Molecular characterization of a neglected food-borne trematode *Artyfechinostomum sufrartyfex* Lane, 1915 (Trematoda: Echinostomatidae): a transcriptomic and *in silico* approach

A Thesis Submitted

To

Sikkim University

In Partial Fulfilment of the Requirement for the

Degree of Doctor of Philosophy

By

Suman Dahal
Department of Zoology,
School of Life Sciences

December 2021

Dedicated to my dear family members, supervisor, teachers since my childhood days and close friends……

DECLARATION

I, **Suman Dahal** hereby declare that the Ph.D. thesis entitled **"Molecular characterization of a neglected food-borne trematode *Artyfechinostomum sufrartyfex* Lane, 1915 (Trematoda: Echinostomatidae): a transcriptomic and *in silico* approach"** submitted by me for the award of Doctor of Philosophy in Zoology, Sikkim University under the supervision of **Dr. Sudeep Ghatani**, Assistant Professor, Department of Zoology is an original research work carried out by me at Department of Zoology, School of Life sciences, Sikkim University, Gangtok. This work has not been submitted for any other degree to this or any other University. All the relevant sources of previous research works have been properly cited.

Place: GANGTOK

Date: 9/12/22

Suman Dahal

Suman Dahal
Roll no. 16PDZ003
Reg no.16SU19574
Ph.D Reg.16/PhD/ZOO/03
Department of Zoology,
School of Life Sciences,
Sikkim University

इल, सामदुर, तादोंग - 737102
क, सिक्किम, भारत
-03592-251212, 251415, 251656
फैक्स - 251067
ाइट - www.cus.ac.in

6th Mile, Samdur, Tadong-737102
Gangtok, Sikkim, India
Ph. 03592-251212, 251415, 251656
Telefax : 251067
Website : www.cus.ac.in

सिक्किम विश्वविद्यालय
SIKKIM UNIVERSITY

(भारत के संसद के अधिनियम द्वारा वर्ष 2007 में स्थापित और नैक (एनएएसी) द्वारा वर्ष 2015 में प्रत्यायित केंद्रीय विश्वविद्यालय)
(A central university established by an Act of Parliament of India in 2007 and accredited by NAAC in 2015)

CERTIFICATE

This is to certify that Ph.D. thesis entitled "**Molecular characterization of a neglected food-borne trematode *Artyfechinostomum sufrartyfex* Lane, 1915 (Trematoda: Echinostomatidae): a transcriptomic and *in silico* approach**" submitted to Sikkim University for the Degree of Doctor of Philosophy in Zoology embodies the original research work carried out by Mr. Suman Dahal at the Department of Zoology, School of Life Sciences, Sikkim University under my supervision. The work described here is original and no part of this thesis has been submitted elsewhere for the award of any Degree, Diploma, Associateship or Fellowship at this or any other University or institution of higher learning. Mr Dahal is conversant with the techniques and literature cited in the thesis and has fulfilled the requirements for the Degree of Doctor of Philosophy in Science (Zoology) at Sikkim University. In character and demeanour, Mr Suman Dahal is fit to submit the thesis for Ph.D. degree.

9/12/2022

Dr. Sudeep Ghatani
Supervisor
Assistant Professor
Department of Zoology
School of Life Sciences
Sikkim University
Gangtok

Dr. Sudeep Ghatani
Assistant Professor
Department of Zoology
SIKKIM UNIVERSITY

Dr. Bhoj Kumar Acharya
Associate Professor & Head
Department of Zoology
School of Life Sciences
Sikkim University
Gangtok

Head
प्राणी विज्ञान विभाग
Department of Zoology
सिक्किम विश्वविद्यालय
Sikkim University

त, सामदुर, तादोंग - 737102
, सिक्किम, भारत
3592-251212, 251415, 251656
स - 251067
ट - www.cus.ac.in

6th Mile, Samdur, Tadong-737102
Gangtok, Sikkim, India
Ph. 03592-251212, 251415, 251656
Telefax : 251067
Website : www.cus.ac.in

सिक्किम विश्वविद्यालय
SIKKIM UNIVERSITY

(भारत के संसद के अधिनियम द्वारा वर्ष 2007 में स्थापित और नैक (एनएएसी) द्वारा वर्ष 2015 में प्रत्यायित केंद्रीय विश्वविद्यालय)
(A central university established by an Act of Parliament of India in 2007 and accredited by NAAC in 2015)

Ref. No: SU/Dept of Zoo/2022/307 Date: 05/12/2022

PLAGIARISM TEST CERTIFICATE

This is to certify that plagiarism check has been carried out for the following Ph.D. thesis with the help of OURIGINAL SOFTWARE and the result is 0% similarity, which is within the permissible limit (below 10% tolerance rate) as per the norm of Sikkim University.

"Molecular characterization of a neglected food-borne trematode *Artyfechinostomum sufrartyfex* Lane, 1915 (Trematoda: Echinostomatidae): a transcriptomic and *in silico* approach"

Submitted by Mr. Suman Dahal under the supervision of Dr. Sudeep Ghatani, Assistant Professor, Department of Zoology, School of Life Sciences, Sikkim University.

Suman Dahal

Suman Dahal
Roll No. 16PDZ003
Signature of the Candidate

Sudeep Ghatani

Dr. Sudeep Ghatani
Assistant Professor
Department of Zoology
School of Life Sciences
Sikkim University

5/12/2022

Dr. Shri Ram
Librarian
Central Library
Sikkim University

Librarian Central Library Sikkim University

Dr. Sudeep Ghatani
Assistant Professor
Department of Zoology
SIKKIM UNIVERSITY

ACKNOWLEDGEMENT

I am grateful and blessed to be able to acknowledge right away that various people have made substantial contributions to the success of this voyage. Many noble beings backed me up, and without their help, my research inquiries would have gone unanswered. As a result, I am delighted to take this occasion to express my gratitude to everyone who has been directly or indirectly engaged in my Ph.D path.

It is a tremendous honour for me to express my heartfelt gratitude to my supervisor Dr. Sudeep Ghatani, Assistant Professor, Department of Zoology, Sikkim University, for his continual direction, insightful suggestions, invaluable counsel, inspiration, strength, shield, and support throughout this journey. It is worth highlighting that this research would not have been feasible without his consistent inspiration, useful advice, unselfish assistance, and insightful suggestions. I consider myself really fortunate to have been given the opportunity to work under the supervision of such an enthusiastic mentor. My supervisor has always encouraged and pushed me to learn from my mistakes, and many of them have given me important career lessons.

I extend my sincere gratitude to Dr. Bhoj Kumar Acharya, Associate professor, Head of the Department of Zoology, Sikkim University for extending the facilities of the Department and for valuable suggestions and support throughout the study. I am very thankful to former Heads of the Department of Zoology, Sikkim University Dr. Basundhara Chettri and Dr. Kshetrimayum Birla Singh for their constant support during my study.

I wish to extend my sincere gratitude to Dr. Bisu Singh, Assistant Professor, Department of Zoology, Sikkim University for his valuable guidance, suggestions and encouragement throughout the study. His suggestions and encouragement supported me a lot throughout my study. My sincere thanks to Dr. Madav Prasad Thapa (Guest faculty), Dr. Kumar Basnet (Former Guest faculty) for their support and encouragement throughout the study.

I am very thankful to Dr Yugal Kishor Prasad, MD (Pediatrics), Director, Shri Shubh Lal Hospital and Research Centre, Sitamarhi, Bihar, India for his valuable support in completion of this study. I also wish to express my sincere thanks to all the hospital staffs at Shri Shubh Lal Hospital and Research Centre for the help extended during field survey and sample collection.

I wish to express my heartfelt gratitude to Prof. Saurabh Raghuvanshi, Department of Plant Molecular Biology, Delhi University, New Delhi, India for extending the facilities at Department of Plant Molecular Biology and his valuable suggestions during the study. I also express my sincere thanks to Miss Pratibha Gour for her kind suggestions and help during the study. Further, I would also like to thank Prof. B. Jayaram, Mr Shashank Shekhar and Mr Dheeraj Chourasia at Indian Institute of Technology, Delhi for their help during this study.

I am very thankful to technical assistant Mr. Bedanta Saikia and former Lab attendant Mrs Sabina Sharma at Department of Zoology, Sikkim University for their constant support and assistant in the laboratory. I am also very thankful to Mr. Diwas Gurung and Mrs. Angel Gurung (MTS), Sikkim University for their help and support.

Friends are very important in our lives. They are always there when we need them and never forget to help, support and motivate us. Therefore, I owe my most heartfelt gratitude to all of my friends, Dr. Sailendra Dewan, Dr. Jiwan Gurung, Mr. Dependra Chamlagai, Miss Prerna Dhakal, Miss Ananta Rai and Mr. Kishor Chettri for their encouragement, unconditional support at every step of my study. I also feel very lucky that I was accompanied by very loving and supportive juniors. I express my heartfelt thanks to all of my academic juniors, Mr. Bishal Thakuri, Miss Tanushree Das, Miss Dipshika Sakia, Mr. Roshan Tamang, Mr. Iswar Chettri, Mr. Dibyan Lama, Mr. Aita Hang Subba, Miss Pempa Lhamu Bhutia and Miss Rabina Mothey for their support and encouragement during the entire journey of my study.

Through the grace of my family and their prayers, I could successfully overcome all struggles and difficult situations during the course of my study. It is a great privilege to express my sincere thanks and deep sense of gratitude to my beloved father, mother, Lt grandfather, grandmother, sister's, brother's, uncle's for their sacrifice, encouragement, inspiration, unconditional love, financial support to complete the study. I would also like to acknowledge funding body DST-SERB for financial grant no. Sanction No. SB/YS/LS-195/2014 dated October 26, 2015.

Suman Bishal

.....

CONTENTS

LIST OF TABLES

Table 15. B-cell epitopes predicted for the diagnostic targets and their respective hits in NCBI and UniProt databases. *Unique epitopes are highlighted in green; other epitopes are highlighted in yellow.

Table 16. Drug targets along with their FDA approved compounds and docking score.

Table 17. Drug targets along with their ZINC molecules and docking score.

Table 18. Details of top five protein target interactions with FDA approved drug molecules (1-5) with binding affinity cutoff of ≤-9.5 kcal/mol. Top five drug targets with ZINC molecules (6-10) with significant binding affinity cutoff of ≤-8.0 kcal/mol.

Appendix Table. ProTSAV results of the 145 protein structures.

LIST OF FIGURES

Figure 12. Representative output result from ProtSAV metaserver for one of the drug target identified in the current study. A blue dot appears in the yellow region.

Figure 13A. Interaction between the predicted target sodium-dependent transporter (uncharacterized) with FDA approved drug molecule Rifamycin (DB11753) exhibiting binding affinity of -11.9 kcal/mol.

Figure 13B. Amino acid residues in the active site of sodium-dependent transporter (uncharacterized) protein within 3.5 Å of the ligand molecule Rifamycin (DB11753).

Figure 13C. Sodium-dependent transporter (uncharacterized) protein with FDA approved drug molecule Rifamycin (DB11753) exhibiting amino acid interactions.

Figure 13D. Interaction between the predicted target Anaerobic glycerol-3 phosphate dehydrogenase subunit A with FDA approved drug molecule Indinavir (DB00224) exhibiting binding affinity of -11.3 kcal/mol.

Figure 13E. Amino acid residues in the active site of Anaerobic glycerol-3 phosphate dehydrogenase subunit A protein within 3.5 Å of the ligand molecule Indinavir (DB00224).

Figure 13F. Anaerobic glycerol-3 phosphate dehydrogenase subunit A with FDA approved drug molecule Indinavir (DB00224) exhibiting amino acid interactions.

Figure 13G. Interaction between the predicted target Tyrosine-tRNA synthetase with FDA approved drug molecule Grepafloxacin (DB00365) exhibiting binding affinity of -11.2 kcal/mol.

Figure 13H. Amino acid residues in the active site of Tyrosine-tRNA synthetase protein within 3.5 Å of the ligand molecule Grepafloxacin (DB00365).

Figure 13I. Tyrosine-tRNA synthetase with FDA approved drug molecule Grepafloxacin (DB00365) exhibiting amino acid interactions.

Figure 13J. Interaction between the predicted target Argininosuccinate synthase with FDA approved drug molecule Rifamycin (DB11753) exhibiting binding affinity of -11.0 kcal/mol.

Figure 13K. Amino acid residues in the active site of Argininosuccinate synthase protein within 3.5 Å of the ligand molecule Rifamycin (DB11753).

Figure 13L. Argininosuccinate synthase with FDA approved drug molecule Rifamycin (DB11753) exhibiting amino acid interactions.

Figure 13M. Interaction between the predicted target Chloride intracellular channel with FDA approved drug molecule Rifamycin (DB11753) exhibiting binding affinity of -10.6 kcal/mol.

Figure 13N. Amino acid residues in the active site of Chloride intracellular channel protein within 3.5 Å of the ligand molecule Rifamycin (DB11753).

Figure 13O. Chloride intracellular channel with FDA approved drug molecule Rifamycin (DB11753) exhibiting amino acid interactions.

Figure 14A. Interaction between the predicted target bifunctional protein Aas with ZINC molecule ZINC02820058 exhibiting binding affinity of -11.4 kcal/mol.

Figure 14B. Amino acid residues in the active site of bifunctional protein Aas protein within 3.5 Å of the ligand molecule ZINC02820058.

Figure 14C. The predicted target bifunctional protein Aas with ZINC molecule ZINC02820058 exhibiting exhibiting amino acid interactions.

Figure 14D. Interaction between the predicted target 50S ribosomal protein L3 glutamine methyltransferase with ZINC molecule ZINC02820058 exhibiting binding affinity of -11.2 kcal/mol.

Figure 14E. Amino acid residues in the active site of 50S ribosomal protein L3 glutamine methyltransferase protein within 3.5 Å of the ligand molecule ZINC02820058.

Figure 14F. 50S ribosomal protein L3 glutamine methyltransferase with ZINC molecule ZINC02820058 exhibiting amino acid interactions.

Figure 14G. Interaction between the predicted target Putative phosphonates utilization ATP-binding protein PhnK with ZINC molecule ZINC02820058 exhibiting binding affinity of -11.2 kcal/mol.

Figure 14H. Amino acid residues in the active site of Putative phosphonates utilization ATP-binding protein PhnK protein within 3.5 Å of the ligand molecule ZINC02820058.

Figure 14I. Putative phosphonates utilization ATP-binding protein PhnK with ZINC molecule ZINC02820058 exhibiting amino acid interactions.

Figure 14J. Interaction between the predicted target Cobalt import ATP-binding protein CbiO with ZINC molecule ZINC12576410 exhibiting binding affinity of -9.7 kcal/mol.

Figure 14K. Amino acid residues in the active site of target Cobalt import ATP-binding protein CbiO within 3.5 Å of the ligand molecule ZINC12576410.

Figure 14L. Cobalt import ATP-binding protein CbiO with ZINC molecule ZINC12576410 exhibiting amino acid interactions.

Figure 14M. Interaction between the predicted target DNA mismatch repair protein MutL with ZINC molecule ZINC03143011 exhibiting binding affinity of -8.5 kcal/mol.

Figure 14N. Amino acid residues in the active site of target DNA mismatch repair protein MutL within 3.5 Å of the ligand molecule ZINC03143011.

Figure 14O. DNA mismatch repair protein MutL with ZINC molecule ZINC03143011 exhibiting amino acid interactions.

ABBREVIATIONS

NTDs- Neglected Tropical Diseases

SDG- Sustainable Development Goals

FBTs- Foodborne Trematodiases

WHO- World Health Organization

FERG- Foodborne Disease Burden Epidemiology Reference Group

CT- Computed Tomography

MRI- Magnetic Resonance Imaging

PCR- Polymerase Chain Reaction

rDNA- Ribosomal DNA

mtDNA- Mitochondrial DNA

I**TS**- Internal Tanscribed Spacer

CO1- Cytochrome c Oxidase Subunit 1

NADH- Nicotinamide Adenine Dinucleotide Hydrogen

NME- New Molecular Entity

CADD- Computer Aided Drug Design

NDA- New Drug Application

FDA- Food and Drug Administration

QSAR- Quantitative Structure Activity Relationship

NCBI- National Centre for Biotechnology Information

IEC- Institutional Ethics Committee

SSL- Shri Subh Lal Hospital

HCL-Hydrochloric Acid

BLAST- Basic Local Alignment Search Tool

BOLDSYSTEM- Barcode of Life Data Systems

MUSCLE- Multiple sequence comparison by log-expectation

EMBL- EBI – European Molecular Biology Laboratory's- European Bioinformatics Institute

SIM- Sequence Identity Matrix

IDS- Identification System

ESPs- Excretory Secretory Proteins

QC- Quality Control

RIN- RNA Integrity Number

SRA- Sequence Read Archive

ORFs- Open Reading Frames

GO- Gene Ontology

TPM- Transcripts Per Million

AADS- Automated Active site detection Docking and Scoring

NMR-Nuclear Magnetic Resonance

THESIS ORGANIZATION

My thesis titled **"Molecular characterization of a neglected food-borne trematode *Artyfechinostomum sufrartyfex* Lane, 1915 (Trematoda: Echinostomatidae): a transcriptomic and *in silico* approach"** is organized into four chapters:

Chapter 1 outlines the public health concerns such as Neglected Tropical Diseases (NTDs) and Food borne trematodiases (FBTs) which is followed by review of literature pertaining to international and national status of *Artyfechinostomum sufrartyfex,* various molecular taxonomic and transcriptomic studies undertaken in helminths and an account on drug resistance in helminth parasites including approaches such as *in silico* drug design and drug repurposing. The research gap relevant in this study and objectives of the research undertaken herein are also defined in this particular chapter.

Chapter 2 deals with morphological and molecular analyses undertaken in the current research work in order to correctly identify the causative agent and establish the source of infection transmission in child patients belonging to northern parts of Bihar, India.

Chapter 3 is related to multi-stage transcriptomic approach undertaken for prediction of *A. sufrartyfex* secretome with a view to identify important species specific diagnostic markers.

Chapter 4 is linked to *in silico* drug design approach with utilization of drug repurposing strategy in order to provide important protein targets along with their ligand molecules against *A. sufrartyfex*.

The thesis has a total of 270 citations.

Chapter 1: General Introduction

1.1. Background

Helminths have afflicted people since before the dawn of recorded history. Helminth eggs have been discovered in mummified human faeces dating back to thousands of years (Cox, 2002; Hotez et al., 2006). Currently, there are around 287 helminth species known to occur in humans, with 95% of them being zoonotic (Taylor et al., 2001). Also, infections credited to parasitic helminths are leading causes of morbidity in developing countries. Their contribution in terms of human and economic loss is enormous (Savioli et al., 1992). Furthermore, the fact that helminth diseases have multiple life cycles, as well as different transmission channels via water, soil, food, insect and other invertebrate vectors, adds to the complexities in delivering robust interventions (Gazzinelli et al., 2012). As per their affiliations to the Neglected Tropical Diseases (NTDs) prevalent around the world, helminths have significant role and contribute largely to these group of diseases (Figure 1). Neglected Tropical Diseases (NTDs) encompass a diverse range of diseases attributed mainly to taxonomic groups such as bacteria, viruses, fungi and parasites that affect the poorest populations in developing nations (Hotez et al., 2007). According to the WHO, there are 20 diseases currently identified as NTDs and affecting almost 1 billion people in nearly 150 countries and territories where they are endemic (Table 1). Despite this, NTDs have previously been placed low on the global health policy agenda and have only recently gained acknowledgment in 2015 with the Sustainable Development Goals (United Nations SDG target 3.3). These diseases, which are mostly prevalent in Sub-Saharan Africa, Asia, and

Latin America, cause long-term disability and poverty (Lammie et al., 2006; Hotez et al., 2009).

Among the various NTDs, Food-borne Trematodiases (FBTs) are of major global concern brought about by digenetic trematodes (flatworms or "flukes"). Infection transmission occurs due to intake of raw or improperly cooked freshwater fish and aquatic invertebrates like crabs, crayfish, snails and clams, water cress containing metacercariae (cysts) of these trematodes (Fried and Abruzzi, 2010). Around 6,000 digenetic trematode species are known, many of them are significant human parasites (Keiser and Utzinger, 2009). Food-borne trematodes include liver flukes (*Clonorchis sinensis, Fasciola gigantica, Fasciola hepatica, Opisthorchis felineus, Opisthorchis viverrini* etc), lung flukes (*Paragonimus* spp. etc) and intestinal flukes (e.g. *Artyfechinosotmum sufrartyfex, Echinostoma* spp., *Fasciolopsis buski* etc) (Keiser and Utzinger, 2009; Prasad et al., 2019). According to the WHO factsheet, 2021, food-borne trematode infections result in the loss of 2 million life years due to disability and death around the world every year. Because public health knowledge and access to health services are typically limited among impacted populations, the full burden of these infections remains unknown. The WHO Food-borne Disease Burden Epidemiology Reference Group (FERG) (2015) estimated annual total of 200,000 illnesses and over 7000 deaths, resulting in more than 2 million disability-adjusted life years around the globe. These infections also have a considerable economic impact, which is connected to losses in the livestock and aquaculture due to decreased animal productivity. The clinical picture, history of appropriate risk factors, eosinophilia detection, ultrasound, computed tomography (CT), or magnetic resonance imaging (MRI) scans are all used to suspect

food-borne trematodiases. To confirm the diagnosis, a variety of diagnostic tests are utilized, including parasitological techniques to detect eggs in faecal samples, immunological techniques identifying parasite specific antibodies in serum samples, and DNA based techniques such as Polymerase Chain Reaction (PCR).

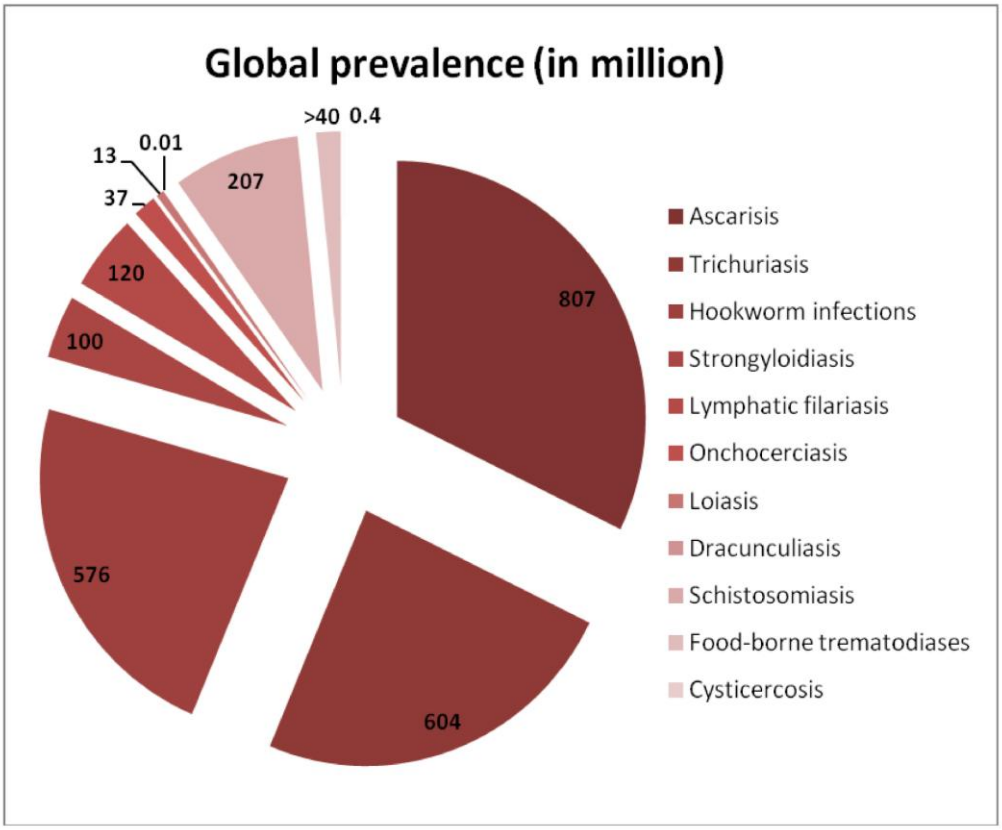


Figure 1. Global prevalence of important human helminthiases (Tandon et al., 2015).

Intestinal flukes, one of the food-borne trematodes, contribute significantly to infections categorized as food-borne intestinal trematodiases. The category of intestinal food-borne trematodes includes a sizable collection of species that cause parasitic zoonoses. Diplostomiasis, echinostomiasis, fasciolopsiasis, gymnophalloidiasis, and heterophyiasis are regarded as the main intestinal food-borne trematodiases (Toledo et al.,

2012). Intestinal flukes of over 70 different species have been isolated from humans (Yu and Mott, 1994). Among these, flukes belonging to the family Echinostomatidae constitute an important group of food-borne intestinal trematodes. The Echinostomatidae family encompasses a diverse collection of cosmopolitan and hermaphroditic digeneans that parasitize a wide range of vertebrate hosts as adults. Human echinostomiasis is caused by several species of echinostomid flukes, one of which, *Artyfechinostomum sufrartyfex* Lane, 1915, is a medically significant but neglected food-borne parasite. Some important symptoms associated with this disease include abdominal pain, diarrhoea, tenesmus, easy fatiguability, reduction in body weight and urinary incontinence (Rim, 1982; Chai and Lee, 2002; Fried and Abruzzi, 2010). Apart from humans, the parasite infects a number of other vertebrate hosts. *A. sufrartyfex* has emerged as an important pathogen in terms of socioeconomic importance with a fairly wide distribution in India and contribute to the rise in food-borne intestinal infection cases in the country. However, majority of the research to date has focused on morphological descriptions of the parasite, case reports, the parasite's life cycle, and epidemiology. This parasite has received very little attention at the molecular level, especially in the direction of diagnostics and therapeutics.

Table 1. Currently recognized NTDs around the globe. *WHO roadmap for neglected tropical diseases (2021-2030).

Diseases	***Causative agent***	***Brief Taxonomic description***	***Current initiative against the disease****
Dracunculiasis	*Dracunculus medinensis*	Nematoda	Subjected for eradication
Yaws	*Treponema pallidum* spp.	Spirochaetes	Subjected for eradication
Leprosy	*Mycobacterium leprae*	Actinobacteria	Subjected for elimination
Human African trypanosomiasis (gambiense and rhodesiense)	*Trypanosoma brucei*	Kinetoplastid (Protozoa)	Subjected for elimination
Onchocerciasis	*Onchocerca volvulus*	Filarial worms (Filarioidea); Nematoda (roundworms)	Subjected for elimination
Chagas disease	*Trypanosoma cruzi*	Kinetoplastid (Protozoa)	Subjected for elimination
Leishmaniasis (visceral and cutaneous)	*Leishmania* spp.	Kinetoplastid (Protozoa)	Subjected for elimination
Lymphatic filariasis	*Wuchereria bancrofti, Brugia malayi and B. timori*	Filarial worms (Filarioidea); Nematoda (roundworms)	Subjected for elimination
Rabies	*Rabies virus*	Lyssavirus	Subjected for elimination
Schistosomiasis	*Schistosoma* spp.	Trematoda (Platyhelminthes)	Subjected for elimination
Soil-transmitted helminthiases	*Ascaris lumbrico1des, Trichuris trichiura, Necator americanus, and Ancylostoma duodenale*	Nematoda	Subjected for elimination
Trachoma	*Chlamydia trachomatis*	Chlamydiales	Subjected for elimination
Buruli ulcer	*Mycobacterium ulcerans*	Actinobacteria	Subjected for control
Dengue	*Dengue virus*	Flavivirus	Subjected for control
Echinococcosis	*Echinococcus granulosus, E. multilocularis*	Cestoda (Platyhelminthes)	Subjected for control
Food-borne trematodiases	*Clonorchis* spp., *Echinostoma* spp., *Opisthorchis* spp., *Fasciola* spp. *and Paragonimus* spp.	Trematoda (Platyhelminthes)	Subjected for control
Mycetoma, chromoblastomycosis and other deep mycoses	*Eumycetoma, Actinomycetoma*	Bacteria, Fungi	Subjected for control
Scabies	*Sarcoptes scabiei*	Arthropod	Subjected for control
Snakebite envenoming	*Snakes*	Serpentes	Subjected for control
Taeniasis/cysticercosis	*Taenia solium, T. saginata*	Cestoda (Platyhelminthes)	Subjected for control

1.2. Review of Literature

1.2.1. *Artyfechinostomum sufrartyfex* - International and national status

The Superfamily Echinostomatoidea, Family Echinostomatidae, Subfamily Himasthlinae includes the genus *Artyfechinostomum* Lane, 1915 (Kostadinova and Jones, 2005). Morphological characters including a circumoral head collar equipped with one or two crowns of massive spines interrupted ventrally distinguishes the family Echinostomatidae Looss 1899 (Kostadinova, 2005). Important characteristic features include muscular and dorsoventrally flattened oval body, tegument with enormous scale-like spines, dorsal collar spines in double row; testes big, deeply branched (Kostadinova, 2005). *A. sufrartyfex* can be identified by distinguishing traits such as small head collar with a crown of 39-45 collar spines on its surface, arrayed in two alternating rows on the dorsal side. Digenean with a lengthy cirrus sac, a submedian ovary, and two deeply lobed big testes and are hermaphroditic adults (Kostadinova and Jones, 2005). An echinostome's life cycle can be complex, requiring the participation of a variety of hosts to complete their sexual and asexual cycles. To complete its life cycle, *A. sufrartyfex* requires a definitive host (vertebrate) and two distinct molluscans as intermediate hosts (Ghatani et al., 2018). The definitive host releases eggs in faeces into unsanitary ponds or rivers, where they grow into miracidial stage and locate the first intermediate host, *Indoplanorbis* exustus or *Gyraulus convexiusculus* (Yu and Mott, 1994; Toledo and Esteban, 2016; Ghatani et al., 2018). Inside the first intermediate host they transform into sporocysts and subsequently redia, which reproduce asexually and develop into daughter redia. After metamorphosis into cercariae, they emerge and seek out the second intermediate host, such as *Pila scutata*, *P. globosa*, *Lymnaea cumingiana*, *Digoniostoma*

pulchella etc (Yu and Mott, 1994; Ghatani et al., 2018; Prasad et al., 2019). After ingestion of the second intermediate host, the encysted metacercariae infect the definitive host usually humans and pigs (Ghatani et al., 2018) (Figure 2).

Echinostoma malayanum (Leiper, 1911) was originally described from a human in Malaysia and was later considered a synonym of *A. sufrartyfex* (syn. *Echinostoma sufrartyfex* Faust, 1929; *Paryphostomum sufrartyfex* Bhalerao, 1931; *Paryphostomum mehrai* Jain, 1957; *Artyfechinostomum rnehrai* Jain, 1960; *Neoartyfechinostomum shubhrai* Agrawal, 1963; *Artyfechinostomum munshii* Deodhar et al., 1968). This species is prevalent in South and South East Asia *A. sufrartyfex* has been reported from Malaysia, Thailand, Indonesia, India and the Philippines (Radomyos et al., 1982; Beaver et al., 1984; Maji et al., 1993; Yu and Mott, 1994; Belizario et al., 2007). Reports of the parasite have also surfaced from Vietnam and Lao PDR (Ngo et al., 2011; Chai et al., 2012). Also metacercariae of this species were discovered in *Pila* sp. snails from Phnom Penh, Cambodia (Sohn et al., 2017).

In India, *A. sufrartyfex* was reported for the first time from a girl in Assam who had died from the infection with the parasite (Lane, 1915). Following that, sporadic cases of human infections with severe clinical symptoms were reported from various parts of the country though sporadically. Another death case owing to infection with this parasite has been reported from Madras (Reddy and Varmah, 1950). In a tribal village near Kolkata, cases of unknown echinostome species have been documented (Maji et al., 1993). There have also been reports of this fluke from the states of Andhra Pradesh, Bihar, Tamil Nadu, and Uttar Pradesh, as well as among the Oraon tribes of West Bengal (Tandon et al., 2008). Apart from humans, previous workers have reported the occurrence

of this parasite from domesticated animals such as pigs (Bhalerao, 1931; Rai and Ahluwalia, 1958; Ahluwalia, 1962), cats and dogs (Dubey et al., 1969) as well as from wild animals such as honey-badger (Srivastava, 1964) and small Indian civet (Amrithraj et al., 1999). *A. sufrartyfex* was also reported from naturally infected pigs in Bengal (Bhattacharya et al., 1972) and Northeastern states of Assam, Meghalaya (Roy and Tandon, 1996; Das et al., 2017), Tripura (Roy and Tandon, 1992) and Mizoram (Borthakur et al., 2007). Also few years back the parasite has been reported from small Indian mongoose in the state of Jammu and Kashmir (Borkataki et al., 2017).

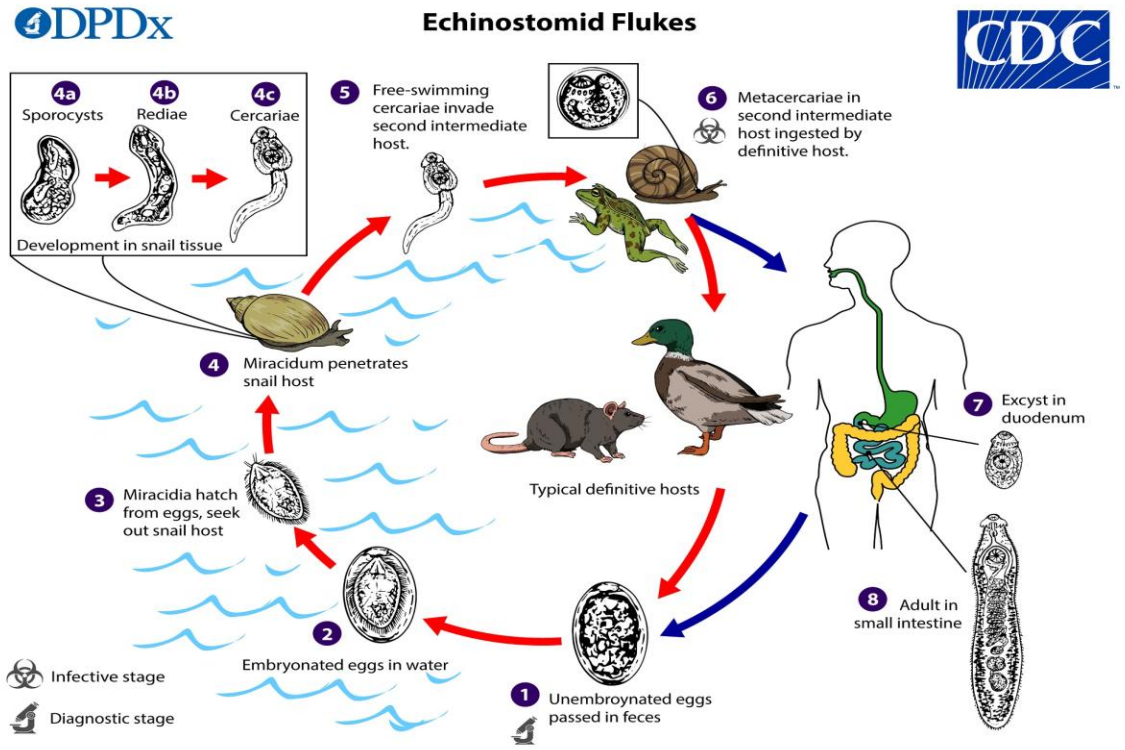


Figure 2. Lifecycle of a typical echinostomid fluke.
(**Source**: https://www.cdc.gov/dpdx/echinostomiasis/index.html)

1.2.2. Molecular taxonomic studies in helminths

The focus in recent years has been on determining acceptable DNA regions to employ for detecting variation at various taxonomic levels. This necessitates the selection of molecular techniques capable of differentiating genetic variants at various hierarchical levels, as well as the examination of the right region of DNA for the level of issues being addressed. The ability to characterize such genetic variation requires adequate and thorough analysis which is strengthened when a variety of genetic loci are used. The advent of DNA sequencing and the resulting quantity of data has greatly boosted the practicality of using PCR-based approaches in helminth studies (Constantine, 2003). PCR technology and DNA sequencing techniques allow species, strains, and populations to be identified from a little amount of tissue at any point of their life cycle (Boore, 1999). Genetic markers such as nuclear ribosomal DNA (rDNA) and mitochondrial DNA (mtDNA) are useful in solving taxonomic complications arising in helminth species (Blair et al., 1996; Nolan and Cribb, 2005; Prasad et al., 2010). These regions are often utilized for phylogenetics and identification of platyhelminth species (Tkach et al., 2000; Kostadinova et al., 2003; Scholz et al., 2004). Ribosomal DNA (rDNA) has been the most extensively used nuclear sequence (Hillis and Davis, 1988; White et al., 1990). Phylogenetic analysis for related taxa (Miller et al., 1996), cryptic species identification (Collins and Paskewitz, 1996) and other population genetic studies (Mukabayire et al., 1999; Wörheide et al., 2002) have utilized the internal transcribed spacer (ITS) regions between the 18S, 5.8S, and 28S genes (or their homologues). Ribosomal DNA sequences are thought to be especially valuable for identifying digenean trematode species (Nolan and Cribb, 2005). Internal transcribed regions, which carry signals for rRNA transcript

processing, are thought to be quite useful for separating closely related species (Adlard et al., 1993; Stevenson et al., 1995; Wilson et al., 1995; Felleisen, 1997; Sugita et al., 1998). On the basis of ITS sequences, isolates of *Fasciola* spp. and *Fascioloides magna* from different territories have been distinguished among the fasciolid digeneans (Adlard et al., 1993). In nematode species, the utility of the ITS2 region for species discrimination has been proven, regardless of the parasite's life cycle stages (Campbell et al., 1995). In a study by Prasad et al. (2010) trematode parasites namely *Paragonimus westermani*, *Fasciolopsis* buski and *Fasciola gigantica* from Northeast India were distinguished based on ITS2 rDNA. A recent study by Le et al. (2017) demonstrated the use of 28S rDNA for species identification of 19 species in the superfamily Opisthorchioidea, in addition they were also able to identify and resolve six common human pathogenic heterophyids.

On the other hand, the mitochondrial genome has been found to demonstrate considerable genetic variation amongst closely related species which is a benefit for its use in taxonomic investigations (Hillis and Davis, 1988). MtDNA sequences have been found to be effective in assessing variation in parasite flarworm species. For example, in the analysis of the carcinogenic liver fluke *Opisthorchis viverrini*, mtDNA genes were utilised as genetic markers (Saijuntha et al., 2008). Mitochondrial sequences, particularly partial regions of mitochondrial protein-coding genes such as cytochrome c oxidase subunit 1 (CO1) and NADH dehydrogenase 1, have proven to be quite useful in identifying *E. granulosus* strains (Bowles et al., 1992). Morgan and Blair (1998) demonstrated the utility of nuclear rDNA, ITS and mitochondrial CO1 and ND1 sequences in distinguishing cryptic species, belonging to the 37 collar-spine *Echinostoma* group. Similarly, molecular investigations of the mtDNA sequences of CO1 were utilized

to differentiate *Echinococcus* species as well as *Echinococcus granulosus* strains into 10 genotypes (G1–G10) (Thompson and McManus, 2002). Several nematode species have also been genetically characterized using mitochondrial CO1 sequences (Okamoto et al., 2007; Yatawara et al., 2007; Fukuda et al., 2011). Utilizing DNA sequencing of the mitochondrial CO1 gene, Saijuntha et al. (2011) investigated genetic variation and relationships of four medically important echinostomes in South-East Asia. Moreover, Ghatani et al. (2014) demonstrated the ITS2 and CO1 regions to be effective for interspecies characterization within members of Family Gastrothylacidae from North-east India. As with rapid development in molecular biology technology, our ability to characterize helminth genetic diversity in terms of taxonomy and phylogeny, as well as individual populations at ever-larger genetic scales (i.e., from proteins to whole genomes) has also increased.

1.2.3. DNA barcoding as a tool for species identification

To help species identification and discovery in huge assemblages of life, DNA barcoding uses sequence diversity in small, standardized gene sections. The major barcode sequence for members of the animal kingdom is a 648-bp section of the mitochondrial CO1 gene (Hebert et al., 2003a). More than 95% of species in test assemblages of various animal groups were found to have distinctive CO1 sequences (Hebert et al., 2003b, 2004; Ward et al., 2005; Hajibabaei et al., 2005), indicating the effectiveness of these libraries in generating species specific identifications. The Barcode of Life Data System (www.barcodinglife.org) is a comprehensive bioinformatics platform which serves as a repository for the specimen and sequence records that serve as the

foundation of all barcode research (Ratnasingham and Hebert, 2007). A study by Moszczynska et al. (2009) demonstrated the use of the barcode region to resolve all specimens to the species level for 504 of 585 specimens of *Diplostomum* and *Tylodelphys*. In a study by Locke et al. (2010) sequences from the barcode region of mtCO1 were used to distinguish *Diplostomum* spp. from diverse fishes of the St. Lawrence River, Canada. In another study, DNA barcoding approach was used to reveal genetic diversity of *Schistosoma haematobium* from Zanzibar, Tanzania (Webster et al., 2013). On the other hand, Zhang et al. (2014) demonstrated the utility of DNA barcoding analysis in cestodes to identify 27 taeniid species including 9 in the genus *Echinococcus*, 2 in *Hydatigera*, 15 in *Taenia*, and 1 in *Versteria*. Another study by Van Steenkiste et al. (2015) revealed the utilization of CO1 barcode region for characterizing 23 families of digeneans and 6 orders of cestodes. A recent study by Gonçalves et al. (2021) evaluated the barcode region in the nematode genera *Anguillicola*, *Caenorhabditis*, *Heterodera*, *Meloidogyne*, *Onchocerca*, *Strongyloides*, and *Trichinella* and found out DNA barcodes as relevant tool for integrative taxonomy of nematodes. Similarly, Nyman et al. (2021) used the DNA barcoding approach to identify different cestode species in northern European marine and freshwater ringed seals, including the first report of *Ligula intestinalis* in seals.

1.2.4. Transcriptomic approach to helminth studies

Variations in the molecular make-up of NTD pathogens are being revealed by genomic research, functional genomics profiles of gene expression at the transcriptome and proteome levels, as well as metabolomic data (Preidis and Hotez, 2015). Large-scale

transcriptome/ proteome data can assist in high-throughput discovery of diagnostic biomarkers in NTD pathogens. Genome sequencing, genome resequencing, transcriptome profiling (RNA-Seq), DNA-protein interactions (ChIP-sequencing), and epigenome characterization are all examples of high-throughput (previously "next-generation") sequencing (de Magalhaes et al., 2010). Accompanied with single base pair resolution, a huge dynamic detection range (>8,000 fold), and low background signals, RNA-seq enables for unbiased transcriptome research (Wang et al., 2009). The application of RNA-seq to examine the immense diversity of RNA species is an apparent and intriguing one, opening up totally new avenues for improving human disease diagnosis and therapy. The RNA-seq method has proven to be a reliable method for analysing transcriptome profiles. The transcriptome, a collection of RNA molecules originating from protein-coding genes whose biological information is required by the organism at a specific time, is the first product of genome expression. Since the introduction of RNA-seq, several researches have been conducted to better understand the transcriptomes of both eukaryotes and prokaryotes. The utility of RNA-seq in providing a huge array of information for the trematode group has been recognized, and studies using RNA-seq as a primary option focused on diverse trematodes have been undertaken all over the world.

Transcriptome analysis of *Schistosoma mansoni*, has led to the discovery of new proteins that could be used as vaccine candidates and therapeutic targets (Verjovski-Almeida et al., 2003). Large-scale screenings of the schistosome transcriptome, proteome, and glycome have revealed a multitude of novel pharmacological, vaccine, and immunomodulatory targets that could be useful in the development of new anti-schistosome therapeutics (Hokke et al., 2007). Several trematodes, including *Schistosoma*

spp., *Fasciola hepatica*, *F. gigantica*, *Fascioloides magna*, *Clonorchis sinensis*, *Opisthorchis viverrini*, *O. felineus*, *Paragonimus westermani*, *P. skrjabini*, and *Fasciolopsis buski*, have had their transcriptomes resolved in recent years (Young et al., 2010 a, b; Young et al., 2011; Almeida et al., 2012; Cantacessi et al., 2012; Li et al., 2016; Pomaznoy et al., 2016; Biswal et al., 2018; Phuphisut et al., 2018). These investigations have added to our understanding of the molecular biology of these worms and provided important clues for the development of new intervention targets. Markers for possible use as vaccine candidates and diagnostic antigens have also been identified in the transcriptome of a cestode, *Taenia pisiformis* (Yang et al., 2012). A study by Zhou et al. (2016) on comprehensive transcriptome analysis in helminths was able to identify 700 differentially expressed genes which changed consistently during helminth infections. In addition, transcriptome analysis of *Echinostoma caproni*, a closely related echinostome parasite, revealed the presence of 180 distinct excretory-secretory (ES) proteins (Garg et al., 2013). Further eight potential therapeutic targets for treating *E. caproni* infections were identified. Similarly, Choudhary et al. (2015) illustrated the utility of *Paramphistomum cervi* transcriptome in identification of Cathepsin L as a vaccine candidate. Transcriptome analysis gives researchers access to an organism's whole gene expression repertoire. Furthermore, the information on genes and gene products obtained from such studies will be critical in understanding various aspects of parasite biology, assisting in the identification of genes responsible for virulence and adaptation in the host, and paving the way for the development of rapid diagnostic tools and drug targets.

1.2.5. *In silico* drug design and drug repurposing

The research and development of new therapeutic agents in the treatment of a variety of diseases is necessitated by the emergence of drug resistance. Drug resistance has previously been identified in helminths, primarily trematodes and nematodes (Geerts and Gryseels, 2000; Geerts and Gryseels, 2001; Wolstenholme et al., 2015; Bourguinat et al., 2017; Jimenez Castro et al., 2019). In *Schistosoma* spp., *Onchocerca volvulus* and *Ascaris lumbricoides* infections, reduced efficacy of currently used anthelmintics such as praziquantel, ivermectin, and benzimidazole has been reported (Alonso et al., 2006; Osei-Atweneboana et al., 2011; Crellen et al., 2016; Krücken et al., 2017; Furtado et al., 2019). Furthermore, a recent case of *Fasciola* spp. resistance to the widely used triclabendazole has surfaced (Fairweather et al., 2020).

The traditional approach of discovering a new molecular entity (NME) is a time-consuming and expensive process. Nonetheless, in the last few decades, an innovative strategy known as *in silico* drug design, which uses computational tools in target identification and prediction of novel medications, has gained traction. The reduced cost-effectiveness of computer-aided drug design (CADD) has sparked interest in additional indications, research and development. The predictive process of discovering novel pharmaceuticals based on the knowledge of a biological target is known as drug design (Madsen, 2002). Predicting whether a specific molecule will attach to a target and, if it will, how strongly will it bind is the primary objective in drug development. In today's drug development process, high-performance computing, data management software, and the internet are facilitating access to vast amounts of data and translating massive complex biological data into usable knowledge. The modern drug discovery method

employs a target-based approach that consists of seven steps: identification of the target protein, validation of the target protein, identification of the hit and lead molecule, optimization of the lead molecule, pre-clinical testing, chemical testing, new drug application (NDA) and Food and Drug Administration (FDA) approval (Figure 3). *In silico* drug design makes use of a diverse set of softwares. Many methodologies such as visualisation, homology, molecular dynamics, energy minimization, molecular docking, and quantitative structure-activity relationship (QSAR) are used. From preclinical discovery to late-stage clinical development, *in silico* drug design aids in the selection of only a potent lead molecule, potentially avoiding late-stage clinical failures and resulting in a significant cost reduction (Wadood et al., 2013). Few studies involving the utilization of *in silico* drug designing in helminths have been taken up worldwide. A study by Yadav et al. (2010) used a homology modelling approach to build the molecular structure of *Brugia malayi* glutathione-S transferase and then simulate its docking with many antifilarials. They gathered structural information and binding site mapping of BmGST for various antifilarials which could aid in the screening and design of new antifilarials or selective inhibitors for filariasis treatment. Using an *in silico* approach, Crowther et al. (2010) provided a list of potential drug targets in seven tropical disease pathogens including *Mycobacterium leprae*, *M. tuberculosis*, *Leishmania major*, *Trypanosoma brucei*, *T. cruzi*, *Plasmodium falciparum*, *P. vivax*, *Toxoplasma gondii*, *Brugia malayi* and *Schistosoma mansoni*. On the other hand, Alvarez et al. (2015) demonstrated the first computer-aided drug design approach against *Fasciola hepatica* cathepsins (*F. hepatica* cathepsin L3), a protease thought to be an attractive target against *F. hepatica's* infective phase, and were able to propose novel compounds such as FhCL3 inhibitors. Further,

Singh et al. (2016) used a homology modelling approach to determine the three-dimensional features of the enzyme SmSITRT2 from *Schistosoma mansoni*. While comparing virtual screening of hSirt2 (human Sirt2) with Smsirt2 (mouse Sirt2), they were able to identify ten lead compounds. Similarly, Ranjit and Surjith (2020) used *in silico* approach to unravel the anthelminthic property of a novel herbal formulation having rhizomes of *Curcuma longa* and *Cassia alata.*

To combat the ever-increasing problem of drug resistance, strategies like drug repurposing, which involves looking at existing, licensed, or exploratory treatments, could be very useful. It helps to alleviate some of the challenges that come with medication development by offering new applications to previously known molecules active against a causative agent. The application of antimalarials against a wide range of trematode illnesses, as well as other broader applications, is by far the most successful anthelmintic drug repurposing success story. Several antimalarial drugs have a broad-spectrum effect on *Schistosoma* species and some liver flukes (Panic et al., 2014). Artemisinin-based therapy and mefloquine also have some anti-schistosomal activity (Utzinger and Keiser, 2004; Keiser and Utzinger, 2007b; 2012). An immunosuppressive agent used for organ transplantation cyclosporine A has been found to possess anti-schistosomal property (Bueding et al., 1981). In vivo activity of OZ78 against the intestinal fluke *Echinostoma caproni*, the liver flukes *Fasciola hepatica* and *Clonorchis sinensis*, and other food-borne trematodes was shown to be highly powerful, whereas testing with *O. viverrini* indicated no activity at all (Panic et al., 2014; Keiser and Utzinger, 2007c). Again, Tribendimidine possesses trematocidal efficacy *in vivo* against *C. sinensis* and *O. viverrini*, in addition to its broad antinematodal activity (Keiser et al.,

2009; Xiao et al., 2013). Furthermore, many other drugs such as doxycycline, flubendazole, paromomycin, tamoxifen etc. have been also repurposed for the treatment of filariasis, taeniasis and neurocysticercosis (Panic et al., 2014).

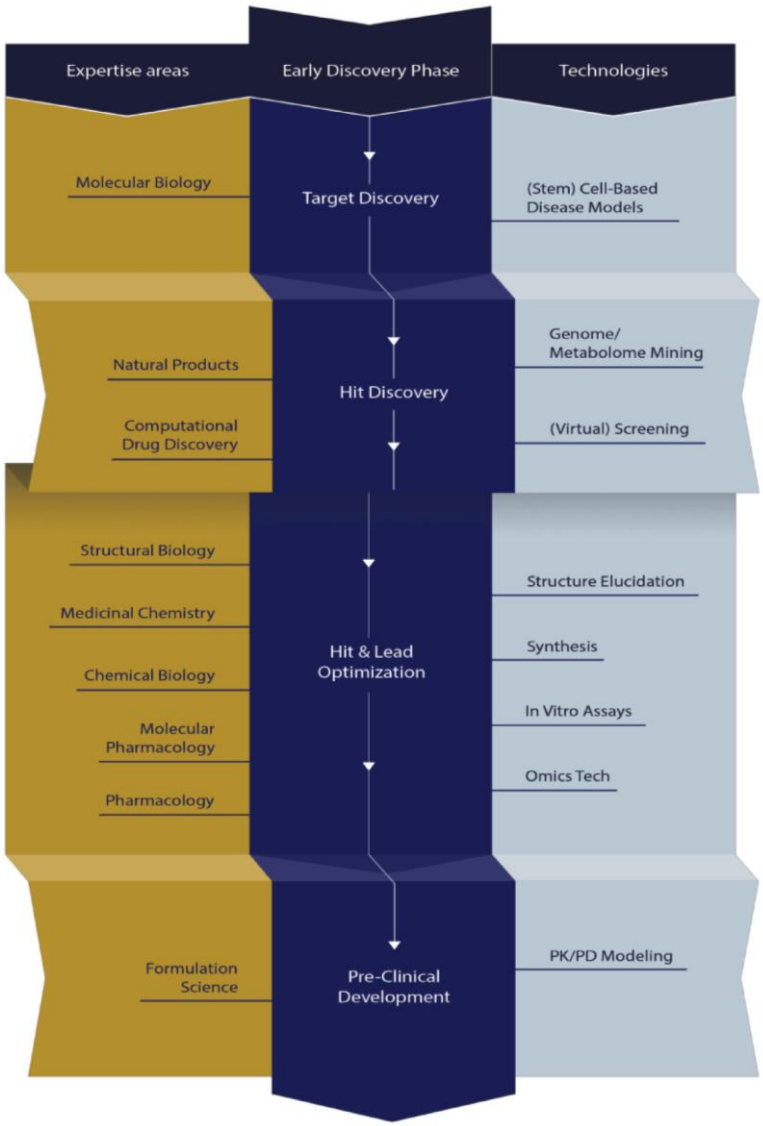


Figure 3. The stages of drug discovery are referred to as the drug discovery pipeline. The first stage is concerned with identifying prospective targets and compounds ('hits') whereas later phases are concerned with turning these hits into viable experimental medications and testing them in clinical trials.
(**Source**: https://www.universiteitleiden.nl/en/science/led3/pipeline)

1.3. Research gap

India has a population of approximately 1.3 billion people, with a large proportion of the population residing in deteriorating and poor surroundings. Such population groups are vulnerable to a number of NTDs along with various commonly occurring diseases. Poverty, malnutrition, lack of adequate food monitoring and sanitation, and decreasing economic conditions all exacerbate intestinal fluke illnesses (Graczyk and Fried, 1998). As a result, infections are more common in locations where traditional cultural norms encourage the use of raw or undercooked foods. According to a recent assessment by Hotez and Damania (2018), there are at least 11 NTDs prevalent in the country with worrying prevalence estimates of each of the NTDs. However, data on FBTs in India was missing in their work, indicating that these NTDs are still being neglected. Although there are frequent reports of this parasite from the nation, there is currently no reliable diagnostic tool available for the accurate diagnosis. Additionally, there is lack of national programmes for monitoring and evaluating the socioeconomic effect of echinostomiasis. Furthermore, the literature on this species is largely composed of morphological descriptions. The nucleotide database of the National Centre for Biotechnology Information (NCBI) contains sparse information about this parasite, with sequences primarily relevant to the ribosomal internal transcribed spacers (rDNA, ITS) and mitochondrial ND1 and CO1 regions. Genetic data on echinostomes is sparse, and full genome sequencing efforts are still in the works, especially for zoonotic echinostomes that infect humans. The detection of operculate, unembryonated, ovally shaped, pale yellow to yellowish brown eggs in faecal specimen is a widely used method for echinostomiasis diagnosis (Esteban et al., 2014). However, there is a great deal of

diversity in egg size among echinostome species, therefore such identification remains disadvantageous in determining the exact species (Huffman and Fried, 1990). At the same time, there is a high level of morphological similarities among adults of entire digenean spectrum which makes accurate identification of species difficult. Besides, this approach requires adequate laboratory facilities, extensive procedures and expertise which may be lacking in rural settings. Apart from this, it usually takes a long time for development of adults from the point of entry of metacercariae into the host and passage of parasite eggs in stools and also for the development of first symptoms. Currently, there are no means to diagnose helminth diseases at early stages of infection. Therefore, there is an urgent need to develop easy diagnostic tools for early detection of the infection and which would also provide a means for effective surveillance and reporting of the disease in the country. Furthermore, even the implementation of WHO programs with a broader approach against echinostomiasis have faced difficulties. Broad range anthelmintic drugs like praziquantel and albendazole, whilst they are currently used to treat a number of trematode diseases, carry a significant risk of developing drug resistance. As a result, this scenario calls for the identification of pharmacological protein targets and potential ligand compounds that can serve as leads for development of alternatives to these already in-use medications.

1.4. Objectives

In view of the public health significance particularly in context of rural India and the dearth of knowledge about *A. sufrartyfex*, the present study proposes to establish markers for easy identification of this species, sequence the transcriptome of this organism from different life cycle stages using RNA-seq approach with a view to identify suitable diagnostic and drug targets and also identify appropriate drug molecules against such targets using *in silico* approach. The current study was therefore undertaken with the following objectives:

1. Establishment of species-specific marker for *A. sufrartyfex* identification.

2. Stage-specific gene expression profiling of the parasite using transcriptomic approach and comparative analysis with related taxa for establishment of markers for diagnostic and control purposes.

3. Comparative analysis of the parasite's critical markers with closely related taxa and determination (*in silico*) of drug/ ligand molecules against the identified proteins/ drug targets.

Chapter 2: An emerging neglected food-borne trematodiasis: *Artyfechinostomum sufrartyfex* infection in children of Bihar, India

2.1. Background

Intestinal flukes are one of numerous food-borne trematodes that contribute significantly to human illness. Members of the Echinostomatidae family are food-borne intestinal flukes that cause the disease echinostomiasis (Haseeb and Eveland, 2000). At least 24 echinostome species are implicated in human echinostomiasis (Toledo and Esteban, 2016). Though the clinical aspects of human echinostomiasis are dependent on the parasite load, the symptoms are much more severe than those of other intestinal trematodes. The most common signs of human echinostomiasis include epigastric and stomach pain, tiredness, diarrhoea, and weight loss (Lane, 1915; Reddy and Varmah, 1950; Grayczk and Fried, 1998; Ghatani et al., 2018). Anaemia, headache, dizziness, stomachache, gastric pain, and loose stools are the most common symptoms reported in light-to-moderate infections. Eosinophilia, abdominal pain, copious watery diarrhoea, anaemia, edema, and anorexia are all symptoms of severe infections. Echinostomes have been observed to induce significant intestine and duodenal erosions, as well as catarrhal irritation (Chai et al., 1994). An increasing public health concern generated by this group of zoonotic parasites in these areas demands much attention and focus. There had previously been a few case reports in India regarding human echinostomiasis caused by the echinostome *Artyfechinostomum sufrartyfex.* This echinostomid fluke is becoming increasingly linked to an increase in food-borne intestinal infections in the country along

with infections reported from other vertebrate animals. Since 2004, child patients with chronic and acute diarrhoea and the passage of a reddish flatworm known locally as "lall kira" or "paterwa kira" in their faeces were being admitted to SSL Hospital and Research Centre Pvt Ltd in Sitamarhi, Bihar. The most common clinical symptom identified in infected children was chronic diarrhoea (lasting more than 14 days) along with severe malnutrition. After thorough questioning and completion of appropriate questionnaires, these patients disclosed a history of eating raw, undercooked, or partially boiled snails, usually from the endemic zone. The natives consider the commonly occurring snail species *Pila globosa* to be a source of strength for weak or ailing children and often consume them as dietary supplement. The primary focus of this study was to identify the intestinal echinostomid fluke passed by the patients in their faeces. Secondly, given the local food habits and practices around these parts of the country efforts were made to identify the source of infection to these child patients and establish the transmission cycle with a view to recognize the population at risk concerned with this infection. Thirdly, with regard to scarce information available about this parasite at both morphological and molecular front, establishing an easy method for identification of the species was a priority.

2.2. Materials and Methods

The study was approved by IEC (Institutional Ethics Committe), Sikkim University (SU/IEC/2017/04) along with necessary consent taken from the participating patients at SSL Hospital and Research Center, Sitamarhi, Bihar (Annexures I, II & III).

2.2.1. Clinical records

Child patients exhibiting symptoms associated with those of intestinal fluke infections were subjected to routine physical examination. Furthermore, these patients also underwent laboratory tests and investigation followed by questionnaires affiliated to food habits. Clinical data for the patients visiting the hospital related with this particular type of infection were maintained in the hospital records since the year 2004 and these records were also retrieved in the current study.

2.2.2. Survey and collection of clinical parasite samples and snails

A. sufrartyfex adults were obtained from the faeces of freshly infected child patients at Shri Subh Lal Hospital (SSL) and Research Center Pvt Ltd in Sitamarhi, Bihar. Ethanol preserved parasite clinical samples (adult stages) were brought to the laboratory of Department of Zoology, Sikkim University. In addition, efforts were made to procure environmental snail samples in order to screen for and recover the infective stage, metacercaria, from foci of infection particularly from those geographical locations where the patients had frequently visited for collection of snails (Figures 4 and 5). The snails were collected from their natural habitats in and around the infection-prone areas, handpicked, placed in sterile plastic containers and transported alive to the laboratory at Department of Zoology, Sikkim University.

2.2.3. Standard pepsin digestion and recovery of metacercaria

The collected snails were subjected to pepsin-HCL solution (0.5% Pepsin; 0.1% HCL) for digestion and recovery of the infective larval stage, metacercaria. Snails were

partially homogenized in a mortar pestle for about 1-1.5 hours, placed in the digesting medium and then incubated at 37˚C in a hot water bath. The resultant material was then filtered through fine sieve mesh, the filtrate was diluted and rinsed 3-4 times with distilled water. The diluted material was then taken in a petri dish and visualized under light microscope to detect the presence of metacercaria.

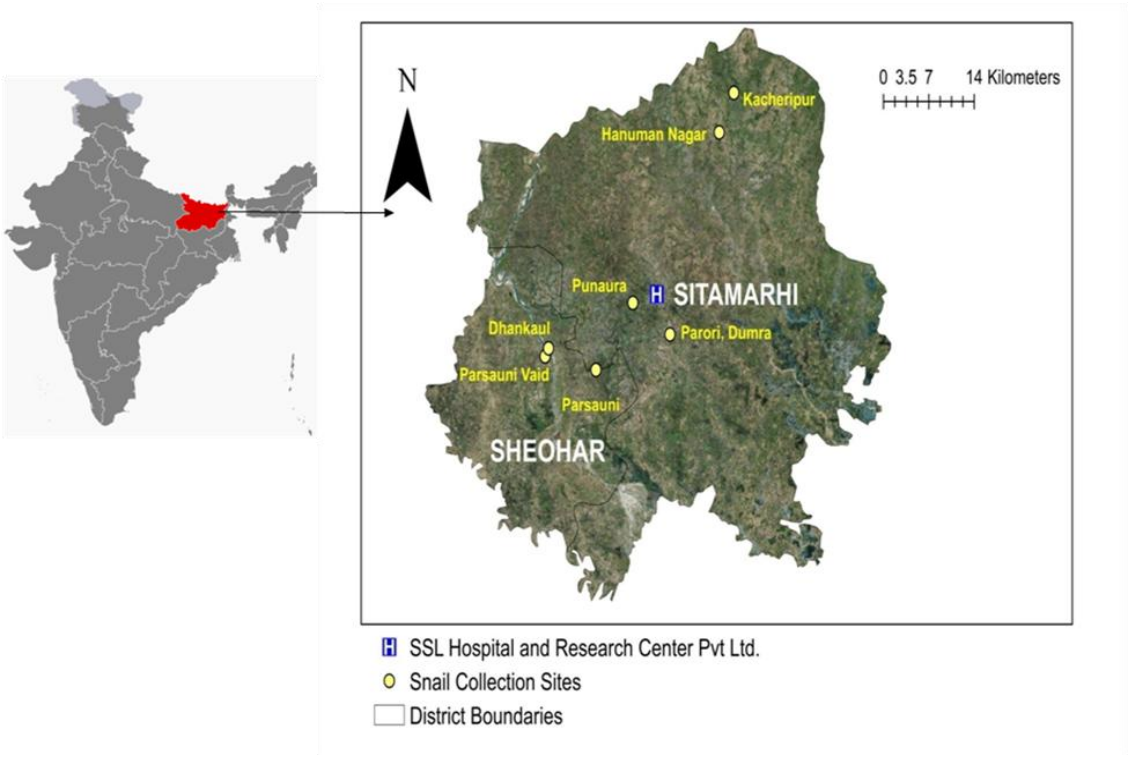


Figure 4. Map representing collection sites of clinical samples from SSL Hospital and Research Center, Sitamarhi (indicated as H) and snails (highlighted in yellow) from two districts of Bihar viz. Sitamarhi and Sheohar.

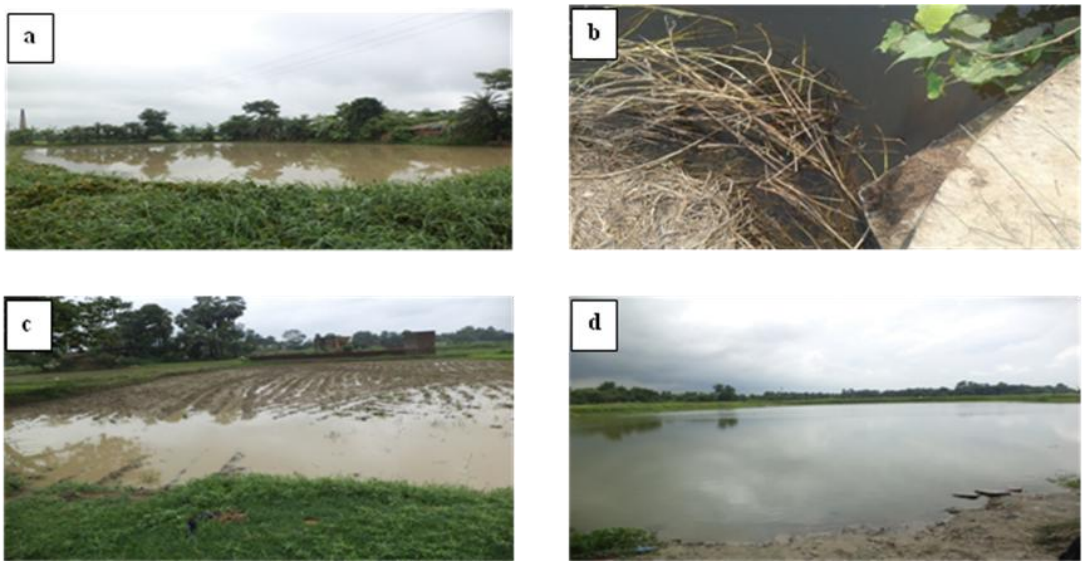


Figure 5. Snail collection sites in and around foci of infection: a) water logged area, b) ditches, c) paddy fields, d) ponds.

2.2.4. Morphological examination

The recovered specimen of adults was processed for morphological characterization using standard Semichon's acetocarmine method and studied using light microscopy. The metacercariae recovered from snails were also studied morphologically.

2.2.5. Molecular Characterization

2.2.5.1. DNA isolation, amplification, sequencing

Total genomic DNA was isolated from individual adult flukes using the standard phenol-chloroform protocol (Sambrook and Russell, 2006). In case of metacercariae, genomic DNA was isolated with the help of DNA isolation Kit (Macherey-Nagel kit, Germany) as per manufacturer's protocol. Around 200 cysts were taken for each reaction. PCR amplification for both adult as well as metacercarial stages were performed in

replicates using universal trematode primers. Three different gene regions were targeted viz. nuclear ribosomal 28S (partial D1-D3 region), internal transcribed spacer 2 and mitochondrial cytochrome c oxidase subunit 1 (Table 2). PCR condition used for amplification was initial denaturation at 95°C (5 mins); 35 cycles: 94°C (1 min), 55°C (1 min), 72°C (1 min); final elongation at 72°C (10 mins). MtDNA barcode region was also amplified following the same PCR condition mentioned above except for annealing temperature set at 49°C. PCR amplicons were purified and then outsourced for sequencing to Macrogen, Korea.

Table 2. Primers utilized for amplification of 28S, ITS2, mtCO1 and mtCO1 (barcode) gene regions.

Gene region	Primer Name	Primer Sequence (5' - 3')	Primer length	Annealing temp	References
28S	DIG12	AAGCATATCACTAAGCGG	18	55˚ C	Tkach et al., 2000
	1500R	GCTATCCTGAGGGAAACTTCG	21		
ITS2	3S	GGTACCGGTGGATCACTCGGCT CGT	26	55˚C	Bowles et al.,1995; Blair et al., 1997
	A28	GGGGATCCTGGTTAGTTTCTTTT CCTCCGC	29		
mt CO1	JB 3	TTTTTTGGGCATCCTGAGGTTTA T	24	55˚C	Tkach et al., 2000
	JB4.5	TAAAGAAAGAACATAATGAAAA TG	24		
mt CO1 (barcode)	DICE1F	TTWCNTTRGATCATAAG	17	49˚C	Moszczynska et al., 2009
	DICE14R	CCHACMRTAAACATATGATG	20		

2.2.5.2. Sequence annotation and submission

Nucleotide sequences for all the gene regions utilized herein viz. 28S, ITS2 and mtCO1 were analyzed through Sequence Scanner 2 (Applied Biosystems) to check the overall trace scores and online servers like nucleotide BLAST for proper authentication. High quality sequences (>40 PHRED score) were only considered for further evaluation.

The sequenced ITS2 region plus the flanking 5.8S and 28S sequences were annotated in the ITS2 database (Keller et al., 2009) using the ‘Annotate’ tool and model ‘Metazoa' in order to retrieve the exact ITS2 sequences for each of the stages. Sequences for ribosomal 28S, ITS2 and mtCO1 were submitted to GenBank, NCBI. MtCO1 barcode sequence along with their chromatogram files and specimen details were submitted to Barcode of Life Data Systems (BOLDSYSTEM).

2.2.5.3. Sequence analysis

For comparative purpose, all available sequences of 28S, ITS2 and CO1 gene regions for different species belonging to the family Echinostomatidae were retrieved from the NCBI nucleotide database (Table 3, 4 and 5). Multiple sequence comparison by log-expectation (MUSCLE) from EMBL-EBI website (www.ebi.ac.uk/Tools/ msa/muscle/) was used for multiple sequence alignment of individual genes. The aligned sequences were then manually edited using the tool BioEdit version 7.2.5 (Hall, 1999). Sequence Identity Matrix (SIM) was generated with the help of the BioEdit software in order to see the genetic similarity between each gene region of the query species sequence with other available sequences within the family Echinostomatidae. After input of the aligned sequences, SIM was computed from the ‘Alignment’ menu with default settings.

2.2.5.4. MtDNA barcode generation

The Barcode of Life Data Systems (BOLDSYSTEMS), developed at Centre for Biodiversity Genomics in Canada, is a cloud based data storage analysis platform that

provides an integrated environment for the use of DNA barcode data (http://www.boldsystems.org/). The mtDNA barcode region was amplified in order to establish an easy method for quick and rapid identification of the species and also to provide a reference for identification in future at the molecular arena. Submission of both the DNA sequence and the specimen data were done to the BOLD Systems version 4 under project name ASU-Trematode Barcoding. In order to check whether the barcodes generated were species-specific or not, the query barcode sequences were subjected to similarity searching against the species level barcode records in BOLD Systems using the BOLD Identification System (IDS).

Table 3. Sequence details of 28S region for species of the Family Echinostomatidae retrieved from NCBI database.

Species name	**Genbank Accession**	**Host**	**Place of origin**
Artyfechinostomum sufrartyfex adult*	MH236132	*Homo sapiens*	India: Bihar
Artyfechinostomum sufrartyfex metacercaria*	MH236133	*Pila globosa*	India: Bihar
Uroproctepisthmium bursicola	KT956938	*Ardea alba*	Ukraine: Kherson region
Artyfechinostomum sufrartyfex	KF781303	*Sus scrofa*	India: Meghalaya
Neoacanthoparyphium echinatoides	KT956922	*Viviparus acerosus*	Slovakia: Danube at Gabcikovo
Moliniella anceps	KT956921	*Planorbarius corneus*	Lithuania: Jodkrante
Echinoparyphium ecurvatum	KT956913	*Radix ovata*	Slovakia
Hypoderaeum conoideum	KT956918	*Anas platyrhynchos*	Ukraine: Kherson region
Echinostoma revolutum	AY222246	*Mesocricetus auratus*	United Kingdom
Petagifer bilobus	KT956945	*Plegadis falcinellus*	Ukraine: Kherson region
Drepanocephalus spathans	MF351546	*Nannopterum brasilianu*	-
Euparyphium capitaneum	KP009620	*Anhinga anhinga*	USA: Mississippi
Ribeiroia ondatrae	KT956956	*Pelecanus erythrorhynchos*	USA: North Dakota
Chaunocephalus ferox	KT447522	*Ciconia nigra*	Ukraine
Isthmiophora melis	KT359583	*Apodemus agrarius*	Poland
Petasiger exaeretus	KY284001	*Rutilus rutilus*	Hungary
Cathaemasia hians	KT956947	*Planorbis planorbis*	Czech Republic
Neopetasiger islandicus	KT956924	*Aechmophorus occidentali*	USA: Near Michigan, North Dakota

***Sequences generated during the present study**

Table 4. Sequence details of ITS2 region for species of the Family Echinostomatidae retrieved from NCBI database.

Species name	**GenBank Accession**	**Host**	**Place of origin**
Artyfechinostomum sufrartyfex adult*	MH237730	*Homo sapiens*	India: Bihar
Artyfechinostomum sufrartyfex metacercaria*	MH237731	*Pila globosa*	India: Bihar
Echinoparyphium mordvilkowi	KJ542640	*Valvata piscinalis*	Lithuania: Ula River
Hypoderaeum conoideum	KM980480	-	China: Liuzhou
Echinoparyphium recurvatum	KJ435282	*Physella cubensis*	-
Euparyphium albuferensis	AJ564384	*Rattus norvegicus*	Spain: Valencia
Artyfechinostomum sufrartyfex	KF781301	*Sus scrofa*	India: Tripura
Echinostoma malayanum	JF412729	*Pomacea canaliculata*	Thailand: Khon Kaen
Echinostoma revolutum	KP342426	*Duck*	China: Liuzhou
Echinostoma caproni	AJ564382	*Rattus norvegicus*	Egypt: Cairo
Echinostoma trivolvis	GQ463127	*Lymnaea elodes*	USA: Indiana
Ribeiroia ondatrae	AY761142	*Helisoma trivolvis*	USA: Shady Lakes, Albequerque
Pegosomum asperum	KX097838	*Egretta alba*	Czech Republic: Prerov
Cathaemasia hians	AY761146	*Biomphalaria glabrata*	-
Drepanocephalus spathans	KY636260	*Nannopterum brasilianus*	-
Petasiger phalacrocoracis	KM972997	*Phalacrocorax carbo*	-

***Sequences generated during the present study**

Table 5. Sequence details of mtCO1 gene for species of the Family Echinostomatidae retrieved from NCBI database.

Species name	**Genbank ID**	**Host**	**Place of origin**
Artyfechinostomum sufrartyfex adult*	MH253673	*Homo sapiens*	India: Bihar
Artyfechinostomum sufrartyfex metacercariae*	MH253674	*Pila globosa*	India: Bihar
Hypoderaeum conoideum	GU324946	-	Thailand
Echinostoma malayanum	GU324941	*Lymnaea rubiginosa*	Thailand
Artyfechinostomum sufrartyfex	KF781304	*Sus scrofa domestica*	India: Meghalaya
Isthmiophora melis	KT359580	*Apodemus agrarius*	Poland
Echinostoma hortense	AF096227	-	-
Isthmiophora hortensis	AB189980	*Procyon lotor*	Japan: Wakayama
Pegosomum asperum	KX097854	*Egretta alba*	Czech Republic: Prerov
Echinostoma miyagawai	KP455599	-	-
Echinostoma revolutum	GQ463026	*Lymnaea elodes*	USA: Indiana
Echinostoma paraensei	AF025825	-	-
Echinoparyphium recurvatum	GU324945	-	Thailand
Echinostoma caproni	AF025830	-	-

***Sequences generated during the present study**

2.3. Results

2.3.1. Clinical scenario

A total of 170 cases of *A. sufrartyfex* infection have been recorded in the hospital from the year 2004 alongwith 11 death cases. The child patients admitted at the hospital exhibited various symptoms such as diarrhea (acute and persistent), watery stool, vomiting, swelling feet and body, loss of appetite, weakness with passage of reddish fluke in stool and vomitus (Table 6). Also general and systemic examination revealed conditions such as edema, dehydration (some/ severe) and malnutrition (Table 7). In addition, biochemical tests in blood revealed a median value of 9800 cmm for total lymphocyte count. In terms of differential lymphocyte count, median values of 60% (Neutrophil), 32% (Leuckocytes) and 0% (Monocytes) were observed. The median value for haemoglobin was found to be 9.5% while erythrocyte sedimentation rate (ESR) had a median value of 13.5 mm/h. On the other hand, serum analysis revealed median values of 1.0 mg% for creatinine, 25 Iu/l for serum glutamic pyruvic transaminase with median values of 135 meq/l, 3.5 meq/l and 95 meq/l for sodium, potasssium and chloride ions, respectively (Table 8). Questionnaires related to food habits revealed that the patients had a history of consuming raw snails. The most popular snail collection sites for the natives were near deep paddy fields with snails usually recovered from water. Snails were rarely collected from the town area. A single case involving transmission of infection without consumption of snail has also been observed. All the remaning 169 cases were either due to intake of raw snails or consumption of snail fluids after breaking their shells (Table 9). Many relapse cases have also been observed possibly due to continued intake of raw snails even after completion of treatment procedure (praziquantel 75 mg/kg in 3 divided

doses for 2 days along with supportive treatment in most cases such as Inj. ceftriaxone, Inj. metronidazole, Inj. hydrocortisone, livosalbutamol, nebulization, IV fluids, O_2 inhalation, zinc etc).

Table 6. Clinical symptoms of patients.

Symptoms	**No. of positive cases (total cases = 170)**
Persitent /chronic diarrhea	150
Acute diarrhea	15
Night blindness	7
Passage of reddish worms in stool/ vomitus	49
Vomiting	153
Fever	67
Loss of appetite	169
Weakness	169
Swelling feet and whole body	78
Rashes	2

Table 7. Generic and systemic examination results.

Signs	**No. of positive cases (total cases = 170)**
Anaemia	170
Febrile	8
Some Dehydration	129
Severe Dehydration	41
Oedema	82
Malnutrition grading (Indian Academy of Pediatrics) I, II, III, IV, with (K) or without (K) with K= with Oedema	
I	11
I(K)	10
II	38
II(K)	20
III	31
III(K)	28
IV	15
IV(K)	22
Not grouped	01
No malnutrition	06
Pedal & generalised edema	82
Glossitis	6
Angular stomatitis	4
Xerosis Conjunctiva	8
Urticaria (Rashes)	2

Table 8. Blood and serum biochemical tests summary.

Parameters		Median values	Normal range
a. Blood			
TLC (cmm)*		9800	3,000 - 9,500
DLC (in %)	N*	60	40–80
	L*	32	20-40
	M*	0	2–10
Hb (in %)		9.5	11-13
ESR mm/h		13.5	3 -13
b. Serum			
Creatinine (mg %)		1.0	0.0–0.7
SGPT Iu/l*		25	5-60
Na^{+} (meq/l)*		135	136 -144
K^{+}(meq/l)*		3.5	3.6- 5.1
Cl^{-}(meq/l)*		95	99-108

*TLC= Total Lymphocyte Count; DLC= Differential Lymphocyte Count *N= Neutrophil *L=Leuckocyte* M=Monocyte; Hb=Haemoglobin; *ESR=Erythrocyte Sedimentation Rate *SGPT= Serum Glutamic Pyruvic Transaminase; *Na^{+}= Sodium ion *K^{+}= Potassium ion* Cl^{-} = Chloride ion

Table 9. Potential source of infection to children in affected areas.

Place of living	Number of cases	Exposure of fresh water bodies or indulgence in local eating habits without ensuring the safety of food & drinks			
		Consuming raw snails recovered from water	Consuming raw snails recovered from dry fields during summer after ploughing	Intake of ghongha pani (snail fluids after breaking their shells)	No history of taking raw snails or ghongha pani (body fluid)
Near bank of pond and deep paddy fields contaminated with human and animal excreta	168	120	42	5	1
Town area	2	1	-	1	-

2.3.2. Snail sample analysis

A total of eight infection hit sites located within two districts of Bihar was surveyed for collection of snail samples following which metacercaria prevalence was determined for the collected snails (Table 10). Survey carried out over a period of two years from various foci of infection led to a total collection of 1841 snails. All the eight surveyed sites were found to show prevalence of metacercariae in snails. Metacercaria prevalence ranged from as high as 48.24% in Punaura to lowest value of 16.12% from Hanumannagar. Further, prevalence of 45.73% was observed in Parsauni, 30.58% in Dhankaul site 2, 30.50% in Parori, 25.69% in Dhankaul site 1, 24.90% in Kacheripur and 23.07% in Parsauni vaid.

2.3.3. Morphological identification

All adult specimens were identified with the help of morphological criteria based on characters considered important from taxonomic standpoint (Kostadinova, 2005) such as: collar spines small, 43 to 45 in double rows, cirrus very long, club shaped, dextral or medians, testes large occupying most of posterior half of the body and intestinal bifurcation just anterior to ventral sucker. Based on these characters, the adult clinical samples were identified as the echinostomid trematode *Artyfechinostomum sufrartyfex*. For the larval stage metacercaria, presence of characteristic echinostome collar spines could be observed. However, the specific identity could not be confirmed based on these characters alone (Figure 6).

Table 10. Prevalence of metacercaria from snails collected from foci of infections (Sitamarhi and Sheohar districts, Bihar).

Site	District	GPS Coordinates	Collection months 2016	Total no. of snails collected	Collection months 2017	Total number of snails collected	No. of infected snails	Prevalence (%)
Punaura	Sitamarhi	N26°35'11.1"	July	85	February	68	220	48.24
		E 085° 27' 39.8"	August	71	July	88		
			October	78	August	66		
Parsauni	Sitamarhi	N 26 °31' 03.3"	July	50	February	38	118	45.73
		E 85° 24' 29.4"	August	41	July	42		
			October	42	August	45		
Hanumanagar,	Sitamarhi	N 26° 45'54.7"	Not surveyed		February		10	16.12
		E 085° 34' 50.1"			July	62		
					August			
Kacheripur,	Sitamarhi	N26° 48'24.9"	August	84	February	62	67	24.90
		E 085° 35' 69.0"	October	73	July	50		
Parsauni vaid	Sheohar	N26° 31' 43.6"	August	54	February	62	54	23.07
		E 085° 19' 81.7"	October	68	July	50		
Parori, Dumra	Sitamarhi	N26° 33' 9.69"			February	56	54	30.50
		E 085° 30' 47.52"	Not surveyed		July	58		
					August	63		
Site 1, Dhankaul	Sheohar	N26° 31' 34.3"	Not surveyed		February	59	46	25.69
		E085° 27' 39.8"			July	62		
					August	58		
Site 2, Dhankaul					February	79	63	30.58
					July	62		
					August	65		

Diagnostic morphological characters pointed out the specimens of both the adult as well as infective metacarcarial stage to be of *A. sufrartyfex*. However, in order to authenticate this finding at the molecular level, the three gene regions viz. 28S, ITS2 and mtCO1 were successfully amplified for both the lifecycle stages and sequences are now available in Genbank with the following accession numbers: MH236132, MH236133; MH237730, MH237731 and MH253673, MH253674, respectively.

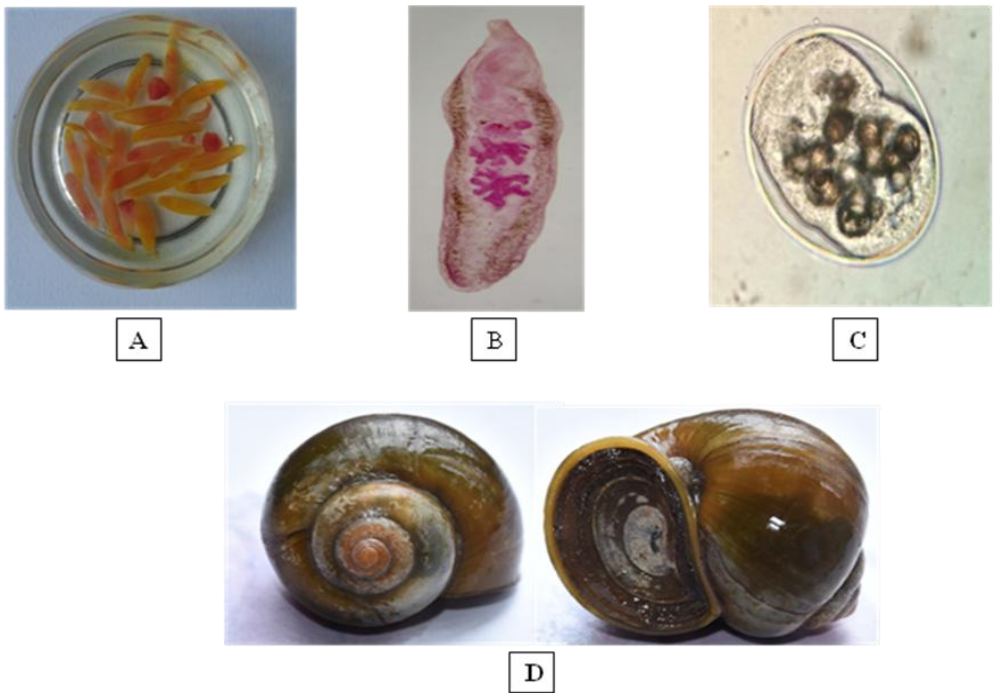


Figure 6. A) Flukes in normal saline after being passed by patient; B) Whole mount adult specimen; C) Metacercariae (under 400x magnification) recovered from *Pila globosa*. D) Second intermediate host *Pila globosa* of the fluke.

2.3.4. Molecular characterization

2.3.4.1. Nucleotide BLAST

The sequences of both the lifecycle stages were subjected to nucleotide BLAST analysis with default parameters. The individual query gene regions viz. 28S, ITS2 and

mtCO1 were found to have sequence length of 1048 bp, 433 bp and 343 bp, respectively. BLAST analysis for 28S gene region of both the lifecycle stages revealed maximum sequence identity with Genbank accession KF781303.1 *Artyfechinostomum sufrartyfex* (99%) from Shillong, India. Similarly, ITS2 gene region of both lifecycle stages revealed maximum sequence identity with the Genbank accession JF412727.1 *Echinostoma malayanum* (99%) from Khon Kaen, Thailand. Finally, the mtCO1 gene region (both the life cycle stages) showed maximum sequence identity (100%) with Genbank accession NC037150.1 *Artyfechinostomum sufrartyfex* from Shillong, India.

2.3.4.2. Intra- and inter- specific variations

Multiple aligned sequences of individual genes were used for sequence identity calculation to see the genetic variations between the query sequences with that of other species within the family Echinostomidae. In case of the 28S gene region, intraspecific variation was observed to range from 0.1% betweeen *A. sufratyfex* adult (Genbank accession MH236132) and *A. sufrartyfex* pig isolate (Genbank accession KF781303) to 0.3% betweeen *A. sufrartyfex* metacercaria sequence (Genbank accession MH236132) and *A. sufrartyfex* pig isolate (Genbank accession KF781303). Interspecific variations ranged from 3.0% with *Patagifer bilobus* (Genbank accession KT956945) to 10.6% with *Uroproctepisthmium bursiocla* (Genbank acession KT956938) (Table 11). For ITS2 gene region, intraspecific variation was observed to range from 0.2% between *A. sufrartyfex* pig isolate (Genbank accession KF781301) and *A. sufratyfex* adult (Genbank accession MH237730) to 0.7% *A. sufrartyfex* pig isolate (Genbank accession KF781301) and *Echinostoma malayanum* (Genbank accession JF412729). Interspecific variation ranged

from 7.4% with *Echinostoma trivolvis* (Genbank accession GQ463127) to 19.5% with *Pegosomum asperum* (Genbank acession KX097838) (Table 12). In case of the mtCO1 gene region, intraspecific variations were observed to range from 0% among *A. sufratyfex* adult (Genbank accession MH253673), *A. sufrartyfex* metacercaria sequence (Genbank accession MH253674 and *A. sufrartyfex* pig isolate (Genbank accession KF781304) to 0.6% among *Echinostoma malayanum* (Genbank accession GU324941), *A. sufrartyfex* adult, *A. sufrartyfex* metacercaria and *A. sufrartyfex* pig isolate. Interspecific variations ranged from 14.5% with *Echinoparyphium recurvatum* (Genbank accession GU324945) to 20.6 % with *Isthmiophora melis* (Genbank acession KT359580) (Table 13). Intra-specific values for the gene regions studied herein showed minor variations with previously published sequence of *A. sufrartyfex* from India and Thailand. Inter-specific values were found to be quite high with other species within the Family Echinostomatidae. In all the three gene regions studied, there was clear distinction in the intra- and inter-specific values. This suggests that the sampled species is indeed *A. sufrartyfex*.

2.3.4.3. *A. sufrartyfex* specific mtDNA barcode

Morphological and molecular analyses confirmed the species of clinical adult samples and the infective metacercarial stage as *A. sufrartyfex*. MtCO1 barcode sequences were then generated and submitted to BOLD database. Unique barcodes for both the lifecycle stages of the parasite was then obtained with Barcode identification number BIN URI- BOLD: ADM2711 for our submissions. The barcode sequences had a length of 777 bp (Figures 7 A & B). Similarilty searching across the BOLD database

revealed the sequences to be highly species-specific with minor intra-specific variation of 0.4-1.3% with other *A. sufrartyfex* sequences within the BOLD database. On the other hand, the closest similarity with other species found to be with that of the echinostomid *Echinostoma miyagawai* with a very low identity value of 84.05%. This difference of 15.95% is well above the species cut-off limit of 2% as established by Herbert et al. (2003) for mtCO1 barcodes of animals.

Table 11. SIM for 28S. Data in green indicate intraspecific values; Data in red indicate interspecific values.

GenBank Accessions	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
MH236132 *A. sufrartyfex	ID																	
MH236133 *A. sufrartyfex	**99.8**	ID																
KF781303 *A. sufrartyfex*	**99.9**	**99.7**	ID															
KT956922 *Neoacanthoparyphium echinatoides*	**96.3**	**96.1**	**96.2**	ID														
AY222246 *Echinostoma revolutum*	**96.4**	**96.2**	**96.3**	96.7	ID													
KT956945 *Patagifer bilobus*	**97.0**	**96.8**	**96.9**	97.2	97.6	ID												
KT956921 *Moliniella anceps*	**96.0**	**95.8**	**95.9**	96.3	96.1	97.1	ID											
KT956913 *Echinoparyphium recurvatum*	**96.1**	**95.9**	**96.0**	96.7	96.5	97.2	97.1	ID										
KT956918 *Hypoderaeum conoideum*	**95.9**	**95.8**	**95.8**	96.6	96.6	97.2	96.8	99.4	ID									
KP009620 *Euparyphium capitaneum*	**90.2**	**90.0**	**90.1**	90.7	90.9	91.3	91.6	90.8	90.8	ID								
MF351546 *Drepanocephalus spathans*	**92.1**	**92.0**	**92.0**	92.4	92.5	92.9	93.6	93.2	93.1	93.7	ID							
KT359583 *Isthmiophora melis*	**92.3**	**92.1**	**92.2**	93.1	93.0	93.7	93.6	93.6	93.7	93.5	96.2	ID						
KY284001 *Petasiger exaeretus*	**92.1**	**92.0**	**92.0**	92.9	92.8	93.6	93.1	93.4	93.5	93.2	96.0	98.1	ID					
KT447522 *Chaunocephalus ferox*	**93.0**	**92.8**	**92.9**	93.9	93.8	94.4	93.9	94.0	94.1	93.6	96.7	96.7	96.8	ID				
KT956924 *Neopetasiger islandicus*	**92.6**	**92.4**	**92.5**	93.7	93.4	94.0	94.1	94.1	94.2	93.4	96.7	96.7	96.4	97.7	ID			
KT956956 *Ribeiroia ondatrae*	**91.4**	**91.2**	**91.3**	92.0	91.6	92.1	91.9	92.0	92.0	90.4	92.4	93.4	93.0	93.7	93.2	ID		
KT956947 *Cathaemasia hians*	**92.0**	**91.9**	**92.0**	92.5	92.6	93.1	93.0	92.6	92.6	91.4	94.4	95.3	95.2	95.0	94.0	95.5	ID	
KT956938 *Uroproctepisthmium bursicola*	**89.6**	**89.4**	**89.5**	89.5	89.9	90.3	90.4	89.7	89.9	85.7	87.9	88.3	88.3	88.6	88.4	86.0	87.6	ID

***Sequences generated in the present study**

Table 12. SIM for ITS2. Data in green indicate intraspecific values; Data in red indicate interspecific values.

GenBank Accessions	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
MH237730 *A sufrartyfex	ID															
MH237731 *A sufrartyfex	**99.4**	ID														
KF781301 *A sufrartyfex*	**99.8**	**99.3**	ID													
JF412729 *Echinostoma malayanum*	**99.4**	**99.8**	**99.3**	ID												
KM980480 *Hypoderaeum conoideum*	**90.0**	**89.8**	**88.9**	**89.5**	ID											
KJ435282 *Echinoparyphium recurvatum*	**89.8**	**89.5**	**88.7**	**89.2**	98.3	ID										
AJ564384 *Euparyphium albuferensi*	**89.8**	**89.5**	**88.7**	**89.2**	98.0	98.8	ID									
KP342426 *Echinostoma revolutum*	**90.9**	**90.6**	**89.8**	**90.3**	87.2	87.8	87.5	ID								
AJ564382 *Echinostoma caproni*	**90.9**	**90.6**	**89.8**	**90.3**	86.6	87.2	86.9	96.0	ID							
GQ463127 *Echinostoma trivolvis*	**92.6**	**92.3**	**91.5**	**92.0**	87.8	88.3	88.1	97.1	97.1	ID						
AY761142 *Ribeiroia ondatrae*	**83.3**	**83.0**	**82.5**	**82.7**	84.5	84.7	85.3	82.1	81.2	82.3	ID					
KY636260 *Drepanocephalus spathans*	**84.3**	**84.0**	**83.2**	**83.8**	86.1	86.6	86.6	84.0	82.2	84.0	90.3	ID				
KX097838 *Pegosomum asperum*	**81.5**	**81.3**	**80.5**	**81.0**	84.4	83.8	83.8	80.6	79.7	80.9	90.0	89.1	ID			
KM972997 *Petasiger phalacrocoracis*	**82.6**	**82.3**	**81.5**	**82.1**	86.6	87.2	87.2	82.0	80.8	82.0	89.7	91.0	89.3	ID		
AY761146 *Cathaemasia hians*	**84.0**	**83.8**	**83.0**	**83.5**	86.6	87.2	87.2	81.7	80.8	82.5	89.1	91.9	87.7	92.8	ID	
KJ542640 *Echinoparyphium mordvilkowi*	**90.3**	**90.0**	**89.2**	**89.8**	96.3	95.1	95.4	88.3	87.2	88.9	84.7	86.4	83.8	86.4	86.4	ID

***Sequences generated in the present study**

Table 13. SIM for mtCO1. Data in green indicate intraspecific values; Data in red indicate interspecific values.

Sequences with GENBANK ID	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
MH253673 *A. sufrartyfex	ID													
MH253674 *A. sufrartyfex	**100**	ID												
KF781304 *A. sufrartyfex*	**100**	**100**	ID											
GU324941 *Echinostoma malayanum*	**99.4**	**99.4**	**99.4**	ID										
AF025825 *Echinostoma paraensei*	**81.1**	**81.1**	**81.1**	**81.6**	ID									
KP455599 *Echinostoma miyagawai*	**83.8**	**83.8**	**83.8**	**84.4**	89.4	ID								
AF025830 *Echinostoma caproni*	**83.3**	**83.3**	**83.3**	**83.3**	90.0	91.1	ID							
GU324945 *Echinoparyphium recurvatum*	**85.0**	**85.0**	**85.0**	**85.5**	89.4	91.6	92.2	ID						
GQ463026 *Echinostoma revolutum*	**83.3**	**83.3**	**83.3**	**83.8**	90.0	91.6	91.1	95.5	ID					
KX097854 *Pegosomum asperum*	**80.5**	**80.5**	**80.5**	**81.1**	86.1	85.5	86.6	85.5	85.5	ID				
GU324946 *Hypoderaeum conoideum*	**85.0**	**85.0**	**85.0**	**85.0**	82.2	84.4	84.4	84.4	82.2	82.7	ID			
KT359580 *Isthmiophora melis*	**79.4**	**79.4**	**79.4**	**80.0**	83.8	86.1	86.1	84.4	83.8	88.3	83.3	ID		
AF096227 *Echinostoma hortense*	**81.1**	**81.1**	**81.1**	**80.5**	84.4	85.0	85.5	82.7	81.6	87.2	84.4	93.3	ID	
AB189980 *Isthmiophora hortensis*	**81.6**	**81.6**	**81.6**	**81.1**	85.0	85.5	86.1	83.3	82.2	87.7	85.0	93.8	99.4	ID

***Sequences generated in the present study**

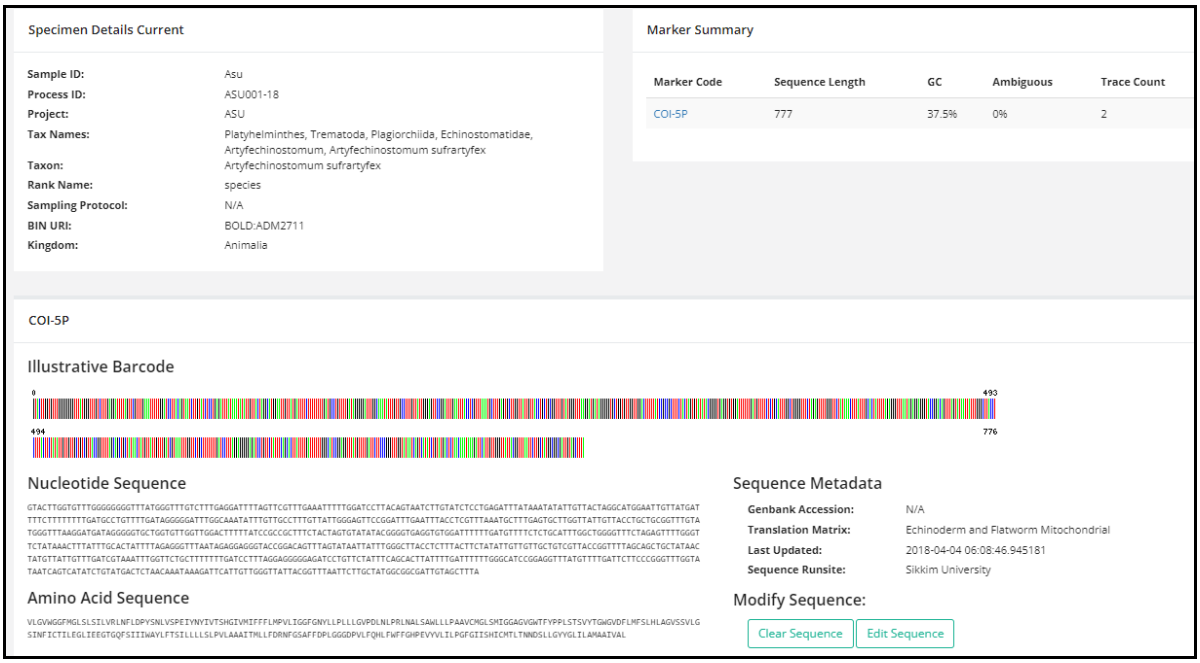


Figure 7A. Barcode sequence details of *A. sufrartyfex* adult generated in the present study (BIN URI- BOLD: ADM2711).

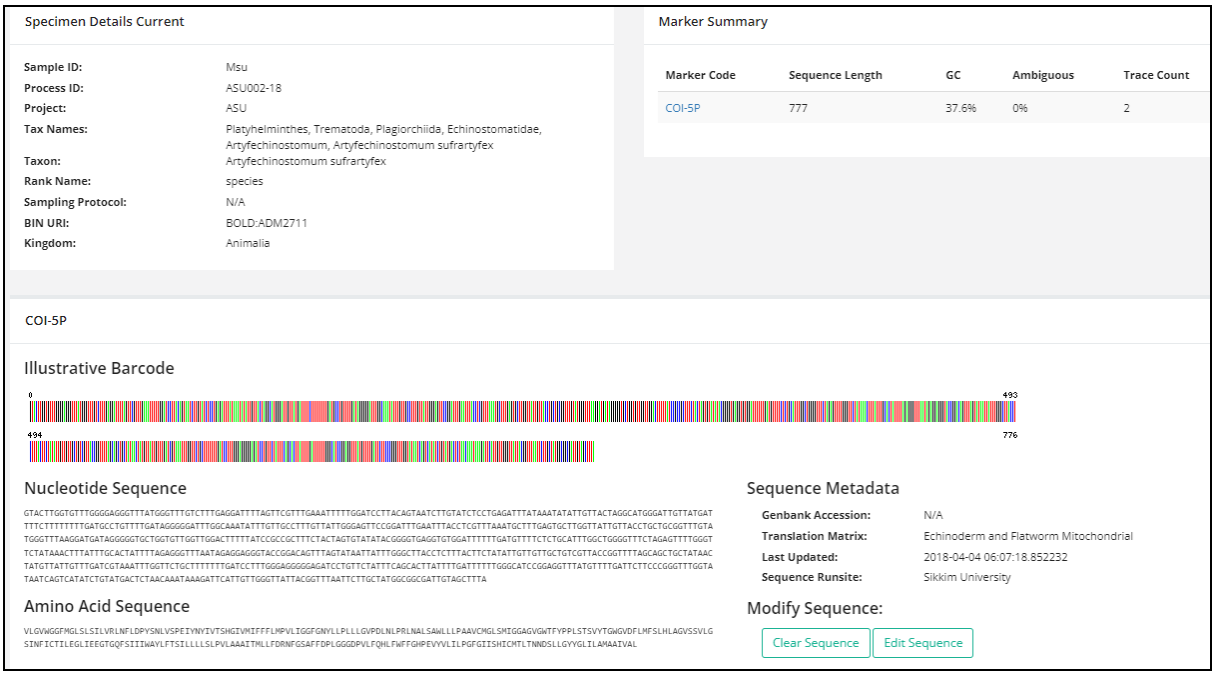


Figure 7B. Barcode sequence details of *A. sufrartyfex* metacercaria generated in the present study (BIN URI- BOLD: ADM2711).

2.4. Discussion

The current study reports a total of 170 cases of *A. sufrartyfex* infection from SSL hospital and Research Center Pvt Ltd, Sitamarhi, Bihar alongwith an alarming figure of 11 death cases. The study also reports the presence of infective larval stage metacercaria of the parasite for the first time in *Pila globosa* examined from infected hit geographical foci of infections from two districts of the state viz. Sitamarhi and Sheohar (Prasad et al., 2019). In case of the snail samples, highest metacercaria prevalence of 48.19% was found from one of the foci of infection Punaura, Sitamarhi. This data indicates that out of every two snails, one was found to harbour the infective metacercarial stage thereby adding to the significantly higher prevalence of the infection in the studied area. Such dense infection among the snails provides higher risk of infection transmission to the at-risk populations. However, the prevalence of metacercaria found in the current study is less compared to that found in Cambodia where the prevalence of the parasite's metacercaria has been shown to be much higher at 70% (Sohn et al., 2017). The infective larval stage metacercariae recovered from *Pila globosa* commonly found in and around the foci of infections clearly show the population at risk particularly those communities directly involved in indulging with food habit including snails as dietary supplements. This may be due to exposure of the patient to both grossly contaminated water and infected snails, harbouring the infective larval stage of the parasites. A single curious case was found where the patient had no history of taking snails or body fluids. The most common symptoms exhibited by the infected patients were diarrhoea (acute and persistent), loss of apetite, vomiting and weakness. General and systemic examination of the patients also revealed malnutrition to be common with Grade II showing highest representation

amongst other grades. Biochemical tests in blood and serum revealed infected patients to have lower haemoglobin level (9.5%) as against the normal range of (11-13%) while the median values for remaining parameters of blood and serum were close to the normal range. This is also indicative of the fact that all the patients were anaemic. In line with the current study, another study has found a significant association between anaemia burden and intestinal helminth infection in rural communities of Edo state, Nigeria (Osazuwa et al., 2011).

Morphological study of the adult specimens revealed important taxonomic characters highly similar to that of *A. sufrartyfex,* while the infective larval stage exhibited characters similar to those of echinostome species. To authenticate this findings molecular approach was undertaken. Sequence analysis in the current study was able to identify the species accurately using molecular markers in the nuclear ribosomal 28S and ITS2 regions as well as mitochondrial CO1 gene. Unique DNA barcodes were also generated from both the lifecycle stages of the fluke so as to provide an easy and accurate method of identification. The study also establishes transmission cycle for the infection through molecular identification of the infective metacercaria from the second intermediate host *Pila globosa*. In view of both morphological study and the different molecular markers utilized in the current study the findings clearly point out that the fluke procured from infected child patients at the hospital is indeed *A. sufrartyfex* and the infective larval stage metacercaria recovered from *Pila globosa* belongs to the same species. Nucleotide BLAST analyses revealed all the three gene regions used in the present study viz. 28S, ITS2 and mtCO1 closely matched Genbank Accessions KF781303.1 *A. sufrartyfex,* JF412727.1 *E. malayanum* and NC037150.1 *A. sufrartyfex,*

respectively. Similarly, sequence analysis utilizing the query sequences from procured specimens also showed least distance with that of *A. sufrartyfex* (Shillong, India) and *E. malayanum* (Thailand) compared to higher genetic variation with other species. It was also observed that the interspecific variation was the highest as compared to intraspecific variations in case of mtCO1 out of the three genes studied herein. Intraspecific variation was nil and interspecific variation was found to range from 14.5%-20.6%. The nuclear ribosomal 28S and ITS2 regions have been used by previous workers for species identification as well as distinguishing cryptic species in the nematodes and trematodes (Adlard et al., 1993; Morgan and Blair, 1998; Tkach et al., 2000; Kostadinova et al., 2003; Scholz et al., 2004; Prasad et al., 2011; Le et al. 2000). Similarly mtCO1 gene regions have also been utilized in discriminating closely related species by previous workers (Thompson and McManus, 2002; Okamoto et al., 2007; Yatawara et al., 2007; Fukuda et al., 2011; Saijuntha et al., 2011; Ghatani et al., 2014).

Owing to the morphological similarities, *A. sufrartyfex* and *E. malayanum* are regarded as synonyms by several workers (Joe, 1963; Mukherjee and Ghosh, 1968; Mohandas, 1971; Premvati and Pande, 1974). However, a study from Thailand utilizing ITS1 and ITS2 gene regions had shown that these two species are distinct (Tantrawatpan et al., 2013). In contrast to this study, the current analysis showed that the two species are indeed same as inferred from the ITS2 as well as mtCO1 gene regions and may be considered as synonyms. The DNA markers analyses showed that the current specimens as well as *A. sufrartyfex* from another part of the country closely matched with *E. malayanum* reported from Thailand with negligible difference. At the same time, the difference with other species was greater enough which clearly indicated that these two

species are indeed same. Based on the result of the present study, it is suggested that the two species should be regarded as synonyms as proposed by many previous workers. This is also suggestive of the fact that *A. sufrartyfex* has a much broader geographical distribution which otherwise was considered to be restricted to India only. Very recently, Pham et al. (2022) have also confirmed that the two species are same through their analysis of the parasite mitochondrial genome.

Many of the workers have used the DNA barcode approach previously with fruitful success (Moszczynska et al., 2009, Locke, 2010; Webster et al., 2013; Zhang et al., 2014; Van Steenkiste et al., 2015; Gonçalves et al., 2021; Nyman et al., 2021). In addition, a review on DNA barcoding in parasites and vectors affecting humans found the barcoding technique to be accurate (Ondrejicka et al., 2014). Generation of unique DNA barcodes for both the lifecycle stages in the current study is bound to act as a reference for identification in future. From identification within the BOLD systems, the intraspecific divergence was found to range as low as 0.4 to 1.3% with *A. sufrartyfex* and the interspecific divergence was considerably higher at 15.95% with *Echinostoma miyagawai* which supports mtCO1 barcode as an exceptional marker for the identification of this species.

Chapter 3: *In silico* identification of important diagnostic candidates from multi-stage transcriptome of *A. sufrartyfex*

3.1. Background

For years, trematode infections and helminthiases in general have been ignored. As a result, there are ineffective monitoring, surveillance, and control systems in place to combat these diseases, owing mostly to a lack of proper diagnostic instruments. Traditional diagnostic methods rely on faeces examinations to detect the presence of eggs, which can be inconvenient at times. Gastroduodenal endoscopy or colonoscopy has detected human echinostomiasis in patients with chronic gastrointestinal symptoms (Jung et al., 2014). Immunological and molecular diagnostic techniques are not completely available for human echinostomiasis at the moment. A definitive test for echinostomes diagnosis utilising molecular approaches is not yet available (Ghatani et al., 2018). For the development of acceptable diagnostic markers, various investigations on the secretome or excretory secretory proteins (ESPs) have been published for different helminth parasites such as *Clonorchis sinensis* (Choi et al., 2003), *Fasciola hepatica* (Gonenc et al., 2004; Robinson et al., 2009), and *Toxocara canis* (Iddawela et al., 2007). The secretome of an organism is defined as the subset of proteins secreted by the cell, often known as excretory/secretory proteins (ESPs). These proteins are critical during parasitic infection because they regulate the host's immune system, allowing the parasite to survive inside the host (Tjalsma et al., 2000). ESPs are laced with short signal peptides (typically 16-30 amino acids long) found at the N-terminus of most newly generated proteins headed for the secretory route (Blobel and Dobberstein, 1975; Kapp et al., 2009).

Given the ESPs' important role in controlling the host immune system and the resultant pathologies associated with the immunological response, identifying the separate components of the ES has been a major focus of diagnostic research (Garg and Ranganathan, 2012). Diagnostic antigens such as DNase II, two serine proteases, two kinds of trypsin (Wang et al., 2014) and 31 kDa antigens (Sun et al., 2015) were identified from *Trichinella spiralis* ES proteins. Moreover, synthetic disaccharide (DiM) glycan was identified as a promising antigenic tool for the serodiagnosis of human toxocariasis (Elefant et al., 2016). A total of 452 ES proteins from human and plant parasitic nematodes were found in a research of genome-wide identification and analysis of ES proteins in nematodes, which were indicated to represent promising targets for parasite control (Gahoi et al., 2019). Subsequently, Verissimo et al. (2019) recently published a list of putative diagnostic targets derived from *Schistosoma japonicum* ES proteins isolated from mature and immature eggs in faeces, as well as *ex vivo* adults.

Further, identification of epitopes in the target proteins is required for the development of effective antibodies capable of detecting and neutralizing bioactive proteins. Computational B-cell epitope prediction is a significant and time-saving option in this regard (Kozlova et al., 2018). On the basis of their spatial configuration B-cell epitopes can be classified into continous (linear or sequential) and discontinous (non linear or conformational) epitopes (Potocnakova et al., 2016). A wide variety of computational tools are avaliable in the public domains for predicting both continous and discontinous epitopes.

In terms of public health, the lack of adequate surveillance and control strategies for helminth illnesses, particularly in rural parts of India, is a serious concern. The current

study is an *in silico* attempt to screen for viable diagnostic targets against the rising food-borne echinostomid *A. sufrartyfex* infection in the country.

3.2. Materials and methods

3.2.1. Collection of parasite and snail samples

The collection of parasite samples for the current objective remains the same and is described in Chapter 2 under the materials and methods section.

3.2.2. Recovery of encysted and excysted metacercariae

The collected snails, *Pila globosa*, were digested and the infective larval stage, metacercaria, was recovered using a pepsin-HCL solution (0.5 % pepsin; 0.1 % HCL). Whole snails were partially homogenized for about 1-1.5 hours in a mortar pestle, then placed in the digesting medium and incubated at 37°C in a hot water bath. The filtrate was diluted and washed 3-4 times with distilled water after it was filtered through fine sieve mesh. To detect the presence of metacercariae, the diluted sample was placed in a petri dish and examined under a light microscope (Olympus CX31, USA) (Prasad et al., 2019). The metacercariae were then placed in a petridish with excystation media containing 0.05% trypsin, 0.5% bile salts, and Earles balanced salts (about 20-25 metacercariae). After an incubation period of one to two hours at 37°C, excysted metacercariae were collected at one- to two-hour intervals (Fried and Reddy, 1997; Saxton et al., 2008).

3.2.3. RNA extraction, library preparation, Illumina sequencing and assembly

TRIzol reagent (Invitrogen, USA) was used to extract total RNA in duplicates from three lifecycle stages, namely encysted metacercaria, excysted metacercaria, and adult stages, according to the manufacturer's protocol. The RNA samples were then outsourced to Bionivid Pvt. Ltd., Bengaluru for transcriptome sequencing and assembly. Only approved samples were taken forward for library creation after passing quality control (QC). For mRNA purification, RNA samples with an RNA integrity number (RIN) of 7.0 and above were employed. The library layout for the *A. sufrartyfex* life cycle stages was paired end. mRNA was isolated from 1 μg of intact total RNA using oligo-dT beads (TruSeq RNA Sample Preparation Kit, Illumina). In the presence of divalent cations, pure mRNA was fragmented at 90°C. Random hexamers and Superscript II Reverse Transcriptase were used to reverse transcribe the fragments (Life Technologies). RnaseH and DNA polymerase were used to make second strand cDNA from the first strand template. Beckman Coulter Agencourt Ampure XP SPRI beads were used to clean the cDNAs. Random fragmentation of the cDNA sample was used to create the sequencing library, which was then followed by 5' and 3' adapter ligation, end-repair, and the insertion of a 'A' base, as well as SPRI cleanup. For the enrichment of the adapter-ligated fragments, the generated cDNA library was amplified using PCR. The NanoDrop spectrophotometer (ThermoFisher Scientific) was used to quantify the various libraries, and a Bioanalyzer was used to evaluate their quality (Agilent Technologies). Following that, the adapter-ligated fragments were PCR amplified and gel purified. The library was put into a flow cell for cluster creation. Through bridge amplification, each fragment was multiplied into unique, clonal clusters. The templates were sequenced using Illumina SBS

technology after cluster creation was completed. For quality control, the NGS QC Toolkit v2.3 (Patel and Jain, 2012) was used, and only reads with (≥70%) HQ bases (≥30 phred) were kept. Trinity assembler was used to perform a de-novo assembly of the pooled transcriptome data (Grabherr et al., 2011). Individual samples were also processed using the Kallisto tool (Bray et al., 2016). Based on sequence and length similarity, CD-HIT-EST (Li and Godzik, 2006) was utilized to group together comparable nucleotide sequences inside the assembly. Only a representative transcript was kept from a set of transcripts that were 80% identical in length and sequence similarity. Furthermore, transcriptome data from individual libraries were first mapped to the 'representative transcripts' using Bowtie to determine the degree of expression of these 'representative transcripts' (Langmead et al., 2009). The read abundance of each transcript in terms of TPM (Transcripts per million) was then estimated using RSEM (Li and Dewey, 2011). Bionivid's amelioration process was used to filter the transcripts in the assembly based on coverage and depth (Bankar et al., 2015). The coverage was set at 70% and the average depth was set at 5X. The raw sequence reads generated in this investigation were submitted to Sequence Read Archive (SRA) at NCBI under the Bioproject accession number PRJNA738306.

3.2.4. Conceptual translation and annotation of transcripts

For assembled transcripts, open reading frames (ORFs) were predicted using CLC Genomics Workbench version 9.0 and only the longest ORF for each transcript was used further. With the help of CLC Genomics Workbench 9.0, these ORFs were conceptually translated into their corresponding protein sequences at default parameters. InterProScan

version 5 was used to classify the conceptually translated protein sequences (Jones et al., 2014). InterPro and gene ontology (GO) terms were retrieved using the InterProScan tool based on their similarity with conserved domains and protein families. According to the gene ontology database, these predicted protein sequences were further classified into GO categories (molecular function, cellular component, and biological process) (Ashburner et al., 2000; Carbon et al., 2021).

3.2.5. Secretome analysis and identification of diagnostic targets

Parasite ESPs (secretomes) are a valuable source of diagnostic antigens for parasitic illness early identification. To identify the secretome, signal peptides that corresponded to either classical or non-classical secretion processes were screened in the conceptually translated protein sequences. SecretomeP version 2.0 (Petersen et al., 2011) was used to predict non-classically secreted proteins (using Neural Network scores 0.9 and other default choices) while SignalP version 4.1 (Bendtsen et al., 2004) was used to detect classically secreted proteins (selecting eukaryote organism categories and other default options). Using TargetP version 1.1 (Emanuelssonet al., 2000) with a specificity of 0.90 and non-plant options, predicted secreted proteins from both the classical and non-classical pathways were further evaluated for sub-cellular localization so as to filter out mitochondrial proteins. Finally, using TMHMM version 2.0 (Krogh et al., 2001), the protein sequences were filtered to eliminate proteins with transmembrane helices. As a result, the *A. sufrartyfex* secretome was derived for proteins that successfully cleared this pipeline. Only those ESPs with higher expression in all three lifecycle stages, based on transcripts per million (TPM) expression values and with a threshold cut-off value of 0.5,

were shortlisted as diagnostic targets. Furthermore, proteins having vertebrate homologs were ruled out.

3.2.6. B-cell epitope prediction

In order to seek for unique epitopes in the predicted diagnostic targets, BepiPred version 2.0 was used to predict linear (continuous) B-cell epitopes for all of the targets using default threshold value of 0.5 (Jespersen et al., 2017). For validation and shortlisting, the predicted epitopes were compared to the epitopes generated by other epitope prediction tools such as Bcpreds (El Manzalawy et al., 2008) and DiscoTope version 2.0 (Kringelum et al., 2012). In the case of Bcpreds, the prediction was made using the default epitope length and a specificity of 75%. Discontinuous epitopes were predicted by DiscoTope version 2.0 utilizing three-dimensional structures of individual diagnostic targets predicted by RaptorX (Källberg et al., 2012), with an epitope identification threshold of -3.7. Only epitopes with common lengths of residues (at least five amino acids or longer) predicted by at least one of the other two tools, Bcpreds and DiscoTope, were evaluated. Furthermore, only those epitopes were picked that were found to be unique after BlastP and UniProt database search. An epitope was considered unique only when the sequence identity was <90% in these database searches (Miles et al., 2017).

3.3. Results

3.3.1. Putative protein annotation

A total of 86,512 transcripts were generated after the primary transcriptome assembly. Following the primary assembly, a total of 74,145 transcripts were generated from the secondary asssembly. Finally, a total of 41,678 representative transcripts could be generated from the final transcriptome assembly for the three lifecycle stages of *A. sufrartyfex* (Table 14). With an average transcript length of 1318.20 and a N50 value of 1826, the transcripts accounted for approximately 54 Mbp. ORFs could be predicted for 30,545 transcripts out of the 41,678 total transcripts. For each transcript, a number of ORFs were predicted; however, only the longest ORF was chosen for further investigation. Conceptually translated protein sequences were blasted against the Swiss-Prot database for additional annotation. The database yielded hits for 29,047 translated sequences, whereas the remaining 1,498 sequences yielded no hits. A cut-off e-value of <1e-10 was used to screen these blast hits, and 9,681 sequences were discovered at this significant level. Finally, a total of 8,160 predicted protein sequences were annotated with GO keywords, comprising 5,663 sequences related to molecular function, 1,074 sequences related to cellular component, and 1,424 sequences related to biological activity. The transcriptome sequences were mostly related with "Protein Binding" (molecular function), "Integral Component of Membrane" (cellular component), and "Proteolysis" (biological process) according to the GO annotations (Figure 8).

Table 14. Transcriptome assembly details.

Assembly details	Primary Assembly	Secondary assembly	Final Assembly
Number of Final Transcripts	86512	74145	41678
Final Transcriptome Length (bp)	85686445 (~85.6 Mbp)	70509785 (~70.5 Mbp)	54940002 (~54.9 Mbp)
Minimum Transcript Length (bp)	301	301	301
Maximum Transcript Length (bp)	31123	31123	31123
Average Transcript Length	990.46	950.97	1318.20
N50	1473	1377	1826
GC%	50.03 %	50.13 %	49.51 %

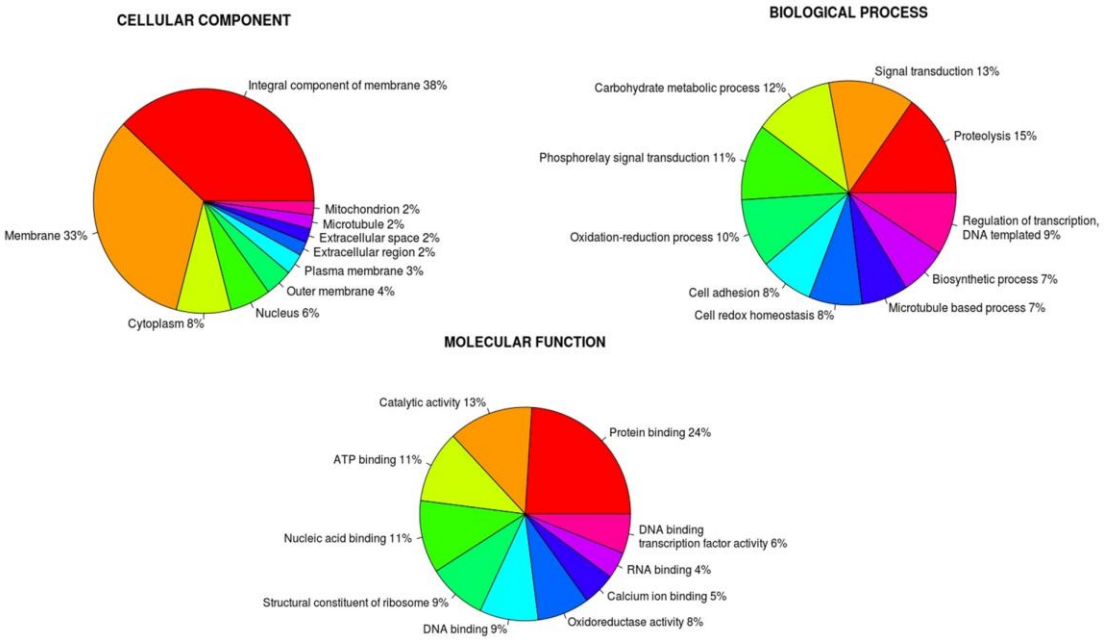


Figure 8. Top 10 GO (Gene Ontology) terms linked with molecular function, cellular component, and biological process that could be mapped onto the assembled transcripts or predicted proteins inferred from *A. sufrartyfex* transcriptome.

3.3.2. Identification of diagnostic markers from secretome

SignalP classified 1,248 (4.08%) of the 30,545 ORFs in the current dataset as classically secreted proteins and 29,297 (95.91%) as non-secretory proteins. After that, the non-secretory proteins were fed into SecretomeP, which identified 1,237 (4.22%) of them as non-classical secretory proteins. Hence, a total of 2,458 (8.04%) proteins were retrieved as classical and non-classical secretory proteins from the two tools. TargetP was used to screen for subcellular proteins, and it was able to identify 157 (6.38 %) of them as mitochondrial proteins with a 95% specificity. After removing these 157 proteins, TMHMM was used to predict transmembrane proteins from the remaining 2,328 probable secretory proteins. Since 598 (25.68%) of the proteins had one or more transmembrane helices, they were excluded from the dataset. Finally, the *A. sufrartyfex* secretome was predicted to have 1,730 proteins (5.66 %) (Figure 9).

Based on their higher TPM values, common occurrence in all the three lifecycle stages of *A. sufrartyfex* and exclusion of vertebrate homologs, 14 putative proteins were eventually identified as potential diagnostic candidates from the secretome dataset (Figure 10). Three of these targets were antioxidant proteins namely, thioredoxin peroxidase, protein dj1 beta and superoxide dismutase. These antioxidant proteins protect helminths from reactive oxygen species produced by host effector cells such as macrophages, neutrophils, and eosinophils. Proteases such as Cathepsin L, cathepsin L-like cysteine proteinase, cathepsin B-like cysteine proteinase and haemoglobinase were also identified as probable targets from the dataset. Parasites secrete these proteases in order to enter their hosts, move through tissues, and breakdown hostglobins. Other targets included glutamine tRNA ligase (probable), 28S ribosomal protein (probable), a

transducer protein guanine nucleotide-binding protein subunit alpha A, a troponin complex protein Troponin I4, an RNA binding protein enhancer of rudimentary homolog, a calcium binding protein stromal interaction molecule homolog and an actin binding protein Huntington interacting protein related 1. tRNA ligases are known to be involved in tRNA aminoacylation. Translational control is primarily controlled by ribosomal proteins. Troponins govern muscle contraction as part of the contractile apparatus. In all eukaryotes, RNA binding proteins are involved in post-transcriptional processes. Transducers play a role in a variety of signalling pathways including G-protein coupled receptors. Calcium binding proteins have a role in calcium signalling in cells. Actin binding proteins are typically involved in cytoskeleton control.

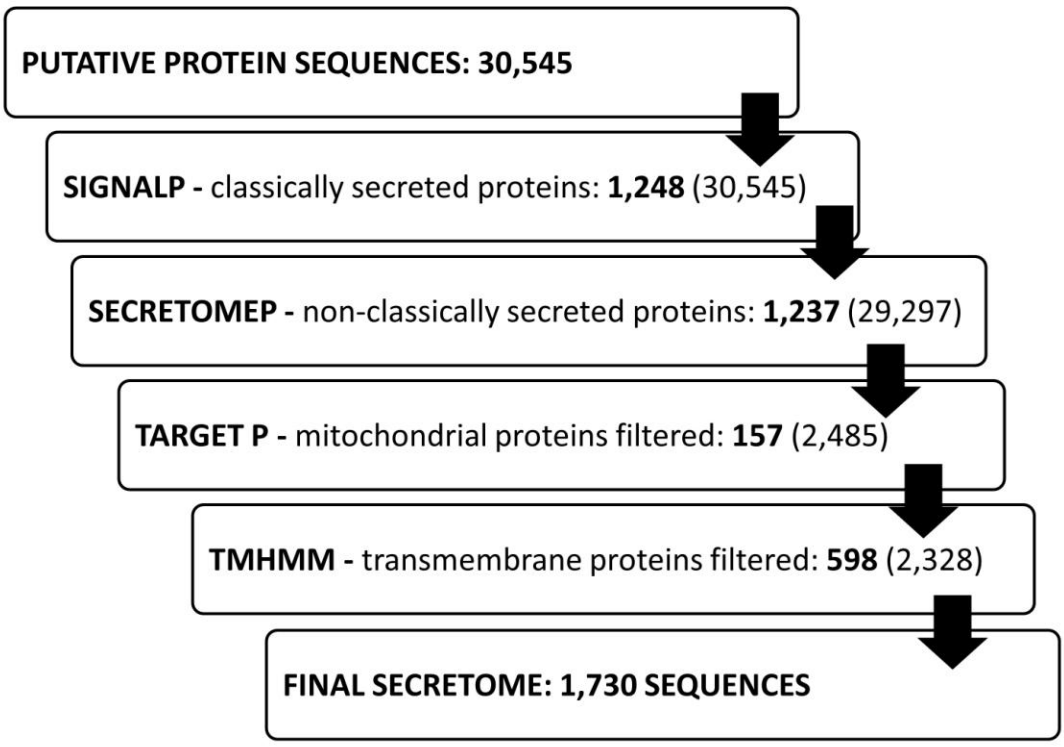


Figure 9. Secretome prediction pipeline from the transcriptome of *A. sufrartyfex*.

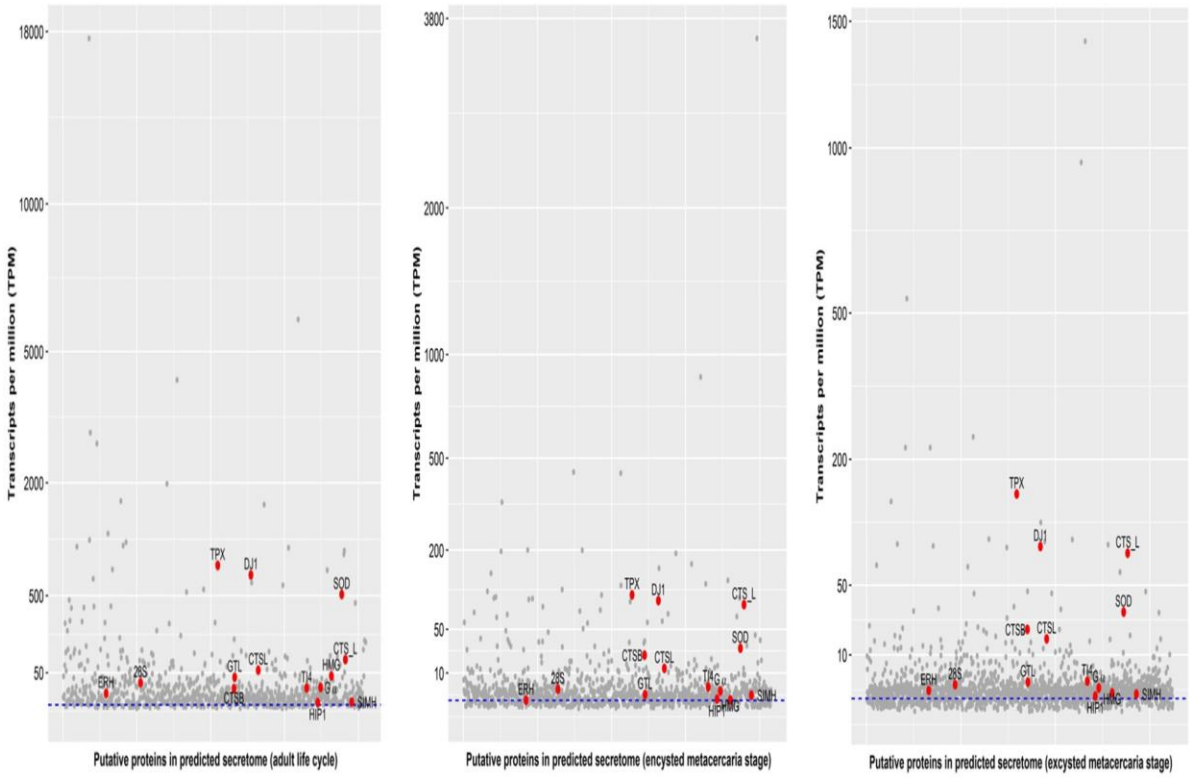


Figure 10. Secretome expression profile of the three life cycle stages of *A. sufrartyfex*. The shortlisted diagnostic targets common to all the three parasite stages are highlighted in red. (TPX= thioredoxin peroxidase, DJ1= protein dj1 beta, SOD= superoxide dismutase, CTS_L= cathepsin L-like cysteine proteinase, CTSL= cathepsin L, HMG= haemoglobinase, GTL= glutamine tRNA ligase (probable), 28S= 28S ribosomal protein (probable), Gα= guanine nucleotide-binding protein subunit alpha A, TI4= Troponin I4, CTSB= cathepsin B-like cysteine proteinase, ERH= enhancer of rudimentary homolog ,STIM= stromal interaction molecule homolog, HIP1= Huntington interacting protein related 1; blue dashed line represents the threshold value of 0.5).

3.3.3. B-cell epitope analysis

Among the antioxidants, four epitopes were found for thioredoxin peroxidase, one epitope for Protein dj-1beta and two epitopes for superoxide dismutase. Of these, only a single epitope of thioredoxin peroxidase was found to be unique as per the criteria set in

the current analysis. In the case of proteases, cathepsin L-like proteinase had five epitopes, of which three were unique. Cathepsin L had four epitopes of which two were unique, cathepsin B-like cysteine proteinase had five epitopes and three were unique, and haemoglobinase had seven epitopes of which four were distinct. Similarly, nine epitopes were found for glutamine tRNA ligase (probable) of which one was unique, four epitopes were found for 28S ribosomal protein (probable) and two epitopes were specific. Stromal interaction molecule homolog had four epitopes and three were found to be unique. Interestingly, all of the epitopes for enhancer of rudimentary homolog and Huntington interacting protein related 1 epitopes identified in the current study were unique. All of these distinct epitopes suggest that these diagnostic targets can be further evaluated for specific *A. sufrartyfex* diagnosis. One of the epitopes of stromal interaction molecule homolog (59.72 %) had the lowest sequence identity with an unnamed protein product of *Echinostoma caproni*, while an epitope of Huntington interacting protein related 1 (88.24 %) had the highest sequence identity with an unnamed protein product of *Echinostoma caproni.* Despite the fact that many epitopes for protein dj1 beta, superoxide dismutase, guanine nucleotide-binding protein G(s) subunit alpha, and troponin I4 were identified, none of them met the study's unique criteria (Table 15).

Table 15. B-cell epitopes predicted for the diagnostic targets and their respective hits in NCBI and UniProt databases. *Unique epitopes are highlighted in green; other epitopes are highlighted in yellow.

Protein name	Epitopes	Protein sequences of the targets with epitopes	NCBI BlastP sequence identity (%)	Uniprot Blast sequence identity (%)
Thioredoxin peroxidase	>Epitope #1* SALSTSGDTMCDRDSPHG HPHPHPHPHGHPPGPPHPP RPFMLQPNMPA	RGYHTAMLKNTLIVPIILLSALS TSGDTMCDRDSPHGHPHPHPHG HPPGPPHPPRPFMLQPNMPAPCF TGQAVVGDEFKTISLSDYKGKW VILAFYPLDFTFVCPTEIIAFSDE MDKFKQRNCEVIFCSTDSVYSH LQWTKMDRKNGGVGKLNFPLL ADKNMSISRSYGVLDEKEGNTY RGNFLIDPKGMLRQITVNDRPV GRSVDEALRLLDAFIFHEEHGE VCPANWKPKGKTIIPTPEGSKAY FSSAN	*NSS*	85.7 (uncharacterized protein *Aspergillus mulundensis*)
	>Epitope #2 YSHLQWTKMDRKNGGVG		94.1 (Thioredoxin-dependent peroxide reductase *Fasciolopsis buski*)	88.2(peroxiredoxin 1 *Paragonimus westermani*)
	>Epitope #3 VLDEKE		100 (titin isoform X1 *Petromyzon marinus*)	*NSS*
	>Epitope #4 WKPKGKTIIPTPEG		92.8(unnamed protein product *Echinostoma caproni*)	92.9 (thioredoxin domain containing protein *Echinostoma caproni*)
Protein dj-1beta	>Epitope #1 >KGSKAMAESP	FFSSTFNMSALLILSEGAEELEA VTVADVLARGNVKVTIGGLQG AHVLECSRGVKIQPSVALNDVA SQLFDVVVMPGGLKGSKAMAE SPLVKKILENHYDNKKIVAAICA APIGLQSHNIGIGKKLTSYPGFE DKLKGFTYCSDRVVVDGNLVTS RGPGTAMEFALKLLEILTNKKTA DDVAKGMLVH	100 (unnamed protein product *Echinostoma caproni*)	100 (dj-1 pfpl domain containing protein) *Echinostoma caproni*
Superoxide dismutase	>Epitope #1 >GDTTNGCISAGAHFNPT NADHGAPSDKIRHVGDL	VLRVTVSVVRFLPSVLSLFQSSA TMVLKAVCVMNGTCGVQGLVK FVQESDNAPVKVTVDITGLKPG KHGFHVHAYGDTTNGCISAGA	94.1 (unnamed protein product *Echinostoma caproni*)	94.1 (Superoxide dismutase (cu -zn) *Echinostoma caproni*)

Table contd...

	>Epitope #2 >NEDDLGLGGHEQSKITG NAGG	HFNPTNADHGAPSDKIRHVGDL GNVEADQSGKAHAEFSDNVISL SGTNSIVGRAMVVHENEDDLGL GGHEQSKITGNAGGRLACGVIG LTE	95.2 (Superoxide dismutase *Fasciola gigantica*)	95.2 (Superoxide dismutase *Fasciola gigantica*)
Cathepsin L-like proteinase	Epitope #1* >DELWHHWKRVYHKEYN Epitope #2* >GPEDGLRRSIWEQ Epitope #3 >VEHIKQHNLRHDLGLVT Epitope #4* >KYKTNVSF Epitope #5 >GYGNYG	VLRLNCETMRLLIITLLAVGAIA SNDELWHHWKRMYQKEYN GPEDGLRRSIWEQNVEHIKQHN LRHDLGLVTYTLGLNQFTDLTF EEFKDIYLSKMPSTSDFISEGTPF KPSNLNVPDSIDWRQLKCVTEV KDQGSCGSCWAFSTTGSMEGQ YMKKYKTNVSFSEQQLVDCST GYGNYGCHGGLMENAYRYLAK FGLEQENDYAYRAQEGNCLYNA TKGVAKTTDYVIVHSGNEARLK ELIGQEGPASVALDVERDFMMY RGGVYQSSSCSSRRLNHGVLAV GYGSENGRDYWIVKNSWGSRW GDGGYIRMARNQNNMCGIATM ASVPHVTHE	87.5 (cathepsin L *Fasciola gigantica*) 83.3 (unnamed protein product *Echinostoma caproni*) 100 (unnamed protein product *Echinostoma caproni*) 87.5 (Alanine--tRNA ligase *Candidatus portiere aleyrodidarum*) 100 (hypothetical protein CBC20_04795 (*Verrucomicrobia* bacterium)	87.5 (unnamed protein product *Echinostoma caproni*) 83.3 (unnamed protein product *Echinostoma caproni*) 100 (unnamed protein product *Echinostoma caproni*) *NSS* *NSS*
Cathepsin L	Epitope #1* >LEQFNYEEPENVRQLYEE Epitope #2 >LKYGKNHANEVDEYRF Epitope #3* >ITKFSDLTAEEFEQQYLST GMRFEASEPSDYVATSDQ APPESFDWRTKGAVTPVQ QQE Epitope #4 >TNWGE	STVVYFCSFLVALVSAIELQHLE QFNYEEPENVRQLYEEFKLKYG KNHANEVDEYRFSVFKDNLKK AQWYQTMEQGTAEYGITKFSDL TAEEFEQQYLSTGMRFEASEPSD YVATSDQAPPESFDWRTKGAVT PVQQQETCGSCWAFSAIGNVEG QWFRKTGELIKLSEQQLVDCDT LDYGCNGGLPMNSYKMLASM GGVEQETDYPYVAVQDTCKLDE NKFVAYVNTSLHLSQDEKEIAA YMAEHGPLSMALNANMLQFYT RGISHPFSFLCPASGLNHAVLGV GYGIEKKKPFWIVKNSWGTNW GEQGYFRIYRGSGCCGINRETTT AIIY	83.3 (unnamed protein product *Echinostoma caproni*) 93.3 (unnamed protein product *Echinostoma caproni*) 66.6(unnamed protein product *Echinostoma caproni*) 100 (mucin-16-like *Crassostrea virginica*)	83.3 (unnamed protein product *Echinostoma caproni*) 93.3 (unnamed protein product *Echinostoma caproni*) 66.7 (unnamed protein product *Echinostoma caproni*) *NSS*

Table contd...

Haemoglobinase	Epitope #1* >SETAT Epitope #2 >SPNNRYP Epitope #3 >SSWAYFCDDPQ Epitope #4* >QEEDDRSLLMKKVNNKP LKSRDSTPAVQA Epitope #5* >RKANSLEEMERARRQM HRVLELA Epitope #6* >HDVTQSVPTPVALDDVD S Epitope #7 >FSIQQVPE	RMMLVGLLLLGLLSSYVLTSET ATPGHWVVLVAGSNGWYNYRH QADLCHAYQLLLKHGIPPSNIIT MMYDDVASSPNNRYPGQLYND YNHIDVYRGVKVDYRGQDVTP LMFRRVLMGDPTLKAAGKKVL ESGPNDNVFIYFTDHGAPNLIAF PNGELHARQLQQVFRQMQTEK RYKNLVMYVEACYSGSMFENIL PSDMNIFVTTAANAQESSWAYF CDDPQIDTCLADHYSYVWMTD SEKHDLRNWTLAEQFQSVKSQL DDSHVMEFGDLKLANSTVASFQ EEDDRSLLMKKVNNKPLKSRDS TPAVQAHLVHLARKMRKANSLE EMERARRQMHRVLELAKMAKE TIDEIVHDVTQSVPTPVALDDVD SQLDCYETVFEHYKTKCFSIQQ VPELAQELHKLDQLCKRGYEA NRIVQAIFDVCG	*NSS* 100 (proteasome activator complex subunit 4B *Oreochromis niloticus*) 90 (Peptidase C13 family protein *Fasciola hepatica*) 60 (DUF1318 domain-containing protein *Verrucomicrobium spinosum*) 68.4 (unnamed protein product *Echinostoma caproni*) *NSS* 100 (gustatory receptor for sugar taste 61a isoform X1 *Drosophila yakuba*)	*NSS* *NSS* 90 (Peptidase C13 family protein *Fasciola hepatica*) 72.2 (uncharacterized protein *Lodderomyces elongisporus*) 68.4 (unnamed protein product *Echinostoma caproni*) 85.7 (uncharacterized protein *Halo belluslimi*) 100 (legumain like *Fasciola gigantica*)
Probable glutamine-tRNA ligase	Epitope #1* >FFRPFLPTLLAFAPMSPA DELQLNEPVGDASKFHKP GENYNSPGYVITKSTKRLL REHLKRTG Epitope #2 >DTNPEKEEEE	NFKCFFRPFLPTLLAFAPMSPAD ELQLNEPVGDASKFHKPGENYN SPGYVITKSTKRLLREHLKRTGG KVVTRFPPEPNGILHIGHAKAIN FNFGYAKKLGGITYLRYDDTNP EKEEEEFFKAIKEMVLWLGFTP YKVTHASDYFEQLYEWAIQLIK LNLAYVCHQRLEEIRGFNPPPSP WRDRPIEESLQLFEDMKNGKID	85.7 (unnamed protein product *Echinostoma caproni*) 100 (hypothetical protein CP533_5612 *Ophiocordyceps camponoti-saundersi*)	85.7 (Glutaminyl trna synthetase *Echinostoma caproni*) 100 (Glutaminyl trna synthetase *Clonorchis sinensis*)

Table contd...

	Epitope #3 >LEEIRGFNPPPSPWRDRPIEESLQLFEDMKNGKI	EGCATLRMKITLGDGKVDPVAYRIKMVPHHRTGSQWCIYPTYDYTHCLCDSIEDITHSLCTKEFQSRRPAYYWLCNSLDLYCPVQWEYGRLNLNYNVVSKRKILKLIEAGIVAGWDDPRLFTLTALRRRGFPPEAINNFCERIGVTMSQTVLDPSTLEACVREYLNDHAPRVMAVLDPLHVTITNWNELYPTEMSVELDVPDFPSDPESKTHKAVLAPEVFIDSTDFQEIPDKGYRRLTPSQSVGLRHAGLVLEATGVLKNDLGKVTKILAVAKRVDACPKPKAFIQWVSSPLTGEARLYDRLFTVKDPDSAKDGFLSVVNPNSLVVMKNAMFERVVQGAPVYSTFQFERIGFFAVDQDTTSDWLVFNRTVPLKADPGKMV	100 (hypothetical protein CRM22_008083 *Opisthorchis felineus*)	100 (Glutaminyl trna synthetase *Clonorchis sinensis*)
	Epitope #4 >GKVDPVAYRIKMVPHHRTGS		100 (unnamed protein product *Echinostoma caproni*)	100 (Glutaminyl trna synthetase *Echinostoma caproni*)
	Epitope #5 >YNVVSKRKILKLIEAGI		94.1 (unnamed protein product *Echinostoma caproni*)	94.1 (Glutaminyl trna synthetase *Echinostoma caproni*)
	Epitope #6 >NELYPTEMSVELDVPDFPSDPESKTHKA		92.5 (unnamed protein product *Echinostoma caproni*)	92.6 (Glutaminyl trna synthetase *Echinostoma caproni*)
	Epitope #7 >TDFQEIPDKGYRRLTP		100 (Glutaminyl-tRNA synthetase *Fasciola hepatica*)	100 (Glutaminyl trna synthetase *Fasciola hepatica*)
	Epitope #8 >KRVDACPKPK		90 (tRNA synthetases class 1 *Fasciolopsis buski*)	90 (Glutaminyl trna synthetase *Fasciola hepatica*)
	Epitope #9 >LFTVKDPDSAKDGFL		93.3 (hypothetical protein CRM22_008083 *Opisthorchis felineus*)	93.3 (Glutaminyl trna synthetase *Clonorchis sinensis*)
Probable 28S ribosomal protein	Epitope #1* >KAFPPKEEPIFNRPVPDKVIRPILYPEDVER	PYWTRILLPVLTSKSSPIYLSAPSMLGSKREKFASIFRRLEGLLQNGVIPYEQRPVWYDVYKAFPPKEEPIFNRPVPDKVIRPILYPEDVER	87.1 (unnamed protein product *Echinostoma caproni*)	87.1 (MRP S23 domain containing protein *Echinostoma caproni*)
	Epitope #2* >FHRYKRLSPINVFKLEDDRS	AKAFHRYKRLSPINVFKLEDDRSSLSRLLKQYEKLKAAHPDLDSDQLFSLAENELQKEGTILHPKE	85 (hypothetical protein FGIG_00803 *Fasciola gigantica*)	85 (MRP S23 domain containing protein *Fasciola gigantica*)

Table contd...

	Epitope #3 >KLKAAHPDLDSDQ		84.6 (hypothetical protein BGP_1892 *Beggiatoa* sp)	100 (uncharacterized protein *Strigamia maritima*)
	Epitope #4 >QKEGTIL		100 (centrosomal protein of 290 kDa isoform X1 *Ooceraea biroi*)	*NSS*
Guanine nucleotide-binding protein G(s) subunit alpha	Epitope #1 >NLSEDAKSRNEANKHIEKLIEKEKKNFK	RAAHLGIIISIMVVGCCTLNNLSEDAKSRNEANKHIEKLIEKEKKNFKSTHRLLLLGAGESGKSTIVKQMRILHIDGFSEREKKEKVEAIRKNLRDAICSIAGAMSSIKPPVELACPENKKLRDYMLEYASKPDFDYPPDFYSNCAKLWKDAGIQETFERSNEYQLIDCAKYFLDKTEEVGRNDYIPSEQDILRCRVLTSGIFETKFSVDKVNFHMFDVGGQREERRKWIQCFNDVTAIIFVAACSSYNMVLREDPSQNRVKESLELLGSIWNNRWLRNISVILFLNKQDLLTEKVLAGKSKIEVYFPHYATYQAPADTLAEYHHENPEVVRARFFFRDEFLRVTSNNNGGRHYCYPHLTCAVDTENIRRVFNDCRDIIQRMHLRQYELL	100 (GTP-binding regulatory protein Gs alpha-S chain *Fasciolopsis buski*)	100 (GTP-binding regulatory protein Gs alpha-S chain *Fasciola gigantica*)
	Epitope #2 >ILHIDGFFSEREKKEK		93.7 (unnamed protein product *Schistosoma margrebowiei*)	93.8 (Guanine nucleotide binding protein *Schistosoma mansoni*)
	Epitope #3 >PPVELACPENKKLRDYMLEYASKPDFDYPP		93.3 (unnamed protein product *Echinostoma caproni*)	93.3 (unnamed protein product *Echinostoma caproni*)
	Epitope #4 >EVGRNDYIPSEQDILR		100 (unnamed protein product *Echinostoma caproni*)	100 (unnamed protein product *Echinostoma caproni*)
	Epitope #5 >GQREERRKWIQ		100 (unnamed protein product *Spirometra erinaceieuropaei*)	100 (Guanine nucleotide binding protein *Schistosoma mansoni*)
	Epitope #6 >NMVLREDPSQNRVKESLE		100 (hypothetical protein T265_09567 *Opisthorchis viverrini*)	100 (Guanine nucleotide binding protein *Schistosoma mansoni*)
	Epitope #7 >LTEKVLAGKSKIEVYFPHYATYQAPADTLAEYHHENPEVVR		97.5 (guanine nucleotide-binding protein G(s) subunit alpha *Paragonimus westermani*)	100 (GTP binding protein Gs alpha S chain *Fasciola gigantica*)
	Epitope #8 >RVTSNNNGGR		100 (hypothetical protein T265_09567 (*Opisthorchis viverrini*)	100 (uncahracterized protein *Opisthorchis viverrini*)

Table contd...

Troponin I 4	Epitope #1 >RQLEGEKYDWEEKLRRQDVEIIE Epitope #2 >NDNKGKFVKPVLKKVSKTESHMARFEKKETGHSLSSFRNQLKSTGHSKYALEEKDETGG	QVLISCFSLSLSLSDVLILQEERKRREEEARREEVAQERQRKQQEKQRQEAEASKKQRKPGQKRRGLGGLSKEKKRMLKQLIMQKAAEIMKAERKKEQEKRDEYVRSKIGTLNLEGLSENELRSKVSQLHEQLRQLEGEKYDWEEKLRRQDVEIIEMTVKANDNKGKFVKPVLKKVSKTESHMARFEKKETGHSLSSFRNQLKSTGHSKYALEEKDETGGKADWRDQLKPKEEESAPAQ	100 (Troponin *Fasciola hepatica*) 100 (Troponin I *Fasciola gigantica*)	100 (Troponin *Opisthorchis viverrini*) 100 (Troponin I *Fasciola gigantica*)
Cathepsin B-like cysteine proteinase	Epitope #1* >LLTLLSTFIIVQATLDSDRYFEPFSDELIRYVNEESGA Epitope #2* >TRFSNVEEFKRHLGAWLETPMQRNSRRPTVRHDGFNGNLPKSFDARKQWPQCPSOGEIRDQ Epitope #3 >NGRLTADL Epitope #4* >TEGIVTGGSNESHIGCEPYPFPKCGHHGSSGTVKPCPSELYPTPECRTECQAGYNKTYDQD	VQSDMWLLTLLSTFIIVQATLDSDRYFEPFSDELIRYVNEESGASWKAAPSTRFSNVEEFKRHLGAWLETPMQRNSRRPTVRHDGFNGNLPKSFDARKQWPQCPSIGEIRDQSSCGSCWAFGATEAMTDRTCIHSNGRLTADLSARDLLSCCAYCGSGCHGGYPAMAWDHWKTEGIVTGGSMESHIGCEPYPFPKCGHHGSSGTVKPCPSELYPTPECRTECQAGYNKTYDQDKVFGRSSYNIRPREKDIMLEIMKNGPVEAAFYVYQDFTAYKGGIYHHVTGRFLGGHAIRMLGWGVEKGEKYWLLANSWNEEWGEKGFFRIRRGVDECGIESMVNAGMPRV	71 (unnamed protein product *Echinostoma caproni*) 76.6 (unnamed protein product *Echinostoma caproni*) 100 (amino acid adenylation domain containing protein *Streptomyces taklimakanensis*) 72.8 (unnamed protein product *Echinostoma caproni*)	77.1 (Pept_c1 domain containing *Echinostoma caproni*) 76.7 (Pept_c1 domain containing protein *Echinostoma caproni*) *NSS* 72.1 (vitellogenic cathepsin B-like protease *Fasciola gigantica*)

Table contd..

	Epitope #5 >NIRPREK		100 (uncharacterized protein *Galendromus occidentalis*)	*NSS*
Enhancer of rudimentary homolog	Epitope #1* >PKRPDSRIWSDYETEEQ	PHMPHTILLLQASPKRPDSRIWSDYETEEQCMESVCKLYEEHLKREHPNAPTITYDIPQLFQFIDQLADLSCLVYHEPTRSYMPHTKDWIKKRVHILLSSQASQ	75 (Endoplasmic reticulum metallopeptidase 1 *Clonorchis sinensis*)	75 (Endoplasmic reticulum metallopeptidase 1 *Clonorchis sinensis*)
	Epitope #2* >EHLKREHPNAPTITYYDIPQ		85 (Endoplasmic reticulum metallopeptidase 1 *Fasciola gigantica*)	85 (Endoplasmic reticulum metallopeptidase 1 *Fasciola gigantica*)
	Epitope #3* >EPTRSYMPHTKDWI		71.4 (unnamed protein product *Protopolystoma xenopodis*)	76.9 (peptidase m28 domain containing protein *Taenia asiatica*)
Stromal interaction molecule homolog	Epitope #1* >AYKRLLDLPEAPRLSSVTEFKRHHLSNSVNEWLLSAERGATSECPRDFYGASPSDERLDHGEEIESNCCLSNADCPHTVAEWDNQQSLKSGSGDPSAGSDVNE	TVLIGRTVAYKRLLDLPEAPRLSSVTEFKRHHLSNSVNEWLLSAERGATSECPRDFYGASPSDERLDHGEEIESNCCLSNADCPHTVAEWDNQQSLKSGSGDPSAGSDVNELALWLQLTYALEMQQYLAKKVKAESHLNNIRQSCKRLHRKRYTILGSVRLVHSDSLDDLERRLVQAKLVLEQLQNEMHERHRRWARIEALLGLKLRSNLGPVRIQQLLTTSSNSARPLSGDQHPLIPGAEYEFRSSEPRETEAPANTEEEKEHGQFLQLKRKETGRVTFRSNAVRTQLPRNYSFVRPFVTLWRKKAGSRVLTSSQTAQ	73.2 (unnamed protein product *Echinostoma caproni*)	73.3 (SOAR domain containing protein *Echinostoma caproni*)
	Epitope #2* >MQQYLAKKVKAES		75 (diguanylate cyclase *Agarivorans* sp.)	83.3 (DNA mismatch repair protein Muts *Hydrogenovibrio marinus*)
	Epitope #3 >NEMHERHRRWAR		76.9 (hypothetical protein PENSPDRAFT_693298 *Peniophora* sp.)	90.9 (SOAR domain containing protein *Echinostoma caproni*)
	Epitope #4* >GLKLRSNLGPVRIQQLLTTSSNSARPLSGDQHPLIPGAEYEFRSSEPRETEAPANTEEEKEHGQFLQLKRKETGRVTFRSNAVRTQLPRNYSF		59.7 (unnamed protein product *Echinostoma caproni*)	65(SOAR domain containing protein *Echinostoma caproni*)

Huntington interacting protein related 1	Epitope #1* >QRLRDTSPSDEQQITVQ Epitope #2* >QEFTRSVGSLTEDQLRVNKQ Epitope #3* >KELKETHNVSEFYRQHSRWTQGF	SMRGILERLMSAVEQFQRLRDTSPSDEQQITVQLEQEMRATMDAIRAAEEKFQEFRTRSVGSLTEDQLRVNKQILNYCSNLMAVVGRLVEIANTVQKELKETHNVSEFYRQHSRWTQGFLSAAK	88.2 (unnamed protein product *Echinostoma caproni*) 86.6 (Huntington interacting protein 1 *Fasciola hepatica*) 82.6 (hypothetical protein FBUS_02009 *Fasciolopsis buski*)	88.2 (unnamed protein product *Echinostoma caproni*) 86.7 (Huntington interacting protein 1 *Fasciola gigantica*) 80 (Huntington interacting protein 1 *Fasciola gigantica*)

3.4. Discussion

A total of 14 feasible diagnostic candidates for *A. sufrartyfex* infection were identified using transcriptome profiling in this investigation. The targets identified in this analysis have shown higher expression values in all the three life cycle stages of *A. sufrartyfex*. One of the identified targets, the antioxidant thioredoxin peroxidase is a well-known detoxifying protein reported in the secretomes of *Brugia malayi*, *Schistosoma mansoni, Fasciola hepatica, Opisthorchis vivverini*, and *Opisthorchis felineus* (Ghosh et al., 1998; Kwatia et al., 2000; Jefferies et al., 2001; Suttiprapa et al., 2008). This protein is found to show diagnositic potential in case of *Echinococcus granulosus*, *Fasciola gigantica* and *Taenia multiceps* infections (Margutti et al., 2008; Zhang et al., 2011; Liu et al., 2019). Furthermore, *Echinococcus granulosus* thioredoxin peroxidase has been proposed as a candidate to improve the diagnostic sensitivity of the ELISA standard diagnostic panel of antigens in the immunodiagnosis of human cystic echinococcosis (Margutti et al., 2008). This protein was again recently discovered to be a potential antigen target in the diagnosis of Asian schistosomiasis, while another thioredoxin-related protein, TRPI, has showed promise as an immunodiagnostic candidate for fasciolosis (Macalanda et al., 2017; Changklungmoa et al., 2020). One of the B-cell epitopes of thioredoxin peroxidase predicted in the study was unique to *A. sufrartyfex*, bolstering its potential for application in specific parasite detection. Another antioxidant protein discovered as a possible diagnostic candidate in the study is protein dj-1 beta. As an oxidative stress sensor, the protein is recognized to play a vital role in the defence against oxidative stress and cell death (Park et al., 2005; Hao et al., 2010). However, no researches have indicated the use of protein dj-1 beta as potential diagnostic candidate in

case of helminth parasites so far. As a result, more research is needed to better unravel its potential as a diagnostic candidate. The present study's B-cell epitope analysis, however, revealed no unique epitopes for the protein. Superoxide dismutase was the third antioxidant protein that was shortlisted as a diagnostic candidate. Superoxide dismutases are an enzymatic defence against reactive oxygen species in immunological effector pathways in parasitic helminths (Maizels et al., 1993; Brophy et al., 1994). They are already reported in *Fasciola hepatica* and *Dicrocoelium dendriticum* among the trematodes, and they are the main antioxidant enzymes in schistosomes engaged in detoxification processes (Sánchez-Moreno et al., 1987). The nematodes *Ascaris suum*, *Toxascaris leonina*, *Toxocara canis*, *T. cati*, *Ancylostoma ceylanicum*, *Onchocerca volvulus*, *Haemonchus contortus*, *Brugia malay*i, and *Necator americanus* have all been found to possess this enzyme (Sanchez-Moreno et al., 1989; Batra et al., 1990; James et al., 1994; Tang et al., 1994; Liddell et al., 1998; Taiwo et al., 1999). Antibody response to superoxide dismutase was also used to diagnose filariasis (Bal and Das, 1995). Some cestodes have been shown to have superoxide dismutase activity, including *Moniezia expansa*, *M. benedeni*, *Avitellina centripunctata*, *Taenia hydatigena,* and *Dipylidium caninum* (Sánchez-Moreno et al., 1989). The high expression of this protein in the secretome of all *A. sufrartyfex* lifecycle phases implies that it is important for the parasite's survival inside the host body, suggesting its potential as a diagnostic marker. However, as was the case with protein dj-1 beta, none of the epitopes of superoxide dismutase were found to be unique to *A. sufrartyfex*.

Proteases have a critical role in tissue invasion, immunological evasion, and anticoagulation in parasitic helminths (McKerrow et al., 2006). *Fasciola hepatica*, a liver

fluke, has the biggest cathepsin L cysteine protease family (Robinson et al., 2009). *Haemonchus contortus*, *Taenia solium*, *Taenia pisiformis*, and *Echinococcus multilocularis* all have cathepsin L proteases. They have been demonstrated to participate in the breakdown of host haemoglobin, fibrinogen, collagen, and IgG antibodies in these parasites (Rhoads and Fetterer, 1995; Baig et al., 2005; Li et al., 2006; Wang et al., 2013). *T. solium* cathepsin L is being investigated as a possible antigen for neurocysticercosis immunodiagnosis (Leon et al., 2013). *P. westermani* has cathepsin L-like cysteine proteases that are involved in the degradation of extracellular matrix proteins and IgG (Chung et al., 1997; Na et al., 2006). They also aid in the degradation of trapped antibodies on the surface of *Echinostoma caproni* (Cortés et al., 2019). Cathepsin B-like cysteine proteases are predominantly produced as digesting enzymes in the worms *Trichuris muris*, *H. contortus* and *Ascaris suum* (Hill and Sakanari, 1997; Shompole and Jasmer, 2001; Jasmer et al., 2015). Hookworm infections have already been linked to haemoglobinases as possible vaccine targets (Pearson et al., 2010). They act as a digestive enzyme in *S. mansoni* and *P. westermani*, allowing nutrients from the host haemoglobin to be absorbed (El Meanawy et al., 1990, Choi et al., 2006). Proteases play a variety of roles in helminth parasites, and they are abundantly expressed in the secretome of *A. sufrartyfex*, suggesting that they could be useful diagnostic indicators. Furthermore, in the present study, several epitopes were revealed to be unique for each of the four proteases, implying that they have species-specific diagnostic capabilities. Another intriguing finding was that one of the discovered haemoglobinase epitopes showed no substantial similarity in the NCBI and UniProt databases.

In helminths, microtubules appear to be the principal chemotherapeutic target.

Because of the evolutionary difference between mammalian hosts and primordial metazoa, they differ from those in the host (Fenell et al., 2008). Troponins, in particular, have been identified as a key target for developing lymphatic filariasis subunit and synthetic peptide vaccines (Gomase et al., 2013; Kushwaha et al., 2019). The immunoprophylactic efficacy of troponins in *Brugia malayi* infection as well as the host's immune responses has already been studied (Kushwaha et al., 2019). As a result, finding Troponin I4 in *A. sufrartyfex* secretome dataset is quite promising. Unfortunately, none of the protein's epitopes were discovered to be unique in the study. In the case of parasitic helminths, no detailed information on the glutamine-tRNA ligase (probable) and 28S ribosomal protein (probable) proteins shortlisted here is available. The former has a catalytic activity, whereas the latter is a structural component of the ribosome and translation. Given how little research has been done on these proteins, identification of distinct epitopes for these two targets is quite intriguing. In the nematode *Caenorhabditis elegans*, the guanine nucleotide-binding protein G(s) subunit alpha plays a variety of roles in development and behaviour (Bastiani and Medel, 2006). In the same nematode, the STIM1 homolog of stromal interaction molecule was found to be essential for effective phagocytosis and apoptosis regulation in a range of cell types (Kraft, 2015). As a result, more research is needed to fully comprehend their diagnostic potential.

Enhancer of rudimentary homolog is a transcriptional co-regulator that is widely expressed and highly conserved across eukaryotes (Wan et al., 2005). Huntington interacting protein related 1 homolog (hipr-1) regulates presynaptic activity (Parker et al., 2007). The identification of these two proteins in *A. sufrartyfex* secretome is promising because no other investigations have identified them as diagnostic candidates. It is also

worth noting that all of the epitopes predicted for these two targets were unique. Overall, the B-cell epitope analysis revealed that 10 of the 14 diagnostic targets have distinct epitopes, implying that they could play a role in *A. sufrartyfex* specific diagnosis. To fully understand these potential targets' diagnostic qualities, empirical studies are further required.

Chapter 4: *In silico* identification of important protein targets and their ligand molecules from the *A. sufrartyfex* transcriptomes

4.1. Background

According to the WHO model list of essential drugs, two benzimidazoles (albendazole and mebendazole), levamisole, pyrantel pamoate, and ivermectin are currently used to treat soil-transmitted helminths (Utzinger and Keiser, 2004; Keiser and Utzinger, 2010). In addition, other drug combinations are used to treat soil-transmitted helminths, the efficacy of which have been studied and documented (Beach et al., 1999; Olds et al., 1999; Mani et al., 2002; Belizario et al., 2003; Fox et al., 2005; WHO, 2006; Ndyomugyenyi et al., 2008; Keiser and Utzinger, 2010). In case of trematode infections, the broad spectrum anthelmintics praziquantel and triclabendazole are the most often used drugs. Other medications, such as albendazole, mebendazole, bithionol, and nitazoxanide, have been used in the past to treat food-borne trematodiases (Keiser and Utzinger, 2010). Again, triclabendazole-clorsulon, triclabendazole-ivermectin and nitroxinil-clorsulon-ivermectin combinations of drugs are used against *Fasciola* species (Stevenson et al., 2002; Hutchinson et al., 2009; Keiser and Utzinger, 2010). Combinations of praziquantel with artemether, artesunate, the 1,2,4-trioxolane OZ78, and tribendimidine, and artesunate-tribendimidine combination have been studied against *Clonorchis sinensis* in rat model (Keiser et al., 2009; Keiser and Utzinger, 2010). However, no combinations of drugs are currently used against food-borne trematodes in

human infections.

In view of the rising concern of drug resistance among helminth parasites, the use of *in silico* drug designing approach may be useful for identification of potential drug targets and their ligand molecules. The discovery and/or development of small compounds for inhibition or activation of specific biomolecules is a common approach to drug development. There are basically two important types of drug design. Firstly, ligand-based drug design which is based on the knowledge of biological active substances that interact with and bind to the target of interest. Secondly, structure-based drug design also known as direct drug design is based on the three-dimensional structure of the biological target, which can be determined using techniques like x-ray crystallography or NMR spectroscopy (Kroemer, 2007). Based on sequence homology with known structures, homology modelling predicts protein structures.

Various approaches for screening huge databases of compounds to choose small molecules with high affinity for a desired target have been developed. These databases might be either general or specialised in nature. Over the last two decades, the number of such databases has multiplied many times. Some of these datasets are owned by individuals and may require licensing. Free databases such as ZINC (Irwin and Shoichet, 2005), Drugbank (https://go.drugbank.com/), PubChem (pubchem.ncbi.nlm.nih.gov), as well as a slew of others, are frequently used in screening research.

In hit identification and lead optimization, computational techniques that "dock" small compounds into the structures of macromolecular targets and "score" their potential complementarity to binding sites are commonly utilized. In the drug development process, the necessity for a quick search for small compounds that may bind to

biologically interesting targets is critical. The docking method entails predicting the conformation and orientation (or posing) of ligands within a certain binding site. In general, docking investigations have two goals: accurate structural modelling and precise activity prediction (Kitchen et al., 2004).

Despite the fact that several studies for drug development and repurposing have been conducted against *Schistosoma* spp. and some soil-transmitted helminth infections, there is insufficient development and deployment of the next generation of anthelminthic agents against other groups of helminths, such as food-borne trematodes. Increased and concerted efforts are needed to identify new therapeutic targets and antihelminthic medicines, particularly for food-borne intestinal flukes like *A. sufrartyfex*. Therefore, in light of increasing reports of drug resistance and the fact that very few effective drugs are available for the treatment of echinostomiasis, the goal of this current study was to find drug targets against *A. sufrartyfex* and their suitable ligand molecules using *in silico* drug designing techniques.

4.2. Materials and methods

4.2.1. 3D modelling of the potential protein targets

All of the 30,545 conceptually translated protein sequences generated from the multi-stage transcriptome of *A. sufrartyfex* were subjected to BlastP against the human database in order to filter out human homologs (cut-off value of less than 30% sequence identity). The filtered sequences were submitted to both RaptorX (Källberg et al., 2012) and Bhageerath H+ (Jayaram et al., 2014) for 3D structure modelling using default parameters.

4.2.2. Quality assesment of the 3D models

RaptorX provides five 3D candidate structures while Bhageerath H+ also provides five 3D candidate structures for a given query protein sequence. The 3D structures acquired from both RaptorX and Bhageerath H+ for the filtered sequences were then submitted to ProtSAV (Protein Tertiary Strucutre Analysis and Validation) metaserver (Singh et al., 2016) for quality assessment. ProSA, dDFIRE, D2N, ProQ, PSN-QA, WhatCheck, QMEAN, ANNOLEA, Naccess, Verify3D, Errat, Procheck and MolProbity are among the quality assessment online servers and standalone programmes available on this metaserver site. It calculates a global quality score based on the given protein model in pdb format and displays the result graphically. Based on their rmsd values (root mean square deviations) from the relevant native structures of the target proteins, graphical representation encompasses three separate classes of colour illustration. The modelled structures are classified into the first class region in 0-2 Å rmsd which is green coloured, the second class region falling between 2-5 Å rmsd is yellow coloured, the third class region in 5-8 Å rmsd is orange coloured, and the fourth class region beyond 8 Å rmsd appears in red. A blue dot in the graphical output symbolizes the ProtSAV score of a modelled structure. Those structures that fall in the orange region are required to be remodelled. Structures with an rmsd value greater than 8 Å rmsd are rejected (Singh et al., 2016) (Figure 11). 3D candidate structures provided by both RaptorX and Bhageerath H+ were then submitted to ProtSAV and only the models falling in the green and yellow regions were considered. Of these, only the best consensus protein model by either of these two tools (RaptorX and Bhageerath H+) was taken forward.

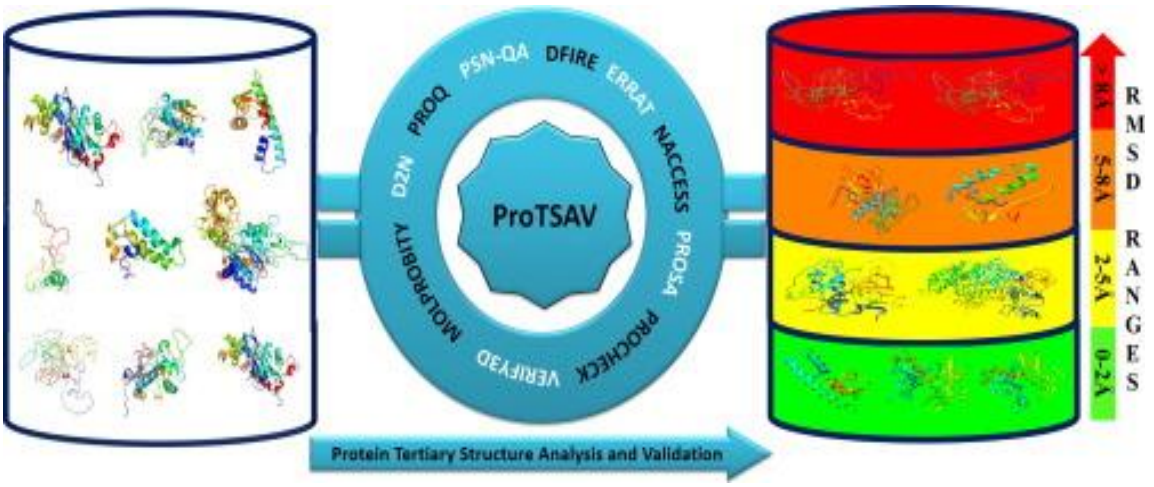


Figure 11. Different colour regions with their rmsd values in ProtSAV metaserver.

4.2.3. Active site prediction

The tool AADS (Automated Active site detection Docking and Scoring) was then used to detect active sites for the quality assessed protein structures. The tool identifies all the active cavities for ligand binding and shorlists the top cavity which can be used in further docking process (Singh et al., 2011). Further, PyMOL molecular graphics System version 2.0 Schrödinger, LLC was used to visualize the amino acid residues within 3.5 Å of the drug molecule obtained from the protein ligand docking complex.

4.2.4. Search for FDA approved molecules

FDA approved compounds with no known human targets were retrieved by uploading the sequences of our chosen 3D protein structures to DrugBank's target sequences section (go.drugbank.com). A total of 9 FDA approved molecules (Drug IDs DB11753, DB00365, DB05154, DB00224, DB11799, DB13997, DB00220, DB08930 and DB00625) against the target proteins with no known interactions with human proteins were identified, and their 3D structures were retrieved. DrugBank is a public

database with a wide range of pharmacological compounds and their targets. DrugBank's most recent release, in 2021, has 14,556 medication entries (https://go.drugbank.com/).

4.2.5. Molecular docking with FDA approved compounds

AutoDock version 4.0 (Morris et al., 2009) was used to perform molecular docking of the FDA approved drug molecules. The cuf off binding energy for the FDA approved drug molecules was kept at ≤-9.5 kcal/mol. AutoDock is a suite of docking tools that anticipates how small compounds, such as substrates or medication candidates, would bind to a 3D-structured receptor. Lower binding energy of G (kcal/mol) indicates that the target protein and ligand molecule are more stable and have better binding (Basu et al., 2020). PyMOL molecular graphics System version 2.0 Schrödinger, LLC was used to visualize the molecular interaction of the docking output. Finally, the protein ligand complex for the top five targets were further analysed to unravel the amino acids involved with the help of Discovery Studio version 2021 (https://discover.3ds.com/ discovery-studio-visualize-download) software.

4.2.6. Molecular docking with ZINC ligands

Those protein structures which remained above the cuf off binding energy of ≤ -9.5 kcal/mol against FDA approved molecules were subjected to RASPD tool (Mukherjee and Jayaram, 2013) in order to identify candidate ligand molecules in the ZINC database. RASPD uses a QSAR-style equation to quickly compute the binding energy of ligands, E (kcal/mol). Each of the target protein models was uploaded as a pdb file, and the results were obtained from the tool. A cut-off binding affinity of ≤ -8.0 kcal/mol was kept for the

ZINC molecules in the current study. The ZINC database is a carefully curated collection of commercially available chemical substances that have been optimized for virtual screening. The database has a library of 7,27,842 molecules, each having a three-dimensional structure, compiled from vendor catalogues (the size of this library is evergrowing). Each molecule in the library has vendor and purchase information and is ready to dock with a variety of docking applications. This database can be downloaded for free (http://zinc.docking.org) in a variety of file formats, including SMILES, mol2, 3D SDF, and DOCK flexibase. Finally, the protein ligand complex for the top five targets were further analysed to unravel the amino acids involved with the help of Discovery Studio version 2021 software.

4.3. Results

4.3.1. Molecular docking with FDA approved compounds

From the 30,545 putative protein sequences, a total of 203 protein sequences non-homologous to humans were successfully modelled. All these 203 structures also passed the quality assessment in ProtSAV. Homology models of these modelled proteins were visualized in the PYMOL software. Of the 203 modelled structures docked with 9 FDA approved drug molecules, a total of 102 docked structures remained below the cut-off binding affinity (≤-9.5 kcal/mol) whereas 101 docked structures were above this threshold.

4.3.2. Molecular docking with ZINC molecules

Consequently, the 101 structures which were above the cut-off value against FDA

drugs were evaluated with the RASPD tool to identify ZINC molecules for these structures. A total of 7 ZINC molecules (ZINC IDs ZINC02895117, ZINC02820058, ZINC12576410, ZINC12376046, ZINC02060288, ZINC08892130 and ZINC03143011) could be identified which were taken forward for molecular docking using the AutoDock tool. Consequently, a total of 43 structures now remained below the cut-off binding affinity of ≤ -8.0 kcal/mol.

Finally, 145 structures showed binding affinity below the significant cut-off level (102 with FDA approved drugs and 43 with ZINC molecules) (Tables 16 and 17). Results of the quality assessment of 3D models of proteins by the ProtSAV meta server showed that all the 145 targets had a root mean square deviation (rmsd) in the green and yellow regions (Figure 12). A total of 55 structures were found in the green region whereas 90 structures were found to be in the yellow region (Appendix Table).

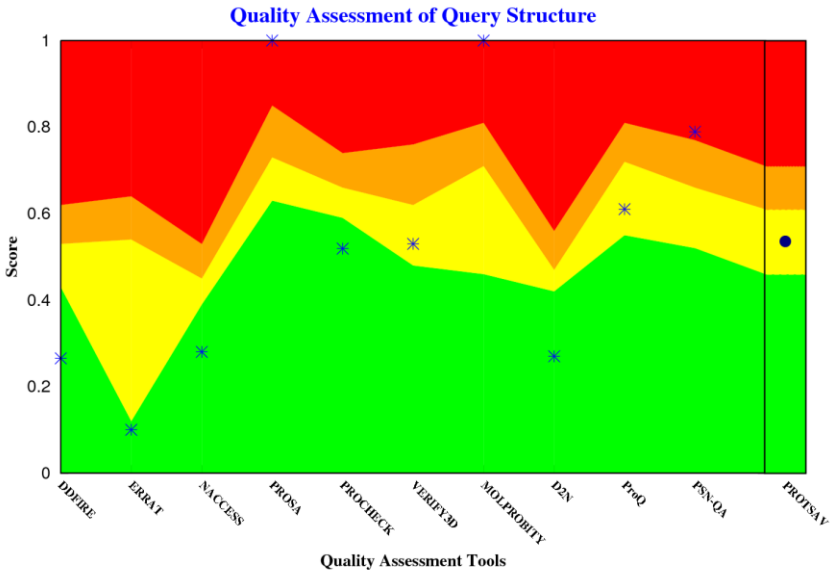


Figure 12. Representative output result from ProtSAV metaserver for one of the drug target identified in the current study. A blue dot appears in the yellow region.

4.3.3. Top drug targets with FDA approved compounds and amino acid interactions

Of the 102 proteins showing significant binding affinity with FDA approved drugs, five proteins that showed the best binding energy was hence shortlisted as the probable drug targets against *A. sufrartyfex* in the present study (Table 18). The proteins were also shortlisted based on their available functional characterization in the GO database. Among the shortlisted drug targets, sodium-dependent transporter (uncharacterized) demonstrated the best binding affinity value of -11.9 kcal/mol with the FDA approved drug Rifamycin (DB11753) (Figure 13A). Conventional hydrogen bond and Pi-Pi stacked interactions of the drug molecule was seen with amino acid residues TYR235, ARG34, GLN232 in the active site of the complex (Figures 13B & C). The next protein target identified was anaerobic glycerol-3-phosphate dehydrogenase subunit A which exhibited binding affinity of -11.3 kcal/mol with the drug Indinavir (DB00224) (Figure 13D). Conventional hydrogen bond, Pi-Pi T shaped, carbon hydrogen bond, Alkyl, Pi-Anion, Pi-Alkyl and Pi-Pi stacked interactions of the drug molecule was seen with amino acid residues ARG59, TYR290, GLU514, ARG470, ARG474, ASP159, PRO160, ARG515 ASP55 in the active site of the complex (Figures 13E & F). Tyrosine-tRNA synthetase was the third best target identified exhibiting binding affinity of -11.2 kcal/mol with the drug Grepafloxacin (DB00365) (Figure 13G). Conventional hydrogen bond, carbon hydrogen bond, unfavourable Acceptor-Acceptor, Alkyl and Pi-Anion interactions of the drug molecule was seen with amino acid residues VAL94, GLY227, GLY228, GLN213, GLY79, TYR209, ASP81, ASP121 ASP55 in the active site (Figures 13H & I). Fourth protein target identified was argininosuccinate synthase showing binding affinity of -11.0 kcal/mol with the drug Rifamycin (DB11753) (Figure 13J).

Conventional hydrogen bond, carbon hydrogen bond, unfavourable Donor-Donor, Pi-Cation and Pi-Anion interactions of the drug molecule was seen with amino acid residues ARG178, GLU197, SER195, GLY30, GLU59, LEU31, ASP32, LYS214 in the active site (Figures 13K & L). Finally, the fifth target identified was chloride intracellular channel exhibiting binding affinity of -10.6 kcal/mol again with the drug Rifamycin (DB11753) (Figure 13M). Conventional hydrogen bond and Pi-Sigma interactions of the drug molecule was seen with amino acid residues SER59, HIS19, ALA124, ILE57, ASP55 in the active site (Figures 13N & O).

.

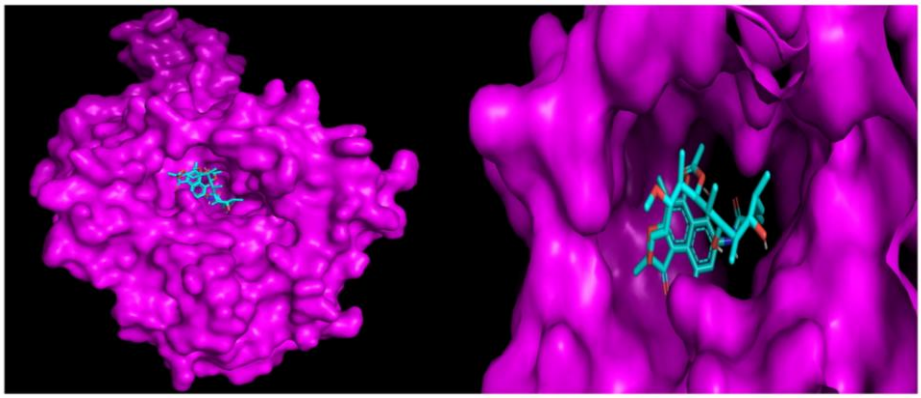


Figure 13A. Interaction between the predicted target sodium-dependent transporter (uncharacterized) with FDA approved drug molecule Rifamycin (DB11753) exhibiting binding affinity of -11.9 kcal/mol.

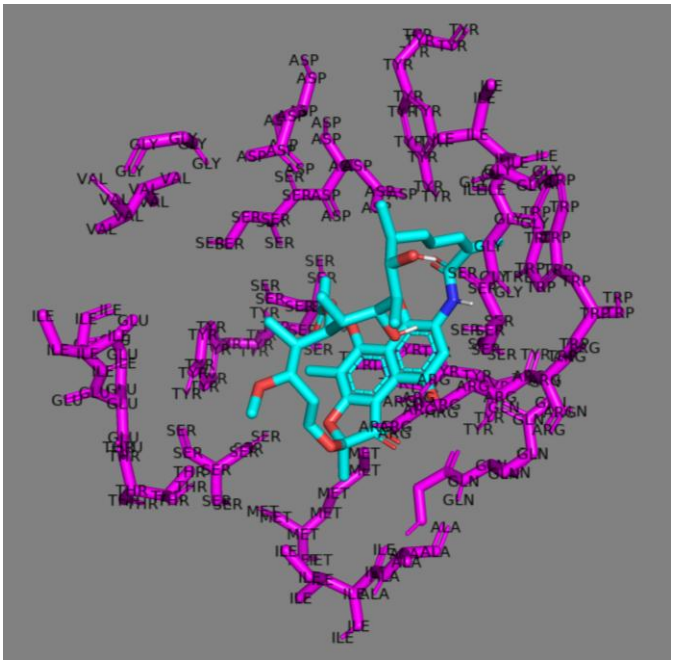


Figure 13B. Amino acid residues in the active site of sodium-dependent transporter (uncharacterized) protein within 3.5 Å of the ligand molecule Rifamycin (DB11753).

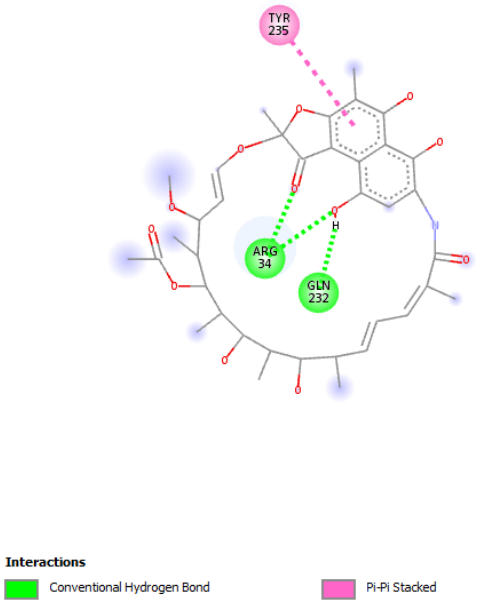


Figure 13C. Sodium-dependent transporter (uncharacterized) protein with FDA approved drug molecule Rifamycin (DB11753) exhibiting amino acid interactions.

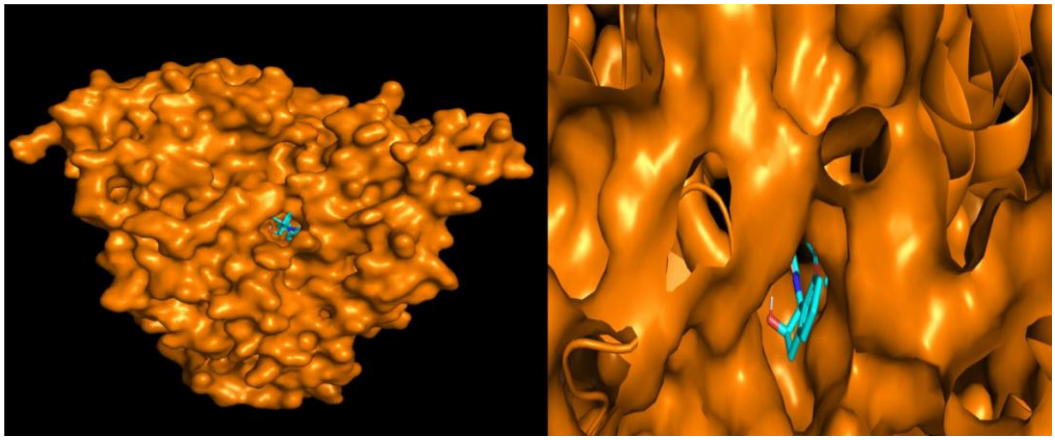


Figure 13D. Interaction between the predicted target anaerobic glycerol-3 phosphate dehydrogenase subunit A with FDA approved drug molecule Indinavir (DB00224) exhibiting binding affinity of -11.3 kcal/mol.

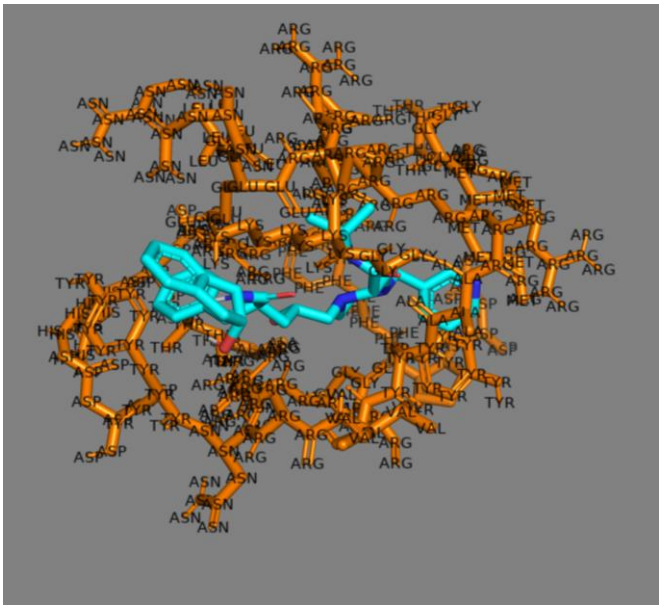


Figure 13E. Amino acid residues in the active site of anaerobic glycerol-3 phosphate dehydrogenase subunit A protein within 3.5 Å of the ligand molecule Indinavir (DB00224).

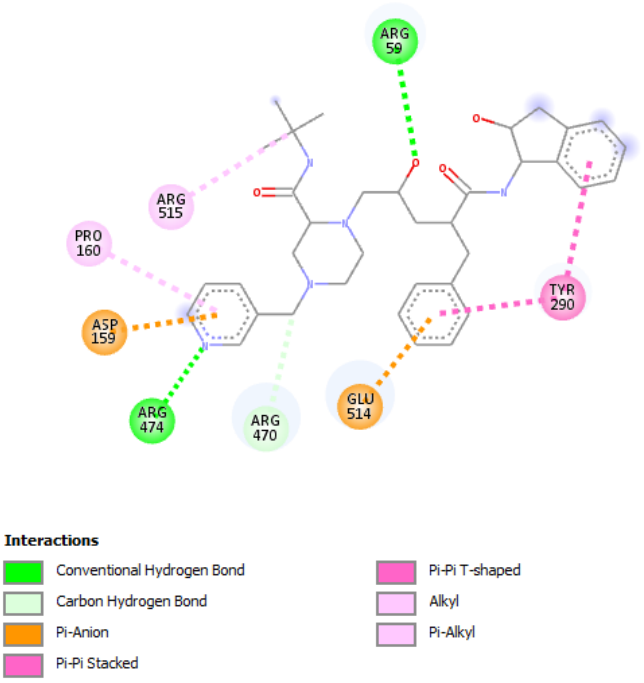


Figure 13F. Anaerobic glycerol-3 phosphate dehydrogenase subunit A with FDA approved drug molecule Indinavir (DB00224) exhibiting amino acid interactions.

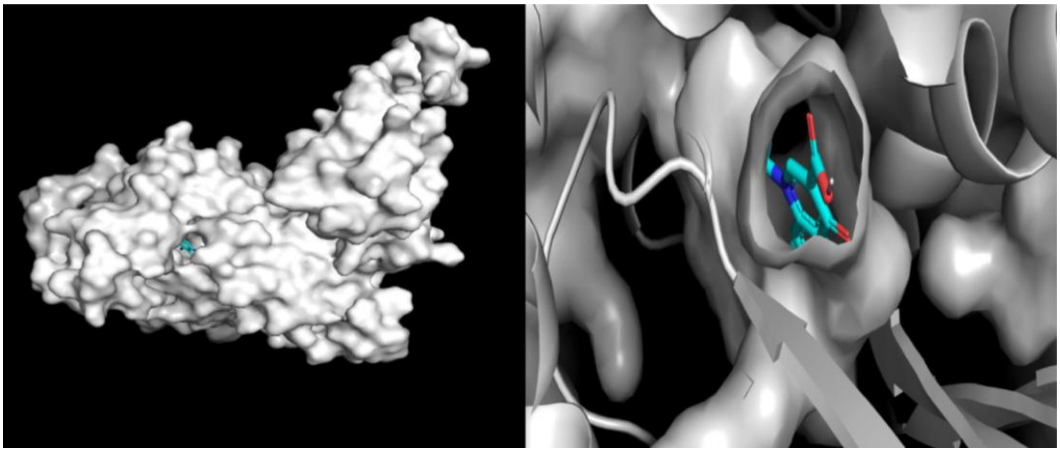


Figure 13G. Interaction between the predicted target tyrosine-tRNA synthetase with FDA approved drug molecule Grepafloxacin (DB00365) exhibiting binding affinity of -11.2 kcal/mol.

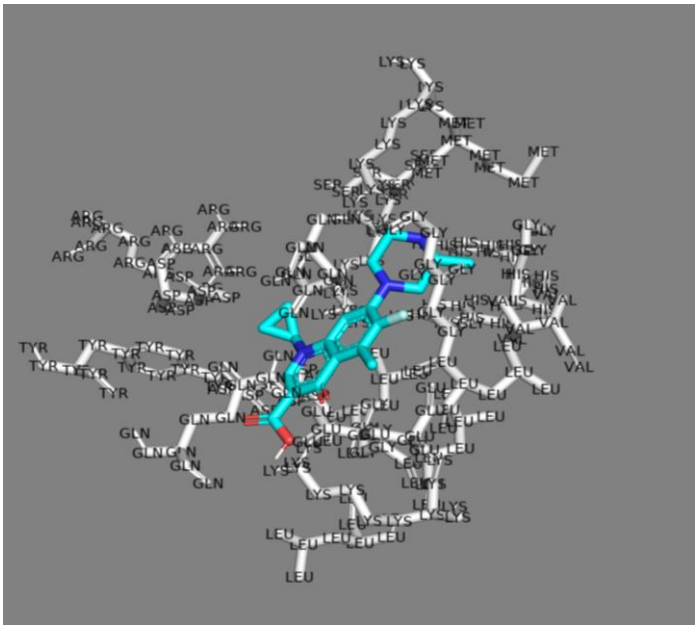


Figure 13H. Amino acid residues in the active site of Tyrosine-tRNA synthetase protein within 3.5 Å of the ligand molecule Grepafloxacin (DB00365).

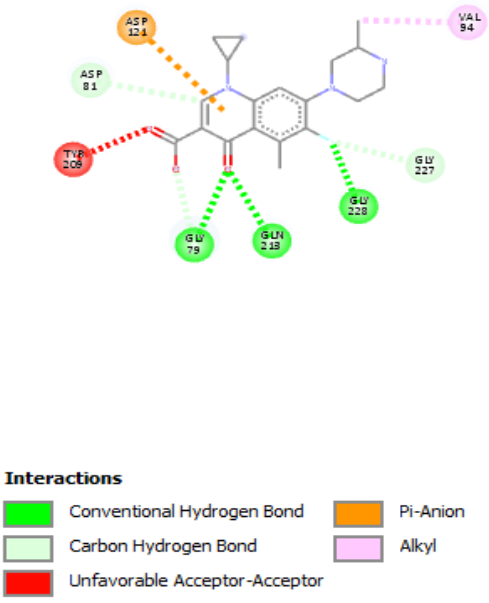


Figure 13I. Tyrosine-tRNA synthetase with FDA approved drug molecule Grepafloxacin (DB00365) exhibiting amino acid interactions.

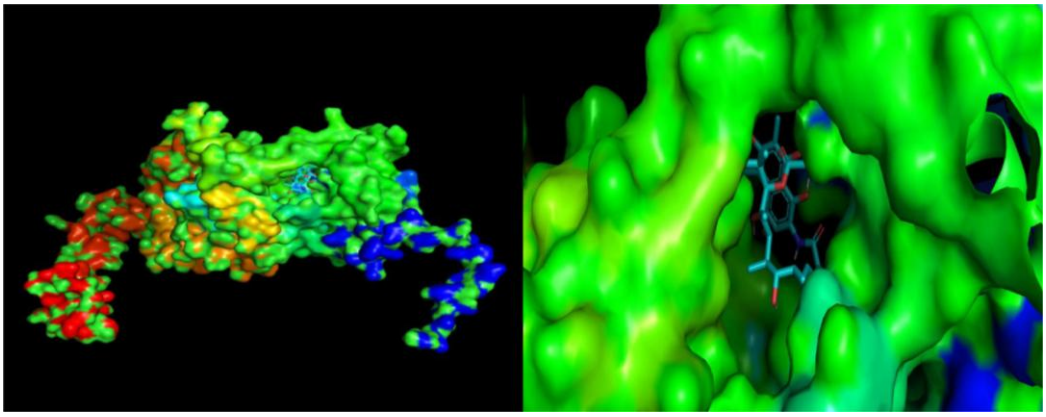


Figure 13J. Interaction between the predicted target Argininosuccinate synthase with FDA approved drug molecule Rifamycin (DB11753) exhibiting binding affinity of -11.0 kcal/mol.

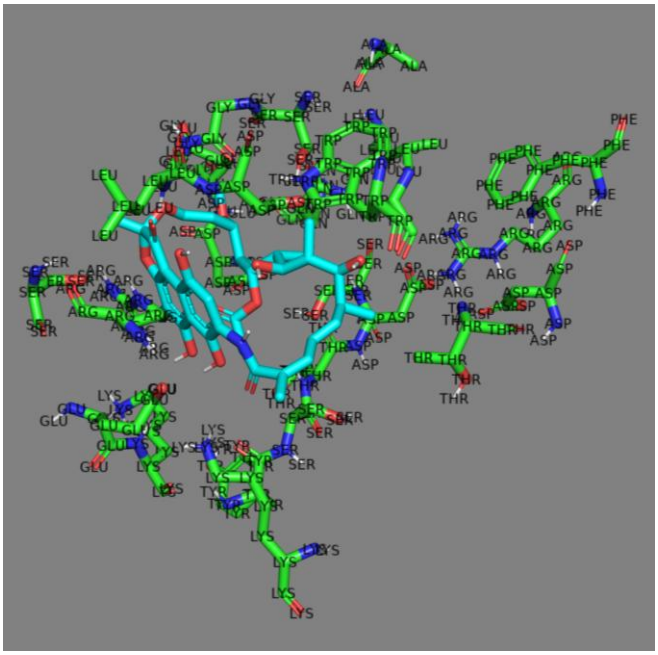


Figure 13K. Amino acid residues in the active site of Argininosuccinate synthase protein within 3.5 Å of the ligand molecule Rifamycin (DB11753).

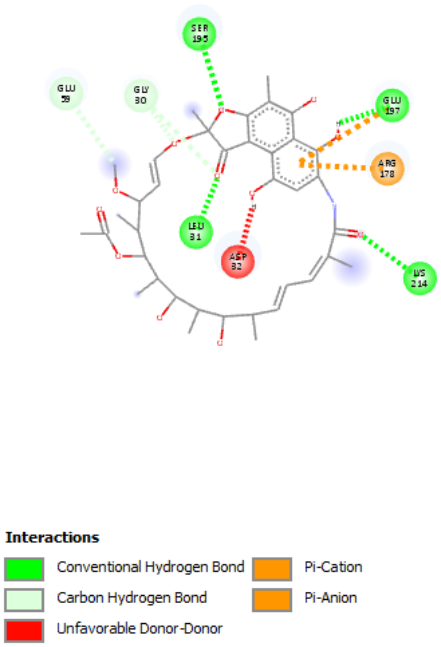


Figure 13L. Argininosuccinate synthase with FDA approved drug molecule Rifamycin (DB11753) exhibiting amino acid interactions.

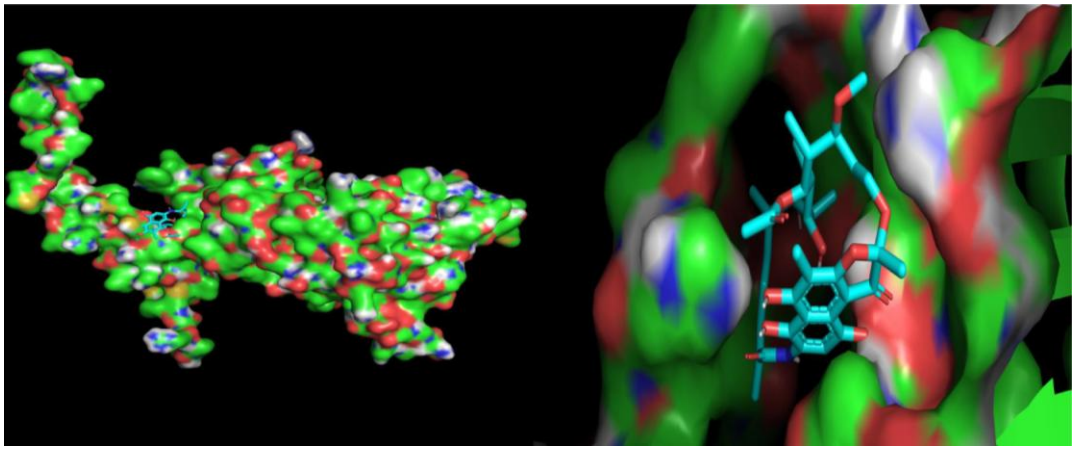


Figure 13M. Interaction between the predicted target Chloride intracellular channel with FDA approved drug molecule Rifamycin (DB11753) exhibiting binding affinity of -10.6 kcal/mol.

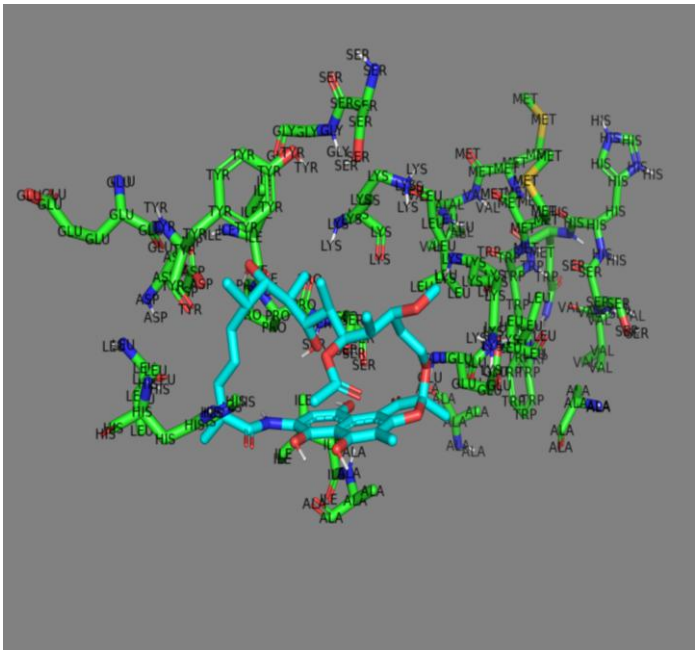


Figure 13N. Amino acid residues in the active site of Chloride intracellular channel protein within 3.5 Å of the ligand molecule Rifamycin (DB11753).

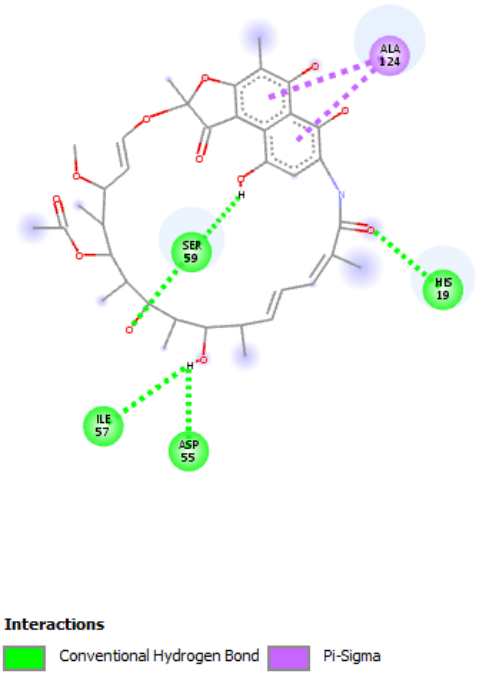


Figure 13O. Chloride intracellular channel with FDA approved drug molecule Rifamycin (DB11753) exhibiting amino acid interactions.

4.3.4. Top drug targets with ZINC ligands and amino acid interactions

Similarly, five top proteins showing best binding affinity with ZINC molecules was hence shortlisted as the probable drug targets against *A. sufrartyfex* (Table 18). Among the shortlisted drug targets, bifunctional protein Aas showed the best binding affinity value of -11.4 kcal/mol with the ZINC molecule ZINC02820058 (Figure 14A) Conventional hydrogen bond, Pi-Pi T shaped, Pi- Anion, Pi- Cation Pi- Alkyl and Pi-Sigma interactions was with amino acid residues VAL255, PHE251, HIS249, ALA320, THR346, VAL345, ARG444, ASP429, VAL365, LYS322, ILE441 in the active site (Figures 14B & C). Second putative target predicted was 50S ribosomal protein L3 glutamine methyltransferase that showed binding affinity of -11.2 kcal/mol with the ZINC molecule ZINC02820058 (Figure 14D). The interactions involved with this complex were conventional hydrogen bond, Pi-Pi T shaped, Pi- Anion, Pi- Cation, Pi-Alkyl and Pi-Donor hydrogen bond interactions with amino acid residues ARG130, TYR63, SER131, PRO377, ASP59, ARG37 in the active site of the complex (Figures 14E & F). Third target identified was putative phosphonates utilization ATP-binding protein PhnK with the ZINC molecule ZINC02820058 showing binding affinity of -11.2 kcal/mol (Figure 14G). Conventional hydrogen bond, carbon hydrogen bond, Pi-Sigma, Pi- Anion, Pi- Cation Pi- Alkyl, Pi-Sigma and Pi-Pi stacked interactions of the ligand molecule was seen with amino acid residues ILE73, TRP107, SER74, LEU103, VAL109, ARG131, ARG104, HIS111, ARG180, ALA134, GLU130 in the active site of the complex (Figures 14H & I). Similarly, cobalt import ATP-binding protein was the next best target exhibiting binding affinity of -9.7 kcal/mol with the ZINC molecule ZINC12576410 (Figure 14J). Conventional hydrogen bond, Pi-Anion, Pi-Cation, Pi-

Alkyl, Pi-Pi Stacked and Pi-Pi T-shaped interactions of the drug molecule was seen with amino acid residues PHE51, ASP168, PHE90, ARG83, ALA87, LEU160, SER109, GLN96 (Figures 14K & L) in the active site of the complex. Finally, the fifth target identified was DNA mismatch repair protein MutL exhibiting binding affinity of -8.5 kcal/mol with the ZINC molecule ZINC03143011 (Figure 14M). Conventional hydrogen bond, Pi-Anion, Pi-Cation, Pi- Alkyl and Pi-Pi Stacked interactions of the drug molecule was seen with amino acid residues ILE20, GLU40, ARG41, TYR44 in the active site of the complex (Figures 14N & O).

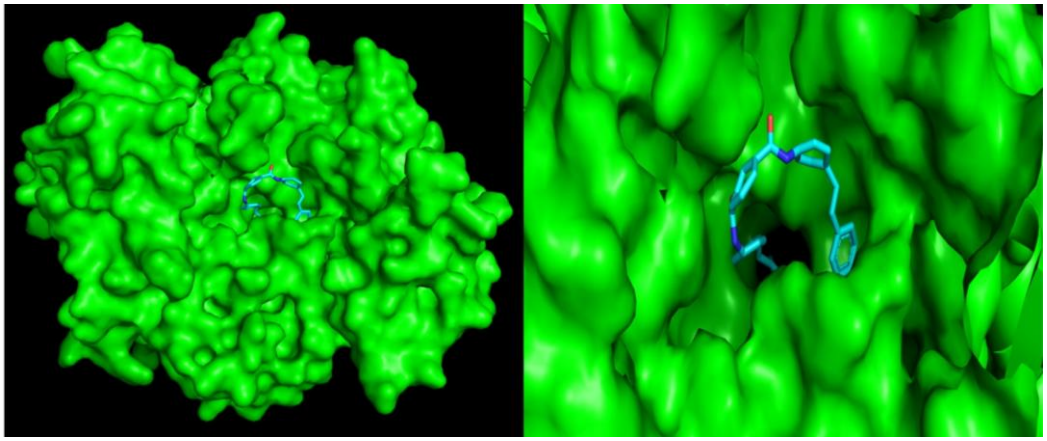


Figure 14A. Interaction between the predicted target bifunctional protein Aas with ZINC molecule ZINC02820058 exhibiting binding affinity of -11.4 kcal/mol.

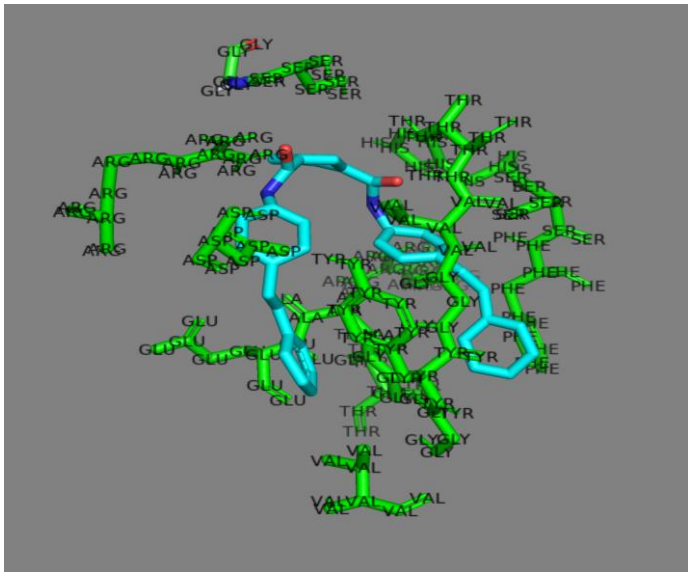


Figure 14B. Amino acid residues in the active site of bifunctional protein Aas protein within 3.5 Å of the ligand molecule ZINC02820058.

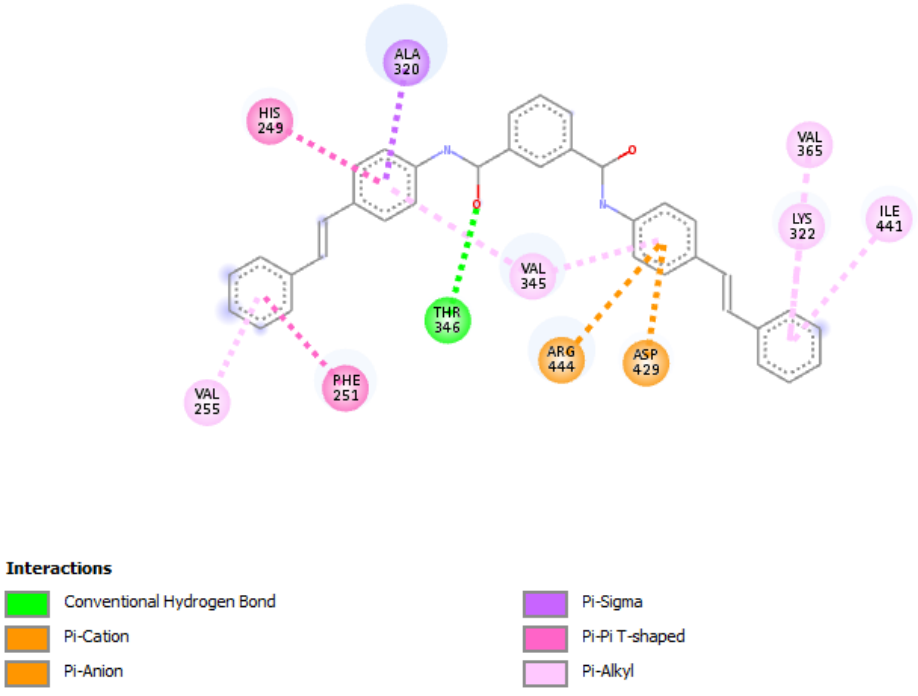


Figure 14C. The predicted target bifunctional protein Aas with ZINC molecule ZINC02820058 exhibiting exhibiting amino acid interactions.

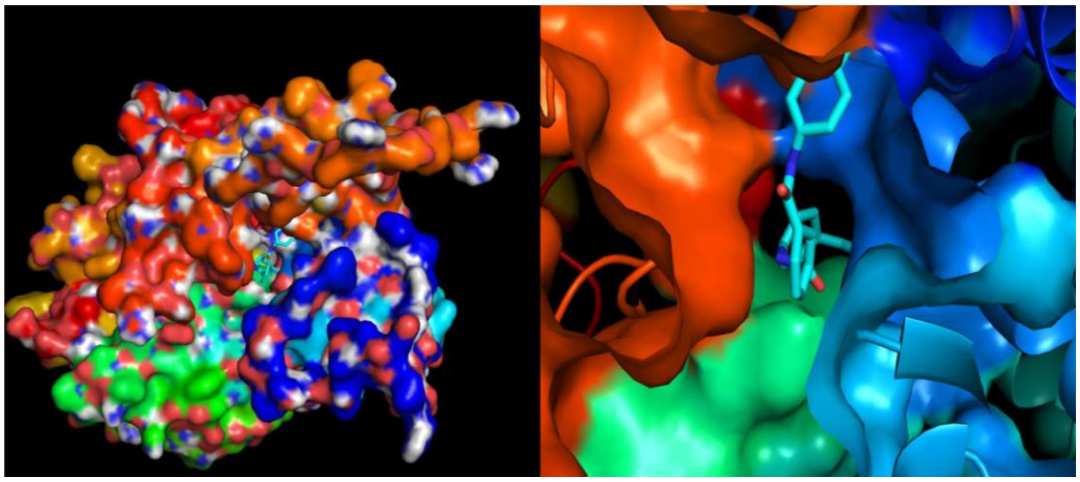


Figure 14D. Interaction between the predicted target 50S ribosomal protein L3 glutamine methyltransferase with ZINC molecule ZINC02820058 exhibiting binding affinity of -11.2 kcal/mol.

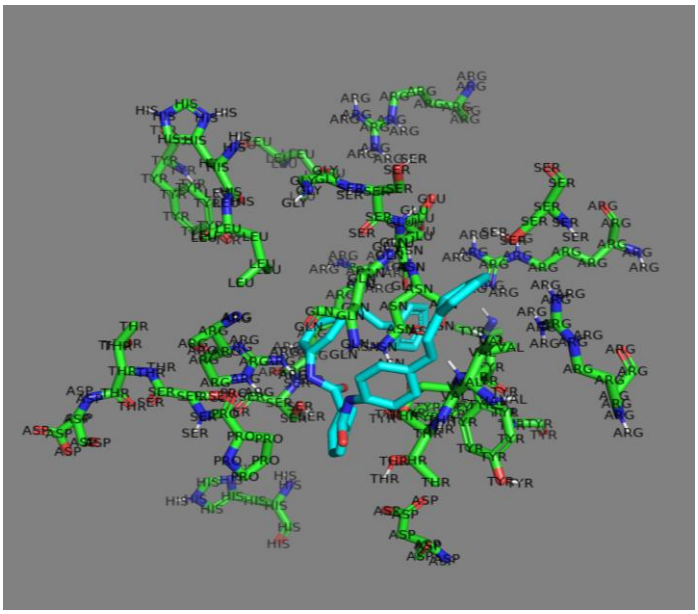


Figure 14E. Amino acid residues in the active site of 50S ribosomal protein L3 glutamine methyltransferase protein within 3.5 Å of the ligand molecule ZINC02820058.

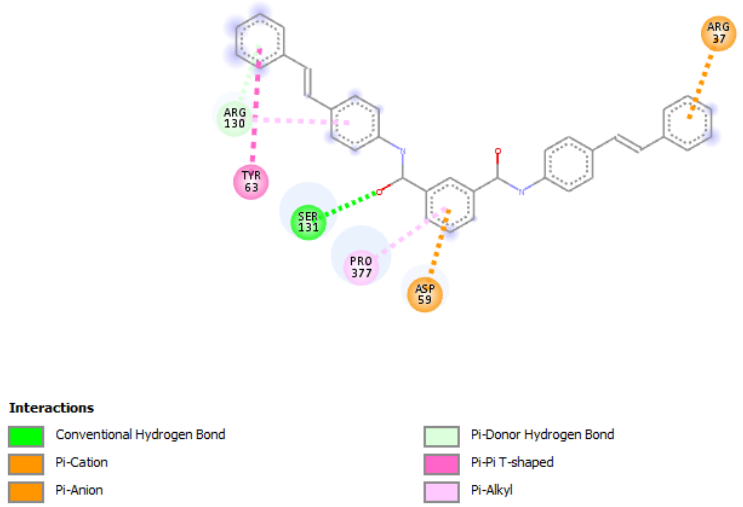


Figure 14F. 50S ribosomal protein L3 glutamine methyltransferase with ZINC molecule ZINC02820058 exhibiting amino acid interactions.

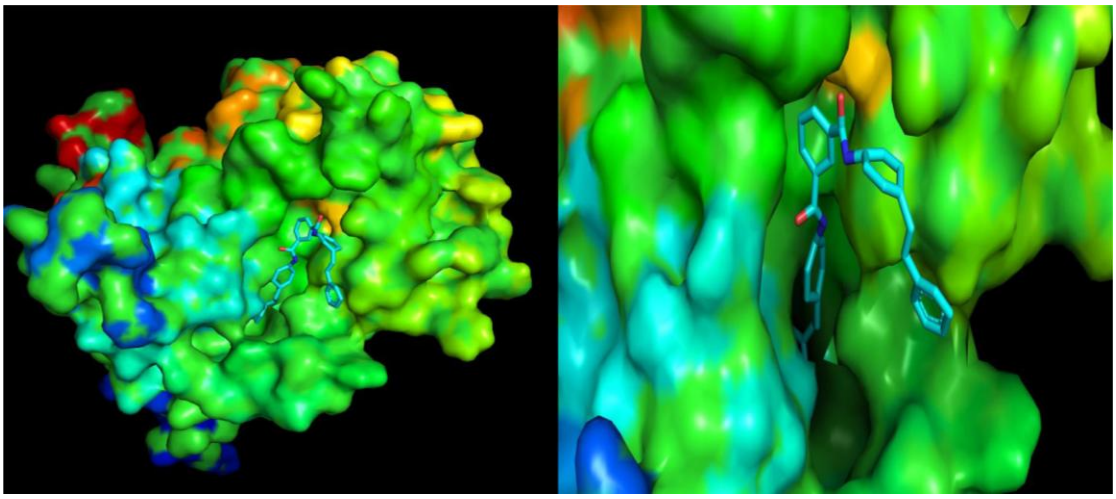


Figure 14G. Interaction between the predicted target Putative phosphonates utilization ATP-binding protein PhnK with ZINC molecule ZINC02820058 exhibiting binding affinity of -11.2 kcal/mol.

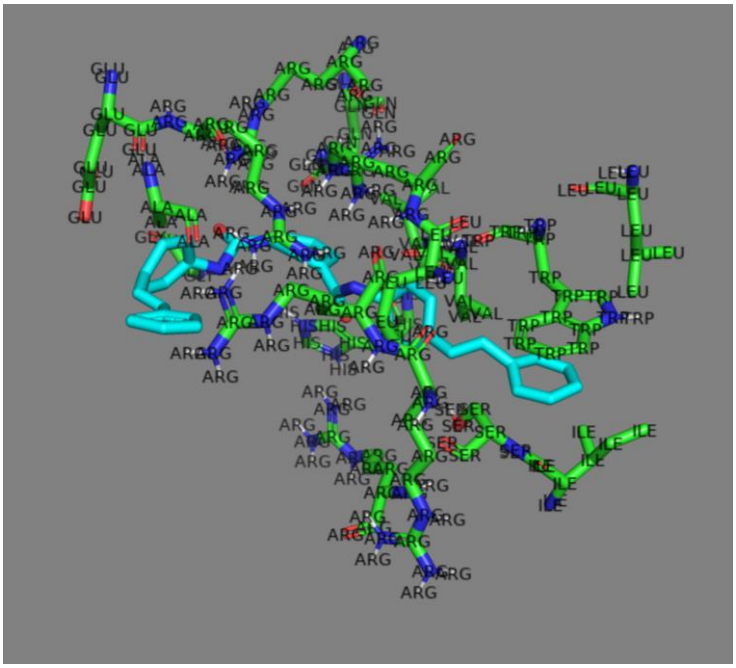


Figure 14H. Amino acid residues in the active site of Putative phosphonates utilization ATP-binding protein PhnK protein within 3.5 Å of the ligand molecule ZINC02820058.

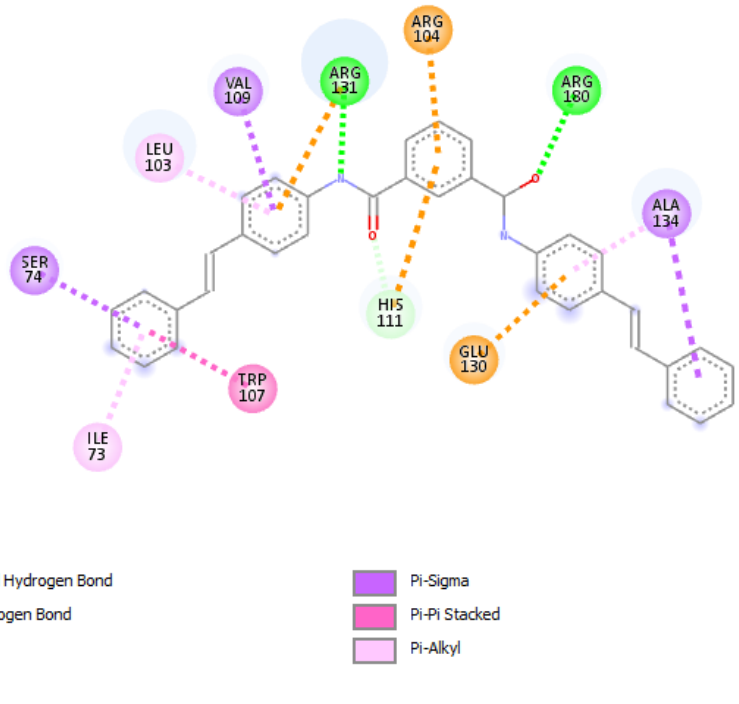


Figure 14I. Putative phosphonates utilization ATP-binding protein PhnK with ZINC molecule ZINC02820058 exhibiting amino acid interactions.

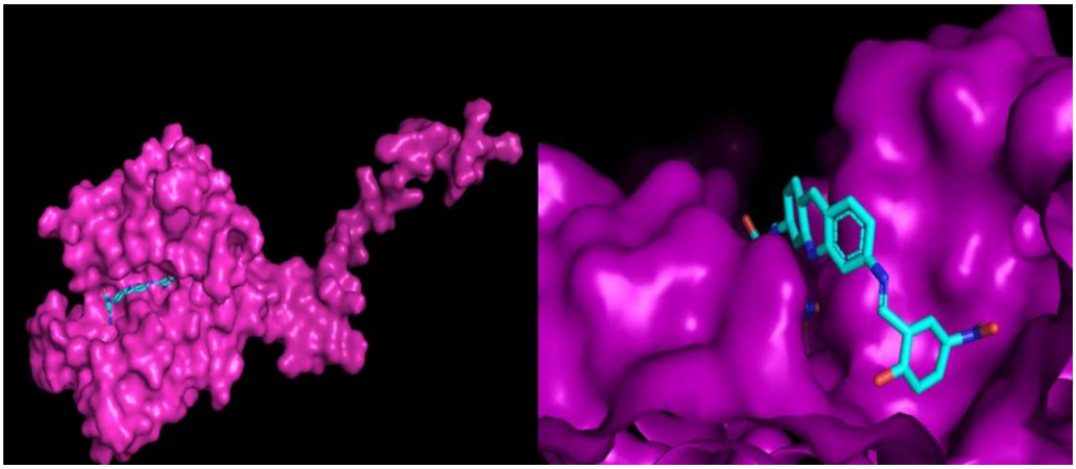


Figure 14J. Interaction between the predicted target Cobalt import ATP-binding protein CbiO with ZINC molecule ZINC12576410 exhibiting binding affinity of -9.7 kcal/mol.

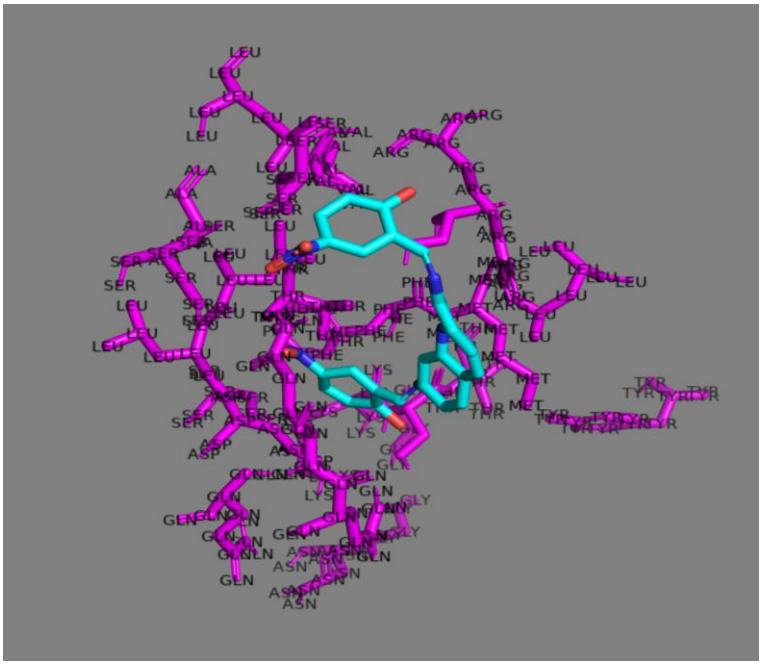


Figure 14K. Amino acid residues in the active site of target Cobalt import ATP-binding protein CbiO within 3.5 Å of the ligand molecule ZINC12576410.

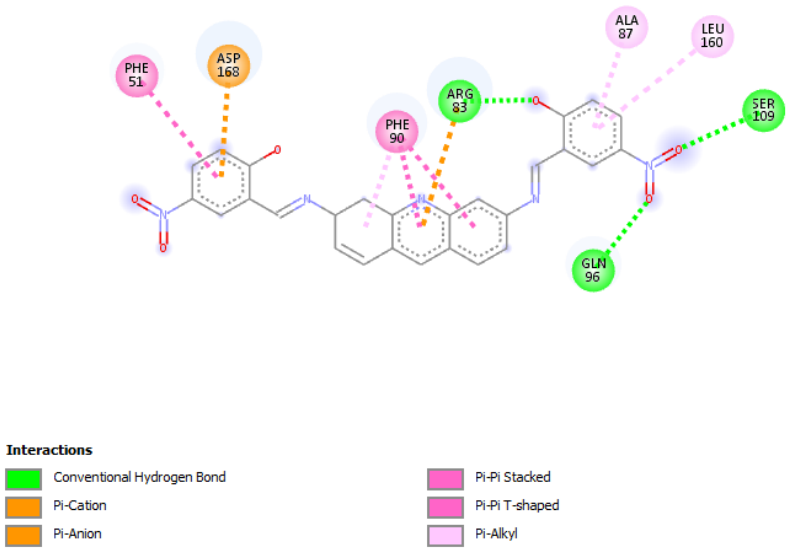


Figure 14L. Cobalt import ATP-binding protein CbiO with ZINC molecule ZINC12576410 exhibiting amino acid interactions.

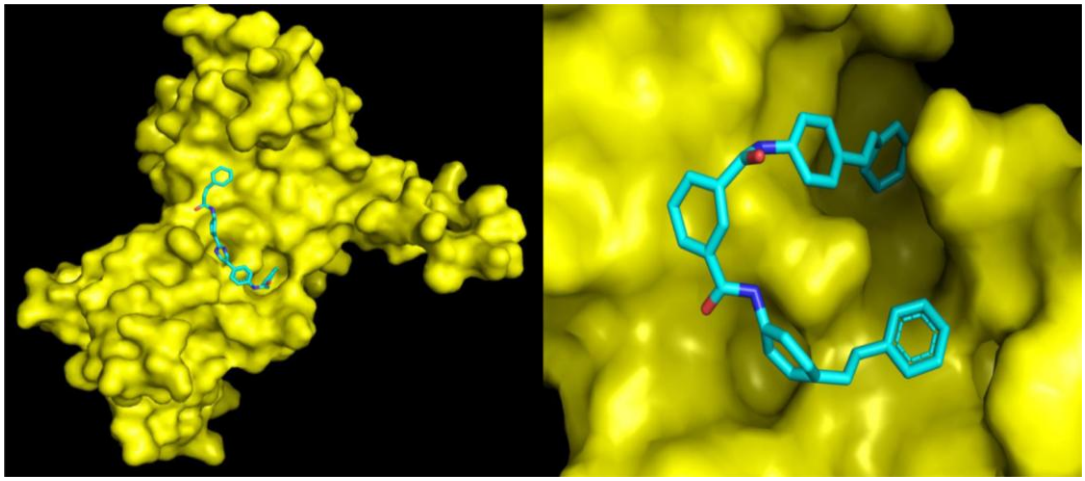


Figure 14M. Interaction between the predicted target DNA mismatch repair protein MutL with ZINC molecule ZINC03143011 exhibiting binding affinity of -8.5 kcal/mol.

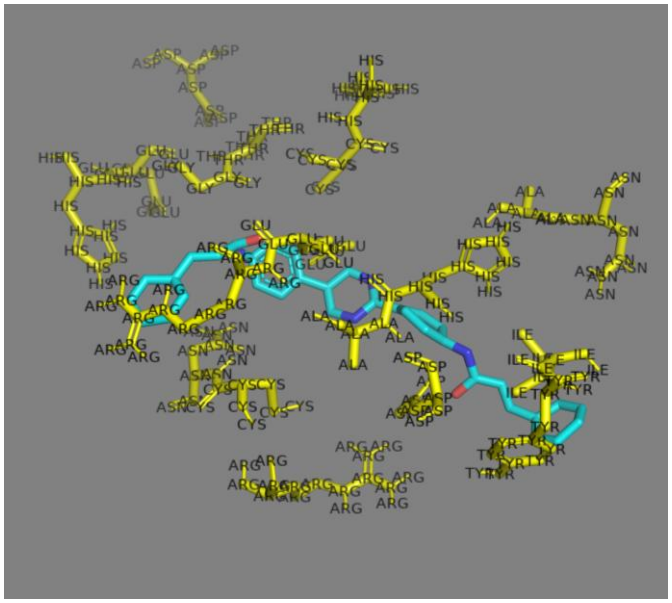


Figure 14N. Amino acid residues in the active site of target DNA mismatch repair protein MutL within 3.5 Å of the ligand molecule ZINC03143011.

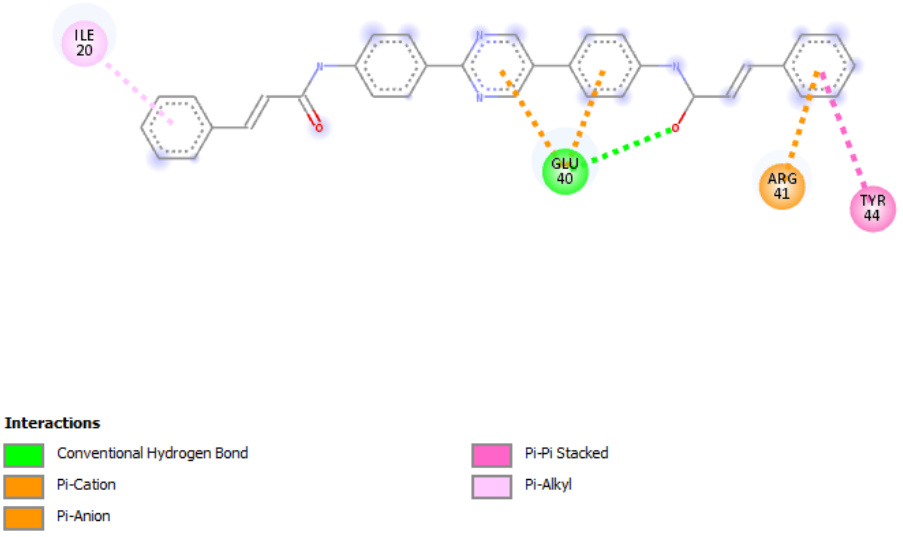


Figure 14O. DNA mismatch repair protein MutL with ZINC molecule ZINC03143011 exhibiting amino acid interactions.

Table 16. Drug targets along with their FDA approved compounds and docking score.

PROTEIN ID	NAME OF THE PROTEIN	FDA APPROVED DRUG ID	BINDING ENERGY (KCAL/MOL)
TRINITY_DN37933_c1_g1_i1	Uncharacterized sodium-dependent transporter MJ1319	DB11753	-11.9
TRINITY_DN53523_c1_g5_i1	Ribonucleoside-diphosphate reductase 1 subunit alpha (EC 1.17.4.1) (Protein B1) (Ribonucleoside-diphosphate reductase 1 R1 subunit) (Ribonucleotide reductase 1)	DB11753	-11.5
TRINITY_DN47007_c0_g1_i1	ADP-ribosyl cyclase/cyclic ADP-ribose hydrolase 1 (EC 3.2.2.6) (2'-phospho-ADP-ribosyl cyclase) (2'-phospho-ADP-ribosyl cyclase/2'-phospho-cyclic-ADP-ribose transferase) (EC 2.4.99.20) (2'-phospho-cyclic-ADP-ribose transferase) (ADP-ribosyl cyclase 1) (ADPRC 1) (CD38H) (Cyclic ADP-ribose hydrolase 1) (cADPr hydrolase 1) (CD antigen CD38)	DB11753	-11.5
TRINITY_DN51304_c6_g6_i1	Anaerobic glycerol-3-phosphate dehydrogenase subunit A (G-3-P dehydrogenase) (EC 1.1.5.3)	DB00224	-11.3
TRINITY_DN52182_c2_g1_i2	Aerobic glycerol-3-phosphate dehydrogenase (EC 1.1.5.3)	DB11753	-11.3
TRINITY_DN53087_c8_g1_i1	Uncharacterized HTH-type transcriptional regulator YjiR	DB11753	-11.3
TRINITY_DN27140_c0_g1_i1	Tyrosine--tRNA ligase (EC 6.1.1.1) (Tyrosyl-tRNA synthetase) (TyrRS)	DB00365	-11.2
TRINITY_DN53053_c2_g1_i1	6-phospho-beta-glucosidase GmuD (EC 3.2.1.86) (Aryl-phospho-beta-D-glucosidase BglD) (Glucomannan utilization protein D)	DB13997	-11.1
TRINITY_DN55161_c0_g1_i1	1,2-phenylacetyl-CoA epoxidase, subunit E (EC 1.-.-.-) (1,2-phenylacetyl-CoA epoxidase, reductase subunit) (1,2-phenylacetyl-CoA monooxygenase, subunit E)	DB11753	-11.1
TRINITY_DN51571_c3_g3_i4	Argininosuccinate synthase (EC 6.3.4.5) (Citrulline--aspartate ligase)	DB11753	-11
TRINITY_DN42811_c0_g1_i1	GDP-6-deoxy-D-talose 4-dehydrogenase (EC 1.1.1.135) (GDP-4-keto-6-deoxy-D-mannose reductase)	DB11753	-11
TRINITY_DN53361_c7_g1_i1	ATP-binding protein SyrD	DB11753	-10.9
TRINITY_DN45788_c5_g1_i1	Protease 2 (EC 3.4.21.83) (Oligopeptidase B) (Protease II)	DB11753	-10.8
TRINITY_DN53246_c2_g2_i1	Iron(3+)-hydroxamate import ATP-binding protein FhuC (EC 3.6.3.34) (Ferric hydroxamate uptake protein C) (Ferrichrome transport ATP-binding protein FhuC) (Iron(III)-hydroxamate import ATP-binding protein FhuC)	DB11753	-10.8
TRINITY_DN45661_c0_g1_i1	Fructokinase (EC 2.7.1.4)	DB11753	-10.8

Table contd...

TRINITY_DN52309_c3_g1_i7	Oxygen-independent coproporphyrinogen III oxidase (CPO) (EC 1.3.98.3) (Coproporphyrinogen III dehydrogenase) (CPDH)	DB11753	-10.8
TRINITY_DN53174_c2_g1_i2	6-phospho-beta-glucosidase GmuD (EC 3.2.1.86) (Aryl-phospho-beta-D-glucosidase BglD) (Glucomannan utilization protein D)	DB11753	-10.7
TRINITY_DN48438_c6_g2_i2	Putative kinase EAE_16955	DB11753	-10.7
TRINITY_DN46299_c1_g1_i1	Endoribonuclease Dcr-1 (Protein dicer-1) (EC 3.1.26.-)	DB11753	-10.7
TRINITY_DN53410_c1_g5_i4	Arginine transport ATP-binding protein ArtP (EC 3.6.3.-)	DB11753	-10.7
TRINITY_DN47557_c7_g2_i1	HTH-type transcriptional regulator TauR	DB08930	-10.6
TRINITY_DN53572_c2_g2_i2	Probable alcohol dehydrogenase (EC 1.1.1.1)	DB00224	-10.6
TRINITY_DN50323_c9_g1_i1	High-affinity branched-chain amino acid transport ATP-binding protein BraG	DB11753	-10.6
TRINITY_DN4138_c0_g1_i1	Oxygen-independent coproporphyrinogen III oxidase (CPO) (EC 1.3.98.3) (Coproporphyrinogen III dehydrogenase) (CPDH)	DB00224	-10.6
TRINITY_DN53556_c0_g3_i1	3-phenylpropionate/cinnamic acid dioxygenase ferredoxin--NAD(+) reductase component (EC 1.18.1.3)	DB11753	-10.6
TRINITY_DN36963_c0_g1_i2	Chloride intracellular channel exc-4 (Excretory canal abnormal protein 4)	DB11753	-10.6
TRINITY_DN53322_c3_g1_i4	Pyridoxal 4-dehydrogenase (tPLDH) (EC 1.1.1.107)	DB08930	-10.6
TRINITY_DN53584_c2_g1_i1	Ribosomal RNA large subunit methyltransferase E (EC 2.1.1.166) (23S rRNA Um2552 methyltransferase) (rRNA (uridine-2'-O-)-methyltransferase)	DB11753	-10.6
TRINITY_DN43567_c0_g1_i1	Ribonucleoside-diphosphate reductase subunit beta (EC 1.17.4.1) (Ribonucleotide reductase small subunit)	DB11753	-10.6
TRINITY_DN52309_c3_g1_i1	Oxygen-independent coproporphyrinogen III oxidase (CPO) (EC 1.3.98.3) (Coproporphyrinogen III dehydrogenase) (CPDH)	DB00224	-10.5
TRINITY_DN52309_c3_g1_i6	Oxygen-independent coproporphyrinogen III oxidase (CPO) (EC 1.3.98.3) (Coproporphyrinogen III dehydrogenase) (CPDH)	DB00224	-10.5
TRINITY_DN43605_c3_g1_i1	Acetylornithine deacetylase (AO) (Acetylornithinase) (EC 3.5.1.16) (N-acetylornithinase) (NAO)	DB11753	-10.5
TRINITY_DN52182_c2_g4_i1	Aerobic glycerol-3-phosphate dehydrogenase (EC 1.1.5.3)	DB08930	-10.5
TRINITY_DN48924_c7_g3_i1	6-phospho-beta-glucosidase BglA (EC 3.2.1.86) (Phospho-beta-glucosidase A)	DB00365	-10.4
TRINITY_DN47921_c4_g2_i1	Thiosulfate sulfurtransferase YnjE (EC 2.8.1.1)	DB11753	-10.4

Table contd...

TRINITY_DN52550_c1_g1_i1	NADH oxidase (NOXase) (EC 1.6.99.3)	DB00224	-10.4
TRINITY_DN52453_c10_g3_i7	Alcohol dehydrogenase YqhD (EC 1.1.1.-)	DB00224	-10.4
TRINITY_DN37427_c2_g1_i1	Alcohol dehydrogenase (EC 1.1.1.1) (EC 1.1.1.4) (EC 1.2.1.3)	DB11799	-10.4
TRINITY_DN47763_c2_g2_i1	Ribosomal protein S12 methylthiotransferase RimO (S12 MTTase) (S12 methylthiotransferase) (EC 2.8.4.4) (Ribosomal protein S12 (aspartate-C(3))-methylthiotransferase) (Ribosome maturation factor RimO)	DB08930	-10.4
TRINITY_DN48169_c3_g1_i1	Uncharacterized peptidase SSP1059 (EC 3.4.-.-)	DB11753	-10.4
TRINITY_DN16205_c0_g1_i1	(R,R)-butanediol dehydrogenase (EC 1.1.1.4)	DB11753	-10.3
TRINITY_DN46819_c12_g1_i1	Pyrimidine 5'-nucleotidase YjjG (EC 3.1.3.5) (House-cleaning nucleotidase) (Non-canonical pyrimidine nucleotide phosphatase) (Nucleoside 5'-monophosphate phosphohydrolase) (dUMP phosphatase)	DB00365	-10.3
TRINITY_DN51782_c10_g1_i1	Fumarate reductase flavoprotein subunit (EC 1.3.5.4) (Flavocytochrome c) (Flavocytochrome c3) (Fcc3)	DB00365	-10.2
TRINITY_DN53480_c0_g3_i1	Aldehyde-alcohol dehydrogenase [Includes: Alcohol dehydrogenase (ADH) (EC 1.1.1.1); Acetaldehyde dehydrogenase [acetylating] (ACDH) (EC 1.2.1.10); Pyruvate-formate-lyase deactivase (PFL deactivase)]	DB13997	-10.2
TRINITY_DN50112_c2_g1_i1	NADPH-dependent butanol dehydrogenase (BDH) (EC 1.1.1.-)	DB13997	-10.2
TRINITY_DN53366_c4_g2_i8	NAD kinase (EC 2.7.1.23) (ATP-dependent NAD kinase)	DB11753	-10.2
TRINITY_DN53178_c5_g1_i1	Bifunctional polymyxin resistance protein ArnA [Includes: UDP-4-amino-4-deoxy-L-arabinose formyltransferase (EC 2.1.2.13) (ArnAFT) (UDP-L-Ara4N formyltransferase); UDP-glucuronic acid oxidase, UDP-4-keto-hexauronic acid decarboxylating (EC 1.1.1.305) (ArnADH) (UDP-GlcUA decarboxylase) (UDP-glucuronic acid dehydrogenase)]	DB11753	-10.2
TRINITY_DN52373_c7_g2_i1	N-hydroxyarylamine O-acetyltransferase (EC 2.3.1.118) (Arylamine N-acetyltransferase) (Arylhydroxamate N,O-acetyltransferase)	DB11753	-10.2
TRINITY_DN44239_c0_g3_i1	Sialic acid transporter NanT (Sialic acid permease) (Sialic acid/H(+) symporter)	DB00220	-10.2
TRINITY_DN53416_c2_g2_i1	Uncharacterized 37.6 kDa protein in cld 5'region (ORF2)	DB11753	-10.2
TRINITY_DN50952_c4_g1_i2	D-alanine--D-alanyl carrier protein ligase (DCL) (EC 6.2.1.-) (D-alanine--poly(phosphoribitol) ligase subunit 1) (D-alanine-activating enzyme) (DAE)	DB11799	-10.1
TRINITY_DN53328_c1_g3_i3	Erythronate-4-phosphate dehydrogenase (EC 1.1.1.290)	DB11753	-10.1

Table contd...

TRINITY_DN44136_c0_g2_i1	GTPase HflX (GTP-binding protein HflX)	DB11753	-10.1
TRINITY_DN51680_c5_g2_i3	Uncharacterized sulfatase YidJ (EC 3.1.6.-)	DB00224	-10.1
TRINITY_DN44420_c2_g2_i1	NADP-dependent alcohol dehydrogenase C (EC 1.1.1.2)	DB11753	-10.1
TRINITY_DN53530_c3_g1_i2	Uracil phosphoribosyltransferase (EC 2.4.2.9) (UMP pyrophosphorylase) (UPRTase)	DB08930	-10.1
TRINITY_DN51710_c2_g1_i2	Selenide, water dikinase (EC 2.7.9.3) (Selenium donor protein) (Selenophosphate synthase)	DB11753	-10.1
TRINITY_DN47572_c0_g1_i1	Ferritin light chain 1 (Ferritin L subunit 1)	DB11753	-10.1
TRINITY_DN53589_c8_g1_i4	Uncharacterized protein YdiJ	DB11799	-10
TRINITY_DN45419_c0_g1_i1	Xylulose kinase (Xylulokinase) (EC 2.7.1.17)	DB00220	-10
TRINITY_DN32575_c0_g1_i1	Acetylornithine deacetylase (AO) (Acetylornithinase) (EC 3.5.1.16) (N-acetylornithinase) (NAO)	DB08930	-10
TRINITY_DN53489_c2_g4_i1	Elongation factor P--(R)-beta-lysine ligase (EF-P--(R)-beta-lysine ligase) (EC 6.3.1.-) (EF-P post-translational modification enzyme A) (EF-P-lysine lysyltransferase)	DB00224	-10
TRINITY_DN53449_c2_g1_i1	Putative TrmH family tRNA/rRNA methyltransferase YacO (EC 2.1.1.-)	DB11753	-10
TRINITY_DN53377_c1_g5_i1	23S rRNA (guanosine-2'-O-)-methyltransferase RlmB (EC 2.1.1.185) (23S rRNA (guanosine2251 2'-O)-methyltransferase) (23S rRNA Gm2251 2'-O-methyltransferase)	DB11799	-10
TRINITY_DN21017_c0_g1_i1	Ribonuclease 3 (EC 3.1.26.3) (Ribonuclease III) (RNase III)	DB00220	-10
TRINITY_DN53530_c3_g1_i1	Uracil phosphoribosyltransferase (EC 2.4.2.9) (UMP pyrophosphorylase) (UPRTase)	DB13997	-10
TRINITY_DN52985_c1_g1_i1	Putative ribose/galactose/methyl galactoside import ATP-binding protein 3 (EC 3.6.3.17)	DB13997	-9.9
TRINITY_DN41347_c0_g1_i1	Argininosuccinate synthase (EC 6.3.4.5) (Citrulline--aspartate ligase)	DB11753	-9.9
TRINITY_DN48674_c0_g1_i1	Gamma-soluble NSF attachment protein (SNAP-gamma) (N-ethylmaleimide-sensitive factor attachment protein gamma)	DB11753	-9.9
TRINITY_DN50525_c5_g3_i4	Multidrug resistance protein MdtK (Multidrug-efflux transporter)	DB00224	-9.9
TRINITY_DN53568_c3_g4_i4	Ribonuclease 3 (EC 3.1.26.3) (Ribonuclease III) (RNase III)	DB00625	-9.9
TRINITY_DN48524_c4_g1_i1	Pantothenate precursors transporter PanS	DB11753	-9.9
TRINITY_DN53232_c3_g2_i3	Diaminobutyrate--2-oxoglutarate aminotransferase (EC 2.6.1.76) (Diaminobutyrate transaminase) (L-2,4-diaminobutyrate:2-ketoglutarate	DB11753	-9.8

Table contd...

	4-aminotransferase) (DABA aminotransferase) (DABA-AT) (L-diaminobutyric acid transaminase)		
TRINITY_DN44483_c1_g2_i1	Glycogen synthase (EC 2.4.1.21) (Starch [bacterial glycogen] synthase)	DB11753	-9.8
TRINITY_DN52331_c8_g1_i5	N-methyl-L-tryptophan oxidase (MTOX) (EC 1.5.3.-)	DB11753	-9.8
TRINITY_DN48199_c5_g1_i1	Nitric oxide reductase FlRd-NAD(+) reductase (EC 1.18.1.-) (Flavorubredoxin reductase) (FlRd-reductase) (FlavoRb reductase)	DB11753	-9.8
TRINITY_DN50525_c5_g3_i1	Multidrug resistance protein MdtK (Multidrug-efflux transporter)	DB00224	-9.8
TRINITY_DN33027_c0_g1_i1	Chloride intracellular channel exc-4 (Excretory canal abnormal protein 4)	DB11753	-9.8
TRINITY_DN52797_c3_g2_i2	Short-chain dehydrogenase/reductase ATR9 (EC 1.-.-.-) (Core atranone cluster (CAC) protein 9)	DB11753	-9.8
TRINITY_DN42869_c0_g1_i1	16 kDa calcium-binding protein (Egg antigen SME16)	DB11753	-9.8
TRINITY_DN50599_c5_g1_i2	DNA topoisomerase 4 subunit B (EC 5.99.1.3) (Topoisomerase IV subunit B)	DB05154	-9.7
TRINITY_DN52324_c7_g6_i2	H(+)/Cl(-) exchange transporter ClcA	DB11753	-9.7
TRINITY_DN53504_c3_g1_i1	Uncharacterized HTH-type transcriptional regulator YdcR	DB11753	-9.7
TRINITY_DN44708_c0_g1_i1	Ribulokinase (EC 2.7.1.16)	DB11753	-9.7
TRINITY_DN52331_c8_g1_i3	N-methyl-L-tryptophan oxidase (MTOX) (EC 1.5.3.-)	DB11753	-9.7
TRINITY_DN51571_c3_g3_i2	Argininosuccinate synthase (EC 6.3.4.5) (Citrulline--aspartate ligase)	DB11753	-9.7
TRINITY_DN50507_c4_g1_i1	Elongation factor Ts (EF-Ts)	DB11753	-9.7
TRINITY_DN50862_c2_g1_i1	Probable N-glycosylase/DNA lyase [Includes: 8-oxoguanine DNA glycosylase (EC 3.2.2.-); DNA-(apurinic or apyrimidinic site) lyase (AP lyase) (EC 4.2.99.18)]	DB00625	-9.7
TRINITY_DN44239_c0_g2_i1	Sialic acid transporter NanT (Sialic acid permease) (Sialic acid/H(+) symporter)	DB00224	-9.7
TRINITY_DN53535_c4_g2_i2	Chelated iron transport system membrane protein YfeB	DB00625	-9.7
TRINITY_DN53506_c0_g1_i1	Xylulose kinase (Xylulokinase) (EC 2.7.1.17)	DB00224	-9.6
TRINITY_DN52453_c10_g3_i8	Alcohol dehydrogenase YqhD (EC 1.1.1.-)	DB11799	-9.6
TRINITY_DN48353_c3_g1_i1	Caffeyl-CoA reductase-Etf complex subunit CarE (EC 1.3.1.108) (Electron transfer flavoprotein large subunit) (ETFLS) (Electron transfer flavoprotein subunit alpha) (Alpha-ETF)	DB11799	-9.6

Table contd...

TRINITY_DN9533_c0_g1_i1	Alcohol dehydrogenase (EC 1.1.1.1) (EC 1.1.1.4) (EC 1.2.1.3)	DB11753	-9.6
TRINITY_DN51533_c3_g1_i1	Methionyl-tRNA formyltransferase (EC 2.1.2.9)	DB00365	-9.6
TRINITY_DN44794_c0_g1_i1	Quinone oxidoreductase (EC 1.6.5.5) (NADPH:quinone reductase) (Zeta-crystallin homolog protein)	DB00365	-9.6
TRINITY_DN52750_c8_g1_i2	Taurine import ATP-binding protein TauB (EC 7.6.2.7)	DB11753	-9.6
TRINITY_DN51597_c1_g3_i2	Proton/glutamate-aspartate symporter (Glutamate-aspartate carrier protein) (Proton-glutamate-aspartate transport protein)	DB11753	-9.5
TRINITY_DN49813_c1_g1_i4	Glutamate-pyruvate aminotransferase AlaC (EC 2.6.1.2)	DB00224	-9.5
TRINITY_DN52026_c3_g1_i3	Fe(3+) dicitrate transport ATP-binding protein FecE (Iron(III) dicitrate transport ATP-binding protein FecE)	DB11753	-9.5
TRINITY_DN43623_c0_g1_i1	ADP-ribosyl cyclase/cyclic ADP-ribose hydrolase 1 (EC 3.2.2.6) (2'-phospho-ADP-ribosyl cyclase) (2'-phospho-ADP-ribosyl cyclase/2'-phospho-cyclic-ADP-ribose transferase) (EC 2.4.99.20) (2'-phospho-cyclic-ADP-ribose transferase) (ADP-ribosyl cyclase 1) (ADPRC 1) (CD38H) (Cyclic ADP-ribose hydrolase 1) (cADPr hydrolase 1) (CD antigen CD38)	DB00224	-9.5
TRINITY_DN52173_c4_g2_i2	Enterobactin synthase component F (EC 2.7.7.-) (Enterochelin synthase F) (Serine-activating enzyme) (Seryl-AMP ligase)	DB11753	-9.5

Table 17. Drug targets along with their ZINC molecules and docking score.

PROTEIN ID	NAME OF THE PROTEIN	RASPD_ZINC ID	BINDING ENERGY (KCAL/MOL)
TRINITY_DN53305_c8_g3_i3	Bifunctional protein Aas [Includes: 2-acylglycerophosphoethanolamine acyltransferase (EC 2.3.1.40) (2-acyl-GPE acyltransferase) (Acyl-[acyl-carrier-protein]--phospholipid O-acyltransferase); Acyl-[acyl-carrier-protein] synthetase (EC 6.2.1.20) (Acyl-ACP synthetase) (Long-chain-fatty-acid--[acyl-carrier-protein] ligase)]	ZINC02820058	-11.4
TRINITY_DN31968_c1_g1_i1	50S ribosomal protein L3 glutamine methyltransferase (L3 MTase) (EC 2.1.1.298) (N5-glutamine methyltransferase PrmB)	ZINC02820058	-11.2
TRINITY_DN50525_c2_g1_i1	Glyoxylate/succinic semialdehyde reductase 1 (AtGLYR1) (AtGR1) (SSA reductase 1) (EC 1.1.1.79) (EC 1.1.1.n11) (Gamma-hydroxybutyrate dehydrogenase) (AtGHBDH)	ZINC12576410	-10.7
TRINITY_DN44792_c2_g1_i1	Bifunctional NAD(P)H-hydrate repair enzyme Nnr (Nicotinamide nucleotide repair protein) [Includes: ADP-dependent (S)-NAD(P)H-hydrate dehydratase (EC 4.2.1.136) (ADP-dependent NAD(P)HX dehydratase); NAD(P)H-hydrate epimerase (EC 5.1.99.6) (NAD(P)HX epimerase)]	ZINC02820058	-10.5
TRINITY_DN42486_c1_g2_i1	Putative phosphonates utilization ATP-binding protein PhnK	ZINC02820058	-10.5
TRINITY_DN53290_c1_g2_i2	Maltodextrin import ATP-binding protein MsmX (EC 3.6.3.-)	ZINC02060288	-10.4
TRINITY_DN10786_c0_g1_i1	Methanol dehydrogenase activator (EC 3.-.-.-) (MDH activator)	ZINC02060288	-10.2
TRINITY_DN37816_c0_g2_i1	Diaminobutyrate--2-oxoglutarate transaminase (EC 2.6.1.76)	ZINC02820058	-10
TRINITY_DN44072_c0_g1_i1	Branched-chain-amino-acid aminotransferase (BCAT) (EC 2.6.1.42) (Transaminase B)	ZINC02820058	-10
TRINITY_DN53535_c4_g2_i1	Chelated iron transport system membrane protein YfeB	ZINC12576410	-9.9
TRINITY_DN46184_c0_g1_i1	Galactose/methyl galactoside import ATP-binding protein MglA (EC 3.6.3.17)	ZINC02820058	-9.9
TRINITY_DN53206_c4_g1_i4	Alpha-xylosidase (EC 3.2.1.177)	ZINC02820058	-9.8
TRINITY_DN37304_c0_g2_i1	Glyoxylate reductase (EC 1.1.1.26)	ZINC12576410	-9.8
TRINITY_DN52927_c8_g4_i1	Cobalt import ATP-binding protein CbiO (EC 3.6.3.-) (Energy-coupling factor transporter ATP-binding protein CbiO) (ECF transporter A component CbiO)	ZINC12576410	-9.7
TRINITY_DN37544_c1_g2_i1	1-acyl-sn-glycerol-3-phosphate acyltransferase (1-AGP acyltransferase) (1-AGPAT) (EC 2.3.1.51) (Lysophosphatidic acid acyltransferase) (LPAAT)	ZINC12576410	-9.7
TRINITY_DN15899_c0_g1_i1	Octaprenyl diphosphate synthase (EC 2.5.1.90) (All-trans-octaprenyl-diphosphate synthase) (Octaprenyl pyrophosphate synthase) (OPP synthase)	ZINC12576410	-9.7

Table contd...

TRINITY_DN44431_c1_g1_i1	Transketolase 2 (TK 2) (EC 2.2.1.1)	ZINC12576410	-9.6
TRINITY_DN49800_c0_g1_i2	Queuine tRNA-ribosyltransferase accessory subunit 2 (Queuine tRNA-ribosyltransferase domain-containing protein 1)	ZINC02820058	-9.6
TRINITY_DN43187_c0_g1_i1	Nicotinate phosphoribosyltransferase (NAPRTase) (EC 6.3.4.21)	ZINC12376046	-9.5
TRINITY_DN48924_c7_g3_i2	Beta-glucoside kinase (EC 2.7.1.85)	ZINC02820058	-9.5
TRINITY_DN51775_c4_g1_i1	Methionine--tRNA ligase (EC 6.1.1.10) (Methionyl-tRNA synthetase) (MetRS)	ZINC02895117	-9.4
TRINITY_DN53020_c0_g1_i1	Outer envelope protein 64, mitochondrial (Mitochondrial outer membrane protein 64) (mtOM64) (Translocon at the outer membrane of chloroplasts 64-V) (AtTOC64-V)	ZINC12576410	-9.4
TRINITY_DN51718_c2_g2_i2	Uncharacterized ABC transporter ATP-binding protein YdiF	ZINC12376046	-9.4
TRINITY_DN43187_c0_g1_i2	Nicotinate phosphoribosyltransferase (NAPRTase) (EC 6.3.4.21)	ZINC02060288	-9.4
TRINITY_DN48109_c2_g2_i1	Xylose import ATP-binding protein XylG (EC 3.6.3.17)	ZINC02820058	-9.4
TRINITY_DN46129_c0_g1_i1	Ribonuclease Oy (RNase Oy) (EC 3.1.27.-)	ZINC02895117	-9.4
TRINITY_DN50814_c0_g1_i1	Testicular acid phosphatase homolog (EC 3.1.3.2) (Acid phosphatase 4)	ZINC12576410	-9.3
TRINITY_DN51896_c8_g2_i1	Aromatic-amino-acid aminotransferase 1 (ARAT-I) (AROAT) (EC 2.6.1.57)	ZINC02060288	-9.3
TRINITY_DN41200_c0_g2_i1	Putative quercetin 2,3-dioxygenase sll1773 (Putative quercetinase) (EC 1.13.11.24) (Pirin-like protein sll1773)	ZINC02820058	-9.2
TRINITY_DN46783_c0_g1_i1	Probable serine/threonine-protein phosphatase PP2A regulatory subunit (Protein phosphatase PP2A regulatory subunit A)	ZINC02820058	-9
TRINITY_DN44084_c0_g1_i1	Poly(U)-specific endoribonuclease-B (EC 3.1.-.-) (Protein endoU-B) (Uridylate-specific endoribonuclease-B) (XendoU-B)	ZINC12576410	-9
TRINITY_DN53087_c8_g1_i2	Uncharacterized HTH-type transcriptional regulator YjiR	ZINC02895117	-8.6
TRINITY_DN55642_c0_g1_i1	DNA mismatch repair protein MutL	ZINC02820058	-8.6
TRINITY_DN51862_c4_g1_i2	Oligopeptide transport ATP-binding protein OppF	ZINC02060288	-8.5
TRINITY_DN16089_c0_g1_i1	4-hydroxyphenylpyruvate dioxygenase (4HPPD) (HPD) (HPPDase) (EC 1.13.11.27)	ZINC08892130	-8.5
TRINITY_DN49362_c15_g1_i1	1,4-alpha-glucan branching enzyme GlgB (EC 2.4.1.18) (1,4-alpha-D-glucan:1,4-alpha-D-glucan 6-glucosyl-transferase) (Alpha-(1->4)-glucan branching enzyme) (Glycogen branching enzyme) (BE)	ZINC12576410	-8.5
TRINITY_DN43492_c0_g1_i1	DNA mismatch repair protein MutL	ZINC03143011	-8.5

Table contd...

TRINITY_DN44003_c0_g1_i1	32 kDa beta-galactoside-binding lectin (Galectin-1)	ZINC02820058	-8.3
TRINITY_DN48390_c0_g1_i1	Lectin L6	ZINC03143011	-8.3
TRINITY_DN42486_c1_g3_i1	Alpha-D-ribose 1-methylphosphonate 5-triphosphate synthase subunit PhnL (RPnTP synthase subunit PhnL) (EC 2.7.8.37)	ZINC12576410	-8.3
TRINITY_DN45099_c0_g1_i1	Calmodulin (CaM)	ZINC02820058	-8.2
TRINITY_DN53318_c7_g2_i1	Protein MalY [Includes: Cystathionine beta-lyase MalY (CBL) (EC 4.4.1.13) (Beta-cystathionase MalY) (Cysteine lyase MalY) (Cysteine-S-conjugate beta-lyase MalY); Maltose regulon modulator]	ZINC12576410	-8.1
TRINITY_DN50571_c2_g1_i3	Poly(U)-specific endoribonuclease-B (EC 3.1.-.-) (Protein endoU-B) (Uridylate-specific endoribonuclease-B)	ZINC02820058	-8

Table 18. Details of the top five protein target interactions with FDA approved drug molecules (1-5) with binding affinity cutoff of ≤-9.5 kcal/mol. Top five drug targets with ZINC molecules (6-10) with significant binding affinity cutoff of ≤-8.0 kcal/mol.

Sl. No.	*PROTEIN LIGAND COMPLEX*	*AUTODOCK SCORE (KCAL/MOL)*	*AMINO ACID RESIDUES INVOVLVED*	*CELLULAR FUNCTION*	*BIOLOGICAL FUNCTION*	*MOLECULAR FUNCTION*
1	Uncharacterized sodium-dependent transporter_DB11753	-11.9	TYR235, ARG34, GLN232	integral component of membrane, plasma membrane	transmembrane transport	sodium symporter activity
2	Anaerobic glycerol-3-phosphate dehydrogenase subunit A_ DB00224	-11.3	ARG59, TYR290, GLU514, ARG470, ARG474, ASP159, PRO160, ARG515	glycerol-3-phosphate dehydrogenase complex	glycerol-3-phosphate metabolic process	sn-glycerol-3-phosphate:ubiquinone oxidoreductase activity
3	Tyrosine-tRNA synthetase _ DB00365	-11.2	VAL94,GLY227,GLY228,GLN213,GLY79,TYR209,ASP81,ASP121	cytoplasm	rRNA catabolic process	-
4	Argininosuccinate synthase_ DB11753	-11.0	ARG178,GLU197, SER195,GLY30,GLU59,LEU31,ASP32,LYS214	cytoplasm	arginine biosynthetic process	argininosuccinate synthase activity
5	Chloride intracellular channel _ DB11753	-10.6	SER59,HIS19,ALA124,ILE57,ASP55	nucleoplasm, cytosol, mitochondrion, membrane	chloride transmembrane transport	glutathione peroxidase activity, chloride channel activity, lipid binding
6	Bifunctional protein Aas_ ZINC02820058	-11.4	VAL255, PHE251, HIS249,ALA320,THR346,VAL345,ARG444,ASP429,VAL365,LYS322,ILE441	plasma membrane	fatty acid metabolic process	phospholipid O-acyltransferase activity
7	50S ribosomal protein L3 glutamine methyltransferase_ ZINC02820058	-11.2	ARG130,TYR63,SER131,PRO377,ASP59,ARG37	Ribosome	DNA methylation on adenine	site-specific DNA-methyltransferase (adenine-specific) activity
8	Putative phosphonates utilization ATP-binding protein PhnK_ ZINC02820058	-11.2	ILE73,TRP107,SER74,LEU103,VAL109,ARG131,	carbon phosphorus lyase complex	organic phosphonate catabolic process	ATPase activity

	Table contd..					
			ARG104,HIS111,ARG180,ALA134,GLU130			
9	Cobalt import ATP-binding protein CbiO_ ZINC12576410	-9.7	PHE51,ASP168,PHE90,ARG83,ALA87,LEU160,SER109,GLN96	plasma membrane	cobalt ion transport	cobalt ion transmembrane transporter activity
10	DNA mismatch repair protein MutL_ ZINC03143011	-8.5	ILE20,GLU40,ARG41,TYR44	mismatch repair complex	mismatch repair, nucleic acid phosphodiester bond hydrolysis	mismatched DNA binding, ATPase activity, ATP binding, endonuclease activity

4.4. Discussion

Helminth infections cause a significant burden in terms of public and animal health, but the available medication arsenal is limited. Given the high cost of drug research and development, as well as the high failure rates and the excessively long time it takes in developing new therapies, drug repurposing can be an effective alternative which can overcome these challenges to a great extent. Because repurposed drugs have already been approved for clinical use, they can save time and money throughout the drug research process. Several cases show that repurposed drugs can be used to treat bacteria, viruses, fungi and parasites (Zheng et al., 2018). Due to lack of medications to treat helminthiases with a decent safety and efficacy profile a few decades ago, this group of diseases caused considerable morbidity in people of all ages in impoverished nations. Food-borne diseases are mostly treated with two drugs: Praziquantel for intestinal fluke infections, clonorchiasis, opisthorchiasis, and paragonimiasis, and Triclabendazole for fascioliasis and paragonimiasis. Current consensus suggests that praziquantel antagonises voltage-gated calcium channels leading to uncontrolled muscle contraction and paralysis, while triclabendazole is known to inhibit microtubule formation (Fairweather, 2020; Thomas and Timson, 2020). Drug resistance or deterioration in their efficacy develops as a result of long-term therapy with the same treatments. Drug-resistant pathogens are becoming increasingly common, posing a significant challenge to the treatment and control of emerging and re-emerging infectious illnesses.

Five of the 102 putative therapeutic targets identified in the current work from the *A. sufrartyfex* transcriptome, which demonstrated better binding affinity with FDA-approved drugs in DrugBank has been shortlisted. Sodium dependent transporter

(uncharacterized), argininosuccinate synthase, chloride intracellular channel proteins, anaerobic glycerol-3-phosphate dehydrogenase and tyrosyl tRNA synthetase are among them. Three FDA approved drugs namely Rifamycin (DB11753), Indinavir (DB00224) and Grepafloxacin (DB00365) which showed the most significant interaction against these targets has been identified herein. Rifamycin is a natural antibiotic and antimycobacterial medication. Adult patients with travellers' diarrhoea caused by *Escherichia coli*, as well as drug-resistant *Mycobacterium* TB, are treated with the drug. Grepafloxacin is a fluoroquinolone antibiotic used to treat diverse gram positive and negative bacterial infections, whilst Indinavir is an antiretroviral protease inhibitor used to treat HIV infection and AIDS. The current investigation revealed that Rifamycin interacts with three of the five targets, namely sodium-dependent transporter (uncharacterized), argininosuccinate synthase, and chloride intracellular channel proteins, implying that this compound has a larger function against *A. sufrartyfex*. A sodium-dependent transporter in the malaria parasite *Plasmodium falciparum* is known to be critically involved in the uptake of inorganic phosphates (Saliba et al., 2006). Few studies have suggested that argininosuccinate synthase is important in the kinetoplastid *Leishmania donovani* for balancing stressful conditions during infection and that it could be used as a therapeutic target (Lakhal-Naouar et al., 2012; Sardar et al., 2016). Chloride intracellular channel proteins have been discovered in the nematode *Caenorhabditis elegans* in response to heat stress (Liang et al., 2017).

Indinavir and Grepafloxacin, on the other hand, have shown strong binding affinity for anaerobic glycerol-3-phosphate dehydrogenase and tyrosyl tRNA synthetase, respectively. CsGPD, a novel *Clonorchis sinensis* glycerol-3-phosphate dehydrogenase

(CsGPD) expressed in the parasite's tegument, has already been discovered as a potential vaccine candidate (Fan et al., 2011). The malaria parasite *Plasmodium falciparum* secretes a tyrosyl-tRNA synthetase that causes pro-inflammatory reactions in human hosts, suggesting that it could be exploited in anti-parasitic treatments (Bhatt et al., 2011). Similarly, *L. donovani* tyrosyl-tRNA synthetase's immunomodulatory role as a protein translation enzyme and host chemokine mimic has offered a platform to investigate it as a possible therapeutic target (Anand and Madhubala, 2016). Although several potential target proteins have been found herein to have strong interactions, studies utilizing the same remain scanty in helminth group.

Similarly, out of the 43 pharmacological targets found in the study that had a substantial binding affinity for ZINC compounds, five with the strongest interaction have been shorlisted. Bifunctional protein Aas, 50S ribosomal protein L3 glutamine methyl transferase, putative phosphonates utilisation PhnK ATP-binding protein, cobalt import ATP-binding protein and DNA mismatch repair protein MutL are among the proteins that have been identified as potential targets. In helminth parasite species, the specific mechanisms and functions of these targets remain unknown. However, the bacteria *Escherichia coli* and *Salmonella typhimurium* have provided some biological information about these proteins. In the bacteria *E. coli*, the bifunctional protein Aas is involved in lysophospholipid acylation (Jackowski et al., 1994). The glutamine methyltransferase of the 50S ribosomal protein L3 methylates glutamine position 150 of the 50S ribosomal protein L3. Use of phosphonates that may or may not be true The PhnK ATP-binding protein is part of an operon that controls alkylphosphonate absorption and carbon phosphorus lyase, although its exact function is unknown (Chen et al., 1990). *Salmonella*

typhimurium cobalt import ATP-binding protein is engaged in the adenosylcobalamin biosynthesis process and is thought to be important for energy coupling to the transport mechanism (Roth et al., 1993). MutL homolog, one of the primary components of the mismatch repair mechanism in the malaria parasite *P. falciparum*, is critical for parasite survival and a potential source of novel therapeutic compounds. The ligands netropsin, daunorubicin, etoposide, ethidium bromide, and nogalamycin have been reported to decrease the ATPase activity of the *P. falciparum* MutL homolog (Tarique et al., 2017). The study has identified ZINC molecules ZINC02820058, ZINC12576410, and ZINC03143011 have a high affinity for these lesser-known targets. These findings are intriguing since they represent new prospects for prospective research.

Conclusion and future perspective

- The present study reports the occurrence of the echinostomid *A. sufrartyfex* among child patients visiting SSL Hospital, Sitamarhi, Bihar.
- The study also reports the presence of infective larval stage metacercaria of *A. sufrartyfex* for the first time in *Pila globosa* from foci of infections in northern Bihar.
- Molecular analysis using three gene markers confirms that the child patients visiting SSL Hospital, Sitamarhi were infected with *A. sufrartyfex* and that the snail *Pila globosa* harbouring the infective metacercariae was the potent source of infection to the affected children.
- The study also provides an easy method in the form of DNA barcodes for rapid identification of *A. sufrartyfex* which could act as a reference for future parasite identification.
- Since, the study was carried out in a small geographical pocket of India, the human infection report presented in the current study may not reflect the true burden of *A. sufrarytyfex* infection in the country.
- The symptoms of echinostomiasis due to *A. sufrartyfex* are common with other diarrhoeal diseases, therefore diagnosis of the disease becomes all the more difficult. Also, the symptoms appear long after the parasite is established in the host. Hopefully, the unique diagnostic markers identified in the study from the parasite transcriptome may help in the development of appropriate immunodiagnostic tools, which, in turn, would help in early detection of the

infection and also contribute as a means for proper surveillance of the disease.

- In view of the ever increasing drug resistance phenomenon in case of parasitic helminths and trematode infections in particular, the study provides numerous protein targets along with their FDA approved drugs and ZINC molecules against *A. sufrartyfex*. Though, the study provides important leads in the form of drug targets and possible ligand molecules, further experimental studies are required to confirm the findings. It is hoped that this finding would help in development of appropriate interventions against *A. sufrartyfex*.

References

Adlard, R. D., Barker, S. C., Blair, D., & Cribb, T. H. (1993). Comparison of the second internal transcribed spacer (ribosomal DNA) from populations and species of Fasciolidae (Digenea). *International Journal for Parasitology, 23*(3), 423-425.

Agrawal, V. (1963). Three new trematodes of the family Echinostomatidae Poche, 1926. *Indian Journal of Helminthology, 15*(2), 127-137.

Ahluwalia, S. S. (1962). Studies on some helminths of the domestic pig (*Sus scrofa domestica*) in western Uttar Pradesh. *Journal of Helminthology, 36*(4), 347-364.

Almeida, G. T., Amaral, M. S., Beckedorff, F. C. F., Kitajima, J. P., DeMarco, R., & Verjovski-Almeida, S. (2012). Exploring the *Schistosoma mansoni* adult male transcriptome using RNA-seq. *Experimental Parasitology, 132*(1), 22-31.

Alonso, D., Muñoz, J., Gascón, J., Valls, M. E., & Corachan, M. (2006). Failure of standard treatment with praziquantel in two returned travelers with *Schistosoma haematobium* infection. *The American Journal of Tropical Medicine and Hygiene, 74*(2), 342-344.

Alvarez, L. H., Feliciano, D. N., González, J. E. H., de Oliveira Soares, R., Gomes, D. E. B., & Pascutti, P. G. (2015). Insights into the interactions of *Fasciola hepatica* cathepsin L3 with a substrate and potential novel inhibitors through in silico approaches. *PLoS Neglected Tropical Diseases*, *9*(5), e0003759.

Amrithraj, M., Bindhu, M., & Pillai, K. M. (1999). *Artyfechinostomum malayanum* (Leiper, 1911)Mendheim, 1943 from a small Indian civet (*Veverricula indica*). *Zoos. Print*, *14*(4), 6.

Anand, S., & Madhubala, R. (2016). Twin attributes of tyrosyl-tRNA synthetase of

Leishmania donovani: a housekeeping protein translation enzyme and a mimic of host chemokine. *Journal of Biological Chemistry, 291*(34), 17754-17771.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., ... & Sherlock, G. (2000). Gene ontology: tool for the unification of biology. *Nature Genetics*, *25*(1), 25-29.

Baig, S., Damian, R. T., Molinari, J. L., Tato, P., Morales-Montor, J., Welch, M., ... & White, A. C. (2005). Purification and characterization of a metacestode cysteine proteinase from *Taenia solium* involved in the breakdown of human IgG. *Parasitology, 131(*3), 411-416.

Bal, M., & Das, M. K. (1995). Antigenicity of filarial superoxide dismutase in human bancroftian filariasis. *Journal of Biosciences, 20*(2), 157-166.

Bankar., KG, Todur., VN, Shukla.,RN,& Vasudevan., M.(2015). Ameliorated de novo transcriptome assembly using Illumina paired end sequence data with Trinity Assembler. *Genome Data*, *15*(5),352-9.

Bastiani, C. and Mendel, J. (2006). Heterotrimeric G proteins in *C. elegans*, *WormBook, ed. The C. elegans Research Community*, WormBook, doi/10.1895/wormbook.1.75.1.

Basu, A., Sarkar, A., &Maulik, U. (2020). Molecular docking study of potential phytochemicals and their effects on the complex of SARS-CoV2 spike protein and human ACE2. *Scientific Reports*, *10*(1), 17699.

Batra., S, Singh., SP, Gupta., S, Katiyar., JC, Srivastava., VM.(1990). Reactive oxygen intermediates metabolizing enzymes in *Ancylostoma ceylanicum* and *Nippostrongylus brasiliensis*. *Free Radical Biology & Medicine,8,*271–274.

Beach, M. J., Streit, T. G., Addiss, D. G., Prospere, R., Roberts, J. M., &Lammie, P. J. (1999). Assessment of combined ivermectin and albendazole for treatment of intestinal helminth and *Wuchereria bancrofti* infections in Haitian schoolchildren. *The American Journal of Tropical Medicine and Hygiene, 60*(3), 479-486.

Beaver, P. C., Jung, R. C., & Cupp, E. W. (1984). *Clinical parasitology*. Lea & Febiger.

Belizario, V. Y., Amarillo, M. E., De Leon, W. U., De los Reyes, A. E., Bugayong, M. G., &Macatangay, B. J. C. (2003). A comparison of the efficacy of single doses of albendazole, ivermectin, and diethylcarbamazine alone or in combinations against *Ascaris* and *Trichuris* spp. *Bulletin of the World Health Organization, 81*, 35-42.

Belizario, V. Y., Geronilla, G. G., Anastacio, M. B. M., de Leon, W. U., Suba-an, A. P., Sebastian, A. C., & Bangs, M. J. (2007). *Echinostoma malayanum* infection, the Philippines. *Emerging Infectious Diseases*, *13*(7), 1130.

Bendtsen, J. D., Jensen, L. J., Blom, N., Von Heijne, G., & Brunak, S. (2004). Feature-based prediction of non-classical and leaderless protein secretion. *Protein Engineering Design and Selection, 17*(4), 349-356.

Bhalerao, G. D. (1931). Two new Trematodes from Reptiles: *Paryphostomum indicum n. sp.* and *Stunkardia dilymphosa ng, n. sp. Parasitology, 23*(1), 99-108.

Bhatt, T. K., Khan, S., Dwivedi, V. P., Banday, M. M., Sharma, A., Chandele, A., ... & Sharma, A. (2011). Malaria parasite tyrosyl-tRNA synthetase secretion triggers pro-inflammatory responses. *Nature communications*, *2*(1), 1-1.

Bhattacharyya, H. M., Das, S. K., Sinha, P. K., Biswas, S. N., & Bose, P. K. (1972). Mortality in pig due to *Paryphostomum sufrartyfex* (Lane, 1915) Bhalerao, 1931. *Indian Veterinary Journal.49*,976-978.

Biswal, D. K., Roychowdhury, T., Pandey, P., & Tandon, V. (2018). De novo genome and transcriptome analyses provide insights into the biology of the trematode human parasite *Fasciolopsis buski*. *PLoS One, 13*(10), e0205570.

Blair, D., Agatsuma, T., Watanobe, T., Okamoto, M., & Ito, A. (1997). Geographical genetic structure within the human lung fluke, *Paragonimus westermani*, detected from DNA sequences. *Parasitology, 115(*4), 411-417.

Blair, D., Campos, A., Cummings, M. P., & Laclette, J. P. (1996). Evolutionary biology of parasitic platyhelminths: the role of molecular phylogenetics. *Parasitology Today, 12*(2), 66-71.

Blobel, G., & Dobberstein, B. (1975). Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *The Journal of Cell Biology, 67*(3), 835-851.

Boore, J. L. (1999). Animal mitochondrial genomes. *Nucleic Acids Research, 27*(8), 1767-1780.

Borkataki, S., Goswami, P., Katoch, R., Kumar, S., & Raghuvanshi, P. (2017). Detection of *Artyfechinostomum sufrartyfex*-a zoonotic parasite from the Small Indian Mongoose *Herpestes auropunctatus* (Mammalia: Carnivora: Herpestidae) in Jammu & Kashmir, India. *Journal of Threatened Taxa*, *9*(11), 10980-10982.

Borthakur, S. K., Rahmani, S., & Sarma, K. (2007). Prevalence of gastrointestinal helminths in pigs in Aizawl. *Journal of Veterinary Parasitology, 21*(2), 173-174.

Bourguinat, C., Keller, K., Xia, J., Lepage, P., McTier, T. L., Woods, D. J., & Prichard, R. K. (2017). Genetic profiles of ten *Dirofilaria immitis* isolates susceptible or

resistant to macrocyclic lactone heartworm preventives. *Parasites & vectors*, *10*(2), 31-38.

Bowles, J., Blair, D., & McManus, D. P. (1992). Genetic variants within the genus *Echinococcus* identified by mitochondrial DNA sequencing. *Molecular and Biochemical Parasitology, 54*(2), 165-173.

Bowles, J., Blair, D., & McManus, D. P. (1995). A molecular phylogeny of the human schistosomes. *Molecular Phylogenetics and Evolution*, *4*(2), 103-109.

Bray., LN., Harold, P., Páll M., and Lior, P.(2016). Near-optimal probabilistic RNA-seq quantification, *Nature Biotechnology, 34*, 525–527.

Brophy, P. M., & Pritchard, D. I. (1994). Parasitic helminth glutathione S-transferases: an update on their potential as targets for immuno-and chemotherapy. *Experimental Parasitology, 79*(1), 89-96.

Bueding, E., Hawkins, J., & Cha, Y. N. (1981). Antischistosomal effects of cyclosporin A. *Agents and Actions*, *11*(4), 380-383

Campbell, A. J., Gasser, R. B., & Chilton, N. B. (1995). Differences in a ribosomal DNA sequence of *Strongylus* species allows identification of single eggs. *International Journal for Parasitology, 25*(3), 359-365.

Cantacessi, C., Mulvenna, J., Young, N. D., Kasny, M., Horak, P., Aziz, A., ... & Gasser, R. B.(2012).A deep exploration of the transcriptome and “excretory/secretory” proteome of adult *Fascioloides magna*. *Molecular & Cellular Proteomics, 11*(11), 1340-1353.

Carbon, S., Douglass, E., Good, B. M., Unni, D. R., Harris, N. L., Mungall, C. J., ... & Pedruzzi, I.(2021). The Gene Ontology resource: enriching a GOld mine. *Nucleic*

Acids Research, 49(D1), D325-D334.

Chai, J. Y., & Lee, S. H. (2002). Food-borne intestinal trematode infections in the Republic of Korea. *Parasitology international, 51*(2), 129-154.

Chai, J. Y., Hong, S. T., Lee, S. H., Lee, G. C., & Min, Y. I. (1994). A case of echinostomiasis with ulcerative lesions in the duodenum. *Korean Journal of Parasitology, 32*(3), 201-204.

Chai, J. Y., Sohn, W. M., Yong, T. S., Eom, K. S., Min, D. Y., Hoang, E. H., ... & Rim, H. J. (2012). Echinostome flukes receovered from humans in Khammouane Province, Lao PDR. *The Korean Journal of Parasitology, 50*(3), 269.

Changklungmoa, N., Kueakhai, P., Sangpairoj, K., Osotprasit, S., Chaiwichien, A., Samrit, T., ... & Chaithirayanon, K. (2020). A novel thioredoxin-related protein 14 from Fasciola gigantica has an immunodiagnostic potential for fasciolosis. *Acta Tropica, 207*, 105471.

Chen, C. M., Ye, Q. Z., Zhu, Z. M., Wanner, B. L., & Walsh, C. T. (1990). Molecular biology of carbon-phosphorus bond cleavage. Cloning and sequencing of the phn (psiD) genes involved in alkylphosphonate uptake and CP lyase activity in *Escherichia coli*. J*ournal of Biological Chemistry, 265*(8), 4461-4471.

Choi, J. H., Lee, J. H., Yu, H. S., Jeong, H. J., Kim, J., Hong, Y. C., ... & Chung, D. I. (2006). Molecular and biochemical characterization of hemoglobinase, a cysteine proteinase, in *Paragonimus westermani*. *The Korean Journal of Parasitology, 44*(3), 187.

Choi, M. H., Park, I. C., Li, S., & Hong, S. T. (2003). Excretory-secretory antigen is better than crude antigen for the serodiagnosis of clonorchiasis by ELISA. *The*

Korean Journal of Parasitology, 41(1), 35-39.

Choudhary, V., Garg, S., Chourasia, R., Hasnani, J. J., Patel, P. V., Shah, T. M., ... & Joshi, C. G. (2015). Transcriptome analysis of the adult rumen fluke *Paramphistomum cervi* following next generation sequencing. *Gene*, *570*(1), 64-70.

Chung., YB, Kong., Y, Yang., HJ, Kang .,SY& Cho., SY(1997;). Cysteine protease activities during maturation stages of *Paragonimus westermani*. *Journal of Parasitology, 83*(5),902–907.

Collins, F. H., & Paskewitz, S. M. (1996). A review of the use of ribosomal DNA (rDNA) to differentiate among cryptic *Anopheles* species. *Insect Molecular Biology, 5(*1), 1-9.

Constantine, C. C. (2003). Importance and pitfalls of molecular analysis to parasite epidemiology. *Trends in Parasitology, 19*(8), 346-348.

Cortés, A., Mikeš, L., Muñoz-Antolí, C., Álvarez-Izquierdo, M., Esteban, J. G., Horák, P., & Toledo, R. (2019). Secreted cathepsin L-like peptidases are involved in the degradation of trapped antibodies on the surface of *Echinostoma caproni*. *Parasitology Research, 118*(12), 3377-3386.

Cox, F. E. (2002). History of human parasitology. *Clinical Microbiology Reviews, 15*(4), 595-612.

Crellen, T., Walker, M., Lamberton, P.H.L., Kabatereine, N.B., Tukahebwa, E.M., Cotton, J.A. & Webster, J.P., (2016). Reduced efficacy of praziquantel against *Schistosoma mansoni* is associated with multiple rounds of mass drug administration. *Clinical Infectious Diseases. 63*, 1151–1159.

Crowther, G. J., Shanmugam, D., Carmona, S. J., Doyle, M. A., Hertz-Fowler, C., Berriman, M., ... &Agüero, F. (2010). Identification of attractive drug targets in neglected-disease pathogens using an in silico approach. *PLoS Neglected Tropical Diseases, 4*(8), e804.

Das, A., Bordoloi, G., Khuman, L., Borah, S. A. N. J. I. B., Chabukdhara, P., Boro, P., & Khargharia, S. (2017). Occurrence Of trematode Artyfechinostomosis in pig of Lakhimpur District of Assam. *International Journal of Agricultural Science Research, 11*, 53-58.

de Magalhaes, J. P., Finch, C. E., & Janssens, G. (2010). Next-generation sequencing in aging research: emerging applications, problems, pitfalls and possible solutions. *Ageing Research Reviews, 9*(3), 315-323.

Deodhar, N. S., Patil-Kulkarni, V. G., & Karyakarte, P. P. (1968). On the occurrence of a new species of genus *Artyfechinostomum* Lane, 1915 in a dog in Bombay. *Indian Journal of Helminthology, 19*(2), 122-125.

Dubey, J. P., Srivastava, H. O., & Sahasrabudhe, V. K. (1969). *Artyfechinostomum sufrartyfex* infection

in cat and dog in India. *Tropical and Geographical Medicine, 21*(2), 210.

El Meanawy, M. A., Aji, T., Phillips, N. F., Davis, R. E., Salata, R. A., Malhotra, I & Davis, A. H. (1990). Definition of the complete *Schistosoma mansoni* hemoglobinase mRNA sequence and gene expression in developing parasites. *The American Journal of Tropical Medicine and Hygiene, 43*(1), 67-78.

Elefant, G. R., Roldán, W. H., Seeböck, A., & Kosma, P. (2016). Evaluation of a di-O methylated glycan as a potential antigenic target for the serodiagnosis of human

toxocariasis. *Parasite Immunology, 38*(4), 236-243.

El- Manzalawy, Y., Dobbs, D., & Honavar, V. (2008). Predicting linear B- cell epitopes using string kernels. Journal of Molecular Recognition: *An Interdisciplinary Journal, 21*(4), 243-255.

Emanuelsson, O., Nielsen, H., Brunak, S., & Von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular biology, 300*(4), 1005-1016.

Esteban, J. G., Muñoz-Antoli, C., Toledo, R., & Ash, L. R. (2014). Diagnosis of human trematode infections. *Digenetic Trematodes,766* 293-327.

Fairweather, I., Brennan, G. P., Hanna, R., Robinson, M. W., & Skuce, P. J. (2020). Drug resistance in liver flukes. *International Journal for Parasitology Drugs and Drug Resistance, 12,* 39–59.

Fan, Y., Wang, X., Deng, C., Huang, Y., Wang, L., Chen, W., ... & Yu, X. (2011). Molecular cloning, expression, and immunolocalization of the NAD+-dependent glycerol 3-phosphate dehydrogenase (GPD) from *Clonorchis sinensis*. *Parasitology Research, 109*(3), 621-626.

Faust, E. C. 1929. *Human Helminthology, Philadelphia,* 616.

Felleisen, R. S. J. (1997). Comparative sequence analysis of 5.8S rRNA genes and internal transcribed spacer (ITS) regions of trichomonadid protozoa. *Parasitology, 115*(02), 111-119.

Fennell, B. J., Naughton, J. A., Barlow, J., Brennan, G., Fairweather, I., Hoey, E., ... & Bell, A. (2008). Microtubules as antiparasitic drug targets. *Expert Opinion on Drug Discovery, 3(*5), 501-518.

Fox, L. M., Furness, B. W., Haser, J. K., Desire, D., Brissau, J. M., Milord, M. D., ... & Beach, M. J.(2005). Tolerance and efficacy of combined diethylcarbamazine and albendazole for treatment of *Wuchereria bancrofti* and intestinal helminth infections in Haitian children. *The American Journal of Tropical Medicine and Hygiene*, *73*(1), 115-121.

Fried, B., & Abruzzi, A. (2010). Food-borne trematode infections of humans in the United States of America. *Parasitology Research*, *106*(6), 1263-1280.

Fried, B., & Reddy, A. (1997). Comparative studies on excystation and morphological features of the metacercariae of *Echinostoma paraensei* and *E. caproni*. *International Journal for Parasitology*, *27*(8), 899-890.

Fukuda, M., Otsuka, Y., Uni, S., Boda, T., Daisaku, H., Hasegawa, H., ... & Bain, O. (2011). Zoonotic onchocerciasis in Hiroshima, Japan, and molecular analysis of a paraffin section of the agent for a reliable identification. *Parasite*, *18*(2), 185-188.

Furtado, L.F.V., Medeiros, C.d.S., Zuccherato, L.W., Alves, W.P., de Oliveira, V.N.G.M.,da Silva, V.J., Miranda, G.S., Fujiwara, R.T., & Rabelo, E.M.L (2019). First identification of the benzimidazole resistance-associated F200Y SNP in the beta- tubulin gene in *Ascaris lumbricoides*. PLoS One 14, e0224108.

Gahoi, S., Singh, S., & Gautam, B. (2019). Genome-wide identification and comprehensive analysis of Excretory/Secretory proteins in nematodes provide potential drug targets for parasite control. *Genomics*, *111*(3), 297-309.

Garg, G. and Ranganathan, S. (2012). Helminth secretome database (HSD): a collection of helminth excretory/secretory proteins predicted from expressed sequence tags (ESTs). *BMC Genomics, 13* Suppl 7S8.

Garg, G., Bernal, D., Trelis, M., Forment, J., Ortiz, J., Valero, M. L., ... & Toledo, R. (2013). The transcriptome of *Echinostoma caproni* adults: further characterization of the secretome and identification of new potential drug targets. *Journal of Proteomics*, *89*, 202-214.

Gazzinelli, A., Correa-Oliveira, R., Yang, G. J., Boatin, B. A., & Kloos, H. (2012). A research agenda for helminth diseases of humans: social ecology, environmental determinants, and health systems. *PLoS Neglected Tropical Diseases, 6*(4), e1603.

Geerts, S., & Gryseels, B. (2000). Drug resistance in human helminths: current situation and lessons from livestock. *Clinical Microbiology Reviews, 13*(2), 207-222.

Geerts, S., & Gryseels, B. (2001). Anthelmintic resistance in human helminths: a review. *Tropical Medicine & International Health, 6*(11), 915-921.

Ghatani, S., Arya, L. K., & Tandon, V. (2018). Emerging trematode zoonoses in India. *Advances in Medico-Veterinary Parasitology: An Indian Perspective*, 340-354

Ghatani, S., Shylla, J. A., Roy, B., & Tandon, V. (2014). Multilocus sequence evaluation for differentiating species of the trematode Family Gastrothylacidae, with a note on the utility of mitochondrial COI motifs in species identification. *Gene*, *548*(2), 277-284.

Ghosh, I., Eisinger, S. W., Raghavan, N., & Scott, A. L. (1998). Thioredoxin peroxidases from Brugia malayi. *Molecular and Biochemical Parasitology, 91(*2), 207-220.

Gomase, V. S., Chitlange, N. R., Changbhale, S. S., & Kale, K. V. (2013). Prediction of *Brugia malayi* antigenic peptides: candidates for synthetic vaccine design against lymphatic filariasis. *Protein and Peptide Letters, 20*(8), 864-887.

Gonçalves, L. T., Bianchi, F. M., Deprá, M., & Calegaro-Marques, C. (2021). Barcoding

a can of worms: testing cox1 performance as a DNA barcode of Nematoda. *Genome, 99*(999), 1-13.

Gonenç, B., Sarimehmetoğlu, H. O., Kara, M., & Kircali, F. (2004). Comparison of crude and excretory/secretory antigens for the diagnosis of *Fasciola hepatica* in sheep by western blotting. *Turkish Journal of Veterinary and Animal Sciences*, *28*(5), 943-949.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... & Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644-652.

Graczyk, T. K., & Fried, B. (1998). Echinostomiasis: a common but forgotten food-borne disease. *The American Journal of Tropical Medicine and Hygiene, 58*(4), 501-504.

Hajibabaei, M., Ivanova, N. V., Ratnasingham, S., Dooh, R. T., Kirk, S. L., Mackie, P. M., & Hebert, P. D. (2005). Critical factors for assembling a high volume of DNA barcodes. *Philosophical Transactions of the Royal Society B: Biological Sciences, 360*(1462), 1959-1967.

Hall, T. A. (1999). BioEdit a user friendly biological sequence alignment editor and analysys program for Windows 95/98/NT. *In*: *Nucleic Acids Symposium Series No. 41*,95-98.

Hao, L. Y., Giasson, B. I., & Bonini, N. M. (2010). DJ-1 is critical for mitochondrial function and rescues PINK1 loss of function. *Proceedings of the National Academy of Sciences, 107(*21), 9747-9752.

Haseeb, M. A., & Eveland, L. K. (2000). Human echinostomiasis: mechanisms of pathogenesis and host resistance. I*n Echinostomes as Experimental Models for*

Biological Research, 83-98. Springer Netherlands.

Hebert, P. D., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003a). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1512), 313-321.

Hebert, P. D., Ratnasingham, S., & de Waard, J. R. (2003b). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London B: Biological Sciences, 270(Suppl 1)*, S96-S99.

Hebert, P. D., Stoeckle, M. Y., Zemlak, T. S., & Francis, C. M. (2004). Identification of birds through DNA barcodes. *PLoS Biology, 2*(10), e312.

Hill, D. E., & Sakanari, J. A. (1997). *Trichuris suis*: thiol protease activity from adult worms. *Experimental Parasitology, 85*(1), 55-62.

Hillis, D. M., & Davis, S. K. (1988). Ribosomal DNA: intraspecific polymorphism, concerted evolution, and phylogeny reconstruction. *Systematic Zoology, 37*(1), 63-66.

Hokke, C. H., Fitzpatrick, J. M., & Hoffmann, K. F. (2007). Integrating transcriptome, proteome and glycome analyses of *Schistosoma* biology. *Trends in Parasitology*, *23*(4), 165-174.

Hotez, P. J., & Damania, A. (2018). India's neglected tropical diseases. *PLoS Neglected Tropical Diseases, 12*(3), e0006038.

Hotez, P. J., Fenwick, A., Savioli, L., & Molyneux, D. H. (2009). Rescuing the bottom billion through control of neglected tropical diseases. *The Lancet, 373*(9674), 1570-1575.

Hotez, P. J., Molyneux, D. H., Fenwick, A., Kumaresan, J., Sachs, S. E., Sachs, J. D., & Savioli, L. (2007). *Control of neglected tropical diseases. New England journal of medicine, 357*(10), 1018-1027.

Hotez, P., Ottesen, E., Fenwick, A., & Molyneux, D. (2006). The neglected tropical diseases: the ancient afflictions of stigma and poverty and the prospects for their control and elimination. *Hot topics in infection and immunity in children III*, 23-33.

Huffman, J. E., & Fried, B. (1990). Echinostoma and echinostomiasis. *Advances in Parasitology, 29*, 215-269.

Hutchinson, G. W., Dawson, K., Fitzgibbon, C. C., & Martin, P. J. (2009). Efficacy of an injectable combination anthelmintic (nitroxynil+ clorsulon+ ivermectin) against early immature *Fasciola hepatica* compared to triclabendazole combination flukicides given orally or topically to cattle. *Veterinary Parasitology, 162*(3-4), 278-284.

Iddawela, R. D., Rajapakse, R. P. V. J., Perera, N. A. N. D., & Agatsuma, T. (2007). Characterization of a *Toxocara canis* species-specific excretory-secretory antigen (TcES-57) and development of a double sandwich ELISA for diagnosis of visceral larva migrans. *The Korean Journal of Parasitology, 45*(1), 19.

Irwin, J. J., & Shoichet, B. K. (2005). ZINC− a free database of commercially available compounds for virtual screening. *Journal of Chemical Information and Modeling, 45*(1), 177-182.

Jackowski, S., Jackson, P. D., & Rock, C. O. (1994). Sequence and function of the aas gene in *Escherichia coli*. J*ournal of Biological Chemistry, 269(*4), 2921-2928.

Jain, G. P. (1957). On the trematode—*Paryphostomum mehrai* Faruqui. *Proceedings of National Academy of Sciences India, 27*, 290-4.

Jain, G. P. (1960). Further observations on *Artyfechinostomum mehrai* (Faruqui)(Echinostomatidae: Trematoda). *Parasitology, 50*(1-2), 7-11.

James., ER, McLean., D, Perler., F.(1994). Molecular cloning of an *Onchocerca volvulus* extracellular Cu-Zn superoxide dismutase. *Infection and Immunity, 62*,713–716.

Jasmer, D. P., Rosa, B. A., & Mitreva, M. (2015). Peptidases compartmentalized to the *Ascaris suum* intestinal lumen and apical intestinal membrane. *PLoS Neglected Tropical Diseases, 9*(1), e3375.

Jayaram, B., Dhingra, P., Mishra, A., Kaushik, R., Mukherjee, G., Singh, A., & Shekhar, S. (2014). Bhageerath-H: a homology/ab initio hybrid server for predicting tertiary structures of monomeric soluble proteins. *BMC Bioinformatics, 15(*16), 1-12.

Jefferies, J. R., Campbell, A. M., Rossum, A. J. V., Barrett, J., & Brophy, P. M. (2001). Proteomic analysis of *Fasciola hepatica* excretory- secretory products. *PROTEOMICS: International Edition, 1*(9), 1128-1132.

Jespersen, M. C., Peters, B., Nielsen, M., & Marcatili, P. (2017). BepiPred-2.0: improving sequence based B-cell epitope prediction using conformational epitopes. *Nucleic acids research, 45*(W1), W24-W29.

Jimenez Castro, P. D., Howell, S. B., Schaefer, J. J., Avramenko, R. W., Gilleard, J. S., & Kaplan, R. M. (2019). Multiple drug resistance in the canine hookworm *Ancylostoma caninum*: an emerging threat?. *Parasites & vectors, 12*(1), 1-15.

Joe, L. K. (1963). Studies on echinostomatidae in Malaya. *Zeitschrift für Parasitenkunde,*

*23(*2), 124-135.

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., ... & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236-1240.

Jung, W. T., Lee, K. J., Kim, H. J., Kim, T. H., Na, B. K., & Sohn, W. M. (2014). A case of *Echinostoma cinetorchis* (Trematoda: Echinostomatidae) infection diagnosed by colonoscopy. *The Korean Journal of Parasitology*, *41*(3), 287.

Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., & Xu, J. (2012). Template-based protein structure modeling using the RaptorX web server. *Nature Protocols*, *7*(8), 1511-1522.

Kapp, K., Schrempf, S., Lemberg, M. K., & Dobberstein, B. (2009). Post-targeting functions of signal peptides. *Protein Transport Into the Endoplasmic Reticulum*, 1-16.

Keiser, J., & Utzinger, J. (2007b). Artemisinins and synthetic trioxolanes in the treatment of helminth infections. *Current opinion in infectious diseases, 20(*6), 605-612.

Keiser, J., & Utzinger, J. (2007c). Food-borne trematodiasis: current chemotherapy and advances with artemisinins and synthetic trioxolanes. *Trends in Parasitology, 23*(11), 555-562.

Keiser, J., & Utzinger, J. (2009). Food-borne trematodiases. *Clinical Microbiology Reviews*, *22*(3), 466-483.

Keiser, J., & Utzinger, J. (2010). The drugs we have and the drugs we need against major helminth infections. *Advances in Parasitology, 73*, 197-230.

Keiser, J., & Utzinger, J. (2012). Antimalarials in the treatment of schistosomiasis.

Current pharmaceutical design, *18*(24), 3531-3538.

Keiser, J., Chollet, J., Xiao, S. H., Mei, J. Y., Jiao, P. Y., Utzinger, J., & Tanner, M. (2009). Mefloquine—an aminoalcohol with promising antischistosomal properties in mice. *PLoS Neglected Tropical Diseases*, *3*(1), e350.

Keller, A., Schleicher, T., Schultz, J., Müller, T., Dandekar, T., & Wolf, M. (2009). 5.8 S-28S rRNA interaction and HMM-based ITS2 annotation. *Gene, 430*(1), 50-57.

Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. Nature reviews *Drug Discovery*, *3*(11), 935-949.

Kostadinova, & Jones, A .(2005). Superfamily Echinostomatoidea Looss, 1899. In: Jones A, Bray RA, Gibson DI (eds). *Keys to the Trematoda, Volume 2*. CABI, Wallingford and London, pp 5-8.

Kostadinova, A .(2005). Family Echinostomatidae. In: Jones A, Bray RA, Gibson DI (eds). *Keys to the Trematoda, Volume 2*. CABI, Wallingford and London, pp 9-64.

Kostadinova, A., Herniou, E. A., Barrett, J., & Littlewood, D. T. J. (2003). Phylogenetic relationships of Echinostoma Rudolphi, 1809 (Digenea: Echinostomatidae) and related genera re-assessed via DNA and morphological analyses. *Systematic Parasitology, 54*(3), 159-176.

Kozlova, E. E. G., Cerf, L., Schneider, F. S., Viart, B. T., NGuyen, C., Steiner, B. T., ... & Machado-de-Ávila, R. A. (2018). Computational B-cell epitope identification and production of neutralizing murine antibodies against Atroxlysin-I. *Scientific reports*, *8*(1), 1-13.

Kraft, R. (2015). STIM and ORAI proteins in the nervous system. *Channels, 9*(5), 245-

Kringelum, J. V., Lundegaard, C., Lund, O., & Nielsen, M. (2012). Reliable B cell epitope predictions: impacts of method development and improved benchmarking. *PLoS Computational Biology, 8*(12), e1002829.

Kroemer, R. T. (2007). Structure-based drug design: docking and scoring. *Current Protein and Peptide science, 8*(4), 312-328.

Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology, 305*(3), 567-580.

Krücken, J., Fraundorfer, K., Mugisha, J.C., Ramünke, S., Sifft, K.C., Geus, D., Habarugira, F., Ndoli, J., Sendegeya, A., Mukampunga, C., Bayingana, C., Aebischer, T., Demeler, J., Gahutu, J.B., Mockenhaupt, F.P., & von Samson-Himmelstjerna, G., (2017). Reduced efficacy of albendazole against *Ascaris lumbricoides* in Rwandan school children. *International Journal for Parasitology; Drugs and Drug Resistance*. 7, 262–271.

Kushwaha, V., Tewari, P., Mandal, P., Tripathi, A., & Murthy, P. K. (2019). Troponin 1 of human filarial parasite *Brugia malayi*: cDNA cloning, expression, purification, and its immunoprophylactic potential. *Parasitology Research, 118*(6), 1849-1863.

Kwatia, M. A., Botkin, D. J., & Williams, D. L. (2000). Molecular and enzymatic characterization of *Schistosoma mansoni* thioredoxin peroxidase. *Journal of Parasitology, 86*(5), 908-915.

Lakhal-Naouar, I., Jardim, A., Strasser, R., Luo, S., Kozakai, Y., Nakhasi, H. L., & Duncan, R. C. (2012). *Leishmania donovani* argininosuccinate synthase is an

active enzyme associated with parasite pathogenesis. *PLoS Neglected Tropical Diseases, 6*(10), e1849.

Lammie, P. J., Fenwick, A., & Utzinger, J. (2006). A blueprint for success: integration of neglected tropical disease control programmes. *Trends in parasitology, 22*(7), 313-321.

Lane, C. (1915). *Artyfechinostomum sufrartyfex*. A New Parasitic Echinostome of Man. *Indian Journal of Medical Research, 2*(4), 977-983.

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology, 10*(3), 1-10.

Le, T. H., Blair, D., & McManus, D. P. (2000). Mitochondrial genomes of human helminths and their use as markers in population genetics and phylogeny. *Acta Tropica, 77*(3), 243-256.

Le, T. H., Nguyen, K. T., Nguyen, N. T. B., Doan, H. T. T., Dung, D. T., & Blair, D. (2017). The ribosomal transcription units of *Haplorchis pumilio* and *H. taichui* and the use of 28S rDNA sequences for phylogenetic identification of common heterophyids in Vietnam. *Parasites & Vectors*, *10*(1), 1-9.

Leiper, R. T. (1911). A new echinostome parasite in man. *Journal of London School of Hygiene & Tropical Medicine, 1*(1), 27-28.

León, N., Padilla, C., Pajuelo, M., Sheen, P., & Zimic, M. (2013). Cathepsin L cysteine protease from *Taenia solium*: its biological role in the infection and potential use for the immunodiagnosis of neurocysticercosis. *Revista Peruana de Medicina Experimental y Salud Publica, 30*(3), 446-454.

Li, B. W., McNulty, S. N., Rosa, B. A., Tyagi, R., Zeng, Q. R., Gu, K. Z., ... & Mitreva, M. (2016). Conservation and diversification of the transcriptomes of adult *Paragonimus westermani* and *P. skrjabini*. *Parasites & vectors*, *9*(1), 1-11.

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*(1), 1-16.

Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22*(13), 1658-1659.

Li., AH, Moon., SU, Park., YK, Na BK, Hwang., MG, Oh., CM, (2006).Identification and characterization of a cathepsin L-like cysteine protease from *Taenia solium* metacestode. *Veterinary Parasitology,141*(3),251–259.

Liang, J., Shaulov, Y., Savage-Dunn, C., Boissinot, S., & Hoque, T. (2017). Chloride intracellular channel proteins respond to heat stress in *Caenorhabditis elegans*. *PloS One*, *12*(9), e0184308.

Liddell, S., & Knox, D. P. (1998). Extracellular and cytoplasmic Cu/Zn superoxide dismutases from *Haemonchus contortus*. *Parasitology*, *116*(4), 383-394.

Liu, Y., Yang, Y., Xu, J., Dong, X., Gu, X., Xie, Y., ... & Yang, G. (2019). Expression and serodiagnostic potential of antigen B and thioredoxin peroxidase from *Taenia multiceps*. *Veterinary Parasitology, 272*, 58-63.

Locke, S. A., McLaughlin, J. D., Dayanandan, S., & Marcogliese, D. J. (2010). Diversity and specificity in *Diplostomum* spp. metacercariae in freshwater fishes revealed by cytochrome c oxidase I and internal transcribed spacer sequences. *International Journal for Parasitology*, *40*(3), 333-343.

Macalanda, A. M. C., Angeles, J. M. M., Moendeg, K. J., Dang, A. T., Higuchi, L., Inoue,

N., ... & Kawazu, S. I. (2017). Evaluation of *Schistosoma japonicum* thioredoxin peroxidase-1 as a potential circulating antigen target for the diagnosis of Asian schistosomiasis. *Journal of Veterinary Medical Science*, 17-0579

Madsen, U. Krogsgaard-Larsen, P, Liljefors, T. (2002). *Textbook of Drug Design and Discovery*. Washington, DC: Taylor & Francis.

Maizels, R. M., Bundy, D. A., Selkirk, M. E., Smith, D. F., & Anderson, R. M. (1993). Immunological modulation and evasion by helminth parasites in human populations. *Nature, 365*(6449), 797-805.

Maji, A. K., Bera, D. K., Manna, B., Nandy, A., Addy, M., & Bandyopadhyay, A. K. (1993). First record of human infection with *Echinostoma malayanum* in India. *Transactions of the Royal Society of Tropical Medicine and Hygiene, 87*(6), 673-673.

Mani, T. R., Rajendran, R., Munirathinam, A., Sunish, I. P., Md. Abdullah, S., Augustin, D. J., & Satyanarayana, K.(2002).Efficacy of co-administration of albendazole and diethylcarbamazine against geohelminthiases: a study from South India. *Tropical Medicine &International Health, 7*(6), 541-548.

Marcilla, A. (2009). Echinostomes: genomics and proteomics. *In The biology of Echinostomes* (pp. 207-228). Springer, New York, NY.

Margutti, P., Ortona, E., Delunardo, F., Tagliani, A., Profumo, E., Riganò, R., ... & Siracusano, A. (2008). Thioredoxin peroxidase from *Echinococcus granulosus*: a candidate to extend the antigenic panel for the immunodiagnosis of human cystic echinococcosis. *Diagnostic Microbiology and Infectious Disease, 60*(3), 279-285.

McKerrow, J. H., Caffrey, C., Kelly, B., Loke, P. N., & Sajid, M. (2006). Proteases in

parasitic diseases. *Annual Review of Pathology: Mechanisms of Disease ,1,* 497-536.

McManus, D. P., Le, T. H., & Blair, D. (2000). Mitochondrial genomes in the parasitic flatworms: form and function. *Current Topics in Biochemical Research, 3,* 27-40.

Miles, S., Navatta, M., Dematteis, S., & Mourglia-Ettlin, G. (2017). Identification of universal diagnostic peptide candidates for neglected tropical diseases caused by cestodes through the integration of multi-genome-wide analyses and immunoinformatic predictions. *Infection, Genetics and Evolution, 54,* 338-346.

Miller, B. R., Crabtree, M. B., & Savage, H. M. (1996). Phylogeny of fourteen *Culex* mosquito species, including the *Culex pipiens* complex, inferred from the internal ranscribed spacers of ribosomal DNA. *Insect Molecular Biology, 5*(2), 93-107.

Mohandas, A. (1971). *Artyfechinostomum sufrartyfex* Lane 1915, a synonym of *Echinostoma malayanum* Leiper, 1911 (Trematoda: Echinostomatidae). *Acta Parasitologica Polonica*, *19*, 361-368.

Morgan, J. A. T., & Blair, D. (1998). Relative merits of nuclear ribosomal internal transcribed spacers and mitochondrial CO1 and ND1 genes for distinguishing among *Echinostoma* species (Trematoda). *Parasitology, 116*(03), 289-297.

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 andAutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry, 30*(16), 2785-2791.

Moszczynska, A., Locke, S. A., Mclaughlin, J. D., Marcogliese, D. J., & Crease, T. J. (2009). Development of primers for the mitochondrial cytochrome c oxidase I

gene in digenetic trematodes (Platyhelminthes) illustrates the challenge of barcoding parasitic helminths. *Molecular Ecology Resources*, *9*(s1), 75-82.

Mukabayire, O., Boccolini, D., Lochouarn, L., Fontenille, D., & Besansky, N. J. (1999). Mitochondrial and ribosomal internal transcribed spacer (ITS2) diversity of the African malaria vector *Anopheles funestus*. *Molecular Ecology, 8*(2), 289-297.

Mukherjee, G., & Jayaram, B. (2013). A rapid identification of hit molecules for target proteins via physico-chemical descriptors. *Physical Chemistry Chemical Physics, 15*(23), 9107-9116.

Mukherjee, R. P., & Ghosh, R. K. (1968, July). On the synonymy of the genus *Artyfechinostomum lane,* 1915 (Trematoda: Echinostomatidae). *In Proceedings of the Indian Academy of Sciences-Section B,68*(1),*52-58.*

Na., BK, Kim., SH, Lee., EG, Kim., TS, Bae., YA,& Kang I.(2006). Critical roles for excretory-secretory cysteine proteases during tissue invasion of *Paragonimus westermani* newly excysted metacercariae. *Cell Microbiology ,8*(6),1034–1046.

Ndyomugyenyi, R., Kabatereine, N., Olsen, A., Magnussen, P., (2008). Efficacy of ivermectin and albendazole alone and in combination for treatment of soil-transmitted helminths in pregnancy and adverse events: a randomized open label controlled intervention trial in Masindi district, western Uganda. *The American Journal of Tropical Medicine and Hygiene,79*, 856–863.

Ngo, H. D., Ermolenko, A. V., & Gerasev, P. I. (2011). *Artyfechinostomum sufratyfex* Lane, 1915 (Trematoda: Echinostomatidae), a parasite of mammals in Southern Vietnam. *Parazitologiia, 45*(1), 54-59.

Nolan, M. J., & Cribb, T. H. (2005). The use and implications of ribosomal DNA

sequencing for the discrimination of digenean species. *Advances in parasitology, 60*, 101-163.

Nyman, T., Papadopoulou, E., Ylinen, E., Wutke, S., Michell, C. T., Sromek, L., ... & Kunnasranta, M. (2021). DNA barcoding reveals different cestode helminth species in northern European marine and freshwater ringed seals. *International Journal for Parasitology: Parasites and Wildlife, 15*, 255-261.

Okamoto, M., Urushima, H., Iwasa, M., & Hasegawa, H. (2007). Phylogenetic relationships of rodent pinworms (genus *Syphacia*) in Japan inferred from mitochondrial CO1 gene sequences. *Journal of Veterinary Medical Science, 69*(5), 545-547.

Olds, G. R., King, C., Hewlett, J., Olveda, R. W. U. G., Wu, G., Ouma, J., ... & Reeve, P. (1999). Double-blind placebo-controlled study of concurrent administration of albendazole and praziquantel in schoolchildren with schistosomiasis and geohelminths. *The Journal of Infectious Diseases, 179*(4), 996-1003.

Ondrejicka, D. A., Locke, S. A., Morey, K., Borisenko, A. V., & Hanner, R. H. (2014). Status and prospects of DNA barcoding in medically important parasites and vectors. *Trends in Parasitology, 30*(12), 582-591.

Osazuwa, F., Ayo, O. M., & Imade, P. (2011). A significant association between intestinal helminth infection and anaemia burden in children in rural communities of Edo state, Nigeria. *North American Journal Of Medical Sciences, 3*(1), 30.

Osei-Atweneboana, M.Y., Awadzi, K., Attah, S.K., Boakye, D.A., Gyapong, J.O., & Prichard, R.K., (2011). Phenotypic evidence of emerging ivermectin resistance in *Onchocerca volvulus*. *PLoS Neglected Tropical Diseases. 5*, e998.

Panic, G., Duthaler, U., Speich, B., & Keiser, J. (2014). Repurposing drugs for the treatment and control of helminth infections. *International Journal for Parasitology: Drugs and Drug Resistance, 4*(3), 185-200.

Park, J., Kim, S. Y., Cha, G. H., Lee, S. B., Kim, S., & Chung, J. (2005). Drosophila DJ-1 mutants show oxidative stress-sensitive locomotive dysfunction. *Gene, 361*, 133-139.

Parker, J. A., Metzler, M., Georgiou, J., Mage, M., Roder, J. C., Rose, A. M., ... & Néri, C. (2007). Huntingtin-interacting protein 1 influences worm and mouse presynaptic function and protects *Caenorhabditis elegans* neurons against mutant polyglutamine toxicity. *Journal of Neuroscience, 27*(41), 11056-11064.

Patel, R. K., & Jain, M. (2012). NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PloS One*, *7*(2), e30619.

Pearson, M. S., Ranjit, N., & Loukas, A. (2010). Blunting the knife: development of vaccines targeting digestive proteases of blood-feeding helminth parasites. *De Gruyter,* 901-911.

Petersen, T. N., Brunak, S., Von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods, 8(*10), 785-786.

Pham, L. T. K., Saijuntha, W., Lawton, S. P., & Le, T. H. (2022). Mitophylogenomics of the zoonotic fluke *Echinostoma malayanum* confirms it as a member of the genus Artyfechinostomum Lane, 1915 and illustrates the complexity of Echinostomatidae systematics. *Parasitology Research*, *121*(3), 899-913.

Phuphisut, O., Ajawatanawong, P., Limpanont, Y., Reamtong, O., Nuamtanong, S.,

Ampawong, S., ... & Adisakwattana, P. (2018). Transcriptomic analysis of male and female *Schistosoma mekongi* adult worms. *Parasites & vectors*, *11*(1), 1-16.

Pomaznoy, M. Y., Logacheva, M. D., Young, N. D., Penin, A. A., Ershov, N. I., Katokhin, A. V., & Mordvinov, V. A. (2016). Whole transcriptome profiling of adult and infective stages of the trematode *Opisthorchis felineus*. *Parasitology International*, *65*(1), 12-19.

Potocnakova, L., Bhide, M., & Pulzova, L. B. (2016). An introduction to B-cell epitope mapping and in silico epitope prediction. *Journal of Immunology Research*,e6730830.

Prasad, P., Tandon, V., & Chatterjee, A. (2010). PCR-based molecular characterization and analysis of food-borne trematode parasites *Paragonimus westermani*, *Fasciolopsis buski* and *Fasciola gigantica* from Northeast India using ITS2 rDNA region. *New Biotechnology, 27*, S27-S28.

Prasad, Y. K., Dahal, S., Saikia, B., Bordoloi, B., Tandon, V., & Ghatani, S. (2019). *Artyfechinostomum sufrartyfex* trematode infections in children, Bihar, India. *Emerging Infectious Diseases, 25*(8), 1571.

Preidis, G. A., & Hotez, P. J. (2015). The newest "omics"—metagenomics and metabolomics—enter the battle against the neglected tropical diseases. *PLoS Neglected Tropical Diseases, 9*(2), e0003382.

Premvati, G., & Pande, V. (1974). On *Artifechinostomum malayanum* (Leiper, 1911) Mendheim, 1943 (Trematoda: Echinostomatidae) with synonymy of allied species and genera. *Proceedings of Helminthological Society of Washington*, *41*(41), 151-160.

Radomyos, P., Bunnag, D., & Harinasuta, T. (1982). *Ecbinostoma ilocanum* (Garrison, 1908) Odhner, 1911, infection in man in Thailand. *Southeast Asian Journal of Tropical Medicine and Public Health, 13*(2), 265-269.

Rai, P., & Ahluwalia, S. S. (1958). A note on common helminths of pigs at Aligarh (Part I). *Proceedings of the National Academy of Sciences of India, Section B. Biological Sciences, 28*(2), 195-199.

Ranjith, D., & Surjith, K. P. (2020). In silico anthelminthic activity of a novel herbal formulation. *Journal of Pharmacognosy and Phytochemistry, 9*(1), 1260-1266.

Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The Barcode of Life Data System (http://www. barcodinglife. org). *Molecular Ecology Notes, 7*(3), 355-364.

Reddy, D. G., & Varmah, K. (1950). *Paryphostomum sufrartyfex* (intestinal fluke) infection in man. *The Indian medical gazette, 85*(12), 546.

Rhoads., ML, Fetterer., RH. (1995) Developmentally regulated secretion of cathepsin L-like cysteine proteases by *Haemonchus contortus*. *Journal of Parasitology, 81*(4),505–512.

Rim, H. J. (1982). Echinostomiasis. Parasitic zoonoses, *Volume 3/the editor [s], Leon Jacobs...[et al.]*.

Robinson, M. W., Dalton, J. P., & Donnelly, S. (2008). Helminth pathogen cathepsin proteases: it's a family affair. T*rends in Biochemical Sciences, 33*(12), 601-608.

Robinson, M. W., Menon, R., Donnelly, S. M., Dalton, J. P., & Ranganathan, S. (2009). An integrated transcriptomics and proteomics analysis of the secretome of the helminth pathogen *Fasciola hepatica*: proteins associated with invasion and infection of the mammalian host. *Molecular and Cellular Proteomics, 8*(8), 1891-

1907.

Roth, J. R., Lawrence, J. G., Rubenfield, M., Kieffer-Higgins, S., & Church, G. M. (1993). Characterization of the cobalamin (vitamin B12) biosynthetic genes of *Salmonella typhimurium*. *Journal of Bacteriology, 175*(11), 3303-3316.).

Roy, B., & Tandon, V. (1992). Seasonal prevalence of some zoonotic trematode infections in cattle and pigs in the north-east montane zone in India. *Veterinary Parasitology*, 41(1-2), 69-76.

Roy, B., & Tandon, V. (1996). Effect of root-tuber extract of *Flemingia vestita*, a leguminous plant, on *Artyfechinostomum sufrartyfex* and *Fasciolopsis buski*: a scanning electron microscopy study. *Parasitology Research, 82*(3), 248-252.

Saijuntha, W., Sithithaworn, P., Wongkham, S., Laha, T., Chilton, N. B., Petney, T. N., ... & Andrews, R. H. (2008). Mitochondrial DNA sequence variation among geographical isolates of *Opisthorchis viverrini* in Thailand and Lao PDR, and phylogenetic relationships with other trematodes. *Parasitology, 135*(12), 1479-1486.

Saijuntha, W., Tantrawatpan, C., Sithithaworn, P., Andrews, R. H., & Petney, T. N. (2011). Genetic characterization of *Echinostoma revolutum* and *Echinoparyphium recurvatum* (Trematoda: Echinostomatidae) in Thailand and phylogenetic relationships with other isolates inferred by ITS1 sequence. *Parasitology Research, 108*(3), 751-755.

Saliba, K. J., Martin, R. E., Bröer, A., Henry, R. I., McCarthy, C. S., Downie, M. J., ... & Kirk, K. (2006). Sodium-dependent uptake of inorganic phosphate by the intracellular malaria parasite. *Nature, 443*(7111), 582-585.

Sambrook, J., & Russell, D. W. (2006). Purification of nucleic acids by extraction with phenol: chloroform. *Cold Spring Harbor Protocols, 2006*(1), pdb-prot4455.

Sánchez-Moreno, M., Garcia-Ruiz, M. A., Garcia-Rejon, L., Valero, A., & León, P. (1989). Superoxide dismutase in cestodes. Isoenzymatic characterization and studies of inhibition by a series of benzimidazoles and by pyrimidine derivatives of recent synthesis. *Arzneimittel-forschung*, *39*(7), 759-761.

Sanchez-Moreno, M., Leon, P., Salas-Peregrin, J. M., & Osuna, A. (1987). Superoxide dismutase in trematodes. *Arzneimittelforschung*, *37*, 903-905.

Sardar, A. H., Jardim, A., Ghosh, A. K., Mandal, A., Das, S., Saini, S & Das, P. (2016). Genetic manipulation of *Leishmania donovani* to explore the involvement of argininosuccinate synthase in oxidative stress management. *PLoS Neglected Tropical Diseases*, *10*(3), e0004308.

Savioli, L., Bundy, D., & Tomkins, A. (1992). Intestinal parasitic infections: a soluble public health problem. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *86*(4), 353-354.

Saxton, T. M., Fried, B., & Peoples, R. C. (2008). Excystation of the encysted metacercariae of *Echinostoma trivolvis* and *Echinostoma caproni* in a trypsin-bile salts-cysteine medium and morphometric analysis of the excysted larvae. *Journal of Parasitology, 94(*3), 669-671.

Scholz, T., Škeříková, A., Shimazu, T., & Grygier, M. J. (2004). A taxonomic study of species of *Bothriocephalus Rudolphi*, 1808 (Cestoda: Pseudophyllidea) from eels in Japan: morphological and molecular evidence for the occurrence of *B. claviceps* (Goeze, 1782) and confirmation of the validity of *B. japonicus*

Yamaguti, 1934. *Systematic Parasitology,* 57(2), 87-96.

Shompole, S., & Jasmer, D. P. (2001). Cathepsin B-like cysteine proteases confer intestinal cysteine protease activity in *Haemonchus contortus*. *Journal of Biological Chemistry, 276*(4), 2928-2934.

Singh, A., Kaushik, R., Mishra, A., Shanker, A., & Jayaram, B. (2016). ProTSAV: a protein tertiary structure analysis and validation server. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 1864*(1), 11-19.

Singh, T., Biswas, D., & Jayaram, B. (2011). AADS-An automated active site identification, docking, and scoring protocol for protein targets based on physicochemical descriptors. *Journal of Chemical Information and Modeling, 51*(10), 2515-2527.

Sohn, W. M., Yong, T. S., Eom, K. S., Sinuon, M., Jeoung, H. G., & Chai, J. Y. (2017). *Artyfechinostomum malayanum*: metacercariae encysted in *Pila* sp. snails purchased from Phnom Penh, Cambodia. *The Korean Journal of Parasitology*, *55*(3), 341.

Srivastava, S. C. (1964). A new host record for *Strongyloides akbari* and *Artyfechinostomum sufrartyfex*. *Indian Journal of Helminthology, 16*(1), 24-26.

Stevenson, C.R., Mahoney, R.H., Fisara, P., Strehlau, G., Reichel, M.P., (2002). The efficacy of formulations of triclabendazole and ivermectin in combination against liver fluke (*Fasciola hepatica*) and gastro-intestinal nematodes in cattle and sheep and sucking lice species in cattle.*The Australian Veterinary Journal, 80*, 698–701.

Stevenson, L. A., Chilton, N. B., & Gasser, R. B. (1995). Differentiation of *Haemonchus placei* from *H. contortus* (Nematoda: Trichostrongylidae) by the ribosomal DNA

second internal transcribed spacer. *International Journal for Parasitology, 25*(4), 483-488.

Sugita, T., Nishikawa, A., & Shinoda, T. (1998). Identification of *Trichosporon asahii* by PCR based on sequences of the internal transcribed spacer regions. *Journal of Clinical Microbiology, 36*(9), 2742-2744.

Sun, G. G., Liu, R. D., Wang, Z. Q., Jiang, P., Wang, L., Liu, X. L., ... & Cui, J. (2015). New diagnostic antigens for early trichinellosis: the excretory-secretory antigens of *Trichinella spiralis* intestinal infective larvae. *Parasitology Research*, *114*(12), 4637-4644.

Suttiprapa, S., Loukas, A., Laha, T., Wongkham, S., Kaewkes, S., Gaze, S., ... & Sripa, B. (2008). Characterization of the antioxidant enzyme, thioredoxin peroxidase, from the carcinogenic human liver fluke, *Opisthorchis viverrini*. *Molecular and Biochemical Parasitology, 160*(2), 116-122.

Taiwo, F. A., Brophy, P. M., Pritchard, D. I., Brown, A., Wardlaw, A., & Patterson, L. H. (1999). Cu/Zn superoxide dismutase in excretory–secretory products of the human hookworm *Necator americanus*: An electron paramagnetic spectrometry study. *European Journal of Biochemistry, 264*(2), 434-438.

Tandon, V., Goswami, L. M., Prasad, P. K., & Chatterjee, A. (2008, November). Molecular characterization of intestinal fluke, *Artyfechinostomum sufrartyfex* (Trematoda: Echinostomatidae) using PCR amplification of rDNA ITS regions. *In "Current Trends in Parasitology, Proceedings of the 20th National Congress of Parasitology, Shillong, India Eds.*,p153.

Tandon, V., Shylla, J. A., Ghatani, S., Athokpam, V. D., & Sahu, R. (2015). Neglected

tropical diseases: Trematodiases—the Indian scenario. *Proceedings of the National Academy of Sciences, India Section B: Biological Sciences, 85*(4), 901-907.

Tang, L. I. A. N. G., Ou, X., Henkle-Dührsen, K., & Selkirk, M. E. (1994). Extracellular and cytoplasmic CuZn superoxide dismutases from Brugia lymphatic filarial nematode parasites. *Infection and Immunity, 62*(3), 961-967.

Tantrawatpan, C., Saijuntha, W., Sithithaworn, P., Andrews, R. H., & Petney, T. N. (2013). Genetic differentiation of *Artyfechinostomum malayanum* and *A. sufrartyfex* (Trematoda: Echinostomatidae) based on internal transcribed spacer sequences. *Parasitology Research, 112*(1), 437-441.

Tarique, M., Chauhan, M., & Tuteja, R. (2017). ATPase activity of *Plasmodium falciparum* MLH is inhibited by DNA-interacting ligands and dsRNAs of MLH along with UvrD curtail malaria parasite growth. *Protoplasma, 254*(3), 1295-1305.

Taylor, L. H., Latham, S. M., & Woolhouse, M. E. (2001). Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 356*(1411), 983-989.

Thomas, C. M., & Timson, D. J. (2020). The mechanism of action of praziquantel: can new drugs exploit similar mechanisms? *Current Medicinal Chemistry, 27*(5), 676-696.

Thompson, R. A., & McManus, D. P. (2002). Towards a taxonomic revision of the genus *Echinococcus*. *Trends in Parasitology, 18*(10), 452-457.

Tjalsma, H., Bolhuis, A., Jongbloed, J. D., Bron, S. & van Dijl, J. M. (2000). Signal

peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. *Microbiology and Molecular Biology Reviews, 64*, 515-547.

Tkach, V., Pawlowski, J., & Mariaux, J. (2000). Phylogenetic analysis of the suborder Plagiorchiata (Platyhelminthes, Digenea) based on partial lsrDNA sequences. *International Journal for Parasitology, 30*(1), 83-93.

Toledo, R., & Esteban, J. G. (2016). An update on human echinostomiasis. *Transactions of the Royal Society of Tropical Medicine and Hygiene, 110*(1), 37-45.

Toledo, R., Esteban, J. G., & Fried, B. (2012). Current status of food-borne trematode infections. *European Journal of Clinical Microbiology & Iinfectious Diseases, 31*(8), 1705-1718.

Utzinger, J., & Keiser, J. (2004). Schistosomiasis and soil-transmitted helminthiasis: common drugs for treatment and control. *Expert Opinion on Pharmacotherapy, 5*(2), 263-285.

Van Steenkiste, N., Locke, S. A., Castelin, M., Marcogliese, D. J., & Abbott, C. L. (2015). New primers for DNA barcoding of digeneans and cestodes (Platyhelminthes). *Molecular Ecology Resources*, *15*(4), 945-952.

Verissimo, C. D. M., Potriquet, J., You, H., McManus, D. P., Mulvenna, J., & Jones, M. K. (2019). Qualitative and quantitative proteomic analyses of *Schistosoma japonicum* eggs and egg-derived secretory-excretory proteins. *Parasites & Vectors*, *12*(1), 173.

Verjovski-Almeida, S., DeMarco, R., Martins, E. A., Guimarães, P. E., Ojopi, E. P., Paquola, A. C., ... & Ashton, P. D. (2003). Transcriptome analysis of the acoelomate human parasite *Schistosoma mansoni*. *Nature Genetics*, *35*(2), 148-

157.

Wadood, A., Ahmed, N., Shah, L., Ahmad, A., Hassan, H., & Shams, S. (2013). *In silico* drug design: An approach which revolutionarised the drug discovery process. *OA Drug Design & Delivery, 1*(1), 3.

Wan, C., Tempel, W., Liu, Z. J., Wang, B. C., & Rose, R. B. (2005). Structure of the conserved transcriptional repressor enhancer of rudimentary homolog. *Biochemistry, 44*(13), 5017-5023.

Wang, L., Cui, J., Hu, D. D., Liu, R. D., & Wang, Z. Q. (2014). Identification of early diagnostic antigens from major excretory-secretory proteins of *Trichinella spiralis* muscle larvae using immunoproteomics. *Parasites & Vectors, 7*(1), 40.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics, 10*(1), 57-63.

Wang., Q, Zhang., S, Luo., X, Hou., J, Zhu., X& Cai X.(2013). Cloning and characterization of a cathepsin L-like cysteine protease from *Taenia pisiformis*. *Veterinary Parasitology,194*(1),26–34.

Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., & Hebert, P. D. (2005). DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 360*(1462), 1847-1857.

Webster, B. L., Culverwell, C. L., Khamis, I. S., Mohammed, K. A., Rollinson, D., & Stothard, J. R. (2013). DNA barcoding of *Schistosoma haematobium* on Zanzibar reveals substantial genetic diversity and two major phylogenetic groups. *Acta Tropica, 128*(2), 206-217.

White, T. J., Bruns, T., Lee, S. J. W. T., & Taylor, J. W. (1990). Amplification and direct

sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols: a guide to methods and applications*, *18*(1), 315-322

Wilson, M. J., Wade, W. G., & Welghtman, A. J. (1995). Restriction fragment length polymorphism analysis of PCR amplified 16S ribosomal DNA of human Capnocytophaga. *Journal of applied bacteriology*, *78*(4), 394-401.

Wolstenholme, A. J., Evans, C. C., Jimenez, P. D., & Moorhead, A. R. (2015). The emergence of macrocyclic lactone resistance in the canine heartworm, *Dirofilaria immitis*. *Parasitology*, *142*(10), 1249-1259.

Wörheide, G., Hooper, J. N., & Degnan, B. M. (2002). Phylogeography of western Pacific *Leucetta* '*chagosensis'*(Porifera: Calcarea) from ribosomal DNA sequences: implications for population history and conservation of the Great Barrier Reef World Heritage Area (Australia). *Molecular Ecology*, *11*(9), 1753-1768.

World Health Organization. (2006). *Preventive chemotherapy in human helminthiasis. Coordinated use of anthelminthic drugs in control interventions: a manual for health professionals and programme managers*. World Health Organization.

Xiao, S.-H., Utzinger, J., Tanner, M., Keiser, J., &Xue, J. (2013). Advances with the Chinese anthelminthic drug tribendimidine in clinical trials and laboratory investigations. *Acta Tropica*, *126*(2), 115–126.

Yadav, M., Singh, A., Rathaur, S., & Liebau, E. (2010). Structural modeling and simulation studies of *Brugia malayi* glutathione-S-transferase with compounds exhibiting antifilarial activity: implications in drug targeting and designing. *Journal of Molecular Graphics and Modelling*, *28*(5), 435-445.

Yang, D., Fu, Y., Wu, X., Xie, Y., Nie, H., Chen, L., ... & Yan, N. (2012). Annotation of

the transcriptome from Taenia pisiformis and its comparative analysis with three Taeniidae species. *PLoS One, 7*(4), e32283.

Yatawara, L., Wickramasinghe, S., Nagataki, M., Rajapakse, R. P. V. J., & Agatsuma, T. (2007). Molecular characterization and phylogenetic analysis of *Setaria digitata* of Sri Lanka based on CO1 and12S rDNA genes. *Veterinary Parasitology, 148*(2), 161-165.

Young, N. D., Hall, R. S., Jex, A. R., Cantacessi, C., & Gasser, R. B. (2010a). Elucidating the transcriptome of *Fasciola hepatica*—a key to fundamental and biotechnological discoveries for a neglected parasite. *Biotechnology Advances*, *28*(2), 222-231

Young, N. D., Campbell, B. E., Hall, R. S., Jex, A. R., Cantacessi, C., Laha, T., ... & Gasser, R. B(2010)b. Unlocking the transcriptomes of two carcinogenic parasites, *Clonorchis sinensis* and *Opisthorchis viverrini*. *PLoS Neglected Tropical Diseases, 4*(6), e719.

Young, N. D., Jex, A. R., Cantacessi, C., Hall, R. S., Campbell, B. E., Spithill, T. W., ... & Gasser, R. B. (2011). A portrait of the transcriptome of the neglected trematode, *Fasciola gigantica*—biological and biotechnological implications. *PLoS Neglected Tropical Diseases, 5*(2), e1004.

Yu, S.H., & Mott, K.E. (1994). Epidemiology and morbidity of food-borne intestinal trematode infections. *Tropical Disease Bulletin, 91*(7),125-152.

Zhang, G., Chen, J., Yang, Y., Liu, N., Jiang, W., Gu, S., ... & Wang, Z. (2014). Utility of DNA barcoding in distinguishing species of the family Taeniidae. *Journal of Parasitology*, *100*(4), 542-546.

Zhang, W., Rogniaux, H., Huang, W., Chauvin, A., & Moreau, E. (2011). Analysis of thioredoxin peroxidase as a promising antigen for diagnosis of *Fasciola gigantica* infection: a preliminary study. *Parasitology International*, *60*(2), 206-208.

Zheng, W., Sun, W., & Simeonov, A. (2018). Drug repurposing screens and synergistic drug combinations for infectious diseases. *British Journal of Pharmacology*, *175*(2), 181-191.

Zhou, G., Stevenson, M. M., Geary, T. G., & Xia, J. (2016). Comprehensive transcriptome meta-analysis to characterize host immune responses in helminth infections. *PLoS Neglected Tropical Diseases*, 10(4), e0004624.

Appendix Table. ProTSAV results of the 145 protein structures.

Protein_ID	TRINITY_ID	NAME OF THE PROTEIN	ProTSAV RESULTS
482425682	TRINITY_DN51775_c4_g1_i1	Methionine--tRNA ligase (EC 6.1.1.10) (Methionyl-tRNA synthetase) (MetRS)	
1273710290	TRINITY_DN53206_c4_g1_i4	Alpha-xylosidase (EC 3.2.1.177)	
7530446034	TRINITY_DN46783_c0_g1_i1	Probable serine/threonine-protein phosphatase PP2A regulatory subunit (Protein phosphatase PP2A regulatory subunit A)	
8947811483	TRINITY_DN44431_c1_g1_i1	Transketolase 2 (TK 2) (EC 2.2.1.1)	
9153676475	TRINITY_DN45788_c5_g1_i1	Protease 2 (EC 3.4.21.83) (Oligopeptidase B) (Protease II)	
787777209	TRINITY_DN51782_c10_g1_i1	Fumarate reductase flavoprotein subunit (EC 1.3.5.4) (Flavocytochrome c) (Flavocytochrome c3) (Fcc3)	
528355	TRINITY_DN50599_c5_g1_i2	DNA topoisomerase 4 subunit B (EC 5.99.1.3) (Topoisomerase IV subunit B)	
528367	TRINITY_DN51304_c6_g6_i1	Anaerobic glycerol-3-phosphate dehydrogenase subunit A (G-3-P dehydrogenase) (EC 1.1.5.3)	
3346328518	TRINITY_DN53589_c8_g1_i4	Uncharacterized protein YdiJ	
7642289554	TRINITY_DN53523_c1_g5_i1	Ribonucleoside-diphosphate reductase 1 subunit alpha (EC 1.17.4.1) (Protein B1) (Ribonucleoside-diphosphate reductase 1 R1 subunit) (Ribonucleotide reductase 1)	
8219309022	TRINITY_DN52985_c1_g1_i1	Putative ribose/galactose/methyl galactoside import ATP-binding protein 3 (EC 3.6.3.17)	
7634602772	TRINITY_DN50952_c4_g1_i2	D-alanine--D-alanyl carrier protein ligase (DCL) (EC 6.2.1.-) (D-alanine--poly(phosphoribitol) ligase subunit 1) (D-alanine-activating enzyme) (DAE)	
3697836620	TRINITY_DN52324_c7_g6_i2	H(+)/Cl(-) exchange transporter ClcA	
6303455774	TRINITY_DN51597_c1_g3_i2	Proton/glutamate-aspartate symporter (Glutamate-aspartate carrier protein) (Proton-glutamate-aspartate transport protein)	
6393818097	TRINITY_DN45419_c0_g1_i1	Xylulose kinase (Xylulokinase) (EC 2.7.1.17)	
9638340556	TRINITY_DN37933_c1_g1_i1	Uncharacterized sodium-dependent transporter MJ1319	
6901918187	TRINITY_DN53506_c0_g1_i1	Xylulose kinase (Xylulokinase) (EC 2.7.1.17)	
6333277324	TRINITY_DN53305_c8_g3_i3	Bifunctional protein Aas [Includes: 2-acylglycerophosphoethanolamine acyltransferase (EC 2.3.1.40) (2-acyl-GPE acyltransferase) (Acyl-[acyl-carrier-protein]--phospholipid O-acyltransferase); Acyl-[acyl-carrier-protein] synthetase (EC 6.2.1.20) (Acyl-ACP synthetase) (Long-chain-fatty-acid--[acyl-carrier-protein] ligase)]	
9334233975	TRINITY_DN47557_c7_g2_i1	HTH-type transcriptional regulator TauR	
3365003550	TRINITY_DN53504_c3_g1_i1	Uncharacterized HTH-type transcriptional regulator YdcR	

8126134150	TRINITY_DN52309_c3_g1_i1	Oxygen-independent coproporphyrinogen III oxidase (CPO) (EC 1.3.98.3) (Coproporphyrinogen III dehydrogenase) (CPDH)	
6299001827	TRINITY_DN52309_c3_g1_i6	Oxygen-independent coproporphyrinogen III oxidase (CPO) (EC 1.3.98.3) (Coproporphyrinogen III dehydrogenase) (CPDH)	
4252855338	TRINITY_DN43187_c0_g1_i1	Nicotinate phosphoribosyltransferase (NAPRTase) (EC 6.3.4.21)	
7344909827	TRINITY_DN50814_c0_g1_i1	Testicular acid phosphatase homolog (EC 3.1.3.2) (Acid phosphatase 4)	
1724124087	TRINITY_DN48924_c7_g3_i1	6-phospho-beta-glucosidase BglA (EC 3.2.1.86) (Phospho-beta-glucosidase A)	
6003180011	TRINITY_DN41347_c0_g1_i1	Argininosuccinate synthase (EC 6.3.4.5) (Citrulline--aspartate ligase)	
9357702524	TRINITY_DN53232_c3_g2_i3	Diaminobutyrate--2-oxoglutarate aminotransferase (EC 2.6.1.76) (Diaminobutyrate transaminase) (L-2,4-diaminobutyrate:2-ketoglutarate 4-aminotransferase) (DABA aminotransferase) (DABA-AT) (L-diaminobutyric acid transaminase)	
9383909297	TRINITY_DN47921_c4_g2_i1	Thiosulfate sulfurtransferase YnjE (EC 2.8.1.1)	
8305025303	TRINITY_DN51571_c3_g3_i4	Argininosuccinate synthase (EC 6.3.4.5) (Citrulline--aspartate ligase)	
6442817744	TRINITY_DN52550_c1_g1_i1	NADH oxidase (NOXase) (EC 1.6.99.3)	
3788291875	TRINITY_DN51862_c4_g1_i2	Oligopeptide transport ATP-binding protein OppF	
6048289859	TRINITY_DN53328_c1_g3_i3	Erythronate-4-phosphate dehydrogenase (EC 1.1.1.290)	
528415	TRINITY_DN27140_c0_g1_i1	Tyrosine--tRNA ligase (EC 6.1.1.1) (Tyrosyl-tRNA synthetase) (TyrRS)	
5242219070	TRINITY_DN44708_c0_g1_i1	Ribulokinase (EC 2.7.1.16)	
4642681529	TRINITY_DN31968_c1_g1_i1	50S ribosomal protein L3 glutamine methyltransferase (L3 MTase) (EC 2.1.1.298) (N5-glutamine methyltransferase PrmB)	
862494428	TRINITY_DN44483_c1_g2_i1	Glycogen synthase (EC 2.4.1.21) (Starch [bacterial glycogen] synthase)	
3587873328	TRINITY_DN53572_c2_g2_i2	Probable alcohol dehydrogenase (EC 1.1.1.1)	
7623160482	TRINITY_DN16089_c0_g1_i1	4-hydroxyphenylpyruvate dioxygenase (4HPPD) (HPD) (HPPDase) (EC 1.13.11.27)	
4140209631	TRINITY_DN52453_c10_g3_i7	Alcohol dehydrogenase YqhD (EC 1.1.1.-)	
2883622216	TRINITY_DN52453_c10_g3_i8	Alcohol dehydrogenase YqhD (EC 1.1.1.-)	
5682100587	TRINITY_DN32575_c0_g1_i1	Acetylornithine deacetylase (AO) (Acetylornithinase) (EC 3.5.1.16) (N-acetylornithinase) (NAO)	
1806384574	TRINITY_DN53361_c7_g1_i1	ATP-binding protein SyrD	
2275259647	TRINITY_DN51896_c8_g2_i1	Aromatic-amino-acid aminotransferase 1 (ARAT-I) (AROAT) (EC 2.6.1.57)	

4759315328	TRINITY_DN37816_c0_g2_i1	Diaminobutyrate--2-oxoglutarate transaminase (EC 2.6.1.76)	
2376536127	TRINITY_DN53480_c0_g3_i1	Aldehyde-alcohol dehydrogenase [Includes: Alcohol dehydrogenase (ADH) (EC 1.1.1.1); Acetaldehyde dehydrogenase [acetylating] (ACDH) (EC 1.2.1.10); Pyruvate-formate-lyase deactivase (PFL deactivase)]	
4604804111	TRINITY_DN44136_c0_g2_i1	GTPase HflX (GTP-binding protein HflX)	
9906567419	TRINITY_DN52331_c8_g1_i3	N-methyl-L-tryptophan oxidase (MTOX) (EC 1.5.3.-)	
5823274829	TRINITY_DN52331_c8_g1_i5	N-methyl-L-tryptophan oxidase (MTOX) (EC 1.5.3.-)	
5913346375	TRINITY_DN48353_c3_g1_i1	Caffeyl-CoA reductase-Etf complex subunit CarE (EC 1.3.1.108) (Electron transfer flavoprotein large subunit) (ETFLS) (Electron transfer flavoprotein subunit alpha) (Alpha-ETF)	
8379208332	TRINITY_DN49813_c1_g1_i4	Glutamate-pyruvate aminotransferase AlaC (EC 2.6.1.2)	
2540442732	TRINITY_DN53020_c0_g1_i1	Outer envelope protein 64, mitochondrial (Mitochondrial outer membrane protein 64) (mtOM64) (Translocon at the outer membrane of chloroplasts 64-V) (AtTOC64-V)	
2797660945	TRINITY_DN50112_c2_g1_i1	NADPH-dependent butanol dehydrogenase (BDH) (EC 1.1.1.-)	
790232341	TRINITY_DN52026_c3_g1_i3	Fe(3+) dicitrate transport ATP-binding protein FecE (Iron(III) dicitrate transport ATP-binding protein FecE)	
928829167	TRINITY_DN37427_c2_g1_i1	Alcohol dehydrogenase (EC 1.1.1.1) (EC 1.1.1.4) (EC 1.2.1.3)	
4320599518	TRINITY_DN51571_c3_g3_i2	Argininosuccinate synthase (EC 6.3.4.5) (Citrulline--aspartate ligase)	
2933585705	TRINITY_DN53318_c7_g2_i1	Protein MalY [Includes: Cystathionine beta-lyase MalY (CBL) (EC 4.4.1.13) (Beta-cystathionase MalY) (Cysteine lyase MalY) (Cysteine-S-conjugate beta-lyase MalY); Maltose regulon modulator]	
9287267390	TRINITY_DN50507_c4_g1_i1	Elongation factor Ts (EF-Ts)	
358104334	TRINITY_DN9533_c0_g1_i1	Alcohol dehydrogenase (EC 1.1.1.1) (EC 1.1.1.4) (EC 1.2.1.3)	
668498950	TRINITY_DN51718_c2_g2_i2	Uncharacterized ABC transporter ATP-binding protein YdiF	
2589049535	TRINITY_DN51680_c5_g2_i3	Uncharacterized sulfatase YidJ (EC 3.1.6.-)	
1102886325	TRINITY_DN50323_c9_g1_i1	High-affinity branched-chain amino acid transport ATP-binding protein BraG	
9255578976	TRINITY_DN47763_c2_g2_i1	Ribosomal protein S12 methylthiotransferase RimO (S12 MTTase) (S12 methylthiotransferase) (EC 2.8.4.4) (Ribosomal protein S12 (aspartate-C(3))-methylthiotransferase) (Ribosome maturation factor RimO)	
1595698752	TRINITY_DN48169_c3_g1_i1	Uncharacterized peptidase SSP1059 (EC 3.4.-.-)	
6587568796	TRINITY_DN48674_c0_g1_i1	Gamma-soluble NSF attachment protein (SNAP-gamma) (N-ethylmaleimide-sensitive factor attachment protein gamma)	

3760744725	TRINITY_DN4138_c0_g1_i1	Oxygen-independent coproporphyrinogen III oxidase (CPO) (EC 1.3.98.3) (Coproporphyrinogen III dehydrogenase) (CPDH)	
7701248468	TRINITY_DN53489_c2_g4_i1	Elongation factor P--(R)-beta-lysine ligase (EF-P--(R)-beta-lysine ligase) (EC 6.3.1.-) (EF-P post-translational modification enzyme A) (EF-P-lysine lysyltransferase)	
6681826044	TRINITY_DN53053_c2_g1_i1	6-phospho-beta-glucosidase GmuD (EC 3.2.1.86) (Aryl-phospho-beta-D-glucosidase BglD) (Glucomannan utilization protein D)	
9216785451	TRINITY_DN53556_c0_g3_i1	3-phenylpropionate/cinnamic acid dioxygenase ferredoxin--NAD(+) reductase component (EC 1.18.1.3)	
1538767709	TRINITY_DN53290_c1_g2_i2	Maltodextrin import ATP-binding protein MsmX (EC 3.6.3.-)	
3201818878	TRINITY_DN44420_c2_g2_i1	NADP-dependent alcohol dehydrogenase C (EC 1.1.1.2)	
7816811536	TRINITY_DN44072_c0_g1_i1	Branched-chain-amino-acid aminotransferase (BCAT) (EC 2.6.1.42) (Transaminase B)	
5768324451	TRINITY_DN53366_c4_g2_i8	NAD kinase (EC 2.7.1.23) (ATP-dependent NAD kinase)	
4210469847	TRINITY_DN48199_c5_g1_i1	Nitric oxide reductase FlRd-NAD(+) reductase (EC 1.18.1.-) (Flavorubredoxin reductase) (FlRd-reductase) (FlavoRb reductase)	
8726550460	TRINITY_DN53087_c8_g1_i2	Uncharacterized HTH-type transcriptional regulator YjiR	
5254767144	TRINITY_DN52182_c2_g1_i2	Aerobic glycerol-3-phosphate dehydrogenase (EC 1.1.5.3)	
4167816983	TRINITY_DN53535_c4_g2_i1	Chelated iron transport system membrane protein YfeB	
3989176136	TRINITY_DN16205_c0_g1_i1	(R,R)-butanediol dehydrogenase (EC 1.1.1.4)	
5027376897	TRINITY_DN51533_c3_g1_i1	Methionyl-tRNA formyltransferase (EC 2.1.2.9)	
9135920245	TRINITY_DN44003_c0_g1_i1	32 kDa beta-galactoside-binding lectin (Galectin-1)	
3029466317	TRINITY_DN50525_c2_g1_i1	Glyoxylate/succinic semialdehyde reductase 1 (AtGLYR1) (AtGR1) (SSA reductase 1) (EC 1.1.1.79) (EC 1.1.1.n11) (Gamma-hydroxybutyrate dehydrogenase) (AtGHBDH)	
930550650	TRINITY_DN36963_c0_g1_i2	Chloride intracellular channel exc-4 (Excretory canal abnormal protein 4)	
8270973614	TRINITY_DN53174_c2_g1_i2	6-phospho-beta-glucosidase GmuD (EC 3.2.1.86) (Aryl-phospho-beta-D-glucosidase BglD) (Glucomannan utilization protein D)	
9914938749	TRINITY_DN43605_c3_g1_i1	Acetylornithine deacetylase (AO) (Acetylornithinase) (EC 3.5.1.16) (N-acetylornithinase) (NAO)	
8188374233	TRINITY_DN46184_c0_g1_i1	Galactose/methyl galactoside import ATP-binding protein MglA (EC 3.6.3.17)	
7866572062	TRINITY_DN53449_c2_g1_i1	Putative TrmH family tRNA/rRNA methyltransferase YacO (EC 2.1.1.-)	
6659383548	TRINITY_DN53246_c2_g2_i1	Iron(3+)-hydroxamate import ATP-binding protein FhuC (EC 3.6.3.34) (Ferric hydroxamate uptake protein C) (Ferrichrome transport ATP-binding protein FhuC) (Iron(III)-hydroxamate import ATP-binding protein FhuC)	

5249306751	TRINITY_DN43187_c0_g1_i2	Nicotinate phosphoribosyltransferase (NAPRTase) (EC 6.3.4.21)	
9565738396	TRINITY_DN41200_c0_g2_i1	Putative quercetin 2,3-dioxygenase sll1773 (Putative quercetinase) (EC 1.13.11.24) (Pirin-like protein sll1773)	
5710901193	TRINITY_DN50862_c2_g1_i1	Probable N-glycosylase/DNA lyase [Includes: 8-oxoguanine DNA glycosylase (EC 3.2.2.-); DNA-(apurinic or apyrimidinic site) lyase (AP lyase) (EC 4.2.99.18)]	
6003879286	TRINITY_DN44794_c0_g1_i1	Quinone oxidoreductase (EC 1.6.5.5) (NADPH:quinone reductase) (Zeta-crystallin homolog protein)	
7473955783	TRINITY_DN45661_c0_g1_i1	Fructokinase (EC 2.7.1.4)	
5869558507	TRINITY_DN48924_c7_g3_i2	Beta-glucoside kinase (EC 2.7.1.85)	
9174802125	TRINITY_DN53178_c5_g1_i1	Bifunctional polymyxin resistance protein ArnA [Includes: UDP-4-amino-4-deoxy-L-arabinose formyltransferase (EC 2.1.2.13) (ArnAFT) (UDP-L-Ara4N formyltransferase); UDP-glucuronic acid oxidase, UDP-4-keto-hexauronic acid decarboxylating (EC 1.1.1.305) (ArnADH) (UDP-GlcUA decarboxylase) (UDP-glucuronic acid dehydrogenase)]	
415551470	TRINITY_DN48438_c6_g2_i2	Putative kinase EAE_16955	
9558650309	TRINITY_DN42811_c0_g1_i1	GDP-6-deoxy-D-talose 4-dehydrogenase (EC 1.1.1.135) (GDP-4-keto-6-deoxy-D-mannose reductase)	
8452601088	TRINITY_DN49362_c15_g1_i1	1,4-alpha-glucan branching enzyme GlgB (EC 2.4.1.18) (1,4-alpha-D-glucan:1,4-alpha-D-glucan 6-glucosyl-transferase) (Alpha-(1->4)-glucan branching enzyme) (Glycogen branching enzyme) (BE)	
4256505929	TRINITY_DN52373_c7_g2_i1	N-hydroxyarylamine O-acetyltransferase (EC 2.3.1.118) (Arylamine N-acetyltransferase) (Arylhydroxamate N,O-acetyltransferase)	
1537252737	TRINITY_DN53530_c3_g1_i2	Uracil phosphoribosyltransferase (EC 2.4.2.9) (UMP pyrophosphorylase) (UPRTase)	
7357520961	TRINITY_DN50571_c2_g1_i3	Poly(U)-specific endoribonuclease-B (EC 3.1.-.-) (Protein endoU-B) (Uridylate-specific endoribonuclease-B)	
9885392899	TRINITY_DN44792_c2_g1_i1	Bifunctional NAD(P)H-hydrate repair enzyme Nnr (Nicotinamide nucleotide repair protein) [Includes: ADP-dependent (S)-NAD(P)H-hydrate dehydratase (EC 4.2.1.136) (ADP-dependent NAD(P)HX dehydratase); NAD(P)H-hydrate epimerase (EC 5.1.99.6) (NAD(P)HX epimerase)]	
2500081070	TRINITY_DN43623_c0_g1_i1	ADP-ribosyl cyclase/cyclic ADP-ribose hydrolase 1 (EC 3.2.2.6) (2'-phospho-ADP-ribosyl cyclase) (2'-phospho-ADP-ribosyl cyclase/2'-phospho-cyclic-ADP-ribose transferase) (EC 2.4.99.20) (2'-phospho-cyclic-ADP-ribose transferase) (ADP-ribosyl cyclase 1) (ADPRC 1) (CD38H) (Cyclic ADP-ribose hydrolase 1) (cADPr hydrolase 1) (CD antigen CD38)	
4390250215	TRINITY_DN46299_c1_g1_i1	Endoribonuclease Dcr-1 (Protein dicer-1) (EC 3.1.26.-)	
8923009203	TRINITY_DN44239_c0_g2_i1	Sialic acid transporter NanT (Sialic acid permease) (Sialic acid/H(+) symporter)	

278150355	TRINITY_DN48390_c0_g1_i1	Lectin L6	
1584305319	TRINITY_DN50525_c5_g3_i1	Multidrug resistance protein MdtK (Multidrug-efflux transporter)	
7691628812	TRINITY_DN50525_c5_g3_i4	Multidrug resistance protein MdtK (Multidrug-efflux transporter)	
528536	TRINITY_DN53568_c3_g4_i4	Ribonuclease 3 (EC 3.1.26.3) (Ribonuclease III) (RNase III)	
5300222172	TRINITY_DN47007_c0_g1_i1	ADP-ribosyl cyclase/cyclic ADP-ribose hydrolase 1 (EC 3.2.2.6) (2'-phospho-ADP-ribosyl cyclase) (2'-phospho-ADP-ribosyl cyclase/2'-phospho-cyclic-ADP-ribose transferase) (EC 2.4.99.20) (2'-phospho-cyclic-ADP-ribose transferase) (ADP-ribosyl cyclase 1) (ADPRC 1) (CD38H) (Cyclic ADP-ribose hydrolase 1) (cADPr hydrolase 1) (CD antigen CD38)	
6509090057	TRINITY_DN37304_c0_g2_i1	Glyoxylate reductase (EC 1.1.1.26)	
9460541228	TRINITY_DN48109_c2_g2_i1	Xylose import ATP-binding protein XylG (EC 3.6.3.17)	
1829600195	TRINITY_DN44239_c0_g3_i1	Sialic acid transporter NanT (Sialic acid permease) (Sialic acid/H(+) symporter)	
1415798932	TRINITY_DN52927_c8_g4_i1	Cobalt import ATP-binding protein CbiO (EC 3.6.3.-) (Energy-coupling factor transporter ATP-binding protein CbiO) (ECF transporter A component CbiO)	
9859495403	TRINITY_DN42486_c1_g2_i1	Putative phosphonates utilization ATP-binding protein PhnK	
4430536167	TRINITY_DN42486_c1_g3_i1	Alpha-D-ribose 1-methylphosphonate 5-triphosphate synthase subunit PhnL (RPnTP synthase subunit PhnL) (EC 2.7.8.37)	
6027916240	TRINITY_DN51710_c2_g1_i2	Selenide, water dikinase (EC 2.7.9.3) (Selenium donor protein) (Selenophosphate synthase)	
3280915228	TRINITY_DN33027_c0_g1_i1	Chloride intracellular channel exc-4 (Excretory canal abnormal protein 4)	
177689629	TRINITY_DN44084_c0_g1_i1	Poly(U)-specific endoribonuclease-B (EC 3.1.-.-) (Protein endoU-B) (Uridylate-specific endoribonuclease-B) (XendoU-B)	
9401399225	TRINITY_DN49800_c0_g1_i2	Queuine tRNA-ribosyltransferase accessory subunit 2 (Queuine tRNA-ribosyltransferase domain-containing protein 1)	
1667898890	TRINITY_DN53322_c3_g1_i4	Pyridoxal 4-dehydrogenase (tPLDH) (EC 1.1.1.107)	
8085342952	TRINITY_DN52750_c8_g1_i2	Taurine import ATP-binding protein TauB (EC 7.6.2.7)	
4219572209	TRINITY_DN53584_c2_g1_i1	Ribosomal RNA large subunit methyltransferase E (EC 2.1.1.166) (23S rRNA Um2552 methyltransferase) (rRNA (uridine-2'-O-)-methyltransferase)	
513549831	TRINITY_DN53377_c1_g5_i1	23S rRNA (guanosine-2'-O-)-methyltransferase RlmB (EC 2.1.1.185) (23S rRNA (guanosine2251 2'-O)-methyltransferase) (23S rRNA Gm2251 2'-O-methyltransferase)	
9843767564	TRINITY_DN52182_c2_g4_i1	Aerobic glycerol-3-phosphate dehydrogenase (EC 1.1.5.3)	
9036456003	TRINITY_DN43492_c0_g1_i1	DNA mismatch repair protein MutL	
6100841513	TRINITY_DN21017_c0_g1_i1	Ribonuclease 3 (EC 3.1.26.3) (Ribonuclease III) (RNase III)	

7055306439	TRINITY_DN52173_c4_g2_i2	Enterobactin synthase component F (EC 2.7.7.-) (Enterochelin synthase F) (Serine-activating enzyme) (Seryl-AMP ligase)	
1497346298	TRINITY_DN52797_c3_g2_i2	Short-chain dehydrogenase/reductase ATR9 (EC 1.-.-.-) (Core atranone cluster (CAC) protein 9)	
7466467303	TRINITY_DN43567_c0_g1_i1	Ribonucleoside-diphosphate reductase subunit beta (EC 1.17.4.1) (Ribonucleotide reductase small subunit)	
528735	TRINITY_DN37544_c1_g2_i1	1-acyl-sn-glycerol-3-phosphate acyltransferase (1-AGP acyltransferase) (1-AGPAT) (EC 2.3.1.51) (Lysophosphatidic acid acyltransferase) (LPAAT)	
13083775	TRINITY_DN46819_c12_g1_i1	Pyrimidine 5'-nucleotidase YjjG (EC 3.1.3.5) (House-cleaning nucleotidase) (Non-canonical pyrimidine nucleotide phosphatase) (Nucleoside 5'-monophosphate phosphohydrolase) (dUMP phosphatase)	
9548923948	TRINITY_DN53535_c4_g2_i2	Chelated iron transport system membrane protein YfeB	
4701532661	TRINITY_DN46129_c0_g1_i1	Ribonuclease Oy (RNase Oy) (EC 3.1.27.-)	
4374684200	TRINITY_DN55161_c0_g1_i1	1,2-phenylacetyl-CoA epoxidase, subunit E (EC 1.-.-.-) (1,2-phenylacetyl-CoA epoxidase, reductase subunit) (1,2-phenylacetyl-CoA monooxygenase, subunit E)	
7395938838	TRINITY_DN48524_c4_g1_i1	Pantothenate precursors transporter PanS	
7245983241	TRINITY_DN47572_c0_g1_i1	Ferritin light chain 1 (Ferritin L subunit 1)	
8836865199	TRINITY_DN53416_c2_g2_i1	Uncharacterized 37.6 kDa protein in cld 5'region (ORF2)	
1540910039	TRINITY_DN53530_c3_g1_i1	Uracil phosphoribosyltransferase (EC 2.4.2.9) (UMP pyrophosphorylase) (UPRTase)	
8321170944	TRINITY_DN15899_c0_g1_i1	Octaprenyl diphosphate synthase (EC 2.5.1.90) (All-trans-octaprenyl-diphosphate synthase) (Octaprenyl pyrophosphate synthase) (OPP synthase)	
3353030725	TRINITY_DN53087_c8_g1_i1	Uncharacterized HTH-type transcriptional regulator YjiR	
6506294583	TRINITY_DN52309_c3_g1_i7	Oxygen-independent coproporphyrinogen III oxidase (CPO) (EC 1.3.98.3) (Coproporphyrinogen III dehydrogenase) (CPDH)	
7923213894	TRINITY_DN55642_c0_g1_i1	DNA mismatch repair protein MutL	
4419491071	TRINITY_DN53410_c1_g5_i4	Arginine transport ATP-binding protein ArtP (EC 3.6.3.-)	
1883543968	TRINITY_DN10786_c0_g1_i1	Methanol dehydrogenase activator (EC 3.-.-.-) (MDH activator)	
3597947641	TRINITY_DN45099_c0_g1_i1	Calmodulin (CaM)	
4016614052	TRINITY_DN42869_c0_g1_i1	16 kDa calcium-binding protein (Egg antigen SME16)	

BIODATA (Publications, Conferences, Trainings and Workshops)

Prasad, Y. K[1]., Dahal, S[1]., Saikia, B., Bordoloi, B., Tandon, V., & Ghatani, S. (2019). Artyfechinostomum sufrartyfex Trematode Infections in Children, Bihar, India. *Emerging infectious diseases*, *25*(8), 1571–1573. https://doi.org/10.3201/eid2508.181427

Dahal, S., Gour, P., Raghuvanshi, S., Prasad, Y. K., Saikia, D., & Ghatani, S. (2022). Multi-stage transcriptome profiling of the neglected food-borne echinostome *Artyfechinostomum sufrartyf*ex reveal potential diagnostic and drug targets. *Acta Tropica*, 233, 106564. doi.org/10.1016/j.actatropica.2022.106564

Saikia, D., Prasad, Y. K., Dahal, S., & Ghatani, S. (2022). *Fasciolopsis buski* detected in Humans in Bihar and Pigs in Assam, India. *Emerging infectious diseases*, *28*(6), 1265–1268. https://doi.org/10.3201/eid2806.220171

Conferences (International and National)

International Conference on "**Chemical Ecology, Environment and Human Health: Emerging Frontiers and Synthesis**" (ICCEEHH, 2019) (Oral Presentation)

National Conference on "**Recent Trends in Biological Research and Future Prospects**", 2018 (Poster Presentation).

Training/Workshops/Seminars	Organizers	Year
Computational R-Bioinformatics and PERL Programming	Department of Science, Technology and Climate Change Sikkim State Council of Science and Technology	2016
Bioinformatics workshop on Genomics, Proteomics, Drug Design and High performance computing	Supercomputing Facility for Bioinformatics and Computational Biology, Indian Institute of Technology, Delhi, India	2016
Forest, Biodiversity and Plant Sciences	Department of Botany, Sikkim University	2016
Next Generation Sequence Analysis	Bionivid Technology Pvt Ltd, Bangalore, India	2017
Ethical and Biosafety Issues	Institutional Biosafety comimttee and Institutional Ethical Committe, Sikkim University	2018
Mathematical Genomics	Indian Statistical Institute, Kolkata, India	2018
Bioinformatics in Eukaryotes Diversity	DBT funded Bioinformatics Centre, Sikkim University	2018
Data Management and Analysis for Biologists using R	Department of Zoology, Sikkim University and ATREE	2019
Awareness on Intellectual Property Rights Issue	Committee on IPR, Sikkim University, Gangtok, India	2019
Hands- on Training Programme on R-software	Department of Horticulture, Sikkim University	2022

Artyfechinostomum sufrartyfex Trematode Infections in Children, Bihar, India

Yugal K. Prasad,[1] Suman Dahal,[1] Barsha Saikia, Bobita Bordoloi, Veena Tandon, Sudeep Ghatani

Eating raw or insufficiently cooked mollusks is a known risk factor for human echinostomiasis. We confirmed identification of *Artyfechinostomum sufrartyfex* trematodes as the causative agent of disease among 170 children in northern Bihar, India. We also identified the snail *Pila globosa* as a potential source of infections in the study area.

Foodborne intestinal trematodiasis, especially that caused by members of the family *Echinostomatidae*, is an emerging yet neglected public health disease. Approximately 24 echinostome species cause human echinostomiasis and are highly endemic to Southeast Asia and the Far East; major foci are located in China, India, Indonesia, South Korea, Malaysia, the Philippines, and Thailand (*1*).

Previously, only 2 deaths attributed to the echinostomid fluke *Artyfechinostomum sufrartyfex* were reported from the states of Assam and Tamil Nadu in India (*2*,*3*). During 2004–2017, several cases of echinostome infection were reported in children at Shri Shubh Lal Hospital and Research Centre in Bihar, India.

The Study

This study was approved by the Institutional Ethics Committee of Sikkim University (SU/IEC/2017/04), Gangtok, India. A total of 170 cases of *A. sufrartyfex* trematode infection occurred in northern Bihar, India, mostly in children ≤12 years of age. The children lived in the districts of Sitamarhi and Sheohar in the state of Bihar. Signs and symptoms were diarrhea (persistent/chronic and acute) with watery or mucus-bound stool, vomiting, loss of appetite, weakness, passage of red worms in stool or vomit, swelling of the feet and the entire body, fever, cough, breathlessness, night blindness, and urticarial rashes (Table 1).

Physical examination showed that most patients were anemic. Clinical laboratory investigations showed leukocytosis and eosinophilia. However, systemic examination showed no adverse effects of the cardiovascular, abdominal, and central nervous systems (Table 2). Levels of serum alanine aminotransferase, bilirubin, blood urea, creatinine, electrolytes, sodium, potassium, and chloride were within reference limits.

These children were immediately hospitalized and kept under careful observation with routine monitoring of stool and vomit for worms. Once worms were observed in samples, the patients were given praziquantel (75 mg/kg in 3 divided doses orally for 2 days) and monitored. At administration of the drug, patients started passing more worms in stool. We recovered >50 worms but ≤300 worms from each child patient. The infection subsided after the standard dose of praziquantel, and most patients recovered from the infection.

However, we observed 11 deaths: 2 patients each during 2004, 2007, 2008, 2012, and 2013 and 1 patient during 2009. Severe acute malnutrition with or without edema and large numbers of worms were major clinical conditions observed for these deaths. Nine children had persistent diarrhea with severe dehydration and shock, and 2 of them had acute diarrhea, severe dehydration, and shock.

The infected patients frequently consumed raw snails. The most prevalent snail species in the study areas was *Pila globosa*, which the children collected from the banks of ponds/ditches and waterlogged paddy fields grossly contaminated with human and animal excreta (168/170 cases, 99%). Therefore, we surveyed as many as 8 sites in 2 districts (Sitamarhi and Sheohar) for snail samples from their natural habitats (Figure 1).

We screened the snails by using a digestion technique with a 0.5% pepsin/0.1% HCl solution and found that the snails were heavily infected with metacercariae, which are the encysted infective stage of the trematode. The prevalence of metacercariae in the snails ranged from 16.12% in Hanumannagar to ≤48.19% in Punaura (Appendix, https://wwwnc.cdc.gov/EID/article/25/8/18-1427-App1.pdf).

To establish the source of infection, we attempted to identify the clinical parasite samples and the metacercariae. We morphologically identified representative parasite samples isolated from the patients (*2*,*4*,*5*). However, we could not identify metacercaria by only morphologic characteristics

Author affiliations: Shri Shubh Lal Hospital and Research Centre, Sitamarhi, India (Y.K. Prasad); Sikkim University, Gangtok, India (S. Dahal, B. Saikia, B. Bordoloi, S. Ghatani); Biotech Park, Lucknow, India (V. Tandon)

DOI: https://doi.org/10.3201/eid2508.181427

[1]These authors contributed equally to this article.

Acta Tropica 233 (2022) 106564

Contents lists available at ScienceDirect

Acta Tropica

journal homepage: www.elsevier.com/locate/actatropica

Multi-stage transcriptome profiling of the neglected food-borne echinostome *Artyfechinostomum sufrartyfex* reveal potential diagnostic and drug targets

Suman Dahal [a], Pratibha Gour [b], Saurabh Raghuvanshi [b], Yugal Kishore Prasad [c], Dipshikha Saikia [a], Sudeep Ghatani [a,*]

[a] Department of Zoology, School of Life sciences, Sikkim University, Gangtok, Sikkim, India
[b] Department of Plant Molecular Biology, University of Delhi, New Delhi, India
[c] Director, Shri Shubh Lal Hospital and Research Centre, Sitamarhi, Bihar, India

ARTICLE INFO

Keywords:
Echinostomiasis
Artyfechinostomum sufrartyfex
Transcriptome
Secretome
Diagnostic targets
Drug targets

ABSTRACT

Lack of effective surveillance and control methods for neglected helminth diseases particularly in context of rural areas in India is a serious concern in terms of public health. With regard to the emerging food-borne echinostomid *Artyfechinostomum sufrartyfex* infection in the country, the current study is an *in silico* attempt to screen for plausible diagnostic and drug targets against the trematode. Transcriptome of adult, encysted and excysted metacercaria stages of the parasite was generated using Illumina sequencing platform. A de-novo assembly strategy utilizing transcriptome data generated from the three lifecycle stages was followed to generate the representative transcripts. Longest open reading frames identified for the transcripts were further conceptually translated into their respective protein sequences. Detailed analysis of this dataset through various bioinformatics pipelines and tools eventually identified 14 credible diagnostic and 10 drug targets along with their FDA-approved and ZINC molecules. Some of the important diagnostic candidates include thioredoxin peroxidase, haemoglobinase, cathepsin L, cathepsin L-like and B-like cysteine proteases. Among the drug targets, uncharacterized sodium dependent transporter and bifunctional protein Aas were identified as top targets exhibiting significant interaction with Rifamycin and ZINC02820058 molecule, respectively. Further, B-cell epitope analysis of the diagnostic targets revealed unique epitopes for 10 of them thus indicating their potential role in specific diagnosis of the parasite. The diagnostic candidates along with a number of lesser known drug targets and their ligand molecules identified in this study provides a reasonable basis for evaluation and development of future intervention strategies against *A. sufrartyfex*.

1. Introduction

Human echinostomiasis is a zoonotic, food-borne, intestinal infection caused by digenetic trematodes under the family Echinostomatidae. About 24 echinostome species particularly endemic to Southeast Asia and the Far East with major foci situated in China, India, Indonesia, Korea, Malaysia, the Philippines and Thailand contribute to the disease (Toledo and Esteban, 2016). We had previously reported several infection cases of the echinostomid *Artyfechinostomum sufrartyfex* among children in India along with establishment of snail *Pila globosa* as the second intermediate host (Prasad et al., 2019). The parasite has been reported time and again from humans and other vertebrate animals in the country with recent reports being from pigs and small Indian mongoose (Das et al., 2017; Borkataki et al., 2017).

The rapid advancements in whole genome, transcriptome and secretome research can be used in development of new anti-parasitic agents and diagnostic candidates. Transcriptome profiling using RNA-seq approach in particular finds its utility in providing a vast array of information for the trematode group as well and studies utilizing transcriptome sequencing focused on various trematodes has already been taken up worldwide. These works have improved our understanding of the biology of these worms and have provided valuable leads toward the discovery of new intervention targets. The importance of excretory secretory proteins (ESPs) in parasitic infections has also been

* Corresponding author.
E-mail address: sghatani@cus.ac.in (S. Ghatani).

https://doi.org/10.1016/j.actatropica.2022.106564
Received 1 February 2022; Received in revised form 1 June 2022; Accepted 9 June 2022
Available online 15 June 2022
0001-706X/

Fasciolopsis buski Detected in Humans in Bihar and Pigs in Assam, India

Dipshikha Saikia,[1] Yugal K. Prasad,[1] Suman Dahal, Sudeep Ghatani

The foodborne intestinal trematode *Fasciolopsis buski* causes the neglected zoonotic disease fasciolopsiasis. We detected *F. buski* infection in 14 pediatric patients in Sitamarhi, Bihar, and in pigs in Sivasagar, Assam, India. Proper diagnostic methods and surveillance are urgently needed to accurately estimate the true burden of this disease in India.

Fasciolopsis buski is a foodborne intestinal trematode that causes the neglected zoonotic disease fasciolopsiasis in humans and pigs. *F. buski* infection is transmitted through ingestion of raw aquatic plants or water carrying encysted metacercaria. Persons with substantial worm loads can have clinical indicators, such as malnutrition, edema, malabsorption, severe diarrhea, ascites, and anemia, and might experience acute intestinal obstruction and ileus (*1–3*). *F. buski* worms are found mostly in Asia and the Indian subcontinent; endemicity is highest in eastern India (*4*). We previously reported multiple cases of infection with *Artyfechinostomum sufrartyfex*, an echinostome trematode, which was diagnosed in children at Shri Shubh Lal (SSL) Hospital and Research Centre in Sitamarhi, Bihar state, India (*5*). We also documented several cases of fasciolopsiasis among SSL patients during 2012–2021 and infection in pigs detected in Sivasagar district, Assam state, India during a 2019–2020 survey.

Infections with this parasite have been reported from diverse regions of India, as well as other parts of Asia (*6–10*). A genetic study suggested that the species found in India is different from species found in China and Vietnam (*11*). To corroborate the genetic distinctions between the strains found in India and those from China and Vietnam, we determined the complete nuclear ribosomal ITS2 and partial mitochondrial cytochrome c oxidase subunit 1 gene (cox1) sequences of *F. buski* from the samples recovered from Bihar and Assam and compared them to sequences from isolates from other regions of India, China, and Vietnam. The institutional ethics committee of Sikkim University in Gangtok, India, approved this work (SU/REG/F-1/03/2018/VOL-1/59).

The Study

During 2012–2021, a total of 14 children 3–12 years of age were brought for treatment to SSL Hospital for reported loose bowel movements, including watery feces and feces tinged with blood and mucus for >15 days, as well as vomiting, flatulence, abdominal discomfort, pain in the abdomen, fever, loss of appetite, weakness, and passage of flat reddish worms, called paterwa or lal keera in local languages. Eight patients were male and 6 female. Among the male patients, 3 were ≤5, 3 were 6–10, and 2 were >10 years of age; among the female patients, 3 were ≤5 and 3 were ≥10 years of age. All of the patients were of low socioeconomic status and resided near ponds or deep-water rice paddies contaminated with human and animal excreta and snail-infested areas. The patients were habituated to consume raw snails, contaminated water, chestnuts, and vegetables irrigated with contaminated water from nearby ditches.

On physical examination, all patients were pale and malnourished. General and systemic examination revealed persistent diarrhea, dehydration, and vomiting in most and anemia in all of the case-patients. Laboratory investigation revealed most of the patients had eosinophilia, and grade II malnutrition was associated with most patients (Table 1). All patients tested negative on tuberculin and HIV tests, and results from routine urine examination, complete blood

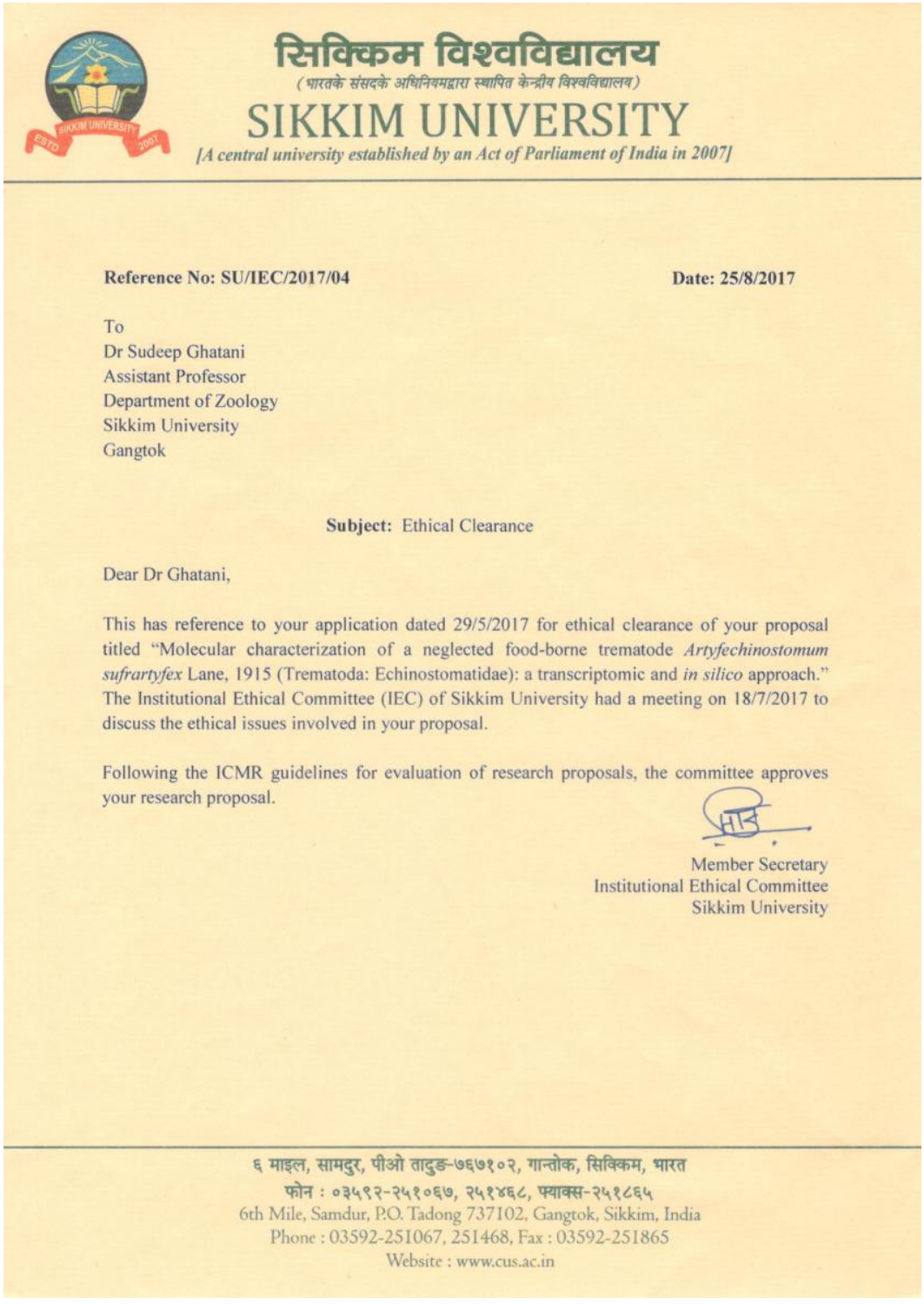
Author affiliations: Sikkim University, Gangtok, India (D. Saikia, S. Dahal, S. Ghatani); Shri Shubh Lal Hospital and Research Centre, Sitamarhi, India (Y.K. Prasad)

DOI: https://doi.org/10.3201/eid2806.220171

[1]These first authors contributed equally to this article.

ANNEXURES

ANNEXURE I



सिक्किम विश्वविद्यालय
(भारतके संसदके अधिनियमद्वारा स्थापित केन्द्रीय विश्वविद्यालय)
SIKKIM UNIVERSITY
[A central university established by an Act of Parliament of India in 2007]

Reference No: SU/IEC/2017/04 **Date: 25/8/2017**

To
Dr Sudeep Ghatani
Assistant Professor
Department of Zoology
Sikkim University
Gangtok

Subject: Ethical Clearance

Dear Dr Ghatani,

This has reference to your application dated 29/5/2017 for ethical clearance of your proposal titled "Molecular characterization of a neglected food-borne trematode *Artyfechinostomum sufrartyfex* Lane, 1915 (Trematoda: Echinostomatidae): a transcriptomic and *in silico* approach." The Institutional Ethical Committee (IEC) of Sikkim University had a meeting on 18/7/2017 to discuss the ethical issues involved in your proposal.

Following the ICMR guidelines for evaluation of research proposals, the committee approves your research proposal.

Member Secretary
Institutional Ethical Committee
Sikkim University

६ माइल, सामदुर, पीओ तादुङ-७६७१०२, गान्तोक, सिक्किम, भारत
फोन : ०३५९२-२५१०६७, २५१४६८, फ्याक्स-२५१८६५
6th Mile, Samdur, P.O. Tadong 737102, Gangtok, Sikkim, India
Phone : 03592-251067, 251468, Fax : 03592-251865
Website : www.cus.ac.in

ANNEXURE II

Consent Form

(To be filled by the participants/parents/guardians and researcher)

I along with my son/daughter/spouse do hereby agree to be a participant in the research work as appended below:

Title: "Molecular characterization of a neglected food-borne trematode *Artyfechinostomum sufrartyfex* Lane, 1915 (Trematoda: Echinostomatidae): a transcriptomic and in silico approach"

In this study the personal information and the **adult parasites from stool** sample (s) will be collected.

As a participant I have been explained and have understood that this type of study has been conducted in several parts of the world and has no adverse effects on the health/social status of the participant.

Dr **Sudeep Ghatani** is to be contacted if additional information is required. (Mobile no. **+919862240590**)

Name & Signature of the participant/ parents/ guardian

Relation with participant:

Date:

Signature of Researcher

ANNEXURE III

सहमति पत्र

(प्रतिभागियों /माता-पिता/ अभिभावकों और शोधकर्ताओं द्वारा भरा जाए)

मैं अपने बेटे / बेटी / पति / पत्नी के साथ नीचे दिए गए अनुसार अनुसंधान कार्य में एक भागीदार बनने के लिए इस प्रकार से सहमत हूं :

शीर्षक: "`Molecular characterization of a neglected food-borne trematode *Artyfechinostomum sufrartyfex* Lane, 1915 (Trematoda: Echinostomatidae): a transcriptomic and in silico approach`"

इस अध्ययन में व्यक्तिगत जानकारी और मल से वयस्क परजीवी एकत्र किए जाएंगे।
प्रतिभागी के रूप में मुझे समझाया गया है और मै यह समझ गया हूँ कि इस प्रकार के अध्ययन दुनिया के कई हिस्सों में किए गए हैं और इसका प्रतिभागी के स्वास्थ्य / सामाजिक स्थिति पर कोई प्रतिकूल प्रभाव नहीं पड़ता है।

यदि अतिरिक्त जानकारी की आवश्यकता है तो डॉ सुदीप घाटानी से संपर्क किया जाए। मोबाइल न.: +919862240590

प्रतिभागी / माता-पिता / अभिभावक के नाम और हस्ताक्षर

प्रतिभागी के साथ संबंध:

दिनांक: शोधकर्ता के हस्ताक्षर