In Silico Identification and Functional Characterization of Conserved miRNAs in the Genome of *Cryptosporidium parvum*

Bioinformatics and Biology Insights
Volume 15: 1–10

DOI: 10.1177/11779322211027665

Md. Irtija Ahsan[1], Md. Shahidur Rahman Chowdhury[2], Moumita Das[1], Sharmin Akter[1], Sawrab Roy[3], Binayok Sharma[2], Rubaiat Nazneen Akhand[4], Mahmudul Hasan[5], Md Bashir Uddin[2] and Syed Sayeem Uddin Ahmed[1]

[1]Department of Epidemiology and Public Health, Sylhet Agricultural University, Sylhet, Bangladesh. [2]Department of Medicine, Sylhet Agricultural University, Sylhet, Bangladesh. [3]Department of Microbiology and Immunology, Sylhet Agricultural University, Sylhet, Bangladesh. [4]Department of Biochemistry and Chemistry, Sylhet Agricultural University, Sylhet, Bangladesh. [5]Department of Pharmaceuticals and Industrial Biotechnology, Sylhet Agricultural University, Sylhet, Bangladesh.

ABSTRACT: *Cryptosporidium parvum*, a predominant causal agent of a fatal zoonotic protozoan diarrhoeal disease called cryptosporidiosis, bears a worldwide public health concern for childhood mortality and poses a key threat to the dairy and water industries. MicroRNAs (miRNAs), small but powerful posttranscriptional gene silencing RNA molecules, regulate a variety of molecular, biological, and cellular processes in animals and plants. As to the present date, there is a paucity of information regarding miRNAs of *C. parvum*; hence, this study was used to identify miRNAs in the organism using a comprehensible expressed sequence tag–based homology search approach consisting of a series of computational screening process from the identification of putative miRNA candidates to the functional annotation of the important gene targets in *C. parvum*. The results revealed a conserved miRNA that targeted 487 genes in the model organism (*Drosophila melanogaster*) and 85 genes in *C. parvum*, of which 11 genes had direct involvements in several crucial virulence factors such as environmental oocyst protection, excystation, locomotion, adhesion, invasion, stress protection, intracellular growth, and survival. Besides, 20 genes showed their association with various major pathways dedicated for the ribosomal biosynthesis, DNA repair, transportation, protein production, gene expression, cell cycle, cell proliferation, development, immune response, differentiation, and nutrient metabolism of the organism in the host. Thus, this study provides a strong evidence of great impact of identified miRNA on the biology, virulence, and pathogenesis of *C. parvum*. Furthermore, the study suggests that the detected miRNA could be a potential epigenomic tool for controlling the protozoon through silencing those virulent and pathway-related target genes.

KEYWORDS: *Cryptosporidium parvum*, microRNA, targets, functions, pathway analysis

RECEIVED: January 21, 2021. **ACCEPTED:** June 4, 2021.

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Md Bashir Uddin, Department of Medicine, Sylhet Agricultural University, Sylhet 3100, Bangladesh. Email: bashir.vetmed@sau.ac.bd

Syed Sayeem Uddin Ahmed, Department of Epidemiology and Public Health, Sylhet Agricultural University, Sylhet 3100, Bangladesh. Email: ahmedssu.eph@sau.ac.bd

Introduction

Cryptosporidium parvum, a fatal apicomplexan protozoon, is one of the causal pathogens of cryptosporidiosis, a potential zoonotic diarrhoeal disease of humans and animals bearing a noteworthy global public health concern.[1-3] From 1984 to 2015, many countries around the world have experienced foodborne cryptosporidiosis outbreaks that have affected a large number of people, of which about 50% of the outbreaks have been caused only by *C. parvum*.[4] Neither the developed nor the developing countries have escaped from the sufferings of cryptosporidiosis, where the ranges of prevalence of the disease are around 1% to 3% and 5% to 10%, respectively.[1,5] Cryptosporidiosis, the second cause of diarrhoea-related childhood mortality, attributes a substantial amount of mortality in children less than 5 years of age in the developing countries.[1,4] In addition, immunocompromised cryptosporidiosis-affected patients, especially those who are HIV-positive with low CD4+ T-lymphocyte counts, suffer from severe and chronic diarrhoea that may even lead to death.[3,6] Like in humans, *C. parvum*, considered as a key threat to livestock industry, affects a considerable number of cattle, sheep, goat, and buffalo population worldwide and causes profuse diarrhoea in adults and calves, which leads to production loss and mortality.[3,7] An epidemiological study showed that 65% of the dairy farms in New Zealand were affected by *C. parvum*.[8] Delafosse et al[9] reported a high mortality attributed by *C. parvum* in calves having high combined scores of diarrhoea and shedding of oocysts. Furthermore, the organism is a headache for water industry across the globe, compelling them to invest a huge amount of money to improve water quality.[3,10] Thus, the disease costs billions of dollars annually, posing a substantial threat to the global economy.[1,4]

Like humans, animals, and other plants, *C. parvum* is a eukaryotic entity, and microRNAs (miRNAs) are known to have integral functional links to the eukaryotic cells.[2,3,11] These miRNAs are tiny endogenous non-coding powerful regulatory RNA molecules, having around 22 nucleotides[12-14]; these modulate gene expression during the posttranscriptional level by binding predominantly to the 3′ untranslated regions (3′ UTRs), however sometimes by an attachment to the coding

regions and 5′ untranslated regions (5′ UTRs) of target messenger RNA (mRNA) or by splitting them through a process called cleavage.[13,15] The enzyme, RNA polymerase II, induces the biogenesis of miRNA by transcribing the miRNA genes into primary miRNAs (pri-miRNAs), which are successively processed to precursor miRNAs (pre-miRNAs) by the Dicer-like 1 (DCL1) enzyme in the nucleus.[16,17] Later, the protein Exportin 5 (XPO5) plays a vital role in the transportation of the precursor miRNAs from the nucleus to cytoplasm, where DCL1 cleaves them to generate a duplex structure composed of a mature miRNA and its star miRNA (miRNA*).[17,18] Afterwards, a conjunction between duplex's single strand and Argonaute (AGO) produces an RNA-induced silencing complex (RISC), which allows miRNAs to recognize their target mRNAs possessing complementary nucleotide sequences.[13,19] Thus, miRNAs are actively involved in a plethora of molecular, biological, and cellular processes such as metabolism, stem cell division, stem cell differentiation, development, apoptosis, cell proliferation, and cell cycle control in animals, plants, and even in virus.[12,13,20] Besides, researchers use them in the diagnosis of diseases and treatments.[20]

Therefore, identification of the conserved miRNAs has received a great deal of interest in the scientific community, and the researchers have adopted several approaches to detect miRNAs in animals and plants, such as genetic screening, directional cloning, high-throughput sequencing, expressed sequence tag (EST), and genome survey sequence (GSS).[12,14,21] However, researchers consider the EST method, a series of screening process of EST sequences (ESTs) containing the conserved miRNA followed by a secondary structure prediction, advantageous because this technique specifies gene expression and requires relatively simple software.[12] Consequently, many researchers have followed the EST approach to identify the conserved miRNAs in animals and plants.[12-14] According to previous research works, miRNAs play crucial roles in the life cycle and pathogenesis of various pathogens such as viruses, nematodes, and fungi by facilitating an evasion from the host defence, development, replication, and adhesion.[11,22,23] However, to the authors' best knowledge, no studies have been conducted on the miRNA of *C. parvum*, despite the organism having remarkable public heath significance and industrial importance. The miRNA might also be involved in a large number of important functions in this organism; hence, this study would contribute to the better understanding of the biology of the organism and consequent invention of innovative control strategies. Therefore, we designed this study to identify the conserved miRNAs in *C. parvum* adopting the EST-based homology approach and their functional characterization through target gene prediction.

Materials and Methods

Retrieval of the ESTs and the reference miRNAs

A total of 60 548 ESTs of the genome of *C. parvum* (Genome Assembly ID: ASM16534v1) were retrieved (January 15, 2020) from the GenBank nucleotide databases at the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/nucleotide/).[24] MiRBase (http://www.mirbase.org/), a reliable public miRNA repository database, released 22.1 version in 2018 containing a total of 38 589 miRNAs of 271 species; however, no miRNAs of protozoa were available in the database.[25] As a protozoon, *C. parvum* was under the Eukaryota domain, which also included algae, slime moulds, animals, and fungi.[26] Therefore, based on the availability in the miRBase database, algae (*Symbiodinium microadriaticum, Ectocarpus siliculosus, Phytophthora infestans, Phytophthora ramorum, Phytophthora sojae*, and *Phaeodactylum tricornutum*), slime moulds (*Dictyostelium discoideum*), and animals (fly, *Drosophila melanogaster* and worm, *Caenorhabditis elegans*), that is, a total of 9 organisms, were selected as candidates for retrieval of the reference miRNA in this study. The rationale of inclusion of fly and worm from the animal group was that there were used as model organisms for miRNA research works.[27] Finally, a total of 1041 mature reference miRNAs (Supplementary File 1) were downloaded (January 14, 2020) from those selected species from the miRBase.

Search for the potential miRNAs contained ESTs in the genome of C. parvum

We performed BLASTn with the query miRNAs against the ESTs in the genome of *C. parvum* (January 15, 2020) using the BLAST version 2.5.0+ in the Linux environment.[28] The e-value was set as 0.1 and all other parameters were kept as default. This step gave the hits of potential miRNA-contained ESTs among the total ESTs of the *C. parvum* genome, and all related information (accession number, query cover, percent identity) of those ESTs was extracted. Subsequently, redundancy was manually removed by deleting ETSs having the same accession number.

Screening for non-coding ESTs containing miRNAs

ESTs, that we got at the end of the previous step after the BLASTn, were further screened through an online BLASTx software at the NCBI. We set 'non-redundant protein sequences' in the database option and kept other algorisms as defaults where a cut-off value of query cover was considered as less than 40% (January 25, 2020).[28] Through this step, the non-coding ESTs were identified and included for the next step; however, the coding ones were excluded from the study.

Identification of conserved miRNAs with the reported parameters

MiRNAs were finally selected based on the following criteria: (1) the predicted mature miRNAs should contain 19 to 23 nucleotides; (2) the secondary structures of pre-miRNA sequences must exhibit a perfect double-stranded hairpin

stem-loop shape and one of its strand should contain mature miRNA; (3) maximum 3 mismatches should be allowed between predicted mature miRNA strand and star miRNA strand; (4) the A + U content of pre-miRNAs should range from 30% to 70%; and (5) the value of minimal folding free energy (MFE) and minimal free energy index (MFEI) of the secondary structures should be negative and high. These criteria of identifying the conserved miRNA were set according to the previous studies.[11,29-33]

Prediction of pre-miRNAs with the secondary structures

A couple of Web servers named miREval version 2.0 and MFold version 2.3 were used to predict the secondary structures of pre-miRNAs.[34,35] In miREval, nucleotide sequences of EST were given as input, and in return, the software provided a number of suggestions of pre-miRNAs with different lengths from where the best pre-miRNA candidate was chosen based on the result of match between reference miRNAs and ESTs from previous step, BLASTn. Afterwards, upon inserting the nucleotide sequence of pre-miRNA as input in MFold, the software built a secondary structure as output and showed additional information of MFE and the number of each of the nucleotides (A, U, G, and C). From the output, the A + U content was calculated by adding the number of A and U nucleotides and then divided by the total number of nucleotides. Consequently, the MFEI was estimated from the value of MFE and A + U according to the formula described by Zhang et al.[36]

Nomenclature and family annotation of the predicted miRNAs

A homology search was performed in the miRBase with the family name 'miR-4968' of the reference miRNA to get the nucleotide sequences of all mature miRNAs (homologous members) of that family. Newly identified miRNAs were annotated according to the standard guidelines recommended by Ambros et al.[37] A sequence alignment between each of the homologous members belonging to the 'miR-4968' family and the predicted miRNA was carried out by Clustal Omega.[38]

Prediction and functional annotation of the putative miRNA targets

To the authors' best knowledge, none of the available software or Web servers had the specific option for the prediction and functional annotation of the putative miRNA targets in case of *C. parvum*. Therefore, TargetScanFly Web server release 7.2 was used to predict the target genes and function analysis of the newly identified miRNA considering *D. melanogaster* as the reference organism and also because its reference miRNA was from that organism.[13,39] Besides, the molecular, biological, and cellular functions of each of the targets were retrieved and recorded (March 21, 2020, to March 28, 2020) from the Flybase (http://flybase.org/) and the Ensembl Genome Browser (https://asia.ensembl.org/index.html).[40,41] In addition, the NCBI BLASTn was used to select the target genes by fixing the organism name (*C. parvum*; taxid: 5807) in the search and setting other parameters as defaults. Once the miRNA-targeted genes in *C. parvum* were predicted through the BLAST tool, their functions were studied from the previous research works and literatures.

Pathway analysis of the putative miRNA targets

In addition, the target genes that popped up in the BLASTn were subjected to the pathway analysis to identify the impact of the detected miRNAs on the biology of the organism. In this study, the pathway analysis was performed manually, without the help of any software, using KEGG pathway database (https://www.genome.jp/kegg/pathway.html).[42] In the database, pathways of the *C. parvum* were searched by putting the organism's name in the search option. Later, the names of all pathways of *C. parvum* were retrieved (March 29, 2020) from the database along with the genes involved in them. Subsequently, a word file (Supplementary File 8) was created from those retrieved information regarding the pathways and their associated genes. Eventually, miRNA-targeted genes were searched in the word file to explore a match with the genes remaining in that file, and thus the relationships and links to the miRNA-targeted genes with the pathways were established.

Results

We designed this study to detect the conserved miRNAs in *C. parvum* protozoon by adopting the EST-based computational homology search approach and to annotate their target genes. The whole procedure, consisting of several searches and screening steps, is shown schematically in Figure 1.

Identification of potential miRNA candidates in C. parvum

After the BLASTn search, a total of 205 ESTs were obtained from 60 548 ESTs of *C. parvum* genome, and after the redundancy check, only 66 unique ESTs of the organism were screened for further investigation (Supplementary file 2). According to the results of BLASTx search of 66 ESTs, only 7 non-coding ESTs were disclosed, which were considered as the potential candidates and qualified for the next step. However, 59 coding ESTs were excluded from the study (Supplementary files 3 and 4). Based on the results of the secondary structure prediction and overall reported criteria, we finally scrutinized a conserved mature miRNA in *C. parvum* (Table 1). The properties of the newly identified miRNA in this study were the following: (1) the length of the mature and pre-miRNA was 20

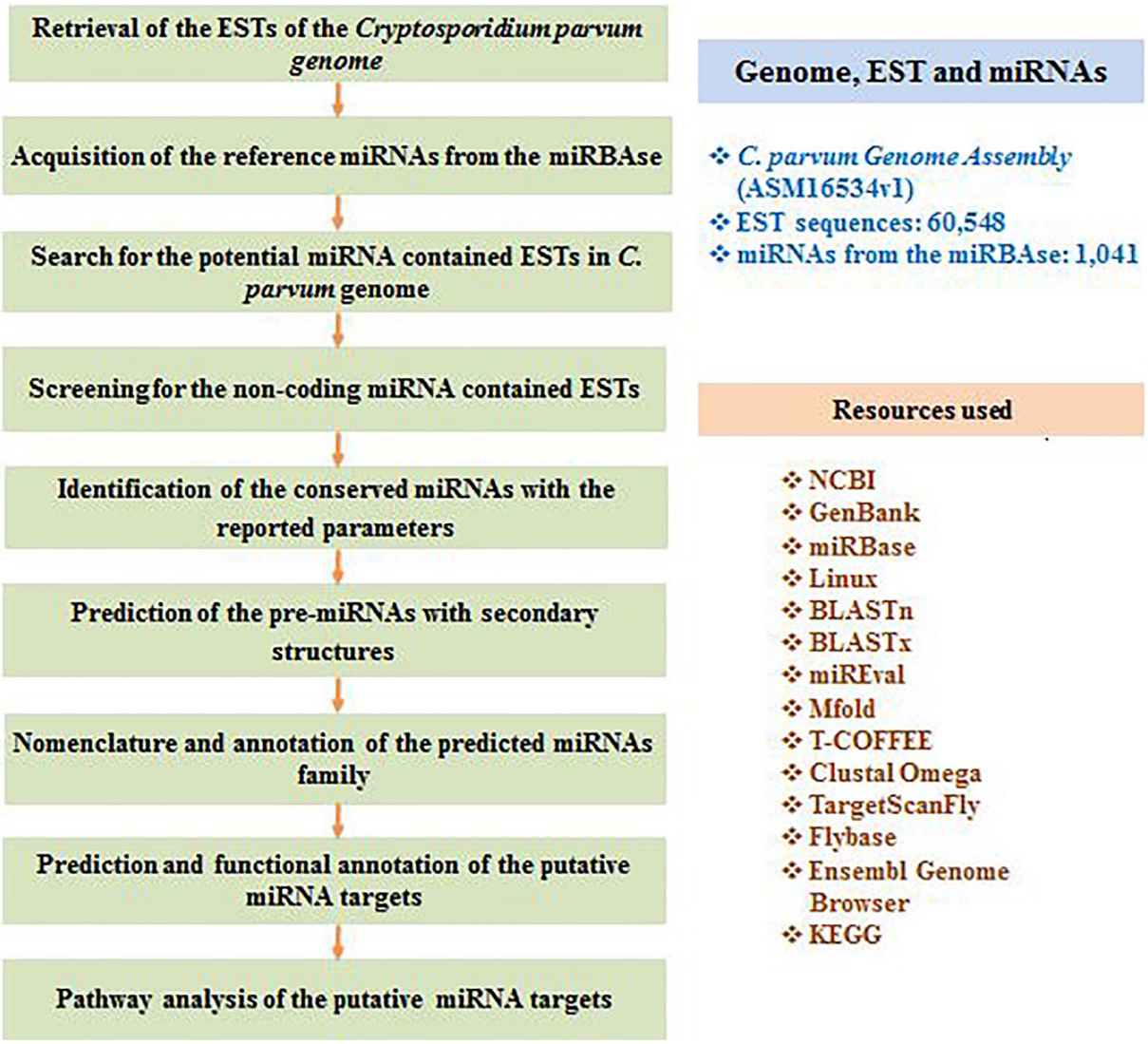

Figure 1. Schematic outline of the methodology for the identification and functional annotation of the putative miRNA in *Cryptosporidium parvum.* miRNA indicates microRNA.

Table 1. Steps involved in the methodology for the identification and functional annotation of the putative miRNA in *Cryptosporidium parvum* and their outputs.

SERIAL	NAME OF THE STEP	NUMBER
1	Retrieval of ESTs of the *C. parvum* genome	60548
2	Retrieval of the reference miRNA sequence from the miRBAse	1041
3	Potential miRNA contained ESTs in the genome of *C. parvum* by BLASTtn	205
4	Potential miRNA contained non-redundant ESTs after removal of redundancy	66
5	Potential non-coding EST candidates after BLASTx	7
6	MiRNAs satisfied secondary hairpin structures and selection criteria	1
7	Number of associated family identified for the putative miRNA	1
8	Number of miRNAs subjected to function prediction	1

Abbreviations: EST, expressed sequence tag; miRNA, microRNA.

and 85 nucleotides, respectively; (2) the secondary structure of the pre-miRNA sequence perfectly folded into a hairpin stem-loop structure where the mature miRNA positioned in one arm of the secondary structure with a small loop involvement; (3) there were 3 mismatches between the mature miRNA strand and the star miRNA strand; (4) the A + U content of the

Table 2. Properties of the putative miRNA of *Cryptosporidium parvum* with suggested criteria.

PROPERTIES	VALUE
NCBI accession number of the EST	GH587382.1
Length of the EST sequence	759 nucleotides
Reference miRNA from the miRBase	dme-miR-4968-3p
Length of the pre-miRNA	85
Number of each of the nucleotides in the pre-miRNA sequence	A = 12, C = 13, G = 25, and U = 35
A + U content	55%
Length of the mature miRNA	20 nucleotides
Mature miRNA sequence	CUGCUGCUGCUGCUGUUGCU
Nucleotide mismatch	03
MFE of the secondary structure	–30.20 kcal/mol
MFEI of the secondary structure	–0.79 kcal/mol
Family of the mature miRNA	miR-4968
Annotation of the identified mature miRNA	cpa-miR-4968-3p

Abbreviations: EST, expressed sequence tag; MFE, minimal folding free energy; MFEI, minimal free energy index; miRNA, microRNA; NCBI, National Center for Biotechnology Information.

pre-miRNA was 55%; and (5) the values of MFE and MFEI of the secondary structure were −30.20 kcal/mol and −0.79 kcal/mol, respectively (Table 2; Supplementary Table 1; Figure 2).

Nomenclature and family annotation of the putative miRNA

The results of family prediction of the conserved miRNA by multiple sequence alignment showed a strong agreement with the 'miR-4968' family, indicating that the identified miRNA in this study belonged to the members under that miRNA family. The conserved miRNA in *C. parvum* was annotated as 'cpa-miR-4968-3p' according to the uniform miRNA annotation guidelines (Figure 3). The sequence alignment of the putative miRNA of *C. parvum* with each of its homologous members from the miRBase is visualized in Figure 3.

Functional characterization of miRNA-targeted gene

The results of target gene prediction revealed that the conserved miRNA targeted 487 genes in the model organism, fly (*D. melanogaster*) (Supplementary file 6). Among 487 targeted genes, 74.33% were involved in the molecular function, biological process, and cellular component; 10.06% played vital roles in the biological process and cellular component; 9.04% took part in the molecular function and biological process; and 1.23% were associated with the molecular function and cellular component; however, 1.44%, 1.23% and 2.67% of total genes were related to the molecular function, biological process, and

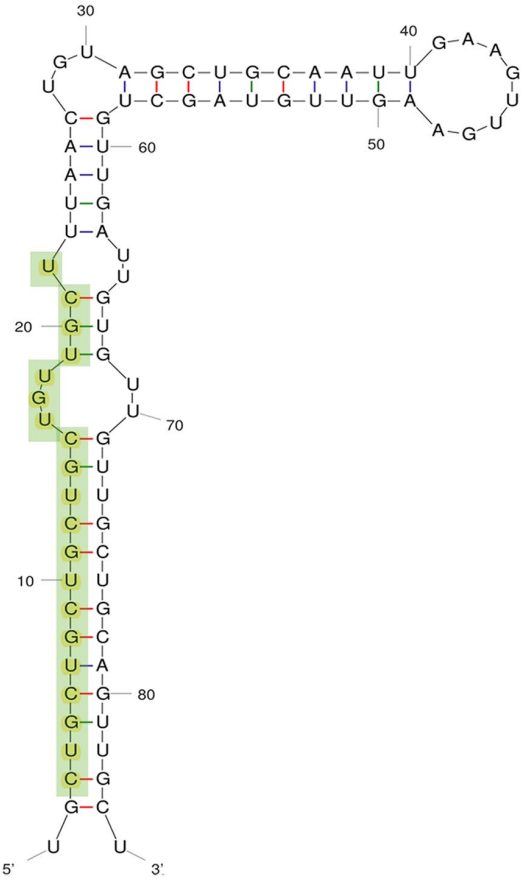

Figure 2. Secondary hairpin stem-loop structure of the pre-miRNA with the location of the mature miRNA of *Cryptosporidium parvum.*
miRNA indicates microRNA.

```
cpa-miR4968          1   CUGCUGCUGCUGCUGUUGCU   20
                         ||||||||||||||||||||
dme-miR4968-3P      21   CUGCUGCUGCUGCUGUUGCU   2
```

```
cpa-miR4968          1    CUGCUGCUGCUGCUGUUGCU   20
                           |      ||   | ||| | |
dme-miR4968-5P      21  UUCGUUGUUUGGCACCGUCGAU   2
```

Figure 3. Multiple sequence alignment for the family prediction of the identified miRNA of *Cryptosporidium parvum*.
miRNA indicates microRNA.

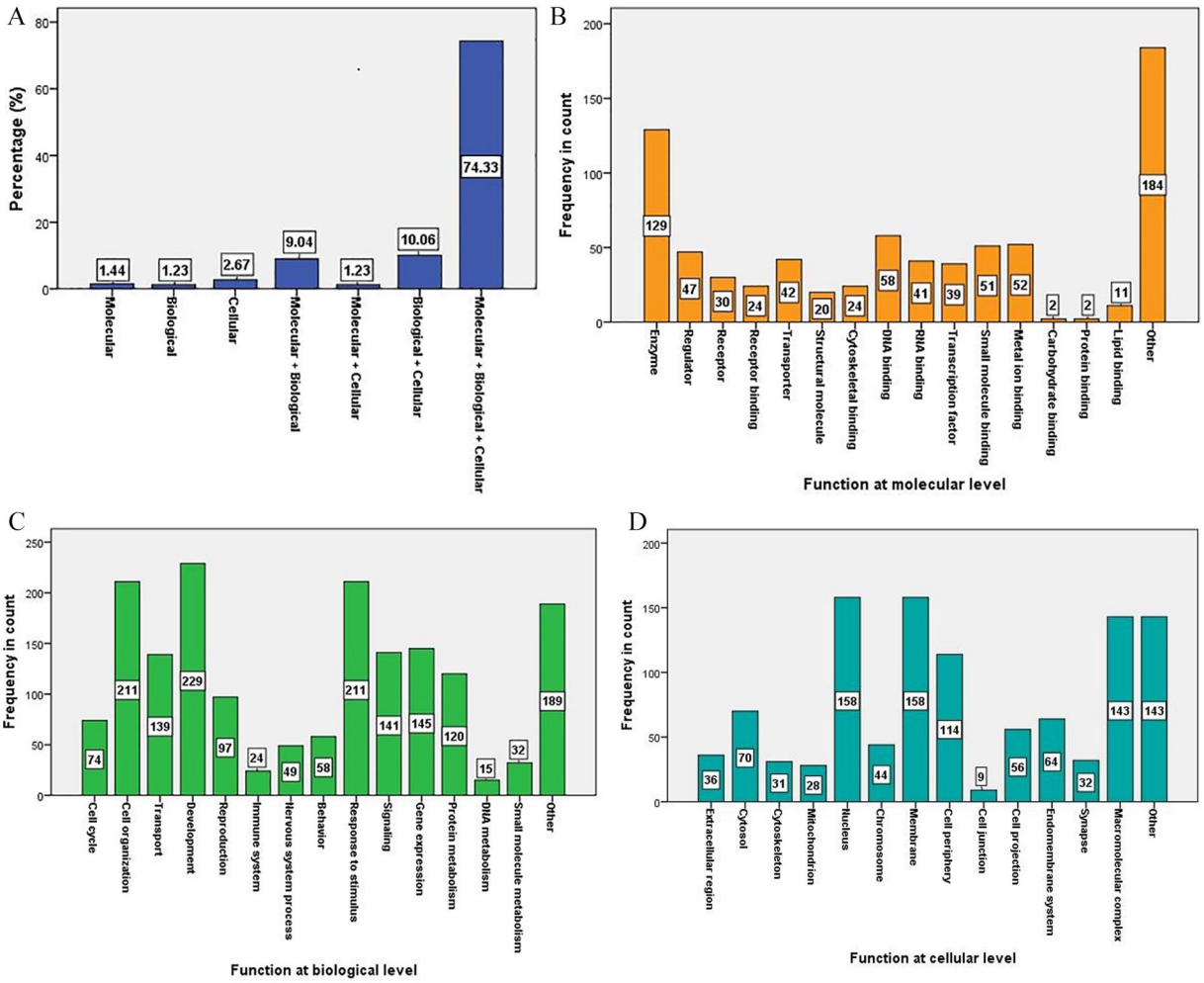

Figure 4. Conserved miRNA-targeted gene function annotation in the model organism, fly (*Drosophila melanogaster*) using TargetScanFly: (A) overall function, (B) molecular function, (C) biological process, and (D) cellular component.
miRNA indicates microRNA.

cellular component, respectively (Figure 4A). Specifically, among the molecular function, a substantial number of targeted genes were involved as enzyme (129), metal ion binding (52), lipid biding (52), small-molecule binding (51), and regulator (47) (Figure 4B). Accordingly, a considerable number of target genes were linked to the biological process as development (229), cell organization (211), response to stimulus (211), signal transduction (141), and gene expression (145) (Figure 4C). In the same way, a high amount of target genes were associated with the cellular component as nucleus (158), membrane (158), cell periphery (114), cytosol (70), and cell projection (56) (Figure 4D). Besides, the results of target gene prediction by the BLAST tool showed that a total of 85 *C. parvum* genes (Supplementary file 7) were targeted by the

Table 3. Conserved miRNA-targeted virulent genes of *Cryptosporidium parvum* and their function.

MIRNA-TARGETED GENE OF *C. PARVUM*	FUNCTION
COWP (*Cryptosporidium* oocyst wall protein)	Protect oocysts in the environment
Aminopeptidase N (aminopeptidase)	Excystation
cgd8_3430 (zincin/aminopeptidase N–like metalloprotease)	Excystation
p23	Locomotion
gp60 (glycoprotein 60)	Adhesion and invasion
gp900 (glycoprotein 900)	Adhesion and invasion
cgd7_4020 (cryptopsoridial mucin)	Adhesion and invasion
cgd3_1540 (Cyst-rich, Thr-rich, mucin)	Adhesion and invasion
cgd4_3710 (Ser/Thr protein kinase)	Invasion, phosphorylation, and cell signalling
cgd7_360 (heat shock protein 70)	Protect from stress and immune attack
cgd7_480 (lactate dehydrogenase)	Intracellular development and survival

Abbreviations: miRNA, microRNA.

conserved miRNA. Interestingly, among the 85 targeted genes, we obtained 11 genes (COWP, Aminopeptidase N, cgd8_3430, p23, gp60, gp900, cgd7_4020, cgd3_1540, cgd7_360, cgd4_3710, and cgd7_480), which were considered as the crucial virulent factors of *C. parvum* that played vital roles in the pathogenesis in case of humans and animals (Table 3).

Pathway analysis of miRNA-targeted genes

According to the results of pathway analysis, among 85 miRNA-targeted genes in *C. parvum*, a total of 20 genes were obtained to have connections with several important pathways of the organism. We obtained cgd3_3850 gene, a centromere-binding factor 5–like PUA domain containing protein, which was associated with the ribosome biogenesis pathway. Besides, we found 8 genes (cgd2_120, cgd2_1070, cgd3_2090, cgd8_430, cgd2_130, cgd6_570, cgd7_5310, and cgd5_2370) that encoded ribosomal proteins and were linked to the ribosome pathway. Furthermore, 2 genes (cgd2_3920 and cgd2_3550) showed their links to the RNA transport pathway, and cgd7_4140 gene was interrelated to the mRNA surveillance pathway. Moreover, we found a connection of cgd8_2030 gene to the basal transcription factor pathway. Besides, pathway analysis showed a linkage of cgd8_2020 and cgd1_3700 genes with the base excision repair (BER) pathway and the ubiquitin-mediated proteolysis pathway, respectively. In addition, we found involvements of 5 genes named cgd8_3430, cgd5_4560, cgd8_4620, cgd3_80, and cgd5_2600 with some nutrient metabolic and biosynthesis pathways such as glutathione metabolism pathway; glycine, serine, and threonine metabolism pathway; purine metabolism pathway; cysteine and methionine metabolism pathway; and phenylalanine, tyrosine, and tryptophan biosynthesis pathway (Table 4).

Discussion

This study was allowed to identify the conserved miRNAs in a zoonotic pathogen *C. parvum*, known to cause paediatric diarrhoea worldwide, leading to a substantial childhood mortality.[1-5] Besides, the previous studies also reported that the organism posed a great impact on the dairy and water industries.[3,7-10] MiRNAs, short powerful nucleic acids having potential gene silencing capacity, were believed to regulate a number of processes at the molecular, biological, and cellular level in animals and plants.[12,13,20] The approach we adopted in this study to detect miRNAs was a popular EST-based computational homology search, which was considered as a simple, comprehensible, and reliable method.[12,13,33]

The length of the conserved mature miRNA identified in this study was 20 nucleotides. The results of the previous studies showed that the nucleotides of a mature miRNA could range from 16 to 27.[11,43] Besides, Jike et al,[14] Zhang et al,[33] and Paul et al[29] found the length of the mature miRNAs to be 20 to 22 nucleotides in the earlier studies that supported the findings of this study. The length of the pre-miRNA (85 nucleotides) of this study agreed with the findings of the previous studies that the range of a pre-miRNA could be from 60 to 263 nucleotides.[14,29,30,32] The A + U content (55%) of the pre-miRNA found in this study was within the regular bound described by Curcio et al[11] as 36.80% to 66.00%, Singh et al[31] as 37.73% to 72.47%, and Zhang et al[33] as 36% to 89.04%, providing a strong support to this study. The previous research works suggested that the formation of a perfect secondary hairpin stem-loop structure by the pre-miRNA was one of the important qualifying pre-requisites in the EST-based computational approach, and highly negative values of the MFE and MFEI of this structure were a true indication of their stringent stability.[13,14,30,31,33] This study showed that the pre-miRNA

Table 4. Involved pathways of the miRNA-targeted genes in *Cryptosporidium parvum*..

MIRNA-TARGETED GENE OF *C. PARVUM*	NAME OF INVOLVED PATHWAY
cgd3_3850 (centromere-binding factor 5–like protein)	Ribosome biogenesis
cgd2_120 (ribosomal protein L29)	Ribosome
cgd2_1070 (40S ribosomal protein S25)	Ribosome
cgd3_2090 (40S ribosomal protein SAe)	Ribosome
cgd8_430 (60S ribosomal protein L7A)	Ribosome
cgd2_130 (60S acidic ribosomal protein LP2)	Ribosome
cgd6_570 (60S ribosomal protein L23)	Ribosome
cgd7_5310 (60S ribosomal protein L24)	Ribosome
cgd5_2370 (60S acidic ribosomal protein LP1–like protein)	Ribosome
cgd2_3920 (eukaryotic initiation factor 4)	RNA transport
cgd2_3550 (low complexity protein)	RNA transport
cgd7_4140 (WD40 domain protein)	mRNA surveillance
cgd8_2030 (TATA-box factor binding protein)	Basal transcription factors
cgd8_2020 (OGG1-like 8-oxoguanine DNA glycosylase)	Base excision repair
cgd1_3700 (conserved hypothetical protein)	Ubiquitin mediated proteolysis
cgd5_4560 (tryptophan synthase)	GSTMP; PTTBP
cgd8_4620 (ISWI-related chromatinic protein)	Purine metabolism
cgd3_80 (S-adenosylhomocysteinase)	Cysteine and methionine metabolism
cgd5_2600 (leucine aminopeptidase)	Glutathione metabolism
cgd8_3430 (zincin/aminopeptidase N–like metalloprotease)	Glutathione metabolism

Abbreviations: miRNA, microRNA; GSTMP, glycine, serine, and threonine metabolism pathway; PTTBP, phenylalanine, tyrosine, and tryptophan biosynthesis pathway.

folded into a perfect hairpin-shaped structure having a stem and a loop, and the mature miRNA was located in the arm with a minor loop involvement, which satisfied the standard criteria that miRNA would not contain a very large loop or bulges recommended in the previous study.[44] Besides, the findings of this study coincided with the findings of Curcio et al,[11] Prakash et al,[30] and Xu et al[32] who identified several miRNAs having 1 or 2 loop involvements. The findings of a highly negative MFE value (−30.20 kcal/mol) of this study came in agreement with the findings of the previous studies.[11,14,30-33] Accordingly, the MFEI value (−0.79 kcal/mol) of this study was in accordance with the findings of Zhang et al[33] and Paul et al[29] who reported the ranges of MFEI values as (−0.40 to −1.07 kcal/mol) and (−0.70 to −1.30 kcal/mol), respectively. Besides, the MFEI value was lower than other reported small RNAs such as transfer RNAs (tRNAs) (−0.64 kcal/mol), ribosomal RNAs (rRNAs) (−0.59 kcal/mol), and mRNAs (−0.65 kcal/mol), indicating a robust evidence of accuracy of the newly detected miRNA of *C. parvum* in this study.[14,36] However, the lower (−0.46 to −0.72 kcal/mol) and higher (−0.85 to −1.40 kcal/mol) MFEI values, than those of present study, were reported by Singh et al[31] and Jike et al,[14] respectively.

The identified conserved miRNA involved 11 virulent genes in *C. parvum*, which were COWP, Aminopeptidase N, cgd8_3430, p23, gp60, gp900, cgd7_4020, cgd3_1540, cgd7_360, cgd4_3710, and cgd7_480. According to the previous research, the COWP target gene was reported to encode *Cryptosporidium* oocyst wall protein, which was one of the fundamental components of oocyst wall providing the rigidity that led to the protection of oocysts from destruction in the environment.[45] Bouzid et al[46] showed that the Aminopeptidase N and cgd8_3430 (zincin/aminopeptidase N–like metalloprotease) target genes facilitated excystation, the process by which infective sporozoites are released in the host from the oocysts. Following an excystation in vivo, the locomotion, adhesion, and invasion were considered as the 3 most important virulent factors associated with the life cycle and concomitant pathogenesis of *C. parvum*.[2,46] The earlier studies stated that P23 was an antigenic gene found on the surface of sporozoites that expedited their motility by the gliding process in the host.[46,47] Besides, the previous research works reported gp60 and gp900 genes as the mucin-like glycoproteins, located at the apical regions of sporozoites, which were found to be associated with the adhesion to enterocytes and subsequent invasion into the

cells.[46,47] Another 2 mucin-encoding genes, cgd7_4020 and cgd3_1540, were narrated to have a structural and functional similarity with the gp900 gene, playing important roles in the attachment and invasion.[47-49] Singh et al[50] showed that the Ser/Thr protein kinase, translated by the cgd4_3710 target gene, was a key virulent factor of *C. parvum* for invasion into the intestinal cells, phosphorylation, and cell signalling, whereas the heat shock protein 70, encoded by the cgd7_360 target gene, was reported to protect the protozoon in the host from the stressful environment, produced due to sudden temperature shifting, nutrient unavailability, and immune attack.[50,51] Besides, Li et al[52] showed that the cgd7_480 gene was responsible for producing lactate dehydrogenase enzyme that facilitated the intracellular growth and survival of *C. parvum*, and an inhibitor of this enzyme could restrict the development of the organism.

Moreover, the conserved miRNA, identified in this study, targeted 20 genes that showed their involvements in several major pathways of *C. parvum*, which were considered to have numerous momentous roles in the biology of the organism.[42] The previous research works reported that the ribosome biosynthesis and ribosome pathways directly took part in the synthesis of ribosomes and their assembly, structures, and functions, which ultimately catalysed protein production by translating genetic information.[42,53,54] Besides, the RNA transport pathway was believed to play a key role in the spatio-temporal articulation of gene expression by transporting tRNAs and rRNAs from the nucleus to cytoplasm.[42,55] Similarly, the earlier studies showed the mRNA surveillance pathway as a quality control mechanism that maintained the expected gene expression by detecting and degrading abnormal, defective, and aberrant mRNAs,[42,56] whereas the basal transcription factor pathway was reported to have gene expression regulation capacity by controlling the rate of transcription from DNA to mRNA.[42,57] Besides, the BER pathway was considered as a dedicated and predominant mechanism for the distorted DNA repair,[42,58] whereas the ubiquitin-mediated proteolysis pathway was believed to play imperative roles in many basic cellular processes such as gene expression, cell cycle, cell division, development, immune responses, and differentiation.[59] In addition, miRNA-targeted genes showed their connections to 5 nutrient metabolic and biosynthesis pathways that were required for the maintenance of various amino acids (glycine, serine, threonine, purine, cysteine, and methionine) and antioxidant (glutathione) metabolisms, and biosynthesis of phenylalanine, tyrosine, and tryptophan by *C. parvum* in the host.[42,60]

Conclusion

Therefore, the identified conserved miRNA had a great impact on the biology of zoonotic protozoon *C. parvum*, as it targeted a number of genes related to the virulence and pathways of the organism. Thus, the detected miRNA could be a potential epigenomic tool for an intricate investigation of cryptosporidiosis as well as controlling the protozoon through silencing those virulence and pathway-related target genes in the organism. However, a laboratory experiment was recommended to study the miRNA of *C. parvum* in future.

Author Contributions

S.S.U.A., M.B.U., and M.H. were involved in conception of the study, designed the study methodology, facilitated the analytical tools, wrote the manuscript, and critically reviewed the article and supervised. R.N.A. was involved in retrieval of data, data curation, contributed to the analysis, and critically reviewed the article. S.R., B.S., M.D., S.A., and M.S.R.C. assisted in the retrieval of data and analysis. M.I.A. assisted in the retrieval of data and analysis, wrote the paper, and critically revised the article.

ORCID iD

Md Bashir Uddin https://orcid.org/0000-0002-0045-9669

Data Availability Statement

Anyone can access the data presented on this paper as supplementary files.

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

1. Mahmoudi MR, Ongerth JE, Karanis P. Cryptosporidium and cryptosporidiosis: the Asian perspective. *Int J Hyg Environ Health*. 2017;220:1098-1109.
2. Tandel J, English ED, Sateriale A, Gullicksrud JA, et al. Life cycle progression and sexual development of the apicomplexan parasite Cryptosporidium parvum. *Nat Microbiol*. 2019;4:2226-2236.
3. Xiao L, Fayer R, Ryan U, Upton SJ. Cryptosporidium taxonomy: recent advances and implications for public health. *Clin Microbiol Rev*. 2004;17:72-97.
4. Ryan U, Hijjawi N, Xiao L. Foodborne cryptosporidiosis. *Int J Parasitol*. 2018;48:1-12.
5. Areeshi MY, Beeching NJ, Hart CA. Cryptosporidiosis in Saudi Arabia and neighboring countries. *Ann Saudi Med*. 2007;27:325-332.
6. Wang RJ, Li JQ, Chen YC, Zhang LX, Xiao LH. Widespread occurrence of Cryptosporidium infections in patients with HIV/AIDS: epidemiology, clinical feature, diagnosis, and therapy. *Acta Trop*. 2018;187:257-263.
7. Budu-Amoako E, Greenwood SJ, Dixon BR, Barkema HW, McClure JT. Foodborne illness associated with Cryptosporidium and Giardia from livestock. *J Food Prot*. 2011;74:1944-1955.
8. Al Mawly J, Grinberg A, Velathanthiri N, French N. Cross sectional study of prevalence, genetic diversity and zoonotic potential of Cryptosporidium parvum cycling in New Zealand dairy farms. *Parasit Vectors*. 2015;8:240.
9. Delafosse A, Chartier C, Dupuy MC, Dumoulin M, Pors I, Paraud C. Cryptosporidium parvum infection and associated risk factors in dairy calves in western France. *Prev Vet Med*. 2015;118:406-412.
10. Ryan U, Paparini A, Monis P, Hijjawi N. It's official – Cryptosporidium is a gregarine: what are the implications for the water industry? *Water Res*. 2016;105:305-313.
11. Curcio JS, Batista MP, Paccez JD, Novaes E, Soares CMA. In silico characterization of microRNAs-like sequences in the genome of Paracoccidioides brasiliensis. *Genet Mol Biol*. 2019;42:95-107.
12. Akter A, Islam MM, Mondal SI, Mahmud Z, et al. Computational identification of miRNA and targets from expressed sequence tags of coffee (Coffea arabica). *Saudi J Biol Sci*. 2014;21:3-12.
13. Hanif Q, Farooq M, Amin I, Mansoor S, Zhang Y, Khan QM. In silico identification of conserved miRNAs and their selective target gene prediction in indicine (Bos indicus) cattle. *PLoS ONE*. 2018;13:e0206154.
14. Jike W, Sablok G, Bertorelle G, Li M, Varotto C. In silico identification and characterization of a diverse subset of conserved microRNAs in bioenergy crop Arundo donax L. *Sci Rep*. 2018;8:16667.
15. Lytle JR, Yario TA, Steitz JA. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. *Proc Natl Acad Sci USA*. 2007;104:9667-9672.

16. Lee Y, Kim M, Han J, et al. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J*. 2004;23:4051-4060.
17. Kurihara Y, Watanabe Y. Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc Natl Acad Sci USA*. 2004;101:12753-12758.
18. Kim YK, Kim B, Kim VN. Re-evaluation of the roles of DROSHA, Export in 5, and DICER in microRNA biogenesis. *Proc Natl Acad Sci USA*. 2016; 113:E1881-E1889.
19. Kobayashi H, Tomari Y. RISC assembly: coordination between small RNAs and Argonaute proteins. *Biochim Biophys Acta*. 2016;1859:71-81.
20. Huang Y, Shen XJ, Zou Q, Wang SP, Tang SM, Zhang GZ. Biological functions of microRNAs: a review. *J Physiol Biochem*. 2011;67:129-139.
21. Zhang BH, Pan XP, Cobb GP, Anderson TA. Plant microRNA: a small regulatory molecule with big impact. *Dev Biol*. 2006;289:3-16.
22. Shao CC, Xu MJ, Alasaad S, et al. Comparative analysis of microRNA profiles between adult Ascaris lumbricoides and Ascaris suum. *BMC Vet Res*. 2014;10:99.
23. Trobaugh DW, Klimstra WB. MicroRNA regulation of RNA virus replication and pathogenesis. *Trends Mol Med*. 2017;23:80-93.
24. Benson DA, Cavanaugh M, Clark K, et al. GenBank. *Nucleic Acids Res*. 2018;46:D41-D47.
25. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. *Nucleic Acids Res*. 2008;36:D154-D158.
26. Adl SM, Simpson AGB, Lane CE, et al. The revised classification of eukaryotes [Erratum in: *J Eukaryot Microbiol*. 60:321]. *J Eukaryot Microbiol*. 2012; 59:429-493.
27. Pal AS, Kasinski AL. Animal models to study microRNA function. *Adv Cancer Res*. 2017;135:53-118.
28. Cock PJ, Chilton JM, Grüning B, Johnson JE, Soranzo N. NCBI BLAST+ integrated into Galaxy. *Gigascience*. 2015;4:39.
29. Paul S, de la Fuente-Jiménez JL, Manriquez CG, Sharma A. Identification, characterization and expression analysis of passion fruit (*Passiflora edulis*) microRNAs. *3 Biotech*. 2020;10:25.
30. Prakash P, Ghosliya D, Gupta V. Identification of conserved and novel microRNAs in Catharanthus roseus by deep sequencing and computational prediction of their potential targets. *Gene*. 2015;554:181-195.
31. Singh N, Srivastava S, Sharma A. Identification and analysis of miRNAs and their targets in ginger using bioinformatics approach. *Gene*. 2016;575:570-576.
32. Xu JH, Li F, Sun QF. Identification of microRNA precursors with support vector machine and string kernel. *Genomics Proteomics Bioinformatics*. 2008;6:121-128.
33. Zhang BH, Wang Q, Wang K, et al. Identification of cotton microRNAs and their targets. *Gene*. 2007;397:26-37.
34. Gao D, Middleton R, Rasko JE, Ritchie W. miREval 2.0: a web tool for simple microRNA prediction in genome sequences. *Bioinformatics*. 2013;29: 3225-3226.
35. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*. 2003;31:3406-3415.
36. Zhang BH, Pan XP, Cox SB, Cobb GP, Anderson TA. Evidence that miRNAs are different from other RNAs. *Cell Mol Life Sci*. 2006;63:246-254.
37. Ambros V, Bartel B, Bartel DP, et al. A uniform system for microRNA annotation. *RNA*. 2003;9:277-279.
38. Sievers F, Higgins DG. Clustal omega. *Curr Protoc Bioinformatics*. 2014; 48:3.13.1-3.13.16.
39. Agarwal V, Subtelny AO, Thiru P, Ulitsky I, Bartel DP. Predicting microRNA targeting efficacy in Drosophila. *Genome Biol*. 2018;19:152.
40. Stalker J, Gibbins B, Meidl P, et al. The Ensembl Web site: mechanics of a genome browser. *Genome Res*. 2004;14:951-955.
41. Thurmond J, Goodman JL, Strelets VB, et al. FlyBase 2.0: the next generation. *Nucleic Acids Res*. 2019;47:D759-D765.
42. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*. 2017; 45:D353-D361.
43. Bortoluzzi S, Bisognin A, Biasiolo M, et al. Characterization and discovery of novel miRNAs and moRNAs in JAK2V617F-mutated SET2 cells. *Blood*. 2012;119:e120-e130.
44. Zhang B, Pan X, Cannon CH, Cobb GP, Anderson TA. Conservation and divergence of plant microRNA genes. *Plant J*. 2006;46:243-259.
45. Templeton TJ, Lancto CA, Vigdorovich V, et al. The Cryptosporidium oocyst wall protein is a member of a multigene family and has a homolog in Toxoplasma. *Infect Immun*. 2004;72:980-987.
46. Bouzid M, Hunter PR, Chalmers RM, Tyler KM. Cryptosporidium pathogenicity and virulence. *Clin Microbiol Rev*. 2013;26:115-134.
47. Sanderson SJ, Xia D, Prieto H, et al. Determining the protein repertoire of Cryptosporidium parvum sporozoites. *Proteomics*. 2008;8:1398-1414.
48. Johnson JK, Schmidt J, Gelberg HB, Kuhlenschmidt MS. Microbial adhesion of Cryptosporidium parvum sporozoites: purification of an inhibitory lipid from bovine mucosa. *J Parasitol*. 2004;90:980-990.
49. Pundir S, Martin MJ, O'Donovan C. UniProt protein knowledgebase. *Methods Mol Biol*. 2017;1558:41-55.
50. Singh P, Mirdha BR, Srinivasan A, et al. Identification of invasion proteins of Cryptosporidium parvum [Erratum in: *World J Microbiol Biotechnol*. 32:72]. *World J Microbiol Biotechnol*. 2015;31:1923-1934.
51. Bajszár G, Dekonenko A. Stress-induced Hsp70 gene expression and inactivation of Cryptosporidium parvum oocysts by chlorine-based oxidants. *Appl Environ Microbiol*. 2010;76:1732-1739.
52. Li K, Nader SM, Zhang X, et al. Novel lactate dehydrogenase inhibitors with in vivo efficacy against Cryptosporidium parvum. *PLoS Pathog*. 2019;15:e1007953.
53. de la Cruz J, Karbstein K, Woolford JL Jr. Functions of ribosomal proteins in assembly of eukaryotic ribosomes in vivo. *Annu Rev Biochem*. 2015;84:93-129.
54. Granneman S, Baserga SJ. Ribosome biogenesis: of knobs and RNA processing. *Exp Cell Res*. 2004;296:43-50.
55. Kindler S, Wang H, Richter D, Tiedge H. RNA transport and local control of translation. *Annu Rev Cell Dev Biol*. 2005;21:223-245.
56. van Hoof A, Wagner EJ. A brief survey of mRNA surveillance. *Trends Biochem Sci*. 2011;36:585-592.
57. Latchman DS. Transcription factors: an overview. *Int J Biochem Cell Biol*. 1997;29:1305-1312.
58. Krokan HE, Bjørås M. Base excision repair. *Cold Spring Harb Perspect Biol*. 2013;5:a012583.
59. Ciechanover A, Orian A, Schwartz AL. The ubiquitin-mediated proteolytic pathway: mode of action and clinical implications. *J Cell Biochem Suppl*. 2000;34:40-51.
60. Rider SD Jr, Zhu G. Cryptosporidium: genomic and biochemical features. *Exp Parasitol*. 2010;124:2-9.