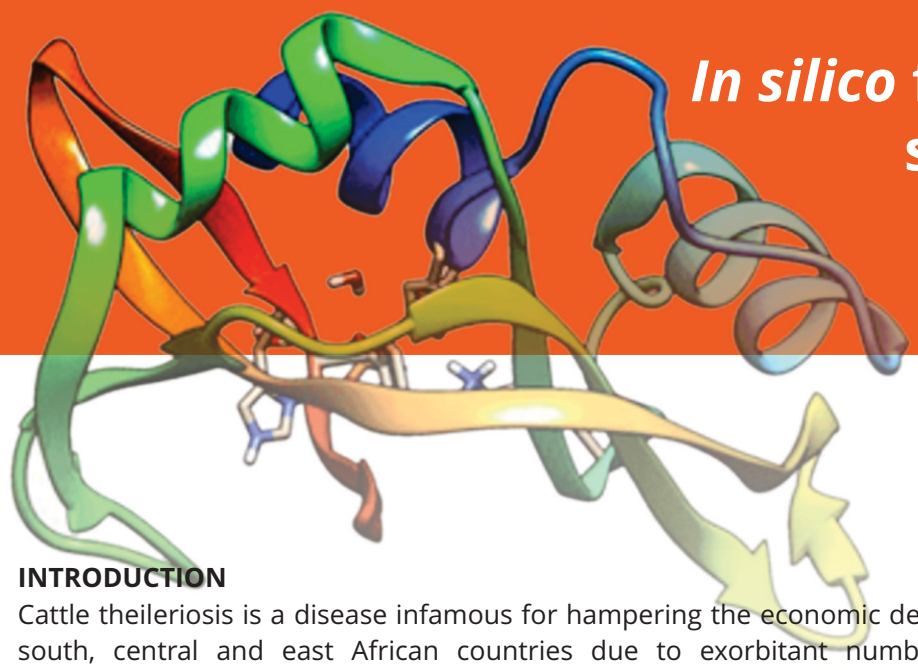

In silico functional prediction and characterization of selected *Theileria parva* hypothetical proteins

Bongiwe P. Mahlobo[1], Fortunate Mokoena[2], Tshepo Matjila[1] and Kgomotso P. Sibeko-Matjila[1]
[1]Department of Veterinary Tropical Diseases, University of Pretoria, Pretoria, South Africa,
[2]Department of Life and Consumer Sciences, University of South Africa, Florida, South Africa

INTRODUCTION

Cattle theileriosis is a disease infamous for hampering the economic development of south, central and east African countries due to exorbitant numbers of cattle mortalities [1]. The disease is caused by *Theileria parva*, a tick-transmitted hemoprotozoan parasite that belongs to the phylum Apicomplexa [2]. Infection of cattle with cattle-derived *T. parva* isolates is responsible for East Coast fever while infections by buffalo-derived isolates results in Corridor disease. However, these causative agents bear similar morphological and serological characteristics and it is not clear why they cause different disease syndromes. Thus, a transcriptome study comparing the cattle-derived and buffalo-derived parasite isolates was performed to detect differentially expressed transcripts (DETs). Identified DETs were found to largely consist of genes encode hypothetical proteins (HPs) (54.4%). These proteins are believed to be crucial in understanding the diseases caused by *T. parva* infections; therefore functional annotation of these proteins is critical.

AIM

The aim of the study was to annotate functions of selected *T. parva* hypothetical proteins encoded by differentially expressed genes, using *in silico* methods.

METHOD/ APPROACH

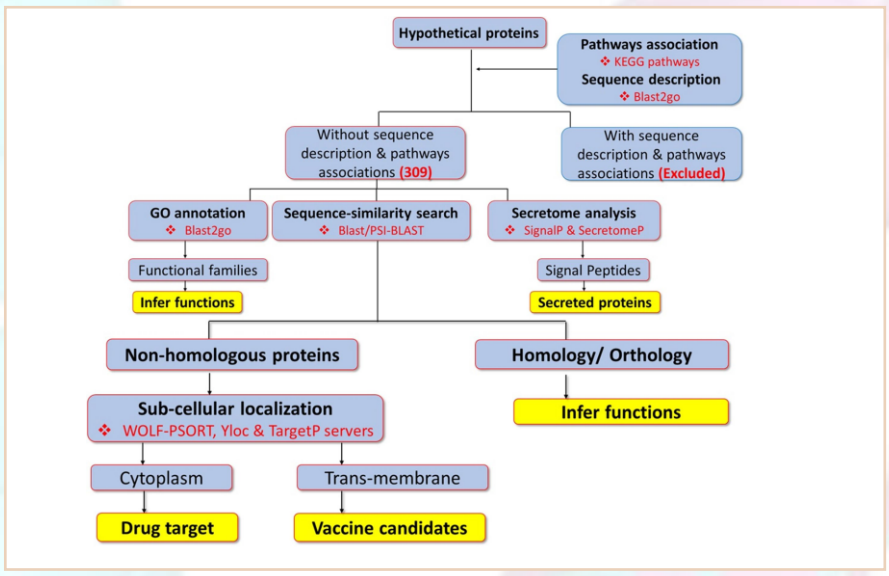

Fig 1: The computational framework adopted for the functional annotation of selected *T. parva* HPs

RESULTS AND DISCUSSION

Classification of hypothetical proteins into functional families **(Fig 2)**

Proteins are allocated to the same superfamily/ family provided they share end-to-end sequence similarity and comprising similar domain architecture[3].

- Forty-three HPs were characterized into seven canonical protein families; most HPs are associated with binding proteins, catalytic activity and transcription factor activity.
- From binding proteins, four were shown to have a Zinc finger domain, and thus may be involved in regulation of apoptosis. Apoptosis contributes to the pathogenesis of a number of diseases[4].
- Five HPs were classified as enzymes. Enzymes facilitate the parasite's survival within the host by carrying out several cellular processes making it viable for the course of infection within the host.

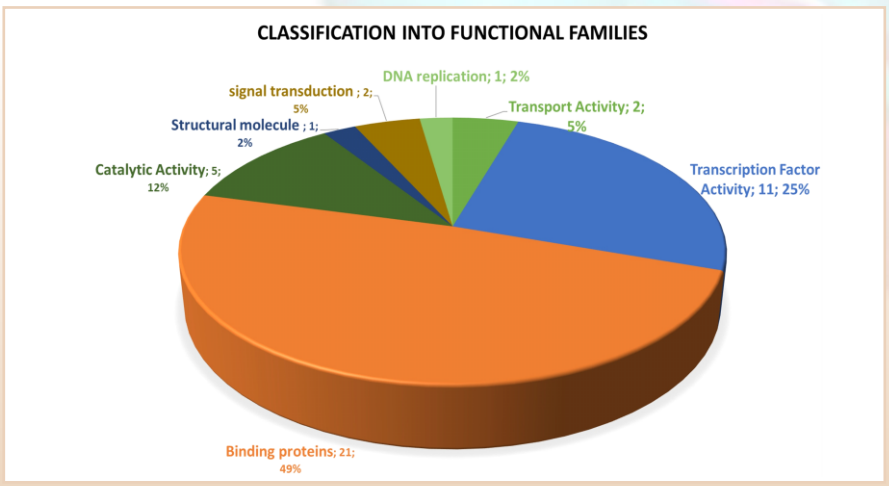

Fig 2. Classification of the 309 HPs into their canonical protein families.

Orthology analysis (Fig 3)

The foundation of function annotation based on orthology is derived from the notion that proteins with similar sequences hint similar functions[5].

- Twenty-two (22) HPs were shown to be orthologous to *T. annulata* proteins belonging to the subtelomeric variable secreted proteins (SVSP) family.
- The SVSPs family is characterized by Nuclear Localization Signal (NLS) identified in 18 HPs, signal peptides detected in 17 HPs and FAINT domain found in 20 HPs. Proteins with NLSs are important as they are likely to contribute to the phenotypic changes of the host cell.

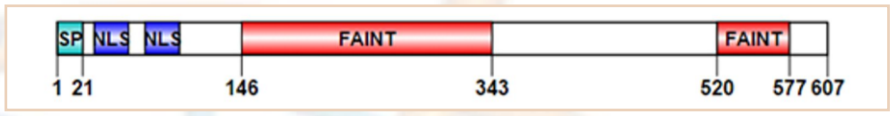

Fig 3. Schematic presentation of TP08_0882, a typical SVSP. This polypeptide of 607 amino acids has a signal peptide (SP; from residue 1-21) responsible for secretion, followed by two nuclear localisation signals (NLS) for protein transportation to the nucleus and lastly two FAINT domains (from residue 146-343 and 520-577).

Homology analysis (Fig 4)

T. parva HPs had homologs, mostly in *T. equi*, *H. sapiens* and *M. muscus.* One homolog of interest was detected in *N. caninum,* belonging to the Acetyltransferase family protein (TP01_0669), known to be involved in virulence-associated functional roles such as invasion, colonization, evasion of host defence and immunomodulation[6,7].

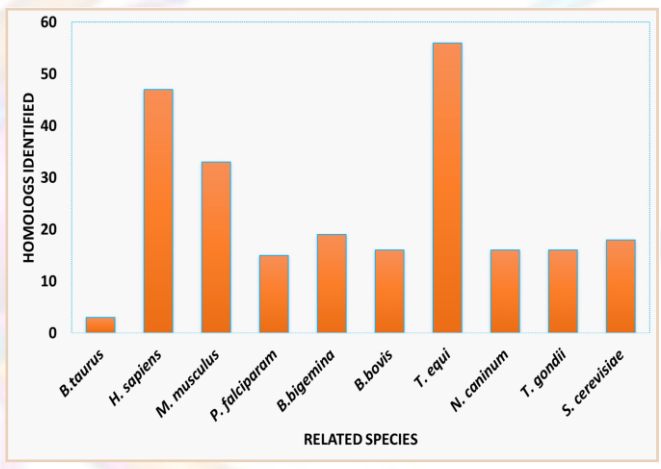

Fig 4. Hypothetical proteins distribution in different related species.

Subcellular localization predictions (Fig 5)

Information obtained from protein sub-cellular localization prediction can be used to infer protein functions and to find novel vaccine and/or drug targets [8]. Three HPs predicted to localize in the cytoplasm were identified as possible therapeutic targets.

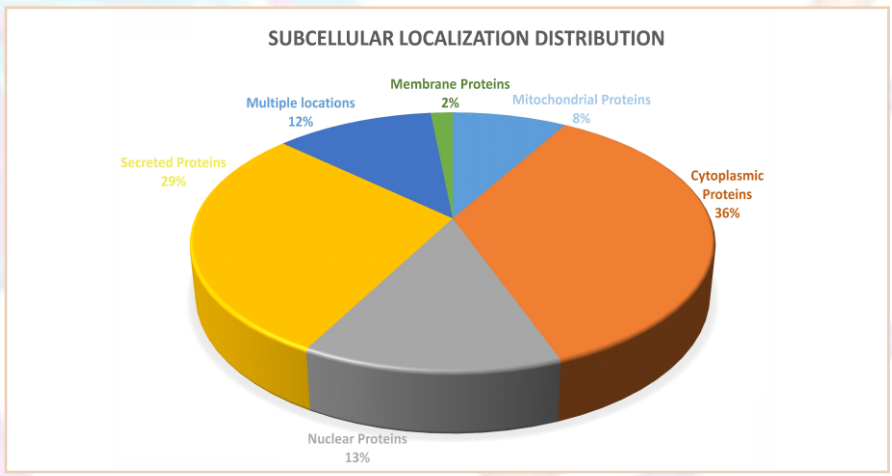

Fig 5. Subcellular localization of all the 309 *T. parva* hypothetical proteins.

CONCLUSION

Using *in silico* approaches, 266 of the 309 *T. parva* HPs investigated were successfully assigned probable functions. Secretome analysis revealed 57 HPs containing signal peptides, suggesting possible interactions with the host. Generally, the results of this study will facilitate a better understanding of the mechanism of pathogenesis of cattle theileriosis caused by *T. parva* and development of more effective disease control strategies.

ACKNOWLEDGEMENTS

Authors' express their sincere gratitude to the National Research Foundation Bioinformatics and Functional Genomics grant and AgriSeta for financial assistance.

REFERENCES

1. Muleya W, Namangala B, Simuunza M, et al. Parasit Vectors. BioMed Central; 2012;5: 255. doi:10.1186/1756-3305-5-255
2. Kuo C-H, Wares JP, Kissinger JC. Mol Biol Evol. 2008;25: 268998. doi:10.1093/molbev/msn213
3. Wu CH, Huang H, Yeh LSL, et al. Comput Biol Chem. 2003;27: 3747. doi:10.1016/S1476-9271(02)00098-1
4. Evan G. Int J cancer. 1997;71: 70911. Available: http://www.ncbi.nlm.nih.gov/pubmed/9180133
5. Trachana K, Larsson TA, et al. Bioessays. 2011;33: 76980. doi:10.1002/bies.201100062
6. Hutchings MI, Palmer T, et al Trends Microbiol. 2009;17: 1321. doi:10.1016/j.tim.2008.10.001
7. Khandavilli S, Homer KA, et al. Mol Microbiol. 2008;67: 54157. doi:10.1111/j.1365-2958.2007.06065.x
8. Kumar K, Prakash A, et al. Gene. Elsevier B.V.; 2014;543: 93100. doi:10.1016/j.gene.2014.03.060