PLOS ONE

RESEARCH ARTICLE

A landscape of gene regulation in the parasitic amoebozoa *Entamoeba* spp

Edgardo Galán-Vásquez [ID]¹*, **María del Consuelo Gómez-García**², **Ernesto Pérez-Rueda** [ID]³*

1 Departamento de Ingeniería de Sistemas Computacionales y Automatización, Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, Ciudad Universitaria, Ciudad de México, México, 2 Laboratorio de Biomedicina Molecular, Escuela Nacional de Medicina y Homeopatía, Instituto Politécnico Nacional, Ciudad de México, México, 3 Unidad Académica Yucatán, Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, Mérida, Yucatán, México

* edgardo.galan@iimas.unam.mx (EG-V); ernesto.perez@iimas.unam.mx (EP-R)

Abstract

Entamoeba are amoeboid extracellular parasites that represent an important group of organisms for which the regulatory networks must be examined to better understand how genes and functional processes are interrelated. In this work, we inferred the gene regulatory networks (GRNs) in four *Entamoeba* species, *E. histolytica*, *E. dispar*, *E. nuttalli*, and *E. invadens*, and the GRN topological properties and the corresponding biological functions were evaluated. From these analyses, we determined that transcription factors (TFs) of *E. histolytica*, *E. dispar*, and *E. nuttalli* are associated mainly with the LIM family, while the TFs in *E. invadens* are associated with the RRM_1 family. In addition, we identified that EHI_044890 regulates 121 genes in *E. histolytica*, EDI_297980 regulates 284 genes in *E. dispar*, ENU1_120230 regulates 195 genes in *E. nuttalli*, and EIN_249270 regulates 257 genes in *E. invadens*. Finally, we identified that three types of processes, Macromolecule metabolic process, Cellular macromolecule metabolic process, and Cellular nitrogen compound metabolic process, are the main biological processes for each network. The results described in this work can be used as a basis for the study of gene regulation in these organisms.

Introduction

The passage of information from DNA to RNA (transcription) is a fundamental mechanism for all organisms [1]. The mechanism for transcription involves a large number of molecules (proteins, enzymes, and DNA sequences, among others) that together orchestrate and carry out the expression of genes in a highly precise, spatially and temporally controlled manner to meet the needs of the cell. Transcription factors (TFs) are essential proteins in this event and are part of the cell's ability to have differential and temporal expressions, increases or decreases in the amounts of transcripts, etc. They do this by interacting with cis-consensus DNA sequences present in gene promoters and with general TFs [2].

TFs may have different types of domains through which they bind to DNA, interact with other proteins to regulate activation, and locate spatially within the cell, and they have motifs through which they undergo different types of posttranslational modifications, such as phosphorylation, acetylation, methylation, SUMOylation, and ubiquitination [3], which are crucial for TF function. Numerous studies on TFs have demonstrated how conserved they are and their specificity for binding to DNA in various species, such as mice [4]. However, protozoa are a group for which very little is known about transcription and TFs.

Particularly in the *Entamoebidae* family, parasitic species have been described, including *Entamoeba histolytica*, the causal agent of intestinal amebiasis and amoebic liver abscess in humans, which causes 100,000 annual deaths [5–9]. *E. nuttalli* infects macaques and different species of monkeys and causes intestinal amebiasis, liver abscesses, and even death. This species is responsible for serious health problems in zoos and in various regions of the planet, such as nature reserves [10–12]. *E. invadens* is a parasite of reptiles (ophidians, saurians, and chelonians) and causes gastrointestinal damage from mild to severe and is the only species of the genus *Entamoeba* that encysts *in vitro* [13]. Finally, *E. dispar* is considered a nonpathogenic species that lives as a commensal in humans; first described in 1993 for Diamond and Clark [14] was posteriorly identified as able to produce liver and intestinal lesions that were occasionally indistinguishable from those produced by *E. histolytica* [15].

These four species of *Entamoeba* are amoeboid extracellular parasites that move via the emission of pseudopods and lack mitochondria; this is why they are located in the early phase of eukaryotic evolution [16]. Only in *E. histolytica* and more recently in *E. invadens* has the presence of mitosomes been described [17,18]. These organisms present two phases in their life cycle: the cyst, which is the infective form, and the trophozoite, which is the invasive form [16,19].

These four species of *Entamoeba* are undoubtedly of great importance in the parasitology not only of humans, but also of other organisms. Therefore, the genomes of these four species have already been sequenced. *E. histolytica* has a genome rich in AT (75%) with a size of 20,800,560 bp and 8,333 genes [20]; *E. dispar* has a genome similar to that of *E. histolytica*, with 8,749 genes and a composition of 76.5% AT and size of 22,955,291 bp. *E. nuttalli* has a genome that contains 74.9% AT, a size of 14,399,953 bp, and 6,193 genes, and of the four mentioned species it has the smallest genome [21]. *E. invadens* has the largest genome of these four species, with 40,888,805 bp, and AT content of 70%, and 11,549 genes [22].

A gene regulatory network (GRN) is a directed graph in which interaction edges connect TFs to target genes (TGs) [23]. This type of network greatly helps in understanding the links between genes and the products they encode, which is a crucial and difficult step in experimental and computational biology [24]. Only a few GRNs have been reconstructed in model organisms [24–27]. As a result, homology-based techniques are frequently used to research GRNs in species that are less well-known [28–31].

In this work, we inferred the GRNs of the four main *Entamoeba* species using a criterion of TF-TG orthology relationships from reference GRNs experimentally described. The reconstructed GRNs were posteriorly analyzed in terms of topology. We consider that the GRN inferences for these strains open the opportunity to explore organisms of public health importance.

Material and methods

The network inference process steps were described in the schematic workflow (Fig 1). Details on each step, including input and output data, are described in follow.
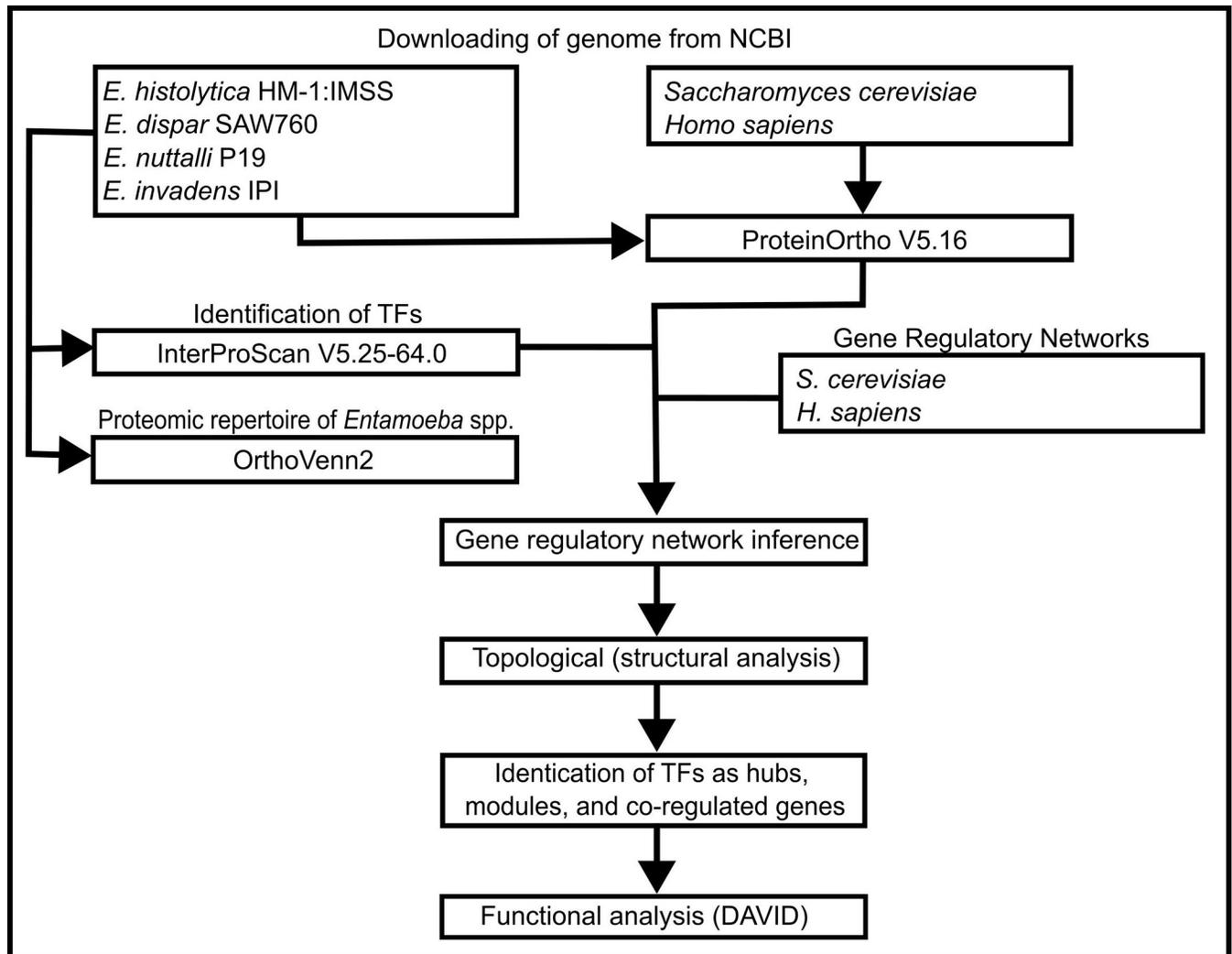
Fig 1. Schematic workflow of the network inference procedure steps. Four Entamoeba genomes were downloaded from NCBI and compared with Orthovenn, to infer the common set of orthologous proteins. Interproscan was used to assign protein domains (Pfam and Supfam). To infer the GRN, two organismal models were used, S.cerevisiae and H. sapiens, and the ProteinOrtho was considered. The GRNs inferred were evaluated in terms of their topological properties, hubs, and functional descriptions.

https://doi.org/10.1371/journal.pone.0271640.g001

Genomes analyzed

Four genomes of *Entamoeba* spp. were downloaded from the NCBI server: *E. histolytica* HM-1:IMSS (GCF_000208925.1), *E. dispar* SAW760 (GCF_000209125.1), *E. nuttalli* P19 (GCF_000257125.1), and *E. invadens* IP1 (GCF_000330505.1). Additionally, the genomes of *Saccharomyces cerevisiae* (GCF_000146045.2), and *Homo sapiens* (GCF_000001405.39) were downloaded to be used for the GRN inferences.

Proteomic repertoire in *Entamoeba* spp.

OrthoVenn2 was used to identify orthologous clusters in the four proteomes, and to perform a functional enrichment analysis for each cluster, we used an E-value of 0.01 as cutoff for all-to-all protein similarity comparisons, and an inflation value of 1.5 for the orthologous clustering employing the Markov Cluster algorithm. The enrichment analysis was considered significant with a P-value less than 0.05 [32].

Identification of TFs

To assess the TF diversity, protein sequences of whole proteomes were used to identify DNA-binding domains (DBDs) associated with regulatory proteins. To do this, InterProScan (v5.25–64.0) [33] was used to map InterPro families and DBDs, using default parameters. Afterwards, 162 Pfam IDs obtained from the TF database and by literature lookup were compiled and identified in the associated predictions (S1 Table).

Reconstruction of GRNs

Two organisms were considered templates for the inferences of the GRNs of *Entamoeba*. The GRN of *S. cerevisiae* was obtained from the YEASTRACT database and is composed of 6,709 nodes and 179,601 edges [34]. The GRN of *H. sapiens* with 2,862 nodes and 8,427 edges was obtained from the TRRUST database v2 [35].

To identify orthologous proteins between the four *Entamoeba* proteomes and the proteomes of *S. cerevisiae* and *H. sapiens*, were used to identify the orthologs using the program ProteinOrtho (V5.16) [36], with the following parameters: E-value of 0.01, a sequence coverage ≥ 50%, and minimal percent identity of best Blast hits of 30%.

To infer the GRNs, we map their interactions considering the following criteria: If the orthologs of the TF and its TG of the model organism (*S. cerevisiae* and / or *H. sapiens*), were found in a new genome, the interaction was assigned using guilt by association. Then, each network was integrated using all the ortholog assignments with the two six reference GRNs. All the network interactions can be inferred by running the scripts, provided as supplementary data 1.

Network structural analysis

To determine the topologies of the reconstructed networks, the following metrics were calculated: the number of edge incidents with other nodes, *i.e.*, the Node degree (K). In GRNs, input degree (Kin) is the number of arrows that enter anode, which corresponds to the TFs that affect a TG, and output degree ($Kout$) is the number of arrows that leave a node, which corresponds to the number of TGs by which a TF is regulated [37,38].

A clustering coefficient measures how connected a node's neighbors are to one another. It is calculated as, "the number of edges connecting a node's neighbors divided by the total number of possible edges between the node neighbors" [39]. The connectivity in an undirected network is the link between two nodes, and this link can be via a direct or indirect edge through intermediate connections. In this context, a connected component is a set of nodes that are linked to each other node by a path, and the component with the most proportions of nodes is called a giant component [38].

Some metrics are proposed to identify the relevance of nodes in a network. Hubs, which are defined as the most connected nodes with other nodes, confer the global structure of the network. Centrality (C), which measures the contribution or importance of nodes, sets node u as more important than another node v if $C(u) > C(v)$. The most relevant centrality metrics are: degree, closeness, betweenness, and eigenvector centrality, which assigns every $v \in V$ of a given graph G a value $C(v) \in R$ [38].

Functional annotation analysis

To determine the function enriched in each network, we used the Database for Annotation, Visualization and Integrated Discovery (DAVID 6.81), a gene functional classification system that integrates a set of functional annotation tools [40]. Each list of genes from the networks

was used to perform an enrichment analysis in Gene Ontology terms, and a statistical significance at a P-value of < 0.05 was set.

Results and discussion

Protein similarities of *Entamoeba* spp.

In order to evaluate how similar, the proteomic repertoire of *Entamoeba* species are, we analyzed the shared orthologous proteins between *E. histolytica*, *E. nuttalli*, *E. invadens*, and *E. dispar* genomes and displayed them in OrthoVenn2 [32]. The OrthoVenn revealed 4,285 clusters of 18,545 orthologous proteins that are shared by all species, accounting for 55.86% of the *E. histolytica* proteome, 51.93% of the *E. dispar* proteome, 70.73% of the *E. nuttalli* proteome, and 41.95% of the *E. invadens* proteome (Fig 2). The main functions associated with these proteins correspond to metabolic processes (GO:0008152), cellular processes (GO:0009987), and macromolecule metabolic processes (GO:0043170). The second longest group of 1,154 clusters includes only proteins of *E. nuttalli* (1,157 proteins), *E. histolytica* (1,205 proteins), and *E. dispar* (1,222 proteins) (Fig 2). These results show that the four species share ~50% of their proteins, which makes sense, as they are organisms of the same genus and all of them can infect a host. However, it is intriguing that *E. dispar*, shares 51.93% of its proteome with *E. histolytica*, a species capable of colonizing the human intestine and even inducing the development of amoebic liver abscesses, the more serious form of disease; this suggests that other factors such as host or environmental factors, in addition to some genetic factors, may define the pathogenicity of these two species. Likewise, the observation that *E. nuttalli*, a species whose host is the macaque, shares a very high percentage of proteins with the other three species is striking, particularly because it has been shown that *E. nuttalli* is the closest species to *E. histolytica* [8,16].

On the other hand, we identified singletons, which are proteins not grouped in any cluster. In this context, *E. histolytica* contains 626 singletons that are mainly related to organic substance metabolic process, primary metabolic process, and cellular metabolic process; *E. dispar*

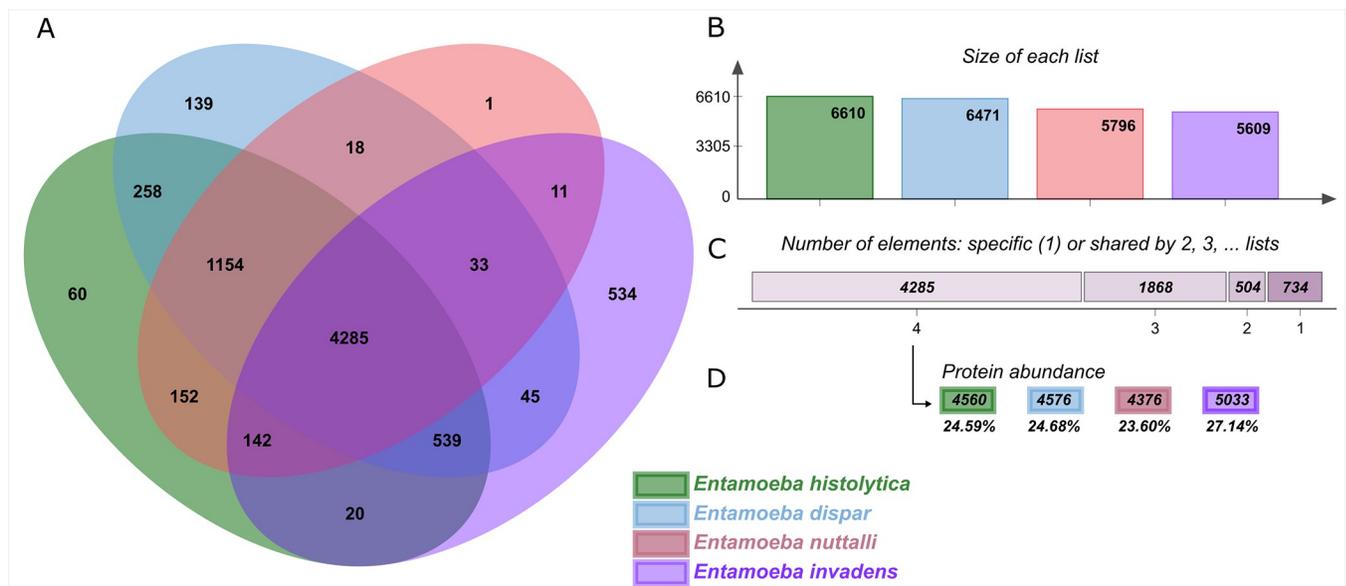**Fig 2. Orthologous proteins shared between Entamoeba strains.** A) Orthologous clusters of whole proteomes. B) The bar plot graph shows the number of orthologous clusters by organism. C) The plot indicates the number of clusters that are organism specific or shared by 2, 3, or 4 organisms. D) For the 4,285 clusters shared by 4 organisms, the protein abundance levels are shown for each organism.

contains 1,137 singletons mainly related to primary metabolic process, organic substance metabolic process, and cellular metabolic process; *E. nuttalli* contains 286 singletons mainly related to biosynthetic process and single-organism signaling. Finally, *E. invadens* contains 2,927 singletons mainly related to single-organism cellular process, cellular response to stimulus, regulation of cell process, and single-organism signaling. These results show that *E. invadens* presents more single proteins than *E. dispar*, *E. histolytica*, and *E. nuttalli*. This may be because *E. invadens* has a much larger genome that is almost twice the size of the genomes of the other three species. Therefore, the genetic information contained in *E. invadens* is probably necessary to infect different species of reptiles, such as turtles, lizards, and snakes [41]. Thus, the processes and mechanisms in which these proteins are involved (cellular process, cellular response stimulus, and regulation of cell process) are different from those of *E. histolytica* and *E. dispar* and to a lesser extent those of *E. nuttalli*. A similar result was previously described for *E. histolytica*, *E. dispar*, *E. invadens*, and *E. moshkowskii* [16]. Therefore, amoebic genetic diversity may vary depending on host species.

Identification of transcription factors

A TF repertoire consists of a set of proteins that regulate gene expression in the cell. Based on the PFAM assignments from InterProScan, we identified a set of 242 TFs in *E. histolytica*, 297 TFs in *E. invadens*, 210 TFs in *E. nuttalli*, and 245 TFs in *E. dispar*, representing 2.9%, 2.57%, 3.39%, and 2.8% of each proteome, respectively. The number of TFs identified in each species of *Entamoeba* seems not to be so different between them, nor to be related to the size of their genomes. However, the number of TFs obtained for each species is within the range of TFs stipulated for other organisms, as it is estimated that TFs constitute between 0.5 and 8% of the genes contained in the genomes of eukaryotic organisms [42].

Interestingly, the TFs predicted in *E. histolytica* are distributed among 11 families, whereas the TFs predicted in *E. invadens* are distributed among 87 families, 103 families in *E. nuttalli*, and 99 families in *E. dispar*. The most abundant family in *E. histolytica*, *E. dispar*, and *E. nuttalli* is the LIM domain (PF004112) (Fig 3), which is a protein structural domain containing two Zn2+ fingers separated by a 2-amino-acid hydrophobic linker. The LIM domain can bind a wide variety of protein targets and is widely distributed among plants, fungi, protozoa, and animals [43,44]. However, to date, only in the species *E. histolytica* has a protein with this domain, called EhLimA, been identified; EhLimA is associated with the actin of the parasite cytoskeleton and membrane [44]. Several possible LIM proteins have been identified in the *E. histolytica* genome, some of which could be relevant proteins in transcription for this *Entamoeba* species [45]. On the other hand, the TFs of *E. invadens* are associated with the RRM_1 family (PF00076) (Fig 3), which is a putative RNA-binding domain of approximately 90 amino acids and is known to bind single-stranded RNAs [46]; it is found abundantly in all life kingdoms [47,48]. Nevertheless, no protein with this domain has been characterized in any *Entamoeba* species to date. Interestingly, the proteins that contain this domain participate in the preprocessing of mRNAs, alternative splicing, stability, edition, and export of mRNAs, and thus they are fundamental in the biological processes of the organism [49]. This contrast in the number of families in which the TFs of these four *Entamoeba* species are distributed may be due to the characteristics of each species, that is, their forms and lifestyles, structures, and life cycles between other factors that have allowed them to specialize according to their needs. For example, *E. invadens* has the ability to infect different types of reptiles (snakes, turtles, and lizards), organisms in which osmotic changes or abrupt carbon source depletion in the intestine can be common, whereas in the human intestine such abrupt changes do not occur [50]. Thus, the type of TF families other *Entamoeba* species require may be different, even having species-
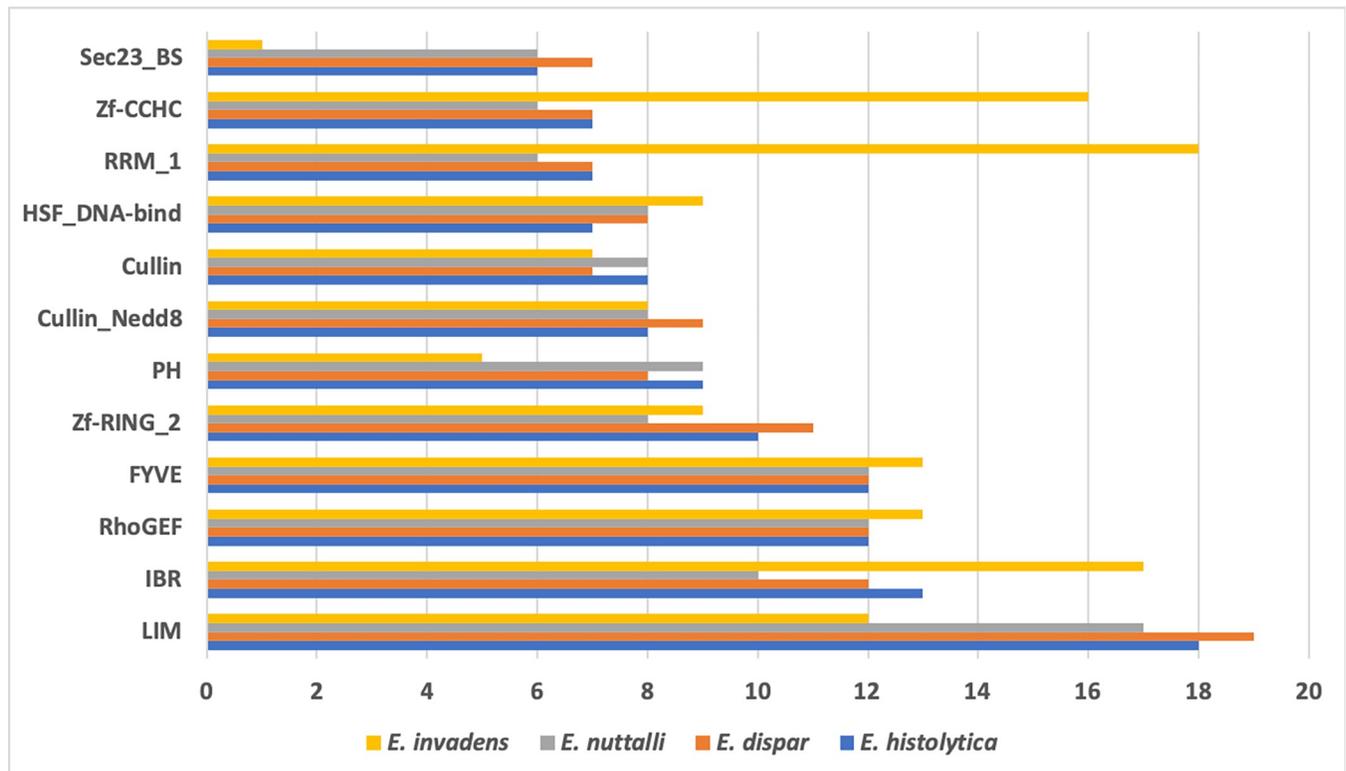
Fig 3. The most abundant families in *Entamoeba* strains. On the X-axis is the number of proteins; the Y-axis indicates the TF family names.

https://doi.org/10.1371/journal.pone.0271640.g003

exclusive TFs. In the case of *E. histolytica* and *E. dispar*, the species are morphologically indistinguishable in both the cyst and the trophozoite forms [51]. Therefore, on the one hand, *E. dispar* must have a series of genes that are repressed or activated and prevent it from causing disease in its host [52], whereas *E. histolytica* seems to have a series of mechanisms, including transcriptional control, that allow it to cause amoebic colitis or a liver abscess in its host, or not. Some of these molecules are enzymes such as glycosidases (sialidase, N-acetylgalactosamidase, and N-acetylglucosaminidase) which are necessary for the invasion of the epithelium, cysteine proteases (CP-A4, CP-A6, EhCPADH, and CP-B1, among others) which kill inflammatory and epithelial cells; or proteins such as the amoebopore that causes cell cytolysis, or the Gal/GalNAc lectin, which is involved in adhesion and the cytopathic effect [53,54] *E. nuttalli*, whose host is the macaque, has a life cycle similar to that of other *Entamoeba* species, and it is also capable of infecting humans but generates an asymptomatic infection. This suggests that this species must also present genomic plasticity and, therefore, specific transcription patterns that require a greater versatility of TFs. Therefore, a finely controlled transcriptional regulation must be carried out in which different TFs are necessary, hence the possible versatility of the families of TFs found in this work. However, we also observed that there are conserved TF families in the four species (Fig 3), in which the similar basic TF repertoire can be found, which may be performing basal transcription in these organisms. Finally, the species *E. histolytica*, *E. dispar*, and *E. nuttalli* present a similar number of TFs from the most abundant families, which coincides with the fact that the three species present a close phylogenetic relationship [51].

In general, few of the TFs identified in this work by sequence analysis have been previously characterized, such as, Nuclear factor Y (NF-Y) that appears at a later time point of *Entamoeba*

encystation [55–57]; the Ehp53, homologous to the tumor suppressor protein p53 [58] from human and *Drosophila melanogaster*. Ehp53 contains seven of the eight DNA-binding residues and two of the four Zn2+-binding sites described for p53. Heterologous monoclonal antibodies against p53 (Ab-1 and Ab-2) recognized a single 53 kDa spot in two-dimensional gels and they inhibited the formation of DNA-protein complexes produced by the interaction of nuclear extracts of *E. histolytica* with an oligonucleotide containing the consensus sequence for the binding of human p53. [58]. In addition, a calcium-sensitive EF-hand protein that binds to the URE3 motif [59], and two proteins that bind to the URE4 sequence (EhEBP1 and EhEBP2) [60], have also been identified.

Finally, we identified by sequence comparisons, members of the Myb-SHAQKYF family. These proteins have been previously identified as differentially expressed in trophozoites under basal cell culture conditions. Members of this group harbor a highly conserved and structured Myb-DBD and a large portion of intrinsically disordered residues. As the Myb-DBD of these proteins harbors a distinctive Q[VI]R[ST]HAQK[YF]F sequence in its putative third α-helix. An NMR structure of the Myb-DBD of EhMybS3 shows that this protein is composed of three α-helices stabilized by a hydrophobic core, similar to Myb proteins of different kingdoms [61]. Therefore, our approach opens the possibility to characterize experimentally diverse TFs with hypothetical functions, predicted in this work.

Regulatory networks

The GRN is defined as a graph $G = (V, A)$, where V is a set of nodes that correspond to genes or proteins in the network and A is a set of edges that correspond to relationships between two nodes. Few GRNs have been reconstructed from experimental data; comparative genomics approaches are usually used for the reconstruction of GRNs in little-known organisms. To this end, the GRN from a model organism can be used as a template to export interactions in the organism of interest; under this approach, orthologous TFs generally regulate the expression of orthologous TGs [28,62]. Therefore, to identify the GRNs of the four *Entamoeba* species, the GRNs of *S. cerevisiae* and *H. sapiens* were used as references. A regulatory association was established when orthologues of a TF-TG relationship in a model organism were found for both a TF and a TG in the target organism (Table 1) (S2 Table) [28,62].

The inferred GRN of *E. histolytica* has 221 nodes and 272 interactions. The regulatory interactions conserved were preferentially assigned from *S. cerevisiae* (248 interactions), and also from the interactions from *H. sapiens* (24 interactions) (Fig 4A). The networks include 28 regulatory proteins: 18 were inferred by homology, and 10 were inferred based on InterPro and Pfam assignments. Of these 28, 4 TFs are self-regulated, *i.e.*, the TF regulates its own gene.

The GRN of *E. dispar* has 583 nodes and 979 interactions. The regulatory interactions conserved were preferentially assigned from *S. cerevisiae* (942 interactions), whereas 34 interactions were inferred from *H. sapiens* (Fig 4B). The networks include 41 regulatory proteins, 17 of which were inferred by homology and 24 were inferred by InterPro and Pfam assignments; from these 41, 7 TFs are self-regulated.

The GRN of *E. nuttalli* includes 382 nodes and 560 interactions. From these, 582 regulatory interactions were inferred from *S. cerevisiae* and 23 interactions from *H. sapiens* (Fig 4C). The network also includes 39 regulatory proteins, of which 6 are self-regulated.

Finally, the GRN of *E. invadens* has 520 nodes and 853 interactions. The regulatory interactions were preferentially assigned from *S. cerevisiae* (805 interactions) and also *H. sapiens* (50 interactions) (Fig 4D). The networks include 38 regulatory proteins, of which 7 TFs are self-regulated. Interestingly, the number of regulatory proteins in *E. histolytica* is lower and they appear to have fewer interactions and nodes compared to the regulatory proteins identified in

Table 1. General properties of the GRNs.

GRN	*E. histolytica*	*E. dispar*	*E. nuttalli*	*E. invadens*
Total number of nodes	221	583	382	520
Total number of interactions	272	979	560	853
Number of TFs	28	41	39	38
Number of TGs	205	542	521	482
Self-regulated*	4	7	6	7
Maximum out degree*	EHI_044890—Helicase (121)**	EDI_297980—Hypothetical protein (284)	ENU1_120230—HSF (195)	EIN_249270—NF-Y alpha (257)
Maximum in-degree*	EHI_131470—NOP10 (4)***	EDI_107330 –Xaa-Pro dipeptidase (7)	ENU1_214880 –Xaa-Pro dipeptidase (6)	EIN_095830—branched-chain-amino-acid aminotransferase (9)
Connected components	4	5	3	5
Average clustering	0.0033	0.0723	0.043	0.0786

*This information can be found in S4 Table.

**The values in parentheses correspond to its Kin.

***The values in parentheses correspond to its Kout.

https://doi.org/10.1371/journal.pone.0271640.t001

E. dispar, *E. nuttalli*, and *E. invadens*. For example, *E. dispar* has 13 more regulatory proteins than *E. histolytica*, but the number of nodes is more than double (521) and the number of interactions is three times higher than in *E. histolytica* (Table 1).

Topological properties of the GRNs

In order to describe the global and local structures of the GRNs of *Entamoeba* spp. strains, the general structures of the four networks were analyzed (Table 1). Networks are structured into connected components (CC), within which the giant component is the one that contains the largest number of nodes in a network. In this context, we identified that the *E. histolytica* network comprises four CCs and the giant component contains 210 nodes and 263 edges; the *E. dispar* network contains five CCs and the giant component has 553 nodes and 945 edges; *E. nuttalli* contains three CCs and the giant component has 379 nodes and 558 edges. Finally, *E. invadens* contains five CCs and one giant component with 510 nodes and 846 edges (Fig 4).

In addition, the clustering coefficient, a measure of the degree to which nodes in a graph tend to cluster together, was calculated. We observed that the maximum clustering coefficient was 1 in *E. invadens* and 0.5 in the other networks. A clustering coefficient of 1 indicates that nodes with neighbors that are related between them form complete graphs, while a clustering coefficient less than 1 is related to few nodes being connected, which is common in the four networks due to limited information.

The input degree (*Kin*) and output degree (*Kout*), which link the number of TFs that control a gene and the number of genes that a TF regulates, were computed. In this context, we identified that 155 in *E. histolytica*, 295 in *E. dispar*, 234 in *E. nuttalli*, and 290 in *E. invadens* are regulated by one TF, *i.e.*, they have an input degree of 1. In this context, the most regulated gene in *E. histolytica* is EHI_131470 (ribosome biogenesis protein Nop10) responsible for ribosome biogenesis and is regulated by four TFs. EDI_107330 (Xaa-Pro dipeptidase) is regulated by seven TFs in *E. dispar*; in *E. nuttalli* ENU1_214880 (Xaa-Pro dipeptidase) is regulated by six TFs. Finally, in *E. invadens* EIN_095830 (branched-chain-amino-acid aminotransferase) is regulated by nine TFs (Table 1).
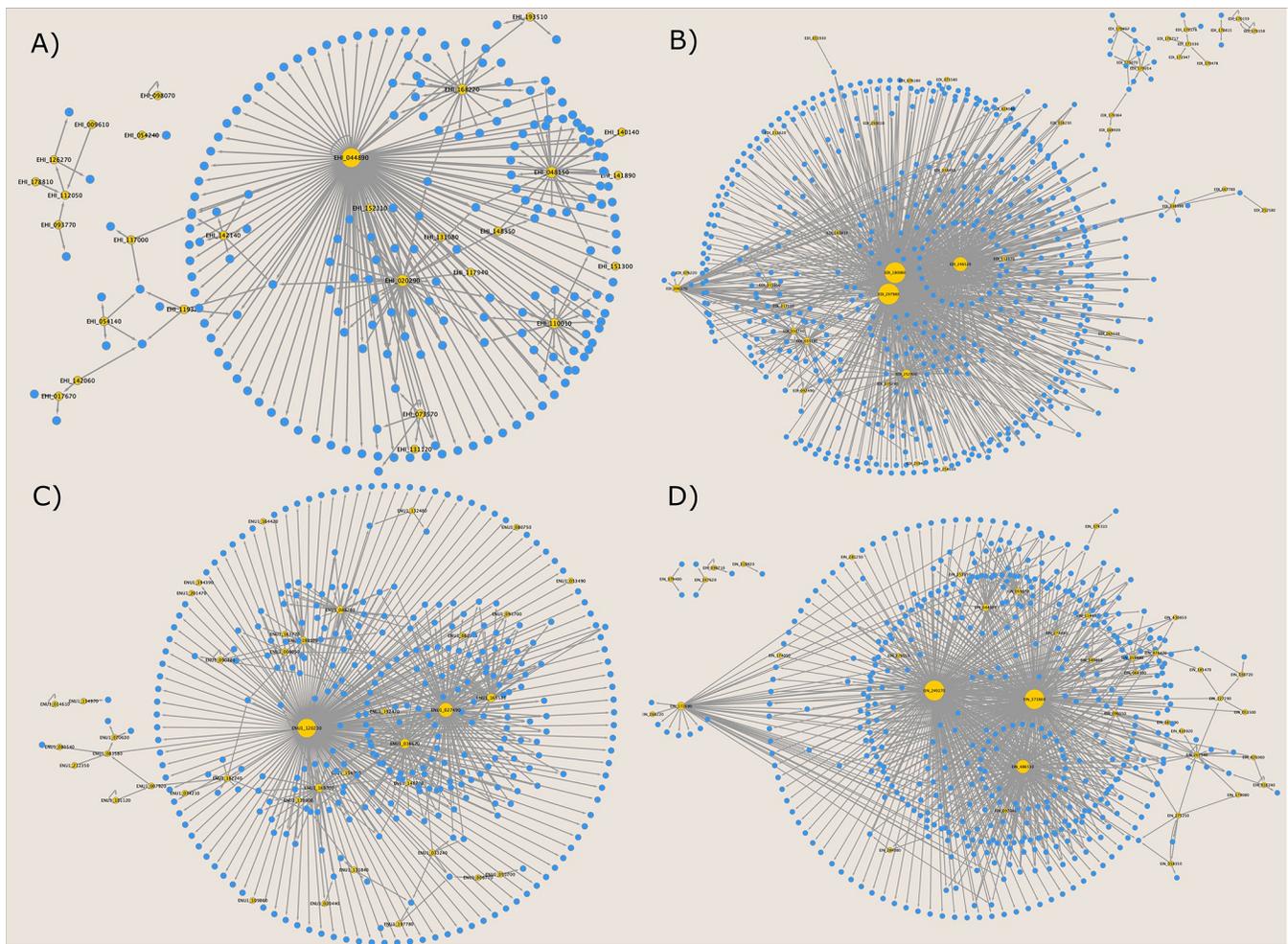
Fig 4. GRNs of A) *E. histolytica*, B) *E. dispar*, C) *E. nuttalli*, and D) *E. invadens*. The yellow nodes are TFs and the blue nodes are TGs; size nodes are proportional to output degree.

https://doi.org/10.1371/journal.pone.0271640.g004

With regard to output degree, the most connected node in the *E. histolytica* network is the putative Helicase EHI_044890, which influences 121 genes. It is interesting that the ortholo-gous proteins of EHI_044890 in other organisms are essential, according to the TDR targets database (https://www.tdrtargets.org). For instance, in *Trypanosoma brucei*, its mutation reduces significant loss of fitness in differentiation of procyclic to bloodstream forms, whereas in *Caenorhabditis elegans*, it is lethal for the embryonic stage. Therefore, we suggest that EHI_044890 is also essential in *E. histolytica*, because of its output degree, functional role, and similarity to other proteins.

EDI_297980 regulates 284 genes and is a hypothetical protein in *E. dispar* network. ENU1_120230 regulates 195 genes and is a putative heat shock transcription factor in *E. nut-talli*. This type of TFs has not been identified in any *Entamoeba* species, except in *E. histolytica*, which has a family of seven EhHSTFs (manuscript in preparation). EhHSTF7 has recently been shown to be the TF responsible for regulating the expression of the multidrug resistance gene EhPgp5 in this species of amoeba [57].

Finally, EIN_249270 regulates 257 genes and is a putative transcription factor NF-Y alpha in *E. invadens*. This TF has been a heterotrimeric protein composed of NF-YA, NF-YB, and NF-YC subunits that bind to the CCAAT box. This factor participates mainly in the regulation

of genes of the cell cycle and metabolism such as gluconeogenesis and appears at a later time point of *Entamoeba* encystation [55–57,63]. In addition, we identified the top 10 most important nodes by the centrality metrics, based on node connectivity as well as the shortest paths between them. In terms of degree centrality, the most important nodes in each network included EHI_044890 (0.5545), which is an isw2p helicase in *E. histolytica* [64]; its homologues in *S. cerevisiae* are chromatin-remodeling factors and yeast ISWI, which is essential for the cell to resist various stresses in vivo, and both homologous show genetic interactions [65]. EDI_297980 (0.4914) is a hypothetical protein in *E. dispar*, and it is homologous to nuclear transcription factor Y, alpha (KEGG). In *E. nuttalli*, ENU1_120230 (0.5118) is orthologous with YGL073W, which is a trimeric heat shock TF [66]. Finally, EIN_249270 (0.4971), which is orthologous to nuclear transcription factor Y, alpha (KEGG), is the most important in *E. invadens*.

Furthermore, we identified the node with the highest closeness score, *i.e.*, that which minimizes the sum of distances to the other nodes. In *E. histolytica* the most important is EHI_131470 (0.0189), for ribose biogenesis protein Nop10, involved in 18S rRNA pseudouridylation and in cleavage of pre-rRNA [67]. In *E. dispar*, EDI_107330 (0.0126) is a putative Xaa-Pro dipeptidase; Xaa-Pro dipeptidase plays a role in collagen metabolism because of the high level of imino acids in collagen (Uniprot). In *E. nuttalli*, ENU1_214880 (0.0167) is also homologous with Xaa-Pro dipeptidase. In *E. invadens* EIN_095830 (0.0166) is a putative branched-chain amino acid aminotransferase.

Betweenness centrality of a node is defined as the sum of the fraction of all-pairs shortest paths that pass through *v*, *i.e.*, the influence of a vertex over the flow of information between every pair of vertices under the assumption that information primarily flows over the shortest paths between them. The most important in *E. histolytica* is EHI_048150 (0.0011519), which encodes the EhCAF1 protein homologous to POP2 in *S. cerevisiae*; POP2 is a nuclease involved in mRNA deadenylation [68,69]. EDI_297980 (0.000999) in *E. dispar* and EIN_249270 (0.0011270) in *E. invadens* are orthologs to nuclear transcription factor Y, alpha. ENU1_027490 (0.00050421) in *E. nuttalli* is homologous to ISW2, a conserved ATP-dependent chromatin-remodeling factor in *S. cerevisiae* [70].

Biological process in the networks

To identify the most abundant functions represented in the networks, they were analyzed with Gene Ontology terms. We identified that the most abundant terms in the four networks are Macromolecule metabolic process (GO:0043170), Cellular macromolecule metabolic process (GO:0044260), and Cellular nitrogen compound metabolic process (GO:0034641), all of which are related to chemical reactions and pathways involving macromolecules and organic and inorganic nitrogenous compounds (Fig 5) (S3 Table).

Additionally, we identified some GO terms associated with one species' network: single-organism biosynthetic process (GO:0044711), organophosphate metabolic process (GO:0019637), and macromolecular complex subunit organization (GO:0043933) in *E. histolytica*; single-organism carbohydrate metabolic process (GO:0005975) and organic substance catabolic process (GO:1901575) in *E. dispar*; generation of precursor metabolites and energy (GO:0006091) in *E. nuttalli*; regulation of protein complex assembly (GO:0043254), regulation of actin filament-based process (GO:0032970), and generation of precursor metabolites and energy (GO:0006091) in *E. invadens*.

Conclusions

The inference of GRNs of *Entamoeba* speciesprovide an excellent opportunity to understand how genes and functional processes are interrelated in these organisms. These networks were
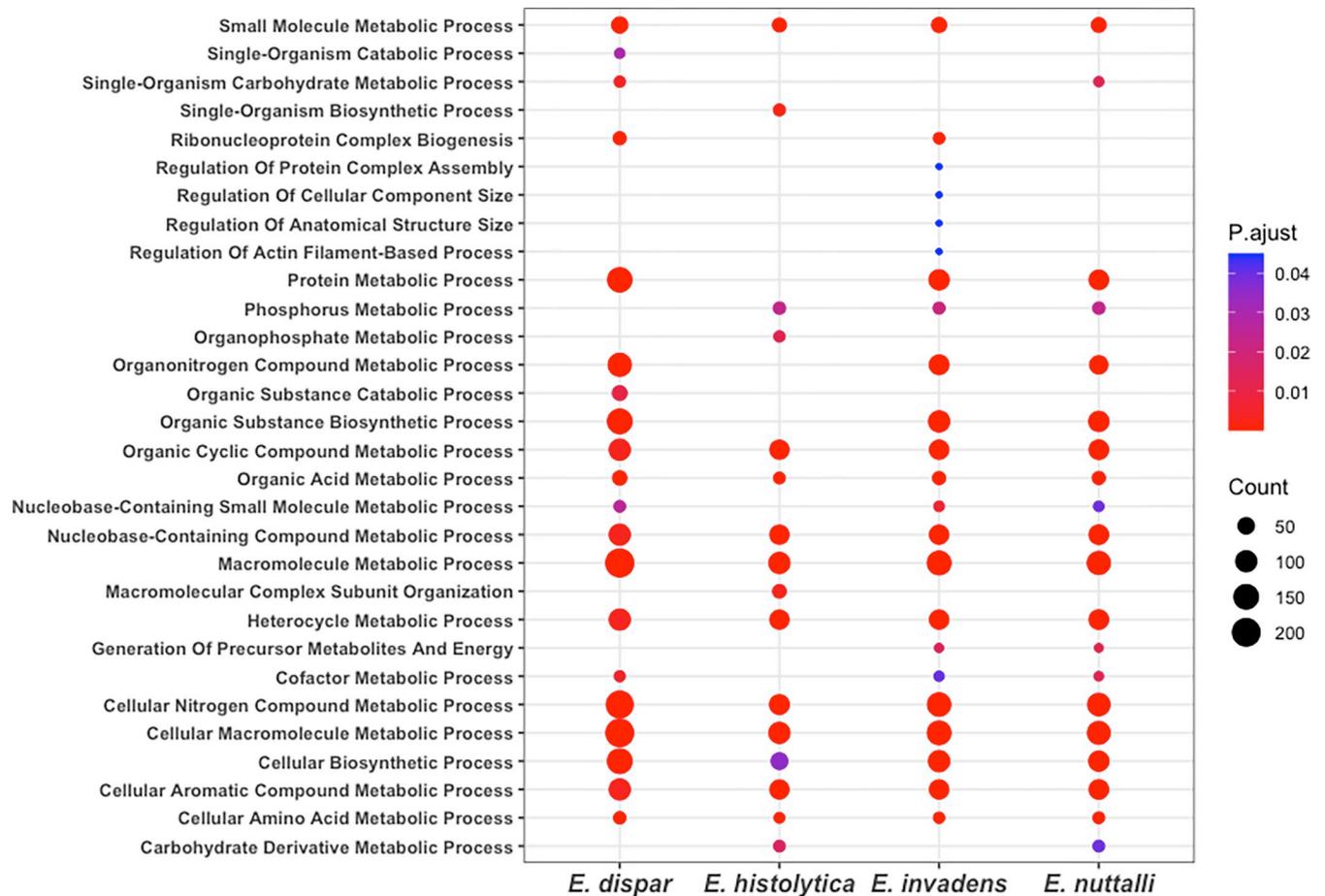
Fig 5. GO enrichment analysis. The dot plot shows the terms (FDR < 0.05) of biological processes identified using DAVID. The size of a dot represents the number of genes associated with the GO term, and the color of dots represents the P-adjusted value.

analyzed in terms of these topological properties to infer the role of TFs in the context of the GRN and the biological functions. From these analyses, we identified that TFs of *E. histolytica*, *E. dispar*, and *E. nuttalli* are associated with the LIM family, whereas the TFs in *E. invadens* are associated with the RRM_1 family. In the context of more connected nodes, we identified that EHI_044890 regulates 121 genes in *E. histolytica*, EDI_297980 regulates 284 genes in *E. dispar*, ENU1_120230 regulates 195 genes in *E. nuttalli*, and EIN_249270 regulates 257 genes in *E. invadens*. Finally, we determined that Macromolecule metabolic process (GO:0043170), Cellular macromolecule metabolic process (GO:0044260), and Cellular nitrogen compound metabolic process (GO:0034641) are the main biological processes for each network. However, there are specific enriched biological processes for each network that determine the differences in the size of each network. The results described in this work can be used for the study of gene regulation in these organisms.

Supporting information

S1 Table. Transcription factors of *Entamoebas* strains.
(XLSX)

S2 Table. Gene regulatory networks of *Entamoebas* strains.
(XLSX)

S3 Table. Gene ontology terms of *Entamoebas* networks.
(XLSX)

S4 Table. Main topological measures.
(XLSX)

S1 Data. Script to build a network from a template.
(ZIP)

Acknowledgments

Author Contributions

Conceptualization: Edgardo Galán-Vásquez, Ernesto Pérez-Rueda.

Data curation: Edgardo Galán-Vásquez.

Formal analysis: Edgardo Galán-Vásquez, Ernesto Pérez-Rueda.

Funding acquisition: Edgardo Galán-Vásquez, Ernesto Pérez-Rueda.

Investigation: María del Consuelo Gómez-García, Ernesto Pérez-Rueda.

Methodology: Edgardo Galán-Vásquez, María del Consuelo Gómez-García, Ernesto Pérez-Rueda.

Supervision: Ernesto Pérez-Rueda.

Validation: María del Consuelo Gómez-García.

Writing – original draft: Edgardo Galán-Vásquez, María del Consuelo Gómez-García, Ernesto Pérez-Rueda.

Writing – review & editing: Edgardo Galán-Vásquez, María del Consuelo Gómez-García, Ernesto Pérez-Rueda.

References

1. Greber BJ, Nogales E. The Structures of Eukaryotic Transcription Pre-initiation Complexes and Their Functional Implications. In: Harris J., Marles-Wright J. (eds) Macromolecular Protein Complexes II: Structure and Function. Subcellular Biochemistry, vol 93. Springer, Cham. 2019. https://doi.org/10.1007/978-3-030-28151-9_5.

2. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The Human Transcription Factors. Cell. 2018; 172, 4, 650–665. https://doi.org/10.1016/j.cell.2018.01.029.

3. Filtz TM, Vogel WK, Leid M. Regulation of transcription factor activity by interconnected post-translational modifications. Trends Pharmacol Sci. 2014; 35:76–85. https://doi.org/10.1016/j.tips.2013.11.005 PMID: 24388790

4. Sönmezer C, Kleinendorst R, Imanci D, Barzaghi G, Villacorta L., Schübeler D,et al. Molecular Co-occupancy Identifies Transcription Factor Binding Cooperativity In Vivo. Mol Cell. 2021; 21; 81(2):255–267. https://doi.org/10.1016/j.molcel.2020.11.015 PMID: 33290745

5. Espinosa A, Paz-y-Miño CG. Discrimination experiments in Entamoeba and evidence from other protists suggest pathogenic amebas cooperate with kin to colonize hosts and deter rivals. Journal of Eukaryotic Microbiology. 2019; 66(2), 354–368. https://doi.org/10.1111/jeu.12673 PMID: 30055104

6.   Mahmood SAF, Bakr HM. Molecular identification and prevalence of Entamoeba histolytica, Entamoeba dispar and Entamoeba moshkovskii in Erbil City, northern Iraq. Polish journal of microbiology. 2020; 69(3), 263.

7.   Babuta M, Bhattacharya S, Bhattacharya A. Entamoeba histolytica and pathogenesis: A calcium connection. PLoS Pathogens. 2020; 16(5), e1008214. https://doi.org/10.1371/journal.ppat.1008214 PMID: 32379809

8.   König C, Honecker B, Wilson IW, Weedall GD, Hall N, Roeder T, et al. Taxon-Specific Proteins of the Pathogenic Entamoeba Species E. histolytica and E. nuttalli. Front. Cell. Infect. Microbiol. 2021; 11:641472. https://doi.org/10.3389/fcimb.2021.641472 PMID: 33816346

9.   Ralston KS, Petri WA. The ways of a killer: how does *Entamoeba histolytica* elicit host cell death? Essays Biochem. 2011; 51:193–210. https://doi.org/10.1042/bse0510193 PMID: 22023450.

10.   Feng M, Cai J, Min X, Fu Y, Xu Q, Tachibana H, et al. Prevalence and genetic diversity of *Entamoeba* species infecting macaques in southwest China. Parasitol Res. 2013; 112: 1529–1536. https://doi.org/10.1007/s00436-013-3299-1 PMID: 23354942

11.   Tachibana H, Yanagi T, Pandey K, Cheng XJ, Kobayashi S, Sherchand JB, et al. An *Entamoeba* sp. strain isolated from rhesus monkey is virulent but genetically different from *Entamoeba histolytica*. Molecular and biochemical parasitology. 2007; 153(2), 107–114. https://doi.org/10.1016/j.molbiopara.2007.02.006 PMID: 17403547

12.   Tachibana H, Yanagi T, Lama C, Pandey K, Feng M, Kobayashi S, et al. Prevalence of *Entamoeba nuttalli* infection in wild rhesus macaques in Nepal and characterization of the parasite isolates. Parasitology International. 2013; 62(2), 230–235. https://doi.org/10.1016/j.parint.2013.01.004 PMID: 23370534

13.   Segovia-Gamboa NC, Chávez-Munguía B, Medina-Flores Y, Cázares-Raga FE, Hernández-Ramírez VI, Martínez-Palomo A, et al. *Entamoeba invadens*, encystation process and enolase. Exp Parasitol. 2010; 125(2):63–9. https://doi.org/10.1016/j.exppara.2009.12.019 PMID: 20045689

14.   Diamond LS, Clark CG. A redescription of *Entamoeba histolytica* 447 Schaudinn, 1903 (emended Walker, 1911) separating it from Entamoeba dispar Brumpt, 1925. J. Eukaryot. Microbiol. 1993; 40, 340–344. https://doi.org/10.1111/j.1550-7408.1993.tb04926.x PMID: 8508172

15.   Oliveira FMS, Neumann E, Gomes MA, Caliari MV. Entamoeba dispar: could it be pathogenic. Tropical Parasitology. 2015; 5(1), 9. https://doi.org/10.4103/2229-5070.149887 PMID: 25709947

16.   Wilson IW, Weedall GD, Lorenzi H, Howcroft T, Hon CC, Deloger M, et al. Genetic diversity and gene family expansions in members of the genus *Entamoeba*. Genome Biol.Evol. 2019; 11(3): 688–705. https://doi.org/10.1093/gbe/evz009 PMID: 30668670

17.   Tovar J, Fischer A, Clark CG. The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite *Entamoeba histolytica*. Mol Microbiol. 1999; 32(5):1013–21. https://doi.org/10.1046/j.1365-2958.1999.01414.x PMID: 10361303.

18.   Siegesmund MA, Hehl AB, van der Giezen M. Mitosomes in trophozoites and cysts of the reptilian parasite *Entamoeba invadens*. Eukaryot Cell. 2011; 10(11):1582–5. https://doi.org/10.1128/EC.05172-11 PMID: 21965513

19.   Tanaka M, Makiuchi T, Komiyama T, Shiina T, Osaki K, Tachibana H. Whole genome sequencing of *Entamoeba nuttalli* reveals mammalian host-related molecular signatures and a novel octapeptide-repeat surface protein. PLoS Negl Trop Dis. 2019; 13(12): e0007923. https://doi.org/10.1371/journal.pntd.0007923 PMID: 31805050

20.   Lorenzi HA, Puiu D, Miller JR, Brinkac LM, Amedeo P, Hall N, et al. New assembly, reannotation and analysis of the *Entamoeba histolytica* genome reveal new genomic features and protein content information. PLoS Negl Trop Dis. 2010; 4(6):e716. https://doi.org/10.1371/journal.pntd.0000716 PMID: 20559563

21.   Tachibana, Yanagi T, Feng M, Bandara KBA, Kobayashi S, Cheng X, et al. Isolation and molecular characterization of *Entamoeba nuttalli* strains showing novel isoenzyme patterns from wild Toque Macaques in Sri Lanka. J. 494 Eukaryot. Microbiol. 2016; 6: 171–180. https://doi.org/10.1111/jeu.12265 PMID: 26333681

22.   Ehrenkaufer GM, Weedall GD, Williams D, Lorenzi HA, Caler E, Hall N, et al. The genome and transcriptome of the enteric parasite *Entamoeba invadens*, a model for encystation. Genome Biol. 2013; 14(7):R77. https://doi.org/10.1186/gb-2013-14-7-r77 PMID: 23889909

23.   Karlebach G, Shamir R. Modelling and analysis of gene regulatory networks. Nat. Rev. Mol. Cell Biol. 2008; 9: 770–780. https://doi.org/10.1038/nrm2503 PMID: 18797474

24.   Jackson CA, Castro DM, Saldi GA, Bonneau R, Gresham D. Gene regulatory network reconstruction using single-cell RNA sequencing of barcoded genotypes in diverse environments. eLife. 2020; 9: e51254. https://doi.org/10.7554/eLife.51254 PMID: 31985403

25. Gerstein MB, Kundaje A, Hariharan M, Landt SG. Yan K-K, Cheng C, et al. Architecture of the human regulatory network derived from ENCODE data. Nature 2012; 489: 91–100. https://doi.org/10.1038/nature11245 PMID: 22955619

26. Chen D, Yan W, Fu L-Y, Kaufmann K. Architecture of gene regulatory networks controlling flower development in *Arabidopsis thaliana*. Nat. Commun. 2018; 9:4534. https://doi.org/10.1038/s41467-018-06772-3 PMID: 30382087

27. Hu Y, Qin Y, Liu G. Collection and curation of transcriptional regulatory interactions in *Aspergillus nidulans* and *Neurospora crassa* reveal structural and evolutionary features of the regulatory networks. Front. Microbiol. 2018; 9:27. https://doi.org/10.3389/fmicb.2018.00027 PMID: 29403467

28. Galán-Vásquez E, Sánchez-Osorio I, Martínez-Antonio A. Transcription factors exhibit differential conservation in bacteria with reduced genomes. PLoS One, 2016; 11(1): e0146901. https://doi.org/10.1371/journal.pone.0146901 PMID: 26766575

29. Lenz AR, Galán-Vásquez E, Balbinot E, de Abreu FP, Souza de Oliveira N, da Rosa LO, et al. Gene regulatory networks of *Penicillium echinulatum* 2HH and *Penicillium oxalicum* 114–2 inferred by a computational biology approach. Frontiers in microbiology. 2020; 11: 2566. https://doi.org/10.3389/fmicb.2020.588263 PMID: 33193246

30. Mercatelli D, Scalambra L, Triboli L, Ray F, Giorgi FM. Gene regulatory network inference resources: A practical overview. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms. 2020; 1863 (6): 194430. https://doi.org/10.1016/j.bbagrm.2019.194430 PMID: 31678629

31. Soberanes-Gutiérrez CV, Pérez-Rueda E, Ruíz-Herrera J, Galán-Vásquez E. Identifying Genes Devoted to the Cell Death Process in the Gene Regulatory Network of *Ustilago maydis*. Frontiers in Microbiology. 2021; 12: 1321.

32. Xu L, Dong Z, Fang L, Luo Y, Wei Z, Guo H, et al. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. Nucleic acids research. 2019; 47 (W1): W52–W58. https://doi.org/10.1093/nar/gkz333 PMID: 31053848

33. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014; 30(9): 1236–1240. https://doi.org/10.1093/bioinformatics/btu031 PMID: 24451626

34. Monteiro PT, Oliveira J, Pais P, Antunes M, Palma M, Cavalheiro M, et al. YEASTRACT+: a portal for cross-species comparative genomics of transcription regulation in yeasts. Nucleic acids research. 2020; 48(D1): D642–D649. https://doi.org/10.1093/nar/gkz859 PMID: 31586406

35. Han H, Cho JW, Lee S, Yun A, Kim H, Bae D, et al. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. Nucleic acids research. 2018; 46(D1): D380–D386. https://doi.org/10.1093/nar/gkx1013 PMID: 29087512

36. Lechner M, Findeiß S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co-) orthologs in large-scale analysis. BMC bioinformatics. 2011; 12(1): 1–9. https://doi.org/10.1186/1471-2105-12-124 PMID: 21526987

37. Barabási AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat. Rev. Genet. 2004; 5: 101–113. https://doi.org/10.1038/nrg1272 PMID: 14735121

38. Junker BH, Schreiber F. (eds). Analysis of Biological Networks. Hoboken, NJ: JohnWiley & Sons, Inc. 2011.

39. Hansen D, Shneiderman B, Smith MA. Analyzing social media networks with NodeXL: Insights from a connected world. Morgan Kaufmann: Burlington. 2010.

40. Jiao X, Sherman BT, Huang D, Stephens R, Baseler MW, Lane HC, et al. DAVID-WS: a stateful web service to facilitate gene/protein list analysis. Bioinformatics. 2012; 28: 1805–1806. https://doi.org/10.1093/bioinformatics/bts251 PMID: 22543366

41. Ojhaa S, Singha N, Bhattacharya A, Bhattacharya S. The ribosomal RNA transcription unit of *Entamoeba invadens*: Accumulation of unprocessed pre-rRNA and a long non coding RNA during encystation. Mol Biochem Parasitol. 2013; 192: 30–38. https://doi.org/10.1016/j.molbiopara.2013.10.002 PMID: 24200639

42. Levine M, Tjian R. Transcription regulation and animal diversity. Nature. 2003; 424(6945):147–51. https://doi.org/10.1038/nature01763 PMID: 12853946

43. Kadrmas J., Beckerle M. The LIM domain: from the cytoskeleton to the nucleus. Nat Rev Mol Cell Biol. 2004; 5: 920–931. https://doi.org/10.1038/nrm1499 PMID: 15520811

44. Wender N, Villalobo E, Mirelman D. EhLimA, a novel LIM protein, localizes to the plasma membrane in *Entamoeba histolytica*. Eukaryot Cell. 2007; 6(9):1646–1655. https://doi.org/10.1128/EC.00177-07 PMID: 17630327

45. Loftus B, Anderson I, Davies R, Alsmark UCM, Samuelson J, Amedeo P, et al. The genome of the protist parasite *Entamoeba histolytica*. Nature. 2005; 433(7028): 865–868. https://doi.org/10.1038/nature03291 PMID: 15729342

46. Maris C, Dominguez C, Allain FHT. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. The FEBS journal. 2005; 272(9): 2118–2131. https://doi.org/10.1111/j.1742-4658.2005.04653.x PMID: 15853797

47. Maruyama K, Sato N, Ohta N. Conservation of structure and cold-regulation of RNA-binding proteins in cyanobacteria: probable convergent evolution with eukaryotic glycine-rich RNA-binding proteins. Nucleic Acids Res. 1999; 27: 2029–2036. https://doi.org/10.1093/nar/27.9.2029 PMID: 10198437

48. Volpon L D'Orso I, Young CR, Frasch AC, Gehring K. NMR structural study of TcUBP1, a single RRM domain protein from Trypanosoma cruzi: contribution of a beta hairpin to RNA binding. Biochemistry. 2005; 44(10):3708–17. https://doi.org/10.1021/bi047450e PMID: 15751947.

49. Shan F, Zhang N. Resonance assignments of La protein RRM domain from Trypanosoma brucei. Biomol NMR Assign. 2021; 15(1):41–44. https://doi.org/10.1007/s12104-020-09980-5 PMID: 33089372

50. Eichinger D. Encystation of entamoeba parasites. Bioessays. 1997; 19(7): 633–639. https://doi.org/10.1002/bies.950190714 PMID: 9230696

51. Cui Z, Li J, Chen Y, Zhang L. Molecular epidemiology, evolution, and phylogeny of *Entamoeba* spp. Infection, Genetics and Evolution. 2019; 75: 104018. https://doi.org/10.1016/j.meegid.2019.104018 PMID: 31465857

52. Davis PH, Chen M, Zhang X, Clark CG, Townsend RR, Stanley SL Jr. Proteomic comparison of *Entamoeba histolytica* and *Entamoeba dispar* and the role of *E. histolytica* alcohol dehydrogenase 3 in virulence. PLoS Negl Trop Dis. 2009; 3(4):e415. https://doi.org/10.1371/journal.pntd.0000415 PMID: 19365541

53. MacFarlane RC, Singh U. Identification of differentially expressed genes in virulent and nonvirulent *Entamoeba* species: potential implications for amebic pathogenesis. Infection and immunity. 2006; 74(1), 340–351. https://doi.org/10.1128/IAI.74.1.340-351.2006 PMID: 16368989

54. Betanzos A, Zanatta D, Bañuelos C, Hérnandez-Nava E, Cuellar P, Orozco E. Epithelial Cells Expressing EhADH, An Entamoeba histolytica Adhesin, Exhibit Increased Tight Junction Proteins. Front Cell Infect Microbiol. 2018; 28;8:340. https://doi.org/10.3389/fcimb.2018.00340 PMID: 30324093

55. Manna D, Lentz CS, Ehrenkaufer GM, Suresh S, Bhat A, Singh U. An NAD+-dependent novel transcription factor controls stage conversion in entamoeba. Elife. 2018; 7:e37912. https://doi.org/10.7554/eLife.37912 PMID: 30375973

56. Manna D, Singh U. (2019). Nuclear Factor Y (NF-Y) Modulates Encystation in Entamoeba via Stage-Specific Expression of the NF-YB and NF-YC Subunits. mBio. 2019; 10(3), e00737–19. https://doi.org/10.1128/mBio.00737-19 PMID: 31213550

57. Bello F, Orozco E, Benítez-Cardoza CG, Zamorano-Carrillo A, Reyes-López CA, Pérez-Ishiwara DG, et al. The novel EhHSTF7 transcription factor displays an oligomer state and recognizes a heat shock element in the Entamoeba histolytica parasite. Microb Pathog 2022; 162;105349. https://doi.org/10.1016/j.micpath.2021.105349 PMID: 34864144

58. Mendoza L, Orozco E, Rodríguez MA, García-Rivera G, Sánchez T, García E, et al. Ehp53, an Entamoeba histolytica protein, ancestor of the mammalian tumour suppressor p53. Microbiology (Reading, England). 2003; 149(Pt 4), 885–893. https://doi.org/10.1099/mic.0.25892-0 PMID: 12686631

59. Gilchrist CA, Holm CF, Hughes MA, Schaenman JM, Mann BJ, Petri WA Jr. Identification and characterization of an Entamoeba histolytica upstream regulatory element 3 sequence-specific DNA-binding protein containing EF-hand motifs. J. Biol. Chem. 2001; 276, pp. 11838–11843. https://doi.org/10.1074/jbc.M007375200 PMID: 11278344

60. Schaenman JM, Gilchrist CA, Mann BJ, Petri WA Jr. Identification of two Entamoeba histolytica sequence-specific URE4 enhancer-binding proteins with homology to the RNA-binding motif RRM. J. Biol. Chem. 2001; 276, pp. 1602–1609. https://doi.org/10.1074/jbc.M006866200 PMID: 11038357

61. Cárdenas-Hernández H, Titaux-Delgado GA, Castañeda-Ortiz EJ, Torres-Larios A, Brieba LG, del Río-Portilla F, et al. Genome-wide and structural analysis of the Myb-SHAQKYF family in Entamoeba histolytica. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics. 2021; 1869(4), 140601. https://doi.org/10.1016/j.bbapap.2021.140601 PMID: 33422669

62. Yu H, Luscombe NM, Lu HX, Zhu X, Xia Y, Han JD, et al. Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. Genome Res. 2004; 14: 1107–1118. https://doi.org/10.1101/gr.1774904 PMID: 15173116

63. Gurtner A, Manni I, Piaggio G. NF-Y in cancer: Impact on cell transformation of a gene essential for proliferation. Biochim Biophys Acta Gene Regul Mech. 2017; 1860(5);604–616. https://doi.org/10.1016/j.bbagrm.2016.12.005 PMID: 27939755

64. Valdés J, Nozaki T, Sato E, Chiba Y, Nakada-Tsukui K, Villegas-Sepúlveda N, et al. Proteomic analysis of *Entamoeba histolytica* in vivo assembled pre-mRNA splicing complexes. Journal of proteomics. 2014; 111: 30–45. https://doi.org/10.1016/j.jprot.2014.07.027 PMID: 25109466

65. Tsukiyama T, Palmer J, Landel CC, Shiloach J, Wu C. Characterization of the imitation switch subfamily of ATP-dependent chromatin-remodeling factors in *Saccharomyces cerevisiae*. Genes & development. 1999; 13(6): 686–697. https://doi.org/10.1101/gad.13.6.686 PMID: 10090725

66. Sorger PK, Pelham HR. Purification and characterization of a heat-shock element binding protein from yeast. The EMBO journal. 1987; 6(10): 3035–3041. https://doi.org/10.1002/j.1460-2075.1987.tb02609.x PMID: 3319580

67. McMahon M, Contreras A, Ruggero D. Small RNAs with big implications: new insights into H/ACA snoRNA function and their role in human disease. Wiley Interdisciplinary Reviews: RNA. 2015; 6(2): 173–189. https://doi.org/10.1002/wrna.1266 PMID: 25363811

68. Daugeron MC, Mauxion F, Séraphin B. The yeast POP2 gene encodes a nuclease involved in mRNA deadenylation. Nucleic acids research. 2001; 29(12): 2448–2455. https://doi.org/10.1093/nar/29.12.2448 PMID: 11410650

69. Ye X, Axhemi A, Jankowsky E. Alternative RNA degradation pathways by the exonuclease Pop2p from *Saccharomyces cerevisiae*. RNA. 2021; 27(4): 465–476. https://doi.org/10.1261/rna.078006.120 PMID: 33408095

70. Yadon AN, Tsukiyama T. DNA looping-dependent targeting of a chromatin remodeling factor. Cell Cycle. 2013; 12(12):1809–1810. https://doi.org/10.4161/cc.25114 PMID: 23708514