In Silico Investigation of Conserved miRNAs and Their Targets From the Expressed Sequence Tags in *Neospora Caninum* Genome

Bioinformatics and Biology Insights
Volume 15: 1–11

DOI: 10.1177/11779322211046729

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

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

ABSTRACT: *Neospora caninum* is a protozoan parasite, the etiologic agent of Neosporosis—a common cause of abortion in cattle worldwide. Herd level prevalence of Neosporosis could be as high as 90%. However, there is no approved treatment and vaccines available for Neosporosis. MicroRNA (miRNA) based prophylaxis and therapeutics could be options for Neosporosis in cattle and other animals. The current study aimed to investigate the genome of *Neospora caninum* to identify and characterize the conserved miRNAs through Expressed Sequence Tags (ESTs) dependent homology search. A total of 1,041 mature miRNAs of reference organisms were employed against 336 non-redundant ESTs available in the genome of *Neospora caninum.* The study predicted one putative miRNA "nca-miR-9388-5p" of 19 nucleotides with MFEI value -1.51 kcal/mol and (A + U) content% 72.94% corresponding with its pre-miRNA. A comprehensive search for specific gene targets was performed and discovered 16 potential genes associated with different protozoal physiological functions. Significantly, the gene "Protein phosphatase" was found responsible for the virulence of *Neospora caninum.* The other genes were accounted for gene expression, vesicular transport, cell signaling, cell proliferation, DNA repair mechanism, and different developmental stages of the protozoon. Therefore, this study finding will provide pivotal information to future aspirants upon Bovine Neosporosis. It will also serve as the baseline information for further studies of the bioinformatics approach to identify other protozoal miRNAs.

KEYWORDS: *Neospora caninum,* neosporosis, miRNA, ESTs, virulence

RECEIVED: April 17, 2021. **ACCEPTED:** August 20, 2021.

TYPE: Original Research

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 AUTHORS:** 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

Bovine neosporosis is an important protozoan disease—having worldwide distribution and causing severe financial damages in the cattle industry.[1,2] The etiologic agent of Neosporosis is *Neospora caninum* – a microscopic protozoon. The pathogen is abortifacient and is considered to be an economically significant pathogen for cattle.[3] Globally, 42% of bovine abortion is attributed to Neosporosis.[4] Herd level prevalence of Neosporosis could be as high as 90% in cattle.[5] The estimated cost per annum for *Neospora* infection is the US$ 1.1 million in the New Zealand beef industry and US$ 546.3 million in the US dairy business.[1] Despite the substantial impact of *Neospora caninum* in the cattle industry, there is no approved treatment and vaccines available for neosporosis.[6] A variety of drug compounds and a combination of drugs have been suggested to control the Neosporosis for subsiding clinical signs and humoral immune response to infants or diminishing the rate of abortion to the pregnant cows. But the therapeutic is not well accepted to date due to lack of experimental validity and failure to present drug efficacy against the bradyzoite stage of the organism.[7] With the limitations of comprehensive research regarding *Neospora caninum*, there is an ardent need to figure out the pragmatic and effective therapeutic strategy or other usefultools to combat Neosporosis in the cattle industry.

MicroRNAs (miRNAs) are non-protein-coding short sequences RNA of around 22 nucleotides (nt) in length. miRNAs are commonly transcribed from DNA sequences. They are found in a wide range of organisms, from protozoans to mammals.[8,9] miRNAs cover 1-3% of the total length of the genome.[10] They play essential biological and physiological regulatory functions in animals[11] and plants.[12] They are essential for the development of animals and have a critical role in different biological processes. miRNAs are pleiotropic and could target hundreds to thousands of genes.[13] miRNAs having a role in post-transcriptional gene regulation and targets specific mRNAs for degradation or translation repression.[14] The pleiotropic nature and role in gene regulation make miRNAs a potential choice for a novel class of therapeutic targets.

In recent days, thus, miRNA research has been focusing on RNA-based therapies. Further, studies are examining the potential of miRNAs for the development of new generation drugs.

The high throughput sequencing, computational biology, and bioinformatics analysis methods accelerated miRNAs research. With enhanced research, thousands of miRNAs have been identified and made available in miRbase (http://www.mirbase.org/) – either recognized through cloning technology or predicted by the computational method. Moreover, in recent days, an Expressed Sequenced Tag (EST) analysis rather than former traditional computational methods has evolved to make conserved miRNA identification convenient.

Neospora caninum is characterizing by different antigenic strains divulging notable genotypic and phenotypic differences.[15] However, very little is known about the genetic diversity of *Neospora caninum* – except few as exemplified genome sequence,[16] phylogenetic analysis,[17] biological and genetic variation between 6 isolates.[18] Moreover, there is no information available about specific microRNAs (miRNAs) of *Neospora caninum* in the database.[19] However, investigating *Neospora caninum* genome for identifying miRNAs with their regulatory target and functions could aid in devising new prophylaxis and therapeutics for nesporiosis. Thus, it is a worthy effort to predict the miRNAs of *Neospora caninum*, including their functions.

The current study aimed to identify plausible miRNAs of *Neospora caninum* through ESTs-based homology search from the NCBI GenBank database. Further utilize *in silico* methods to perform the function analysis of *Neospora caninum* involved with molecular, cellular, and biological functions.

Materials and Methods

Retrieval of ESTs in the genome of Neospora caninum

Initially, the genome of *Neospora caninum* (GenBank Assembly Accession: GCA_000208865.2; Assembly ID: ASM20886v2) was employed to search for ESTs from GenBank database of NCBI (http://www.ncbi.nlm.nih.gov/). The genome based BLAST search suggested available ESTs of *Neospora caninum* which were further retrieved from the EST database of NCBI. This database is open access - annotated nucleotide sequences are publicly available.

Acquisition of a reference set of miRNAs

In the absence of insufficient information about protozoal miRNA in the database, we decided to pick up the reference organisms considering the taxonomy of *Neospora caninum* from miRBase 22.1 (January 2020)[20] database. In order to look for potential miRNAs of *Neospora caninum*, a total of 1,041 formerly known mature miRNAs from miRBase database were utilized. These mature sequences were designated as query sequences employed against the retrieved ESTs of *Neospora caninum.*

Search for potential miRNAs of Neospora caninum

The query miRNAs were subjected to nucleotide BLAST (BLASTn) (https://blast.ncbi.nlm.nih.gov/Blast.cgi) of NCBI against *Neospora caninum* sequences. Python language-based command line for Linux platform was established for BLASTn and addressing by "tblastn -subject -query -out -evalue -word_size 10 -qcov_hsp_perc." The search considered 80% query cover and 90% identity match and no set E-value. Considering all these criteria with the BLASTn, a set of EST sequences had been primarily identified that was stated to be the possible miRNA candidates for *Neospora caninum.*

Removal of protein-coding sequences

As miRNAs characteristically must not code for any protein,[21] the next step was to extract non-protein-coding sequences through BLASTx.[21] This set of sequences were further screened to isolate pre-miRNA through assessing the previously reported properties.[22]

Identification of novel miRNAs with stated parameters

Following the procedure described by previous authors,[23,24] the putative miRNA precursors were identified. The same procedures were used in different other studies.[25,26] The putative miRNA precursors must fulfill subsequent criteria: (i) the predicted mature miRNA must not allow above three nucleotide substitutions in comparison with reference sequences (ii) the pre-miRNA sequence must form a stem-loop hairpin structure (iii) the mature miRNA must be placed at the single arm of secondary structure (iv) above three mismatches were not acceptable between predicted and reference sequences (v) higher the A + U percentage better the quality of precursor miRNA (present study considered above 70%) (vi) the anticipated secondary hairpin structure should have high negative MFE (Minimal Folding Free Energy, ΔG in kcal/mol) and MFEI (Minimal Folding Free Energy Index) values.[26,27]

Prediction of pre-miRNA and secondary structure of mature miRNA

An algorithm sorted out the potential pre-miRNAs aspirants with miREval 2.0.[26] To predict appropriate hairpin-like secondary structures that must contain the minimum length of 85 nucleotides, the potential pre-miRNAs were subjected to the Mfold web server.[27,28] Typically the putative secondary structure should possess high negative MFE and MFEI values, which confirmed the best-fitted pre-miRNAs to pick up. MFEI was calculated using the equations given below[29]:

Table 1. Steps involved in the characterization of putative miRNA from ESTs of *Neospora caninum* genome.

NO.	STEPS	NO. OF ESTS
01	Retrieval of EST sequences from NCBI	25,095
02	Retrieval of previously known mature miRNA sequences from miRBase	1,041
03	Search for potential miRNAs through BLASTn	339
04	Non-redundant ESTs	336
05	Candidate ESTs with non-protein coding sequences after Blastx	93
06	Putative pre-miRNA sequences with Hair-loop secondary structure	01
07	Identification of newly identified miRNA family	01
08	Prediction and functional analysis of putative miRNA targets	01
09	Final candidate of conserved newly identified miRNA in NC genome	01

Abbreviations: EST, expressed sequence tag; miRNA, microRNA; NC, *Neospora caninum*.

$$\text{Adjusted Minimal Folding Free Energy (AMFE)} = (\text{MFE} \div \text{Lenth of RNA sequence}) * 100$$

$$\text{MFEI (Minimal Folding Free Energy Index)} = [(\text{MFE} / \text{Length of RNA sequence}) \times 100] / (\text{G} + \text{C})$$

Finally, the potential miRNA of *Neospora caninum* was extracted from all the non-coding mature ESTs qualifying earlier mentioned preconditioned criteria.

Nomenclature and family annotation of predicted microRNAs

By using miRBase database, the final retrieved miRNA family was again subjected to local blast to validate the family annotation of predicted miRNA. This final miRNA was addressed according to miRNA nomenclature.[30] For verification, the sequence alignment of homologous sequences of putative miRNA families was done by miRBase search output and used Clustal Omega.[31] A comparative analysis of base nucleotide between putative mature miRNAs and its homolog miRNA was also conducted.

Functional annotation of gene targets by putative miRNAs

The function annotation of the newly identified miRNA family was accomplished by TargetScanFly database version 7.2.[32] The miRNA category of the database provided the available target genes of the respective family. The molecular, biological, and cellular functions involved with each particular gene were also analyzed and documented. Moreover, the single specific gene holding NCBI gene Identification number, Ensembl ID (https://asia.ensembl.org/index.html), UniProtKB (https://www.uniprot.org/), and Alliance of Genomic Resources ID (https://www.alliancegenome.org/) were tabulated for convenience.

Results

Retrieval of ESTs in the genome of Neospora caninum

All available ESTs of *Neospora caninum* were downloaded from NCBI GenBank database. There were a total of 25,095 ESTs of *Neospora caninum* available in the GenBank database (Table 1). The schematic outline of the entire methodology of identifying miRNA of *Neospora caninum is* shown in Figure 1.

Acquisition of a reference set of miRNAs

Neospora caninum is classified under the kingdom of Chromista or superfamily Chromalveolata of eukaryotes along with superphylum Alveolata and phylum Apicomplexa.[33,34] We sorted out six organisms in miRBase, namely *Phaeodactylum tricornutum, Phytophthora sojae, Phytophthora ramorum, Phytophthora infestans, Ectocarpus siliculosus* and *Symbiodinium microadriaticum* that were also categorized into the superphylum Alveolata or superfamily Chromalveolata. Hence, those organisms got prioritized during the retrieval of mature miRNAs in this study. Besides, only mycetozoa *Dictyostelium discoideum* in miRbase was considered to retrieve mature miRNA as mycetozoa exhibited the characteristics of both *protozoans* like organism such as *Neospora caninum* and fungi.[35] Moreover, mature miRNAs of two model organisms such as *Caenorhabditis elegans*[33] and *Drosophila melanogaster*[34] were also considered in the study. However, in order to look for potential miRNAs of *Neospora caninum*, a total of 1,041 mature miRNAs of miRBase 22.1 (as of January 15, 2020)[20] database were retrieved and considered for further study (S1 File).

Search for potential miRNAs of Neospora caninum

There were 339 ESTs obtained through BLASTn (S2 File), which was counted 336 after the redundancy check (S3 File). These sequences were enforced to allow not exceeding four mismatches and the lowest preset E-value.

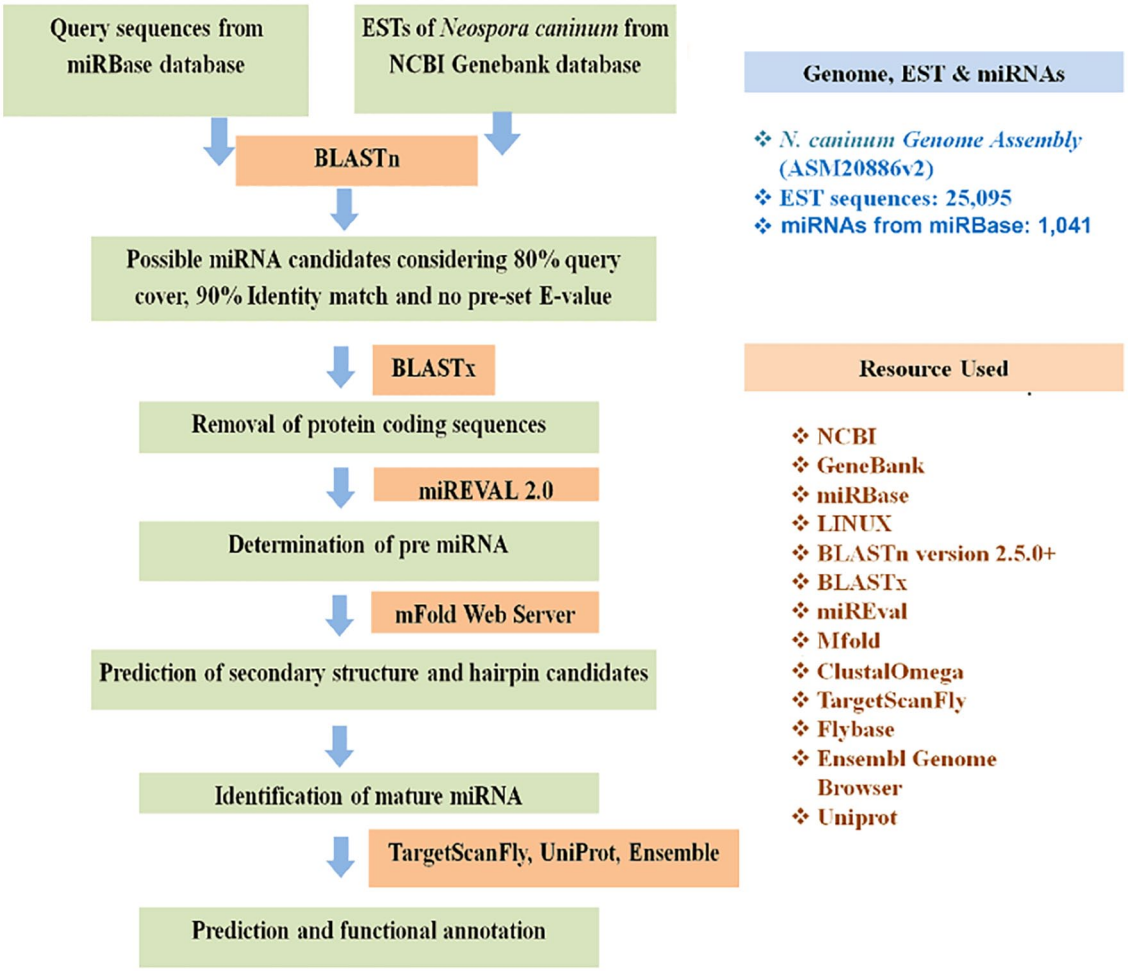

Figure 1. The schematic outline of entire methodology of identifying miRNA of *Neospora caninum*.

Removal of protein-coding sequences

All the non-redundant sequences were employed through BLASTx search to search for non-protein-coding ESTs. Hypothetical proteins were discarded for the final data set. Finally, 93 non-protein-coding ESTs had been attained (S4 File), having the probability of being a potential miRNA candidate for *Neospora caninum*.

Prediction of pre-miRNA with secondary structure

At the next step, sequences were rendered through miREval to predict potential pre-miRNA structures. The sequences were further attempted to Mfold Web Server for extrapolation of hairpin-like secondary structure. Furthermore, sequences were tabulated with the length of pre-miRNA, length of mature miRNA, the number of mismatches between query and subject sequences, % of A + U, minimal folding free energy, minimal folding free energy index, and the location of mature miRNA in the secondary structure (S3 Table).

Among all the non-protein-coding ESTs, 92 sequences were not the best-fitted candidates to be the potential miRNA for *Neospora caninum* (S5 File). Eventually, one potential EST candidate (CF938952.1) was explored as the precursor of the putative miRNA of *Neospora caninum* (S1 Table) against the reference miRNA dme-miR-9388-5p from miRBase. The secondary structure of pre miRNA achieved from Mfold web server formed a hairpin looped configuration (Figure 2). In the 85 nucleotides long pre-miRNA, the mature miRNA was sited in one arm of pre miRNA (gray box in Figure 2). The investigation also displayed the length of 19 nucleotides with no record of substitutions. The amount of A + U percentage in total length was 72.94% indicates a better precursor. The minimal folding free energy (MFE) and minimal folding free energy index (MFEI) depicted -29.60 and -1.51, respectively (Table 2).

Nomenclature and family annotation of predicted microRNA

The local blast result of miRBase revealed that miRNA of *Neospora caninum* genome belongs to "miR-9388-5p" family of mature miRNA. Therefore, the predicted miRNA was named as "nca-miR-9388-5p" (Figure 3). Result of comparative nucleotide base composition between putative miRNA, and homologous families obtained from miRBase is presented in Figure 4.

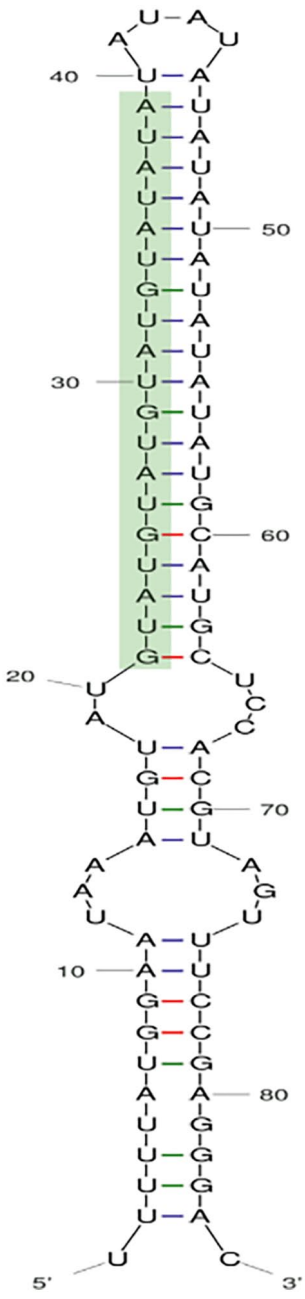

Figure 2. Secondary structures of final putative miRNA by using Mfold software.

The overall nucleotide percentage was documented as 31.58% Adenine, 47.37% Uracil, and 21.05% Guanine (Figure 4). The final miRNA candidate of the *Neospora caninum* genome did not contain cytosine. However, the uracil percentage of the reference miRNA family was displayed slightly lower than the resulting miRNA.

Prediction and functional analysis of putative Neospora caninum miRNA

From the TargetScanFly 7.2 web server, a total of 424 genes were found, making the best hit for target specification for by the putative miRNA of *Neospora caninum*. The functions involved with these genes were tabulated as molecular function, biological process, and cellular component (S2 Table), and the number of total genes involved with all three functions was 77.12%. Besides 84.67%, 85.61% and 81.13% genes played a role in molecular and biological; biological and cellular, and molecular and cellular functions, respectively. Moreover, in the KEGG database (https://www.genome.jp/kegg/pathway.html) for pathway analysis found no information about *Neospora caninum*. Therefore, to establish a probable relationship between obtained and expected genes functions, earlier literature was thoroughly reviewed describing the functions linked with any of the genes of either *Neospora* or other equipotential protozoans. After a comprehensive review of the literature, the study found 16 responsible genes (Table 3) that were linked with molecular, cellular, and biological functions. In particular, different sub-functions under the biological process were revealed as development, cell organization/biogenesis, response to the stimulus, cell cycle/ proliferation, transport/ localization, reproduction, immune system, nervous system process, behavior, signaling, gene expression, protein metabolism, DNA metabolism, small molecule metabolism (Figure 5). The most significant part of the gene pull (213 genes) was found to be engaged in the developmental function of the reference family. They contributed to different organ parts and body system development. Being the second highest, a total of 200 genes played a role in cell organization or biogenesis. Commonly they performed at mitotic sister chromatid segregation, myosin II filament organization, histone modification, and mitochondrial translational termination. There were 192 genes responded to various biological stimuli like detection of chemical stimulus concerning sensory perception, phototransduction, different signaling pathways, transmembrane transport, response to the odorant, cellular response to DNA damage stimulus, and DNA repair. The genes are involved with molecular function as enzyme, regulator, receptor, receptor binding, transporter, structural molecule, cytoskeletal binding, DNA binding, RNA binding, transcription factor, small molecule binding, metal ion binding, lipid binding and carbohydrate-binding (Figure 6). The highest pull of genes are engaged with different enzymatic activities (129) under molecular function. GTPase activity, serine-type endopeptidase activity, acyl-CoA dehydrogenase activity, butyryl-CoA dehydrogenase activity, glutathione transferase activity and endo-peptidase activity. Also, 87 were found to be involved with DNA/RNA binding activities. However, 64 genes are involved in the olfactory receptor, transmembrane signaling receptor, photoreceptor activity, hormone activity, ligand activity, insulin receptor binding, and G-protein coupled receptor activity. The cellular function includes extracellular region, cytosol, cytoskeleton, mitochondrion, nucleus, chromosome, membrane, cell periphery, cell junction, cell projection, endomembrane system, synapse, macromolecular complex (Figure 7). In total, 178 genes act for the dendrite membrane, plasma membrane, mitochondrial inner membrane, SNARE complex,

Table 2. Properties of putative miRNA of *Neospora caninum* with suggested criteria.

PROPERTIES	VALUE
NCBI accession number of EST	CF938952.1
Length of EST sequence	594 nucleotides
Reference miRNA from miRBase	dme-miR-9388-5p
Length of pre-miRNA	85
Number of each of nucleotides pre-miRNA sequence	A = 27, C = 8, G = 15 and U = 35
A + U content	72.94%
Length of mature miRNA	19 nucleotides
Mature miRNA sequence	GUAUGUAUGUAUGUAUAUA
Nucleotide mismatch	01
MFE of secondary structure	–29.60 kcal/mol
MFEI of secondary structure	–1.51 kcal/mol
Family of mature miRNA	miR-9388
Annotation of identified mature miRNA	nca-miR-9388-5p

Abbreviations: EST, expressed sequence tag; miRNA, microRNA; MFE, Minimal Folding Free Energy; MFEI, Minimal Folding Free Energy Index.

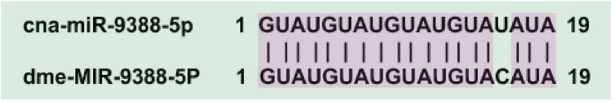

Figure 3. Alignments of the putative miRNA of *Neospora caninum* genome and its homologues from respective microRNA family.

an integral component of membrane, basolateral plasma membrane, and postsynaptic density membrane. Another pool with 136 genes play a role in developing nuclear envelope, nucleoplasm, cdc73 or ESC complex, histone methyltransferase complex, nuclear euchromatin of the entire nucleus. Also, 118 genes took part in the function associated with the different macromolecular complexes like alpha DNA polymerase complex, ubiquitin ligase complex, hetero-trimeric G-protein complex, gamma subunit complex, large mitochondrial subunit, and polysome, Smc5-Smc6 complex.

Discussion

The present study conducted an *in silico* ESTs based homology search to find out novel miRNA of *Neospora caninum* from a large volume of primary EST sequences available in the public domain. Approaching miRNA identification and characterization of *Neospora caninum* predicted one putative miRNA "nca-miR-9388-5p" of 19 nucleotides with MFEI value -1.51 kcal/mol and 72.94% of (A + U) content corresponding with its pre-miRNA. A comprehensive search for specific gene targets was performed resulting 16 potential genes associated with different protozoal physiological function. Significantly, the gene "Protein phosphatase" was found responsible for the virulence of *Neospora caninum*. The other genes are accounted for gene expression, vesicular transport, cell signaling, cell proliferation, DNA repair mechanism, and the developmental stage of different protozoa.

Though in the early days, miRNAs identification was only restricted to the cloning method, miRNAs identification is now a much simpler approach due to the advancement of computational biology and bioinformatics.[53] The present work was initiated with the acquisition of EST sequences and template miRNAs of *Neospora caninum*. We collected available ESTs of *Neospora caninum* from the NCBI Genebank database and reference miRNA sequences from miRBase to perform BLASTn following earlier studies.[25,36,54,55] We did not set up a cut-off point for E- value during the BLAST search due to retrieving the best matching region between subject EST and reference sequences. Thus, the sequences that displayed 100% similarity with very minimum mismatches were allowed to be the expected output sequences. The nucleotide length of pre-miRNA of *Neospora caninum* was counted as 85, which supports the results of the previous studies that ranged from 60-265.[21,26,56] The predicted miRNA carried 19 nucleotides within the range of 16-22 nucleotides suggested by the previous studies.[23,56-58] Earlier studies have shown that different plants and vertebrates pre-miRNAs containing higher (A + U)% than (G + C)% could have a better function in biochemical signals in miRNA biogenesis.[23,59] The current study considered above 70% as cut off of (A + U) % to find out the best-predicted miRNA. The way of predicting secondary structure from the information of primary sequences depends on reducing the free energy of a molecule by exploiting the quantity of satisfactory

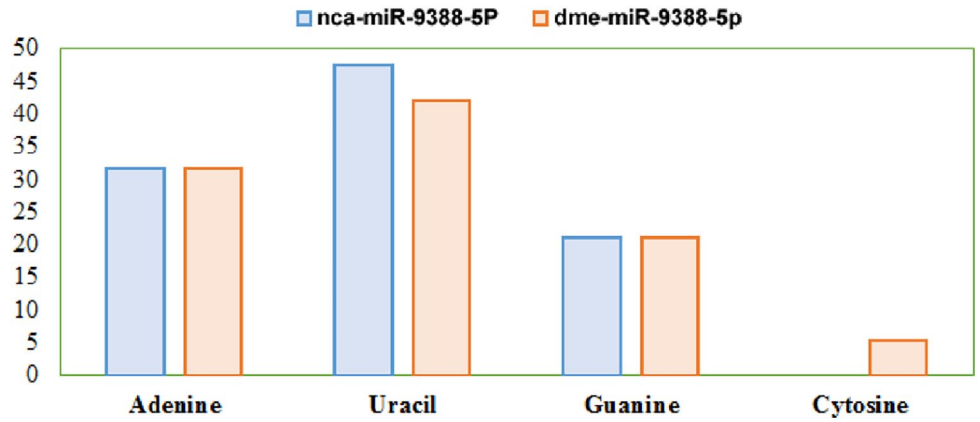

Figure 4. Overall nucleotide compositions (%) of putative miRNA from *Neospora caninum* compared with the homologues mRNA of miRBase.

Table 3. Putative miRNA targets in the crucial cellular process of *Neospora caninum.*

GENE NAME	NCBI ACC. NO	FUNCTION	REFERENCE
Protein phosphatase	49780	Role in virulence	Jike et al[36]
Formin 3	35256	Tachyzoite Growth and Motility	Loong and Mishra[37]
TBP-associated factor 1(Taf1)	40813	Cell proliferation, cell growth, cell cycle, transcription of many genes	Zuker and Stiegler[38]
Syntaxin	36173	Neurotransmission, membrane fusion and internal membrane compartment reconstruction	Schuster et al[39]
		Vesicular transport	Bonnet et al[40]
AKT1	41957	Cell signaling and promotion of cell survival	Harfe[41]
RNA-binding protein 4	41668	Developmental stage of *Trypanosoma* especially in epimastigote	Zhang et al[42]
RNA-binding protein 6	39919	Developmental stage of *Trypanosoma* especially in epimastigote	Johnston and Hobert[43]
Poly-(ADP-ribose) polymerase	3355109	DNA repair mechanisms and controlling power in different phases of cell growth	Reinhart et al[44]
Pumilio	41094	Gene expression in post-transcriotional stage	Hakimi and Cannella[45]
p21-activated kinase	44039	Cell morphology, cell-cycle and gene transcription	Gaidatzis et al[46]
		Polarity, motility and phagocytosis	Dahiya et al[47]
Chaperonin containing TCP1 subunit 2	31838	microtubule polymerizing	Johnston and Hobert[43]
Calpain C	32597	cell regulatory and differentiation processes	Liu et al[48]
Casein kinase II beta subunit	32132	Induction of mitogenic pathways	Al-Bajalan et al[49]
Ubiquitin conjugating enzyme 7	44054	Inhibit erythrophagocytosis	Avendaño-Borromeo et al[50]
ADP ribosylation	40506	Vesicular transport	Dacks and Doolittle[51]
DNZDHHC/NEW1 zinc finger protein 11	34503	Transcriptional mediator in various growth factors	Dacks and Doolittle[52]

Abbreviation: miRNA, microRNA; ADP, Adenosine diphosphate.

base-pairing interactions.[38,39] MFE is calculated by assessing the minimum energy values acquired by complementary base pairs diminished by the stacking energy of consecutive base pairs or amplified by the destabilizing energy accompanying with non-complementary bases.[38,39] According to Bonnet and co-workers,[40] maximum miRNAs demonstrate an affinity to show minimal folding free energy (MFE) compared to tRNAs and rRNAs. This characteristic is suitable to form a

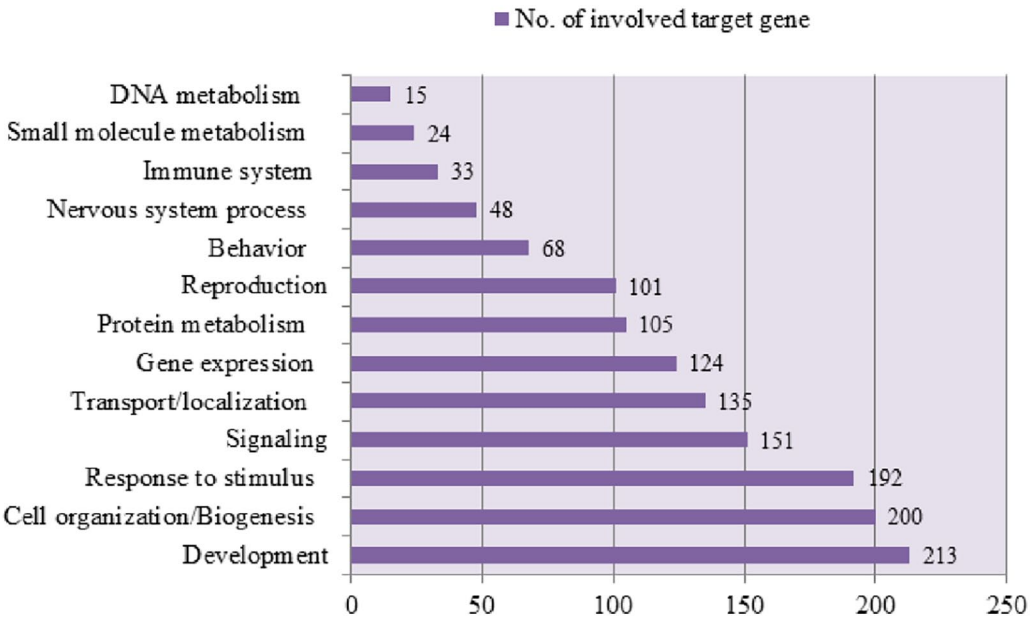

Figure 5. Biological Function of predicted target.

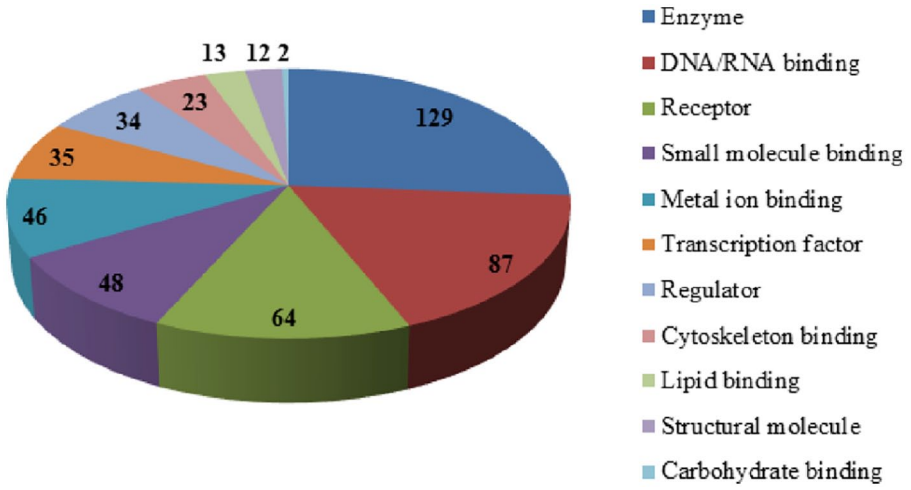

Figure 6. Molecular function of predicted target.

stable secondary structure by the pre-miRNAs. However, above 90% of miRNAs contain an MFEI value 0.85 that makes a clear demarcation from other RNAs like tRNAs, mRNA or rRNAs.[29] The final annotated miRNA of the current study having the MFE and MFEI values -29.60 and -1.51, which is in line with the findings of.[29] As a prerequisite of being miRNA, it must be located at the one arm of the secondary hairpin structure of pre-miRNA.[29] The maximum nucleotide involvement in the stem is the best structure for being a miRNA. The predicted miRNA of this study constituted with 19 nucleotides, is entirely located at the stem of the secondary hairpin structure of pre-miRNA without forming any internal loop or bulges.

Since the discovery of the non-protein-coding micro RNAs, the most pressing question is how these small molecules run their activities in body regulatory mechanisms? To understand the function of micro RNAs, identifying the responsible target genes under the micro RNA family is immensely important.[41,42] Formerly, several approaches based on the abnormal appearance of targeted miRNAs were popularly used for identifying the target genes of worm development.[43,44] First, the miRNA of *Caenorhabditis elegans* was found responsible for transferring one larval developmental stage to another.[10] Later with the progression of science, thousands of miRNA targets have been placed in the database performing through different computational approaches.[45-47] However, the KEGG pathway database lacked information about *Neospora caninum*, which forced us to search for the functions from previous literature. Likewise, some common genes were popped up having relation to molecular, biological and cellular functions of reference fly family "dme-miR-9388-5p" which also had the significance of different protozoal physiology. The present study only searched for protozoal

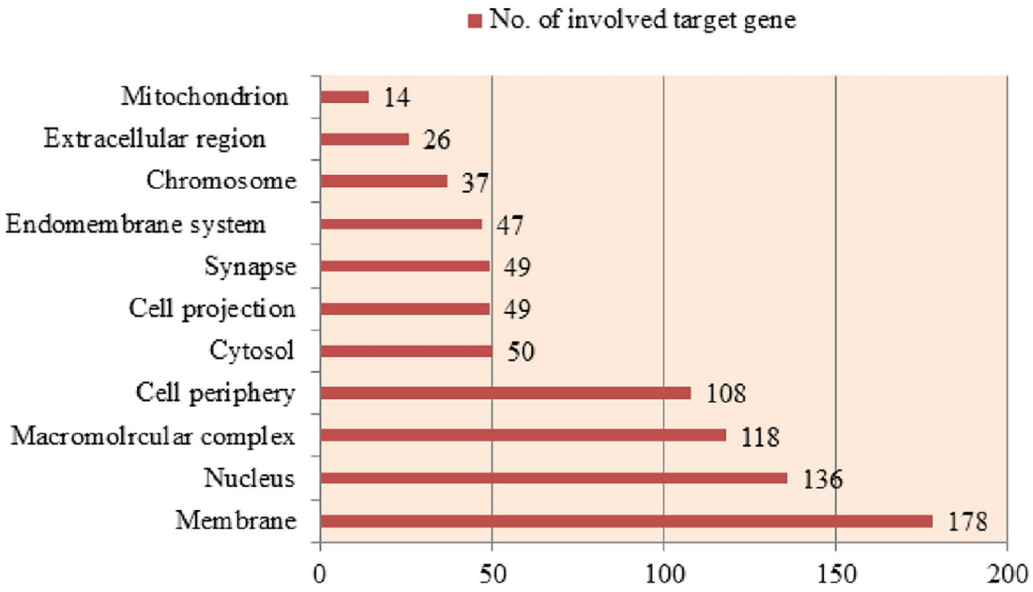

Figure 7. Cellular functions of predicted target.

gene function because *Neospora caninum* is a protozoan parasite that might share common genes with similar gene functions with other protozoa. One out of sixteen genes found in the present study is linked directly with *Neospora caninum* responsible for the virulence of the organism. The rest of the gene connected with other protozoans such as *Toxoplasma*, *Entamoeba*, *Trichomonas*, *Trypanosoma*, and *Leishmania. "Formin 3"* was identified in *Toxoplasma gondii* playing role in tachyzoite growth and gliding motility for invasion into the host cell.[48] As the Apicomplexan parasite *Toxoplasma* and *Neospora* share high similarities between the morphology, genomes and transcriptomes, - they must have some standard regulatory functions in infecting the host cell.[49] The gene recognized for host cell invasiveness in *Toxoplasma* might act in the same way in *Neospora. "TBP-associated factor 1 (Taf1)"* gene, a constituent of Transcription factor IID (TFIID), was found to be associated with cell proliferation, cell growth, cell cycle, cell cycle checkpoint, and various other processes in *Entamoeba histolytica*.[50] The function of the *"Syntaxin"* gene had to be related to vesicular transport of eukaryotic cells in various protozoan parasites like *Entamoeba, Giardia, Trichomonas* and *Trypanosoma*.[51] The other functions of this gene were neurotransmission, vesicular transport, membrane fusion, and even internal membrane compartment reconstruction connected with *Giardia*, *Trypanosoma*, and different kinds of algae.[52] This experiment of function analysis also documented the same evidence as cell cycle proliferation, cell organization/biogenesis, transport/localization, development and signaling. Another gene *"AKT1"* identified in this study was found responsible for cell signaling and promoting cell survival in *Leismania spp*.[60] Besides, two RNA binding proteins *"RBP4"* and *"RBP6"* play a significant role in the developmental stage of *Trypanosoma*, especially in epimastigotes.[57] The present study identified both the proteins. In *Trypanosoma cruzi* "Poly-(ADP-ribose) polymerase" enzyme has a vital role in DNA repair mechanisms and controlling power in different phases of cell growth[61] which is found to play a similar function in the case of "dme-miR-9388-5p" family. The "Pumilio," a RNA-binding protein, modulates gene expression mainly in post-transcriptional stage in Trypanosomes.[56] Besides in *Plasmodium*, a significant contribution of this protein has been reported. *Plasmodium* Puf2 is one of the key associates of the Pumilio family that is responsible for playing a vital role in the organism's developmental progression, especially during the sporozoite transmission from mosquito saliva to human.[58] The current study found "Protein Phosphatase" in *Neospora caninum* that has a significant effect on the virulence of the organism. The protein is proved to exist at profuse volume in a highly virulent strain of the organism.[59] Virulence in *Neospora caninum* is defined as the capability of prompting pathogenicity in laboratory organism mice and the trans-placental transmission from mother to offspring.[15] Some other previous studies also demonstrated the "Protein Phosphatase" as an invasion protein of other Apicomplexan parasites as example *Toxoplasma* and *Plasmodium* facilitating the entry of parasite into targeted host cell.[62-65]

Since the discovery of the first miRNA, from gene cloning-based primitive identification method to modern computational approach, thousands of miRNAs of plants and animal kingdom have been discovered. However, the microRNA database is still missing different important organisms and parasites. The current study endeavors to enrich the microRNA database by identifying the potential miRNA of *Neospora caninum* via EST-based homology search. In this study, we identified one putative mature miRNA of *Neospora caninum* and sixteen possible target genes. Although the only gene "Protein Phosphatase" was explored directly associated with *Neospora caninum.* It is known that the different organisms from common family origin can have homologous genes and exhibit

similar functions in different protozoa. Thus, identified genes in this study associated with other protozoa might function similarly for *Neospora caninum*. However, further wet-lab validation will be necessary to clarify the definite functions and relationship between genes. Nevertheless, the present experiment has been able to fill the research gaps on the protozoan parasite *Neospora caninum*. Further, the findings of this study may serve as the baseline in elucidating many unsolved questions related to the miRNAs biology of *Neospora caninum*.

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.I.A, S.A., and M.S.R.C. assisted in the retrieval of data and analysis. M.D. 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

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

1. Reichel MP, Alejandra Ayanegui-Alcérreca M, Gondim LF, Ellis JT. What is the global economic impact of Neospora caninum in cattle—the billion dollar question. *Int J Parasitol*. 2013;43:133-142. doi:10.1016/j.ijpara.2012.10.022.
2. Trees AJ, Davison HC, Innes EA, Wastling JM. Towards evaluating the economic impact of bovine neosporosis. *Int J Parasitol*. 1999;29:1195-1200. doi:10.1016/S0020-7519(99)00093-4.
3. Derdour S-Y, Hafsi F, Azzag N, et al. Prevalence of the main infectious causes of abortion in dairy cattle in Algeria. *J Vet Res*. 2017;61:337-343. doi:10.1515/jvetres-2017-0044.
4. Dubey JP. Review of Neospora caninum and neosporosis in animals. *Korean J Parasitol*. 2003;41:1-16.
5. Woodbine KA, Medley GF, Moore SJ, Ramirez-Villaescusa A, Mason S, Green LE. A four year longitudinal sero-epidemiology study of Neospora caninum in adult cattle from 114 cattle herds in south west England: associations with age, herd and dam-offspring pairs. *BMC Vet Res*. 2008;4:35. doi:10.1186/1746-6148-4-35.
6. Marugan-Hernandez V. Neospora caninum and Bovine Neosporosis: current vaccine research. *J Comp Pathol*. 2017;157:193-200. doi:10.1016/j.jcpa.2017.08.001.
7. Sánchez-Sánchez R, Vázquez P, Ferre I, Ortega-Mora LM. Treatment of toxoplasmosis and neosporosis in farm ruminants: state of knowledge and future trends. *Curr Top Med Chem*. 2018;18:1304-1323.
8. Aravin AA, Lagos-Quintana M, Yalcin A, et al. The small RNA profile during Drosophila melanogaster development. *Dev Cell*. 2003;5:337-350. doi:10.1016/S1534-5807(03)00228-4.
9. Bentwich I, Avniel A, Karov Y, et al. Identification of hundreds of conserved and nonconserved human microRNAs. *Nat Genet*. 2005;37:766-770. doi:10.1038/ng1590.
10. Zhao Y, Srivastava D. A developmental view of microRNA function. *Trends Biochem Sci*. 2007;32:189-197. doi:10.1016/j.tibs.2007.02.006.
11. Sengar GS, Deb R, Singh U, et al. Differential expression of microRNAs associated with thermal stress in Frieswal (Bos taurus x Bos indicus) crossbred dairy cattle. *Cell Stress Chaperones*. 2018;23:155-170.
12. Szweykowska-Kulinska Z, Jarmolowski A. Post-transcriptional regulation of MicroRNA accumulation and function: new insights from plants. *Mol Plant*. 2018;11:1006-1007.
13. Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell*. 2009;136:215-233.
14. Tiwari N, Kumar V, Gedda MR, et al. Identification and characterization of miRNAs in response to *Leishmania donovani* infection: delineation of their roles in macrophage dysfunction. *Front Microbiol*. 2017;8:314. doi:10.3389/fmicb.2017.00314.
15. Al-Qassab S, Reichel MP, Ellis J. A second generation multiplex PCR for typing strains of Neospora caninum using six DNA targets. *Mol Cell Probes*. 2010;24:20-26. doi:10.1016/j.mcp.2009.08.002.
16. Reid AJ, Vermont SJ, Cotton JA, et al. Comparative genomics of the apicomplexan parasites Toxoplasma gondii and Neospora caninum: Coccidia differing in host range and transmission strategy. *PLoS Pathog*. 2012;8:e1002567. doi:10.1371/journal.ppat.1002567.
17. Ellis J, Luton K, Baverstock PR, Brindley PJ, Nimmo KA, Johnson AM. The phylogeny of Neospora caninum. *Mol Biochem Parasitol*. 1994;64:303-311. doi:10.1016/0166-6851(94)00033-6.
18. Schock A, Innes EA, Yamane I, Latham SM, Wastling JM. Genetic and biological diversity among isolates of Neospora caninum. *Parasitology*. 2001;123:13-23. doi:10.1017/S003118200100796X.
19. Zheng Y, Cai X, Bradley JE. microRNAs in parasites and parasite infection. *RNA Biol*. 2013;10:371-379. doi:10.4161/rna.23716.
20. Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res*. 2019;47:D155-D162. doi:10.1093/nar/gky1141.
21. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25:3389-3402. doi:10.1093/nar/25.17.3389.
22. Fu X, Zhu W, Cai L, et al. Improved pre-miRNAs identification through mutual information of pre-miRNA sequences and structures. *Front Genet*. 2019;10:119. doi:10.3389/fgene.2019.00119.
23. Akter A, Islam MdM, Mondal SI, et al. Computational identification of miRNA and targets from expressed sequence tags of coffee (Coffea arabica). *Saudi J Biol Sci*. 2014;21:3-12. doi:10.1016/j.sjbs.2013.04.007.
24. Singh N, Srivastava S, Sharma A. Identification and analysis of miRNAs and their targets in ginger using bioinformatics approach. *Gene*. 2016;575:570-576.
25. 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.
26. Gao D, Middleton R, Rasko JEJ, Ritchie W. miREval 2.0: a web tool for simple microRNA prediction in genome sequences. *Bioinformatics*. 2013;29:3225-3226. doi:10.1093/bioinformatics/btt545.
27. Reeder J, Höchsmann M, Rehmsmeier M, Voss B, Giegerich R. Beyond Mfold: recent advances in RNA bioinformatics. *J Biotechnol*. 2006;124:41-55. doi:10.1016/j.jbiotec.2006.01.034.
28. Wiese KC, Hendriks A. Comparison of P-RnaPredict and mfold—algorithms for RNA secondary structure prediction. *Bioinformatics*. 2006;22:934-942. doi:10.1093/bioinformatics/btl043.
29. 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.
30. Griffiths-Jones S, Grocock RJ, Van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res*. 2006;34:D140-D144.
31. Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics*. 2003;2:Unit 2.3.
32. Agarwal V, Subtelny AO, Thiru P, Ulitsky I, Bartel DP. Predicting microRNA targeting efficacy in Drosophila. *Genome Biol*. 2018;19:152. doi:10.1186/s13059-018-1504-3.
33. Kaletta T, Hengartner MO. Finding function in novel targets: C. elegans as a model organism. *Nat Rev Drug Discov*. 2006;5:387-398. doi:10.1038/nrd2031.
34. Roberts DB. Drosophila melanogaster: the model organism. *Entomol Exp Appl*. 2006;121:93-103. doi:10.1111/j.1570-8703.2006.00474.x.
35. Mueller GM. *Biodiversity of Fungi: Inventory and Monitoring Methods*. Amsterdam, The Netherlands: Elsevier; 2011.
36. 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.
37. Loong SNK, Mishra SK. Unique folding of precursor microRNAs: quantitative evidence and implications for de novo identification. *RNA*. 2007;13:170-187. doi:10.1261/rna.223807.
38. Zuker M, Stiegler P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res*. 1981;9:133-148.
39. Schuster P, Fontana W, Stadler PF, Hofacker IL. From sequences to shapes and back: a case study in RNA secondary structures. *Proc Biol Sci*. 1994;255:279-284.
40. Bonnet E, Wuyts J, Rouze P, Van de Peer Y. Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. *Bioinformatics*. 2004;20:2911-2917. doi:10.1093/bioinformatics/bth374.

41. Harfe BD. MicroRNAs in vertebrate development. *Curr Opin Genet Dev*. 2005;15:410-415. doi:10.1016/j.gde.2005.06.012.
42. Zhang B, Wang Q, Pan X. MicroRNAs and their regulatory roles in animals and plants. *J Cell Physiol*. 2007;210:279-289.
43. Johnston RJ, Hobert O. A microRNA controlling left/right neuronal asymmetry in Caenorhabditis elegans. *Nature*. 2003;426:845-849. doi:10.1038/nature02255.
44. Reinhart BJ, Slack FJ, Basson M, et al. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature*. 2000;403:901-906. doi:10.1038/35002607.
45. Hakimi M, Cannella D. Apicomplexan parasites and subversion of the host cell microRNA pathway. *Trends Parasitol*. 2011;27:481-486. doi:10.1016/j.pt.2011.07.001.
46. Gaidatzis D, van Nimwegen E, Hausser J, Zavolan M. Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC Bioinformatics*. 2007;8:69. doi:10.1186/1471-2105-8-69.
47. Dahiya N, Sherman-Baust CA, Wang T-L, et al. MicroRNA expression and identification of putative miRNA targets in ovarian cancer. *PLoS One*. 2008;3:e2436. doi:10.1371/journal.pone.0002436.
48. Liu W, Apagyi K, McLeavy L, Ersfeld K. Expression and cellular localisation of calpain-like proteins in Trypanosoma brucei. *Mol Biochem Parasitol*. 2010;169:20-26.
49. Al-Bajalan MMM, Xia D, Armstrong S, Randle N, Wastling JM. Toxoplasma gondii and Neospora caninum induce different host cell responses at proteome-wide phosphorylation events; a step forward for uncovering the biological differences between these closely related parasites. *Parasitol Res*. 2017;116:2707-2719. doi:10.1007/s00436-017-5579-7.
50. Avendaño-Borromeo B, Narayanasamy RK, García-Rivera G, et al. Identification of the gene encoding the TATA box-binding protein-associated factor 1 (TAF1) and its putative role in the heat shock response in the protozoan parasite Entamoeba histolytica. *Parasitol Res*. 2019;118:517-538. doi:10.1007/s00436-018-6170-6.
51. Dacks JB, Doolittle WF. Molecular and phylogenetic characterization of syntaxin genes from parasitic protozoa. *Mol Biochem Parasitol*. 2004;136:123-136. doi:10.1016/j.molbiopara.2004.02.014.
52. Dacks JB, Doolittle WF. Novel syntaxin gene sequences from Giardia, Trypanosoma and algae: implications for the ancient evolution of the eukaryotic endomembrane system. *J Cell Sci*. 2002;115:1635-1642.
53. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. 2004;116:281-297.
54. 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.
55. 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. doi:10.1371/journal.pone.0206154.
56. Dallagiovanna B, Pérez L, Sotelo-Silveira J, Smircich P, Duhagon MA, Garat B. Trypanosoma cruzi: molecular characterization of TcPUF6, a Pumilio protein. *Exp Parasitol*. 2005;109:260-264.
57. Romaniuk MA, Cervini G, Cassola A. Regulation of RNA binding proteins in trypanosomatid protozoan parasites. *World J Biol Chem*. 2016;7:146-157. doi:10.4331/wjbc.v7.i1.146.
58. Gomes-Santos CSS, Braks J, Prudêncio M, et al. Transition of plasmodium sporozoites into liver stage-like forms is regulated by the RNA binding protein Pumilio. *PLoS Pathog*. 2011;7:e1002046. doi:10.1371/journal.ppat.1002046.
59. Unjit K. *The Effects of Development and Virulence on Gene Expression in the Protozoan Parasite Neospora Caninum* [PhD thesis]. Liverpool, UK: University of Liverpool; 2018.
60. Varela-M RE, Ochoa R, Muskus CE, Muro A, Mollinedo F. Identification of a RAC/AKT-like gene in *Leishmania* parasites as a putative therapeutic target in leishmaniasis. *Parasites Vectors*. 2017;10:458.
61. Vilchez Larrea SC, Haikarainen T, Narwal M, et al. Inhibition of poly(ADP-ribose) polymerase interferes with Trypanosoma cruzi infection and proliferation of the parasite. *PLoS One*. 2012;7:e46063. doi:10.1371/journal.pone.0046063.
62. Dubey JP, Buxton D, Wouda W. Pathogenesis of bovine neosporosis. *J Comp Pathol*. 2006;134:267-289. doi:10.1016/j.jcpa.2005.11.004.
63. Dubey JP, Schares G, Ortega-Mora LM. Epidemiology and control of neosporosis and Neospora caninum. *Clin Microbiol Rev*. 2007;20:323-367. doi:10.1128/CMR.00031-06.
64. Innes EA, Wright S, Bartley P, et al. The host-parasite relationship in bovine neosporosis. *Vet Immunol Immunopathol*. 2005;108:29-36. doi:10.1016/j.vetimm.2005.07.004.
65. Innes EA. The host-parasite relationship in pregnant cattle infected with Neospora caninum. *Parasitology*. 2007;134:1903-1910. doi:10.1017/S0031182007000194.