Construction of Metabolic Pathways based on Whole Genome Sequencing Reveals Laterally-Transferred Wolbachia-like Sequences in the Setaria Digitata Genome

Nadarajah Rashanthy
University of Colombo
Mahil Sharada Alwis Kothalawala
University of Colombo
Thamarahansi Shashiprabha Mugunamalwaththa
University of Colombo
Kasun de Zoysa
University of Colombo
Naduviladath Vishvanath Chandrasekharan
University of Colombo
Yasanthi Illika Nilmini Silva Gunawardene
University of Kelaniya
Prashanth Suravajhala
Birla Institute of Scientific Research
Ranil Samantha Dassanayake (rsdassanayake@chem.cmb.ac.lk)
University of Colombo

Research Article

Keywords: Setaria digitata, Nematodiasis, KEGG, BLAST2GO, Human Lymphatic Filariasis

Posted Date: June 11th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-579175/v1

Abstract

Setaria digitata is a *Wolbachia*-free filarial parasite that causes cerebrospinal nematodiasis in non-permissive hosts such as goats, sheep and horses leading to substantial economic losses in animal husbandry. Due to its similarity to *Wuchereria bancrofti*, primary causing agent of human lymphatic filariasis (HLF), *S. digitata* can be used as a model organism to study the biology of HLF. This study was mainly aimed to bring functional analysis of metabolic pathways in *S. digitata*. A draft genome of 78,774,594 bases making a total of 2,075 contigs was generated. 'BLAST2GO' functional annotation resulted in 28112 BLAST hits with an e-value lower than 1e-4 and a sequence similarity higher than 30%. Out of a total of 2075 contigs, 1280 contigs were used to generate a total of 6055 GO annotations at a mean level of 6.488 with standard deviation of 2.675. Overall, 89.1% of mapped reads were annotated by at least one of the three categories of the GO function classification. Moreover, 111 enzymes associated with 95 distinct metabolic pathways were identified. We suggest that *S. digitata* may have evolved its own sequences to code for haem, riboflavin, and FAD in the absence of *Wolbachia*.

1. Introduction

Setaria digitata is an ivory coloured, slimy filarial parasitic worm with a coiled tapering tail. It is classified under class Secernentea, order Spirurida and family *Setariidae. S. digitata* naturally resides in the peritoneal cavity of grazing hoofed animals. (Shin et al., 2002; Shiny et al., 2011) They cause cerebrospinal nematodiasis, a neuropathological disorder that causes dysfunction of the central nervous system leading to lumbar paralysis with eventual death of non-permissive domesticated hosts such as goats, sheep and horses. As a result, this causes substantial economic losses in animal husbandry in South East Asia and the Far East. (Wickramatunga et al., 2020) However, they are not parasitic in their natural hosts such as cattle and buffaloes, but may cause mild disease conditions like fibrinous peritonitis. Human infections have also been reported in many studies and these can cause allergic reactions, eye lesions, abscesses, enlarged lymph nodes and lung inflammation. (Taylor et al., 1999; Gunawardene et al., 2015; Sundar et al., 2015).

Most of the filarial nematodes are mutually associated with *Wolbachia.* Their endosymbiotic relationship carries the genes required for the metabolism of haem, riboflavin, FAD, glutathione and nucleotides whereas its filarial host does not (Werren et al., 2008; Fenn et al., 2006). Although from earlier studies on filarial genomes have already identified five biosynthetic pathways coded by *Wolbachia* that are involved in haem/riboflavin/FAD/glutathione and nucleotide synthesis, only a few of them are fully encoded by *Wolbachia*. For example, FAD and glutathione pathways are found to be complete in all nematode genomes, whereas riboflavin and haem biosynthesis pathways are missing which otherwise would allow them to be carried out by protein coding genes in the *Wolbachia* genome. Filarial nematodes also do not have most enzymes required for purine biosynthesis and the first enzyme required for pyrimidine synthesis. Therefore, filarial nematodes cannot synthesize nucleotides *de novo* (Desjardins et al., 2013). Nevertheless, Wolbachia has the complete *de novo* nucleotide synthesis pathway and thus supplies the host with nucleotides during periods of high metabolic demand (Darby et al., 2012). They also contain

purine-pyrimidine interconversion pathways. (Lee, 2002). As it is not clear on the endosymbiont relationships, whether or not the genes responsible for the biosynthetic pathways should be present in its genome remains to be ascertained. (Voronin et al., 2015). It was earlier shown that the *Wolbachia* free filarial nematode *Loa loa* have evolved their own DNA sequences to code for haem and riboflavin biosynthetic pathways. For some pathways, they have gained partial gene sequences from *Wolbachia*, indicating horizontal gene transfer (HGT) within *Loa loa* and *Wolbachia* at some point during the evolution. (Desjardins et al., 2013) Therefore, independent survival of *S. digitata* also can be explained by HGT events leading to coding of these pathways. (Mcnulty et al., 2010) .

The use of anti-*Wolbachia* chemotherapy against filarial parasites is a novel approach for filarial disease control and eradication. In anti-*Wolbachia* chemotherapy, development, reproduction, and survival of the filarial nematodes are interfered by eliminating *Wolbachia* endosymbiont through antibiotic treatment. (Desjardins et al., 2013; Mcnulty et al., 2010; Lefoulon et al., 2016). However, *S. digitata* may not be responsive to antibiotics since it does not have *Wolbachia*. *S. digitata* is also considered as a model organism for HLF, due to their close resemblance to *Wuchereria bancrofti*, the primary causative agent of HLF, in morphology, histology and antigenicity (Perumal et al., 2015). Therefore, generation of a draft genome of *S. digitata* and complete functional analysis will pave the way to understand not only the biology of this organism but also to identify novel drug targets and/or vaccine candidates for human filariasis as well as other filarial diseases. Further, the reconstruction of complete metabolic pathways of *S. digitata* has not been undertaken yet and hence this study was undertaken to decipher the knowledge on how *Setaria digitata* acquires this metabolism. We discuss in great details the results, annotation of the genome heralding pathways.

2. Methodology

2.1. Collection of adult *S. digitata* worms

Adult worms of *S. digitata* were collected from the peritoneal cavity of cattle (*Bos taurus*) of the western and central provinces of Sri Lanka. The worms were washed thoroughly in PBS (pH 7.4) and preserved in 80% ethanol at − 20°C prior to analysis. As the worms were collected from already slaughtered animals, there was no need of ethics approval.

2.2. Extraction of genomic DNA

Genomic DNA of adult *S. digitata* worms was extracted using the DNA Micro kit, QIAGEN. The DNA was quantified using Qubit (Version 2.0) (Thermo Fisher Scientific, Waltham, Massachusetts, USA).

2.3. Genome sequencing

The sequencing was carried out by a sequence service provider Macrogen Ltd. Sequencing libraries were constructed from the extracted DNA using the TruSeq™ DNA PCR-Free Kit. Purified libraries were loaded onto an Illumina HiSeq4000 for paired end sequencing. Sequence data (base call files) were converted to obtain the FASTQC raw reads.

2.4. Genome assembly and annotation

The reads were filtered before assembly with a base quality greater than or equal to Q20 and the obtained 150bp reads were analyzed for K-mers using JELLYFISH (http://www.cbcb.umd.edu/software/jellyfish). Once the optimum k-mer size was identified, a *de novo* draft assembly was built with scaffolding using SOAP denovo2 (http://soap.genomics.org.cn/soapdenovo.html). Further downstream analysis (gene prediction and annotation) of the generated contigs/scaffolds were done using MARKER software (http://www.yandell-lab.org/software/maker.html) University of Virginia (UVA) FASTA algorithm. The NCBI's non-redundant protein database (nr) was used to BLAST the contigs with an e value of 1e-3 used as the threshold.

2.5. Functional Annotation using the BLAST2GO Tool

BLAST2GO is a functional annotation tool and pipeline was used for an extensive annotation and data mining of novel datasets using Gene Ontology, Enzyme Commission, InterPro and KEGG databases (Conesa et al., 2005; Conesa et al., 2008). The annotation is based on homologous mapping using BLAST as it integrates visualization and statistical software, including InterPro, enzyme codes, KEGG pathways, GO direct acyclic graphs (DAGs) and GO Slim. BLAST2GO first performs a BLASTX search against the NCBI non-redundant (NR) database. The generated BLAST hits were then subjected to downstream annotations. (Zamora et al., 2013) The 2075 contigs generated were used for downstream annotations.

2.6. Analysis of Wolbachia like endosymbiont DNA in the Setaria digitata genome

FASTA file containing 2075 *S. digitata* contigs was searched against several *Wolbachia* specific reference sequences such as *Wolbachia* surface protein (WSP), *Wolbachia*-specific 16S rRNA and *Wolbachia* MLST genes (*coxA, gatB, fbpA, ftsZ*, and *hcpA*) using UVA FASTA36 program (Pearson and Lipman, 1988)

3. Results

After the initial sequencing using the Illumina platform, a total of 14,735,628,242 bases in 97,586,942 reads were obtained with a GC % of 31.67% and a Q20 value of 96.09%. This was followed by quality control and pre-processing measures which yielded a total of 11,933,095,844 bases in 79,292,174 reads with a GC % of 31.77% and a Q20 value of 99.18%. A draft genome of 78,774,594 bases belonging to a total of 2,075 contigs was generated with the longest contig being 1,212,773 bp and the shortest 70 bp. The average contig size was 37,817 bp, and N50 value was 113,368. The GC % in the assembled draft genome was found to be 31.45%. As in a typical random library, it is expected to see a roughly normal distribution of GC content where the central peak corresponds to the overall GC content. An unusually shaped distribution could indicate a contaminated library or some other kinds of a biased subset. The histogram obtained showed a normal distribution and therefore, it was concluded that there was no contamination.

Out of 2075 total contigs/sequences used through the pipeline, 530 contigs did not generate a BLAST hit. Here BLASTx was performed against the NCBI non-redundant database with a cut-off of 1e-3. Out of 1545 BLAST hit generated contigs, 110 contigs generated a BLAST hit without further downstream GO annotation. Remaining 1435 BLAST hit generated contigs were mapped to retrieve GO terms. Out of 1435 mapped contigs, only 1280 contigs generated GO annotations while 155 contigs were only mapped. Data distribution of BLAST2GO analysis is shown in Fig. 1. Percentage of sequence similarities associated with 28112 hits are distributed within the range of 32% – 100%. A rule of thumb is that; two sequences are homologous if they are more than 30% identical over their length. All of the obtained hits have a sequence similarity higher than 30%, which was acceptable.

4. Discussion

The distribution of BLAST hits by species is shown in Fig. 2. Most of the hits belonged to filarial nematodes including *Brugia malayi* (4998 hits), *Wuchereria bancrofti* (3311 hits), *Onchocerca flexuosa* (2292 hits), *Loa loa* (2129 hits), *Brugia pahangi* (1940 hits) and *Onchocerca ochengi* (1673). Based on the Fig. 2, it is clear that a majority of the *S. digitata* sequences show a sequence homology towards filarial nematodes harbouring *Wolbachia*, like *Brugia malayi Wuchereria bancrofti*, *Brugia pahangi* and *Onchocerca ochengi*. In this study, 6055 total annotations have been done at a mean level of 6.488 with a standard deviation of 2.675. Overall, 1280 of 1435 (89.1%) mapped BLAST hits were annotated by at least one of the three categories of the GO function classification. The 1280 mapped contigs were then mapped to 6055 GO terms, among which 2297 (grouped in 9 sub-categories) (Fig. 3), 2206 (grouped in 8 sub-categories) (Fig. 4) and 1552 (grouped in 5 subcategories) (Fig. 5). The GO terms could be grouped to the biological process category, molecular function category and the cellular component category, respectively. Homologous sequences were found only for *coxA* and *gatB* via FASTA36 sequence similarity analysis. Sequence identity was higher than 50% with e-value less than 10^{-4}, and therefore we considered that sequence similarity is significant. As an alternative method of categorizing contigs/sequences by biochemical function, sequences were assigned to biological pathways using the KEGG database. A total of 246 contigs had been clustered into 95 pathways, in which the most over-represented pathways are biosynthesis of antibiotics (22 genes), Phosphatidylinositol signaling system (11 genes) and Purine metabolism (10 genes).

As stated above, previous genomic analysis of filarial nematodes and *Wolbachi*a endosymbiont have led to the identification of five biosynthetic pathways (haem, riboflavin, FAD, glutathione and nucleotide synthesis) present in *Wolbachia* that provide metabolites needed by their filarial hosts. Since, *S. digitata* does not harbor *Wolbachi*a, they should have their own biosynthetic pathways for haem, riboflavin and nucleotides and therefore, there has to be a set of genes in *S. digitata* genome that encode enzymes involved in these pathways. These genes can either be laterally transferred from *Wolbachia* or *S. digitata* have developed their own genes. However, no transfer relating to these metabolic pathways were apparent in previous genomic analysis (Senanayake et al., 2020). Our genome sequencing and subsequent analyses showed complete metabolic pathway reconstructions of *S. digitata* to determine

how *S. digitata* acquires these metabolites. The KEGG pathway analysis of the *S. digitata* genomic dataset revealed key genes that encode for enzymes involved in the above-mentioned nucleotide (purine, pyrimidine) and riboflavin synthesis pathways (Table 1), however, not a single enzyme involved in haem synthesis was found. According to the BLAST results of the BLAST2GO functional analysis, most of the *S. digitata* sequences have a higher sequence similarity to *Wolbachia* containing filarial nematodes like *Brugia malayi, Wuchereria bancrofti, Brugia pahangi* and *Onchocerca ochengi* than *Wolbachia* free filarial nematodes like *Loa loa* and *Onchocerca flexuosa*. Based on these results, it can be concluded that *S. digitata* is more closely related to *Wolbachia* containing nematodes than *Wolbachia* free nematodes. According to the FASTA36 sequence similarity analysis, partial sequences of *Wolbachia* marker genes (coxA and gatB) were found within the *S. digitata* genome.

Conclusions

Setaria digitata is a parasitic nematode which infects cattle. We have sequenced its genome and identified the pathways providing evidence for the presence of *Wolbachia* like DNA sequences in the *S. digitata* genome. Since 90% of the filarial nematodes studied to date contain *Wolbachia*, the sequencing similarity and downstream analyses suggested that the ancestors of *S. digitata* may have been colonized with *Wolbachia* in the distant past, and through the HGT, *Wolbachia* like DNA may have been brought into the nuclear genome prior to endosymbiont loss. Some of these HGTed sequences may have undergone genetic changes that are not detected *in silico* or they may have become nonfunctional pseudogenes, thus unable to produce functional proteins. In spite of the studies revealing that all nematode genomes contain FAD and glutathione pathways, and the riboflavin and haem biosynthesis pathways, genes encoding enzymes for FAD and glutathione pathways were not observed with only one gene encoding an enzyme involved in riboflavin synthesis was found in *S. digitata*. Moreover, complete nucleotide synthesis pathway and haem synthesis pathway were also not observed. This suggests that *S. digitata* may have evolved its own genes that encode enzymes involved in these biosynthetic pathways just like *Wolbachia* free filarial nematodes *Loa loa*. Transcriptome analyses is currently underway in our laboratories to ascertain this.

Declarations

Acknowledgement:

NR wishes to acknowledge Dr Prashanth Suravajhala's group members, staff for providing her training in Genomic Bioinformatics analysis at his laboratories in Department of Biotechnology and Bioinformatics, Birla Institute of Scientific Research, Statue circle, Jaipur 302001, RJ, India under the NSF Overseas Special Training Programme (OSTP) in 2019.

Authors' contributions:

• Nadarajah Rashanthy : Methodology, Formal analysis, Writing- Original draft preparation

• M.S.A.Kothalawala: Formal analysis, Writing- Original draft preparation.

• Thamarahansi Shashiprabha Mugunamalwaththa: Writing- Reviewing and Editing

• Yasanthi Illika Nilmini Silva Gunawardene : Writing - Review & Editing

• Naduviladath Vishvanath Chandrasekharan: < Writing - Review & Editing Kasun de Zoysa: Writing- Reviewing and Editing

• Prashanth Suravajhala: Formal analysis, Writing- Original draft preparation. Reviewing and Editing

• Ranil Samantha Dassanayake : Conceptualization, Supervision, Writing- Original draft preparation, Project administration

Funding:

The research was funded by a grant (AP/3/2017/CG/34) awarded to Professor R.S. Dassanayake by the University of Colombo, Sri Lanka, to promote collaborative research. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Conesa, A. & Götz, S. *Int. J. Plant Genome*. **218** (1), 1–12 (2008).
2. Conesa, A. *et al*. *J. Bio inform.* **21** (18), 3674–3676 (2005).
3. Darby, A. C. *et al*. Analysis of Gene Expression from the Wolbachia Genome of a Filarial Nematode Supports both Metabolic and Defensive Roles within the Symbiosis. Genome Research. **22** (12), 2467–2477 (2012).
4. Desjardins, C. A. *et al*. Genomics of Loa loa, a Wolbachia-free filarial parasite of humans. *Nat. Publ. Gr.* **45** (5), 495–500 (2013).
5. Ekblom, R. & Wolf, J. B. W. A field guide to whole-genome sequencing, assembly and annotation. *Evol. Appl.* (2014).
6. Fenn, K. *et al*. Phylogenetic relationships of the Wolbachia of nematodes and arthropods. *Journal.plos.org*. **2** (10), 887–889 (2006).
7. Gunawardene, Y. I. N. S., Dassanayake, R. S. & Perumal, A. N. I. Setaria digitata in advancing our knowledge of human lymphatic filariasis. *J. Helminthol*. (2015).
8. Lee, D. L. The Biology of Nematodes; Taylor & Francis: London (2002).
9. Lefoulon, E. *et al.* Breakdown of coevolution between symbiotic bacteria Wolbachia and their filarial hosts. *PeerJ.* (2016).

10. Mcnulty, S. N. *et al*. Endosymbiont DNA in endobacteria-free filarial nematodes indicates ancient horizontal genetic transfer. *PLoS One*. **5** (6) (2010).
11. Perumal, A., Gunawardene, Y. & Dassanayake, R. *J.Helminthol.* **90** (2), 129– 138 (2015). (Please check this)
12. Senanayake, K. S. *et al.* The genome of Setaria digitata: a cattle nematode closely related to human filarial parasites. *Genome Biology and Evolution*. **12** (2), 3971-3976 (2020).
13. Shin, S., Cho, K. & Wee, S. Ocular infection of cattle with Setaria digitata. *Journal of Veterinary Medical Science*. 3–6.
14. Shiny, C., Krushna, N. S. A., Haripriya, K. & Narayanan, R. B. Existence of endosymbiont Wolbachia in Setaria digitata. *Int. J. Medical Eng. Informatics*. **3** (2), 108–121 (2002).
15. Sundar, S. T. B. & Souza, P. E. D. Morphological characterization of *Setaria* worms collected from cattle. *J. Parasite. Dis*. **39** (3), 572–576 (2015).
16. Taylor, M. & Hoerauf, A. *Parasitol today*. **15** (11), 437–442 (1999)
17. Voronin, D., Abeykoon, A. M. L. L., Gunawardene, Y. I. S. & Dassanayake, R. S. Absence of Wolbachia endobacteria in Sri Lankan Isolates of the nematode parasite of animals Setaria digitata. *Vet. Parasitol*. **207** (3–4), 350–354 (2015).
18. Werren, J. H., Baldo, L. & Clark, M. E. Wolbachia: master manipulators of invertebrate biology. *Nat. Rev*. 741-751 (2008).
19. Wickramatunga, P. G. T. S., Gunawardene, Y. I. N. S., Wijesinghe, K. J., Ellepola, A. N. B. & Dassanayake, R. S. RNAi-mediated silencing of *ARV1* in Setaria digitata impairs in-vitro microfilariae release, embryogenesis and adult parasite viability. *Vet. Parasitol.* **284**, 109189 (2020).
20. Zamora, G. O. M., De Novo sequencing, assembly and analysis of the genome and transcriptome of the nematode Panagrolaimus superbus. Ph.D. Thesis, National University of Ireland (2013).

Table

Table 1
Enzymes identified through KEGG pathway analysis that involve in the Purine, Pyrimidine and Riboflavin metabolism in *S. digitata.*

Metabolic pathway	Identified enzymes
Purine metabolism	adenylosuccinate lyase (EC:4.3.2.2)
	nucleoside-triphosphate phosphatase (EC:3.6.1.15)
	guanylate cyclase (EC:4.6.1.2)
	bis(5'-adenosyl)-triphosphatase (EC:3.6.1.29)
	ribose-phosphate diphosphokinase (EC:2.7.6.1)
	nucleotide diphosphatase(EC:3.6.1.9)
	adenosinetriphosphatase (EC:3.6.1.3)
	adenylosuccinate synthase (EC:6.3.4.4)
	3',5'-cyclic-nucleotide phosphodiesterase (EC:3.1.4.17)
	pyruvate kinase (EC:2.7.1.40)
Pyrimidine metabolism	dihydroorotate dehydrogenase (quinone) (EC:1.3.5.2)
	nucleotide diphosphatase (EC:3.6.1.9)
	UMP/CMP kinase (EC:2.7.4.14)
	thymidylate synthase (EC:2.1.1.45)
	UMP kinase (EC:2.7.4.22)
	dCMP deaminase (EC:3.5.4.12)
Riboflavin metabolism	Nucleotide diphosphatase (EC:3.6.1.9)

Figures

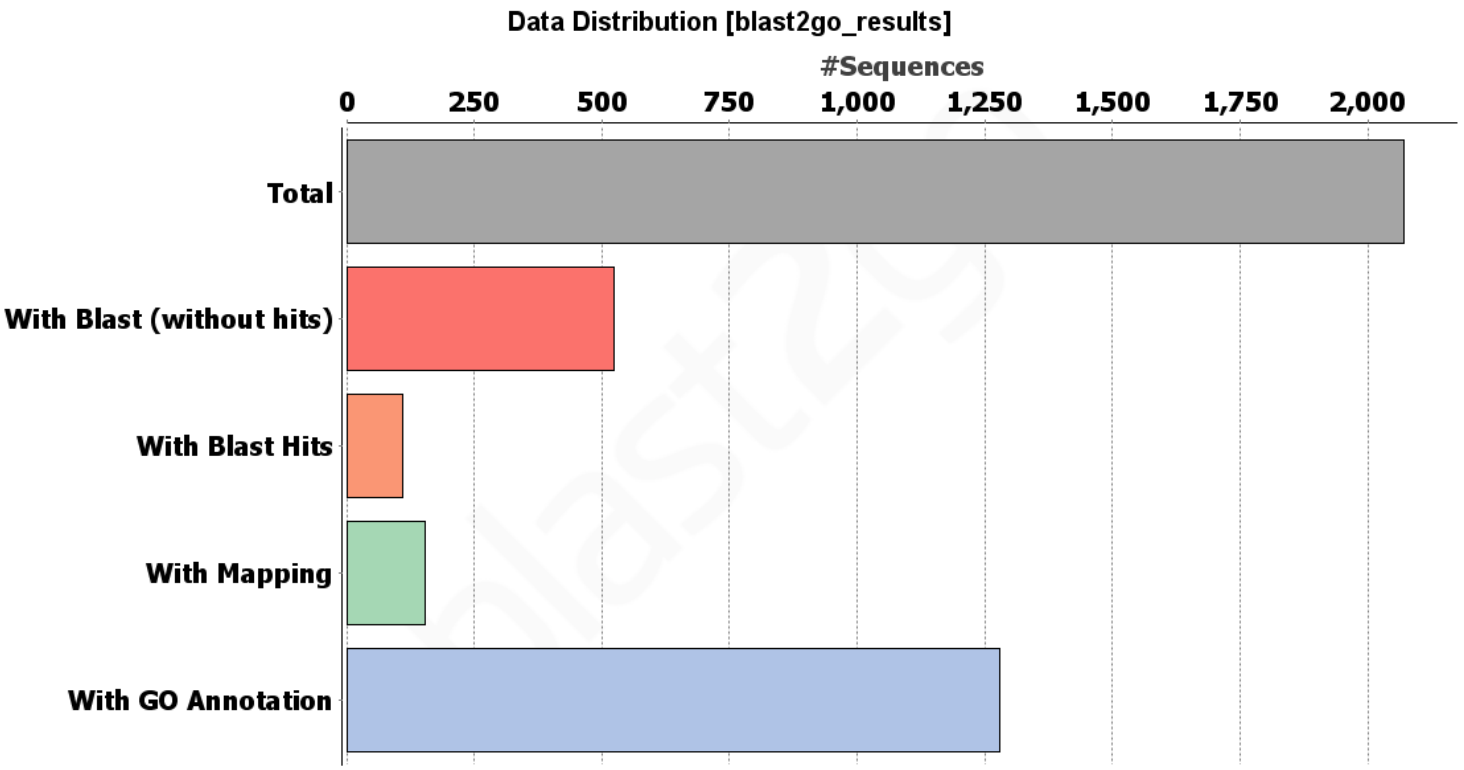

Figure 1

Data distribution of the BLAST2GO analysis of the S. digitata genome dataset post BLAST2GO annotation. Out of 2075 total contigs, 530 contigs did not generate a BLAST hit. Out of 1545 BLAST hit generated contigs, 110 contigs generated a BLAST hit without further downstream GO annotation. Out of remaining 1435 mapped contigs, only 1280 contigs generated GO annotations.

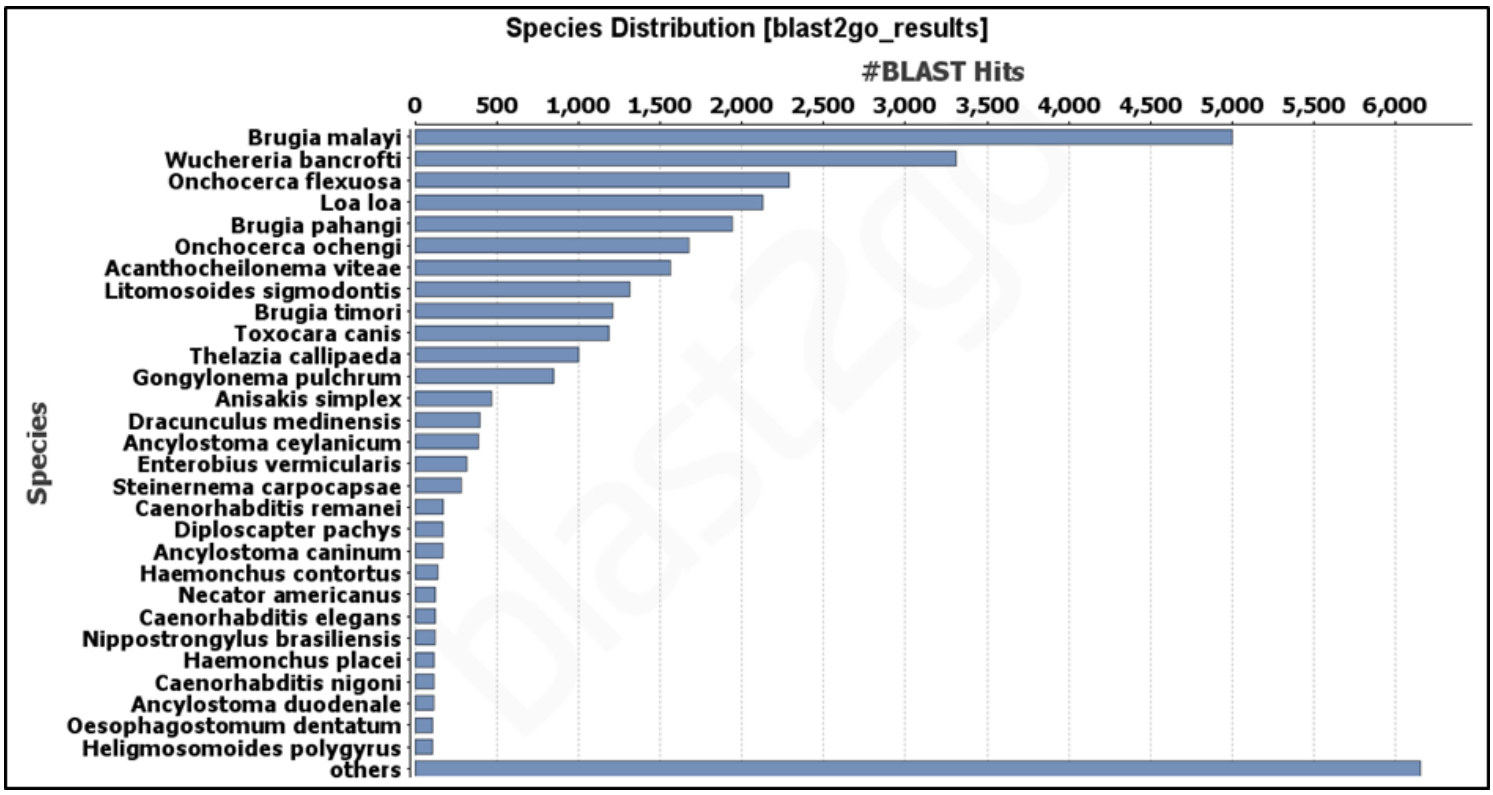

Figure 2

Species Distribution of the S. digitata sequences which returned BLAST hits following BLASTX against NR with a cut-off of 1e-3. Majority of the BLAST hits were belonged to filarial nematodes with Wolbachia, like Brugia malayi (4998 hits), Wuchereria bancrofti (3311 hits), Brugia pahangi (1940 hits) and Onchocerca ochengi (1673 hits).

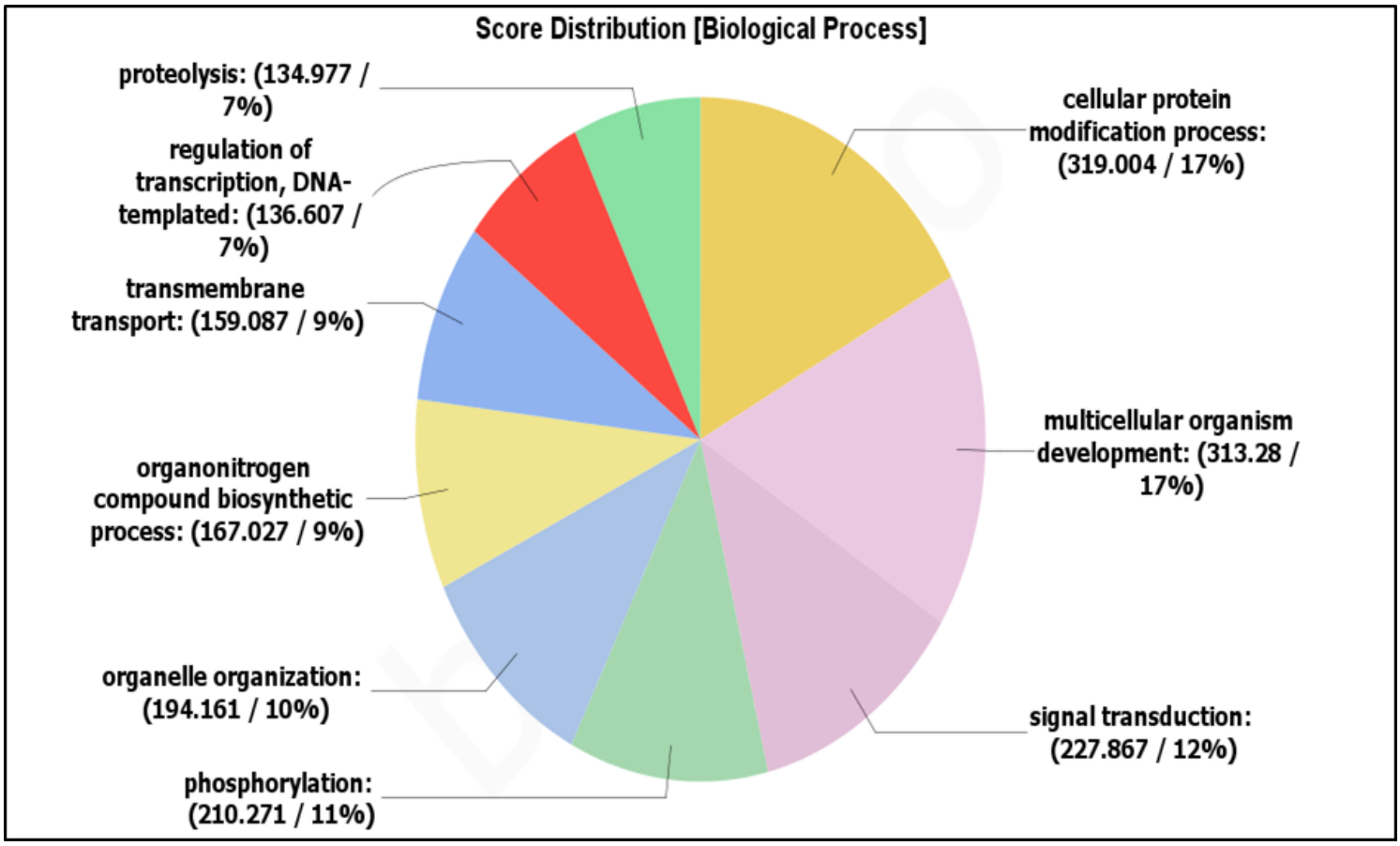

Figure 3

Score distribution of Biological processes. 2297 GO terms has been grouped in 9 subcategories. With regard to the ‘biological process’ category, the most annotated GO terms were ‘cellular protein modification process’ (17%). ‘Multicellular organism development’ (17%), ‘signal transduction’ (12%) and ‘phosphorylation’ (11%).

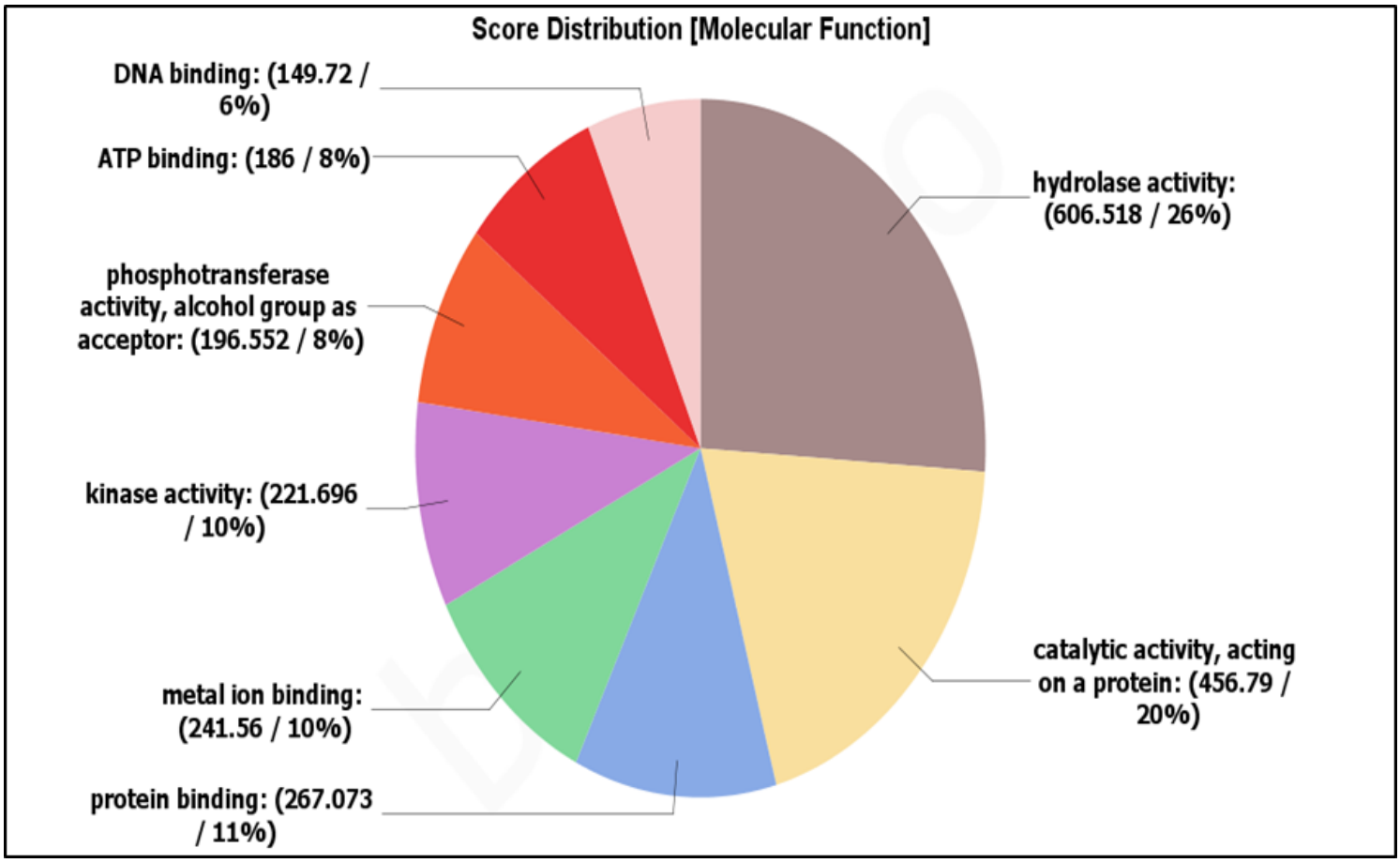

Figure 4

Score distribution of Molecular functions. 2206 GO terms has been grouped in 8 subcategories. In the ‘molecular function’ category, most annotated GO terms were ‘hydrolase activity’ (26%), ‘catalytic activity’ (20%), and ‘protein binding’ (11%).

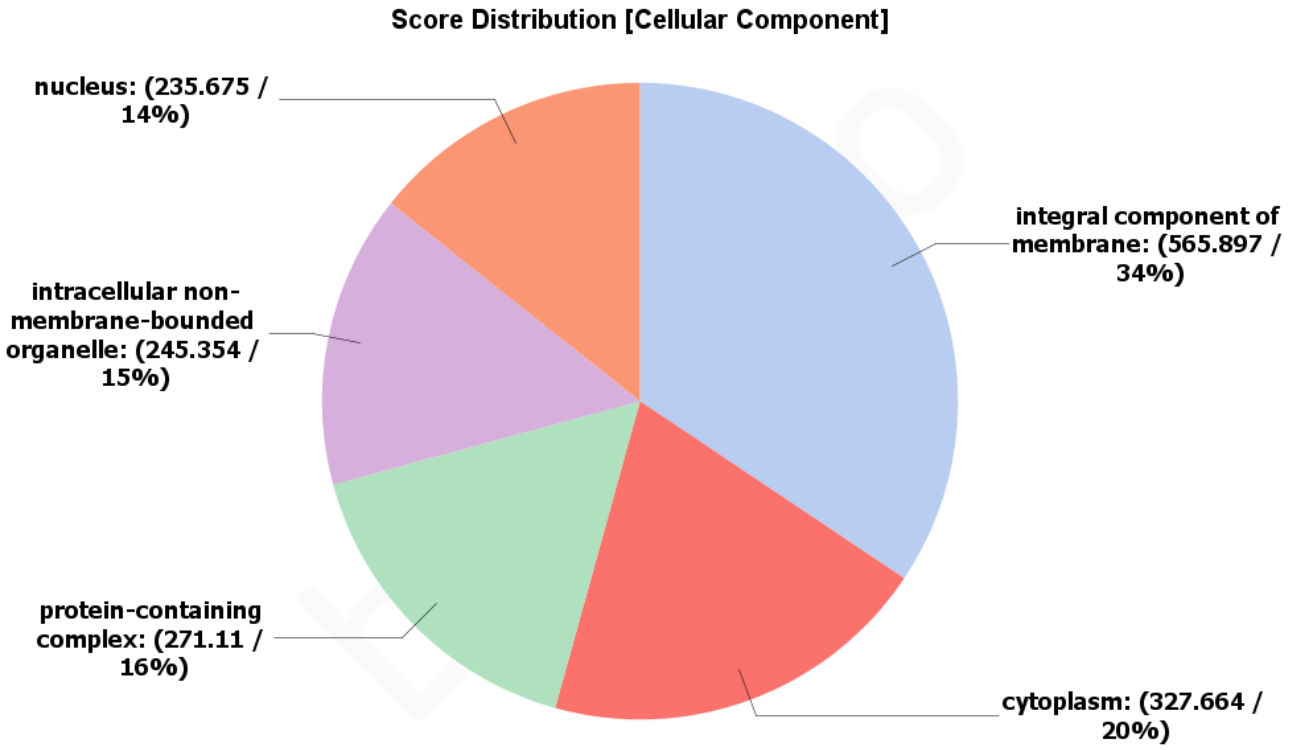

Figure 5

Score distribution of Cellular Component. 1552 GO terms has been grouped in 5 subcategories. Within ‘cellular components’ category, most annotated GO terms were ‘integral component of membrane’ (34%), ‘cytoplasm’ (20%), ‘protein-containing complex’ (16%), ‘intracellular non-membrane bounded organelle’ (15%) and ‘nucleus’ (14%).