International Journal for Parasitology 49 (2019) 933–943

Contents lists available at ScienceDirect

International Journal for Parasitology

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

Genome-wide analysis of *Anisakis simplex* sensu lato: the role of carbohydrate metabolism genes in the parasite's development ☆

Elżbieta Łopieńska-Biernat [a,*], Łukasz Paukszto [b], Jan Paweł Jastrzębski [b], Kamil Myszczyński [c], Iwona Polak [a], Robert Stryiński [a]

[a] Department of Biochemistry, Faculty of Biology and Biotechnology, University Warmia and Mazury in Olsztyn, Oczapowskiego 1A, 10-719 Olsztyn, Poland
[b] Department of Plant Physiology, Genetics and Biotechnology, Faculty of Biology and Biotechnology, University Warmia and Mazury in Olsztyn, Oczapowskiego 1A, 10-719 Olsztyn, Poland
[c] Department of Botany and Nature Protection, Faculty of Biology and Biotechnology, University of Warmia and Mazury in Olsztyn, Plac Łódzki 1, 10-727 Olsztyn, Poland

ARTICLE INFO

Article history:
Received 15 March 2019
Received in revised form 19 June 2019
Accepted 24 June 2019
Available online 24 September 2019

Keywords:
Anisakis simplex sensu lato
Draft genome
Carbohydrate metabolism genes
Trehalose
Glycogen

ABSTRACT

Anisakis simplex sensu lato is a parasitic nematode which can cause gastric symptoms and/or allergic reactions in humans who consume raw and undercooked fish. Anisakiasis poses a growing health problem around the globe because it causes non-specific symptoms and is difficult to diagnose. This genome-wide study was undertaken to expand our knowledge of *A. simplex* s.l. at the molecular level and provide novel data for biological and biotechnological research into the analyzed species and related nematodes. A draft genome assembly of the L3 stage of *A. simplex* s.l. was analyzed in detail, and changes in the expression of carbohydrate metabolism genes during the parasite's life cycle were determined. To our knowledge, this is the first genome to be described for a parasitic nematode of the family Anisakidae to date. We identified genes involved in parasite-specific pathways, including carbohydrates metabolism, apoptosis and chemo signaling. A total of 7607 coding genes were predicted. The genome of *A. simplex* s.l. is highly similar to genomes of other parasitic nematodes. In particular, we described a valuable repository of genes encoding proteins of trehalose and glycogen metabolism, and we developed the most comprehensive data set relating to the conversion of both saccharides which play important roles during the parasite's life cycle in a host environment. We also confirmed that trehalose is synthesized at the expense of glycogen. Trehalose anabolism and glycogen catabolism were the predominant processes in stages L4 and L5, which could confirm our and other authors' previous reports that trehalose is synthesized at the expense of glycogen. The *A. simplex* s.l. genome provides essential data for post-genomic research into the biology of gastrointestinal and allergic anisakiasis in humans and the biology of other important parasitic helminths.

© 2019 Australian Society for Parasitology. Published by Elsevier Ltd. All rights reserved.

1. Introduction

The life cycle of many pathogens involves transitions between distinct developmental stages which are necessary for transmission. In helminth parasites, the transmissible infective stage plays the most important role because it enables survival outside the host (Antonovics et al., 2017). A robust knowledge of molecular processes which control stage transition is required to develop transmission-blocking therapies and novel diagnostic methods (Lv et al., 2015). However, the genomic and genetic resources required for the research into parasitic nematodes are scarce. The need to develop a workable model of a parasitic nematode is widely recognized, but the majority of helminth species infecting humans are not suitable for experimental studies. Experiments conducted on *Anisakis simplex* s.l., a gastrointestinal parasitic nematode of sea mammals, can fill this knowledge gap and contribute to the introduction of a new model organism for research into allergies, anthelmintic drugs and vaccines (Pravettoni et al., 2012; Ivanović et al., 2017).

Empirical studies into the distribution and abundance of *Anisakis* larvae in intermediate and final hosts have contributed important insights into the biological life cycle of these parasites. *Anisakis simplex* sensu lato (s.l.) has a complex development cycle. L3s have been identified in paratenic hosts (i.e., fish, cephalopods), final hosts (marine mammals i.e., seals, whales) and accidental hosts

☆ Nucleotide sequence data reported in this paper are available in GenBank under accession numbers. MH425576-78, MF069073-113, MG210776, MG557620-25, KX280018-24, KP326558-61, KP640594, KJ789948-55, KJ560557 and KM496564.

* Corresponding author. Fax: +48 895352015.
E-mail address: ela.lopienska@uwm.edu.pl (E. Łopieńska-Biernat).

https://doi.org/10.1016/j.ijpara.2019.06.006
0020-7519/© 2019 Australian Society for Parasitology. Published by Elsevier Ltd. All rights reserved.

(humans). The role of the L3 in various hosts has to be explored to elucidate the mechanisms of parasite transmission. Analyses of the parasite's metabolism and survival strategy are conducted at the genomic level with the use of molecular tools. In the final host organism, the parasite undergoes three consecutive development stages: L4, the juvenile stage and the adult stage (Grabda, 1976; Klimpel et al., 2004). In this study, we relied on the above developmental stages to analyze carbohydrate metabolism in *A. simplex* s.l. Additionally, we believe this is the first study where the adult stage was analyzed.

Anisakis cannot complete its life cycle in humans which act as accidental hosts, but they can cause diseases and allergies. The consumption of raw or undercooked fish leads to food-borne zoonoses which are most commonly caused by the larvae of Anisakidae nematodes of the genera *Anisakis, Contracaecum* and *Pseudoterranova* (Sakanari and McKerrow, 1989; Audicana et al., 2002; Klimpel and Palm, 2011). *Anisakis simplex* s.l. is responsible for most of all registered episodes of human anisakiasis and allergic reactions around the world (Yil et al., 2005). Almost 3000 cases are reported annually (Baird et al., 2014). The results of a quantitative risk assessment suggest that anisakiasis is a highly underestimated disease. The prevalence of anisakiasis in Europe is likely to increase from 500 to 7700–8320 cases per year (Bao et al., 2017).

Anisakiasis has emerged as an important zoonotic infection in various parts of the world, reminding us that effective disease control requires persistent vigilance. The species of the genus *Anisakis* display unique and biologically interesting complexity during interactions with their hosts. Significant progress has been made towards understanding these interactions in detail. The discussed parasites are important for several reasons: (i) the presence of larvae in fish products decreases their commercial value and leads to foodborne zoonoses (Baird et al., 2014; Sanchez-Alonso et al., 2018) (ii) *A. simplex* s.l. infections produce gastrointestinal symptoms and hypersensitive states (Daschner et al., 2012), (iii) *A. simplex* s.l. is analyzed for diagnostic purposes, (iv) *A. simplex* s.l. has therapeutic potential in immune-mediated diseases (Mehrdana and Buchmann, 2017), and (v) the species can be used in cancer research (Nieuwenhuizen, 2016; Messina et al., 2017).

Genomic research into *Anisakis* spp. began with the generation of an analysis of expressed sequence tags, which contributed to the rapid discovery of genes in other eukaryotes (Yu et al., 2007). Genomic functional sequences in the parasite were studied by bioinformatic characterization of transcripts in two *Anisakis* spp., *Anisakis simplex* and *Anisakis pegreffii,* with emphasis on the potential allergens and their origins from different parts of the parasite's body (Baird et al., 2016; Cavallero et al., 2018). The existing databases contain contigs and partial data without annotations because the information gained from free-living *Caenorhabditis elegans* nematodes is not suitable for research into the parasitic organisms.

To date, 60% of sequences of human parasites have been deposited in databases. However, data relating to *Anisakis* are still fragmentary, which hampers research into effective drugs and new generation vaccines (Brindley et al., 2009).

L3s represent a critical stage in the life cycle of *Anisakis* parasites because they are transmitted by fish that act as vectors of infection in mammals. Genetic variation could be responsible for the parasite's remarkable ability to adapt to different climatic regions and host species (Kuhn et al., 2016).

A thorough understanding of the molecular basis of parasitism, in particular parasitic infections in humans and animals, is one of the most important challenges in biology. Helminths generally have much more complex genomes than model organisms such as yeasts or fruit flies (Hotez and Kamath, 2009). A synthetic approach where nematodes undergo a transition from a free-living to a parasitic lifestyle will be the ultimate test of our ability to understand the genetic and genomic basis of nematode parasitism (Viney, 2017). The need to develop a workable model of parasitic nematodes is widely recognized, but the majority of helminth species infecting humans are not suitable for experimental studies due to the absence of references to the model. In contrast, *A. simplex* s.l., a gastrointestinal helminth which also causes allergies in humans, can be a suitable organism not only for studying anisakiasis, but also for allergies and immune-mediated diseases, as mentioned before. However, the only source of information about the *A. simplex* s.l. genome is the GenBank database, where only raw data and scaffolds are present. New sequencing of the *A. simplex* s.l. genome has to be done to overcome the problems associated with previous genomic studies based mostly on homology to *C. elegans*, a phylogenetically distant organism, in the absence of more suitable reference genomes. Additionally, Gilabert et al. (2016) demonstrated significant differences not only in the genomic sequence of various nematodes species, but also their sister species. The presence of three sibling species in the *Anisakis simplex* complex (Mattiucci et al., 2014) emphasizes the importance of the need for such research into *Anisakis*.

The genome of *A. simplex* isolated from Baltic Sea fish has no published studies to date. This study made, to our knowledge, the first attempt to sequence and fully describe the genome of *A. simplex* s.l. subspecies from the Baltic Sea region.

This study describes the assembly and annotation of the draft genome of larvae of *A. simplex* s.l. with other parasitic nematodes. The results of this genome-wide analysis provide valuable insights for future research into anthelmintic resistance and parasite biology based on *A. simplex* s.l., the most important model organism for studying allergies caused by helminths. Carbohydrate metabolism plays a very important role in the transmission and survival of gastrointestinal parasites in hosts (Komuniecki and Harris, 1995). Caenorhabditis *elegans* is the only nematode that have been extensively studied in this area, but most of the analyses were focused only on the trehalose metabolism (Pellerone et al., 2003). In this study, genome-wide analysis of *A. simplex* s.l. was done to describe the metabolism of glycogen and trehalose, the most important saccharides in the life cycle of parasitic nematodes. These interconnected metabolic pathways have been studied in *A. simplex* s.l. at the molecular level, to our knowledge for the first time.

2. Material and methods

2.1. Parasite material

Anisakis simplex s.l. L3s were isolated from Baltic herring (*Clupea harengus membras*). Only larvae longer than 1 cm were selected for the experiment. Larvae were rinsed three times in sterile saline solution (0.9% NaCl) and washed for 30 min in bactericidal and fungicidal solutions (Iglesias and Adroher, 1997). They were identified at species level as *A. simplex* type I under a stereoscopic microscope. Larvae were divided into two groups. The first group was stored separately for DNA and RNA extraction (see Sections 2.2 and 2.3), and the second group was used for in vitro cultures. Larvae were cultured in the RPMI-1640 medium (R8758, Sigma Aldrich, USA) enriched with 20% fetal bovine serum (F7524, Sigma Aldrich) and 1% pepsin (P7125, Sigma Aldrich), in a six-well plate (BD Biosciences, Poland), at pH 4 and 37 °C with 5% CO_2, according to the method described by Iglesias et al. (2001). The in vitro culture until the 'adult' development stage of *A. simplex* s.l. lasted 35 days.

2.2. DNA isolation

Total genomic DNA was isolated from L3s of *A. simplex* s.l. using the Genomic Max AX Direct Kit according to the manufacturer's

protocol (A&A Biotechnology, Poland). Two independent biological replicates containing genomic DNA from three L3s of *A. simplex* s.l. were analyzed. The total DNA concentration was determined using the Quant-iT™ Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific, USA).

2.3. RNA extraction and reverse transcription PCR

Three independent biological replicates containing total RNA from three parasites representing different development stages (L3, L4/6 days, L4/12 days, L5, adult) of *A. simplex* s.l. were analyzed in six independent experiments. RNA was obtained with the RNA Total Mini Kit (A&A Biotechnology). RNA samples were treated with the DNase before use according to the manufacturer's protocol (A&A Biotechnology). The purity was checked spectrophotometrically in the NanoDrop 1000 system (Thermo Fisher Scientific). The quality of the extracted RNA was verified with the use of an Agilent Bioanalyzer 2100 (Agilent Technologies, Germany). cDNA synthesis was performed with oligo$(dT)_{18}$ primers and reverse transcriptase (20 U/μl) (Transcriba Kit, A&A Biotechnology).

2.4. Genome sequencing and quality control

The DNA integrity was verified with the use of an Agilent Bioanalyzer 2100 (Agilent Technologies). The library was prepared with the Nexter XT DNA Library Prep Kit (Illumina, USA) according to the manufacturer's instructions. Sequencing was performed on a MiSeq (Illumina) device in paired-end mode with a single read length of 250 bases. A MiSeq Reagent Kit v2 (500 cycle) was used to obtain the required amount of DNA (from 200 ng of the genomic template). The sequences were saved in FASTQ format. The quality of raw reads was checked to remove adapter sequences and low quality sequences (<20 PHRED) in the Trimmomatic program (Bolger et al., 2014). The GC content was determined to control the randomness of sequencing and to analyze nucleotide distribution.

2.5. Genome assembly and mitogenome identification

Clear reads were assembled de novo in the SOAP program (http://soap.genomics.org.cn) with k-mer 31 coverage, and in the Velvet program (http://www.ebi.ac.uk/~zerbino/velvet) with k-mer 47 coverage. The sequencing statistics were calculated in the assemblathon2 program (https://assemblathon.org/assemblathon2). The final number of contigs was obtained after de novo assembly in the Velvet program and the SOAP program. Some of the resulting gaps were eliminated in the contigs generated in low coverage areas in the GapCloser program (https://sourceforge.net/projects/soapdenovo2/files/GapCloser/). In the next step, bacterial artifacts that could be sequenced together with *A. simplex* s.l. material were eliminated. For this purpose, the de novo assembled genome was scanned in the BLASTn program and compared with *Shigella* genomic sequences. The reads were mapped to the *A. simplex* s.l. previous mitogenome in Geneious v. 8.0 (95% quality, 50 bp coverage). Fifteen other *A. simplex* s.l. references (Supplementary Table S1) were used to compare mitogenome haplotypes in LASTZ (https://www.bx.psu.edu/~rsharris/lastz/) and to identify potential single nucleotide variants (SNVs).

2.6. Prediction of repetitive elements

The assembled genome was analyzed in PHOBOS v. 3.3.12 (http://www.ruhr-uni-bochum.de/spezzoo/cm/cm_phobos.htm) to investigate the distribution of simple sequence repeats (SSRs) in the *A. simplex* s.l. genome. Exact repeats with 2–6 units and a minimum length of 15 bp were regarded as SSRs. The distribution of SSRs in the *A. simplex* s.l. genome was analyzed and nucleotide repeat frequency was plotted as a function of the number of repeats using R function barplot3d (http://addictedtor.free.fr/graphiques/sources/source_161.R).

2.7. Circos

The transcripts corresponding to selected *C. elegans* proteins were aligned to the *C. elegans* genome in the GMAP program (https://cbsu.tc.cornell.edu/lab/labsoftware.aspx or http://gmap3.net/). The results (inner ring) and *A. simplex* s.l. reads (outer ring) were visualized in Circos (http://circos.ca/) to determine the quality distribution of the reads mapped to the coding sequence. The results were generated using Geneious v. R8 (http://www.geneious.com) (Kearse et al., 2012).

2.8. Comparative analysis

In the next stage, *A. simplex* s.l. contigs were compared against proteomic databases of *Ascaris suum*, *Toxocara canis*, *Brugia malayi* and *C. elegans* by pairwise alignment in the BLASTx program with an *E*-value of 10^{-5}. The number of common contigs equivalent to all databases was determined. Then after pairwise comparison between *Anisakis simplex* s.l. and four other nematode species, the Venn diagram with the numbers of homologous proteins was plotted in R (https://www.r-project.org/).

2.9. Functional analysis and protein-coding gene prediction

A pool of contigs was scanned in the cd-hit-est program (http://weizhongli-lab.org/cd-hit/) to eliminate redundant sequences with identity greater than 97%. Then, the contigs selected from alignments to *A. suum* were used as patterns for analysis in BLAST2GO (*E*-value $<10^{-5}$). Only the processes that were directly related to L3s of *A. simplex* were selected from the annotations in the Biological Process category. Selected Gene Ontology (GO) terms were divided into 10 subcategories related to transport, response to oxidative stress, apoptosis, basic metabolism, and trehalose and glycogen metabolism. In addition, a new metabolic pathway for glycogen and trehalose transformation, which could be specific for the family Anisakidae, was proposed based on the known starch and sucrose pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG) and The Comprehensive Enzyme Information System (BRENDA) databases.

Based on the individual sequences allocated to selected GO terms, a pool of contigs was selected and identified to carry out the protein-coding gene prediction. In the first stage, exon–intron structure templates were identified with the Augustus algorithm in Geneious v. R8 (Stanke and Morgenstern, 2005) and next, genes prediction was done based on determined gene products.

2.10. Gene expression by real-time quantitative PCR (RT-qPCR)

The designed primers had a predicted melting temperature of 60 ± 5 °C, length of 20 nucleotides (nt), and GC content of >48%, where 100–200 bp amplicons spanned introns. The primers (Supplementary Table S2) were designed based on the coding regions of *tpp, tps, tre, gs, gp* sequences of *A. simplex* s.l. Primer pairs were tested by PCR amplification of the genetic material from *A. simplex* s.l. L3s. A single product band with the expected size after gel electrophoresis (2% agarose gel) was regarded as a positive result. The melting curve analysis after each RT-qPCR assay definitively confirmed the specificity of the primers. Standard curves were run for housekeeping genes *ef-1α* and *ppi12* to confirm the efficiency ($R^2 = 0.99$, $E = 99\%$) and linear regression of fluorescence data

during the exponential phase of RT-qPCR. The amplification efficiency (*E*) of *tpp, tps, tre, gp, gs* primer pairs was estimated with the use of Lin Reg PCR software (Ramakers et al., 2003). Primer pairs had *E*-values of >0.9. The final concentration of primers in the RT PCR Mix SYBR B (A&A Biotechnology) was 250 nM in a total reaction volume of 20 μl. The relative expression in all stages was calculated using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001), with the use of two housekeeping genes, *ef-1α* and *ppi12*, for which expression values were obtained separately for each stage and treated as controls. All reactions were carried out in an Applied Biosystems 7500 Fast Real-Time PCR System. Data were obtained and analyzed using Applied Biosystems software. Significant differences between samples were identified with a Student's *t*-test at $P < 0.05$.

2.11. Data accessibility

The results of the whole genome shotgun project were deposited in European Nucleotide Archive (ENA) (accession no ERS2790326). The described coding sequences of *A. simplex* s.l. were submitted to GenBank under accession numbers MH425576-78, MF069073-113, MG210776, MG557620-25, KX280018-24, KP326558-61, KP640594, KJ789948-55, KJ560557 and KM496564 (Supplementary Table S3).

3. Results

3.1. Description of a new version of the A. simplex s.l. genome

3.1.1. Genomic assembly and mapping to reference

We sequenced 32,262,754 (2 × 16,131,337) paired-end reads of the *A. simplex* s.l. genome with a read length of 100–250 bp. Approximately 70% of the sequences had a PHRED score of >35 (Fig. 1). A total of 26,479,344 reads were obtained after quality control (elimination of adapter sequences and sequences shorter than 50 bp). Two de novo assemblies were performed simultaneously in SOAP and Velvet programs for quality optimization. The assemblers produced 59,943 and 69,529 contigs, respectively (49,180 in the PRJEB496 project) (International Helminth Genomes Consortium, IHGC 2019), which is why Velvet-generated pool of contigs with enhanced statistical parameters (scaffold N50 – 6.9 kb; total size of contigs – 143 bp; maximum contigs size – 90.7 kb, where these values in the PRJEB496 project were: scaffold N50 – 9.3 kb; total genome size – 127 Mb; maximum contigs size 427.7 kb) (IHGC, 2019) was selected for further analysis. *Anisakis simplex* s.l. contigs were filtered de novo to identify bacterial contaminants, and the resulting 57,402 contigs were used in further analysis. Genome size and GC content are species-specific metabolic parameters (Supplementary Table S4). The size of the *A. simplex* genome corresponds to the average values for the phylum Nematoda, i.e., PRJEB496 *Anisakis simplex* (127 Mb), *Brugia malayi* (93.7 MB) and *Trichinella spiralis* (64 MB). GC content (37.4%) also corresponds to the range of values characteristic for helminths (30.5%–43.9%) (Lv et al., 2015), and similar value was obtained in PRJEB496 project for *Anisakis simplex* (36.7%) (Supplementary Table S4). We relied on extensive evidence from genomic analyses of *A. simplex* L3s to curate a de novo model of protein-coding genes. The mean coding sequence length of *A. simplex* s.l. (1234 bp) was similar to that of *T. canis* (1156 bp), and the average gene length was specific for the parasitic clade III (429 bp for *A. simplex* and 434 bp for *B. malayi*) (Wormbase, https://parasite.wormbase.org/), where the high value of the mean intron length characteristic for other nematodes was present also in *A. simplex* s.l. (286.7 bp) (Supplementary Table S4). This data was used to verify open reading frames and confirm exon/intron and gene boundaries. Genome data from de novo predictions of L3s and homology-based searches were used to annotate approximately 170 genes in the NCBI database. Genes were selected based

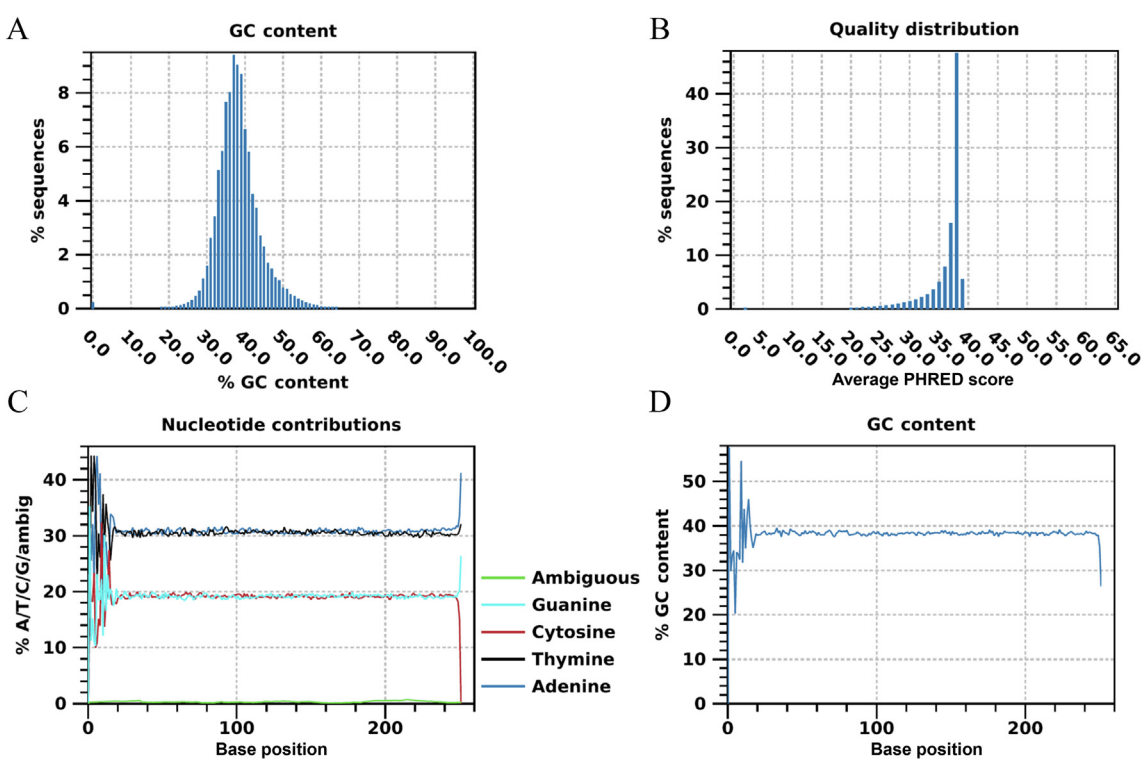


Fig. 1. A representative example of quality control metrics of sequenced reads of the *Anisakis simplex* sensu lato genome as indicated by FastQC. (A) GC (%) content distribution over all sequences of genomic data; (B) Phred quality score distribution over all reads in each base for genomic data; (C) The distribution of nucleotides over all sequenced reads; (D) GC (%) content distribution over all base positions.

on >90% coverage with reference data (Supplementary Table S3). The resulting data have numerous applications in functional genomics for designing probes and identifying putative protein sequences in proteomics (PAIRWISE alignment with selected homologs).

3.2. Mitogenome and SSRs

Mapping to a previously sequenced mitogenome (Kim et al., 2006; Yamada et al., 2017; Mohandas et al., 2014) supported the development of a more comprehensive and complete mitogenome with a length of 13,938 bp, where 11 genes code for proteins, 22 for tRNA, two for rRNA, and lacking a gene of mitochondrially encoded ATP synthase membrane subunit 8 that encodes a subunit of mitochondrial ATP synthase (ATP8). A comparison of the *A. simplex* mitogenome (KU899549) with 15 other haplotypes of the *A. simplex* mitogenome from the NCBI database revealed 95.3%–99.3% similarity (Supplementary Table S1). Four synonymous SNVs that were not present in the remaining haplotypes were also identified in COX2, CYTB, ND2 and ND1 genes (Fig. 2). Using the reference-based approach, we detected 1654 potential SSRs, including 733 di-SSR (44.32%), 862 tri-SSR (52.12%), 51 tetra-SSR (3.08%), two penta-SSR (0.12%) and six hexa-SSR (0.36%). The most abundant SSR motif (CCG) was identified in the tri-SSR and di-SSR repeats (Fig. 3).

3.3. Comparison of the A. simplex s.l. genome with the C. elegans genome and transcriptome

We also compared the abundance and distribution of *A. simplex* s.l. contigs with the *C. elegans* genome. All reads were mapped to a complete set of chromosomes of *C. elegans* which is the closest relative of *A. simplex* s.l. with a fully sequenced and annotated genome. *Anisakis simplex* s.l. contigs were simultaneously used to find the corresponding proteins in *C. elegans* in the Protein Ensemble Database (local BLASTx with an e-value of 10^{-5}). The identified proteins were used to select the related transcripts of *C. elegans*. A total of 13,047 transcript sequences were obtained for *A. simplex* s.l. and aligned to the *C. elegans* genome using the 'exonerate with est2genome' model (Mott, 1997). The distribution of the *A. simplex* s.l. genome reads and transcripts mapped/aligned to the *C. elegans* genome were compared with a customized Perl script which counted the frequency of reads/transcripts with a 100 kb window. The results were presented in a Circos graph (Fig. 4). Both distribution histograms were standardized for comparison.

3.4. Comparative genomics of A. simplex s.l.

The *A. simplex* s.l. genome was compared with *A. suum*, *T. canis*, *B. malayi* and *C. elegans*. In all compared organisms, 1104 orthologues were identified as conserved orthologs of Nematoda.

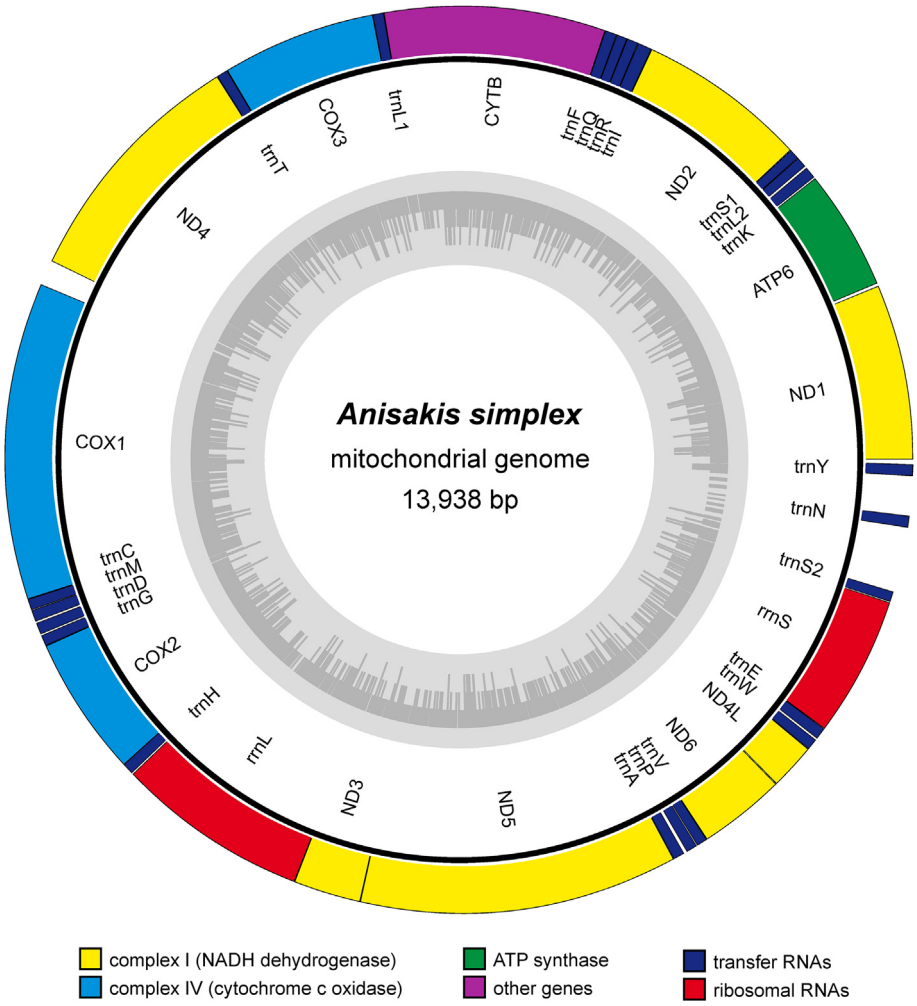


Fig. 2. The circular map of the mitogenome of *Anisakis simplex* sensu lato. Different colors represent different gene families. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

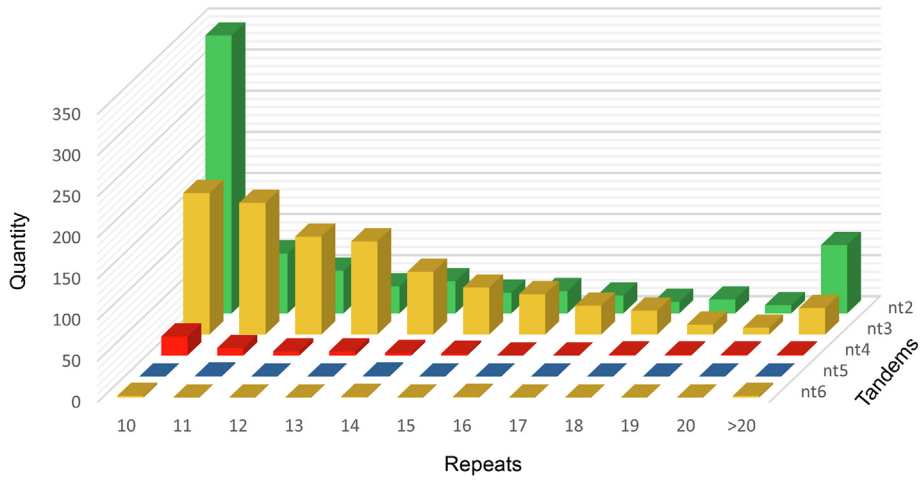


Fig. 3. The distribution of simple sequence repeats in the *Anisakis simplex* sensu lato genome. Di-, tri-, tetra-, penta- and hexa-nucleotide repeats were analyzed and their frequencies were plotted as functions of the number of repeats. nt, nucleotide.

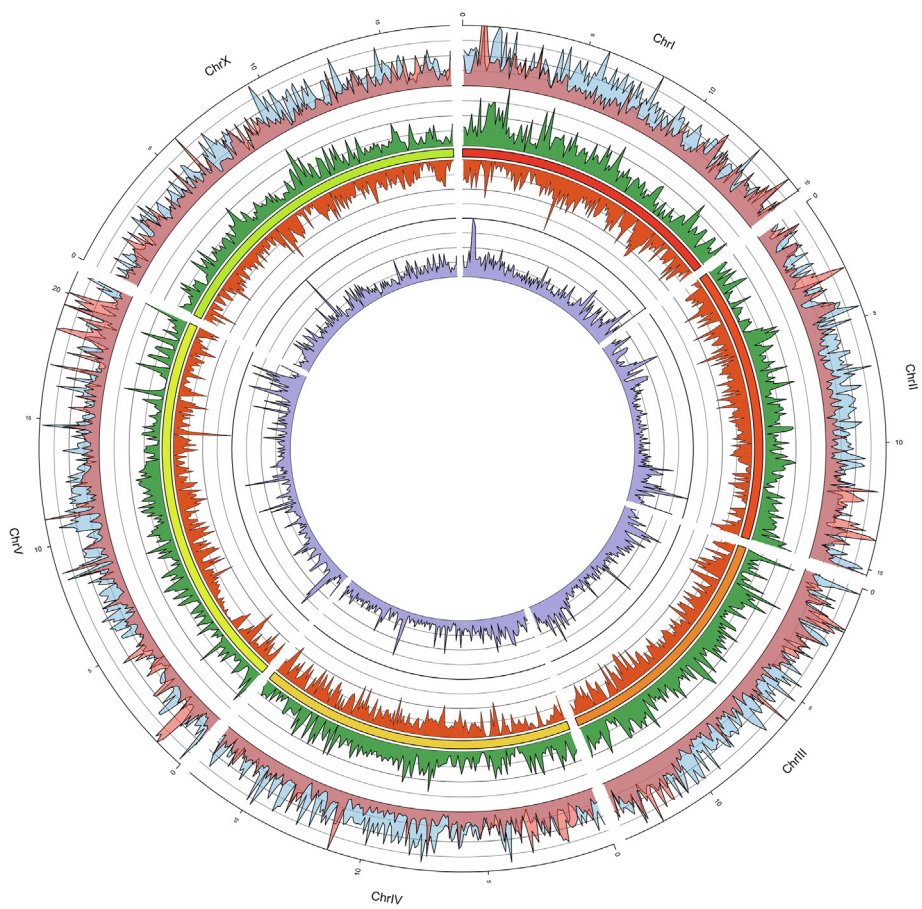


Fig. 4. Comparison of abundance and distribution of the *Anisakis simplex* sensu lato reads mapped to the *Caenorhabditis elegans* genome and *C. elegans* transcripts aligned to its genome. The innermost circle represents *A. simplex* s.l. reads mapped to the *C. elegans* genome. Two middle tracks represent *C. elegans* transcripts aligned to the *C. elegans* genome on reverse (inner of the two circles) and forward (ourter of the two circles) strands. The outermost circle represents alignment of *C. elegans* transcripts to both strands of the *C. elegans* genome (blue (light grey) histogram) covered with *A. simplex* s.l. reads mapped to the *C. elegans* genome (red (medium grey) histogram). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Contigs were compared with the reference genome of *A. simplex* s.l. (Wormbase, https://parasite.wormbase.org/), to reveal 7607 unique sequences, which could point to the presence of Anisakidae-specific genes. These results can be used to identify new genes. Homologs were not detected in the remaining nematode species. *Anisakis simplex s.l.* harbored 2573; 2119; 3611 and 2440 unique sequences that were homologous with the families Ascaridae, Filariidae, Rhabditidae and Toxocaridae, respectively (Fig. 5). The highest number of homologs was identified in the genome of *A. suum*. A comparison with *C. elegans* confirmed

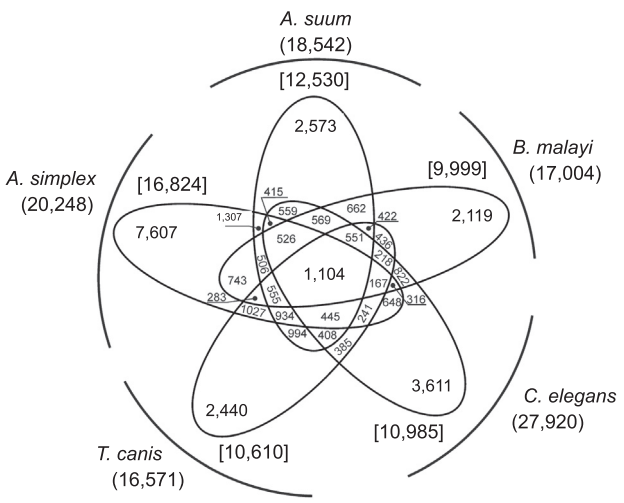


Fig. 5. Venn diagram with the numbers of homologous protein between *Anisakis simplex* sensu lato and four other nematode species (*Ascaris suum, Caenorhabditis elegans, Brugia malayi* and *Toxocara canis)* after pairwise comparison.

the presence of evolutionary distance, and only 30% of homologous genes were identified. *Anisakis simplex* s.l. harbored 12,530 homologous sequences with *A. suum*. The number of homologs would be higher if the entire phylum Nematoda was sequenced. A total of 60% of homologous sequences were identified in *A. simplex* s.l., *A. suum* and *T. canis* which are bound by phylogenetic relationships and belong to the parasitic clade III (Fig. 5).

The contig sequences from the assembly process were used to identify potential coding regions in precursor polypeptides. Five protein databases (*A. simplex*, *A. suum*, *B. malayi*, *C. elegans* and *T. canis*) were scanned with BLASTx (cut-off = 10e-5) to reveal 16,824; 12,530; 9999; 10,985 and 10,610 peptide homologs, respectively (Fig. 5). Cavallero et al. (2018) predicted 19,403 peptides in a comparison of transcriptomes specific for *A. simplex* sensu stricto (s.s.) and *A. pegreffii*.

3.5. Functional classification based on GO and KEGG assignments

The subset of genes evaluated with GO programs provides insights into the functioning of cellular and metabolic pathways in different stages of the *A. simplex* s.l. life cycle. In the group of predicted contigs, 3344 were annotated in the InterProScan database, 1680 were annotated in the GO database, and 1300 was annotated in the KEGG database. InterProScan revealed protein encoding genes annotated in PANTHER (741), PFAM (2265) and SMART (338), including 242 oxidoreductases, 668 transferases, 677 hydrolases, 105 lyases, 88 isomerases and 73 ligases.

Contigs denoting the presence of a potential coding region (based on the results of alignment to four Nematoda references: *A. suum, B. malayi, C. elegans* and *T. canis*) were scanned in BLAST2GO and classified into three main categories: biological process, cellular component and molecular function. As a result, 4676; 8740 and 16,015 contigs were assigned to GO terms. In the group of biological processes, 1295 contigs were annotated to the development of nematode larvae, 424 to metabolic processes, 240 to proteolysis, 180 to innate immune responses, and 276 to the molting cycle, production of collagen and cuticulin-based cuticle (Fig. 6A). The most frequently predicted genes were identified in the nucleus (821), cytoplasm (745), plasma membrane (373), mitochondrion (236) and cytosol (200) (Fig. 6B). More than 680 contigs were assigned to binding processes (ATP, proteins, metals, nucleotides, zinc ions, GTP, DNA). The key identified molecular functions were hydrolase activity (169 contigs), catalytic activity (113 contigs), GTPase activity (148 contigs), protein serine/threonine kinase activity (128 contigs) and calcium ion binding (127 contigs) (Fig. 6C).

The KEGG analysis revealed 1641 genes annotated to 119 metabolic pathways. The highest number of genes were identified for five metabolic pathways: purine metabolism (762 sequences assigned to 42 groups of enzymes), thiamin metabolism (607/1), biosynthesis of antibiotics (168/74), aminobenzoate degradation (111/3) and starch and sucrose metabolism (107/20) (Fig. 6D).

All contigs were scanned to determine the functions of protein products. The contigs classified to the biological process category were directly assigned to an appropriate category. Selected processes were classified into 11 categories related to the biology of parasitic nematodes (cuticle/apoptosis, ubiquitin, receptor, immunity, transport, development, oxidative stress, mitogen activated protein kinase (MAPK), carbohydrate metabolism, neuromuscular transporters drug target, housekeeping genes) and compared with the Nematoda reference data in BLASTp. Genes were selected based on >90% coverage with reference data (Supplementary Table S3).

4. Discussion

The metabolism of carbohydrates (typically glucose) is one of the main biochemical processes that generate energy which is temporarily stored in cells as ATP. Carbohydrates play important roles in development, morphogenesis, and immunity. Cell–cell interactions and host–pathogen interactions are typical examples of the key roles played by carbohydrates in living organisms (Committee on Assessing the Importance and Impact of Glycomics and Glycosciences (https://www.nap.edu/initiative/committee-on-assessing-the-importance-and-impact-of-glycomics-and-glycosciences). Previous attempts at cloning single genes, analyzing their expression (Łopieńska-Biernat et al., 2015) and explaining the molecular mechanisms underpinning carbohydrate metabolism did not bring the anticipated results. The proper molecular analysis of carbohydrate metabolism in *A. simplex* s.l. became possible only with the onset of genomic sequencing and availability of reference genomes of other parasitic nematodes. In this study, genome-wide analysis of *A. simplex* s.l. was done to describe the metabolism of glycogen and trehalose, the most important saccharides in the life cycle of parasitic nematodes. These interconnected metabolic pathways have been studied in *A. simplex* s.l. at the molecular level, to our knowledge for the first time. Through this the identified fragment of starch and sucrose pathways (KEGG), based on the genomic sequences of *A. simplex* s.l., was the premise for more comprehensive description of a glycogen and trehalose metabolism pathway in this nematode (Supplementary Fig. S1).

Most of 22 parasite-specific enzymes identified in *Haemonchus contortus* participate in amino acid and carbohydrate metabolism (Laing et al., 2013). The transcripts of *A. simplex* and *A. pegreffii* also confirmed the expression of 34.5% of the genes involved in sugar metabolism, physiology and transport (Cavallero et al., 2018). The presence of 21 trehalose and glycogen coding sequences, in particular the expression of five mRNA genes (Fig. 7A) encoding trehalose and glycogen in *A. simplex* s.l., indicate that these genes play crucial physiological roles in parasitic nematodes.

In this study we described the complete exon–intron structure of 21 unigenes and isoforms involved in glycogen and trehalose pathways in *A. simplex* s.l. (Fig. 7A, B). All sequences have been deposited in the GenBank NCBI database (Supplementary Table S1). Three genes were assigned to the trehalose pathway, nine genes were allocated to the glycogen pathway, and two transporter genes have been identified. These are among others: hexok-

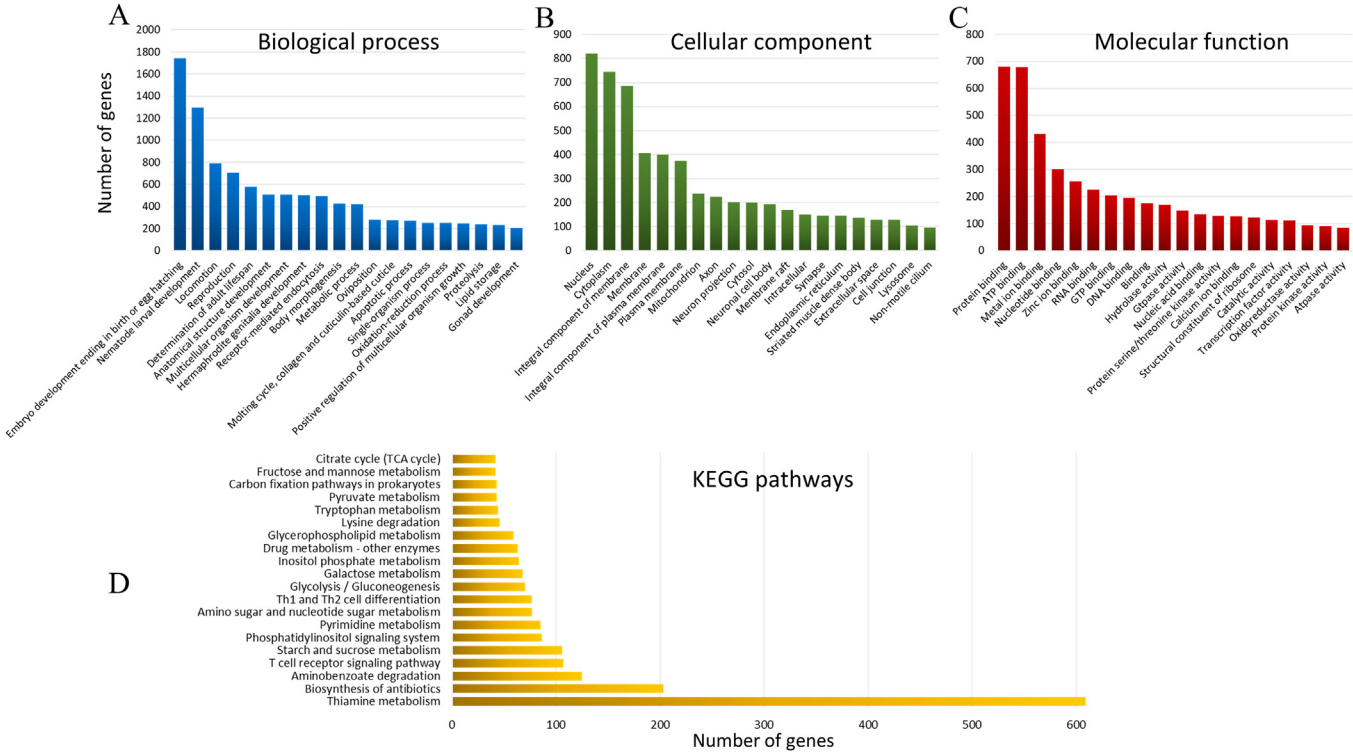


Fig. 6. Prediction of *Anisakis simplex* sensu lato gene function according to Gene Ontology categories. (A) Biological process, (B) cellular component, and (C) molecular function. The Kyoto Encyclopedia of Genes and Genomes (KEGG) classification of metabolic pathways is also shown (D).

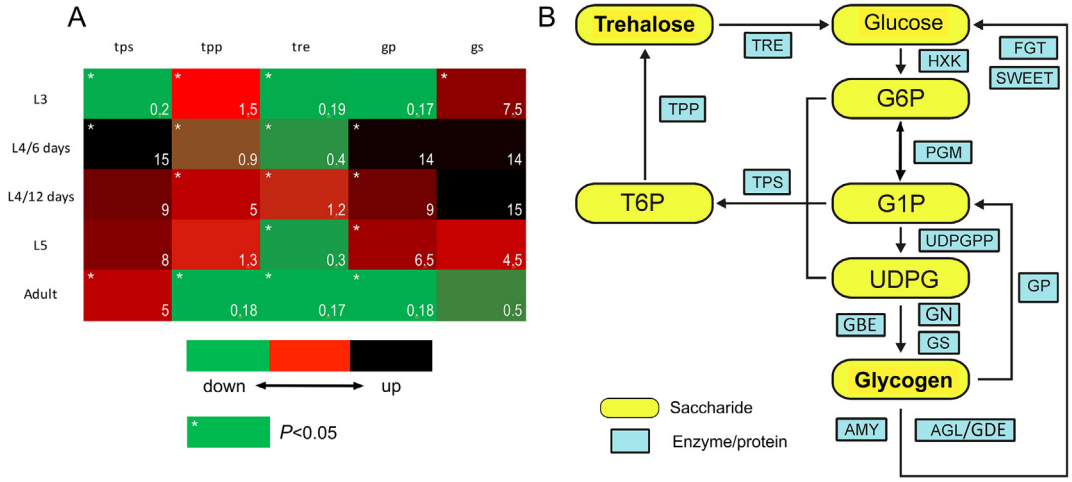


Fig. 7. Expression of mRNA trehalose and glycogen metabolism enzymes and pathway of carbohydrates metabolism of *Anisakis simplex* sensu lato. (A) Differentially expressed mRNAs of trehalose and glycogen metabolism enzymes between developmental stages of *A. simplex* s.l. Color differences indicate up- or down-regulation of gene expression. Additionally, values of relative gene expression (a.u.) were added. The expression of selected genes in all stages was calculated using the $2^{-\Delta\Delta CT}$ method, where the control value was the mean of expression of two housekeeping genes, *ef-1α* and *ppi12*, obtained separately for each stage. Significant differences between samples were identified with a Student's *t*-test at $^{*}P < 0.05$. (B) The glycogen and trehalose metabolism pathway in L3s of *A. simplex* s.l. T6P, trehalose 6-phosphate; G6P, glucose 6-phosphate; G1P, glucose 1-phosphate; UDPG, Uridine diphosphate glucose; TPP, trehalose-6-phosphate phosphatase, 3.1.3.12; TPS, trehalose-6-phosphate synthase, 2.4.1.15; TRE, trehalase, 2.4.1.28; HXK, hexokinase, 3.2.1.28; PGM, phosphoglucomutase, 5.4.2.2; UDPGPP, UTP-glucose-1-phosphate uridylyltransferase, 2.7.7.9; GBE, 1,4-alpha-glucan branching enzyme, 2.4.1.18; GN, glycogenin glucosyltransferase, 2.4.1.186; GS, glycogen synthase; 2.4.1.11; AMY, alpha-amylase, 3.2.1.1; AGL/GDE, amylo-alpha-1,6-glucosidase, 3.2.1.33/4-alpha-glucanotransferase, 2.4.1.25; GP, glycogen phosphorylase, 2.4.1.1; FGT, facilitated glucose transporters; SWEET, "sugars will eventually be exported" transporter family. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

inase and a multi-functional enzyme acting as 1,4-α-glucan, 4α-D-glucosyltransferase, and amylo-1,6-glucosidase with independent catalytic sites (*AGL*/GDE). The structures of the genes responsible for saccharide transport, the facilitated glucose transporters (*fgt*) and the SWEET (sugars will eventually be exported transporter) family transporters (*sweet1*), have also been described in *A. simplex* s.l. for the first known time (Fig. 7B).

Parasites typically have multiple gene copies. Gene families are established by duplication and the accumulation of mutations in copies. This process leads to the generation of genes encoding new proteins that are structurally similar but have new properties. A pseudogene is created when a mutation in a gene copy deactivates that gene. A loss of 5000 protein domains was reported in *Schistosoma japonicum*. For example, the invadolysin metallopro-

tease was detected in 12 copies in *S. japonicum*, but only one copy was identified in *C. elegans* and fruit flies (Zhou et al., 2009). In *A. simplex* s.l., five saccharide pathway genes, trehalose-6-phosphate synthase (TPS, two copies), trehalase (TRE, five copies) (Łopieńska-Biernat et al., 2019a), phosphoglucomutase (PGM, two copies), hexokinase (HX, two copies) and FGT (four copies), were detected and classified as created as a result of duplication, which is a common feature in parasites that facilitates adaptation to variable environmental conditions. Analyses of the mRNA of genes encoding trehalose metabolism revealed only one active isoform of the TPS gene. It is possible that TPS2 occurs as a pseudogene, which is often the case in parasites. In the future, transcriptomic studies of *A. simplex* s.l. conducted with the use of the RNA interference method (RNAi) will support detailed analyses of the expression of active genes and the identification of the relevant null mutant strain of *A. simplex* s.l.

Specific gene expression is related to the functions and pathogenicity of different development stages in parasites. These changes should be compared and described during the development of *A. simplex*. Non-feeding L3s rely only on endogenous sources of energy, and their metabolic rate is decreased to enable survival. Gene transcription and protein biosynthesis occur only in the L4 stage, and these processes probably involve carbohydrates. Sugars play a very important role in anaerobic metabolism, in particular glycogen phosphorylase (GP) and glycogen synthase (GS) (Komuniecki and Harris, 1995). In this study, the mRNA expression of five genes encoding trehalose and glycogen metabolism was compared in five developmental stages of *A. simplex* s.l. cultured in vitro in an environment characteristic of the final host or an accidental host of this parasite (Fig. 7A). The *tps*, *tpp* and *tre* genes play an important role in trehalose metabolism, whereas *gf* and *gs* genes encode glycogen catabolism and anabolism. Significant changes were observed in the L4 stage, where the expression of trehalose and glycogen genes increased more than eight-fold (Fig. 7A). Trehalose anabolism and glycogen catabolism were the predominant processes in stages L4 and L5, which could confirm our and other authors' previous reports that trehalose is synthesized at the expense of glycogen (Behm, 1997; Solomon et al., 2000; Łopieńska-Biernat et al., 2006; Łopieńska-Biernat et al., 2019b). The results of studies into the carbohydrate metabolism of plant and animal pathogens indicate that glycogen phosphorolysis (glycogen phosphorylase and glycogen debranching enzyme) plays an important role in pathogenesis (Corral-Ramos and Roncero, 2015; Gupta et al., 2017). This situation can be explained by the observation that the loss of amyloglucosidase or glycogen phosphorylase activity in *Magnaporthe oryzae* changes trehalose synthesis by decreasing the expression of TPS1 several-fold during infection (Badaruddin et al., 2013). The presence or absence of glucose could induce or inhibit virulence, respectively, due to the fact that glucose is the product of metabolic pathways discussed above, which are the most important in nematode physiology. Therefore, reduction in the glucose supply may weaken the nematode organisms and reduce virulence, and inversely, increase in the glucose supply may strengthen the nematode organism and enhance virulence. This factor should also be considered in nematodes. Glycogen and trehalose accumulation are interconnected and respond to stimuli that are at least partially overlapping in the pathways of different species (Lillie and Pringle, 1980; Wharton et al., 2002). Trehalose and glycogen metabolism are interconnected because they rely on the same substrate, UDP-glucose (Fig. 7B). This link implies that the metabolism of both sugars can be regulated. Two isoforms of PGM also control trehalose and glycogen pathways, and support the synthesis of substrates (G1P or G6P) for these saccharides. Pioneering studies have been conducted into glycogen metabolism genes. Additionally, previous studies have demonstrated that the trehalose synthesis pathway can be a potential drug target because humans do not have trehalose anabolism enzymes (Cross et al., 2015; Łopieńska-Biernat et al., 2018).

In *A. simplex* s.l., carbohydrate metabolism also involves phosphoglycerate mutases (PGAMs) (MG210789) which participate in glycolytic and gluconeogenic pathways in reversible isomerization of 3-phosphoglycerate. Both pathways can be controlled by 2-phosphoglycerate. PGAMs are members of two distinct protein families that are or are not dependent on the 2,3-bisphosphoglycerate cofactor (iPGAM) (Mercaldi et al., 2012). Cofactor-independent phosphoglycerate mutase is a glycolytic enzyme responsible for the transport of carbohydrates, metabolism, catalytic activity and development (Blackburn et al., 2014). Additionally, GDP-D-glucose phosphorylase has been recently identified in mammals and *C. elegans*. This enzyme could prevent the mis-substitution of mannose residues with glucose in glycoproteins and other glycoconjugates (Adler et al., 2011). In *A. simplex* s.l., this gene (MG210751) should be further studied to broaden our understanding of the molecular basis of GDP-D-Glc-forming activity of GDP-D-Man pyrophosphorylase, the major mannosyl donor for the synthesis of N-linked glycoproteins and glycosylphosphatidylinositol membrane anchors, and to determine whether GDP-D-Glc phosphorylase is needed for functional glycoprotein biosynthesis. These processes are very important in studies into an antigenic variation of parasitic glycoproteins.

Multicellular organisms metabolize sugars such as glucose and saccharose for translocation between cells, tissues and organs. Sugar uptake and efflux across cell membranes are among the most important processes for cell growth and development. Two main sugar transporter superfamilies of the major facilitator superfamily (MFS) have been identified: (i) human GLOB glucose uniporters (FGT) and (ii) the SWEET family of sugar transporters (Reddy et al., 2012) (Supplementary Table S3). The SWEET plays an important role in pathogenesis and sensitivity. In *C. elegans*, the SWEET (*swt-1*) homolog participates in glucose and trehalose transport and plays important physiological role. *Caenorhabditis elegans* nematodes where *swt-1* expression was inhibited by RNAi were characterized by fewer offspring, changes in life span and lipid content (Xuan et al., 2013). The SWEET and FGT sugar transporters act as uniporters and facilitate the import and efflux of sugars from cells. In animals, carbon skeletons and energy are required for the efflux of cellular glucose (Chen et al., 2010). In trematodes *Schistosoma masoni* and *Echinococcus granulosus*, glucose can be transported across the tegument, and transporter proteins are expressed in these species (Zhang et al., 2013). The data relating to the expression of the glucose transporter in *C. elegans* (FGT1) indicate that this transporter is not glucose-specific. It is less specific than in humans, and it probably transports other hexoses such as fructose, galactose and mannose (Kitaoka et al., 2013). Silencing of the FGT1 gene in *C. elegans*, as well as *swt-1*, leads to an extension of lifespan by limiting glucose supply and inhibiting glycolysis (Feng et al., 2013), which creates new opportunities for research into parasites.

The identification of *fgt* isoforms and the *sweet* transporters in *A. simplex* s.l. (Supplementary Table S3) could provide valuable information for description of the transport mechanism across the cuticle in parasitic nematodes and for determination of whether trehalose, an important disaccharide, could be of exogenous origin. Transporter proteins have not been identified in *A. simplex* s.l. to date, and our results pave the way to conduct new research.

The present findings revealed significant biological changes during the development of *A. simplex* s.l.: (i) environmental information processing is converted to genetic information processing; (ii) energy production switches from carbohydrate degradation to protein metabolism; and (iii) molecular adaptation in free-living larvae leads to parasitism in the invasive larval stage.

The results of this study expand our understanding of molecular mechanisms of developmental processes in parasitic helminths. Our study contributes novel information about the genome of *A. simplex* s.l., a ubiquitous zoonotic pathogen. The presented genomic data can be used to elucidate complex nutrition and metabolic mechanisms in roundworms, host-parasite interactions, development and maturation processes, as well immune evasion strategies. In this study, a genome-wide analysis of *A. simplex* s.l. helped to describe the metabolism of glycogen and trehalose, the most important saccharides in the life cycle of parasitic nematodes. These interconnected metabolic pathways have been studied in *A. simplex* s.l. at the molecular level for the first known time. Hopefully, our findings will contribute to specific and targeted research and the development of new anthelmintic drugs and vaccines for humans and livestock. The resulting data (i.e. 162 full gene sequences) have numerous applications in functional genomics and identification of putative protein sequences in proteomics. Greater effort is required to translate genomic data into functional data to improve our understanding of the biology of *A. simplex* s.l. in the parasitic stage and identification of new targets for parasite control.

Acknowledgements

This work was supported by the Dean of the Faculty of Biology and Biotechnology, University of Warmia and Mazury in Olsztyn, Poland (12.610.012.300).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijpara.2019.06.006.

References

Adler, L.N., Gomez, T.A., Clarke, S.G., Linster, C.L., 2011. A novel GDP-D-glucose phosphorylase involved in quality control of the nucleoside diphosphate sugar pool in *Caenorhabditis elegans* and mammal. J. Biol. Chem. 286, 21511–21523.

Antonovics, J., Wilson, A.J., Forbes, M.R., Hauffe, H.C., Kallio, E.R., Leggett, H.C., Longdon, B., Okamura, B., Sait, S.M., Webster, J.P., 2017. The evolution of transmission mode. Phil. Trans. R. Soc. B 372, 20160083.

Audicana, M.T., Ansotegui, I.J., Fernández de Corres, L., Kennedy, M.W., 2002. *Anisakis simplex*: dangerous—dead and alive?. Trends Parasitol. 18, 20–25.

Badaruddin, M., Holcombe, L.J., Wilson, R.A., Wang, Z.Y., Kershaw, M.J., Talbot, N.J., 2013. Glycogen metabolic genes are involved in trehalose-6-phosphate synthase-mediated regulation of pathogenicity by the rice blast fungus *Magnaporthe oryzae*. PLoS Pathog. 9, (10). https://doi.org/10.1371/journal.ppat.1003604 e1003604.

Baird, F.J., Gasser, R.B., Jabbar, A., Lopata, A.L., 2014. Foodborne anisakiasis and allergy. Mol. Cell Probes 28, 167–174.

Baird, F.J., Su, X., Aibinu, I., Nolan, M.J., Sugiyama, H., Otranto, D., Lopata, A.L., Cantacessi, C., 2016. The *Anisakis* transcriptome provides a resource for fundamental and applied studies on allergy-causing parasites. PLoS Negl. Trop. Dis. 10, (7) e0004845.

Bao, M., Pierce, G.J., Pascual, S., González-Muñoz, M., Mattiucci, S., Mladineo, I., Cipriani, P., Bušelić, I., Strachan, N.J.C., 2017. Assessing the risk of an emerging zoonosis of worldwide concern: anisakiasis. Sci. Rep. 7, 43699.

Behm, C.A., 1997. The role of trehalose in the physiology of nematodes. Int. J. Parasitol. 27, 215–229.

Blackburn, E.A., Fuad, F.A.A., Morgan, H.P., Nowicki, M.W., Wear, M.A., Michels, P.A.M., Fothergill-Gilmore, L.A., Walkinshaw, M.D., 2014. Trypanosomatid phosphoglycerate mutases have multiple conformational and oligomeric states. Biochem. Biophys. Res. Commun. 450, 936–941.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Brindley, P.J., Mitreva, M., Ghedin, E., Lustigman, S., 2009. Helminth genomics: the implications for human health. PLoS Negl. Trop. Dis. 3, (10) e538.

Cavallero, S., Lombardo, F., Su, X., Salvemini, M., Cantacessi, C., D'Amelio, S., 2018. Tissue-specific transcriptomes of *Anisakis simplex* (sensu stricto) and *Anisakis pegreffii* reveal potential molecular mechanisms involved in pathogenicity. Parasit. Vector 11, 31.

Chen, L.-Q., Hou, B.-H., Lalonde, S., Takanaga, H., Hartung, M.L., Qu, X.-Q., Guo, W.-J., Jung-Kim, J.-G., Underwood, W., Chaudhuri, B., Chermak, D., Antony, G., White, F.F., Somerville, S.C., Mudgett, M.B., Frommer, W.B., 2010. Sugar transporters for intercellular exchange and nutrition of pathogens. Nature 468, 527–532.

Corral-Ramos, C., Roncero, M.I.G., 2015. Glycogen catabolism, but not its biosynthesis, affects virulence of *Fusarium oxysporum* on the plant host. Fungal Genet. Biol. 77, 40–49.

Cross, M., Rajan, S., Chekaiban, J., Saunders, J., Hamilton, C., Kim, J.-S., Coster, M.J., Gasser, R.B., Hofmann, A., 2015. Enzyme characteristics of pathogen-specific trehalose-6-phosphate phosphatases. Sci Rep. 7. https://doi.org/10.1038/s41598-017-02220-2.

Daschner, A., Cuellar, C., Rodero, M., 2012. The *Anisakis* allergy debate: does an evolutionary approach help?. Trends Parasitol. 28, 9–15.

Feng, Y., Williams, B.G., Koumanov, F., Wolstenholme, A.J., Holman, G.D., 2013. FGT-1 is the major glucose transporter in *C. elegans* and is central to aging pathways. Bioch. J. 456, 219–229.

Gilabert, A., Curran, D.M., Harvey, S.C., Wasmuth, J.D., 2016. Expanding the view on the evolution of the nematode dauer signaling pathways: refinement through gene gain and pathway co-option. BMC Genomics 17, 476.

Grabda, J., 1976. Studies on the life cycle and morphogenesis of *Anisakis simplex* (Rudolphi, 1809) (Nematoda: Anisakidae) cultured *in vitro*. Acta Ichthyol. Piscat. 6, 119–141.

Gupta, A.K., Singh, A., Singh, S., 2017. Glycogen as key energy storehouse and possibly responsible for multidrug resistance in *Mycobacterium tuberculosis*. In: Gunjan, A., Andaleeb, S., Vipin Chandra, K. (Eds.), Drug Resistance in Bacteria, Fungi, Malaria, and Cancer. Springer International Publishing, Switzerland, pp. 263–285.

Hotez, P.J., Kamath, A., 2009. Neglected tropical diseases in sub-saharan Africa: review of their prevalence, distribution, and disease burden. PLoS Negl. Trop. Dis. 3 (8), e412.

Iglesias, L., Adroher, F.J., 1997. Some factors which influence the *in vitro* maintenance of *Anisakis simplex* (Nematoda). Folia Parasitol. 44, 297–301.

Iglesias, L., Valero, A., Benítez, R., Adroher, F.J., 2001. *In vitro* cultivation of *Anisakis simplex*: pepsin increases survival and moulting from fourth larval to adult stage. Parasitology 1123, 285–291.

International Helminth Genomes Consortium, 2019. Comparative genomics of the major parasitic worms. Nature 51, 163–174.

Ivanović, J., Baltić, M.Ž., Bošković, M., Kilibarda, N., Dokmanović, M., Marković, R., Janjić, J., Baltić, B., 2017. *Anisakis* allergy in human. Trends Food Sci. Technol. 59, 25–29.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.

Kim, K.H., Eom, K.S., Park, K.K., 2006. The complete mitochondrial genome of *Anisakis simplex* (Ascaridida: Nematoda) and phylogenetic implications. Int. J. Parasitol. 36, 319–328.

Kitaoka, S., Morielli, A.D., Zhao, F.-Q., 2013. FGT-1 is a mammalian GLUT2-like facilitative glucose transporter in *Caenorhabditis elegans* whose malfunction induces fat accumulation in intestinal cells. PLoS One 8, (6) e68475.

Klimpel, S., Palm, H.W., 2011. Anisakid Nematode (Ascaridoidea) Life cycles and distribution: increasing zoonotic potential in the time of climate change? In: Mehlhorn, H. (Ed.), Progress in Parasitology vol. 2. Parasitology Research Monographs. Springer, Berlin, Heidelberg, pp. 201–222.

Klimpel, S., Palm, H.W., Ruckert, S., Piatkowski, U., 2004. The life cycle of *Anisakis simplex* in the Norwegian Deep (northern North Sea). Parasitol. Res. 94, 1–9.

Komuniecki, R., Harris, B.G., 1995. Carbohydrate and energy metabolism in helminths. In: Marr, J.J., Muller, M. (Eds.), Biochemistry and Molecular Biology of Parasites. Academic Press Limited, London, pp. 19–23.

Kuhn, T., Cunze, S., Kochmann, J., Klimpel, S., 2016. Environmental variables and definitive host distribution: a habitat suitability modelling for endohelminth parasites in the marine realm. Sci. Rep. 6, 30246.

Laing, R., Kikuchi, T., Martinelli, A., Tsai, I.J., Beech, R.N., Redman, E., Holroyd, N., Bartley, D.J., Beasley, H., Britton, C., et al., 2013. The genome and transcriptome of *Haemonchus contortus*, a key model parasite for drug and vaccine discovery. Genome Biol. 14, R88.

Lillie, S.H., Pringle, J.R., 1980. Reserve carbohydrate metabolism in *Saccharomyces cerevisiae*: responses to nutrient limitation. J. Bacteriol. 1, 1384–1394.

Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C}_{T}$ method. Methods 24, 402–408.

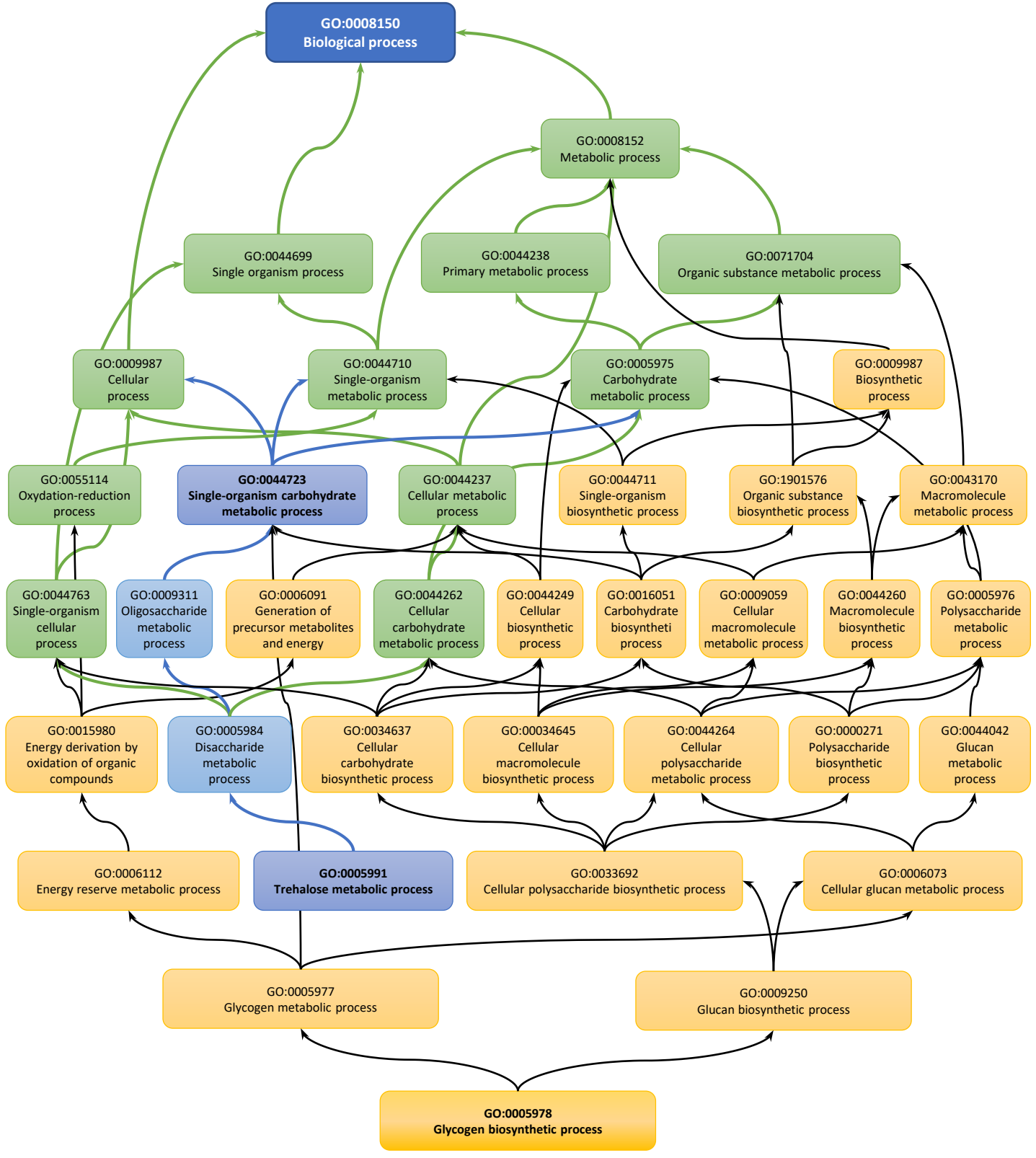
Łopieńska-Biernat, E., Molcan, T., Paukszto, Ł., Jastrzębski, J.P., Myszczyński, K., 2018. Modelling studies determing the mode of action of anthelmintics inhibiting *in vitro* trehalose-6-phosphate phosphatase (TPP) of *Anisakis simplex* s.l. Ex. Parasitol. 184, 46–56.

Łopieńska-Biernat, E., Paukszto, Ł., Jastrzębski, J.P., Makowczenko, K., Stryiński, R., 2019a. Genes expression and *in silico* studies of function of trehalases, a highly dispersed *Anisakis simplex* specific gene family. Int. J. Biol. Macromol. 129, 957–964.

Łopieńska-Biernat, E., Stryiński, R., Dmitryjuk, M., Wasilewska, B., 2019b. Infective larvae of *Anisakis simplex* (Nematoda) accumulate trehalose and glycogen in response to starvation and temperature stress. Biology Open 8 (3), bio040014.

Łopieńska-Biernat, E., Zaobidna, E.A., Dmitryjuk, M., 2015. Expression of genes encoding the enzymes for glycogen and trehalose metabolism in L3 and L4 larvae of *Anisakis simplex*. J. Parasitol. Res 438145.

Łopieńska-Biernat, E., Żółtowska, K., Rokicki, J., 2006. The content of carbohydrates in larval stages of *Anisakis simplex* (Nematoda, Anisakidae). Helmintologia 43, 125–129.
Lv, Z., Wu, Z., Zhang, L., Ji, P., Cai, Y., Luo, S., Wang, H., Li, H., 2015. Genome mining offers a new starting point for parasitology research. Parasitol. Res. 114, 399–409.
Mattiucci, S., Cipriani, P., Webb, S.C., Paoletti, M., Marcer, F., Bellisario, B., Gibson, D. I., Nascetti, G., 2014. Genetic and morphological approaches distinguish the three sibling species of the *Anisakis simplex* species complex, with a species designation as *Anisakis berlandi* n. sp. for *A. simplex* sp. C (Nematoda: Anisakidae). J. Parasitol. 100, 199–214.
Mehrdana, F., Buchmann, K., 2017. Excretory/secretory products of anisakid nematodes: biological and pathological roles. Acta Vet. Scand. 59, 42.
Mercaldi, G.F., Pereira, H.M., Cordeiro, A.T., Michels, P.A., Thiemann, O.H., 2012. Structural role of the active-site metal in the conformation of *Trypanosoma brucei* phosphoglycerate mutase. FEBS J. 279, 2012–2021. https://doi.org/10.1111/j.1742 4658.2012.08586.x.
Messina, C.M., Pizzo, F., Santulli, A., Bušelić, I., Boban, M., Orhanović, S., Mladineo, I., 2017. *Anisakis pegreffii* (Nematoda: Anisakidae) products modulate oxidative stress and apoptosis-related biomarkers in human cell lines. Parasite Vectors 9, 607.
Mohandas, N., Jabbar, A., Podolska, M., Zhu, X.Q., Littlewood, D.T., Jex, A.R., Gasser, R. B., 2014. Mitochondrial genomes of Anisakis simplex and Contracaecum osculatum (sensu stricto)–comparisons with selected nematodes. Infect. Genet. Evol. 21, 452–462.
Mott, R., 1997. EST_GENOME: a program to align spliced DNA sequences to unspliced genomic DNA. Comput. Applic. 13, 477–478.
Nieuwenhuizen, N.E., 2016. *Anisakis* – immunology of a foodborne parasitosis. Parasite Immunol. 38, 548–557.
Pellerone, F.I., Archera, S.K., Behm, C.A., Grant, W.N., Lacey, M.J., Somerville, A.C., 2003. Trehalose metabolism genes in *Caenorhabditis elegans* and filarial nematodes. Int. J. Parasitol. 33, 1195–1206.
Pravettoni, V., Primavesi, L., Piantanida, M., 2012. *Anisakis simplex*: current knowledge. Eur. Ann. Allergy Clin. Immunol. 44, 150–156.
Ramakers, C., Ruijtera, J.M., Lekanne Depreza, R.H., Moorman, A.F.M., 2003. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci. Lett. 339, 62–66.
Reddy, V.S., Shlykov, M.A., Castillo, R., Sun, E.I., Saier Jr, M.H., 2012. The major facilitator superfamily (MFS) revisited. FEBS J. 279, 2022–2035.
Sakanari, J.A., McKerrow, J.H., 1989. Anisakiasis. Clin. Microbiol. Rev. 2, 278–284.
Sanchez-Alonso, I., Carballeda-Sangiao, N., Gonzalez-Munoz, M., Navas, A., Arcos, S. C., Mendizabal, A., Tejada, M., 2018. Pathogenic potential of *Anisakis* L3 after freezing in domestic freezers. Food Control, 8461e69.
Solomon, A., Salomon, R., Paperna, I., Glazer, I., 2000. Desiccation stress of entomopathogenic nematodes induces the accumulation of a novel heat-stable protein. Parasitology 121, 409–416.
Stanke, M., Morgenstern, B., 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33 (Web Server issue), W465–W467. https://doi.org/10.1093/nar/gki458.
Viney, M., 2017. How can we understand the genomic basis of nematode parasitism?. Trends Parasitol. 33, 6.
Wharton, D.A., Goodall, G., Marshall, C.J., 2002. Freezing rate affects the survival of a short-term freezing stress in *Panagrolaimus davidi*, an Antarctic nematode that survives intracellular freezing. Cry. Lett. 2, 5–10.
Xuan, Y.H., Hu, Y.B., Chen, L.-Q., Sosso, D., Ducat, D.C., Hou, B.-H., Frommer, W.B., 2013. Functional role of oligomerization for bacterial and plant SWEET sugar transporter family. PNAS 110 (39), E3685–E3694.
Yamada, A., Ikeda, N., Ono, H., 2017. The complete mitochondrial genome of *Anisakis pegreffii* Campana-Rouget & Biocca, 1955, (Nematoda, Chromadorea, Rhabditida, Anisakidae) – clarification of mitogenome sequences of the *Anisakis simplex* species complex. Mitochondrial DNA Part B 2, 240–241.
Yil, C.J., Murrell, K.D., Lymbery, A.J., 2005. Fish-borne parasitic zoonoses: status and issues. Int. J. Parasitol. 35, 1233–1254.
Yu, H.S., Park, S.K., Lee, K.H., Jeong, H.J., 2007. *Anisakis simplex*: analysis of expressed sequence tags (ESTs) of third-stage larva. Ex. Parasitol. 117, 51–56.
Zhang, M., Hong, Y., Han, Y., Han, H., Peng, J., Qiu, C., Yang, J., Lu, K., Fu, Z., Lin, J., 2013. Proteomic analysis of tegument-exposed proteins of female and male *Schistosoma japonicum* Worms. J. Proteome Res. 12, 5260–5270.
Zhou, Y., Zheng, H., Chen, X., Zhang, L., Wang, K., Guo, J., Huang, Z., Zhang, B., Huang, W., Jin, K., et al., 2009. The *Schistosoma japonicum* genome reveals features of host-parasite interplay. Nature 460, 345–351.


GO:0008150
Biological process
GO:0008152
Metabolic process
GO:0044699
Single organism process
GO:0044238
Primary metabolic process
GO:0071704
Organic substance metabolic process
GO:0009987
Cellular process
GO:0044710
Single-organism metabolic process
GO:0005975
Carbohydrate metabolic process
GO:0009987
Biosynthetic process
GO:0055114
Oxydation-reduction process
GO:0044723
Single-organism carbohydrate metabolic process
GO:0044237
Cellular metabolic process
GO:0044711
Single-organism biosynthetic process
GO:1901576
Organic substance biosynthetic process
GO:0043170
Macromolecule metabolic process
GO:0044763
Single-organism cellular process
GO:0009311
Oligosaccharide metabolic process
GO:0006091
Generation of precursor metabolites and energy
GO:0044262
Cellular carbohydrate metabolic process
GO:0044249
Cellular biosynthetic process
GO:0016051
Carbohydrate biosyntheti process
GO:0009059
Cellular macromolecule metabolic process
GO:0044260
Macromolecule biosynthetic process
GO:0005976
Polysaccharide metabolic process
GO:0015980
Energy derivation by oxidation of organic compounds
GO:0005984
Disaccharide metabolic process
GO:0034637
Cellular carbohydrate biosynthetic process
GO:00034645
Cellular macromolecule biosynthetic process
GO:0044264
Cellular polysaccharide metabolic process
GO:0000271
Polysaccharide biosynthetic process
GO:0044042
Glucan metabolic process
GO:0006112
Energy reserve metabolic process
GO:0005991
Trehalose metabolic process
GO:0033692
Cellular polysaccharide biosynthetic process
GO:0006073
Cellular glucan metabolic process
GO:0005977
Glycogen metabolic process
GO:0009250
Glucan biosynthetic process
GO:0005978
Glycogen biosynthetic process

Supplementary Table S2. The primers of trehalose and glycogen metabolism genes used for real-time quantitative PCR assays for verification of mRNA levels in *Anisakis simplex* sensu lato developmental stages.

Gene name	GeneBank Accession no.	Primer sequence (5'-3')	
tps	KJ560557	For. ACTACGCATCACAAGCAACG	Rev. CGCTTGACGTATCAATGGAA
tpp	KJ560558	For. ATCGGATGATACGCTGCAAG	Rev. TACGAAAGCATTCCAACGTG
tre	KJ406707	For. AAAGGTCTGCCAACGTCACT	Rev. TTAACCACTGTGTCGCCATC
gs	KJ789949	For. GCCACCGATTTGCACACACAA	Rev. CAGCCACGCAAATTCCTCA
gp	KJ560559	For. GGATTCGCACACAACAATACTA	Rev. RCTGCGTCTTCTCGATCTCTTG
ppi 12	KM496568	For. GGCACGATATTCCACAGGAT	Rev. CTCCATAGATCGATGCACCA
ef-1α	KP326558	For. TCCTCAAGCGTTGTTATCTGTT	Rev. AGTTTTGCCACTAGCGGTTCC

Supplementary Table S4. The comparison of the *Anisakis simplex* sensu lato draft genome and genomes of other nematodes published to date.

Description	**Clade III**					**Clade V**	
	Anisakis simplex	***Anisakis simplex***	***Toxocara canis***	***Brugia malayi***	***Ascaris suum***	***Haemonchus contortus***	***Caenorhabditis elegans***
Project	(Velvet/ SOAP)	PRJEB496	PRJEB533	PRJNA10729	PRJNA80881	PRJEB506	Laing et al., 2013
Genome size (Mb)	144/72	127	317	90-95	250	60	100
Scaffold N50 (kb)	6.9 /1.3	9.3	375	94	408	83	17.5
Genome GC content %	37.42/37.4	36.7	49.9	30.5	37.9	41.3	35.4
Number of contigs	59,943/69,529	49,180	51,969	26,589	30,842	51,174	n.a.
Haploid chromosome #	n.a.	n.a.	n.a.	6	12	6	6
Total size of contigs (Mb)	143/72	127	n.a.	n.a.	n.a.	930.9	n.a.
Mean size of contigs (bp)	2,397/1,035	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Minimum size of contigs (bp)	200/400	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Maximum size of contigs (kb)	90,7/18,105	427,7	n.a.	n.a.	n.a.	n.a.	n.a.

Mean exon number per gene	10.6	n.a.	7.4	5.9	6.4	9.8	6.5
Mean exon length (bp)	148.8	n.a.	156	160	153	126	202
Mean intron length (bp)	286.7	n.a.	1,133	280	1023	601	320
Mean gene size (bp)	4,294	n.a.	8,416	n.a.	6,536	n.a.	1,956
Mean CDS length (bp)	1,234	n.a.	1,156	n.a.	n.a.	n.a.	n.a.

CDS, coding DNA sequence ; n.a., not available