Annals of Parasitology 2021, 67(4), 671–682
doi: 10.17420/ap6704.383 Copyright© 2021 Polish Parasitological Society

Original paper

Transcriptome analysis of *Nosema assamensis* infecting muga silkworms (*Antheraea assamensis*) reveals insights into candidate pathogenicity related genes and molecular pathways required for pathogenesis

Anupama JAGADISH[1,2], Himanshu DUBEY[1], Indumathi KAMATCHI[1], Appukuttannair R. PRADEEP[1], Gangavarapu SUBRAHMANYAM[3], Rakesh K. MISHRA[1], Kangayam M. PONNUVEL[1]

[1]Genomic Division, Seri Biotech Research Laboratory, Carmelaram Post, Kodathi, Bengaluru-560035, India
[2]Department of Biotechnology, School of Sciences, Jain University, Bengaluru-560 027, India
[3]Central Muga Eri Research and Training Institute (CMER&TI), Lahdoigarh-785700, Jorhat (Assam), India

Corresponding Author: Kangayam M. Ponnuvel; e-mail: kmpvel@gmail.com

ABSTRACT. Muga silkworms are often prone to many diseases since, these are non-domesticated and are reared outdoors. Microsporidia, an obligate intracellular pathogen with spore as its active form, causes pebrine disease in these silkworms. The study has attempted to categorise the transcript data of the *Nosema* obtained from the infected muga silkworm using gene ontology and KEGG pathway studies. A total of 2850 unigene sets were identified out of which 2739 unigenes were placed under biological, cellular as well as molecular function categories based on the gene ontology (GO) terms. 1620 out of these unigenes sets found their orthologous partner in the corresponding *Nosema bombycis* transcriptome. The unigenes were found to be enriched under organic substance metabolic process, organic cyclic compound binding and intracellular anatomical structure for biological process, molecular function and cellular components respectively. The KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis indicated majority of the enzymes were found to be mapped under purine and thiamine metabolic pathways, indicating an increase in the energy metabolism required to establish the infection in the silkworms. The putative virulence genes identified in this study are PTP2, PTP3, SWP12 and SWP26 which were found to be expressed in other *Nosema* species indigenous to India, indicating a probable conservation of these genes, which are primarily involved in establishing host pathogen interactions. The expression of these genes was in detectable levels in the infected silkworm samples. These genes may be validated further through bioassay in order understand their roles in establishing the infection and propagation of the spores. The identified virulence genes may be further targeted to develop diagnostic tools for identification of the pathogen at early stages of infection.

Keywords: muga, transcriptome, *Nosema*, silkworms, virulence genes

Introduction

Muga silk is an indigenous variety of silk produced by *Antheraea assamensis*, a saturniid wild silkworm reared in north eastern states of India. Pebrine, is a devastating disease affecting not only domesticated silkworms (viz., *Bombyx mori*) but also wild silkworms such as muga and tasar. These silkworms are reared by the local tribe in the north eastern states of India and are often polyphagous, hence making it difficult to rear them under controlled housing conditions [1]. Microsporidia, an obligate intracellular parasite, currently classified under the group of fungi is the causal agent for the pebrine disease; possessing some unique abilities conserved through evolution [2]. The pebrine disease is highly devastating and causes immense damage to the muga crop (40%) leading to

economic losses to the local tribe and population in north eastern regions [1]. High mortality (52–100%) of muga silkworms were observed in cross-infectivity studies (cross infected with *N. bombycis*, *N. mylitta*, *N. ricini* and *N. assamensis*) especially in the month of August–September [3]. These microsporidians are in possession of a unique apparatus called as "polar tube" which acts like an anchor to hook the pathogen to the infecting host cell, mainly to establish the infection by forming a passage for delivering the infective cargo of "sporoplasm" which hosts the necessary apparatus required for its propagation [4]. Their genome size is highly variable and is compact due to loss of various metabolic related genes making them extremely dependent on the host for their survival [5]. Analysis of transcriptome data is of extreme significance and the studies by Wu et al. [6] are indicative that the transcriptome data provides an understanding on the molecular mechanism involved in host pathogen interactions, the gene expressions and immune response. The spore germination is the key step in infection and happens in the highly alkaline conditions as described in the literature [7–9]. A detailed study on the sporoplasm (the infecting cargo delivered to the host) morphology and its transcriptome study has revealed that the sporoplasm gain entry into host cell via phagocytosis. Further the transcriptome study also revealed that there were 541 genes which were upregulated mainly the 10 transporter genes whereas the other energy related metabolic genes were downregulated indicating the pathogens dependency on the invading host for obtaining its majority of energy [10]. Transcriptome data characterisation on germinated and ungerminated spores of *N. bombycis* revealed 66 genes to be differentially expressed, the data presents genes expressed during germination. These data sets provide target functional genes which can be exploited for further understanding the germination and infection process [11]. A transcriptome study on the detailed response to *N. bombycis* infection in *B. mori* has revealed up to 124 genes that were modulated related to basal metabolic pathways. The genes associated with juvenile hormone synthesis were upregulated indicating the response of the host to the invading pathogen. It was also noted that the serine protease cascade melanisation pathway was inhibited by the serpins secreted by the *N. bombycis*. The study also highlighted the increase in production of antimicrobial peptides, induction of toll and JAK/STAT pathway. The study in general provided various details on the host parasite interaction [12].

There have been many studies associated with the transcriptome analysis of *N. bombycis*, however, there are not many studies associated with the microsporidia infecting muga silkworms. The present study is believed to provide initial insights into the transcriptome data on the infecting microsporidia *N. assamensis* [13], particularly on the genes that are expressed upon the invasion of host. The available transcriptome data was also used in identification of putative virulence factors based on the earlier reference on putative virulence factors referred in honey bee infecting microsporidia, *N. apis* based on its genome sequencing and comparative genomics data. The genes were further validated for its presence in other *Nosema* species infecting silkworm's native to India.

Materials and Methods

Raw data pre-processing and transcript assembly

The raw reads were obtained from the RNA sequencing data of muga silkworm generated in our lab. Total RNA was isolated from mid gut and fat body tissues using RNeasy animal blood tissue system (Qiagen, US) using manufacturers protocol. High quality RNA was used for sequencing library preparation using NEB Next® ultra™ directional RNA library Prep kit (NEB, UK). The pair end sequencing was performed on Illumina Hiseq platform. Since muga silkworms are reared outdoors the probability of getting microsporidia infection is highly likely, hence the raw reads were aligned against *Nosema* genome (NCBI genome accession no. GCA_000383075.1). Seven samples from fat body and midgut were found to contain reads matching with *Nosema* sp. The raw reads from these samples were extracted, and the quality of these reads were examined by FASTQc tool (https://www.bioinformatics.babraham.ac.uk/ projects/fastqc/). High-quality reads were de novo assembled using Trinity v2.11.0 [14]. The assembled reads were further clustered by CD-HIT [15] to obtain the unigenes sets.

The assembled unigenes of *N. assamensis* were compared with *N. bombycis* in order to check for homology between these 2 species using BLAST tool. In addition, to find the homology of *N. assamensis* unigene sets with other *Nosema* spp. at

amino acid level, the TransDecoder tool [16] was used to translate the unigene sets. The translated amino acid sequences were used in OrthoVenn 2 (https://orthovenn2.bioinfotoolkits.net/home) for checking for the presence of orthologous genes with other *Nosema* species (*N. antheraea*, *N. apis*, *N. bombycis*, *N. cerenae*, *N.* sp. YNPr) [17].

Functional annotation and pathway analysis

To obtain functional annotation information, unigenes sets were compared with NCBI non-redundant database using the Blast software [18]. The Gene Ontology (GO), including molecular functions, biological processes, and cellular components, and metabolic pathways analysis (KEGG database) [19] were performed using omics box version 1.4 [20].

In silico expression of virulence related genes

To check the in silico expression analysis, raw reads were mapped to assembled unigene set using HISAT, StringTie program [21]. The FPKM values of the all the PTP1(Polar Tube Protein 1), PTP3 (Polar Tube Protein 3), SWP4 (Spore Wall Protein 4), SWP9 (Spore Wall Protein 9), SWP12 (Spore Wall Protein 12), SWP13 (Spore Wall Protein 13), SW and ADCP (Spore Wall and Associated Complex Protein) and SWP30 (Spore Wall Protein 30) were plotted as a heatmap generated through heat mapper tool (www.heatmapper.ca).

Identification and validation of pathogenicity related genes

The putative virulence genes identified in Chen et al. [22] were used and compared against the unigenes sets from transcriptome in this study. Four genes showed a significant match with the putative virulence factor sequences mentioned in the said article. The identified genes were used to design primers for qPCR analysis.

To validate the expression of the identified putative virulence genes the infected muga silkworms were collected from CMER&TI, Assam. The silk moth was further screened for pebrine (microsporidia) infection using pathogen specific *β-tubulin* primer [23]. RNA was extracted from infected silk moth using RNAiso Plus (TAKARA). For cDNA synthesis the PrimeScript 1st strand cDNA synthesis kit (TAKARA Bio) method was used following manufacturers recommended protocol. The RNA was treated with DNase prior to synthesis of cDNA using DNase I Amplification grade (Invitrogen by Thermo Fischer Scientific, USA).

Quantitative Real Time PCR reaction mix consisted of 0.1 µl (100µM) each of forward and reverse primers, 5 µl of the SYBR green master mix, 0.1 µl of ROX and the volume made up to 10 µl with water after accounting for the template addition of 1 µl. The reaction parameters were as follows: 95°C for 2 min followed by 35 cycles of 95°C for 30 s, 60°C for 30 s and 72°C for 35 s. The Ct values were noted and were used in calculation of the copy number based on the standard curve of *β-tubulin*. The analysis was performed on 3 technical replicates for each gene. The statistical analysis for the different genes was performed to check for the significance of variation in expression between the genes (*β-tubulin*, PTP2, PTP3, SWP12 and SWP26) based on the Tukey HSD post -hoc test following one-way ANOVA test (https://astatsa.com /OneWay_ Anova_with_TukeyHSD/).

Cross species presence of putative virulence genes

To check the conservation of the identified virulence genes in other *Nosema* sp., the designed primers were also checked for amplification in conventional PCR using DNA samples obtained from the isolated *Nosema* spores from infected *B. mori*, *A. mylitta* (tasar silkworm) and *Samia cynthia ricini* (eri silkworm). The genomic DNA was isolated form spores using initial percoll gradient centrifugation for spore isolation and subsequent DNA extraction using the tissue DNA extraction kit (Qiagen) [23].

PCR reaction mix consisted of 0.25 µl (250µM) each of forward and reverse primers, 5 µl of the emerald master mix (TAKARA Bio, USA) and the volume made up to 10 µl with water after counting for template 1 µl. The reaction parameters were as follows: 94°C for 2 min followed by 30 cycles of 94°C for 30 s, 60°C for 30 s and 72°C for 35 s. The PCR mixtures were run on 1.2% agarose gel at 100V for 30 min with ethidium bromide of 3 µl in the gel.

Results

Assembly and annotation

A total of 5990 contigs with length of ≥ 200 bps were generated. The average contig lengths obtained were observed to be ~618 bps. Further clustering of these transcripts with ≥ 95% of sequence identity yielded a total of 2850 unigenes

Table 1. Assembly statistics of the raw data reads obtained from the transcriptome study

Assembly statistics	
Contigs Generated:	5990
Maximum Contig Length:	4692
Minimum Contig Length:	201
Average Contig Length:	618
Total Contigs Length:	3701767
Total Number of Non-ATGC Characters:	0
Percentage of Non-ATGC Characters:	0
Contigs >= 100 bp:	5990
Contigs >= 200 bp:	5990
Contigs >= 500 bp:	2794
Contigs >= 1 Kbp:	871
Contigs >= 10 Kbp:	0
N50 value:	759
At the level of 90% identity	2850
At the level of 95% identity	2850

(Tab. 1). These unigenes were further utilised for functional annotation and pathway studies to derive conclusions. The total assembled unigenes sets are available from the Corresponding Author.

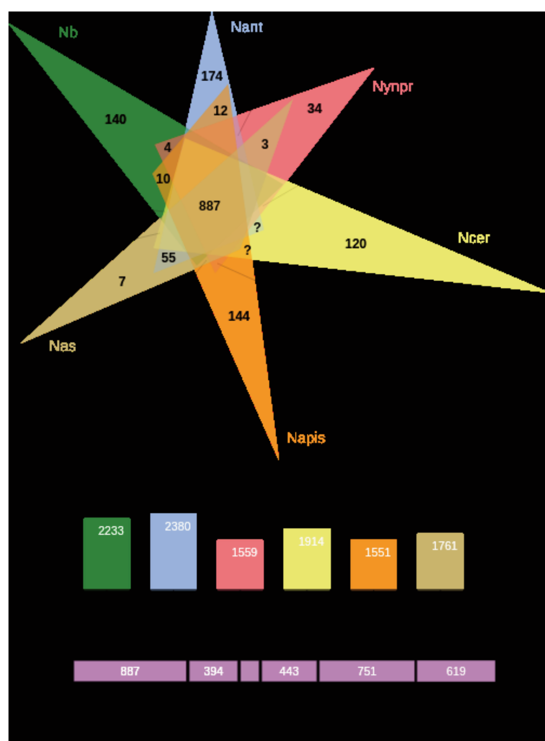

Figure 1. Venn chart of the transcripts in comparison with other *Nosema* spp.
Nb: *N. bombycis*, *Nant*: *N. antheraea*, *Nynpr*: *N.* sp. YNPr isolate, *Ncer*: *N. cerenae*, *Napis*: *N. apis*, *Na*: *N. assamensis*

The homology identification of assembled unigenes with *N. bombycis* yielded a total of 1620 unigenes that were found to be homologous to the genes in *N. assamensis* (Tab. 2; details available from the Corresponding Author).

Out of 2850 unigenes, only 2396 were translated to protein sequences of ≥ 100aa using TransDecoder tool. These sequences were used for ortholog analysis among 5 different *Nosema* spp. (*N. apis*, *N. antheraea*, *N.* sp. YNPr, *N. cerenae* and *N. bombycis*) yielded 449 singletons and 7 clusters (15 proteins) found unique to *N. assamensis* (Fig. 1).

Functional annotation and gene ontology

To ascertain the biological roles of the identified 2850 unigenes the OMICS box software was utilized and the unigenes were BLAST searched against NCBI-nr database (e-value 10^{-5}) and functional domain information were obtained by InterProscan search [24]. The BLAST search yielded a total of 2739 functional annotations of these genes; the remaining reads did not give any hits in the NCBI database. Out of these annotations 141 genes encoded for putative proteins and total of 606 genes encoded for hypothetical proteins. Among the functionally annotated genes 50 coded for transporter proteins which mostly consisted of glucose, sugar transporters and others were protein transporters. 20 spore wall protein genes and 6 polar tube protein related genes were observed which form the core structure of the *Nosema* pathogen. 86 nucleotide binding proteins were observed along with few DNA binding proteins in the RNA transcript pool extracted from the infected muga moth samples.

The unigenes sets were further mapped under different categories *i.e.*, biological process, molecular function and cellular process based on the GO terms. Within the biological process category, most representative category was organic substance metabolic process (769/21%) (GO:0071704) and cellular metabolic process (763/21%) (GO:004423 7) followed by primary metabolic process (733/20%) (GO:0044238), nitrogen compound metabolic process (702/19%) (GO:0006807), biosynthetic process (361/10%) (GO:0009058), establishment of localisation (177/5%) (GO:0051234) and cellular component organisation or biogenesis (133/4%) (GO:0071840) (Fig. 2a).

For molecular function, the most enriched GO

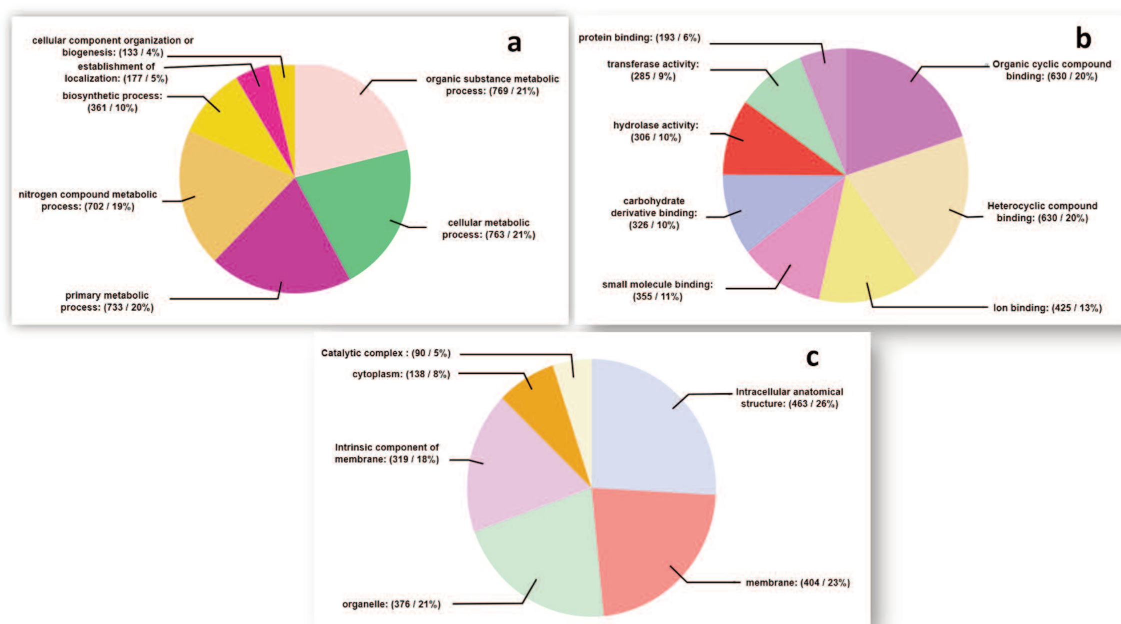

Figure 2. a – Gene enrichment of the identified unigene sets at level 3 for biological process,
b – Gene enrichment of the identified unigene sets at level 3 for molecular function,
c – Gene enrichment of the identified unigene sets at level 3 for cellular components

components were mapped under organic cyclic compound binding (GO:0097159) and heterocyclic compound binding (630/20%) (GO:1901363), followed by ion binding (425/13%) (GO:0043167) (Fig. 2b).

Intracellular anatomical structure (463/26%) (GO:0005622) topped the major components of unigenes available at the cellular components GO enrichment category, followed by membrane proteins (404/23%) (GO:0016020), organelle (376/21%) (GO:0043226), intrinsic component of membrane (319/18%) (GO:0031224) (Fig. 2c). Among the enriched GO terms, the major category was molecular function followed by biological process and cellular components.

KEGG pathway analysis

All the identified unigene sets were further subjected to KEGG pathway analysis for functional annotation of the genes and to further understand their biological functions. The unigenes were mapped to 27 different pathways. Out of these pathways, 36 unigenes were mapped for purine metabolism indicating that the majority of the genes were under this category with the majority of the enzymes being mapped to phosphatase. Another 36 transcripts were found to be mapped under thiamine metabolism. A major enzyme mapped to the enzyme phosphatase was found to be associated with purine metabolism pathway. Further 3 unigenes were mapped under mTOR signalling pathway, relaxin pathway and PI3K-Akt signalling pathway. Three genes were mapped to amino acid metabolism and glutathione metabolism. Two enzymes were found to be mapped to amino sugar and nucleotide sugar metabolism required for nucleotide biosynthesis. Remaining unigenes were mapped to fatty acid elongation, alanine, aspartate and glutamate metabolism (Fig. 3). Among the identified pathways the genes were enriched towards, purine metabolism, thiamine metabolism and amino acid metabolism.

Table 2. Homology comparison of transcripts of *N. assamensis* with *N. bombycis*

Total unigenes of *N. assamensis* obtained	2850 unigenes
Total unigenes homologous with unigenes from *N. bombycis*	1620 unigenes
Unigenes specific to *N. assamensis*	1230 unigenes

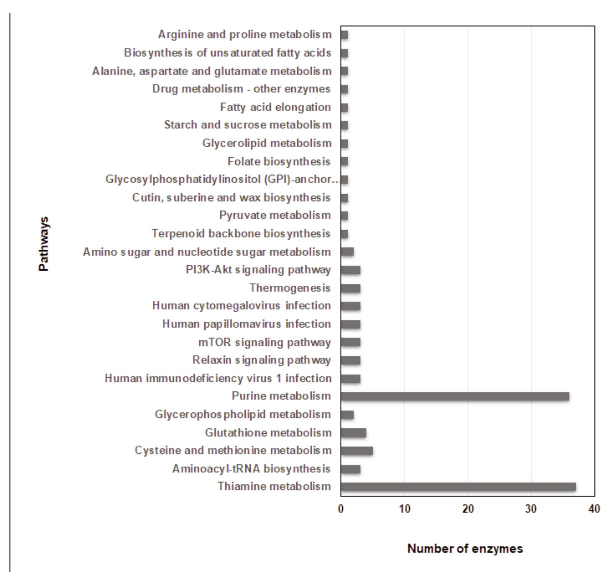

Figure 3. Enzymes identified in KEGG analysis shows the distribution of the unigenes, across different metabolic pathways. Purine metabolism and thiamine metabolism has many genes/ enzymes mapped under them, indicating the abundance of these 2 metabolic pathways in the infection

Identification and validation of putative virulence genes

The putative virulence factors identified by Chen et al. [22] were used to find their homologs in *N. assamensis* unigene sets. These putative virulence factors were characterized based on the functions associated with the genes. A total of four unigenes showed significant hits (e-value 10^{-20}, bit score ≥ 100). These unigenes were corresponding to PTP2 (Polar Tube Protein 2), PTP3 (Polar Tube Protein 3), SWP12 (Spore Wall Protein 12) and SWP26 (Spore Wall Protein 26) which have been indicated in various earlier studies as some of the important components of microsporidia in invading and establishing the infection in the host cell.

The putative virulence genes were identified as being expressed through qPCR assay in the cDNA samples of the infected muga moth (Tab. 3). With reference to *β-tubulin* gene the copy numbers were found to be in the decreasing order of expression starting with PTP3, SWP26, SWP12 and PTP2 (Fig. 4). The critical value for Tukey-Kramer HSD Q-statistic was established based on the 5 genes including *β-tubulin* and ν=10 degrees of freedom for the error term, for significance level α=0.01 and 0.05 (*P*-values) in the Studentized Range distribution.

The Tukey HSD post-hoc statistical analysis indicated no significant differences between the

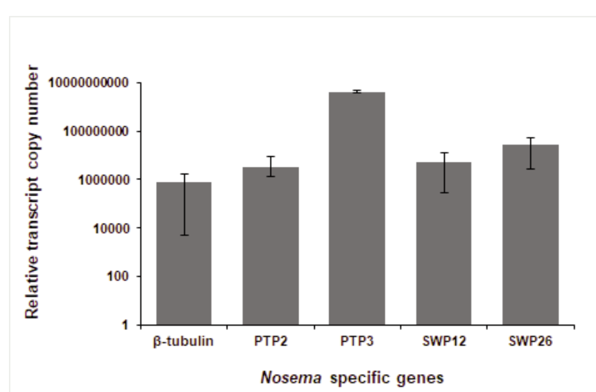

Figure 4. qPCR analysis of the expression of the putative virulence genes in the infected muga moth cDNA samples. The genes were found to express in the infected sample in comparison to the expression of *β-tubulin*. The difference in expression of the genes were analysed using Tukey-HSD post- hoc analysis which indicated no significant difference in the level of expression

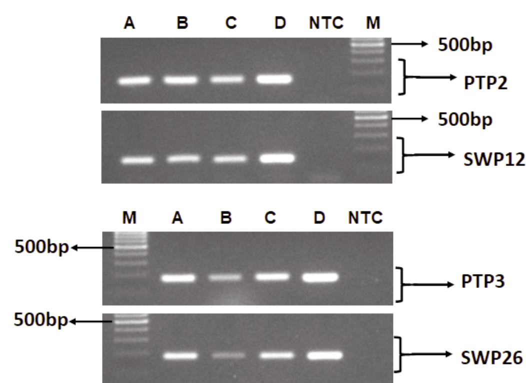

Figure 5. Conventional PCR check for the presence of the virulence genes in the spore DNA of different species of *Nosema*. The genes were found to be expressed is all the species at the genomic level, and seems that these genes may be conserved
A – *Nosema assamensis*; B – *Nosema bombycis*; C – *Nosema mylitta*; D – *Nosema ricini*; NTC – Non template control; M – Marker

expression of the genes with the p-values noted for the *Q*-statistic Q_{ij}.

The presence of these genes was also checked in different *Nosema* sp. infecting *B. mori*, *A. assamensis*, *A. mylitta* and *S. ricini* in conventional PCR via spore DNA as template (Fig. 5). Amplification was noted for all the genes in all the species indicating a probable conservation of these genes in the said species.

In silico expression of the PTP1, PTP3, SWP4, SWP9, SWP12, SWP13, SW and ADCP (spore wall

Table 3. Primer sequences for putative virulence factors/genes identified in this study. The primers were designed based on corresponding transcripts of *N. assamensis*

Primers	Sequence
PolarTubeProtein2_FP	5' GAAGCCACCATCTGAACAATTC 3'
PolarTubeProtein2_RP	5' CTTCAGCAGTTTGTGGAGTAGA 3
PolarTubeProtein3_FP	5' CAGCACCACAATCGAGTTCTA 3'
PolarTubeProtein3_RP	5' GGTGTAGCTGAAGGAGATGAAG 3'
SporeWallProtein12_FP	5' GACGGGAAATGGAAGTGTTAATG 3'
SporeWallProtein12_RP	5' GAGAGCCTGTCTACTTCTTCTTG 3'
SporeWallProtein26_FP	5' ACATGGCTAGTGAAGCCAATAA 3'
SporeWallProtein26_RP	5' GAAACGGCAGCATCAAGAAAC 3'

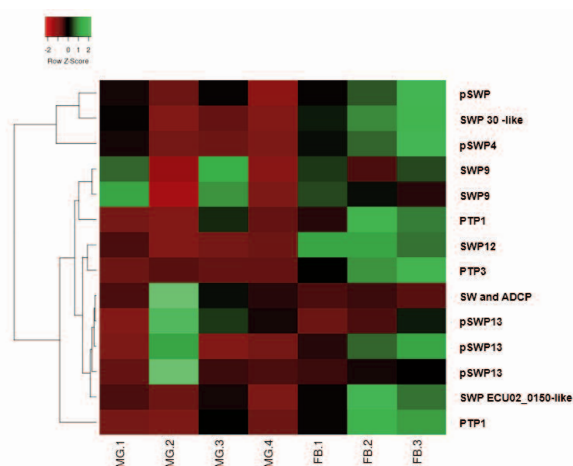

Figure 6. Heat map for SWP and PTP samples expressed in 7 different sample sets from mid gut and fat body. The identified genes were found to be expressed well in case of FB.2 and 3 samples when compared with other samples, MG – midgut; FB – fat body

and associated complex protein) and SWP30, showed a greater number of reads for the above said genes in fat body tissues (Fig. 6).

Discussion

A. assamensis produces golden colour silk indigenous to India, reared in the north eastern states. These silkworms are frequently attacked by the microsporidia, which are the causal agent of the pebrine disease, typically leading to poor cocoon quality. Microsporidia are known obligate parasites which rely on their host for their survival. Despite possessing unique infection apparatus, the organisms are not different rather are primitive and are related to fungi in contrast to their earlier link to protists. Currently there is no information on the genes or transcripts for microsporidian pathogen infecting muga silkworms. This is the first study which aims to provide a brief account on the probable genes that are expressed during the infection process.

In the current study 2739 unigenes were annotated in NCBI and were enriched for GO and KEGG pathways in order to understand the genes that are expressed by *Nosema* during the infection in muga silkworms. The homology search of *N. assamensis* unigenes set yielded 1620 reciprocal BLAST hist against *N. bombycis* gene sequences. Further to analyse the orthologous genes against other *Nosema* spp. such as *N. apis*, *N. antheraea*, *N.* sp. YNPr, *N. cerenae* including *N. bombycis*, revealed that out of 2396 translated protein sequences, 1761 sequences matched with selected *Nosema* species. However, 7 clusters (containing 15 proteins) and 449 singletons (~19%) were found to be specific to *N. assamensis*. Such *Nosema* specific orthologous genes have been observed in earlier studies which give them their identity. It has been observed that 8% genes in *N. antheraea*, 15.7% in *N. bombycis* and 30.5% in *N. cerenae* were reported to be species specific [25]. These specific genes need to be further explored and understand if they provide the desired host specificity to the specific microsporidians.

Organic substance metabolic process and cellular metabolic process being the primary enriched biological process, indicates the significance of metabolism a priority in the

propagating microsporidia inside the host after the spore germination. Organic substances metabolism may involve many substances such as carbohydrate metabolism, lipid metabolism etc., which provide the necessary energy requirements to propagating microsporidia. Dolgikh et al. [26] reported that energy metabolism in *N. grylli* occurs through Embden-Meyerhof Pathway. It was also noted that the absence of hexokinase enzyme in these spores may be the key factor for their survival for a long time in the environment resulted from the low energy metabolism [26]. *Enterocytozoon bieneusi*, a microsporidium infecting the humans do not possess any genes for glycolysis, so it is completely dependent on host to meet ATP requirements for its propagation [27].

Microsporidia do not possess a fully functional mitochondria, instead a mitochondrial remnant as observed in the case of *Trachipleistophora hominis* [28]. Although the microsporidia lack a fully functional mitochondria existence of an alternative respiratory chain in merogonial and early sporogonial stages and also its accumulation in the mature spores have been identified in *Antonospora locustae*, which might provide the desired energy metabolism for the survival of the pathogen especially in sporogenesis. Although the microsporidia lack Krebs cycle, a respiratory chain, they possess genes for glycolysis, ADP/ATP carriers to import the energy molecules from the host cell [29–31]. Microsporidia once inside a host cell rely on the host for the energy requirement and are dependent on the host for their survival, as many genes associated with metabolism have been lost during evolution. Only some of these genes are required for their growth and reproduction have been retained during evolution [32]. Hence, an increase in the organic substance metabolic process may be indicative of a successful establishment of infection in host further modulating the host metabolism to derive energy for its own survival.

Enrichment of organic cyclic compound binding/heterocyclic compound binding under molecular function is indicative of probable nucleotide interactions. An increase in transporter proteins have been observed in the past in various studies indicating that the NTT (nucleotide transporter proteins), as observed in case of *T. hominis*, exploits the ATP generated by the host by transporting these energy rich molecules via purine nucleotide transport proteins located in the plasma membrane [33].

Many of the GO enrichment under cellular component nestled under intracellular anatomical structure indicating the probable intracellular transport of the secretary proteins by microsporidia since, at level 5 of GO analysis (data not shown) 74% of the enrichment was under the intracellular membrane bound organelles and 26% of ribosomes, which may be indicative of increased protein translation coupled with the transport of essential cargo through membrane bound organelles [34].

The KEGG pathway analysis revealed enrichment of the enzymes under the purine metabolism, thiamine pathway and amino acid metabolism. All of which are indicative on the increased metabolic activities for increased energy production to meet the requirements of the propagating microsporidia. As indicated in earlier literature, purine biosynthesis is a major pathway noted in most of the transcriptome studies of microsporidia, transcriptome study and experimental evidence have suggested that microsporidia manipulate their host cell in order to increase their cellular and biological processes for nucleotide biosynthesis which further enhances the potential of ATP/nucleotide import for their survival [35]. Phosphatase being the major enzyme under purine metabolism, can be mentioned as the infection associated protein [11].

It has been demonstrated earlier that PTPs and SWPs are involved in spore adherence and infection to host cell. From our data we were able to locate four transcripts that have been termed as putative virulence factors previously [22]. We found 2 PTP proteins (PTP2 and 3) and 2 SWP (SWP12 and 26) associated with the virulence factors identified from the transcriptome data which matches with the sequences identified in the corresponding genes from the *Nosema* infecting honey bees. The qPCR validation of these genes was performed on SYBR green platform and the genes were found to be expressed in the infected muga moths. The orthologs of these genes were also observed to be present in other *Nosema* species infecting other silkworms via conventional PCR, indicating that the genes may be required for microsporidia hence may be conserved across the *Nosema* genus, providing the required virulence to the infecting pathogen to establish a stable host pathogen interaction. The polar tubes and spore wall proteins often interact to form associations which has been studied in the past giving them the advantage to establish a stable infection [36]. These interactions are found to be necessary for their assembly to orderly orient,

arrange and anchor to the anchoring disk, which is the major machinery required for delivering the sporoplasm [36,37]. Polar tube proteins have been indicated to be present in an unpolymerized state inside polar tubes, which during infection gets concentrates at the growing tip of the polar tube during eversion [37–39]. PTP2 is multigenic and is present in phylogenetically related pathogen families but are absent in mammalian microsporidia. PTP2 protein is highly conserved in properties such as their molecular weight, basic isoelectric point, high lysine content and cysteine residues [40]. PTP3 might be a scaffolding protein playing important in the interactions during the formation of polar tube [41]. So far there are six polar tubes that have been identified which are observed to play an important role during microsporidia infections [42]. From the earlier studies it has been found that the polar tube is not involved in pierce or break of the plasma membrane of host instead pushes the host plasma membrane while still surrounded with the host cell membrane creating a microenvironment for the extrusion of the infective cargo [43–46]. Similarly, spore wall proteins such as EnP1 (*Encephalitozoon cuniculi* ECU01_0820) have been observed to interact with the host cell surface indicating its probable role in regulating host cell adherence and infection [47]. Spore wall protein 5 (SWP5) is found to interact with polar tube proteins (PTP2 and PTP3) in *N. bombycis* potentially maintaining the structural integrity of spore wall and regulation of infections in silkworms [37]. SWP12 is found to be observed in both inside and outside of the microsporidian spore [48]. Whereas SWP26 was found to be expressed in endospore and plasma membrane was observed to be present thinly in the mature spores [49]. The in silico expression analysis depicted the expression of the SWP and PTP commonly expressed across the samples and their expression profile revealed variable expression for the identified genes.

The significance of the earlier data can be linked to the probable correlation of the virulence factors to the infection, which is viewed as essential weapons of microsporidia which equip them with advantages to infect and establish infection in the host cells.

The transcript resources generated in this study will be immensely beneficial for the future research work to further investigate and understand the functional aspects of the genes involved in *Nosema* infection. In addition to this, our preliminary study in virulence genes warrants an in-depth analysis of spore wall protein and polar tube protein during infection process. Further studies may also target in obtaining experimental evidence on the association of many putative virulence genes. The proving of the presence of these virulence genes, their probable variations within species may provide target for the development of species-specific markers, identification and distinction between virulent and avirulent strains and probable RNAi studies.

Acknowledgements

The funding for this study was from the grant provided by CSB under the Project-AIT-5872.

The authors would like to acknowledge the funding and support from Central Silk Board, Bengaluru, India.

References

[1] Tikader A., Vijayan K., Beera S. 2013. Muga silkworm, *Antheraea assamensis* (Lepidoptera: Saturniidae) – an overview of distribution, biology and breeding. *European Journal of Entomology* 110: 293–300. doi:10.14411/eje.2013.096

[2] Han B., Weiss L.M. 2017. Microsporidia: obligate intracellular pathogens within the fungal kingdom. *Microbiology Spectrum* 5: article number 10. doi:10.1128/microbiolspec.FUNK-0018-2016

[3] Chakrabarty S., Manna B., Saha A.K., Bindroo B.B. 2012. Comparative study on effect different types of *Nosema* sp. (Microsporidia: Nosematidae) on mulberry and vanya silkworms. *Journal of Biological Sciences* 12: 1–14. doi:10.3923/jbs.2012.1.14

[4] Jaroenlak P., Cammer M., Davydov A., Sall J., Usmani M., Liang F.X., Ekiert D.C., Bhabha G. 2020. 3-dimensional organization and dynamics of the microsporidian polar tube invasion machinery. *PLOS Pathogens* 16: e1008738. doi:10.1371/journal.ppat.1008738

[5] Keeling P. 2009. Five questions about Microsporidia. *PLOS Pathogens* 5: e1000489. doi:10.1371/journal.ppat.1000489

[6] Wu P., Wang X., Qin G., Liu T., Jiang Y.F., Li M.W., Guo X.J. 2011. Microarray analysis of the gene expression profile in the midgut of silkworm infected with cytoplasmic polyhedrosis virus. *Molecular Biology Reports* 38: 333–341. doi:10.1007/s11033-010-0112-4

[7] Undeen A.H., Vander Meer R.K. 1990. The effect of ultraviolet radiation on the germination of *Nosema algerae* Vávra and Undeen (Microsporida: Nosematidae) spores. *Journal of Protozoology* 37:

194–199. doi:10.1111/j.1550-7408.1990.tb01127.x

[8] Undeen A.H. 1990. A proposed mechanism for the germination of microsporidian (Protozoa: Microspora) spores. *Journal of Theoretical Biology* 142: 223–235. doi:10.1016/S0022-5193(05)80223-1

[9] Undeen A.H., Frixione E. 1990. The role of osmotic pressure in the germination of *Nosema algerae* spores. *Journal of Protozoology* 37: 561–567. doi:10.1111/j.1550-7408.1990.tb01265.x

[10] He Q., Luo J., Xu J-Z., Wang C., Meng X., Pan G.Q., Li T., Zhou Z.Y. 2020. Morphology and transcriptome analysis of *Nosema bombycis* sporoplasm and insights into the initial infection of Microsporidia. *mSphere* 5: doi:10.1128/mSphere.00958-19

[11] Liu H., Li M., He X., Lu X. 2016. Transcriptome sequencing and characterization of ungerminated and germinated spores of *Nosema bombycis*. *Acta Biochimica et Biophysica Sinica* 48: 246–256. doi:10.1093/abbs/gmv140

[12] Ma Z., Li C., Pan G., Li Z., Han B., Xu J., Lan X., Chen J., Yang D., Chen Q., Sang Q., Ji X., Li T., Long M., Zhou Z. 2013. Genome-wide transcriptional response of silkworm (*Bombyx mori*) to infection by the microsporidian *Nosema bombycis*. *PLOS ONE* 8: e84137. doi:10.1371/journal.pone.0084137

[13] Subrahmanyam G., Esvaran V.G., Ponnuvel K.M., Hassan W., Chutia M., Das R. 2019. Isolation and molecular identification of microsporidian pathogen causing nosemosis in muga silkworm, *Antheraea assamensis* Helfer (Lepidoptera: Saturniidae). *Indian Journal of Microbiology* 59: 525–529. doi:10.1007/s12088-019-00822-0

[14] Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N., Regev A. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology* 29: 644–652. doi:10.1038/nbt.1883

[15] Li W., Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22: 1658–1659. doi:10.1093/bioinformatics/btl158

[16] Tang S., Lomsadze A., Borodovsky M. 2015. Identification of protein coding regions in RNA transcripts. *Nucleic Acids Research* 43: e78. doi:10.1093/nar/gkv227

[17] Xu L., Dong Z., Fang L., Luo Y., Wei Z., Guo H., Zhang G., Gu Y.Q., Coleman-Derr D., Xia Q., Wang Y. 2019. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Research* 47: W52–W58. doi:10.1093/nar/gkz333

[18] Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden T. L. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10: article number 421. doi:10.1186/1471-2105-10-421

[19] Kanehisa M., Sato Y., Kawashima M., Furumichi M., Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research* 44: D457–D462. doi:10.1093/nar/gkv1070

[20] Götz S., García-Gómez J.M., Terol J., Williams T.D., Nagaraj S.H., Nueda M.J., Robles M., Talón M., Dopazo J., Conesa A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* 36: 3420–3435. doi:10.1093/nar/gkn176

[21] Pertea M., Kim D., Pertea G.M., Leek J.T., Salzberg S.L. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols* 11: 1650–1667. doi:10.1038/nprot.2016.095

[22] Chen Y.P., Pettis J.S., Zhao Y., Liu X., Tallon L.J., Sadzewicz L.D., Li R., Zheng H., Huang S., Zhang X., Hamilton M.C., Pernal S.F., Melathopoulos A.P., Yan X., Evans J.D. 2013. Genome sequencing and comparative genomics of honey bee microsporidia, *Nosema apis* reveal novel insights into host-parasite interactions. *BMC Genomics* 14: 451. doi:10.1186/1471-2164-14-451

[23] Esvaran V.G., Mohanasundaram A., Mahadeva S., Gupta T., Ponnuvel K.M. 2019. Development and comparison of real-time and conventional PCR tools targeting β-tubulin gene for detection of *Nosema* infection in silkworms. *Journal of Parasitic Diseases* 43: 31–38. doi:10.1007/s12639-018-1053-4

[24] Quevillon E., Silventoinen V., Pillai S., Harte N., Mulder N., Apweiler R., Lopez R. 2005. InterProScan: protein domains identifier. *Nucleic Acids Research* 33: W116–W120. doi:10.1093/nar/gki442

[25] Keeling P.J., Corradi N., Morrison H.G., Haag K.L., Ebert D., Weiss L.M., Akiyoshi D.E., Tzipori S. 2010. The reduced genome of the parasitic microsporidian *Enterocytozoon bieneusi* lacks genes for core carbon metabolism. *Genome Biology and Evolution* 2: 304–309. doi:10.1093/gbe/evq022

[26] Dolgikh V.V., Sokolova J.J., Issi I.V. 1997. Activities of enzymes of carbohydrate and energy metabolism of the spores of the microsporidian, *Nosema grylli*. *Journal of Eukaryotic Microbiology* 44: 246–249. doi:10.1111/j.1550-7408.1997.tb05707.x

[27] Keeling P.J., Corradi N. 2011. Shrink it or lose it: balancing loss of function with shrinking genomes in the microsporidia. *Virulence* 2: 67–70. doi:10.4161/viru.2.1.14606

[28] Williams B.A.P., Hirt R.P., Lucocq J.M., Embley T.M. 2002. A mitochondrial remnant in the microsporidian *Trachipleistophora hominis*. *Nature* 418: 865–869. doi:10.1038/nature00949

[29] Cuomo C.A., Desjardins C.A., Bakowski M.A.,

Goldberg J., Ma A.T., Becnel J.J., Didier E.S., Fan L., Heiman D.I., Levin J.Z., Young S., Zeng Q., Troemel E.R. 2012. Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth. *Genome Research* 22: 2478–2488. doi:10.1101/gr.142802.112

[30] Dolgikh V.V., Senderskiy I.V., Pavlova O.A., Naumov A.M., Beznoussenko G.V. 2011. Immunolocalization of an alternative respiratory chain in *Antonospora* (*Paranosema*) *locustae* spores: mitosomes retain their role in microsporidial energy metabolism. *Journal of Clinical Microbiology* 10: 588–593. doi:10.1128/EC.00283-10

[31] Williams B.A., Haferkamp I., Keeling P.J. 2008. An ADP/ATP-specific mitochondrial carrier protein in the microsporidian *Antonospora locustae*. *Journal of Molecular Biology* 375: 1249–1257. doi:10.1016/j.jmb.2007.11.005

[32] Wadi L., Reinke A.W. 2020. Evolution of microsporidia: an extremely successful group of eukaryotic intracellular parasites. *PLoS Pathogens* 16: e1008276. doi:10.1371/journal.ppat.1008276

[33] Heinz E., Hacker C., Dean P., Mifsud J., Goldberg A.V., Williams T.A., Nakjang S., Gregory A., Hirt R.P., Lucocq J.M., Kunji E.R.S., Embley T.M. 2014. Plasma membrane-located purine nucleotide transport proteins are key components for host exploitation by microsporidian intracellular parasites. *PLoS Pathogens* 10: e1004547. doi:10.1371/journal.ppat.1004547

[34] Beznoussenko G.V., Dolgikh V.V., Seliverstova E.V., Semenov P.B., Tokarev Y.S., Trucco A., Micaroni M., Di Giandomenico D., Auinger P., Senderskiy I.V., Skarlato S. O., Snigirevskaya E.S., Komissarchik Y.Y., Pavelka M., De Matteis M.A., Luini A., Sokolova Y.Y., Mironov A.A. 2007. Analogs of the Golgi complex in microsporidia: structure and avesicular mechanisms of function. *Journal of Cell Science* 120: 1288–1298. doi:10.1242/jcs.03402

[35] Dean P., Hirt R.P., Embley T.M. 2016. Microsporidia: why make nucleotides if you can steal them? *PLoS Pathogens* 12: e1005870. doi:10.1371/journal.ppat.1005870

[36] Yang D., Pan L., Chen Z., Du H., Luo B., Luo J., Pan G. 2018. The roles of microsporidia spore wall proteins in the spore wall formation and polar tube anchorage to spore wall during development and infection processes. *Experimental Parasitology* 187: 93–100. doi:10.1016/j.exppara.2018.03.007

[37] Li Z., Pan G., Li T., Huang W., Chen J., Geng L., Yang D., Wang L., Zhou Z. 2012. SWP5, a spore wall protein, interacts with polar tube proteins in the parasitic microsporidian *Nosema bombycis*. *Eukaryotic Cell* 11: 229–237. doi:10.1128/EC.05127-11

[38] Yang D., Pan L., Peng P., Dang X., Li C., Li T., Long M., Chen J., Wu Y., Du H., Luo B., Song Y., Tian R., Luo J., Zhou Z., Pan G. 2017. Interaction between SWP9 and polar tube proteins of the microsporidian *Nosema bombycis* and function of SWP9 as a scaffolding protein contribute to polar tube tethering to the spore wall. *Infection and Immunity* 85: e00872-16. doi:10.1128/IAI.00872-16

[39] Xu Y., Weiss L.M. 2005. The microsporidian polar tube: a highly specialised invasion organelle. *International Journal for Parasitology* 35: 941–953. doi:10.1016/j.ijpara.2005.04.003

[40] Delbac F., Peuvel I., Metenier G., Peyretaillade E., Vivares C.P. 2001. Microsporidian invasion apparatus: identification of a novel polar tube protein and evidence for clustering of genes in three *Encephalitozoon* species. *Infection and Immunology* 69: 1016–1024. doi:10.1128/IAI.69.2.1016-1024.2001

[41] Bouzahzah B., Nagajyothi F., Ghosh K., Takvorian P.M., Cali A., Tanowitz H.B., Weiss L.M. 2010. Interactions of *Encephalitozoon cuniculi* polar tube proteins. *Infection and Immunology* 78: 2745–2753. doi:10.1128/IAI.01205-09

[42] Lv Q., Wang L., Fan Y., Meng X., Liu K., Zhou B., Chen J., Pan G., Long M., Zhou Z. 2020. Identification and characterization a novel polar tube protein (NbPTP6) from the microsporidian *Nosema bombycis*. *Parasites and Vectors* 13: article number 475. doi:10.1186/s13071-020-04348-z

[43] Magaud A., Achbarou A., Desportes Livage I. 1997. Cell invasion by the microsporidium *Encephalitozoon intestinalis*. *Journal of Eukaryotic Microbiology* 44: 81s–81s. doi:10.1111/j.1550-7408.1997.tb05795.x

[44] Schottelius J., Schmetz C., Kock N.P., Schüler T., Sobottka I., Fleischer B. 2000. Presentation by scanning electron microscopy of the life cycle of microsporidia of the genus *Encephalitozoon*. *Microbes and Infection* 2: 1401–1406. doi:10.1016/s1286-4579(00)01293-4

[45] Bigliardi E., Sacchi L. 2001. Cell biology and invasion of the microsporidia. *Microbes and Infection* 3: 373–379. doi:10.1016/s1286-4579(01)01393-4

[46] Bohne W., Böttcher K., Gross U. 2011. The parasitophorous vacuole of *Encephalitozoon cuniculi*: biogenesis and characteristics of the host cell-pathogen interface. *International Journal of Medical Microbiology* 301: 395–399. doi:10.1016/j.ijmm.2011.04.006

[47] Southern T.R., Jolly C.E., Lester M.E., Hayman J.R. 2007. EnP1, a microsporidian spore wall protein what enables spores to adhere to and infect host cells *in vitro*. *Eukaryotic Cell* 6: 1354–1362. doi:10.1128/EC.00113-07

[48] Chen J., Geng L., Long M., Li T., Li Z., Yang D., Ma C., Wu H., Ma Z., Li C., Pan G., Zhou Z. 2013. Identification of a novel chitin-binding spore wall protein (NbSWP12) with a BAR-2 domain from *Nosema bombycis* (microsporidia). *Parasitology* 140:

1–9. doi:10.1017/S0031182013000875
[49] Li Y., Wu Z., Pan G., He W., Zhang R., Hu J., Zhou Z. 2009. Identification of a novel spore wall protein (SWP26) from microsporidia *Nosema bombycis*. *International Journal for Parasitology* 39: 391–398. doi:10.1016/j.ijpara.2008.08.011

Received 30 April 2021
Accepted 01 August 2021