

# EVOLUTIONARY GENOMICS OF *GLOSSINA MORSITANS* IMMUNE-RELATED CLIP DOMAIN SERINE PROTEASES AND SERINE PROTEASE INHIBITORS

Sarah Mwangi, Edwin Murungi, Mario Jonas and Alan Christoffels

South African National Bioinformatics Institute, University of the Western Cape, Private Bag X17, Modderdam Road, Bellville, Cape Town, South Africa

Email: Sarah Mwangi – [sarah@sanbi.a.za](mailto:sarah@sanbi.a.za); Edwin Murungi - [edwin@sanbi.ac.za](mailto:edwin@sanbi.ac.za); Mario Jonas - [mario@sanbi.ac.za](mailto:mario@sanbi.ac.za); Alan Christoffels - [alan@sanbi.ac.za](mailto:alan@sanbi.ac.za)

## ABSTRACT

Several species of haematophagous tsetse flies (genus *Glossina*) are vectors for trypanosomes, the parasitic protozoans that cause Human African Trypanosomiasis (HAT). Although there was a reduced incidence of HAT in the mid 1960's, decreased disease surveillance has led to a resurgence of HAT in sub-Saharan Africa. Despite being efficient vectors for HAT transmission, the prevalence of *G. morsitans* infection by trypanosomes in the wild is surprisingly minimal. The precise mechanisms by which *G. morsitans* remain refractory to trypanosome infection are largely unknown although it has been demonstrated that *G. morsitans* mounts a strong immune response to invading pathogens. This study identifies *G. morsitans* immune-related CLIP domain serine proteases and their inhibitors, serine protease inhibitors (serpin) genes. It further establishes their evolutionary relationships with counterparts in *Drosophila melanogaster*, *Anopheles gambiae*, *Bombyx mori*, *Manduca sexta* and *Culex quinquefasciatus*. Multiple sequence alignments show conservation of most secondary structure elements for both CLIPs and serpins. Amino acid composition of the serpin reactive site loop (RSL) indicates that the *G. morsitans* serpins act through an inhibitory mechanism to the target serine protease. Similar to *D. melanogaster* and unlike *A. gambiae*, the transcriptome data suggest that *G. morsitans* does not contain gene expansions in their CLIP-domain serine protease and serpin families. The presence of alternatively spliced variants in the *G. morsitans* serpins transcriptome data mirrors that of the *D. melanogaster* transcriptome.

Keywords: *Glossina morsitans*, *Drosophila melanogaster*, CLIP domain serine protease, serpin.