Supplementary Information to:

A functional study of genes essential for autogamy and nuclear reorganization in *Paramecium*

Jacek K. Nowak et al.

SUPPLEMENTARY MATERIALS AND METHODS

Spotting and hybridization of dedicated Megabase DNA microarrays

PCR fragments (100ng/µL) were spotted on Corning Gaps II slides, according to (21), with the Genetac G3 arrayer (Genomic Solutions). Quality control was performed before and after spotting by scanning each slide with an Axon 4000B scanner and slides with excessive autofluorescence were discarded.

cDNAs from *P. tetraurelia* d4-2 at different stages of autogamy (Figure S1) were synthesized and indirectly labeled by Cy3- or Cy5-dye coupling using the Fairplay Microarray Labeling Kit (Stratagene). As a tentative hybridization reference, *Eco*RI-digested total genomic DNA from the same strain was indirectly labeled with a 1:4 mixture of dTTP and amino allyl-dUTP (AA-dUTP; GE Healthcare), using the Prime-a-gene Labeling kit (Promega). Cy5- or Cy3-dye coupling was carried out with the Fairplay Microarray Labeling Kit (Stratagene).

Hybridization was performed using the Genetac Hybstation (Genomic solutions) or manually at 42°C in Denhardt Solution containing 70% formamide and 0.7% SDS. Slides were submitted to 3 successive washes (2X SSC 0.1% SDS / 0.5X SSC 0.1% SDS / 0.1X SSC, respectively) before scanning. Three technical replicates were performed for each RNA preparation: two with Cy3-labeled cDNAs and one dye-swap with Cy5 (GEO submission #GSE12988).

Image acquisition and processing

Hybridized slides were scanned using a GenePix 4000B scanner equipped with a laser to excite Cy3 and Cy5 (10 µm resolution, 100% laser power and AutoPMT). The resulting 16-bit images were analyzed using GenePix Pro 6.0® software. Raw data overview and pre-processing were performed using MAnGO software (16), before morphological background subtraction. The signals obtained by hybridization of total genomic DNA were very heterogeneous among the arrays and, therefore, were not used as an inter-array reference. Instead, single-color normalization was performed using the normalizeQuantiles function of the R limma package (5) (see GSE12988_rawdatatrieparID.xls file submitted to GEO). For each array, we calculated the arithmetic mean of normalized signals for all spots corresponding to a given probe (GEO submission #GSE12988). For further statistical analysis, individual normalized signals (3-6 spots per array and 3 replicate arrays) were averaged for each time-point, leading to a mean profile for each probe throughout the time-course experiment. K-means clustering (11) with 20 groups was computed from all mean profiles, by using $1-r^2$ (where r is the Pearson correlation) as a similarity measure and a centroid-based search as a cluster initialization method (Table S1). Among the 20 groups of probes obtained following K-means clustering, some showed similar profiles and were pooled into four expression clusters. (Figure 1A). Within each cluster, probes with the largest induction or repression factors were selected by applying the following filters: T0/TV>2 for cluster I (30 probes), T11/TV>2 for cluster II (23 probes), T20/TV>2 for cluster III (20 probes) and T0/TV<0.5 for cluster IV (31 probes).

Silencing constructs

All RNAi plasmids are derivatives of vector L4440 (26) and carry a target gene fragment inserted between two convergent T7 promoters (Megabase inserts used in this study are listed in Table S6). Control RNAi plasmids were p0ND7c (8) and pICL7a (9), which target non essential *ND7* and *ICL7a* genes, respectively.

GFP fusion constructs

Plasmid p344-GFP encoding a C-terminal GFP fusion to Ptmb344p was obtained in two steps. First, a 2627-bp fragment containing the whole *PTMB.344c* coding sequence with its upstream 96 bp and downstream 121 bp (bp 695245-692938 from accession number CR548612) was cloned into pCRScriptAmpSK+ vector (Stratagene). Then, the EGFP coding sequence optimized for *Paramecium* codon usage (18) was inserted into a *Hpa*I site near the 3'-end of the gene (between codons encoding L769 and T770).

For the construction of plasmid p220-GFP5' encoding an N-terminal GFP fusion to Ptmb.220p, a 5153-bp fragment carrying the whole *PTMB.220* coding sequence with its upstream 527 bp and downstream 182 bp (bp 483140-488292 from gene accession number CR548612) was cloned into pCRScriptAmpSK+. Then, the EGFP coding sequence was inserted into an *Afl*II site near the 5'-end of the gene (between codons encoding Q14 and V15).

Plasmid p170-ZCΔRIX encoding an N-terminal GFP fusion to Ptmb.170p was constructed by inserting a 4800-bp fragment containing the *PTMB.170c* open reading frame and its potential 251-bp terminator region (bp 368611- 363812 from accession number CR548612) between the *BamH*I and *Pst*I sites of plasmid pZCΔRIX (kindly provided by E. Meyer & S. Malinsky), directly downstream of the GFP coding sequence expressed under the control of the constitutive promoter of *P. primaurelia* surface antigen G^{156} gene (bp 1-290 from accession number X03882).

Before microinjection, all plasmids were purified using a QIAfilter Plasmid Maxi Kit (Qiagen) and linearized within the vector sequence with *Xmn*I (p344-GFP), *Sca*I (p220-GFP5') or *Bgl*I (p170-ZCΔRIX). They were filtered through a 0.22 µm Ultrafree-MC filter (Millipore), precipitated with ethanol and dissolved in filtered water to a final concentration of 5 µg/µL. All plasmid sequences are available upon request.

LIST OF SUPPLEMENTARY FIGURES

Figure S1. Autogamy time-course of strain d4-2 *(see legend below figure)*

Figure S2. Biological replicates of the autogamy time-course in strain 51 *(see legend below figure)*

Figure S3. Northern blot analysis of full-size mRNA in cells submitted to RNAi and control cultures *(see legend below figure)*

Figure S4. Protargol-staining of *PTMB.08c*-silenced cells *(see legend below figure)*

Figure S5. Multiple alignment of Ptmb.422p with proteins sharing similar Guanylate-Binding Protein (GBP) domains

Sequences were aligned using ClustalW (7, 14) with default parameters. Alignment was colored using Boxshade at http://www.ch.embnet.org/software/BOX_form.html (grey: similar residues; black: identical residues; fraction of aligned residues that must agree for shading: 0.4).

Consensus sequences of the G-domain motifs are based on motifs found for GBP proteins (19) and structural annotations are based on the structure of human Guanylate Binding Protein-1 (PDB accession code 1F5N, see (20)).

Abbreviations and accession numbers of all sequences are as follows:

Tt: *Tetrahymena thermophila* TTHERM_01211770 - XP_001018996, TTHERM_00657240 - XP_001024024

At: *Arabidopsis thaliana* AT5G46070 - NP_199419

Hs: *Homo sapiens* GBP1 - NP_002044, GBP3 - NP_060754, GBP5 - NP_443174 and GBP4 - NP_443173

Mm: *Mus musculus* GBP2 - NP_034390

Dr: *Danio rerio* LOC570622- XP_699217

Dd: *Dictyostelium discoideum* DDB0235201 - XP_640552

Figure S6. Alignment of Cks proteins from *P. tetraurelia* and other organisms

Sequences were aligned using ClustalW and colored using Boxshade (fraction of aligned residues that must agree for shading: 0.5).

The different structural motifs and important residues are marked according to their annotation in (6): orange – residues interacting with CDK2; blue – residues from the anion binding site. Positions of β strands and α helices are indicated.

Abbreviations and accession numbers are as follows:

Ce: *Caenorhabditis elegans* Cks-1 - NP_501457, Cks-2 - NP_490896

At: *Arabidopsis thaliana* CKS1 - NP_180363, CKS2 - NP_180364

Hs: *Homo sapiens* CKS1B - NP_001817, CKS2 - NP_001818

Dm: *Drosophila melanogaster* Cks30A - NP_476947, Cks85A - NP_649817

Sc: *Saccharomyces cerevisiae* Cks1 - NP_009693

Sp: *Schizosaccharomyces pombe* suc1 - NP_595431

Figure S7. Multiple alignment of Ptmb.344p and Msh2 proteins from different species

Sequences were aligned using ClustalW and colored using Boxshade (fraction of aligned residues that must agree for shading: 0.5). All structural annotations are based on the structure of *E. coli* MutS protein (PDB accession code 1E3M_A, see (13)).

Abbreviations and accession numbers are as follows:
At: *Arabidopsis thaliana* - NP_566804
Zm: *Zea mays* - Q9XGC9
Hs: *Homo sapiens* - NP_000242
Dr: *Danio rerio* - NP_998689
Dm: *Drosophila melanogaster*: - NP_523565
Ce: *Caenorhabditis elegans* - NP_491202
Mg: *Magnaporthe grisea* - XP_368365
Nc: *Neurospora crassa* - XP_959643
Sc: *Saccharomyces cerevisiae* - NP_014551
Dd: *Dictyostelium discoideum* - XP_643399
Tb: *Trypanosoma brucei* - XP_823311
Tc: *Trypanosoma cruzi* - XP_819877
Pf: *Plasmodium falciparum* - XP_001348428
Py: *Plasmodium yoelii yoelii* - XP_728672
Tt: *Tetrahymena thermophila* - XP_001013224
Pt: *Paramecium tetraurelia* – PTMB.344c

Figure S8. Immunolabeling of *PTMB.170c*-silenced autogamous cells with anti-γ tubulin antibodies *(see legend below figure)*

Figure S9. Alignment of the helicase domain of Ptmb.220p with other similar DExH-box helicases

Full protein sequences were aligned using T-Coffee (17) with default parameters. The alignment was colored using Boxshade (fraction of aligned residues that must agree for shading: 0.3). Only the region encompassing the helicase domain is shown, since N and C-terminal parts of the proteins did not give significant alignment.

Structural annotation is based on the structure of *Bacillus caldotenax* UvrB protein (PDB accession code 1D9Z_A, see (25)).

Consensus amino acid sequence of the 8 helicase motifs constituting the helicase domain of DEAH family genes based on (24).

Abbreviations and accession numbers are as follows:
Tt: *Tetrahymena thermophila* Ema1p - BAF49175
Eh: *Entamoeba histolytica* EHI_090040 - XM_643619
At: *Arabidopsis thaliana* NIH- NP_172152; HVT1 - NP_850154
Hs: *Homo sapiens* DHX29 - NP_061903; DHX57 - NP_945314; TDRD9 - NP_694591; RHA - NP_001348
Dm: *Drosophila melanogaster* CG1582 - NP_572663

LIST OF SUPPLEMENTARY TABLES

Table S1 *(Excel file)*. K-means clustering of Megabase microarray data
All probes spotted on customized plates are listed in the first column (GEO submission # GPL7296). Mean values of the normalized hybridization signals obtained for each time-point are displayed in the following four columns (linear values). Probes are ordered according to their K-means group (last column).

Tables S2-S5 *(Excel file)*. Lists of probes in clusters I to IV and human curated annotation of corresponding genes
For each gene shown on these tables, expert human curation of gene models was performed using the Apollo software (15) for scaffold_1, which corresponds to the Genoscope assembly of the Megabase chromosome (2). PTMB numbers refer to Megabase accession numbers (27), GSPATG to Genoscope automatic annotation and PTETG to human curated gene models that can be found in the ParameciumDB database at the following URL: http://paramecium.cgm.cnrs-gif.fr/ (1).
Statistical validation of individual gene expression profiles was performed using whole-genome microarrays as described in the legend to Tables S7-S10. The genes covered by the Megabase microarray probes, that followed the characteristic expression patterns identified for each cluster, are indicated by a « + » symbol in the last column of Tables S2-S5.

Table S6. Northern blot hybridization probes and feeding inserts used in this study

Tables S7-S10 *(Excel file)*. Statistical analysis of expression profiles on NimbleGen microarrays for genes included in clusters I to IV
Whole-genome oligonucleotide microarrays were designed and hybridized at NimbleGen, with six 50-mer oligonucleotide probes for each gene (GEO submission # GPL7221; Arnaiz et al. submitted). cDNAs were prepared from two independent autogamy time-course experiments performed with 51 (Figures S2B and C). Raw signals from the 12 hybridizations (6 for each biological replicate) were normalized simultaneously using the normalizeBetweenArrays function from the Limma package (23) (GEO submission # GSE17998). For each microarray, the expression level of each gene was taken as the median of the normalized signals obtained for its 6 individual probes. Mean expression levels (mV, mS, m0, m5, m11 and m20) are log2 values corresponding to the mean of the two replicates for each time-point.
To compare changes in expression level between time-points, the difference between mean expression levels was calculated to give logFC. Statistical significance of changes in expression level was estimated using linear modeling and empirical Bayes statistics implemented in the lmFit and eBayes functions from the Limma package (22). To minimize the detection of false positives, adjusted p-values were computed after correction for multiple testing using the Benjamini and Hochberg method (3).

Table S11. Ohnologs and other paralogs of silenced genes
Ohnologs resulting from the known WGDs were identified using the data available in ParameciumDB. Other paralogs were found by Blastp searches and proteins showing amino acid identity higher than 40% are indicated.

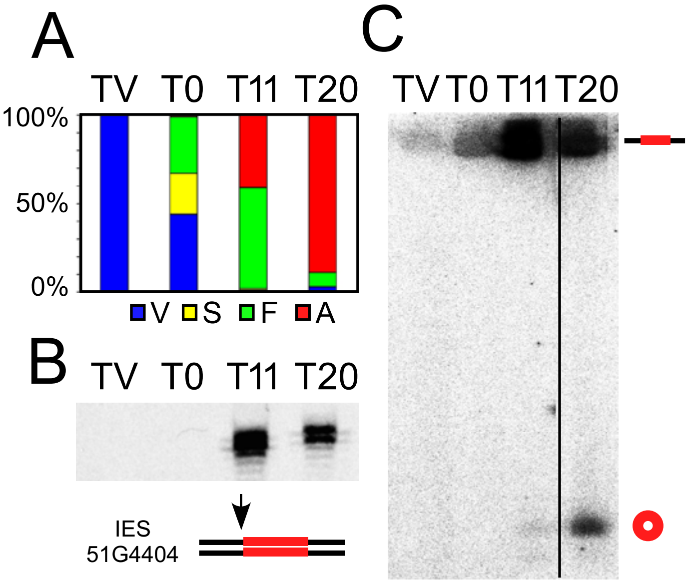


Figure S1. Autogamy time-course of strain d4-2

A. Histograms showing the progression of autogamy during starvation of d4-2. For each time-point (TV: vegetative culture; T0: 50% of cells with fragmented MAC; T11 and T20: 11 and 20 hours following T0, respectively), cells were stained with DAPI to visualize old and new MACs. V: vegetative parental MAC; S: skein formation; F: fragmented old MAC but no detectable developing new MACs; A: fragmented old MAC + 2 visible anlagen.

B. Double-strand break detection at the left end of IES 51G4404 by LMPCR. Experimental details can be found in (10).

C. Detection of circular excised IES 51G4404 molecules by Southern blot hybridization of total genomic DNA. Experimental procedure was described in (4).

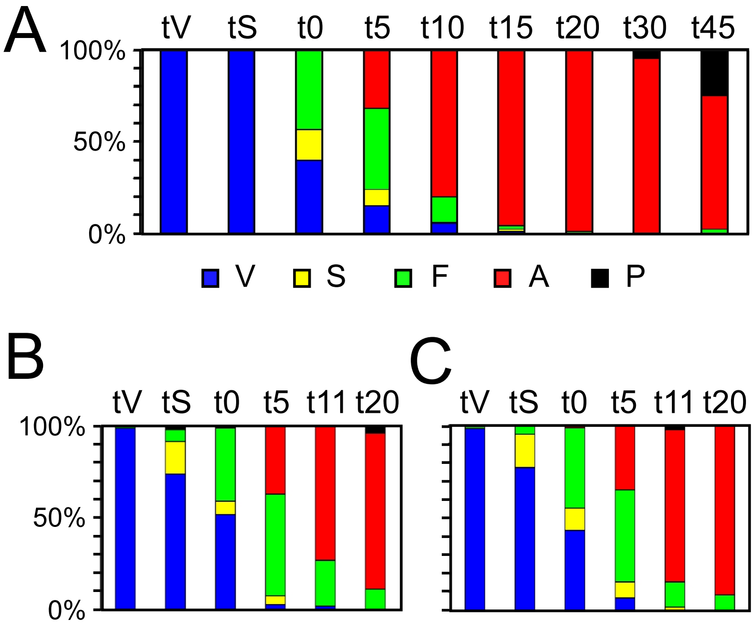


Figure S2. Biological replicates of the autogamy time-course in strain 51

A. Progression of autogamy in the culture of 51 mt8 used for the Northern blot analysis of mRNAs shown in Figure 1.

B and C. Progression of autogamy in two independent cultures of 51 mt8 used in NimbleGen microarray experiments.

Indicated time-points: tV for vegetative cultures ; tS for starved or meiotic cells with intact parental MAC; t0 for 50% of cells with fragmented MAC; the following time-points refer to hours following t0.

At each time-point, cells were stained with DAPI to monitor autogamy stages. The color code is the same as in Figure S1 for V, S, F and A. P: post-karyonidal division (1 anlagen + fragmented old MAC).

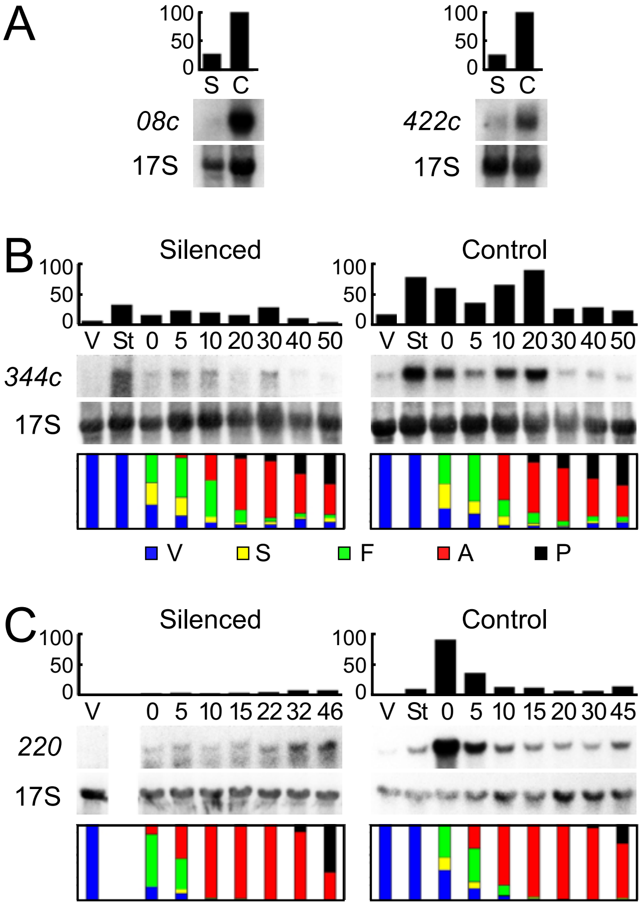


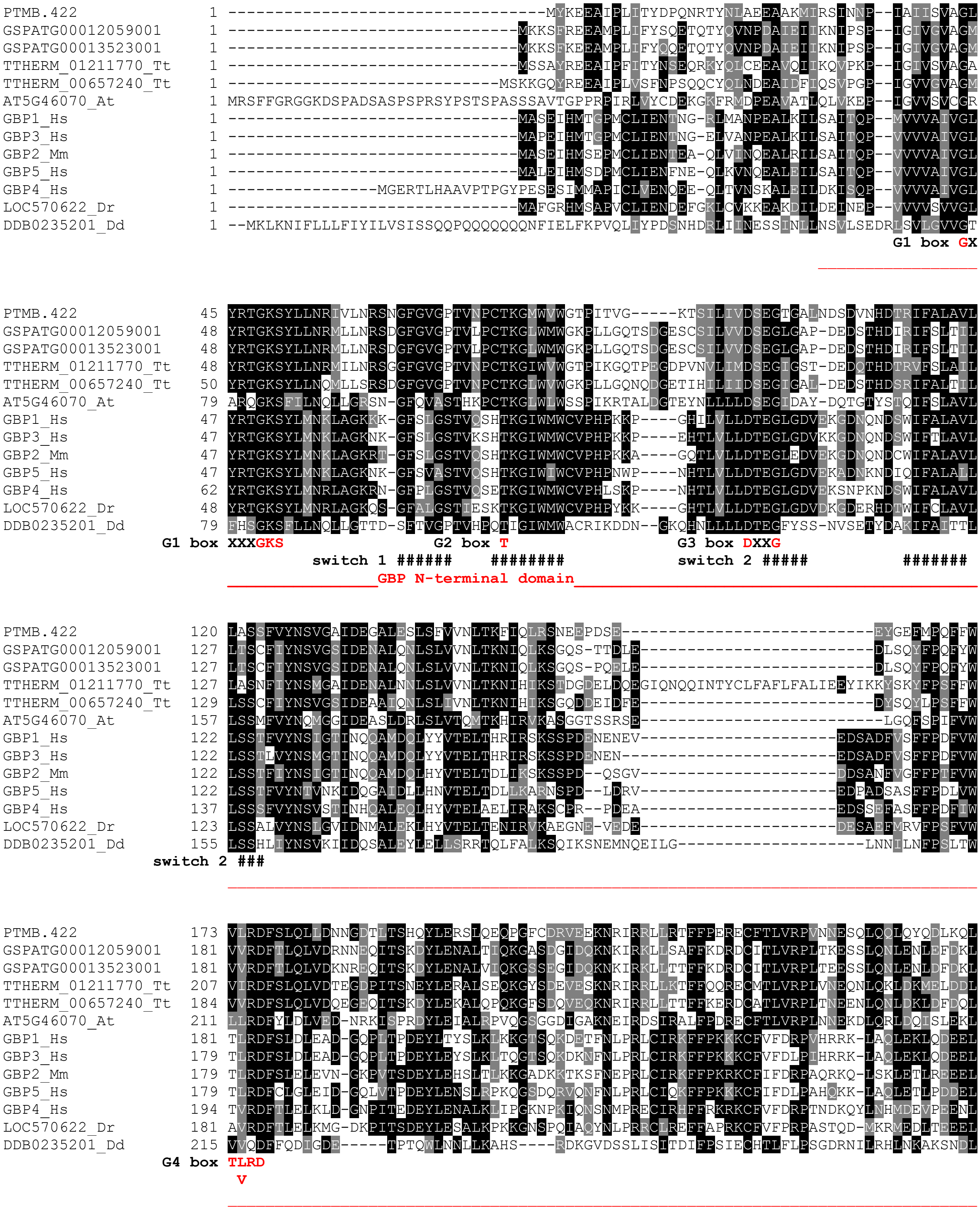
Figure S3. Northern blot analysis of full-size mRNA in cells submitted to RNAi and control cultures
A. Analysis of total RNA extracted from vegetative cells following 26-30 hours of incubation in *PTMB.08c* (left panel) or *PTMB.422c* (right) silencing medium (S). Control RNAi (C) was against *ND7* gene. Hybridization probes are indicated on the left of each panel. Normalized amounts of each mRNA relative to 17S rRNA are shown on top, with an arbitrary unit for vertical axis
B. Autogamy time-course of *PTMB.344c* (Silenced) and *ND7* (Control) silenced cells. Northern blots of total RNA were hybridized successively with a *PTM344c* probe and a 17S rDNA probe. Normalized amounts of *PTMB344c* mRNA are shown on top. V: vegetative cells; St: starved cells; following time-points are in hours. Histograms at the bottom represent the progression of autogamy in the cultures (same color code as in Figure S2).
C. Autogamy time-course of *PTMB.220* silenced cells (left). Control time-course (right) was performed with a culture grown in standard *K. pneumoniae* medium (see Figure S2A).
In all panels, the signal corresponding to full-length mRNA was normalized relative to 17S rRNA signal (histograms on top of each panel). The silencing of *PTMB.186* could not be checked by northern blot hybridization because the whole coding sequence was inserted into the RNAi plasmid and bacterial RNA gave a strong hybridization signal when probed with a *PTMB.186* fragment. *PTMB.170* silencing was not quantified either, due to the very low transcription level detected in wild-type cells.

Figure S4. Protargol-staining of *PTMB.08c*-silenced cells

After 10 hours of incubation in *PTMB.08c*-silencing medium, 41 cells were fixed and stained with Protargol protein-staining reagent to document the loss of DNA and the absence of the whole macronuclear structure. 27% of cells were small and did not have any visible nuclei, 20% harbored remnants of nuclei in the form of few dispersed dots (see bottom left cell on the picture) and 53% exhibited apparently normal MACs (top and right cells).

Unannotated arrows point at some small nuclear structures, most probably corresponding to the small dots visible in DAPI-stained cells (see Figure 2 panels a-a’). OA: oral apparatus; A: anterior, P: posterior.

Nowak et al. Figure S5



```
PTMB.422               1 ----------------------------------MYKEEAIPLITYDPQNRTYNLAEEAAKMIRSINNP--IAIISVAGL
GSPATG00012059001      1 -------------------------------MKKSFREEAMPLIFYSQETQTYQVNPDAIEIIKNIPSP--IGIVGVAGM
GSPATG00013523001      1 -------------------------------MKKSFKEEAMPLIFYQQETQTYQVNPDAIEIIKNIPSP--IGIVGVAGM
TTHERM_01211770_Tt     1 -------------------------------MSSAYREEAIPFITYNSEQRKYQLCEEAVQIIKQVPKP--IGIVSVAGA
TTHERM_00657240_Tt     1 -----------------------------MSKKGQYREEAIPLVSFNPSQQCYQLNDEAIDFIQSVPGP--IGVVGVAGM
AT5G46070_At           1 MRSFFGRGGKDSPADSASPSPRSYPSTSPASSSAVTGPPRPIRLVYCDEKGKFRMDPEAVATLQLVKEP--IGVVSVCGR
GBP1_Hs                1 -------------------------------MASEIHMTGPMCLIENTNG-RLMANPEALKILSAITQP--MVVVAIVGL
GBP3_Hs                1 -------------------------------MAPEIHMTGPMCLIENTNG-ELVANPEALKILSAITQP--VVVVAIVGL
GBP2_Mm                1 -------------------------------MASEIHMSEPMCLIENTEA-QLVINQEALRILSAITQP--VVVVAIVGL
GBP5_Hs                1 -------------------------------MALEIHMSDPMCLIENFNE-QLKVNQEALEILSAITQP--VVVVAIVGL
GBP4_Hs                1 ----------------MGERTLHAAVPTPGYPESESIMMAPICLVENQEE-QLTVNSKALEILDKISQP--VVVVAIVGL
LOC570622_Dr           1 -------------------------------MAFGRHMSAPVCLIENDEFGKLCVKKEAKDILDEINEP--VVVVSVVGL
DDB0235201_Dd          1 --MKLKNIFLLLFIYILVSISSQQPQQQQQQQNFIELFKPVQLIYPDSNHDRLIINESSINLLNSVLSEDRLSVLGVVGT
                                                                                               G1 box GX
                                                                                               _________

PTMB.422              45 YRTGKSYLLNRIVLNRSNGFGVGPTVNPCTKGMWVWGTPITVG-----KTSILIVDSEGTGALNDSDVNHDTRIFALAVL
GSPATG00012059001     48 YRTGKSYLLNRMLLNRSDGFGVGPTVLPCTKGLWMWGKPLLGQTSDGESCSILVVDSEGLGAP-DEDSTHDIRIFSLTIL
GSPATG00013523001     48 YRTGKSYLLNRMLLNRSDGFGVGPTVLPCTKGLWMWGKPLLGQTSDGESCSILVVDSEGLGAP-DEDSTHDIRIFSLTIL
TTHERM_01211770_Tt    48 YRTGKSYLLNRMILNRSGGFGVGPTVNPCTKGIWVWGTPIKGQTPEGDPVNVLIMDSEGIGST-DEDQTHDTRVFSLAIL
TTHERM_00657240_Tt    50 YRTGKSYLLNQMLLSRSDGFGVGPTVNPCTKGLWVWGKPLLGQNQDGETIHILIIDSEGIGAL-DEDSTHDSRIFALTIL
AT5G46070_At          79 ARQGKSFILNQLLGRSN-GFQVASTHKPCTKGLWLWSSPIKRTALDGTEYNLLLLDSEGIDAY-DQTGTYSTQIFSLAVL
GBP1_Hs               47 YRTGKSYLMNKLAGKKK-GFSLGSTVQSHTKGIWMWCVPHPKKP----GHILVLLDTEGLGDVEKGDNQNDSWIFALAVL
GBP3_Hs               47 YRTGKSYLMNKLAGKNK-GFSLGSTVKSHTKGIWMWCVPHPKKP----EHTLVLLDTEGLGDVKKGDNQNDSWIFTLAVL
GBP2_Mm               47 YRTGKSYLMNKLAGKRT-GFSLGSTVQSHTKGIWMWCVPHPKKA----GQTLVLLDTEGLEDVEKGDNQNDCWIFALAVL
GBP5_Hs               47 YRTGKSYLMNKLAGKNK-GFSVASTVQSHTKGIWIWCVPHPNWP----NHTLVLLDTEGLGDVEKADNKNDIQIFALALL
GBP4_Hs               62 YRTGKSYLMNRLAGKRN-GFPLGSTVQSETKGIWMWCVPHLSKP----NHTLVLLDTEGLGDVEKSNPKNDSWIFALAVL
LOC570622_Dr          48 YRTGKSYLMNRLAGKQS-GFALGSTIESKTKGIWMWCVPHPYKK----GHTLVLLDTEGLGDVDKGDERHDTWIFCLAVL
DDB0235201_Dd         79 FHSGKSFLLNQLLGTTD-SFTVGPTVHPQTIGIWMWACRIKDDN--GKQHNLLLLDTEGFYSS-NVSETYDAKIFAITTL
                  G1 box XXXGKS                G2 box T                  G3 box DXXG
                                  switch 1 ######    ########            switch 2 #####          #######
                         ________________GBP N-terminal domain___________________________________________

PTMB.422             120 LASSFVYNSVGAIDEGALESLSFVVNLTKFIQLRSNEEPDSE---------------------------EYGEFMPQFFW
GSPATG00012059001    127 LTSCFIYNSVGSIDENALQNLSLVVNLTKNIQLKSGQS-TTDLE-------------------------DLSQYFPQFYW
GSPATG00013523001    127 LTSCFIYNSVGSIDENALQNLSLVVNLTKNIQLKSGQS-PQELE-------------------------DLSQYFPQFYW
TTHERM_01211770_Tt   127 LASNFIYNSMGAIDENALNNLSLVVNLTKNIHIKSTDGDELDQEGIQNQQINTYCLFAFLFALIEEYIKKYSKYFPSFFW
TTHERM_00657240_Tt   129 LSSCFIYNSVGSIDEAAIQNLSLIVNLTKNIHIKSGQDDEIDFE-------------------------DYSQYLPSFFW
AT5G46070_At         157 LSSMFVYNQMGGIDEASLDRLSLVTQMTKHIRVKASGGTSSRSE--------------------------LGQFSPIFVW
GBP1_Hs              122 LSSTFVYNSIGTINQQAMDQLYYVTELTHRIRSKSSPDENENEV---------------------EDSADFVSFFPDFVW
GBP3_Hs              122 LSSTLVYNSMGTINQQAMDQLYYVTELTHRIRSKSSPDENEN-----------------------EDSADFVSFFPDFVW
GBP2_Mm              122 LSSTFIYNSIGTINQQAMDQLHYVTELTDLIKSKSSPD--QSGV---------------------DDSANFVGFFPTFVW
GBP5_Hs              122 LSSTFVYNTVNKIDQGAIDLLHNVTELTDLLKARNSPD--LDRV---------------------EDPADSASFFPDLVW
GBP4_Hs              137 LSSSFVYNSVSTINHQALEQLHYVTELAELIRAKSCPR--PDEA---------------------EDSSEFASFFPDFIW
LOC570622_Dr         123 LSSALVYNSLGVIDNMALEKLHYVTELTENIRVKAEGNE-VEDE---------------------DESAEFMRVFPSFVW
DDB0235201_Dd        155 LSSHLIYNSVKIIDQSALEYLELLSRRTQLFALKSQIKSNEMNQEILG--------------------LNNILNFPSLTW
                switch 2  ###
                         ________________________________________________________________________________

PTMB.422             173 VLRDFSLQLLDNNGDTLTSHQYLERSLQEQPGFCDRVEEKNRIRRLLRTFFPERECFTLVRPVNNESQLQQLQYQDLKQL
GSPATG00012059001    181 VVRDFTLQLVDRNNEQITSKDYLENALTIQKGASDGIDQKNKIRKLLSAFFKDRDCITLVRPLTKESSLQNLENLEFDKL
GSPATG00013523001    181 VVRDFTLQLVDKNREQITSKDYLENALVIQKGSSEGIDQKNKIRKLLTTFFKDRDCITLVRPLTEESSLQNLENLDFDKL
TTHERM_01211770_Tt   207 VIRDFSLQLVDTEGDPITSNEYLERALSEQKGYSDEVESKNRIRRLLKTFFQQRECMTLVRPLVNEQNLQKLDKMELDDL
TTHERM_00657240_Tt   184 VVRDFSLQLVDQEGEQITSKDYLEKALQPQKGFSDQVEQKNRIRRLLTTFFKERDCATLVRPLTNEENLQNLDKLDFDQL
AT5G46070_At         211 LLRDFYLDLVED-NRKISPRDYLEIALRPVQGSGGDIGAKNEIRDSIRALFPDRECFTLVRPLNNEKDLQRLDQISLEKL
GBP1_Hs              181 TLRDFSLDLEAD-GQPLTPDEYLTYSLKLKKGTSQKDETFNLPRLCIRKFFPKKKCFVFDRPVHRRK-LAQLEKLQDEEL
GBP3_Hs              179 TLRDFSLDLEAD-GQPLTPDEYLEYSLKLTQGTSQKDKNFNLPRLCIRKFFPKKKCFVFDLPIHRRK-LAQLEKLQDEEL
GBP2_Mm              179 TLRDFSLELEVN-GKPVTSDEYLEHSLTLKKGADKKTKSFNEPRLCIRKFFPKRKCFIFDRPAQRKQ-LSKLETLREEEL
GBP5_Hs              179 TLRDFCLGLEID-GQLVTPDEYLENSLRPKQGSDQRVQNFNLPRLCIQKFFPKKKCFIFDLPAHQKK-LAQLETLPDDEL
GBP4_Hs              194 TVRDFTLELKLD-GNPITEDEYLENALKLIPGKNPKIQNSNMPRECIRHFFRKRKCFVFDRPTNDKQYLNHMDEVPEENL
LOC570622_Dr         181 AVRDFTLELKMG-DKPITSDEYLESALKPKKGNSPQIAQYNLPRRCLREFFAPRKCFVFPRPASTQD-MKRMEDLTEEEL
DDB0235201_Dd        215 VVQDFFQDIGDE-----TPTQWLNNLLKAHS----RDKGVDSSLISITDIFPSIECHTLFLPSGDRNILRHLNKAKSNDL
                  G4 box TLRD
                          V
                         ________________________________________________________________________________
```

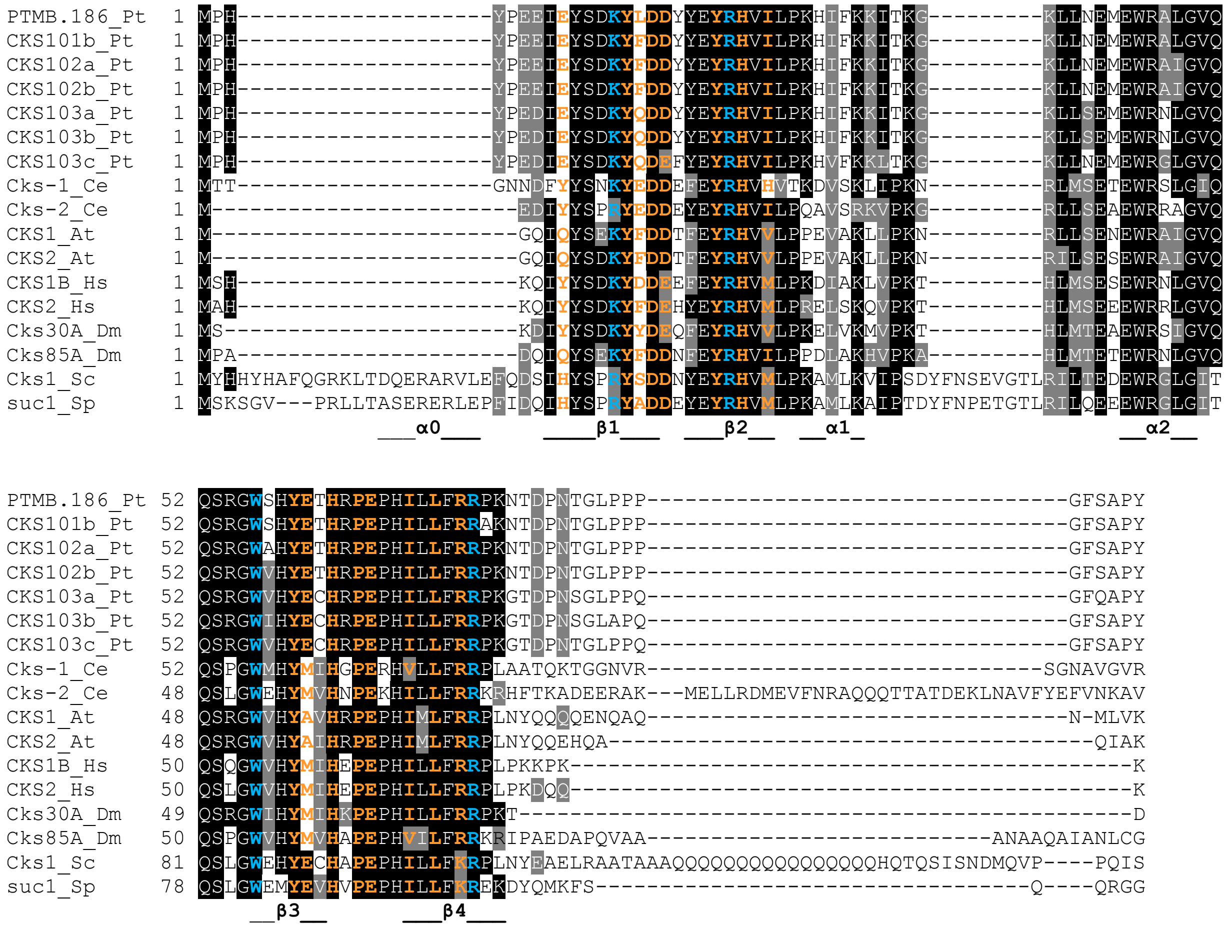
```
PTMB.422            253 RPEFYEQAISLRKMLFSKVKPKQFR--GKTLTGGMYINLINSYLQAMNSGCVATIESSWTYVFREECRKLQAEAFEIYEK
GSPATG00012059001   261 RPEFFEQVINLRKKILNRIRPKTMN--GKTLSGQMYCDLIKSYVTAINNGAVPAIESAWTYICKNECQKAVAEAFDTYEQ
GSPATG00013523001   261 RPEFFEQVINLRKKILNRIRPKMLN--GKTLSGQMYCDLIRSYVTAINNGAVPAIESAWTYICKNECQKAVAEAFDTYEQ
TTHERM_01211770_Tt  287 RPEFFEQVIQLRKKILQRIKPKNLN--GKLLNGEMIMTLCDSYVRAINNGAVPNIENAWTYVCKDECYKASQKAFESYEK
TTHERM_00657240_Tt  264 RPEFFEQVINLRKSILCRMKCKQLR--GKNLTGEMYCDLIKSYVFAINNGAVPNVENAWSYICKNECHKAQQSAIENYDQ
AT5G46070_At        290 RPEFGAGLDAFTKFVFEKTRPKQLG--GTVMTGPILVGITQSYLDALNNGAVPTITSSWQSVEETECRRAYDSGVEAYMA
GBP1_Hs             259 DPEFVQQVADFCSYIFSNSKTKTLSG-GIQVNGPRLESLVLTYVNAISSGDLPCMENAVLALAQIENSAAVQKAIAHYEQ
GBP3_Hs             257 DPEFVQQVADFCSYIFSNSKTKTLSG-GIKVNGPRLESLVLTYINAISRGDLPCMENAVLALAQIENSAAVQKAIAHYDQ
GBP2_Mm             257 CGEFVEQVAEFTSYILSYSSVKTLCG-GIIVNGPRLKSLVQTYVGAISNGSLPCMESAVLTLAQIENSAAVQKAITHYEE
GBP5_Hs             257 EPEFVQQVTEFCSYIFSHSMTKTLPG-GIMVNGSRLKNLVLTYVNAISSGDLPCIENAVLALAQRENSAAVQKAIAHYDQ
GBP4_Hs             273 ERHFLMQSDNFCSYIFTHAKTKTLRE-GIIVTGKRLGTLVVTYVDAINSGAVPCLENAVTALAQLENPAAVQRAADHYSQ
LOC570622_Dr        259 DSEFLQQANTFCSYIYNSTEPKTVKG-GRTITGTALGNLAEVYVEAIRSGKVPCLENAVVSLAKIQNVRAVEQALQIYMT
DDB0235201_Dd       286 NPIYIKELNELKSTLFKNDNDNGGGGKSKLTGPSISTLLRLLIEVANGNKFPTVPSIWSGFIKQQQQSALEDSIIAYKD
                                                       GBP, C-terminal domain

PTMB.422            331 QIKETLLS----KLPTTLEDVKALNKICKSKAIAHMGNNAE--IIP---YVNELKQKIKGKFENLLAFNETESMALCKQY
GSPATG00012059001   339 ILKENLHN----RFPISNEDLKAFNRSLKEQAFALFKKKCV--GEY-EEFKNELSKRIKNRFAAVKQENDREGSRMCSQF
GSPATG00013523001   339 ILKENLHN----RFPISNEDLKAFNRSLREQAFALFKKKCV--GDA-DEYKLELQRRIKQRFAAVKQENDREGSRMCSQF
TTHERM_01211770_Tt  365 TIKEMIGS----KIPTTNDEIKIAHRTAKETALEIFKKRAI--GDVSKEHLEELNKKMKARVQQIRATNEKESYQECQNF
TTHERM_00657240_Tt  342 VLREVLHN----KIPCSLEDLKIYHKMAKESSVSIFKKKAV--GEVSEEYLNEVIKKIKQKFNIVRQENERESARNCQSF
AT5G46070_At        368 AFDQSKAP----EEG----ALREEHEEAVRKALAIFNSNAV--GNG--SARKKFEDLLHKDLKKKFEDYKKNAFMEADLR
GBP1_Hs             338 QMGQKVQL----PTET-LQELLDLHRDSEREAIEVFIRSSF--KDVDHLFQKELAAQLEKKRDDFCKQNQEASSDRCSAL
GBP3_Hs             336 QMGQKVQL----PAET-LQELLDLHRVSEREATEVYMKNSF--KDVDHLFQKKLAAQLDKKRDDFCKQNQEASSDRCSAL
GBP2_Mm             336 QMNQKIQM----PTET-LQELLDLHRPIESEAIEVFLKNSF--KDVDQKFQTELGNLLVAKRDAFIKKNMDVSSARCSDL
GBP5_Hs             336 QMGQKVQL----PMET-LQELLDLHRTSEREAIEVFMKNSF--KDVDQSFQKELETLLDAKQNDICKRNLEASSDYCSAL
GBP4_Hs             352 QMAQQLRL----PTDT-LQELLDVHAACEREAIAVFMEHSF--KDENHEFQKKLVDTIEKKKGDFVLQNEEASAKYCQAE
LOC570622_Dr        338 EMLSMAQF----PLD--PEELSSIHTLAEKKAIEEFINMSF--SDNDQIYQQELMGKIHNEYQDLCQQNHQASLMQCQEV
DDB0235201_Dd       366 KMSSITNEDSFSKPPLCLKEFEKLESQSNNYATNLYKQLLFGLEEAYKPGQQKLERLLSDNHQYFLKENYFKIQNYCQRI

PTMB.422            402 LVQQFAPIDQKLKMSEYKTYSEFEKDLKLFNIFLDEHGPKVVKRKQLFVEFIQKMTTEGATFFVQNIQQDFEGYKKFSEE
GSPATG00012059001   412 IQQEFQPIDRKLKLGEYRSFGEYEKDIKMFYNFFIENGPRVGTRNQIILEFLQRALIEGSNLFIRQYSQETEMMKNVAQE
GSPATG00013523001   412 IQQEFQPIDRKLKMGEYKSFGEYEKDIKMFYNFFIENGPRVGTRNQIILEFLQRALIEGSNLFIRQYAQETDMMKNVAFE
TTHERM_01211770_Tt  439 INREFATIDRKLKSNQYKRFSDFESDIKLFYSFLVNHGPITVNRNVIFLEFLQRSINEAANIFVKNLEQEVDIQRSCNSE
TTHERM_00657240_Tt  416 IQQEFQSIERKLKMNEYKNYPEYEQDLKLFYNYFMENGPKTVNRQQFVSDFIHRATSDGAQYFIKQFSHEQDIQKTLSQE
AT5G46070_At        436 CTSTIQRMEKQLRAACHASNANMDNVVKVLEARLAEYEASCHGP------------------------------------
GBP1_Hs             411 LQVIFSPLEEEVKAGIYSKPGGYRLFVQKLQDLKKKYYEEPRKG------------------------------------
GBP3_Hs             409 LQVIFSPLEEEVKAGIYSKPGGYCLFIQKLQDLEKKYYEEPRKG------------------------------------
GBP2_Mm             409 LEDIFGPLEEEVKLGTFSKPGGYYLFLQMRQELEKKYNQAPGKG------------------------------------
GBP5_Hs             409 LKDIFGPLEEAVKQGIYSKPGGHNLFIQKTEELKAKYYREPRKG------------------------------------
GBP4_Hs             425 LKRLSEHLTESILRGIFSVPGGHNLYLEEKKQVEWDYKLVPRKG------------------------------------
LOC570622_Dr        410 LEDVFDPLEMNISNGSYLRPGGYREYRALLAQLTSSYRARTKSQ------------------------------------
DDB0235201_Dd       446 YNEMIESFENQLESIKIPISSKHFNHQTSILKNEIINQYKEITK------------------------------------

PTMB.422            482 QINKLQNGLNTIKLDSQKQVEQQNQQIKQLNQQNTEITQKYKTQCDELEKLKQKNTQSTQQMKTQFENDFNKLQTSFEEL
GSPATG00012059001   492 TQKKLELELKEARQDGLKDKNNLLMKLAQIESEKTEIELREQVARDNLEELKIQKDQVERDLKLESENEKNELMRQIQEL
GSPATG00013523001   492 TQKKLEQELKEARQDSLKDKNNLLMKLAQVESEKSDLELREQVARDNLEELKIQKEQVERDLKLESENEKIELTRQIQEL
TTHERM_01211770_Tt  519 VQRRNETELVELRKKTDNERVLLERKVEQLLKEKDSIATDKVKNERAVKEIEEEKQKMEKEYKELLNSERVELNDQINNL
TTHERM_00657240_Tt  496 VQKKLEFELQELKQRMHSEINNYEIKKSQLEFEKTEIEMREQVLRENLAIVKSEKEKLEYEFKEQATSEKQECVKVINDQ
AT5G46070_At        480 ----------------------GKWQKLSVFLQQSLEGPIYDLTKRLIDSIAIEKNSLAMKFRS-VEDAMKHLKQQLDDS
GBP1_Hs             455 ----------------------IQAEEILQTYLKSKESMTDAILQTDQTLTEKEKEIEVERVKAESAQASAKMLQEMQRK
GBP3_Hs             453 ----------------------IQAEEILQTYLKSKESVTDAILQTDQILTEKEKEIEVECVKAESAQASAKMVEEMQIK
GBP2_Mm             453 ----------------------LQAEAMLKNYFDSKADVVETLLQTDQSLTEAAKEVEEERTKAEAAEAANRELEKKQKE
GBP5_Hs             453 ----------------------IQAEEVLQKYLKSKESVSHAILQTDQALTETEKKKKEAQVKAEAEKAEAQRLAAIQRQ
GBP4_Hs             469 ----------------------VKANEVLQNFLQSQVVVEESILQSDKALTAGEKAIAAERAMKEAAEKEQELLREKQKE
LOC570622_Dr        454 ----------------------IMSEEVLCMFLKGKEDVGNMILQADESLSAAEQEKEVQMLKNEILQQRQRGLEEQNRL
DDB0235201_Dd       490 -----------------------QYNLNRNQNQNNYFEIFLNSLEKDIERLINGKQLQNRNEIEILLTKSIDMISERYKS

PTMB.422            562 KQKCQTAEEALKSAEREQFILQSESKKKEALLEQKISHLQKLNDEFNKKEKQYDSSLMTSKSEFSVKLRELTASYEQTLS
GSPATG00012059001   572 KGQVLKAEEFNKDLERTTLFGNSEFEKERALLEQKITFFEKLVNEMNAKEIDYQNEIKNLRKEYSLQSKDQQSKADQSVR
GSPATG00013523001   572 KGQVLKAEEFNKDMERNTLFGNSEFEKERALLEQKITFFEKLVNEMNAKEVDYQNEIKNLRKEHSLQSKDQQTKSDQTVR
TTHERM_01211770_Tt  599 RNKSKITEETLKKVERELLLKESSFQKEKALLEQRAQHFEKLCEEYLAKDKATENMMQSTHSGFNTQMKEQQTKYERIIK
TTHERM_00657240_Tt  576 KQKLIMQEEMLSELERNKILSQSEFEKEKALLTQRIQFYEKQMEELSKKDKDQSNDIKSLKKDYNSQIKDISNKYESLNK
AT5G46070_At        537 ERYKLEYQKRYDESNNDKKKLEDIYRERITKLQGENSSLNERCSTLVKTVESKKEEIKEWIRNYDQIVLKQKAVQEQLSS
GBP1_Hs             513 NEQMMEQKERSYQEHLKQLTEKMENDRVQLLKEQERTLALKLQEQEQLLKEGFQKESRIMKNEIQDLQTKMRRRKACTIS
GBP3_Hs             511 YQQMMEEKEKSYQEHVKQLTEKMERERAQLLEEQEKTLTSKLQEQARVLKERCQGESTQLQNEIQKLQKTLKKKTKRYMS
GBP2_Mm             511 FELMMQQKEKSYQEHVKKLTEKMKDEQKQLLAEQENIIAAKLREQEKFLKEGFENESKKLIREIDTLKQNKSSGK-CTIL
GBP5_Hs             511 NEQMMQERERLHQEQVR----QMEIAKQNWLAEQQKMQEQQMQEQAAQLSTTFQAQNRSLLSELQHAQRTVNNDDPCVLL
GBP4_Hs             527 QQQMMEAQERSFQEYMAQMEKKLEEERENLLREHERLLKHKLKVQEEMLKEEFQKKSEQLNKEINQLKEKIESTKNEQLR
LOC570622_Dr        512 QEQAFKDMERTHQEHVNQIVHQMEREQERIRRDNERVLEAKLKEREALLKQGFEEEASRMQKEINSLKTDMSKQEESKPS
DDB0235201_Dd       547 YMNQLDLPITLEQLKNEDKKHYNQLLNSKFKQITEVAHEESYYQPFKSMAEKQLHQLFQYFQEINEQKILEKSKFESDQS
```

```
PTMB.422              642 EVRAQNLQLNEKLQETSDQLMQLEIKYENYYQVQTHKEESQSNAIKKLMQSQQTLQEQINIEITKNKQLQDDLLSTQMQV
GSPATG00012059001     652 QLQQKLNDMQEKLNEMENELVEKETNFENDFKKFEHKERSLTKQNLELNEQIQALTRELR-EYRRNEEQVQLMLKNDANN
GSPATG00013523001     652 QLQQKLSDLQEKLNEMENELIEKESNFENDFKKFEHKERSLTKQNLEQNEQIQALTREIR-EYKRNEEQMQQMLKSDANN
TTHERM_01211770_Tt    679 ELENKLTNLKESNFELNEKISNQSQNFSQEVETLQTRERGLLRQIEELQTEISSTKKQLS-EASTNGLKNLGDLKEQFNN
TTHERM_00657240_Tt    656 SLTQKINELQERIYELENELTNRQQKFDSEKRKWEVNENNLSRLLEESQLQIQKLQNEIR-DLRNYEEQHAQFLQNDMSK
AT5G46070_At          617 EMEVLRTRSTTSEARVAAAREQAKSAAEETKEWKRKYDYAVGEARSALQKAASVQE-----RSGKETQLREDALREEFSI
GBP1_Hs               593 --------------------------------------------------------------------------------
GBP3_Hs               591 HK-LKI--------------------------------------------------------------------------
GBP2_Mm               590 --------------------------------------------------------------------------------
GBP5_Hs               587 --------------------------------------------------------------------------------
GBP4_Hs               607 LL-KILDMASNIMIVTLPGAS-KLLGVGTKYLGSRI--------------------------------------------
LOC570622_Dr          592 TMRKVLDGIGTAATLFLPGFIPKAAGIGLSFLSRLF--------------------------------------------
DDB0235201_Dd         627 ILDFKKSVSQIILPIEEESLVKSMNEFKIRSINSYKN-------------------------------------------

PTMB.422              722 ESNEVGEQYKQQVKQLEIKLQEQKQIQ---LKSDAILNQQIQFLQMELEEKKNTINDLKKSHDSAMLALEQSQNTQNQQQ
GSPATG00012059001     731 QVNEMVEKVQNLEDLIKSKDDQLKQARSQSEKDKALMQQKMEFMEVQLDEYKKQIEENKKSHEAIMKAFENSSNESTYKI
GSPATG00013523001     731 QVNELLEKVQGLEDLIKTKDDQLKQAKSHSEKDKALMQQKMEFMEVQLDEYKKQIEENKKSHEAIMKAFENSSNESTYKI
TTHERM_01211770_Tt    758 ETNDLKIQNTQLQEELSKTLDKHKQDVSELRKREALLQQQIEFVRHELDESRSQLVEAKKAHESTLQALEINNMGDE-ES
TTHERM_00657240_Tt    735 NHTELQEKMVYYEELIKEKEDTLINFKNNFQKEKALLQQKIEFLQVQFEESKQQLEENKRAHEAIMKALESSSLDSSGKL
AT5G46070_At          692 TLANKDEEITEKATKLEKAEQSLTVLRSDLKVAESKLES-----------------------------------------
GBP1_Hs                   --------------------------------------------------------------------------------
GBP3_Hs                   --------------------------------------------------------------------------------
GBP2_Mm                   --------------------------------------------------------------------------------
GBP5_Hs                   --------------------------------------------------------------------------------
GBP4_Hs                   --------------------------------------------------------------------------------
LOC570622_Dr              --------------------------------------------------------------------------------
DDB0235201_Dd         664 --------------------------------------------------------------------------------

PTMB.422              799 LNKSIIDLKEQHMIEIRQAEHTSQQLRKQLQQQLDQLQGELQEAEMRGDVFLQEKNKLQEELSESYQVQDELKQKIQQQL
GSPATG00012059001     811 DAAKMNDLREQHKRELKSVENEYENTKKRLQQQIDQLNERNNELELKVKFETGDLIKELENLKEQLATTEEQRNKLLEQN
GSPATG00013523001     811 DAAKMNDLREQHKRDLKNIENEYESVKKRLQQQVDQLNERNSELELKVKFETGDLINEIENLKEQLQTSEEQKNKLLEQN
TTHERM_01211770_Tt    837 GQKQILELKDTHMKQIRAMEAEALSQKKTLQSQIDALNKELSDLELKYKVSIHDYETEVKTLSDQLTESQAIRTQLSTKL
TTHERM_00657240_Tt    815 DNRQLQELKENHTREIKALENDFEKAKKRYIQQIDQLNEKNSELEMRNKFEEADFQKEIEQLKEQVQELKNVNQRLEEKN
AT5G46070_At          731 -----------------------------------------------FEVELASLRLTLSEMTDKLDSANKKALAYEKEA
GBP1_Hs                   --------------------------------------------------------------------------------
GBP3_Hs                   --------------------------------------------------------------------------------
GBP2_Mm                   --------------------------------------------------------------------------------
GBP5_Hs                   --------------------------------------------------------------------------------
GBP4_Hs                   --------------------------------------------------------------------------------
LOC570622_Dr              --------------------------------------------------------------------------------
DDB0235201_Dd         664 -------------------------------------------------------------RLEPFRNSNSYSTQLNKLE

PTMB.422              879 KEMESNKHTQFREKELRMNQRIKQLEEELSQCKQQLQNTGNLDKNSIEQQVNELRNYYEMEKDVLERRIHEERQKADQKY
GSPATG00012059001     891 KMLDGQKLQILKEQELRYQKKIKQLEQAIDEADAKIAKEVNLAQAKAEESLTQLKNFYEIERERLERRICEEKEKSDKKF
GSPATG00013523001     891 KTLDGQKLSILKEQELRYQKKIKQLEQAMDEADTKTAREVNQAQAKAEESLTQLKNFYEIERERLERRICEEKEKSDKKF
TTHERM_01211770_Tt    917 SESEEMLKLSSKDMEKKFIQKIRELEDTNEAIKDRSQSEIRENQAKSEESLQQLKNMFEIERETLEKRLIEEKEKREKSY
TTHERM_00657240_Tt    895 SVLEDEKQKSFKEVEERYINRIRLLEDEIEEIKQNAIKDIRESQAKSEESLAQLKNIYEIERETLERRLAEDKEKWDKRY
AT5G46070_At          764 NKLEQEKIRMEQKYRSEFQR-FDEVKERCKAAEIEAKRATELADKARTDAVTSQKEKSESQRLAMER-------------
GBP1_Hs                   --------------------------------------------------------------------------------
GBP3_Hs                   --------------------------------------------------------------------------------
GBP2_Mm                   --------------------------------------------------------------------------------
GBP5_Hs                   --------------------------------------------------------------------------------
GBP4_Hs                   --------------------------------------------------------------------------------
LOC570622_Dr              --------------------------------------------------------------------------------
DDB0235201_Dd         683 SSLEKQIELKLNENVTQMKWIVMESLDTVRVMMISRTPIFWFQHSYEN--------------------------------

PTMB.422              959 QILFEEQEQKMRDEQQQYEEEIETLKDELRDLEINLTTQQQQYDNEIELKNKQLNSMENILNETKEQLVQLQNTFQTQVE
GSPATG00012059001     971 QAAQEEFYHKLRETEQNYEEEIENLKDDLRDQVQQYTNTIQQYDHEIALKQQTIEIFEKHIKETKEQLISLQNNNNATLE
GSPATG00013523001     971 QAAQEEFYHKLRETEQNYEEEIETLKDDLRDQVQQYTNTIQQYDHEIALKQQTIEIFEKHIKEIKEQLISLQNNNNTTLE
TTHERM_01211770_Tt    997 NQVVEEYEQKLREEQAALEEEIENLRDDIKDYEIQVINMQGQYEHELDLKQKTIENLEKMLQETKAQLLSSQASYQSQLE
TTHERM_00657240_Tt    975 QLSCDEYQEQIREAQHTYEEEIENLKDDLREQEAQYHNVVQQYEHELALKQQTIETLEKYTKETKESLLNLQSNNNTTLE
AT5G46070_At          830 -LAQIERAERQVENLERQKTDLEDELDRLRVSEMEAVSKVTILEARVEEREKEIGSLIKETNAQRAHNVKSLEKLLDEER
GBP1_Hs                   --------------------------------------------------------------------------------
GBP3_Hs                   --------------------------------------------------------------------------------
GBP2_Mm                   --------------------------------------------------------------------------------
GBP5_Hs                   --------------------------------------------------------------------------------
GBP4_Hs                   --------------------------------------------------------------------------------
LOC570622_Dr              --------------------------------------------------------------------------------
DDB0235201_Dd         731 ---------------EAYREAEKRIIDKIKSKPLRDKVIKQFINNDLKQQLNSLFKLSTILLTLGIILIFSILISFSTIK
```

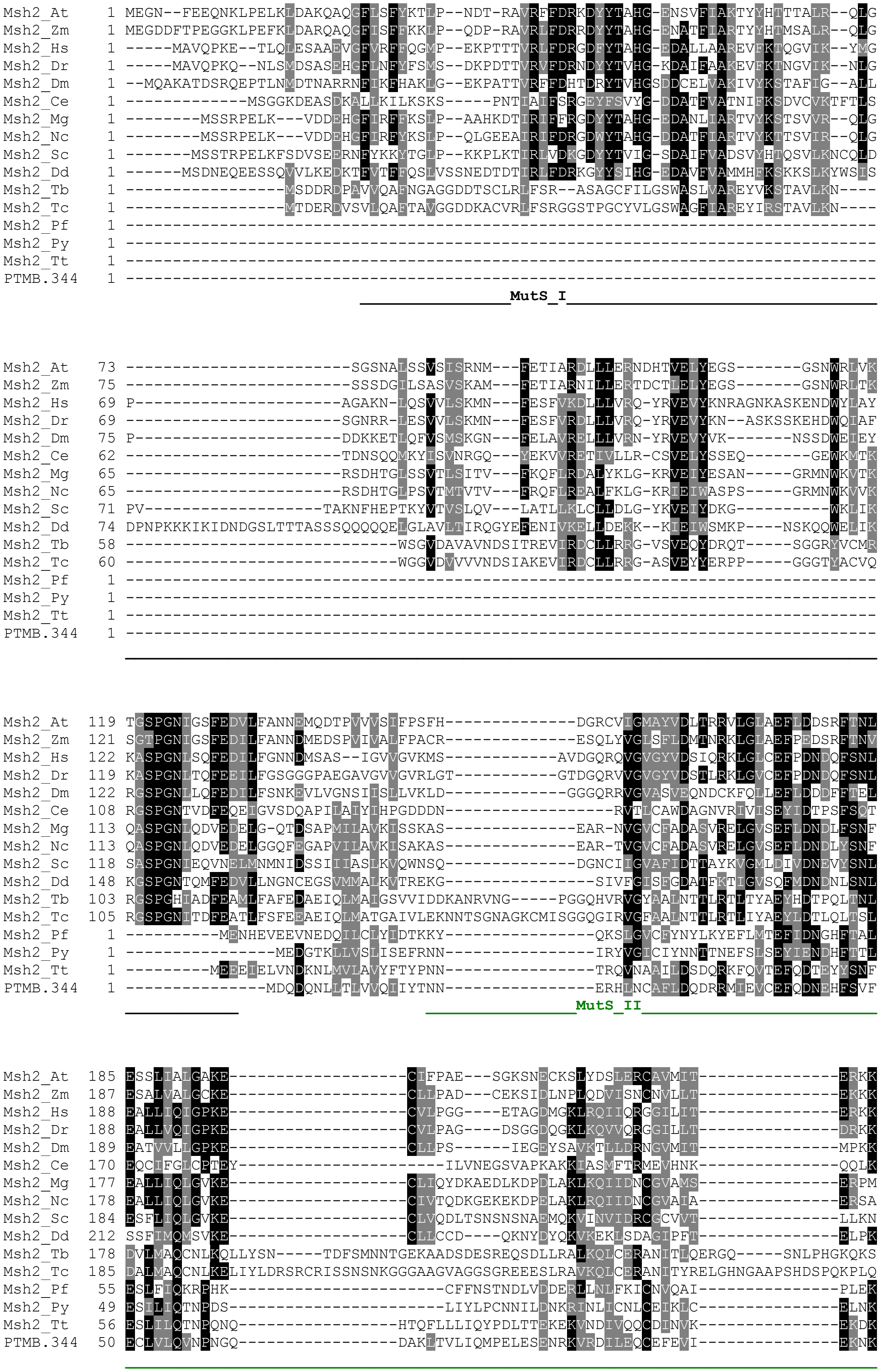
Nowak et al. Figure S5

```
PTMB.422               1039 QRINNLNVTIQSLESQLTNQQQNNQQLVKENQLMTQKLENLDVKLQQKLSEFKQLKEDQEKEKTQLQESLQDLRRKYTAT
GSPATG00012059001      1051 QQMNSFTTERKSLIQKIDVLSSQLNNLQKEHMALQQKKDLLENEKSRKEQQFEQSRKEWQEEKKEQLERLEETKLRLQKM
GSPATG00013523001      1051 QQMNSFTTERKSLISKNDVLSSQLNNLQKEHMALQQKKDLLENEKTRKEQQFEQSRKEWQEEKRELIDRLEETKQRLQKM
TTHERM_01211770_Tt     1077 QQMASYNADRKELQQKYETVQKNLQLSDNKILTLTQKLESSEASYENKKKELAQLKQDNIIERQALEEKLEETRKKYQEL
TTHERM_00657240_Tt     1055 QYINNFNAERKTFIQKVENLTQELSKKEKENIALQQKKENLEANLKKRESQLTQAKVEMQQERTENSNYIDDLKQKLQQA
AT5G46070_At            909 KAHIAANRRAEALSLELQAAQAHVDNLQQELAQARLKETALDNKIRAASSSHGKRSRFEDVVDMDIGEGSDRILR-----
GBP1_Hs                     --------------------------------------------------------------------------------
GBP3_Hs                     --------------------------------------------------------------------------------
GBP2_Mm                     --------------------------------------------------------------------------------
GBP5_Hs                     --------------------------------------------------------------------------------
GBP4_Hs                     --------------------------------------------------------------------------------
LOC570622_Dr                --------------------------------------------------------------------------------
DDB0235201_Dd           796 K-------------------------------------------------------------------------------

PTMB.422               1119 CDEYLEKKINYEKAIALSAQQNEFFAKKVEELERQLESCNLKYEERIKIQKQEWTQELSDRLSKLNEEKQQIESKSTQLK
GSPATG00012059001      1131 NDEFLEKKIEYGRETALIQQQNEFLQKKIEDLQRQLDTQQSRFDEKLKQQKNEYLVELEQKLERAQEEKTAIETKYEKNK
GSPATG00013523001      1131 NDEFLEKKIEYGRETALTQQQNEFLQKKIEDLQKQIDTQQSRFDEKIKQQKNEYSVELEQKLERAQEEKTAIETKYEKIK
TTHERM_01211770_Tt     1157 SDEHLQTKINSEKSIALSSQQNEFYQKKVQELQVALESSTQRFEERLKTSKAEWAQETLEKIQNLQDEKVNLESKINALK
TTHERM_00657240_Tt     1135 SDEQIQQKIEFSREIALSNQRNEFLSKKNEDLARQNETLIQHYEEKLKILKQEMNQELNEKTDKLQQQRDQMEEKYEKVK
AT5G46070_At            984 -------------------------TNKRARSTRGDDHGPTDEGDEDFQSHQDNGEEEEEDYRKLTVQNLKHELTKYDCG
GBP1_Hs                     --------------------------------------------------------------------------------
GBP3_Hs                     --------------------------------------------------------------------------------
GBP2_Mm                     --------------------------------------------------------------------------------
GBP5_Hs                     --------------------------------------------------------------------------------
GBP4_Hs                     --------------------------------------------------------------------------------
LOC570622_Dr                --------------------------------------------------------------------------------
DDB0235201_Dd               --------------------------------------------------------------------------------

PTMB.422               1199 KQLREKETQFLKAQQDLEKETALSTEKIVYLEQKLREHEQQTSSENSNAAMQLKQLREQFSLLKSKSSSDIEQLKSQLTN
GSPATG00012059001      1211 KQLKEAEYQYNKQASTLEREKAILTEKLGHLEARKNELESKLKDETATITQQQAQLREQIAQEKKQLQQELEKYKQLNLQ
GSPATG00013523001      1211 KQLKENEYQYNKQASTLEREKAILTEKLGQLDSRKNELESKIKDESASVAQYQTALREQIAAEKKSLQQELEKYKQFNLQ
TTHERM_01211770_Tt     1237 KSTKEFETQYQKRVTDLEKEKAIYVEKLQYTESKLKETEDKYMIETEQLKDQLSKAKEFTSTEKKMLIQEIEKLKSQNVQ
TTHERM_00657240_Tt     1215 KTLKEIEATYNKQLSQIEKEKAIYQEKLANIESKKSELDMKYQAENQNYVMQILQLKEQYSLEKKTLLNEADKYKQFNLQ
At_NP_199419          1040 HLLLNRGHQNKKEILALYEAHVLPKKALAREEERKKQREVTSS------------------------------------
GBP1_Hs                     --------------------------------------------------------------------------------
GBP3_Hs                     --------------------------------------------------------------------------------
GBP2_Mm                     --------------------------------------------------------------------------------
GBP5_Hs                     --------------------------------------------------------------------------------
GBP4_Hs                     --------------------------------------------------------------------------------
LOC570622_Dr                --------------------------------------------------------------------------------
DDB0235201_Dd               --------------------------------------------------------------------------------

PTMB.422               1279 LEFEKQELQANYEKDKILWQGKTQFLESQRESLKQELADAMRKFETTIQTLQRQRSLERNDHNQDITEMLNQIERKYQDQ
GSPATG00012059001      1291 LEQDKSEIYTSYERDKALWEGKFQFLEQQKEQAKQDLVDALKKFEMTLMHLQRARSNEKDEQENNLNELLLSVERKYQAQ
GSPATG00013523001      1291 LEQEKSEIHTSYERDKALWEGKFQFLEQQKEQAKQDLMDALKKFEMTLMHLQRARSNEKDEQENNLNELLLSVERKYQSQ
TTHERM_01211770_Tt     1317 LDNEKHELNTSYERDKVLWEGKFSFLQQQKEQAKQDLIEAQKKFEQTLTHLQKQRQNEKNEREQSVSDMLMQIERKYQTQ
TTHERM_00657240_Tt     1295 LEQDKNELITNYERDKALWEGKFAFLEQQKEQAKQDLADALKKFEMTLMHLQKARSNEKDEHENNLNELLISMEKKYQNQ
AT5G46070_At                --------------------------------------------------------------------------------
GBP1_Hs                     --------------------------------------------------------------------------------
GBP3_Hs                     --------------------------------------------------------------------------------
GBP2_Mm                     --------------------------------------------------------------------------------
GBP5_Hs                     --------------------------------------------------------------------------------
GBP4_Hs                     --------------------------------------------------------------------------------
LOC570622_Dr                --------------------------------------------------------------------------------
DDB0235201_Dd               --------------------------------------------------------------------------------

PTMB.422               1359 VKDIQQQHQKKCDDYQERIERLEKELKQSQSKELLDQQSKIG--QQFERKTAELLENEKRLLSTIEELKQERDQRILEHH
GSPATG00012059001      1371 IEEANQTHQRIVQDYEDKIRRLQKEVKTHKDKILIDQHGKIGNQLLSEKKFAEMLDNEKRLQQEIENIKQDRDQKILDYQ
GSPATG00013523001      1371 IEEANQTHQRIVQDYEDKIRRLQKEVKTHRDKILIDQHGKIGNQLLSEKKFAEMLDNEKRLQQEIENIKQDRDYKIFEYQ
TTHERM_01211770_Tt     1397 ISELQETSNTKLSAANQKITKLEAELKNSLDQTMMESRQKNSNAIFTEKRLAELLESEKKLSSELDIIRKEREAEKTEAQ
TTHERM_00657240_Tt     1375 INELNENNQRSVQDYEDKIRRLNKEIKQLKENMLVEKHGKLGNQMLNEKKLADFIENEKKLLLEIEQLKNERDQKIIDYQ
AT5G46070_At                --------------------------------------------------------------------------------
GBP1_Hs                     --------------------------------------------------------------------------------
GBP3_Hs                     --------------------------------------------------------------------------------
GBP2_Mm                     --------------------------------------------------------------------------------
GBP5_Hs                     --------------------------------------------------------------------------------
GBP4_Hs                     --------------------------------------------------------------------------------
LOC570622_Dr                --------------------------------------------------------------------------------
DDB0235201_Dd               --------------------------------------------------------------------------------
```

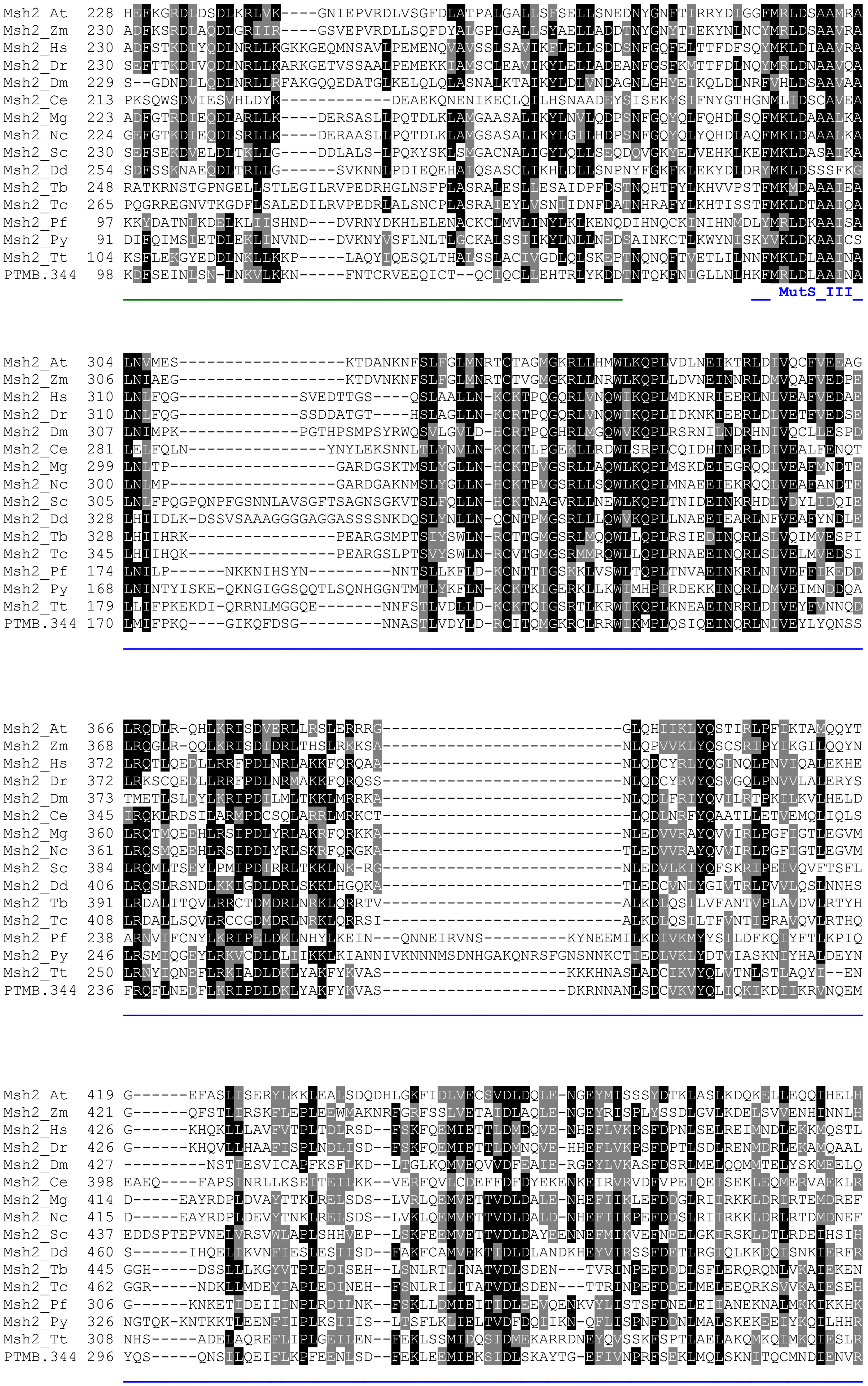
```
PTMB.422           1437 EEVEQEKEYWRNKINELEERQRESDKKQSQLIFYHEKERAKWSQEKDYIMQQKMELQDQLSRLEKKKELLLKENEKMKNS
GSPATG00012059001  1451 RMLEQERESLKAKITELETKYKEVESKRSTLIFEFEKERAKWNLDRDHLNNLKNELTDQLDRLRKKEELLLRDNEKLKNE
GSPATG00013523001  1451 KMLEQERENLKAKIVELETKYKEVENKRSTLIFEFEKERAKWNLDRDHLNNIKTELSDQLEKLSDKKEQLLRENEKLKNE
TTHERM_01211770_Tt 1477 RKLDQERETLKTKMLQIEEKFREAEKKRNTLIFEHEKERAKWNLEKDHLLNQKNELQENLNKIEKKKEILLRENERLKND
TTHERM_00657240_Tt 1455 RQLDTDKEVLKQKIAEIEQKYKEAEQKRGTLIFEHEKERAKWNLDRDHLLVQKNELQDHIDKLEKKKELLLRENEKLRNE
AT5G46070_At            --------------------------------------------------------------------------------
GBP1_Hs                 --------------------------------------------------------------------------------
GBP3_Hs                 --------------------------------------------------------------------------------
GBP2_Mm                 --------------------------------------------------------------------------------
GBP5_Hs                 --------------------------------------------------------------------------------
GBP4_Hs                 --------------------------------------------------------------------------------
LOC570622_Dr            --------------------------------------------------------------------------------
DDB0235201_Dd           --------------------------------------------------------------------------------

PTMB.422           1517 SKSLRKYN----PNQTLNNSYLN------------KQASDKKIP-----SQEVVNMSYDVTKSLEDKDPSQKENLSTSTT
GSPATG00012059001  1531 QRATRRSV--AAHNMTSNNIMTG-----------NKYRNPINNVSTIGLQKLSPTHSNNTSTSSANISVLKKTNLADITN
GSPATG00013523001  1531 QRATRRSV--AAHNMTSNNIKTG-----------NTYRNPINNISTIGLQKLSPTHSNNTSTSSANISVLKKNNLADITN
TTHERM_01211770_Tt 1557 SRQSRRQF-ERLSDQNSNIFLTQ-----------NRQMMNANQPSFLNMSRNMGSERKSPMRSARSRERSSLNDYSYANN
TTHERM_00657240_Tt 1535 SRVTRRSINLVGTGIQSNTALSSSKSGIKRSGTTNQQQSPVSNTSILS-KKNSMIQENNQFMSNSNLNQQK--NLNDITN
AT5G46070_At            --------------------------------------------------------------------------------
GBP1_Hs                 --------------------------------------------------------------------------------
GBP3_Hs                 --------------------------------------------------------------------------------
GBP2_Mm                 --------------------------------------------------------------------------------
GBP5_Hs                 --------------------------------------------------------------------------------
GBP4_Hs                 --------------------------------------------------------------------------------
LOC570622_Dr            --------------------------------------------------------------------------------
DDB0235201_Dd           --------------------------------------------------------------------------------

PTMB.422           1576 FGSFKKYYQMMNSKQVS----------------------PSKTKEISDI-------------------------------
GSPATG00012059001  1598 FEKVVPQTQQSFQNNKY-----------------YFYGNQQGQNDDSMIGQSDFQK------------------------
GSPATG00013523001  1598 YEKTAPLTQQSFQNSKF-----------------YYYGNQQGQNDDSMIGQCEFQK------------------------
TTHERM_01211770_Tt 1625 SVDCGEDNGSSDNNTNN----------------ILENLNLNSSRENQPIQNTPPANGFKSFMQVLQEKSGLNNENKPPVS
TTHERM_00657240_Tt 1612 FAGIMPYDPSRQMNQKLTSLNSQEDEAQQNQFLQYQYLMQDQQKAQQLISSNNMQYNNPNMSSPQAPMFNPSNYGQQNIY
AT5G46070_At            --------------------------------------------------------------------------------
GBP1_Hs                 --------------------------------------------------------------------------------
GBP3_Hs                 --------------------------------------------------------------------------------
GBP2_Mm                 --------------------------------------------------------------------------------
GBP5_Hs                 --------------------------------------------------------------------------------
GBP4_Hs                 --------------------------------------------------------------------------------
LOC570622_Dr            --------------------------------------------------------------------------------
DDB0235201_Dd           --------------------------------------------------------------------------------

PTMB.422                -------------------
GSPATG00012059001       -------------------
GSPATG00013523001       -------------------
TTHERM_01211770_Tt 1689 PSKKENVLI----------
TTHERM_00657240_Tt 1692 PQNSQPFQIEKFKYFNSNV
AT5G46070_At            -------------------
GBP1_Hs                 -------------------
GBP3_Hs                 -------------------
GBP2_Mm                 -------------------
GBP5_Hs                 -------------------
GBP4_Hs                 -------------------
LOC570622_Dr            -------------------
DDB0235201_Dd           -------------------
```



```
PTMB.186_Pt   1 MPH--------------------YPEEIEYSDKYLDDYYEYRHVILPKHIFKKITKG---------KLLNEMEWRALGVQ
CKS101b_Pt    1 MPH--------------------YPEEIEYSDKYFDDYYEYRHVILPKHIFKKITKG---------KLLNEMEWRALGVQ
CKS102a_Pt    1 MPH--------------------YPEEIEYSDKYFDDYYEYRHVILPKHIFKKITKG---------KLLNEMEWRAIGVQ
CKS102b_Pt    1 MPH--------------------YPEEIEYSDKYFDDYYEYRHVILPKHIFKKITKG---------KLLNEMEWRAIGVQ
CKS103a_Pt    1 MPH--------------------YPEDIEYSDKYQDDYYEYRHVILPKHIFKKITKG---------KLLSEMEWRNLGVQ
CKS103b_Pt    1 MPH--------------------YPEDIEYSDKYQDDYYEYRHVILPKHIFKKITKG---------KLLSEMEWRNLGVQ
CKS103c_Pt    1 MPH--------------------YPEDIEYSDKYQDEFYEYRHVILPKHVFKKLTKG---------KLLNEMEWRGLGVQ
Cks-1_Ce      1 MTT--------------------GNNDFYYSNKYEDDEFEYRHVHVTKDVSKLIPKN---------RLMSETEWRSLGIQ
Cks-2_Ce      1 M------------------------EDIYYSPRYEDDEYEYRHVILPQAVSRKVPKG---------RLLSEAEWRRAGVQ
CKS1_At       1 M------------------------GQIQYSEKYFDDTFEYRHVVLPPEVAKLLPKN---------RLLSENEWRAIGVQ
CKS2_At       1 M------------------------GQIQYSDKYFDDTFEYRHVVLPPEVAKLLPKN---------RILSESEWRAIGVQ
CKS1B_Hs      1 MSH----------------------KQIYYSDKYDDEEFEYRHVMLPKDIAKLVPKT---------HLMSESEWRNLGVQ
CKS2_Hs       1 MAH----------------------KQIYYSDKYFDEHYEYRHVMLPRELSKQVPKT---------HLMSEEEWRRLGVQ
Cks30A_Dm     1 MS-----------------------KDIYYSDKYYDEQFEYRHVVLPKELVKMVPKT---------HLMTEAEWRSIGVQ
Cks85A_Dm     1 MPA----------------------DQIQYSEKYFDDNFEYRHVILPPDLAKHVPKA---------HLMTETEWRNLGVQ
Cks1_Sc       1 MYHHYHAFQGRKLTDQERARVLEFQDSIHYSPRYSDDNYEYRHVMLPKAMLKVIPSDYFNSEVGTLRILTEDEWRGLGIT
suc1_Sp       1 MSKSGV---PRLLTASERERLEPFIDQIHYSPRYADDEYEYRHVMLPKAMLKAIPTDYFNPETGTLRILQEEEWRGLGIT
                                   ___α0___     _____β1___  ____β2__  __α1_                 __α2__

PTMB.186_Pt  52 QSRGWSHYETHRPEPHILLFRRPKNTDPNTGLPPP---------------------------------GFSAPY
CKS101b_Pt   52 QSRGWSHYETHRPEPHILLFRRAKNTDPNTGLPPP---------------------------------GFSAPY
CKS102a_Pt   52 QSRGWAHYETHRPEPHILLFRRPKNTDPNTGLPPP---------------------------------GFSAPY
CKS102b_Pt   52 QSRGWVHYETHRPEPHILLFRRPKNTDPNTGLPPP---------------------------------GFSAPY
CKS103a_Pt   52 QSRGWVHYECHRPEPHILLFRRPKGTDPNSGLPPQ---------------------------------GFQAPY
CKS103b_Pt   52 QSRGWIHYECHRPEPHILLFRRPKGTDPNSGLAPQ---------------------------------GFSAPY
CKS103c_Pt   52 QSRGWVHYECHRPEPHILLFRRPKGTDPNTGLPPQ---------------------------------GFSAPY
Cks-1_Ce     52 QSPGWMHYMIHGPERHVLLFRRPLAATQKTGGNVR-------------------------------SGNAVGVR
Cks-2_Ce     48 QSLGWEHYMVHNPEKHILLFRRKRHFTKADEERAK---MELLRDMEVFNRAQQQTTATDEKLNAVFYEFVNKAV
CKS1_At      48 QSRGWVHYAVHRPEPHIMLFRRPLNYQQQQENQAQ---------------------------------N-MLVK
CKS2_At      48 QSRGWVHYAIHRPEPHIMLFRRPLNYQQEHQA--------------------------------------QIAK
CKS1B_Hs     50 QSQGWVHYMIHEPEPHILLFRRPLPKKPK--------------------------------------------K
CKS2_Hs      50 QSLGWVHYMIHEPEPHILLFRRPLPKDQQ--------------------------------------------K
Cks30A_Dm    49 QSRGWIHYMIHKPEPHILLFRRPKT------------------------------------------------D
Cks85A_Dm    50 QSPGWVHYMVHAPEPHVILFRRKRIPAEDAPQVAA---------------------------ANAAQAIANLCG
Cks1_Sc      81 QSLGWEHYECHAPEPHILLFKRPLNYEAELRAATAAAQQQQQQQQQQQQQQQQHQTQSISNDMQVP----PQIS
suc1_Sp      78 QSLGWEMYEVHVPEPHILLFKREKDYQMKFS----------------------------------Q----QRGG
                    __β3__      ____β4___
```



```
Msh2_At    1 MEGN--FEEQNKLPELKLDAKQAQGFISFYKTLP--NDT-RAVRFFDRKDYYTAHG-ENSVFIAKTYYHTTTALR--QLG
Msh2_Zm    1 MEGDDFTPEGGKLPEFKLDARQAQGFISFFKKLP--QDP-RAVRLFDRRDYYTAHG-ENATFIARTYYHTMSALR--QLG
Msh2_Hs    1 -----MAVQPKE--TLQLESAAEVGFVRFFQGMP--EKPTTTVRLFDRGDFYTAHG-EDALLAAREVFKTQGVIK--YMG
Msh2_Dr    1 -----MAVQPKQ--NLSMDSASEHGFLNFYFSMS--DKPDTTVRVFDRNDYYTVHG-KDAIFAAKEVFKTNGVIK--NLG
Msh2_Dm    1 --MQAKATDSRQEPTLNMDTNARRNFIKFHAKLG--EKPATTVRFFDHTDRYTVHGSDDCELVAKIVYKSTAFIG--ALL
Msh2_Ce    1 -------------MSGGKDEASDKALLKILKSKS-----PNTIAIFSRGEYFSVYG-DDATFVATNIFKSDVCVKTFTLS
Msh2_Mg    1 --------MSSRPELK---VDDEHGFIRFFKSLP--AAHKDTIRIFFRGDYYTAHG-EDANLIARTVYKSTSVVR--QLG
Msh2_Nc    1 --------MSSRPELK---VDDEHGFIRFYKSLP--QLGEEAIRIFDRGDWYTAHG-DDATFIARTVYKTTSVIR--QLG
Msh2_Sc    1 -------MSSTRPELKFSDVSEERNFYKKYTGLP--KKPLKTIRLVDKGDYYTVIG-SDAIFVADSVYHTQSVLKNCQLD
Msh2_Dd    1 ------MSDNEQEESSQVVLKEDKTFVTFFQSLVSSNEDTDTIRLFDRKGYYSIHG-EDAVFVAMMHFKSKKSLKYWSIS
Msh2_Tb    1 -----------------MSDDRDPAVVQAFNGAGGDDTSCLRLFSR--ASAGCFILGSWASLVAREYVKSTAVLKN----
Msh2_Tc    1 -----------------MTDERDVSVLQAFTAVGGDDKACVRLFSRGGSTPGCYVLGSWAGFIAREYIRSTAVLKN----
Msh2_Pf    1 --------------------------------------------------------------------------------
Msh2_Py    1 --------------------------------------------------------------------------------
Msh2_Tt    1 --------------------------------------------------------------------------------
PTMB.344   1 --------------------------------------------------------------------------------
                                     _______________MutS_I__________________________________

Msh2_At   73 ------------------------SGSNALSSVSISRNM---FETIARDLLLERNDHTVELYEGS-------GSNWRLVK
Msh2_Zm   75 ------------------------SSSDGILSASVSKAM---FETIARNILLERTDCTLELYEGS-------GSNWRLTK
Msh2_Hs   69 P----------------------AGAKN-LQSVVLSKMN---FESFVKDLLLVRQ-YRVEVYKNRAGNKASKENDWYLAY
Msh2_Dr   69 -----------------------SGNRR-LESVVLSKMN---FESFVRDLLLVRQ-YRVEVYKN--ASKSSKEHDWQIAF
Msh2_Dm   75 P----------------------DDKKETLQFVSMSKGN---FELAVRELLLVRN-YRVEVYVK-------NSSDWEIEY
Msh2_Ce   62 -----------------------TDNSQQMKYISVNRGQ---YEKVVRETIVLLR-CSVELYSSEQ-------GEWKMTK
Msh2_Mg   65 -----------------------RSDHTGLSSVTLSITV---FKQFLRDALYKLG-KRVEIYESAN-----GRMNWKVTK
Msh2_Nc   65 -----------------------RSDHTGLPSVTMTVTV---FRQFLREALFKLG-KRIEIWASPS-----GRMNWKVVK
Msh2_Sc   71 PV-------------------TAKNFHEPTKYVTVSLQV---LATLLKLCLLDLG-YKVEIYDKG----------WKLIK
Msh2_Dd   74 DPNPKKKIKIDNDGSLTTTASSSQQQQQELGLAVLTIRQGYEFENIVKELLDEKK--KIEIWSMKP----NSKQQWELIK
Msh2_Tb   58 -----------------------------WSGVDAVAVNDSITREVIRDCLLRRG-VSVEQYDRQT-----SGGRYVCMR
Msh2_Tc   60 -----------------------------WGGVDVVVVNDSIAKEVIRDCLLRRG-ASVEYYERPP-----GGGTYACVQ
Msh2_Pf    1 --------------------------------------------------------------------------------
Msh2_Py    1 --------------------------------------------------------------------------------
Msh2_Tt    1 --------------------------------------------------------------------------------
PTMB.344   1 --------------------------------------------------------------------------------
             ________________________________________________________________________________

Msh2_At  119 TGSPGNIGSFEDVLFANNEMQDTPVVVSIFPSFH--------------DGRCVIGMAYVDLTRRVLGLAEFLDDSRFTNL
Msh2_Zm  121 SGTPGNIGSFEDILFANNDMEDSPVIVALFPACR--------------ESQLYVGLSFLDMTNRKLGLAEFPEDSRFTNV
Msh2_Hs  122 KASPGNLSQFEDILFGNNDMSAS--IGVVGVKMS------------AVDGQRQVGVGYVDSIQRKLGLCEFPDNDQFSNL
Msh2_Dr  119 KASPGNLTQFEEILFGSGGGPAEGAVGVVGVRLGT-----------GTDGQRVVGVGYVDSTLRKLGVCEFPDNDQFSNL
Msh2_Dm  122 RGSPGNLLQFEDILFSNKEVLVGNSIISLLVKLD-------------GGGQRRVGVASVEQNDCKFQLLEFLDDDFFTEL
Msh2_Ce  108 RGSPGNTVDFEQEIGVSDQAPILAIYIHPGDDDN------------------RVTLCAWDAGNVRIVISEYIDTPSFSQT
Msh2_Mg  113 QASPGNLQDVEDELG-QTDSAPMIIAVKISSKAS--------------EAR-NVGVCFADASVRELGVSEFLDNDLFSNF
Msh2_Nc  113 QASPGNLQDVEDELGGQFEGAPVILAVKISAKAS--------------EAR-TVGVCFADASVRELGVSEFLDNDLYSNF
Msh2_Sc  118 SASPGNIEQVNELMNMNIDSSIIIASLKVQWNSQ--------------DGNCIIGVAFIDTTAYKVGMLDIVDNEVYSNL
Msh2_Dd  148 KGSPGNTQMFEDVLLNGNCEGSVMMALKVTREKG----------------SIVFGISFGDATFKTIGVSQFMDNDNLSNL
Msh2_Tb  103 RGSPGHIADFEAMLFAFEDAEIQLMAIGSVVIDDKANRVNG-----PGGQHVRVGYAALNTTLRTLTYAEYHDTPQLTNL
Msh2_Tc  105 RGSPGNITDFEATLFSFEEAEIQLMATGAIVLEKNNTSGNAGKCMISGGQGIRVGFAALNTTLRTLIYAEYLDTLQLTSL
Msh2_Pf    1 ----------MENHEVEEVNEDQILCLYIDTKKY----------------QKSLGVCFYNYLKYEFLMTEFIDNGHFTAL
Msh2_Py    1 ----------------MEDGTKLLVSLISEFRNN----------------IRYVGICIYNNTTNEFSLSEYIENDHFTTL
Msh2_Tt    1 ---------MEEEIELVNDKNLMVLAVYFTYPNN----------------TRQVNAAILDSDQRKFQVTEFQDTEYYSNF
PTMB.344   1 ---------------MDQDQNLLTLVVQIIYTNN----------------ERHLNCAFLDQDRRMIEVCEFQDNEHFSVF
             ____________                    ________________MutS_II_________________________

Msh2_At  185 ESSLIALGAKE-------------------CIFPAE---SGKSNECKSLYDSLERCAVMIT---------------ERKK
Msh2_Zm  187 ESALVALGCKE-------------------CLLPAD---CEKSIDLNPLQDVISNCNVLLT---------------EKKK
Msh2_Hs  188 EALLIQIGPKE-------------------CVLPGG----ETAGDMGKLRQIIQRGGILIT---------------ERKK
Msh2_Dr  188 EALLVQIGPKE-------------------CVLPAG----DSGGDQGKLKQVVQRGGILLT---------------DRKK
Msh2_Dm  189 EATVVLLGPKE-------------------CLLPS------IEGEYSAVKTLLDRNGVMIT---------------MPKK
Msh2_Ce  170 EQCIFGLCPTEY----------------------ILVNEGSVAPKAKKIASMFTRMEVHNK---------------QQLK
Msh2_Mg  177 EALLIQLGVKE-------------------CLIQYDKAEDLKDPDLAKLKQIIDNCGVAMS---------------ERPM
Msh2_Nc  178 EALLIQLGVKE-------------------CIVTQDKGEKEKDPELAKLRQIIDNCGVAIA---------------ERSA
Msh2_Sc  184 ESFLIQLGVKE-------------------CLVQDLTSNSNSNAEMQKVINVIDRCGCVVT---------------LLKN
Msh2_Dd  212 SSFIMQMSVKE-------------------CLLCCD----QKNYDYQKVKEKLSDAGIPFT---------------ELPK
Msh2_Tb  178 DVLMAQCNLKQLLYSN-----TDFSMNNTGEKAADSDESREQSDLLRALKQLCERANITLQERGQ-----SNLPHGKQKS
Msh2_Tc  185 DALMAQCNLKELIYLDRSRCRISSNSNKGGGAAGVAGGSGREEESLRAVKQLCERANITYRELGHNGAAPSHDSPQKPLQ
Msh2_Pf   55 ESLFIQKRPHK-----------------------CFFNSTNDLVDDERLLNLFKICNVQAI---------------PLEK
Msh2_Py   49 ESILIQTNPDS-----------------------LIYLPCNNILDNKRINLICNLCEIKLC---------------ELNK
Msh2_Tt   56 ESLILQTNPQNQ-----------------HTQFLLLIQYPDLTTEKEKVNDIVQQCDINVK---------------EKDK
PTMB.344  50 ECLVLQVNPNGQ-----------------DAKLTVLIQMPELESENRKVRDILEQCEFEVI---------------EKNK
             ________________________________________________________________________________
```



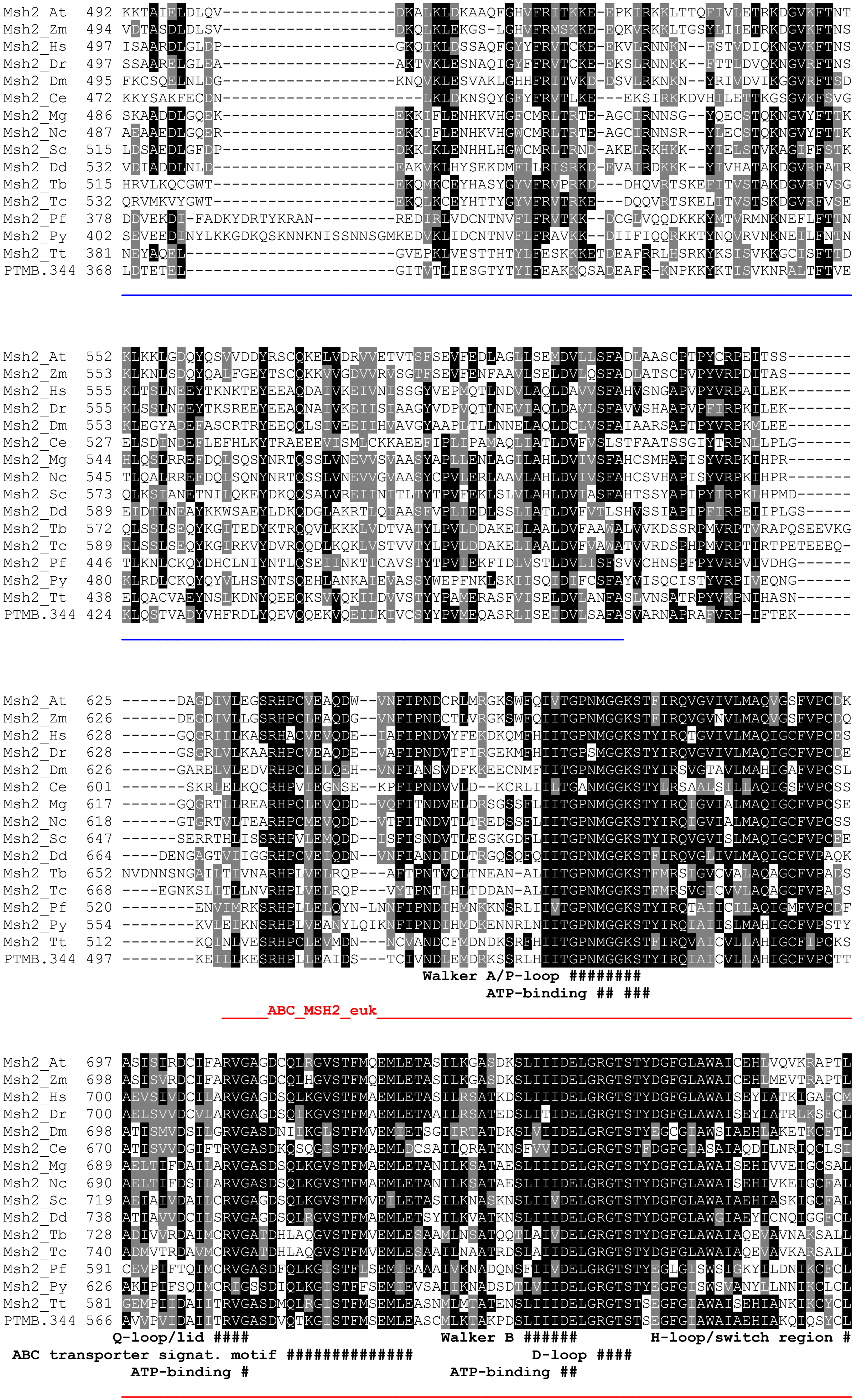
```
Msh2_At   228 HEFKGRDLDSDLKRLVK----GNIEPVRDLVSGFDLATPALGALLSFSELLSNEDNYGNFTIRRYDIGGFMRLDSAAMRA
Msh2_Zm   230 ADFKSRDLAQDLGRIIR----GSVEPVRDLLSQFDYALGPLGALLSYAELLADDTNYGNYTIEKYNLNCYMRLDSAAVRA
Msh2_Hs   230 ADFSTKDIYQDLNRLLKGKKGEQMNSAVLPEMENQVAVSSLSAVIKFLELLSDDSNFGQFELTTFDFSQYMKLDIAAVRA
Msh2_Dr   230 SEFTTKDIVQDLNRLLKARKGETVSSAALPEMEKKIAMSCLEAVIKYLELLADEANFGSFKMTTFDLNQYMRLDNAAVQA
Msh2_Dm   229 S--GDNDLLQDLNRLLRFAKGQQEDATGLKELQLQLASNALKTAIKYLDLVNDAGNLGHYEIKQLDLNRFVHLDSAAVAA
Msh2_Ce   213 PKSQWSDVIESVHLDYK------------DEAEKQNENIKECLQILHSNAADEYSISEKYSIFNYGTHGNMLIDSCAVEA
Msh2_Mg   223 ADFGTRDIEQDLARLLK----DERSASLLPQTDLKLAMGAASALIKYLNVLQDPSNFGQYQLFQHDLSQFMKLDAAALKA
Msh2_Nc   224 GEFGTKDIEQDLSRLLK----DERAASLLPQTDLKLAMGSASALIKYLGILHDPSNFGQYQLYQHDLAQFMKLDAAALKA
Msh2_Sc   230 SEFSEKDVELDLTKLLG----DDLALS-LPQKYSKLSMGACNALIGYLQLLSEQDQVGKYELVEHKLKEFMKLDASAIKA
Msh2_Dd   254 SDFSSKNAEQDLTRLLG------SVKNNLPDIEQEHAIQSASCLIKHLDLLSNPNYFGKFKLEKYDLDRYMKLDSSSFKG
Msh2_Tb   248 RATKRNSTGPNGELLSTLEGILRVPEDRHGLNSFPLASRALESLLESAIDPFDSTNQHTFYLKHVVPSTFMKMDAAAIEA
Msh2_Tc   265 PQGRREGNVTKGDFLSALEDILRVPEDRLALSNCPLASRAIEYLVSNIIDNFDATNHRAFYLKHTISSTFMKLDTAAIQA
Msh2_Pf    97 KKYDATNLKDELKLIISHND---DVRNYDKHLELENACKCLMVLINYLKLKENQDIHNQCKINIHNMDLYMRLDKAAISA
Msh2_Py    91 DIFQIMSIETDLEKLINVND---DVKNYVSFLNLTLGCKALSSIIKYLNLLNEDSAINKCTLKWYNISKYVKLDKAAICS
Msh2_Tt   104 KSFLEKGYEDDLNKLLKKP-----LAQYIQESQLTHALSSLACIVGDLQLSKEPTNQNQFTVETLILNNFMKLDLAAINA
PTMB.344   98 KDFSEINLSN-LNKVLKKN-----FNTCRVEEQICT--QCIQCLLEHTRLYKDDTNTQKFNIGLLNLHKFMRLDLAAINA
```

```
Msh2_At   304 LNVMES------------------KTDANKNFSLFGLMNRTCTAGMGKRLLHMWLKQPLVDLNEIKTRLDIVQCFVEEAG
Msh2_Zm   306 LNIAEG------------------KTDVNKNFSLFGLMNRTCTVGMGKRLLNRWLKQPLLDVNEINNRLDMVQAFVEDPE
Msh2_Hs   310 LNLFQG-------------SVEDTTGS----QSLAALLN-KCKTPQGQRLVNQWIKQPLMDKNRIEERLNLVEAFVEDAE
Msh2_Dr   310 LNLFQG-------------SSDDATGT----HSLAGLLN-KCRTPQGQRLVNQWIKQPLIDKNKIEERLDLVETFVEDSE
Msh2_Dm   307 LNIMPK-------------PGTHPSMPSYRWQSVLGVLD-HCRTPQGHRLMGQWVKQPLRSRNILNDRHNIVQCLLESPD
Msh2_Ce   281 LELFQLN---------------YNYLEKSNNLTLYNVLN-KCKTLPGEKLLRDWLSRPLCQIDHINERLDIVEALFENQT
Msh2_Mg   299 LNLTP------------------GARDGSKTMSLYGLLN-HCKTPVGSRLLAQWLKQPLMSKDEIEGRQQLVEAFMNDTE
Msh2_Nc   300 LNLMP------------------GARDGAKNMSLYGLLN-HCKTPVGSRLLSQWLKQPLMNAEEIEKRQQLVEAFANDTE
Msh2_Sc   305 LNLFPQGPQNPFGSNNLAVSGFTSAGNSGKVTSLFQLLN-HCKTNAGVRLLNEWLKQPLTNIDEINKRHDLVDYLIDQIE
Msh2_Dd   328 LHIIDLK-DSSVSAAAGGGGAGGASSSSNKDQSLYNLLN-QCNTPMGSRLLLQWVKQPLLNAEEIEARLNFVEAFYNDLE
Msh2_Tb   328 LHIIHRK----------------PEARGSMPTSIYSWLN-RCTTGMGSRLMQQWLLQPLRSIEDINQRLSLVQIMVESPI
Msh2_Tc   345 LHIIHQK----------------PEARGSLPTSVYSWLN-RCVTGMGSRMMRQWLLQPLRNAEEINQRLSLVELMVEDSI
Msh2_Pf   174 LNILP------NKKNIHSYN---------NNTSLLKFLD-KCNTTIGSKKLVSWLTQPLTNVAEINKRLNIVEFFIKEDD
Msh2_Py   168 LNINTYISKE-QKNGIGGSQQTLSQNHGGNTMTLYKFLN-KCKTKIGERKLLKWIMHPIRDEKKINQRLDMVEIMNDDQA
Msh2_Tt   179 LLIFPKEKDI-QRRNLMGGQE-------NNFSTLVDLLD-KCKTQIGSRTLKRWIKQPLKNEAEINRRLDIVEYFVNNQD
PTMB.344  170 LMIFPKQ----GIKQFDSG---------NNASTLVDYLD-RCITQMGKRCLRRWIKMPLQSIQEINQRLNIVEYLYQNSS
```

```
Msh2_At   366 LRQDLR-QHLKRISDVERLLRSLERRRG--------------------------GLQHIIKLYQSTIRLPFIKTAMQQYT
Msh2_Zm   368 LRQGLR-QQLKRISDIDRLTHSLRKKSA--------------------------NLQPVVKLYQSCSRIPYIKGILQQYN
Msh2_Hs   372 LRQTLQEDLLRRFPDLNRLAKKFQRQAA--------------------------NLQDCYRLYQGINQLPNVIQALEKHE
Msh2_Dr   372 LRKSCQEDLLRRFPDLNRMAKKFQRQSS--------------------------NLQDCYRVYQSVGQLPNVVLALERYS
Msh2_Dm   373 TMETLSLDYLKRIPDILMLTKKLMRRKA--------------------------NLQDLFRIYQVILRTPKILKVLHELD
Msh2_Ce   345 IRQKLRDSILARMPDCSQLARRLMRKCT---------------------------LQDLNRFYQAATLLETVEMQLIQLS
Msh2_Mg   360 LRQTMQEEHLRSIPDLYRLAKRFQRKKA--------------------------NLEDVVRAYQVVIRLPGFIGTLEGVM
Msh2_Nc   361 LRQSMQEEHLRSIPDLYRLSKRFQRGKA--------------------------TLEDVVRAYQVVIRLPGFIGTLEGVM
Msh2_Sc   384 LRQMLTSEYLPMIPDIRRLTKKLNK-RG--------------------------NLEDVLKIYQFSKRIPEIVQVFTSFL
Msh2_Dd   406 LRQSLRSNDLKKIGDLDRLSKKLHGQKA--------------------------TLEDCVNLYGIVTRLPVVLQSLNNHS
Msh2_Tb   391 LRDALITQVLRRCTDMDRLNRKLQRRTV--------------------------ALKDLQSILVFANTVPLAVDVLRTYH
Msh2_Tc   408 LRDALLSQVLRCCGDMDRLNRKLQRRSI--------------------------ALKDLQSILTFVNTIPRAVQVLRTHQ
Msh2_Pf   238 ARNVIFCNYLKRIPELDKLNHYLKEIN---QNNEIRVNS---------KYNEEMILKDIVKMYYSILDFKQIYFTLKPIQ
Msh2_Py   246 LRSMIQGEYLRKVCDLDLIIKKLKIANNIVKNNNMSDNHGAKQNRSFGNSNNKCTIEDLVKLYDTVIASKNIYHALDEYN
Msh2_Tt   250 LRNYIQNEFLRKIADLDKLYAKFYKVAS--------------------KKKHNASLADCIKVYQLVTNLSTLAQYI--EN
PTMB.344  236 FRQFLNEDFLKRIPDLDKLYAKFYKVAS--------------------DKRNNANLSDCVKVYQLIQKIKDIIKRVNQEM
```

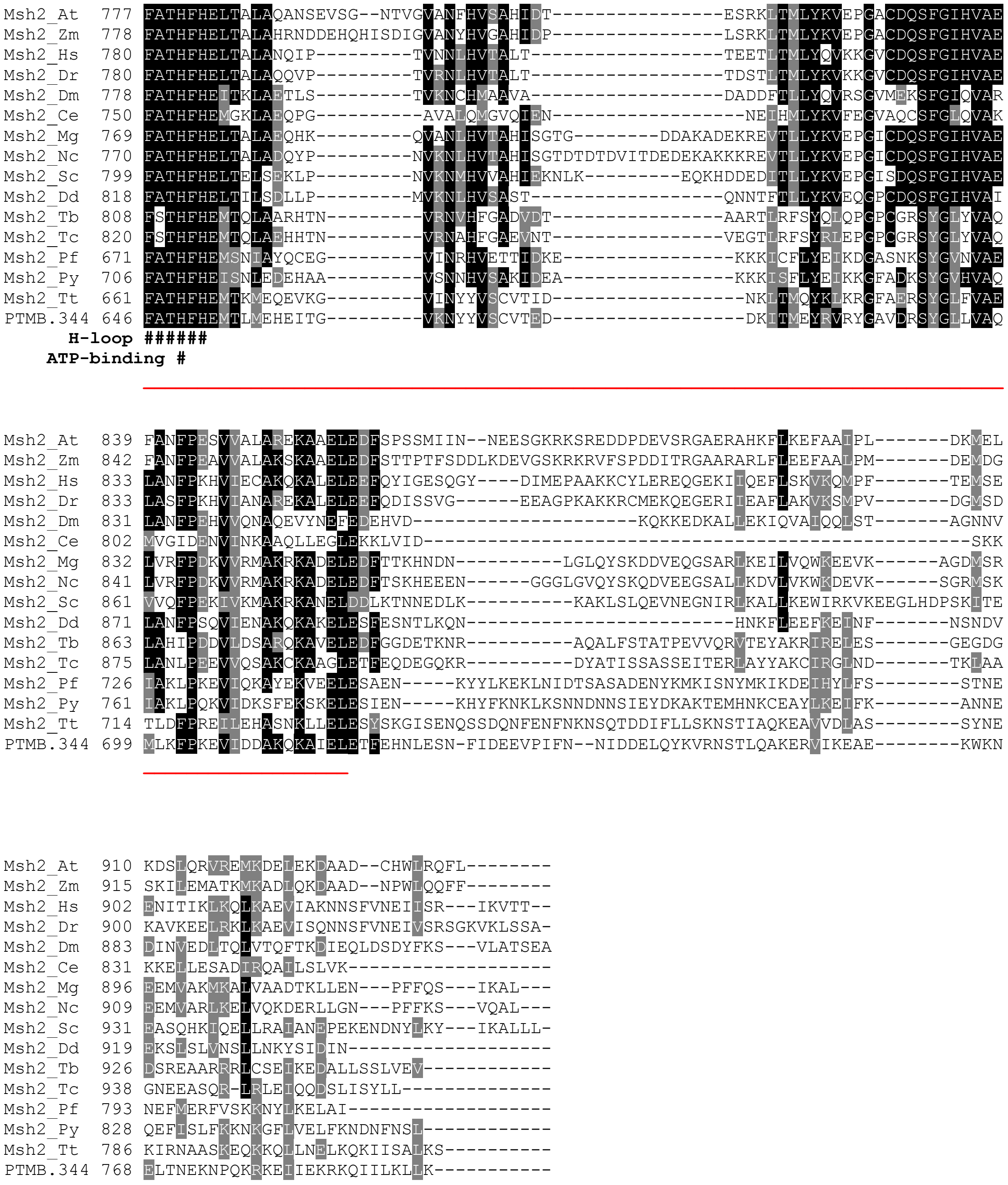
```
Msh2_At   419 G------EFASLISERYLKKLEALSDQDHLGKFIDLVECSVDLDQLE-NGEYMISSSYDTKLASLKDQKELLEQQIHELH
Msh2_Zm   421 G------QFSTLIRSKFLEPLEEWMAKNRFGRFSSLVETAIDLAQLE-NGEYRISPLYSSDLGVLKDELSVVENHINNLH
Msh2_Hs   426 G------KHQKLLLAVFVTPLTDLRSD--FSKFQEMIETTLDMDQVE-NHEFLVKPSFDPNLSELREIMNDLEKKMQSTL
Msh2_Dr   426 G------KHQVLLHAAFISPLNDLISD--FSKFQEMIETTLDMNQVE-HHEFLVKPSFDPTLSDLRENMDRLEKAMQAAL
Msh2_Dm   427 ---------NSTIESVICAPFKSFLKD--LTGLKQMVEQVVDFEAIE-RGEYLVKASFDSRLMELQQMMTELYSKMEELQ
Msh2_Ce   398 EAEQ----FAPSINRLLKSEITEILKK--VERFQVLCDEFFDFDYEKENKEIRVRVDFVPEIQEISEKLEQMERVAEKLR
Msh2_Mg   414 D-----EAYRDPLDVAYTTKLRELSDS--LVRLQEMVETTVDLDALE-NHEFIIKLEFDDGLRIIRKKLDRIRTEMDREF
Msh2_Nc   415 D-----EAYRDPLDEVYTNKLRELSDS--LVKLQEMVETTVDLDALD-NHEFIIKPEFDDSLRIIRKKLDRLRTDMDNEF
Msh2_Sc   437 EDDSPTEPVNELVRSVWLAPLSHHVEP--LSKFEEMVETTVDLDAYEENNEFMIKVEFNEELGKIRSKLDTLRDEIHSIH
Msh2_Dd   460 S------IHQELIKVNFIESLESIISD--FAKFCAMVEKTIDLDLANDKHEYVIRSSFDETLRGIQLKKDQISNKIERFR
Msh2_Tb   445 GGH-----DSSLLLKGYVTPLEDISEH--LSNLRTLINATVDLSDEN---TVRINPEFDDDLSFLERQRQNLVKAIEKEN
Msh2_Tc   462 GGR-----NDKLLMDEYIAPLEDINEH--FSNLRILITATVDLSDEN---TTRINPEFDDELMELEEQRKSVVKAIESEH
Msh2_Pf   306 G------KNKETIDEIIINPLRDILNK--FSKLLDMIEITIDLEEVQENKVYLISTSFDNELEIIANEKNALMKKIKKHK
Msh2_Py   326 NGTQK-KNTKKTLEENFIIPLKSIIIS--LTSFLKLIELTVDFDQIIKN-QFLISPNFDDNLMALSKEKEEIYKQILHHR
Msh2_Tt   308 NHS-----ADELAQREFLIPLGEILEN--FEKLSSMIDQSIDMEKARRDNEYQVSSKFSPTLAELAKQMKQIMKQIESLR
PTMB.344  296 YQS-----QNSILQEIFLKPFEENLSD--FEKLEEMIEKSIDLSKAYTG-EFIVNPRFSEKLMQLSKNITQCMNDIENVR
```

Nowak et al. Figure S7



```
Msh2_At  492 KKTAIELDLQV-------------------DKALKLDKAAQFGHVFRITKKE-EPKIRKKLTTQFIVLETRKDGVKFTNT
Msh2_Zm  494 VDTASDLDLSV-------------------DKQLKLEKGS-LGHVFRMSKKE-EQKVRKKLTGSYLIIETRKDGVKFTNS
Msh2_Hs  497 ISAARDLGLDP-------------------GKQIKLDSSAQFGYYFRVTCKE-EKVLRNNKN--FSTVDIQKNGVKFTNS
Msh2_Dr  497 SSAARELGLEA-------------------AKTVKLESNAQIGYFFRVTCKE-EKSLRNNKK--FTTLDVQKNGVRFTNS
Msh2_Dm  495 FKCSQELNLDG-------------------KNQVKLESVAKLGHHFRITVKD-DSVLRKNKN--YRIVDVIKGGVRFTSD
Msh2_Ce  472 KKYSAKFECDN----------------------LKLDKNSQYGFYFRVTLKE---EKSIRKKDVHILETTKGSGVKFSVG
Msh2_Mg  486 SKAADDLGQEK-------------------EKKIFLENHKVHGFCMRLTRTE-AGCIRNNSG--YQECSTQKNGVYFTTK
Msh2_Nc  487 AEAAEDLGQER-------------------EKKIFLENHKVHGWCMRLTRTE-AGCIRNNSR--YLECSTQKNGVYFTTK
Msh2_Sc  515 LDSAEDLGFDP-------------------DKKLKLENHHLHGWCMRLTRND-AKELRKHKK--YIELSTVKAGIFFSTK
Msh2_Dd  532 VDIADDLNLD--------------------EAKVKLHYSEKDMFLLRISRKD-EVAIRDKKK--YIVHATAKDGVRFATR
Msh2_Tb  515 HRVLKQCGWT--------------------EKQMKCEYHASYGYVFRVPRKD---DHQVRTSKEFITVSTAKDGVRFVSG
Msh2_Tc  532 QRVMKVYGWT--------------------EKQLKCEYHTTYGYVFRVTRKE---DQQVRTSKELITVSTSKDGVRFVSE
Msh2_Pf  378 DDVEKDI-FADKYDRTYKRAN---------REDIRLVDCNTNVFLFRVTKK--DCGLVQQDKKKYMTVRMNKNEFLFTTN
Msh2_Py  402 SEVEEDINYLKKGDKQSKNNKNISSNNSGMKEDVKLIDCNTNVFLFRAVKK--DIIFIQQRKKTYNQVRVNKNEILFNTN
Msh2_Tt  381 NEYAQEL-----------------------GVEPKLVESTTHTYLFESKKKETDEAFRRLHSRKYKSISVKKGCISFTTD
PTMB.344 368 LDTETEL-----------------------GITVTLIESGTYTYIFEAKKQSADEAFR-KNPKKYKTISVKNRALTFTVE


Msh2_At  552 KLKKLGDQYQSVVDDYRSCQKELVDRVVETVTSFSEVFEDLAGLLSEMDVLLSFADLAASCPTPYCRPEITSS-------
Msh2_Zm  553 KLKNLSDQYQALFGEYTSCQKKVVGDVVRVSGTFSEVFENFAAVLSELDVLQSFADLATSCPVPYVRPDITAS-------
Msh2_Hs  555 KLTSLNEEYTKNKTEYEEAQDAIVKEIVNISSGYVEPMQTLNDVLAQLDAVVSFAHVSNGAPVPYVRPAILEK-------
Msh2_Dr  555 KLSSLNEEYTKSREEYEEAQNAIVKEIISIAAGYVDPVQTLNEVIAQLDAVLSFAVVSHAAPVPFIRPKILEK-------
Msh2_Dm  553 KLEGYADEFASCRTRYEEQQLSIVEEIIHVAVGYAAPLTLLNNELAQLDCLVSFAIAARSAPTPYVRPKMLEE-------
Msh2_Ce  527 ELSDINDEFLEFHLKYTRAEEEVISMLCKKAEEFIPLIPAMAQLIATLDVFVSLSTFAATSSGIYTRPNLLPLG------
Msh2_Mg  544 HLQSLRREFDQLSQSYNRTQSSLVNEVVSVAASYAPLLENLAGILAHLDVIVSFAHCSMHAPISYVRPKIHPR-------
Msh2_Nc  545 TLQALRREFDQLSQNYNRTQSSLVNEVVGVAASYCPVLERLAAVLAHLDVIVSFAHCSVHAPISYVRPKIHPR-------
Msh2_Sc  573 QLKSIANETNILQKEYDKQQSALVREIINITLTYTPVFEKLSLVLAHLDVIASFAHTSSYAPIPYIRPKLHPMD------
Msh2_Dd  589 EIDTLNEAYKKWSAEYLDKQDGLAKRTLQIAASFVPLIEDLSSLIATLDVFVTLSHVSSIAPIPFIRPEIIPLGS-----
Msh2_Tb  572 QLSSLSEQYKGITEDYKTRQQVLKKKLVDTVATYLPVLDDAKELLAALDVFAAWALVVKDSSRPMVRPTVRAPQSEEVKG
Msh2_Tc  589 RLSSLSEQYKGIRKVYDVRQQDLKQKLVSTVVTYLPVLDDAKELIAALDVFVAWATVVRDSPHPMVRPTIRTPETEEEQ-
Msh2_Pf  446 TLKNLCKQYDHCLNIYNTLQSEIINKTICAVSTYTPVIEKFIDLVSTLDVLISFSVVCHNSPFPYVRPVIVDHG------
Msh2_Py  480 KLRDLCKQYQYVLHSYNTSQEHLANKAIEVASSYWEPFNKLSKIISQIDIFCSFAYVISQCISTYVRPIVEQNG------
Msh2_Tt  438 ELQACVAEYNSLKDNYQEEQKSVVQKILDVVSTYYPAMERASFVISELDVLANFASLVNSATRPYVKPNIHASN------
PTMB.344 424 KLQSTVADYVHFRDLYQEVQQEKVQEILKIVCSYYPVMEQASRLISEIDVLSAFASVARNAPRAFVRP-IFTEK------


Msh2_At  625 ------DAGDIVLEGSRHPCVEAQDW--VNFIPNDCRLMRGKSWFQIVTGPNMGGKSTFIRQVGVIVLMAQVGSFVPCDK
Msh2_Zm  626 ------DEGDIVLLGSRHPCLEAQDG--VNFIPNDCTLVRGKSWFQIITGPNMGGKSTFIRQVGVNVLMAQVGSFVPCDQ
Msh2_Hs  628 ------GQGRIILKASRHACVEVQDE--IAFIPNDVYFEKDKQMFHIITGPNMGGKSTYIRQTGVIVLMAQIGCFVPCES
Msh2_Dr  628 ------GSGRLVLKAARHPCVEAQDE--VAFIPNDVTFIRGEKMFHIITGPSMGGKSTYIRQVGVIVLMAQIGCFVPCDE
Msh2_Dm  626 ------GARELVLEDVRHPCLELQEH--VNFIANSVDFKKEECNMFIITGPNMGGKSTYIRSVGTAVLMAHIGAFVPCSL
Msh2_Ce  601 -------SKRLELKQCRHPVIEGNSE--KPFIPNDVVLD--KCRLIILTGANMGGKSTYLRSAALSILLAQIGSFVPCSS
Msh2_Mg  617 ------GQGRTLLREARHPCLEVQDD--VQFITNDVELDRSGSSFLIITGPNMGGKSTYIRQIGVIALMAQIGCFVPCSE
Msh2_Nc  618 ------GTGRTVLTEARHPCMEVQDD--VTFITNDVTLTREDSSFLIITGPNMGGKSTYIRQIGVIALMAQIGCFVPCSS
Msh2_Sc  647 ------SERRTHLISSRHPVLEMQDD--ISFISNDVTLESGKGDFLIITGPNMGGKSTYIRQVGVISLMAQIGCFVPCEE
Msh2_Dd  664 ----DENGAGTVIIGGRHPCVEIQDN--VNFIANDIDLTRGQSQFQIITGPNMGGKSTFIRQVGLIVLMAQIGCFVPAQK
Msh2_Tb  652 NVDNNSNGAILTIVNARHPLVELRQP---AFTPNTVQLTNEAN-ALIITGPNMGGKSTFMRSIGVCVALAQAGCFVPADS
Msh2_Tc  668 ----EGNKSLITLLNVRHPLVELRQP---VYTPNTLHLTDDAN-ALIITGPNMGGKSTFMRSVGICVVLAQAGCFVPADS
Msh2_Pf  520 --------ENVIMRKSRHPLLELQYN-LNNFIPNDIHMNKKNSRLIIVTGPNMGGKSTYIRQTAIICILAQIGMFVPCDF
Msh2_Py  554 --------KVLEIKNSRHPLVEANYLQIKNFIPNDIHMDKENNRLNIITGPNMGGKSTYIRQIAIISLMAHIGCFVPSTY
Msh2_Tt  512 --------KQINLVESRHPCLEVMDN---NCVANDCFMDNDKSRFHIITGPNMGGKSTFIRQVAICVLLAHIGCFIPCKS
PTMB.344 497 --------KEILLKESRHPLLEAIDS---TCIVNDLEMDRKSSRLHIITGPNMGGKSTYIRQIAICVLLAHIGCFVPCTT
                                              Walker A/P-loop ########
                                                     ATP-binding ## ###
                             ABC_MSH2_euk


Msh2_At  697 ASISIRDCIFARVGAGDCQLRGVSTFMQEMLETASILKGASDKSLIIIDELGRGTSTYDGFGLAWAICEHLVQVKRAPTL
Msh2_Zm  698 ASISVRDCIFARVGAGDCQLHGVSTFMQEMLETASILKGASDKSLIIIDELGRGTSTYDGFGLAWAICEHLMEVTRAPTL
Msh2_Hs  700 AEVSIVDCILARVGAGDSQLKGVSTFMAEMLETASILRSATKDSLIIIDELGRGTSTYDGFGLAWAISEYIATKIGAFCM
Msh2_Dr  700 AELSVVDCVLARVGAGDSQIKGVSTFMAEMLETAAILRSATEDSLITIDELGRGTSTYDGFGLAWAISEYIATRLKSFCL
Msh2_Dm  698 ATISMVDSILGRVGASDNIIKGLSTFMVEMIETSGIIRTATDKSLVIIDELGRGTSTYEGCGIAWSIAEHLAKETKCFTL
Msh2_Ce  670 ATISVVDGIFTRVGASDKQSQGISTFMAEMLDCSAILQRATKNSFVVIDELGRGTSTFDGFGIASAIAQDILNRIQCLSI
Msh2_Mg  689 AELTIFDAILARVGASDSQLKGVSTFMAEMLETANILKSATAESLIIIDELGRGTSTYDGFGLAWAISEHIVVEIGCSAL
Msh2_Nc  690 AELTIFDSILARVGASDSQLKGVSTFMAEMLETANILKSATAESLIIIDELGRGTSTYDGFGLAWAISEHIVKEIGCFAL
Msh2_Sc  719 AEIAIVDAILCRVGAGDSQLKGVSTFMVEILETASILKNASKNSLIIVDELGRGTSTYDGFGLAWAIAEHIASKIGCFAL
Msh2_Dd  738 ATIAVVDCILSRVGAGDSQLRGVSTFMAEMLETSYILKVATKNSLIIIDELGRGTSTYDGFGLAWGIAEYICNQIGGFCL
Msh2_Tb  728 ADIVVRDAIMCRVGATDHLAQGVSTFMVEMLESAAMLNSATQQTLAIVDELGRGTSTYDGFGLAWAIAQEVAVNAKSALL
Msh2_Tc  740 ADMVTRDAVMCRVGATDHLAQGVSTFMVEMLESAAILNAATRDSLAIIDELGRGTSTYDGFGLAWAIAQEVAVKARSALL
Msh2_Pf  591 CEVPIFTQIMCRVGASDFQLKGISTFLSEMIEAAAIVKNADQNSFIIVDELGRGTSTYEGLGISWSIGKYILDNIKCFCL
Msh2_Py  626 AKIPIFSQIMCRIGSSDIQLKGISTFFSEMIEVSAIIKNADSDTLVIIDELGRGTSTYEGFGISWSVANYLLNNIKCLCL
Msh2_Tt  581 GEMPIIDAIITRVGASDMQLRGISTFMSEMLEASNMLMTATENSLIIIDELGRGTSTSEGFGIAWAISEHIANKIKCYCL
PTMB.344 566 AVVPVIDAIITRVGASDVQTKGISTFMSEMLEASCMLKTAKPDSLIIIDELGRGTSTSEGFGIAWAIAEHIAKQIQSYCL
            Q-loop/lid ####                     Walker B ######        H-loop/switch region #
 ABC transporter signat. motif ##############             D-loop ####
              ATP-binding #                      ATP-binding ##
```

Nowak et al. Figure S7



```
Msh2_At   777 FATHFHELTALAQANSEVSG--NTVGVANFHVSAHIDT----------------ESRKLTMLYKVEPGACDQSFGIHVAE
Msh2_Zm   778 FATHFHELTALAHRNDDEHQHISDIGVANYHVGAHIDP----------------LSRKLTMLYKVEPGACDQSFGIHVAE
Msh2_Hs   780 FATHFHELTALANQIP---------TVNNLHVTALT------------------TEETLTMLYQVKKGVCDQSFGIHVAE
Msh2_Dr   780 FATHFHELTALAQQVP---------TVRNLHVTALT------------------TDSTLTMLYKVKKGVCDQSFGIHVAE
Msh2_Dm   778 FATHFHEITKLAETLS---------TVKNCHMAAVA------------------DADDFTLLYQVRSGVMEKSFGIQVAR
Msh2_Ce   750 FATHFHEMGKLAEQPG----------AVALQMGVQIEN-------------------NEIHMLYKVFEGVAQCSFGLQVAK
Msh2_Mg   769 FATHFHELTALAEQHK---------QVANLHVTAHISGTG--------DDAKADEKREVTLLYKVEPGICDQSFGIHVAE
Msh2_Nc   770 FATHFHELTALADQYP---------NVKNLHVTAHISGTDTDTDVITDEDEKAKKKREVTLLYKVEPGICDQSFGIHVAE
Msh2_Sc   799 FATHFHELTELSEKLP---------NVKNMHVVAHIEKNLK---------EQKHDDEDITLLYKVEPGISDQSFGIHVAE
Msh2_Dd   818 FATHFHELTILSDLLP---------MVKNLHVSAST------------------QNNTFTLLYKVEQGPCDQSFGIHVAI
Msh2_Tb   808 FSTHFHEMTQLAARHTN--------VRNVHFGADVDT-----------------AARTLRFSYQLQPGPCGRSYGLYVAQ
Msh2_Tc   820 FSTHFHEMTQLAEHHTN--------VRNAHFGAEVNT-----------------VEGTLRFSYRLEPGPCGRSYGLYVAQ
Msh2_Pf   671 FATHFHEMSNIAYQCEG--------VINRHVETTIDKE-----------------KKKICFLYEIKDGASNKSYGVNVAE
Msh2_Py   706 FATHFHEISNLEDEHAA--------VSNNHVSAKIDEA-----------------KKKISFLYEIKKGFADKSYGVHVAQ
Msh2_Tt   661 FATHFHEMTKMEQEVKG--------VINYYVSCVTID-------------------NKLTMQYKLKRGFAERSYGLFVAE
PTMB.344  646 FATHFHEMTLMEHEITG--------VKNYYVSCVTED-------------------DKITMEYRVRYGAVDRSYGLLVAQ
       H-loop ######
    ATP-binding #
```

```
Msh2_At   839 FANFPESVVALAREKAAELEDFSPSSMIIN--NEESGKRKSREDDPDEVSRGAERAHKFLKEFAAIPL-------DKMEL
Msh2_Zm   842 FANFPEAVVALAKSKAAELEDFSTTPTFSDDLKDEVGSKRKRVFSPDDITRGAARARLFLEEFAALPM-------DEMDG
Msh2_Hs   833 LANFPKHVIECAKQKALELEEFQYIGESQGY----DIMEPAAKKCYLEREQGEKIIQEFLSKVKQMPF-------TEMSE
Msh2_Dr   833 LASFPKHVIANAREKALELEEFQDISSVG------EEAGPKAKKRCMEKQEGERIIEAFLAKVKSMPV-------DGMSD
Msh2_Dm   831 LANFPEHVVQNAQEVYNEFEDEHVD---------------------KQKKEDKALLEKIQVAIQQLST-------AGNNV
Msh2_Ce   802 MVGIDENVINKAAQLLEGLEKKLVID---------------------------------------------------SKK
Msh2_Mg   832 LVRFPDKVVRMAKRKADELEDFTTKHNDN----------LGLQYSKDDVEQGSARLKEILVQWKEEVK------AGDMSR
Msh2_Nc   841 LVRFPDKVVRMAKRKADELEDFTSKHEEEN------GGGLGVQYSKQDVEEGSALLKDVLVKWKDEVK------SGRMSK
Msh2_Sc   861 VVQFPEKIVKMAKRKANELDDLKTNNEDLK----------KAKLSLQEVNEGNIRLKALLKEWIRKVKEEGLHDPSKITE
Msh2_Dd   871 LANFPSQVIENAKQKAKELESFESNTLKQN-------------------------HNKFLEEFKEINF-------NSNDV
Msh2_Tb   863 LAHIPDDVLDSARQKAVELEDFGGDETKNR----------AQALFSTATPEVVQRVTEYAKRIRELES-------GEGDG
Msh2_Tc   875 LANLPEEVVQSAKCKAAGLETFEQDEGQKR----------DYATISSASSEITERLAYYAKCIRGLND-------TKLAA
Msh2_Pf   726 IAKLPKEVIQKAYEKVEELESAEN-----KYYLKEKLNIDTSASADENYKMKISNYMKIKDEIHYLFS--------STNE
Msh2_Py   761 IAKLPQKVIDKSFEKSKELESIEN-----KHYFKNKLKSNNDNNSIEYDKAKTEMHNKCEAYLKEIFK--------ANNE
Msh2_Tt   714 TLDFPREILEHASNKLLELESYSKGISENQSSDQNFENFNKNSQTDDIFLLSKNSTIAQKEAVVDLAS--------SYNE
PTMB.344  699 MLKFPKEVIDDAKQKAIELETFEHNLESN-FIDEEVPIFN--NIDDELQYKVRNSTLQAKERVIKEAE--------KWKN
```

```
Msh2_At   910 KDSLQRVREMKDELEKDAAD--CHWLRQFL--------
Msh2_Zm   915 SKILEMATKMKADLQKDAAD--NPWLQQFF--------
Msh2_Hs   902 ENITIKLKQLKAEVIAKNNSFVNEIISR---IKVTT--
Msh2_Dr   900 KAVKEELRKLKAEVISQNNSFVNEIVSRSGKVKLSSA-
Msh2_Dm   883 DINVEDLTQLVTQFTKDIEQLDSDYFKS---VLATSEA
Msh2_Ce   831 KKELLESADIRQAILSLVK-------------------
Msh2_Mg   896 EEMVAKMKALVAADTKLLEN---PFFQS---IKAL---
Msh2_Nc   909 EEMVARLKELVQKDERLLGN---PFFKS---VQAL---
Msh2_Sc   931 EASQHKIQELLRAIANEPEKENDNYLKY---IKALLL-
Msh2_Dd   919 EKSLSLVNSLLNKYSIDIN-------------------
Msh2_Tb   926 DSREAARRRLCSEIKEDALLSSLVEV------------
Msh2_Tc   938 GNEEASQR-LRLEIQQDSLISYLL--------------
Msh2_Pf   793 NEFMERFVSKKNYLKELAI-------------------
Msh2_Py   828 QEFISLFKKNKGFLVELFKNDNFNSL------------
Msh2_Tt   786 KIRNAASKEQKKQLLNELKQKIISALKS----------
PTMB.344  768 ELTNEKNPQKRKEIIEKRKQIILKLLK-----------
```

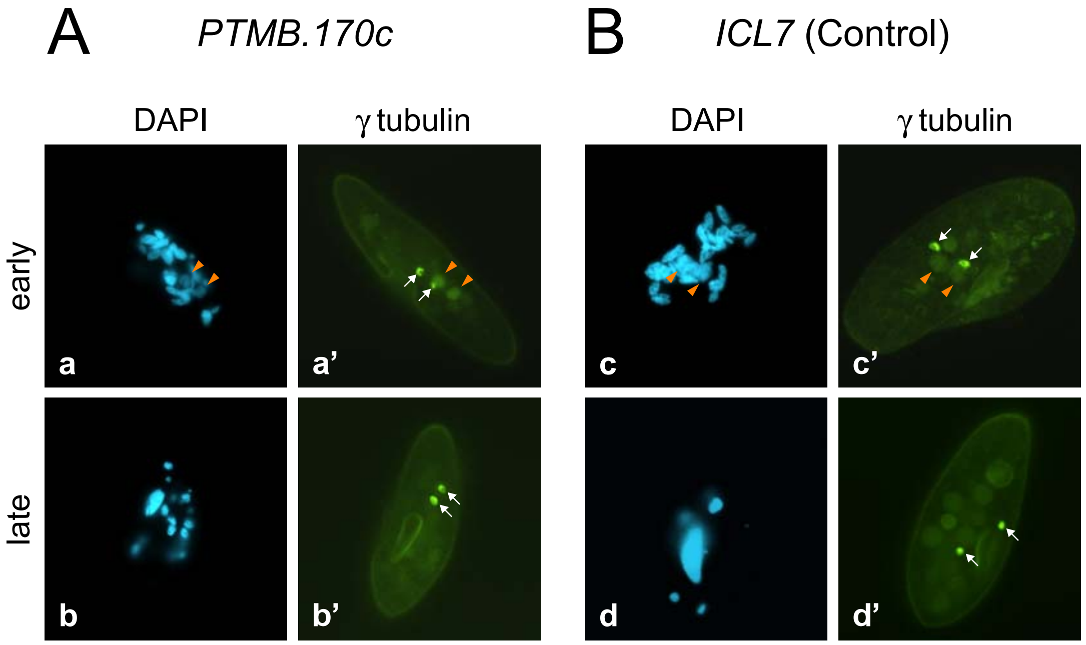
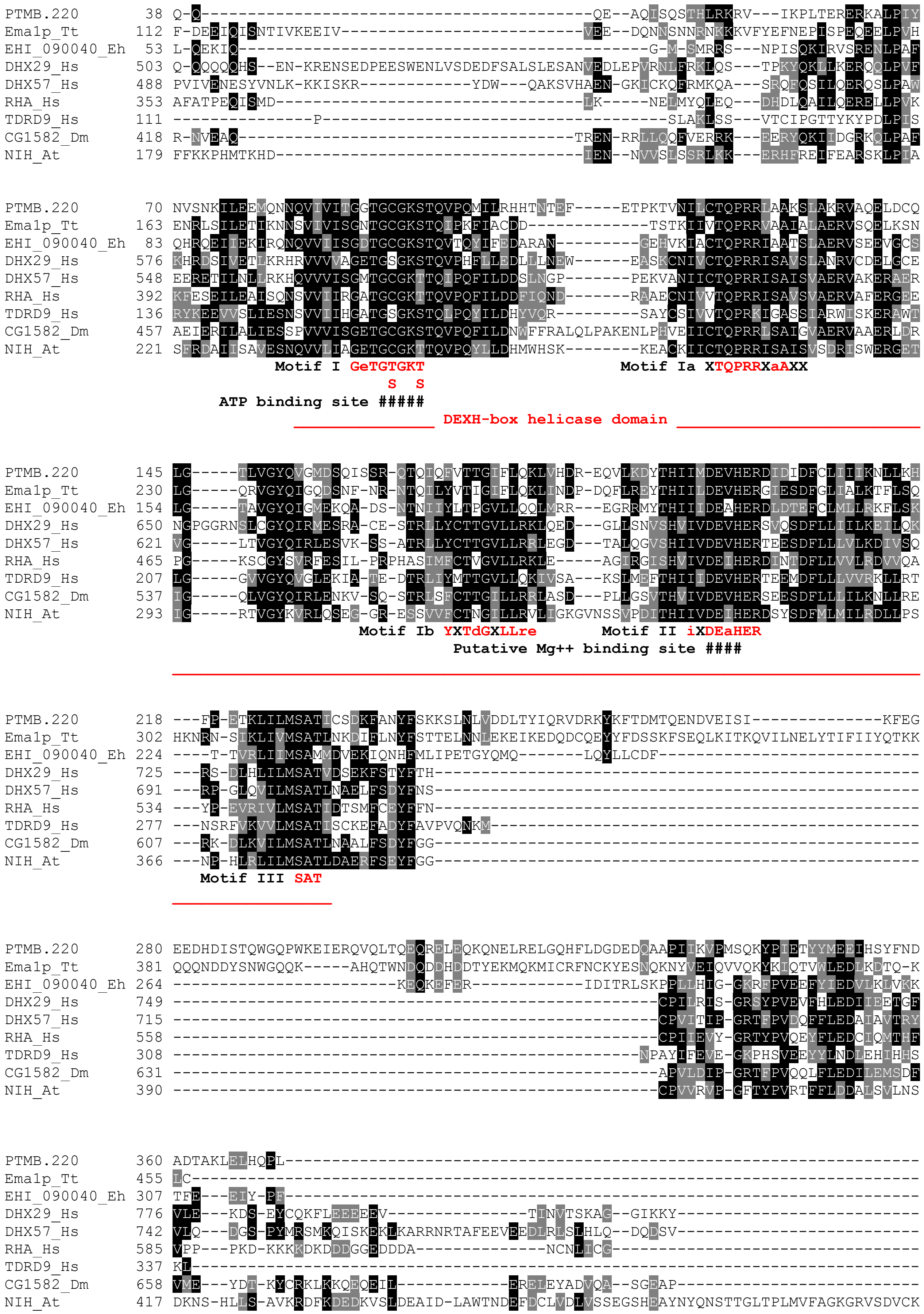
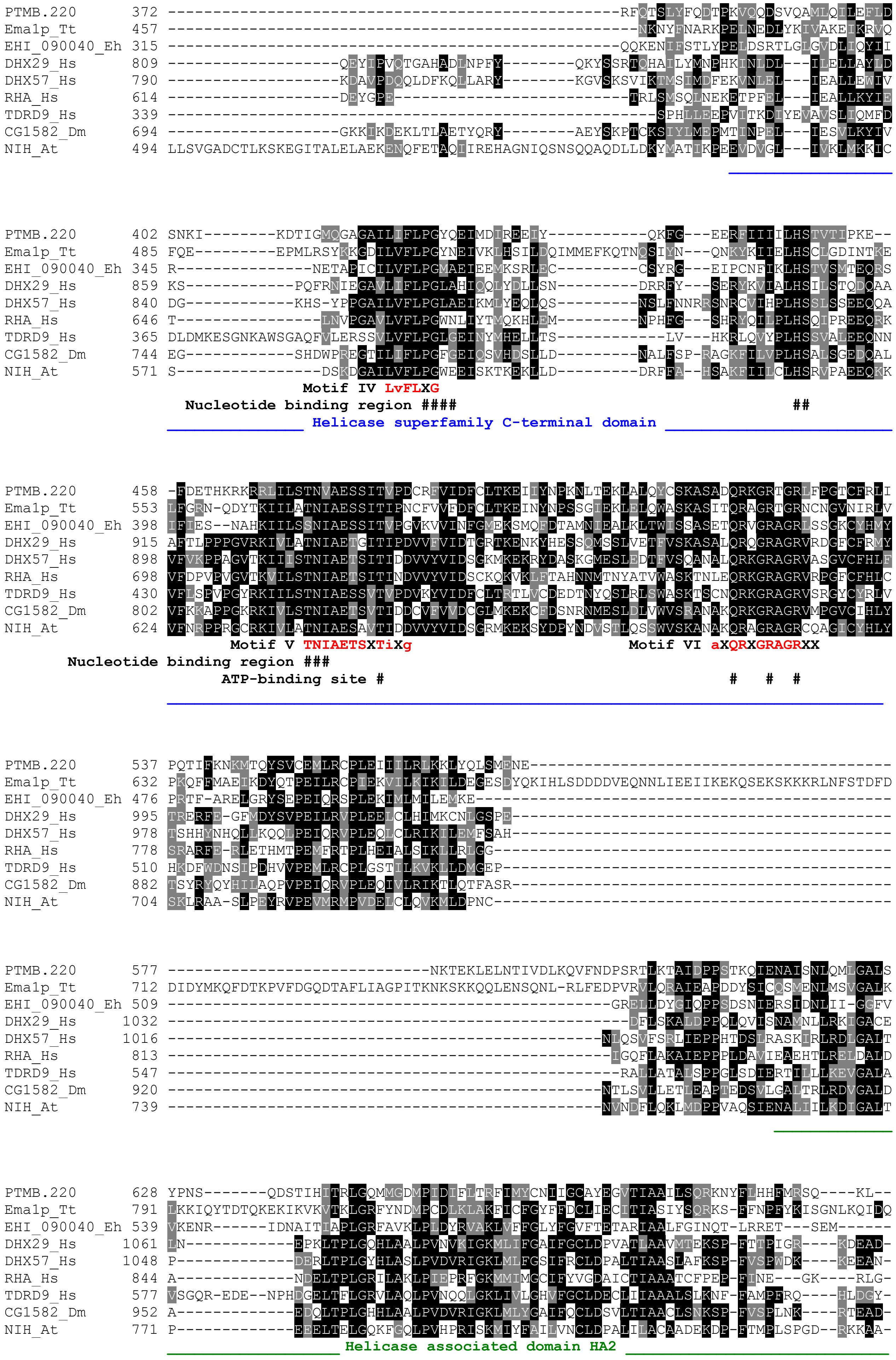


Figure S8. Immunolabeling of *PTMB.170c*-silenced autogamous cells with anti-γ tubulin antibodies
A. Silencing of *PTMB.170c*.
B. Control silencing of the non-essential *ICL7* gene.
Paramecium 51 mt7 cells were grown for 10 vegetative divisions in the presence of bacteria overproducing dsRNA homologous to each target gene. During autogamy, fixed cells were labeled with anti-γ tubulin antibody γ-4 to reveal MICs (panels a‘-d’) and early new MACs (12). DAPI was used to stain all nuclei (panels a-d).
Top panels (early): cells were fixed after the second division of the zygotic nucleus, when MAC development starts. At this stage, anti-γ tubulin antibody labels new MICs (white arrows) and early new MACs (orange arrowheads). No significant difference can be seen between both silencing experiments.
Bottom panels (late): cells were isolated 24 hours later, after karyonidal division. Normally looking MICs are decorated with anti-γ tubulin antibody. Developing MACs can only be detected by DAPI staining.

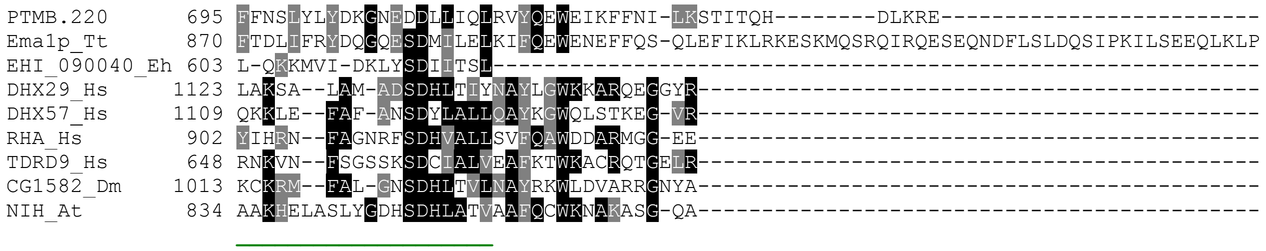
Nowak et al. Figure S9



```
PTMB.220       38 Q-Q------------------------------------------QE--AQISQSTHLRKRV---IKPLTERERKALPIY
Ema1p_Tt      112 F-DEEIQISNTIVKEEIV--------------------------VEE--DQNNSNNRNKKKVFYEFNEPISPEQEELPVH
EHI_090040_Eh  53 L-QEKIQ--------------------------------------------G-M-SMRRS---NPISQKIRVSRENLPAF
DHX29_Hs      503 Q-QQQQQHS--EN-KRENSEDPEESWENLVSDEDFSALSLESANVEDLEPVRNLFRKLQS---TPKYQKLLKERQQLPVF
DHX57_Hs      488 PVIVENESYVNLK-KKISKR------------YDW---QAKSVHAEN-GKICKQFRMKQA---SRQFQSILQERQSLPAW
RHA_Hs        353 AFATPEQISMD---------------------------------LK-----NELMYQLEQ---DHDLQAILQERELLPVK
TDRD9_Hs      111 ---------------P-------------------------------------SLAKLSS---VTCIPGTTYKYPDLPIS
CG1582_Dm     418 R-NVEAQ------------------------------------TREN-RRLLQQFVERRK---EERYQKIIDGRKQLPAF
NIH_At        179 FFKKPHMTKHD---------------------------------IEN--NVVSLSSRLKK---ERHFREIFEARSKLPIA

PTMB.220       70 NVSNKILEEMQNNQVIVITGGTGCGKSTQVPQMILRHHTNTEF-----ETPKTVNILCTQPRRLAAKSLAKRVAQELDCQ
Ema1p_Tt      163 ENRLSILETIKNNSVIVISGNTGCGKSTQIPKFIACDD-------------TSTKIIVTQPRRVAAIALAERVSQELKSN
EHI_090040_Eh  83 QHRQEIIEKIRQNQVVIISGDTGCGKSTQVTQYIFEDARAN---------GEHVKIACTQPRRIAATSLAERVSEEVGCS
DHX29_Hs      576 KHRDSIVETLKRHRVVVVAGETGSGKSTQVPHFLLEDLLLNEW------EASKCNIVCTQPRRISAVSLANRVCDELGCE
DHX57_Hs      548 EERETILNLLRKHQVVVISGMTGCGKTTQIPQFILDDSLNGP-------PEKVANIICTQPRRISAISVAERVAKERAER
RHA_Hs        392 KFESEILEAISQNSVVIIRGATGCGKTTQVPQFILDDFIQND-------RAAECNIVVTQPRRISAVSVAERVAFERGEE
TDRD9_Hs      136 RYKEEVVSLIESNSVVIIHGATGSGKSTQLPQYILDHYVQR---------SAYCSIVVTQPRKIGASSIARWISKERAWT
CG1582_Dm     457 AEIERILALIESSPVVVISGETGCGKSTQVPQFILDNWFFRALQLPAKENLPHVEIICTQPRRLSAIGVAERVAAERLDR
NIH_At        221 SFRDAIISAVESNQVVLIAGETGCGKTTQVPQYLLDHMWHSK--------KEACKIICTQPRRISAISVSDRISWERGET
                             Motif I GeTGTGKT                    Motif Ia XTQPRRXaAXX
                                         S  S
                       ATP binding site #####
                               _______________ DEXH-box helicase domain __________________________

PTMB.220      145 LG-----TLVGYQVGMDSQISSR-QTQIQFVTTGIFLQKLVHDR-EQVLKDYTHIIMDEVHERDIDIDFCLIIIKNLLKH
Ema1p_Tt      230 LG-----QRVGYQIGQDSNF-NR-NTQILYVTIGIFLQKLINDP-DQFLREYTHIILDEVHERGIESDFGLIALKTFLSQ
EHI_090040_Eh 154 LG-----TAVGYQIGMEKQA-DS-NTNIIYLTPGVLLQQLMRR---EGRRMYTHIIIDEAHERDLDTEFCLMLLRKFLSK
DHX29_Hs      650 NGPGGRNSLCGYQIRMESRA-CE-STRLLYCTTGVLLRKLQED---GLLSNVSHVIVDEVHERSVQSDFLLIILKEILQK
DHX57_Hs      621 VG-----LTVGYQIRLESVK-SS-ATRLLYCTTGVLLRRLEGD---TALQGVSHIIVDEVHERTEESDFLLLVLKDIVSQ
RHA_Hs        465 PG-----KSCGYSVRFESIL-PRPHASIMFCTVGVLLRKLE-----AGIRGISHVIVDEIHERDINTDFLLVVLRDVVQA
TDRD9_Hs      207 LG-----GVVGYQVGLEKIA-TE-DTRLIYMTTGVLLQKIVSA---KSLMEFTHIIIDEVHERTEEMDFLLLVVRKLLRT
CG1582_Dm     537 IG-----QLVGYQIRLENKV-SQ-STRLSFCTTGILLRRLASD---PLLGSVTHVIVDEVHERSEESDFLLLILKNLLRE
NIH_At        293 IG-----RTVGYKVRLQSEG-GR-ESSVVFCTNGILLRVLIGKGVNSSVPDITHIIVDEIHERDSYSDFMLMILRDLLPS
                                      Motif Ib YXTdGXLLre          Motif II iXDEaHER
                                                Putative Mg++ binding site ####
                  ________________________________________________________________________________

PTMB.220      218 ---FP-ETKLILMSATICSDKFANYFSKKSLNLVDDLTYIQRVDRKYKFTDMTQENDVEISI--------------KFEG
Ema1p_Tt      302 HKNRN-SIKLIVMSATLNKDIFLNYFSTTELNNLEKEIKEDQDCQEYYFDSSKFSEQLKITKQVILNELYTIFIIYQTKK
EHI_090040_Eh 224 ----T-TVRLIIMSAMMDVEKIQNHFMLIPETGYQMQ-------LQYLLCDF----------------------------
DHX29_Hs      725 ---RS-DLHLILMSATVDSEKFSTYFTH----------------------------------------------------
DHX57_Hs      691 ---RP-GLQVILMSATLNAELFSDYFNS----------------------------------------------------
RHA_Hs        534 ---YP-EVRIVLMSATIDTSMFCEYFFN----------------------------------------------------
TDRD9_Hs      277 ---NSRFVKVVLMSATISCKEFADYFAVPVQNKM----------------------------------------------
CG1582_Dm     607 ---RK-DLKVILMSATLNAALFSDYFGG----------------------------------------------------
NIH_At        366 ---NP-HLRLILMSATLDAERFSEYFGG----------------------------------------------------
                     Motif III SAT
                  _________________

PTMB.220      280 EEDHDISTQWGQPWKEIERQVQLTQEQRELEQKQNELRELGQHFLDGDEDQAAPIIKVPMSQKYPIETYYMEEIHSYFND
Ema1p_Tt      381 QQQNDDYSNWGQQK-----AHQTWNDQDDHDDTYEKMQKMICRFNCKYESNQKNYVEIQVVQKYKIQTVWLEDLKDTQ-K
EHI_090040_Eh 264 ------------------------KEQKEFER------------IDITRLSKPPLLHIG-GKRFPVEEFYIEDVLKLVKK
DHX29_Hs      749 ----------------------------------------------------CPILRIS-GRSYPVEVFHLEDIIEETGF
DHX57_Hs      715 ----------------------------------------------------CPVITIP-GRTFPVDQFFLEDAIAVTRY
RHA_Hs        558 ----------------------------------------------------CPIIEVY-GRTYPVQEYFLEDCIQMTHF
TDRD9_Hs      308 --------------------------------------------------NPAYIFEVE-GKPHSVEEYYLNDLEHIHHS
CG1582_Dm     631 ----------------------------------------------------APVLDIP-GRTFPVQQLFLEDILEMSDF
NIH_At        390 ----------------------------------------------------CPVVRVP-GFTYPVRTFFLDDALSVLNS

PTMB.220      360 ADTAKLELHQPL--------------------------------------------------------------------
Ema1p_Tt      455 LC------------------------------------------------------------------------------
EHI_090040_Eh 307 TFE---EIY-PF--------------------------------------------------------------------
DHX29_Hs      776 VLE---KDS-EYCQKFLEEEEEV---------------TINVTSKAG--GIKKY--------------------------
DHX57_Hs      742 VLQ---DGS-PYMRSMKQISKEKLKARRNRTAFEEVEEDLRLSLHLQ--DQDSV--------------------------
RHA_Hs        585 VPP---PKD-KKKKDKDDDGGEDDDA--------------NCNLICG---------------------------------
TDRD9_Hs      337 KL------------------------------------------------------------------------------
CG1582_Dm     658 VME---YDT-KYCRKLKKQEQEIL------------ERELEYADVQA--SGEAP--------------------------
NIH_At        417 DKNS-HLLS-AVKRDFKDEDKVSLDEAID-LAWTNDEFDCLVDLVSSEGSHEAYNYQNSTTGLTPLMVFAGKGRVSDVCK
```



```
PTMB.220       372 --------------------------------------------------RFQTSLYFQDTPKVQQDSVQAMLQILEFLD
Ema1p_Tt       457 ----------------------------------------------------NKNYFNARKPELNEDLYKIVAKEIKRVQ
EHI_090040_Eh  315 --------------------------------------------------QQKENIFSTLYPELDSRTLGLGVDLIQYII
DHX29_Hs       809 -------------------QEYIPVQTGAHADLNPFY--------QKYSSRTQHAILYMNPHKINLDL---ILELLAYLD
DHX57_Hs       790 -------------------KDAVPDQQLDFKQLLARY--------KGVSKSVIKTMSIMDFEKVNLEL---IEALLEWIV
RHA_Hs         614 -------------------DEYGPE--------------------------TRLSMSQLNEKETPFEL---IEALLKYIE
TDRD9_Hs       339 ------------------------------------------------------SPHLLEEPVITKDIYEVAVSLIQMFD
CG1582_Dm      694 -------------------GKKIKDEKLTLAETYQRY--------AEYSKPTCKSIYLMEPMTINPEL---IESVLKYIV
NIH_At         494 LLSVGADCTLKSKEGITALELAEKENQFETAQIIREHAGNIQSNSQQAQDLLDKYMATIKPEEVDVGL---IVKLMKKIC


PTMB.220       402 SNKI--------KDTIGMQGAGAILIFLPGYQEIMDIREEIY-----------QKFG---EERFIIIILHSTVTIPKE--
Ema1p_Tt       485 FQE---------EPMLRSYKKGDILVFLPGYNEIVKLHSILDQIMMEFKQTNQSIYN---QNKYKIIELHSCLGDINTKE
EHI_090040_Eh  345 R---------------NETAPICILVFLPGMAEIEEMKSRLEC---------CSYRG---EIPCNFIKLHSTVSMTEQRS
DHX29_Hs       859 KS------------PQFRNIEGAVLIFLPGLAHIQQLYDLLSN---------DRRFY---SERYKVIALHSILSTQDQAA
DHX57_Hs       840 DG------------KHS-YPPGAILVFLPGLAEIKMLYEQLQS---------NSLFNNRRSNRCVIHPLHSSLSSEEQQA
RHA_Hs         646 T----------------LNVPGAVLVFLPGWNLIYTMQKHLEM---------NPHFG---SHRYQILPLHSQIPREEQRK
TDRD9_Hs       365 DLDMKESGNKAWSGAQFVLERSSVLVFLPGLGEINYMHELLTS------------LV---HKRLQVYPLHSSVALEEQNN
CG1582_Dm      744 EG------------SHDWPREGTILIFLPGFGEIQSVHDSLLD---------NALFSP-RAGKFILVPLHSALSGEDQAL
NIH_At         571 S----------------DSKDGAILVFLPGWEEISKTKEKLLD---------DRFFA--HSAKFIILCLHSRVPAEEQKK
                                  Motif IV LvFLXG
                     Nucleotide binding region ####                                        ##
                                  Helicase superfamily C-terminal domain


PTMB.220       458 -FDETHKRKRRLILSTNVAESSITVPDCRFVIDFCLTKEIIYNPKNLTEKLALQYCSKASADQRKGRTGRLFPGTCFRLI
Ema1p_Tt       553 LFGRN-QDYTKIILATNIAESSITIPNCFVVFDFCLTKEINYNPSSGIEKLELQWASKASITQRAGRTGRNCNGVNIRLV
EHI_090040_Eh  398 IFIES--NAHKIILSSNIAESSITVPGVKVVINFGMEKSMQFDTAMNIEALKLTWISSASETQRVGRAGRLSSGKCYHMY
DHX29_Hs       915 AFTLPPPGVRKIVLATNIAETGITIPDVVFVIDTGRTKENKYHESSQMSSLVETFVSKASALQRQGRAGRVRDGFCFRMY
DHX57_Hs       898 VFVKPPAGVTKIIISTNIAETSITIDDVVYVIDSGKMKEKRYDASKGMESLEDTFVSQANALQRKGRAGRVASGVCFHLF
RHA_Hs         698 VFDPVPVGVTKVILSTNIAETSITINDVVYVIDSCKQKVKLFTAHNNMTNYATVWASKTNLEQRKGRAGRVRPGFCFHLC
TDRD9_Hs       430 VFLSPVPGYRKIILSTNIAESSVTVPDVKYVIDFCLTRTLVCDEDTNYQSLRLSWASKTSCNQRKGRAGRVSRGYCYRLV
CG1582_Dm      802 VFKKAPPGKRKIVLSTNIAETSVTIDDCVFVVDCGLMKEKCFDSNRNMESLDLVWVSRANAKQRKGRAGRVMPGVCIHLY
NIH_At         624 VFNRPPRGCRKIVLATNIAESAVTIDDVVYVIDSGRMKEKSYDPYNDVSTLQSSWVSKANAKQRAGRAGRCQAGICYHLY
                          Motif V TNIAETSXTiXg                       Motif VI aXQRXGRAGRXX
       Nucleotide binding region ###
                         ATP-binding site #                                    #   #  #


PTMB.220       537 PQTIFKNKMTQYSVCEMLRCPLEIIILRLKKLYQLSMENE----------------------------------------
Ema1p_Tt       632 PKQFFMAEIKDYQTPEILRCPIEKVILKIKILDEGESDYQKIHLSDDDDVEQNNLIEEIIKEKQSEKSKKKRLNFSTDFD
EHI_090040_Eh  476 PRTF-ARELGRYSEPEIQRSPLEKIMLMILEMKE----------------------------------------------
DHX29_Hs       995 TRERFE-GFMDYSVPEILRVPLEELCLHIMKCNLGSPE------------------------------------------
DHX57_Hs       978 TSHHYNHQLLKQQLPEIQRVPLEQLCLRIKILEMFSAH------------------------------------------
RHA_Hs         778 SRARFE-RLETHMTPEMFRTPLHEIALSIKLLRLGG--------------------------------------------
TDRD9_Hs       510 HKDFWDNSIPDHVVPEMLRCPLGSTILKVKLLDMGEP-------------------------------------------
CG1582_Dm      882 TSYRYQYHILAQPVPEIQRVPLEQIVLRIKTLQTFASR------------------------------------------
NIH_At         704 SKLRAA-SLPEYRVPEVMRMPVDELCLQVKMLDPNC--------------------------------------------


PTMB.220       577 -----------------------------NKTEKLELNTIVDLKQVFNDPSRTLKTAIDPPSTKQIENAISNLQMLGALS
Ema1p_Tt       712 DIDYMKQFDTKPVFDGQDTAFLIAGPITKNKSKKQQLENSQNL-RLFEDPVRVLQRAIEAPDDYSICQSMENLMSVGALK
EHI_090040_Eh  509 -------------------------------------------------GRELLDYGIQPPSDSNIERSIDNLII-GGFV
DHX29_Hs      1032 ---------------------------------------------------DFLSKALDPPQLQVISNAMNLLRKIGACE
DHX57_Hs      1016 ------------------------------------------------NLQSVFSRLIEPPHTDSLRASKIRLRDLGALT
RHA_Hs         813 -------------------------------------------------IGQFLAKAIEPPPLDAVIEAEHTLRELDALD
TDRD9_Hs       547 --------------------------------------------------RALLATALSPPGLSDIERTILLLKEVGALA
CG1582_Dm      920 ------------------------------------------------NTLSVLLETLEAPTEDSVLGALTRLRDVGALD
NIH_At         739 ------------------------------------------------NVNDFLQKLMDPPVAQSIENALIILKDIGALT


PTMB.220       628 YPNS-------QDSTIHITRLGQMMGDMPIDIFLTRFIMYCNIIGCAYEGVTIAAILSQRKNYFLHHFMRSQ----KL--
Ema1p_Tt       791 LKKIQYTDTQKEKIKVKVTKLGRFYNDMPCDLKLAKFICFGYFFDCLIECITIASIYSQRKS-FFNPFYKISGNLKQIDQ
EHI_090040_Eh  539 VKENR------IDNAITIAPLGRFAVKLPLDYRVAKLVFFGLYFGVFTETARIAALFGINQT-LRRET---SEM------
DHX29_Hs      1061 LN------------EPKLTPLGQHLAALPVNVKIGKMLIFGAIFGCLDPVATLAAVMTEKSP-FTTPIGR----KDEAD-
DHX57_Hs      1048 P-------------DERLTPLGYHLASLPVDVRIGKLMLFGSIFRCLDPALTIAASLAFKSP-FVSPWDK----KEEAN-
RHA_Hs         844 A-------------NDELTPLGRILAKLPIEPRFGKMMIMGCIFYVGDAICTIAAATCFPEP-FINE---GK----RLG-
TDRD9_Hs       577 VSGQR-EDE--NPHDGELTFLGRVLAQLPVNQQLGKLIVLGHVFGCLDECLIIAAALSLKNF-FAMPFRQ----HLDGY-
CG1582_Dm      952 A-------------EDQLTPLGHHLAALPVDVRIGKLMLYGAIFQCLDSVLTIAACLSNKSP-FVSPLNK----RTEAD-
NIH_At         771 P-------------EEELTELGQKFGQLPVHPRISKMIYFAILVNCLDPALILACAADEKDP-FTMPLSPGD--RKKAA-
                                      Helicase associated domain HA2
```



```
PTMB.220       695 FFNSLYLYDKGNEDDLLIQLRVYQEWEIKFFNI-LKSTITQH--------DLKRE-------------------------
Ema1p_Tt       870 FTDLIFRYDQGQESDMILELKIFQEWENEFFQS-QLEFIKLRKESKMQSRQIRQESEQNDFLSLDQSIPKILSEEQLKLP
EHI_090040_Eh 603 L-QKKMVI-DKLYSDIITSL------------------------------------------------------------
DHX29_Hs      1123 LAKSA--LAM-ADSDHLTIYNAYLGWKKARQEGGYR--------------------------------------------
DHX57_Hs      1109 QKKLE--FAF-ANSDYLALLQAYKGWQLSTKEG-VR--------------------------------------------
RHA_Hs         902 YIHRN--FAGNRFSDHVALLSVFQAWDDARMGG-EE--------------------------------------------
TDRD9_Hs       648 RNKVN--FSGSSKSDCIALVEAFKTWKACRQTGELR--------------------------------------------
CG1582_Dm     1013 KCKRM--FAL-GNSDHLTVLNAYRKWLDVARRGNYA--------------------------------------------
NIH_At         834 AAKHELASLYGDHSDHLATVAAFQCWKNAKASG-QA--------------------------------------------
```

Nowak et al. Table S6

Gene	Coordinates of hybrydization probe	Coordinates of silencing insert
Cluster I - early induction peak		
PTMB.08c	29612 … 30529	30580 … 31351
PTMB.96	nt	208991 … 209866
PTMB.104	nt	220433 … 221343
PTMB.157c	nt	334832 … 336442
PTMB.170c	367695 … 368605	365562 … 366800
PTMB.182	nt	401645 … 403784
PTMB.186c	414934 … 415237	414934 … 415237
PTMB.220	485312 … 486442	483829 … 484842
PTMB.236c	nt	521796 … 523272
PTMB.238c	nt	529894 … 532410
PTMB.239c	nt	532496 … 533275
PTMB.285c	609853 … 610264	610922 … 612118
PTMB.422c	873575 … 874983	876703 … 878481
PTMB.344c	693077 … 693866	694411 … 695111
Cluster II - late induction peak		
PTMB.169c	362850 … 363407	nt
PTMB.219	481729 … 482273	nt
PTMB.443c	932116 … 932675	931872 … 932705
Cluster III - gradually induced genes		
PTMB.10c	36175 … 36761	nt
PTMB.143c	316219 … 316717	nt
Cluster IV - repressed genes		
PTMB.142c	314022 … 314891	nt
PTMB.176c	388926 … 389542	nt
PTMB.400c	798492 … 799038	nt

All coordinates refer to Megabase chromosome (Acc. No. CR548612).

nt – not tested

Table S11. Ohnologs and other paralogs of silenced genes.

Gene	Ohnologs from WGD						Other proteins with AA identity >40%	
	recent WGD		intermediary WGD		old WGD			
	Acc. number (AA identity)	*Co-silenced**	*Acc. number (AA identity)*	*Co-silenced*	*Acc. number (AA identity)*	*Co-silenced*	*Acc. number (AA identity)*	*Co-silenced*
PTMB.08c	-	-	-	-	-	-	-	-
PTMB.96	GSPATG00003747001 (94%)	Yes (11; 363nt)	-	-	-	-	GSPATG00027670001 (54%) GSPATG00032765001 (54%)	No No
PTMB.104	GSPATG00003737001 (31%)	No	-	-	-	-	-	-
PTMB.157c	-	-	-	-	-	-	GSPATG00003686001 (54%) GSPATG00005648001 (50%) GSPATG00011788001 (50%) GSPATG00024302001 (46%) GSPATG00008096001 (46%)	No No No No No
PTMB.170c	-	-	GSPATG00000774001 (52%) GSPATG00008594001 (51%)	No No	GSPATG00008008001 (30%)	No	-	
PTMB.182	GSPATG00003662001 (91%)	Yes (18; 547nt)	-	No	GSPATG00005730001** GSPATG00007856001** GSPATG00008012001** GSPATG00035986001**	No No No No	-	
PTMB.186c	PTETG800015001 (98%)	Yes (4; 149nt)	-	-	-	-	GSPATG00005766001 (97%) GSPATG00007882001 (97%) GSPATG00004184001 (89%) GSPATG00010146001 (89%) GSPATG00024666001 (88%)	No*** No*** No No No
PTMB.220	GSPATG00003619001 (93%)	Yes (6; 193nt)	-	-	-	-	-	-
PTMB.236c	GSPATG00003603001 (79%)	Yes (9; 274nt)	GSPATG00008659001 (62%) GSPATP00000836001 (59%)	No	-	-	-	-
PTMB.238c	GSPATG00003600001 (73%)	Yes (10; 294nt)	-	-	-	-	-	-
PTMB.239c	-	-	-	-	-	-	-	-
PTMB.285c	GSPATG00003553001 (93%)	Yes (3; 100nt)	GSPATG00000904001 (65%)	No	-	-	-	-

PTMB.422c	GSPATG00003412001 (96%)	Yes (15; 590nt)	-	-	-	-	GSPATG00012059001 (40%) GSPATG00013523001 (40%) GSPATG00015493001 (39%)	No
PTMB.344c	-	-	-	-	-	-	-	-
PTMB.443c	-	-	-	-	-	-	GSPATG00034046001 (49%) GSPATG00037262001 (48%)	No

---

* between parenthesis: identical stretches ≥23nt (= Paramecium siRNA length) within RNAi insert and total length covered by identical stretches. Comparison ot silencing inserts with paralogous genes was done using Smith-Waterman alignment tool (http://paramecium.cgm.cnrs-gif.fr/cgi-bin/other/alignmen)

** conserved domains only

*** GSPATG00005766001 and GSPATG00007882001 each share one identical ≥23nt stretch with PTMB.186c silencing insert covering 23 and 29nt, respectively.

SUPPLEMENTARY REFERENCES

1. **Arnaiz, O., S. Cain, J. Cohen, and L. Sperling**. 2007. ParameciumDB: a community resource that integrates the *Paramecium tetraurelia* genome sequence with genetic data. Nucleic Acids Res **35**:D439-444 Epub 2006 Nov 2016.
2. **Aury, J. M., O. Jaillon, L. Duret, B. Noel, C. Jubin, B. M. Porcel, B. Segurens, V. Daubin, V. Anthouard, N. Aiach, O. Arnaiz, A. Billaut, J. Beisson, I. Blanc, K. Bouhouche, F. Camara, S. Duharcourt, R. Guigo, D. Gogendeau, M. Katinka, A. M. Keller, R. Kissmehl, C. Klotz, F. Koll, A. Le Mouel, G. Lepere, S. Malinsky, M. Nowacki, J. K. Nowak, H. Plattner, J. Poulain, F. Ruiz, V. Serrano, M. Zagulski, P. Dessen, M. Betermier, J. Weissenbach, C. Scarpelli, V. Schachter, L. Sperling, E. Meyer, J. Cohen, and P. Wincker**. 2006. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. Nature. **444**:171-178.
3. **Benjamini, Y., and Y. Hochberg**. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc Ser B **57**:289–300.
4. **Bétermier, M., S. Duharcourt, H. Seitz, and E. Meyer**. 2000. Timing of developmentally programmed excision and circularization of *Paramecium* internal eliminated sequences. Mol. Cell. Biol. **20**:1553-1561.
5. **Bolstad, B. M., R. A. Irizarry, M. Astrand, and T. P. Speed**. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics **19**:185-193.
6. **Bourne, Y., M. H. Watson, M. J. Hickey, W. Holmes, W. Rocque, S. I. Reed, and J. A. Tainer**. 1996. Crystal structure and mutational analysis of the human CDK2 kinase complex with cell cycle-regulatory protein CksHs1. Cell. **84**:863-874.
7. **Chenna, R., H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins, and J. D. Thompson**. 2003. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res **31**:3497-3500.
8. **Garnier, O., V. Serrano, S. Duharcourt, and E. Meyer**. 2004. RNA-mediated programming of developmental genome rearrangements in *Paramecium tetraurelia*. Mol. Cell. Biol. **24**:7370-7379.
9. **Gogendeau, D., C. Klotz, O. Arnaiz, A. Malinowska, M. Dadlez, N. G. de Loubresse, F. Ruiz, F. Koll, and J. Beisson**. 2008. Functional diversification of centrins and cell morphological complexity. J Cell Sci. **121**:65-74.
10. **Gratias, A., G. Lepere, O. Garnier, S. Rosa, S. Duharcourt, S. Malinsky, E. Meyer, and M. Betermier**. 2008. Developmentally programmed DNA splicing in *Paramecium* reveals short-distance crosstalk between DNA cleavage sites. Nucleic Acids Res **36**:3244-3251.
11. **Hartigan, J. A., and M. A. Wong**. 1979. A K-means clustering algorithm. Applied Statistics **28**:100-108.
12. **Klotz, C., F. Ruiz, N. Garreau de Loubresse, M. Wright, P. Dupuis-Williams, and J. Beisson**. 2003. Gamma-tubulin and MTOCs in *Paramecium*. Protist **154**:193-209.
13. **Lamers, M. H., A. Perrakis, J. H. Enzlin, H. H. Winterwerp, N. de Wind, and T. K. Sixma**. 2000. The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch. Nature **407**:711-717.
14. **Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins**. 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23**:2947-2948.
15. **Lewis, S. E., S. M. Searle, N. Harris, M. Gibson, V. Lyer, J. Richter, C. Wiel, L. Bayraktaroglir, E. Birney, M. A. Crosby, J. S. Kaminker, B. B. Matthews, S. E.**

Prochnik, C. D. Smithy, J. L. Tupy, G. M. Rubin, S. Misra, C. J. Mungall, and M. E. Clamp. 2002. Apollo: a sequence annotation editor. Genome Biol **3**:RESEARCH0082 Epub 2002 Dec 0023.

16. **Marisa, L., J. L. Ichante, N. Reymond, L. Aggerbeck, H. Delacroix, and M. H. Mucchielli-Giorgi**. 2007. MAnGO: an interactive R-based tool for two-colour microarray analysis. Bioinformatics **23**:2339-2341 Epub 2007 Jun 2322.
17. **Notredame, C., D. G. Higgins, and J. Heringa**. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol **302**:205-217.
18. **Nowacki, M., W. Zagorski-Ostoja, and E. Meyer**. 2005. Nowa1p and Nowa2p: novel putative RNA binding proteins involved in *trans*-nuclear crosstalk in *Paramecium tetraurelia*. Curr Biol. **15**:1616-1628.
19. **Praefcke, G. J., and H. T. McMahon**. 2004. The dynamin superfamily: universal membrane tubulation and fission molecules? Nat Rev Mol Cell Biol. **5**:133-147.
20. **Prakash, B., L. Renault, G. J. Praefcke, C. Herrmann, and A. Wittinghofer**. 2000. Triphosphate structure of guanylate-binding protein 1 and implications for nucleotide binding and GTPase mechanism. EMBO J. **19**:4555-4564.
21. **Rickman, D. S., C. J. Herbert, and L. P. Aggerbeck**. 2003. Optimizing spotting solutions for increased reproducibility of cDNA microarrays. Nucleic Acids Res **31**:e109.
22. **Smyth, G. K.** 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol **3**:Article3.
23. **Smyth, G. K., and T. Speed**. 2003. Normalization of cDNA microarray data. Methods **31**:265-273.
24. **Tanner, N. K., and P. Linder**. 2001. DExD/H box RNA helicases: from generic motors to specific dissociation functions. Mol Cell **8**:251-262.
25. **Theis, K., P. J. Chen, M. Skorvaga, B. Van Houten, and C. Kisker**. 1999. Crystal structure of UvrB, a DNA helicase adapted for nucleotide excision repair. Embo J **18**:6899-6907.
26. **Timmons, L., and A. Fire**. 1998. Specific interference by ingested dsRNA. Nature **395**:854.
27. **Zagulski, M., J. K. Nowak, A. Le Mouel, M. Nowacki, A. Migdalski, R. Gromadka, B. Noel, I. Blanc, P. Dessen, P. Wincker, A. M. Keller, J. Cohen, E. Meyer, and L. Sperling**. 2004. High coding density on the largest *Paramecium tetraurelia* somatic chromosome. Curr Biol **14**:1397-1404.