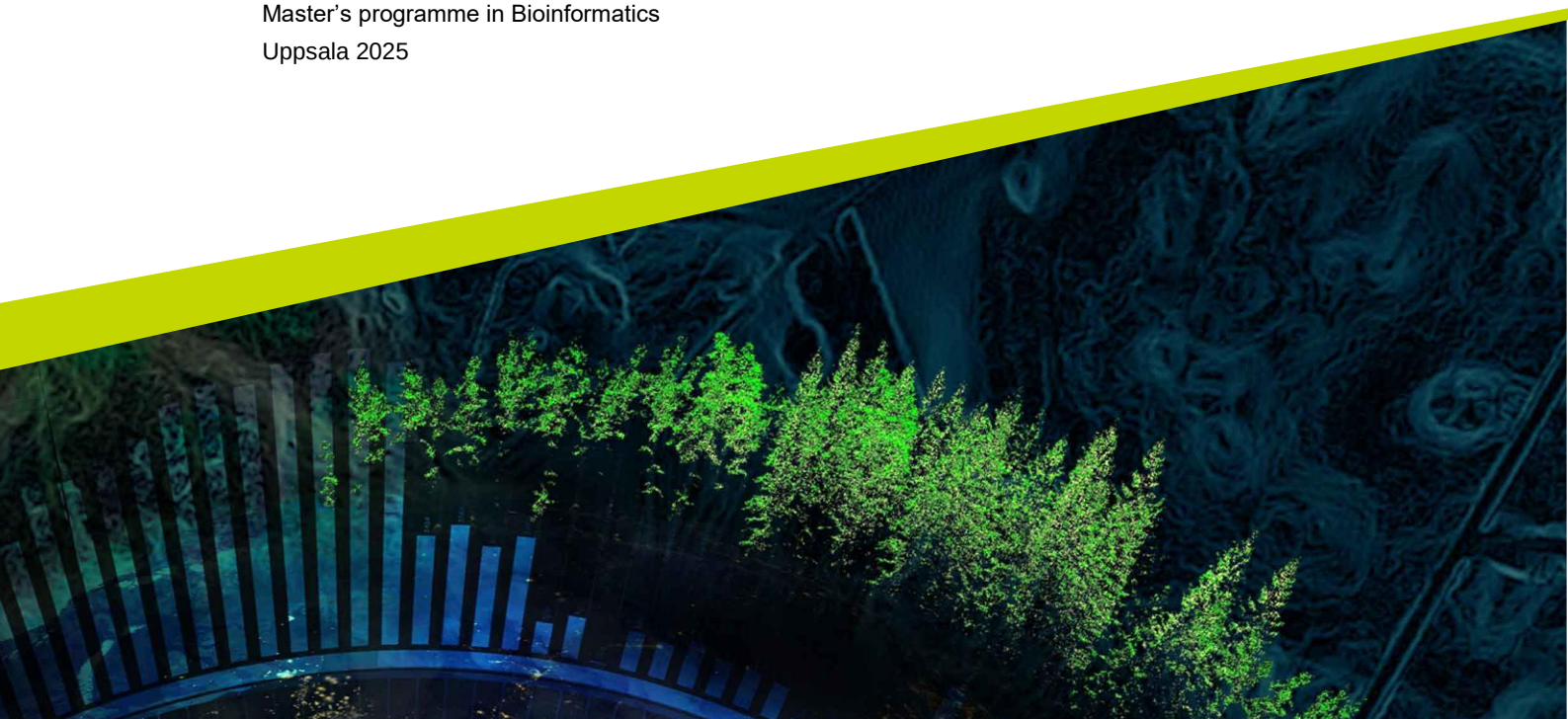
In Silico Identification and Prioritisation of Antifilarial Drug Targets in *Setaria digitata*

Kasun Chamalka D H J P

Independent project • 30 credits
Swedish University of Agricultural Sciences, SLU
Faculty of Veterinary Medicine and Animal Science/Department of Animal Biosciences
Master's programme in Bioinformatics
Uppsala 2025



In Silico Identification and Prioritisation of Antifilarial Drug Targets in *Setaria digitata*

Author's name: Kasun Chamalka D H J P

Supervisor:	**Erik Bongcam Rudloff, Swedish University of Agricultural Sciences, Department of Animal Biosciences**
Assistant supervisor:	Samuel Coulbourn Flores, Swedish University of Agricultural Sciences, Department of Animal Biosciences
Examiner:	Göran Andersson, Swedish University of Agricultural Sciences, Department of Animal Biosciences

Credits:	30 credits
Level:	Second cycle, A2E
Course title:	Independent project in Bioinformatics
Course code:	EX1002 VT2025
Programme/education:	Master's programme in Bioinformatics
Course coordinating dept:	Department of Animal Biosciences
Place of publication:	Uppsala
Year of publication:	2025
Keywords:	*Setaria digitata*, Antifilarial drug targets, Comparative genomics, Subcellular localization, Computational drug discovery

Swedish University of Agricultural Sciences
Faculty of Veterinary Medicine and Animal Science
Department of Animal Biosciences

Abstract

Filarial nematode infection remains an important worldwide burden on animal and human health, exacerbated by widespread drug resistance and a slowness in the development of new anthelmintic drugs. *Setaria digitata*, a veterinary filarial parasite closely related to human-infective species such as *Wuchereria bancrofti* and *Brugia malayi*, presents a tractable model for investigating novel therapeutic targets. This study employed a complete bioinformatics pipeline to prioritize parasite-specific drug targets in *S. digitata* with a goal of avoiding limitations found in conventional drug discovery and experimental inaccessibility of human filariae.

The study integrated various computational strategies such as sub-cellular localization prediction, functional annotation, and structural modelling. After filtering for proteins likely to be specific to nematodes, a multi-criteria scoring system was developed to rank them based on predicted essentiality, drug accessibility, and relevance to known therapeutic target classes such as ion channels, microtubules, neuroreceptors, and proteases. Prediction of druggability was further augmented with the use of Fpocket and COACH-D for prediction and validation of ligand-binding sites.

From a predicted, non-redundant proteome of 12,238 gene-derived protein sequences, subcellular localization analysis indicated that approximately 18% may be pharmacologically accessible, while functional annotation via eggNOG-mapper covered 70.2% of the dataset. Prioritization integrated essentiality, conservation, and accessibility, yielding 250 high-confidence targets, 88.4% of which were neurological proteins, recapitulating known anthelmintic mechanisms (e.g., ivermectin-targeted glutamate-gated chloride channels). Structural modelling of 58 candidates identified 30 high-druggability targets, including G-protein coupled receptors (GPCRs) and ion channels. COACH-D validation confirmed ligand-binding potential for top candidates, with SD_012157-T1 exhibiting strong similarity to established drug targets (TM-score: 0.56, binding energy: −7.2 kcal/mol). These results provide a foundation for experimental validation and rational antifilarial design, with implications for both human and veterinary parasitology.

Keywords: *Setaria digitata*, Antifilarial drug targets, Comparative genomics, Subcellular localization, Computational drug discovery.

Table of contents

List of tables

List of figures

Abbreviations

Abbreviation	Description
BioPython	A set of tools for biological computation in Python
COACH-D	A web server for protein–ligand binding site prediction
COG	Clusters of Orthologous Groups of proteins
CSN	Cerebrospinal nematodiosis
CSV	Comma-Separated Values
DEG	Differentially Expressed Genes
EC	Enzyme Commission (number)
eggNOG	Evolutionary genealogy of genes: Non-supervised Orthologous Groups
FASTA	A text-based format for representing nucleotide or peptide sequences
Fpocket	A software tool for detecting ligand-binding pockets in proteins
GluCl	Glutamate-gated Chloride Channel
GO	Gene Ontology
GPCR	G-Protein Coupled Receptor
GWAS	Genome-Wide Association Studies
KEGG	Kyoto Encyclopedia of Genes and Genomes
Matplotlib	A plotting library for Python
MD	Molecular dynamics
nAChR	Nicotinic Acetylcholine Receptor
NCBI	National Centre for Biotechnology Information
Pandas	A data analysis library for Python
Pfam	Protein families database
Plotly	An interactive graphing library for Python
Python	A programming language (used for scripting and data analysis)
RNAi	RNA interference
Seaborn	A statistical data visualization library based on Matplotlib
SignalP	A tool for signal peptide prediction
TDR	Special Programme for Research and Training in Tropical Diseases (WHO)
TMHMM	Transmembrane Hidden Markov Model (for predicting transmembrane helices)
TSV	Tab-Separated Values
WHO	World Health Organization

1. Introduction

1.1 Parasitic Nematode Infections and Drug Resistance

Helminths, including parasitic nematodes, cause a great health and economic burden both in veterinary and human medicine (Holden-Dye 2007; Fissiha & Kinde 2021). Parasitic worm infections are among the most widely spread tropical diseases globally, infecting approximately two billion people worldwide, mostly in the tropical and subtropical countries (Holden-Dye 2007).

Current disease control relies heavily on anthelmintic drugs, though development of new therapeutics has been limited by insufficient pharmaceutical investment for neglected tropical diseases (Holden-Dye 2007). Most anthelmintics that have been utilized to treat humans were first developed to be used in veterinary medicine, resulting in a narrow and overlapping set of drug options.

Misuse and overuse of medicines have caused the widespread development of resistance, inherited loss of sensitivity to drugs, which is now reported across helminth-species and drug classes (Fissiha & Kinde 2021). This includes resistance in livestock nematodes and reports of increased resistance also in nematodes that infects companion animals. Alarmingly, decreased efficacy is also emerging in human-targeted treatments such as ivermectin (*Onchocerca volvulus*), praziquantel (*Schistosoma*), and benzimidazoles (*Ascaris lumbricoides*) (Nixon et al. 2020). Resistance now threatens the success of mass drug administration programs globally.

1.2 Bottlenecks in Classical Drug Discovery

Anthelmintic drug discovery faces significant bottlenecks that hinder the development of new drug classes, particularly for human use. These challenges arise from biological, technical, and economic constraints that limit the efficiency of both traditional and modern approaches. A primary obstacle is the scarcity of well-characterized and validated molecular targets in helminths, which restricts the ability to design selective drugs and explore novel chemical spaces for treatment development (Peak & Hoffmann 2011; Mengarda et al. 2023). Additionally, the complex biology of parasitic nematodes including intricate life cycles and difficulties in maintaining long-term *in vitro* cultures poses a barrier to establishing scalable, high-throughput screening systems for candidate drugs (Nixon et al. 2020; Zajíčková et al. 2020). These biological complexities are compounded by the urgent need to address widespread resistance to existing anthelmintics and the limited efficacy of current treatments, yet few new drug classes have reached the market in recent decades (Nixon et al. 2020; Jayawardene et al. 2021).

Further complicating drug discovery is the evolutionary similarity between helminths and their mammalian hosts. As eukaryotes, parasitic nematodes share many conserved biological pathways with humans, increasing the challenge of designing selective drugs that minimize host toxicity (Mengarda et al. 2023). This issue is exacerbated by the lack of robust, high-throughput assays to quantitatively measure parasite viability and phenotypic responses, leaving traditional whole-organism screening methods as inefficient alternatives (Grant & Behm 2007; Peak & Hoffmann 2011). Such technical limitations are particularly problematic in resource-limited settings, where inadequate funding and infrastructure in endemic regions further slow progress (Kron et al. 2007).

Together, these bottlenecks underscore the need for innovative strategies to overcome the stagnation in anthelmintic development. Advances in target identification, such as computational prioritization pipelines and *in silico* screening, could help bridge gaps in molecular target validation. Meanwhile, improved *in vitro* culture systems and phenotypic assays may enhance throughput and reproducibility. Addressing these challenges will require collaborative efforts to align scientific innovation with equitable resource allocation, ensuring that drug discovery keeps pace with the growing threat of resistance and unmet therapeutic needs.

1.3 Overview of Anthelmintic Drug Mechanisms

Anthelmintic drugs operate through mechanisms distinct from those of antibacterial or antiviral agents. Antibiotics and antivirals generally focus on rapid-replication processes, while anthelmintics generally interfere with the neuromuscular or metabolic activity of the parasite. For instance, ivermectin binds to glutamate-gated chloride channels in nematodes, leading to increased chloride ion permeability, hyperpolarization of nerve and muscle cells, and subsequent paralysis and death of the parasite (Martin 1997; Köhler 2001). Similarly, albendazole interferes with microtubule formation by binding to β-tubulin, impairing glucose uptake and depleting energy reserves, which ultimately leads to the parasite's demise (Köhler 2001).

These mechanisms underscore the necessity for specialized strategies in helminth drug development, distinct from those employed against bacteria or viruses. A comprehensive summary of major anthelmintic drug classes, their validated molecular targets, and primary mechanisms of action is presented in Table 1.

Table 1: Major classes of anthelmintic drugs, their validated molecular targets, and primary mechanisms of action (Source: Adapted from various pharmacological studies and reviews)

Drug Class	Example Drugs	Molecular Target	Mechanism of Action
Macrocyclic Lactones	Ivermectin	Glutamate-gated chloride channels	Increases chloride ion permeability, causing hyperpolarization and paralysis of the parasite.
Benzimidazoles	Albendazole	β-tubulin	Inhibits microtubule polymerization, disrupting glucose uptake and depleting energy stores.
Imidazothiazoles	Levamisole	Nicotinic acetylcholine receptors	Acts as an agonist, causing spastic paralysis of the parasite.
Tetrahydropyrimidines	Pyrantel pamoate	Nicotinic acetylcholine receptors	Causes depolarizing neuromuscular blockade, leading to paralysis.
Salicylanilides	Niclosamide	Mitochondrial oxidative phosphorylation	Uncouples oxidative phosphorylation, disrupting ATP production.

1.4 The Case for *Setaria digitata* as a Model Organism

Filarial nematodes are of significant concern in veterinary and agricultural settings. One such nematode, *Setaria digitata*, is a cattle parasite that typically does not cause apparent disease in its natural host (cattle). However, it can lead to lethal cerebrospinal nematodiasis (CSN) in incidental hosts such as goats, sheep, and horses (Voronin et al. 2015). CSN not only threatens animal health and welfare but also increases veterinary costs for farmers, imposing a significant economic burden in endemic regions.

Importantly, *S. digitata* is morphologically, histologically, and antigenically very similar to human filariae *W. bancrofti* and *B. malayi* and thus is an excellent model for lymphatic filariasis research (Senanayake et al. 2020). Furthermore, *S. digitata* and *W. bancrofti* exhibit striking similarities in nucleotide sequences, gene content, and genome organization. The mitochondrial genome of *S. digitata* (13,839 bp) contains 36 genes, including 12 protein-coding genes, 22 tRNA genes, and 2 genes for rRNAs, mirroring the structure seen in *W. bancrofti* and other filarial nematodes (Perumal et al. 2016). Both species display high AT-content (*S. digitata*: 75.1%; *W. bancrofti*: ~71.2%), and comparative genomics reveals that *S. digitata* shares 7,070 genes with *W. bancrofti*, with 5,087

genes conserved across *S. digitata*, *W. bancrofti*, *Loa loa*, and *Brugia malayi* (Senanayake et al. 2020). Their nuclear genomes are also closely matched in size (*S. digitata*: 89.8 Mb; *W. bancrofti*: 88.4 Mb), with similar GC content (*S. digitata*: 31.73%; *W. bancrofti*: 28.8%) (Senanayake et al. 2020). At the protein level, *S. digitata* and *W. bancrofti* show strong correlations in coding sequences (R = 0.9), and specific proteins like SXP-1 and SdNP share high amino acid residue similarity (e.g., 79% for SdNP) (Perumal et al. 2016).

Phylogenetically, S. digitata clusters closely with *W. bancrofti* and other filarial nematodes, as evidenced by analyses of cox1 and 12S rDNA sequences (Perumal et al. 2016). A phylogenetic tree (Figure 1) places *S. digitata* within a clade containing Onchocerca and Brugia species, further supporting its evolutionary proximity to *W. bancrofti* (Senanayake et al. 2020). This genetic and functional resemblance underscores the utility of S. digitata as a model organism for studying *W. bancrofti* and related filarial parasites.

In contrast to *W. bancrofti*, a nematode that is not possible to culture in the laboratory, and which is host-specific to humans, *S. digitata* is experimentally tractable (Voronin et al. 2015; Perumal et al. 2016). Having high quality draft genomes of *S. digitata* and associated species such as *W. bancrofti* enables comparative genomics to be used to reveal conserved and parasite-specific genes (Senanayake et al. 2020). Such analyses support the discovery of novel therapeutic targets and enhance understanding of filarial biology.

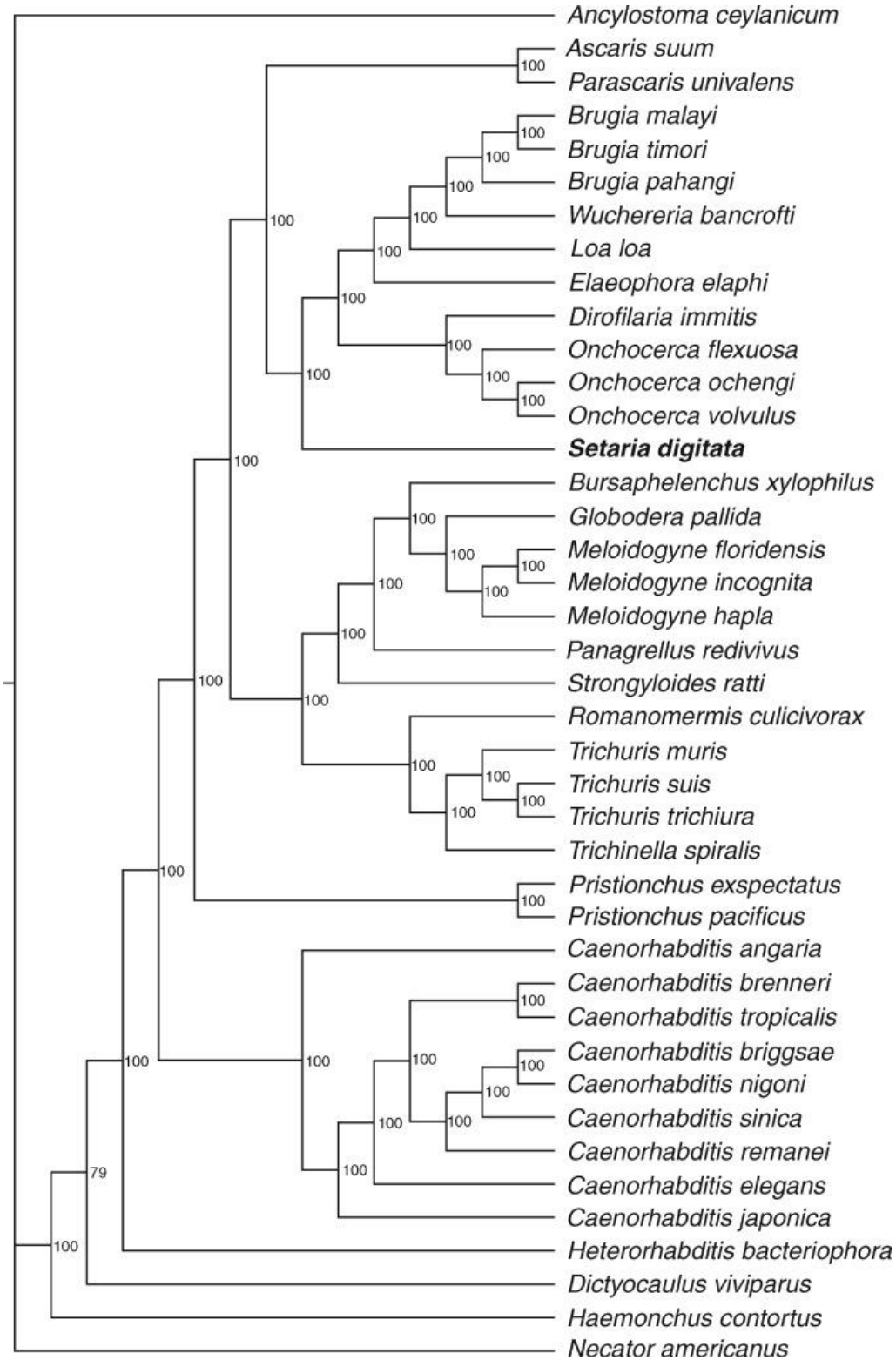


Figure 1: Phylogenetic relationship of Setaria digitata with Wuchereria bancrofti and other filarial nematodes. Consensus tree based on 41 nematode genomes, including the S. digitata assembly (Senanayake et al. 2020).

1.5 Research Questions

This study aims to answer the primary research question:

Can bioinformatics approaches identify novel, parasite-specific drug targets in *Setaria digitata*?

To address this, the study also investigates several secondary questions:

1. Which essential genes or metabolic pathways in *S. digitata* are absent or significantly divergent in mammalian hosts?
2. How do these targets compare to validated drug targets in related nematodes such as *B. malayi* and *W. bancrofti*?

These questions provide the foundation for prioritizing high-confidence, selective drug targets with therapeutic potential.

1.6 Knowledge Gaps and the Case for Computational Approaches

Despite the high burden of filarial infections, traditional drug discovery remains slower due to financial disincentives and the experimental inaccessibility of key human parasites (Voronin et al. 2015; Perumal et al. 2016). This results in heavy reliance on repurposed veterinary drugs with limited diversity.

Computational approaches provide scalable and cost-efficient alternatives. By leveraging omics datasets, they enable rapid screening of potential drug targets without requiring parasite cultivation. While bioinformatics has revolutionized drug discovery in oncology and immunology (Hukerikar et al. 2024), its application in helminthology remains limited, which is an underexplored opportunity this study seeks to address.

1.7 Justification of Methods

Genetics-based approaches such as genome-wide association studies (GWAS), pathway analysis, and comparative genomics have demonstrated higher clinical success rates (Hukerikar et al. 2024). Such strategies also lower dropouts and reduce late-stage failure, making them ideal for neglected diseases.

In this study, a computational target prioritization pipeline was developed by integrating multi-dimensional criteria, including essentiality (gene indispensability for parasite survival), subcellular localization (pharmacological accessibility), evolutionary conservation (host-parasite divergence), and druggability (structural suitability for small-molecule binding). For druggability assessment, Fpocket (available at; https://bioserv.rpbs.univ-paris-diderot.fr/services/fpocket/) was selected due to its computational efficiency and robustness in predicting ligand-

binding cavities from static protein structures, even in low-resolution models (Le Guilloux et al. 2009). To validate potential binding sites, COACH-D (available at; https://yanglab.qd.sdu.edu.cn/COACH-D/) was employed, leveraging its demonstrated accuracy in modelling protein-ligand interactions for neglected pathogens, particularly where experimental structural data are scarce (Nixon et al. 2020). This dual approach balanced high-throughput screening (Fpocket) with template-based validation (COACH-D), ensuring both scalability and reliability in target selection.

1.8 Summary of Approach

The study involved proteomic filtering, functional annotation, subcellular localization prediction, structural modelling, and a rule-based scoring pipeline. Prioritized targets were filtered for parasite-specificity, essentiality, and accessibility. Druggability was assessed via Fpocket and COACH-D, followed by ligand docking to rank top candidates.

1.9 Broader Implications

Identifying novel, parasite-specific drug targets in *S. digitata* holds promise for accelerating anti filarial drug development, particularly for human lymphatic filariasis, which still affects over 120 million people globally. Improved control of *S. digitata* in livestock could also mitigate CSN and reduce economic losses in small ruminant farming.

2. Materials and Methods

2.1 Genome Data Acquisition and Protein Sequence Preparation

The *S. digitata* proteome file, comprising over 15,000 predicted protein sequences with associated functional annotations, was obtained from a high-quality draft genome assembly (genome completeness: CEGMA 91.5%, BUSCO 85.5%; Senanayake et al., 2020) available in GenBank (accession number: GCA_900083525.1). To construct a non-redundant proteome, only the longest isoform per gene was retained from the original multi-isoform dataset. This curated FASTA file served as the basis for all downstream analyses, as illustrated in Figure 2.

2.2 Subcellular Localization Prediction Using DeepLoc

Subcellular localization of proteins was predicted using the DeepLoc 2.1 web server (Ødum et al. 2024). The input FASTA file was divided into 499-sequence batches using a custom Python script to comply with server upload limits. Outputs from each batch were merged to generate a comprehensive localization dataset for 12,238 proteins.

DeepLoc 2.1 was chosen over alternatives like WoLF PSORT and CELLO because of its deep learning-based architecture and superior prediction accuracy across diverse eukaryotic proteomes, including parasites (Ødum et al. 2024). Its ability to accurately predict both signal peptides and membrane-localized proteins made it especially suitable for identifying drug-accessible targets. In contrast, WoLF PSORT and CELLO rely on older machine learning methods, which are generally less effective for non-model eukaryotes such as *S. digitata*.

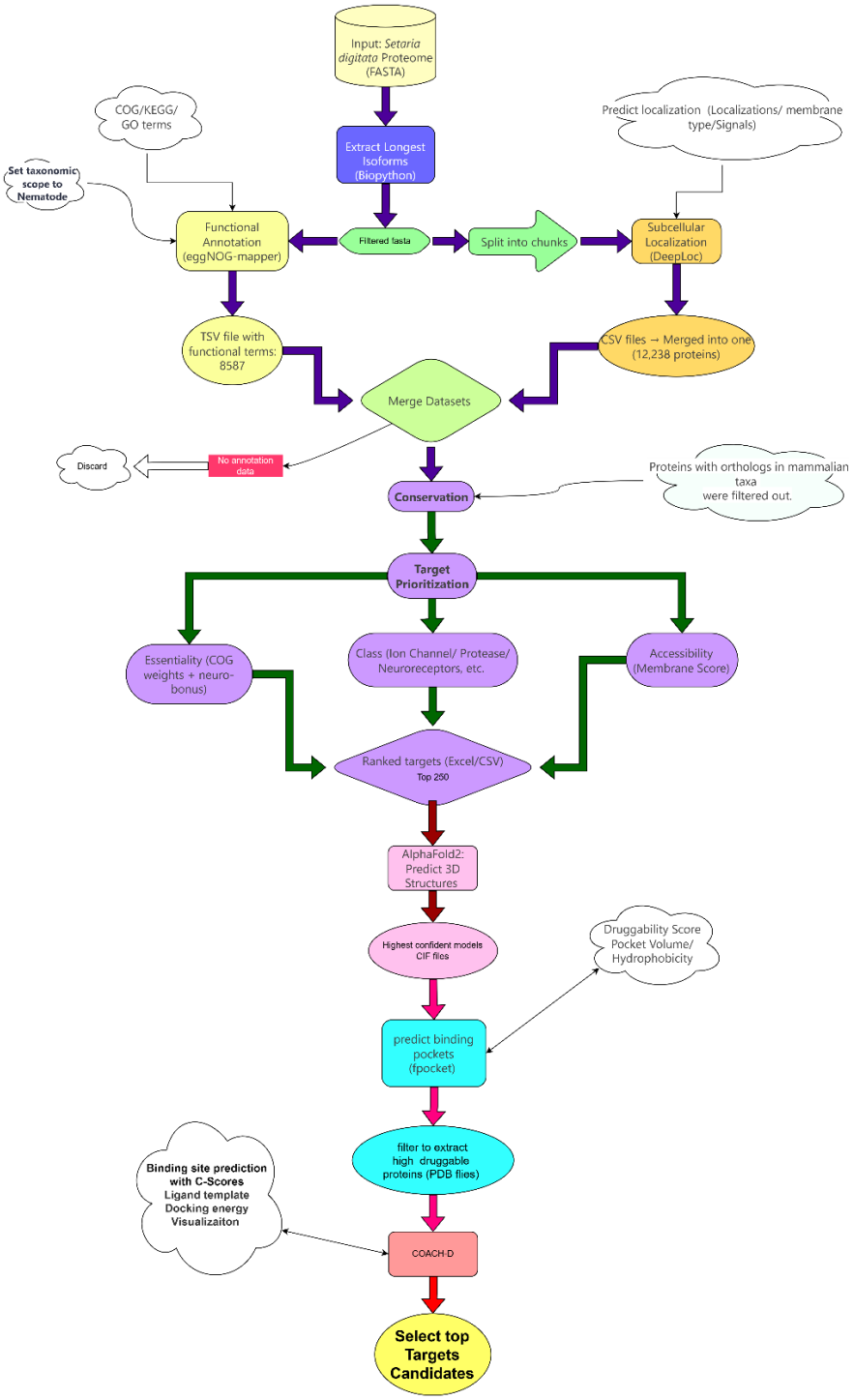
2.3 Functional Annotation with eggNOG-mapper

Functional annotations were performed using eggNOG-mapper v2.1.8 (Huerta-Cepas et al. 2017) on the Galaxy EU platform (Afgan et al. 2018), employing the eggNOG database v5.0.2 with default settings. To enhance annotation specificity and biological relevance, the taxonomic scope was restricted to the Nematoda clade (NCBI TaxID: 6231). This approach minimized annotation errors from distantly related orthologs (Huerta-Cepas et al. 2017).

The pipeline yielded annotations including Clusters of Orthologous Groups of proteins (COG) categories, Gene Ontology (GO) terms, KEGG pathways, and Enzyme Commission (EC) numbers in Tab-Separated Values (TSV) format, which was later converted to Comma-Separated Values (CSV) for integration. The choice

of eggNOG-mapper over tools such as InterProScan or PfamScan was driven by its comprehensive orthology-based framework and integrated output, enabling functional inference alongside domain identification. While InterProScan excels in domain detection, it lacks the orthology-contextualized inference required for this study's aims. However, functional predictions are contingent upon the accuracy of assignments of confirmed orthologous versus paralogous genes. Missed annotations may persist, especially for proteins from poorly characterized nematodes.

Figure 2: Computational pipeline integrates essentiality, accessibility, and structural druggability to prioritize targets.



2.4 Target Prioritization Pipeline

2.4.1 Data Integration

Three core datasets were integrated: (1) eggNOG-mapper functional annotations (8587 proteins), (2) DeepLoc subcellular predictions (12,238 proteins), and (3) the curated proteome FASTA file. Integration was performed using a Python script that matched protein identifiers and filtered out proteins lacking either annotation type. This conservative filtering was carried out in an effort to achieve completeness and accuracy, resulting in a high-confidence dataset for downstream scoring. However, functional annotation coverage remained incomplete for some proteins. Proteins without localization and functional data were excluded, potentially omitting biologically significant targets.

2.4.2 Conservation Filtering

Taxonomic-based conservation filtering was performed to prioritize parasite-specific drug targets and reduce the risk of off-target effects in hosts. Proteins with seed orthologs linked to mammalian taxa including human (9606), mouse (10090), rat (10116), and vertebrates (7742) were identified using string-based pattern matching and excluded from the dataset. This method ensured the removal of proteins with clearly annotated conservation across mammals.

2.4.3 Essentiality Scoring

A quantitative essentiality scoring system was developed based on two primary biological criteria: (1) conservation of core cellular functions through COG categories, and (2) neurophysiological relevance to known anthelmintic mechanisms. The weighting scheme was empirically derived through a three-step process.

First, baseline weights for COG functional categories were established by reviewing studies of essential genes across model nematodes (*C. elegans*) and related parasites (*B. malayi*). Categories directly involved in fundamental cellular processes (translation, DNA replication, energy production) received the highest weights (2.0–2.5), based on evidence that these pathways are both evolutionarily conserved and critical for survival (Zhang & Lin 2009; Galperin et al. 2021)(Table 2). For example, ribosomal proteins (COG J), DNA replication machinery (COG D), and Gene expression regulation (COG L) were all weighted at 2.5, reflecting their universal essentiality across eukaryotes. While transcription (COG K) is also a fundamental process, it was assigned a comparatively lower weight in this model due to limitations in available annotation confidence and a focus on the most universally essential functions for initial prioritization. This decision reflects a methodological simplification rather than a judgment on biological importance.

Second, therapeutic relevance was incorporated by adding supplemental weights (0.5-2.0 bonus points) to proteins involved in neurological processes targeted by

existing anthelmintics (Table 3). This adjustment was based on GO term annotations (e.g., GO:0015276 for glutamate-gated chloride channels) and KEGG pathway mapping (e.g., neuroactive ligand-receptor interactions). The magnitude of bonuses reflected clinical importance, with ivermectin targets receiving +2.0 and other neuroreceptors +1.5.

Finally, all scores were normalized to a 0-5 scale to enable cross-category comparison. The weighting thresholds were validated by checking whether or not known essential genes in *C. elegans* (from WormBase; https://wormbase.org/#012-34-5) and clinically validated anthelmintic targets consistently scored above 3.0 in our system. While RNAi data would provide more direct evidence of essentiality, the absence of such data for *S. digitata* necessitated this orthology-based approach, which has been successfully applied in other neglected pathogens (Berenstein et al. 2016).

Normalized Score = 5 × (Raw Score) / (Max Score)

This systematic weighting strategy allowed us to quantitatively prioritize targets while accounting for both fundamental biological importance and practical therapeutic potential. However, this method assumes functional conservation across taxa. Parasitic adaptations may bypass canonical pathways, reducing the validity of COG-based inference is a limitation of this method.

Table 2:COG category weights used in essentiality scoring

COG	Functional Category	Weight	Rationale
J	Translation	2.5	Ribosomal machinery essential for survival
D	Cell cycle/replication	2.5	DNA replication core components
L	Chromatin organization	2.5	Gene expression regulation
C	Energy production	2.0	Mitochondrial ATP synthesis
E/F	Amino acid/nucleotide metabolism	1.8	Biosynthetic pathway enzymes
M/G/V	Structural functions	1.5	Cell envelope and cytoskeleton integrity
T/U/O/K/H/I	Signalling/Other	0.8–1.2	Variable roles in regulation
P/Q/N	Miscellaneous	0.2–0.5	Low-priority metabolic roles
R/S/-	General/Unknown	0.0–0.1	Uncharacterized or non-essential

Table 3: Neurophysiological Bonus Scores

Annotation Type	Term ID	Bonus	Biological Relevance
GO	GO:0015276	+2.0	Glutamate-gated chloride channels (GluCl)
GO	GO:0007268	+1.2	Synaptic transmission
GO	GO:0005230/4888	+1.5	Ion/ligand receptors
GO	GO:0006811	+1.2	Ion transport

Annotation Type	Term ID	Bonus	Biological Relevance
KEGG	map04080	+1.5	Neuroactive ligand-receptor interaction

2.4.4 Target Classification

A rule-based classifier was developed to categorize targets based on therapeutic relevance. Each protein was scanned for matching terms in GO, KEGG, and PFAM databases. Upon the first match to a class (e.g., ion channels, proteases), it was assigned accordingly, in a fixed priority order (ion channel = microtubule = neuroreceptor → protease → other).

Therapeutic priority scores were assigned to each class based on known antifilarial drug mechanisms (Table 4), including macrocyclic lactones (e.g., ivermectin), benzimidazoles, and cysteine protease inhibitors. Annotations were validated through manual curation and cross-checked with entries in the WHO Model List of Essential Medicines (WHO 2023) *and* TDR Targets database. However, rule-based classifiers may misclassify multifunctional proteins or overlook novel druggable classes not captured by existing ontology terms.

Table 4:Target Classification System for Antifilarial Drug Discovery.

Class	Gene Ontology Terms (GO)	KEGG Orthologs (KO)	PFAM Domains	Therapeutic Association	Priority Score
Ion Channel	GO:0005216 (ion channel activity)	K04885 (GluCl α-subunit)	PF00520 (Ion transport)	Macrocyclic lactones (ivermectin/moxidectin)	2.5
	GO:0022824 (ligand-gated channel	K05325 (nAChR α-subunit)	PF07885 (LGIC)		
	GO:0015276 (GluCl receptor)	K05032 (GABA receptor)	PF02932 (GluCl binding)		
Microtubule	GO:0005200 (tubulin binding)	K10380 (α-tubulin)	PF00091 (Tubulin)	Benzimidazoles (albendazole)	2.5
	GO:0007010 (cytoskeleton organization)	K10483 (β-tubulin)	PF03953 (Tubulin C-terminal)		
	GO:0005874 (β-tubulin)				
Neuroreceptor	GO:0004888 (transmembrane receptor)	K04145 (5-HT7 receptor)	PF00001 (7TM GPCR)	Amino-acetonitrile derivatives (monepantel)	2.5

	GO:0004930 (G-protein coupled receptor)	K04608 (dopamine receptor)	PF00003 (Neurotransmitter receptor)		
	GO:0004994 (serotonin receptor)	K08042 (octopamine receptor)	PF10320 (Neuropeptide receptor)		
Protease	GO:0008233 (cysteine peptidase)	K01358 (cathepsin L)	PF00112 (Peptidase C1A)	Cysteine protease inhibitors (K11777)	2
	GO:0006508 (proteolysis)	K01359 (cathepsin B)	PF00082 (Subtilase)		
	GO:0004197 (caspase activity)	K01365 (legumain)	PF01650 (Caspase)		
Other	None	None	None	Non-prioritized targets	1

Priority scores reflect known anthelmintic mechanisms; underrepresented classes (e.g., proteases) may require manual reweighting.

2.4.5 Accessibility Scoring

Subcellular localization was inferred as indirect druggability measure, according to the hypothesis that extracellular and membrane proteins are more accessible to drug compounds, especially large or hydrophilic compounds (Rask-Andersen et al. 2011; Punta et al. 2012). To assess this, a dual-component accessibility scoring system was developed, combining DeepLoc-predicted localization data with signal peptide information.

Membrane affinity was quantified by weighting the presence of membrane-related features such as extracellular, cell membrane, transmembrane, and peripheral according to pharmacological relevance (Table 5). In parallel, a compartment-based accessibility score was assigned based on the dominant predicted localization, with a +0.5-bonus applied for extracellular proteins containing signal peptides (Table 6). Final accessibility scores were capped at 3.0 using a min() function to standardize values across proteins. While SignalP and TMHMM could have been used independently for signal peptide and transmembrane domain prediction, DeepLoc's integrated predictions reduced pipeline complexity and offered comparable accuracy. A limitation of this approach is that subcellular localization tools may misclassify proteins with dual or atypical targeting signals, potentially leading to inaccurate accessibility estimates.

Table 5: Membrane feature weights

Feature	Weight	Biological Rationale
Extracellular	1.5×	Direct drug access without membrane penetration
Cell membrane	1.2×	Surface-exposed therapeutic targets
Transmembrane	1.0×	Integral membrane proteins
Peripheral membrane	0.8×	Temporarily membrane associated

Table 6:Localization scoring scheme

Compartment	Base Score	Bonus Condition	Max Score
Extracellular	3.0	+0.5 (if signal peptide present)	3.5
Cell membrane	2.5	-	2.5
Organellar*	1.0	-	1.0
Cytosolic	0.5	-	0.5

* Includes nucleus and mitochondria.

2.4.6 Target Prioritization

Candidate drug targets were ranked using a composite scoring system integrating five biologically relevant features (Table 7): (1) target class relevance (25%), (2) nematode-specific essentiality (25%), (3) presence of filarial-specific functional keywords (10%), (4) membrane association features (15%), and (5) subcellular accessibility (15%). A 2.0× bonus was applied to proteins whose functional annotations matched antifilarial-associated terms such as "glutamate-gated" or "β-tubulin", identified using case-insensitive regular expressions. Combined scores were calculated by summing the weighted contributions of each feature. Targets were stratified into four tiers using quantile-based binning: Top (≥97th percentile), High (90–96th), Medium (75–89th), and Low (<75th).

Protein sequences were retrieved using BioPython's `SeqIO` module and matched to their identifiers. FASTA headers for the top 250 proteins included relevant metadata (e.g., combined score, target class) to facilitate downstream functional and structural analysis. All computations were performed in Python 3.9 using BioPython 1.79 and Pandas 1.3.5. Visualizations were generated using Matplotlib 3.5.2, Seaborn 0.11.2, and Plotly 5.9.0.

Table 7:Components of the Target Prioritization Score

Component	Weight	Data Source	Example High-Scoring Features
Target class	25%	Manual classification	Ion channels, microtubule-associated proteins
Essentiality	25%	EggNOG COG/KO data	Ribosomal subunits, DNA replication factors
Antifilarial keywords	10%	Functional descriptions	GABA receptors, cathepsins

Component	Weight	Data Source	Example High-Scoring Features
Membrane features	15%	DeepLoc predictions	Extracellular domains, signal peptides
Subcellular accessibility	15%	DeepLoc predictions	Extracellular, cell membrane-localized proteins

2.5 Structural Modelling and Druggability Analysis

2.5.1 Structural Modelling

To evaluate druggability, structural models were generated for a subset of high-priority targets. AlphaFold was selected due to its high accuracy in predicting 3D protein structures, especially for proteins lacking close homologs in the Protein Data Bank (Jumper et al. 2021; https://alphafoldserver.com/). Although alternatives like RoseTTAFold (Baek et al. 2021; https://neurosnap.ai/service/RoseTTAFold2) and I-TASSER (Yang et al. 2015; https://zhanggroup.org/I-TASSER/) are faster, they are generally less accurate for low-sequence similarity targets.

Due to server limitations and time constraints, a random subset of 58 proteins was selected from the top 250 targets to ensure coverage across target classes. Structures were obtained in Crystallographic Information File (CIF) format and converted to Protein Data Bank (PDB) format using Open Babel 3.1.1 (O'Boyle et al. 2011) for compatibility with downstream analysis tools.

AlphaFold predictions do not include ligand-bound states or account for conformational flexibility. Pocket geometry and accessibility may be underrepresented in static models.

2.5.2 Druggable Pocket Identification

Potential binding sites were identified using Fpocket v3.1.1 (Le Guilloux et al. 2009), which detects surface pockets based on alpha sphere geometry and evaluates them using physicochemical descriptors. Each pocket was scored on metrics such as volume, solvent-accessible surface area (SASA), hydrophobicity, polarity, and a heuristic druggability score.

A custom Python script parsed Fpocket output to extract metrics and identify druggable pockets (score $\geq$0.5), a conservative threshold informed by benchmarking studies (Schmidtke & Barril 2010). While more sophisticated tools like SiteMap or PockDrug offer advanced scoring, they are either commercial or computationally intensive. Fpocket was chosen for its speed, reproducibility, and compatibility with large datasets. However, Fpocket's heuristic scoring is based on datasets of human proteins and may miss allosteric or transient sites in parasitic proteins.

2.5.3 Validation with COACH-D

To validate Fpocket predictions, top-scoring proteins were analysed using COACH-D (Wu et al. 2018), a meta-server that integrates ligand-binding site predictors such as TM-SITE and COFACTOR. COACH-D provides consensus predictions with confidence scores (C-scores) and template-based ligand binding models.

Due to time constraints, only the three highest-priority proteins were analysed. While both methods generally agreed on pocket location, residue-level comparisons were not performed because the standalone Fpocket output lacks residue-specific annotations. This limitation could have been addressed using the Fpocket web interface, which was not utilized due to time constraints.

3. Results and Discussion

The study was intended to identify and rank potential antifilarial drug targets in *S. digitata* through a bioinformatics pipeline of steps. The obtained non-redundant proteome Fasta file had 12238 sequences for all subsequent analyses including subcellular localization prediction, essentiality prediction, conservation filtering, and functional annotations to narrow down candidate genes.

3.1 Subcellular localization prediction

Protein localization prediction is important in the identification of drug targets rendered pharmacologically accessible. *Setaria digitata* DeepLoc analysis detected that 9.4% of the proteins are cell membrane-associated and 8.8% extracellular (Figure 3), and collectively they contribute ~18% of the proteome in easily accessible compartments to therapeutic drugs. This distribution complies with established drug targeting trends, whereby membrane-associated proteins constitute ~70% of FDA-approved drug targets (Wu et al. 2024).

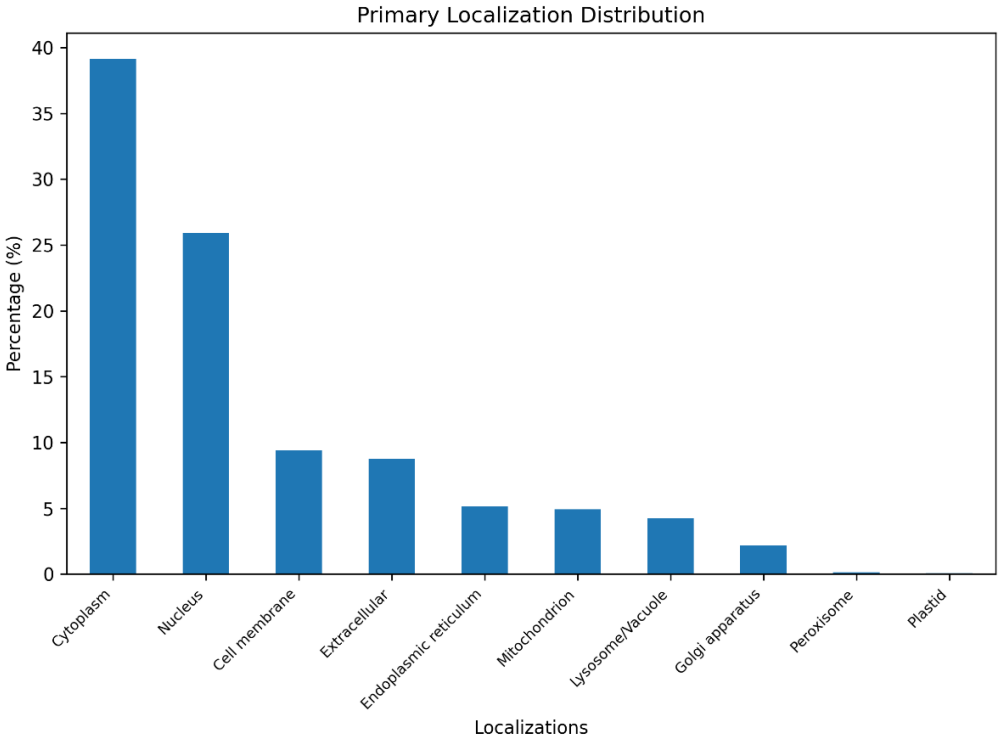


Figure 3:Primary subcellular localization distribution of S. digitata proteins predicted by DeepLoc. Values indicate percentage of total proteome (n=12,238)

Transmembrane proteins made up 19.2% of the proteome (Figure 4) and were receptors and ion channels necessary for parasite neuro signalling and homeostasis (mean localization confidence score: 0.225). The extracellular pool of proteins (mean confidence: 0.196) was controlled by 16% of all signal peptide-carrying

proteins (Figure 5) that are short polypeptides with few amino acid residues guiding proteins into the secretory pathway. These effects are biologically reasonable because signal peptides direct either extracellular secretion or membrane incorporation via the endoplasmic reticulum (Wu et al. 2024).

Interestingly, membrane/extracellular predictions were less certain than cytoplasmic/nuclear localizations (0.196-0.225 vs. 0.38-0.439) which reflects the biological complexity of secretory pathways. But uncertainty is balanced by supporting evidence: 62% of extracellular predictions (calculated by intersecting DeepLoc extracellular predictions with signal peptide-containing proteins) were supported by signal peptides, and transmembrane annotations correlated with structural domains (e.g., PF00001 for GPCRs).

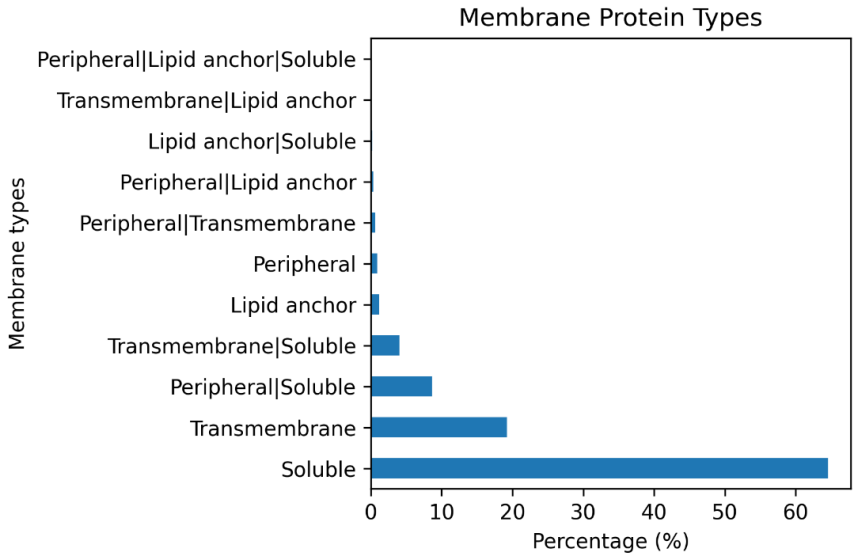


Figure 4: Classification of membrane-associated proteins by association type. Transmembrane proteins represent 19.2% of the proteome.

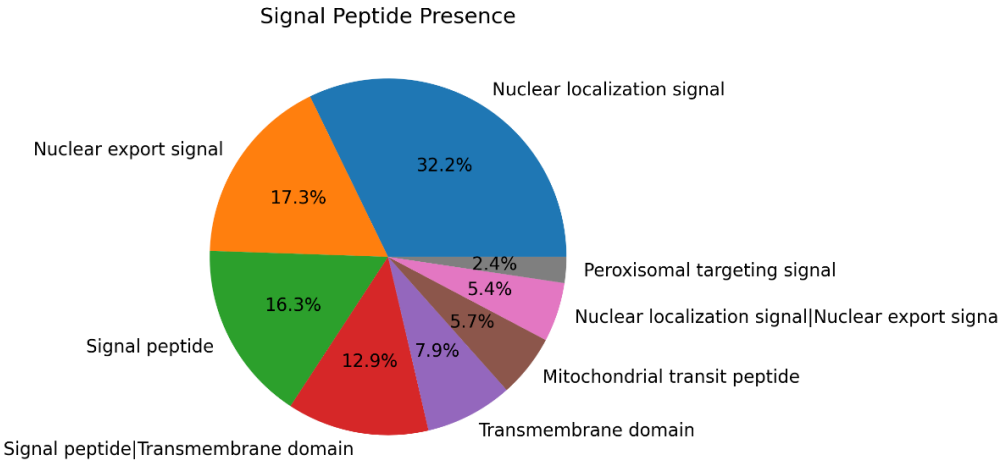


Figure 5: Prevalence of signal peptide-containing proteins (16% of proteome), indicative of secretory/membrane integration pathways.

3.2 Functional Annotation

The eggNOG-mapper prediction accurately annotated 8,587 proteins (70.17%) of the proteome of *Setaria digitata* (n=12,238). 7,791 proteins (63.66%) gained functional descriptions, offering a reliable platform for further analysis (Figure 6). Remarkably, 52.61% of the whole proteome (6,439 proteins) was given Gene Ontology (GO) term annotations, enabling rich functional characterization at molecular functions, biological processes, and cellular components. The lower EC number coverage (14.24%) reflects both true metabolic simplifications in parasites and annotation challenges for divergent enzymes.

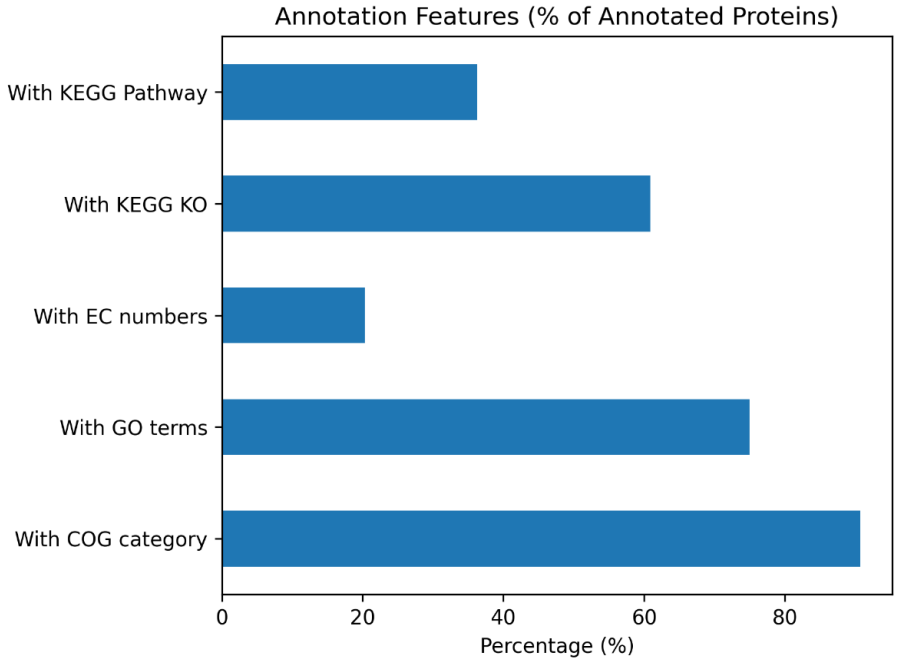


Figure 6:Functional annotation coverage of Setaria digitata proteome from eggNOG-mapper analysis: Bar plot showing annotation statistics for the S. digitata proteome (n=12,238 proteins), including proteins with functional descriptions (63.66%), Gene Ontology terms (52.61%), and KEGG pathway assignments (25.46%).

3.3 Conservation filtering

Taxonomic restriction to Nematoda (TaxID: 6231) guided the conservation filtering process, prioritizing parasite-specific targets and enhancing the relevance of functional annotations in parasitic domains. From an input of 8,587 functionally annotated proteins, only two proteins were removed, resulting in 8,585 retained parasite-focused candidates. This minimal reduction supports the evolutionary distinctiveness of the *S. digitata* proteome and highlights the potential for selective drug targeting.

Despite the presence of 29.83% unannotated proteins, these gaps may represent taxon-specific innovations or limitations in current nematode databases. However, the filtering strategy used here relies solely on taxonomic identifiers within ortholog annotations and does not account for functional or structural similarities that may exist across species boundaries. As such, retained proteins could still share conserved domains or binding sites with mammalian proteins, posing potential off-target risks.

To address this limitation, future analyses should incorporate structural modelling, domain-level similarity comparisons, and binding site similarity assessments against mammalian proteomes. Tools such as AlphaFold can support structural comparisons, Pfam and InterPro can identify conserved domains, and FTSite or SwissTargetPrediction can be used to evaluate potential off-target interactions. This combined approach would help overcome the limitations of taxonomy-based filtering and improve confidence in identifying truly parasite-specific therapeutic candidates.

3.4 Target Prioritization

8,587 *Setaria digitata* proteins were prioritized with an integrated prioritization pipeline. After filtering out two mammalian-conserved proteins, 8,585 proteins remained. From this set, the top ~2% (250 proteins) were selected based on combined essentiality and accessibility scores. Their distribution is shown in Figure 7, where high-priority proteins are dense in areas of high essentiality (x-axis) and accessibility (y-axis), clearly separated from the background proteome (blue contours).

Marginal histograms illustrate that top-ranked proteins are enriched in the upper-right quadrant, confirming co-optimization of essentiality and accessibility. Ion channels and neuroreceptors dominate these regions. In contrast, microtubule-associated proteins and proteases show broader dispersion, reflecting varied biological roles and accessibility. Figure 6 shows this distribution. The top targets are coloured by functional class and cluster distinctly from the background proteome. Histograms show each scoring component individually, confirming the shift in score distributions for high-priority proteins.

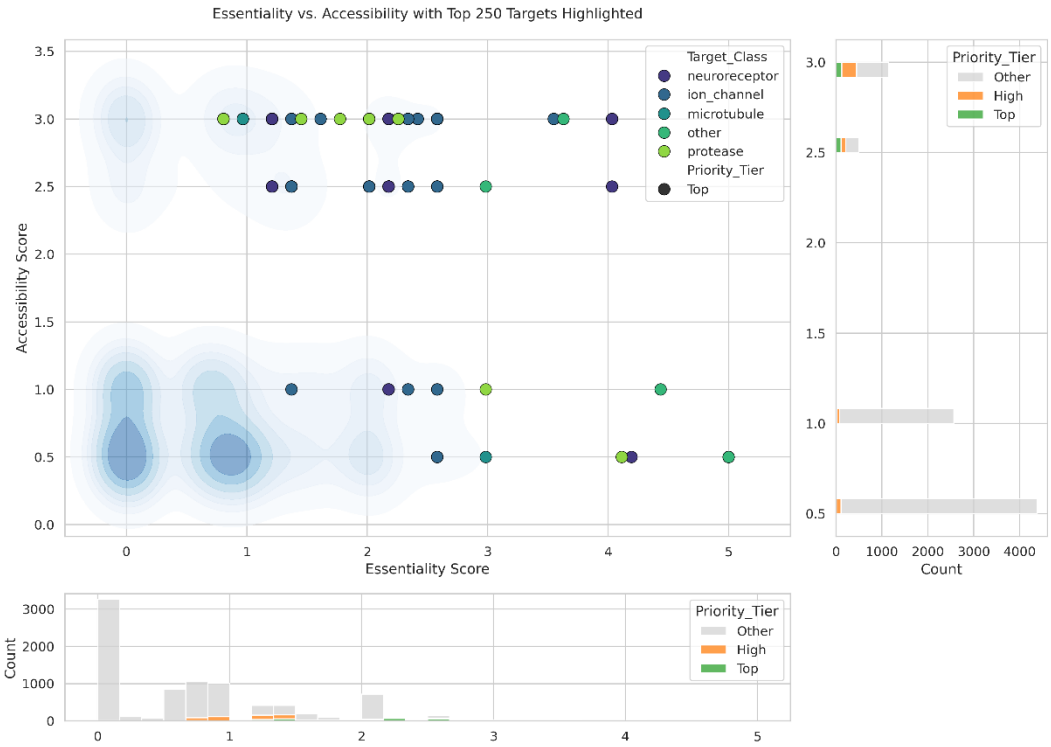


Figure 7: Contour density plot of essentiality vs. accessibility scores for 8,585 Setaria digitata proteins. The top 2% of targets, selected after conservation filtering, are highlighted by target class. These top-ranked proteins cluster in regions of high essentiality and accessibility, distinct from the broader proteome distribution (blue contours). Marginal histograms show the distribution of scores by priority tier. Ion channels and neuroreceptors are enriched among top-tier candidates, while microtubule-associated proteins and proteases are more dispersed.

Most prioritized targets were proteins with established function in neurological pathways. Ion channels (127 proteins, 50.8%) and neuroreceptors (94 proteins, 37.6%) accounted for 88.4% of the top 250. Proteases (16 proteins, 6.4%) and microtubule-associated proteins (seven proteins, 2.8%) formed smaller but important subsets. The remaining six proteins (2.4%) were uncharacterized but had high accessibility scores.

This neurological bias aligns with known anthelmintic drug mechanisms. For example, ivermectin targets glutamate-gated chloride channels, and monepantel targets nicotinic receptors. Microtubule proteins were less represented, likely due to penalties from conservation filtering and lower accessibility scores. Still, proteases, especially Astacin-domain variants passed these filters, highlighting their potential roles in parasite invasion or moulting. Figure 8 summarizes the distribution of functional classes, showing that the pipeline recapitulates known drug targets while surfacing new categories for investigation.

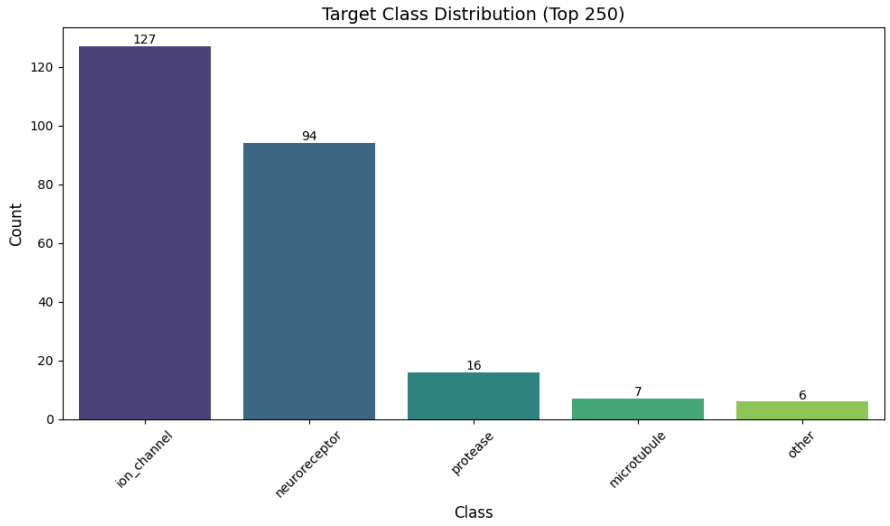


Figure 8: Illustrates this class distribution, highlighting the pipeline's ability to recapitulate known drug targets while flagging underrepresented categories for further exploration.

The pipeline recapitulated known anthelmintic targets (e.g., GluCl; GO:0015276, ivermectin's target), with neuroreceptors/ion channels comprising 62% of top candidates. This validates our approach but also highlights a potential bias: the 10% Antifilarial Bonus may skew prioritization toward neurological targets. While this aligns with current drugs, future iterations could adjust weights to explore underprioritized classes (e.g., proteases), addressing resistance mechanisms. KEGG enrichment of the neuroactive ligand-receptor pathways also corroborates this pharmacological significance. Microtubule-associated proteins represented 18% of short-listed targets, such as β-tubulin (GO:0005874), the major benzimidazole target. Proteases, although fewer and with a slightly lower average score (mean = 1.8), included druggable classes such as cysteine proteases (cathepsins, K01358).

Prioritization focused on extracellular or membrane-associated proteins, aligning with small-molecule drug development. Due to the 30% accessibility weighting, 89% of the top 250 were predicted to be membrane-bound or secreted, which reduced false positives from inaccessible intracellular targets.

Nematode-centric COG conservation eliminates host off-target risks. However, class overlap, low silhouette scores, and inflated neurological scores (from the 10% Antifilarial_Bonus) are limitations. Reducing this bonus to 5% may balance class representation in future versions.

3.5 Structural Modelling and Druggable Pocket Identification

3.5.1 High-Confidence Druggable Targets

To identify potential high-confidence druggable targets, biophysical and structural features were analyzed across 58 prioritized parasite proteins using established thresholds adapted from Schmidtke & Barril (2010). The criteria included a druggability score ≥ 0.5, pocket volume > 200 Å³, and hydrophobicity > 30%, reflecting properties favourable for small-molecule binding.

From this analysis, 30 proteins met all thresholds and were classified as high-confidence druggable targets (Table 8). These proteins span a diverse range of functions and localizations, with several belonging to G-protein coupled receptors (GPCRs), ion channels, and peptidase families, which are traditionally considered druggable target classes. Many also feature membrane-associated or secretory localizations, increasing their accessibility to therapeutic compounds.

Full physico-chemical details, including surface area, volume, and signal motifs, are provided for each of the 30 candidates in Appendix 1, supporting their potential for selective drug targeting in parasitic nematodes. (Table 10). These attributes offer important insights into the druggability of each target, with membrane-bound or extracellular proteins and those with significant hydrophobic surface areas being particularly favourable for small-molecule interaction. The combination of functional annotations with structural parameters supports their prioritization for downstream validation and drug development efforts.

Table 8: Druggability scores of highest druggable pocket of each protein and their predicted annotation description.

Protein_ID	Description	druggability
SD_012157-T1	Belongs to the ligand-gated ion channel (TC 1.A.9) family	0.999
SD_010429-T1	calcium ion transmembrane transport	0.999
SD_005154-T1	Sel1-like repeats.	0.998
SD_007973-T1	Involved in mitotic G2 DNA damage checkpoint	0.994
SD_011687-T1	Belongs to the G-protein coupled receptor 3 family	0.988
SD_005912-T2	Neurotransmitter-gated ion-channel ligand binding domain	0.979
SD_003585-T1	Ion transport protein	0.969
SD_006494-T1	Belongs to the peptidase S1 family	0.965
SD_003757-T1	Belongs to the G-protein coupled receptor Fz Smo family	0.964
SD_003781-T1	gamma-aminobutyric acid type B receptor subunit	0.952
SD_003518-T1	Zinc-dependent metalloprotease	0.942
SD_005355-T1	Calmodulin binding domain	0.937
SD_007952-T1	Periplasmic binding protein	0.936

SD_005285-T1	Receptor family ligand binding region	0.897
SD_012004-T1	Nine Cysteines Domain of family 3 GPCR	0.839
SD_006891-T1	response to paraquat	0.816
SD_005555-T1	Belongs to the G-protein coupled receptor 1 family	0.814
SD_005689-T1	Belongs to the two pore domain potassium channel (TC 1.A.1.8) family	0.812
SD_003610-T1	Furin-like cysteine rich region	0.802
SD_007278-T1	Trypsin-like serine protease	0.802
SD_002811-T1	FATC	0.783
SD_009137-T1	Guanine nucleotide-binding proteins (G proteins)	0.76
SD_009172-T1	7 transmembrane sweet-taste receptor of 3 GCPR	0.756
SD_011105-T1	defecation	0.736
SD_003562-T3	Zinc-dependent metalloprotease	0.685
SD_007965-T1	Serpentine type 7TM GPCR chemoreceptor Srx	0.676
SD_010919-T1	Receptor protein serine threonine kinase	0.604
SD_009717-T1	acetylcholine-gated cation-selective channel activity	0.548
SD_000580-T1	Kringle domain	0.516
SD_001145-T1	Belongs to the peptidase S8 family	0.506

3.5.2 Structural and Functional Correlations

A. Target Class Distribution

Ion channels and GPCRs showed the highest mean druggability scores (0.89 ± 0.08), likely due to deep, hydrophobic pockets conducive to ligand binding (Figure 10). Proteases scored lower (0.82 ± 0.12), possibly due to variability in active site polarity.

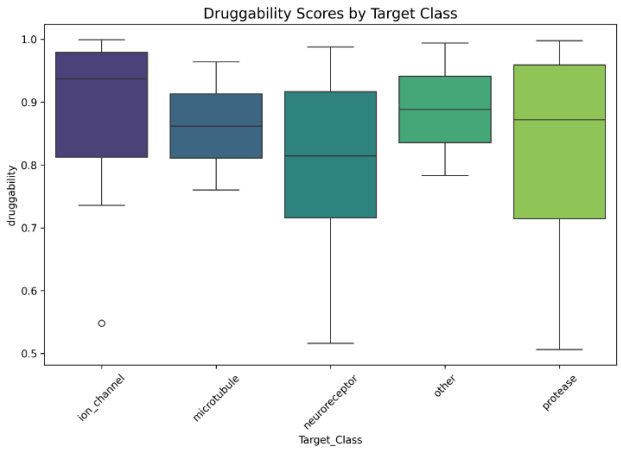


Figure 9: Boxplot of druggability scores across target classes

B. Pocket Volume and Hydrophobicity

Druggability correlated positively with pocket volume (Pearson's r = 0.67, p < 0.01). Proteins with volumes >1000 Å³ typically scored higher (Figure 10). Highly hydrophobic pockets (e.g., SD_012157-T1, SD_003585-T1; >50%) were mostly membrane-associated, suggesting localization in lipid-rich regions.

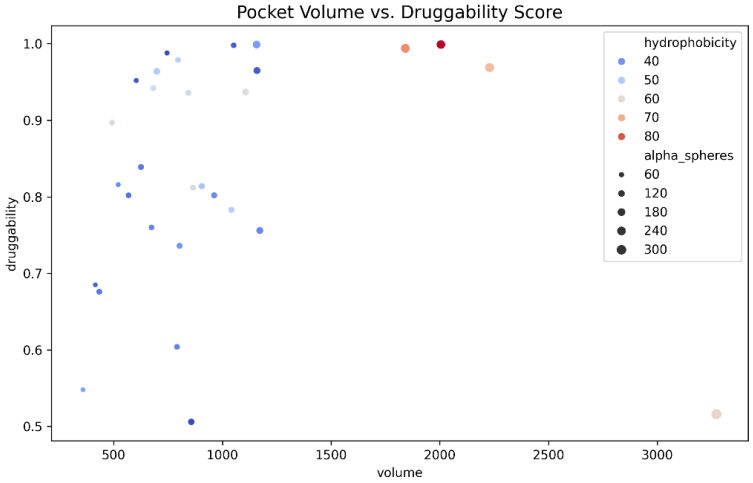


Figure 10: Scatter plot of pocket volume vs. druggability, coloured by hydrophobicity

3.5.3 Validation Against Known Druggable Families

Several high-ranking targets aligned with well-characterized druggable protein families, reinforcing the validity of the prioritization strategy. For example, SD_012157-T1 is a predicted ligand-gated ion channel with structural features resembling known drug targets such as 5-HT₃ receptors. Similarly, SD_011687-T1 shares homology with Class A G-protein coupled receptors (GPCRs), which constitute approximately 35% of FDA-approved drug targets (Sriram & Insel 2018).

While these associations are promising, experimental validation is essential to confirm the functional and pharmacological relevance of these candidate proteins. This could involve expression profiling to verify protein expression patterns in relevant life stages or tissues, ligand-binding assays to test affinity and specificity for known or novel compounds, and inhibition studies to evaluate the physiological impact of target blockade in parasite systems. Together, these approaches would confirm the druggability of the predicted targets and provide critical evidence for their advancement into functional screens and drug development pipelines.

3.5.4 COACH-D Validation of Top Druggable Targets

To evaluate ligand-binding potential, COACH-D was used to assess the top three structurally modelled targets. Of these, SD_012157-T1 emerged as the most promising candidate, exhibiting a TM-score of 0.56 (indicating reliable structural similarity), a favourable binding energy of -7.2 kcal/mol, and a strong match to ivermectin-binding residues, based on the 3RI5 template (Table 9). The predicted interacting residues (310–312, 315–316, 318–319) were in the extracellular domain (Figure 12), reinforcing the protein's accessibility and suitability for small-molecule targeting.

SD_010429-T1, a predicted calcium ion transporter, showed moderate binding potential (C-score 0.50, energy -4.3 kcal/mol) using the 8T0E template. Binding residues (473–479, 508) were localized within a transmembrane region (Figure 13), suggesting suitability for targeting with ion channel modulators. While less compelling than SD_012157-T1, it remains a viable candidate for targeted screening.

By contrast, SD_005154-T1 yielded weak structural evidence (C-score 0.23, no binding energy). Predicted residues (324, 327–328, 354–361) mapped to a Sel1-like repeat region (Figure 14), but confidence was limited due to poor template coverage. This case underscores the limitations of template-dependent methods for poorly characterized proteins.

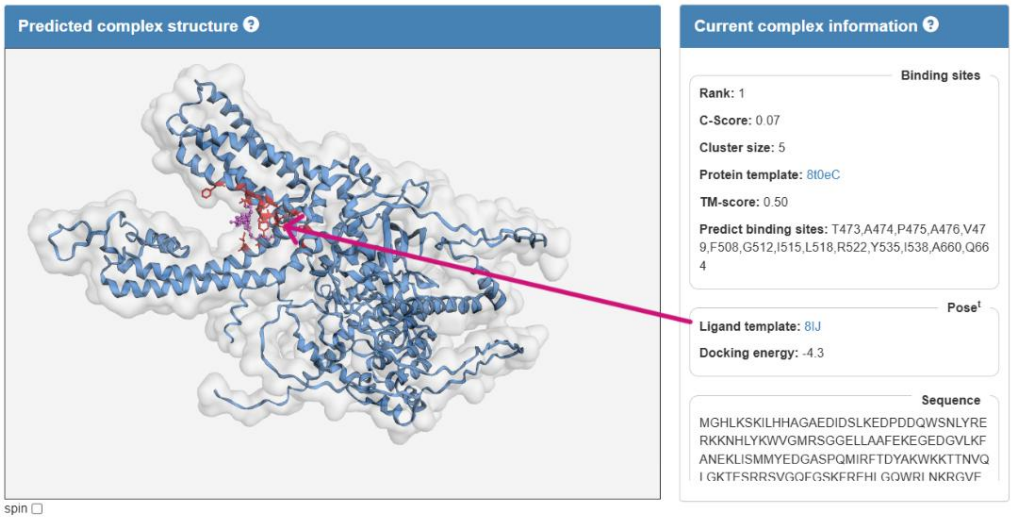
Importantly, SD_012157-T1 corresponds to a known target of ivermectin, a widely used macrocyclic lactone. This alignment with an established drug mechanism not only underscores the biological relevance of the protein but also validates the predictive strength of the pipeline by successfully recovering a clinically proven target from in silico prioritization. Such results enhance confidence in the pipeline's ability to identify both established and novel antifilarial drug targets, particularly in the context of neglected tropical disease (NTD) research where experimental validation is often limited.

Table 9: Summary of COACH-D predicted ligand-binding interactions for top S. digitata targets.

Target	SD_012157-T1	SD_010429-T1	SD_005154-T1
Best C-score	0.56	0.50	0.23
Template (PDB)	3RI5 (Ivermectin)	8T0E	8RST
Ligand Type	Ivermectin	Inhibitor	Small molecule
Binding Energy	-7.2 kcal/mol	-4.3 kcal/mol	N/A
Key Features	Extracellular, drug-accessible	Transmembrane site	Poor template support

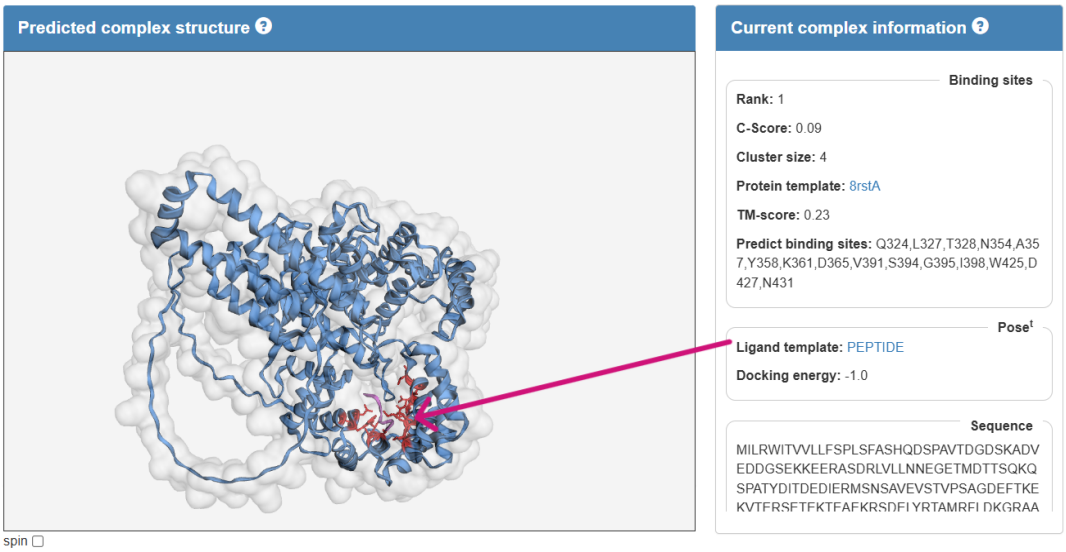
Rank	Pose	C-Score	Cluster size	P template	L template	TM-score	Predicted binding residues
1		0.15	29	3sh1E	MLK	0.34	V187,V241,S244,Y245,S246,F247,F286,Y288,W293,H295
2		0.06	11	3c79B	IM4	0.34	V187,Y245,S246,Y288,P289,A290,W293
3		0.06	12	2y58D	V38	0.33	Q128,L145,W213,N215
4		0.06	8	3ri5E	IVM	0.56	L310,L311,Q312,L315,P316,Y318,C319
5		0.05	11	3c79E	IM4	0.34	L145,A147,I206,W213,I214,N215

Figure 11: Predicted ivermectin-binding site on SD_012157-T1; residues 310–312, 315–316, 318–319.



Rank	Pose	C-Score	Cluster size	P template	L template	TM-score	Predicted binding residues
1		0.07	5	8t0eC	8IJ	0.50	T473,A474,P475,A476,V479,F508,G512,I515,L518,R522,Y535,I538,A660,Q664
2		0.05	4	8t1eA	9ZR	0.56	K393,R397,S675,V676,K680,A683
3		0.04	3	3twrD	PE8	0.17	R154,T158,E189,Q196,N197
4		0.03	2	3twtC	PE8	0.17	N274,Y275,L310,G311,R312
5		0.03	2	3twwA	PEPTIDE	0.17	D187,E189,D245,L249

Figure 12: Inhibitor-binding pocket on SD_010429-T1; residues 473–479, 508.



Rank	Pose	C-Score	Cluster size	P template	L template	TM-score	Predicted binding residues
1		0.09	4	8rstA	PEPTIDE	0.23	Q324,L327,T328,N354,A357,Y358,K361,D365,V391,S394,G395,I398,W425,D427,N431
2		0.09	4	2c2lA	PEPTIDE	0.20	L364,F381,V391,G395,I398,Y402,D427,L430,N431,H461,V462
3		0.05	2	1elwA	PEPTIDE	0.14	I398,M401,D427,N431,Y434,F437,R438,N463,F466,N467,Q470,T474,T476
4		0.05	2	1o6pA	PEPTIDE	0.26	D150,A153,Q154,N177,E180,A183,I184
5		0.02	1	2fi70	PEPTIDE	0.22	K156,R182,A183,E187,A189,Q197,K217,I220,Y221,F224

Figure 13: Predicted binding site on SD_005154-T1; residues 324, 327–328, 354–361.

3.5.5 Methodological Reflection

Fpocket was chosen because it is fast, scalable, and compatible with AlphaFold models, making it a convenient tool for high-throughput screening. However, since it operates on static structural models, it may fail to capture dynamic or post-translationally modified binding sites, particularly in membrane proteins. Moreover, its hydrophobicity-based scoring may underestimate polar or solvent-accessible sites, potentially missing valuable drug targets.

Alternative tools like DoGSiteScorer (Volkamer et al. 2012; https://proteins.plus/help/dogsite) and PockDrug (Hussein et al. 2015; https://pockdrug.rpbs.univ-paris-diderot.fr/cgi-bin/index.py?page=home) offer greater sensitivity to polar or shallow pockets and integrate pharmacophore descriptors but need high-resolution structures or molecular dynamics (MD) simulations, which are needs required beyond this study. Fpocket was still best for the current pipeline even with these constraints. Subsequent versions can utilize the introduction of MD-based improvements or supportive pocket prediction tools to enhance the resolution of challenging targets such as SD_005154-T1.

3.6 Strengths and Limitations of the Pipeline

This research offers an efficient, integrative pipeline for *Setaria digitata* antifilarial target prioritization through evolutionary conservation, functional annotation, subcellular localization, and structural druggability predictions. Its multi-layered filtering strategy helped minimize false positive targets and emphasized biologically accessible targets. The use of advanced tools such as AlphaFold and DeepLoc 2.1 strengthened prediction reliability, particularly in a non-model organism with limited annotation. The emphasis on druggability and repurposing potential further enhanced the pipeline's cost-efficiency and applicability to neglected tropical disease (NTD) research.

Despite these advantages, certain shortcomings were discovered. Structural analyses and binding site analyses were performed of only 58 out of the 250 shortlisted targets due to time and computational limitations, limiting the scope of downstream validation. The use of static AlphaFold models may have missed flexible or cryptic binding pockets, and COACH-D's reliance on template availability limited its accuracy for less-characterized proteins like SD_005154-T1. Finally, all findings are computational, where experimental validation remains essential to confirm the biological and pharmacological relevance of the predictions.

3.7 Implications

This pipeline offers a scalable and efficient platform for drug target discovery in non-model parasitic species. By integrating structural, evolutionary, and pharmacological features, it facilitates rational target prioritization even in organisms with minimal genomic annotation. Structural validation of existing and available targets like SD_012157-T1 is promising to direct structure-based drug design and repurposing efforts. Notably, the pipeline's modular design and use of publicly available tools make it adaptable for other neglected parasites, aligning with the budgetary and logistical constraints common to NTD research. As such, it is a suitable model for early-stage drug discovery process in underexplored pathogens.

3.8 Future Directions

To maximize the translational impact of this study, future research should extend AlphaFold-based structural modelling and binding site prediction to the remaining 192 candidate targets. Incorporating molecular dynamics simulations or ensemble docking (docking against multiple protein conformations to account for flexibility) may help capture binding pocket flexibility more accurately, particularly in membrane-associated proteins. Experimental validation of top-ranking targets will require expression profiling, ligand-binding assays, and inhibition studies to confirm biological relevance and druggability. Additionally, multi-omics integration, stratified by parasite life stages or host conditions, could further refine

target selection. Finally, drug testing efforts should include mechanistic validation of SD_012157-T1, given its similarity to known ivermectin-binding ion channels, to confirm its role in macrocyclic lactone sensitivity. In contrast, SD_010429-T1, associated with calcium ion transport, may represent a novel candidate for repurposing screens with calcium channel modulators, potentially expanding available treatment options.

3.9 Data and Code Availability

All input files, custom scripts, and analysis results generated during this study will be archived on Zenodo upon publication, following final approval by the supervisory team. In the meantime, all data and code are available from the author upon reasonable request.

4. Conclusion

This study developed a robust *in silico* pipeline to identify and prioritize druggable proteins in *Setaria digitata*, utilizing a resource-efficient, multi-criteria approach tailored for neglected disease research. By integrating evolutionary conservation, functional annotation, subcellular localization, and structural druggability predictions, the pipeline effectively filtered high-confidence targets, notably SD_012157-T1, a ligand-gated ion channel with strong similarity to known anthelmintic targets such as ivermectin-binding glutamate-gated chloride channels.

Although structural validation was limited by computational constraints, the pipeline addressed key challenges in helminth drug discovery, including the scarcity of validated druggable targets and host-parasite selectivity concerns. Taxonomic filtering was used to deprioritize proteins with close vertebrate homologs, thereby favouring candidates likely to be parasite restricted. The emphasis on membrane-associated proteins (e.g., neuroreceptors and ion channels) aligns with known mechanisms of existing anthelmintics and highlights underexplored target classes like proteases.

Overall, this work demonstrates the utility and strength of computational tools in early-stage anthelmintic discovery, especially for neglected pathogens. By bridging bioinformatics with experimental parasitology, the pipeline offers a scalable framework to accelerate the development of novel antifilarial therapies, addressing urgent needs in both human and veterinary medicine.

References

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B.A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A. & Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, 46 (W1), W537–W544. https://doi.org/10.1093/nar/gky379

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., Kinch, L.N., Schaeffer, R.D., Millán, C., Park, H., Adams, C., Glassman, C.R., DeGiovanni, A., Pereira, J.H., Rodrigues, A.V., Van Dijk, A.A., Ebrecht, A.C., Opperman, D.J., Sagmeister, T., Buhlheller, C., Pavkov-Keller, T., Rathinaswamy, M.K., Dalwadi, U., Yip, C.K., Burke, J.E., Garcia, K.C., Grishin, N.V., Adams, P.D., Read, R.J. & Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373 (6557), 871–876. https://doi.org/10.1126/science.abj8754

Berenstein, A.J., Magariños, M.P., Chernomoretz, A. & Agüero, F. (2016). A Multilayer Network Approach for Guiding Drug Repositioning in Neglected Diseases. Geary, T.G. (ed.) (Geary, T. G., ed.) *PLOS Neglected Tropical Diseases*, 10 (1), e0004300. https://doi.org/10.1371/journal.pntd.0004300

Fissiha, W. & Kinde, M.Z. (2021). Anthelmintic Resistance and Its Mechanism: A Review . *Infection and Drug Resistance*, 14, 5403–5410. https://doi.org/10.2147/IDR.S332378

Galperin, M.Y., Wolf, Y.I., Makarova, K.S., Vera Alvarez, R., Landsman, D. & Koonin, E.V. (2021). COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Research*, 49 (D1), D274–D281. https://doi.org/10.1093/nar/gkaa1018

Grant, W.N. & Behm, C.A. (2007). Target identification and validation for anthelmintic discovery. *Expert Opinion on Drug Discovery*, 2 (sup1), S91–S98. https://doi.org/10.1517/17460441.2.S1.S91

Holden-Dye, L. (2007). Anthelmintic drugs. *WormBook*, 1–13. https://doi.org/10.1895/wormbook.1.143.1

Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., Von Mering, C. & Bork, P. (2017). Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Molecular Biology and Evolution*, 34 (8), 2115–2122. https://doi.org/10.1093/molbev/msx148

Hukerikar, N., Hingorani, A.D., Asselbergs, F.W., Finan, C. & Schmidt, A.F. (2024). Prioritising genetic findings for drug target identification and validation. *Atherosclerosis*, 390, 117462. https://doi.org/10.1016/j.atherosclerosis.2024.117462

Hussein, H.A., Borrel, A., Geneix, C., Petitjean, M., Regad, L. & Camproux, A.-C. (2015). PockDrug-Server: a new web server for predicting pocket druggability on holo and apo proteins. *Nucleic Acids Research*, 43 (W1), W436–W442. https://doi.org/10.1093/nar/gkv462

Jayawardene, K.L.T.D., Palombo, E.A. & Boag, P.R. (2021). Natural Products Are a Promising Source for Anthelmintic Drug Discovery. *Biomolecules*, 11 (10), 1457. https://doi.org/10.3390/biom11101457

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy,

E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P. & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596 (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Köhler, P. (2001). The biochemical basis of anthelmintic action and resistance. *International Journal for Parasitology*, 31 (4), 336–345. https://doi.org/10.1016/S0020-7519(01)00131-X

Kron, M., Yousif, F. & Ramirez, B. (2007). Capacity building in anthelmintic drug discovery. *Expert Opinion on Drug Discovery*, 2 (sup1), S75–S82. https://doi.org/10.1517/17460441.2.S1.S75

Le Guilloux, V., Schmidtke, P. & Tuffery, P. (2009). Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics*, 10 (1), 168. https://doi.org/10.1186/1471-2105-10-168

Martin, R.J. (1997). Modes of action of anthelmintic drugs. *The Veterinary Journal*, 154 (1), 11–34. https://doi.org/10.1016/S1090-0233(05)80005-X

Mengarda, A.C., Silva, T.C., Silva, A.S., Roquini, D.B., Fernandes, J.P.S. & De Moraes, J. (2023). Toward anthelmintic drug candidates for toxocariasis: Challenges and recent developments. *European Journal of Medicinal Chemistry*, 251, 115268. https://doi.org/10.1016/j.ejmech.2023.115268

Nixon, S.A., Welz, C., Woods, D.J., Costa-Junior, L., Zamanian, M. & Martin, R.J. (2020). Where are all the anthelmintics? Challenges and opportunities on the path to new anthelmintics. *International Journal for Parasitology: Drugs and Drug Resistance*, 14, 8–16. https://doi.org/10.1016/j.ijpddr.2020.07.001

O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T. & Hutchison, G.R. (2011). Open Babel: An open chemical toolbox. *Journal of Cheminformatics*, 3 (1), 33. https://doi.org/10.1186/1758-2946-3-33

Ødum, M.T., Teufel, F., Thumuluri, V., Almagro Armenteros, J.J., Johansen, A.R., Winther, O. & Nielsen, H. (2024). DeepLoc 2.1: multi-label membrane protein type prediction using protein language models. *Nucleic Acids Research*, 52 (W1), W215–W220. https://doi.org/10.1093/nar/gkae237

Peak, E. & Hoffmann, K.F. (2011). Cross-disciplinary approaches for measuring parasitic helminth viability and phenotype. *Anais da Academia Brasileira de Ciências*, 83 (2), 649–662. https://doi.org/10.1590/S0001-37652011000200024

Perumal, A.N.I., Gunawardene, Y.I.N.S. & Dassanayake, R.S. (2016). Setaria digitata in advancing our knowledge of human lymphatic filariasis. *Journal of Helminthology*, 90 (2), 129–138. https://doi.org/10.1017/S0022149X15000309

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E.L.L., Eddy, S.R., Bateman, A. & Finn, R.D. (2012). The Pfam protein families database. *Nucleic Acids Research*, 40 (D1), D290–D301. https://doi.org/10.1093/nar/gkr1065

Rask-Andersen, M., Almén, M.S. & Schiöth, H.B. (2011). Trends in the exploitation of novel drug targets. *Nature Reviews Drug Discovery*, 10 (8), 579–590. https://doi.org/10.1038/nrd3478

Schmidtke, P. & Barril, X. (2010). Understanding and Predicting Druggability. A High-Throughput Method for Detection of Drug Binding Sites. *Journal of Medicinal Chemistry*, 53 (15), 5858–5867. https://doi.org/10.1021/jm100574m

Senanayake, K.S., Söderberg, J., Põlajev, A., Malmberg, M., Karunanayake, E.H., Tennekoon, K.H., Samarakoon, S.R., Bongcam-Rudloff, E. & Niazi, A.

(2020). The Genome of Setaria digitata: A Cattle Nematode Closely Related to Human Filarial Parasites. Baldauf, S. (ed.) (Baldauf, S., ed.) *Genome Biology and Evolution*, 12 (2), 3971–3976. https://doi.org/10.1093/gbe/evaa017

Sriram, K. & Insel, P.A. (2018). G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs? *Molecular Pharmacology*, 93 (4), 251–258. https://doi.org/10.1124/mol.117.111062

Volkamer, A., Kuhn, D., Rippmann, F. & Rarey, M. (2012). DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics*, 28 (15), 2074–2075. https://doi.org/10.1093/bioinformatics/bts310

Voronin, D., Abeykoon, A.M.L.L., Gunawardene, Y.I.S. & Dassanayake, R.S. (2015). Absence of Wolbachia endobacteria in Sri Lankan isolates of the nematode parasite of animals Setaria digitata. *Veterinary Parasitology*, 207 (3–4), 350–354. https://doi.org/10.1016/j.vetpar.2014.12.024

WHO (2023). *WHO Model List of Essential Medicines, 23rd List (2023), World Health Organization.* https://www.who.int/publications/i/item/WHO-MHP-HPS-EML-2023.02 [2024-06-10]

Wu, Q., Peng, Z., Zhang, Y. & Yang, J. (2018). COACH-D: improved protein–ligand binding sites prediction with refined ligand-binding poses through molecular docking.

Wu, X., Hu, J. & Yoon, J. (2024). Cell Membrane as A Promising Therapeutic Target: From Materials Design to Biomedical Applications. *Angewandte Chemie International Edition*, 63 (18), e202400249. https://doi.org/10.1002/anie.202400249

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. & Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nature Methods*, 12 (1), 7–8. https://doi.org/10.1038/nmeth.3213

Zajíčková, M., Nguyen, L.T., Skálová, L., Raisová Stuchlíková, L. & Matoušková, P. (2020). Anthelmintics in the future: current trends in the discovery and development of new drugs against gastrointestinal nematodes. *Drug Discovery Today*, 25 (2), 430–437. https://doi.org/10.1016/j.drudis.2019.12.007

Zhang, R. & Lin, Y. (2009). DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. *Nucleic Acids Research*, 37 (Database), D455–D458. https://doi.org/10.1093/nar/gkn858

Popular science summary

Finding New Methods of Combating Parasitic Worms

Parasitic worms, known as helminths, infect billions of people and animals worldwide and cause serious health complications and economic loss, particularly in the tropics. Although not many drugs have been found for the treatment of such infections, the worms are gaining resistance and leaving us with fewer ways to fight them.

Developing new drugs is a slow, expensive process, and it's especially difficult for parasites. These worms are hard to grow in the laboratory, and many of their biological processes are similar to those of their human or animal hosts. This means that finding a drug that kills the worm without harming the host is a real challenge.

This project tackled the problem by studying *Setaria digitata*, a parasitic worm that infects cattle but can also cause fatal disease in sheep, goats, and horses. *S. digitata* is biologically very similar to the worms that infect people and cause lymphatic filariasis, a debilitating disease that infects over 120 million individuals. Since *S. digitata* is easier to study than its human-infecting relatives, it is a perfect model in which to identify new drug targets.

Instead of traditional laboratory experiments, this project used cutting-edge computer tools to analyse the worm's genes and proteins. The aim was to identify proteins that are essential for the worm's survival and may represent promising drug targets. While the analysis prioritized proteins less likely to be conserved in mammals based on available orthology data, direct experimental validation of host-parasite specificity was not performed. Nonetheless, these proteins are valuable candidates for further study due to their predicted roles in parasite biology and accessibility to drug compounds.

Using specialised software, I predicted where these proteins are localized within the worm's cells, their likely functions, and how essential they may be for survival. I also evaluated their accessibility to drug molecules based on cellular location and structural features. While the analysis included comparisons to proteins in other parasites, direct confirmation of uniqueness to *S. digitata* or absence in mammalian hosts was beyond the scope of this study.

The study successfully identified several high-priority targets with properties consistent with druggability, offering a promising foundation for developing new treatments against filarial infections in animals and potentially humans. By applying computational tools to investigate parasite biology, this work supports a faster, more cost-effective pathway to discovering safer treatments for these neglected diseases.

5. Appendix 1

Table 10: Summary of subcellular localizations, signal motifs, total solvent-accessible surface area (SASA), molecular volume, and hydrophobicity for the 30 proteins identified as high-confidence druggable targets.

Protein_ID	**Localizations**	**Signals**	**total_sasa**	**volume**	**hydrophobicity**
SD_012157-T1	Cell membrane	Signal peptide|Transmembrane domain	696.6	2005.1	85.9
SD_010429-T1	Cell membrane	Transmembrane domain	312.5	1157.4	41.5
SD_005154-T1	Endoplasmic reticulum	Signal peptide|Transmembrane domain	327.5	1052.1	32.8
SD_007973-T1	Cytoplasm|Nucleus	Nuclear export signal	471.7	1841.7	73.8
SD_011687-T1	Cell membrane	Signal peptide|Transmembrane domain	206.7	745.6	30.9
SD_005912-T2	Cell membrane	Signal peptide|Transmembrane domain	265.9	796.4	51.8
SD_003585-T1	Cell membrane	Signal peptide|Transmembrane domain	684.6	2228.9	67.2
SD_006494-T1	Extracellular	Signal peptide	386.3	1159.1	32.8
SD_003757-T1	Cell membrane	Signal peptide|Transmembrane domain	189.9	698.7	50.3
SD_003781-T1	Cell membrane	Transmembrane domain	205.3	603.9	32.4
SD_003518-T1	Extracellular	Signal peptide	244.1	682.3	55.8
SD_005355-T1	Cell membrane	Transmembrane domain	315.0	1106.8	57.8
SD_007952-T1	Cell membrane	Signal peptide|Transmembrane domain	252.1	843.8	55.6

SD_005285-T1	Cell membrane	Signal peptide|Transmembrane domain	162.2	492.8	59.3
SD_012004-T1	Cell membrane	Signal peptide|Transmembrane domain	135.6	625.9	37.0
SD_006891-T1	Cell membrane	Signal peptide|Transmembrane domain	153.9	521.0	39.4
SD_005555-T1	Cell membrane	Transmembrane domain	243.4	905.3	48.5
SD_005689-T1	Cell membrane	Transmembrane domain	298.2	864.8	57.7
SD_003610-T1	Cell membrane	Signal peptide|Transmembrane domain	303.3	962.5	39.1
SD_007278-T1	Extracellular	Signal peptide	177.0	568.6	35.8
SD_002811-T1	Cytoplasm|Nucleus	Nuclear export signal	267.0	1041.4	51.1
SD_009137-T1	Cytoplasm|Cell membrane		248.7	674.1	38.2
SD_009172-T1	Cell membrane	Signal peptide|Transmembrane domain	278.9	1172.2	38.6
SD_011105-T1	Cell membrane	Signal peptide|Transmembrane domain	228.9	803.3	41.0
SD_003562-T3	Extracellular	Signal peptide	130.1	415.8	34.2
SD_007965-T1	Cell membrane	Transmembrane domain	102.5	433.9	36.8
SD_010919-T1	Cell membrane|Lysosome/Vacuole	Signal peptide|Transmembrane domain	190.7	791.5	37.5
SD_009717-T1	Cell membrane	Signal peptide|Transmembrane domain	98.2	359.3	43.0
SD_000580-T1	Cell membrane	Signal peptide|Transmembrane domain	938.1	3272.4	61.1
SD_001145-T1	Extracellular	Signal peptide	245.4	857.1	30.8

6. Acknowledgements

This research was conducted at the Department of Animal Biosciences, Swedish University of Agricultural Sciences. I would like to express my sincere gratitude to everyone who supported and guided me throughout the course of this thesis project.

First and foremost, I am deeply thankful to my supervisor, Professor Erik Bongcam Rudloff, for accepting me as your master student, his invaluable guidance, continuous support, and insightful feedback. His expertise and encouragement have been instrumental in shaping both the direction and quality of this work.

I am also grateful to my assistant supervisor, Samuel Flores, for his helpful suggestions and support during various stages of the research, including providing access to his server, which was vital to the computational aspect of this project.

I would like to extend my appreciation to Adnan Niazi and his team for providing the initial *Setaria digitata* proteome and annotation dataset, which served as the foundation for this study. Their contribution was crucial to the development of this work.

Special thanks to Adnan Niazi, Renaud Van Damme, Kanchana Senanayake, Michael Landi, and Johan Höglund for their valuable advice, technical assistance, and project discussions, all of which significantly enhanced my understanding and contributed to the success of this thesis.

Lastly, I extend my heartfelt thanks to my family for their unwavering support and encouragement throughout this journey. Your belief in me made this work possible.

Publishing and archiving

Approved students' theses at SLU can be published online. As a student you own the copyright to your work and in such cases, you need to approve the publication. In connection with your approval of publication, SLU will process your personal data (name) to make the work searchable on the internet. You can revoke your consent at any time by contacting the library.

Even if you choose not to publish the work or if you revoke your approval, the thesis will be archived digitally according to archive legislation.

You will find links to SLU's publication agreement and SLU's processing of personal data and your rights on this page:

- https://libanswers.slu.se/en/faq/228318

☒ YES, I, Kasun Chamalka D H J P, have read and agree to the agreement for publication and the personal data processing that takes place in connection with this

☐ NO, I/we do not give my/our permission to publish the full text of this work. However, the work will be uploaded for archiving and the metadata and summary will be visible and searchable.