

Review

Not peer-reviewed version

---

# Genome-Wide Classification of Myb Domain-Containing Protein Families in *Entamoeba invadens*

---

[Patricia Cuellar](#) , Elizabeth J. Castañeda-Ortiz , César Rosales-Zarza , Carlos E Martínez-Rodríguez , Israel Canela-Pérez , [Marío Alberto Rodríguez](#) , [Jesús Valdés](#) , [Elisa Azuara-Liceaga](#) \*

Posted Date: 4 January 2024

doi: 10.20944/preprints202401.0289.v1

Keywords: MYB-DBD containing proteins; encystation-excystation; transcriptional regulation; protozoan; Myb Recognition Element



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

# Genome-Wide Classification of Myb Domain-Containing Protein Families in *Entamoeba invadens*

Patricia Cuellar <sup>1</sup>, Elizabeth J. Castañeda-Ortiz <sup>1</sup>, César Rosales Zarza <sup>2</sup>, Carlos E Martínez-Rodríguez <sup>3</sup>, Israel Canela-Pérez <sup>4</sup>, Mario Alberto Rodriguez <sup>5</sup>, Jesús Valdés <sup>4</sup> and Elisa Azuara Liceaga <sup>1,\*</sup>

<sup>1</sup> Posgrado en Ciencias Genómicas, Universidad Autónoma de la Ciudad de México, San Lorenzo 290, Colonia del Valle Sur, Benito Juárez, C.P. 03100. Mexico City, Mexico e-mail: elisa.azuara@uacm.edu.mx

<sup>2</sup> Licenciatura Ciencias Genómicas, Universidad Autónoma de la Ciudad de México, San Lorenzo 290, Colonia del Valle Sur, Benito Juárez, C.P. 03100. Mexico City, Mexico. e-mail: adrian.rosales@estudiante.uacm.edu.mx

<sup>3</sup> Academia de Matemáticas, Universidad Autónoma de la Ciudad de México, Prol. San Isidro 151, San Lorenzo Tezonco, Iztapalapa, C.P. 09790. Mexico City, Mexico. e-mail: carlos.martinez@uacm.edu.mx

<sup>4</sup> Departamento de Bioquímica CINVESTAV-IPN, Av. Instituto Politécnico Nacional 2508, C.P. 07360, Mexico City, Mexico. e-mail: jvaldes@cinvestav.mx

<sup>5</sup> Departamento de Infectómica y Patogénesis Molecular CINVESTAV-IPN, Av. Instituto Politécnico Nacional 2508, C.P. 07360, Mexico City, Mexico. e-mail: marodri@cinvestav.mx

\* Correspondence: elisa.azuara@uacm.edu.mx; Tel.: (52)55-5850-1901

**Abstract:** *Entamoeba histolytica*, the causative agent of amebiasis, is the third leading cause of death among parasitic diseases globally. Its life cycle includes encystation, which has been mostly studied in *Entamoeba invadens*, responsible of reptilian amebiasis. However, the molecular mechanisms underlying this process are not fully understood. Therefore, we focused on the identification and characterization of Myb proteins, that regulate the expression of encystation-related genes in various protozoan parasites. Through bioinformatic analysis, we identified 47 genes in *E. invadens* encoding MYB-domain-containing proteins. These were classified into single-repeat 1R (19), 2R-MYB proteins (27), and one 4R-MYB protein. The *in silico* analysis suggests that these proteins are multifunctional, participating in transcriptional regulation, chromatin remodeling, telomere maintenance, and splicing. Transcriptomic data analysis revealed expression signatures of *eimyb* genes, suggesting a potential orchestration in the regulation of early and late encystation-excystation genes. Furthermore, we identified 3135 probable target genes associated with reproduction, meiotic cell cycle, ubiquitin-dependent protein catabolism, and endosomal transport. In conclusion, our findings suggest that *E. invadens* Myb proteins regulate stage-specific proteins and a wide array of cellular processes. This study provides a foundation for further exploration of the molecular mechanisms governing encystation and unveils potential targets for therapeutic intervention in amebiasis.

**Keywords:** MYB-DBD containing proteins; encystation-excystation; transcriptional regulation; protozoan; Myb recognition element

## 1. Introduction

*Entamoeba histolytica*, a unicellular protozoan that causes dysentery as the primary symptom of colonic invasion, is one of the most common parasitic causes of death worldwide [1]. This organism has two distinct life stages: an invasive trophozoite form and a latent cyst that is resistant to environmental changes. The process of infection starts when a person consumes contaminated food or water, then the cysts excyst in the small intestine and release the motile trophozoite [2]. Gene regulation is critical for environmental adaptation as well as for cyst conversion and pathogen transmission. Encystation-excystation pathways have been attempted in this pathogen, however, *Entamoeba invadens* is still the model system for studying *in vitro* encystment development [3–5]. The

genome of *E. invadens* is 40.88 MB long and is therefore the largest among the Entamoeba species [6–8]. This genome codifies 11,549 transcripts and regulates transcription through an EiCPM-GL motif (*E. invadens* core promoter motif-GAAC-like) localized 30 nt upstream from the start codon. This element resembles a fusion of the GAAC-like and Inr elements of *E. histolytica* [9]. Remarkably, no TATA box has been found, most likely due to the AT-rich nature of the genome, which makes bioinformatic searches challenging; however, a TATA-binding protein (TBP) has been identified [6,10]. Additionally, a novel transcription factor ERM-BP (Encystation Regulatory Motif- Binding Protein), a Nuclear factor Y (NF-Y) and recently a EiHbox1 have been described as transcription factors involved in encystment of these parasites [11,12]. Understanding the molecular mechanisms of gene expression regulation is crucial to characterize differentiation from trophozoites to cysts. Transcriptome analysis during encystation through RNAseq showed that almost 50% of all *E. invadens* genes modify their expression during this differentiation process. Besides phospholipase D, Rab, BspA, phosphatases, and cyst-wall formation-related genes that overexpress during encystation, it was also observed that genes coding for MYB-DBD-containing proteins are present in this protozoan and have differential expression patterns [6]. Forty-four genes encoding for MYB-DBD-containing proteins have been previously identified in this parasite, nine containing a SHAQKYF motif, and 23 annotated as Myb putative or hypothetical [10,13], without any further characterization of these proteins. Changes in the expression of these proteins have been documented during cyst formation, indicating that distinct gene expression is regulated by a particular gene set at different stages of encystation. However, little is known about these transcription factors in *E. invadens*. MYB-DBD-containing proteins, from hereafter referred to as Myb proteins, have a domain related to the MYB-DBD of human c-Myb. These proteins have been described as transcription factors, coactivators, telomere-binding proteins, ribosomal binding proteins, or splicing factors [14–18]. The MYB-DBD is approximately 52 amino acid residues in length and forms a helix-turn-helix conformation with three regularly spaced tryptophan or aromatic residues with up to four imperfect conserved repeats (R) in tandem, which form a hydrophobic core [19]. Four major subfamilies of Myb proteins—1R-MYB/Myb-related (1R), 2R-MYB (2R), 3R-MYB (3R), and 4R-MYB (4R)—are distinguished based on the number of MYB repeats in such proteins [17]. Recently, Myb proteins have been studied in unicellular organisms such as *Dictyostelium discoideum* [20,21], *Euplotes aediculatus* [22], *Trichomonas foetus* [23], *Trypanosoma brucei* and *Leishmania amazonensis* [24,25], *Plasmodium falciparum* [26], *Babesia bovis* [27], and *E. histolytica* in which a wide genome analysis has been conducted [28]. In *Trichomonas vaginalis* three different Myb proteins (TvMyb1, TvMyb2 and TvMyb3) regulate the expression of the adhesion protein AP65 (Ong et al., 2006; Hsu et al., 2009). Myb transcription factors are also important regulators of cell differentiation for example Myb2 in *Giardia lamblia* regulates the expression of cyst wall proteins [31–33], and BFD1 controls Bradyzoite differentiation in *Toxoplasma gondii* [34] as does EhMybdr in *E. histolytica* [35]. To understand the importance of Myb proteins in *E. invadens* differentiation, we performed a genomic survey of these transcription factors using c-Myb and EhMyb10 (a 2R-MYB protein of *E. histolytica*) as queries. In this study, we showed that *E. invadens* has 47 MYB-DBD-containing proteins instead of 44, as initially stated. These genes modulate their expression during encystation processes, and the proteins encoded have 1, 2, and 4 imperfect conserved repeats (R) in their MYB-DNA binding domain. Therefore, these proteins may play a crucial role as transcription modulators in *E. invadens*, enabling the invasion and formation of cysts in its reptilian host. Understanding the function and regulation of Myb proteins in *E. invadens* will allow the development of novel chemotherapeutics that could prevent cyst conversion and, consequently, disease transmission in their human counterparts.

## 2. Materials and Methods

### 2.1. Genomic data and identification of EiMyb-encoding proteins in *E. invadens*

Myb proteins were searched through a PSI-BLAST against the *E. invadens* IP1 genome (taxid: 33085) annotated in AmoebaDB (<https://amoebadb.org/>) [36,37] using human c-Myb (access number P10242 UniProt database) and EhMyb10 sequences (access number EHI\_129790 from AmoebaDB) as

queries. The EiMyb protein sequences were retrieved from AmoebaDB and used as queries for BLASTp searches until unique MYB-DBD-containing proteins were obtained.

## 2.2. EiMyb protein classification

The number of repeats in the MYB-DBD (1R, 2R, 3R, or 4R) was identified using InterProScan (<http://www.ebi.ac.uk/interpro/search/sequence-search>) and UniProt (<https://www.uniprot.org/>). Proteins with incomplete or distantly spaced repeats were discarded and not included in further analysis. Logos were obtained using WebLogo 3 (<https://weblogo.threeplusone.com/>) [38].

## 2.3. Multiple Sequence Alignment and Phylogenetic Analysis of EiMyb Proteins

MYB-DBD amino acid sequences were aligned using ClustalW and manually edited using Bioedit 7.0.5.3. Phylogenetic analysis was performed without the Gblocks tool using Phylogenyfr (<https://www.phylogeny.fr/>) [39].

## 2.4. Amino acid sequence analysis of EiMyb proteins

The molecular weight (MW) and isoelectric point (pI) were calculated using the ProtParam tool (<https://web.expasy.org/protparam/>). Protein transmembrane helices were predicted using the TMHMM server 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>). The nuclear localization signals (NLS) were determined using <http://www.moseslab.csb.utoronto.ca/NLStradamus/> and the PSORT program (<https://www.psort.org/>). Protein domain organization was performed using DOG 1.0 (<https://dog.biocuckoo.org/>) [40]. Protein structures were obtained from the alphafold.ebi.ac.uk data base database (last accessed: 12/05/2023), and the MYB-DBDs were visualized using the PyMOL program.

## 2.5. *eimyb* gene expression analysis in *E. invadens*

Expression patterns of *eimyb* genes were examined using the available *E. invadens* transcriptome data in AmoebaDB (<https://amoebadb.org/amoeba/app/search/transcript/GenesByRNASeqEvidence>). A heat map of *eimyb* genes and transcripts per million (TPM) distribution was obtained by hierarchical cluster analysis using the pheatmap package in R software (version 3.4.3.2). The colors in the graph indicate the magnitude of gene expression in the sample [ $\text{Log}_2(\text{TPM})$ ]. Boxplot was built using the ggplot2 package in R software (version 3.4.3.2).

## 2.6. Identification of Myb recognition elements (MRE) in the promoter regions of *E. invadens* genes

The region from -500 to +10 nucleotides relative to the transcription initiation site for each of the 12,007 ORFs of *E. invadens* was searched using the AmoebaDB DNA motif pattern tool. The presence of regular motifs using the Myb Recognition Element (MRE) [CT]AAC[GT]G and a C-rich sequence [CA]CCCC, previously detected in *E. histolytica* gene promoters [28,35] was analyzed using the Streme tool of the MEME Suite version 5.5.0 [41].

## 2.7. Analysis of enriched gene ontologies

Gene ontology analysis using AmoebaDB and REVIGO (<http://revigo.irb.hr/>) software [42]. The scatter plots were built using the ggplot2 package in R software (version 3.4.3.2).

# 3. Results and Discussion

## 3.1. Myb proteins in *E. invadens*

To identify all ORFs that encode MYB-DBD proteins in the genome of *E. invadens*, we performed a PSI-BLAST search using the amino acid sequence of the MYB-DBDs from the human c-Myb and *E. histolytica* EhMyb10. We identified 47 genes encoding EiMyb proteins in *E. invadens* genome;

therefore, this organism possesses more proteins than its close relative *E. histolytica*, which has 32 Myb proteins [28]. This could be because these transcription factors may regulate multiple vital functions to mediate reptilian invasion and cyst-trophozoite conversion. The 47 EiMyb proteins were retrieved from AmoebaDB and classified by the number of MYB-DBD repeats (R) using the InterPro and UniProt databases (Table 1). Forty-four of the identified EiMyb proteins match those reported by Ehrenkauf et al. (2013) [6] and de Cadiz et al. (2013) [13] in their RNAseq analysis. Furthermore, we identified three more eimyb genes that were not identified in previous studies, probably because of their divergence in the MYB-DBD region. Nineteen proteins were found with only one R1/R2 repeat (1R-MYB), and 27 proteins had two repeats (2R-MYB). Lastly, one 4R-MYB encoded protein was identified in *E. invadens* (Table 1) making it the first report of a four-repeat Myb protein in the *Entamoeba* genus.

**Table 1.** MYB DBD-containing proteins in *E. invadens* retrieved from AmoebaDB and classified according to their number of DBD-MYB repeats and motifs.

Myb subfamily	Groups	Number of members
1R-MYB	SHAQKYF (CCA1-like )	9
	Bdp1-like	4
	TRF-like	2
	Ada2-like (Transcriptional adapter putative-Ada2)	2
	Myb-related	1
	Zuotin-like	1
2R-MYB	Myb transcription factors	13
	Trichome differentiation protein GL1 related	6
	Werewolf transcription factors related	3
	Hypothetical proteins	3
	R2R3-Myb transcription factors	2
4R-MYB	Snap190	1
<b>Total</b>		<b>47</b>

The size of EiMyb proteins (aa) as well as computed parameters, including MW, pI, NLS, and subcellular localization are listed in Table S1. EiMyb proteins lengths ranged between 103 and 663 amino acids that weighed from 12.13 to 77.44 kDa with an average weight of 23.6 kDa. All proteins are defined with DNA binding function in GO terms (Molecular Function GO: 0003677). When determining the subcellular localization, we observed that most proteins are predicted nuclear. Our analysis revealed that 18 proteins have classical-type monopartite NLSs, accounting for 41% of proteins with 4–7 residues; 15 proteins have bipartite signals, comprising 17 amino acid residues (34.1%); and 11 proteins are NLS-free. Only two proteins have a transmembrane domain (Table S1), suggesting that these proteins must be embedded in the nuclear membrane to develop their function.

### 3.2. 1R-MYB subfamily in *E. invadens*

The 1R-MYB-MYB subfamily, also referred to as Myb-related proteins, is a highly heterogeneous subfamily with several roles as TFs, chromatin remodeling proteins, and telomeric repeat-binding

proteins [43–45]. 1R-MYBs usually contain other domains, reflecting their functional diversity. Of the 19 1R-MYB proteins in *E. invadens*, 17 were annotated as hypothetical proteins, and only two were annotated as putative transcriptional adapters (Table S1). The top result that showed a strong resemblance to a Blast search and the existence of the identified domains enabled us to name them EiMyb proteins (Table 2). The lengths of these 19 1R-MYB proteins ranged from 103 amino acids to 531 amino acids with an average of 239 amino acids (Table 2). Additionally, the pI varied, ranging from 6.35 to 10.13, indicating that their functions may be distinct from one another (Table S1). The amino acid sequence alignment shows that the MYB-DBD domain is highly divergent (Figure 1A). Generally, the MYB-DBD conserves the three-spaced tryptophan residues; however, in *E. invadens*, most 1R-MYB proteins have the first tryptophan conserved, and hydrophobic amino acids substitute the second and third tryptophans (Figure 1A). The MYB-DBD is located at the N-terminal and central regions of the proteins, except for three proteins in which is located in the C-terminal region (Figure 1C). Some of these proteins harbor the SANT domain (Swi3, Ada2, human N-CoR, and the transcription factor Bdp), and thus are MYB-related [46]. SANT domains are mainly found in plants and can interact with histone tails through their acidic residues and recruit remodeling complexes [47]. Additional protein domains found in these proteins include TRFH, ADA2-like ZZ, and the DnaJ-domain (Table 2 and Figure 1C).

**Table 2.** EiMYB proteins of *E. invadens* named and classified according to their homology to *H. sapiens* and *A. thaliana*.

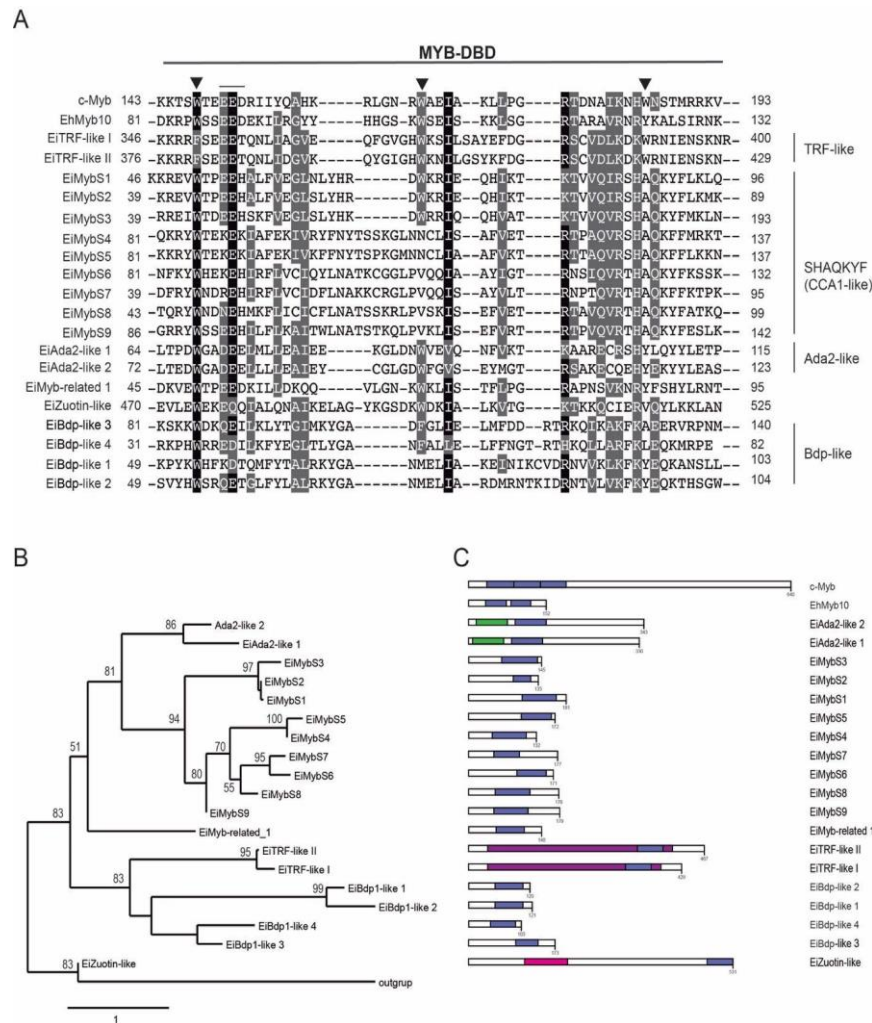
1R-MYBS										
Group	Gene ID	Gene (pb)	mRNA	Protein name	DBD-MYB	InterProScan domains	CD-Search domains	<i>H. sapiens</i> c-Myb	<i>A. thaliana</i>	<i>E. histolytica</i>
TRF-like	EIN_023650	1290	1290	EiTRF-like I	348-392	Telomeric Repeat Binding Factor 1// TM Helix	SANT_TRF/SANT Superfamily	29.07% 7e-12 29.90% 3e-17 (TERF1)	48.08% 3e-13 (CAD531509.1)	49.89% 5e-133 (EHI_148140)
	EIN_079420	1404	1404	EiTRF-like II	378-422	Telomeric Repeat Binding Factor 1// TM Helix	SANT_TRF/SANT Superfamily	26.32% 4e-14 35.87 8e-17 (TERF1)	21.54% 2e-15 (OAP03200.1)	55.13% 5e-164 (EHI_001110)
CCA-like (SHAQKYF)	EIN_086260	540	540	EiMybS1	48-92	Myb-DNA Binding/SANT Superfamily	Myb-DNA Binding/SANT Superfamily/RSC8 Chromatin remodeling	22.95% 4e-14	64.71% 6e-25 (AAF23291.1)	63.10% 2e-45 (EHI_092160)
	EIN_087120	537	537	EiMybS2	46-90	Myb-DNA Binding/SANT Superfamily	Myb-DNA Binding/SANT Superfamily	21.92% 1e-14	45.37% 6e-22 (AAF81310.1)	52.43% 2e-35 (EHI_092160)
	EIN_031250*	601	534	EiMybS3	41-85	Myb-DNA Binding/SANT Superfamily	Myb-DNA Binding/SANT Superfamily	25.45% 2e-14	58.06% 6e-20 (OAP07468.1)	54.55% 4e-35 (EHI_136420)
	EIN_095950	519	519	EiMybS4	83-133	Myb-DNA Binding/SANT Superfamily	SANT Superfamily	16.13% 1e-11	20.83% 3e-12 (CAA0383923.1)	38.22% 1e-18 (EHI_155580)
	EIN_224050	516	516	EiMybS5	83-133	Myb-DNA Binding/SANT Superfamily	SANT Superfamily	25.64% 2e-09	27.78% 7e-12 (NP_00107786.1)	39.10% 6e-17 (EHI_155580)
	EIN_020720	438	438	EiMybS6	56-106	Myb-DNA Binding/SANT Superfamily	SANT Superfamily	-	24.29% 2e-12 (CAA0367555.1)	62.02% 5e-36 (EHI_051440)
	EIN_469690	399	399	EiMybS7	41-91	Myb-DNA Binding/SANT Superfamily	SANT Superfamily	-	21.51% 4e-16 (AAM63125.1)	59.84%

	EIN_081930	408	408	EiMybS8	45-95	Myb-DNA Binding/SANT Superfamily	SANT Superfamily	-	29.63% 5e-14 (CAA0198797.1)	51.88% 1e-33 (EHI_013340)	5e-38 (EHI_051440)
	EIN_407300	546	546	EiMybS9	88-138	Myb-DNA Binding/SANT Superfamily	SANT Superfamily	23.64 % 3e-11	19.77% 5e-12 (NP_001330337.1)	46.92% 6e-19 (EHI_038640)	
Ada2-like	EIN_359680*	1229	993	EiAda2-Like 1	66-108	ADA2-like ZZ	Histone acetyltransferase complex SAGA/ADA, subunit ADA2 [Chromatin structure and dynamics]	18.75% 3e-13	30.59% 1e-62 (CAD5328117.1)	64.85% 9e-146 (EHI_142140)	
	EIN_390470	1032	1032	EiAda2-like 2	74-116	ADA2-like ZZ	Histone acetyltransferase complex SAGA/ADA, subunit ADA2 [Chromatin structure and dynamics]	20.83% 1e-09	34.38% 2e-60 (CAD5328118.1)	46.86% 9e-92 (EHI_142140)	
Myb-related	EIN_020090	423	423	EiMyb-related 1	47-91	Myb-DNA Binding/SANT Superfamily	Myb-DNA Binding/SANT Superfamily	29.1% 4e-17	32% 4e-18 (NP_201038.1)	25.00% 6e-19 (EHI_009930)	
Zuotin	EIN_182440*	1658	1596	EiZuotin-like	472-521	DNAJ domain	ZUO1 Superfamily / SANT Superfamily	16.26% 2e-21	30.24% 2e-29 (AAG51437.1)	52.35% 8e-62 (EHI_128200)	
Bdp1-like	EIN_223710	366	366	EiBdp1-like 1	45-111	Transcription factor TFIIIB component B', Myb domain	SANT/Myb-like DNA-binding domain-containing protein	- 28.95% 0.017 TFIIIB	31.58% 8e-05 (CAB43631.1)	34.29% 2e-10 (EHI_074810)	
	EIN_034860	363	363	EiBdp1-like 2	48-110	Transcription factor TFIIIB component B'', Myb domain	SANT/Myb-like DNA-binding domain-containing protein	- 30.88% 5e-20 TFIIIB	25.69% 4e-20 (CAD5330371.1)	ND	
	EIN_314460	522	522	EiBdp1-like 3	84-126	Transcription factor TFIIIB component B'', Myb domain	SANT/Myb-like DNA-binding domain-containing protein	17.65% 6e-15	23.70% 5e-20 (CAD5330371.1)	65.68% 3e-37 (EHI_009820)	
	EIN_096130	312	312	EiBdp1-like 4	30-90	Transcription factor TFIIIB component B'', Myb domain	BDP1 super family	-	30.77% 3e-06 (CAB43631.1)	40.59% 3e-12 (EHI_009820)	

2R-MYBS										
I	EIN_284910	519	519	EiMyb1	28-121	Myb-DNA Binding/SANT Superfamily	Transcription repressor MYB5; Provisional	33.63% 6e-44	34.35% 1e-42 (NP_001330339.1)	42.55% 2e-43 (EHI_063550)
	EIN_178740	531	531	EiMyb2	26-119	Myb-DNA Binding/SANT Superfamily	Transcription repressor MYB5; Provisional	30.89% 4e-46	33.33% 1e-43 (NP_190575.1)	42.25% 1e-39 (EHI_063550)
-	EIN_047330	474	474	EiMyb3	18-112	Myb-DNA Binding/SANT Superfamily	SANT DNA binding domain / Transcription repressor MYB5; Provisional	38.39% 2e-40	31.45% 2e-42 (NP_190575.1)	35.26% 1e-41 (EHI_063550)
	EIN_206260	450	450	EiMyb4	16-110	Myb-DNA Binding/SANT Superfamily	PLN03091 super family hypothetical protein; Provisional	34.17% 6e-45	35.42% 1e-45 (NP_190575.1)	34.09% 6e-41 (EHI_063550)
	EIN_169560	438	438	EiMyb5	14-109	Myb-DNA Binding/SANT Superfamily	REB1 superfamily	25.19% 1e-44	37.93% 5e-43 (NP_190575.1)	37.40% 1e-37 (EHI_063550)
	EIN_168610	447	447	EiMyb6	16-110	Myb-DNA Binding/SANT Superfamily	PLN03091 super family hypothetical protein; Provisional	27.83% 1e-42	37.93% 5e-43 (NP_190575.1)	39.69% 1e-37 (EHI_098070)
	EIN_207200	447	447	EiMyb7	17-79	Myb-DNA Binding/SANT Superfamily	REB1 superfamily	29.29% 4e-45	36.36% 1e-44 (NP_190575.1)	42.34% 5e-41 (EHI_098070)
	EIN_022390	495	495	EiMyb8	29-122	Myb-DNA Binding/SANT Superfamily	REB1 superfamily	31.03% 5e-46	30.71% 4e-43 (NP_190575.1)	42.31% 2e-39 (EHI_063550)
	EIN_080130	504	504	EiMyb9	23-114	Myb-DNA Binding/SANT Superfamily	REB1 superfamily	32.65% 4e-46	33.10% 1e-41 (NP_190575.1)	38.40% 2e-40 (EHI_063550)
II	EIN_276810*	754	702	EiMyb10	31-124	Myb-DNA Binding/SANT Superfamily	Transcription repressor MYB5; Provisional	38.53% 4e-42	33.56% 6e-42 (NP_190575.1)	31.48% 2e-38 (EHI_063550)
	EIN_307410	507	507	EiMyb11	20-113	Myb-DNA Binding/SANT Superfamily	Transcription repressor MYB5; Provisional	30.58% 5e-46	31.72 7e-46 (VYS56784.1)	47.47% 4e-39 (EHI_063550)
	EIN_307180	468	468	EiMyb12	19-113	Myb-DNA Binding/SANT Superfamily	Transcription repressor MYB5; Provisional	38.61% 5e-43	32.88% 5e-44 (NP_190575.1)	40.27% 2e-41 (EHI_063550)
III		468	468					38.53%	34.04%	40.29%

	EIN_308550			EiMyb13	19-112	Myb-DNA Binding/SANT Superfamily	Transcription repressor MYB5; Provisional	1e-43	4e-43 (NP_190575.1)	1e-40 (EHI_063550)
-	EIN_095310	474	474	EiMyb14	30-124	Myb-DNA Binding/SANT Superfamily	PLN03091 super family hypothetical protein; Provisional	37.07% 3e-40	33.87% 1e-40 (NP_195443.1)	51.52% 2e-40 (EHI_153350)
	EIN_399710*	1119	918	EiMyb15	161-255	Myb-DNA Binding/SANT Superfamily	REB1 superfamily	35.66% 2e-43	34.27% 2e-42 (NP_195443.1)	68.29% 1e-48 (EHI_098070)
	EIN_490880	477	477	EiMyb16	41-140	Myb-DNA Binding/SANT Superfamily	REB1 superfamily	38.89% 2e-41	35.83% 1e-42 (NP_195443.1)	39.09% 3e-39 (EHI_063550)
	EIN_310240	444	444	EiMyb17	31-125	Myb-DNA Binding/SANT Superfamily	REB1 superfamily	45.05% 4e-44	36.36% 3e-44 (NP_195443.1)	44.25% 9e-40 (EHI_098070)
IV	EIN_425380	486	486	EiMyb18	32-124	Myb-DNA Binding/SANT Superfamily	Transcription repressor MYB5; Provisional	28.93% 1e-43	33.05% 6e-42 (OAO92063.1)	44.14% 4e-35 (EHI_098070)
	EIN_046410	504	504	EiMyb19	39-132	Myb-DNA Binding/SANT Superfamily	REB1 superfamily	34.58% 2e-43	34.91% 1e-41 (OAO92063.1)	51.25% 3e-47 (EHI_063550)
	EIN_183110	456	456	EiMyb20	31-124	Myb-DNA Binding/SANT Superfamily	PLN03091 super family hypothetical protein; Provisional	33.07% 4e-46	37.01% 3e-46 (VYS56784.1)	38.84% 1e-38 (EHI_098070)
	EIN_183730	459	459	EiMyb21	31-124	Myb-DNA Binding/SANT Superfamily	PLN03091 super family hypothetical protein; Provisional	38.10% 2e-42	33.08% 2e-42 (CAA0401764.1)	64.10% 2e-39 (EHI_130060)
	EIN_169190	453	453	EiMyb22	15-108	Myb-DNA Binding/SANT Superfamily	Transcription repressor MYB5; Provisional	29.82% 2e-41	35.77% 2e-41 (AAS58517.1)	46.48% 7e-39 (EHI_168310)
	EIN_359630	453	453	EiMyb23	29-122	Myb-DNA Binding/SANT Superfamily	PLN03091 super family hypothetical protein; Provisional	35.45% 4e-43	39.05% 2e-42 (CAA0383923.1)	57.58% 4e-44 (EHI_129790)
V	EIN_379820	453	453	EiMyb24	29-122	Myb-DNA Binding/SANT Superfamily	PLN03091 super family hypothetical protein; Provisional	34.43% 5e-44	32.41% 1e-43 (VYS56784.1)	53.62% 5e-42 (EHI_129790)
-	EIN_168860	495	495	EiMyb25	29-126	Myb-DNA Binding/SANT Superfamily	PLN03091 super family hypothetical protein; Provisional	29.13% 9e-46	33.93% 3e-42 (CAD5329766.1)	46.23% 1e-32 (EHI_092700)

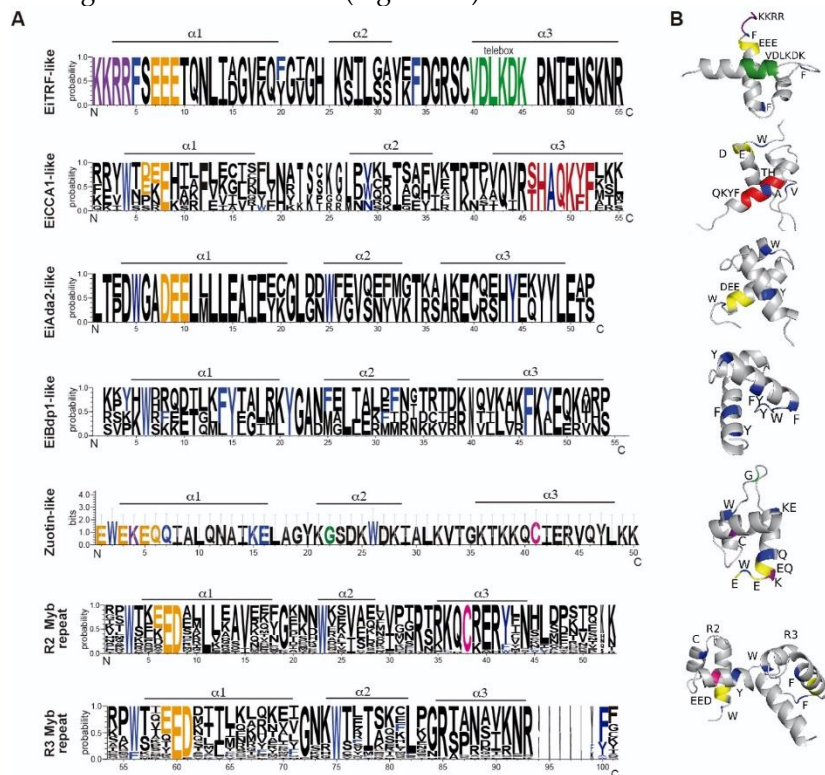
-	EIN_405040	615	615	EiMyb26	65-458	Myb-DNA Binding/SANT Superfamily	REB1 superfamily	28.23% 4e-40	30.86% 1e-43 (AAS58517.1)	43.37% 5e-36 (EHI_053000)
-	EIN_248780	714	714	EiCDC5-like	10-102	CDC5L_II		22.78% 5e-39 58.71% 8e-41 (CDC5)	53.45% 2e-55 (OAP18307.1 CDC5)	78.48% 2e-22 (EHI_000550)
<b>4R-MYBS</b>										
<b>SNAP-like</b>	EIN_267690	1992	1992	EiSNAP-like I	R1 436-599	SANT/Myb domain	SANT/Myb domain	28.22% 2e-11 39.08% 7e-08 (SNAP190)	31.22% 3e-18	29.97% 1e-185 (EHI_130710)



**Figure 1. 1R-MYB proteins of *E. invadens*.** **A)** ClustalW alignment of the MYB-DBD region. Arrowheads indicate conserved tryptophan residues. The acidic patch is underlined. Groups are indicated at right. Numbers indicate the MYB-DBD position of each protein shown in the alignment. **B)** Phylogenetic tree of the 1R-MYB proteins obtained in phylogeny.fr. Bootstrap values >50% (from 1,000 replicates) are shown near the individual branches. EiAda2-like 2 (EIN\_390470), EiAda2-like 1 (EIN\_359680), EiMybS3 (EIN\_031252), EiMybS2 (EIN\_087120), EiMybS1 (EIN\_086260), EiMybS5 (EIN\_224050), EiMybS4 (EIN\_095950), EiMybS7 (EIN\_469690), EiMybS6 (EIN\_020720), EiMybS8 (EIN\_081930), EiMybS9 (EIN\_407300), EiMyb-related\_1 (EIN\_020090), EiTRF-like II (EIN\_079420) EiTRF-like I (EIN\_023650), EiBdp1-like 1 (EIN\_223710), EiBdp1-like 2 (EIN\_034860), EiBdp1-like 4 (EIN\_096130), EiBdp1-like 3 (EIN\_314460), EiZuotin-like (EIN\_182440), Outer group (*Mus musculus* P52651). **C)** Schematic representation of 1R-EiMyb proteins according to their size and domains. Blue, MYB-DBD; Green, ADA2-like ZZ; Pink, DNAJ and purple, TRF 1 domains. c-Myb and EhMyb10 are used as reference.

The *E. invadens* 1R-MYB proteins were then subjected to a phylogenetic study. Different clades with strong support values were identified by the tree topology: Zuotin, transcription factor III B (Bdp-like), Adaptor 2 (Ada2-like), telomeric repeat-binding factors (TRF-like), and circadian clock-associated (CCA1-like) (Figure 1B). CCA1-like is the largest subgroup with nine members with the conserved SHAQK(Y/F) in the third helix of the MYB-DBD, as in *E. histolytica* proteins, and with high identity with CCA1 proteins from *A. thaliana* [48] (Table 2). These proteins were dubbed EiMybS proteins (EiMybS1 to EiMybS9). EiMybS7 and EiMybS9 have a THAQK(Y/F) motif, where a threonine substitutes the serine (Figure 1A). The SHAQKYF-MYB proteins are common in plants, algae, and *D. discoideum*, indicating a restricted distribution in only some phyla. Studies in plants have shown that some SHAQKYF-MYBs are sequence-specific TFs that regulate the expression of

clock-regulated genes and stress responses [49]. We performed multiple alignments and generated separate sequence logos for the MYB-DBD (Figure 2A).



**Figure 2. Sequence logos of the *E. invadens* Myb proteins.** Multiple alignments of MYB domains were performed with ClustalW software and visualized with WebLogo 3. **A)** The Y-axis score indicates the probability for each position in the sequence. Black lines illustrate the position of the three  $\alpha$ -helices in MYB-DBD. Blue: conserved hydrophobic residues; yellow: acidic patch. EiTRF-like: purple, amino-linker; green, telebox EiCCA1-like: red, SHAQKYF EiAda2-like EiBdf1-like R2: conserved cysteine of the KQCRER motif shown in magenta, and R3 repeat. **B)** Molecular structures of the MYB domain of 1R and 2R-MYB proteins obtained by AlphaFold and visualized with PyMOL.

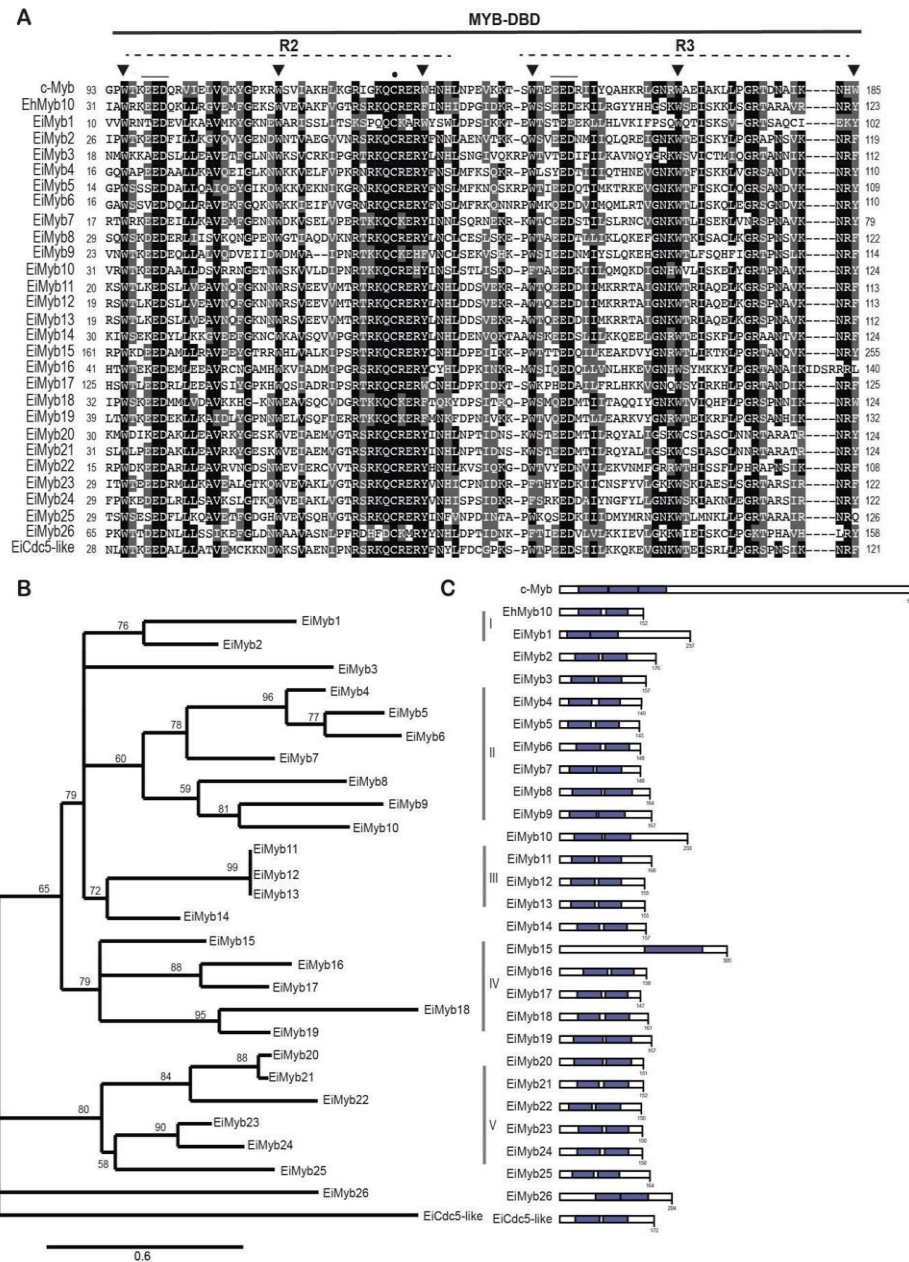
The SHAQKYF motif is localized in the third  $\alpha$ -helix and probably because of the diversity of the CCA1-like subgroup, a second helix is not clearly observed the CCA1-like conserves the acidic patch as well as the hydrophobic residues involved in the stability of the HTH structure (Figure 2A and B). The TRF-like subgroup is formed by two proteins that conserve basic amino acids in the first positions (KKRR) and the telebox motif LKDKWRN (Figure 2A), which is involved in the recognition of telomeric DNA [14] and were named EiTRF-like I and EiTRF-like II due to their high identity with TRF proteins (Table 2). The telebox motif suggests the presence of a conserved mechanism of telomeric protection in these early-branched parasites. When analyzing the molecular structure, the telebox motif conforms the first portion of the third helix that stabilizes DNA binding (Figure 2B) and therefore could be implicated in telomere recognition. This parasite possesses only two TRF-like proteins, whereas *E. histolytica* preserves three (EhTRF-like I, II, and III) [50]. This leads us to hypothesize that gene duplication occurred in *E. histolytica*. The Ada2-like subgroup is formed by two proteins named EiAda2-like 1 and 2 that contain the ADA2-like ZZ domain. ADA2 proteins are transcriptional coactivators of the SAGA complex involved in chromatin remodeling and transcriptional regulation; they also stabilize complexes formed by direct interactions between activators and general factors in eukaryotes and were identified in *P. falciparum* [51]. Interestingly, EiAda2-like proteins may have a similar role in *E. invadens* as protein ADA2, which is a component of complexes with histone acetyltransferase. The logo sequences for the MYB-DBD region for each group (Figure 2A) show the acidic patch, and the first and second conserved tryptophans that conform the clearly distinguishable HTH and a well-structured hydrophobic core (Figure 2B).

In addition, a 1R-MYB protein resembles a Zuotin protein because of the presence of a characteristic DNAJ domain and is dubbed EiZuotin-like. Although this protein has two MYB-DBD repeats, it was classified as 1R because the second repeat is imperfect. EiZuotin-like could be related to MIDA1, a Zuotin protein in the fungus that contains two repeats of the DBD-MYB and a DNAJ domain [52]. Zuotin proteins have *in vitro* binding activity to tRNA and Z-DNA [53,54] and are also ribosome-associated proteins [55]. The 3-D of the DBD-MYB region of Ei-Zuotin-like shows two helix structures but a not so defined hydrophobic core, however Zuotin proteins harbor, besides the DNAJ domain, an evolutionary conserved 4HB domain besides that serves as a linker to the SANT domain and contributes to its stability [56]. Finally, four proteins were classified as EiBdp-like. The Bdp1 protein is one of the three subunits of the TFIIIB complex and is also termed B". Recruitment of Pol III and promoter opening during transcription initiation depend on Bdp1. The C-terminal region of Bdp1 contains a conserved SANT domain, which normally functions as a DNA-binding module. When transcription begins, Bdp1 is situated within the Pol III active site cleft [57]. The four EiBdp-like proteins identified in *E. invadens* are significantly different from the human and *Saccharomyces cerevisiae* proteins in the sequences flanking the MYB-DBD. Three of these EiBdp-like proteins were not previously reported because of the divergence of their MYB-DBD that can be observed in the logo generated (Figure 2A).

However, the molecular structure shows two long well-defined helices and one short one that indicates a clear HTH structure related with a stable hydrophobic core (Figure 2B). Finally, one protein was not grouped but considered Myb-related because of its high identity with *Arabidopsis* Myb transcription factors. In summary, all these proteins could function as transcriptional factors, telomere recognition proteins, transcription coactivators, ribosome-associated proteins, or DNAJ molecular chaperones. In contrast, *E. histolytica* only has 17 1R-MYB proteins that include CCA1-like, TRF, and HAT-related (ADA2), which suggests that *E. invadens* requires a greater number of transcriptional regulators, probably because of the diversity of environments and hosts in which it develops.

### 3.3. 2R-MYB subfamily in *E. invadens*

The most prevalent subfamily in *E. invadens*, as well as in *E. histolytica*, is the 2R-MYB subfamily, also known as R2R3-Myb proteins. This 2R-MYB subfamily comprises 27 ORFs; interestingly, they have more similarity to plant 2R-MYB proteins than to *H. sapiens* c-Myb (Table 2). These proteins are mostly annotated as transcription factors related to *A. thaliana* Myb proteins (transcription factor MYB, putative; transcription factor WEREWOLF, putative; trichome differentiation protein GL1, putative; r2r3-MYB transcription factor, putative or C-MYB, putative) in AmoebaDB, although there are three proteins annotated as hypothetical. In 2R-MYB are the most abundant Myb proteins in plants, and are involved in a variety of biological activities, including seed development, morphogenesis, meristem formation, secondary cell wall production, and hormonal signal transmission [58,59]. Although these proteins greatly conserve their MYB-DBD, their N and C-termini are divergent, often having residues in disordered regions that may undergo post-translational modifications and therefore could affect the transcription factor stability or localization [60]. These proteins contain two repeats in their MYB-DBD. The size of these proteins ranges from 145 to 305 amino acids and is similar to that of their *E. histolytica* counterparts, with a molecular weight of 17.21 kDa to 36.47 kDa, respectively. The pI of R2R3-Myb proteins varied from 6.18 to 10.08 (Table S1). These proteins are predicted to be localized in the nucleus, and in some cases, a nuclear localization signal was predicted, supporting their role in transcriptional regulation (Table S1). These proteins contain two repeats in their MYB-DBD. Alignment analysis of the MYB-DBD revealed that the first and second tryptophan residues of repeats 2 and 3 are conserved; nevertheless, substitutions with aromatic residues are present in the third tryptophan of both repeats, often replaced by tyrosine or phenylalanine residues (Figure 3A).

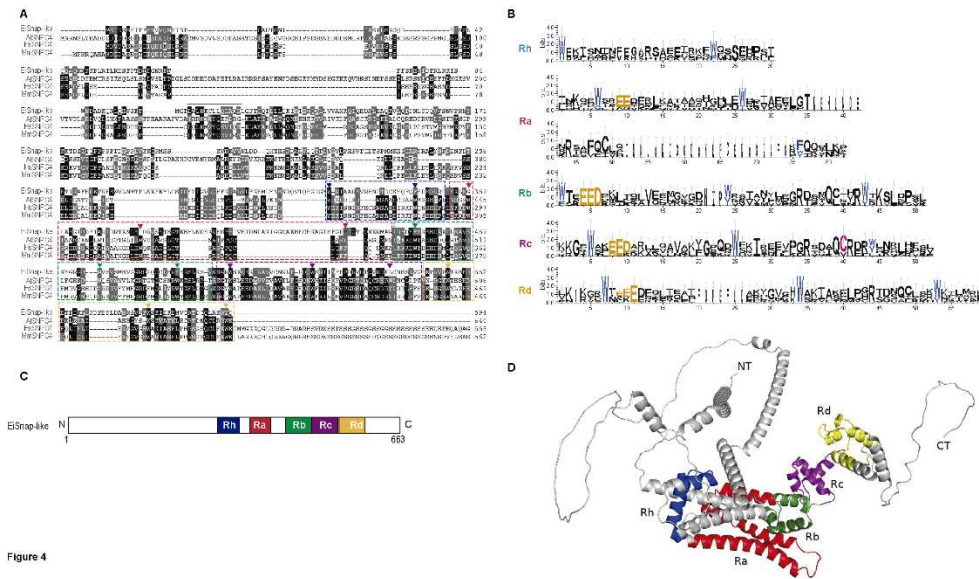


**Figure 3. 2R-MYB proteins of *E. invadens*.** **A)** ClustalW alignment of the MYB-DBD region. Arrowheads indicate conserved tryptophans. Numbers indicate the MYB-DBD position of each protein shown in the alignment. The acidic patch is underlined, and R2, and R3 repeats are indicated as dotted lines. The black circle indicates the conserved cysteine residue in the R2 repeat. **B)** Phylogenetic tree of the 2R-MYB proteins obtained in phylogeny.fr. Bootstrap values >50% (from 1,000 replicates) are shown near the individual branches. Groups are indicated at the right. EiMyb1 (EIN\_284910), EiMyb2 (EIN\_178740), EiMyb3 (EIN\_047330), EiMyb4 (EIN\_206260), EiMyb5 (EIN\_169560), EiMyb6 (EIN\_168610), EiMyb7 (EIN\_207200), EiMyb8 (EIN\_022390), EiMyb9 (EIN\_080130), EiMyb10 (EIN\_276810), EiMyb11 (EIN\_307410), EiMyb12 (EIN\_308550), EiMyb13 (EIN\_308550), EiMyb14 (EIN\_095310), EiMyb15 (EIN\_399710), EiMyb16 (EIN\_490880), EiMyb17 (EIN\_310240), EiMyb18 (EIN\_425382), EiMyb19 (EIN\_046410), EiMyb20 (EIN\_183110), EiMyb21 (EIN\_183730), EiMyb22 (EIN\_169190), EiMyb23 (EIN\_359630), EiMyb24 (EIN\_379820), EiMyb25 (EIN\_168860), EiMyb26 (EIN\_405040), EiCdc5-like (EIN\_248780). **C)** Schematic representation of Myb proteins according to their size and domains. Blue, MYB-DBD. c-Myb and EhMyb10 are used as references.

The highly conserved patch of acidic residues such as glutamic or aspartic acid, is common to all Myb-related domains and is also in 2R-MYB proteins in *E. invadens* (Figure 3A). These acidic residues are positioned in the first of the alpha-helices within each of the two repeats that comprise the MYB-DBD (Figures 3B). In c-Myb, the acidic residues are relevant for transcriptional activity, chromatin binding, and interaction with the H4 histone N-terminal tail (Ko et al., 2008). A conserved cysteine residue in the third helix of the R2 domain of all the *E. invadens* 2R-MYB proteins was also present, forming the QCRER motif (Figure 3A), as in the *E. histolytica* R2R3Myb proteins. This motif can be observed in the third helix of the R2 repeat near to the acidic residues localized in the first helix (Figure 2B). The conserved cysteine is relevant for REDOX-dependent DNA binding in mammals, plants, and other eukaryotic organisms [61]. Next, we performed a phylogenetic analysis of the 2R-MYB proteins (Figure 3B), which were further divided into five subgroups (I, II, III, IV, and V), except for three protein sequences that could not be grouped. 2R-Mybs were dubbed according to their position in the phylogenetic analysis and other present domains (Table 2). In most cases, the MYB-DBD is located in the middle of the polypeptide and comprises almost the total length of the protein (Figure 3C). Finally, the protein access number EIN\_248780 presents a high identity with the CDC5 protein from *H. sapiens* and *A. thaliana* and is therefore dubbed EiCDC5-like. CDC5 proteins have two MYB repeats followed by a third imperfect MYB-like repeat, or D3 domain. In *S. cerevisiae*, the ortholog of CDC5 has been reported to play a role in pre-mRNA splicing [16], but it also functions as a transcription factor in plants that recognize the DNA-binding consensus CTCAGCG, showing multiple roles in transcriptional regulation [62].

#### 3.4. 4R MYB-DBD protein

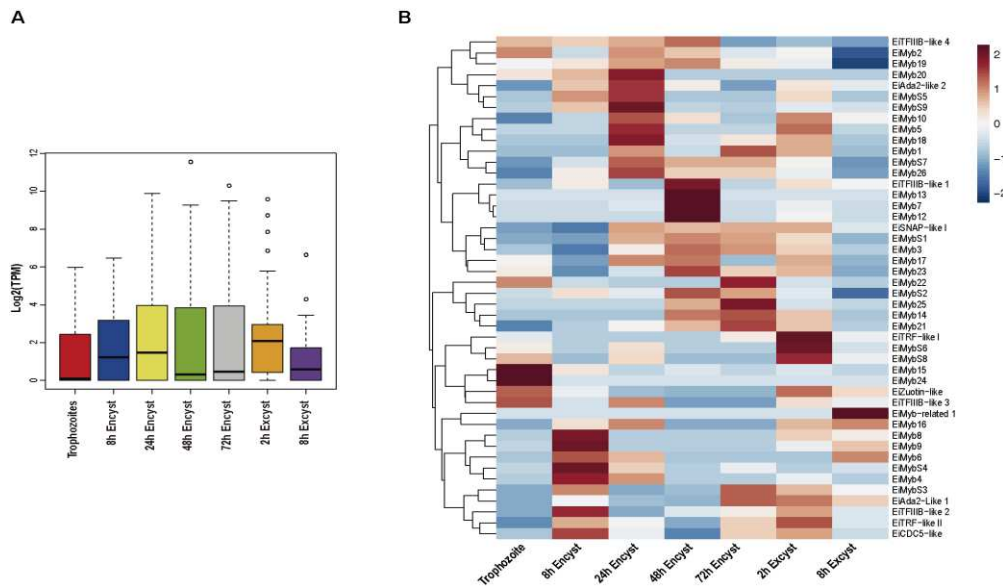
With 663 amino acids, EIN\_267690 encodes the largest Myb protein found in *E. invadens* and interestingly it has not detectable nuclear localization signals (Table S1). In AmoebaDB, this protein is annotated as snap190 putative, with 26.47% identity to c-Myb and 22.9% and 23.03% identity to SNPC4 from *H. sapiens* and *A. thaliana*, respectively (Table 2). The MYB-DBD from EiSnap-like exhibits substantial conservation of amino acid residues, which are essential for the sequence-specific binding of the promoter region of snRNA genes [63,64] (Figure 4A,B). The MYB-DBD from EiSnap-like has four MYB repeats: Ra, Rb, Rc, and Rd, and an additional half MYB repeat (Rh) situated N-terminal to Ra according to the nomenclature used for the HsSNAPc4 (Figure 4C) [65]. The 3-D structure shows a mostly helicoidal conformation and a non-structured NT and CT regions that could contribute regulating EiSNAP-like (Figure 4D). 4R-MYB has been reported as the small nuclear RNA (snRNA)-activating protein complex subunit that participates in the transcription initiation of snRNAs in plants [63]. Both RNA polymerase II and III snRNA gene transcription requires the complex SNAPc, in which SNAP190 proteins participate. Most eukaryotes have SNAPc, which can have three or five subunits depending on the species [65,66]. Interestingly, the SNAP proteins have been identified in the Excavata group, including *G. lamblia*, *L. major*, *T. brucei*, and *Naegleria gruberi*, with 64% identity [63]. As was mentioned earlier, no 4R-MYB proteins have been previously discovered in *E. histolytica*; however, the protein encoded by the locus EHI\_130710 is considered its ortholog in the AmoebaDB database; therefore, it would be interesting to identify if it indeed possesses a 4R-MYB as well as the genes that are regulated by these proteins in both parasites.



**Figure 4.** 4R-MYB protein of *E. invadens*. **A)** ClustalW alignment of the amino-terminal region of EiSnap-like and SNAPc orthologs from *A. thaliana*, *H. sapiens*, and *Mus musculus*. Arrowheads indicate the conserved tryptophans, and the dotted line indicates the four adjacent MYB repeats Ra, Rb, Rc, and Rd (red, green, brown and yellow dotted boxes) with an additional half MYB repeat (Rh) in front of Ra (blue dotted box). **B)** Sequence logos generated from the multiple sequence alignment of the analyzed ortholog 4R-MYB proteins. **C)** Schematic diagram of EiSnap-like domains visualized with Dog 2.0 **D)** Three-dimensional structure of EiSnap-like protein performed in AlphaFold and visualized with PyMOL.

### 3.5. Expression analysis of the *eimyb* genes during trophozoite differentiation

Focused on cyst differentiation, the transcriptome dataset obtained through RNAseq data from AmoebaDB was analyzed. We analyzed the expression profiles of *eimyb* genes from encysting parasites (at 8, 24, 48, and 72 h after transfer to encystation media) and from excysting parasites (2 and 8 h after induction of excystation) [6]. When analyzing the median and distribution of expression values of all *eimyb* genes in trophozoite differentiation, we observed an upregulation during late encystation (24–72 h) (Figure 5A). The expression patterns of the 47 *eimyb* genes in *E. invadens* under encysting conditions was visualized using a heatmap analysis. We observed that only 11 are expressed in the trophozoite stage, with *eimyb15* and *eimyb24* having the greatest expression (Figure 5B). In addition, *eimyb24* is a trophozoite-specific gene. In *E. histolytica*, its ortholog is EhMyb10, which means that EhMyb10 could be essential for the parasite and therefore a potential target for therapy development. Forty-six *eimyb* genes modulate their expression during cyst differentiation; therefore, we searched for signatures that could suggest stage specific Myb proteins (Figure 5B). During early encystation (h), 18 *eimyb* genes are expressed, with *eimyb9* and *eimyb4* being the most expressed. During encystation progression (24, 48, and 72 hours), 23, 19, and 18 *eimyb* genes are expressed (Figure 5B). At 24 hours, *eimyb9*, *eimyb18* and *eimyb20* have the greatest expression. In late encystation (48 h), *eimyb7*, *eimyb12*, and *eimyb13* are most expressed. Interestingly, these three genes appear as a specific signature for this encystment time (Figure 5B). At 72 hours of encystation *eimyb22*, *eimyb25* expresses the most. On the other hand, excystation is an important process that ensures *E. invadens* dissemination; interestingly, during early excystation times (2 h), the greatest number of *eimyb* genes is expressed (25 *eimyb* genes).



**Figure 5. Expression profile of *E. invadens* Myb genes during encystation and excystation. A)** Boxplot showing the number of *eimyb* genes expressed in each condition analyzed during trophozoite-cyst differentiation. The middle lines in the boxplot represent the median, and circles represent outliers. **B)** Hierarchical clustering heatmap of *eimyb* genes; each column represents a gene, and each row represents a condition. The colors in the graph represent the sample's level of gene expression [Log<sub>2</sub>(TPM)]. Blue signifies that the gene expression is low in the sample, whereas red shows that the gene is strongly expressed. Data were obtained from AmoebaDB.

This could be due to the parasite's need to reactivate transcription and initiate reptilian host invasion, as previous studies have shown that from the total transcriptome, 1,025 and 1,032 genes are upregulated at 2 h and 8 h, respectively [13]. At 2 h excystation, *eimybs6*, *eimybs8*, and *eitrF-like 1* showed the greatest expression. At 8 h excystment, only 7 *eimyb* genes are expressed, and *eimyb-related 1* has the greatest expression and is specific to this stage time (Figure 5B). Altogether, these data suggest that while widely expressed *eimybs* may control the transcription of a large number of genes, a specific set of *EiMyb* proteins is required to modulate the spatial-temporal expression patterns during trophozoite-cyst differentiation. Therefore, it is important to study the genes that are regulated through this *Myb* selective expression. In agreement, we did not observe a constitutive expression pattern of any of the *eimyb* genes, which reinforces their specific role during parasite development. This could explain why, in other studies, only a subset of cyst-specific genes is induced when a single *eimyb* gene is overexpressed [35]. Interestingly, the gene that codes for *EiCDC5-like*, a protein similar to *CDC5* that participates in splicing, suggesting that splicing might be a necessary process in early encystation-excystation. This is interesting because almost 26% of the expressed genes contain introns (1536 from 5894 genes with introns from genome annotation) [6] suggesting the necessary participation of the spliceosome in these stages. Lastly, *eitrF-like I* and *II*, are expressed in specific stages (Figure 5B) in which replication occurs, as nuclear division is necessary for encystment, therefore these proteins could be required for telomeric protection. *TRF-like* proteins have been identified and characterized in *T. brucei*, *T. cruzi*, *L. major* and *E. histolytica*, where their role as telomere DNA-binding proteins has been reported to provide a possible function in telomere-end protection [24,25,50].

### 3.6. Presence of the *Myb* recognition element (MRE) and the C-rich sequence in *E. invadens* gene promoters

To identify the target genes of *EiMyb* proteins, we searched for *in silico* *Myb* recognition elements in *E. invadens* gene promoters through two DNA sequences previously identified in *E. histolytica*: the canonical *Myb* recognition element (MRE) and a C-rich sequence [28,35]. In this analysis, 2,559 genes had the canonical MRE in their promoter region; 1,700 genes are annotated as hypothetical; and 859 have predicted functions. In contrast, 288 genes have a C-rich sequence in the

promoter region (192 hypothetical and 96 with predicted functions). The MRE and C-rich sequences were confirmed through STREME (Table 3). Interestingly, the signatures of both sequences had slight differences depending on the encystation or excystation stage (Table 3). Subsequently, we observed that 815 and 838 genes modified at least 2-fold their expression during encystment and excystment, while 99 and 100 genes modified their expression during encystment and excystment, respectively, for MRE and C-rich sequences (Table 3). Further experimental analysis could confirm that these signatures are recognized by EiMyb proteins.

**Table 3.** MRE and CCCCC motif search in *E. invadens* gene promoters (-500 +50 pb).

Motif consensus sequence	Modified sequence	Motif containing genes*	E-value	STREME confirmed	Stage related genes
[T/C]AAC[G/T]G	CAACTG	2559 (21.31%)	2.0 e-036	2541 (99.29%)	Trophozoite
	DCAACTG	815 (6.78%)	1.3 e-011	807 (99.01%)	Encystation
	CAACTG	838 (6.97%)	5.5 e-011	834 (99.52%)	Excystation
[CA]CCCCC	MCCCCC	288 (2.39%)	1.1 e-007	284 (98.6%)	Trophozoite
	ACCCCCA	99 (0.82%)	6.8 e-003	97 (97.97%)	Encystation
	CCCCCC	100 (0.83%)	1.8 e-001	98 (98.0%)	Excystation

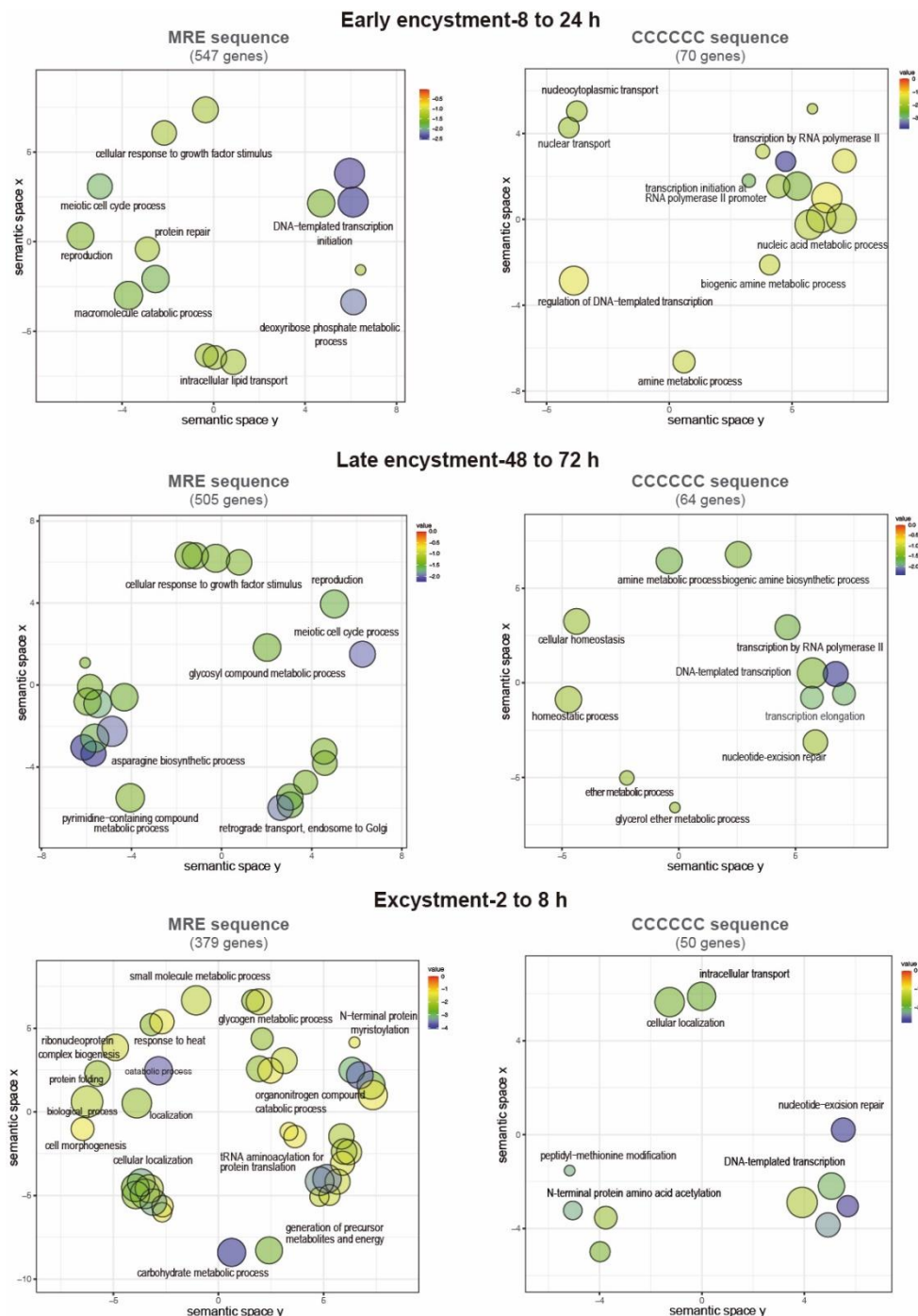
\* Search performed against 12,007 ORFs identified in AmoebaDB. p-value < 0.05.

### 3.7. Functions of the putative EiMyb target genes

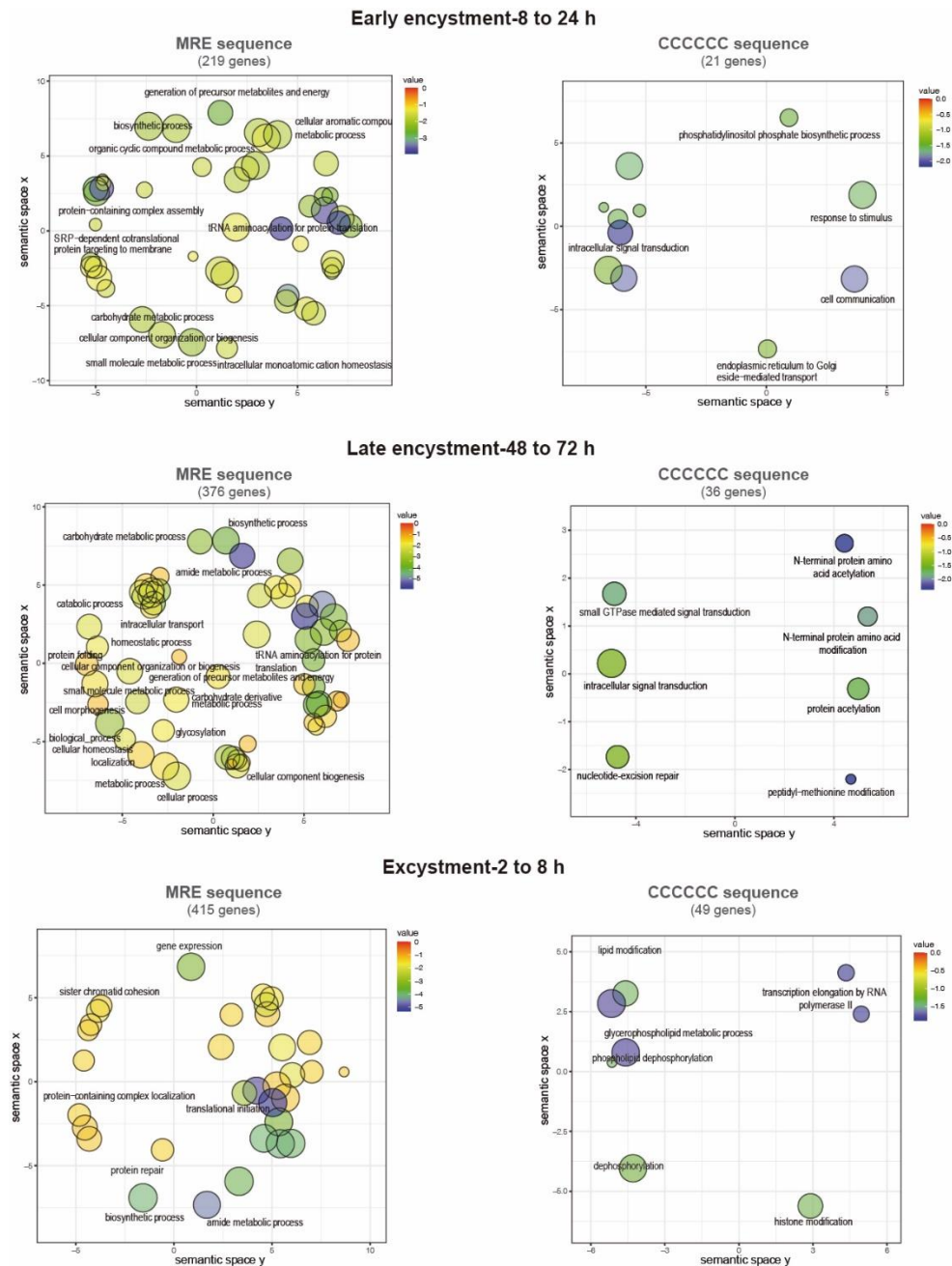
Term enrichment analysis was performed on the previous genes to identify GO categories related to biological processes. Notably, 547 MRE-containing genes upregulate during early encystment and are related to catabolism because cellular metabolism decreases in this stage (Figure 6). An interesting finding was that in early and late encystment (505 total upregulated genes), many DNA and RNA processing genes are upregulated, probably to prepare the cell for encystment and regulate its gene expression. Nuclear division is an important event during encystment to generate tetra-nucleated cysts, and for this, DNA replication must be present, which is represented by meiotic nuclear division and DNA repair upregulated genes. In *E. invadens*, encystation is accomplished by multinucleation events that could benefit the parasite by allowing genomic changes and recombination [67]. In agreement, during encystation, it has been observed that meiotic-related genes are expressed [6]. We also found genes related to the secretion process and exocytosis that could be related to the transportation of cyst-forming components to the cell membrane (Figure 6) [68]. During excystment, 379 genes were upregulated, and an increase in metabolism-related genes was observed, as well as genes related to temperature stimulus responses and the reproductive process. The sexual pathway is induced by the stress response to starvation, as in many eukaryotes such as yeast and *Dictyostelium* [67]. This finding is in agreement with the overexpression of meiotic and homologous recombination genes reported by Ehrenkaufer et al., (2013) during stage conversion. Furthermore, downregulated genes are represented by a metabolic process decrease related to glucose and energy uptake as well as organelle and protein biogenesis (219, 376 and 415 downregulated genes in early and late encystment and excystment, respectively) (Figure 7).

On the other hand, the gene ontology terms associated with the C-rich sequence showed genes involved in cyst formation and were also enriched in genes involved in post-translational modifications (70 and 64 genes upregulated in early and late encystment, respectively). During excystment, among the 47 upregulated genes, we found some related to transcription initiation, probably reactivating the transcription of many genes during this process. Finally, the downregulated genes (21, 36 and 49 genes in early and late encystment and excystment, respectively)

were mostly related to intracellular signal transduction and transcription initiation. We suggest that MYB-DBD proteins could recognize both an MRE element and a C-rich sequence to regulate gene expression in *E. invadens*; however, its genome is approximately 70% AT-rich, which could be related to an increase in the number of MRE identified (2,559 vs. 288 genes).



**Figure 6.** GO annotations of upregulated genes during cyst differentiation in *E. invadens*. Biological process annotations associated with genes containing the MRE and the C-element in their promoters are visualized using a two-dimensional semantic space scatterplot. The spatial organization is based on semantic similarity. The number of node labels is minimized to allow visualization of the node colors on the scatterplot. The score equals the p value for each GO annotation term node. Blue nodes indicate more significant p values and red nodes indicate less significant p values.



**Figure 7. GO annotations of downregulated genes during cyst differentiation in *E. invadens*.** Biological process annotations associated with genes containing the MRE and the C-element in their promoters visualized using two-dimensional semantic space scatterplot. The spatial organization is based on semantic similarity. The number of node labels is minimized to allow visualization of the node colors on the scatterplot. The score equals p value for each GO annotation term node. Blue nodes indicate more significant p values and red nodes indicate less significant p values.

Finally, in *E. histolytica*, the expression of Myb transcription factors in trophozoite is related to invasive ameba (Naiyer et al., 2019). It may be that EiMyb proteins are also related to the ability to infect and invade all tissues of reptilian hosts. On the other hand, encystation in *E. invadens* is triggered by glucose starvation, which in *E. histolytica* is related to the overexpression of some Myb proteins. It is important to mention that 41.95% of the *E. invadens* proteome is common among other species, and the difference in the proteome could be related to the ability to infect different species of reptiles [69]. Therefore, it is understandable that the parasite needs a greater family of transcription factors to respond selectively to the host.

#### 4. Conclusions

In this study, we searched for *E. invadens* MYB-domain-containing proteins, and forty-seven genes encoding for MYB-domain-containing proteins were identified and classified, as well as thoroughly described in this work. Most proteins have domains that are involved in transcription initiation, such as ADA-2, SWI complex I, and Reb1, among others. Expression analysis during encystation-excystation obtained from the AmoebaDB database showed that genes encoding MYB-domain-containing proteins were differentially expressed, some of them only in the trophozoite stage, while others mainly in the cyst stage. This indicates that Myb domain-containing proteins may regulate the expression of stage-specific proteins and a great variety of cellular processes in this parasite. The elucidation of the function and regulation of EiMyb proteins in the *E. invadens* stage transition may lead to the discovery of targets for the development of new chemotherapeutics that interfere with cyst conversion. Also, knowing how Myb proteins tune cyst conversion could help elucidate how the process is executed in *E. histolytica* and promote encystation in vitro through Myb overexpression or repression.

**Supplementary Materials:** The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Table S1: Characteristics of all the full length EiMyb proteins of *Entamoeba invadens* classified according to their repeat number.

**Author Contributions:** P.C.: Investigation, Methodology, Analysis, Validation, Visualization, Writing - Original Draft. E.J.C.O.: Investigation, Methodology. C.R.Z.: Investigation, Methodology. : Methodology, Formal analysis. C.E.M.: Methodology, Formal analysis. M.A.R.: Writing-Review and Editing, Funding acquisition. J.V.: Writing-Review & Editing, Funding acquisition. E.A.L.: Conceptualization, Analysis, Investigation, Writing-Review & Editing, Supervision, Project Administration, Funding acquisition.

**Funding:** This work was supported by CCyT-UACM (grant CCyT-2021-8) and CONAHCYT CF-2019-194163 and postdoctoral fellowships to PC and IC (333090 and 545233, respectively).

**Data Availability Statement:** We encourage all authors of articles published in MDPI journals to share their research data. In this section, please provide details regarding where data supporting reported results can be found, including links to publicly archived datasets analyzed or generated during the study. Where no new data were created, or where data is unavailable due to privacy or ethical restrictions, a statement is still required. Suggested Data Availability Statements are available in section "MDPI Research Data Policies" at <https://www.mdpi.com/ethics>.

**Acknowledgments:** The authors are grateful to Alfredo Barberi for graphic design.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### References

1. Shirley, D.-A.T.; Farr, L.; Watanabe, K.; Moonah, S. A Review of the Global Burden, New Diagnostics, and Current Therapeutics for Amebiasis. *Open Forum Infect Dis* 2018, 5, doi:10.1093/ofid/ofy161.
2. Haque, R.; Huston, C.D.; Hughes, M.; Houpt, E.; Petri, W.A. Amebiasis. *New England Journal of Medicine* 2003, 348, 1565–1573, doi:10.1056/NEJMra022710.
3. Eichinger, D. Encystation in Parasitic Protozoa. *Curr Opin Microbiol* 2001, 4, 421–426, doi:10.1016/S1369-5274(00)00229-0.
4. Eichinger, D. Encystation of Entamoeba Parasites. *BioEssays* 1997, 19, 633–639, doi:10.1002/bies.950190714.
5. Wesel, J.; Shuman, J.; Bastuzel, I.; Dickerson, J.; Ingram-Smith, C. Encystation of Entamoeba Histolytica in Axenic Culture. *Microorganisms* 2021, 9, 873, doi:10.3390/microorganisms9040873.
6. Ehrenkaufer, G.M.; Weedall, G.D.; Williams, D.; Lorenzi, H.A.; Caler, E.; Hall, N.; Singh, U. The Genome and Transcriptome of the Enteric Parasite Entamoeba Invadens, a Model for Encystation. *Genome Biol* 2013, 14, R77, doi:10.1186/gb-2013-14-7-r77.
7. Wilson, I.W.; Weedall, G.D.; Lorenzi, H.; Howcroft, T.; Hon, C.-C.; Deloger, M.; Guillén, N.; Paterson, S.; Clark, C.G.; Hall, N. Genetic Diversity and Gene Family Expansions in Members of the Genus Entamoeba. *Genome Biol Evol* 2019, 11, 688–705, doi:10.1093/gbe/evz009.
8. Manna, D.; Lozano-Amado, D.; Ehrenkaufer, G.; Singh, U. The NAD<sup>+</sup> Responsive Transcription Factor ERM-BP Functions Downstream of Cellular Aggregation and Is an Early Regulator of Development and Heat Shock Response in Entamoeba. *Front Cell Infect Microbiol* 2020, 10, doi:10.3389/fcimb.2020.00363.

9. Manna, D.; Ehrenkaufer, G.M.; Singh, U. Regulation of Gene Expression in the Protozoan Parasite *Entamoeba Invadens*: Identification of Core Promoter Elements and Promoters with Stage-Specific Expression Patterns. *Int J Parasitol* 2014, 44, 837–845, doi:10.1016/j.ijpara.2014.06.008.
10. Singh, U.; Ehrenkaufer, G. Recent Insights into *Entamoeba* Development: Identification of Transcriptional Networks Associated with Stage Conversion. *Int J Parasitol* 2009, 39, 41–47, doi:10.1016/j.ijpara.2008.09.004.
11. Manna, D.; Singh, U. Nuclear Factor Y (NF-Y) Modulates Encystation in *Entamoeba* via Stage-Specific Expression of the NF-YB and NF-YC Subunits. *mBio* 2019, 10, doi:10.1128/mBio.00737-19.
12. Meenakshi; Balbhim, S.S.; Sarkar, S.; Vasudevan, M.; Ghosh, S.K. Three-amino Acid Loop Extension Homeodomain Proteins Regulate Stress Responses and Encystation in *Entamoeba*. *Mol Microbiol* 2023, 120, 276–297, doi:10.1111/mmi.15121.
13. De Cádiz, A.E.; Jeelani, G.; Nakada-Tsukui, K.; Caler, E.; Nozaki, T. Transcriptome Analysis of Encystation in *Entamoeba Invadens*. *PLoS One* 2013, 8, e74840, doi:10.1371/journal.pone.0074840.
14. Bilaud, T.; Koering, C.E.; Binet-Brasselet, E.; Ancelin, K.; Pollice, A.; Gasser, S.M.; Gilson, E. The Telobox, a Myb-Related Telomeric DNA Binding Motif Found in Proteins from Yeast, Plants and Human. *Nucleic Acids Res* 1996, 24, 1294–1303, doi:10.1093/nar/24.7.1294.
15. Lipsick JS. One Billion Years of Myb. *Oncogene* 1996, 13, 223-35.
16. Burns, C.G.; Ohi, R.; Krainer, A.R.; Gould, K.L. Evidence That Myb-Related CDC5 Proteins Are Required for Pre-mRNA Splicing. *Proceedings of the National Academy of Sciences* 1999, 96, 13789–13794, doi:10.1073/pnas.96.24.13789.
17. Dubos, C.; Stracke, R.; Grotewold, E.; Weisshaar, B.; Martin, C.; Lepiniec, L. MYB Transcription Factors in *Arabidopsis*. *Trends Plant Sci* 2010, 15, 573–581, doi:10.1016/j.tplants.2010.06.005.
18. Collier, S.E.; Voehler, M.; Peng, D.; Ohi, R.; Gould, K.L.; Reiter, N.J.; Ohi, M.D. Structural and Functional Insights into the N-Terminus of *Schizosaccharomyces Pombe* Cdc5. *Biochemistry* 2014, 53, 6439–6451, doi:10.1021/bi5008639.
19. Ogata, K.; Hojo, H.; Aimoto, S.; Nakai, T.; Nakamura, H.; Sarai, A.; Ishii, S.; Nishimura, Y. Solution Structure of a DNA-Binding Unit of Myb: A Helix-Turn-Helix-Related Motif with Conserved Tryptophans Forming a Hydrophobic Core. *Proceedings of the National Academy of Sciences* 1992, 89, 6428–6432, doi:10.1073/pnas.89.14.6428.
20. Otsuka, H.; Van Haastert, P.J.M. A Novel Myb Homolog Initiates *Dictyostelium* Development by Induction of Adenylyl Cyclase Expression. *Genes Dev* 1998, 12, 1738–1748, doi:10.1101/gad.12.11.1738.
21. Tsujioka, M.; Zhukovskaya, N.; Yamada, Y.; Fukuzawa, M.; Ross, S.; Williams, J.G. *Dictyostelium* Myb Transcription Factors Function at Culmination as Activators of Ancillary Stalk Differentiation. *Eukaryot Cell* 2007, 6, 568–570, doi:10.1128/EC.00373-06.
22. Lv, J.; Yang, T.; Yang, H.; Li, Z.; Qin, P.; Zhang, X.; Liang, X.; Li, J.; Chen, Q. Identification of Myb Genes in *Euplotes Aediculatus* May Indicate an Early Evolutionary Process. *Gene* 2013, 530, 266–272, doi:10.1016/j.gene.2013.06.072.
23. Alonso, A.M.; Schcolnicov, N.; Diambra, L.; Cóceres, V.M. In-Depth Comparative Analysis of *Trichomonas Foetus* Transcriptomics Reveals Novel Genes Linked with Adaptation to Feline Host. *Sci Rep* 2022, 12, 10057, doi:10.1038/s41598-022-14310-x.
24. da Silva, M.S.; Perez, A.M.; da Silveira, R. de C. V.; de Moraes, C.E.; Siqueira-Neto, J.L.; Freitas-Junior, L.H.; Cano, M.I.N. The *Leishmania Amazonensis* TRF (TTAGGG Repeat-Binding Factor) Homologue Binds and Co-Localizes with Telomeres. *BMC Microbiol* 2010, 10, 136, doi:10.1186/1471-2180-10-136.
25. Li, B.; Espinal, A.; Cross, G.A.M. Trypanosome Telomeres Are Protected by a Homologue of Mammalian TRF2. *Mol Cell Biol* 2005, 25, 5011–5021, doi:10.1128/MCB.25.12.5011-5021.2005.
26. Gissot, M.; Briquet, S.; Refour, P.; Boschet, C.; Vaquero, C. PfMyb1, a *Plasmodium Falciparum* Transcription Factor, Is Required for Intra-Erythrocytic Growth and Controls Key Genes for Cell Cycle Regulation. *J Mol Biol* 2005, 346, 29–42, doi:10.1016/j.jmb.2004.11.045.
27. Alzan, H.F.; Knowles, D.P.; Suarez, C.E. Comparative Bioinformatics Analysis of Transcription Factor Genes Indicates Conservation of Key Regulatory Domains among *Babesia Bovis*, *Babesia Microti*, and *Theileria Equi*. *PLoS Negl Trop Dis* 2016, 10, e0004983, doi:10.1371/journal.pntd.0004983.
28. Meneses, E.; Cárdenas, H.; Zárate, S.; Brieba, L.G.; Orozco, E.; López-Camarillo, C.; Azuara-Liceaga, E. The R2R3 Myb Protein Family in *Entamoeba Histolytica*. *Gene* 2010, 455, 32–42, doi:10.1016/j.gene.2010.02.004.
29. Hsu, H.-M.; Ong, S.-J.; Lee, M.-C.; Tai, J.-H. Transcriptional Regulation of an Iron-Inducible Gene by Differential and Alternate Promoter Entries of Multiple Myb Proteins in the Protozoan Parasite *Trichomonas Vaginalis*. *Eukaryot Cell* 2009, 8, 362–372, doi:10.1128/EC.00317-08.
30. Ong, S.-J.; Hsu, H.-M.; Liu, H.-W.; Chu, C.-H.; Tai, J.-H. Multifarious Transcriptional Regulation of Adhesion Protein Gene Ap65 - 1 by a Novel Myb1 Protein in the Protozoan Parasite *Trichomonas Vaginalis*. *Eukaryot Cell* 2006, 5, 391–399, doi:10.1128/EC.5.2.391-399.2006.

31. Huang, Y.-C.; Su, L.-H.; Lee, G.A.; Chiu, P.-W.; Cho, C.-C.; Wu, J.-Y.; Sun, C.-H. Regulation of Cyst Wall Protein Promoters by Myb2 in *Giardia Lamblia*. *Journal of Biological Chemistry* 2008, 283, 31021–31029, doi:10.1074/jbc.M805023200.
32. Sun, C.-H.; Palm, D.; McArthur, A.G.; Svärd, S.G.; Gillin, F.D. A Novel Myb-Related Protein Involved in Transcriptional Activation of Encystation Genes in *Giardia Lamblia*. *Mol Microbiol* 2002, 46, 971–984, doi:10.1046/j.1365-2958.2002.03233.x.
33. Yang, H.; Chung, H.J.; Yong, T.; Lee, B.H.; Park, S. Identification of an Encystation-Specific Transcription Factor, Myb Protein in *Giardia Lamblia*. *Mol Biochem Parasitol* 2003, 128, 167–174, doi:10.1016/S0166-6851(03)00072-0.
34. Waldman, B.S.; Schwarz, D.; Wadsworth, M.H.; Saeij, J.P.; Shalek, A.K.; Lourido, S. Identification of a Master Regulator of Differentiation in *Toxoplasma*. *Cell* 2020, 180, 359–372.e16, doi:10.1016/j.cell.2019.12.013.
35. Ehrenkauf, G.M.; Hackney, J.A.; Singh, U. A Developmentally Regulated Myb Domain Protein Regulates Expression of a Subset of Stage-Specific Genes in *Entamoeba Histolytica*. *Cell Microbiol* 2009, 11, 898–910, doi:10.1111/j.1462-5822.2009.01300.x.
36. Aurrecochea, C.; Barreto, A.; Brestelli, J.; Brunk, B.P.; Caler, E. V.; Fischer, S.; Gajria, B.; Gao, X.; Gingle, A.; Grant, G.; et al. AmoebaDB and MicrosporidiaDB: Functional Genomic Resources for Amoebozoa and Microsporidia Species. *Nucleic Acids Res* 2011, 39, D612–D619, doi:10.1093/nar/gkq1006.
37. Alvarez-Jarreta, J.; Amos, B.; Aurrecochea, C.; Bah, S.; Barba, M.; Barreto, A.; Basenko, E.Y.; Belnap, R.; Blevins, A.; Böhme, U.; et al. VEuPathDB: The Eukaryotic Pathogen, Vector and Host Bioinformatics Resource Center in 2023. *Nucleic Acids Res* 2023, doi:10.1093/nar/gkad1003.
38. Crooks, G.E.; Hon, G.; Chandonia, J.-M.; Brenner, S.E. WebLogo: A Sequence Logo Generator. *Figure 1. Genome Res* 2004, 14, 1188–1190, doi:10.1101/gr.849004.
39. Dereeper, A.; Guignon, V.; Blanc, G.; Audic, S.; Buffet, S.; Chevenet, F.; Dufayard, J.-F.; Guindon, S.; Lefort, V.; Lescot, M.; et al. Phylogeny.Fr: Robust Phylogenetic Analysis for the Non-Specialist. *Nucleic Acids Res* 2008, 36, W465–W469, doi:10.1093/nar/gkn180.
40. Ren, J.; Wen, L.; Gao, X.; Jin, C.; Xue, Y.; Yao, X. DOG 1.0: Illustrator of Protein Domain Structures. *Cell Res* 2009, 19, 271–273, doi:10.1038/cr.2009.6.
41. Bailey, T.L. STREME: Accurate and Versatile Sequence Motif Discovery. *Bioinformatics* 2021, 37, 2834–2840, doi:10.1093/bioinformatics/btab203.
42. Supek, F.; Bošnjak, M.; Škunca, N.; Šmuc, T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS One* 2011, 6, e21800, doi:10.1371/journal.pone.0021800.
43. Broccoli, D.; Smogorzewska, A.; Chong, L.; de Lange, T. Human Telomeres Contain Two Distinct Myb-Related Proteins, TRF1 and TRF2. *Nat Genet* 1997, 17, 231–235, doi:10.1038/ng1097-231.
44. Alabadi, D.; Oyama, T.; Yanovsky, M.J.; Harmon, F.G.; Más, P.; Kay, S.A. Reciprocal Regulation Between TOC1 and LHY / CCA1 Within the Arabidopsis Circadian Clock. *Science (1979)* 2001, 293, 880–883, doi:10.1126/science.1061320.
45. Ramalingam, A.; Kudapa, H.; Pazhamala, L.T.; Garg, V.; Varshney, R.K. Gene Expression and Yeast Two-Hybrid Studies of 1R-MYB Transcription Factor Mediating Drought Stress Response in Chickpea (*Cicer Arietinum* L.). *Front Plant Sci* 2015, 6, doi:10.3389/fpls.2015.01117.
46. Aasland R, S.A.G.T. The SANT Domain: A Putative DNA-Binding Domain in the SWI-SNF and ADA Complexes, the Transcriptional Co-Repressor N-CoR and TFIIIB. *Trends Biochem Sci*. 1996, 21, 87–95.
47. Boyer, L.A.; Latek, R.R.; Peterson, C.L. The SANT Domain: A Unique Histone-Tail-Binding Module? *Nat Rev Mol Cell Biol* 2004, 5, 158–163, doi:10.1038/nrm1314.
48. Carre, I.A. MYB Transcription Factors in the Arabidopsis Circadian Clock. *J Exp Bot* 2002, 53, 1551–1557, doi:10.1093/jxb/erf027.
49. Schaffer, R.; Ramsay, N.; Samach, A.; Corden, S.; Putterill, J.; Carré, I.A.; Coupland, G. The Late Elongated Hypocotyl Mutation of Arabidopsis Disrupts Circadian Rhythms and the Photoperiodic Control of Flowering. *Cell* 1998, 93, 1219–1229, doi:10.1016/S0092-8674(00)81465-8.
50. Rendón-Gandarilla, F.J.; Álvarez-Hernández, V.; Castañeda-Ortiz, E.J.; Cárdenas-Hernández, H.; Cárdenas-Guerra, R.E.; Valdés, J.; Betanzos, A.; Chávez-Munguía, B.; Lagunes-Guillen, A.; Orozco, E.; et al. Telomeric Repeat-Binding Factor Homologs in *Entamoeba Histolytica*: New Clues for Telomeric Research. *Front Cell Infect Microbiol* 2018, 8, doi:10.3389/fcimb.2018.00341.
51. Fan, Q.; An, L.; Cui, L. PfADA2, a Plasmodium Falciparum Homologue of the Transcriptional Coactivator ADA2 and Its in Vivo Association with the Histone Acetyltransferase PfGCN5. *Gene* 2004, 336, 251–261, doi:10.1016/j.gene.2004.04.005.
52. Shoji, W.; Inoue, T.; Yamamoto, T.; Obinata, M. MIDA1, a Protein Associated with Id, Regulates Cell Growth. *Journal of Biological Chemistry* 1995, 270, 24818–24825, doi:10.1074/jbc.270.42.24818.
53. Zhang, S.; Lockshin, C.; Herbert, A.; Winter, E.; Rich, A. Zuotin, a Putative Z-DNA Binding Protein in *Saccharomyces Cerevisiae*. *EMBO J* 1992, 11, 3787–3796, doi:10.1002/j.1460-2075.1992.tb05464.x.

54. Wilhelm, M.L.; Reinbolt, J.; Gangloff, J.; Dirheimer, G.; Wilhelm, F.X. Transfer RNA Binding Protein in the Nucleus of *Saccharomyces Cerevisiae*. *FEBS Lett* 1994, 349, 260–264, doi:10.1016/0014-5793(94)00683-0.
55. Yan, W. Zuotin, a Ribosome-Associated DnaJmolecular Chaperone. *EMBO J* 1998, 17, 4809–4817, doi:10.1093/emboj/17.16.4809.
56. Shrestha, O.K.; Sharma, R.; Tomiczek, B.; Lee, W.; Tonelli, M.; Cornilescu, G.; Stolarska, M.; Nierzwicki, L.; Czub, J.; Markley, J.L.; et al. Structure and Evolution of the 4-Helix Bundle Domain of Zuotin, a J-Domain Protein Co-Chaperone of Hsp70. *PLoS One* 2019, 14, e0217098, doi:10.1371/journal.pone.0217098.
57. Hu, H.-L.; Wu, C.-C.; Lee, J.-C.; Chen, H.-T. A Region of Bdp1 Necessary for Transcription Initiation That Is Located within the RNA Polymerase III Active Site Cleft. *Mol Cell Biol* 2015, 35, 2831–2840, doi:10.1128/MCB.00263-15.
58. Jiang, C.-K.; Rao, G.-Y. Insights into the Diversification and Evolution of R2R3-MYB Transcription Factors in Plants. *Plant Physiol* 2020, 183, 637–655, doi:10.1104/pp.19.01082.
59. Singh, V.; Kumar, N.; Dwivedi, A.K.; Sharma, R.; Sharma, M.K. Phylogenomic Analysis of R2R3 MYB Transcription Factors in Sorghum and Their Role in Conditioning Biofuel Syndrome. *Curr Genomics* 2020, 21, 138–154, doi:10.2174/1389202921666200326152119.
60. Millard, P.S.; Kragelund, B.B.; Burow, M. R2R3 MYB Transcription Factors – Functions Outside the DNA-Binding Domain. *Trends Plant Sci* 2019, 24, 934–946, doi:10.1016/j.tplants.2019.07.003.
61. Guehmann, S.; Vorbrueggen, G.; Kalkbrenner, F.; Moelling, K. Reduction of a Conserved Cys Is Essential for Myb DNA-Binding. *Nucleic Acids Res* 1992, 20, 2279–2286, doi:10.1093/nar/20.9.2279.
62. Hirayama, T.; Shinozaki, K. A Cdc5+ Homolog of a Higher Plant, *Arabidopsis Thaliana*. *Proceedings of the National Academy of Sciences* 1996, 93, 13371–13376, doi:10.1073/pnas.93.23.13371.
63. Thiedig, K.; Weisshaar, B.; Stracke, R. Functional and Evolutionary Analysis of the Arabidopsis 4R-MYB Protein SNAPc4 as Part of the SNAP Complex. *Plant Physiol* 2021, 185, 1002–1020, doi:10.1093/plphys/kiaa067.
64. Sun, J.; Li, X.; Hou, X.; Cao, S.; Cao, W.; Zhang, Y.; Song, J.; Wang, M.; Wang, H.; Yan, X.; et al. Structural Basis of Human SNAPc Recognizing Proximal Sequence Element of SnRNA Promoter. *Nat Commun* 2022, 13, 6871, doi:10.1038/s41467-022-34639-1.
65. Wong, M.W.; Henry, R.W.; Ma, B.; Kobayashi, R.; Klages, N.; Matthias, P.; Strubin, M.; Hernandez, N. The Large Subunit of Basal Transcription Factor SNAP c Is a Myb Domain Protein That Interacts with Oct-1. *Mol Cell Biol* 1998, 18, 368–377, doi:10.1128/MCB.18.1.368.
66. Henry, R.W.; Mittal, V.; Ma, B.; Kobayashi, R.; Hernandez, N. SNAP19 Mediates the Assembly of a Functional Core Promoter Complex (SNAP c) Shared by RNA Polymerases II and III. *Genes Dev* 1998, 12, 2664–2672, doi:10.1101/gad.12.17.2664.
67. Krishnan, D.; Ghosh, S.K. Cellular Events of Multinucleated Giant Cells Formation During the Encystation of *Entamoeba Invadens*. *Front Cell Infect Microbiol* 2018, 8, doi:10.3389/fcimb.2018.00262.
68. Reiner, D.S.; McCaffery, M.; Gillin, F.D. Sorting of Cyst Wall Proteins to a Regulated Secretory Pathway during Differentiation of the Primitive Eukaryote, *Giardia Lamblia*. *Eur J Cell Biol* 1990, 53, 142–153.
69. Galán-Vásquez, E.; Gómez-García, M. del C.; Pérez-Rueda, E. A Landscape of Gene Regulation in the Parasitic Amoebozoa *Entamoeba Spp*. *PLoS One* 2022, 17, e0271640, doi:10.1371/journal.pone.0271640.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.