Characterization of the highly diverse genomic regions of an important African vector of schistosomiasis

Ryan Burd*[1], Jared Cayton*[1], Tom Pennance[1], Jacob A. Tennessen[2], Fredrick, Rawago[3], Maurice Odiere[3], George Owino[3], Michelle Steinauer[1]

[1]Western University of Health Sciences College of Osteopathic Medicine of the Pacific-Northwest. [2]Harvard T.H. Chan School of Public Health, Boston, MA. [3]Kenya Medical Research Institute, Kisumu, Kenya.

*Equally contributing authors

INTRODUCTION

Schistosomiasis is a chronic inflammatory disease caused by infection with parasitic blood flukes (*Schistosoma* spp.) that impact the lives of millions in endemic regions of South America, Asia and Africa (Steinmann et al., 2006). Chronic tissue inflammation and fibrosis leading to severe organ necrosis and eventual organ failure of those infected with schistosomes is caused by trapped schistosome eggs in the host tissues, laid by reproducing adult schistosomes (Colley *et al*., 2014). Eggs that do not get trapped are released into the environment with the feces or urine of the host

Snail intermediate hosts are required for the transmission of schistosomes from one vertebrate definitive host to another. Snails become infected when eggs released by the host reach freshwater, hatch, and the free-swimming schistosome larvae (miracidia) infect aquatic snails. In snails, schistosomes that successfully infect them clonally reproduce, resulting in the daily release of thousands of free-swimming, short lived, infective larvae, called cercariae. These cercariae are equipped to penetrate the skin of vertebrate hosts, causing infection and completing the parasites lifecycle.

Current schistosomiasis control efforts focus on the at-risk human population through mass drug administration (MDA) with Praziquantel (Colley *et al*., 2020). However, even with adequate treatment, people become rapidly reinfected, because infected snails remain in the environment.

Control of snail populations, typically through the application of molluscicides, has contributed to the reduction of reinfection rates following MDA in some countries (Allan *et al*., 2020). However, efficient application of molluscicides requires substantial logistical planning, community sensitization, human effort and material resources (Gryseels *et al.*, 2006; King *et al.,* 2015; Allan *et al*., 2020), proves to be ineffective in some endemic settings, and is indiscriminately toxic to other freshwater wildlife (Sokolow *et al*., 2018, Knopp *et al*., 2019).

An alternative approach to effective schistosomiasis control is needed to reduce or eliminate transmission to humans. Understanding inherent genetic resistance mechanisms in intermediate host snail species is likely to yield novel strategies of control.

PATHOGEN RECOGNITION RECEPTORS (PRRs)

Pathogen recognition receptors (PRRs) are proteins that play a critical component of the innate immune system by detecting molecules released by invading pathogens or damaged cells to then activate and recruit various immune cells to infection sites (Figure 1). **It is hypothesized that evolutionary host defense and reciprocal pathogenic innovation has driven PRRs to a high degree of genetic polymorphism**, a phenomenon known as balancing selection (Fijarczyk et al., 2016; Lundberg et al., 2020).

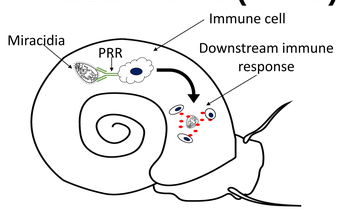

Figure 1: - Representing the role of pathogen recognition receptors (PRRs) in active snail immunity

PRRs have been demonstrated as active components of the immune system of *Biomphalaria glabrata*, the South American vector of *Schistosoma mansoni* (Pila et al., 2017; Adema et al., 2017), however PRRs in African vectors such as *B. sudanica*, which are responsible for the majority of global transmission, are unknown. Several types of PRR have been identified in *B. glabrata*, most of which contain transmembrane domains including toll-like receptors, Peptidoglycan recognition-binding proteins (PGRP), Gram-negative binding protein (GNBP) and Fibrinogen-related proteins (FREP).

AIMS & HYPOTHESIS

Aim: To identify candidate PRR genes in the *B. sudanica* genomes of 5 inbred lines by annotating genes in genomic regions exhibiting high nucleotide diversity and containing peptides with predicted transmembrane domains.

Hypothesis: We predict that novel PRRs in the *B. sudanica* genome, as well as those closely related to *B. glabrata*, will be identified in highly diverse regions of the genome because of the high nucleotide polymorphism in these genes mediated by balancing selection.

METHODS AND RESULTS

Sequence the Genome of Five *B. sudanica* Lines

Five inbred (for at least 3 generations) *B. sudanica* lines originating from Lake Victoria, Kenya sequenced with PacBio Sequel II and Illumina paired-end reads

Sequenced *B. sudanica* genomes ~0.96 gigabases (Gb), with final alignment containing combined 6,815 scaffolds and contigs

Nucleotide Diversity Between Lines

Mean inter-line nucleotide diversity (Π) was calculated across genome in windows of either 10 (Figure 2), 30 and 100Kb and selected for further analysis if:

- 10Kb: >0.01 Π (i.e. >1% nucleotide diversity)
- 30 & 100Kb: If contained in the top 50 most diverse contig regions for each window size that were not included in 10Kb windows with >0.01 Π

75 of 6,815 genome regions met the minimum Π threshold of 1% in 10Kb windows (n = 50) or were contained in the top 50 most diverse regions of 30 & 100Kb windows (n = 17)

Regions in three contigs with previously identified orthologous PRR genes in *B. glabrata* were included following selection (Table 1).

Predict Open Reading Frames in High-Diversity Regions

For each window that met our selection criteria of high nucleotide diversity, 1 Mb of surrounding nucleotide sequence had open reading frames (ORFs) identified using GENSCAN predictive peptide software (Burge and Karlin, 1997).

4578 ORFs predicted from the 67 genome regions annotated

Identify Orthologous Open Reading Frames and Predicting Transmembrane Domains with TMHMM

For each predicted ORF, orthologous peptides/proteins searches were performed using established databases (NCBI BLAST (Altschul et al., 1990) and Vectorbase for *B. glabrata* specific orthologs)

1399 ORFs (26%) with BLAST ortholog with E-value <1E-99

1286 ORFs (28%) with Vectorbase ortholog with E-value <1E-99

Peptide/Protein structures of ORFs were predicted using TMHMM to predict the presence, and number of, transmembrane domains (TMDs)

566 ORFs were predicted to contain between 1-50 TMDs

Immune suspected peptides were determined based on BLAST/Vectorbase results showing predicted receptor/membrane component functions and presence of TMDs (whilst ignoring highly repetitive sequences and transposon/POL predicted funtions).

818 or 17.9% of ORFs demonstrated immune function – 414 have at least 1 TMD

Interesting peptides (Table 1) include established PRRs in *B. glabrata* and novel peptides that have PRR functionality in *B. sudanica*

- **PTC2**, TLR-4 (c208), **Fibronectin type 3** (c359)
- **TLR-7** (s3064), **TLR-8**, **C3 complement protein** (c274)

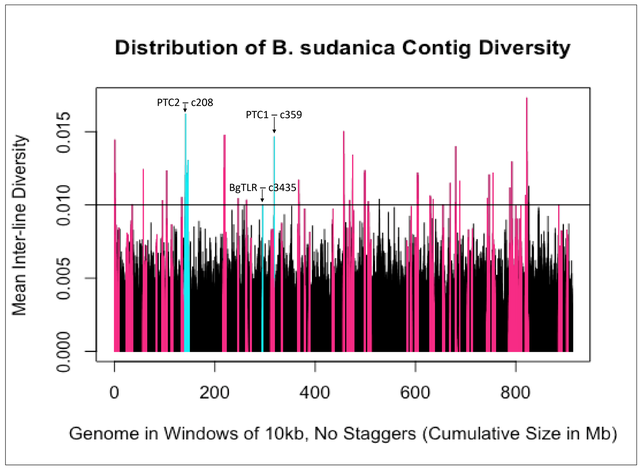

Figure 2: Distribution of highly diverse *B. sudanica* genome regions in windows of 10kb with known candidate pathogen recognition receptor genes; PTC1, PTC2 and BgTLR, labeled

Contig/Scaffold	Gene	Total ORFs	Predicted Peptides w/TMDs	Immune suspected peptides	Features of interest	Predicted Function	Reference
With previously known PRRs in *B. glabrata*							
Contig 208	PTC2	339	43	28	PTC2, TLR4, immunglobulin function	host-parasite interaction	Tennessen et al. 2020
Contig 359	PTC1/GRC	55	5	4	Fibronectin type 3 protein binding	host-parasite interaction	Tennessen et al. 2015
Contig 3435	BgTLR	27	2	6	C-type lectin binding mannose	Innate immune system	Pila et al. 2016
Novel regions with potential PRRs							
Contig 48		79	20	0	TLR3, CD180 like, leucine rich reapeats, transporters	innate immune system - pattern recognition, protein-protein interaction, secondary transporters	
Contig 69		85	14	2	Nuclear receptor subfamily, neurotensin receptor, GPCR	Inhibition of NF-kappa-B transactivation of IL2	
Contig 101		55	13	1	Cell wall integrity, GPCR	transmembrane signaling receptor activity	
Contig 205		62	3	1	Mucin-SAC-like	Mucus secretion	
Contig 274		82	13	2	C3 complement protein, TLR8	Innate immune system - pattern recognition	
Contig 499		81	8	15	TLR-4, Lectin binding repeats, Fibronectin	Innate immune system - pattern recognition	
Contig 540		60	8	2	Cell surface proteins	protein dephosphorylation	
Contig 1475		39	2	1	dependent K transport protein	immunoglobulin,Na+/Pi-cotransporter	
Contig 3586		78	1	3	Mucin-SAC-like, craniofacial development protein 2-like	mucus secretion	

Table 1: - Most interesting peptides from previously identified *B. glabrata* PRRs and novel PRRs unique to *B. sudanica*. Features of interest and predicted functions are included with total ORFs, peptides with TMDs and peptides with suspected immune function

Comparison to Control Regions

TMDs are overrepresented in immune-suspected peptides in highly divergent regions as compared to control regions*

30 randomly selected contigs investigated with 17 of 327 total peptides demonstrating PRR structure and function

414 / 4578 = **9.04%** of peptides in highly divergent regions with immune suspected function and TMDs

17 / 327 = **5.20%** of peptides in control regions with immune-suspected function and TMDs

*statistically significant with p<.05 using Fischer's exact test

CONCLUSIONS

Our candidate gene yield, their respective features and functions supports our hypothesis that the diversity-based approach is capable of identifying PRRs in the *B. sudanica* vector.

Our compiled list of candidate PRR genes contains orthologs of known PRR genes in the closely-related *B. glabrata* such as PTC1 and PTC2 which show 8-fold and 15-fold effects on odds of infection by *S. mansoni* respectively (Tennessen et al., 2020)

Novel peptides suspected of PRR structure and function in *B. sudanica* have been discovered including TLR-7, TLR-8 and C3 Complement protein, as well as many novel genes with unknown functions

Our compiled list of candidate PRR genes suspected of immune function provides a foundation guiding future gene resistance and knockout and genome-wide association studies (GWAS) for *Biomphalaria* species. Additional future work includes further interrogation of the nucleotide diversity dataset and directed investigation for *B. glabrata* immune genes SOD1, RADres, FREP3, HSP90, BgGRN, Knight marker, Prx4, Catalase and Phox

ACKNOWLEDGEMENTS

- Stephanie Bollmann and Centre for Quantitative Life Sciences - Oregon State University
- Western University of Health Sciences Summer Student Research Fellowship
- KEMRI
- NIH NIAID R01AI141862

REFERENCES

Adema CM, Hillier LW, Jones CS, et al. Whole genome analysis of a schistosomiasis-transmitting freshwater snail. *NATURE COMMUNICATIONS*. 2017;8:15451. doi:10.1038/ncomms15451

Allan F, Ame SM, Tian-Bi Y-NT, et al. Snail-Related Contributions from the Schistosomiasis Consortium for Operational Research and Evaluation Program including Xenomonitoring, Focal Mollusciciding, Biological Control, and Modeling. *The American journal of tropical medicine and hygiene*. 2020;103(1_Suppl):66-79. doi:10.4269/ajtmh.19-0831

Altschul SF, Gish W, Miller W, Meyers EW, Lipman DJ. Basic Local Alignment Search Tool. *JMB Online (Journal of Molecular Biology)*. 1990;215(3):403-410. Accessed November 27, 2021. https://search-ebscohost-com.proxy.westernu.edu/login.aspx?direct=true&db=edo&AN=epi2728048&site=eds-live

Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. *Journal of molecular biology*. 1997;268(1):78-94. doi:10.1006/jmbi.1997.0951

Colley D.G., Bustinduy, A.L., Secor, W.E., King, C.H. (2014) Human Schistosomiasis. *Lancet*. 383: 2253-64. DOI: 10.1016/ S0140-6736(13)61949-2

Fijarczyk A, Dudek K, Babik W. Selective Landscapes in newt Immune Genes Inferred from Patterns of Nucleotide Variation. *GENOME BIOLOGY AND EVOLUTION*. 2016;8(11):3417-3432. doi:10.1093/gbe/evw236

Gryseels B, Polman K, Clerinx J, Kestens L. Human schistosomiasis. *Lancet*. 2006;368(9541):1106-1118. doi:10.1016/s0140-6736(06)69440-3

King, C. H., Sutherland, L. J., Bertsch, D. (2015) Systematic Review and Meta-analysis of the Impact of Chemical-Based Mollusciciding for Control of *Schistosoma mansoni* and *S. haematobium* Transmission. *PLOS Neglected Tropical Disease*. 9(12): e0004290. DOI: 10.1371/journal.pntd.0004290

Knopp S, Person B, Ame SM, et al. Evaluation of integrated interventions layered on mass drug administration for urogenital schistosomiasis elimination: a cluster-randomised trial. *LANCET GLOBAL HEALTH*. 2019;7(8):E1118-E1129. DOI: 10.1016/S2214-109X(19)30189-5

Lundberg M, Zhong X, Konrad A, Olsen R-A, Raberg L. Balancing selection in Pattern Recognition Receptor signalling pathways is associated with gene function and pleiotropy in a wild rodent. *MOLECULAR ECOLOGY*. 2020;29(11):1990-2003. doi:10.1111/mec.15459

Pila EA, Li H, Hambrook JR, Wu X, Hanington PC. Schistosomiasis from a Snail's Perspective; Advances in Snail Immunity. *TRENDS IN PARASITOLOGY*. 2017;33(11):845-857. doi:10.1016/j.pt.2017.07.006

Steinmann P, Keiser J, Bos R, Tanner M, Utzinger J. Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk. *The Lancet Infectious diseases*. 2006;6(7):411-425. doi:10.1016/S1473-3099(06)70521-7

Tennessen JA, Bollmann SR, Peremyslova E, et al. Clusters of polymorphic transmembrane genes control resistance to schistosomes in snail vectors. *ELIFE*. 2020;9:e59395. doi:10.7554/eLife.59395