different physical properties of
various protein regions (Shafrir and Guy 2004). These alignment programs have the advantage of working well for water-soluble or globular proteins, but this is not the case for transmembrane proteins that occur in two different physiochemical environments inside and outside the membrane bilayer (Shafrir and Guy 2004). Furthermore, as a result of the R. microplus genome being largely unannotatable (Wang et al. 2007); a large portion of this species biology remains unknown. Furthermore, much of the current data available is obtained from ESTs which contain a number of truncated sequences. This could also be a contributing factor to the large amount of seemingly unannotatable transcripts. Moreover, due to the lack of validated proteome data, correct open reading frames cannot be confirmed and only predictions based on open reading frames of homologous proteins or peptides are possible in most cases. Additionally, 70% of the tick genome consists of repetitive DNA (Ullmann et al. 2005) and many gene duplications and/or genome duplications have been suggested as being a major influence on tick evolution (Ribeiro et al. 2006). Thus, similar sequences can develop very different functions over time resulting in the incorrect homology based prediction of the function of many tick proteins.

Some of these problems are expected to be addressed with the availability of the complete R. microplus genome (Bellgard et al. 2012). This will allow determination of full length transcripts that will improve functional annotation, localization, as well as topology predictions. Additionally, large-scale proteomics will be needed to gain insight into the complex and dynamic protein networks of this organism that cannot be described with conventional DNA-based techniques. Moreover, a comprehensive transcriptomic and proteomic profile for R. microplus, across its life cycle, will enable characterization of constitutive and discrete expression of genes. Together, a full systems biology approach will lead to a better understanding of tick biology and consequently the improvement of tick control strategies.
Acknowledgements

This study was funded by the Red Meat Research Development Trust, the University of Pretoria Research Development Programme and the Technology and Human Resources for Industry Programme. We would also like to thank the Onderstepoort Veterinary Institute (OVI) for cattle management and tick rearing and the proteomics unit at the Central Analytical Facility of Stellenbosch University for LC-MS/MS analysis.

References


Rafati, S., F. Ghaemimanesh, and F. Zahedifard. 2006. Comparison of potential protection induced by three vaccination strategies (DNA/DNA, Protein/Protein and DNA/Protein) against *Leishmania major* infection using signal peptidase type I in BALB/c mice. Vaccine 24:3290-3297.

1989. Cloning and expression of a protective antigen from the cattle tick Boophilus  
Norfolk, UK: Caister Academic.  
Ribeiro, J. M. C., F. Alarcon-Chaidez, I. M. B. Francischetti, B. J. Mans, T. N. Mather, J. G.  
from Ixodes scapularis ticks. Insect Biochem Mol Cell 36:111-129.  
protective "concealed" antigen from Boophilus microplus. Purification, localization, and  
Rodríguez-Mallon, A., E. Fernández, P. E. Encinosa, Y. Bello, L. Méndez-Pérez, L. Ruiz, D.  
tick antigen shows high vaccine efficacy against the dog tick, Rhipicephalus sanguineus.  
Vaccine 30:1782-1789.  
Rodriguez, M., M. L. Penichet, A. E. Mouris, V. Labarta, L. L. Luaces, R. Rubiera, C. Cordoves,  
P. A. Sanchez, E. Ramos, and A. Soto. 1995. Control of Boophilus microplus  
populations in grazing cattle vaccinated with a recombinant Bm86 antigen preparation.  
Vet Parasitol 57:339-349.  
appendiculatus gut (Ra86) and salivary gland cement (Trp64) proteins as candidate
antigens for inclusion in tick vaccines: Protective effects of Ra86 on infestation with adult

Sarewicz, M. and A. Osyczka. 2015. Electronic connection between the quinone and
cytochrome c redox pools and its role in regulation of mitochondrial electron transport

laboratory air identified as source of extreme background signals in nanoelectrospray
mass spectrometry. J. Mass Spectrom. 38, 523– 525

*Rhipicephalus (Boophilus) microplus* embryo proteins as target for tick vaccine. Vet
Immunol Immunop 148:149-156.

Bioinformatics 20:758-769.

vaccine antigen for the control of cattle tick infestations in Indian situation. Vaccine
32:3488-3494.

Sheu, G. C., Y. Y. Wo, S. M. Yao, F. Y. Chou, T. C. Hsu, C. L. Ju, Y. Cheng, S. N. Cang, and C.
H. Lu. 2001. Characteristics and potency of an acellular pertussis vaccine composed of
pertussis toxin, filamentous hemagglutinin, and pertactin. J Microbiol Immunol Infect
34:243-251.

Siddiqi, S., A. Mani, and S. Siddiqi. 2010. The identification of the SNARE complex required for


Zhou, J., H. Gong, Y. Zhou, X. Xuan, and K. Fujisaki. 2006. Identification of a glycine-rich protein from the tick *Rhipicephalus haemaphysaloides* and evaluation of its vaccine potential against tick feeding. Parasitol Res 100:77-84.