

Article

# Genome-Wide Analysis of Known and Potential Tetraspanins in *Entamoeba histolytica*

Kentaro Tomii <sup>1,\*</sup> , Herbert J. Santos <sup>2</sup>  and Tomoyoshi Nozaki <sup>2,\*</sup> 

<sup>1</sup> Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology (AIST), 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan

<sup>2</sup> Department of Biomedical Chemistry, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan; hjsantos@m.u-tokyo.ac.jp

\* Correspondence: k-tomii@aist.go.jp (K.T.); nozaki@m.u-tokyo.ac.jp (T.N.)

Received: 29 August 2019; Accepted: 31 October 2019; Published: 3 November 2019



**Abstract:** Tetraspanins are membrane proteins involved in intra- and/or intercellular signaling, and membrane protein complex formation. In some organisms, their role is associated with virulence and pathogenesis. Here, we investigate known and potential tetraspanins in the human intestinal protozoan parasite *Entamoeba histolytica*. We conducted sequence similarity searches against the proteome data of *E. histolytica* and newly identified nine uncharacterized proteins as potential tetraspanins in *E. histolytica*. We found three subgroups within known and potential tetraspanins, as well as subgroup-associated features in both their amino acid and nucleotide sequences. We also examined the subcellular localization of a few representative tetraspanins that might be potentially related to pathogenicity. The results in this study could be useful resources for further understanding and downstream analyses of tetraspanins in *Entamoeba*.

**Keywords:** tetraspanin; *Entamoeba*; membrane protein; subcellular localization; intron

## 1. Introduction

The human intestinal parasite, *Entamoeba histolytica*, is a pathogen that causes amoebiasis. This disease is characterized by colitis, often manifested by symptoms such as diarrhea and dysentery. It affects millions worldwide and causes up to 73,800 deaths annually [1]. Extraintestinal complications of amoebiasis arise when the parasite establishes colonization of other organs such as the liver, leading to amoebic liver abscess formation. The virulence of this parasite is due to its ability to inflict damage to host cells and tissues, which involve processes that are directly caused or regulated by membrane proteins. Among them are the galactose and N-acetyl-d-galactosamine specific lectin [2], adhesin [3], and the lysine and glutamic acid enriched protein (KERP1) [4], involved in parasite attachment to colonic epithelial cells, and rhomboid protease [5], which contribute to immune evasion.

Tetraspanins are membrane proteins with four transmembrane helices, two extracellular domains, and three short intracellular regions. They possess a 'CCG' motif, which is part of the well-conserved disulfide bonds in their extracellular domain 2 (EC2). Tetraspanins play important roles in organizing various cell-to-cell interactions and signal pathways [6]. Some tetraspanins are capable of forming stable interactions with other proteins such as receptor and signaling proteins, including other tetraspanins, leading to the establishment of functional complexes in the plasma membrane called tetraspanin microdomains [7]. Some of the known tetraspanins are integrated to the membranes of extracellular vesicles including CD9, CD63, CD81, CD82, and CD151 [8]. Mammalian tetraspanins have also been found to be involved in immune cell migration [9], in controlling membrane architecture [10], in cancer cell invasion and metastasis [11], as well as in mediating viral and non-viral membrane fusions [12].

Among parasites, tetraspanins have been reported to be involved in tegument formation of some helminths such as *Opisthorchis viverrini* [13] and *Schistosoma mansoni* [14] and are promising candidates for vaccine development [15]. In the unicellular parasite *Trichomonas vaginalis*, the tetraspanin TvTsp1 was found to be associated to exosomes [16] while TvTsp6, localized in the flagella, is involved in sensory reception and migration during infection [17]. In addition to the reported pathogenesis-associated membrane proteins in *Entamoeba*, it is likely that tetraspanins may also play a role in infection and virulence, however they have not been well characterized so far. Thus, we performed a genome-wide analysis to elucidate and scrutinize potential tetraspanins in *E. histolytica* and in other species of *Entamoeba*. Here, we provide a refined list of tetraspanins including both known and potential ones revealed by our bioinformatic analysis of the proteome in *E. histolytica* and in four other species of genus *Entamoeba*. We analyzed the evolutionary relationships of the candidate tetraspanins, as well as examined the subcellular localization of a few representative candidates. The list and results in this study could be useful resources for further understanding and analyses of tetraspanins in *Entamoeba*.

## 2. Materials and Methods

### 2.1. Sequence Data of *Entamoeba histolytica* and Other *Entamoeba*

As a reference of amino acid sequences, we used the proteome of *Entamoeba histolytica* (Strain: ATCC 30459/HM-1: IMSS; Proteome ID: UP000001926), consisting of 7959 proteins, provided by UniProt [18]. In this study, we regarded eight proteins (see Results and Table 1) reported in previous studies [19,20] as “known” *E. histolytica* tetraspanins. Nucleotide sequences, including introns, of tetraspanin genes in *E. histolytica* and in four other *Entamoeba* species (i.e., *E. dispar*, *E. invadens*, *E. moshkovskii* and *E. nuttalli*) and amino acid sequences of tetraspanins in other *Entamoeba* were retrieved from AmoebaDB (41 Released) [21].

For the four tetraspanin sequences in *E. histolytica*, we used the alternative sequences (UniProtKB accession numbers (ACs): A0A175JUA3 (for TSPAN3), A0A175JDI1 (for TSPAN11), A0A175JRI8 (for TSPAN6) and A0A175JZG7 (for TSPAN17)), instead of the original ones as they likely lack their N-terminal portions. We used C4LYT6 (for TSPAN10) whose length is 197 amino acids, instead of A0A175JJX9 (191 amino acids), and used A0A175JXX9 (for TSPAN13) whose second residue is Ala, instead of C4M9J5 (for EHI\_107790) whose second residue is Pro. The information on these entries can be found in Table 1.

### 2.2. Computational Analysis of Sequence Data

To detect uncharacterized but potential tetraspanins in *E. histolytica*, we performed similarity searches using the FASTA36 package [22] with an efficient substitution matrix, MIQS [23,24], which outperforms conventional matrices in terms of both detection sensitivity and alignment quality. For amino acid sequence comparisons, we used *ssearch36* with an expectation threshold of 0.001. Filtering low-complexity regions of amino acid sequences by the PSEG program [25] was done using the University of Virginia (UVA) FASTA server [26]. To compare amino acid sequences to the nucleotide sequences, we used *tfastx36/tfasty36* in the package.

We used FAMSA (ver. 1.2.5) [27], which also used MIQS as a default matrix, to construct multiple sequence alignments of tetraspanins. We performed the neighbor-joining method using the JTT substitution model with the “estimate” option for heterogeneity among sites, at the MAFFT server [28], based on a multiple sequence alignment to estimate a phylogenetic tree. We used, as an outgroup protein, a putative trans 2,3-enoyl-coa reductase, C4LTB8 (CL6EHI\_045030) with four transmembrane helices. For phylogenetic tree calculations, we used only the gap-free regions (151 residue positions in total) found in the multiple sequence alignment. We resampled 1000 times for calculating the neighbor-joining bootstrap support values. We used RNAfold (Version 2.4.8) [29] with a default setting to predict the secondary structure of intronic RNAs.

### 2.3. Amoeba Culture and Plasmid Construction

Trophozoites of *Entamoeba histolytica* HM-1: IMSS clone 6 [30] were axenically cultured using Diamond BI-S-33 medium [30]. Genomic DNA was isolated by first washing the cells with 1 × phosphate-buffered saline (PBS), followed by lysis using cell lysis buffer C1 (Qiagen, Hilden, Germany). The resulting lysate was centrifuged at 1300× *g* for 10 min at 4 °C. The pellet containing nuclei was washed and reacted in ice cold buffer C1 before centrifugation at 1300× *g* for 10 min at 4 °C. Buffer G2 (Qiagen, Hilden, Germany) was added to the washed pellet and thoroughly mixed by inversion, followed by treatment with Protease K (Qiagen, Hilden, Germany) and RNase (Qiagen, Hilden, Germany). The mixture was incubated for 1 h at 50 °C. Then, 1:1 TE-saturated phenol–chloroform was added followed by gentle mixing by inversion. The resulting mixture was incubated for 5 min on ice. Afterwards, the mixture was centrifuged at 8600× *g* for 10 min at 4 °C. The upper phase was transferred to a new tube and reacted with chloroform. The mixture was incubated for 2 min on ice, then centrifuged at 8600× *g* for 10 min at 4 °C. The upper phase was collected and treated with Pellet Paint (Novagen, Madison, WI, USA), 3 M sodium acetate (pH 5.2) and ethanol to precipitate DNA. The genomic DNA pellet was washed with 70% ethanol. After removal of residual ethanol, the pellet was dissolved in TE buffer before storage at −80 °C.

All polymerase chain reaction amplifications were performed using Tks Gflex DNA polymerase (Takara, Shiba, Japan). Primer sets used are summarized in Table S1. Amplified DNA fragments were digested by *Bgl*III (New England Biolabs, Beverly, MA, USA), and ligated into pEhEx-HA [31] using Ligation-Convenience Kit (Nippongene, Tokyo, Japan). Transfection by lipofection, selection, and maintenance of amoebic transformants were performed as previously described [32].

### 2.4. Immunoblot Analysis and Immunofluorescence Assay (IFA)

Trophozoites of representative hemagglutinin (HA)-tagged tetraspanin-expressing strains were lysed, and approximately 20 µg proteins were separated by denaturing sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) [33]. A strain transfected with the pEhEx-HA empty vector was used as control. Proteins were transferred to polyvinylidene difluoride (PVDF) membranes and were stained with anti-HA monoclonal antibody (16B12, Biolegend, San Diego, CA, USA) and anti-cysteine synthase 1 (CS1; an enzyme involved in sulfur-containing amino acid metabolism) antiserum (loading control) as described previously [34]. IFA was performed as previously described [32], with some modifications. Cells were double-stained with mouse anti-HA antibody and one of the following rabbit antisera; anti-adenosine-5′-phosphosulfate kinase (APSK) [35] as mitosome marker, anti-binding immunoglobulin protein (BiP) [36] as ER marker, and anti-vacuolar protein sorting-associated protein 26 (Vps26) [37] as endosome marker, diluted 1:500 (except for anti-APSK antiserum, which was diluted 1:300), respectively in 0.2% saponin and 1% bovine serum albumin in PBS. Samples were visualized using a confocal laser scanning microscope, LSM780 (Carl Zeiss Microscopy, Oberkochen, Germany).

## 3. Results

### 3.1. Identification of Potential Tetraspanins in *Entamoeba histolytica*

Two articles which analyzed evolutionary relationships of tetraspanins including ones in *E. histolytica*, have been published so far [19,20], both of which reported six candidate *E. histolytica* proteins as tetraspanins. Among those proteins, two [C4LUK2 (TSPAN1 in this study) and C4M992 (TSPAN4)] were reported in both studies. One of the proteins used in the former study [19], EAL45935 (EHI\_139370; UniProtKB AC: C4M6W3) is a high mobility group (HMG) box domain containing protein, according to the annotation provided by UniProt, and contains only a single cysteine residue. In addition, we could not find any statistically significant similarity between the protein and a plausible tetraspanin in *E. histolytica*. Thus, we did not regard C4M6W3 as a candidate tetraspanin in this study. One of the proteins reported in the later study [20], EHI5A\_058850 (EHI\_127260; UniProtKB AC:

C4LWZ8) is a Plecksin homology domain protein, according to the annotation by UniProt, and lacks the CCG motif. Thus, we also excluded this protein as a candidate tetraspanin in *E. histolytica*, though we found weak but significant similarity between this protein and C4M9D7 (TSPAN12).

Using *sssearch36* with the MIQS matrix and the latest proteome of *E. histolytica*, we could identify nine additional uncharacterized proteins as potential tetraspanins in *E. histolytica* in addition to the eight remaining proteins derived from the previous studies. Table 1 lists the refined set of 17 proteins including both known and potential tetraspanins in *E. histolytica*. The lengths of the proteins are around 200 (or more) amino acids and are similar to each other (191 to 225 amino acids), except for two, TSPAN5 (259 aa) and TSPAN17 (254 aa). The proteins listed in Table 1 have been annotated to possess four transmembrane helices by Phobius [38] in Pfam [39], except for five proteins, TSPAN3, 6, 11, 13, and 17. However, they are not registered as a member of the tetraspanin family (PF00335) in Pfam, except for one, TSPAN4 (/C4M992).

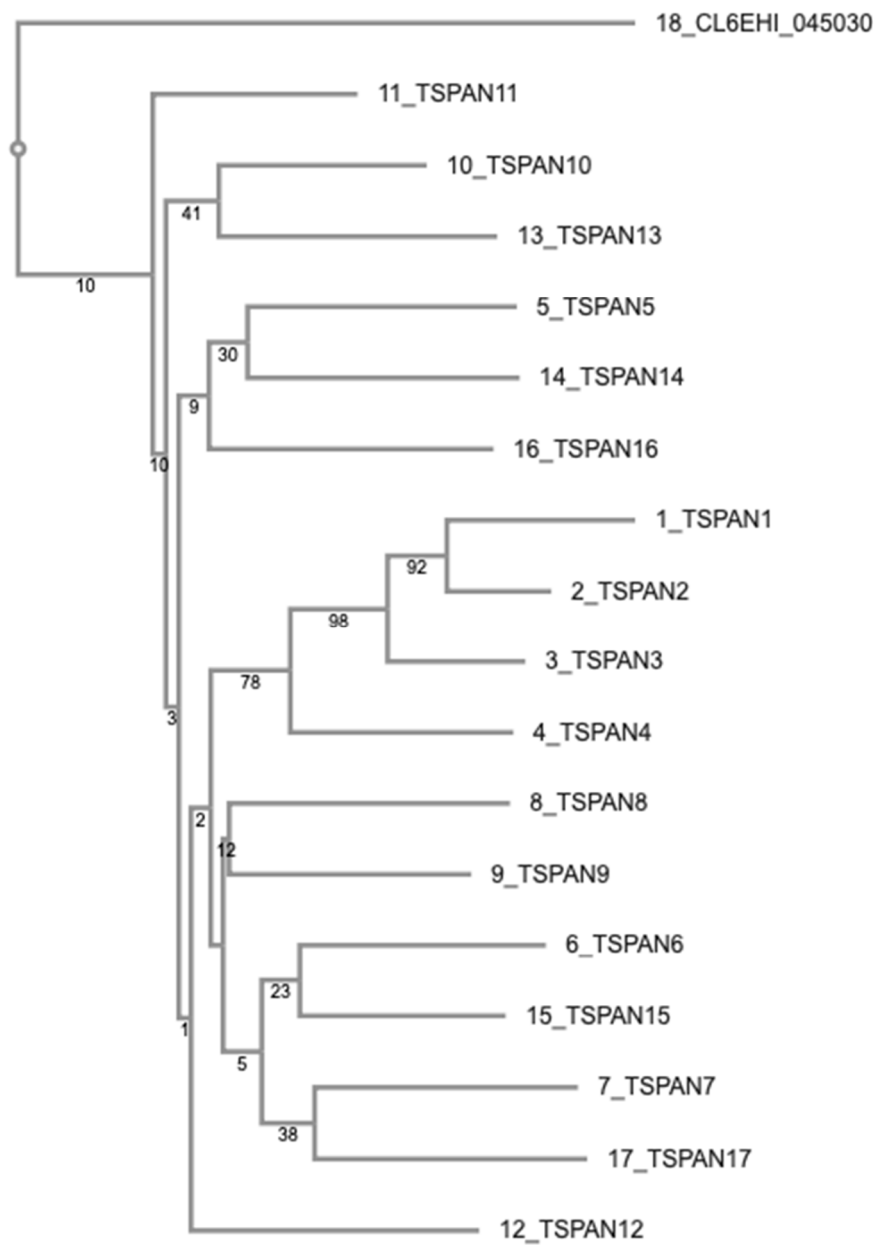
**Table 1.** A list of known and potential tetraspanins in *E. histolytica*.

Gene Name(s)	Entry Name (UniProtKB)	Length (aa)	in Previous Studies <sup>1</sup>	Names in This Study
EHI_010390	C4LYT6_ENTHI	197	–	TSPAN10
EHI_016390 (CL6EHI_016390)	C4M3W1_ENTHI	196	EHI5A_042920	TSPAN15
EHI_022890 (CL6EHI_022890)	C4LUK2_ENTHI	206	EAL48097 EHI5A_127460	TSPAN1
EHI_040610 (CL6EHI_040610)	C4M5N8_ENTHI	213	–	TSPAN16
EHI_055360 (CL6EHI_055360)	C4M9Q1_ENTHI	222	EHI5A_158290	TSPAN14
EHI_075690 (CL6EHI_075690)	C4M992_ENTHI	218	EAL43389 EHI5A_106540	TSPAN4
EHI_091490 (CL6EHI_091490)	C4M9D7_ENTHI	225	–	TSPAN12
EHI_097600 (CL6EHI_097600)	C4M931_ENTHI	209	–	TSPAN9
CL6EHI_107790	A0A175JXX9_ENTHI	195	EHI5A_153470	TSPAN13
EHI_133990 (CL6EHI_133990)	C4M1M8_ENTHI	200	–	TSPAN7
CL6EHI_139380	A0A175JUA3_ENTHI	218	–	TSPAN3
EHI_147520 (CL6EHI_147520)	C4M6M8_ENTHI	259	EAL49140	TSPAN5
CL6EHI_152810	A0A175JDI1_ENTHI	201	–	TSPAN11
EHI_164840 (CL6EHI_164840)	C4M1X7_ENTHI	191	–	TSPAN8
CL6EHI_168280	A0A175JRI8_ENTHI	193	–	TSPAN6
EHI_174220 (CL6EHI_174220)	C4M269_ENTHI	208	EAL47116	TSPAN2
CL6EHI_198000	A0A175JZG7_ENTHI	254	(EAL42794)	TSPAN17

<sup>1</sup> IDs used in the two previous studies [19,20].

### 3.2. Possible Subgroups of *Entamoeba Tetraspanins*

We tried to analyze evolutionary relationships of the 17 proteins in Table 1. Most of them are highly diverged so that evolutionary relationships among them are obfuscated. Figure 1 shows a neighbor-joining tree based on a multiple sequence alignment of the 17 proteins and C4LTB8 (CL6EHI\_045030), as an outgroup, calculated by FAMSAs (Figure 2). Although the neighbor-joining bootstrap support values are not so high in most cases, we were able to recognize, to some extent, characteristic residues that appear to be associated within certain subgroups (Table 2).



**Figure 1.** Phylogenetic tree of 17 tetraspanins in *E. histolytica*. Bootstrap support values are placed at each branch point.

```

TSPAN1 (1) -----MKVIOFIVLGGWIFTLIAAGLWIGVVFVGP-IRK--M
TSPAN2 (1) -----MEKQRIYNLSLLVLGVVVVLASTALIFLTFGP-IRK--M
TSPAN3 (1) -----MFSLSKFLLLITGIIIVLSGAALFVLIYIGP-VHN--M
TSPAN4 (1) -----MVSKLELSTTIAAAVYVLIIGFFLIVASIVFT--KD-IN
TSPAN5 (2) -----MKVQIAGLIYQFSIKKIHDITLCIGIAFFLFSLHVT-IFH
TSPAN6 (2) -----MNTSNTLCFLLIFFLYFICGAMVALLGLTRY-NTG---
TSPAN7 (2) -----MAHTPLLLLSLLGIIILCCFFPIQFYFNL-----
TSPAN8 (2) -----MNYSTFSIIQLISTAVVIGIEVYAI-IYC---
TSPAN9 (2) -----MGEHQIRITPRVYFSLLIIFMALYSIFIGTTIY-SNG---
TSPAN10 (3) -----MIPSRQQILKYINISIIIGIFLVFHGITNALFSK---
TSPAN11 (3) -----MATGFSNFTIIVIIISSITATVGFIMI-GHG-CI
TSPAN12 (3) -----MPSTKRAFSTTSIFLFTLGTVGLLISVVLV-----V
TSPAN13 (3) -----MPFGITIIIVIGVGMGLYLDKLFEL-----
TSPAN14 -----MIIFPTIVVTLISMFSYSMIYQSSLLIG-QYGDII
TSPAN15 -----MKNKQSKKILCISQIVMSLLFLLIGLITS-----
TSPAN16 MNVFQNESIVIQFIETVIDRYKQGNQRDKILFIGEIIISIMSLFYLVYAILLP-----
TSPAN17 -----MEKSRIFGICFGIFLIISGISVEIVSAFGG-----
    
```

### EC1

```

TSPAN1 (1) EQNNTLAKSS---LNYSRNTFVVFCTVAALMGLLGIILCVYSLIPKPTMGIVSY-----
TSPAN2 (1) PNTNKLNSSD---LKFYSKTFVFLIMISLAVLFMVAGGMMIATSFVQKDWMCLLSY-----
TSPAN3 (1) SELKNIKSSNH--LYMSKRTYQIEMMACGGIAAFVGIISILAVFIEQKWLRFIT-----
TSPAN4 (1) AFARSAEGKD-GDLTYFTKASCVYLIVAGAICMLSGSLGMVGALEKQWSMMLYF-----
TSPAN5 (2) VQYNSLPHVQ-----SLTTIVTCSFLQAI I I LSSLIQILHSTHNIPTSMPFHY-----
TSPAN6 (2) --YNL--SPA----VHHPLSIGILMILIGISMCLASVSTIISLAKKNIAGIISL-----
TSPAN7 (2) -LDPDIEFNE-----KISSLFISVGCIFG-ILLFIST-TIVSCCFNKGKICTIIG-----
TSPAN8 (2) --YNIHSAD----SIYPEKYPVLYLYFSMIGAI FLNIPSLLTGILQSKIASIFAI-----
TSPAN9 (2) --TTQQPSKK----LPSKYMFLTNVIVGICEIFIGLFGIIAFSVRGKYFLILL-----
TSPAN10 (3) -----HPEN----ELSSKSVTHSSLITGIIIVISYSIYCFKQYINYNKKEMYLLI-I--
TSPAN11 (3) CFYNQISSKH----ALGDYSVSSLSLICGVVAVIVAIIVFYLYQIHSSPSFIIAF-----
TSPAN12 (3) LLKNSMHFIP----SGIKWILSFSICGSSVALITSLIGIYSGFQLKKIFIIYI-----
TSPAN13 (3) --FSILSYS-----FSIYLYAGS-ALFLLSISAFIHAVAKYNKPVLIIL-----
TSPAN14 VGHRNNDAMP----FALNTMAICFVFILALVLFVYICISII LAHSCDNKVITLLSI-VCS
TSPAN15 -----IKYMK-----QDFIGIRLISVGFISFIVTGIIGVISGYFMSDTPYKPKFS-----
TSPAN16 -ILQDLYYHD----FYFNECAFIVVATITG---LIFACIRLIATSLDNKMLGKICS-----
TSPAN17 -FNKIIDSCK--SMIKYSKTFFRGFNYAIGASLFCIGILLIIASCLMNGFFINGGG-----
    
```

### EC2

```

TSPAN1 (1) ---IGLALAVISALATVGYFGYT---YFHMRNAT--NAS---MSSSLALWEEKFGCCGW
TSPAN2 (1) ---IVIGLGMCAFSEFGVYFAVT---YGSFVKTT--NTS---LKT-LAHWEDIFDCCGW
TSPAN3 (1) ---IIYVLLGLTALGMMGLYGYS---YFLLTAS--DST---MSSHLTQYENAFQCCGW
TSPAN4 (1) ---MTLIVSLITWVLEGVFGMI---TFGLTAAS--EAT---MSSTIATWELRFHCCGW
TSPAN5 (2) ---FMLFVRLISLTSIILIV---SLHQFNKIN--DSS-----HHLSLIESQYSCCGW
TSPAN6 (2) ---SLFVLISIMLTVQAIMCGID---YTVLKVD--KIE----EYDLQVIEEQFNCCGW
TSPAN7 (2) -----SILVLLIGGFIIISQT---MMKQNDFS-EDIN-----GLHNIELKSNCCGW
TSPAN8 (2) ---IFDIIAIVLLFIVYIIGQI---EQLGVSSVS--EQD-----KITEIEKEANCCGW
TSPAN9 (2) ---VGVIISTFLLFIITSFYGIT---SIRFDYLDTM--GSD-----NMITIEYENCCGW
TSPAN10 (3) ---SIGCLIIMYCILSIFTNSI---TYSINSII--KKE-----QLIGVENEFECCGW
TSPAN11 (3) ---SLFITLLMSTVTFIAFI---DTYSVRSLV--KEN----TSKMSLEYEFECGCGY
TSPAN12 (3) -GGCVVVTILFLGTTVTSFIGVA---FPDWSLTHFW-NGLN----SDQKAFTEVAYRCCGW
TSPAN13 (3) LCACIFVALIELWTLSTAYQNM L--HTPKCMENLY--DST---TKGIRREINLKFACCGW
TSPAN14 ILTSLLSILVIILLTVSIIINTKKSFP SFLSTEGESEQDTNEEGKKEFRNQLEYKFSCCGF
TSPAN15 ---ILLECFTVLFIVLIITSIR---LIYNVKKVN--DQS-----LQLHSVERNFKCCGW
TSPAN16 ---FIPIIILYMLVALLSINSQR---YITSHGLI--SPN-----NLVQVEFKHCCGW
TSPAN17 -----IINVLFILFVIGSISYFS---YDLRQLF--TMD-----EEKAKPFQIPYDCCGW
    
```

Figure 2. Cont.

```

TSPAN1 (1) DVYR---QKP-LCASDVGSL-NPNMTCKIKTQPLRNASLQNFVAVGCFE---IVVLFVS
TSPAN2 (1) NEYR---SM-PPCASASGS--TSRITCKTVTQPIAGASLQNLIVGFSE---FLLLVVG
TSPAN3 (1) KTSR---DF-PPCDSAVGY--QQQLTCYNATQKVRASLQNFISGGVY---LLLITLL
TSPAN4 (1) KTLF-ADASK---CAYPAAF--SQGHTCIQEFNSIWNLSIAGFAVVAAF---SIFAIVQ
TSPAN5 (2) FSIN---TM--TCKSLAAI--RADRTCSSTIGVSFHNVLSIALVLNIV---EIIIVSLSH
TSPAN6 (2) KDIR---QN---CISDIGL--LMKQNCFETVGKNVISMFGIDTFISVL---FLVFT
TSPAN7 (2) KSIN---LDG---CYAKN-IT-DSIKSCYDEIGYPIQLIFIHFLVFVYMC---YVIYIFIT
TSPAN8 (2) KTLK---IEG---CQAKD-I--SKAVPCYKIVGEPFLESVNFAIEIDGFA-MKAIVLLIC
TSPAN9 (2) KTLK---LYG---C--GSSI--DSQETCYSLTGVGYEKLLKSTFSNLLF---QLLFFVYF
TSPAN10 (3) KEPR---NY---CSSK--I---LPTCYHVLITPHNT-YKTITSLITL---LEIIHIVI
TSPAN11 (3) KSQR---EY---CQYLHKP-----TCYSRIVKPFMT-FRMLCSIAFF---SLLIETIS
TSPAN12 (3) KAYS---QN--DCAYYPIP-GEVQAPCYAKLYFPYNGFIIMTLILGAL-ASVMCVSFIG
TSPAN13 (3) KSYN---ANS---CSFTSTD---RKVACYDEMIYPINVNYIMMIVLITL---ATLFLLLFI
TSPAN14 DNAT---VQY--ECHKFTPL---ELTCKTTLYEPFMKLMYYLFPVISA---TLILCVII
TSPAN15 NHYY-----DCNVTPR---GRNCYKAVGIGLKS-LGYMSIIFLL---LILFIES
TSPAN16 LASN---VDN---CGAIPKT---STTCFERLVHPLNSFFKHNAIRCSF-----LCVFE
TSPAN17 KLTNITLNN-STCQSPIGI--KTGKTCFQVLKDIYADYQEIIVMMYS-----IVLILIS

```

```

TSPAN1 (1) ---AVMIVIL-----QQKTKDTTGGYTPV-----
TSPAN2 (1) ---IIFIYLI-----GG-SLNSNAGYNKL--DQ-----
TSPAN3 (1) ---IIFIAIV-----TFCSGEVSNSYSKV--DGEENVSDVID-----
TSPAN4 (1) ---MVLACTL-----YRYLDNLFYSYL--NNQHGDNKDE-----
TSPAN5 (2) ---LFISYISHR---IITQODEETENPDDI--ISNYLTTNQTIDINEETILLNNSFSK
TSPAN6 (2) ---FFLIFVS-----YQSLKTKPKQA--HVN-----
TSPAN7 (2) ---LLLIICS-----STSSSNKQNL---QRTEYSNNENLKED-----
TSPAN8 (2) ---VTLVFII-----FHRNQSKVKQA-----
TSPAN9 (2) ---LIEVI-----GELNQLRNKQKEDDL--YAEPFIDPDNE-----
TSPAN10 (3) ---LLIGTCI-----SYKPTKRKEDF--YSMYTNIGEK-----
TSPAN11 (3) ---FGIFLCL-----WEDGLEAQPQDMI---TLHISRVD-----
TSPAN12 (3) ---MAVYFII-TFKVGNSSSSAASKKDDY-QRVVDEFSENIEE-----
TSPAN13 (3) ---VVLFFQL-----FTFPAFPAPAKDKNT---EPLME-----
TSPAN14 ---TCINLTL-----FMYQRKEYRHDEW---GDTFYS-----
TSPAN15 ---LLFCL-----YWI TKSNAHLEDQ---ETDPLLPNEDDLAPF-----
TSPAN16 ---MVIALCM-----YFFTKDDFIESQY---EWTN-----
TSPAN17 ESSYRLLNETL-----QQQNTNQEVSI LPRITDEITDNNEEYANDS--NDNINSP

```

```

TSPAN1 (1) -----
TSPAN2 (1) -----
TSPAN3 (1) -----
TSPAN4 (1) -----
TSPAN5 (2) QHHVNRPLSIYNTTESDNDNDNSNLI
TSPAN6 (2) -----
TSPAN7 (2) -----
TSPAN8 (2) -----
TSPAN9 (2) -----
TSPAN10 (3) -----
TSPAN11 (3) -----
TSPAN12 (3) -----
TSPAN13 (3) -----
TSPAN14 -----
TSPAN15 -----
TSPAN16 -----
TSPAN17 LLINSQEGITQRISESTTTNSEDYLV

```

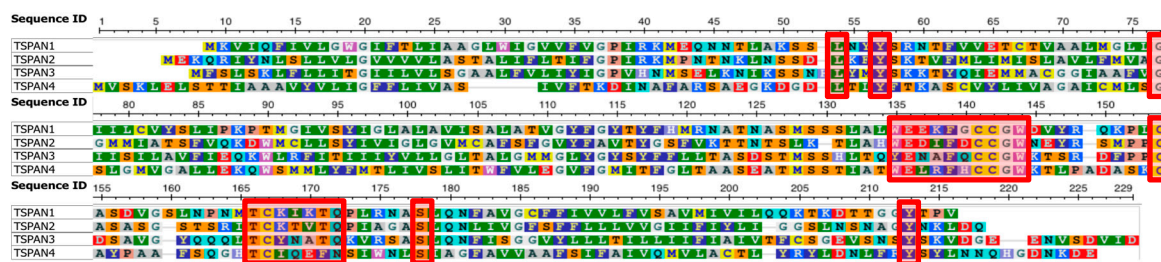
Figure 2. Multiple sequence alignment of 17 tetraspanins in *E. histolytica*. Putative membrane-spanning regions provided by UniProt are in green. The CCG motifs are highlighted in yellow. Locations of EC1 and EC2 are displayed respectively. Numbers in parentheses indicate subgroup numbers.

**Table 2.** Subgroups and motifs of 17 tetraspanins in *E. histolytica*.

TSPAN Name	Entry Name (UniProtKB)	Sequence Motifs					Subgroup
		IE	FxCCG	CCGW	TC	CY	
TSPAN1	C4LUK2_ENTHI	×	○	○	○	×	1
TSPAN2	C4M269_ENTHI	×	○	○	○	×	1
TSPAN3	A0A175JUA3_ENTHI	×	○	○	○	○	1
TSPAN4	C4M992_ENTHI	×	○	○	○	×	1
TSPAN5	C4M6M8_ENTHI	○	YSCCG	○	○	×	2
TSPAN6	A0A175JRI8_ENTHI	○	○	○	NC	×	2
TSPAN7	C4M1M8_ENTHI	○	SNCCG	○	SC	○	2
TSPAN8	C4M1X7_ENTHI	○	ANCCG	○	PC	○	2
TSPAN9	C4M931_ENTHI	○	NNCCG	○	○	○	2
TSPAN10	C4LYT6_ENTHI	×	○	○	○	○	3
TSPAN11	A0A175JDI1_ENTHI	×	○	CCGY	○	○	3
TSPAN12	C4M9D7_ENTHI	×	YRCCG	○	PC	○	3
TSPAN13	A0A175JXX9_ENTHI	×	○	○	AC	○	3
TSPAN14	C4M9Q1_ENTHI	×	○	CCGF	○	×	–
TSPAN15	C4M3W1_ENTHI	×	○	○	NC	○	(3?)
TSPAN16	C4M5N8_ENTHI	×	HHCCG	○	○	×	–
TSPAN17	A0A175JZG7_ENTHI	×	YDCCG	○	○	×	(2?)

×: absent, ○: present.

The result of our phylogenetic analysis suggests that there are several clades of tetraspanins in *E. histolytica*, one of which appears to have diverged relatively recently from others. This subgroup (subgroup 1) consists of TSPAN1, TSPAN2, TSPAN3, and TSPAN4. These four proteins are relatively conserved within this subgroup, compared with the other tetraspanins in *E. histolytica*. We found conserved residues associated with this subgroup (Figure 3). Among the four proteins, “ΩExxFxCCGW–C–TCxxx [Q/N]”, where Ω and x represent aromatic and any types of amino acids respectively, around the CCG motif in EC2 is completely conserved. In addition, Leu and Tyr in extracellular domain 1 (EC1), Gly in the second transmembrane helix, Ser at the end of EC2, and Tyr in the C-terminal intracellular regions, are uniquely conserved among the four proteins. These characteristics of having shared residues in proteins belonging to subgroup 1 are almost conserved among orthologous proteins in four other *Entamoeba* species (i.e., *E. dispar*, *E. invadens*, *E. moshkovskii* and *E. nuttalli*) (Figure S1), thus, they might have implications to their function.



**Figure 3.** Multiple sequence alignment of subgroup 1 tetraspanins in *E. histolytica*. Conserved residues and regions mentioned in the main text are indicated with red boxes. The figure was generated using the NCBI MSA Viewer [40].



Characteristic residues are also found in the other subgroups of *Entamoeba* candidate tetraspanins. We defined “Subgroup 2”, comprising TSPAN5, TSPAN6, TSPAN7, TSPAN8, and TSPAN9 (Table 2), based on the uniquely conserved sequence in their EC2. The proteins in subgroup 2, as well as their orthologous proteins in *E. dispar*, *E. invadens*, and *E. nuttalli*, possess an almost conserved sequence of “IE<sub>xxx</sub> [N/S]CCGW–C–C<sub>xxxx</sub>G” in their EC2, though some proteins in *E. nuttalli* have slightly different sequences from this pattern (Figure S2). Although TSPAN15 and TSPAN17 are in the same clade with the proteins in subgroup 2 based on the phylogenetic tree (Figure 1), they lack this sequence pattern. Moreover, TSPAN15 might be related to subgroup 3, as discussed in Section 3.3 below. TSPAN5 is not in the same clade with proteins of this subgroup, however, it possesses the same sequence pattern and its gene contains no introns (see Section 3.3). Thus, we tentatively regarded TSPAN5 as a member of subgroup 2.

A third subgroup named “Subgroup 3” consisting of four proteins, TSPAN10, TSPAN11, TSPAN12, and TSPAN13, was formed because these proteins do not possess “IE” but instead share “CY” in EC2. The four proteins in subgroup 3, and their orthologs in the other four species of *Entamoeba* share a completely conserved segment in their EC2, with the following sequence “[E/N]xx[F/Y]xCCGΩK–C–CYxx[I/L/M]xxP”. Also charged residues tend to be clustered in the C-terminal intracellular regions among the subgroup 3 proteins (Figure S3).

### 3.3. Gene Structure of Tetraspanins in *Entamoeba histolytica*

Out of the 17 genes which potentially encode tetraspanins in *E. histolytica*, 10 were identified to contain intron(s). Among them, eight genes consist of one intron and two exons, while two genes have two introns and three exons. All 12 introns in the 10 genes have 5’GT/AG3’ splice sites, as well as the strongly conserved 5’ sequence, 5’GTTTG [41]. These gene structures are almost conserved within *Entamoeba*, except for *E. invadens* (Table 3 and Discussion Section).

**Table 3.** Orthologous relationships and gene organizations of tetraspanins in *Entamoeba*.

Gene Name(s)/TSPAN Name	#Intron/Length (bp)	<i>E. dispar</i>	<i>E. invadens</i>	<i>E. moshkovskii</i>	<i>E. nuttalli</i>	Subgroup
		#Intron/Length (bp), Gene Name, Length (aa)				
EHI_022890 (CL6EHI_022890) TSPAN1	0/–	0/– EDI_248410 (206)	1/43 EIN_226110 (212)	0/– EMO_114210 (206)	0/– ENU1_002880 (206)	1
EHI_174220 (CL6EHI_174220) TSPAN2	1/53	1/51 EDI_018160 (208)	1/141 EIN_092430 (154)	1/56 EMO_051310 (208)	1/51 ENU1_208670 (208)	
CL6EHI_139380 TSPAN3	0/– <sup>1</sup>	1/55 EDI_284040 (219)	1/42 EIN_227240 (189)	1/67 EMO_007720 (222)	0/– ENU1_174960 (184)	
EHI_075690 (CL6EHI_075690) TSPAN4	0/–	0/– EDI_076910 (218)	1/38 EIN_371310 (217)	0/– EMO_068760 (218)	0/– ENU1_204710 (218)	
EHI_147520 (CL6EHI_147520) TSPAN5	0/–	0/– EDI_063570 (259)	0/– EIN_164970 (287)	0/– EMO_039780 (269)	0/– ENU1_095870 (259)	2
CL6EHI_168280 TSPAN6	1/45 <sup>2</sup>	0/– EDI_252640 (172)	0/– EIN_169070 (199)	1/42 EMO_099170 (193)	0/– ENU1_172260 (172)	
EHI_133990 (CL6EHI_133990) TSPAN7	0/–	–	0/– EIN_052950 (188)	0/– EMO_052920 (206)	0/– ENU1_202640 (201)	
EHI_164840 (CL6EHI_164840) TSPAN8	0/–	0/– EDI_282680 (191)	3/21, 77, 57 EIN_469620 (202)	0/– EMO_010630 (193)	0/– ENU1_142790 (191)	
EHI_097600 (CL6EHI_097600) TSPAN9	1/64	1/64 EDI_032730 (209)	1/171 EIN_228410 (179)	1/46 EMO_092880 (209)	1/64 ENU1_056490 (209)	

Table 3. Cont.

Gene Name(s)/TSPAN Name	#Intron/Length (bp)	<i>E. dispar</i>	<i>E. invadens</i>	<i>E. moshkovskii</i>	<i>E. nuttalli</i>	Subgroup
		#Intron/Length (bp), Gene Name, Length (aa)				
EHI_010390 TSPAN10	1/46	1/47 EDI_255180 (197)	–	1/44 EMO_033170 (199)	1/47 ENU1_058930 (197)	3
CL6EHI_152810 TSPAN11	1/55 <sup>3</sup>	1/48 EDI_093040 (176)	2/83, 80 EIN_153280 (201)	2/52, 44 EMO_021620 (201)	1/54 ENU1_015090 (176)	
EHI_091490 (CL6EHI_091490) TSPAN12	2/50, 53	2/50, 54 EDI_217940 (226)	3/70, 148, 75 EIN_153850 (240)	1/44 EMO_103930 (192)	2/49, 53 ENU1_017640 (225)	
CL6EHI_107790 TSPAN13	1/54 <sup>4</sup>	1/54 EDI_235170 (215)	2/261, 59 EIN_085580 (215)	1/48 EMO_090110 (215)	1/54 ENU1_021210 (215)	–
EHI_055360 (CL6EHI_055360) TSPAN14	2/48, 55	1/55 EDI_044480 (209)	0/– EIN_224490 (222)	2/61, 40 EMO_001580 (222)	2/48, 58 ENU1_087800 (222)	
EHI_016390 (CL6EHI_016390) TSPAN15	1/59	1/54 EDI_187220 (196)	1/94 EIN_328370 (199)	1/93 EMO_005920 (143)	1/57 ENU1_177210 (196)	(3?)
EHI_040610 (CL6EHI_040610) TSPAN16	0/–	0/– EDI_013520 (216)	1/96 EIN_152150 (221)	0/– EMO_033580 (222)	0/– ENU1_060540 (213)	–
CL6EHI_198000 TSPAN17	1/59 <sup>5</sup>	1/60 EDI_157940 (254)	1/41 EIN_371570 (245)	1/50 EMO_050060 (249)	1/59 ENU1_018010 (214)	(2?)

<sup>1</sup> Based on EHI\_139380, <sup>2</sup> Based on EHI\_168280 (located in the 5' UTR), <sup>3</sup> Based on EHI\_152810, <sup>4</sup> Based on EHI\_107790, <sup>5</sup> Based on EHI\_198000.

The genes encoding proteins of subgroup 3 tend to possess more introns than others. In fact, EHI\_091490 (TSPAN12) has two introns and three exons. Some orthologous genes of EHI\_152810 (TSPAN11) and EHI\_107790 (TSPAN13) in other *Entamoeba* also have two introns and three exons, though EHI\_152810 and EHI\_107790 only have a single intron. Note that alternative sequences of these genes are not registered in AmoebaDB, (CL6EHI\_152810/TSPAN11 and CL6EHI\_107790/TSPAN13) but are deposited in UniProtKB are longer. Hence, it is likely that these two genes also have at least one more intron in their longer forms. The rest of subgroup 3, EHI\_010390 (TSPAN10), and its orthologous genes in *Entamoeba* have one intron and two exons.

We also found sequence similarity between introns of subgroup 3 genes in *E. histolytica*. The second intron of EHI\_091490 (TSPAN12) and the intron of EHI\_107790 (TSPAN13) show over 70% identity at the corresponding position, i.e., at N-terminal regions of the third transmembrane helix, of both TSPAN12 and TSPAN13 (Figure 4). In addition, we can see similar relationships between EHI\_010390 (TSPAN10) and EHI\_016390 (TSPAN15). Their introns, which show sequence similarity, are located at the same position encoding three/two amino acids before the CCG motif in both genes (Figure 5), though TSPAN15 is in the same clade with proteins in subgroup 2 in the phylogenetic tree. These results suggest relatively close relationships between proteins in subgroup 3 (and also TSPAN15), though their amino acid sequences are highly diverged.

```

                200      210      220      230      240      250
TSPAN12 TTCATTT-AGTATTTGTGGT-TCATCAGTGGCATTAAATTACATCATTAAATTGGAATATAT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
TSPAN13 GGAAGTGCATTATTTTATTGTCAATATCTGCATTTATTCATGCAGTA---GCAAAATAT
      120      130      140      150      160      170

                260      270      280      290      300      310
TSPAN12 AGTGGATTCCAATTAAGAAAATATTCATATATATTGTTGTTTTATTATTT-TTAACAT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
TSPAN13 AAT-----AAACCAGTATTAAT---TATTGTTGTTTTAACAATTATTAAGTT
      180      190      200      210

                320      330      340      350      360      370
TSPAN12 AATAAAAATAATTATTAATAAAATAAATAGTATATTGGTGGATGTGTTGTAGTTACTATT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
TSPAN13 AATTATAAGAAATAAAATAATGATAAATAGTTGTTATGTGCTTGTATTTTTGTTGCTTTA
      220      230      240      250      260      270

                380      390      400      410      420
TSPAN12 CTTT--TTCTTGGTACTACAGTAACAAGTTTTATTGGA-GTTGCATT-TCCTGATTGGAG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
TSPAN13 ATTGAATTATGGACATTA-AGTACTGCTTATCAAAATATGCTTCATACTCCAAAATGTAT
      280      290      300      310      320      330

```

Figure 4. Local alignment of nucleotide sequences between TSPAN12 and TSPAN13 in *E. histolytica*. Letters in magenta represent intron sequences.

```

                300      310      320      330      340
TSPAN10 TTCAATAACAACATATTCAATTAATTCAATAATAAAAAAGAACAA---CTT--ATTGGT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
TSPAN15 -TCTATTAGATTA-ATTTA--TAATGTAAGAAAGTTAATGACCAATCTCTTCAATTACA
      270      280      290      300      310      320

                350      360      370      380      390
TSPAN10 ----GTTGAAAATGAAGGTTTGTGTTATTTTATTTCTTT-----TTAAGAAAATTAATTA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
TSPAN15 TTCAGTTGAACGTAATGTTTGTTTTATGTTTTTTTATAAAATCTTTATGACATTGATGA
      330      340      350      360      370      380

                400      410      420      430      440
TSPAN10 T-TTAATAGTTT-----AATTGTTGTGGATGGAAAGA--ACCAAGAAATTATTGTT-CT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
TSPAN15 TGTTTCATATTTTCAGTTCAAATGTTGTGGGTGGAATCATTATTATGATTGTAATGTTACA
      390      400      410      420      430      440

                450      460      470      480      490      500
TSPAN10 TCGAAGATATTACCAACATGTTATCATGTA-----TTAATTACACCACATAACTTA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
TSPAN15 ACACCTAGAGGAAGGAATTGTTATAAAGCAGTTGGAATTGGTTTGAAGTCATTAGG-TTA
      450      460      470      480      490      500

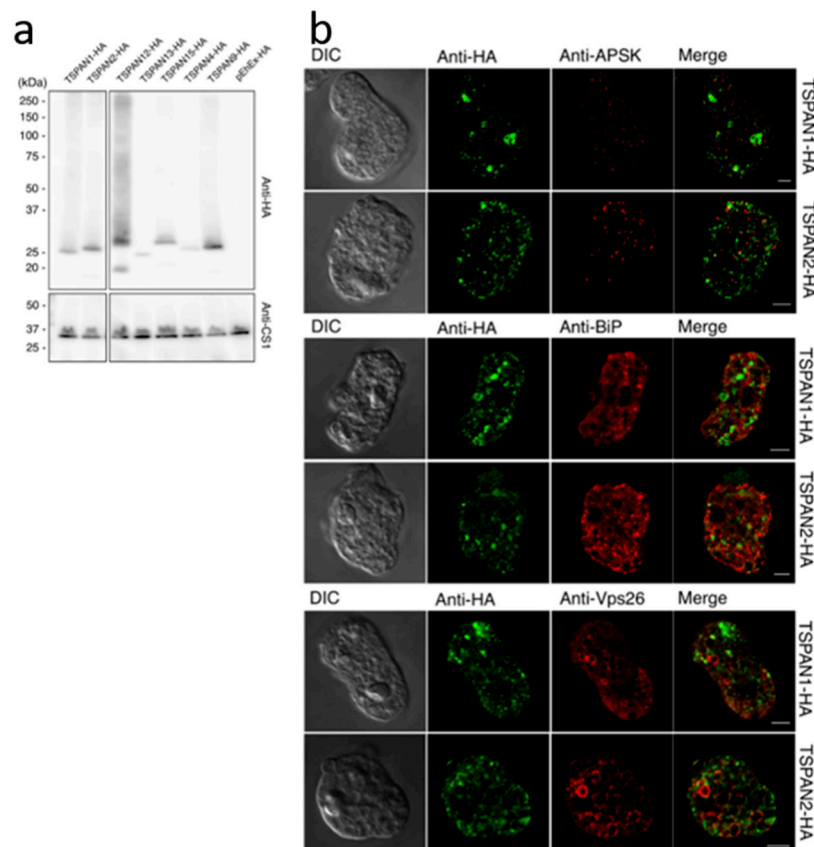
```

Figure 5. Local alignment of nucleotide sequences between TSPAN10 and TSPAN15 in *E. histolytica*. Letters in magenta represent intron sequences.

#### 3.4. Subcellular Localization Analysis of Representative Tetraspanin Candidates

We expressed a few representative tetraspanin candidates in *E. histolytica* trophozoites, with hemagglutinin (HA) tag at the carboxyl-terminus (Figure 6a), including two proteins belonging

to subgroup 1, TSPAN1 and TSPAN2, which are similar to each other (~35% identity in 200 aa). Both were originally found in the proteome of the mitosome, a class of mitochondrion-related organelle present in *Entamoeba* [32]. To verify the localization of these proteins, we performed immunofluorescence assay. Micrographs of cells stained with anti-HA antibody showed mostly vesicular signals of TSPAN1 and TSPAN2 (Figure 6b), rather than the punctate signal pattern associated with mitosome localization [32,42–47]. The anti-HA signals did not colocalize with the mitosome marker (anti-APSK: mitosomal matrix enzyme, [35]) in double-stained cells (Figure 6b, top panel), suggesting TSPAN1 and TSPAN2 are not localized to mitosomes. We then assessed whether the anti-HA signals colocalize with either ER (anti-BiP; ER lumen chaperone, [36]) or endosome (anti-Vps26: retromer-like complex component, [37]) markers (Figure 6b, middle and bottom panels respectively), however the immunofluorescence micrographs likewise suggest that TSPAN1 and TSPAN2 are not localized to either the ER or endosomes. Although the signals of TSPAN1 and TSPAN2 did not show colocalization with the organellar markers used, both proteins similarly showed mostly vesicular patterns suggesting they are localized to certain vesicular compartments in the cytosol. We also performed IFA using trophozoites that were expressing other representative tetraspanins (Figure S5). The anti-HA micrographs show a variety of anti-HA signal patterns, suggesting these candidate amoebic tetraspanins localize to different compartments in the cell.



**Figure 6.** Expression of TSPAN1-HA and TSPAN2-HA in *E. histolytica*. (a) Immunoblot analysis of total cell lysate from representative tetraspanin-expressing and pEhEx-HA control strains respectively. Approximately 20  $\mu$ g of protein was separated by SDS-PAGE and transferred to PVDF membranes. The membranes were reacted with anti-HA antibody (top panel) and anti-CS1 antiserum (bottom panel, loading control) respectively. (b) Immunofluorescence assay micrographs of TSPAN1-HA and TSPAN2-HA double-stained with mouse anti-HA antibody (green) and rabbit anti-APSK (red, top panel), anti-BiP (red, middle panel), and anti-Vps26 (red, bottom panel) antisera respectively. Scale bar, 5  $\mu$ m.

Another previously reported amoebic tetraspanin is TSPAN4 (C4M992), which was identified as differentially more highly expressed in the non-pathogenic A1<sup>NP</sup> strain than the pathogenic B2<sup>P</sup> cell line. In addition, overexpression of TSPAN4 in B2<sup>P</sup> clones caused a loss of the pathogenic phenotype, based on the significantly reduced amoebic liver abscesses formed when injected to gerbils, compared with the respective controls [48]. Association of TSPAN4 to subgroup 1, indicates that the other proteins in this cluster namely TSPAN1, 2, and 3, may also be potentially related with pathogenicity.

Meanwhile, the expression level of almost all candidate tetraspanins did not significantly change ( $\leq 2.0$ -fold change) when trophozoites were subjected to different stress conditions such as glucose starvation [49], cysteine deprivation [50], and paraquat treatment [51] as previously reported in three independent transcriptomic studies. However, we found that TSPAN17 was marginally upregulated (2.05-fold and 2.14-fold) after 24 h and 48 h of cysteine deprivation [50], respectively, and TSPAN7, was slightly upregulated (2.42-fold) after 12 h of paraquat treatment [51]. These data may suggest possible involvement of TSPAN17 and TSPAN7 in response to cysteine-depletion and oxidative stress, respectively. Interestingly, TSPAN7 and TSPAN17 are neighbors in the phylogenetic tree (Figure 1).

#### 4. Discussion

Based on our computational analysis of proteomic and genomic data, we showed that *E. histolytica* possesses at least 17 putative tetraspanins, though the similarity search method used in this study is classical (but effective) [23,24]. We suppose that examining whether those proteins are tetraspanins or not is an important issue to be addressed. Although we have not experimentally verified functions of the potential tetraspanins, one (TSPAN16/C4M5N8) of them shows weak similarity (~18% identity in 185 aa) to TvTSP8, a tetraspanin that mediates parasite–parasite interactions in *Trichomonas vaginalis* [20].

According to the classification results of extracellular domain 2 (EC2) in the previous study [19], all 17 proteins in *E. histolytica* are the ancient 4c type. We found that there are three possible subgroups of tetraspanins in *E. histolytica*, and that uniquely conserved residues are present within each subgroup. Such residues are mainly clustered around the CCG motif in EC2. These may reflect the involvement of different interacting partners of each subgroup. Even so, the result of our phylogenetic analysis is not so robust, as tetraspanins in *E. histolytica* are highly diverged. In fact, phylogenetic trees can vary, except for subgroup 1, depending on MSAs, parameters, and methods used in calculations. Therefore, classification of tetraspanins in *E. histolytica* may be updated or improved with further analysis.

It is also plausible that the C-terminal intracellular regions of potential tetraspanins in *E. histolytica* may have crucial roles related to their subcellular localization. In this study, we found three distinctive patterns in the C-terminal regions of potential tetraspanins. In subgroup 1, Tyr in the C-terminal regions is uniquely conserved. We recognize similarity between a sequence around this tyrosine and a tyrosine-based sorting signal, Yxx $\Phi$ , where  $\Phi$  represents a bulky hydrophobic amino acid [52]. Leu/Val occurred at the corresponding position to  $\Phi$  in tetraspanins of subgroup 1. This tyrosine-based signal is involved in targeting to proteins of various intracellular compartments. The position, i.e. location at the C-terminal region of a protein, and distance from the membrane of YxxL/V are likely to be suitable for such a signal.

In subgroup 3, we found a cluster of charged residues, consisting of two or three positively charged residues and negatively charged residue(s), in their C-terminal regions. However, the functional implications of these conserved residues are not known. In addition, recently, it was reported that the C-terminal regions of tetraspanins in *Trichomonas vaginalis* regulate their subcellular localization [20]. As compared to the wild type, three *T. vaginalis* tetraspanins (TvTSP2, TvTSP3 and TvTSP5) show different localization upon truncation of the C-terminal region that includes a stretch of Ser residues. Two tetraspanins in *E. histolytica* also possess a similar stretch: 200SSSSAAS206 in TSPAN12 (C4M9D7) and 175SSTSSSS181 in TSPAN7 (C4M1M8), though this is a protein of subgroup 2. These portions in the C-terminal regions of potential tetraspanins in *E. histolytica* may likewise contain information related to their subcellular localization.

Our analysis of introns in *Entamoeba* tetraspanins showed that the members of subgroup 1, especially TSPAN1 and TSPAN2 are similar to each other, suggesting that these genes were relatively recently diverged from a common ancestor. However, only the TSPAN2 gene has an intron. Secondly, according to the previous study of 14 introns in *E. histolytica*, the average length was determined to be 65 nucleotides [53]. In this study we found that the average length among a total of 12 introns in the 10 candidate tetraspanin genes in *E. histolytica*, is about 53 nucleotides. Specifically, the average intron lengths among different *Entamoeba* species are 49 (from 11 introns in *E. dispar*), 53 (from 13 introns in *E. moshkovskii*), and 54 (from 11 introns in *E. nuttalli*) nucleotides, respectively, which is consistent with the previous observation. However, the number and length of introns of potential tetraspanin genes in *E. invadens* appear to be higher (18 introns) and longer (89 nucleotides on average) than in other species. These observations suggest that frequencies of intron generation (or elimination) in *Entamoeba* may be higher than previously considered [53] and may indicate the most recent divergence of *E. invadens* in *Entamoeba*. As shown above, there is a striking similarity between the introns of EHI\_091490 (TSPAN12) and EHI\_107790 (TSPAN13). We also found that these two similar introns can possess secondary structures (Figure S4), suggesting that there may be a relationship between the pre-mRNA secondary structure and splicing, as previously reported [54].

## 5. Conclusions

Based on sequence similarities to 'known' tetraspanins, we newly identified nine uncharacterized proteins as potential tetraspanins and provided a refined list of 17 tetraspanins in *E. histolytica*. Computational analysis of those sequences revealed existence of possible subgroups, and characteristic features associated with each subgroup contained in both amino acid and nucleotide sequences. Such features may have implications on the function and evolution of tetraspanins in *E. histolytica*.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4425/10/11/885/s1>: Table S1. List of primer sets used in this study; Figure S1: Multiple sequence alignment of subgroup 1 tetraspanins in *Entamoeba*; Figure S2: Multiple sequence alignment of subgroup 2 tetraspanins in *Entamoeba*; Figure S3: Multiple sequence alignment of subgroup 3 tetraspanins in *Entamoeba*; Figure S4: Predicted secondary structure of introns from representative tetraspanins of subgroup 3; Figure S5: Immunofluorescence assay micrographs of other representative tetraspanin candidates.

**Author Contributions:** Conceptualization, K.T., H.J.S., and T.N.; formal analysis, K.T. and H.J.S.; investigation, K.T. and H.J.S.; writing—original draft preparation, K.T. and H.J.S.; writing—review and editing, K.T., H.J.S., and T.N.; project administration, T.N.; funding acquisition, K.T. and T.N.

**Funding:** This research was partially supported by the Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) from AMED under Grant Number JP19am0101110 to K.T., the Grants-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (JSPS) under Grant Numbers 17K19416 and 18H02650 to T.N., a grant for research on emerging and re-emerging infectious diseases to T.N., a grant for Science and Technology Research Partnership for Sustainable Development (SATREPS) from AMED and Japan International Cooperation Agency (JICA) to T.N., and a grant from the International Medical Research and Development Fund of the National Center for Global Health and Medicine (NCGM) Japan to T.N.

**Acknowledgments:** We would like to thank Mihoko Imada and Kumiko Shibata of The University of Tokyo for providing technical support.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Lozano, R.; Naghavi, M.; Foreman, K.; Lim, S.; Shibuya, K.; Aboyans, V.; Abraham, J.; Adair, T.; Aggarwal, R.; Ahn, S.Y.; et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **2012**, *380*, 2095–2128. [[CrossRef](#)]
2. Petri, W.A., Jr.; Haque, R.; Mann, B.J. The bittersweet interface of parasite and host: Lectin-carbohydrate interactions during human invasion by the parasite *Entamoeba histolytica*. *Annu. Rev. Microbiol.* **2002**, *56*, 39–64. [[CrossRef](#)] [[PubMed](#)]

3. Arroyo, R.; Orozco, E. Localization and identification of an *Entamoeba histolytica* adhesin. *Mol. Biochem. Parasitol.* **1987**, *23*, 151–158. [[CrossRef](#)]
4. Seigneur, M.; Mounier, J.; Prevost, M.C.; Guillén, N. A lysine- and glutamic acid-rich protein, KERP1, from *Entamoeba histolytica* binds to human enterocytes. *Cell. Microbiol.* **2005**, *7*, 569–579. [[CrossRef](#)] [[PubMed](#)]
5. Baxt, L.A.; Baker, R.P.; Singh, U.; Urban, S. An *Entamoeba histolytica* rhomboid protease with atypical specificity cleaves a surface lectin involved in phagocytosis and immune evasion. *Genes Dev.* **2008**, *22*, 1636–1646. [[CrossRef](#)] [[PubMed](#)]
6. Yeung, L.; Hickey, M.J.; Wright, M.D. The many and varied roles of tetraspanins in immune cell recruitment and migration. *Front. Immunol.* **2018**, *9*, 1644. [[CrossRef](#)]
7. Hemler, M.E. Tetraspanin functions and associated microdomains. *Nat. Rev. Mol. Cell. Biol.* **2005**, *6*, 801–811. [[CrossRef](#)]
8. Andreu, Z.; Yáñez-Mó, M. Tetraspanins in extracellular vesicle formation and function. *Front. Immunol.* **2014**, *5*, 442. [[CrossRef](#)]
9. Jiang, X.; Zhang, J.; Huang, Y. Tetraspanins in cell migration. *Cell Adhes. Migr.* **2015**, *9*, 406–415. [[CrossRef](#)]
10. Bari, R.; Guo, Q.; Xia, B.; Zhang, Y.H.; Giesert, E.E.; Levy, S.; Zheng, J.J.; Zhang, X.A. Tetraspanins regulate the protrusive activities of cell membrane. *Biochem. Biophys. Res. Commun.* **2011**, *415*, 619–626. [[CrossRef](#)]
11. Vences-Catalán, F.; Levy, S. Immune targeting of tetraspanins involved in cell invasion and metastasis. *Front. Immunol.* **2018**, *9*, 1277. [[CrossRef](#)] [[PubMed](#)]
12. Hantak, M.P.; Qing, E.; Earnest, J.T.; Gallagher, T. Tetraspanins: Architects of viral entry and exit platforms. *J. Virol.* **2019**, *93*. [[CrossRef](#)] [[PubMed](#)]
13. Piratae, S.; Tesana, S.; Jones, M.K.; Brindley, P.J.; Loukas, A.; Lovas, E.; Eursitthichai, V.; Sripa, B.; Thanasuwan, S.; Laha, T. Molecular characterization of a tetraspanin from the human liver fluke, *Opisthorchis viverrini*. *PLoS Negl. Trop. Dis.* **2012**, *6*, e1939. [[CrossRef](#)] [[PubMed](#)]
14. Gobert, G.N.; Tran, M.H.; Moertel, L.; Mulvenna, J.; Jones, M.K.; McManus, D.P.; Loukas, A. Transcriptional changes in *Schistosoma mansoni* during early schistosomula development and in the presence of erythrocytes. *PLoS Negl. Trop. Dis.* **2010**, *4*, e600. [[CrossRef](#)] [[PubMed](#)]
15. Mekonnen, G.G.; Pearson, M.; Loukas, A.; Sotillo, J. Extracellular vesicles from parasitic helminths and their potential utility as vaccines. *Expert Rev. Vaccines* **2018**, *17*, 197–205. [[CrossRef](#)] [[PubMed](#)]
16. Twu, O.; de Miguel, N.; Lustig, G.; Stevens, G.C.; Vashisht, A.A.; Wohlschlegel, J.A.; Johnson, P.J. *Trichomonas vaginalis* exosomes deliver cargo to host cells and mediate host: parasite interactions. *PLoS Pathog.* **2013**, *9*, e1003482. [[CrossRef](#)]
17. De Miguel, N.; Riestra, A.; Johnson, P.J. Reversible association of tetraspanin with *Trichomonas vaginalis* flagella upon adherence to host cells. *Cell. Microbiol.* **2012**, *14*, 1797–1807. [[CrossRef](#)]
18. The UniProt Consortium. UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* **2017**, *45*, D158–D169. [[CrossRef](#)]
19. Huang, S.; Yuan, S.; Dong, M.; Su, J.; Yu, C.; Shen, Y.; Xie, X.; Yu, Y.; Yu, X.; Chen, S.; et al. The phylogenetic analysis of tetraspanins projects the evolution of cell-cell interactions from unicellular to multicellular organisms. *Genomics* **2005**, *86*, 674–684. [[CrossRef](#)]
20. Coceres, V.M.; Alonso, A.M.; Nievas, Y.R.; Midlej, V.; Frontera, L.; Benchimol, M.; Johnson, P.J.; de Miguel, N. The C-terminal tail of tetraspanin proteins regulates their intracellular distribution in the parasite *Trichomonas vaginalis*. *Cell. Microbiol.* **2015**, *17*, 1217–1229. [[CrossRef](#)]
21. 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. [[CrossRef](#)] [[PubMed](#)]
22. Pearson, W.R. Finding Protein and nucleotide similarities with FASTA. *Curr. Protoc. Bioinform.* **2016**, *53*, 3.9.1–3.9.25. [[CrossRef](#)]
23. Yamada, K.; Tomii, K. Revisiting amino acid substitution matrices for identifying distantly related proteins. *Bioinformatics* **2014**, *30*, 317–325. [[CrossRef](#)] [[PubMed](#)]
24. Tomii, K.; Yamada, K. Systematic Exploration of an Efficient Amino Acid Substitution Matrix: MIQS. *Methods Mol. Biol.* **2016**, *1415*, 211–223. [[CrossRef](#)]
25. Wootton, J.C.; Federhen, S. Analysis of compositionally biased regions in sequence databases. *Methods Enzymol.* **1996**, *266*, 554–571. [[CrossRef](#)]

26. FASTA Sequence Comparison. Available online: [https://fasta.bioch.virginia.edu/fasta\\_www2/](https://fasta.bioch.virginia.edu/fasta_www2/) (accessed on 1 June 2018).
27. Deorowicz, S.; Debudaj-Grabysz, A.; Gudyś, A. FAMSA: Fast and accurate multiple sequence alignment of huge protein families. *Sci. Rep.* **2016**, *6*, 33964. [[CrossRef](#)]
28. Molecular Phylogeny by the NJ/UPG Methods. Available online: <https://mafft.cbrc.jp/alignment/server/phylogeny.html> (accessed on 12 September 2019).
29. RNAfold Web Server. Available online: <http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi> (accessed on 27 March 2019).
30. Diamond, L.S.; Harlow, D.R.; Cunnick, C.C. A new medium for the axenic cultivation of *Entamoeba histolytica* and other *Entamoeba*. *Trans. R. Soc. Trop. Med. Hyg.* **1978**, *72*, 431–432. [[CrossRef](#)]
31. Nakada-Tsukui, K.; Okada, H.; Mitra, B.N.; Nozaki, T. Phosphatidylinositol-phosphates mediate cytoskeletal reorganization during phagocytosis via a unique modular protein consisting of RhoGEF/DH and FYVE domains in the parasitic protozoon *Entamoeba histolytica*. *Cell. Microbiol.* **2009**, *11*, 1471–1491. [[CrossRef](#)]
32. Mi-ichi, F.; Yousuf, M.A.; Nakada-Tsukui, K.; Nozaki, T. Mitosomes in *Entamoeba histolytica* contain a sulfate activation pathway. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 21731–21736. [[CrossRef](#)]
33. Marumo, K.; Nakada-Tsukui, K.; Tomii, K.; Nozaki, T. Ligand heterogeneity of the cysteine protease binding protein family in the parasitic protist *Entamoeba histolytica*. *Int. J. Parasitol.* **2014**, *44*, 625–635. [[CrossRef](#)]
34. Nozaki, T.; Asai, T.; Kobayashi, S.; Ikegami, F.; Noji, M.; Saito, K.; Takeuchi, T. Molecular cloning and characterization of the genes encoding two isoforms of cysteine synthase in the enteric protozoan parasite *Entamoeba histolytica*. *Mol. Biochem. Parasitol.* **1998**, *97*, 33–44. [[CrossRef](#)]
35. Mi-ichi, F.; Makiuchi, T.; Furukawa, A.; Sato, D.; Nozaki, T. Sulfate activation in mitosomes plays an important role in the proliferation of *Entamoeba histolytica*. *PLoS Negl. Trop. Dis.* **2011**, *5*, e1263. [[CrossRef](#)] [[PubMed](#)]
36. Hanadate, Y.; Saito-Nakano, Y.; Nakada-Tsukui, K.; Nozaki, T. Endoplasmic reticulum-resident Rab8A GTPase is involved in phagocytosis in the protozoan parasite *Entamoeba histolytica*. *Cell. Microbiol.* **2016**, *18*, 1358–1373. [[CrossRef](#)] [[PubMed](#)]
37. Nakada-Tsukui, K.; Saito-Nakano, Y.; Ali, V.; Nozaki, T. A retromerlike complex is a novel Rab7 effector that is involved in the transport of the virulence factor cysteine protease in the enteric protozoan parasite *Entamoeba histolytica*. *Mol. Biol. Cell* **2005**, *16*, 5294–5303. [[CrossRef](#)]
38. Käll, L.; Krogh, A.; Sonnhammer, E.L. A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **2004**, *338*, 1027–1036. [[CrossRef](#)]
39. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.J.; Salazar, G.A.; Smart, A.; et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **2019**, *47*, D427–D432. [[CrossRef](#)]
40. NCBI MSA Viewer. Available online: <https://www.ncbi.nlm.nih.gov/tools/msaviewer/> (accessed on 27 March 2019).
41. Wilihoeft, U.; Campos-Góngora, E.; Touzni, S.; Bruchhaus, I.; Tannich, E. Introns of *Entamoeba histolytica* and *Entamoeba dispar*. *Protist* **2001**, *152*, 149–156. [[CrossRef](#)]
42. Makiuchi, T.; Mi-ichi, F.; Nakada-Tsukui, K.; Nozaki, T. Novel TPR-containing subunit of TOM complex functions as cytosolic receptor for *Entamoeba* mitosomal transport. *Sci. Rep.* **2013**, *3*, 1129. [[CrossRef](#)]
43. Santos, H.J.; Imai, K.; Makiuchi, T.; Tomii, K.; Horton, P.; Nozawa, A.; Ibrahim, M.; Tozawa, Y.; Nozaki, T. A novel Mitosomal  $\beta$ -barrel Outer Membrane Protein in *Entamoeba*. *Sci. Rep.* **2015**, *5*, 8545. [[CrossRef](#)]
44. Santos, H.J.; Imai, K.; Hanadate, Y.; Fukasawa, Y.; Oda, T.; Mi-ichi, F.; Nozaki, T. Screening and discovery of lineage-specific mitosomal membrane proteins in *Entamoeba histolytica*. *Mol. Biochem. Parasitol.* **2016**, *209*, 10–17. [[CrossRef](#)]
45. Makiuchi, T.; Santos, H.J.; Tachibana, H.; Nozaki, T. Hetero-oligomer of dynamin-related proteins participates in the fission of highly divergent mitochondria from *Entamoeba histolytica*. *Sci. Rep.* **2017**, *7*, 13439. [[CrossRef](#)] [[PubMed](#)]
46. Santos, H.J.; Hanadate, Y.; Imai, K.; Nozaki, T. An *Entamoeba*-specific mitosomal membrane protein with potential association to the Golgi apparatus. *Genes* **2019**, *10*. [[CrossRef](#)] [[PubMed](#)]
47. Santos, H.J.; Imai, K.; Makiuchi, T.; Tomii, K.; Horton, P.; Nozawa, A.; Okada, K.; Tozawa, Y.; Nozaki, T. Novel lineage-specific transmembrane  $\beta$ -barrel proteins in the endoplasmic reticulum of *Entamoeba histolytica*. *FEBS J.* **2019**, *286*, 3416–3432. [[CrossRef](#)] [[PubMed](#)]



48. Meyer, M.; Fehling, H.; Matthiesen, J.; Lorenzen, S.; Schuldt, K.; Bernin, H.; Zaruba, M.; Lender, C.; Ernst, T.; Ittrich, H.; et al. Overexpression of differentially expressed genes identified in non-pathogenic and pathogenic *Entamoeba histolytica* clones allow identification of new pathogenicity factors involved in amoebic liver abscess formation. *PLoS Pathog.* **2016**, *12*, e1005853. [[CrossRef](#)] [[PubMed](#)]
49. Baumel-Alterzon, S.; Weber, C.; Guillén, N.; Ankri, S. Identification of dihydropyrimidine dehydrogenase as a virulence factor essential for the survival of *Entamoeba histolytica* in glucose-poor environments. *Cell. Microbiol.* **2013**, *15*, 130–144. [[CrossRef](#)] [[PubMed](#)]
50. Husain, A.; Jeelani, G.; Sato, D.; Nozaki, T. Global analysis of gene expression in response to L-Cysteine deprivation in the anaerobic protozoan parasite *Entamoeba histolytica*. *BMC Genom.* **2011**, *12*, 275. [[CrossRef](#)]
51. Husain, A.; Sato, D.; Jeelani, G.; Soga, T.; Nozaki, T. Dramatic increase in glycerol biosynthesis upon oxidative stress in the anaerobic protozoan parasite *Entamoeba histolytica*. *PLoS Negl. Trop. Dis.* **2012**, *6*, e1831. [[CrossRef](#)]
52. Marks, M.S.; Ohno, H.; Kirchhausen, T.; Bonracino, J.S. Protein sorting by tyrosine-based signals: Adapting to the Ys and wherefores. *Trends Cell. Biol.* **1997**, *7*, 124–128. [[CrossRef](#)]
53. Weedall, G.D.; Hall, N. Evolutionary genomics of *Entamoeba*. *Res. Microbiol.* **2011**, *162*, 637–645. [[CrossRef](#)]
54. Rogic, S.; Montpetit, B.; Hoos, H.H.; Mackworth, A.K.; Ouellette, B.F.; Hieter, P. Correlation between the secondary structure of pre-mRNA introns and the efficiency of splicing in *Saccharomyces cerevisiae*. *BMC Genom.* **2008**, *9*, 355. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).