A systematic, functional genomics and reverse vaccinology approach to the identification of vaccine candidates in the cattle tick, *Rhipicephalus microplus.*

Christine Maritz-Olivier[1*], Willem van Zyl[2], Christian Stutzer[2]

[1] Department of Genetics, Faculty of Natural and Agricultural Sciences, University of Pretoria, South Africa.

[2] Department of Biochemistry, Faculty of Natural and Agricultural Sciences, University of Pretoria, South Africa.

* Corresponding author: Christine Maritz-Olivier, Tel: +27 012 420 3945, Fax: +27 012 362 5327, E-mail address: christine.maritz@up.ac.za.

Abstract

In the post-genomic era, reverse vaccinology is proving promising in the development of vaccines against bacterial and viral diseases, with limited application in ectoparasite vaccine design. In this study, we present a systematic approach using a combination of functional genomics (DNA microarrays) techniques and a pipeline incorporating in silico prediction of subcellular localization and protective antigenicity using VaxiJen for the identification of novel anti-tick vaccine candidates. A total of 791 candidates were identified using this approach, of which 176 are membrane-associated and 86 secreted soluble proteins. A preliminary analysis on the antigenicity of selected membrane proteins using anti-gut antisera yielded candidates with an IgG binding capacity greater than previously identified epitopes of Bm86. Subsequent vaccination trials using recombinant proteins will not only validate this

approach, but will also improve subsequent reverse vaccinology approaches for the identification of novel anti-tick vaccine candidates.

Introduction

The availability of the *Ixodes scapularis* genome and nearly completed genome of *Rhipicephalus microplus*, in combination with the fast expanding amount of transcriptomic data, is enabling scientists to venture into a post-genomic era for anti-tick vaccine design. By utilizing the principles of genome-based vaccine development (reverse vaccinology), it is possible to access all the proteins that are encoded by an organism using available genome or transcriptome data in combination with computational analysis, rather than starting with the organism itself (Moriel et al. 2008). Past experience has however indicated some essential contributing factors to protective immunity that one needs to consider when devising a genome-wide vaccine screening strategy. These include host factors such as the presence of B- and T-cell epitopes and the type of immune response elicited, parasite/pathogen factors such as expression level of antigens and their subcellular localization, as well as chemical and physical properties of an antigen such as its post-translational modifications and aggregation status when used for vaccination (Fig. 1).

Host contributing factors

Principles underlying protective immunogenicity in humans against viral diseases, bacterial pathogens and intracellular eukaryotic parasites such as *Plasmodium* are well described. However, very limited information is available for ectoparasites, such as ticks and
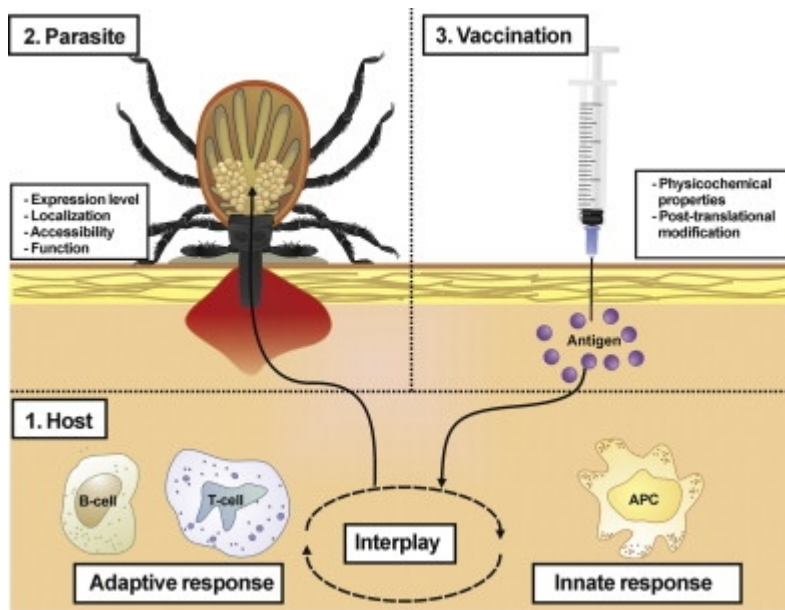
Fig. 1. Factors underlying immunogenicity. Immunogenicity is influenced by a great number of factors, these include: (1) Factors associated with host immunity, (2) parasite proteins and (3) the vaccine antigen. Abbreviation: APC- Antigen presenting cell. Adapted from (Flower et al. 2010).

their associated hosts. In the host, a parasite/pathogen is faced by both the innate defense mechanisms (mediated by tissue-residing macrophages and dendritic cells, as well as mobile phagocytic cells) and the adaptive immune response (mediated by large sets of molecules and cells that confer either regulatory or effector functions). Research has revealed that innate immunity sets the scene for the adaptive response; and that innate and adaptive immunity have to interact vigorously (via antigen presenting cells) to confer protective immunity (Fig. 1) (Zepp 2010). Protective immunity against a vast amount of diseases is mediated via the adaptive immune response, meaning it is antibody -, cytokine-, B- and/or T-cell dependent (Zepp 2010). Evidence for the possibility of induced protective immunity against ticks was provided in 1988 by Opdebeeck and colleagues, who indicated that membrane fractions from *R. microplus* midgut tissue protected cattle by 91% against tick challenge, and that the levels of IgG1 and complement-fixing antibodies related to the levels of protection induced by

vaccination (Opdebeeck et al. 1988; Jackson and Opdebeeck 1990). Furthermore, the only commercialized anti-tick vaccine based on the midgut protein, Bm86, induces a strong IgG-mediated response and the IgG alone (or with the aid of complement) is enough to damage the tick gut (Kemp et al. 1989; Rand et al. 1989; Vargas et al. 2010). For a review on the identification of Bm86 and its use as a vaccine, refer to De La Fuente et al. (De La Fuente et al. 2007). In 2009, Piper et al. showed that both *Bos indicus* and *Bos taurus* cattle infested with *R. microplus* displayed a strong adaptive immune response. *Bos indicus* displayed a strong T-cell mediated response, while in *Bos taurus* breeds, high levels of inflammatory molecules, IgG1 and elevated MHCII gene expression was observed. In conclusion, it became evident that an acquired T-cell response is critical to the development of tick-specific IgG and most probably to host resistance to infestation (Piper et al. 2009).

Recognition of epitopes by T-cells, B-cells and soluble antibodies forms the basis of the immune response. To date, various in silico T- and B-cell epitope predictors have been designed, and used with great success for bacterial and viral pathogens. For anti-tick vaccine development, the greatest limitation of these programs lies therein that they use human or murine major histocompatibility alleles, and that very limited data is available for bovine alleles. It should also be noted that evaluating an antigen's similarity to host proteins is troublesome, as it may have similar epitopes. These may go undetected when using conventional similarity-based BLAST searches. As a result, vaccination could lead to the production of cross-reactive, high-affinity antibodies that recognize the host during vaccination and can to date only be determined during vaccine trials. As 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) and 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,

may lead to the identification of truly novel tick protective antigens. VaxiJen was developed to allow antigen classification solely based on the physicochemical properties of proteins without recourse to sequence alignment (Doytchinova and Flower 2007a; Doytchinova and Flower 2007b). 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 2007b; Flower et al. 2010). As whole antigen prediction is best used in conjunction with other methods such as membrane topology and subcellular localization prediction, we made use of both during the strategy of this study (Fig. 2) (Vivona et al. 2008).
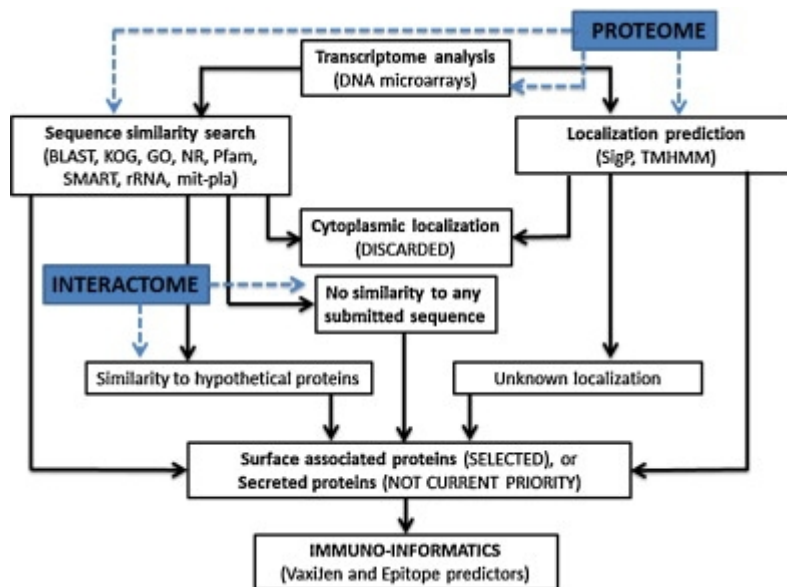

Fig. 2. Strategy used for the identification of anti-tick vaccine candidates using a functional genomics and in silico reverse vaccinology approach. As discussed, proteomics will greatly aid in confirming predicted open reading frames and the expression levels of their respective encoded proteins, thereby improving similarity searches. Insight into the interactome will greatly assist in the annotation of proteins, as protein function and tertiary structure are more conserved than primary protein sequence. Thus elucidation of biochemical pathways and functional protein complexes may be possible. Adapted from (Rappuoli and Bagnoli 2011)

Parasite/pathogen contributing factors

In a reverse vaccinology approach to identify promising protein-based vaccines, in silico analysis remains the central step. Typically, the first step entails the prediction and localization of genes within the genome, followed by analysis of their expression profiles, sequence similarity to the host (assessed in the current strategy during initial sequence similarity searches, Fig. 2) and the subcellular localization of the proteins within a cell (Flower 2008; Flower et al. 2010).

In cases where a genome is lacking, transcriptome analyses via DNA microarrays or RNAseq allow for the identification of expression patterns throughout the life cycle of the pathogen/parasite and in combination with proteomics, the associated protein expression levels. To date, the most widely used method for transcriptome analysis is DNA microarray, a high-throughput technique that has been successfully applied in *R. microplus* for investigation of acaricide-induced gene expression in larvae (Saldivar et al. 2008), organ-specific responses to pathogen infection in male adults (Mercado-Curiel et al. 2011), as well as host vector responses in feeding larvae and adult females on different cattle breeds (Rodriguez-Valle et al. 2010). This technique was incorporated into the current strategy for rational vaccine discovery (Fig. 2).

For proteins to be accessible to the host immune system, it is presumed that they are expressed during a reasonable period of the life cycle and that they are secreted or presented on membranes or external surfaces of the parasite. Currently, localization to the cell's surface is thought to be a major determinant of immunogenicity (Flower 2008). As it is known that membrane fractions from *R. microplus*, and not secreted proteins, confer protection to cattle, emphasis was placed on membrane-associated proteins in this study (Opdebeeck et al. 1988; Jackson and Opdebeeck 1990). An obstacle faced during prediction of subcellular

compartmentalization or membrane-association is the under-estimated complexity of the cell, and the lack of tools for the prediction of transient, permanent or multiple localizations, as well as localization to organelles or multi-protein complexes (Flower et al. 2010). To date, SignalP remains one of the best methods for signal sequence prediction, as it uses both neural networks and a hidden Markov model to allow discrimination between uncleaved signal anchors and cleaved signal peptides (Fig. 2) (Flower 2008). The more conventional approach, which is based on the assumption that the subcellular localization of protein homologs is similar, is still used today with success. Briefly, proteins are localized using global homology searches (such as BLAST) or protein motif/family-based identification (such as PROSITE and Pfam) (Sigrist et al. 2010; Punta et al. 2011). In addition to localization, membrane topology also needs to be evaluated for proteins to ensure that exoplasmic regions are selected for subsequent expression of recombinant proteins. In this regard, TMHMM is a powerful program that makes use of a hidden Markov model, with a prediction accuracy of 97-98% for transmembrane helices and the ability to discriminate between soluble and membrane proteins with both specificity and sensitivity better than 99% (Sonnhammer et al. 1998; Krogh et al. 2001).

Although the function a protein performs in the parasite is irrelevant to its status as an antigen (a protein recognized and recalled by the host), targeting a protein of vital function during vaccination remains sensible. To date, proteins involved in adhesion, invasion, secretion, signaling and evading host responses, as well as lipoproteins are regarded as key players in the host-pathogen/host-parasite interface and therefore good vaccine candidates (Vivona et al. 2008). However, as the vast majority of proteins expressed in *R. microplus* remain unannotated, candidate selection based on functional annotation is hampered in the rational design of anti-tick vaccines.

Vaccine antigen contributing factors

Although biotechnology has improved significantly over the past few years, the expression of recombinant protein antigens remains a serious limitation to the production of vaccines. In bacterial pathogens such as *Neisseria meningitidis*, *Streptococcus pneumoniae* and *Porphyromonas gingivalis*, the number of vaccine candidates that can successfully be expressed range between 30-60% of the initially identified number of candidates. Of the latter, only 1-4% of the candidates confer immunity in vivo (Flower et al. 2010). This can in part be attributed to the physicochemical properties and the post-translational modifications of the antigen. Whether peptide or protein, the properties of an antigen result from the primary and secondary protein structure (as well as their associated modifications), thereby influencing its solubility, charge, aggregation status and stability. Of greater importance, is the contribution of these properties to the quality of the immune response, including binding capability of antibodies, dynamics of the peak/priming response and generation of immune memory cells (Flower 2008; Zepp 2010).

Glycosylation is regarded as one of the most important factors when manufacturing vaccines as it is known that polysaccharides can serve as a first signal for B-cell activation. Following internalization of the glycosylated protein, the protein component is presented to T-helper cells, thereby promoting antibody switching from IgM to IgG and generation of memory B-cells (Zepp 2010). Preliminary evidence that protective antigens in *R. microplus* midgut are either glycoproteins, or are dependent on carbohydrates for their specificity came from the study by Lee *et al*., who showed that sodium metaperiodate (periodate) treatment of gut membrane fractions completely abolished their protective ability in cattle (Lee et al. 1991). This was further supported by studies done with Bm86 expressed in *Pichia pastoris* (that allows for glycosylation), where antibodies directed against the carbohydrate determinants of Bm86 were found not to be protective (Willadsen and McKenna 1991;

Garcia-Garcia et al. 1998a; Garcia-Garcia et al. 1998b). Therefore, the true role of glycosylation in tick vaccine efficacy remains to be determined.

Finally, the use of crude extracts or purified recombinant antigen, as well as the choice in adjuvant, can also contribute to the efficacy of the final vaccine formulation (Zepp 2010; Leroux-Roels 2010). Examples of the effect of these factors on immunization of cattle with *R. microplus* antigens have been published (Almazan et al. 2011; Jackson and Opdebeeck 1995).

In this study, we aimed at systematically selecting promising vaccine candidates from *R. microplus* by following a functional genomics and immuno-informatic approach. Our pipeline included the following steps: identification of transcripts with sufficient levels of expression throughout most of the period of infestation, in silico and manual curation of transcripts to determine subcellular localization, ranking of candidates using VaxiJen, evaluation of immunogenicity by means of synthetic peptides and ultimately, future expression of promising candidates for small-scale vaccination trials (Fig. 2). Should they confer protective immunity to *R. microplus*, this pipeline will be applicable for the identification of additional, novel anti-tick vaccine candidates. Combining the latter with candidates that do not confer protection during cattle trials will allow for the improvement of the current strategy.

Materials and Methods

Tick rearing and sample collection

R. microplus (Mozambique strain) larvae, hatched at 25°C (75-85% humidity), were allowed to feed on Holstein-Friesian cattle at the University of Pretoria Biomedical Research

Centre (UPBRC), Onderstepoort veterinary campus (South Africa). Ticks were collected on day 4, 5, 9, 13, 15 and 20 following infestation. Collected life stages were assessed under light microscope and whole ticks, as well as selected immature stages and adult tissues, were snap-frozen in TRI REAGENT® (Molecular Research Center, Inc.) and stored at -70°C. Adult tissues were collected according to the method by Nijhof et al. (Nijhof et al. 2010). Ethical clearance was obtained from the South African Department of Agriculture, Forestry and Fisheries as well as the University of Pretoria's Animal Use and Care Committee (Project approval number EC022-10).

Isolation of total RNA and cDNA synthesis

Total RNA was isolated utilizing manufacturer's guidelines for TRI REAGENT® and RNA purity and integrity were assessed using the Bioanalyzer 2100 micro-fluidics system (Agilent technologies, USA). A reference RNA pool consisting of equivalent amounts of RNA from immature and mature life stages and tissues was prepared. Based on attachment and feeding status, test groups selected to represent each life stage were: larvae (day 4), nymphs (day 9) and tissues (salivary gland, midgut and ovary) collected from partially fed females (day 20). cDNA synthesis was performed using Superscript™ III (Invitrogen™ life technologies, USA), a poly-dT primer (5'-(T)$_{25}$VN-3'; N=ATGC; V=AGC), random nonamers and aminoallyl dUTP for Cy3/Cy5 dye coupling. The cDNA template concentration was determined using the Nanodrop-1000 (Thermo Fisher Scientific, USA) and template was labeled with Cy3 (reference pool) or Cy5 (test sample).

Microarray of R. microplus tissues

Previously, a sequence database consisting of 13 477 contiguous sequences was assembled from available EST data from Genbank (http://www.ncbi.nlm.nih.gov/nucest) and the gene index of *R. microplus* (BmiGI release 2.1) (Wang, Guerrero et al. 2007) using the online bioinformatic tools cd-hit-est (http://www.bioinformatics.org/cd-hit/) and cap3 (http://genome.cs.mtu.edu/cap/cap3.html). The VecScreen tool (http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen_docs.html) was used for detection and removal of vector sequences from EST data prior to assembly. The final sequence dataset was used for array design using the Agilent 8x15k microarray and eArray microarray design platforms (https://earray.chem.agilent.com/earray/). A set of 60 mer probes was designed and probe quality was assessed from base composition scores, in order to select unique probes that were subsequently randomly distributed across the array. The custom array was chemically synthesized by Agilent technologies (USA). An equimolar amount of Cy3-labeled cDNA, from a common reference pool, was hybridized with Cy5-labeled individual test cDNA. Biological and technical replication was employed for each test sample and hybridization was performed at 65°C for 17 hours. Each slide was washed and rinsed in stabilization and drying solution (Agilent Technologies, USA) then dried and scanned with the GenePix^{TM} 4000B microarray laser scanner (Molecular devices Inc., USA).

Microarray data analysis and functional annotation

Fluorescence intensities of Cy3 and Cy5 were extracted using the GenepixPro feature extraction software (v6.0, Axon Molecular Devices, USA) at default parameters. Following visual evaluation of spot quality, normalization within slides was performed using the Limma package in R (http://CRAN.R-project.org), employing the locally weighted scatterplot smoothing (LOWESS) technique, followed by Gquantile normalization between slides. From

the normalized data, the intensity values of each spot in each Cy5-labeled test group for larvae, nymphs, ovaries, midgut and salivary glands were calculated. Based on inspection of spot intensities, a minimum signal intensity threshold of 1000 was chosen for further evaluation of differentially expressed genes. The Limma package was also used to calculate the fold change expression between each group pair-wise comparison, to identify significantly differentially expressed transcripts with p-values adjusted for multiple comparisons false discovery rates. Functional annotation of affected transcripts was performed using the desktop cDNA Annotation System (dCAS, v.1.4.3) (Guo et al. 2009). BLAST searches were performed for each transcript against the following databases: GO, KOG, Mit-Pla, NR, Pfam, RRNA and SMART (http://exon.niaid.nih.gov). For prediction of putative reading frames, Prot4EST was used to obtain the best putative reading frame for all assembled contigs (Wasmuth and Blaxter 2004). Using the immuno-informatic web-based tool, VaxiJen (http://www.ddg-pharmfac.net/vaxijen/vaxijen/vaxijen.html), proteins were ranked according to their likelihood of being protective antigens. Bm86, the only commercially available anti-tick vaccine, obtained a score of 0.7698, therefore only proteins with a VaxiJen score of above 0.7 were considered. For prediction of membrane spanning transcripts the TMHMM web server (http://www.cbs.dtu.dk/services/TMHMM/, v.2.0) was used and secreted transcripts were identified using the SigP server (http://www.cbs.dtu.dk/services/SignalP-3.0/, v.3.0). Final classification of transcripts was as either non-secreted (including truncated transcripts lacking identifiable signal sequences), membrane-associated and secreted soluble transcripts, based on their SigP and TMHMM results. For the identification of potentially GPI-anchored proteins, a GPI anchoring site had to be predicted by at least 2 of the following four programs: PredGPI (http://gpcr.biocomp.unibo.it/predgpi/pred.htm), GPI-SOM (http://gpi.unibe.ch/), MemType2L (http://www.csbio.sjtu.edu.cn/bioinf/MemType/) and BigPI

(http://mendel.imp.ac.at/gpi/gpi_server.html). Array validation via qPCR and full annotation analysis has been submitted for publication elsewhere (Stutzer et al, submitted).

Synthetic peptide design

In order to evaluate the proposed reverse vaccinology strategy, transcripts were selected that were predicted to membrane-associated proteins. Six programs were used to predict antigenic sites (i.e. epitopes) and consensus areas were chosen for peptide design. These programs include: the method developed by Kolaskar and Tongaonkar that identifies epitopes based on the amino acid physicochemical properties, the method developed by Emini that identifies surface exposed regions in the protein and Bepipred that identifies linear B-cell epitopes. These programs are part of a suite of predictors that are available on-line (http://tools.immuneepitope.org/tools/bcell/iedb_input). Additional MHCII binding antigen predictors MHCPred (http://ddg-pharmfac.net/MHCPred) and ProPred (http://www.imtech.res.in/raghava/propred) were used to predict the top antigenic regions of the selected membrane-associated proteins. Finally, secondary structures that play a role in antigenicity, such as beta-turns, were identified using the online predictor BetaTurns (http://www.imtech.res.in/raghava/betaturns). Following epitope prediction, peptides of at least 9 amino acids in length were synthesized by GenScript (USA). As previously identified antigenic regions were available for Bm86, these were included as positive controls (Patarroyo et al. 2002; Patarroyo et al. 2009; Freeman et al. 2010). Therefore, using sequence information for the Mozambique strain of *R. microplus*, peptides were synthesized: SSVCSDFGNEFCRNA (Peptide 1), CDCGEWGAMNKTTR (Peptide 2) and LSKHVLRKLQACEH (Peptide 3). A C-terminal linker sequence (GGGC) was added to

each peptide for potential conjugation to carrier proteins such as BSA or KLH (Keyhole limpet hemocyanin).

Isolation of R. microplus midgut membrane proteins and immunization of BALB/c mice

Twenty replete female *R. microplus* ticks were dissected and their midguts were removed and stored in 10 mM phosphate-buffered saline (PBS, 150 mM NaCl, pH 7.4) and protease inhibitor cocktail (Sigma, MO, USA). Tissues were homogenized via mechanical shearing using needles and pulse-sonicated ten times using a Virsonic^{TM} sonifier (1 s pulse, 1 s rest) at 3 W output. Samples were subsequently centrifuged at 100,000 g for 1 hour after which the pellet was washed with PBS and centrifuged again at 100,000 g for 1 hour. The resultant supernatant was discarded and the pellet finally resuspended in PBS. Protein concentration was measured using the Nanodrop-1000 (Thermo Fisher Scientific, USA).For the production of antisera, 3 six-week old BALB/c female mice were used. Each mouse was immunized subcutaneously with 100 μg *R. microplus* gut membrane protein in PBS mixed with Montanide ISv50 adjuvant (1:1 v/v) (SEPPIC, France) on weeks 1, 4 and 6. Prior to the immunizations, naive sera were collected and stored at -70°C. Animals were sacrificed on week 7 for complete blood collection and sera were stored at -70°C. Ethical clearance was obtained from the South African Department of Agriculture, Forestry and Fisheries as well as the University of Pretoria's Animal Use and Care Committee (Project approval number EC022-10).

ELISA of polyclonal antisera using synthetic peptides

An enzyme-linked immunosorbent assay was performed in order to measure the reactivity of the pooled polyclonal antisera against the predicted antigenic synthetic peptides. The lyophilized peptides were dissolved in Tris-buffered saline (TBS, 25 mM Tris–HCl, 150 mM NaCl, and pH 7.4) and in the case of acidic or basic peptides that did not dissolve readily, 100 mM NaOH and HCl were added respectively. Experiments were performed in quadruplicate and each well was loaded with 13 nmol of the appropriate peptide. The plate was dried to ensure complete adsorption of the peptides and subsequently blocked overnight at 4 °C using TBSC (25 mM Tris–HCl, 150 mM NaCl, pH 7.4, 0.5% w/v casein). The plate was then washed four times using TBSC and 50 µl of 1:25 pooled antisera were added to each well. The plate was incubated for 1h at room temperature, washed four times with TBSC and incubated for 45 min at room temperature with 1:1000 horseradish peroxidase-conjugated goat α-murine IgG (Sigma, MO, USA). Following a final washing step, developing buffer (10 ml citrate, 10 mg O-phenylene diamine and 8 mg H_2O_2, pH 4.5) was added and the reaction was monitored at 450 nm using the Multiscan Plus reader (Thermo, France). Pooled pre-immune sera were used as negative controls and these were performed in duplicate.

Results

Gene expression and immuno-informatic analysis of immature life stages and adult tissues of R. microplus ticks

From the Harvard gene index project site (De Miranda Santos, Valenzuela et al. 2004) and Genbank nucleotide sequence database (http://www.ncbi.nlm.nih.gov/nucest), some 60

000 ESTs and 13 643 singleton *R. microplus* sequences were downloaded and assembled, obtaining 13 477 contiguous sequences used for microarray chip design, though intensity values for 13 456 were finally calculated (data not shown).

The overall distribution of transcripts showed that the majority of genes regulated above an arbitrary intensity threshold were shared among all the life stages and tissues (Fig. 3, Table 1). Interestingly, the second largest set of transcripts identified was uniquely

Table 1. Summary of predicted antigenic transcripts for selected tissue and life stage comparisons. Indicated is the number of transcripts identified from microarray analysis, as well as the number of transcripts that are predicted antigens with their predicted cellular fate.

Tissue and life stage comparison	Total number of transcripts[a]	VaxiJen score >0.7[b]	Non-secreted intracellular[c]	Membrane-associated[d]	Secreted soluble[e]
All life stages	3135	566	398	130	38
L+ all tissues	187	33	23	8	2
N+ all tissues	183	20	15	3	2
L+G+SG	86	20	16	3	1
L+G+O	26	3	3	0	0
L+O+SG	67	10	5	5	0
N+G+SG	81	15	14	1	0
N+G+O	73	11	9	1	1
N+O+SG	64	15	8	1	1
L+G	44	9	7	2	0
L+O	55	9	5	3	1
L+SG	94	24	14	5	33
N+G	103	15	8	6	4
N+O	124	21	16	2	2
N+SG	72	20	1	6	1
Total [f]	4394	791	542	176	86

[a] Total number of transcripts expressed above an intensity threshold of 1000

[b] Total number of transcripts predicted to have a protective antigen probability score higher than 0.7 using the VaxiJen server.

expressed in ovaries. Considering transcripts that were up-regulated in at least one immature life stage and an adult tissue, expression analysis indicated that 33% of the total complement of sequences was regulated above the intensity threshold (Fig. 3, Table 1). These transcripts were subjected to further immuno-informatic analysis using the VaxiJen web server and 791 transcripts were predicted to be protective antigens (Table 1).
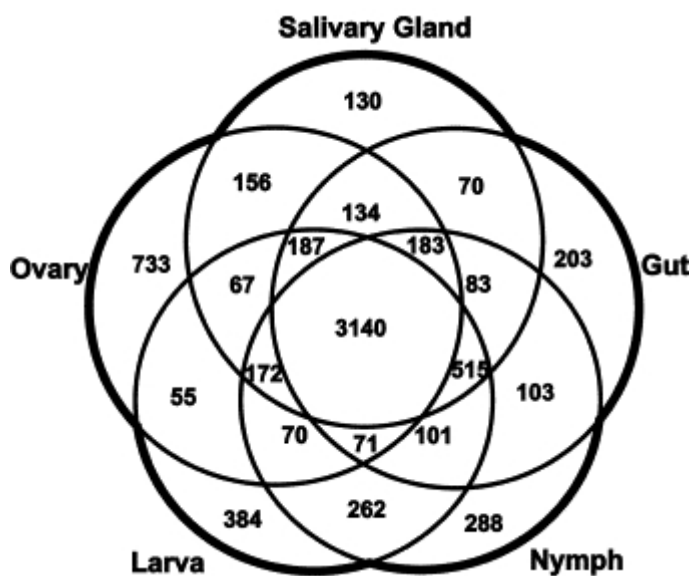

Fig. 3: Distribution of transcripts regulated between immature life stages and adult tissues of _Rhipicephalus microplus_ ticks. Venn diagram indicating the number of genes that are unique to or shared between larvae, nymphs, salivary glands, midguts and ovaries above threshold.

The comparison that contained the largest subset of probable antigens comprised of 566 transcripts that were shared between all life stages and tissues (Table 1). Sequence
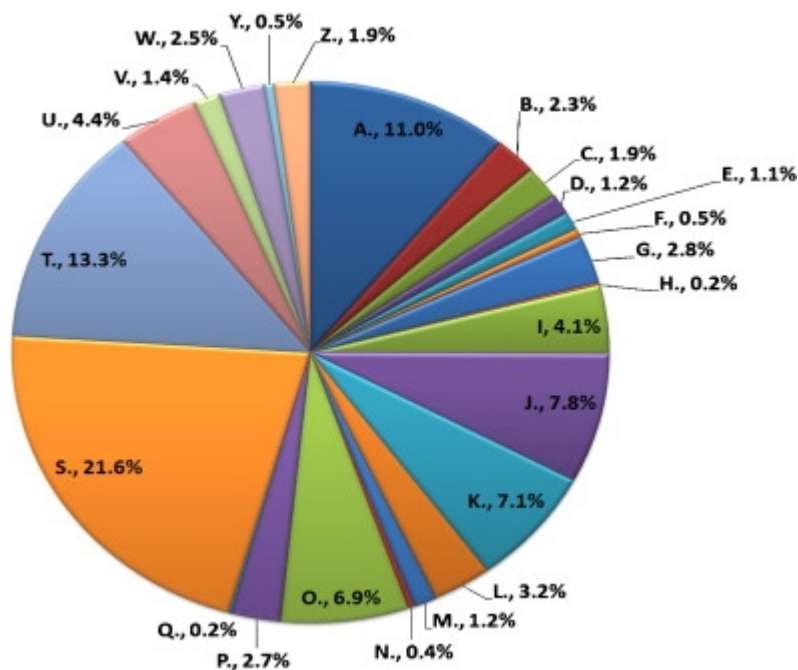

Fig. 4. Functional distribution of predicted antigenic transcripts shared between immature life stages and adult tissues of *Rhipicephalus microplus* ticks. Indicated are the percentages of genes shared between all immature life stages and adult tissues above threshold, with a VaxiJen score greater than 0.7. Transcripts are classified according to their eukaryotic orthologous functional groups (KOGs): **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 and metabolism; **Q-** Secondary metabolites biosynthesis, transport and catabolism; **S-** Function unknown (also includes transcripts with only general functional predictions); **T-** Signal transduction mechanisms; **U-** Intracellular trafficking, secretion, and vesicular transport; **V-** Defense mechanisms; **W-** Extracellular structures; **Y-** Nuclear structure; **Z-** Cytoskeleton.

annotation and functional classification according to the eukaryotic gene ontology (KOG), indicated that almost 22% of the antigenic transcripts could not be functionally annotated (Fig. 4). Some of the major functional classes that were represented included: transcripts involved in RNA modification and processing (11%); translation, ribosomal structure and biogenesis (7.8%); transcription (7.1%) and posttranslational modification, protein turnover, chaperones (6.9%). The greatest compliment of annotatable transcripts identified (13.3%) relate to signal transduction mechanisms (Fig. 4). The latter highlights the essential role signal transduction plays in basic metabolism and cellular function. Two smaller classes that were identified contained transcripts related lipid transport and metabolism (4.1%), as well as intracellular trafficking (4.4%) (Fig. 4).

Analysis of the topology of the 566 predicted protective antigens that are expressed in all life stages and adult tissues showed that 398 transcripts had no identifiable membrane spanning regions or signal sequences for secretion (Table 1). However, 130 transcripts could be identified that showed membrane topology in regards to TMHMM and GPI analysis. Analysis of the predicted antigenicity scores of these transcripts showed that more than 50% obtained a similar or better score than that predicted for Bm86 at 0.7698 (results not shown). Following signal peptide prediction, 89 transcripts were shown to be putatively secreted (results not shown). Of the latter, only 38 transcripts were identified that had no predicted membrane localization for this comparison (Table 1). An additional 20 membrane proteins were identified from the various comparisons with the nymphal immature life stage. In contrast, 26 additional transcripts were identified as membrane-associated from the remaining comparisons with larvae (Table 1). Overall, 176 putative membrane-associated proteins were identified, with or without identifiable signal peptides, for all the comparisons considered (Table 1).

Table 2: Selected properties of synthetic peptides. Indicated are the antigens represented by each peptide, the peptide's size, charge and composition as percentage hydrophilic, hydrophobic and neutral amino acids.

Antigen	Peptide number	Length	Nett charge[a]	Amino acid composition (%)		
				Hydrophilic	Hydrophobic	Neutral
Bm86	1	19	-1	16	16	68
	2	18	0	22	11	67
	3	18	4	33	22	44
Antigen 1	1	12	2	17	8	75
Antigen 2	1	22	2	9	32	59
	2	19	2	11	16	74
Antigen 3	1	18	1	28	28	44
Antigen 4	1	20	2	20	40	40
	2	19	-1	37	21	42
	3	17	3	18	29	53
Antigen 5	1	16	0	38	12	50

[a] Nett charge based on amino acid composition at neutral pH

Evaluation of predicted membrane antigen proteins using synthetic peptides

In order to evaluate the current methodology, five predicted membrane-associated antigenic proteins were selected and representative synthetic peptides designed using bioinformatic tools. The chemical properties of these peptides are shown in table 2. By performing a multiple comparison ANOVA (Holm-Sidak method), it can be seen that three peptides (peptide 1 from Antigens 2, 3 and 4) showed better recognition (p-value <0.001) by polyclonal antisera from BALB/c mice immunized with a crude extract of tick midgut membrane proteins, compared to peptide Bm86-2 (Fig. 5). Apart from peptide 2 from antigen 4, which showed very little recognition by the antisera, all other peptides showed statistically similar recognition compared to the three Bm86 peptides.
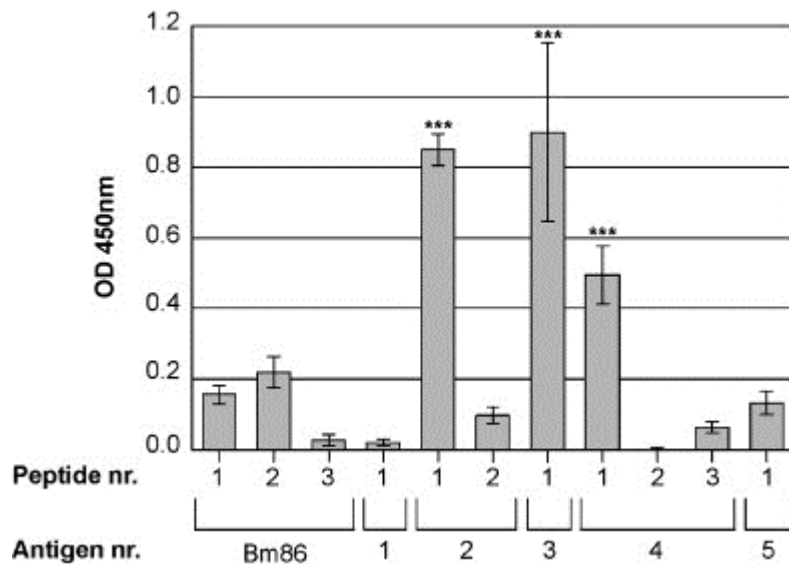
Fig. 5. Evaluation of the reactivity of selected membrane proteins using antisera from mice challenged with crude midgut membrane proteins. Indicated are ELISA results obtained for several peptides representing predicted antigens (1-5) using antisera from BALB/c mice immunized with a crude *R. microplus* gut membrane extract. Asterisks indicate a significant difference (P value < 0.001) between results obtained for predicted antigen peptides compared to previously validated Bm86 epitopes that was used as positive controls.

Discussion

Vaccine design in the post-genomic era holds the promise of being able to combine genomic and transcriptome data with in silico approaches, in order to predict novel potential vaccine candidates. Considering the vast amount of transposons and retro-transposable elements in the genome of *R. microplus* (Guerrero et al. 2010), as well as inter-strain sequence variation and differential control of gene expression (Kamau et al. 2011), acquired resistance to protein-based vaccines will be a significant obstacle in the foreseeable future. This has already been reported in the field with Bm86 resistant tick strains (De La Fuente et al. 2000). This can likely be ascribed to either antigen variability that occurred via antigenic drift (the acquisition of point mutations during replication), major

recombination/reassortment of genetic material between related strains (antigenic shift), differential gene expression or a combination of the latter.

With the expansion of databases containing information on validated antigens and an improved understanding of host immunology, the strategy of reverse vaccinology is continually being improved. In the field of anti-tick vaccine development, not all of the available tools are equally applicable, as they are predominantly based on human vaccines against bacterial, viral and intracellular pathogens. Moreover, the tick transcriptome remains largely un-annotatable and the current transcriptomic data may contain a vast amount of truncated transcripts. This could affect the predictability of protein localization and topology, as membrane-associated proteins could be predicted as soluble due to a lack of identifiable membrane sequences or signal peptides. Similarly, secreted soluble proteins could be predicted as non-secreted intracellular proteins (Table 1). Finally, proteomic data to verify the encoded open reading frames and subcellular localization of proteins is currently insufficient. In this regard, the newly established Cattle Tick Database will be an invaluable resource, as the basis for a systematic attempt at annotating the full complement of genes and proteins of *R. microplus* (Bellgard et al., 2012).

Despite these obstacles, we describe the identification of some 791 promising vaccine candidates that were selected based on the criteria of being expressed in all out the life stages of *R. microplus* at levels comparable to that of Bm86. The candidates obtained similar scores in VaxiJen, an alignment-independent computational tool that has been successfully applied in the prediction of other parasitic antigens. However, the predicted antigens should also be available for immune surveillance, thereby giving preference to membrane- and secreted antigens. The approach for predicting subcellular localization in combination with VaxiJen also resulted in the inclusion of Bm86, shown in literature to be membrane-bound and to confer protective immunity (Rand et al. 1989). As most of the predicted membrane associated

vaccine candidates function in vital biological processes such as signal transduction, trafficking and transport (Fig. 4), targeting these antigens may have significant downstream effects resulting in severe impairment of biological functions in the tick. In addition, immunization against membrane proteins also offers the possibility to interfere with the tick-pathogen interface. Therefore, in order to fully exploit these proteins and the various processes they mediate as tick control targets, an understanding of their function and mode of action is vital. Interactome studies using conventional two-hybrid systems and/or proteomics approaches are indispensable tools for enhancing rational design of vaccines (Fig. 2). Targeting epitopes that are masked as a result of complex formation in vivo may be futile, whereas targeting protein-interacting domains may disrupt interactions.

For the preliminary evaluation of the current approach, several predicted membrane-associated antigens were selected and the reactivity of representative epitopes was assessed using antisera from mice immunized with *R. microplus* midgut membrane proteins. This lead to the identification of 3 epitopes with higher recognition by the antisera than those previously confirmed for Bm86 (Fig. 5). However, antigenicity alone does not guarantee that immunization with a particular protein will in fact confer anti-tick resistance in cattle trials. A good example is 5'-nucleotidase isolated from *R. microplus* midgut fractions, which strongly elicited an antibody response in sheep and cattle, but did not confer significant protection in cattle following tick challenge (Hope et al. 2010). It is interesting to note that the VaxiJen score obtained for this protein was lower than the threshold value of 0.7. As no significant differences in antisera recognition of the other peptides (compared to the three Bm86 controls) were observed, the antigens they represent cannot unequivocally be disregarded as potential vaccine candidates.

In this study, a total of 176 membrane-associated and 86 secreted soluble proteins were identified of which the most promising candidates are currently being expressed for

vaccine trials. Both protective and non-protective antigens, evaluated during vaccination trials, will be valuable in improving the predictive potential of VaxiJen and by extension the current strategy (Fig. 6). If successful, this approach will offer a pipeline for the identification of additional anti-tick vaccine candidates.
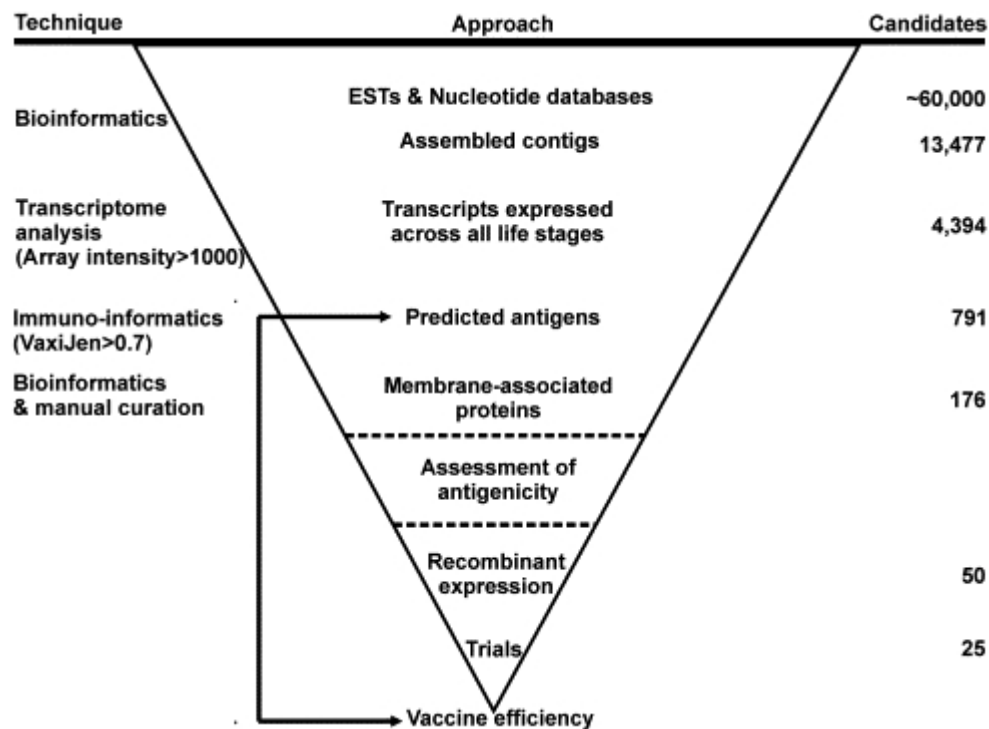

Fig. 6. Systematic analysis and evaluation of transcripts regulated between immature life stages and adult tissues of *Rhipicephalus microplus* ticks. Indicated is the systematic evaluation of transcriptomic data for the identification of potential vaccine candidates using a combination of bioinformatic and immuno-informatic approaches. Indicated by the black arrow, trial data will be used to expand the available antigen database, as well as improve the current reverse vaccinology approach.

Acknowledgements

UP-BRC) for cattle management, Professor Fourie Joubert (UP, Department of Bioinformatics) for bioinformatic support, as well as Professor Dave Berger and Mr Nicky Olivier (UP, Department of Genetics) for advice and technical assistance during microarray analysis.

References

Almazan, C., Moreno-Cantu, O., Moreno-Cid, J.A., Galindo, R.C., Canales, M., Villar, M., de la Fuente, J., 2011. Control of tick infestations in cattle vaccinated with bacterial membranes containing surface-exposed tick protective antigens. Vaccine doi:10.1016/j.vaccine.2011.10.102

Bellgard, M.I., Moolhuijzen, P.M., Guerrero, F.D., Schibeci, D., Rodriguez-Valle, M., Peterson, D.G., Dowd, S.E., Barrero, R., Hunter, A., Miller, R.J., Lew-Tabor, A.E., 2011. CattleTickBase: An integrated Internet-based bioinformatics resource for *Rhipicephalus (Boophilus) microplus*. Int J Parasitol doi:10.1016/j.ijpara.2011.11.006

De la Fuente, J., Rodriguez, M., Garcia-Garcia, J.C., 2000. Immunological control of ticks through vaccination with *Boophilus microplus* gut antigens. Ann N Y Acad Sci 916: 617-621.

De la Fuente, J., Almazán, C., Canales, M., Pérez de la Lastra, J.M., Kocan, K.M., Willadsen, P., 2007. A ten-year review of commercial vaccine performance for control of tick infestations on cattle. Anim Health Res Rev 8, 1: 23-8.

De Miranda Santos, I.K.F., Valenzuela, J. G., Ribeiro, J.M., de Castro, M., Costa, J.N., Costa, A.M., da Silva, E.R., Neto, O.B., Rocha, C., Daffre, S., Ferreira, B.R., da Silva, J.S., Szabo, M.P., Bechara, G.H., 2004. Gene discovery in *Boophilus microplus*, the cattle

tick: the transcriptomes of ovaries, salivary glands, and hemocytes. Ann N Y Acad Sci 1026: 242-246.

Doytchinova, I.A., Flower, D.R., 2007a. Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. Vaccine 25, 5: 856-866.

Doytchinova, I.A., Flower, D.R., 2007b. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics 8: 4.

Flower, D.R., 2008. Bioinformatics for Vaccinology. West Sussex, UK., Wiley-Blackwell.

Flower, D.R., Macdonald, I.K., Ramakrishnan, K., Davies, M.N., Doytchinova, I.A., 2010. Computer aided selection of candidate vaccine antigens. Immunome Res 6 Suppl 2: S1.

Freeman, J.M., Davey, R.B., Kappmeyer, L.S., Kammlah, D.M., Olafson, P.U., 2010. Bm86 midgut protein sequence variation in South Texas cattle fever ticks. Parasit Vectors 3: 101.

Garcia-Garcia, J.C., Montero, C., Rodríguez, M., Soto, A., Redondo, M., Valdes, M., Mendez, L., de la Fuente, J., 1998a. Effect of particulation on the immunogenic and protective properties of the recombinant Bm86 antigen expressed in Pichia pastoris. Vaccine 16, 4: 374-380.

Garcia-Garcia, J.C., Soto, A., Nigro, F., Mazza, M., Joglar, M., Hechevarria, M., Lamberti, J., de la Fuente, J., 1998b. Adjuvant and immunostimulating properties of the recombinant Bm86 protein expressed in *Pichia pastoris*. Vaccine 16, 9-10: 1053-1055.

Guerrero, F.D., Moolhuijzen, P., Peterson, D.G., Bidwell, S., Caler, E., Bellgard, M., Nene, V.M., Djikeng, A., 2010. Reassociation kinetics-based approach for partial genome

sequencing of the cattle tick, *Rhipicephalus*, *Boophilus microplus*. BMC Genomics 11: 374.

Guo, Y., Ribeiro, J.M., Anderson, J.M., Bour, S., 2009. dCAS: a desktop application for cDNA sequence annotation. Bioinformatics 25, 9: 1195-1196.

Hope M, Jiang X, Gough J, Cadogan L, Josh, P., Jonsson, N., Willadsen, P., 2010. Experimental vaccination of sheep and cattle against tick infestation using recombinant 5'-nucleotidase. Parasite Immunol 32, 2: 135-142.

Jackson, L.A., Opdebeeck, J.P., 1990. Humoral immune responses of Hereford cattle vaccinated with midgut antigens of the cattle tick, *Boophilus microplus*. Parasite Immunol 12, 2: 141-151.

Jackson, L.A., Opdebeeck, J.P., 1995. The effect of various adjuvants on the humoral immune response of sheep and cattle to soluble and membrane midgut antigens of *Boophilus microplus*. Vet Parasitol 58, 1-2: 129-141.

Kamau, L., Skilton, R.A., Odongo, D.O., Mwaura, S., Githaka, N., Kanduma, E., Obura, M., Kabiru, E., Orago, A., Musoke, A., Bishop, R.P., 2011. Differential transcription of two highly divergent gut-expressed Bm86 antigen gene homologues in the tick *Rhipicephalus appendiculatus*, Acari: Ixodida. Insect Mol Biol 20, 1: 105-114.

Kemp, D.H., Pearson, R.D., Gough, J.M., Willadsen, P., 1989. Vaccination against *Boophilus microplus*: localization of antigens on tick gut cells and their interaction with the host immune system. Exp Appl Acarol 7, 1: 43-58.

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305, 3: 567-580.

Lee, R.P., Jackson, L.A., Opdebeeck, J.P., 1991. Immune responses of cattle to biochemically modified antigens from the midgut of the cattle tick, *Boophilus microplus*. Parasite Immunol 13, 6: 661-672.

Leroux-Roels, G., 2010. Unmet needs in modern vaccinology: adjuvants to improve the immune response. Vaccine 28 Suppl 3: C25-36.

Mercado-Curiel, R.F., Palmer, G.H., Guerrero, F.D., Brayton, K.A., 2011. Temporal characterisation of the organ-specific *Rhipicephalus microplus* transcriptional response to *Anaplasma marginale* infection. Int J Parasitol 41, 8: 851-860.

Moriel, D.G., Scarselli, M., Serino, L., Mora, M., Rappuoli, R., Masignani, V., 2008. Genome-based vaccine development: a short cut for the future. Hum Vaccin 4, 3: 184-188.

Nijhof, A.M., Balk, J.A., Postigo, M., Rhebergen, A.M., Taoufik, A., Jongejan, F., 2010. Bm86 homologues and novel ATAQ proteins with multiple epidermal growth factor (EGF)-like domains from hard and soft ticks. Int J Parasitol 40, 14: 1587-97.

Opdebeeck, J.P., Wong, J.Y., Jackson, L.A., Dobson, C., 1988. Hereford cattle immunized and protected against *Boophilus microplus* with soluble and membrane-associated antigens from the midgut of ticks. Parasite Immunol 10, 4: 405-410.

Patarroyo, J.H., Portela, R.W., De Castro, R.O., Pimentel, J.C., Guzman, F., Patarroyo, M.E., Vargas, M.I., Prates, A.A., Mendes, M.A., 2002. Immunization of cattle with synthetic peptides derived from the *Boophilus microplus* gut protein (Bm86). Vet Immunol Immunopathol 88, 3-4:163-72.

Patarroyo, J.H., Vargas, M.I., Gonzalez, C.Z., Guzman, F., Martins-Filho, O.A., Afonso, L.C., Valente, F.L., Peconick, A.P., Marciano, A.P., Patarroyo, A.M. 5th, Sossai, S., 2009. Immune response of bovines stimulated by synthetic vaccine SBm7462 against *Rhipicephalus*, *Boophilus microplus*. Vet Parasitol 166, 3-4: 333-339.

Piper, E.K., Jonsson, N.N., Gondro, C., Lew-Tabor, A.E., Moolhuijzen, P., Vance, M.E., Jackson, L.A., 2009. Immunological profiles of *Bos taurus* and *Bos indicus* cattle infested with the cattle tick, *Rhipicephalus*, *Boophilus microplus*. Clin Vaccine Immunol 16, 7: 1074-1086.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E.L., Eddy, S.R., Bateman, A., Finn, R.D., 2011. The Pfam protein families database. Nucleic Acids Res: 1-12.

Rand, K.N., Moore, T., Sriskantha, A., Spring, K., Tellam, R., Willadsen, P., Cobon, G.S., 1989. Cloning and expression of a protective antigen from the cattle tick *Boophilus microplus*. Proc Natl Acad Sci U S A 86, 24: 9657-9661.

Rappuoli, R., Bagnoli, F., 2011. Vaccine design: Innovative approaches and novel strategies. Norfolk, UK, Caister Academic Press.

Rodriguez-Valle, M., Lew-Tabor, A., Gondro, C., Moolhuijzen, P., Vance, M., Guerrero, F.D., Bellgard, M., Jorgensen, W., 2010. Comparative microarray analysis of *Rhipicephalus, Boophilus microplus* expression profiles of larvae pre-attachment and feeding adult female stages on *Bos indicus* and *Bos taurus* cattle. BMC Genomics 11.

Saldivar, L., Guerrero, F.D., Miller, R.J., Bendele, K.G., Gondro, C., Brayton, K.A., 2008. Microarray analysis of acaricide-inducible gene expression in the southern cattle tick, *Rhipicephalus, Boophilus microplus*. Insect Mol Biol 17, 6: 597-606.

Sigrist, C.J., Cerutti, L., de Castro, E., Langendijk-Genevaux, P.S., Bulliard, V., Bairoch, A., Hulo, N., 2010. PROSITE, a protein domain database for functional characterization and annotation. Nucleic Acids Res 38: D161-166.

Sonnhammer, E.L., von Heijne, G., Krogh, A., 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol 6: 175-182.

Vargas, M., Montero, C., Sanchez, D., Perez, D., Valdes, M., Alfonso, A., Joglar, M., Machado, H., Rodriguez, E., Mendez, L., Lleonart, R., Suarez, M., Fernandez, E., Estrada, M.P., Rodriguez-Mallon, A., Farnos, O., 2010. Two initial vaccinations with the Bm86-based Gavacplus vaccine against *Rhipicephalus*, *Boophilus microplus* induce similar reproductive suppression to three initial vaccinations under production conditions. BMC Vet Res 6: 43.

Vivona, S., Gardy, J.L., Ramachandran, S., Brinkman, F.S., Raghava, G.P., Flower, D.R., Filippini, F., 2008. Computer-aided biotechnology: from immuno-informatics to reverse vaccinology. Trends Biotechnol 26, 4: 190-200.

Wang, M., Guerrero, F.D., Pertea, G., Nene, V.M., 2007. Global comparative analysis of ESTs from the southern cattle tick, *Rhipicephalus*, *Boophilus microplus*. BMC Genomics 8: 368.

Wasmuth, J.D., Blaxter, M.L., 2004. prot4EST: translating expressed sequence tags from neglected genomes. BMC Bioinformatics 5: 187.

Willadsen, P., McKenna, R.V., 1991. Vaccination with 'concealed' antigens: myth or reality? Parasite Immunol 13, 6: 605-616.

Zepp, F., 2010. Principles of vaccine design-Lessons from nature. Vaccine 28 Suppl 3: C14-24.