Gene 424 (2008) 1–10

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Putative DEAD and DExH-box RNA helicases families in *Entamoeba histolytica*

Laurence A. Marchat [a], Esther Orozco [b], Nancy Guillen [c,d], Christian Weber [c,d], César López-Camarillo [e,*]

[a] *ENMH-IPN, Programa Institucional de Biomedicina Molecular, México DF, Mexico*
[b] *CINVESTAV-IPN, Departamento de Patología Experimental, México DF, Mexico*
[c] *Institut Pasteur, Unité Biologie Cellulaire du Parasitisme, Paris, France*
[d] *INSERM U786, Paris, France*
[e] *Universidad Autónoma de la Ciudad de México, Posgrado en Ciencias Genómicas, Av. San Lorenzo #290, Col. del Valle, CP 03100, México DF, Mexico*

ARTICLE INFO

Article history:
Received 20 April 2007
Received in revised form 23 October 2007
Accepted 24 July 2008
Available online 9 August 2008

Keywords:
Protozoan parasite
RNA metabolism
Multigene family
Comparative genomics

ABSTRACT

RNA helicases catalyze the unwinding of double-stranded RNA structures to perform numerous genetic processes. These enzymes are characterized by the presence of a conserved helicase domain with specific helicases motifs whose amino acid sequence allows the differentiation between DEAD and DExH-box RNA helicase families. Taking advantage of the availability of the complete genome sequence of *Entamoeba histolytica*, the protozoan responsible for human amoebiasis, we have performed a genomic survey for DEAD and DExH-box RNA helicases encoding genes in this organism. By extensive *in silico* analysis, we identified 20 EhDEAD and 13 EhDExH-box RNA helicases, which contain almost all the conserved helicase motifs. Additionally, several EhDEAD and EhDExH proteins present specific N- and C-terminal domains that could be related to subcellular localization or function. Phylogenetic studies and sequences analysis suggested that this large EhDEAD/DExH-box RNA helicases family has been generated by gene or internal regions duplication, mutation events, introns formation and motif deletions. Interestingly, *EhDexh1 and EhDeaxh10* genes seem to be formed by gene fusion of two ancestral bacterial genes, a mechanism that appears to be evolutionary conserved in the eukaryotic lineage of orthologous proteins. Finally, RT-PCR assays, microarrays and proteomics data analysis showed that several *EhDead* are differentially expressed in relation to distinct culture conditions. These computational and experimental data provide new information on the evolution of EhDEAD/EhDExH-box RNA helicases and their potential relevance for RNA metabolism in *E. histolytica*.

1. Introduction

RNA helicases have been identified in many organisms, from viruses to human, including bacteria and plants (Linder and Daugeron, 2000; Tanner and Linder, 2001). These enzymes are responsible for ATP-dependent unwinding of RNA secondary structures, a required step for transcription, splicing, ribosome biogenesis, pre-mRNA processing, mRNA export, degradation and translation (de la Cruz et al., 1999; Linder and Daugeron, 2000; Tanner and Linder, 2001; Linder and Stutz, 2001). RNA helicases are also regulatory factors for cellular growth and differentiation (Schmid and Linder, 1992; Lohman and Bjornson, 1996). In addition, they act as RNPases and remodel RNA–proteins interactions (Linder et al., 2001; Schwer, 2001).

RNA helicases exhibit a highly conserved helicase domain of about 290–360 amino acids (aa), suggesting that orthologous genes evolved from a common ancestor (Aubourg et al., 1999). Most RNA helicases belong to the DEAD-box family of the SFII superfamily, whose prototype is the eukaryotic eIF-4A translation initiation factor (Benz et al., 1999; Caruthers et al., 2000). DEAD-box proteins are characterized by the conserved DEAD motif II (Asp-Glu-Ala-Asp) and seven additional functional motifs called I, Ia, Ib, III, IV, V and VI that are required for ATP hydrolysis, as well as RNA-binding and unwinding activities (Gorbalenya and Koonin, 1993; Linder et al., 2001; Rocak and Linder 2004). Additionally, upstream of motif I, most DEAD-box proteins possess a conserved aromatic residue (Phe) and a Q motif with six residues including an invariant Gln, which forms a loop-helix-loop structure involved in the interaction with motif I and ATP (Tanner, 2003; Tanner et al., 2003; Cordin et al., 2004). The SFII superfamily also includes the DExH-box family, whose members share high similarities in sequences with DEAD-box proteins but exhibit a specific DExH motif II (Asp-Glu-x-His). The change from Asp to His is responsible for a reduced helicase activity in DExH-box proteins (Pause and Sonenberg, 1992). Motifs Ia, Ib, IV and V in DExH-box proteins are also markedly different from those of DEAD-box proteins (Abdelhaleem et al., 2003) and Q motif has not been reported in DExH-box helicases.

Abbreviations: aa, amino acid; BLAST, basic local alignment search tool; bp, base pair; MDR, multidrug resistance phenotype; nt, nucleotide; ORF, open reading frame; TIGR, The Institute for Genomic Research.

* Corresponding author. Tel.: +55 5488 6661x5352; fax: +55 55 59 01 87.
E-mail address: genomicas@yahoo.com.mx (C. López-Camarillo).

0378-1119/$ – see front matter.
doi:10.1016/j.gene.2008.07.042

These signature sequences represent useful tools for the prediction of new RNA helicases in genomes databases. Extensive *in silico* analysis allowed the identification of 55 putative RNA helicase genes in *Arabidopsis thaliana*, whereas *Caenorhabditis elegans* and *Drosophila melanogaster* have 32 and 29 homologous genes, respectively (Boudet et al., 2001); *Homo sapiens* has 36 DEAD and 14 DExH-box (Abdelhaleem et al., 2003) and *Saccharomyces cerevisiae* has 26 DEAD-box and 7 DExH-box RNA helicases (de la Cruz et al., 1999). Recently, 29 RNA helicases have been reported in the protozoan parasite *Leishmania major* (Ivens et al., 2005) and a family of 22 DEAD-box helicases was found in *Plasmodium falciparum* (Tuteja and Pradhan, 2006).

To date, there are only few reports about RNA metabolism and posttranscriptional gene expression regulation in *Entamoeba histolytica*, the protozoan responsible for human amoebiasis. In our group, we have demonstrated that *EhPgp5* mRNA stability is a regulatory event in the *E. histolytica* multidrug resistance (MDR) phenotype (López-Camarillo et al., 2003). We have also identified the pre-mRNA 3′ end processing machinery in this parasite (López-Camarillo et al., 2005) and characterized the poly(A) polymerase EhPAP (Garcia-Vivas et al., 2005). To gain insights into molecules and pathways participating in *E. histolytica* RNA metabolism, here we have studied RNA helicases that are considered as key factors for all genetic processes involving RNA. Taking advantage of the helicase domain conservation through evolutionary scale, we have screened the *E. histolytica* genome databases to find sequences coding for open reading frames (ORF) with significant similarity to DEAD and DExH-box proteins. Our analysis led to the *in silico* characterization of 33 putative EhDEAD and EhDExH-box RNA helicases, which are currently used as a guide for experimental studies. In addition, we demonstrated that several *EhDead* and *EhDexh* genes are transcribed through RT-PCR assays and microarrays data analysis.

2. Materials and methods

2.1. Screening of databases and sequences analysis

To search for *E. histolytica* RNA helicases, we initially screened the parasite genome at The Institute for Genomic Research (TIGR, http://www.tigr.org/tdb/e2k1/eha1/) and Sanger (http://www.sanger.ac.uk/Projects/E_histolytica/) databases using the conserved eukaryotic RNA helicase domain (PF00271). The existence of selected *E. histolytica* genes was then confirmed by analysis of the recently actualized *E. histolytica* Pathema database (http://pathema.tigr.org/tigr-scripts/Entamoeba/PathemaHomePage.cgi). Putative orthologous proteins were selected from BLAST analysis against all databases using the Blosum 62 scoring matrix (http://us.expasy.org/tools/blast/) and the following criteria: (i) at least 20% identity and 35% homology to the query sequence; (ii) *e*-value lower than 0.002; and (iii) absence of stop codons in the coding sequence. Predicted aa sequences were aligned by ClustalW software (http://www.ch.embnet.org/software/ClustalW.html), allowing gap penalties of 10 to maximize proteins homology. Structural domains and sequence patterns were predicted by ScanProsite (http://us.expasy.org/tools/scanprosite/), Motif Scan (http://myhits.isb-sib.ch/cgi-bin/motif_scan), Pfam (http://www.sanger.ac.uk/Software/Pfam/search.shtml), SMART (http://smart.embl-heidelberg.de/) and PSORT II (http://psort.ims.u-tokyo.ac.jp/). Pairwise comparisons were performed using BioEdit biological sequences alignment editor (version 7.0.5.3) and GeneDoc (http://www.psc.edu/biomed/genedoc/) softwares to determine identity/homology percentages. Dot plot analyses were conducted using the PLALIGN software (http://bioinfo.hku.hk/FASTA/plalign.htm). Data about predicted introns were obtained from TIGR and Sanger databases. An intron located between two codons was considered in phase 0; one located within a codon was considered in phase 1 (lying after the first nucleotide (nt)) or 2 (lying after the second nt). The phases of two introns surrounding an exon defined the exon type.

2.2. Phylogenetic analysis

The phylogenetic tree for EhDEAD and EhDExH-box RNA helicases was constructed using the Neighbor-Joining distance method with a cut off value of 50% (Saitou and Nei, 1987) as implemented in the MEGA package version 3.1 (Kumar et al., 2004). We also generated phylogenetic trees for each *E. histolytica* RNA helicase aligned with orthologues from different species. The robustness of the phylogenetic inferences was tested by bootstrapping method, involving 500 replications of the data based on the criteria of 50% majority-rule consensus.

2.3. E. histolytica cultures, DNA and RNA isolation, cDNA preparation

Trophozoites of clone A (strain HM1:IMSS) axenically cultured in TYI-S-33 medium (15% serum) (Diamond et al., 1978) were harvested in logarithmic growth phase to obtain genomic DNA with the Wizard® Genomic DNA Purification Kit (Promega) and total RNA using the Trizol reagent (Invitrogen). cDNA was then synthesized from 1 μg of total RNA, 100 ng of oligo (dT_{18}), 100 mM DTT, 10 mM dNTPs, 40 U of SUPERase-in (Ambion) and 200 U of Superscript II reverse transcriptase (Invitrogen) in first-strand buffer, for 1 h at 42 °C. To remove the excess of RNA template, 2 U of RNase H (Invitrogen) were added and the mixture was incubated at 37 °C for 15 min (López-Camarillo et al., 2003).

Table 1
Repertoire of putative DEAD and DExH-box proteins in *E. histolytica*

Entamoeba histolytica		Homologous proteins						
Protein name	Accession number[a]	Protein Size	Protein name	Specie	Accession number[b]	*e*-value	*H*	*I*
EhDEAD1	XP_653330	578	DDX3Y	*Hs*	O15523	1e−121	60	44
EhDEAD2	XP_654333	535	DDX5	*Br*	Q6TNU9	1e−121	70	48
EhDEAD3	XP_650493	722	PRP11	*Sp*	Q9P7C7	1e−132	57	41
EhDEAD4	XP_650751	585	U5snRNP	*At*	P93008	1e−110	58	38
EhDEAD5	XP_649929	705	DDX50	*Hs*	Q9BQ39	8e−89	55	34
EhDEAD6	XP_648201	600	CG9143	*Dm*	Q9V915	7e−63	61	31
EhDEAD7	XP_654408	432	CG9253-PA	*Dm*	Q9VIF6	e−114	73	52
EhDEAD8	XP_657561	450	T26G10.1	*Ce*	P34580	2e−74	57	41
EhDEAD9	XP_657423	474	DDX55	*Br*	Q9PBZ0	7e−73	54	35
EhDEAD10	XP_652222	558	DDX31	*Hs*	Q9H8H2	1e−99	58	37
EhDEAD11	XP_650379	542	HAS1	*Sp*	Q09916	e−139	72	54
EhDEAD12	XP_651045	624	none	*At*	Q9FFT9	e−113	56	38
EhDEAD13	XP_655288	684	DDX54	*Mm*	Q8K4L0	1e−119	54	38
EhDEAD14	XP_651630	500	DDX51	*Hs*	Q8N8A6	1e−38	48	29
EhDEAD15	XP_651424	551	DBP9	*Sp*	O60080	1e−86	58	36
EhDEAD16	XP_657280	391	eIF4A-15	*Ha*	Q6T8C6	e−126	77	57
EhDEAD17	XP_657009	384	DDX48	*Hs*	P38919	e−132	78	58
EhDEAD18	XP_655794	419	BAT1	*Rn*	Q63413	1e−106	68	49
EhDEAD19	XP_649638	419	Sub2	*Sc*	Q07478	3e−95	65	46
EhDEAD20	XP_651823	440	DBP5	*Cn*	Q5KBP5	1e−46	54	35
EhDExH1	XP_655520	1804	U5snRNP200	*Dm*	Q9VUV9	0.0	57	35
EhDExH2	XP_654243	1192	DHX57	*Hs*	Q6P158	3e−99	50	30
EhDExH3	XP_659556	909	DHX8	*Hs*	O86YB2	9e−86	53	34
EhDExH4	XP_650643	757	DHX8	*Sp*	O42642	4e−92	59	37
EhDExH5	XP_654225	845	PRP16	*Hs*	Q92620	0.0	73	53
EhDExH6	XP_655062	664	DHX8	*Hs*	Q14562	e−127	58	38
EhDExH7	XP_656898	675	DHX15	*Hs*	O43143	e−176	67	48
EhDExH8	XP_657573	811	DHX16	*Hs*	O60231	0.0	68	49
EhDExH9	XP_649168	845	DHX8	*Hs*	Q14652	0.0	78	59
EhDExH10	XP_657324	1799	U5snRNP200	*Dm*	Q9VUV9	0.0	54	35
EhDExH11	XP_648655	1062	SKIV2L2	*Hs*	O15477	1e−127	49	32
EhDExH12	XP_652111	977	SKIV2L2	*Hs*	P42285	0.0	65	48
EhDExH13	XP_648711	1214	DHX57	*Hs*	Q6P158	4e−67	50	31

At, Arabidopsis thaliana; Br, Brachydanio rerio; Ce, Caenorhabditis elegans; Cn, Cryptococcus neoformans; Dm, Drosophila melanogaster; Ha, Helianthus annuus; Hs, Homo sapiens; Mm, Mus musculus; Rn, Rattus norvegicus; Sc, Saccharomyces cerevisiae; Sp, Schizosaccharomyces pombe. Homology (*H*) and identity (*I*) values are expressed in percentage. Protein size is expressed in amino acid numbers.

[a] GenBank database.

[b] UniProt Knowledgebase database.

2.4. PCR assays

We designed five pairs of specific primers corresponding to regions that flank predicted introns in *EhDead3, EhDead5, EhDead9, EhDead12* and *EhDead20* genes. *Actin* sense and antisense specific primers were used as an internal control in all experiments. PCR were performed with 1 μg genomic DNA or 1.5 volume of the reverse transcription mixture, using 2.5 U of *Taq* DNA polymerase (Invitrogen). Amplified products were separated through 12% denaturing PAGE, ethidium bromide stained and observed in a Gel-Doc apparatus (Bio-Rad).

3. Results

3.1. E. histolytica has a large DEAD and DExH-box RNA helicases family

By genomic survey and *in silico* analysis of the *E. histolytica* genome, we detected 33 genes that encode for putative RNA helicases, including 20 EhDEAD and 13 EhDExH-box proteins (Table 1). BLAST analyses of the 33 *E. histolytica* RNA helicases presented *e*-values lower than 1e−46 as well as 29–59% identity and 49–78% homology with eukaryotic RNA helicases, confirming that they are orthologous proteins (Table 1). Multiple alignment of the whole data set showed that the helicase domain of all EhDEAD-box helicases is highly conserved and contained the nine consensus motifs described in other organisms including the conserved Phe residue close to Q motif (Fig. 1B). Moreover, distances between each motif are in the reported range (Fig. 1C) (Pause and Sonenberg, 1992; Cordin et al., 2004). However, in EhDEAD3, EhDEAD4, EhDEAD6 and EhDEAD11, the Phe residue is replaced by Trp or Tyr, another aromatic aa, whereas it is absent in EhDEAD10 and EhDEAD14. Intriguingly, some motifs were not found in several proteins, including motif I in EhDEAD9 and motif V in EhDEAD12 (Fig. 1B). In most EhDExH-box helicases, the helicase domain is also well conserved, although motifs Ia, Ib, IV and V are missing in EhDExH1, EhDExH10, EhDExH11 and EhDExH12 proteins (Fig. 1B). The absence of some helicase motifs in some EhDEAD and EhDExH proteins suggests that deletions/insertions events could be shaping the formation of *EhDead* and *EhDexh* gene families. However, we do not know whether these changes alter enzymes functions.

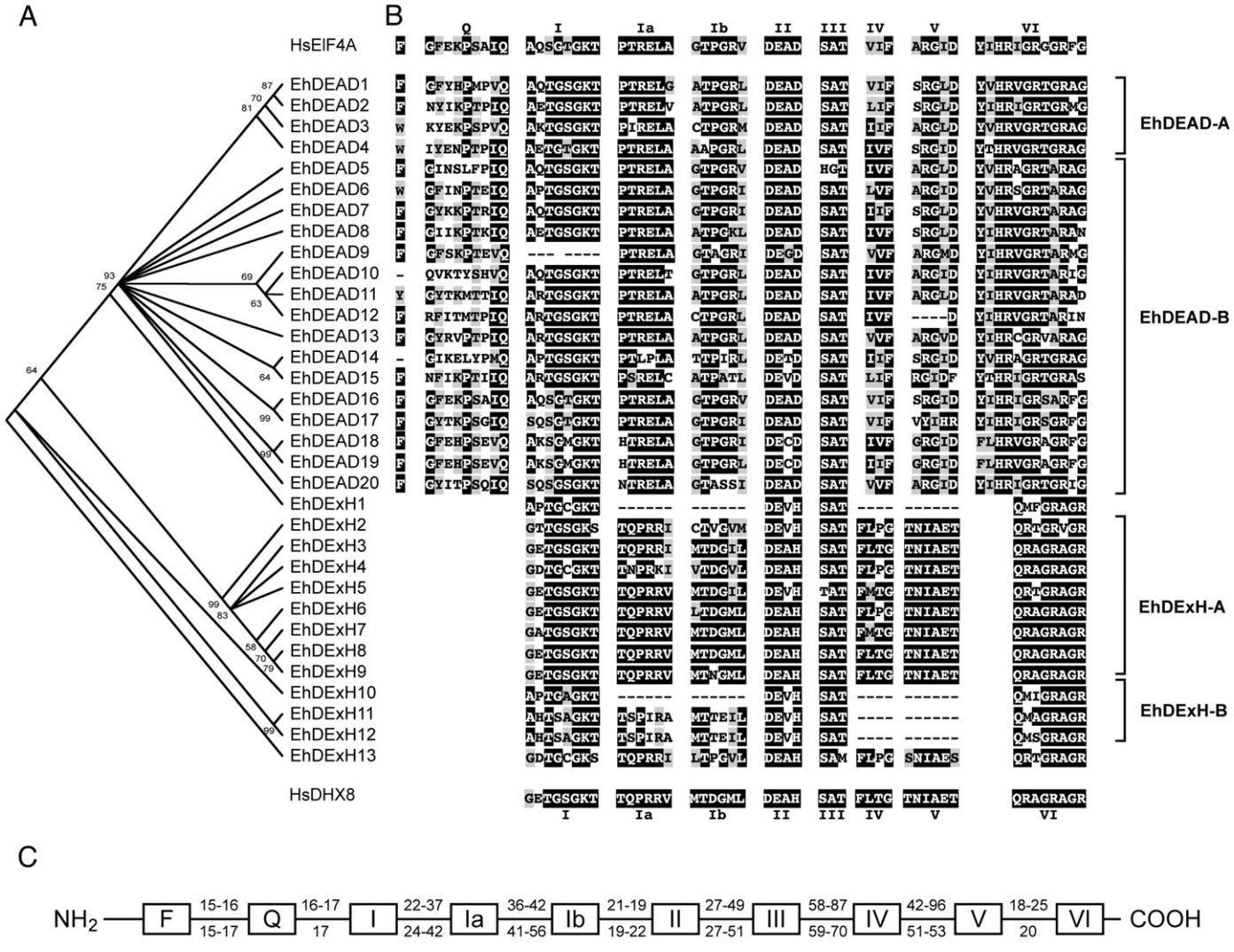

Fig. 1. DEAD and DExH-box RNA helicases families in *E. histolytica*. (A) Phylogenetic tree derived from the alignment of complete predicted proteins sequences. Numbers indicate the bootstrap values (percentage of 500 replications) for each fork. (B) Alignment of conserved helicase motifs. Sequences were aligned by ClustalW. HseIF4A and HsDHX8, motifs for *H. sapiens* DEAD and DExH-box helicases prototypes, respectively. Subfamilies based on phylogenetic study (see Section 3.2 for details) are indicated at the right. (C) Schematic representation of the DEAD-box helicase domain. Numbers indicate the distance between each motif in consensus (up) and parasite (down) helicase domain.

3.2. Most EhDEAD and EhDExH-box RNA helicases can be clustered in four subfamilies

Phylogenetic inference of the 33 *E. histolytica* RNA helicases confirmed the evolutionary separation of DEAD and DExH-box proteins families (Fig. 1A). DEAD-box proteins emerged from two distinct nodes and were clustered into EhDEAD-A (EhDEAD1 to EhDEAD4) and EhDEAD-B (EhDEAD5 to EhDEAD20) subfamilies, which probably evolved from distinct ancestors at an early stage of the parasite evolution. EhDExH-box proteins also formed two separate EhDExH-A (EhDExH2 to EhDExH9) and EhDExH-B (EhDExH10 to EhDExH12) subfamilies, suggesting that they came from distinct ancestors. Additionally, two EhDExH proteins were located in two separate arms, probably because they have evolved through a distinct evolutionary process. EhDExH13 was detected alone at the bottom of the tree, close to EhDExH-B subfamily members. In contrast, EhDExH1 appeared to be closely related to EhDEAD-B subfamily, suggesting that EhDExH1 and EhDEAD-box RNA helicases could have recently evolved from a common ancestor (Fig. 1A). Independent phylogenetic analysis of each EhDEAD and EhDExH proteins showed that *E. histolytica* proteins are clustered with homologues from other protozoan, such as *Trypanosoma*, *Leishmania*, *Giardia* or *Plasmodium*, whereas RNA helicases from higher eukaryotic organisms form a distinct group (data not shown).

Sequence comparisons of each full length *E. histolytica* RNA helicase sequence against the others confirmed the evolutionary relation among EhDEAD and EhDExH families' members. Particularly, EhDEAD16 and EhDEAD17 proteins, which appeared closely related in the phylogenetic tree, share 46% identity and 71% homology, suggesting that they are paralogous RNA helicases. The same observation can be made for EhDEAD18 and EhDEAD19 that exhibit 73% identity and 86% homology. In contrast, groups formed by EhDEAD9, EhDEAD10, EhDEAD11 and EhDEAD12, and by EhDEAD14 and EhDEAD15 in the phylogenetic tree, do not correspond to paralogous helicases, since proteins only share 16–24% identity and 28–44% homology. Based on complete amino acid sequences comparison, no paralogous proteins could be identified among DExH helicases.

The high sequence similarity between EhDEAD/EhDExH proteins and characterized orthologous proteins suggest that they may have similar function in *E. histolytica* (Table 1). Thus, EhDEAD1, whose helicase activity has been experimentally confirmed (Lopez-Camarillo et al., 2008), is homologous to human DDX3Y, which plays a role in mRNA translation and cell cycle, respectively (Foresta et al., 2001). The paralogous helicases, EhDEAD16 and EhDEAD17 could be involved in translation as the initiation factor EIF4A-15 factor (Caruthers et al., 2000), while EhDEAD18 and EhDEAD19, are homologues of factors involved in transcription elongation, pre--mRNA splicing and mRNA nuclear export (Stresser and Hurt, 2001; Stresser et al., 2002; Rondon et al., 2003). EhDExH3, EhDExH4 and EhDExH9 correspond to the human DHX8 helicase that facilitates nuclear export of spliced mRNA by releasing RNA from the spliceosome (Ohno and Shimura, 1996). EhDExH11 and EhDExH12 are orthologous to the human SKIV2L2 helicase, which is proposed to regulate various aspects of RNA metabolism in the cells (Chen et al., 2001) (Table 1). Although they do not share enough aa identity to be considered as paralogous proteins, EhDExH1 and EhDExH10 are *E.*

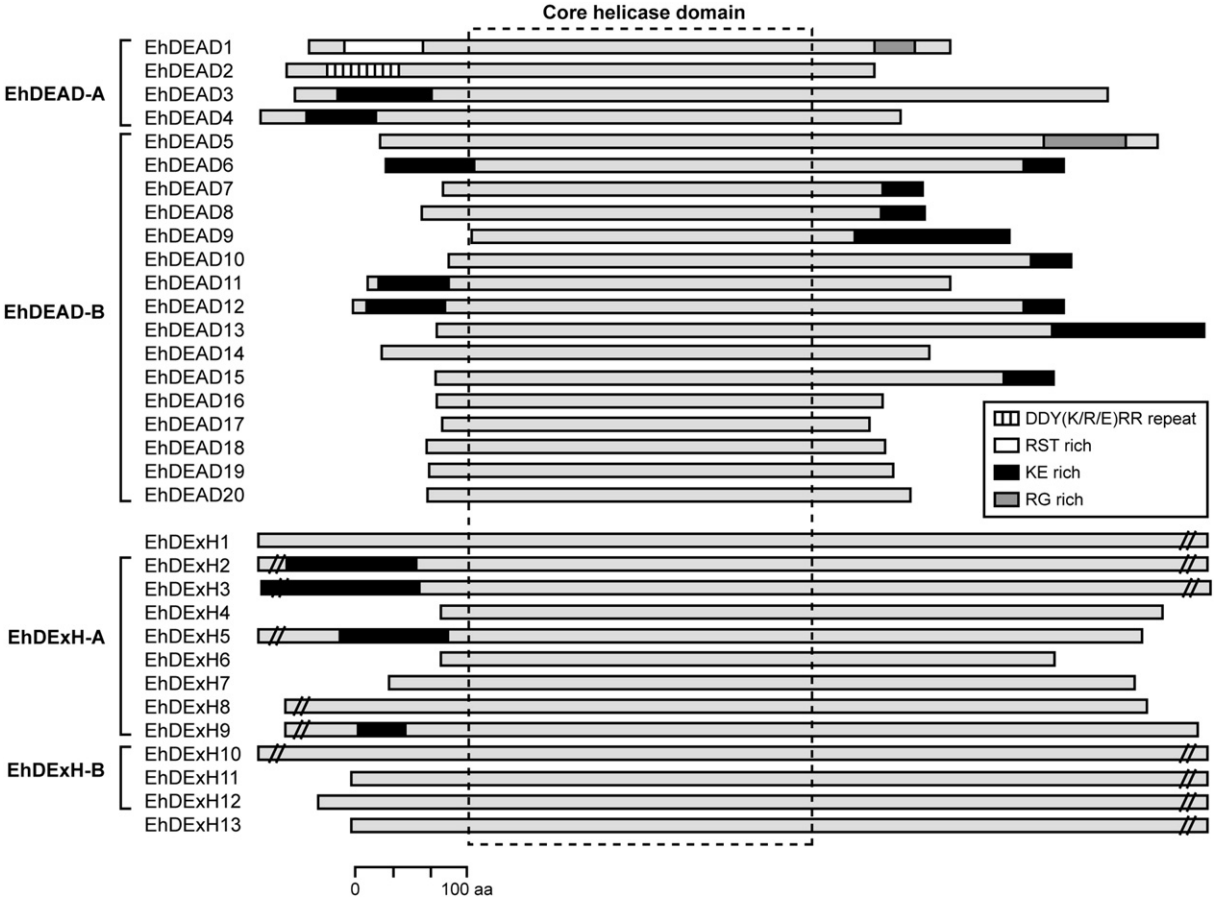

Fig. 2. Schematic representation of amino acids content in N- and C-terminal extensions from EhDEAD and EhDExH-box proteins. The core helicase domain position is indicated by the discontinuous line. The scale is shown at the bottom.

histolytica orthologues of the eukaryotic U5snRNP-200 kDa factor involved in pre-mRNA splicing (Laggerbauer et al., 1998).

3.3. EhDEAD and EhDExH-box RNA helicases present specialized extensions

The study of helicase domain flanking regions revealed the presence of specific regions that could be related to subcellular localization and function (Fig. 2). 11 EhDEAD and four DExH-box helicases have KE rich regions (33–60% KE residues) in their N- or C-terminal extensions. These sequences may be related to well known nuclear localization signals, such as the tract of four basic aa or the bipartite nuclear targeting sequence (that is characterized by two basic aa followed by a spacer of 10 aa and at least three basic aa in a final tract of five aa), suggesting that these proteins may have a nuclear localization. The DDY(K/R/E)RR tandem repeat identified in the EhDEAD2 N-terminus could also correspond to nuclear localization signals. Interestingly, PSORT II analysis also predicted that EhDEAD4, EhDEAD5, EhDEAD9, EhDEAD10, EhDEAD11, EhDExH12 and EhDExH13 were nuclear proteins with a significant score of 70–87%. On the other hand, EhDEAD and EhDExH-box proteins that do not have these specialized extensions were not predicted as nuclear proteins by PSORT II. The other classes of extensions included a RG-rich region, which could be a variant of the RGG box involved in RNA-binding and a RST-rich region with unknown function (Fig. 2).

3.4. EhDEAD-box proteins exhibit internal repeats

Comparison of each protein with itself through dot plot analyses allowed the detection of internal repeats within several *E. histolytica* RNA helicases (Fig. 3). In EhDEAD11 we identified two overlapped direct repeats of about 70 aa at 65–135 aa and 85–155 aa positions ($p<0.0001$). EhDEAD3 and EhDEAD17 have two overlapped repeats that span almost all the protein (75–575 and 130–630 aa, and 0–260 and 75–325 aa positions, respectively) but they share less aa sequence similarity ($p<0.01$). Additionally, the last 125 aa of EhDEAD5 C-terminal end is composed of multiple short repetitive direct regions ($p<0.001$ and $p<0.01$) (Fig. 3). This analysis suggested that internal sequences duplications could represent a key event in the generation and diversification of the EhDEAD-box RNA helicases gene family.

3.5. EhDexh1 and EhDexh10 genes appear to result from gene fusion events

Gene fusion leading to the formation of multidomain proteins is one of the major routes of protein evolution (Yanai et al., 2002). To better understand the evolution of the large EhDEAD and EhDExH-box families, we searched for data that could indicate gene fusion events in *E. histolytica* RNA helicases. Interestingly, by using dot plot and Pfam tools, we found that EhDExH1 contains two direct repeats (repeat 1 and 2) of about 750 aa ($p<0.0001$) than span almost all the protein sequence (Fig. 4). Both repeats contain four RNA helicase motifs (motifs I, II, III and VI) as well as an additional Sec63 motif of unknown function that has been previously found in yeast Sec63p, Brr2p and HFM1p helicases (Ponting, 2000). Sec63p is required for the translocation of SRP-dependent precursors into the yeast endoplasmic reticulum *in vivo* (Young et al., 2001); Brr2p is a RNA-dependent ATPase RNA helicase involved in the facilitation and disruption of snRNA interactions, that is required for disruption of U4/U6 base-pairing in native snRNPs to activate the spliceosome for catalysis (Raghunathan and Guthrie, 1998); HFM1p is a meiosis specific DNA helicase involved in the conversion of double-stranded breaks to later recombination intermediates and in crossover control (Nakagawa and Ogawa, 1999). Moreover, the EhDExH1 molecular

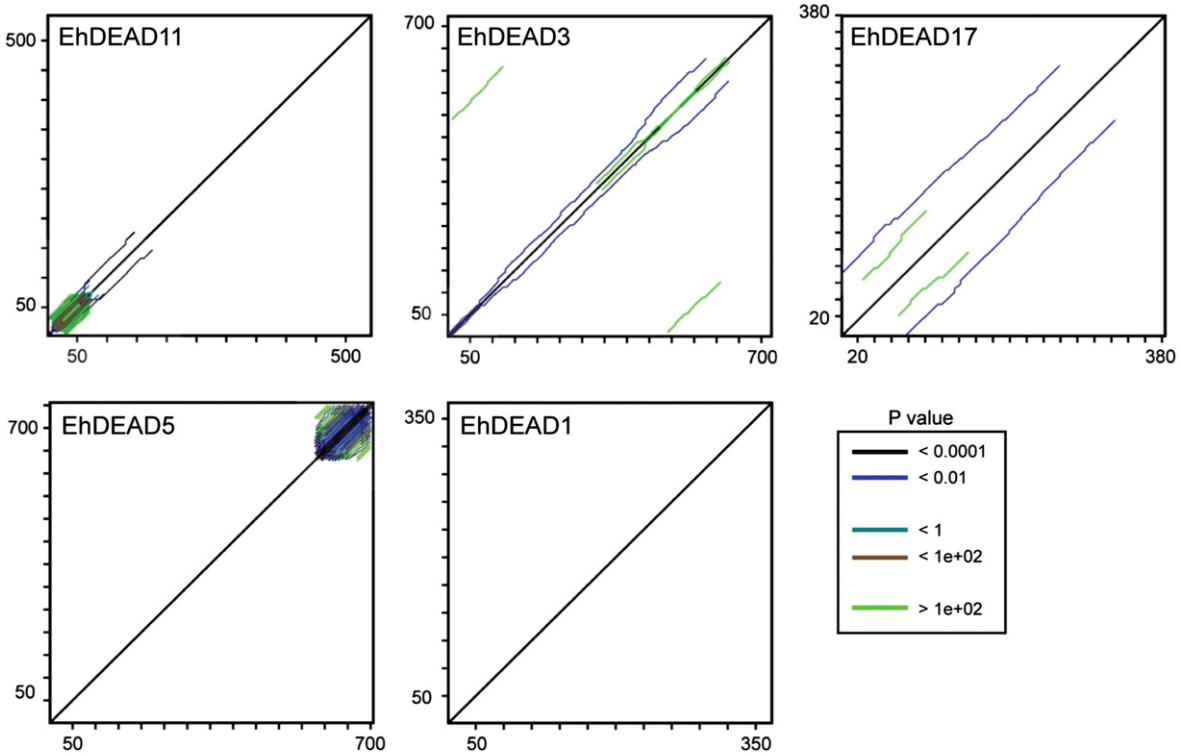

Fig. 3. Dot plot analysis of EhDEAD helicases. Each protein sequence was aligned with itself using the PLALIGN program and the results were plotted from a point file with the coordinates of the points in common position. The presence of direct internal repeats was revealed by diagonal rows in the graph. The graph corresponding to EhDEAD1 is representative of proteins without internal repeats. *p* value, probability of statistical error.

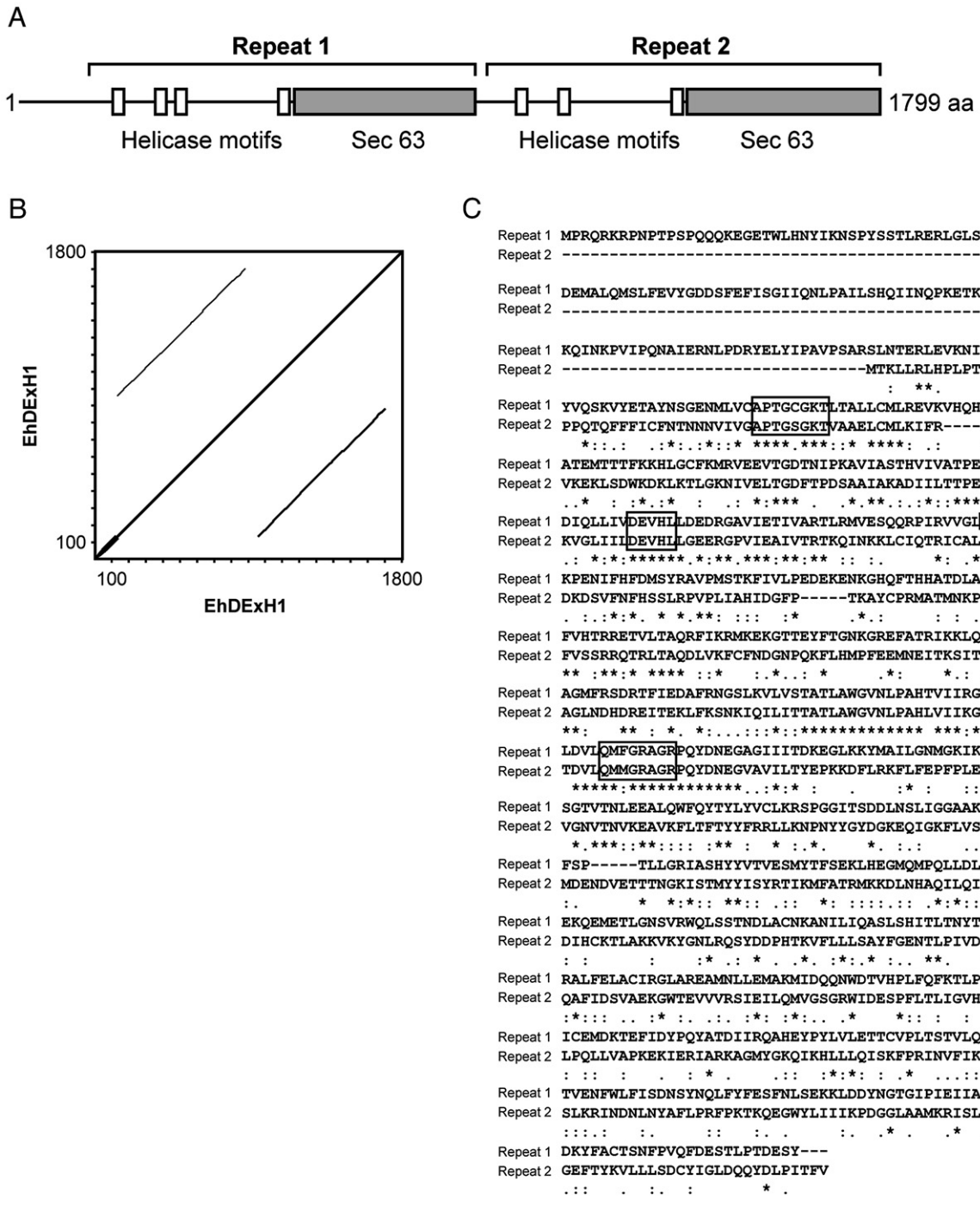

Fig. 4. Gene fusion event in EhDExH1-box protein. (A) Molecular organization of the predicted aa sequence. (B) Dot plot matrix analysis. (C) Sequence comparison of the two internal repeats using the ClustalW software. Asterisk, identical aa; single and double point, homologous aa; open box, consensus helicase motifs.

length (1799 aa) is about twice the conventional RNA helicases size (Fig. 4). The same observations were made for the EhDExH10 protein that is formed by two direct repeats of 550 aa (data not shown). This molecular organization suggested that *EhDexh1* and *EhDexh10* genes were formed from the fusion of two independent RNA helicase genes with the simultaneous loss of intergenic sequences (Fig. 5A).

To strengthen this hypothesis, a similar gene fusion event was explored by BLAST analysis through multiple species using EhDExH1 repeat 1 and 2 as query (Table 2). Interestingly, we detected patterns of co-occurrence of the two repeats in several eukaryotic orthologous proteins through evolutionary scale that are codified by genes with the same molecular organization, suggesting that gene fusion is an evolutionary conserved event in the eukaryotic lineage of EhDExH1 orthologues. In contrast, each EhDExH1 repeat corresponds to stand-alone proteins in archae-bacteria, suggesting that gene fusion event has occurred early in the evolution of archae-bacteria to eukaryotic lineage (Fig. 5B). Identity and homology values were high, leaving no doubt that they are true orthologues (Table 2). In addition, phylogenetic tree reconstruction for each EhDExH1 repeat showed that eukaryotic fused proteins formed a separate cluster from the non-fused bacterial and archaeal orthologues (Fig. 5C and D), suggesting a horizontal gene transfer between bacteria and eukaryotes. The same observations were made for the paralogous

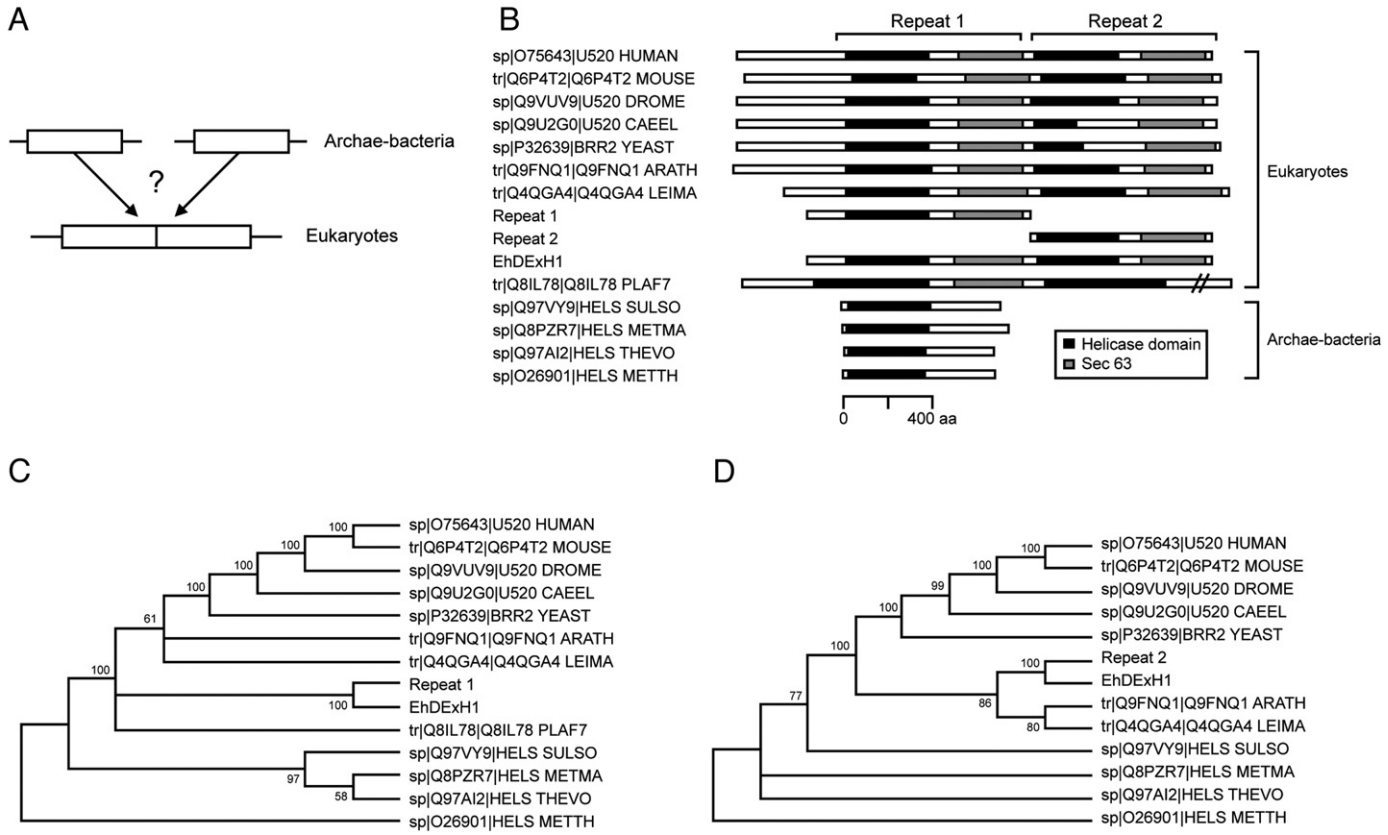

Fig. 5. Gene fusion event through the evolution of EhDExH1-box orthologues. (A) Hypothetical model for gene fusion event through DExH-box helicases evolution. (B) Molecular organization of prokaryotic and eukaryotic orthologues of putative EhDExH1 repeats 1 and 2. (C and D) Phylogenetic analysis of prokaryotic and eukaryotic orthologues of EhDExH1 repeats 1 (C) and 2 (D). Numbers indicate the bootstrap values (percentage of 500 replications) for each fork. The scale is at the bottom.

EhDExH10 protein (data not shown). It has been proposed that the selective advantage of domain fusion lies in the increased efficiency of coupled catalytic activities (Marcotte et al., 1999). However, experimental data are required to test this assumption for EhDExH1 and EhDExH10.

3.6. Several EhDead genes present introns

We next analyzed the molecular structure of the whole set of *E. histolytica* RNA helicases encoding genes (Fig. 6). No introns were found in *EhDexh* genes, whereas five *EhDead* genes were predicted to contain one or two introns within the core helicase domain. Notably, we observed that most introns are located over the 5′ end of genes, which is a common feature of intron-sparce genomes (Mourier and Feffares, 2003). Both introns are separated by 147 nt in the *EhDead3* gene. In agreement with the previous reports of small introns in *E. histolytica* genes (Loftus et al., 2005; Davies et al., 2007), the mean length of introns is 65 nt, with a minimum of 38 nt for *EhDead12 and* a maximum of 127 nt for *EhDead3* (Table 3). Considering that the addition of an intron can occur with an equal probability for the three codon sites, the expected value for each intron phase is 33%. Here, we determined that phase 0 is over-represented (83%) in the 6 introns found in *EhDead* genes, whereas only one intron is in phase 1 (17%) and none is in phase 2. The *EhDead3* gene has an exon of type 0–1 (Table 3). Comparative analysis of genomic sequences suggested that mutation events in the *EhDead9* gene sequence (nt change or deletion) seem to have led to the formation of a non encoding sequence that was reported as an intron in TIGR and Sanger *E. histolytica* databases. Additionally, this change induced the loss of motif I in EhDEAD9 as described above. In contrast, introns of *EhDead3*, *EhDead5*, *EhDead12* and *EhDead20* genes, seem to result from the addition of non encoding DNA sequences since the distance between flanking helicase motifs were augmented (Fig. 6). With the exception of the second intron of *EhDead3*, most predicted 5′ splice and 3′ splice sites are not in agreement with the GUUUGU-UAG splice sites recently proposed for *E. histolytica* genes from a large scale computational study of splicing (Davies et al., 2007) (Table 3).

In order to experimentally confirm the existence of predicted introns, we performed PCR assays using genomic DNA and cDNA from *E. histolytica* trophozoites growing under standard axenic conditions,

Table 2
Identification of homologous proteins using repeat 1 and 2 from EhDExH1

Organism	Accession number[a]	Repeat 1			Repeat 2		
		E-value	*I*	*H*	*E*-value	*I*	*H*
Dm	Q9VUV9	e−150	36	57	e−152	34	57
At	Q9FNQ1	e−148	36	56	0.0	47	65
Ce	Q9U2G0	e−148	36	55	e−138	33	55
Hs	O75643	e−146	35	56	e−163	37	59
Mm_2	Q6P4T2	e−146	35	56	e−163	37	59
Lm	Q4QGA4	e−140	34	54	0.0	41	61
Sc	P32639	e−126	32	50	2e−97	33	52
Mm_1	Q8PZR7	9e−40	28	46	1e−35	26	45
Ss	Q97VY9	3e−38	31	51	4e−41	29	51
Mt	O26901	5e−37	28	45	5e−32	28	47
Tv	Q97AI2	2e−37	29	47	6e−47	29	51

At, Arabidopsis thaliana; Ce, Caenorhabditis elegans; Dm, Drosophila melanogaster; Hs, Homo sapiens; Lm, Leishmania major; Mm_1*, Methanosarcina mazei;* Mm_2*, Mus musculus; Mt, Methanobacterium thermoautotrophicum; Sc, Saccharomyces cerevisiae; Ss, Sulfolobus solfataricus; Tv, Thermoplasma volcanium.* Homology (*H*) and identity (*I*) values are expressed in percentage.

[a] UniProt Knowledgebase database.

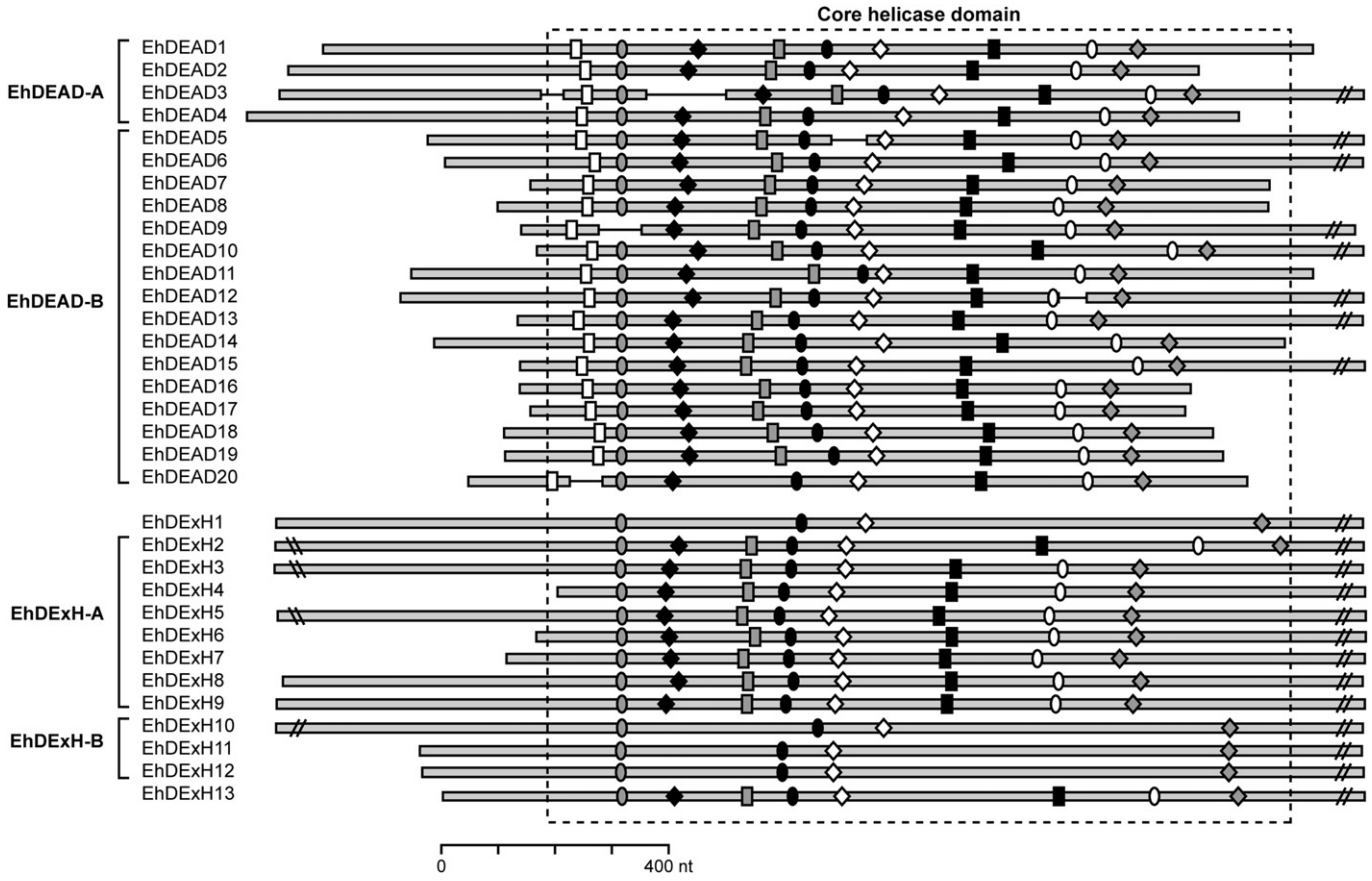

Fig. 6. Molecular organization of *EhDead* and *EhDexh*-box genes. Subfamilies based on phylogenetic study (see Section 3.2 for details) are indicated at the right. Grey bars, encoding regions. Black lines, introns. The core helicase domain position is indicated by the discontinuous line. The scale is shown at the bottom; □, Q motif; ⬯, motif I; ♦, motif Ia; ▨, motif Ib; ⬮, motif II; ◊, motif III; ■, motif IV; ○, motif V; ◈, motif VI.

respectively (Fig. 7). For *EhDead3* and *EhDead12* genes, PCR amplification of genomic DNA generated the expected 522 bp and 318 bp fragments, respectively. When we used cDNA, we obtained smaller fragments that probably correspond to spliced transcripts, confirming the existence of predicted introns in *EhDead3* and *EhDead12* genes. RT-PCR results also indicated that *EhDead3* and *EhDead12* genes are transcribed in trophozoites growing in standard culture conditions, whereas we did not detect any amplified product for *EhDead5, EhDead9 and EhDead20* genes.

4. Discussion

The knowledge of genomes sequences offers an invaluable tool to predict and study proteins families and molecular machineries. The recent completion of the *E. histolytica* genome sequence generated a large raw data set that can be used to gain insights into the molecular biology of this organism. The present work and the accompanying paper by López-Camarillo et al. (2008) are the first reports about RNA helicases in this protozoan parasite. The EhDEAD and EhDExH-box families of *E. histolytica* are similar in size to RNA helicases families described in other eukaryotic organisms (de la Cruz et al., 1999; Boudet et al., 2001; Abdelhaleem et al., 2003), including the protozoan parasites *L. major* (Ivens et al., 2005) and *P. falciparum* (Tuteja and Pradhan, 2006). The 20 EhDEAD and 13 EhDExH-box proteins described here bring new and interesting data about the molecular machinery responsible for the unwinding of RNA secondary structures to produce the functional conformation required for the realization of numerous genetic processes such as transcription, splicing, ribosome biogenesis, pre-mRNA processing, mRNA export, degradation and translation in *E. histolytica*.

E. histolytica RNA helicases possess the conserved motif II (DEAD or DExH) and almost all the additional motifs that are characteristics of these proteins in other eukaryotes. Phylogenetic analysis, as well as sequences and molecular characterization of proteins and genes, allowed us to propose a model for the evolution of EhDEAD and EhDExH-box helicases families. It is possible that EhDEAD-A and EhDEAD-B subfamilies members have evolved from a common gene ancestor without any intron. A first evolutionary event could have led to the separation of EhDEAD-A proteins, a second evolutionary event allowing the formation of EhDEAD-B in a distinct arm. Then, subsequent gene duplication events could have led to the formation of paralogous *EhDead* box genes (*EhDead16-EhDead17* and *EhDead18-EhDead19*). Simultaneously, mutation events and non coding DNA sequence addition, could have created introns in several *EhDead* genes

Table 3
Characteristics of predicted introns in *EhDead* genes

Gene		Intron				
Name	Size (aa)	Location (nt)	Phase	Size (nt)	5′ splicing site	3′ splicing site
EhDead3	1344	454–500	0	47	tacgT**TT**GT**T**	AATAGAatgg
		649–775	1	127	aaac**GTTTGT**	AAT**TAG**atgt
EhDead5	2168	727–782	0	56	tggaAG**T**GT**T**	CCAGCAtggg
EhDead9	1496	139–212	0	74	attgT**T**CAAA	GAAAGAgaaa
EhDead12	1910	1156–1193	0	38	gagaG**G**T**A**T**T**	GGA**T**TAgata
EhDead20	1370	181–230	0	50	tttgT**TT**GT**T**	AATAGAaaag
Consensus splice sites[a]					nnnnGTTTGT	NNNTAGnnnn

Exon sequences are shown in lower case letters; intron sequences are shown in upper case letters; nucleotides conserved in the consensus splice sites are in bold, n, any nucleotide.

[a] Davies et al., 2007.

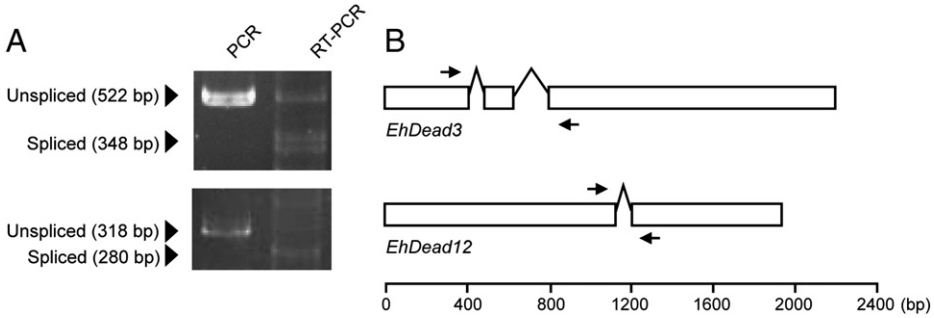

Fig. 7. Comparative PCR and RT-PCR products for *EhDead3 and EhDead12* genes. (A) Genomic DNA (lane 1) and cDNA (lane 2) were PCR amplified using specific exonic primers flanking predicted introns. Products were analyzed in 12% polyacrylamide electrophoresis gel and ethidium bromide stained. (B) Schematic representation of *EhDead3* and *EhDead12* genes. Open box, exon; caret, intron; Arrow, oligonucleotides.

It is possible that *EhDead9* and *EhDead12* genes do not encode functional helicases since motifs II and V, respectively, have been lost during intron formation event. Similarly, EhDExH proteins may be the result of successive evolutionary events that have led to the separation of EhDExH-A and EhDExH-B groups, whereas EhDExH1 and EhDExh-13 remained isolated.

The fact that most EhDEAD/EhDExH proteins do not share high sequence identity suggested that gene duplication within the parasite genome was not the only event that generated this large RNA helicases encoding genes family. Another route for protein evolution is the formation of multidomains proteins by gene fusion. We demonstrated that EhDExH1 and EhDExH10 have two helicase domains that could result from the fusion of two ancestral prokaryotic genes. Moreover, the clustering of homologous fused proteins from phylogenetically distant species showed that this mechanism is an evolutionary conserved process in the eukaryotic lineage of orthologous DExH1 and EhDExH10-box proteins. It also suggested a positive selection for the multidomain architectures (Yanai et al., 2002).

It has been proposed that physically close genes resulting from local genome duplications tend to remain very similar, being homogenized both by unequal recombination and gene conversion (Graham, 1995). This suggests that paralogous helicases could be codified by genes organized in tandem. Therefore, scaffolds 3 and 5 corresponding to *EhDead16* and *EhDead17* genes could be located in the same chromosome region, whereas scaffolds 22 and 283 (*EhDead18* and *EhDead19* genes) may map to another chromosome region.

Finally, to further characterize *E. histolytica* RNA helicases genes expression and function, we demonstrated here that *EhDead3* and *EhDead12* genes are expressed through RT-PCR assays in basal culture conditions. Additionally, we also detected *EhDead1* transcript in the same conditions (Lopez-Camarillo et al., 2008). BLAST analysis of a cDNA library containing 1450 cDNA sequences enriched with mRNA obtained from the highly virulent *E. histolytica* HM1:IMSS strain (Weber et al., 2006) evidenced *EhDead5, EhDead12, EhDead13, EhDead14, EhDead16, EhDead18, EhDead19* and *EhDead20* mRNA expression that could be related to parasite virulence (Supplementary data Table S1). We also took advantage of microarrays results recently published to search whether the regulation of *EhDead* and *EhDeaxh* genes transcription was modulated in several stress conditions, such as heat shock exposure (Weber et al., 2006) or invasion of the intestinal ceacum of mice (Gilchrist et al., 2006). *EhDead5* and *EhDead16* were down regulated by heat shock and interestingly, *EhDead16* was also down regulated during intestinal infection. In addition, EhDEAD16 was found associated to phagosome fractions analyzed by proteomics (Marion et al., 2005) (Supplementary data Table S1). EhDEAD16 is the homologue of eIF4A, a factor that is part of the protein complex involved in the recognition of 5′ UTR sequences in most cellular mRNAs. This complex, called eIF4F, mediates the interaction of mRNA with translation factors. EhDEAD5 is homologous to a RNA helicase that unwinds double strand RNA in a 5′ to 3′ direction, and introduces a secondary structure into a single strand RNA. These results suggested that down regulation of several RNA helicases expression could be a common response of trophozoites submitted to stress conditions, such as heat shock exposure or tissue invasion, in order to control changes in distinct aspects of RNA metabolism, including mRNA translation and turnover, among others, with a concomitant temporary reduction of cell growth or cell cycle arrest, until trophozoites were in more favorable conditions. The relation between RNA helicases expression and cell growth has been well documented in other organisms. The URH49 helicase mRNA level decreases three to six-fold when HeLa cells enter quiescence (Pryor et al., 2004). A nucleolar RNA helicase has been identified as an upregulated gene in human colonic carcinoma by fluorescent mRNA differential display (Ojala et al., 2002). Similarly, the DDX39 DEAD-box RNA helicase is overtranscribed in human lung squamous cell carcinoma and promotes colony formation of HeLa cells (Sugiura et al., 2007).

The fact that the expression of most *EhDead* and *EhDexh* helicase genes was not detected in the experimental data presented here and in previous works, does not mean that they are not expressed in trophozoites. It is possible that they are expressed at low levels or in other specific conditions. Various works showed that RNA helicases from other organisms are differentially transcribed in distinct tissues or development stages, suggesting that each RNA helicase may be particularly important for a specific molecular mechanism. For example, expression levels of the HelTc RNA helicase mRNA are almost eight times higher in *Trypanosoma cruzi* trypomastigotes than in epimastigotes (Diaz Añel et al., 2000). In *Drosophila*, the smaller transcript of the twister RNA helicase gene is more abundant in 0–2 h embryos whereas the larger transcript reaches its highest levels in 6–8 h embryos (Seago et al., 2001). In human, the DEAD-box RNA helicases p68, PL10 and GRTH are expressed at much higher levels in testes than in other tissues (Leroy et al., 1989; Lemaire and Heinlein, 1993; Sheng et al., 2003).

In conclusion, we reported that *E. histolytica* has a large EhDEAD and EhDExH-box RNA helicases family that has been shaped by gene duplication, internal regions duplication, gene fusion, introns formation and motif deletions. The high conservation of the helicase domain suggested that EhDEAD and EhDExH-box proteins could be functional RNA helicases orthologues, however their catalytic activity remains to be studied. Finally, mRNA expression analysis suggested that several *EhDead* and *EhDexh* genes are differentially transcribed in distinct experimental conditions.

Acknowledgements

This work was supported by CONACyT (Mexico), SIP-IPN (Mexico), COFAA-IPN (Mexico), UACM (Mexico) and European Community

grants. We thank Alfredo Padilla and Sollange Archer for their help in the artwork. Our thanks are also to Nathalie Marchau for reviewing the English for grammar.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2008.07.042.

References

Abdelhaleem, M., Maltais, L., Wain, H., 2003. The human DDX and DHX gene families of putative RNA helicases. Genomics 81, 618–622.

Aubourg, S., Kreis, M., Lecharny, A., 1999. The DEAD box RNA helicase family in *Arabidopsis thaliana*. Nucleic Acids Res. 27, 628–636.

Benz, J., Trachsel, H., Baumann, U., 1999. Crystal structure of the ATPase domain of translation initiation factor 4A from *Saccharomyces cerevisiae*—the prototype of the DEAD box protein family. Struct. Fold Des. 7, 671–679.

Boudet, N., Aubourg, S., Toffano-Nioche, C., Kreis, M., Lecharny, A., 2001. Evolution of intron/exon structure of DEAD helicase family genes in *Arabidopsis*, *Caenorhabditis* and *Drosophila*. Genome Res. 11, 2101–2114.

Caruthers, J.M., Johnson, E.R., McKay, D.B., 2000. Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. Proc. Natl. Acad. Sci. U. S. A. 97, 13080–13085.

Chen, C.Y., et al., 2001. AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. Cell 107, 451–464.

Cordin, O., Tanner, N.K., Doere, M., Linder, P., Banroques, J., 2004. The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. EMBO J. 23, 2478–2487.

de la Cruz, J., Kressler, D., Linder, P., 1999. Unwinding RNA in *Saccharomyces cerevisiae*: DEAD-box proteins and related families. Trends Biochem. Sci. 24, 192–198.

Davies, C.A., Brown, M.P.S., Singh, U., 2007. Functional characterization of spliceosomal introns and identification of U2, U4 and U5 snRNAs in the deep-branching eukaryote *Entamoeba histolytica*. Eukaryot. Cell 6, 940–948.

Diaz Añel, A.M., et al., 2000. mRNA encoding a putative RNA helicase of the DEAD-box gene family is up-regulated in trypomastigotes of *Trypanosoma cruzi*. J. Eukaryot. Microbiol. 47, 555–560.

Foresta, C., Moro, E., Ferlin, A., 2001. Y chromosome microdeletions and alterations of spermatogenesis. Endocrinol. Rev. 22, 226–239.

Garcia-Vivas, J., Lopez-Camarillo, C., Azuara-Liceaga, E., Orozco, E., Marchat, L.A., 2005. *Entamoeba histolytica*: cloning and expression of the poly(A) polymerase EhPAP. Exp. Parasitol. 110, 226–232.

Gilchrist, C.A., et al., 2006. Impact of intestinal colonization and invasion on the *Entamoeba histolytica* transcriptome. Mol. Biochem. Parasitol. 147, 163–176.

Gorbalenya, A.E., Koonin, E.V., 1993. Helicases: amino acid sequence comparisons and structure–function relationships. Curr. Opin. Struct. Biol. 3, 419–429.

Graham, G.J., 1995. Tandem genes and clustered genes. J. Theor. Biol. 175, 71–87.

Ivens, A.C., et al., 2005. The genome of the kinetoplastid parasite, *Leishmania major*. Science 309, 436–442.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. 5 (2), 150–163.

Laggerbauer, B., Achsel, T., Lührmann, R., 1998. The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplices *in vitro*. Biochemistry 95, 4188–4192.

Lemaire, L., Heinlein, U.A., 1993. High-level expression in male germ cells of murine P68 RNA helicase mRNA. Life Sci. 52, 917–926.

Leroy, P., Alzari, P., Sassoon, D., Wolgemuth, D., Fellous, M., 1989. The protein encoded by a murine male germ cell-specific transcript is a putative ATP-dependent RNA helicase. Cell 57, 549–559.

Linder, P., Daugeron, M.C., 2000. Are DEAD-box proteins becoming respectable helicases? Nat. Struct. Biol. 7, 97–99.

Linder, P., Stutz, F., 2001. mRNA export: travelling with DEAD box proteins. Curr. Biol. 11, R961–R963.

Linder, P., Tanner, N.K., Banroques, J., 2001. From RNA helicases to RNPases. Trends Biochem. Sci. 26, 339–341.

Loftus, B., et al., 2005. The genome of the protist parasite *Entamoeba histolytica*. Nature 433, 865–868.

Lohman, T.M., Bjornson, K.P., 1996. Mechanisms of helicase-catalyzed DNA unwinding. Annu. Rev. Biochem. 65, 169–214.

López-Camarillo, C., Luna-Arias, J.P., Marchat, L.A., Orozco, E., 2003. *EhPgp5* mRNA stability is a regulatory event in the *Entamoeba histolytica* multidrug resistance phenotype. J. Biol. Chem. 278, 11273–11280.

López-Camarillo, C., Orozco, E., Marchat, L.A., 2005. *Entamoeba histolytica*: comparative genomics of the pre-mRNA 3′ end processing machinery. Exp. Parasitol. 110, 184–190.

López-Camarillo, C., García-Hernández, M.L., Marchat, L.A., Luna-Arias, J.P., Hernandez de la Cruz, Olga, Mendoza, L., Orozco, E., 2008. *Entamoeba histolytica* EhDEAD1 protein is a conserved DEAD-box RNA helicase with ATPase and ATP-dependent RNA unwinding activities. Gene 14, 19–31.

Marcotte, E.M., Pellegrini, M., Ng, H.L., Rice, D.W., Yeates, T.O., Eisenberg, D., 1999. Detecting protein function and protein–protein interactions from genome sequences. Science 285, 751–753.

Marion, S., Laurent, C., Guillen, N., 2005. Signalization and cytoskeleton activity through myosin IB during the early steps of phagocytosis in *Entamoeba histolytica*: a proteomic approach. Cell Microbiol. 7, 1504–1518.

Mourier, T., Feffares, D.C., 2003. Eukaryotic intron loss. Science 300, 1393.

Nakagawa, T., Ogawa, H., 1999. The *Saccharomyces cerevisiae* MER3 gene, encoding a novel helicase-like protein, is required for crossover control in meiosis. EMBO J. 18, 5714–5723.

Ohno, M., Shimura, Y., 1996. A human RNA helicase-like protein, HRH1, facilitates nuclear export of spliced mRNA by releasing the RNA from the spliceosome. Genes Dev. 10, 997–1007.

Ojala, P., Sundstrom, J., Gronroos, J.M., Virtanen, E., Talvinen, K., Nevalainen, T.J., 2002. mRNA differential display of gene expression in colonic carcinoma. Electrophoresis 23, 1667–1676.

Pause, A., Sonenberg, N., 1992. Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. EMBO J. 11, 2643–2654.

Ponting, C.P., 2000. Proteins of the endoplasmic-reticulum-associated degradation pathway: domain detection and function prediction. Biochem. J. 351, 527–535.

Pryor, A., Tung, L., Yang, Z., Kapadia, F., Chang, T.H., Johnson, L.F., 2004. Growth-regulated expression and G0-specific turnover of the mRNA that encodes URH49, a mammalian DExH/D box protein that is highly related to the mRNA export protein UAP56. Nucleic Acids Res. 32, 1857–1865.

Raghunathan, P.L., Guthrie, C., 1998. RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. Curr. Biol. 8, 847–855.

Rocak, S., Linder, P., 2004. DEAD-BOX proteins: the driving forces behind RNA metabolism. Nat. Rev. Cell Biol. 5, 232–241.

Rondon, A.G., Jimeno, S., Garcia-Rubio, M., Aguilera, A., 2003. Molecular evidence that the eukaryotic THO/TREX complex is required for efficient transcription elongation. J. Biol. Chem. 278, 39037–39043.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Schmid, S.R., Linder, P., 1992. D-E-A-D protein family of putative RNA helicases. Mol. Microbiol. 6, 283–291.

Schwer, B., 2001. A new twist on RNA helicases: DExH/D box proteins as RNPases. Nat. Struct. Biol. 8, 113–116.

Seago, J.E., Chernukhin, I.V., Newbury, S.F., 2001. The *Drosophila* gene twister, an orthologue of the yeast helicase SKI2, is differentially expressed during development. Mech. Dev. 106, 137–141.

Sheng, Y., Tsai-Morris, C.H., Dufau, M.L., 2003. Cell-specific and hormone-regulated expression of gonadotropin-regulated testicular RNA helicase gene (GRTH/Ddx25) resulting from alternative utilization of translation initiation codons in the rat testis. J. Biol. Chem. 278, 27796–27803.

Stresser, K., Hurt, E., 2001. Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p. Nature 413, 648–652.

Stresser, K., et al., 2002. TREX is a conserved complex coupling transcription with messenger RNA export. Nature 417, 304–308.

Sugiura, T., Nagano, Y., Noguchi, Y., 2007. DDX39, upregulated in lung squamous cell cancer, displays RNA helicase activities and promotes cancer cell growth. Cancer Biol. Ther. 26 6(6) (Epub).

Tanner, N.K., Linder, P., 2001. DExD/H box RNA helicases: from generic motors to specific dissociation functions. Mol. Cell 8, 251–262.

Tanner, N.K., 2003. The newly identified Q motif of DEAD box helicases is involved in adenine recognition. Cell Cycle 2, 18–19.

Tanner, N.K., Cordin, O., Banroques, J., Doere, M., Linder, P., 2003. The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. Mol. Cell 11, 127–138.

Tuteja, R., Pradhan, A., 2006. Unraveling the 'DEAD-box' helicases of *Plasmodium falciparum*. Gene 376, 1–12.

Weber, C., et al., 2006. Stress by heat shock induces massive down regulation of genes and allows differential allelic expression of the Gal/GalNAc lectin in *Entamoeba histolytica*. Eukaryot. Cell 5, 871–875.

Yanai, I., Wolf, Y.I., Koonin, E.V., 2002. Evolution of gene fusions: horizontal transfer versus independent events. Genome Biol. 3 research 0024.1–research 0024.13.

Young, B.P., Craven, R.A., Reid, P.J., Willer, M., Stirling, C.J., 2001. Sec63p and Kar2p are required for the translocation of SRP-dependent precursors into the yeast endoplasmic reticulum in vivo. EMBO J. 20, 262–271.