

Probing small ribosomal subunit RNA helix 45 acetylation across eukaryotic evolution

Bortolin-Cavaillé Marie-Line⁽¹⁾, Quillien Aurélie^{(1)(#)}, Thalalla Gamage Supuni^{(2)(#)}, Thomas Justin M.⁽²⁾, Sas-Chen Aldema⁽⁴⁾⁽⁵⁾, Sharma Sunny⁽³⁾⁽⁶⁾, Plisson-Chastang Célia⁽¹⁾, Vandel Laurence⁽¹⁾⁽⁷⁾, Blader Patrick⁽¹⁾, Lafontaine Denis L.J.⁽³⁾, Schwartz Schraga⁽⁴⁾, Meier Jordan L.⁽²⁾ and Cavaillé Jérôme^{(1)(*)}

⁽¹⁾ Molecular, Cellular and Developmental Biology unit (MCD), Centre de Biologie Intégrative (CBI), University of Toulouse; UPS; CNRS; 118 route de Narbonne, 31062 Toulouse, France

⁽²⁾ Chemical Biology Laboratory, National Cancer Institute, Frederick, Maryland 21702, United States.

⁽³⁾ RNA Molecular Biology, Fonds National de la Recherche Scientifique (FRS/FNRS), Université Libre de Bruxelles (ULB), B-6041 Gosselies, Belgium.

⁽⁴⁾ Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel.

⁽⁵⁾ Present address: The Shmunis School of Biomedicine and Cancer Research, The George S. Wise Faculty of Life Sciences, Tel Aviv University.

⁽⁶⁾ Present address: Department of Cell Biology and Neurosciences, Rutgers, The State University of New Jersey, Piscataway, NJ-08904.

⁽⁷⁾ Present address: Université Clermont Auvergne, CNRS, INSERM, iGReD, Clermont-Ferrand, France

(*) Corresponding author: Jerome.cavaille@univ-tlse3.fr; tel +33(0)561335927; fax: +33(0)561335886

(#) equally contribution

Abstract

NAT10 is an essential enzyme that catalyzes the formation of N⁴-acetylcytidine (ac⁴C) in eukaryotic transfer RNA (tRNA) and 18S ribosomal RNA (rRNA). Recent studies in human cells suggested that rRNA acetylation is dependent on SNORD13, a non-canonical box C/D small nucleolar RNA (SNORD) predicted to base-pair with 18S rRNA via two antisense elements. However, the selectivity of SNORD13-dependent cytidine acetylation and its relationship to NAT10's essential function in pre-rRNA processing remain to be defined. Here, we used CRISPR-Cas9 genome editing to formally demonstrate that SNORD13 is required for acetylation of a single cytidine residue of human and zebrafish 18S rRNA. In-depth characterization revealed that SNORD13-dependent ac⁴C is dispensable for yeast or human cell growth, ribosome biogenesis, translation, and the development of multicellular metazoan model organisms. This loss of function analysis inspired a cross-evolutionary survey of the eukaryotic rRNA acetylation 'machinery' that led to the characterization of many novel SNORD13 genes in phylogenetically-distant metazoans and more deeply rooted photosynthetic organisms. This includes an atypical SNORD13-like RNA in *D. melanogaster* which appears to guide ac⁴C to 18S rRNA helix 45 despite lacking one of the two rRNA antisense elements. Finally, we discover that *C. elegans* 18S rRNA is not acetylated despite the presence of an essential NAT10 homolog. Altogether, our findings shed light on the molecular mechanisms underlying SNORD13-mediated rRNA acetylation across the eukaryotic tree of life and raise new questions regarding the biological function and evolutionary persistence of this highly conserved rRNA base modification.

Introduction

N4-acetylcytidine (ac⁴C) in eukaryotic and archaeal rRNA was first documented in the 1970's, but the catalytic machinery and specific sites of this base modification have been defined only recently (1,2). In eukaryotes, ac⁴C occurs at two positions within rRNA (helix 34 and 45) where it is catalyzed by the nucleolar GCN5-related RNA acetyltransferase NAT10, also referred to as Kre33 or Rra1 in yeast (3-5). In addition, recent studies showed that certain archaeal species harbour >100 ac⁴C residues in their rRNA, including at helix 45, placing cytidine acetylation on par with prevalent nucleoside modifications such as pseudouridines and 2'-O-methylations in these organisms (6,7). Beside its involvement in rRNA modification, NAT10/Kre33 (assisted by its specific co-factor THUMP1/Tan1) can modify C12 of serine and leucine tRNAs (5,8) and has been suggested to modify eukaryotic mRNAs (9,10). Orthogonal validation has, however, called into question the involvement of NAT10 in mRNA modification (6). In addition to modifying RNA, NAT10 may also be implicated in protein acetylation (11-13). Moreover, NAT10 is essential for pre-rRNA processing reactions leading to synthesis of 18S rRNA (5), and unsurprisingly it was associated with human diseases (14,15). The importance of helix 45 ac⁴C for cell homeostasis has not been addressed.

In yeast, the ability of Kre33 to specifically acetylate 18S rRNA was demonstrated to rely on two box C/D small nucleolar RNAs (SNORD), namely snR4 and snR45, which form imperfect base pairing interactions surrounding the targeted cytidine. Indeed, knockout of snR4 and snR45 led to the specific disappearance of 18S rRNA acetylation at SSU-C1280 (helix 34) and SSU-C1773 (helix 45), respectively, while leaving tRNA acetylation unaffected (16). SNORDs are associated with four core proteins (SNU13, NOP56, NOP58, and Fibrillarin) and form nucleolar ribonucleoprotein (RNP) complexes known to mediate site-specific ribose methylation of rRNA, U6 spliceosomal snRNA and tRNAs; some of them are also involved in early nucleolar pre-

rRNA processing (17). The possible contribution of snoRNA-associated proteins in rRNA acetylation remains undefined. Of note, SNORD13 (formerly U13) displays two rRNA complementarity tracts reminiscent of those present on snR45 (**Figure 1**), strongly implicating it as a human counterpart (5,18). Consistent with this notion, inhibiting SNORD13 expression in a cancer cell line using antisense oligonucleotides causes a ~50% decrease in levels of ac⁴C in 18S rRNA (5). However, the nucleotide-specificity of this inhibition and the residual level of acetylation did not afford to easily assess the impact of the modification on cell growth or ribosome biogenesis. Finally, three SNORD13-like RNA species were also reported in *A. thaliana* (snoR105, snoR108 and snoR146) but their roles in rRNA acetylation still awaits investigation (19).

Outside their potential to form two imperfect base-pairing interactions with 18S rRNA, vertebrate SNORD13 and its *S. cerevisiae* and *A. thaliana* homologs are not highly conserved. This explains why, although SNORD13 was discovered thirty years ago (18,20), its relationship with snR45 was only recognized recently (16). This suggests two distinct evolutionary hypotheses: either SNORD-mediated rRNA acetylation emerged early in a common ancestor of plants, fungi and animals and diverged thereafter, or alternatively, SNORD13 arose independently in these different lineages via convergent evolution. Irrespective of the scenario, the possibility also exists that SNORD13 has been lost in some lineages and, accordingly, helix 45 acetylation may rely on a stand-alone enzyme, or no longer be present in rRNA in certain species. This poses the question as to how SNORD-guided cytidine acetylation is conserved across the eukaryotic tree of life, and what cross-organismal commonalities in the nucleic acid acetylation machinery and its rRNA substrates govern this process.

Here, we use CRISPR-Cas9 genome editing to delete SNORD13 in human cells, confirming the hypothesis that this snoRNA is specifically required for acetylation of a single cytidine residue in

small subunit ribosomal RNAs (SSU-ac⁴C1842 located in rRNA helix45). Using this cell line and related model systems, we further establish that loss of SSU-ac⁴C1842 does not lead to obvious defects in terms of cell growth, protein synthesis, ribosome biogenesis, or the development of metazoan model organisms (zebrafish, flies). Finally, considering that Kre33/NAT10 is essential for pre-rRNA processing (5) while ac⁴C in helix 45 appears dispensable, we surveyed the conservation of the ribosomal RNA acetylation machinery across the eukaryotic tree of life. This cross-evolutionary analysis revealed that SNORD-mediated rRNA acetylation is widely conserved in Metazoans, but defined *C. elegans* as a multicellular organism in which rRNA acetylation has been lost. Our findings illuminate the molecular mechanisms underlying SNORD13-mediated rRNA acetylation across evolution and raise new questions regarding the role of highly conserved SSU-ac⁴C1842.

Results

Genetic deletion of human SNORD13 gene causes loss of SSU-ac⁴C1842 - As a first step to probe directly the functional role of SNORD13 in rRNA modification, the human SNORD13 gene was inactivated in the haploid human chronic myelogenous leukemia-derived HAP1 cell line (21). In order to avoid potential transcriptional interference with the adjacent TTI2 gene, the 3'-end region of SNORD13 was targeted for CRISPR-Cas9-mediated deletion (**Figure 2A**). PCR analysis identified three independent SNORD13-deficient clones (#12/#20/#30) harboring the expected deletion (not shown), which rendered SNORD13 undetectable by Northern blot (**Figure 2B**). Three additional clones (#13/#19/#22) were identified that were exposed to CRISPR/Cas9 without any deletion at SNORD13, providing a set of matched control cell lines subjected to identical clonal selection procedures and/or putative off-target effects. Unless otherwise specified, in subsequent experiments SNORD13-KO and WT are used to refer to a

stoichiometric mixture of these three SNORD13-deficient and SNORD13-expressing clones, respectively.

To define the site-specific activity of SNORD13, we took a dual approach to analyze our mutant and WT cell lines. First, we directly measured ac⁴C levels in helix 45 using a mung bean nuclease protection assay coupled to reversed phase high performance liquid chromatography (RP-HPLC, (5)). Briefly, total RNA extracted from SNORD13-KO and WT cells were hybridized with an antisense DNA oligonucleotide overlapping helix 45 and its flanking sequences and were treated with mung bean nuclease. Protected rRNA fragments were then gel-purified and nucleoside hydrolysates obtained after nuclease P1 digestions were subjected to RP-HPLC analysis. As shown in **Figure 2C**, cytosine acetylation at helix 45 was completely abolished in SNORD13-KO cells as compared to WT. As orthogonal validation, RNA isolated from SNORD13-KO and WT cells was analyzed using ac⁴C sequencing chemistry (22). In this approach total RNA is treated with sodium cyanoborohydride under acidic conditions, causing chemical reduction of the acetylated residue. This reduced nucleobase causes misincorporation of non-cognate NTPs during reverse transcription (RT), enabling ac⁴C sites to be detected as C-to-T mutations upon cDNA sequencing. Applying this approach to the two cell lines, we detected SNORD13-dependent misincorporation at SSU-ac⁴C1842 but not at SSU-ac⁴C1337 (**Figure 2D**). These experiments validated our knockout strategy and unambiguously defined SNORD13 as a human counterpart of yeast snR45 required for guiding formation of ac⁴C at the terminal stem-loop of human 18S rRNA.

SNORD13 deletion does not disrupt cell growth, pre-rRNA processing or protein synthesis in human cells - Next, we assessed the phenotypic consequences of SNORD13 deletion and/or SSU-ac⁴C1842 loss. Analysis of proliferation by cell counting revealed that loss of SNORD13 did not impact cell growth significantly (**Figure 2E**) and that cell death rates and

modalities were within normal range (**Figure 2F**). We next examined the levels of pre-rRNA precursors and mature rRNAs in SNORD13-KO cells by Northern blot (**supplementary Figure 2A-B**), RNase H mapping (**supplementary Figure 2C**) and pulse chase experiments (**supplementary Figure 2D**). These analyses revealed that loss of SNORD13 and/or SSU-ac⁴C1842 does not interfere with normal pre-rRNA processing. An involvement of SNORD13 and/or SSU-ac⁴C1842 on ribosomal subunit assembly and global translation was also investigated by performing polysomal analysis. In this assay, whole cytoplasmic extracts from SNORD13-KO and WT cells were subjected to sucrose density gradient fractionation (10-50%). As shown in **supplementary figure 2E**, the distribution and amplitude of peaks corresponding to 40S, 60S, 80S monosomes and polysomes were virtually identical in SNORD13-KO and WT cells. Moreover, detection of nascent protein production by puromycin labelling (SUnSET assay) revealed identical, translation-dependent incorporation patterns in SNORD13-KO and WT cells, again indicating that overall protein synthesis is unlikely to be severely impaired in SNORD13-deficient cells (**Figure 2G**). To test whether SSU-ac⁴C1782 plays a more subtle role in fine-tuning aspects of ribosome function such as translational fidelity, we transfected WT and SNORD13-KO cells with firefly luciferase reporter genes bearing in-frame stop codons (UAA, UAG) or point mutations (R218S) that abrogate enzymatic luciferase activity. These reporters enable any decrease in translational accuracy to be detected via a corresponding increase in firefly luciferase activity. Our data indicate that stop codon read-through and amino acid misincorporation in SNORD13-KO cells was similar between WT and KO cells (**Figure 2H-left panel**). Finally, when using a set of Renilla-IRES-Firefly bicistronic vectors, we also showed that the efficiency of internal ribosome entry site (IRES)-mediated translation was apparently unaffected in SNORD13-KO cells (**Figure 2H-right panel**). Overall, these findings suggest that SNORD13 and SSU-ac⁴C1842 are not essential for ribosome biogenesis or function; at least in standard laboratory growth conditions.

SNORD13-deficient zebrafish embryos develop normally - In order to understand the biological relevance of SNORD13 at the whole organism level, a 189 bp-long, CRISPR-Cas9 mediated deletion overlapping the *Danio rerio* (zebrafish) *snord13* gene locus on chromosome 8 was generated (**Figure 3A**). Crossing of individual heterozygotes led to the generation of homozygous adult *snord13* knockout fish of normal appearance (not shown). To exclude any maternal contribution, three different homozygous females were crossed with heterozygous males. As expected, *snord13* and ac⁴C in helix 45 were no longer detectable in RNA isolated from homozygous knockout fish, as judged by Northern blotting (**Figure 3B**) and ac⁴C sequencing (**Figure 3C**), respectively. Homozygous *snord13* individuals in the progeny did not present any apparent morphological defects at 30 hours post fertilization (30 hpf) and developed normally until adulthood (**Supplementary Figure 3A**). To explore the possible occurrence of more subtle defects, such as activation of the nucleolar stress pathway in response to impaired ribosome biogenesis (23), we performed immunostaining against activated Caspase-3. No increase in apoptotic cells was observed in the developing brain (**Figure 3D-E**) or spinal cord (**Figure 3F-G**) of *snord13*-deficient embryos, as compared to WT. Accordingly, the level of expression of the pro-apoptotic *bcl2*, *bax* and *puma* genes, as judged by RT-qPCR, was not significantly altered in *snord13* mutant embryos (**supplementary Figure 3B**). Finally, we performed immunostaining in WT and mutant embryos for the pan neural marker HuC/D which labels newly differentiated neurons in different structures including olfactory epithelium, pineal gland or optic tectum and the spinal cord. Again, no obvious difference could be detected between WT and mutant embryos (**Figure 3D-G**). Beside central nervous system development, we also sought to explore whether *snord13* mutant displayed defects in another process such as vasculogenesis. We found that the vasculature of the trunk which is composed of the dorsal aorta (DA), the cardinal vein (CV) and intersegmental vessels was also normal in 30 hpf *snord13*-deficient embryos, as revealed by *in situ* hybridization against the key vascular transcription factor *fli1a* (**Figures 3H-I**). Similarly, the expression levels of various markers of

the neural system (*lhx1a*, *pax6a* and *otx5*) and of the dorsal aorta trunk vasculature (*cdh5* and *fli1a*) were unchanged between WT and snord13-KO embryos, validating the lack of phenotypes observed (**Supplementary Figure 3B**). In conclusion, neither SNORD13 nor 18S rRNA ac⁴C in helix 45 appear to play any essential roles in embryonic zebrafish development.

A cross-evolutionary survey of eukaryotic ribosomal RNA acetylation machinery – To more fully apprehend the evolutionary origin of SNORD-mediated rRNA acetylation, we revisited the phylogenetic distribution of SNORD13 throughout eukaryotes, focusing primarily on Metazoan (animals) and Archaeplastida (organisms capable of photosynthesis). Using BLAST searches with low stringency algorithm parameters together with manual inspection, we identified many novel candidate SNORD13 homologs outside the vertebrate lineage, including Urochordata, Cephalochordata, Echinodermata, Arthropoda, Mollusca, Brachiopoda, Cnidaria and even Porifera (**Figure 4A/B**). As depicted in **Figure 4B**, many of the newly-identified SNORD13 genes are located within introns, both in the sense or antisense orientation with respect to transcription of their host-gene transcript. In Archaeplastida, SNORD13 was also uncovered in phylogenetically distant land plants, including in the freshwater green alga *C. braunii* (**Figure 4C**). A full list of newly-identified SNORD13 and proposed snoRNA:rRNA duplexes is provided in **Supplementary Data 1** and **Supplementary Figure 1**, respectively. These sequence analyses were accompanied by experimental profiling of ac⁴C at helix 45 in representative eukaryotic species. As listed in **Figure 4A**, acetylated cytidine was readily detected in the vast majority of organisms we examined, including unicellular eukaryotes such as slime molds (*P. polycephalum*), ciliates (*P. tetraurelia*, *P. falciparum*) or microalgae (*P. tricornutum*). We found, therefore, that SNORD13 is much more broadly distributed than initially thought and SNORD-mediated RNA acetylation is a widespread principle in eukaryotes. Consistent with this conclusion, NAT10 homologs were found in almost all organisms we queried, with the notable exception of *E. gracilis* (**Figure 4A, Supplementary data 2**).

Atypical SNORD13-like homologs in Diptera – SNORD13-like genes are present in many groups of Arthropoda (**Figure 5A, Supplementary data 1**), including *D. melanogaster* and *B. mori* where they had previously been identified as SNORDs of unknown functions (named Or-CD1 and Bm-10, respectively; (24,25)). In Diptera for which a genome annotation is available, Or-CD1 can be found either in large introns positioned in the 5'-UTR of two transcript isoforms generated from two alternative transcription sites (e.g. Swim, Tinagl1_2), or within short introns embedded in the open reading frame of the host-gene (e.g. Zfp15, Rps11) (**Figure 5B**). To our surprise, the published sequence of Or-CD1, as well as that of all newly-identified SNORD13-like in Diptera, lack conserved antisense rRNA sequence at their 5'-end (**Figure 5C**). One potential explanation would be that the sequence upstream of the C-box is removed by exonucleolytic trimming of the spliced-out host-intron, very likely due to the absence of a 5'-capped structure which normally plays a protective role against degradation (26). Although plausible for small introns, such as those in Zfp15 or Rps11 genes, this hypothesis cannot account for all observations. For example, many Or-CD1-like in Arthropoda are located in introns (**Figure 4B**), yet they still retain a conserved rRNA complementarity in their 5'-end segment (**Figure 5C, Supplementary Figure 1**). As an alternative, we reasoned that being intronic may not necessarily rule out independent transcription from distinct promoter regions. This is supported by the fact that SNORD13 can be found positioned either in the sense or antisense orientation relative to transcription of their host-gene (**Figure 4B/5B**), a scenario that is not compatible with post-transcriptional processing from introns.

To probe the mode of expression of Or-CD1 in Arthropoda, total RNAs extracted from *D. melanogaster*, *L. humile* (ants) and *A. mellifera* (bees) were subjected to immunoprecipitation with R1131 antibodies that recognize the 2,2,7 trimethylguanosine (TMG) cap structure, enabling its use as a proxy for independent transcription. Consistent with the expected size for small RNAs lacking 5'-end extension in *D. melanogaster*, Northern blot analyses detected a

radioactive signal migrating at about 110 nucleotides (**Figure 5D, see also Figure 5F**). These RNA species were specifically immunoprecipitated with the R1131 antibodies indicating that, despite being located within an intron, Or-CD1 gene is independently transcribed. This conclusion is in line with prior studies showing that Or-CD1 represents one of the very few SNORD genes that recruit RNA polymerase II and the Little Elongation Complex (LEC;(27)). In agreement with phylogenetic comparison that identifies two conserved antisense rRNA elements (**Figure 5C**), Or-CD1 displays a 5'-capped extension in both *L. humile* and *A. mellifera* (**Figure 5D**), as evidenced by the sizes of RNAs recovered in the immunoprecipitation pellet fraction. This therefore identifies a second expression strategy by which independent transcription from introns can also yield “regular” extended forms of Or-CD1 (**Figure 5E**).

Thus, Or-CD1 in Diptera - but not in Arthropoda - displays only one antisense rRNA sequence (**Figure 5C**). In order to test if this remarkable fly SNORD13 was capable of guiding rRNA acetylation, a strain bearing a P-element inserted in the Or-CD1 gene was retrieved from the Bloomington Drosophila Stock Center (Dmel\PсноRNA:Or-CD1G9117). After demonstrating that Or-CD1 was no longer detected in the mutant adult flies (**Figure 5F**), we showed that acetylation in helix45 was abolished (**Figure 5G**). Thus, SSU-C1968 in *D. melanogaster* can be targeted for acetylation by a SNORD13 homolog displaying a single antisense element. Of note, Or-CD1 adult knockout flies are viable, fertile and do not manifest gross abnormalities (not shown), thus strengthening the notion that N4-acetylcytidine at helix 45 is largely dispensable for viability, at least under standard laboratory conditions. Based on our findings, we propose to rename Or-CD1 as SNORD13.

Life without 18S rRNA acetylation: the cases of *C. elegans* and genetically modified $\Delta snR4\Delta snR45$ mutant yeast. A key role in ribosome biology of 18S rRNA helix 45, which in mature small ribosomal subunits lies directly adjacent to the decoding site, is implied by its

extreme sequence conservation in many phylogenetically-distant organisms belonging to all known eukaryotic supergroups. Furthermore, when mutations in helix 45 arise, they are often accompanied by compensatory base changes (**Supplementary Figure 4A**). The 5'-CCG-3' motif - wherein the second C is targeted for acetylation - was established as a key specificity sequence determinant for NAT10 (3,6). Remarkably, a few well-studied organisms lack this conserved motif or even the targeted substrate cytidine, including flagellated protozoan parasites (*T. vaginalis*, *T. brucei*, *L. major*), protozoan algae (*E. gracilis*), ciliate protozoa (*T. thermophila*), slime mold (*P. polycephalum*), corn smut fungus (*U. maydis*) or even the nematode (*C. elegans*) (**Figure 4A**). Other examples can be found in **Supplementary Figure 4B**. Only in the slime mold *P. polycephalum* did our ac⁴C sequencing reaction provide evidence for helix 45 acetylation within a non-consensus 5'-UCG-3' motif, suggesting that NAT10 may have a distinct biochemical specificity in this organism (**Supplementary Figure 5**). Of note, helix 34 of *P. polycephalum* is apparently not acetylated despite the presence of a similar 5'-UCG-3'. Given its importance as animal model and the prior determination that it carries an essential NAT10 gene (named *nath-10*)(28), we next focused on *C. elegans*. As shown in **Figure 6A-6B**, helix 45 of *C. elegans* 18S rRNA harbors a 5'-CUG-3' motif also found in other nematodes, notably those belonging to the Clades IV and V. By definition, this motif variant that lacks a middle cytidine cannot be acetylated canonically. Nonetheless, since NAT10 is present in the worm genome, we verified the absence of any other ac⁴C residue within helix 45 (**Figure 6C**, not shown). More surprisingly, we found that the 5'-CCG-3' motif in helix 34 was also devoid of acetylation (**Figure 6C**). The lack of rRNA acetylation in *C. elegans* elsewhere than in helices 34 and 45 was further examined using anti-ac⁴C immuno-northern blotting, in which signals for acetylated RNA were found to be limited to tRNAs (**Figure 6D**). It is important to note that ac⁴C-free ribosomes are not necessarily a universal feature of nematodes. Indeed, helix 45 was found to be acetylated in *S. carpocapsae*, an entomopathogenic nematode of the *Steinernematidae* family (**Supplementary Figure 5**). The evolutionary forces that drove

disappearance of rRNA acetylation in *C. elegans* remain unknown. Nonetheless, our results indicate that the changes in robustness of vulval cell-fate specification observed when the essential *nath10* gene was mutated (28,29) did not reflect loss of rRNA modification but more likely the involvement of *nath10* in pre-rRNA processing. A similar conclusion was reached recently when it was shown that the involvement of Fibrillarin in neural crest cell maturation depends on its role in pre-rRNA processing rather than in methylation (doi.org/10.1101/2021.11.25.469989). This is similar to prior observations in *S. cerevisiae* where Kre33 was shown to be essential but snR45 and/or SSU-ac⁴C1773 were not (16). As further validation of ac⁴C's dispensability in rRNA we interrogated a double mutant yeast strain in which snR4 and snR45 genes were simultaneously deleted, producing rRNA completely devoid of ac⁴C. As observed in **Figure 6E**, growth of double Δ snR45 Δ snR4 mutants, as assayed in liquid or solid media, was undistinguishable from that of its wild-type counterpart whether at 30°C or higher sub-optimal temperatures. The lack of obvious phenotypes for Δ snR4 Δ snR45 mutants at 30°C was also previously reported (29). Collectively, these data are in agreement with previous findings showing that Δ Kre33 yeasts re-expressing a catalytic-dead form of Kre33 (H545A or R637A) grow normally, as compared to Δ Kre33 yeasts re-expressing a WT form of Kre33 (16). Thus, we confirm that the absence of rRNA acetylation is unlikely to contribute to the essential function of Kre33. This latter lies very likely with its modification-independent requirement for pre-rRNA processing events leading to the production of mature 18S rRNA (5). This conclusion does not, however, rule out some roles for tRNA acetylation, particularly under suboptimal growth conditions. Indeed, *tan1* mutant strains show reduced growth rates at higher temperature (39°C), presumably due to decreased stability of unmodified serine tRNAs (30).

Discussion

In order to interrogate the function of a single SNORD13-dependent cytidine acetylation in helix 45 of 18S rRNA, we have combined a cross-evolutionary survey of rRNA acetylation machinery across the eukaryotic tree of life and loss-of-function studies in four distinct models: budding yeasts, human cells, zebrafish and fly. This study represents one of the very few examples addressing the physiological roles of a SNORD in multicellular organisms bearing a germline-derived knockout (31-35). Although this peculiar rRNA ac⁴C is highly conserved, its loss surprisingly does not appear to grossly impact ribosome biogenesis or function, or even embryonic development. Consistent with this observation, we identify *C. elegans* as an evolutionary example of dispensable metazoan rRNA acetylation. Finally, the discovery of a 5' truncated form of SNORD13 in *D. melanogaster* suggests that the mechanism by which SNORDs assist NAT10 in catalyzing rRNA ac⁴C is likely more diverse than previously anticipated and in some case may rely on a single antisense element.

An important question raised by our study concerns the specific role of ac⁴C in 18S rRNA helix 45. In humans, the ribosomal protein eL41 (formerly RpL41) is positioned at the interface between the 40S and 60S subunits (36,37) and directly contacts ac⁴C1842 (38). This led to the proposal that acetylation may mediate functionally important crosstalk between the two subunits (38). However, our experimental settings failed to uncover any robust defects in ribosomal subunit assembly or protein synthesis in human SNORD13-KO cells. This is perhaps not surprising since eL41 knockout in yeast does not affect translation unless formation of another inter-subunit contact is compromised (39). A prominent role of ac⁴C in eL41 assembly and/or RNA::protein interaction is also quite unlikely since eL41 is correctly incorporated into ribosomes of *T. vaginalis* (40), *E. gracilis* (41), *L. donovani* (42) and *T. brucei* (42), all of which lack acetylation in helix 45 (**Supplementary Figure 6**). Even more tellingly, the relative

positioning of eL41 in ribosome and the 3-D architecture of helix 45 in the SSU of the archaea *T. kodakarensis* are not affected when TkNAT10 gene is deleted (6). Altogether, these observations point to a subtle molecular role for ac⁴C, possibly in fine tuning ribosome structure and/or function in sub-optimal environmental contexts and/or sensitized genetic backgrounds. Unravelling the precise physiological relevance of 18S rRNA acetylation in multicellular organisms, if any, represents a conceptually and technically formidable, but important, task for future studies.

Our bioinformatics search has considerably expanded the inventory of SNORD13 sequences in metazoans, identifying homologs well beyond vertebrates where they were previously thought to be restricted. However, we were unable to identify SNORD13 counterparts in many major phyla including Hemichordata, Annelida, Nematoda, Fungi (other than budding yeasts), algae or unicellular eukaryotes. Although the most parsimonious hypothesis would be that rapidly-evolving SNORD13 sequences make them difficult to detect, one cannot exclude the possibility that SNORD13 may have been lost in these lineages, which would imply that in these organisms rRNA acetylation may be catalyzed by NAT10 alone. More sophisticated structural and biochemical analyses will be critical to solve this question. An additional related question is whether acetylation of helix 34 is also strictly dependent on an antisense SNORD, as shown only in *S. cerevisiae* to date (16). The characterization of RNA-guided rRNA acetylation systems in Archaea, should they exist, represents an additional avenue that may lend invaluable evolutionary and mechanistic insights into this process (6,7).

Single nucleotide resolution cross-linking analysis has previously described how SNORDs guide the acetyltransferase to its rRNA substrate by use of two short base-pairing interactions on each side of the substrate cytosine, forming a “modification pocket” akin to H/ACA snoRNAs (SNORA) acting in pseudouridylation (16). During the course of this study, we unexpectedly

identified independently-transcribed SNORD13 gene in *D. melanogaster* that produces RNA species lacking 5'-sequence complementary to rRNA (**Figure 7**). The same may also hold true in other Diptera for which SNORD13 is embedded within a very short intron (e.g. *L. sericata*, *M. domestica*). In this case, it is indeed more than likely that exonucleolytic processing of the host-intron produces uncapped RNAs with the C- and D-boxes brought together by a short terminal 5'-3' terminal stem, as commonly found for other intron-located SNORD (26). The genomic constraints and selection pressure that led to changes in SNORD13 expression strategies across evolution are unknown, yet they have inherent mechanistic repercussions since the deposition of ac⁴C in helix 45 in *D. melanogaster* does not strictly require two antisense elements flanking the rRNA ac⁴C site. This finding contrasts with conclusions drawn from mutagenesis of yeast snR45 where the 5' antisense element appears essential for acetylation (16). It therefore implies an alternative mode of action of SNORD13 in *D. melanogaster* and, also more broadly, in Diptera. Finally, the overall structure of SNORD13 in Diptera is indistinguishable from that of the canonical intron-located SNORDs which mediate rRNA ribose methylation. It thus raises the intriguing question as to whether other intronic SNORDs in these organisms, particularly those of still elusive function, might also mediate acetylation on still unidentified RNA targets.

Altogether, our data point to interlaced questions regarding SNORD13 sequence diversity across evolution, the molecular mechanisms underlying site-specific rRNA acetylation and the physiological importance of ac⁴C in rRNA in cell homeostasis, development and physiology. In our view, tackling these complex issues through an evolutionary perspective will be undoubtedly fruitful and may produce more surprises since, paradoxically not to mention ironically, observations made here with two widely-studied models - *C. elegans* and *D. melanogaster* - depart considerably from those previously obtained from vertebrates and budding yeasts.

Materials and Methods

Unless otherwise noted, all techniques for cloning and manipulating nucleic acids were performed according to standard protocols.

Generation and characterization of SNORD13-deficient HAP1 cells - HAP1 cells were grown at 37°C with 5% CO₂ in IMDM (Gibco; 4.5 g/L glucose) supplemented with 10% fetal bovine serum (PAN biotech), 1 mM Sodium Pyruvate (Gibco) and 1% penicillin-streptomycin (Sigma-Aldrich). The human SNORD13 gene was disrupted via CRISPR/Cas9-mediated deletion. Two sgRNAs selected using the CRISPR design tool at <https://zlab.bio/guide-design-resources> were cloned into pX459 V2.0 (Addgene #62988) plasmid. HAP1 cells seeded in six-well dishes (70-80% confluency) were co-transfected using lipofectamine 2000 (Thermo Fisher Scientific) with pX459_sgRNA1 and pX459_sgRNA2 (500 ng each). 48 hours post-transfection, cells were transiently treated with puromycin (2 µg/mL; ~ 36 hours) before being subjected to clonal selection by limiting dilution. Deletion events were validated by PCR (primer sequences are listed in **supplementary data S3**) and three clones per genotype were randomly chosen for further analyses. The rate of cell growth was assayed using Beckman Coulter Z1 particle counter. Cells were seeded into a 24-well plate and counted in quadruplicate at each time points. Six independent proliferation assays were performed. Apoptosis was detected using Annexin V-FITC/propidium iodide staining according to the manufacturer's instructions (BioLegend). Cells were analyzed using FacsVerse (Becton Dickinson) with acquisition of 20,000 total events (BD FacsSuite). Two independent experiments were performed.

RNA extraction and rRNA analyses - Total RNA was prepared using TRI reagent (MRC) according to the manufacturer's instructions and treated with RQ1 RNase-free DNase (Promega) and proteinase K (Sigma). For Northern blot analyses of low molecular-weight

species, total RNA (10 μg) was fractionated by electrophoresis on a 6% acrylamide, 7 M urea denaturing gel and electrotransferred onto a Hybond N membrane (Amersham GE Healthcare) as described previously (43). For Northern blot analyses of higher molecular-weight species, 4 μl of total RNA (2.5 $\mu\text{g}/\mu\text{l}$) were added to 20 μl of "glyoxal mix" (60% DMSO, BTPE 1.2X, 5% glycerol, 8% glyoxal solution (Sigma) and 40 $\mu\text{g}/\text{ml}$ Ethidium bromide), heat-denatured at 55°C for 60 min, chilled on ice for 10 min and separated by electrophoresis on a 1.2% agarose gel (Pipes 10 mM, Bis-tris 300 mM, EDTA 10 mM). Capillarity transfer was performed onto Hybond N+ membrane. In all cases, transfer was followed by UV light irradiation and membranes were pre-hybridized for 1 hour in 5x SSC, 0.1% SDS, 5x Denhardt's, 150 mg/ml yeast tRNA. Hybridization was carried out overnight at 50°C with 5'-[³²P]-labeled-DNA oligonucleotide probes (100 000 cpm/ml). Membranes were washed twice with 0.1XSSC, 0.1% SDS at room temperature (2x10 min) before autoradiography. Pulse chase experiments were performed as described in (44). The day before, 12-well plates are seeded with 500,000 cells. Cells are then incubated in methionine-free DMEM (Invitrogen) for 30 min at 37°C, labeled for exactly 15 min with L-methyl ³H methionine (50mCi/ml) and rinsed twice with IMDM containing unlabeled methionine. They were then incubated for 0 min (rinsed immediately twice with cold PBS), 30 min, 1h, 1h30, 2h, 2h30 in regular IMDM. After the corresponding chase time, cells were rinsed twice with cold PBS and lysed with Tri-Reagent (MRC). RNA samples were glyoxal denatured and separated on a 1.2% agarose gel and passively transferred to a nylon membrane as described above. The membrane was exposed 5 weeks to Biomax KODAK MS films with a KODAK BioMax Transcreen LE. RNase H mapping was performed as described in (44). A water solution containing 4 μg of total RNA was heat-denatured (95°C for 3 min) in the presence of an antisense oligonucleotide matching the 3'-end of 18S rRNA (RNase H_2, 100 μM). After allowing annealing by cooling down to room temperature, the RNA/oligonucleotide mixture was diluted to 30 μl final with a reaction mix (1X RNase H reaction buffer, 65 μM DTT, 0.5U/ μl RNasin (Promega), 50 U RNase H (New England Biolabs) and

incubated at 37°C for 30 min. The reaction was stopped by adding 0.3 M sodium acetate pH 5.2/0.2 mM EDTA. RNA was phenol extracted, ethanol precipitated, loaded onto a 15% acrylamide gel and analyzed by Northern blot (³²P-labeled 3'18S oligo-probe) as described above. For primer extension, two µg of total RNA extracted from adult flies were mixed with a 5'-³²P-labeled oligo (~100000 cpm) and cDNA synthesis, generated according to the manufacturer's instructions, were resolved onto a 12 % denaturing acrylamide gel.

Sucrose gradient sedimentation and Sunset assay - Exponentially growing WT and SNORD13-Ko cells (~ 20 million per genotype) were treated with 100 µg/ml of cycloheximide (CHX, SIGMA) for 10 min at 37°C, rinsed twice with cold PBS/CHX (100µg/ml), collected by centrifugation (400 g, 5 min) and resuspended in 5 ml of cold Buffer A (10 mM Hepes, 1,5 mM MgCl₂, 100mM KCl, CHX 100µg/ml, 1mM DTT). Cells were then centrifuged (400g, 10 min at 4°C), suspended in 700 µl of Buffer A and incubated 10 min on ice. Cells were then mechanically disrupted with a cold Dounce homogenizer (B) and centrifuged for 10 min (1000 g at 4°C). The cytoplasmic fraction (supernatant) was collected and centrifuged twice at 10 000 g (10 min, 4°C). About 1,5 mg of protein was loaded on a 10–50% sucrose gradient and centrifuged for 2h45min at 39 000 rpm in an Optima L-100XP Ultracentrifuge (Beckman-Coulter) using a SW41 rotor (Beckman). Gradients were then monitored at 260 nm and fractions were collected from the top using a Foxy Jr. fraction collector (Teledyne ISCO). Sunset assay was performed as described in (45). Exponentially growing cells were treated with 1 µg/ml of puromycin for 20 min. As a control, cycloheximide (100 µg/ml) was also added 15 min before puromycin treatment. Total cell extracts were then processed for western blot using an anti-puromycine antibody (Millipore, clone 12D10 #MABE343) as well as Anti-CDK9 (C12F7) mAb (cellsignal.com) as gel loading control.

Immunoprecipitations - Thirty microliters of rabbit R1131 anti-trimethylguanosine (3 mg/ml, a gift of Dr. R. Luhrmann) were incubated with gentle agitation for 120 min. at 4°C with 80µl of Proteine A Sepharose® 4B, Fast Flow from *S. aureus* (PAS, Sigma #P9424) in 1 ml of NET-150 buffer (50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 0.05% Igepal® CA-630, SIGMA). Mouse IgG (Jackson Immunoresearch #315-005-008) was used as negative controls. PAS-Ig pellets were washed three times with 1 ml of NET-150 buffer. Thirty micrograms of total RNAs was then added to PAS-R1131 (or PAS-IgG) in 0.5 ml of NET-150 and incubated with gentle agitation for 60 min. at 4 °C. Pellets were then collected by centrifugation and washed seven times in 1 ml of NET-150. RNAs from pellet and supernatant were extracted by SDS/phenol extraction and analyzed by Northern blot as described below.

Searching for SNORD13-like and 18S rRNA sequences in eukaryotic genomes - Using vertebrate SNORD13 sequences as queries, iterative BLAST searches with low stringency algorithm parameters, together with manual inspection of some hits, were conducted using eukaryotic genome databases found at <https://blast.ncbi.nlm.nih.gov/Blast.cgi>, <https://rnacentral.org/>, <https://www.echinobase.org/entry/>, <http://mgbase.qnlm.ac/home>, <http://www.insect-genome.com/>, <https://bipaa.genouest.org/is/>, <https://i5k.nal.usda.gov/webapp/blast/>, <https://spaces.facsci.ualberta.ca/ephybase/>. Eukaryotic 18S rRNA sequences were retrieved either at <https://www.arb-silva.de/> or <https://www.ncbi.nlm.nih.gov/nucleotide/>.

Mung bean nuclease protection assay and RP-HPLC - Mung bean nuclease (MBN) protection assay was performed exactly as described before (46). 500 pmol of the synthetic deoxy oligonucleotide (5'-TAATGATCCTTCCGCAGGTTACCTACGGAAACCTTGTTACGAC TTTTAC-3') was incubated with 50 pmol of 18S rRNA and 5% of DMSO in 0.3 volume of hybridization buffer (250 mM of HEPES, 500 mM of KCl at pH 7). The mixture was incubated at 90°C for 5 minutes and then slowly cooled down to 45°C over 2 h. After hybridization, 35 units

of mung bean nuclease (New England Bio Labs (NEB)) and 0.02 mg/ml RNase A (Sigma-Aldrich) along with appropriate amount of 10x MBN buffer (NEB) were added to start digestion. The digestion was carried out at 35°C for 1 h. The protected fragment (RNA-DNA hybrid) was extracted from the reaction mixture by phenol/chloroform extraction, followed by overnight ethanol precipitation. The protected rRNA fragment was separated from the complementary DNA oligonucleotides on a denaturing 7 M Urea 15% PAGE gel. Bands were visualized by ethidium bromide staining and the rRNA band was excised and eluted using the D-Tube™ Dialyzers according to the manufacturer's protocol for electro elution (Novagen). Eluted rRNA fragment was digested to nucleosides using P1 nuclease and alkaline phosphatase as described before (46). Nucleosides were next analyzed by RP-HPLC on a Supelcosil LC-18-S HPLC column (25 cm x 4.6 mm, 5 µm) equipped with a pre-column (4.6 x 20 mm) at 30°C on an Agilent 1200 HPLC system, using a protocol described previously in (16).

Targeted ac4C sequencing - Nucleotide resolution sequencing of ac4C in 18S rRNA (helix 45 and helix 34) was performed as previously described (6). RNA samples were treated with sodium cyanoborohydride (100 mM in H₂O) or vehicle (H₂O) in a final reaction volume of 100 µL. Reactions were initiated by addition of 1 M HCl to a final concentration of 100 mM and incubated for 20 minutes at room temperature. Reactions were stopped by neutralizing the pH by the addition of 30 µL 1 M Tris-HCl pH 8.0. After incubation, reactions were adjusted to 200 µL with H₂O, ethanol precipitated, desalted with 70% ice-cold ethanol, briefly dried on Speedvac, resuspended in H₂O, and quantified by using a Nanodrop 2000 spectrophotometer. Cellular total RNA from individual reactions (200 pg) was incubated with specific h45/h34 rev primer (4.0 pmol) in a final volume of 20 µL. Individual reactions were heated to 65 °C for 5 min and transferred to ice for 3 min to facilitate annealing in SuperScript III reaction buffer (Invitrogen). After annealing, reverse transcriptions were performed by adding 5 mM DTT, 200 units SuperScript III RT, 500 µM dNTPs (use 250 µM dGTP) and incubating for 60 min at 55 °C.

Reaction was quenched by increasing the temperature to 70 °C for 15 min and store at 4 °C. cDNA (2 µL) was used as template in 50 µL PCR reaction with Phusion Hot start flex (New England Biolabs). Reaction conditions: 1X supplied HF buffer, 2.5 pmole each forward and reverse primer, 200 µM each dNTPs, 2 units Phusion hot start enzyme, 2 µL cDNA template. PCR products were run on a 2% agarose gel, stained with SYBR safe and visualized on UV transilluminator at 302 nm. Bands of the desired size were excised from the gel and DNA extracted using QIA-quick gel extraction kit from Qiagen and submitted for Sanger sequencing (GeneWiz) using the forward PCR primer. Processed sequencing traces were viewed using 4Peaks software. Peak height for each base was measured and the percent misincorporation was determined using the equation: “Percent misincorporation = (peak intensity of T)/(sum of C and T base peaks)*100%”.

Dual luciferase reporter assay - 100,000 WT and SNORD13-KO cells were seeded into a 24-well plate for 24 hours. Cells were transfected using lipofectamine 2000 (Invitrogen) using 0.06 ng Renilla vector (internal control) and 500 ng of reporter Firefly vectors. 24 hours post-transfection, cells were harvested, lysed and luciferase activity was monitored using the Dual-Luciferase Reporter Assay kit (Promega) according to the manufacturer’s recommendations. Luciferase detection was assayed on Centro LB 960 Microplate Luminometer (Mikrowin 2000 Software). Five-Six independent transfections were performed with each measurement made in triplicate.

Yeast cells and media conditions - *S.cerevisiae* cells deleted of snR4 and snR45 or wild-type cells were grown at 30°C in standard YEP medium (1% yeast extract, 2% Bacto Peptone) supplemented with 2% dextrose (YPD). For Liquid growth assay, yeast cells were inoculated from a fresh YPD-agar plate into liquid YPD and grown for 48-72 hours at 30°C or 37°C. Optical density at 600 nm (OD₆₀₀) of cells was then measured and used to dilute cells into a 96 well

plate (Cat # 167008, Thermo Scientific) at equal ODs of ~0.02OD (n=3 or 12 for 30°C and 37°C samples, respectively). Cells were then grown for 90 hours in an Epoch 2 microplate spectrophotometer (BID EPOCH2 Microplate Spectrometer from BioTek) at 30°C and 37°C and OD was measured every 30 minutes. Measured OD values were analyzed in R². For presentation of growth curves, a regression line representing the mean of biological replicates was calculated using a generalized additive model. For Spot assay, yeast cells were inoculated from a fresh YPD-agar plate into liquid YPD and grown for 72 hours at 30°C and 37°C. Optical density at 600 nm (OD₆₀₀) was subsequently measured and cells were serially diluted in 2% YPD, such that OD of cells varied from 0.02 OD to 0.00016 OD (5 fold dilutions). 3 µl of each diluted sample was used to spot cells on a YPD-agar plate, which was transferred to 30°C and 37°C. After 48 hours of incubation pictures were taken of the colonies. Each strain was grown in duplicates.

Ethics statement and Zebrafish care - Fish were handled in a facility certified by the French Ministry of Agriculture (approval number A3155510). The project has received an agreement number APAFIS#7124-20161 00517263944 v3. Anesthesia and euthanasia procedures were performed in Tricaine Methanesulfonate (MS222) solutions as recommended for zebrafish (0.16 mg/ml for anesthesia, 0.30 mg/ml for euthanasia). All efforts were made to minimize the number of animals used and their suffering, in accordance with the guidelines from the European directive on the protection of animals used for scientific purposes (2010/63/UE) and the guiding principles from the French Decret 2013–118. Embryos were raised and staged according to standard protocols and the Recommended Guidelines for Zebrafish Husbandry Conditions (47,48)

Generation of Zebrafish snord13 mutants - The two guide RNAs (gRNA) were designed using CHOPCHOP CRISPR Design *website* (<http://chopchop.cbu.uib.no>). The designed oligos

were annealed and ligated into the gRNA plasmid pX459 digested by *BbsI* (Thermo Scientific). The gRNAs were generated using the MEGAscript T7 transcription kit (Ambion) with PCR fragments containing T7 promoter. Transcripts were purified by phenol-chloroform extraction and Ethanol precipitation. 1 nl of a solution containing 10 μ M EnGen Cas9 NLS (NEB) and 100 ng/ μ l of gRNAs was injected at the one-cell stage. WT, heterozygous, and homozygous SNORD13 animals were then identified by PCR. Sequence primers can be found in **(supplementary data S3)**.

RNA extraction, Reverse transcription and real-time PCR - Total RNAs from 15 wild type and 15 mutant zebrafish embryos were extracted using TRI reagent (MRC) and treated with RQ1-RNase free DNase I (Promega) and Proteinase K (Sigma) as recommended by manufacturer's instructions. Total RNAs were converted into cDNA using GoScript RTase (Promega) random hexamer primers for 60 min at 42 °C according to manufacturer's instructions. cDNAs were then diluted 20-fold and quantified by qPCR using SYBER green (Bio-rad) and specific primers. Data were acquired on CFX96 Real-Time PCR detection System (Bio-rad). Samples were analysed in triplicates and the expression level was calculated relative to zebrafish housekeeping gene *EF1 α* . Sequence primers can be found in **(supplementary data S3)**.

Immunostaining and *in situ* hybridization - Embryos were fixed overnight at 4°C in BT-FIX, after which they were immediately processed or dehydrated and stored at -20°C until use. After fixation or rehydration, embryos were washed twice with Phosphate Buffered Saline/1% Triton X-100 (PBST), permeabilized with PBST/0.5% Trypsin for 30 sec and washed twice again with PBST. After blocking with PBST/10% Fetal Calf Serum (FCS)/1% bovine serum albumin (BSA) (hereafter termed 'blocking solution') for at least 1 h, embryos were incubated with antibodies directed against either cleaved Caspase-3 (Asp175) (Cell Signaling Technology), or HuC/D

(Molecular Probes), in blocking solution overnight at 4 °C followed by 5 washing steps with PBST. Embryos were then incubated with the appropriate Alexa Fluor-conjugated secondary antibodies (Molecular Probes) for at least 2 h at room temperature and washed three times. Fluorescent in situ hybridization was carried out as previously described (49). *fli1a* riboprobe preparation has been described elsewhere (50). Embryos were dissected, flat-mounted in glycerol and images were recorded on the confocal microscope TCS SP8 (Leica Microsystems) with an L 25 × /0.95 W FLUOSTAR VIZIR objective (zoom X1.25) using the scanner resonant mode.

Comparison of Ribosome 3D structures - Atomic models of ribosomes from the following species: *H. sapiens*, PDB: 6EK0 (Natchiar et al, 2017); *P. falciparum*, pdb: 6OKK and 3J79 (51); *T. gondii*, PDB: 5XXU (40); *T. vaginalis*, PDB: 5XYI (40), *L. donovani*, PDB: 6AZ1, 6AZ3 (42); *E. gracilis*, PDB: 6ZJ3 (41) were displayed and aligned in UCSF Chimera (52) using the “Matchmaker” option, with the human 18S rRNA chain as reference. Prior to this structural alignment, cryo-EM maps of *P. falciparum* and *T. gondii* (EMD-2660 and EMD-6780, respectively) were visually inspected. Using UCSF Chimera and Coot (53), unattributed densities above C2065 in *P. falciparum* and C1764 in *T. gondii* cryo-EM map allowed to unambiguously replace unmodified residues by their ac⁴C counterparts in the corresponding atomic models.

Data availability statement - The data that support the findings of this study are available from the corresponding author [JC] upon reasonable request.

Disclosure declaration - The authors declare that they have no competing interests.

Acknowledgments

We thank the many colleagues who kindly provided us with biological samples, particularly EMBRC-France (N. Turque, R. Lasbleiz), the Botanical garden of Toulouse, Insectosphere, Botanic-Blagnac, A. Mattout, S.Tournier, I. Massou, R. Jeanson, L. Distefano, P.Valenti, A. Dussutour, M. Betermier, J. Santos, K. Mochizuki, G. Allorent and P.M. Delaux. We are also grateful to Aurore Laire for excellent zebrafish care. This work was supported by Agence Nationale de la Recherche (ANR-18- CE12-0008-01). Research in the lab of JLM was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research (ZIA-BC011488). Research in the lab of D.L.J.L. is supported by the Belgian Fonds de la Recherche Scientifique (F.R.S./FNRS), the Université Libre de Bruxelles (ULB), the European Joint Programme on Rare Diseases (EJP-RD 'RiboEurope' and 'DBAGencure'), the Région Wallonne (SPW EER 'RIBOncancer'), the Internationale Brachet Stiftung, and the Epitran COST action (CA16120).

References

1. Bruenger, E., Kowalak, J.A., Kuchino, Y., McCloskey, J.A., Mizushima, H., Stetter, K.O. and Crain, P.F. (1993) 5S rRNA modification in the hyperthermophilic archaea *Sulfolobus solfataricus* and *Pyrodicticum occultum*. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, **7**, 196-200.
2. Thomas, G., Gordon, J. and Rogg, H. (1978) N4-Acetylcytidine. A previously unidentified labile component of the small subunit of eukaryotic ribosomes. *J Biol Chem*, **253**, 1101-1105.
3. Ito, S., Akamatsu, Y., Noma, A., Kimura, S., Miyauchi, K., Ikeuchi, Y. and Suzuki, T. (2014) A single acetylation of 18 S rRNA is essential for biogenesis of the small ribosomal subunit in *Saccharomyces cerevisiae*. *J Biol Chem*, **289**, 26201-26212.
4. Ito, S., Horikawa, S., Suzuki, T., Kawauchi, H. and Tanaka, Y. (2014) Human NAT10 is an ATP-dependent RNA acetyltransferase responsible for N4-acetylcytidine formation in 18 S ribosomal RNA (rRNA). *J Biol Chem*, **289**, 35724-35730.
5. Sharma, S., Langhendries, J.L., Watzinger, P., Kotter, P., Entian, K.D. and Lafontaine, D.L. (2015) Yeast Kre33 and human NAT10 are conserved 18S rRNA cytosine acetyltransferases that modify tRNAs assisted by the adaptor Tan1/THUMP1. *Nucleic acids research*, **43**, 2242-2258.
6. Sas-Chen, A., Thomas, J.M., Matzov, D., Taoka, M., Nance, K.D., Nir, R., Bryson, K.M., Shachar, R., Liman, G.L.S., Burkhart, B.W. *et al.* (2020) Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping. *Nature*, **583**, 638-643.
7. Coureux, P.D., Lazennec-Schurdevin, C., Bourcier, S., Mechulam, Y. and Schmitt, E. (2020) Cryo-EM study of an archaeal 30S initiation complex gives insights into evolution of translation initiation. *Commun Biol*, **3**, 58.
8. Johansson, M.J. and Bystrom, A.S. (2004) The *Saccharomyces cerevisiae* TAN1 gene is required for N4-acetylcytidine formation in tRNA. *RNA*, **10**, 712-719.
9. Arango, D., Sturgill, D., Alhusaini, N., Dillman, A.A., Sweet, T.J., Hanson, G., Hosogane, M., Sinclair, W.R., Nanan, K.K., Mandler, M.D. *et al.* (2018) Acetylation of Cytidine in mRNA Promotes Translation Efficiency. *Cell*, **175**, 1872-1886 e1824.
10. Tardu, M., Jones, J.D., Kennedy, R.T., Lin, Q. and Koutmou, K.S. (2019) Identification and Quantification of Modified Nucleosides in *Saccharomyces cerevisiae* mRNAs. *ACS Chem Biol*, **14**, 1403-1409.
11. Liu, H.Y., Liu, Y.Y., Yang, F., Zhang, L., Zhang, F.L., Hu, X., Shao, Z.M. and Li, D.Q. (2020) Acetylation of MORC2 by NAT10 regulates cell-cycle checkpoint control and resistance to DNA-damaging chemotherapy and radiotherapy in breast cancer. *Nucleic acids research*, **48**, 3638-3656.
12. Liu, X., Tan, Y., Zhang, C., Zhang, Y., Zhang, L., Ren, P., Deng, H., Luo, J., Ke, Y. and Du, X. (2016) NAT10 regulates p53 activation through acetylating p53 at K120 and ubiquitinating Mdm2. *EMBO reports*, **17**, 349-366.
13. Shen, Q., Zheng, X., McNutt, M.A., Guang, L., Sun, Y., Wang, J., Gong, Y., Hou, L. and Zhang, B. (2009) NAT10, a nucleolar protein, localizes to the midbody and regulates cytokinesis and acetylation of microtubules. *Experimental cell research*, **315**, 1653-1667.

14. Balmus, G., Larrieu, D., Barros, A.C., Collins, C., Abrudan, M., Demir, M., Geisler, N.J., Lelliott, C.J., White, J.K., Karp, N.A. *et al.* (2018) Targeting of NAT10 enhances healthspan in a mouse model of human accelerated aging syndrome. *Nature communications*, **9**, 1700.
15. Larrieu, D., Britton, S., Demir, M., Rodriguez, R. and Jackson, S.P. (2014) Chemical inhibition of NAT10 corrects defects of laminopathic cells. *Science*, **344**, 527-532.
16. Sharma, S., Yang, J., van Nues, R., Watzinger, P., Kotter, P., Lafontaine, D.L.J., Granneman, S. and Entian, K.D. (2017) Specialized box C/D snoRNPs act as antisense guides to target RNA base acetylation. *PLoS genetics*, **13**, e1006804.
17. Watkins, N.J. and Bohnsack, M.T. (2012) The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA. *Wiley interdisciplinary reviews. RNA*, **3**, 397-414.
18. Bachellerie, J.P., Michot, B., Nicoloso, M., Balakin, A., Ni, J. and Fournier, M.J. (1995) Antisense snoRNAs: a family of nucleolar RNAs with long complementarities to rRNA. *Trends in biochemical sciences*, **20**, 261-264.
19. Kim, S.H., Spensley, M., Choi, S.K., Calixto, C.P., Pendle, A.F., Koroleva, O., Shaw, P.J. and Brown, J.W. (2010) Plant U13 orthologues and orphan snoRNAs identified by RNomics of RNA from Arabidopsis nucleoli. *Nucleic acids research*, **38**, 3054-3067.
20. Tyc, K. and Steitz, J.A. (1989) U3, U8 and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus. *EMBO J*, **8**, 3113-3119.
21. Essletzbichler, P., Konopka, T., Santoro, F., Chen, D., Gapp, B.V., Kralovics, R., Brummelkamp, T.R., Nijman, S.M. and Burckstummer, T. (2014) Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. *Genome Res*, **24**, 2059-2065.
22. Thomas, J.M., Briney, C.A., Nance, K.D., Lopez, J.E., Thorpe, A.L., Fox, S.D., Bortolin-Cavaille, M.L., Sas-Chen, A., Arango, D., Oberdoerffer, S. *et al.* (2018) A Chemical Signature for Cytidine Acetylation in RNA. *J Am Chem Soc*, **140**, 12667-12670.
23. Langhendries, J.L., Nicolas, E., Doumont, G., Goldman, S. and Lafontaine, D.L. (2016) The human box C/D snoRNAs U3 and U8 are required for pre-rRNA processing and tumorigenesis. *Oncotarget*, **7**, 59519-59534.
24. Huang, Z.P., Zhou, H., He, H.L., Chen, C.L., Liang, D. and Qu, L.H. (2005) Genome-wide analyses of two families of snoRNA genes from *Drosophila melanogaster*, demonstrating the extensive utilization of introns for coding of snoRNAs. *RNA*, **11**, 1303-1316.
25. Li, D., Wang, Y., Zhang, K., Jiao, Z., Zhu, X., Skogerboe, G., Guo, X., Chinnusamy, V., Bi, L., Huang, Y. *et al.* (2011) Experimental RNomics and genomic comparative analysis reveal a large group of species-specific small non-message RNAs in the silkworm *Bombyx mori*. *Nucleic acids research*, **39**, 3792-3805.
26. Kiss, T. and Filipowicz, W. (1995) Exonucleolytic processing of small nucleolar RNAs from pre-mRNA introns. *Genes & development*, **9**, 1411-1424.
27. Smith, E.R., Lin, C., Garrett, A.S., Thornton, J., Mohaghegh, N., Hu, D., Jackson, J., Saraf, A., Swanson, S.K., Seidel, C. *et al.* (2011) The little elongation complex regulates small nuclear RNA transcription. *Molecular cell*, **44**, 954-965.
28. Dubeau, F. and Felix, M.A. (2012) Role of pleiotropy in the evolution of a cryptic developmental variation in *Caenorhabditis elegans*. *PLoS biology*, **10**, e1001230.

29. Dudnakova, T., Dunn-Davies, H., Peters, R. and Tollervey, D. (2018) Mapping targets for small nucleolar RNAs in yeast. *Wellcome open research*, **3**, 120.
30. Dewe, J.M., Whipple, J.M., Chernyakov, I., Jaramillo, L.N. and Phizicky, E.M. (2012) The yeast rapid tRNA decay pathway competes with elongation factor 1A for substrate tRNAs and acts on tRNAs lacking one or more of several modifications. *RNA*, **18**, 1886-1896.
31. Ding, F., Li, H.H., Zhang, S., Solomon, N.M., Camper, S.A., Cohen, P. and Francke, U. (2008) SnoRNA Snord116 (Pwcr1/MBII-85) deletion causes growth deficiency and hyperphagia in mice. *PloS one*, **3**, e1709.
32. Hebras, J., Marty, V., Personnaz, J., Mercier, P., Krogh, N., Nielsen, H., Aguirrebengoa, M., Seitz, H., Pradere, J.P., Guiard, B.P. *et al.* (2020) Reassessment of the involvement of Snord115 in the serotonin 2c receptor pathway in a genetically relevant mouse model. *eLife*, **9**.
33. Lee, J., Harris, A.N., Holley, C.L., Mahadevan, J., Pyles, K.D., Lavagnino, Z., Scherrer, D.E., Fujiwara, H., Sidhu, R., Zhang, J. *et al.* (2016) Rpl13a small nucleolar RNAs regulate systemic glucose metabolism. *The Journal of clinical investigation*, **126**, 4616-4625.
34. Skryabin, B.V., Gubar, L.V., Seeger, B., Pfeiffer, J., Handel, S., Robeck, T., Karpova, E., Rozhdestvensky, T.S. and Brosius, J. (2007) Deletion of the MBII-85 snoRNA gene cluster in mice results in postnatal growth retardation. *PLoS genetics*, **3**, e235.
35. Soeno, Y., Fujita, K., Kudo, T., Asagiri, M., Kakuta, S., Taya, Y., Shimazu, Y., Sato, K., Tanaka-Fujita, R., Kubo, S. *et al.* (2013) Generation of a mouse model with down-regulated U50 snoRNA (SNORD50) expression and its organ-specific phenotypic modulation. *PloS one*, **8**, e72105.
36. Ben-Shem, A., Garreau de Loubresse, N., Melnikov, S., Jenner, L., Yusupova, G. and Yusupov, M. (2011) The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science*, **334**, 1524-1529.
37. Natchiar, S.K., Myasnikov, A.G., Kratzat, H., Hazemann, I. and Klaholz, B.P. (2017) Visualization of chemical modifications in the human 80S ribosome structure. *Nature*, **551**, 472-477.
38. Sharma, S. and Lafontaine, D.L.J. (2015) 'View From A Bridge': A New Perspective on Eukaryotic rRNA Base Modification. *Trends in biochemical sciences*, **40**, 560-575.
39. Tamm, T., Kisly, I. and Remme, J. (2019) Functional Interactions of Ribosomal Intersubunit Bridges in *Saccharomyces cerevisiae*. *Genetics*, **213**, 1329-1339.
40. Li, Z., Guo, Q., Zheng, L., Ji, Y., Xie, Y.T., Lai, D.H., Lun, Z.R., Suo, X. and Gao, N. (2017) Cryo-EM structures of the 80S ribosomes from human parasites *Trichomonas vaginalis* and *Toxoplasma gondii*. *Cell research*, **27**, 1275-1288.
41. Matzov, D., Taoka, M., Nobe, Y., Yamauchi, Y., Halfon, Y., Asis, N., Zimmermann, E., Rozenberg, H., Bashan, A., Bhushan, S. *et al.* (2020) Cryo-EM structure of the highly atypical cytoplasmic ribosome of *Euglena gracilis*. *Nucleic acids research*, **48**, 11750-11761.
42. Shalev-Benami, M., Zhang, Y., Rozenberg, H., Nobe, Y., Taoka, M., Matzov, D., Zimmerman, E., Bashan, A., Isobe, T., Jaffe, C.L. *et al.* (2017) Atomic resolution snapshot of *Leishmania* ribosome inhibition by the aminoglycoside paromomycin. *Nature communications*, **8**, 1589.

43. Cavaille, J. and Bachellerie, J.P. (1998) SnoRNA-guided ribose methylation of rRNA: structural features of the guide RNA duplex influencing the extent of the reaction. *Nucleic acids research*, **26**, 1576-1587.
44. Montellese, C., Montel-Lehry, N., Henras, A.K., Kutay, U., Gleizes, P.E. and O'Donohue, M.F. (2017) Poly(A)-specific ribonuclease is a nuclear ribosome biogenesis factor involved in human 18S rRNA maturation. *Nucleic acids research*, **45**, 6822-6836.
45. Martineau, Y., Azar, R., Muller, D., Lasfargues, C., El Khawand, S., Anesia, R., Pelletier, J., Bousquet, C. and Pyronnet, S. (2014) Pancreatic tumours escape from translational control through 4E-BP1 loss. *Oncogene*, **33**, 1367-1374.
46. Yang, J., Sharma, S., Watzinger, P., Hartmann, J.D., Kotter, P. and Entian, K.D. (2016) Mapping of Complete Set of Ribose and Base Modifications of Yeast rRNA by RP-HPLC and Mung Bean Nuclease Assay. *PLoS one*, **11**, e0168873.
47. Alestrom, P., D'Angelo, L., Midtlyng, P.J., Schorderet, D.F., Schulte-Merker, S., Sohm, F. and Warner, S. (2020) Zebrafish: Housing and husbandry recommendations. *Laboratory animals*, **54**, 213-224.
48. Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B. and Schilling, T.F. (1995) Stages of embryonic development of the zebrafish. *Developmental dynamics : an official publication of the American Association of Anatomists*, **203**, 253-310.
49. Quillien, A., Moore, J.C., Shin, M., Siekmann, A.F., Smith, T., Pan, L., Moens, C.B., Parsons, M.J. and Lawson, N.D. (2014) Distinct Notch signaling outputs pattern the developing arterial system. *Development*, **141**, 1544-1552.
50. Lawson, N.D., Scheer, N., Pham, V.N., Kim, C.H., Chitnis, A.B., Campos-Ortega, J.A. and Weinstein, B.M. (2001) Notch signaling is required for arterial-venous differentiation during embryonic vascular development. *Development*, **128**, 3675-3683.
51. Wong, W., Bai, X.C., Brown, A., Fernandez, I.S., Hanssen, E., Condrón, M., Tan, Y.H., Baum, J. and Scheres, S.H. (2014) Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. *eLife*, **3**.
52. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry*, **25**, 1605-1612.
53. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography*, **66**, 486-501.

Figure - Legends

Figure 1 - Schematic of SNORD13-dependent ribosomal RNA acetylation. A) Through an incompletely defined antisense-mediated mechanism, human SNORD13 assists the RNA cytidine acetyltransferase NAT10 in modifying C1842 in the universally conserved helix 45 at the 3' extremity of 18S rRNA. **B)** Helix 45 is enriched with post-transcriptional RNA modifications. RNA modifying enzymes and their modified nucleotides (lollipops) are indicated: SSU-ac⁴C1842 (yellow), SSU-m⁶A1832 (green) and SSU-m₂⁶A1850 and SSU-m₂⁶A1851 (orange). **C)** Human SNORD13 base-pairs with 18S rRNA sequences on each side of the cytidine to be acetylated. Note that SNORD13-rRNA duplexes tolerate G-U wobble base-pairs, mismatches or even bulged nucleotides and the length of the intervening loops, from either rRNA or SNORD side, also differ from one organism to another (**Supplementary figure S1**).

Figure 2 - The specific lack of SNORD13-mediated acetylation in helix 45 does not alter cell growth or translation. A) Schematic representation of SNORD13 locus at human chromosome 8. The relative location of the two DNA sequences targeted by sgRNAs is shown (PAM sequence is underlined). **B)** Northern blot showing that SNORD13 is no longer detected in clones bearing deletions. * indicate truncated SNORD13 form which is routinely detected and corresponding to RNA species whose 5'-end is positioned 5-6 nucleotides upstream of the C-box (not shown). **C)** RP-HPLC chromatograms of nucleosides obtained from helix 45 of WT (black) and SNORD13 (red) cells. Peak corresponding to SSU-ac⁴c1842 is indicated. **D)** Sanger DNA sequencing of RT-PCR products obtained after borohydride treatments of total RNA extracted from WT and SNORD13-KO cells. % of misincorporation (C-to-U) at SSU-C1842 (helix 45) and SSU-C1337 (helix 34) is indicated below each electropherogram. **E)** Growth curves of WT and SNORD13-KO cells as judged by cell counting (n=6 for each genotype). **F)** Apoptosis levels determined by flow cytometry using propidium and Annexin V

staining of WT and SNORD13-KO cells (n=2 for each genotype). **G**) SUnSET assay. Incorporation of puromycin into the elongating peptide was assayed in WT and SNORD13-KO cells (western blotting with anti-puromycin antibodies). Ponceau S staining of the membrane and detection of CDK9 (western blotting) were used as gel loading controls. Cells treated with the translation inhibitor cycloheximide (CHX) were also used as negative controls. **H**) Luciferase reporter gene assays. WT and SNORD13-KO cells were transiently co-transfected by two plasmids expressing Renilla luciferase gene (LucR; internal control) or Firefly luciferase gene (lucF; translation reporter) carrying either an in-frame stop-codons (UAA, UAG), a detrimental mutation (R-to-S) or an IRES from FGF1A, FGF2 and CMV. Histograms show the normalized luciferase activity (lucF-to-lucR ratio). Data are expressed as mean +/- s.e.m. and represent 5-6 independent experiments with triplicate measurements.

Figure 3 - Phenotypic analyses of SNORD13-deficient zebrafish embryos. **A)** Schematic representation of the zebrafish *snord13* locus. The relative location of the two DNA sequences targeted by sgRNAs is shown (PAM sequence is underlined). **B)** Northern blotting showing that SNORD13 is no longer detected in embryos bearing deletion events. *: truncated SNORD13 form which is routinely detected. SNORD118 was used as a gel loading control. **C)** Histograms show percentage of misincorporation (C-to-U ratio) measured in borohydride-treated RNA samples prepared from WT and SNORD13-KO embryos (n= 3). **D-G** Confocal projections of WT (**D, F**) and SNORD13-KO (**E, G**) embryos after immunostaining against cleaved Caspase-3 (red) and the pan neural marker HuC/D (green). **D-E:** Dorsal view of the zebrafish brain at 30 hpf, neurons from the olfactory epithelium (oe), tectum (t), pineal gland (pg), optic tectum (ot) and trigeminal ganglion (tg) can be detected. **F-G:** Lateral view of the spinal cord (sc) at 30 hpf. **H-I** Confocal projections of WT (**H**) and SNORD13-KO (**I**) embryos after *in situ* RNA hybridization using antisense riboprobes directed against *fli1a* transcripts. White brackets indicate the dorsal aorta (DA) and the cardinal vein (CV) and the white arrows point to

intersegmental vessels. Pictures are representative of three independent experiments (at least 6 embryos imaged per genotype).

Figure 4 - Cross-evolutionary survey of eukaryotic rRNA acetylation machinery. **A)** This table summarizes the conservation of the 5'-CCG-3' motif, the acetylation status of helix 45 and the detection of SNORD13 and NAT10 genes in commonly used models or organisms for which we experimentally assayed the presence of ac⁴C at helix 45. n.d.: not determined. n/a: not applicable (genome is not available). **B)** Simplified phylogenetic tree of Metazoan. The genomic organization (intronic vs intergenic) of newly-identified SNORD13 genes (red arrows) is shown. Note that intronic SNORD13 can be positioned in the sense or antisense orientation with respect to their host-genes (black arrow). **C)** Simplified phylogenetic tree of Archaeplastida. Acetylation status of helix 45 and the presence of SNORD13 in the genome of representative species are indicated. Note that while rRNA in algal species is acetylated, we failed to identify obvious SNORD13 counterparts. A full list of newly-identified SNORD13 sequences can be found in **Supplementary data S1**.

Figure 5 - Atypical SNORD13-related RNA guides rRNA acetylation in *D. melanogaster*. **A)** Simplified phylogenetic tree of Arthropoda. Representative species where SNORD13-like were identified are indicated. **B)** Simplified phylogenetic tree of Diptera. The genomic arrangements of newly-identified SNORD13 genes (blue arrow) are shown. **C)** Multiple sequence alignment of representative eukaryotic SNORD13 sequences. The small red arrow indicates the relative position of the mature 5'-end of Or-CD1 in Diptera (see also panel F) while the two horizontal black arrows in opposite orientation depict the terminal 5'-3' stem structure. Note that SNORD13 sequences in Diptera - but not in other species - lack one of the two conserved antisense rRNA elements (denoted by horizontal blue bars). Vertical black, grey and white bars highlight Diptera, other Arthropoda and non-Arthropoda species, respectively. **D)** Immunoprecipitation by anti-

trimethyl cap (R1131) antibodies. OR-CD1 in various insects (as indicated above the panels) was detected by Northern blots. The same membrane was hybridized with an antisense oligo probes that recognized uncapped 5.8S rRNA used as negative controls. Input RNA (I); RNA recovered from the pellet (P), RNA recovered from the supernatant (S). The theoretical size of SNORD13 (nt) in each of the species studied is indicated in parentheses. **E)** Schematic representation of two distinct modes of expression: independent transcription gives rise to short (top) or long (bottom) forms of 5' capped SNORD13. **F)** A mutant KO fly strain harboring an inserted P-element in the Or-CD1 gene (*Dmel*\{EP}snoRNA:Or-CD1G9117) does not express OR-CD1 as assayed by primer extension. Note that the 5'-end of the cDNA product confirms that fly Or-CD1 lacks one 18S rRNA complementarity. The expected size (nt) of OR-CD1 with ~4-5 nucleotides upstream of the C-box (as shown on the panel) is indicated in parentheses. ³²P labeled primer (P). **G)** Misincorporation (C-to-U) at SSU-C1968 (helix 45) as judged by Sanger DNA sequencing of RT-PCR products obtained after borohydride treatments of total RNA extracted from WT and Or-CD1-KO adult flies.

Figure 6 – Life without rRNA acetylation **A)** Simplified phylogenetic tree of Nematoda. Sequence contexts of the acetylated cytidine in representative species of each clade are indicated. **B)** Schematic representations of helix 45. Green bases represent compensatory base changes, as compared to human helix 45, while red bases provoke irregularity in the geometry of helix 45. The triplet sequence motif is framed. **C)** Sanger DNA sequencing of RT-PCR products obtained after borohydride treatments of total RNA extracted from *C. elegans*. **D)** Ac⁴C content in total RNA extracted from adult *C. elegans* was visualized by Northern blotting-based assay using anti-Ac⁴C antibodies. Ethidium bromide (etBr) staining indicates the relative position of rRNA and tRNA species. **E)** Top-panels: WT and Δ snR4 Δ snR45 yeast strains were grown at 30°C (left) or 37°C (right) in liquid YPD. Optical density at 600 nm (OD₆₀₀) was measured every 30 minutes. Growth curves of individual biological samples (thin lines; n=3 or

12 for 30 °C and 37 °C samples, respectively) and the mean of replicates (thick lines) are presented along with a confidence interval (gray area). Bottom-panels: WT and Δ snR4 Δ snR45 yeast strains were spotted on a YPD-agar plate and incubated for 48 hours at 30°C or 37°C. Cells were grown in duplicates with a 5-fold dilution between distinct concentrations.

Figure 7 – Comparative schematic representation of rRNA modification systems. A)

Through the formation of a perfect (or near perfect) RNA duplex (~10-20 bp), SNORDs guide Fibrillarin/Nop1 to their target rRNA nucleotides, i.e. the nucleotide to be 2'-O-methylated (2'-O-Me) is paired to the fifth position upstream of the conserved D (or D') box. **B)** Through the formation of two perfect (or near perfect) RNA duplexes (~4-10 bp each), SNORAs guide Dyskerin/Cbf5 to their target rRNA nucleotides, i.e. the uridine to be isomerized into pseudouridine (ψ) remains unpaired and it is located at ~ 14-16 nt from the conserved H (or ACA) box. **C)** snR45, snR4 and SNORD13 contain two imperfect antisense elements matching either side of the cytidine to be acetylated (ac^4C) and, through still unknown mechanisms, guide cytosine acetylation by NAT10/Kre33. **D)** In *D. melanogaster*, SNORD13 targets rRNA acetylation at helix 45 through the use of a single antisense rRNA sequence.

Supplementary figure S1 - Examples of SNORD13-rRNA duplexes across evolution.

Schematic representations of base-pairing interactions between SNORD13 and 18S rRNA (only a few representative examples are shown). The targeted C for acetylation is written in green.

Supplementary figure S2 - Loss of SNORD13 does not affect pre-rRNA synthesis and processing. A-B)

Northern blot - Total RNA prepared from WT and SNORD13-KO clones, as well as from a pool of the three WT and KO clones, were analysed by Northern blot using probe 1 (5-ITS1) and probe 2 (ITS2) that detect main rRNA precursors intermediates as depicted in panel B). Cytoplasmic 18S and 28S rRNAs were also visualized by ethidium bromide staining of

a denaturing 1.2% agarose gel. Relative quantification of rRNA intermediates based on three independent Northern blots did not reveal any obvious abnormalities (not shown). **C)** RNase H assay - Total RNA extracted from WT or SNORD13-KO cells were annealed with antisense DNA oligonucleotide (black line), RNase H digestion products were then resolved onto a 12% acrylamide gels and visualized by Northern blot with an antisense oligo-probe matching the 3'-end of 18S (red line). The ladder-like pattern reflecting heterogeneity of the 3'-end at 18S in SNORD13-KO cells was indistinguishable from that of observed in WT cells, indicating that the accuracy of cleavages that delineates the 3' extremity of 18S rRNA is not impaired in SNORD13-KO cells. **D)** Pulse chase experiments - WT and SNORD13-KO cells were pulse-labelled with L-[methyl-³H]-methionine and harvested by chase times as indicated above the panel. Freshly-synthesized rRNA intermediates were detected by autoradiography after being separated by electrophoresis on denaturing 1.2% gel and transferred to a nylon membrane. Dynamics of rRNA synthesis in the SNORD13-KO cells was in the normal range, as shown by the timing of appearance and lifespan of the detected methyl-³H labelled rRNA precursors. **E)** Polysome profiling - Whole cytoplasmic extracts prepared from WT and (left) SNORD13-KO (right) cells were fractionated by ultracentrifugation on a 10-50% sucrose gradient. Absorbance (254 nm) of each fraction was recorded and the peaks corresponding to free 40S and 60S subunits, 80S ribosomes (monosomes) and polysomes are indicated.

Supplementary figure S3 - Deletion of *snord13* gene does not perturb zebrafish development. **A)** Bright field images of WT (up) and homozygous (down) *snord13* mutant embryos derived from the crossing of homozygous *snord13* females with heterozygous males. Pictures are representative of three independent crosses. **B)** Relative mRNA expressions determined by RT-qPCR in 30 hpf WT and SNORD13-KO embryos (3 independent experiments with at least 15 animals per condition). The tested genes are listed below the histogram (black and red histograms correspond to WT and KO, respectively).

Supplementary figure S4 – Phylogenetic conservation of helix 45 across evolution.

Sequence **(A)** and schematic representations **(B)** of helix 45 in representative species belonging to each eukaryotic supergroup. Green bases represent compensatory base changes, as compared to human helix 45, while red bases alter the geometry of helix 45. The triplet sequence motif is framed.

Supplementary figure S5 – Probing rRNA acetylation in *T. thermophila*, *S. carpocapsae*

and *P. polycephalum*. Sanger DNA sequencing of RT-PCR products obtained after borohydride treatments of total RNA extracted from *P. polycephalum* (left) *T. thermophila* (middle), *S. carpocapsae* (right). % of misincorporation at helix 45 is indicated below each electropherogram.

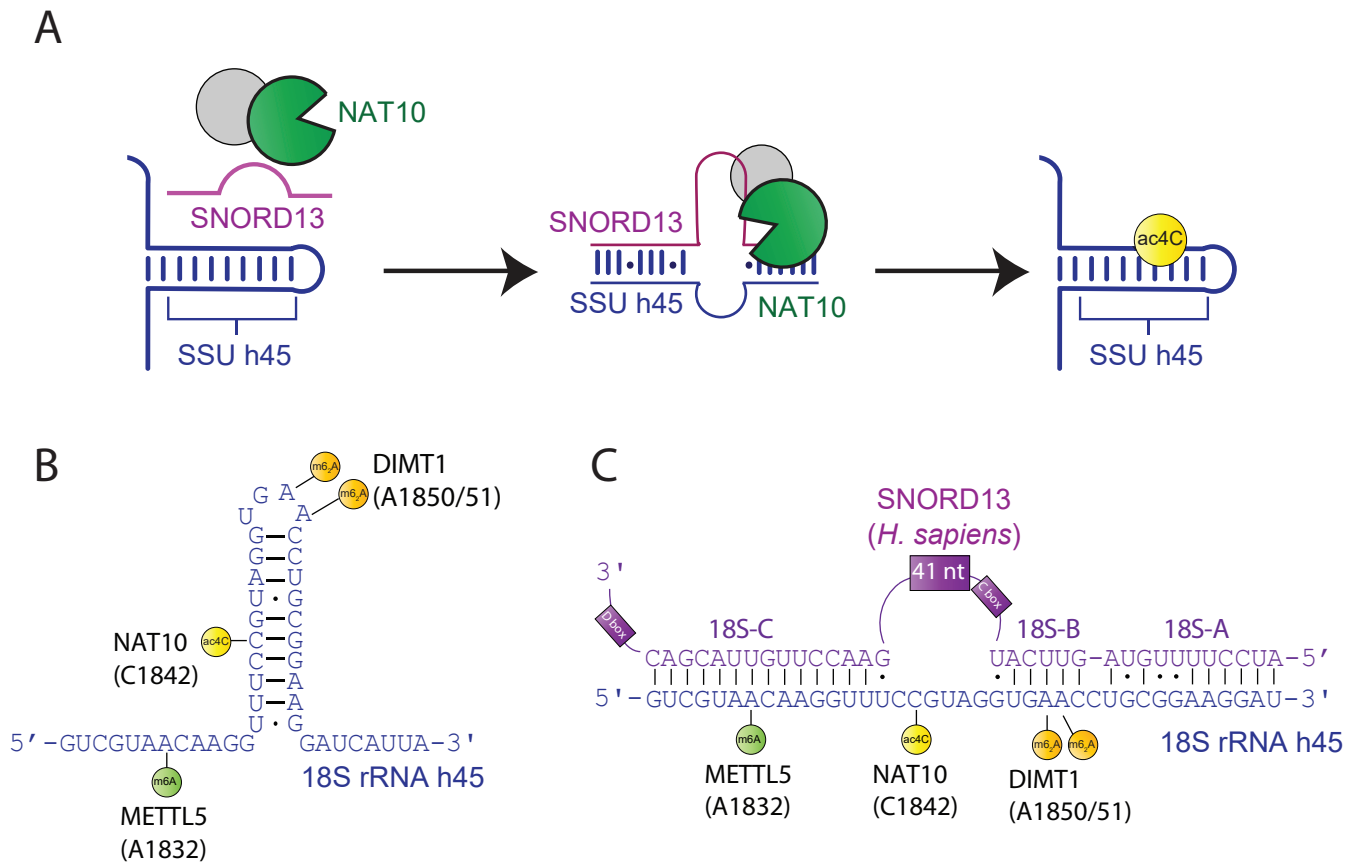
Supplementary figure S6 – rRNA h45 acetylation does not seem to affect Rpl41

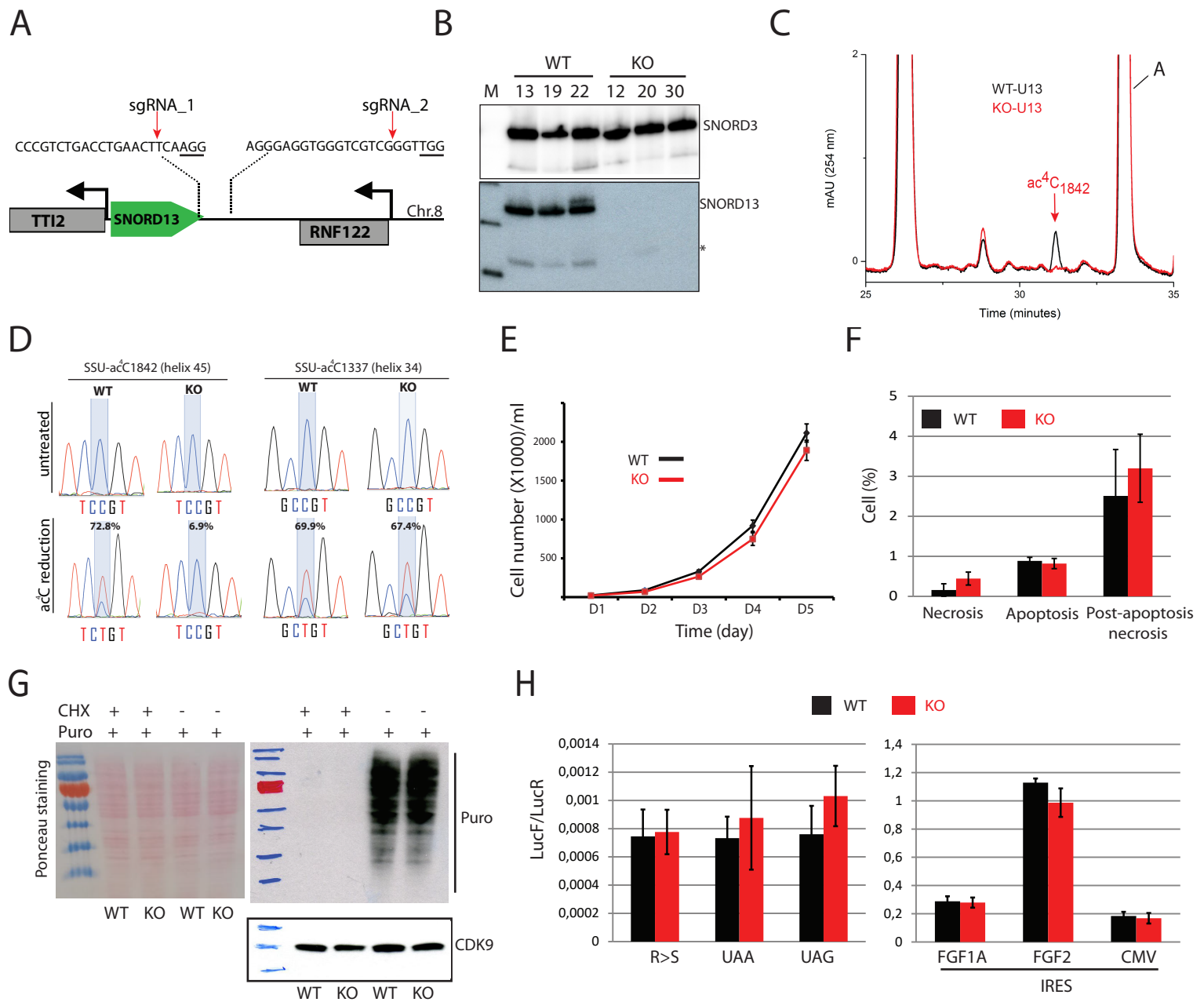
positioning. Spatial alignment of Ribosomal 3D structures of various species, carrying either acetylated C (left panel) or unmodified U or G (right panel) on rRNA helix 45 (h45) revealed no change in the location of ribosomal protein Rpl41 (eL41). For each aligned species (indicated on the top left of each vignette), the modified/unmodified nucleotide of interest is shown both on the structure as well as on the 3-nucleotide sequence by an underlined character. For spatial cueing, lateral chain of an N-terminal Arginine of Rpl41, one of the closest aminoacids of the nucleotide of interest, is fully displayed.

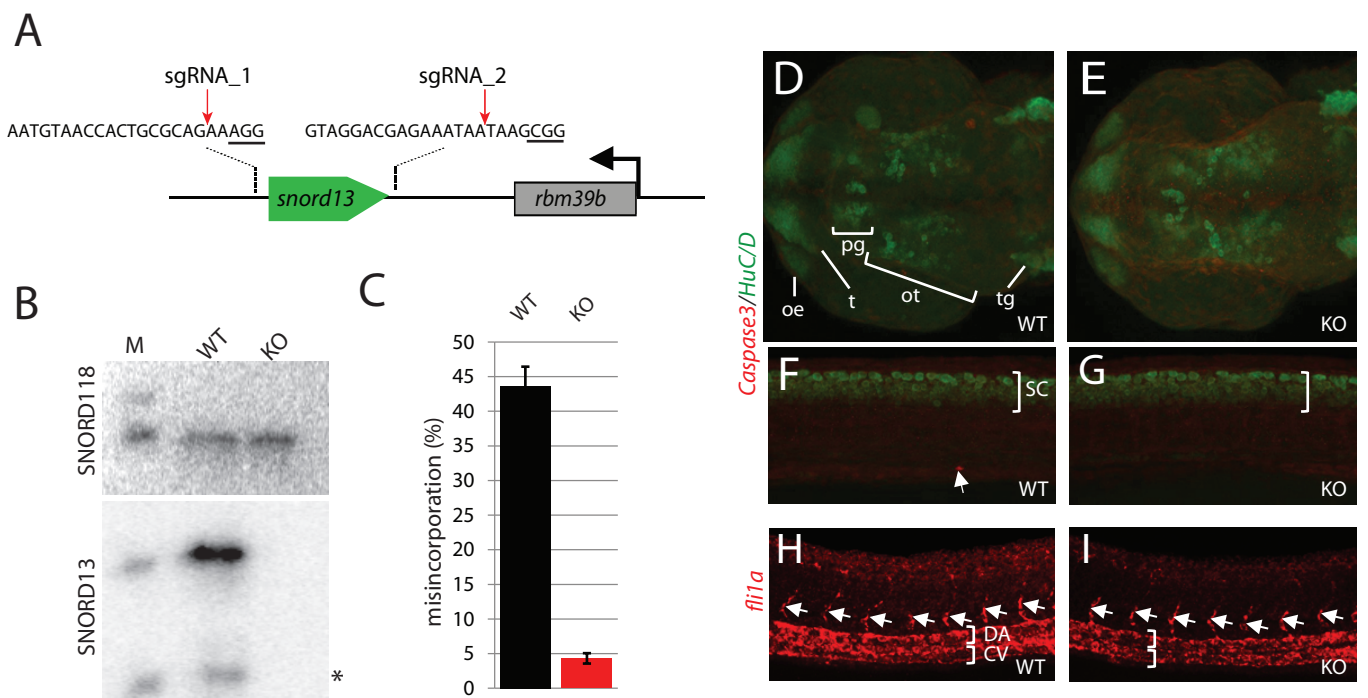
Supplementary data S1 – Examples of newly-identified SNORD13 sequences.

Supplementary data S2 – The primary structure of NAT10 across evolution

Supplementary data S3 – Sequences of primers



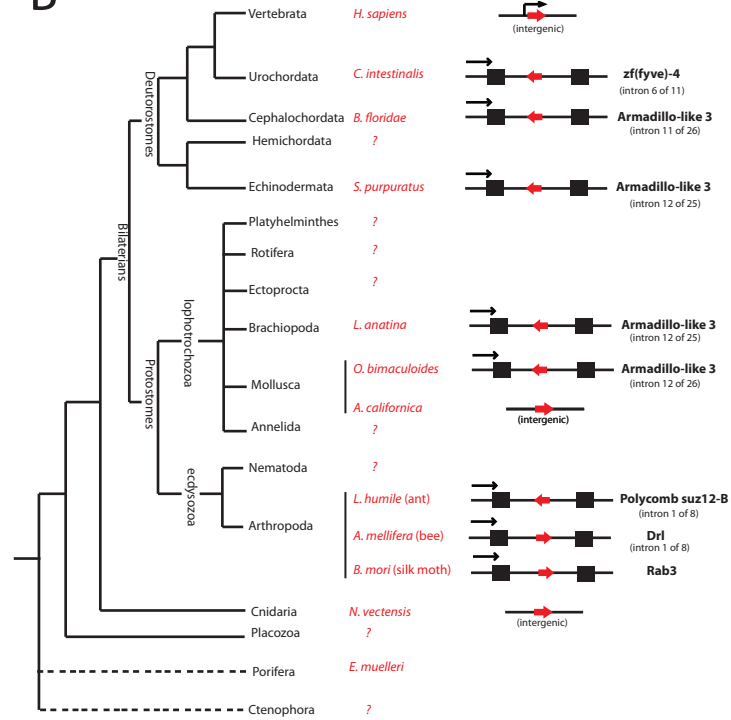




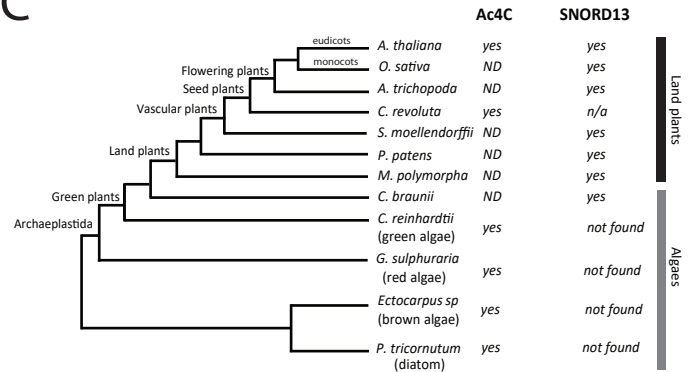
A

| Organisms | Helix 45 | Ac4C | SNORD | NAT10 |
|---|----------|----------------------------|-----------------------------------|-----------|
| <i>H. sapiens</i> | CCG | yes (Sharma et al 2015) | SNORD13 (U13) | yes |
| <i>M. musculus</i> (mouse) | CCG | n.d. | SNORD13-like (this study) | yes |
| <i>R. norvegicus</i> (rat) | CCG | yes (Thomas et al 1978) | SNORD13-like (this study) | yes |
| <i>G. gallus</i> (chicken) | CCG | yes (Thomas et al 1978) | GGN86 (Makarova et al 2011) | yes |
| <i>X. laevis</i> (frog) | CCG | n.d. | SNORD13-like (this study) | yes |
| <i>D. rerio</i> (zebrafish) | CCG | yes (this study) | SNORD13-like (this study) | yes |
| <i>P. marinus</i> (lamprey) | CCG | n.d. | SNORD13-like (this study) | yes |
| <i>C. intestinalis</i> (sea squirt) | CCG | n.d. | SNORD13-like (this study) | yes |
| <i>B. floridae</i> (Florida lancelet) | CCG | n.d. | SNORD13-like (this study) | yes |
| <i>S. purpuratus</i> (purple sea urchin) | CCG | n.d. | SNORD13-like (this study) | yes |
| <i>H. attenuata</i> (star fish) | CCG | yes (this study) | n/a | n/a |
| <i>A. mellifera</i> (bee) | CCG | yes (this study) | SNORD13-like (this study) | yes |
| <i>A. gambiae</i> (mosquito) | CCG | n.d. | not found | yes |
| <i>D. melanogaster</i> (fruit fly) | CCG | yes (this study) | OR-CD1 (this study) | yes |
| <i>B. mori</i> (domestic silk moth) | CCG | n.d. | Bm_10 (this study) | yes |
| <i>A. labyrinthica</i> (spider) | CCG | yes (this study) | n/a | n/a |
| <i>C. scutellaris</i> (ant) | CCG | yes (this study) | n/a | n/a |
| <i>I. scapularis</i> (deer tick) | CCG | n.d. | SNORD13-like (this study) | yes |
| <i>C. elegans</i> (nematode) | CUG | no (this study) | not found | yes |
| <i>S. carpocapsae</i> (nematode) | CCG | yes (this study) | not found | yes |
| <i>H. aspera</i> (snail, mollusk) | CCG | yes (this study) | n/a | n/a |
| <i>M. galloprovincialis</i> (mussel, mollusk) | CCG | yes (this study) | SNORD13-like (this study) | yes |
| <i>A. californica</i> (aplysia, mollusk) | CCG | n.d. | SNORD13-like (this study) | yes |
| <i>L. anatina</i> (brachiopoda) | CCG | n.d. | SNORD13-like (this study) | yes |
| <i>O. bimaculoides</i> (octopus) | CCG | n.d. | SNORD13-like (this study) | yes |
| <i>A. viridis</i> (anemona, cnidaria) | CCG | yes (this study) | n/a | n/a |
| <i>C. hemisphaerica</i> (cnidaria) | CCG | yes (this study) | SNORD13-like (this study) | yes |
| <i>P. gracilis</i> (cnidaria) | CCG | yes (this study) | n/a | n/a |
| <i>N. vectensis</i> (Starlet sea anemone, cnidaria) | CCG | n.d. | SNORD13-like (this study) | yes |
| <i>D. gigantea</i> (coral, cnidaria) | CCG | n.d. | SNORD13-like (this study) | yes |
| <i>D. australiensis</i> (sponge) | CCG | yes (this study) | n/a | n/a |
| <i>Haliclona</i> sp (sponge) | CCG | yes (this study) | n/a | n/a |
| <i>E. muelleri</i> (sponge) | CCG | n.d. | SNORD13-like (this study) | yes |
| <i>U. maydis</i> (corn smut fungus) | CUG | no | not found | yes |
| <i>A. bisporus</i> (mushroom from Paris) | CCG | yes (this study) | not found | yes |
| <i>S. pombe</i> (fission yeast) | CCG | yes (Taoka et al 2014) | not found | yes |
| <i>N. crassa</i> (filamentous fungus) | CCG | n.d. | not found | yes |
| <i>P. ansarina</i> (filamentous fungus) | CCG | n.d. | not found | yes |
| <i>A. niger</i> (filamentous fungus) | CCG | n.d. | not found | yes |
| <i>S. cerevisiae</i> (budding yeast) | CCG | yes (Ito et al 2014) | snR45 (Sharma et al 2017) | yes |
| <i>D. discoideum</i> (social amoeba) | CCG | yes (McCarroll et al 1983) | not found | yes |
| <i>P. polycephalum</i> (slime mod) | UCG | yes (this study) | not found | yes |
| <i>A. thaliana</i> (eudicot) | CCG | yes (Sharma et al 2015) | snoR105/108/2279 (Kim et al 2010) | yes |
| <i>O. sativa</i> (monocot) | CCG | n.d. | SNORD13-like (this study) | yes |
| <i>A. trichopoda</i> (ancestor of angiosperms) | CCG | n.d. | SNORD13-like (this study) | yes |
| <i>N. colorata</i> (water lily) | CCG | n.d. | SNORD13-like (this study) | yes |
| <i>N. revoluta</i> (Japanese sago palm) | CCG | yes (this study) | n/a | n/a |
| <i>S. moellendorffii</i> (lycophyte) | CCG | n.d. | SNORD13-like (this study) | yes |
| <i>P. patens</i> (moss) | CCG | n.d. | SNORD13-like (this study) | yes |
| <i>H. cupressiforme</i> (moss) | CCG | yes (this study) | n/a | n/a |
| <i>M. polymorpha</i> (liverwort) | CCG | n.d. | SNORD13-like (this study) | yes |
| <i>C. braunii</i> (freshwater algae) | CCG | n.d. | SNORD13-like (this study) | yes |
| <i>C. reinhardtii</i> (green algae) | CCG | yes (this study) | not found | yes |
| <i>G. sulphuraria</i> (red algae) | CCG | yes (this study) | not found | yes |
| <i>C. crispus</i> (red algae) | CCG | n.d. | not found | yes |
| <i>Ectocarpus</i> sp (brown algae) | CCG | yes (this study) | not found | yes |
| <i>P. tricornutum</i> (diatom, microalgae) | CCG | yes (this study) | not found | yes |
| <i>T. thermophila</i> (ciliata) | CUG | no (this study) | not found | yes |
| <i>P. tetraurelia</i> (ciliata) | CCG | yes (this study) | not found | yes |
| <i>P. falci-parum</i> (agent of malaria, parasite) | CCG | yes (this study) | not found | yes |
| <i>T. gondii</i> (agent of toxoplasmosis, parasite) | CCG | yes (N. Gao, pers. com.) | not found | yes |
| <i>G. intestinalis</i> (flagellated protozoan parasite) | CCG | n.d. | not found | yes |
| <i>T. vaginalis</i> (flagellated protozoan parasite) | CGG | no | not found | yes |
| <i>E. gracilis</i> (flagellated green algae) | CUG | no | not found | not found |
| <i>T. brucei</i> (flagellated protozoan parasite) | CUG | no | not found | yes |
| <i>L. major</i> (flagellated protozoan parasite) | CUG | no | not found | yes |

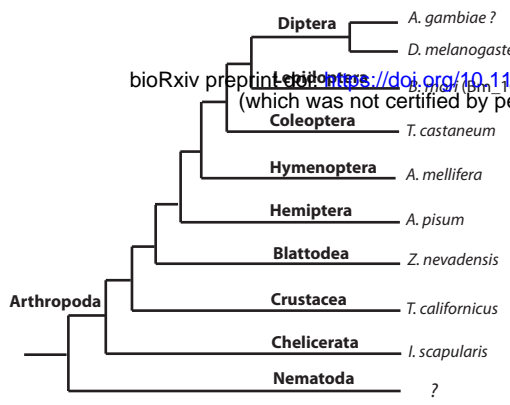
B



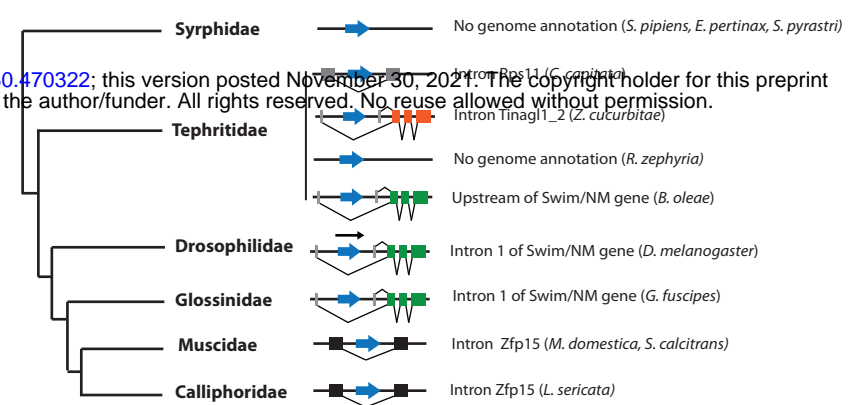
C



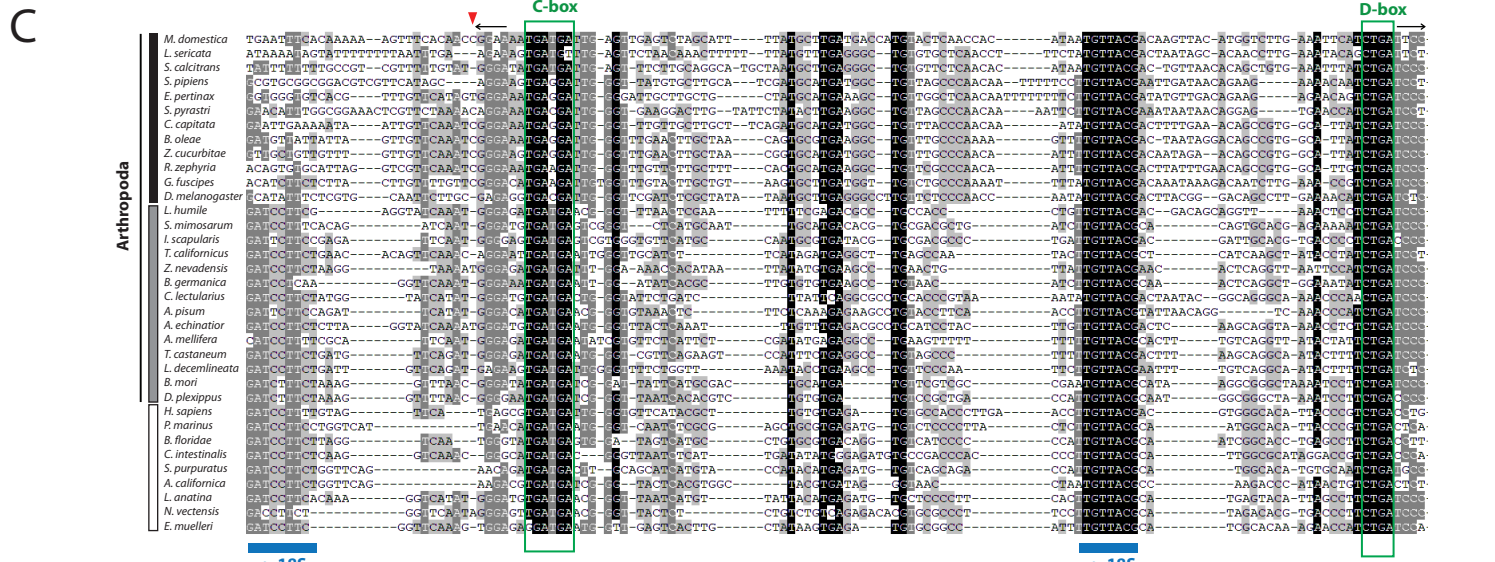
A



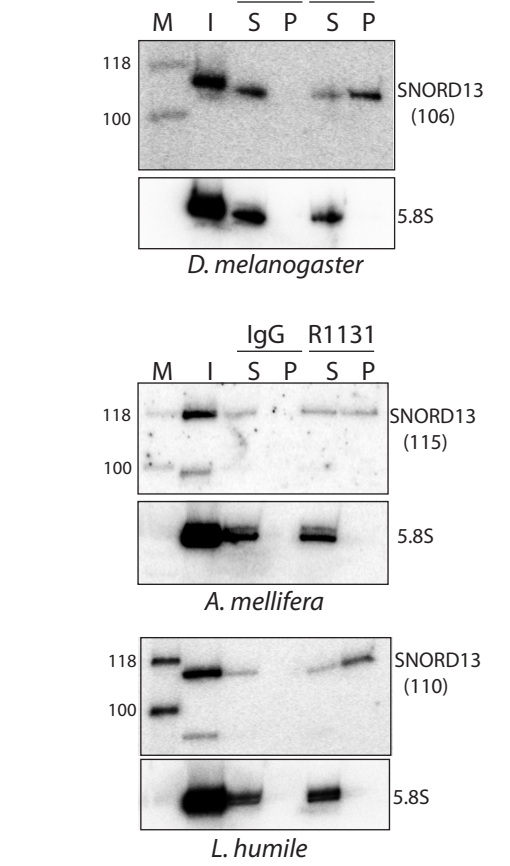
B



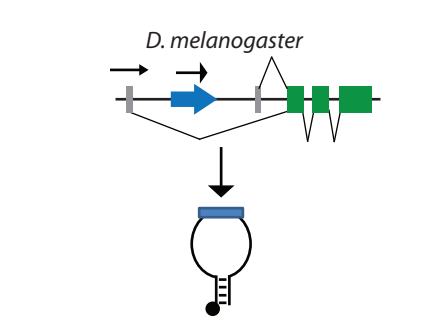
C



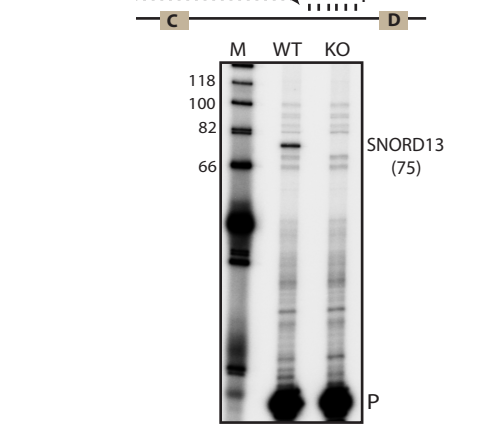
D



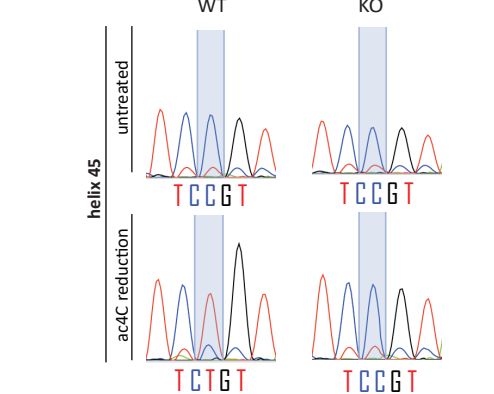
E



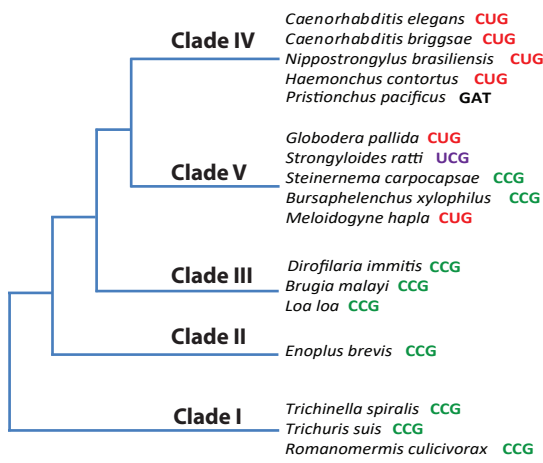
F



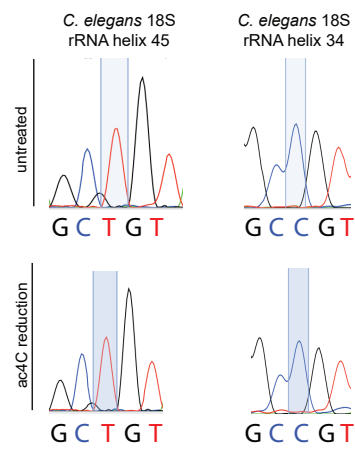
G



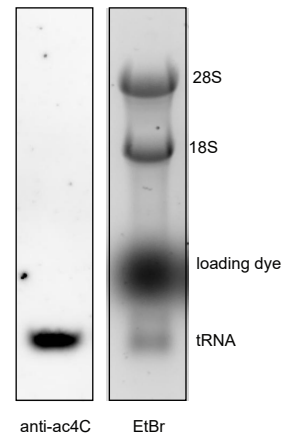
A



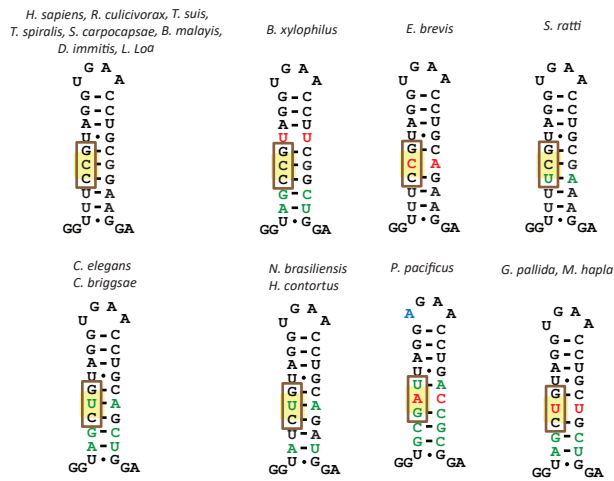
C



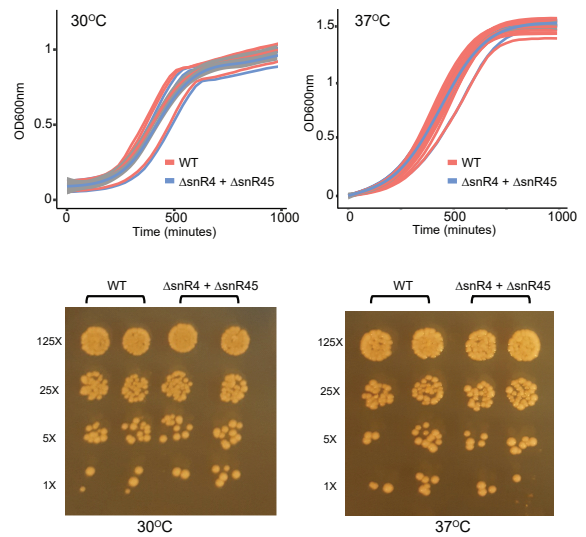
D

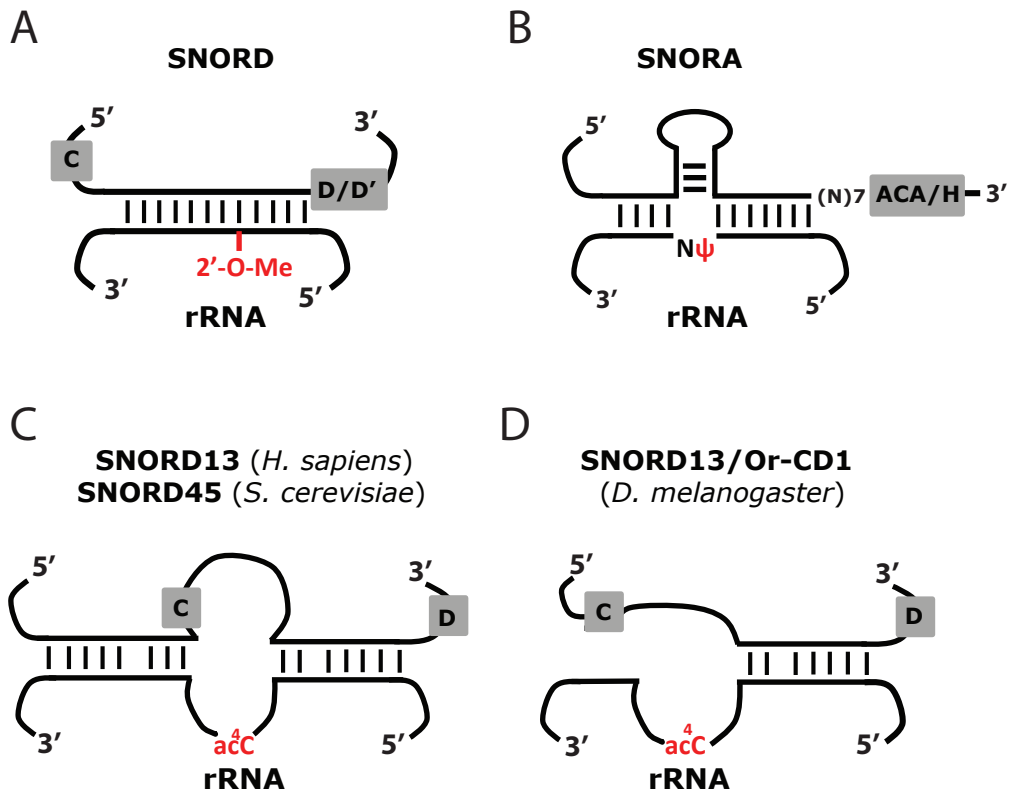


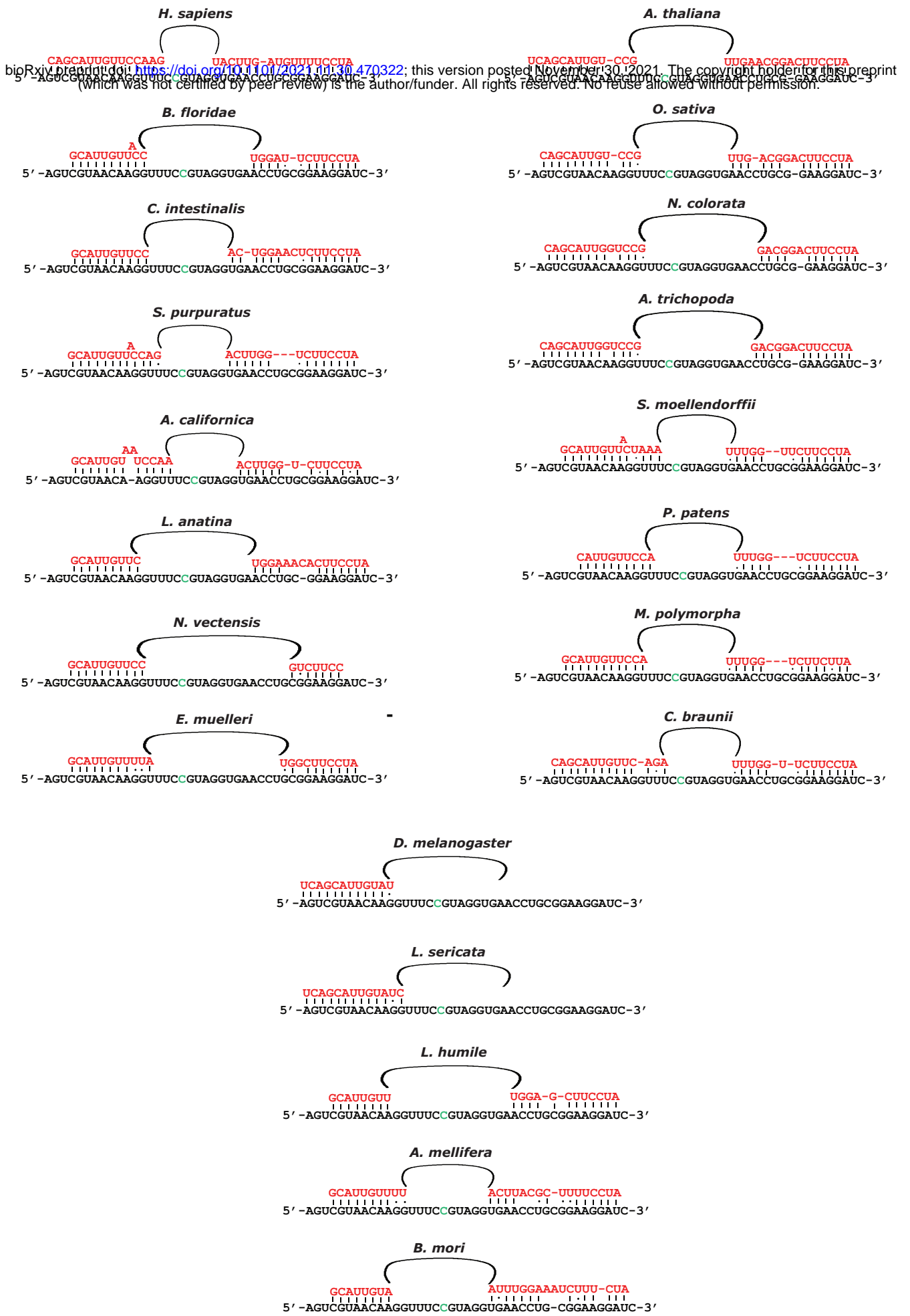
B

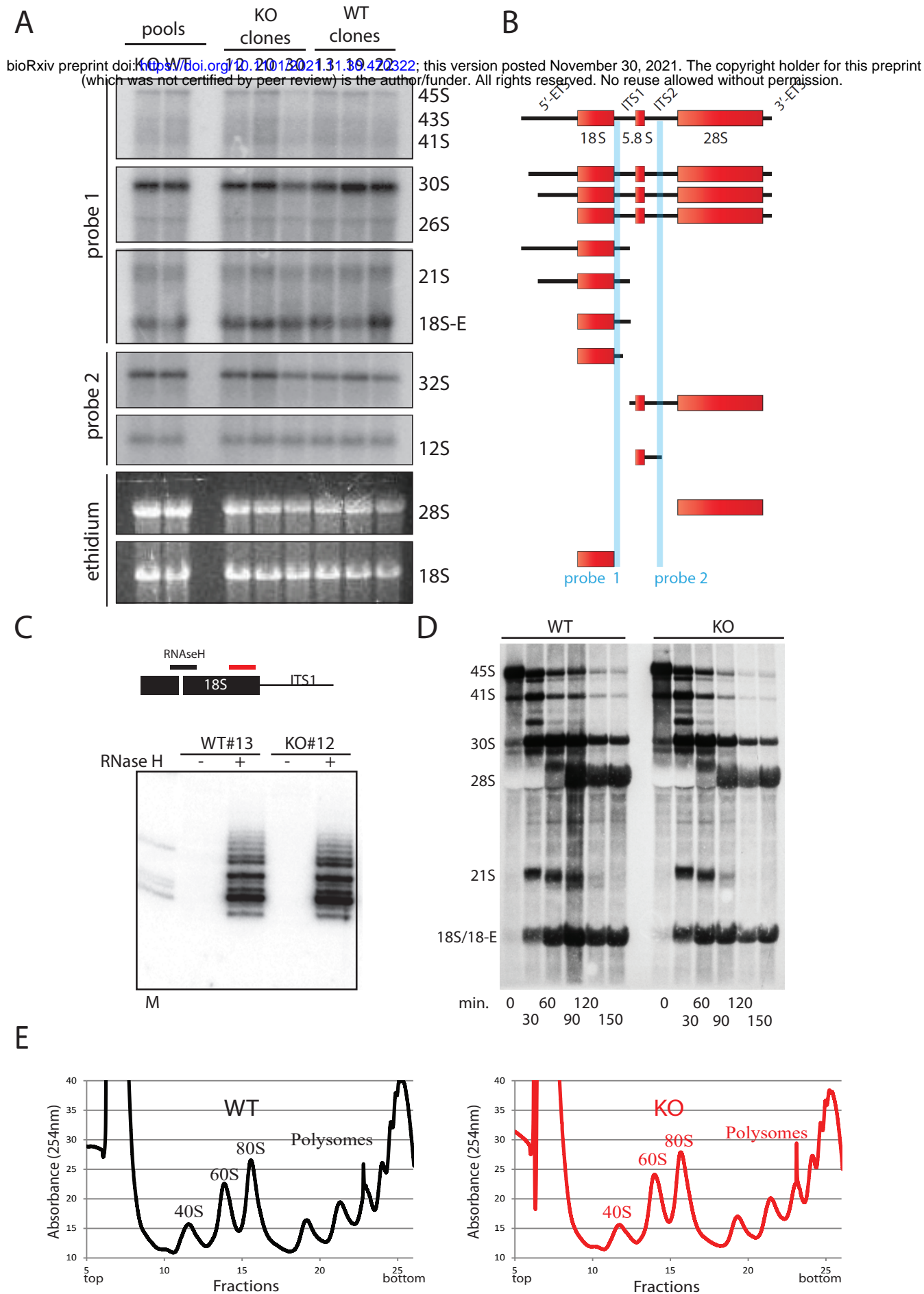


E

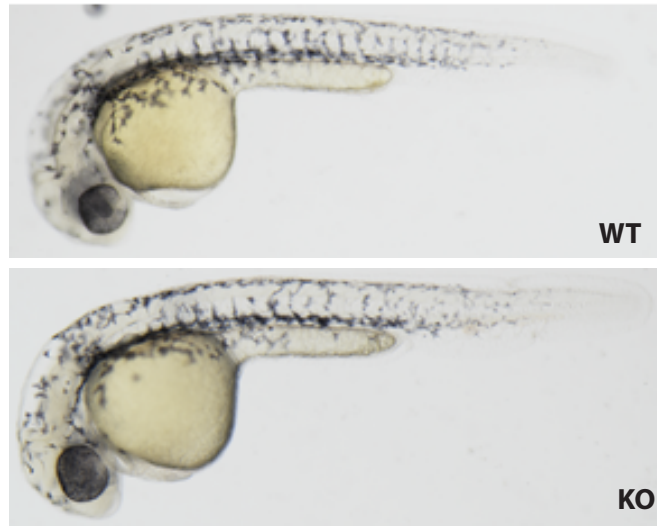




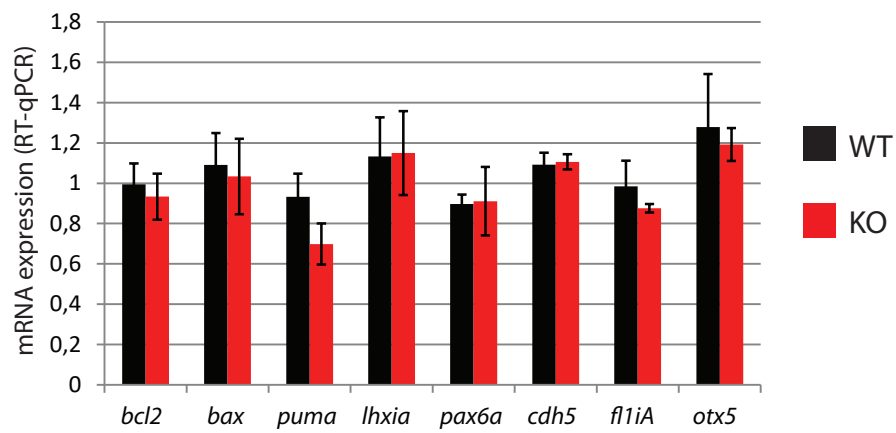




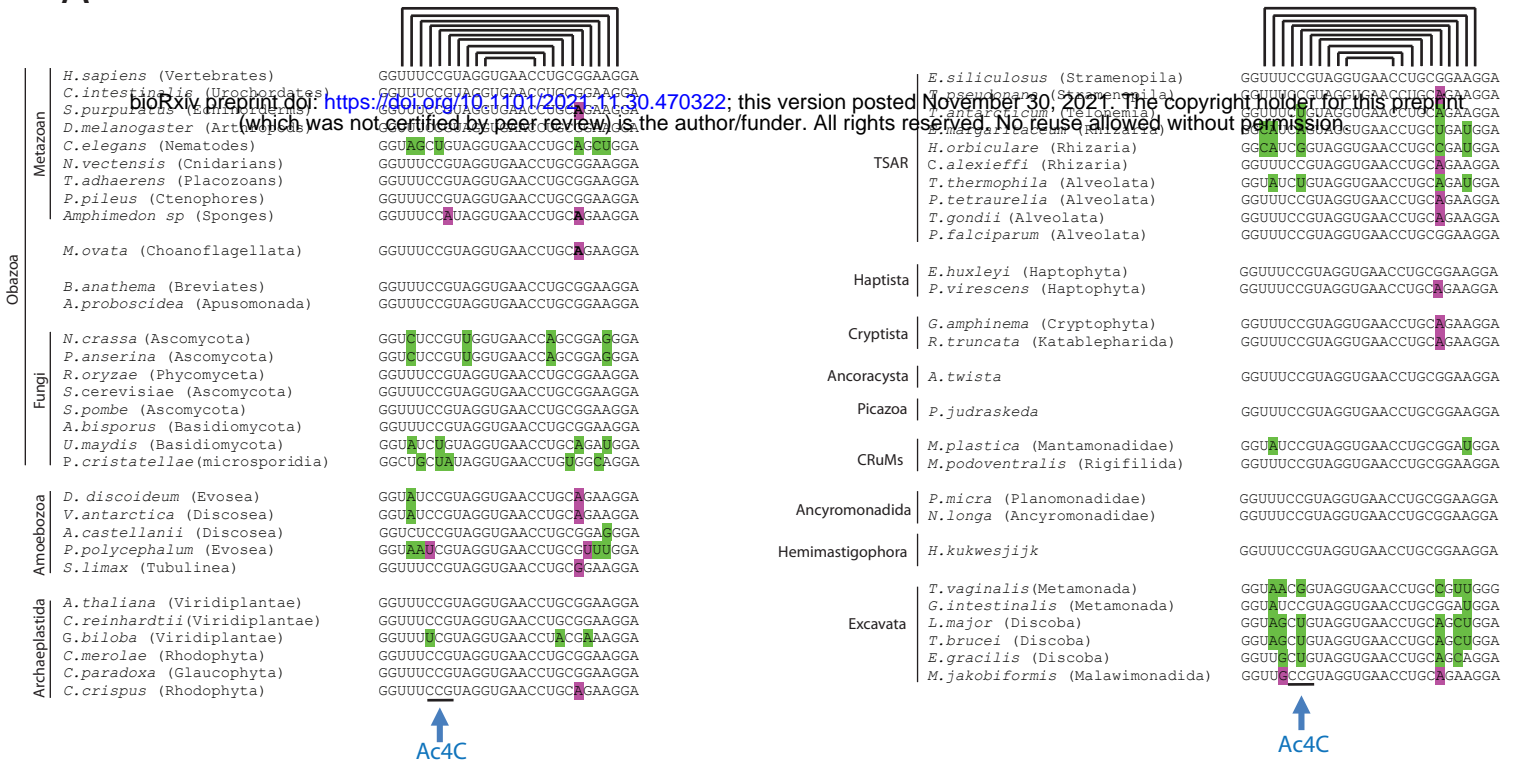
A



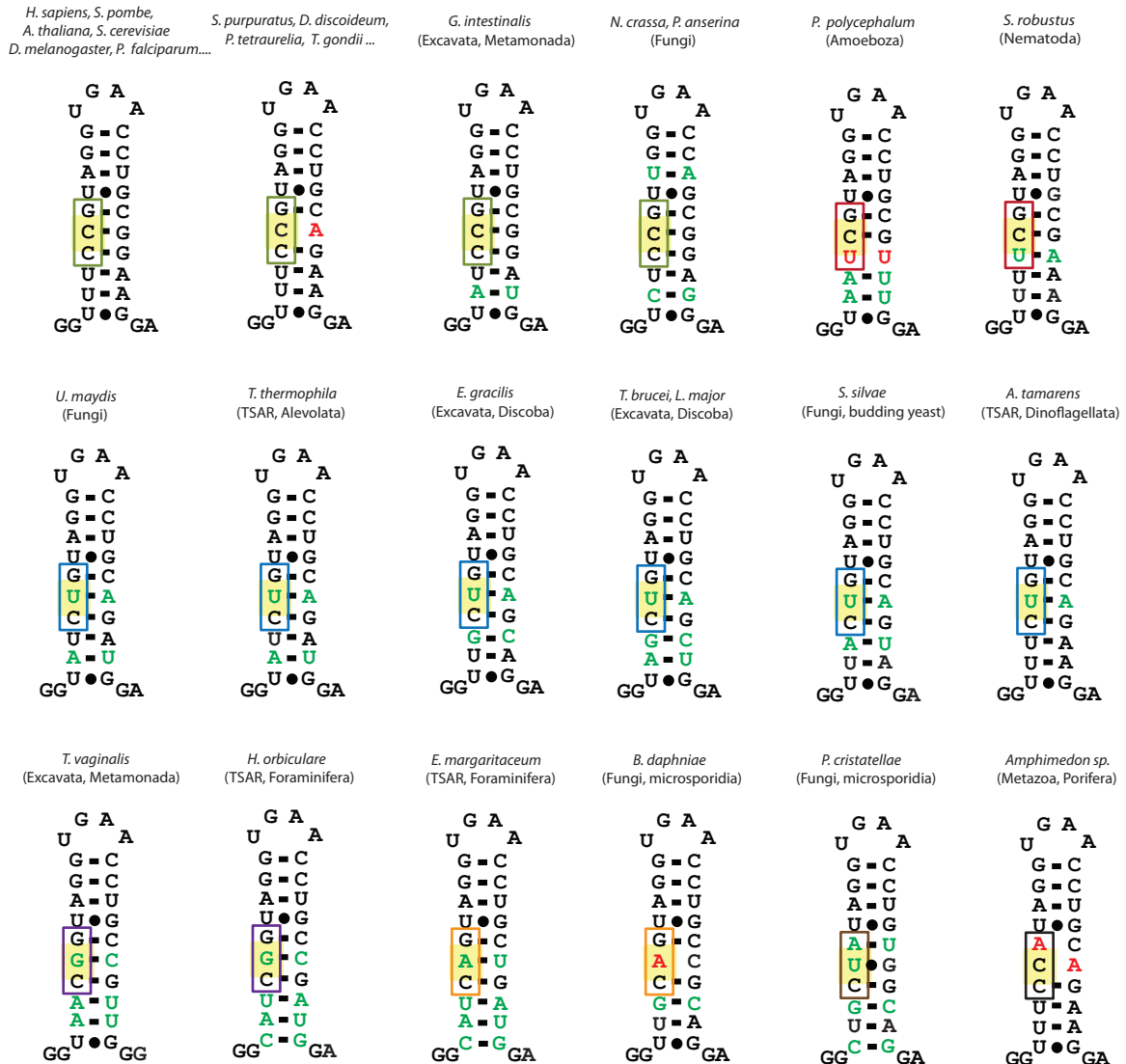
B

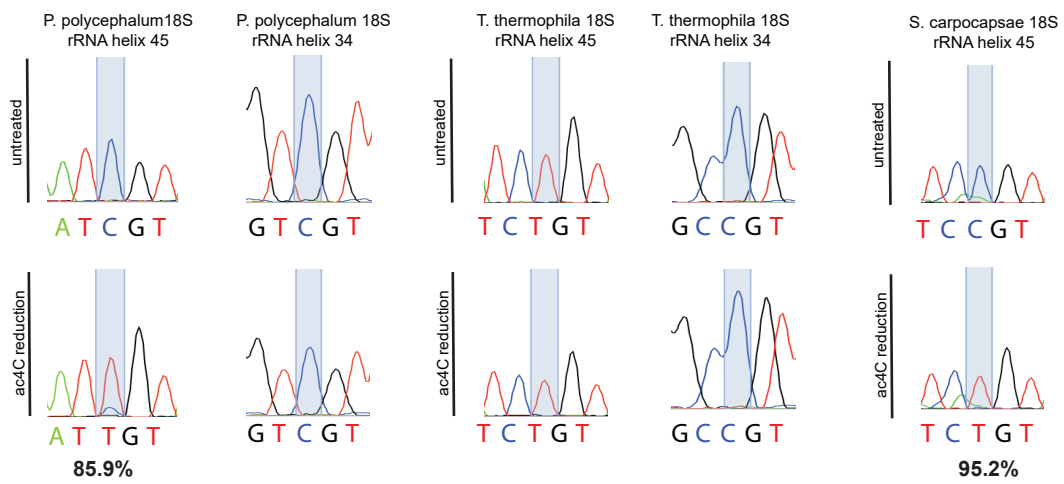


A



B

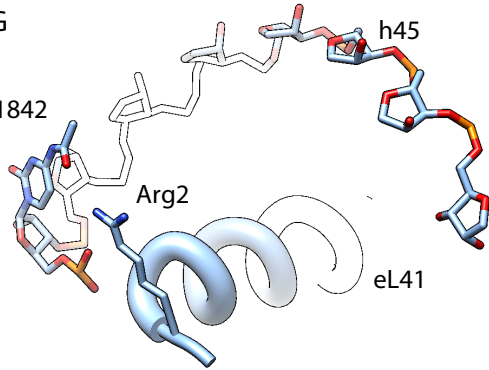




H. sapiens

CCG

ac⁴C1842

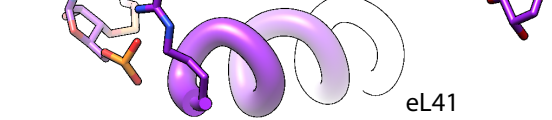


T. vaginalis

CGG

G1553

Arg2

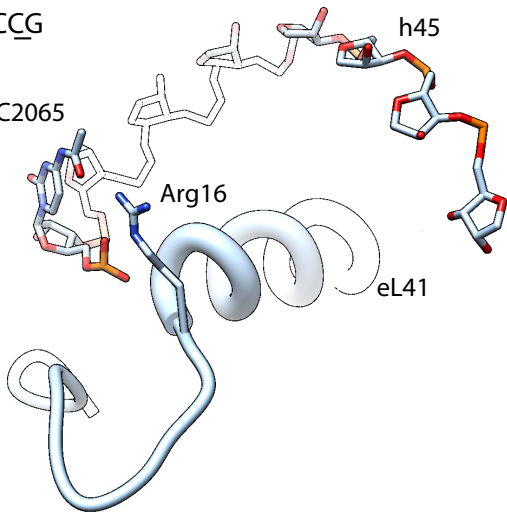


P. falciparum

CCG

ac⁴C2065

Arg16

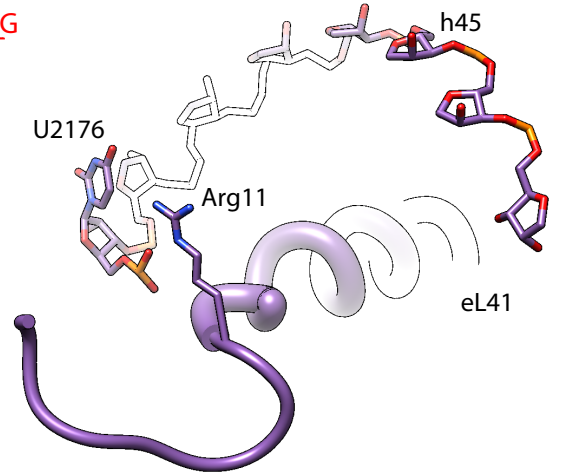


L. donovani

CUG

U2176

Arg11

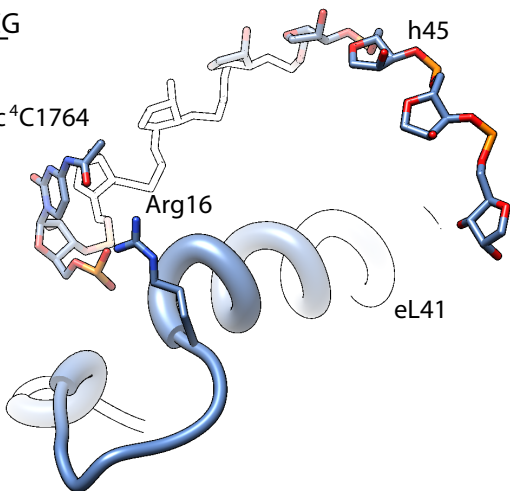


T. gondii

CCG

ac⁴C1764

Arg16

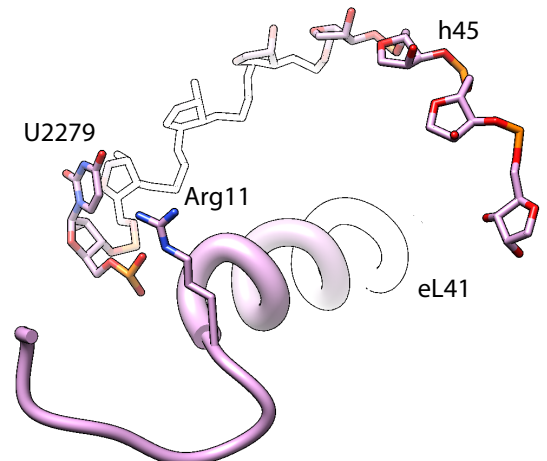


E. gracilis

CUG

U2279

Arg11



Supplementary data S1

- **Vertebrata** – A few representative species are listed

```
>Homo_sapiens
GATCCTTTTGTAGTTCATGAGCGTGATGATTGGGTGTTTCATACGCTTGTGTGAGATGTGC
CACCCCTTGAACCTTGTACGACGTGGGCACATTACCCGTCTGACC
>Callithrix_jacchus (monkey)
GATCCTTTTGTAGTTCATGAGCGTGATGATTGGGTGTTTCATACGCTTGTGTGAGATGTGC
CACCCCTTGAACCTTGTACGACGTGGGCACATTACCCGTCTGACC
>Mus_musculus (mouse)
GATCCTTTCTGGTTCATAAAGCGTGATGATTGGGTGTTTCACGCCATTGCGTGACATGTGC
CGCCATAAACCTTGTACGACGTGGGCACATTACCCGTCTGACA
>Oryctolagus_cuniculus (rabbit)
GATCCTTTTTCGTAGTTCATAAAGCGTGATGATTGGGTGTTTCACGCAATTGCGTGAGATGT
GCCACCCTCGAAACCTTGTACGACGCCGGCACATGACCCGTCTGACG
>Erinaceus_europaeus (hedgehog)
GATCCTTTTGTAGTTCATAAGTGTGATGATTGGGTGTTTCATGCGACTGCGTGAGGTGTGC
CACCCATAAACCTTGTACGACGTGGGCACATTACCCGTCTGACA
>Tursiops_truncatus (dolphin)
GATCCTTTTGTAGTTCATAAGTGTGATGATTGGGTGTTTCACGCGCTTGTGTGACATGTGC
TTCCTTCCAACCTTGTACGACTGTGGCACATTACCCGTCTGACA
>Ceratotherium_simum (Rhinoceros)
GATCCTTTTGTAGTTCATAAAGCGTGATGATTGGGTGTTTCACGCGTCTGTGTGAGATGTGC
CTCCCAAAACCTTGTACGACGTGGGCACATCACCCTCTGACG
>Canis_lupus_familiaris (dog)
GATCCTTTTGTAGTTCATAAAGCGTGATGATTGGGTGTTTCACGCCGTTGTGTGACATGTGC
CTCCCTCAAACCTTGTACGACGCCGGCACATTACCCGTCTGACG
>Pteropus_vampyrus (megabat)
GATCCTTTTGTAGTTCATAAAGCGTGATGATCGGGTGTTCATGCTGTTGTGTGAGATGTGC
CGCCCATCTAACCTTGTACGACATTAGGCACATGACCCGTCTGACG
>Loxodonta_africana (Elephant)
GATCCTTTTTCGGTTCATAAAGCGTGATGATTGGGTGTTTCATACGCTCTGTGTGATATGTGC
CTCCCCAAACCTTGTACGACGTGGGCACATTACCCGTCTGACG
>Dasypus_novemcinctus (Armadillo)
GATCCTTTTGTGGTTCATAAAGTATGATGAATGGGTGTTTCACGCGCTTGTGTGAGATGTGC
CTCCCTCAAACCTTGTACGACGTGGGCACATTACCCGTCTGACA
>Monodelphis_domestica (Opossum)
GATCCTTCTGGTTCATACAAGTGTGACTGGGTGTTTCACGCGTTTGTGTGAGATGTGCCG
CCCTTACTTTTCTTGTACGCATAGGTACATTACCCGTCTGATT
>Ornithorhynchus_anatinus (Platipus)
GATCCTTTTCGGTTCATTAATAATGATGATTGGGTGTTTCACGCGCCTGCGTGAGATGTGCC
GCCCTTTGTTACGCATGCGCACATTACCCTTCTGATC
>Gallus_gallus (chicken)
GATCCTTCCGGTTCATAGCAAATGATGAATGGGAGTTGCACGCGGCTGCGTGACGTGTGC
GCCCTTTGTTACGACGTGCACAGCCCCTTCTGAGC
>Anolis_carolinensis (lizard)
GATCCTTCCAGTTCCTTCATGTGATGAGTGGGTGTTTCATGCGCATGCGTGAGATGTGCC
GTCCCAATTGTTACGACAGGGCACAGTACCCATCTGATC
>Xenopus_tropicalis (frog)
GATCCTTCTGGTTCCTTACATGTGATGAGTGGGTATCACACGGCTGTGTGAGATGTGCCT
TCCCGACTTGTACGACTCGGCATAGTACCCATCTGATC
>Takifugu_Rubripes (fugu)
GATCTTTCTGGTTCATGCATGTGATGATTGGTCAAATCACGGATCAATCTGTGAAAGGTG
CAGCCTTCTTGTACGCTATGCGCCTCAGCCTCTCTGATC
>Danio_rerio (Zebrafish)
GATCCTTGAGGTTTCATACAACATGATGATTGGTCTATCACGACTGTTTCGTGAAATGTGCA
GCCCTTTGTTACGACGTGGCACATTGACCCGTCTGATT
>Petromyzon_marinus (lamprey)
```

GATCCTTCCTGGTCATTGAACATGATGAATGGGTCAATCTCGCGAGCTGCGTGAGATGTG
TCTCCCCTTACTCTTGTTACGCAATGGCACATTACCCGTCTGACT

• Urochordata

>*Ciona_intestinalis*

GATCCTTCTCAAGGTCAAACGGGCATGATGACGGGTTAATCTCATTGATATATGGGAGATGTGCCGACCCACCCC
TTGTTACGCATTGGCGCATAGGACCGTCTGACCCA

• Cephalochordata

>*Branchiostoma_floridae*

GATCCTTCTTAGGTCAATGGGTATGATGAGTGGATAGTCATGCCTGTGCGTGACAGGTGTCATCCCCCATTTGTT
ACGCAATCGGCACCTGAGCCTTCTGACCTT

>*Branchiostoma_belcheri*

GATCCTTCTTAGGTCAATAGGTATGATGAGTGGTTCCCTCATGCCTGTGCATGACAGGTGTTGTCCCACCATTGTT
ACGCATTCAACACCTGAGCCCTCTGACCTT

>*Branchiostoma_lanceolatum*

GATCCTTCTTAGGTCAATAGGTATGATGAGTGGATACTCATGCCTGTGCGTGACAGGTGTCATCCCCCATTTGTT
ACGCAATCGGCACCTGAGCCCCCTGACCTT

• Echinodermata

>*Strongylocentrotus_purpuratus*

GATCCTTCTGGTTCAGAACAGATGATGACTTGCAGCATCATGTACCATACATGAGATGTGTCAGCAGACCATTGTT
TACGCATGGCACATGTGCAATCTGATGCC

>*Lytechinus_variegatus*

GATCCTTCTGGTTCAGAACAGGTGATGACTTGCATGATCATGTAACTATACATGAGATGTGTCAGCAGACCATTG
TTACGCAAGGCACATATGCAATCTGATGCT

>*Patiria_miniata*

GATCCTTCTAAGGTCAAACGGGAGATGATGATTGGTTGATCATGTCTTAGTACATGAGACGTGTGACCATTTTAT
CCTTGTTACATACCATCACGTTGACCATCTGATCCC

>*Asterias_rubens*

GATCCTTCTAAGGTCAAACGGGAGATGATGATTGGTTATTTCATGCTCTATACATGAGACATGTCACCATTTATCCT
TGTTACATTCTGACATGAAACCATCTGATCCC

>*Anneissia_japonica*

GATCCTTCCATAATCATGGAATAATGATGAAAGGTATTTTCATGCACCATGCATGAGAGGGGCTGACCTCCCTACT
TGTTACGTGAAGCACCTCACCTATCTGAATCC

>*Acanthaster_planci*

GATCCTTCTAAGGTCAATTGGGAGGTGATGATTGGTTTATCATGTCTGTACATGAAACGTGTGCCCATGGTTAC
GCACTGACACGTTAGCCATCTGATCCC

• Mollusca

>*Mytilus_galloprovincialis*

GGCCTTCCTGGTTTATACATTATGATGAATAGGGTAATCATGTGACTGCATGAGACATG
TCTCCTTCTAACTTGTTACGCAGAAGGCATGTTCCCATCTGAATAA

>*Pecten_maximus*

GACCTTCTTTACTGATGGTTCATTAAAGAATGATGAATGGGTTTTTCAGGCGTCTGTCTGACACATGTCGTCTCC
ATTGTTACGCATGACATGTTACCCAACCTGACTCC

>*Pomacea_canaliculata*

GATCCTTCTGGTTCAAACGCATGATGAATGGGAAGTCACGCGACTGCGTGATGGTGCCAACCTCCAGTTACGCA
ATGGCACATCCCGTCTGACTTG

>*Crassostrea_virginica*

GTTAATGAAACCTTCTGGTTCTGTCAAATGATGACTGGCTGGGACATGCAACATGCATGACACATGCCACCGACT
TGTTACGTTAGGCATGTTTACCCTCTGATTCA

>*Aplysia_californica*

GATCCTTCTGGTTCAGAAGACGTGATGATCGGGTACTCACGTGGCTACGTGATAGGGTAACTAATGTTACGCCA
AGACCCATAACTGTCTGACTCT

>Mizuhopecten_yessoensis
GACCTTCTTTACTGATGGTTCAATTTAGAATGATGAAAGGTTTCTCAGGCGTCTGTCTGAGACATGTCACCTCCAT
TGTTACGCACGACATGTTACCTCTCTGACTCT
>Octopus_bimaculoides
GATCCTTCTGTGTTTCTGAGAAATGATGAATCGGGTTATCACGTAATGATATGTGAGACATGTCGCCGTTTACCT
TGTTACGACTATACGACATGACACCCTGTCTGATTCT
>Octopus_sinensis
GATCCTTCTGAGTTCTGAGAAATGATGAATCGGGTTATCACGTAATGATACGTAATGAGACATGTCGCCGTTTACCT
TGTTACGACTATACGACATGACACCCTGTCTGATTCT
>Batillaria_attramentaria
GATCCTTCTGTTTCTGAAACACATGATGAATGGAGTACATGCGTCTGCATGACAGTGCCAGCCTCCTTGTTACGCT
CAGGCACTCAGCCATCTGACTCA
>Crassostrea_gigas
GTTAATGAAACCTTCTGGTTCTGTTACATGATGACTGGTGGGGACATGCTCCATGCATGAGACATGCCGCCGGCT
TGTTACGCTAGGCATGTGTACCTTCTGACTCA
>Lottia_gigantea
GGTAATGACCTTCTAGGTCATATGGGATATGATGAACCGGTTAATCTCAGGCATTGTGGAGATATGTGTCAGCCA
ACCTTGTTACGCTCTGACACAGAGAACCATCTGATCCC
>Gigantopelta_aegis
GGCCTTTTGGTTTCAATAAAATGATGAATGGGTTATTCATGTGTCTACATGAGATATGTCTCCAGCATTGTTACG
CACAGACATATTACCTCTCTGATA

• Brachiopoda

>Lingula_anatina
GATCCTTCTCACAAGGTCATATGGGATGTGATGAACCGGTTAATCATGTTATTACATGAGATGTGCTCCCCTTAC
TTGTTACGCATGAGTACATTAGCCTTCTGATCCC

• Cnidaria

>Clytia_hemisphaerica
GATCCTTCTGGTTCAACGGAATGATGAGTCGGTCTTCACTTTACGTAAGTGATACGTGTTTACGCCAACCTTG
TTACGCACGGGCACGTCACCGAGTCTGATCC
>Acropora_millepora
GAATCCTTCTCAGGATTTAACAAGGGATATGATGAGTGGTTACTTCACTGGTACGTGAGACGTGTCACCCTTT
CGTTACGACACCGACACGAGACCTCTCTGATCCC
>Acropora_digitifera
GAATCCTTCTCAGGATTTAACAAGGGATATGATGAGTGGTACTTCACTGGTACGTGAGACGTGTCACCATTT
CGTTACGACACCGACACGAGACCTCTCTGATCCC
>Nematostella_vectensis
GACCTTCTGGTTCAATAGGGAGTTGATGAACCGGTTACTCTCTGTCTGTCAGAGACACGTGCGCCCTTCCCTTGT
ACGCATAGACACGTGACCTTCTGATCCC
>Pocillopora_damicornis
TATCCTTCTGGTTTCAATGAGGAGATGATGATTGGTCCAATCACTGTTCCATCAGTGAGACGTGTCCTCAACTTGT
TACGCATAGACACGTGACCTCTCTGATCCT
>Stylophora_pistillata
CATCCTTCTGGTTTCAATGAGGAGATGATGATCGGTTCCAATCACTGCTCTATCAGTGAGATGTGTCCTCATCTTGT
TACGCATAGACACGTGACCTCTCTGATCCT
>Dendronephthya_gigantea
GATTCTTCAATGGTTTCAATAGGGAGATGATGATTGGTTAATCACCTCATTAGGTGATAAGTGTCGTCATTTTTTG
TTACGCATAGACACTTGCCCATCTGATCCC
>Paramuricea_clavata
TCCTTCAATGGTTTCAATAGGGAGATGATGATTGGTTAATCACCTCATTAGGTGAGAAGTGTCGTCATAATTTGTT
ACGCATAGACACTTGACCTCTCTGAT
>Exaiptasia_pallida
GACTTCTGGTTTCAATGCGGAAGTGAATGGGTTTTCACTTTTCTTAGTGAGACGTGTCCTCTCTGTTTAC
GCACGACACGTAACCCATCTGATCCC
>Actinia_tenebrosa
GCCTTCAATGGTTTCAATGCGGAAGTGAATGGGTTAATCGCTAGTATTAGTGAGACGTGTCACCCACCAATGTT
ACGCACAGACACGTGACCTCCTGATCCC

>Actinia_equinia

GGCCTTCATGGTTCATGCGGGAAGTGATGAACTGGGTAATCACTAGTATTAGTGAGACGTGTCACCCACCAATGT
TACGCAAAGACACGTGACCCTCCTGATCCC

• Porifera

>Ephydatia_muelleri

GATCCTTCGGTTCAAAGTGGAGAGGATGAATGGTTGAGTCACTTGCTATAAGTGAGATGTGCGGCCATTTTGTTA
CGCATCGCACAAAGAACCATCTGATCCA

• Arthropoda

>Lucilia_sericata

AGAAAGTGATGTTTTGAGTTCTAACAACTTTTTTTATGTTTGAGGGCTGTGTGCTCAACC
TTTCTATGTTACGACTAATAGCACAACTTGAAATACAGCTGATTCT

>Stomoxys_calcitrans

GGGATATGATGATTGAGTTTCTTGACAGGCATGCTAATGCTTGAGGGCTGTGTTCTCAACA
CATAATGTTACGACTGTTAACACAGCTGTGAAATTTATCTGATCCC

>Syrirta_pipiens

AGGAAGTGAGGATTGGGTTATGTGCTTGATCGATGCATGATGGCTGTTAGCCCAACAAT
TTTTCTTGTACGAATTGATAACAGAAGAAAACAATCTGATCCT

>Eristalis_pertinax

GGGAAATGAGGATTGGGATTGCTTGCTGCTATGCATGAAAGCTGTTGGCTCAACAATTT
TTTTCTTGTACGATATGTTGACAGAAGAGAACAGTCTGATCCC

>Scaeva_pyrastris

AGGAAATGACGATTGGGTGAAGGACTTGTATTCTATACTTGAAGGCTGTTAGCCCAACAA
AATTCTTGTACGAAATAATAACAGGAGTGAACCATCTGATCCT

>Ceratitis_capitata

GGGAAATGAGGATTGGGTTTGTGCTTGCTTCAGATGCATGATGGCTGTTTACCCAAACAA
ATATGTTACGACTTTTGAAACAGCCGTGGCATTATCTGATCCC

>Bactrocera_tryoni

GGGAAATGAGGATTGGGTTTGAACCTTGCTAACTATGCATGATGGCTGTTTGCCCAACAAT
TTTGTTACGACTAATAGAACAGCCGTGGCATTATCTGATCCC

>Bactrocera_oleae

GGGAAATGAGGATTGGGTTTGAACCTTGCTAACAGTGCCTGAAGGCTGTTTGCCCAAAAAGT
TTTGTTACGACTAATAGGACAGCCGTGGCATTATCTGATCCC

>Bactrocera_latifrons

GGGAAATGAGGATTGGGTTTGAACCTTGCTAGCAATGCGTGAAGGCTGTTTGCCCAACAAT
TTTGTTACGACTATTAGAACAGCCGTGGCATTATCTGATCCC

>Rhagoletis_zephyria

GGGAAATGAAGATTGGGTTTGTCTTCTGCTTCACTGCATGAAGGCTGTTGCGCCCAACAAT
TTTGTTACGACTTATTTGAACAGCCGTGGCATTGTCTGATCCC

>Glossina_austeni

GGGGCGTGAAGATTGTGGTTTGTACTTGCTGTAAGTGCTTGATGGTTGTCTGCCCAAAAC
TTAATGTTACGACAAATAAAGACAACCTTGAAACCGTCTGATCCC

>Glossina_pallidipes

GGGGCGTGAAGATTGTGGTTTGTACTTGCTGTAAGTGCTTGATGGTTGTCTGCCCAAAAC
TTAATGTTACGACAAATAAAGGCAACCTTGAAACCGTCTGATCCC

>Glossina_morsitans

GGGGCGTGAAGATTGTGGTTTGTACTTGCTTTAAATGCTTGATGGTTGTCTGCCCAAAAC
TTAATGTTACGACAAATAAAGGCAACCTTGAAACCGTCTGATCCC

>Glossina_brevipalpis

GGGAAGTGACGATTGTGGCATGTACTTTCTTTAAATGTTTGATGGTTGTCTGAGCACAATA
TTAATGTTACGACTGATAAGACAGCCTTGAAACCATCTGATCCC

>Glossina_palpalis

GGGACATGAAGATTGTGGTTTGTACTTGCTGTAAGTGCTTGATGGTTGTCTGCCCAAAAT
TTTATGTTACGACAAATAAAGACAATCTTGAAACCGTCTGATCCC

>Glossina_fuscipes fuscipes

GGGACATGAAGATTGTGGTTTGTACTTGCTGTAAGTGCTTGATGGTTGTCTGCCCAAAAT
TTTATGTTACGACAAATAAAGACAATCTTGAAACCGTCTGATCCC

>*Drosophila_melanogaster*
GAGAGGTGACGATTGGGTTTCGATCTCGCTATATAATGCTTGAGGGCCTTGTTCTCCCAAC
CAATATGTTACGACTTACGGGACAGCCTTGAAAACATCTGATCTC

>*Drosophila_ananassae*
GAGAGGTGAAGATTGGGCTAGGTCTCGCGAACTGATGCTTGAGGGCCTTGTTCTCCCAAC
CAATATGTTACGACTTACGGAACAGCCTTGAAACGATCTGATCTC

>*Drosophila_simulans*
GAGAGGTGACGATTGGGTTTCGATCTCGCTATAAAAATGCTTGAGGGCCTTGTTCTCCCAAC
CAATATGTTACGACTTACGGGACAGCCTTGAAAACATCTGATCTC

>*Drosophila_sechellia*
CGAGAGGTGACGATTGGGTTTCGATCTCGCTATAAAAATGCTTGAGGGCCTTGTTCTCCCAA
CCAATATGTTACGACTTACGGGACAGCCTTGAAAACATCTGATCTC

>*Drosophila_yakuba*
GAGAGGTGACGATTGGGTTTCGATCTCGCTATATAATGCTTGAGGGCCTTGTTCTCCCAAC
CAATATGTTACGACTTACGGGACAGCCTTGAAAACATCTGATCTC

>*Drosophila_erecta*
GAGAGGTGACGATTGGGTTTCGATCTCGCTATATAATGCTTGAGGGCCTTGTTCTCCCAAC
CAATATGTTACGACTTACGGGACAGCCTTGAAAACCTCTGATCTC

>*Drosophila_pseudoobscura*
GAGAAGTGAGGATTGGGCCAGGTCTCGCAATACTATGCTTGAGGGCCTTGTTCTCCCAAC
AAATATGTTACGACTTGTGGAACAGCCTTGAAAACATCTGATCTC

>*Drosophila_persimilis*
GAGAAGTGAGGATTGGGCCAGGTCTCGCAATACTATGCTTGAGGGCCTTGTTCTCCCAAC
AAATATGTTACGACTTGTGGAACAGCCTTGAAAACATCTGATCTC

>*Drosophila_mojavensis*
GAGAGGTGAAGATTGGGTTATTCTTTCCTTAACCATGCTTGAAGGTTCTGTTTGCCCAAC
CAAATATGTTACGATTACAAAACAGCCTTGAAAACATCTGATCTC

>*Drosophila_virilis*
GAGAGGTGAAGATTGGGTTACTCTTGCCTTAACCATGCTTGAGGGTCTTGTTTCGCCCAAC
TTATACGTTACGACTAACAGAACAGCCTTGAGAACATCTGATCTC

>*Drosophila_grimshawi*
GAGAAGTGAGGATTGGGTTAATCTTGCCTTAACCATGCTTGAGGGTCTTGTTTCGCCCAAC
ATTTTGTACGACTAACAGAACAGCCTTAGGATAACATCTGATCTC

>*D. willistoni*
GAGGTGTGAGGATTGGGTTTAGTTCTAGCATTTCATGCTTGATCGCTGTGTTTCGCCCA
ACTGTTTATGTTACGACTAAAGAACAGCCATGAGAAACATCTGATCTC

>*Frankliniella_occidentalis*
GATCCTTCTAGGTTTAAACGGGAATTGATGAATGGATTATCAAACGATACATTTGTTTGAA
GCCTGAGTCTATCTTGTACGACTTGTTCAGGCGAAACCTATCTGATCCCTC

>*Melanaphis_sacchari*
GATTCTTCCAGATTCATATGGGACATGATGAACGGGTTAATCTCTTCTTGAAGAGAAGC
CTGTGCCCTCACACTTGTTACGTATTAACAGGTCAAACCTATCTGATCCT

>*Diaphorina_citri*
GATCCTTCTGTGATTGAAATGGGATATGATGAATGGGGTATGTCACGTGACTCCGTGAA
GCCTACACTCCTTACCTTGTTACGATTAGGTAGGCGAAACCTCTCTGATCCC

>*Cimex_lectularius*
GATCCTTCTATGGTATCATATGGGATGTGATGACTGGGTATTCTGATCTTATTCAGGCGC
CTGCACCCGTAATAATGTTACGACTAATACGGCAGGGCAAAAACCAACTGATCCC

>*Nilaparvata_lugens*
GATCCTTTCAGTTCAATGAAAGATGATGAATAAGGTAAACATCCATGTGGGTGAAACCTG
AATCTTTCCTTGTTACGCTTATTCAGGTCAAACCTCTCTGATTTT

>*Aphis_gossypii*
GATTCTTCCAGATTCATATGGGATATGATGAACGGGTCTTATCTCTTCTTATGGAGAATC
CTGTGCCTTCAAACCTTGTTACGTATTTATCAGGTTAAACCCATCTGATCCT

>*Diuraphis_noxia*
GATTCTTCCAGATTCATATGGGACATGATGAATGGGTCTAAACTCTCTCAAAGAGAAGC
CTGAGCCTTCAAACCTTGTTACGTATTTAACAGGTCAAACCTATCTGATCCT

>*Myzus_persicae*
GATTCTTCCAGATTCATATGGGAAATGATGAACGGGTCTAATCTCTTCTCAAAGAGAAGC
CTGTGCCCTCAAACCTTGTTACGTATTAACAGGTCAAACCTTCTGATCCC

>*acyrthosiphon_pisum*
GATTCTTCCAGATTCATATGGGACATGATGAACGGGTGTAAACTCTTCTCAAAGAGAAGC
CTGTACCTTCAACCTTGTTACGTATTAACAGGTCAAACCCATCTGATCCC

>*Bemisia_tabaci*
GATTCTTCTGAGTTCATTGGGAAATGATGATCGGTCAATTTATGACTCCTACTGATGACGC
CTGAAGCCATTTGTTACGACTTGGTCAGGTGAAACCCATCTGATCCC

>*Sipha_flava*
GATTCTTCCAGATTCATATGGGACATGATGAACGGGTTTAAACTTCATTTCGTGAAGAATC
CTGTGCCTTTAAACTTGTTACGATTTTCATCAGGTCAAACCTATCTGATCCC

>*Gerris_buenoi*
ATGTAGATATTTGTTTGTATGGGGATATGATGATTGGTTGTAACCGGTGTTCTGCGCGAA
GCCTGGACCTCTTTATTAACCTTGTTACGATTAACCCAGGTAAATACCATCTGATCCT

>*Laodelphax_striatellus*
GATCCTTTTCAAGTTCATGAAAGATGATGAATAAGGTAATCATCCTTCATGGGTGAAGCCT
GAGTCTTTACTTGTACGCATATTCAGGTCAAACCTCTCTGATTTT

>*Ixodes_scapularis*
GATTCTTCCGAGATTCATGAGTGGGAGTGTGAGTCGTGGGTGTTTCATGCCAATGCGTGATA
CGTGCGACGCCCTGATTGTTACGACGATTGCACGTGACCCCTCTGACCCC

>*Varroa_destructor*
GATCCTTCTGACCAGTTCAGAAGCATGATGAGTCGGGAGTCATACCTTCCATGGTGACAC
GTGTGACGCTATAACCTTGTTACGCATTCGCACGGGACCCGCCTGATTCT

>*Varroa_jacobsoni*
GATCCTTCTGACCAGTTCAGAAGCATGATGAGTCGGGAGTCATACCTTCCATGGTGACAC
GTGTGACGCTATAACCTTGTTACGCATTCGCACGGGACCCGCCTGATTCT

>*Dermacentor_silvarum*
GATCCTTTAACGGTTCCTACTGGGGAGTGTGAGTTGTGGGTTTCGCACGCCTCATGCGTGA
TACGTGCCACGCCCTTTAACCTTGTTACGACGACAGGCACGTAGCCCTTCTGACCCC

>*Galendromus_occidentalis*
GATCCTTCTGACTAGATTCAGAAGCATGATGAGTCGGGATTCTGACACCCAGTGGTGAC
ACGTGTGACGCCAATCCTTGTTACGCAAACGCACGCGACCCGCCTGATTCTCG

>*Rhipicephalus_sanguineus*
GATCCTTTTTGGTTCCTACTGGGGAATGATGAGCTGTGGGTTTCGCACGCCTCATGCGTGAT
ACGTGCCACGCCCTTTTACCTTGTTACGACGATTGGCACGTAGCCCTTCTGACCCC

>*Rhipicephalus_microplus*
GATCTGTTGCGGTTCTACTGGGGAATGATGAGTTGTGGGTCTGCATGCCTCGTGCGTGAT
ACGTGCCGCGTCTTTTACCTTGTTGCGACGATTGGCACGTAGCCCTTCTGACCCC

>*Stegodyphus_mimosarum*
GATCCTTTTACAGATCAATGGGATGTGATGAGTCGGGTCTCATGCAATTGCATGACACGT
GCGACGCTGATCTTGTTACGCACAGTGCACGAGAAAAATCTGATCCC

>*Stegodyphus_dumicola*
GATCCTTTTACAGATCAATGGGATaTGATGAGTCGGaTCTCATGCAAgTGCATGACACGT
GCGACGCTtGATgTTGTTACGCACAGTGCACGAGAAAAgTCTGATCCC

>*Loxosceles_reclusa*
GATCCTTTCTGGTTCAAAGGGAAGTGTGAGTCGGTCTTACGTTGATATACGTGAAACGG
GTACCGCTGTTTGTACGACGTTTGGCCGTAAGCCGTCTGATCCC

>*Zootermopsis_nevadensis*
GATCCTTCTAAGGTAATGGGAGATGATGATTTGGAAAACACATAATTATATGTGAAG
CCTGAAGTGTATTGTTACGAACACTCAGGTAAATTCATCTGATCCC

>*Blattella_germanica*
GATCCTCAAGGTTCAAATGGGAAATGATGAATTGGATATCACGCTTGTGTGTGAAGCCTG
TAACATCTTGTTACGCAAACCTCAGGCTGGAAATATCTGATCCC

>*Centruroides_sculpturatus*
GATCCTTTTGCATTTTACAAGTAATGGGAAATGATGAGTTGGGGTTCACGATTCTTCGT
GACACGTGCGCCACCTTCTGTTACGCATTGCACGTGACCTTCTGATCCC

>*Ostrinia_furnacalis*
GATCTTTCAAAGGTTAAATGGGGGATGATGATCGGATTGGACACGCGACTGCGTGACGT
CCGCCCGGACTGTTACGCTTGGGCGGGCCAAAATCCTACTGACTCC

>*Vanessa_tameamea*
GATCTTTCTAAAGGTTACAACGGGGAATGATGATCGGGCTAATCATGCGTCTGCGTGAAG
TCCGCCACCGATTGTTACGCATTACGGCGGGCCAAAATCCTTCTGACCCC

>Trichoplusia_ni
GATCTTTCCAAAGGTTAAACGGGGGATGATGATCGGATTGGACTTGCCTGTCATGATGT
CCACCACCGAATGTTACGCTTTGGTGGGCTAAAATCCTTCTGACTCC

>Pieris_rapae
TATCCTTTCCAGGACCAAACGGGGGATGATGATCGGTTTATTACGCGTCTGCGTGACG
TCCGTCACCGATTGTTACGCTTAGACGGGCTAAAACCTTCTGACCCC

>Papilio_machaon
GATCTTTCCAAAGGTTTAAACGGGGAATGATGAATTTCGGTTAATCACGCGGCTGCGTGAAG
TCCGTCACCGAATGTTACGCATTTGGCGGGCTAAAACCTTACTGACCCC

>Plutella_xylostella
GATCTTTCCAAAGGTATACGGGGGATGATGATCGGATTGGACATGCGCCTGCATGAAGTC
CGCAGCCGATTGTTACGCTTAGCGGGCTAAAATCCTTCTGACCCC

>Zerene_cesonia
GTATCCTTTCCAGGACCAAACGGGGGATGATGATCGGTTGAATCACGTGTCTACGTGATG
TCCGTCACCGAATGTTACGCATTAGACGGGCTAAAACCTTCTGACCCC

>Galleria_mellonella
GATCTTTCTAAAGGTTTTAAATGGAGTATGATGATCGGTTTTTTTACACGTCTGTGTGAT
GTCTGCCACCGAAATGTTACGCTTTGGCGGGCTAAAATCCTTCTGACTCC

>Danaus_plexippus
GATCTTTCTAAAGGTTTTAAACGGGGAATGATGATCGGGTTAATCACACGTCTGTGTGATG
TCCGCTGACCATTGTTACGCAATGGCGGGCTAAAATCCTTCTGACCCC

>Amyelois_transitella
GATCTTTCCAAAGGTTAAACGGGGAATGATGATCGGAATTTTTACGCGACTGCGTGAAGT
CCGCAGCCGATTGTTACGCTTGTGCGGGCGAAAATCTTTCTGACCCC

>Hyposmocoma_kahamanoa
GATCTTTCTAAAGGTTAAACGGGGGATGATGATCGGATGGACACGCGTCTGCGTGATGT
CCGCCACCGAATGTTACGCATTAGGCGGGCCAAAATCCTTCTGACCCC

>Spodoptera_litura
GATCTTTCCAAAGGTTAAACGGGGAATGATGATCGGATTGGACATGCGTCTGCATGATGTC
CGTTGCCGATTGTTACGCTTTGACGGGCTAAAATCCTACTGACCCC

>Pararge_aegeria
GATCTTTCTGTAAGTTTTTAAACGGGGGATGATGATTTCGGATTGAACACGCGTCTGCGTG
AGGTCGCCCGCCGATTGTTACGCATTTCGAGCGGGCTAAAATCCTTCTGACCCC

>Bicyclus_anyna
GATCTTTCCAAAGGTTAAACGGGGGATGATGATCCGGGTTTACACGCGTCTGCGTGATG
TCCGCCACCGTACTGTTACGCTTTACGGCGGGCTAAAATCCATCTGACCCC

>Bombyx_mori
GATCTTTCTAAAGGTTTTAAACGGGATATGATGATCGGATTATTCATGCGACTGCATGATGT
TCGTCGCCGAATGTTACGCATAAGGCGGGCTAAAATCCTTCTGATCCC

>Bombyx_mandarina
GATCTTTCTAAAGGTTTTAAACGGGATATGATGATCGGATTATTCATGCGTCTGCATGATGT
TCGTCGCCGAATGTTACGCATAAGGCGGGCTAAAATCCTTCTGATCCC

>Manduca sexta
GATCTTTCTAAAGGTTTAAACGGGGGATGATGATCCGATTAATCACGCGTCTGCGTGAAGTC
CGCCATCGATTGTTACGCATTTGGCGGGCTAAAATCAGACTGACCCC

>Heliconius_melpomene
GATCTTTCTGTAAGTTTTTACGGGGAATGATGATCGGGATGATCACGCGTCTGCGTGAGG
TCCGCCACCGAATGTTACGCAATACGGCGGGCTCAAATCCTTCTGACCCC

>Melitaea_cinxia
GATCTTTCTAAAGGTTAAACGGGGAATGATGATCGGGTTTCATCACGCGTCTGCGTGATGT
CCGCCACCGATTGTTACGCATTTCAGGCGGGCTAAAATCCTTCTGACCCC

>Heliothis_virescens
GATCTTTCCAAAGGTTAAACGGGGAATGATGATCGGATTGGACTTGCATTGCATGATGTC
CGTCGCCGAATGTTACGCTTATGGCGGGCTAAAATCCTTCTGACCCC

>Anoplophora_glabripennis
GATCCTTCTGATTGTTTCAGATGAGAAGTGATGTTTTGGGGTTTTTTCAGAAGATTATTTCTG
AAGTCTGCTCCCCAGTTTTTTTTGTTACGACTTGGAGCAGACGATACTTCTGATCTC

>Leptinotarsa_decehlineata
(GATCCTTCTGATTGTTTCAGATGAGAAGTGATGATTGGGGTTTTTCTGGTTAAATACCTGAA
GCCTGTTCCCAATTCTTGTACGAATTTTGTACGGCAATACTTTTCTGATCTC

>Diabrotica_virgifera

GATCCTTCTGATAGTTTCAGATGAGAGATGATGAATGGGTTTTTCAGAAACAAAATTTCTG
AAGTCTGTAACCTCCACTTGTACGATTTGGCAGACGATACCTACCTGATCTC
>*Tribolium_castaneum*
GATCCTTCTGATGTTTCAGATGGGAGATGATGAATGGGTCGTTTCAGAAAGTCCATTTCTGAG
GCCTGTAGCCCTTTTTGTTACGACTTTAAGCAGGCAATACTTTTTCTGATCCC
>*Aethina_tumida*
GATCCTTCTGATTGTTTCAGATGGGAGATGATGAACGGGTTCTTGCAGGGACGTCCCTGAA
GCCTGTTCCCAAACCTTGTACGATTTTGGCAGGCCACTACCCATCTGATCCC
>*Sitophilus_oryzae*
GATCCTTCTGATCGTTTCAGATGGGAGATGATGATTGGGATTTAACAGAATTTTTTCTGAA
GCCTGTACCCATATTGTTACGACTATGGCAGGCAATCCTCATCTGATCCC
>*Dendroctonus_ponderosae*
GATCCTTTTACAATGGTGAAATTTGGAAATGATGAACGGGGACGATCAGATTTTTTGTCTGA
AGTCTGCACCTATTTGTTACGACTTTTGGCAGACCATCCCTATCTGATCC
>*Anoplophora_glabripennis*
GATCCTTCTGATTGTTTCAGATGAGAAGTGATGTTTGGGGTTTTTCAGAAGATTATTTCTG
AAGTCTGCTCCCCAGTTTTTTTTGTTACGACTTGGAGCAGACGATACTCTTCTGATCTC
>*Tigriopus_californicus*
GATCCTTCTGAACACAGTTCAAACAGGAATTGATGAATTGGGTTGCATCTTCATAGATGA
GGCTTGAGCCAATACTTGTACGCTCATCAAGCTATACTATCTGATCCT
>*Catajapyx_aquilonaris*
GATCCTTCTAAAGTTCAAATGGGAAGTGATGAGCTGGTTTTCTTACCACCATGGTGAGGC
CTGCACCACAAAATATACGACACTTGCAGGATAAACCTATCTGATCCC
>*Cyphomyrmex_costatus*
GATCCTTCAATCAGGTATCAAAATGGGAGATGATGAATGGGTTCACTCAAATTTGTTTGA
GACGCCTGCACCCTACTTGTACGACGAAAGCAGGTTAAACCCCTCTGATCCC
>*Dinoponera_quadriceps*
GATCCTTCTTAAGTTTGAATGGGAAATGATGAATGGGTTTTGCTCAAATTTGTTTGAAGC
CCTGAACCTTACTTTGATACGACATTCGGCAGGCTAAACCTCTCTGATCCC
>*Nylanderia_fulva*
GATCCTTCAATCAGGTATCAAAATGGGATGTGATGAATGGGTTCAACTCGAAATTATTTCGA
GACGCCTGCGTCCCCTATGTTACGACACTAGCAGGTTAAACCTCTCTGATCCC
>*Odontomachus_brunneus*
GATCCTTCTTGAGTTTGAATGGGAAATGATGAATAGGTTTTGGCTCAAATTTATTGAGATG
CCTGCATCTTACTTTGTTACGACTTCCAGCAGGCTAAACCTCTCTGATCCC
>*Pseudomyrmex_gracilis*
TATCCTTTCGCGTTTTTAAATTTGGGAGATAATGAATGGGTTTTATTTCAAATTTTATTGGA
GACACCTGCACCCTGTATTGTTACGACATTTGCAGGTAAAACCTGTCTGATCC
>*Trachymyrmex_cornetzi*
GATCCTTCAATCAGGTATCAAAATGGGATGTGATGAATGGGTTTACTCAAATTTGTTTGA
GACGCCTGCACCCTATTTGTTACGACGCAAGCAGGTTAAACCTCTCTGATCCC
>*Temnothorax_curvispinosus*
GATCCTTCAATCAGGTATCAAAATGGGAGGTGATGAACGGGTTTACTCAAATTTATTGGA
GACGCCTGCAACCCTAGTTGTTACGACGCTAGCAGGTTAAACCCCTCTGATCCC
>*Trachymyrmex_septentrionalis*
GATCCTTCAATCAGGTATCAAAATGGGACGTGATGAATGGGTTCACTCAAATTTATTGGA
GACGCCTGCACCCTCTTGTACGACGCAAGCAGGTTAAACCCCTCTGATCCC
>*Trachymyrmex_zeteki*
GATCCTTCAATCAGGTATCAAAATGGGATATGATGAATGGGTTTACTCAAATTTGTTTGA
GACGCCTGCACCCTACTTGTGTTACGACCCAAGCAGGTTAAACCTCTCTGATCCC
>*Vollenhovia_emeryi*
GATCCTTCAATCAGGTATCAAAATGGGAAGTGATGACTGGGTTTAAAGCTCAAGATTATTT
GAGACTCCTGCGACCCGTAATTGTTACGACTATGCAGGCTAAACCTCTCTGATCCC
>*Linepithema_humile*
GATCCTTTCGAGGTATCAAATGGGAGATGATGAACGGGTTTTAACTCGAATTTTTCGAGACG
CCTGCCACCCTGTTGTTACGACGACAGCAGGTTAAACTCCTCTGATCCC
>*Ooceraea_biroi*
GATCCTTCAATGTCAGGAATCAAACCTGGGATATGATGAATAGGTTTTACCTCAAATTTGTT
GAGACGCCTGTAACCTTTTTTAAAGTTACGACATCGGCAGGCTAAACCGCTCTGATCCC
>*Camponotus_floridanus*
GATCCTTCAATCAGTTTTCAAATGGGAAATGATGAATGGGTTTTACTCAAATTTATTGAG

ACGCCTGCATCCCTTACTTTGTTACGACACTAGCAGGCAAACCTCTCTGATCCC
>Harpegnathos_saltator
GATCCTTCTTCAGTTTCAATGGGAAATGATGAATAGGTTGGCTCAAGTTATTTGAGACGC
CTGCACCTTACTTTGTTACGACTTTTCTGCAGGTTAAACCTCTCTGATCCC
>Wasmannia_auropunctata
GATCCTTCGATCAGGTATCAAAATGGGAGGTGATGAACGGGTTTTGACTCAAAATTTGTTT
GAGACGCCTGCACCCTAGTTGTTACGACGTTAGCAGGTTAAACCCCTCTGATCCC
>Acromyrmex_echinatior
GATCCTTCTCTTAGGTATCAAAATGGGATGTGATGAATGGGTTTACTCAAATTTGTTTGA
GACGCCTGCATCCTACTTGTGTTACGACTCAAGCAGGTAAACCTCTCTGATCCC
>Monomorium_pharaonis
GATCCTTCTCGAGGTATCAAAATGGGAGATGATGAATGGGTTTTGACTCAAAGTTATTTGA
GATGCCTGCACCCTGATTGTTACGACACAAGCAGGCTAAACTCTCTGATCCC
>Solenopsis_invicta
GATCCTTCTCGAGGTATCAAAGGGAGATGATGAATGGGTTTATCTCAAATTTATTTGAG
ACGCCTGCACCCTGACTGTTACGACGAGAGCAGGCTAAACCCCTCTGATCCC
>Pogonomyrmex_barbatus
GATCCTTCAATCAGGTATCAAGATGGGAGATGATGAACGGGTTTTGACTCAATATTATTTG
AGACGCCTGCACCCTGTTTGTGTTACGACGCAAGCAGGCTAAACCCATCTGATCCC
>Pogonomyrmex_barbatus
GATCCTTCAATCAGGTATCAAGATGGGAGATGATGAACGGGTTTTGACTCAATATTATTTG
AGACGCCTGCACCCTGTTTGTGTTACGACGCAAGCAGGCTAAACCCATCTGATCCC
>Cardiocondyla_obscurior
GATCCTTCTTTAAGGTGAAAATGGGATATGATGAACGGGTTTAGCTCGAATTTGTTTGA
ACGCCTGCAGCCTTTATTGTTACGACGATTGCAGGTTAAACCTCTCTGATCCC
>Bombus_terrestris
GATCCTTCTTTAAGGAAAATGGGAAATGATGATTAGGTTATCTCATTTTTCATATGAGA
AGCCTGCATTCTTACATACTTGTGTTACGCTTTATGCAGGTCAAACCTCTCTGATCCC
>Bombus_impatiens
GATTCTTCTTTAAGGAAAATGGGAAATGATGATTAGGTTATTCTCGTTTTCCATATGAGA
AGCCTGCATTCTTTACATACTTGTGTTACGCTTTATGCAGGTAAACCTCTCTGATCCC
>Cephus_cinctus
GATCCTTCTGTGATTACAGAACGGGGTGTGATGATATGGTTTTGCTCGTTCTGTTACGAGA
CGCCTGCGGCCACAATCCTTGTGTTACGATTATAAGGCAGGCTAAACCGCTCTGATCCC
>Trichogramma_pretiosum
GATCCTTCTGGATTACAGAACGGGAGATGATGACTGAGTAAGTCGCTCGTTACCCATACGA
GATGTCTGTTGCTCGGAACCTTGTGTTACGCAATCGGCAGACTATGACACTCTGATCCC
>Fopius_arisanus
GATCCTTCTGTGATTACAGAACGGGAGATGATGACTCGTTTTGGCTCGCAATTGCGAGATG
CCTGTTACGGCAATCCTTGTGTTACGATTATAAGGCAGGCAAACCTCTCTGATTCC
>Chelonus_insularis
GATCCTTCTCGGATTCAAACGAGAGATGATGATTGTTAGGACTCGTTACATACGAGAAG
CCTGCACGGAACTTGTGTTACGATTACGAGGCTAAACGGCTCTGATCTC
>Neodiprion_lecontei
GATCCTTCTGAGATTACAGAACGGGAAATGATGACGCGGTCGACCCTGTTTCATACGGGAC
GCCTGCACCGCAAACCTTGTGTTACGATAACGCAGGTCAAACCTCTCTGATCCC
>Orussus_abietinus
GATCCTTCTGTGATTACAGAACGGGGTGTGATGACTGGGTTGTTCTCGTTTTTCATACGAGA
CGCCTGCACCGGATTCTTGTGTTACGATTCTAGCAGGTCAAACCTCTCTGATCCC
>Diachasma_alloeuum
GATCCTTCTGTGATTACAGAACGGGAGATGATGACTCGTTTTGGCTCGCATTTGCGAGATG
CCTGCTACGGCAAACCTTGTGTTACGATCAAGCAGGCGAAACCTCTCTGATTCC
>Belonocnema_treatae
GATCCTTCTGTGATTACAGATGGGAGATGATGATTTGGTTTTAGCTCGTTTCGATACGAGAAG
CCTGCGCAACCAAACCTTGTGTTACGACAGTGCAGGCAAACCGCTCTGATCCC
>Vespa_mandarinia
GATCCTTCTGTTGATCAGAATGGGAAATGATGATCTGGTTTTCTCTCGTTACGAGATG
CCTGCACCATATTTGTTACGACTTTTGGCAGGTTGAACCTCTCTGATCCC
>Eufriesea_mexicana
GATCTTTCTAAAGTCAATGGGAAATGATGAAGAGGTTATTCTCATTCTTATATGAGAAGC
CTGCACAAATCTTATACTTGTGTTACAATTTAAAGCAGGTTAAACCTTTCTGATCCC

>Megalopta_genalis
GATTCTTCCCAATCAAAGGGATATGATGACTGGGTGAATCTCGTGCTGATACGATACGC
CTGCACCGTTTCATTTCTTGTTACGTTTAAAGCAGGTTAAACCCATCTGATCCC

>Nomia_melanderi
GTATCCTTCTTAGGTGTAAAAGGGATATGATGATTGGGATCTTCTCGTTTTTATACGAGA
TGCCTGCACTTTATACTTGTTACGATTAAAGCAGGTTAAACCTCTCTGATCCC

>Apis_mellifera
CATCCTTTTCGCATTCAATGGGAGATGATGAATATCGTGTTCTCATTCTCGATATGAGAG
GCCTGAAGTTTTTTTTTTGTTACGCACCTTTGTCAGGTTATACTATTCTGATCCC

>Apis_florea
CATCCTTCTAGCAGTCAAATGGGATATGATGAATGACGTGTTCTCATTCTAGATATGAGA
GGCCTGAAGTTTTTTTTTTGTTACGACATTTATTACAGGTTATAACCATTCTGATCCC

>Apis_dorsata
CATCCTTCCCTGCGATCAATGGGAGATGATGAATGTCGTATTCTCATTCTCGATATGAGAG
GCCTGAGTTTTTTTTTTCTTGTTACGTAGTTTGTGTCAGGTTATAACCATTCTGATCCC

>Apis_cerana
CATCCTTCCGATGTTTTAATGGGAGATGATGAATGTCGTGTTCTCATTCTCGATATGAGAG
GCCTGAAATTTTTTTTTTTGTTACGCAATTTGTCAGGTTAAACCTCTGATCCC

>Dufourea_novaeangliae
GATCTTTCCAGTTCAAATGGGATGTGATGATTGGGTAATTCTCGTTTTTATATGAGACGC
CTGCATGCTCTTAATTTTGTACGCTTAAGCAGGTTAAACCTTCTGATCCC

>Ceratina_calcarata
GATTCTTCCCTGAATTTCAAATGGGAAGTGATGATAAGGTTTTTCTCGTCCGAATACGAGA
AGCCTGCGGATTTTTATCCTTGTTACGTTCAACGCAGGTTAAACCTCTCTGATCCC

>Melipona_quadrifasciata
GATTCTTCTGTAAGGAAATGTGGGAAATGATGATCTAGGTTATGCTCATCTTTAATATGA
GAAGCCTGCATTTGTTATACTTGTTACATTTTATGCAGGTTAAACCTCTCTGATCCC

>Habropoda_laboriosa
GTTTCATATATTTCTTACATGGGAGGTGATGATTAGGTTACTCTCATCCAGATATGAGAAG
CCTGCTTTTCTAAATTTCTTGTTACGCTTTAAAGCAGGTTAAACCTTTCTGATCCC

>Copidosoma_floridanum
GATCCTTCTGGGATTGAGAACGGGAAATGATGACCGAGGTGAATCTCGTTTTGATACGAG
ACGTCTGTGCACTCGAACTTGTTACGCATATGGCAGACTATGACCCTCTGATCCC

>Microplitis_demolitor
GATCCTTCTGTGATTGAGAACAAGAAATGATGACTCGTGAGAGCTCGTCTGATACGAGAT
GCCTGCAACGGACATCCTTGTTACGATTTAGCAGGCCAAACATTTCTGATCTT

>Nasonia_vitripennis
GATCCTTCTGTGATTGAGAACGGGATATGATGACTGAGTCTAGCTCGTTCTATACGAGAT
GTCTGCGCGCTCGCAAACCTTGTTACGCTTACGGCAGACCATGACCCTCTGATCCC

>Polistes_canadensis
GATCCTTCTGTTGATCAGAATGGGACGTGATGATTTGGTTTTGCTCGTTCGATACGAGATG
CCTGCACCGACTCTTGTTACGACTTCCGGCAGGATAAACCTCTCTGATCCC

>Polistes_dominula
GATCCTTCAAAGGTCATATGGGAAATGATGATTTGGTTTTGCTCGTTGTTACGAGATGCC
TGCACCGACATTTGTTACGACTATTGCAGGTTAAACCTCTCTGATCCC

>Athalia_rosae
GATCCTTCTGAGATTGAGAACGGGAGGTGATGACTCGGGGCGGCTTGTGATTGATACGAG
ATGCCTACACCGCAAACCTTGTTACGATACTGTAGGTCAAACCCCTCTGATCCC

>Megachile_rotundata
GATCCTTCCGTTCAAAGGGATATGATGAATCGGTCTATCTTGTCTTTTACAAGAAGC
CTGCACCGTGTACGACGTGGCAGGTTAAACCTCTCTGATCCC

- **Archeplastida - A few representative species are listed**

>Syntrichia_caninervis
GATCCTTCTGGTTTTACAACATGATGAATGGACAATTCTCGCGTCTGCGTGACAGGTGCC
CGTTACATACTTGTTAAACAGGGCACCCCTCCATATCTGACCC

>Physcomitrella_patens
GATCCTTCTGGTTTTCAAATGATGAGAGGAGGAGTCCGCGTCTGCGGAGAGGTGCC

CGCCCCAATACCTTGTTACAAAGGGCAACCCCCCTGTCTGACCC
>Marchantia_polymorpha
GATTCTTCTGGTTTTTAAATATGATGAATGGCTTATTCTCGCGCTTGCAGAGAGAGGTGTC
CGCAGCACCTTGTTACGCATGGACACCGCTGTTCTGATCCTT
>Amborella_trichopoda
GATCCTTCAGGCAGAACGGGAAGATGAGGAGTGGTATTTTCTCACTTGTGAGATGAGTGT
CTGCCCCATGCCTGGTTACGACAAATGGGCACTCACATTCTAATCCTATCTGATTCC
>Nymphaea_colorata
GATCCTTCAGGCAGCAAGGGAAGGTGATGAGTGGGTCTAATCTCGCTGATAACATGCGGG
ATGGGTGTTGCCCCAACGCCTGGTTACGACGCTCGGGCACCCCAATCTTAACCATCTG
ATTCC
>Selaginella_moellendorffii
GATCCTTCTTGGTTTTACTGGGGGATGATGAGTGGTAAATGCTCGGTTTCTCCACTGTGA
GAGGGCTTGCCCGCCCAAATCATTGTTACGCAAGGGGAGTCTCCTATTTTGGCCTCT
GACCT
>Chara_braunii
GATCCTTCTTGGTTTTATGAAATGATGATGAGTGGTACTCACCTCCCTTTTGCATGTGA
GAGGTGGCCCTCACCCAGACTTGTTACGACCGGTGCGCCGCTAAGCCTATCTGACCCCT
>Arabidopsis_thaliana
GATCCTTCAGGCAAGTTAAAGGGGATATGATGAATGGTAAAACTCGCTTATATTGCGAG
AAGAGCGTTCCGCCCAAAGCCTGTTACGACTATTGGGCGCTCTCTTTTTTACACAATCTG
ATCCCT
>Oryza_Sativa_Japonica_Group
GATCCTTCAGGCAGTTAAAGGGAGATGATGAGTGGTAAACATCTCGCTAATTAATGTTGCG
AGAGGGGTGTCTGCCCAAACCAAGCCTGTTACGACAACAGGGCACTCCAATATTTTACCA
TCTGATCCC
>Theobroma_cacao_Matina_1.6
GATCCTTCAGGCAGTAAACGGGATATGATGAGTGGTCACTACTCACAAGCGCTGTGAGAT
GGGTGTCCGCCCAAACGCCTGGTTACGACAACATGGGCACTCCCCTCTGTACCGCTGAT
CCC
>Arabidopsis_halleri
GATCCTTCAGGCAAGTTAAAGGGGAAATGATGAATGGTAAAGTACTCGCTTCTATTGCGAG
AAGAGCGTTCCGCCCAAAGCCTGTTACGACAATTGGGCGCTCTCATTTTTTACACCGTCTG
ATCCC
>Beta_vulgaris
GATCCTTCAGGCAGTAAATGGAATATGAGGAGTGGTCAATTTCTCGCTTATATAAATTGCG
AGAGGGGTGTCCGCCCAAAGCCTGGTTACGACAATTGGGCATCCCCATTCTTAACCATCT
GATTCTT
>Brassica_oleracea
GATCCTTCAGGTTTTAAAGGGGAAGTGAAGATTGGGTTTTAATCTCGCATATTTTATTGCGT
GAAGAGCGTTCGCTCAAACCTTTTTACGACAATTGGGCGCTCCAATTTTCTTATCATCT
GATCCC
>Arabidopsis_lyrata
GATCCTTCAGGCAAGTTAAAGGGGAAATGATGAATGGTAAATACTCGCTTCTATTGCGAG
AAGAGCGTTCGCCCCAAGCCTGTTACGACAATTGGGCGCTCTCATTTTTTTACACAATCT
GATCCC
>Brassica_napus
GATCCTTCAGGTTTTAAAGGGGAAGTGAAGATTGGGTTTTAATCACGCATATTTAATTGTGT
GAAGAGCGTTCGCTCAAACCTTTTTACGACAATTGGGCGCTCCAATTTTCAATATATCTGATCC
GATCCC
>Brassica_rapa
GATCCTTCAGGTTTTAAAGGGGAAGTGAAGATTGGGTTATTCTCGCATATATGCGTGAAGA
GCGTTCGCTCAAACCTTTTTACGACAATTGGGCGCTCCAATTTTCAATATATCTGATCC
C
>Actinidia_chinensis
GATCCTTCAGGCAGTTTTAAAGGGGATGTGATGAGTGGTCAACTCGCAATAAAAAATGC
GAGATGGGTGCTACGGCACAAAGCCTGGTTACGACAATGCGGGCACTCTCATCTGAACC
AATCTGATCCC
>Corchorus_capsularis
GATCCTTCAGGCAGTAAACGGGATATGATGAGTGGTCACTACTCGCACGGATGCGAGAT
GGGTGTCCGCCACGCCTGGTTACGACAACCTGGGCACTCCCATTTTTTACCATCTGATCCC

>*Vigna_radiata*
GATCCTTCAGGCAGTTAACGGGATGTGATGAGTGGTAAATGCTTGCATATAACATGCAAG
AGGGGTGTCCGCCCCAGCCTGGTTACGACTATGGGGCACTCTCTTTTTTGACCTATCTGAT
CCC

>*Vigna_angularis*
GATCCTTCAGGCAGTTAATGGGATATGATGAGTGGTAAATGCTTGCCTTAATATGCAAG
AGGGGTGTCCGCCCCAGCCTGGTTACGACTATGAGGCACTCTCTTTTTTGACCTATCTGAT
CCC

>*Cynara_cardunculus*
GATCCTTCAGGCAAGTTTTAATGGGAAGTGATGAGTGGTTCGCTACTCGCATTTAATATGC
GAGAAGGGGTGTTTCGCCTAAAGCCTTACGACAATTGGCCACCTCCAATTTTTAACCATCT
GATCCCT

>*Glycine_max*
GATCCTTCAGGCAGTTAACGGGATATGAGGAGTGGTAAATCTCGCACAAAATATGCGAG
TGGGGTGTCCACCCAGCCTGGTTACGACACTAGGACACTCTCTTTTTTAACCAATCTGAT
CCC

>*Daucus_carota*
GATCCTTCAGGCAGATAATGGGATGTGATGAGCGGTGAGAATATTGCAATAAAAAATGCA
ATATGGGTGTCTAGCCCAAGCCTGTTACGACAACATAGGCGCTCTCGTTTTTAACCAATC
TGATCCC

>*Cucumis_sativus*
GATCCTTCAGGCAGTAACGGGATATGATGAGTGGTTGGCTCTCGCTAATAATAATGCGAGT
TGGGTGTCCGCCCCAAGCCTGGTTACGACAATTGGGCACTCCCATCTCTTCCAATCTGA
TCCC

>*Medicago_truncatula*
GATCCTTCAGGCAGAAACGGGATATGATGAGTGGTTGAGTCTCGCACATAATATGCGAGT
TGGGTGTTTCGCCCCAAGCCTGGTTACGACAATTGGGCACTCTCATTTAATCAATCTGATC
CC

>*Phaseolus_vulgaris*
GATCCTTCAGGTTTATTGGATGTGATGAGTGGTAAAGACTTGCATACAATATGCAAGAGG
GGTGTCCGCCCCAGCCTTGTACGACTGGGGGCACTCAATCTTTAGACCAATCTGATCCT

>*Lupinus_angustifolius*
GATCCTTCAGGCAGTTAACGGGATATGATGAGTGGTTAAATGTGCAATAAAAAATGCGA
GTTGGGTGTTTCGCCCCAAGCCTGGTTACGACAATTGGGCACTCCCATCTTATTATTCCAATCTG
ATCCC

>*Prunus_persica*
GATCCTTCAGGCAGAAACGGGATGTGATGAGTGGTCAGTATTCGCAATAATTATGCGAGA
GATGGGTGTCCGCCCCAAGCCTGGTTACGACAATTGGGCACTCCCATTAATCTTTGTACCATC
TGATCCC

>*Helianthus_annuus*
GATCCTTCAGGCAAGTTTTGAAGGGAGATGAGGAGTGGTCAATGCTTGCATTTAAATATG
CAAGATGGGGTGTCCGCCCCAAGCCTGTTACGACAATTGGGCACTCCATTTTTAAACCA
TCTGATCCC

>*Manihot_esculenta*
GATCCTTCATGGCAGTTAACGGGATGTGATGAGTGGTCATTTTTTCGCGATTAGCTTTGCG
AGATGGGTGTCCGCCCCAATGCCAGTTACGACAATACGGGCACTCCCATTAATACCAATC
TGATCCC

>*Solanum_tuberosum*
GATCCTTCAAGCAAGTTTTAATGGGATATGATGAGTGGTCATTACTCGCAATAAAAAATAAA
ATATGCGAGAAGAGCGTTCCAGCCCAGCCTGGTTACGACAATTTGGGCGCTCTCATTTTTT
ACCAATCTGATCCC

>*Prunus_dulcis*
GATCCTTCAGGCAGAAACGGGATGTGATGAGTGGTCAGTATTCGCAATAATTATGCGAGA
GATGGGTGTCCGCCCCAAGCCTGGTTACGACAATTGGGCACTCCCATTAATCTTTGTACCATC
TGATCCC

>*Solanum_lycopersicum*
GATCCTTCAGGCAAGTTTTAATGGATGTGATGAGTGGTCATAACTCGCAATAAAAAATGCG
AGAAGAGCGTTCCAGCCCAGCCTGGTTACGACAATTTGGGCGCTCTTAATTTTTACCAATC
TGATCCT

>*Theobroma_cacao_Belizian_Criollo_B97-61_B2*
GATCCTTCAGGCAGTAAACGGGATATGATGAGTGGTCATTACTCACAAGCGCTGTGAGAT

```
GGGTGTCCGCCTCAACGCCTGGTTACGACAACATGGGCACTCCCCTCTGTACCCCTGAT
CCC
>Capsicum_annuum
GATCCTTCAGGCAAGTTTTAATGGGATATGATGAGTGGTCATTACTCGCAATTTAAAAAT
ATGCGAGATGGGCGTCCAGGCTCAGCCTCGTTACGACATTTTGGGCGCTCTCATTTTTAA
CCAATCTGATCCC
>Prunus_avium
GATCCTTCAGGCAAGAAACGGGATGTGATGAGTGGTCAGTATTCGCAATAATTATGCGAGA
GATGGGTGTCCGCCCAAAGCCTGGTTACGACAATTGGGCACCCTAATCTTTGTACCATC
TGATCCC
>Nicotiana_attenuata
GATCCTTCAGGCAAGTTTTAAAGGGGATATGATGAGTGGTCAATGCTCGCATTTTAAATATG
CGAGAAGGGCGTCCAGCTCAGCCTGTTACGACAATTTGGGCGCTCTCATTTTACCAATCT
GATCCC
>Gossypium_raimondii
GATCCTTCAGGCAAGAAACGGGATATGATGAGTGGTCATTACTCGCAAGCGTTGCGAGAT
GGGTGTCCGCCCAAAGCCTGGTTACGACACTTGGGCACTCCCATTTTTTACCATCTGATC
CC
>Olea_europaea_var._sylvestris
GATCCTTCAGGCAAGTTTTACGGGGATGTGATGAGTGGTCGACACTCGCACTGAACATGC
GAGATGGGTGTCCAGCCTAAGCCTGGTTACGACAATATGGGCACTCTCATTTTTACCAAT
CTGATCCC
>Malus_domestica
GATCCTTCAGGCAAGAAACGGGATGTGATGAGTGGTCATAATTCGCAAATTTAAATGCGAG
AGATGGGTGTCTGCCCAAAGCCTGTTACGACAATTGGGCACTCTCCTGTTTTTACCATCT
GATCCC
>Pistacia_vera
GATCCTTCAGGCGGTTTTAACGGGGATGTGATGAGTGGTCATCACTCGCTATAACTTTGCG
AGATGGGTGTCCGACCTATGCCTGGTTACGACAATCTGGGCACTCCCATCTTTACCAATC
TGATCCC
>Populus_trichocarpa
GATCCTTCAGGGTTTTAACGGGATATGATGAGTGGTCAATGCTCGCATATAAACATGCGAG
ATGGGTGTCCAGCCCACTCCCTGGTTACGACAACAGGGCACTCCCATCTTACCTATCTG
ATCCC
>Citrus_clementina
GATCCTTCAGGCAGTAACGGGGATGTGATGAGTGGTCATCGCTCGCTAAAAGTCTGCGAG
ATGGGTGTCCGCCCAAAGCCTGGTTACGACAATCTGGGCACTCCCATCTTTACCGATCT
GATCCC
>Coffea_canephora
GATCCTTCAGGCAGTTTTACGGGGATGTGATGAGCGGTCAAAGCTCGCATTTTAAACATGCG
AGATGGGTGCCCGGCCCAAAGCCTGGTTACGACAACCCGGGCGCTCCCATATTTACCGAT
CTGATCCC
>Ipomoea_triloba
GATCCTTCAGGCAAGTTTTAAGGGGATGTGATGAGTGGTTAATGCTCGCAGTTTGTGCGA
GAAGGGTGCCCGCCTTGCTGTTACGACATATTGGGCATTCCAATGTTTTTACCATCTGAT
TCC
>Vitis_vinifera
GATCCTTCAGGCAGTTAACGGGATATGATGAGTGGTCAACGTTTCGCAACAAATATGCGAG
ATGGGTGTCCGCCCTAGCCTGGTTACGACAATTGGGCACTCTCATCTTTTACCAATCTGAT
CCC
>Trifolium_pratense
GATCCTTCAGGCAGAAACGGGGATATGATGAGTGGTCTGAGTCTCGCAATAAAAAATGCGA
GTTGGGTGTTTCGCCCAAAGCCTGGTTACGACAATTGGGCACTCTCCTTTAATCAATCTGA
TCCC
>Brachypodium_distachyon
GATCCTTCAGGCAGCAAAGGGAAATGATGAGCGGTAACGGCTTGCAAATTTCACTGCGAG
AGGGTGCCCGCCCAACTAAGCCTGTTACGACAATGGGGCACCCCAATGATTTACCATCT
GATCCC
>Eragrostis_tef
GATCCTTCAGGCAGTCAAAGGGATATGATGAGTGGTAAAAGCTCGCAAATTTATCTGCGA
GAGGGGTGCTGCCCAAATAAGCCTGTTACGACAACAGAGCACTCACATGCTTACCA
```


TCTGATCCC

>*Triticum_dicoccoides*

GATCCTTCAGGCAGTAAAGGGAGATGATGAGTGGTATCAGCTCGCTACAATGCGAGAGG
GGTGTCCGCCCAACCAAGCCTGTTACGACAATGGGGCGCCCCAGTGACTTACCATCTGA
TCCC

>*Oryza_nivara*

GATCCTTCAGGCAGTAAAGGGAGATGATGAGTGGTAAACATCTCGCTAATTAATGTTGCG
AGAGGGGTGTCTGCCCCAACCAAGCCTGTTACGACAACAGGGCGCTCCAATATTTTACCA
TCTGATCCC

>*Oryza_glumipatula*

GATCCTTCAGGCAGTAAAGGGAGATGATGAGTGGTAAACATCTCGCTAATTAATGTTGCG
AGAGGGGTGTCTGCCCCAACCAAGCCTGTTACGACAACAGGGCACTCCGATATTTTACCA
TCTGATCCC

>*Oryza_glaberrima*

GATCCTTCAGGCAGTAAAGGGAGATGATGAGTGGTAAACATCTCGCTAATTAATGTTGCG
AGAGGGGTGTCTGCCCCAACCAAGCCTGTTACGACAACAGGGCACTCCAATATTTTACCA
TCTGATCCC

>*Oryza_longistaminata*

GATCCTTCAGGCAGTAAAGGGAGATGATGAGTGGTAAACATCTCGCTAATTAATGTTGCG
AGAGGGGTGTCTGCCCCAACCAAGCCTGTTACGACAACAGGGCACTCCAATATTTTACCA
TCTGATCCC

>*Oryza_barthii*

GATCCTTCAGGCAGTAAAGGGAGATGATGAGTGGTAAACATCTCGCTAATTAATGTTGCG
AGAGGGGTGTCTGCCCCAACCAAGCCTGTTACGACAACAGGGCACTCCAATATTTTACCA
TCTGATCCC

>*Oryza_meridionalis*

GATCCTTCAGGCAGTAAAGGGAGATGATGAGTGGTAAACATCTCGCTAATTAATGTTGCG
AGAGGGGTGTCTGCCCCAACCAAGCCTGTTACGACAACAGGGCACTCCAATATTTTACCA
TCTGATCCC

>*Oryza_brachyantha*

GATCCTTCAGGCAGTAAAGGGAGATGAGGAGTGGTAAATCTCGCTATTTTCCGCTGCG
AGAGGGGTGCCCCGCCCAACCAAGCCTGTTACGACAATTGGGGCATCCCCGTGTCTACCA
TCTGATCCC

>*Oryza_punctata*

GATCCTTCAGGCAGTAAAGGGAGATGATGAGTGGTAAACATCTCGCTAGTTATGTTGCGA
GAGGGGTGTCTGCCCCAACTAAGCCTGTTACGACAACAGGGCACTCCAGTCATTTACCAT
CTGATCCC

>*Oryza_sativa_Indica_Group*

GATCCTTCAGGCAGTAAAGGGAGATGATGAGTGGTAAACATCTCGCTAATTAATGTTGCG
AGAGGGGTGTCTGCCCCAACCAAGCCTGTTACGACAACAGGGCACTCCAATATTTTACCA
TCTGATCCC

>*Oryza_rufipogon*

GATCCTTCAGGCAGTAAAGGGAGATGAGGAGTGGTAAACATCTCGCAACTTTTACTGCGA
GAGGGGTGCCCCGCCCAACCAAGCCTGTTACGACAACAGGGGCATCCCCATCATTTTAC
CATCTGATCCC

>*Ananas_comosus*

GATCCTTCAGGCAGTAAACGGGATGTGATGAGCGGTAACAGCTCGCATTCTGCGAGAGGG
GCACCCGCCCCATAAGCCTGTTACGACAAACAGGGGCGCTCTCGTAATTTACCAATCTG
ATCCC

>*Saccharum_spontaneum*

GATCCTTCAGGCAGTAAAGGGAAATGATGAGTGGTAAAGCTCGCGTTTTTTCAATTGC
GAGAGGGGTGCTGCCCCAAATAAGCCTGTTACGACAACAGGGGCGCTCACACTGTTTACC
ATCTGATCCC

>*Panicum_hallii_FIL2_Phallii*

GATCCTTCAGGCAGTCAAAGGGAAATGAGGAGTGGTATAAGCTCGCAATTTATTATGCGA
GAGGGGTGCTGCCCCAACTAAGCCTGTTACGACAACAGGGGCACCCCAATGAATTTACC
ATCTGATCCC

>*Hordeum_vulgare*

GATCCTTCAGGCAGCAAAGGGAGATGATGAGCGGTGTCAGCTCGCTATATGCGAGAGGGG
TGTCCGCCCAACTAAGCCTGTTACGACAATGGGGCACCCCAATGACTTACCATCTGATC
CC

```
>Sorghum_bicolor
GATCCTTCAGGCAGTTAAAGGGAAATGATGAGTGGTAGAAGCTCGCGTTTTTTTATTGCG
AGAGGGGTGCCTGCCCCAAATAAGCCTGTTACGACAACAGGGCACTCACACTGTTTACCA
TCTGATCCC
>Triticum_turgidum
GATCCTTCAGGCAGCAAAGGGAGATGATGAGTGGTATCAGCTCGCTACAATGCGAGAGG
GGTGTCCGCCCCAACCAAGCCTGTTACGACAATGGGGCGCCCCAGTGACTTACCATCTGA
TCCC
>Aegilops_tauschii
GATCCTTCAGGCAGCCAAAAGGGATGTGATGAGTGGTATCAGCTCGCTATAATGCGAGAG
GGGTGTCCGCCCCAACTAAGCCTGTTACGACAATGGGGCGCCCCAGTGACTTACCATCTG
ATCCC
>Triticum_urartu
GATCCTTCAGGCAGCAAAGGGATGTGATGAGTGGTATCAGCTCGCTATGTGCGAGAGGG
GTGTCCGCCCCAACTAAGCCTGTTACGACAATGGGGCACCCCAGTGACTTACCATCTGAT
CCC
>Panicum_hallii_HAL2
GATCCTTCAGGCAGTCAAAGGGAAATGAGGAGTGGTATAAGCTCGCAATTTATTATGCGA
GAGGGGTGCCTGCCCCAACTAAGCCTGTTACGACAACAGGGGCACCCCAATGAATTTACC
ATCTGATCCC
>Musa_acuminata
GATCCTTCAGGCAGTAAACGGGATGTGATGAGTGGTAAAAGCTCGCGCTGTTTGCGTGAA
GAGCAGCCGCCAGAGCCTGTTACGACAACAGGGCCTGCTCCCATCTTTACCCTTCTGAT
CCC
>Triticum_aestivum
GATCCTTCAGGCAGCAAAGGGATGTGATGAGTGGTATCAGCTCGCTATAATGCGAGAGG
GGTGTCCGCCCCAACTAAGCCTGTTACGACAATGGGGCGCCCCAGTGACTTACCATCTGA
TCCC
>Leersia_perrieri
GATCCTTCAGGCAGTTAAAGGGAGATGAGGAGTGGTAAACAGCTCGCCATTTTTACTGCGA
GAGGGGTGCCCGCCCCAACCAAGCCTGTTACGACAACAGGGCATTCCCTTCATCTTACCA
TCTGATCCC
>Zea_mais
GATCCTTCAGGCAGTTAAAGGGATATGATGAGTGGTAAAACTCGGTTTTTTTTATTCCGA
GAGGGGTGCCTGCCCCAAATAAGCCTGTTACGACAACATGGCGCTCACATTTTTTACCAT
CTGATCCC
>Eragrostis_curvula
GATCCTTCAGGCAGTCAAAGGGATATGATGAGTGGTAAAAGCTCGCAAATTTATCTGCGA
GAGGGGTGCCTGCCCCAAATAAGCCTGTTACGACAACAGAGCACTCACATGCTTCACCA
TCTGATCCC
>Setaria_italica
GATCCTTCAGGCAGTTAAAGGGACATGATGAGTGGTAAAACTCGCGATTTTTACTGCGAG
AGGGGTGCCTGCCCCAAATAAGCCTGTTACGACAACAGGGCACTCACAGCGTTTACCATC
TGATCCC
>Dioscorea_rotundata
GATCTTTCAGGCGGTCCAATGGGAGGTAATGAGTGGTAAAAGCTAGCGCTTTGATGCGTG
ATGGGTGTGCAAAGCCTGTTACGACAATGAGCACACTCTAATCTTTGCCAATATGATCCC
```

- **Fungi (budding yeasts)** – See also Sharma et al (2017) PLoS Genet

```
>S_cerevisiae
ATGACCTTCCAAGTTTTTAAAAGAATACGATGATATTATTTGCGTTTTCAAATCGAACAAT
TCTTCTCGGAGCGATCTGAGGTTTTAATGGAGATAGCGGTTCTGCGCAACCCATTGATC
TTGTTACATTCTTAAGAATGACAAGGACGCTTTTATAAAATTCTGATTCTTT
>K_lactis
ATGACCTTCCAAGTTTTTAAAAGAAGACGATGATACTATTTGCGCTTTTTTAATTCTTACA
ATTCTTTCGTGGATGGGAATTATTTTCTGTTTCATGGTTTTAATGGAGAATGTGGTGCCCT
CGCAAGCCTTATAATCTTGTTACATTCTTTGAATGATTTGGGACACTTTTTTATAAAATTC
TGATTCTTT
>C_glabatra
```

ATGACCTTCCAAGTTTTATTAAGAAGACGATGAAATTATTTGCGTGAATCCTATTTCGAACA
ATTCTTCTTGCATCCTATCGTGAGGTTTTAATGGAGAAAGCGGTACCAACGCAGACCTT
ATTTCTTGTTACATCCCTAGTTTCTTTTCGAGGAGCTAGTTTTCTGGACGCTACTAGAAAA
TTCTGATTCTT
>S_eubayanus
ATGACCTTCCAAGTTTTTAAAAGAACACGATGATACTATTTGCGTTTTCAAATCGAACAAT
TCTTCTCGGAGCAATCTGAGGTTTTAATGGAGATAGCGGTTCTGCGCAGCCAATTGATC
TTGTTACATTCTTAAGAATGACAAGGACGCTTTTATAAAAATTCTGATTCTTT
>C_albicans
ATGACCTTCCAAGTTTTTATACTTACATGATGAAATCTCGCGTTATATTCAACGATTTCGGC
CCTTCTGTTGACTGGTTTTCTTTTATGAGAAGTTGAGTCTTTTTTCGGAGGATTCTTTTTT
GTAAGTTCATGGGAAACTGGTTAGTAGAAAATGACTTGACCGTCAGGGTAATGAAGTTG
ATGAAGGTGTCTGGCGCAACCATTATCTTGTACATATATTTTCAGACGCTATAGAGAACT
TCTGACCTA
>A_gossypii
ATGACCTTCCAAGTTTTTATAAGAAGTGGATGATATGATTTGCGTAGCAACTCGAACAAT
TCCTCGCCTGTGATCAGGCTGAGGTTTTAATGGAGAAAGTGGTGCCCGCAGCAAATAA
TCTTGTACATTCTTTGAATGCCTGGGACACTTGCTATAAAAATTCTGATTCTT
>L_fermentati
ATGACCTTCCAAGTTTTTATAAGAAGAAGATGAAATTATTTGCGTTTTCAATTCGAACAAT
TCTTCTCAGTGTAGCTTAGCTGTACTTTGAGGTTTTGATGGAGAATGTGGTTCTGCGCAG
CTTTTTAATCTTGTACATTCAATGAATGCTCAAGGACACTTTTTTAGAAAATTCTGATTC
TTT
>S_passalidarum
ATGACCTTCCAAGTTTTTATACTTACATGATGAAATCTCGCGTTTTTAAATCATTTCGACAAT
TCCCTTTCGAGGTTGAGTTGTTCCCTTCGGGAACGATTCTTTGGAAAATGCTTCTTTCC
AGGAGCTGGGTTTTAATGGTGCAGGTGTCCGGCGCAACAATTCTTGTACTAAATTTACG
GACGCTGAGGAGAGCTTCTGACCCC
>C_lusitaniae
ATGACCTTCCAAGTTTTTATAACAAATGATGAAGTCTCGCGTGCATCTATACTATTCCCTC
TGGAGTTGCAGTCGGAGCTTTCGGGCTCTGACTCCTCCGGAATGCTTTTTCCGTACGGGA
AAGTTGGGTTTGTGTGGTGAAGTGTGGCGCAACCTTTCTTGTACATTTGACAAAACA
CTGAAAGAATTTCTGATCCAA
>C_jadinii
ATGAACCTTCCATAGTTTTTAACTAACGTGATGAAATCTCGCGTCTATCTCTTCTAACGA
TTCTTCCCTCGGAGGTGTGAAAGCACATTTTCGAGAATGGATGCCTTTTAGGCGTTCTGGG
TTTCGATGGGGATGGGCATTCGACGCAACCTTTTCTTGTACATAACTATTTGGATGCTC
TAAGGAGAAGTTCTGATTTAC
>A_rubescens
ATGACCTTCCAAGTTTTTATACTATGATGAAACTTAGCACTTCTGAAACTCTGACCGATT
CCCTTTCCTATGAAATTGCCGGTGTCTCAGCGCCGGCTTGGGTTTTGAATGGAGTCAACG
CTTGGTGCACAATTCTTGTACAACTTAAAGCGTTTGTCTATTTCTGATCCTT
>W_anomalus
ATATCCTTCCAAGTTTTTATACTATATGATGAAGTCTCGCGTTTTCTTCTAACAAATCCCC
AAGGTTGAGTACTGGCAACGTTATTCTTTGAATGGTGTGGTTTTAAATCATATCTGGG
TTTTAATGGAGATGGTGTTCGGCGCAAATTTATCTTGTACAAATTAATTATTAGGACGC
CTAAATGAGAAATTCTGACCCCT
>S_stipitis
ATGACCTTCCAAGTTTTTATACTCATATGATGAAATCTCGCGTTTCTATCCATGACAGATC
CCTCTGGAGTTGAGTTGAAGTTCTGCTTCGATTCTCCGGAACGGCCTCTTGTACGAGA
GACCTGGGTTCTACTGGTGTAGAGCGTTTGGCGCAACCATTCTTGTACATAAAGAAAAGA
CGCTGAAAGAATTCTGACCCC
>D_hansenii
ATGACCTTCCACGTTTTTATACTCATATGATGAAATCTCGCGTTATTCAAGACAATTCCCTTT
AGAGTTGTAGTCTTTCTTTTCGGGAGAGATTCTCTGAAATTGCTTTCTCATACGAGAGAG
TAGGGTTTTATTGGTTAAGGTGTTTGGCGCAAATCTTGTACAAAAATATCACAGA
CGCTTAAGAGAATTCTGAACCTC
>B_inositolovora
ATGAACCTTCCAAGTTTTTATACTCATATGATGAAGTCTCGCGTTTCTCATGACAGCTGG
CCCCTCTCCTGAAGTTGAGTTACCTCACGGTATTCTTTGGGGGGAAATTTAGTCCTTG
TATGAGGGCTAGGGTGCTAATGGTGTGATTTCGGCGCACAAACCATAATCTTGTACAA

TTTTATTTAACGGATGCTTTGGAGAATTTCTGACCCC

>P_tannophilus

ATGACCTTCCAAGTTTTATTTACATGATGAAATCTCGCGTACTCCATATTATGACTTTTC
CCTTCTTCTTTTTTTTTATTCCGTATTTCAAGCTGGTTTTGTTTTCGCCTTACACAGGCTTGA
GACAAGATTAGATTTGGAAGGAATTTCTATATGAAGAAAAATTTCTTTTCTTATGGGA
AGGAAAATGGGTTTGTAGTGGGTAATGGCGATTGGCGCAAACCTTTCTTGTACACAATT
AATTTTTACAGTCGCTATAAATGAGATGTTCTGACCCCTT

>O_polymorpha

ATTATCCTTCCCTAGTTTTTACTAAGAAATGATGAAAACCTAGGCGCTGTTTTTTAACACAC
CCGGTTTTCCCTCCTTCGGGAATAAGTTGTTCTTCGGGATGGCTTTGGGAAAGCGTGGGTG
AAAGCGTTTGGCGCAAATTTATCTTGTACAAACATTTGTTGCAAGACGCTTTTGTACCTT
CTGAATT

>P_membranifaciens

ATGACCTTCCCTAGTTTTTACTATAACATGATGAAAACCTAGGTGTTGTCTGTTGACACATC
CGGCATCCCTCCTTCGGAAGTGTCTTTATGGCAGCCGGGATGGTATGGGTGACTGCGT
TTGGCACAATTTCTTGTACATAAAAACGACGCTTTCATTTTGTACATTCTGACCCC

>L_elongisporus

ATGACCTTCCAAGTTTTTATACTACATGATGAAATCTCGCGTTTATTCATTCATTGATTAA
AGTCCCTTTCTCAAAGTTGTAGTCGGGGTCTTCTTTTCGTTTGTAGAGGGGGCTCTTGATTC
CTTTGAGAGATGTGTCTCTTGACGGAGGCTGGGTATTTAAATGATGAAAGGTGTCTGGCG
CAACCTTATCTTGTACATTTTACAGACGCTTAAAGAGAACTTCTGACCCCT

>T_phaffii

ATGACCTTCCAAGTTTTTAAAAGGATATGATGAAGTTATTTGCATTTTTAATCGAACAAAT
TCTTCTCAATATAATTGAGGTTTTAATGGAGATTGCGGTGCCTGTGCAGAAAATTATCTT
GTTACATTCAATTGAATGAATCAGGACGCTATATTAGAAAATTCTGATCCTT

>Y_lypolytica

ATGACCTTCCACGTTTTTAAACGTTTTATGATGAAAATTATGCGTGCTAGCTGATTTGCCCC
CTCCTCGGAGGTAGTCCTTGACCGGACGGCGGCATCTTATGTTGATAGCGTTTGGCGCAC
TTACCTTGTTTTAAACAAACGCGGTAGTACTTTCTGATCCTT

>N_fulvescens

ATGACCTTCCACGTTTTTATAATGTGTATGATGAAAACCTACGCGTTTATCAACTTAGCTT
GCCCCCTGGTTCGAGTCTACTACGTTGATAATTTGGAGGAAAGGAACCTGTTTTTTTTTAC
TAGGTTATTGACCGGTGGAATCCTGGGCGTCCCTGACCGGACTGGGCGTAGCATGGTGAT
AGCGTTCGGGCGCAACATTACCTTGTTTTTACGAACGCTTTTGTACTTTCTGATCCCT

>L_starkeyi

ATGACCTTCCAGCGTTTTACGTTGTACATGATGAAAACCAGGCGTTAGCTGTTCGGATATG
TGTCGGGTGATTTAGTTTCGCTTTGCGACCGTATGTATGTGATGACGGCGTTTCGACGCA
ACCTCTTGTTACGTTACCAGGACGCTTTTGGTACCTTCTGATTTTTT

>E_cymbalariae

ATGACCTTCCAAGTTCTATAAAAAGGTGATGATATTATTTGCATTTCTAAATCGAACAAAT
TCTTCTCGTGAGCTTTATGCACCTCGAGGTTTTAATGGAGAATGTGGTGCCTGCGAGCC
TTATAATCTTGTACATTCAATTGAATGCCAAGGGACACTATAATCGTTAGATAAATCTGA
TTTTT

>V_polyspora

ATGAACCTTCCAAGTTTTTAAAAGAGTGCAGTACTATTTGCATTTATTAAATCGAAC
AATTCTTCTCGGTGGACTAGATAATATCTAATTTACTGAGGTTTTAATGGAGAATGCGGT
TCCTATGCAGCCAATATAATCTTGTACATTCAATTGAATGACAGGACGCTAATATTATG
AAATTCTGATTCTT

>T_blattae

ATGACCTTCCAAGTTTTTATAAGAATACGATGATATTATTTGCGTAATAAGCTAACAAAT
CTTCTCGTGGGTTTTCGGCTCTGTGAGGTTTTAATGGGGAATGCGGTTCCCGCGCAGCCA
ACCAATCTTGTACATTTTTTTTTAAAATGATTTTTTGGGACGCTACAATATAAAAATTCTG
ATTCTTT

>L_thermotolerans

ATGACCTTCCAAGTTTTTATAAGAAGTGGATGATATGATTTGCGTAGCAACTCGAACAAAT
TCCTCGCCTGTGATCAGGCTGAGGTTTTAATGGAGAAAGTGGTGCCCGCGCAGCAAATAA
TCTTGTACATTCTTTGAATGCCTGGGACACTTGCTATAAAAATTCTGATTCTTT

>T_delbrueckii

ATGACCTTCCAAGTTTTTATAAGAAGTGGATGATATGATTTGCGTAGCAACTCGAACAAAT
TCCTCGCCTGTGATCAGGCTGAGGTTTTAATGGAGAAAGTGGTGCCCGCGCAGCAAATAA
TCTTGTACATTCTTTGAATGCCTGGGACACTTGCTATAAAAATTCTGATTCTTT

bioRxiv preprint doi: <https://doi.org/10.1101/2021.11.30.470322>; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

>Z_rouxii

```
ATGACCTTCCAAGTTTTAAAAGAAGAAGATGAAATTATTTGCGTCTATTAAATCGTACAA
TTCTTCCCTCGGGTTTTAAAGCTCGTGGGGTTTTAATGGTGAAGACGGTTCCCGCGCAGC
AATTTTTCTTGTTACACCAATTTAATGGTTTTATAGGGACGTCTCTAGGAAATTCTGAT
TCTTT
```

Supplementary data S2

>H_sapiens

MRQLQKKIKNGTLNIKQDDPFELFIAATNIRYCYYNETHKILGNTFGMCVLDQDFEALTPN
LLARTVETVEGGGLVIVILLRTMNSLKQLYVTMTMDVHSRYRTEAHQDVVGRFNERFILSLA
SCKKCLVIDDQLNILPISSHVATMEALPPQTPDESLGPSDELRELKESLQDTQPVGVLV
DCCKTLDQAKAVLKFIEGISEKTLRSTVALTAARGRGKSAALGLAIAGAVAFGYNSNIFVT
SPSPDNLHRTLFEFVFKGFDALQYQEHLDYEIIQSLNPEFNKAVIRVNVFREHRQTIQYIH
PADAVKLGQAEVLVIDEAAAIPPLVKSLGPLYLVMFMASTINGYEGTGRSLSLKLIQQLR
QQSAQSQVSTTAENKTTTTARLASARTLYEVSLQESIRYAPGDAVEKWLNLDLCLDCLNI
TRIVSGCPLPEACELYVNRDTLFCYHKASEVFLQRLMALYVASHYKNSPNDLQMLSDAP
AHHLFCLLPPVPPTQNALPEVLAVIQVCLEGEISRQSILNSLSRGKKASGDLPWTVSEQ
FQDPDFGGLSGGRVVRIVVHPDYQGMGYGSRALQLLQMYEGRFPCEEKVLQETPQEIHT
VSSEAVSLLEEVITPRKDLPLLLKLNERPAERLDYLGVSYGLTPRLLKFWKRAFVPVY
LRQTPNDLTGEHSCIMLKTLDDEDEADQGGWLAAFWKDFRRRFLALLSYQFSTFSPSLAL
NIIQNRNMGKPAQPALSREELEALFLPYDLKRLEMYSRNMVDYHLIMDMI PAISRIFYLN
QLGDLALSAAQSALLLGLGQHSVDQLEKEIELPSGQLMGLFNRIIRKVVKLFNEVQEK
AIEEQMVAAKDVVMEPTMKTLSDDLDEAAKEFQEKHKKEVGLKLSMDLSEYIIRGDDEEW
NEVLNKAGPNASIIISLKSDDKRRKLEAKQEPKQSKKLNRETKNKKDMKLRKK

>M_musculus

MNRKKVDNRIRILIEINGVAERQSLFVVVVGDRGKDQVVILHHMLSKATVKARPSVLWCYK
KELGFSSHRKKRMRFSHNKIMSGTLYLKQDDPFELFVAATNIRYCYYNETHKILGNTFGM
CVLQDFEALTPNLLARTVETVEGGGLVIVILLRTMNSLKQLYVTMTMDVHSRYRTEAHQDVV
GRFNERFILSLASCKKCLVIDDQLDILPISSHVASIEALPPQAPDENLSPAALLELLELKE
SLQDTQPVGVLDCCCKTLDQAKAVLKFIEGISEKTLRSTVALTAARGRGKSAALGLAIAG
AVAFGYNSNIFVTSPSPDNLHRTLFEFVFKGFDALQYQEHLDYEVQSLNPEFNKAVIRVNV
FREHRQTIQYIHPADAVKLGQAEVLVIDEAAAIPPLVKSLGPLYLVMFMASTINGYEGTG
RSLSLKLIQQLRQQSAQSQVSTTAENKTTTTARLASARTLHEVSLQESIRYAPGDAVEKW
LNLDLCLDCLNITRIVSGCPLPEACELYVNRDTLFCYHKASEVFLQRLMALYVASHYKN
SPNDLQMLSDAPAHHLFCLLPPVPPTQNALPEVLAVVQVCLEGEISRQSILNSLSRGKKA
SGDLPWTVSEQFQDPDFGGLSGGRVVRIVVHPDYQGMGYGSRALQLLQMYEGRFPCEE
EKVLETPQEIHTVSSEAVSLLEEVITPRKDLPLLLKLNERPAERLDYLGVSYGLTPRLL
KFWKRAFVPVYLRQTPNDLTGEHSCIMLKTLDDEDEAEQGAWLAAFWKDFRRRFLALLS
YQFSTFSPALSLNIIQNRNVAKSALPALGREHLEALFLPYDLKRLEMYSRNMVDYHLIMD
LIPAIRSRYFLNQLGDLALSAAQSALLLGLGQHSVDQLEKEIELPSGQLMGLFNRIIR
KVVKLFNDVQEKAIIEEQMVAVKDVVMEPTMKTLSDDLDEAAKEFQEKHKKEVGLKDMDL
SQYVIRGDDEEWNEVLSKAGQNASIVSLKSDDKRRKLETKQEPKQS

>R_novergicus

MHRKKVDNRIRILIEINGVSEQRSLFVVVVGDRGKDQVVILHHMLSKATVKARPSVLWCYK
KELGFSSHRKKRMRQLQKKIKSGTLNIKQDDPFELFVAATNIRYCYYNETHKILGNTFGM
CVLQDFEALTPNLLARTVETVEGGGLVIVILLRTMNSLKQLYVTMTMDVHSRYRTEAHQDVV
GRFNERFILSLASCKKCLVIDDQLNILPISSHVASIEALPPQAPDENLSPAALLELQELKE
SLQDTQPVGVLDCCCKTLDQAKAVLKFIEGISEKTLRSTVALTAARGRGKSAALGLAIAG
AVAFGYNSNIFVTSPSPDNLHRTLFEFVFKGFDALQYQEHLDYEVQSLNPEFNKAVIRVNV
FREHRQTIQYIHPADAVKLGQAEVLVIDEAAAIPPLVKSLGPLYLVMFMASTINGYEGTG
RSLSLKLIQQLRQQSAQSQVSTTAENKTTTTARLASARTLHEVSLQESIRYAPGDAVEKW
LNLDLCLDCLNITRIVSGCPLPEACELYVNRDTLFCYHKASEVFLQRLMALYVASHYKN
SPNDLQMLSDAPAHHLFCLLPPVPPTQNALPEVLAVVQVCLEGEISRQSILNSLSRGKKA
SGDLPWTVSEQFQDPDFGGLSGGRVVRIVVHPDYQGMGYGSRALQLLQMYEGRFPCEE
EKVLETPQEIHTVSSEAVSLLEEVITPRKDLPLLLKLNERPAERLDYLGVSYGLTPRLL
KFWKRAFVPVYLRQTPNDLTGEHSCIMLKTLDDEDEAERGAWLAAFWTFRRRFLALLS
YQFSTFSPALSLNIIQNRNIAKSAPPVGLGREHLEALFLPYDLKRLEMYSRNMVDYHLIMD
LIPAIRSRYFLNQLGDLALSAAQSALLLGMGLQHSVDQLEKEIELPSGQLMGLFNRIIR
KVVKLFNDVQEKAIIEEQMVAVKDVVMEPTIKTLDLDEAAKEFQEKHKKDVGLKDMDL
SQYIIRGDDEEWNEVLSKAGQNASIVSLKSDDKRRKLETKQEPKQSKKLNRETKNKKDMKLRKK

>G_gallus

MQRKVDNRIRIVLIEINGVAERQGLFVVVVGDRGKDQVVILHHMLSKATVKARPSVLWCYK
KELGFSSHRKKRMRQLQKKIKSGTLNIKDDDPFELFIAATNIRYCYYNETHKILGNTFGM
CVLQDFEALTPNLLARTVETVEGGGIVVILLRTMNSLKQLYVTMTMDVHSRYRTEAHQDVV

GRFNERFILSLASCKNCIVIDDLNLPISSHVANITPVPPQSQEDSLRPQDLELKE
SLQDTQPVGVLDGCKTLDQAKAVLKFIEAISEKTLRSTVALTAARGRGKSAALGLAIVG
AVAFGYSNIFVTSPSPDNLHTLFEFIFKGF DALQYQEHL DYE I IQSLNPEFNKAVVRVNV
FKEHRQTIQYIHPADSVKLGQAELVVIDEAAAIP LPLVKKLLG P YL VFMAS TINGYEGTG
RSLSLKLIQQLRQSSAQAVTMTAENKSTATAKLASARTLHEVSLHESIRYAPGDPVEKW
LNDLLCLDCLSI TRI I SGCLPEDCDLYVNRDTLFCYHRASELFLQRLMALYVASHYKN
SPNDLQMLSDAPAHHLFCLLPVPPTQNSLPEVLAVVQVCMERGEISRQSIMNSLSRGKKA
SGDLIPWTISEQFQDPDFGALSGGRVVR IAVHPDYQGMGYGSRALNLLQMYEYEGKFPCL
EKTVPKPEIATVSETVSLLEEVVTPRKDLPLLLKLSERQAESLDYLGVS YGLTPRLL
KFWKRAGFVPVYLRQTPNDLTGEHSCIMMKMLNEEDVEQEPWLTAFWKDFRRRFLSLSY
QFSTFSPSLALNILQNKNIKQSQPPI SRAELES AFIPYDLKRLEMYSRNMVDYHLIMDM
VPTIARMFFLNQLGDISLSAAQSALLLGIQLQHKSLDQLEKEVELPSSQLMGLFNRI IRK
VVQLFNTVQEKAVEEQMVATKDIVMEPTIKSLNDDLEEAKEFQEKHKQEVGKLEKEMDLT
QYIIRGDDEEWNEVLSKAGQNASIISVKSEKRRKLETVRGHKQKQKFKKTKDIKQKRKK
>D_erio

MATFRKKVDNRIRVQIENGVAEHRSMFVI VGDGRDQV VILH HMLSKATVRARPSVLWC
YKKELGFSNRKKRMRQLQKKIKTGTLNLDQDDPFELFVAATNIRYCYYNETHKILGNTF
GMCVLQDFEALTPNLLARTVETVEGGGIVVILLRTMNSLKQLYTMTMDVHSRYRTEAHQD
VIGRFNERFILSLSSCKSCVVIDDLNLPISSHATNIKPIPPKKQDDGLSPREQLKDL
KESLQDTQPVGVLDGCKTLDQAKAVLKFIEAISEKTLRSTVALTAARGRGKSAALGLAM
AGAVAFGYSNIFVTSPSPDNLHTLFEFVFKGF DALQYQEHL DYE I IQSLNPEFNKAVVRV
NIFKEHRQTIQYIHPADAVKLGQAELLVIDEAAAIP LPLVKKLLG P YL VFMAS TINGYEG
TGRSLSLKLIQQLRQSSSESQONLSAENRNTSTARLAAARSLHEVTLHESIRYGGQDHVE
RWLNDLLCLDCLSVPRIISGCLPQTCDLYVNRDTLFCYHKASEAFLQRLMALYVASHY
KNSPNDLQMLSDAPAHHLFCLLPVPPTQNSLPEVLAVLQVCEGEISRQILSSLSRGK
KASGDLIPWTISEQFQDPDFGALSGGRIVRIAVNPDYQGMGYGTRALQLLQRYEYEGQFPL
WDEQEQTMTSAITSVSEAVSLLEEVVSPRKDLPLLLKLNERRAERLDYLGVS YGLTPQ
PLKFWKAGFVPVYLRQTPNDLTGEHSCIMLKDNLNSEDLDQNLQWLSAFWKDFRRRFLSL
MSFQFSKFSSTLALNILQNKSAKDDSTAVLTSAE LAATFTPYDLKRLEMYSRNMVDYHLI
MDMIPVIARMHFLKQFGDISLSVAQSALLLGIQLQHKSVDELEKEIGLPSSQLMGLFNRI
IRKVVQFFNTLQERAVEAEMVATKDIGLEPTVQTLNEDLNDAAKEFDEKHKKDI EKIKDM
NLSEYMIRGDDEEWDQVLKAGHTAIVSIKGDKKRKL EGTNQE QREGKMKKIKDKRKG
FGKKT

>X_laevis

MQRKKIDNRIRVLIENGVAQKHSFFVVDGHDGKDQV VILH HMLSKATVRARPSVLWCYK
KELGFSSSHRRKKMRKLQKRIKSGSLNIKEDDPFELFVAATNIRYCYYNETHKILGNTFGM
CVLQDFEALTPNLLARTVETVEGGGIVVLLLR TMS SLKQLYTMTMDVHSRYRTEAHQDVV
VRFNERFILSLASCKTCMVIDDLNLPISSHIANIKPIPPKTQDDSLSP EEQELQDVKL
SLQDTQPVGVLDNCKTLDQAKAVLKFIEAVSEKTLRSTVTLTAARGRGKSAALGLA IAG
AVAFGYSNIFITSPSPDNLHTLFEFIFKGF DALQYQEHL DYE I IQSLNPEFEKAVVRVNV
FKEHRQTIQYIHPADAVKLGQAELVVIDEAAAIP LPLVKKLMG P YL VFMAS TINGYEGTG
RSLSLKLIQQLRQSSADSQVTLTAENKSTASNKLVSARTLHEVTLQESIRYAPGDHVEKW
LNDMLCLDCMNI SRI I SGCLPETCELYYINRDTLFCYHKASESFLQRLVALYVASHYKN
SPNDLQMMSDAPAHHLFCLLPVPPTQNSLPEVLAVVQVCEGEISRQILSSLSRGKKA
SGDLIPWTISEQFQDPDFGALSGGRVVR IAVHPDYQGMGYGSRALRLVQMYEYEGQFPCLE
ENAANKPQEITSISGEAVSLLEESLTPRKNLPLLLRLSERPAEKLDYLGVS YGLTPKLL
KFWKRAGFVPVYLRQTPNDLTGEHSCVMLKTLHEDEESDQEPWLTAFWKDFQKRFLSLS
YQFSTFTPALALNILQNRNIKKQSENLI SRSQLEGALTPYDLKRLEMYSRNMVDYHLIMD
LIPSLARMYFLQQLGGLSLSAAQSALLLGIQLQHKGVDQLEKEIELPSSQLMGLFNRI IR
KVVQLFNRIQERAVEEEMVT TKEVMEPTL KSLQEDLEEAGKEFQEKHKQEV EKLKGINL
SEYVIRGDDEEWNEVLKKTGQNASVISMKSDMKRKL EKEPEKEGFKQKLLKKNKEKKQKFA

>P_marinus

MSRKKVDNRIRVLIENGFSQKYRTL FVVVDGDKGDQV VILH HMLSKAAVKARPSVLWCYK
KELGFSSSHRRKKMRQLQKKLKTGTL DVKHDDPFELFVAATNIRYCYYNETHKILGNTFGM
CVLQDFEAMTPNLLARTIETVEGGGLAVILLRTMNSLKQLYTMTMDVHTRFRTE SHQDVV
GRFNERFILSLASCKTCMVIDDLNLP LSTHTL CISPLPPKPQDESLS PSEIELGDLKE
SLQDTQPVGVLVSCCKTLDQAKAVLKFIEAISEKTLRSTVALTAARGRGKSAAMGLAIAS
AVAFGYSNIFVTSPSPDNLHTLFEFLFKGF DALQYQEHL DYE I IQSTNPEFNKAI VRVNI
YKEHRQTIQYIHPGDAAKLGQAELVVIDEAAAIP LPLVKSLG P YL VFMAS TINGYEGTG
RSLSLKLIQQLRQSSQSQVMTKAENKNTAVSKLASARSLHEVTLDESIRYAPGDPVEKW
LNDLLCLDCLNIQRVSVGCPVPTCDLYVNRDTLFCYHKASEAFLHRLMSLYVASHYKN

SPNDLQMLSDAPAHHLFCLLPVQATQNALPEVLAVIQVCLEGEISQQSIMSSSLARGKRA
AGDLIPWTIISEQFQDPDFGSLSGGRIVRIAVHPDYQGMGYGSRALRLLQMYEYEGKFPCLD
ETASQTQGHISSIASDAVGLLEETLTPRRDLPLLLHKLSEKPTERLDYLGVSFGLTPQLL
KFWKRSQFTPVYLRQTPNDLTGEHSCIMLKVLNEDEQMERPENWLAAYWKDFRRRFVTL
SYQFSGFLPAVSLNILQNRNVKEKPAQALSKESEATFSPYDLKRLMYSRNMVDYHLIM
DLVPGIARLYFTHRLPDMSLSAQAQLALLLVGLQHKAVEQLEKEIDL PVSQMLGLFNRIM
RKAVQVFTAIQETAIEEQMTKVKEVMEPTVKSLAEDLEDAAKEFAEKHKKDLVQIKDMN
LTEYVIRGDDKEWNEALKKVGPSASVVSITGEKRKSAEKDSSHDHKSKKSFQKGPGLG
KPGKPGKPGGPGKAGKSRGGFKKQK

>C_intestinalis

MRKKVDNRIRVLVENGVA TKHRTMFFIVGDKCRNQVILHHMLSKAQVRARPSVLWCYK
ELGFSTHRKKMKQIQKRISSTGLDKENDPFELFVAATNVRYCYNETHKILGSTYGM
VLQDFEALTPNLLARTIETVEGGGII CFLLRSM TSLKQLYTMTMDVHARYRTE SHSDVVG
RFNERFLLSVASCNNCLIIDDELNVLPVSTNSQNI EAVPPKPAEDSLTPSEKELVDLKE
LQDTQPVGCVINCKTIDQKALLKFVDAISEKTLRSTVVM TAARGRGSATLGLAISAA
VAFGYSNIFVTSPPENLKTLEFEVFKGFDAL EYQEHLDYELIQSTNPDFNKAIVRVNVF
REHRQTIQYIHPSDSAKLSQAELLVIDEAAA IPLPLVQSL LGPYLVFMSSTINGYEGTGR
SLSLKLKQLRIQSATRNTATTQTTASPLGRNLSEITLEESVRYAPTDP IESWLNELLC
LDCGSSSFSKI STGCPPPSACDLYYVNRD TLF CYHKASEAFLQRLMSLYVASHYKNTPN
LQMLSDAPAHHIFCLLPVSAEQSLPDVLCVLCLEGEISKSSVSSLSRGKRAAGDL
IPWTVTEQFQDHFASLSGARVVRIATHPDYQGMGYGSRALELLEKYYEGKMTSLDENET
PAEEVNIVNEEEVEAMDEVLT PRKSLPPLLWKLSERRPEQLNYIGVSYGLTAPLLKFW
AGYTSVYLRQTANDLTGEHSCIMLKALDTEVDKSWQDAYFDDFRRFISLLHFQFSSFKS
STALQVLRKSKKTTKSNPLTKSELDFFLT SYDTKRLTLYSRNMADYHLVMDLLPTLARL
VFLEKLVVPLSVAQKAIILAIGLQHKSVETVASEIDL PSSQVLGLFNRAVRKITTHLSEI
ESAAA EKELKSTLGDENAADAVNMQPLEESLDSELNAAAYEVVQKQKEELKQTLASMDLK
QYIEIGGSDEAWDALSSGKKQNI SVKSAKRPLPTSEEPDTKDKKKRKFKKKVKSA

>B_floridae

MVLKKVDNRIRVLIENGVVQHHRGLFVIVGDKGKQVILHHMLSKAQVKARPSVLWCYK
KELGFSTHRKKRMKQLQKKIRNGTLDVKEDDPFELFIAATNVRYCYAETHKILGNTYGM
CVLQDFEALTPNLLCRTIETVEGGGIIACL LLSMNSLRQLYTMAMDVHSRYRTEAHRDVV
GRFNERFLLSLSSETCMVIDDQNLNIPIS SHILNISAVPSRSKEDSLNPAEVELKGLKE
SLQDTQPVGSLVNCKTLDQAKALLKFI EAISEKTLRSTVAMTAARGRGSAAALGLA IAG
AVAFSYSNIFVTSPPENLNTLFEFVFKGFDALDYQEHMDYELVQSTNPDFHKAVRVNI
YREHRQTIQYLHPSDAQKLGQAELVVIDEAAA IPLPLVKS LMGPYLVFMAS TVNGYEGTG
RSLSLKLIQQLRQONATFGGNSEAKAAAAMATKTEGGASATGRTLHELVLTESIRYASED
PVEAWLNRLCLDVSSVPRVHSGCIPASCDLYYVNRD TLF CYHKASEAFLQRI MSYVS
AHYKNTPNDLQMLSDAPAHHLFVLLPPVDPDQSALPEVLCVVQVCLEGEISKQSIMSGLS
RGKRASGDLIPWTVSQQFQDTEFASLSGGRVVR IATHPDYQGMGYGMRALQQLDMYYQ GK
VPSLSEDTGTAQGLPAVDKETVSLLEEKLGPRKNLP LLSKLSERPAERLDYLGVSYGLT
APLLKFWKKAGYAPVYLRQTPNDLTGEHSCIMLKRIQEDDEGQD TDWLSAFWTFRRRFL
SLLSYQFRTFQPALSLSILQAKKDVKTDT SVLTQGDLYAHLTQYDLKRLDLYSRNMVDYH
LIMDLLPALARMYFLQKLDTSLSVAQSA ILLGLGLQHKTVDDLERELELPASQVMGLFNR
LVRKIVQFLSSIEEQAVESELLAPKEVMEPTAQSLKTELAEEAKEFEDKQKELKLLKD
MDLEKYVIKGETEWESALTSGLKSGVSVASGKKRKEEAAPTPESTKKKSKKFKKDKKF
KKT

>S_purpuratus

MVRKKIDNRIRVQIENGVTQGHRSFLVVGDKGKQVILHHMLSKAVVKARPSVLWCYK
KELGFSSHRKKRMKQLQKKIKGGMVDVSDDDPFELFIAATNIRYCYSETHKILGNTYGM
CVLQDFEAMTPNLLARTIETVEGGGIVVILLRSVSSLQQLYTMSMDVHSRYRTE SHQDVV
ARFNERFLLSLSASDCMVDQQLQILPLSQHTLNIEAIPPRLRXVEDSQSPA EIELKEL
KASLQDTQPVGALINCTKTLDQAKALLKFVESISEKTLRSTVSM TAARGRGSAAALGLAI
ATAVAFGYSNIFVTAPSPENIRTLFEFIFKGF DGLDYQEHLDYELVQSTNPEFNKAVIRV
NIFREHRQTIQYIHPGDAHKLSQAELVVIDEAAA IPLPLVKELLGPYLVFMSSTINGYEG
TGRSLSLKLQQLRQONATFGVSSAANKASTSLSTGSSASGRVLHEVTLEESIRYSEGD
SVEKWLNELLC LNVAEVQRISSGCPLPADCDLYYINRD TLF SYHKASEVFLQRLMSIYVA
AHYKNTPNDLQLLSDAPAHHIFCLLPVKASQSGLPEILCVLQVCLEGEISKETIMSSLS
RGKRASGDLIPWTISQQFQDNDFP SLSGGRVVR IATHPDYQGMGYGSRAMSLLQKYYQGE
FPSMAETPETKTITTVKEAVGLLEEVIVPRKNLP LLSKLSERPAEKLDYIGVSFGLTA
DLLKFWRKAGYLPVYLRQTANDLTGEHSCIMLKTVQESDTEGANQETWLTAFWTFRKR
MSLCAYQFRAFKPSLALS VLFCKAYNKMKPQPM TATELELLFSKYDIQRLELYSRNMADY

HLITDILLPSVTRLVFLKKNINFQLSAVQFSILLGLGLQHKTVDVLEKETEVPVSQLLALFN
RTVRKVVTELVNSVLEREVEKDMMEKVVNMEPTRQTMAQELDEAAAVIEAKQKKELAAALK
GLDLSRYAIRGQSEWDDALKSGKKKNLISIKIANQEKRAAEKSKSTEKRRKGDRESLGGG
DDDQEMSKKKKKKKFTPGKGSKN

>A_mellifera

MVRKKIDNRIRILIEINGVITGYRTMFI I IGEKARDQVVLHMLSKS I IKARPSVLWCYK
KELGFSSHRKKKMSLQKKVKSGLDVNEDDPFELFVVSTNIRYCYHETHRILGNTYGM
CVLQDFEALTPNLLARTIETVEGGGLIIFLLQSVNSLKQLYTMNMDVHQFRTEAHQNI I
GRFNERFLLSLASCKRCLVDDQLNVPLSSSHNLKIEPIQKLSSETLSELDTLKESLKD
TQPISSLVNCCSTLDQAKVLLKFI ECISEKTLRSTVSLTAARGRGKSAALGLAVAAA I TF
GYSNIYISSPSPENLNTLFEFIFKGF DALGYQEHL DYGLVQSTNPEFNKAI IRVNI FRDH
RQTVQYIHPDHAHKL NQAELLVIDEAAA I PLPVKAMLG PYLIFLASTINGYEGTGRSLS
LKLLQQLRNQTISSNSHEKQKNEK I IGRQLHELTLEESIRYKPGDSIEQWLCDLLCLNV
IINTPILSGCPSPDICELYIINRDTLFSYHKASELFLQRLVALYVASHYKNSPNDLQMS
DAPAHHLFCLLGPIDSNNKILPEILVVIQICLEGEINKNTINDGLVRGRRAGDLIPWTI
AQYQDQDFPALAGARIVRIATHPEYQSMGYGSRALELLKKYEMKIVNINETSLOTCMT
EISKVKNEEVNLLKEKIEPRVSLP LLLKLSERRPENLDYIGVSFGVTEPLLKFKYKRANF
IPVYLRQTTNDITGENSCIMLYKINSDDITW LQAYW NDFRKRFI SLLSFSFN TYSPLA
LSILFNKSITSEATTLTKNILDIYFTSYDLKRLSMYSNNMADYHLIMDLLPSLARLYFLN
MDDIHL SAVQSAILLGLGLQHKTVDTLAKELDLPTTQLLGLFNRIIRKFIQYLNK I TEN
FIETMMKIEGNNETIQLNPI NGKSLYDELENAAKELKAKQKAELEK LKENLEQYAIKG
TEADWNVALSEKSKNLISIKSGEKRELEHNILEENKQFKKKKKHFSKI

>A_gambiae

MVKKKIDNRIRVMIENGVKLGHR TMFVI VGNKARDQVP ILYDILTKASVKARPTVLWCYK
NKDEAISNHGKKRAK I QSGKIDINESDLF DAFRVATTIHGRYKDTHTILGKTYGVCVL
QDFEALTPNLMARTVETVEGGGLI ILLLKT I SSLKQLYTMSMDVHKRYRTEAHQNVTCRF
NERLILSLADCSRCLLVNDDLTVLPLSSRTADV KPIDVASIENVQSEQLAELKESLADAP
PAGPLVNL CRTYDQAKAVAQFIDALAEKQLKPPTSLTAGRGRGKSAAMGLA IAGSIAFGY
VNVYVTS P SPENLITLFEF I LKGF DVLEYQEHTDYTI IRSTNPNFNKAI IRINITRNNRQ
TIQYISPTDAHLLNAADLLI IDEAAA I PLPMVRAMLGPYL VFMAS TINGYEGTGRSLSLK
LLSQLQKETNAPLP I KLEESIRYSPGDPVESWLTSL LCLDATIVQNMNSGCPPPDECQLY
YVDRDALFSYHKA AETFLQRI VSI CVSSHYKNSPNDLQMSDAPAHHLFCLLGP I VKNQ
LPEILVVIQVCLGEISSTTLANSLVRGMKASGDLIPWNITEQFGDHEFPKLSGARIVRI
ATHPNYQRMGYGKRALNQLKHYYEGRFPIESDGLGDEQGGDDNGIETIDDEVDLLKEVI
TPRKKIPTLLKAL TERKPEMLDYLGT SYGLTGELLRFWKSQKFVVPVYLSQKENELTGEHS
CIMLCPINSTVERVETNEWLN DYFLDFRRRVLKLLGKSFNKFTTSLALS LLENKAVKMEG
KALTQTTIDEVFLPHDVQRLEMYIHNQVEYKLIMDLTSDLASLYFQ GKMSGAH I ETLHKA
ILLGCGLQHKTI DKLMEELNMP SNQVLAKFYDCVKKLTNYILRTMESTIESGMAKPSELN
MGENLVPLKQSLDDEFAEDVKTLEKQKKE LAKLKLNL DQYAIKGTDEEWSKVLSSNKS
TIVSIKSGEKRLNENADDSMSAETGDVSAFQKKKAKFGKKQKSRNNKG

>D_melanogaster

MVKKKIDNRIRVMIENGVKLGHR TMFVI VIGDKARDQVP ILYDILTKSTVKARPTVLWCYK
NKDEAISNHGKKRAK I AVGKVDVNEADLFD SFRVATTIHGRY YSETHAVLGR TYGVCVL
QDFEALTPNLLARTVETVEGGGLI ILLLKT I QSLKQLYTMSMDVHKRFRTEAHQTVTCRF
NERLILSLADCSRCLLVNDDLTVLPLSSKTINVEPVNPAGAGRSPNEASLKEKESLTLV
QPAGALVNLCKTYDQANAVAQFIEALVDKQLKPPMSLTAARGRGKSAALGLSIAAAVAFG
YVNIYVTS P SPENLITLFEFV LKGFDALEYQEHADYTI IRSTNADYKKA I IRINITRSSR
QTIQYIAPSDTHLLNAADLLL IDEAAA I PLPLVKKMIGPYLVFMAS TINGYEGTGRSLSL
KLISQLQKDNAR PPLKLEESIRYQENDDIEKWLINLLCLDASTVPSISSGCPTPDACEL
YYVDRDALFSYHKA AEFHLRLVSIYVSSHYKNTPNDLQMSDAPAHHLFCLLGPVQRM
ALPEILVVIQVALEGQISAQSISDSLGRGKKAAGDLIPWNVAEQYGD RDPKLCGVRIVR
VATHPNYQRMGYGKRAIQLLKDY YARKHTNLEDGPVASKDAGKGI EEEVEEELSLKEQI
RPRSRIPTLLQRLHERVPEHVDYIGTSYGLTTELLKFWKNAGFVVPVYLSQKSNELTAEHS
CIMLHTPNATPWLGLYQDFRRRVLKLMGKTFREFETKLC LALLKNKSVDT EGSALKVLD
KPMLDVYFLPHDLQRLESYARQQSEFRLIIDLTDIAQLYFQGRIDGLQLDLVQQGILLA
LGVQGKTVDALGLELNMPGNQLLAKFFDAMKRCNQCFRSVLEEHIEGMLREADLSKGEE
LQPLTSLDKELDQTAQKLSKQQRKELKRLKAEQLDEFQIKGTEEDWSKALETNGTGGGS
GLLSVKS GVKRLDGP I ETREDGD LAAPLSKKKKKNNPKQRRSQGKSLI

>B_mori

MVKKKIDNRIRVMIENGVKLGHR TMFL IIGDKSRDQVP ILYDMLVKSTVKARPTVLWCYK
NKDEAISNHGRKRAK I AAGKLEVSEESLFD AFRVATTIHGRY YSESHAMLGQTYGVCVL

QDFEALTPNLMARTIETVEGGGLIIFLLKTMDSLRLQHSITMDVHSRFRKTEAYDTPVNRFR
NERFLLSLADNPRCLVLDLSTVLPISSKTAQIDPVDATPERVNPVKLTELISLSDSPPA
GPLVALTRTYDQATALIALINTLADKQSRPPHCLTAARGRGKSAACGLAVAAVAALGYVN
IYVTSPPHENLITLFEFLLRGLDACLYQEHIDYNIVRSTNPDFKKAIVRVNIARNRQTV
QYITPDDHSLLSAADLVLVDEAAAAIPLVHVAAAAQKAPLALLSSTVSGYEGTGRALSCLK
FAQLQTQHNAPPIKLDPEIRYRTGDPISWLNLSLLEAPPSSMGVGGAPPAACELRV
NRDALFCYHKAEEAFLHRLVSIYVASHYKNSPNDLQLLADAPAHCLFVLLAPPRAAALP
ELLCVLQCLLEGNISDKSVRDSLGRGRKAAGDLIPWNI CEQFGDKDFPKLSGARIVRIAT
HPAYQRMGYKRALQYLAAYYSGDIPCLDEDASDAETDRQTDAQSTLQETIAPRSKPPT
LLKRLTEVRPNLDYLGTSFGLTEDLLRFWKSQKYVPVYLSQKANELTGEHSCIMLRALR
DAAWLGAYSADFTRRAARLLPRTLRLPAALALSTLVTDVSVKTNKPVLT KDLEIQYLTGH
DLRLEAYSKQQADYRLITDLLQPLAELVFHTKSTIKLDAVQKVI F I AMGFQMKDVDDIA
SELGLPGSQILAKFYEASKKINAALNAVLEDSAAKEIGIEDKAASGAAQPVQQSLQDELS
KAAKDLEKQKELSKLMGEDLSQYRIKGSDDLWQALASTKQKTILSVKSGEKRLGEDM
KEIDDLDLQTNKKKKKKKSFKN

>I_scapularis

MTRKKIDNRLRVL IENGVHTGHR TMFVVVGEKGRDQVVI LHHMLSKTTVKARPSVLWCYK
KELGFSTHRKKRMRQLQKKIKSGKLDVNEEDPFELFVAATNIRYCYHETHKILGQTYGM
LVLQDFEAMTANVLARTMETVEGGGLVLLLLQSISLRLQLYTLAMDVHARYRTEAHQEVV
GRFNERFILSLSSCEACLVDRLNILPVSTSAAGIAPVPRAPQDPEASELASLRASLED
SQPLGALVAVCSTCDQARALLKCVDAVAEKTFRTTVAVTAARGRGKSAALGLAVAAVAF
GYSNVFVTAPSPENLKTLEFVVLKGFDAKYEHEHQDYEVVQSSQPELNKAVRVTVFHEE
HRQVVQVWHPGDSHLLGQAE LLVVDEAAAIPLPLVQKLLGPYLVFLSSTVNGYEGTGRSL
SLKLIQQLRQQAAGDAGKSSSGGRVLHEVSLSESIRYRPGDPVEAWLNRLCLDANVPP
ASGCPGDCQLFYVNRDTLFSYHRASEFLQOVVSLCVASHYKNSPNDLQMLSDAPAHH
LFVLLPPVQPDRALPQVLCVLCVQCYEGGISRASVSDNFARGRRAGDLIPWTVSQFQD
EEFAGLAGARIVRIATHPEYQRMGYGLRALELLTEYHLGHTPSLEEAEEAATPGAAGADET
GAEAAERERLEPRKRLPPLLLRLSEKRPERLDYLGVSYGLTADLLRFWKKAGYVPVYLRQ
TPSELTGEHSTIMLRCLQEGAQWLPQLWADFRRRFLSLLGYQFRVLKPATALSVLNRPLE
DSVHPRVPSLAEVEREFTAYDLRRLLELYARNLVDHHLVTDLLPALARMHFLKALGDVHLS
AVQSAIILLGLGLQHRTVDEL SRELDLPTSQLLGLFNRTVRKCATFLGALLEKSLDAALPR
PSPALLPLQPVVQSLGDEL SGAAREIEARQKELKKLQOMDLSQYSIKGS DSEWKDALQQ
PGKQLVSIKSINKRPQETDTPKEPKQKKKKKQK

>C_elegans

MRTKLDGRI RTQIENGVASGHRSMFAVVGDKARDQVPILYHILSKSTVSARPNVLWCYK
ELSFSTHRAKKAKMKKATTTISGSLPDADPFDFVISTTQIRYCYNETEKILGNTFGVL
VLQDFEAMTPNLLARTIETIEGGGMVILLMQSVRSLRQLYTI SMDVHNRYRTEAHNEITA
RFNRFILSLASCSSVLVLDLQRLVLPISSHIENVEAIPASQKKIQSESDAELASLKEAM
KETKPIGPLLSRARTACQAKALLRFLDVITEKQSNVTCSLTAGRGRGKSAAVGLSLAGAI
AFGYTNIFVTSPSPENLKTLEFVVKGFDA LDYQEH TDYELIQSANPEFKNCLVRINVFR
EHKQTIQYISPTDVQKLGQCELIVIDEAAAIPLPLVKELISGPYISFLSSTINGYEGTGR
SLSLKLLQQLRQQAAGGAEKGSASNKGKTLHEMHMEESIRYKPGDKIEKWLNRLLCLD
ATNCQLKLECGTPPPAACELYIVNRDTLFSFHDAEAFLLQVMAIFVSAHYKNSPNDLQOM
LSDAPAHNLFVLMAPIDKSRKTIPEVLAVVQVLCLEGRLDSDNIQNGLESKRAAGDLLPW
TVSQQFMDKQFGTL CGGRIVRVAVHPDYQSMGYGGRVQLIEQYYLGLATSLDEEEKAPA
PPSKTVIKQVKDGHTVELLEERIEPRADLPPLLQRLDERKPERLDYLGVSFGLTVPLLK
WKRNEFVPVYIRQNSNDITGEHTCIIKLGLEHGGSDSDEEPSATWLPVYWREFRRRIVNL
LSFDFSSFPQMALSLQLKNKHVEKQMKRLVIERSELAIHLSNTDLRRMSQYGRNMVDS
HIITDILPIVAKLNFEQRLPQELKLAVTQSAIILLARGLQHKHFEDISKELDLPMNQIFAL
LTKAIRRIGDWFDEVCE TAVRENLDKEAEASAANKPTSSLPKAVPLANLEDELESAAKEI
RARHDRDRKALLAELGNE LQKYEIIQDEKELAEAYESVNMKYANKLVSVKSKRTAIQAAI
PDAKDPANKNAKKKKRFS SGGRR

>S_carpocapsae

MVRTKLDNRIRTLIENGVATGHRSMFAIVGDKGRDQIVILHHILSKAVVAARPSILWCYK
KDLGFSSHRKKRMEVQKKLALGKINVREDDPFELFVSSTEIRYCYAESHKILGNTYGM
LVLQDFEAITPNLLARAIETVEGGGIIVFLLRSVNSLRQLYTMMDVHSRYRTEAHSEIV
ARFNRFILSLASCNKCAVIDDQNLVLPISSHIKKIEAIPSSSEKSLTKEEEEELKELKMA
MADSKPIGQLLSKCRNTNCQKVLRLLDVITEKSLKATCSITAARGRGKSAALGLA IAGA
IGFGYTNVFTVTSPPENLKTLEFIVKGFDA MDYEEHSDYELIQSTNPEFNKALVRINIT
RDHRQTIQYFHPSDAGMLGQAE LLVVIDEAAAIPLPLVKELISGPYLVFLASTINGYEGTG
RSLSLKLLQQLREQSSGAAKKTEKGGKLATATARMLHEL SLEESIRYKPGDQIELWLNRI

LCLDATNIHRLISGTPPPKDCELYVNRDTLFSFHKGSEQFLRNMMAINVSAHYKNTPNL
LQMLSDAPAHHLFVLMGPVRDEQTHLPEILAVVQVCIEGSLTKNSVTSSELRGRRRAAGDL
LPWTISQQFLDNDFPTLSGARIVRIAVHPDYQSMGYGSRALELLMQYYSGEVSSLSSEKSL
KSSASSKIRSVEDDKIGLLEEQIAPRPREDLPPLLHRLTERPAERLDYVGVSFGLTVNLL
KFWKASFPVYLRQTNDLTGEYTCIMLRSLAPQDDESSRDSWLASYWKEFRTRLISLL
GFEFKFSPHLAISLLQLKNPFVEKQMERKPYSRKELEFLSNTDLKRIAQYARNMIDHH
LVTDLLPTVSKLYFNEKISDEVKINIVQSAILLGLGLQHKTVDRMVCELEMPANQILALF
SKTMRRSLSDHFDSCMNALESQIGKKEQVVEGTPKGNVEMEAMPISLEEDLRDAEEDIKR
RQKRDRAKLESELGKSFKEYAIKGTDEQWANALSTVKLDSAKKSAIITVGSQMEKNSSAI
EDLTKEAPRSACKNLKRKRKSGGKL

>M_galloprouncialis

MVRKKIDNRLRILIEINGVAEKHRSMFVVVGDHGRDQVVILHHMLSKATVRARPSVLWCYK
KELGFSSHRKKRMKQLQKKIKSGKLDVKEDDPFELFVASTNIRYCYAETHKILGQTFGM
CILQDFEALTPNLLARTIETVEGGGVVILLKTMASLKQLYTMSMDVHARYRTEESHQDVV
GRFNERFLLSLTSCKNMVIDDQFNMLPVSSHILNIQPVPAKSMEDALSPQEQELRDLKD
SLQDTQPVGSIVNCKTLDQAKAVLKFEVAISEKTLRSTAVMTAARGRGKSAALGLAIAT
AVSFGYSNIFVTAPSPENLKTVEFVFKGFDTLEYQEHLDYEIIQSTNPDFNKAIVRANI
FREHRQTIQYIHPSDSHKLGQAELVCIDEAAAIPPLVKNLLGPYLVFMSSTINGYEGTG
RSLSLKLVQQLRQQSATFSGSSEVQKAI SMATKSDSVSSATGRALHEIVLND SIRYANGD
PVEKWLNDLLCLDATNVSRIITSGCPVPECDLYVNRDTLFSYHKASEGFLHRIMALYVS
SHYKNTPNDLQLLSDAPAHHIFALLGPVDPNQSSLPVLCILQICLEGDISSTIMNSLS
RGKRASGDLIPWTISQQFQDQDFAGLSGARVVRIATHPDYQSMGYGTRAMNLLQYYKGE
MISLSETKDIEDNVTVPVDEQDVSLVDERLTTPRKNLPPLLMKLTERRKPEKLDYLGVSFGLT
AGLLKFWKSKFTPVYLRQTPNDLTGEHSCIMLKMHTEDNPEDRSEPWLAAFWKDFRRR
FVSLLSYQFRFTFKPSMALNILQEKTFKSTGKKKMSLAELEAHFTKYDVKRELELYAQNMD
YHLVIDLLPAISRFLFGLIDTHLSVVQSGLLLGLGLQHRTVDELEKELDLPSSQLLGLF
NRIIRKIVQNLNEILEADVEKDMVARKEITLEPVKQTMDEELV

>A_californica

MVRKKIDNRIRVMIENGVAQKHRSMVVLVGDGRKQVVILHHMLTKATLKARPSVLWCYK
KELGFSSHRKKRMKELQKKINSKLSAKEDDPFELFISSTSIRYCYAETHKILGNTYGM
CVLQDFEALTPNLLARTVETVEGGGVIVLLLRMTSSLKQLYTMTMDVHLRFRTEAHQDVV
GRFNERFLLSLASCETCVVLDKLNLI PVSTHALALEPLPAKSLEENLTAQEI ELEEELKT
SIQDMVTTGKI IKLCRTLDQGGQFLKFVDAIVEKTLRSTVVMATAARGRGKSAALGLAMS
ASVGFYYSNIFVTAPSPENLKTDFDFIFKGF DALDYQEHIDYEIIQSTNPDFNKAVVRVN
IFRDHRQTIQYIHPSDAHKLGQAELVCIDEAAAIPPLVKNLLGPYLVFMSSTVNGYEGT
GRSLSLKLVQQLRQQVSTYGGSVKEARLANMAKQGVDSASGRVLHEVELKESIRYASG
DPVEAWLNRLCLDASSVPRSITGCPPLASCNLYVNRDTLFSYHPASEKFLHRIMALYV
SSHYNKSPNDLQLLSDAPAHHIFVLLGPVQSANTELPEVLCVLQVSLEGEISKSTIMNSL
QRGQRGSGDLIPWTLSQQFQDQDFPGLSGARVVRIATHPDYQGMGYGSRALTLMQEYYEG
KFPNLKESNPDDAQESSNDGGDRGDEELKQVDEEVTLLNEHLAPRKNLPPLLVKLTDR
RPEPLQYLGVSYGLTSDLFRFWKRSYVYLRQTANELTGEHSCIMVKTISVDEKEDKQ
SWLPSFWKDFRKRLLVALLAYQFRSFKPALALNILQNKNFKTEEKAVSLRRSELLQFVTY
DIQRLELYSSNMVDYHLIVDLSLARLFFTNHVP IHLVSVQSSILLALGLQHKTVEDVE
KDIELPVSQILALLNRTIRKRVQHFNEILEKEVEATMAEKKEIVMEPTLTSVDQELKEAG
QQFKQEHKQNVATLVGADLSHFAIKGSDKDWQEAIKGNSKSVISVKSHLEKRRKSPA
EADTDQMSNKKQKKRKKQREKS

>A_californica

MVRKKIDNRIRVMIENGVAQKHRSMVVLVGDGRKQVVILHHMLTKATLKARPSVLWCYK
KELGFSSHRKKRMKELQKKINSKLSAKEDDPFELFISSTSIRYCYAETHKILGNTYGM
CVLQDFEALTPNLLARTVETVEGGGVIVLLLRMTSSLKQLYTMTMDVHLRFRTEAHQDVV
GRFNERFLLSLASCETCVVLDKLNLI PVSTHALALEPLPAKSLEENLTAQEI ELEEELKT
SIQDMVTTGKI IKLCRTLDQGGQFLKFVDAIVEKTLRSTVVMATAARGRGKSAALGLAMS
ASVGFYYSNIFVTAPSPENLKTDFDFIFKGF DALDYQEHIDYEIIQSTNPDFNKAVVRVN
IFRDHRQTIQYIHPSDAHKLGQAELVCIDEAAAIPPLVKNLLGPYLVFMSSTVNGYEGT
GRSLSLKLVQQLRQQVSTYGGSVKEARLANMAKQGVDSASGRVLHEVELKESIRYASG
DPVEAWLNRLCLDASSVPRSITGCPPLASCNLYVNRDTLFSYHPASEKFLHRIMALYV
SSHYNKSPNDLQLLSDAPAHHIFVLLGPVQSANTELPEVLCVLQVSLEGEISKSTIMNSL
QRGQRGSGDLIPWTLSQQFQDQDFPGLSGARVVRIATHPDYQGMGYGSRALTLMQEYYEG
KFPNLKESNPDDAQESSNDGGDRGDEELKQVDEEVTLLNEHLAPRKNLPPLLVKLTDR
RPEPLQYLGVSYGLTSDLFRFWKRSYVYLRQTANELTGEHSCIMVKTISVDEKEDKQ
SWLPSFWKDFRKRLLVALLAYQFRSFKPALALNILQNKNFKTEEKAVSLRRSELLQFVTY

DIQRLELYSSNMVDYHLIVDLLSDLARLFFTNHVP IHL SVVQSS ILLALGLQHKTVEDVE
KDIELPVSQILALLNRTIRKRVQHFNE ILEKEVEATMAEKKEIVMEPTLTSVDQELKEAG
QQFKQEHLQNVATLVGADLSHFAIKGSDKDWQEAIKGNSKSVISVKSHLEKKRKS PAEAD
TDQMSNKKQKKRKKQREKS

>O_bimaculoides

MTRKKIDNRITLIENGVAKKHRTMFVVVADQGHQDQIVVLHMHMLTKAEVKARPSVLWCYK
KELGFSSHGKRRMRKMQKKIKAGRIATMSEKDADAFELFLASTNIRYCYKETHKILGN
TYGMCVLQDFEALTPNLLARTVETVEGGGLIVILLRTMNSLKQLYTMTMDVHRSRFRTE SH
QDVTGRFNERFLLSLTTC SQCLVIDNQYNILPLSSLTSTVTPLPKLSKEESLTQSERELV
ALKESLQDTQPVGAI INCARTLDQAKGLLKFIDSVSEKTLRNTVVM TAARGRGKSAALGL
AMATAVGFYSNIFVTS SPENLKTLEFIFKGFVDVLEYQEHIDYEI IQSSNPDFNKAIV
RVNIFRDHRQTIQYIHPD AHKLGQAELVCIDEAAAIP LPLVKELIGPYLVFMSSTVNGY
EGTGRSLSLKLIQQLRQQAATFSGTEVKKVTTVESTTNSLSGRALYEVQLNESIRYADGD
PVEKWLNDLLCLDVTNVRITSGCPLPQDCDLYYVSRDTLFSFHKGSEAF LHRIMSLYVS
SHYKNTPNDLQMLSDAPAHHIFVLLGPMESTKQALPEVLCVVMQCLEGDISKAAVMSGLR
QGKRASGDLIPWTISQQFQDNDFASLSGVRVRIATHPDYQKMGYGSRALQLLSQYYKGE
IVGLNEDVEEVPKDIEPVDEEELDLIVEKVPVRKNLP LLLLK LTERPPEQLDYMGVSFGL
TADLLKFWKRNGFVPVYLRQTANDLTGEHSCIMLKTLHDKEDESNTLYDKKWL LSFWKDF
RRRFISLCSYQFRFFSCSTALSILQQKHCKEEAVALAKQELDFKLT VFDLKRLELYSQNM
IDYHLIMDLMP TIAELYFLQRLPELHL SAVQQALLL GIGLQHKTVD DLEKNLGLAASQLL
GLYNRIIRKVVQHLNSIVEKSVEQTLVKEHKITMEPVPQSLDDELDAAAVEIKEEMRKKI
APILEESELSQFAVKGDEDQWNLALSTGKKNFKSISIKSIRQKRQLKEVEENTSQKKSKK
FKKSKKHK

>L_anatina

MVRKKIDNRIRIMIENGV AERHRS LFVIVGDHGRDQV VILHMHMLSKAAIKARPSVLWCYK
KELGFSSHRKRRMRQLQKKIKSGKLDVKEDDPFELFVAATNIRYTYSETHKILGN TFGM
CVLQDFEALTPNLLARTVETVEGGGLV LLLRTMTSLRQLYTMTMDVHARFRTE SHQD VV
GRFNERFLLSLASCKRCMVIDDQLNVLPVSSHTLQIQPVPPKTSDDPLPPSEQELKELKE
SLQDTQPVGVLVDSCKTLDQAKAVLKFVEAISEKTLRSTVSLTAARGRGKSAALGLAIAA
AVGFGYSNIFVTS SPENLKT VFEFVFKGFDSLEYQEHLDYELVQSTNPEFNNAIVRVNI
FREHRQTIQYIHPGDAHKLGQAELVCIDEAAAIP LPLVKNLLGPYLVFMASSTVNGYEGTG
RSLSLKLLQQLRQQNVTFGKQATDQQK GAPGLASATGRILHEISL TESIRYGDGDAVEQ
WLNEVLCLDASNVRISSGCPVPENC DLYYVNRDTLFSYHKASEAFLQRI MGLYVASHYK
NTPNDLQMLSDAPAHHIFVLLGPVDPTQNTLPEVLCV LQVCLEGEISKSNIMNSLSRGKR
ASGDLIPWTISQQFQDHD FQP LSGGRVRIATHPDYQGMGYGTRALEILQHYYEGKIPSL
VEDAVVEEAVPTVSKDAVGLLEER IAPRKNLP LLLLKLSERPAEKLDYLGVS YGLTAGLL
KFWKRSGFIPVYLRQTPNDLTGEHSCIMLKV LNEGEKTEQEEESWLMAFWKDFRRRFLSL
LSYQFRTFHPSLSLSILQQKQCKEKESVISA EELDNYFTKYDLKRLELYSQNMVDYHLIM
DLLPSIARLHFLNKTGVHISAVQAA I LVSMGLQHKSVEDVEKDMEVPSSQLLGLFN R VIR
KMVQQFNQIAEEAVEKDMVARKEVIMEPVKNSLDQELGEAAA EAVAKKQKKDLELLK DADL
SQYTIKGESEGDWENALTKGT KGLLSVKSGVQEKRKSVDSTSF EQQTPGKHKHKKKQKHHK
H

>N_vectensis

MVLKKVDNRRLILIENGVQTRHRS LFVVVGDKGRDQV VILHMHMLSKAAVKARPSVLWCYK
KELGFSSHRKRRMRKQLQKKIKSGTLD MNKDDPFELFVSSSTNIRYCYKETHKILGN TYGM
CVLQDFEALTPNLLARTVETVEGGGIVVILLRTMSSLKQLYTMSMDVHARYRTE SHQD VV
GRFNERFILSLSSCKSCMVIDDQLN ILPLTSHSLNIKAVPPKPKKEESLTPSEVELKELKD
SLQDTQPVGSLINCCRTLDQAKALLKFIEAISEKTLRRTTVSLTAARGRGKSA AIGLALAS
AVAFGYSNMFVTS SPENLHTLFEFVFKGF DALDYQEHIDYELVQSTNPEFNKAVVRINI
FRDHRQTIQYIHPSDAHKLGQAELLCIDEAAAIP LPLVKNLLGPYLVFMSSTINGYEGTG
RSLSLKLLQQLRQQSV PQVTSTKTAEAGGATTAVMGRMLSEV TLEESIRYAPGDAVEKW
LNKLLCLDATVAQNITGGCPLPESCELYYVNRDTLFSYHKASEVFLQRLMALYVASHYKN
TPDDLQLLSDAPAHHIFCLLGPVDPSSPALPEVYCVLQVCLEGDISKSTIMSSLSRGKRA
SGDLIPWTVSQQFNDHEFASLSGGRVRIATHPDVQGMGYGRKALSSLVQYYEGKIPNLS
ENGQTETSHLGIPTESEADNLLAESITPRANLP LLLLK LGERPAERLDYLGVS YGLTAQL
LRFWKRSCFVPVYLRQTANDLTGEHSCIMLRILGEIDAEQSDDWL KAYTVDFRRRFLSLL
SFQFRSFPASLGLSILQDKSLKDHNVLSRAELDFTLTNYDMKRLDMYSRNMVDHHLIMD
LVPAVARFYFLRKLDISLSAIQSA ILLGVGLQHKTDFDQLEKELELPVGQLLGLFN R ITRK
VVQHLNSVVEAAIEGEIVKQKDVHMEPLAKSLNQDLEEA AKEME VKQKKELEKLVDF S
EYAISGKEDDWQEA LSGKGT HNLVSI PSSKRKSLPTSDSGQQKHKHKKSKKHRHSTK

>D_gigantea

MVLKKIDNRIRVLIENGVQLRHRSLVFI VGDKGRDQVVILHHMLSKATIKARPTVLWCYK
KELGFSSHRKKRMKQLQKKIKSGTLDLNKDDPFELFISSTNIRYCYYNETHKILGNTYGM
CIIQDFEALTPNLLARTVETVEGGGIIVVLLRTMTSLKQLYTMTMDVHARYRTEESHQDVV
GRFNERFILSLSSCKQCMVIDDRLNVLPISSQTLKITPLTPVTKEESVSPNDVELRELKS
SLEDTPPVGVIVNCKTLDQAKAVLKFI EAI SEKTLRRTTVALTAARGRGKSAALGLAMAA
AIAFGYSNIFVTPSPENLHTLFEFVFKGFDALDYQEHLDYELVQSTNPEFNKAVVRVNV
FREHRQTIQYIHPGDSQKLGQAELVVVDEAAAIP LPLVKSLGPLYLVFLSSTINGYEGTG
RSLSLKLI SQRVQSSGLGTRTTSESKNSNDTNTGSSASRILSEVSLNESIRYSPGDAVE
DWNHLLCLDATI VPKFSSGCPLPDACDLYVNRDITLFSYHKASEVFLQRIMSLYVSSH
KNTPNLQLLSDAPAHHIFCLLGPVDPAKKTLPEVFCVIQVCLGEISKSSIMNSLARGK
RASGDLPWTISQQFNDFEFP SLSGGRIVRIATHPDYQGMGYGKRALQLLIQYYEGKIAD
LSESTVQPTDQEI QPTDLNESGDLLKETIGPRNNLPPLLLKLNERCAERLDYLGVSYGLT
NDLLRFWKKSSFVPVYIRQTANELTGEHSCIMLRDLHHVDENGTSGWLLAFWTDFFFFRFL
SLLSFQFGLFSATLALSVLQNKVFGVSNQGLTQSELYSVLTKYDIKRLDMYSKNLVDYHL
ITDLLPAVSRLFQKTNFHLSGVQSAIILMGIQLQHKTMEDLEKELQLPVSQLLGLFSRV
IRKVVKNFNEI LESAVESEMTKEKEIIMEPLAKDLNEELEIEAEI KEKQKEKTEMLKSM
DLSQYAIAGKDNNDWNEALKGKGPSQMISIASDKPEKRKSLALEESTSSKNKRKKQKKVRT
G

>E_muelleri

MSRKKVDNRIRVLIENSVALRHRSMFVVVGDKAKDQVSSQPRGSFVATANTIAFNSSSSH
RKKRMRQLQKKLKSGLTDTKDDPFELFVAATTIRYCYYSETHKILGSTYGMCVLQDFEAL
TPNLLARTIETVEGGGMVVFLLKSMASLKQLYTTLMDIHARYRTEAHNDVVGRFNERFLL
SLGDCSHCMVIDDQLNLLPVSSHMLKLEPVRAKTKEESMTPEQLDLQALKHSLADTQPVG
VLVECAKTLQAKALLQFVEAI SEKTLRRTTVALTASRGRGKSAALGMAIAAAVAFGYSNI
FVTSPSPENLKT VFEFVLKGF DALHYVEHEDYEVIQSTNPEFNEAVVRVNI FHEHRQTIQ
YIHPADSHKLEVASSELLVCIDEAAAIP LPLVKNLLGPLYVFMAS TINGYEGTGRSLSLKLI
DQVQSSVTVGSASSSGGSAFAGRTLREVTLEEP IRYSGADPVERWLHGLLCLDASSIRR
VGVGCPPPQDCELYHVNRDITLFSYHKASEEFLHRLMALYVASHYKNTPNDLQLLSDAPRP
PRLCPPWARGPFHVL PARDPVCAAQFVYVDVSSCDLTRQFADNNFASLSGARVVR IATH
PDFQSMGYGSRALSLSDYYEGRFPLSEEEADRQSGVVNVALDNDEEGEDGGLMEGGG
ATGQAAPAVDEL DGAEAQRGWTKPGRGIWNDGKLFKVAPHGSCGFWKKAGYLPVYLRQTE
NELTGEHSAIMLKALASSSSGSSDWL KAYTIDFRGRFASLLSYQFRQFKPVVALSVLHTA
SCFDTRSACAYEELTRVLSQQDLKRLDLYSRNMADHHLITDLLPTLAKLYFWSKVPFNLS
AVQASILLGMGLQHKTVLEVLEKELGLPSSQIMGLFNRSVRKFS SHFTEVVVEGVVGRGIPL
PTDAGMAPLEHTVDQELSEEVKAFKKKQLDLKEQLE TADLEQYSVSGKEVDWAGVTTDKP
GIISMARTGHKRLPTKPSHKEKEGKAKRKKQK

>U_maydis

MRKQLDPRIPTLIRNNVALNHRSFVIVGDKGDQIVNLHFLLSQSRVQSRPNVLWCYKK
DLGFTTHRKKREKIKNDIKRGIREKGQDPFELFVSLTDIRYCYKDTPKILGQTYGML
ILQDFEAVTPNMLARTIETVEGGGVVILLKTMSSLRQLYSLGMDVHRNYSNASDDDPV
ARFNERFLLSLGANPD TLLDDELNVLPLSKAKDIQPLPETSTGAGTGVGSTVRKGERV
EADEELAELKQVRETKVVGQVVRHAKTLDQAKAVLTILDILASSSLSTTVALTAARGR
KSAALGLCIAAAVAHGYSNIFVTSPSPENLKT LFEFVFKGLDALGYDEVADWDLQRGTGE
WKDVVVRVNI FRGHRQTIQYIQPDHQVQLQAELVVIDEAAAIP LPLVRNLMGPYLVFLS
STINGYEGTGRSLSLKLIQQLRDSARGIADATDADDSTASSKAVRKQAKGGLNTGRAAG
AALAARSLKEVELKEPI RYSRGDKIESWLHQLLCLDASLTRLSSAALKAKGCPHPSSCDL
YMNVRDALFSYHPASEVFLQRMMALYVASHYKNSPNDLQLMSDAPGHRLFVLLAPLKGNE
GGLPEPLCVVQVALEGNISRGVVLNLSRGTREAGDLIPWLVAQQFQDADFASLSGARVV
RIAVHPDYARMGYGARALQALEAFYSQGLLDVDNVRDDLDDGETFAAVRDRKISKDANLL
QGDEIRVRDAARMPALLQRLSERRPEQLDWLGVSYGLTPQLFKFWKKAGYTPLVWRQIAN
DLTGEYTTVQLKSLDTSISTTGSAWLGSLAADFRKRFISLLSYRFREFSTITALTVLEAA
TQGARLAADDDSLPASSTSLSLGAAELR TLLTPFDMKRLDSYSNNMVEMSVVDLLPTLG
ALYFNNRLRAVREDEATASAI DAQDDEEELRLSGLQSSLLLAIGLQRKTPDEISAELRLP
LQQAMALFVKT VRLLVKSLRKVEKRDIVRSMPELSSGLEARAPLRKKANSARGTDQSDDW
TALKGDLQSELRDAGREFLAGNKESEAYGKLDQEDDHE SKDGADQQEDKEEEGEEQD
DVDEDDDDDDDELDTAKQKLIDSMDLSKYAIKDGADGANWSQAEAEVANMLRKNNGND
LKGFN TTVSVKGTKRAADDKVEASPKKAKASNKSGERKKQKRR

>S_pombe

MPKKALDSRIPTLIKNGCQEKQRSFFVVVGDRARDQVVNLHWLLSQSKVAARPNVLWMYK
KDLLGFTSHRKKRENKIKKEIKRGIRDPNSEDPFELFCSITNIRYCYKESKILGQTYG
MLVLQDFEALTPNLLARTIETVEGGGIIVVLLHKLNSLQLYTMSMDIHSRYRTEAHSDV

TARFNERFILSLGNCENCLVIDDELNVLPISGGKNVKALPPTLEEDNSTQNSIKELQESL
GEDHPAGALVGVTKTLDQARAVLTFVESIVEKSLKGTVSLTAGRGRGKSAALGLAIAAAI
AHGYSNIFITSPSPENLKTLEFVFKGFDALNYEEHVDYDIIQSTNPAYHNAIVRVNIFR
DHRQTIQYISPEDSNVLGQAELVVIDEAAAIPPLVLRKLI GPYLVFMASTINGYEGTGRS
LSLKLQQLREQSRIYSGSGNKSQSHISGRTLKEISLDEPIRYAMGDRIELWLNKLL
CLDAASYVSRMATQGFPHSECSLYRVSRTLFSYHPISEAFLQRMMSLYVASHYKNSPN
DLQLMSDAPAHQLFVLLPPVDLKNPKLPDPICVIQLALEGSI SRESIMNSLSRGQRAGGD
LIPWLISSQFQDENFAALGGARIVRIAVSPEHVKMGYGTAMQQLLHEYFEGKFI SASEEF
KAVKHSKLRIGDEEIENTALQTEKIHVRDAKTMPLLLKLSELQPEPLHYVGVSYGLTPS
LQKFWKREGYCLYLRQTANDLTGEHTCVMRLRVEGRDSEWLGAFQNFYRRFLSLLGYQ
FREFAAITALSVLDACNNGTKYVVNSTSKLTNEEINNVSFESYDLKRLESYSNNLLDYHVI
VDLLPKLAHLYFSGKFPDSVKLSPVQQSVLLALGLQYKTIDTLEKEFNLP SNQLLAMLVK
LSKKIMKCIDEIETKIDIEEELGSNKKTESNSKLPEFTPLQQSLEEEELQEGADEAMLALR
EKQRELINAIIDLEKYAIRGNEEDWKAENQIQKTNGKGARVVS IKGEKRKNSLSDASDKK
TKEKPSSKKKFRK

>A_bisporus

MRKQLDPRIPILINNGVKKNHRSFIVLVGDKGRDQIVNLHYLLSQARVSARPSVLWCYKK
ELGFTSHRKKREAKIKRDVKGIREPNEQNPFIFVTVTDIRYTYKESHKILGNTY GMC
VLQDFEAITPNLLARTVETVEGGGLIVLLLKTMSLKLQLYTMTMDVHSRYRTSSHDSVIA
RFNERFILSLGSCDDCLFLDDELNVLPISRGKDI TPLSDTPGKEKEISSELKDKDSLAD
TKPAGDLVKLSKTIDQAKAILTFIDAIAEKTLSSVTTLTAGRGRGKSAALGLAMAAALAH
GYSNIFVTSPSPENLKTLEFVFKGLDALGYEEHLDYDIAQSTNPDFNKAIVRVNVFRGH
RQTIQYIQPEDAHVLGQAELVVIDEAAAIPPLVLRNLIGPYLVFMASTINGYEGTGRSLS
LKLIIQQLRESTRPSLTKDLAAGTSKDDDSTAATTIASSSKPLSKAPPKSRTLREIKLEI
PIRYANGDKIEKWLNLGLCLDAAIVPKSIGQCPHPKCELYVNRDRLF SYHPASELFL
QRMALYVASHYKNPNLQLMSDAPAHHLFVLLPPLKDDNDGGSGGVGNETYLPPELVV
IQVALEGSISKNVILEGLRAGGDMIPWLISSQFQETFRFGMLSGARVVR IACHPDYA
NMGYGSRALQTLNSFYSGELFNLDESHTVQSELPPDALVNIDP SDSLLTEKPSIRAVNS
MPPLLQRLTERKPENLDYLGVSYGLTSPLLRFWK RAGYVPLYLRQTQSDLTGEHTCVMVR
GLNSSTEPELEWLSEFTKDFRFRFVLSLSYKFREFG SVTALS VLEAAHSGIRAKEGESKD
EIQANELSLLSPFDSKRLESYANNMLDYHVILDLMP IVAGLYFLKRLGEEVKLTAVQSA
ILLAVGLQRKSVEDVETELQLPVSQALALFAKL IKKISKRLVDIQKAAISAELPLQSTGV
VGISRADDGTKEIKEKKATEGWKSVEVLEDELNEAGDEATKRLREKQREMIDALDLQKY
TIDDVEMDWSSAEAQIRGGGTNGYSTVVS IKSGAEGGKSSGQKRKAVEEAEDDSKSGKE
AKKTRRGKNRSTSLRLGK

>N_crassa

MSQKVIDSRIPALIRNGMQEKKRSFFVVG DHSKEAIVHLYYILSNANMKHNKSVLWAYK
SKLLGFTSHRKKRENKIKKEIKRGVREANSEDPFELFISLNDIRYVYKETE KILGNTY G
MCILQDFEAMTPNILARTIETVEGGGLVILLKGMNSLKLQLYTLSMDVHSRYRTEAHDDV
VARFNERFILSLGSCSCLVVDDELNVLPISGGKAVKALPRPTDDEEVNPAKKELEKIKE
SLEDTQPIGSLVQLARTTDQAKALLTFVDAIAEKTLRNTVTLTAARGR GKSAAMGVAIAA
AVAYGYSNIFITSPSPENLKTLEFVFKGFDALDYKDHADYSIIQSTNPEFNKAIVRVNI
HRNHRQTIQYIRPQDAHVLGQAELLVIDEAAAIPPLVKKLMGPYLVFMASTISGYEGTG
RSLSLKLIKQLREQSRTGAGVKDAVEVDRTSGKAAKDTEVFTGGRSLKEITLSEPIRYA
QGDVPEKWLNTLLCLDATLPRSKLNTQGC DPDSQCELLHVNRDRLF SFHPVSEKFLQQMV
ALYVASHYKNPNLQLMSDAPAHFLVLTAPIVEGRLPEPLCVIQVSLEGKISKQSILN
SLSRGSQPAGDLIPWLVSSQFQDDEFASLSGARVVR IATNPDYVSMGYGSRALQLLDVYY
EGKFANLSEEAGTVVPKSAPRVTD AELANASLFD DIKVRDMNELPPLFAKLAERRPEPLE
YIGVSYGLTQPLHKFWKRASFAPVYLRQTPNDLTGEHTCVMRLRPLGDGTDPSWLGFAND
FHRRFLSLLSYKFRFPAILALTI EESASVGA KLDP SNEPTLEKAELDKLLTPFDHKRL
ESYANGLLDYHVLDLIP SIAQLFFEGRIRSNLQLSGLQQA ILLAIGLQRKEIDA IAGEL
TLPSSQLLAIFIKIMRKISTHLASLVEGAI AKEMPQQIGVSAENASAAHDEI IENNFVP
LDQNLDDLAEGGDQAMKAIRAKQRELIDSLPLDQYEIDGGDAEWKEAEKRVAKATKEGK
SNPLVSVKTKRKAEPVYEEKSEKRDRGDKKGGKVRHDKRK

>P_anserina

MFSYLLLAYIAHLDQYQLTFLPHRDSRIPTLIRNGLQEKKRSFFVVG DHAKDAIVHLYY
IMSSMEVRHNKQVLWAYKNKLLGFTSHRKKRENKIKKEIKRGIREANSEDPFELFISLND
IRYVYKETE KILGNTY GMLILQDFEAMTPNILARTIETVEGGGLVLLKGMNSLKLQLY
TMSMDVHSRYRTEAHDDVVARFNERFILSLGSCDSCLVIDDELNVLPISGGKGVKALPPP
EEDEPLSKTKLELEKIKDSLQDTQPIGSLVKLAKTTDQAKALLTFVDAIAEKTLRNTVTL
TAARGR GKSAAMGVAIAAAVAYGYSNIFITSPSPENLKTLEFVFKGFDALDYKDHADYS

IIQSTNPDFNKAIVRVNIHRNHRQTIQYFRPQDAHVLGQAELVVIDEAAAIPLPLVKKLM
GPYLVFMAS TVSGYEGTGRSLSLKLIKQLREQSMTGGNPNNGSGTAEVDRASGKTAKETAG
IGGRSLKEITLSEPIRYAQGDSVEKWLNTLLCLDATLPRSKLSTQGCPDPSQCELLHVNR
DTLFSFHPVSEFLQQMVALYVASHYKNSPNDLQLMSDAPAHFLVLTGPISEGRLEPEPL
CVIQVSLEGGKISRQSILNSLARGSQPAGDLIPWLVSQQFQDDEFASLSGARVVRIATNPD
YVSMGYGSKALELLVSYEYEGKFANLSEDDDDVTMAQAAPRVTRAELEERGLFDDIKVRDMS
ELPPLLPNWQRG

>A_niger

MPRKAIDSRI PALIRNGVQEKRSFFVVGDRAKDVI VHLHYIMSSVDVKQNKSVLWAYK
KDLLGFTSHRKKREAIKKEVKGIREPNQEDPFELFITLNQIRYVYKETEKILGNTYG
MCVLQDFEAMTPNLLARTVETVEGGGIVL LLLKGMNSLKQLYTLSMDIHSRYRTEAHDDV
TARFNERFILSLGSCDSCLVVDDELNVLPI SGGKSVKPLPPPESIDDTNIGTKKELKEIK
ESLAESQPIGSLVNLARTVDQAKALLTFADAIAEKTLRSTVALTAARGRGKSAALGVAIA
AAVAHGYSNIFITSPSPENLKTLEFEVFKGFDALGYLDHVDYTI LQSTNPDFNKAIVRVN
IHRQHRQTIQYIQPQDAHVLGQAELVVIDEAAAIPLPLVKKLMGPYLVFMAS TINGYEGT
GRSLSLKLIQQLREQSRGGVKVSGEDDIDVADRATGKAAKNIDKSLAGRSLREITLSEPI
RYAPGDSVEKWLNVKCLDATLPKSRMNTQGTTPHPSQCQLLQVNRDTLFSFHPVSEKFLQ
QMMALYVASHYKNTPNDLQLMSDAPAHQLFVLVPPIDEEATKLEPEPLCVIQVALEGRISR
QSVLNSLSRGQRAGGDLIPWLVSQQYQDEDFAGLSGARIVRIATNPEYLSMGYGSRALEL
LIDFYEGKFTSLSEDIPEPQEEEMVRVTDEELTNSNLLDDNVKVRDIHTMPPLFGKLSERR
PDALDYVGVSFGLTPSLHKFWKRSSFSPVYLRQTPNDLTGEHSCVMLR TLTSGSNDAAWL
GAFSRDFHKRFSLLSYQFREFPVLSLSICESANS GAKLDPSISQSPLTKAELDAAFSP
FDLKR LDSYANNLLDYHVILDMVPTIAEYFSGRLSGKVSLSGVQQSILLAIGLQRKNLD
TIEKELTLPSSQLLAMFLKIRKVVSTYFRNLI EGAVAETMPAEKAPSARVTTEAHDELDP
ERFKPLETSLDDELREGGQEVNEELREKQRALIDALPLDKYEINHGSAAWEEAEKQIRSG
GASTVSVKTSKQTKRKKGESAREIYDQEIDSKRQKI IKKGTEGKKKKH

>S_cerevisiae

MAKKAIDSRI PSLIRNGVQTKQRSIFVIVGDRARNQLPNLHLYMMSADLKMNFV LWAYK
KLLGFTSHRKKRENKIKKEIKRGTRVEMDPFESFISNQNIRYVYKESSEKILGNTYG
MCILQDFEALTPNLLARTIETVEGGGIVVILLKSMSSLKQLYTMTMDVHARYRTEAHGDV
VARFNERFILSLGSPNCLVVDDELNVLPLSGAKNVKPLPPKEDDELPPKQLELQELKES
LEDVQPAGSLVSLKTVNQAHAILSFIDAISEKTLNFTVALTAGRGRGKSAALGISIAAAA
VSHGYSNIFVTSPPENLKTLEFEIFKGF DALGYQEHIDYDI IQSTNPDFNKAIVRVDIK
RDHRQTIQYIVPQDHQVLGQAELVVIDEAAAIPLPIVKNLLGPYLVFMAS TINGYEGTGR
SLSLKLIQQLRNQNTSGRESTQTAVVSRDNKEKDSHLHSQSRQLREISLDEPIRYAPGD
PIEKWLNKLLCLDVTLIK NPRFATRGTTPHPSQCNL FVVNRDTLFSYHPVSENFLEKMMAL
YVSSHYKNSPNDLQLMSDAPAHKLFVLLPPI DPKDGGRIPDPLCVIQIALEGEISKE SVR
NSLSRGQRAGGDLIPWLISQQFQDEEFASLSGARIVRIATNPEYASMGYGSRAIELLRDY
FEGKFTDMSEDVRPKDYSIKRVSDKELAKTNLLKDDVKLRDAKTL PPLLLKLSEQPPHYL
HYLGVSYGLTQSLHKFWKNSNFVPVYLRQTANDLTGEHTCVMLNVLEGRESNWLVEFAKD
FRKRFLSLLSYDFHKFTAVQALSVIESKKAQDLS DDEKHDNKELTRTHLDDIFSPFDLK
RLDSYNNLLDYHVIDMIPMLALLYFGDKMGDSVKLSSVQSAILLAIGLQRKNIDTIAK
ELNLP SNQTIAMFAKIMRKMSQYFRQLLSQSIEETLPNIKDDAIAEMDGE EIKNYNAAEA
LDQMEEDLEEAGSEAVQAMREKQKELINSLNLDKYAINDNSEEWAESQKSLEIAAKAKGV
VSLKTGKRTTEKAEDIYRQEMKAMKKPRKSKKAAN

>D_discoideum

MVRKKVDTRIRTLIENG VATNHRSFVIVGDNQKQV PNLHYILSKSIVKARPSVLWCYK
EDLGF SKHKKKMKLKEKSKSSGIDSVNQEDPFDVFI STTNIRYSYSESHKILGNTFGM
LVLQDFEAITPNLLARTIETVEGGGLIVL LLLKTMDSLKQLYTMTMDVHTRFRSENSKGEV
VCRFNERFLLSLGKSEQCLVMDDELNIP I SSQSRSIEAKQQILETPEQVELREFKQQVK
DTDIAGALI ENTKTMDQATALLTFIDSI SEKTLRSTVTLTAGRGRGKSAALGLAISAAVA
FGYSNIFVSSPTPENLNTLFQFVFKGFD S MEYVEHVDYELVKSTNPEFHDAI IRVNI FR
HRQTIQYIQPQDYQKLGQAELVVIDEAAAIPLPLVKNLLGPYLVFMSSTINGYEGTGRSL
SLKLIKQLREQSTVVSNSNKALNSITGRMLREIELNEPIRYSARDPIERWLNELLCLDS
TIAKSSPTGCPHPSACQLYVNRDTLFSYHKASEAFLQKMVGLFVSSHYKNSPNDLLLMS
DAPDHHLFVLLGPI DENNTTGLPEILCAVQVSLEGEIAKESILNSIKRGYQASGDLIPWT
LTQQYQDEDFPRLSGVRIVRIATHPDYQKMGYGSKALELLTQYYQGEITNLDEVNQEEEE
EDEEKVDQNKGSSKISTVKKDNASLLTEVIRPKSNIPPLLFKLSERKPEKLHYMGVSYGL
TQQLYQFWSKSKYLPVYLR LTSNDITGEHTC IMLRELNNEQNNTICKDGLWQSFHQDFKK
RFINLLGYEFRNFNSSIALNILEYKKNNTVAVASSTTKNELTQNEMLLFSSYDLKRLE
SYSNNIVDYHVVIDLLPSLSKLYFTNKLIKIEDISLIQSAILLALGLQHKTVDNLIGELNL

ASNQVLSLNFQTMRKINTELKQKQEKFIQDSMPKLSIVAPRTGQFKSFKGDIKETDMIPL
AEDMESELEKGAEEVVNKLKHQLENDSSLSKYLVKGNDEDWSKALKPGSIPNSITIKRKS
DENEETDKKENNKKSKTKNNNNNNNNNNKKNVNNQKSKK

>P_polycephalum

MVKKKVD SRLRTLIENGVK TGHRSMFVI VGDNARDQVVNLHLYLVSKASVQARPSVLWCYK
KDLGFSTNREKRMKQRKAMFRKQADPDMNDPFDLFISSTSIRYTYSESHKVLGNTYGM
MVLQDFEAITPNLLARTIETVAGGGVIVVILKTM TSLKQLYTMTMDVHNRFRQDESEDDV
GRFNERFLLSISVCPTCLVIDDELNLPVSSFARTIKPIEGRMVGDE TIAQRELRELKNS
LKDVHPVSELVGVAKTVDAQAKAIMTFVESITEKTLRSTVVLTAAARGRKSAAALGVAMAAA
IAFGYSNIFVTPSPENLATA LFEFVFKGFDALEYKEHIDYELLRSTNEQFNKAIVRVNI
KSHRQTIQYIQPQDYQKLGQAELVVIDEAAAIPLPFVKNNLLGPYLVFLSSTINGYEGTGR
SLSLKLIKQLREGSAAGNSGGRVLRMCKL TEP IRYTPGDEVEKWLNGLLCLDATVAPP
IA TGYPDPAVCQLYVNRDTLFSFHKQAETFLQRMVALYVSSHYKNSPNDLLMSDAPGHHL
FVLLGPVDVTKNALPEIMCVIQVALEGRIVKDAVIAGLARGVRSAGDVIPWTISQQFQDY
EFGRLSGARIVRIATHPDYHRMGYGTRALELLASYYQGDIIISLNEEDEKEEEEEEEEE
QGTLLTEAIKPRKLLPPLLTQLSERPPEKLHYMGVSYGVTQLYRFWSKLSFLPVYMRLT
INETTGENTCVMLRELQNAGLDINAGWLDAFANDFKRRFITLLGFEFRRFPASLALAILK
KEHYVELESNPNEEVSNEELARKDIKEFFTDYDMKRLHAYA QELVDYHVITDLLPALARFY
FLDRFGSAKVAPTQAQILIAMGLQYKNVDETSKEMGIPANQVLALFNKSI RRI SRFIRGM
QEEEEEELEGGKEETTENGTLKEEAKKEEAKKEEGNEEEEDDENKKQVRDSMLESLKLGAY
SIKGSTDDWKSALKGKTNLSSITLSTKAKGDAQDNKFAKRPNGKKEKPKGQKRSKH

>A_thaliana

MRKKVDERIRTLIENGVKLRHRSMFVIIGDKSRDQIVNLHHMLSKAVIKCNPSVLWCYKD
KLDISSHKQKRKSKQLKRLRERGLDPEKLDAFSRLLDVGRVTHCLYKDSERILGNTFGMC
ILQDFEALTPNLLARTIETVEGGGLVVLILRSLTSLTSLCTMVMVDVHDFRTESSHSEAA
RFNERFLLSLASCKACVVMDDDELNLP LSSHIRSI TQVPTEK DSEGLSEAERDLKSLKED
LSDDFPVGLSLIKKCTLDQKAVVTFDDA ILDKALRSIVALIASRGRGKSAALGLAVAGA
VAAGYSNIYITAPSPDNLKTFFEFVCKGFDAL EYKEHLDYDVVKSANPDFKKA VVRINIF
KQHRQTIQYIQPHEHEKLSQVELLVIDEAAAIPLPVVKSLLLGPYLVFLSSTVSGYEGTGR
SLSLKLLQQLDEQSRPATGLESGCLFKKIELTESIRYSGSDPIESWLNGLLCLDVATC
LPNPACHPSPSQCDLYVNRDTLFSYHKDSELFLQRMALCVSSHYKNSPNDLQLLADAP
AHHLFVLLGPVDESQNKIPDILCVIQVCELGKISENSALQSLRDGHSPYGDQIPWKFC EQ
FRDTEFPFGSGARIVRIAVHPNAMKMGYGSAAVELLTRYFEGQIAPISEAEDKVDVEHAP
IKVTEAAEKVSMLEEQVKPRTNLP LLLVPLHRRRPEKLHYIGVSFGLTLDLFRFWRKHN
F APFYSAVTGEHTCMLLKLKNDLEVNESDELGFFTPFYKDFKIRFSKLLSDKFKKMDYK
LAMSVLNPKINFAEVDSSGSSGGFLKTLNGLSPYDMERLRAYTENLTDNFNVYDI CKT
LAHQYFEEKLPVSLSYVQASILLCLGLQETDFSSIERQMQLERGQIHSLLLKVARELYKY
LNGVAGKEIKSALPRLKERELTAHNVSVD DDIREGAKQVEEQMKKEKIEGLMDSELQQYV
IGDKEAEALQHSKISSSGIISVKSTKSENENGFDKSTKKRSSDKRSSSSSKSKSSKKRKS
LKE

>O_sativa

MRKKVDERIRTLIENGVRERQRSMFVI VGDKSRDQIVNLNYMLAKSRVKS RPSVLWCYRD
KLEISSHKKRAKQIKKLMQRGLMDPEKADPFLFLETSDITYCLYKDSERVLGNTFGMC
ILQDFEALTPNLLARTIETVEGGGLIILLRSLSSLTSLYTMVMVDVHERFRTE SHTQAAA
RFNERFLLSIASCKSCVVMDDDELNLP ISSHMKFIQPV TNNEDSEGLSERERELKDLKDQ
FREDFPVGPLIGKFTMDQKAVINFLDSILDKSLRSTVALLAARGRKSAAALGLA IAGA
IAAGYSNIFVTAPSPENLKT LFDVCKGMNALEYKEHLHYDVVKSADPELKKATI QIN
VY KQHRQTIQYLKPHDHGKLSQVELLVIDEAAAIPLP IVKSLLLGPYLVFLSSTVNGYEGTGR
SLSLKLLQQLESQSQPSAPNNGPNSSRLFKKIELNESIRYASGDIESWLN D L LCLDLAN
SIPNISRLPHPKCEDLYVNRDTLFSYHKESEIFLQRMALYVASHYKNSPNDLQLMADA
PAHHLFVLLGPVDESQNLPDILCVVQVCELGQISRKSAMKSLSEGRSPSGDQIPWKFC EQ
QFQDNVFP SLSGARIVRIAVHPSAVRLGYGSAAVDLLTRYEGQMTLFAEDEEENE EPEV
RITEAAEKASLLEETVKPRANLP LLLVHLRERRPEKLHYLGVSFGLTQELFRFWRKHN
FY PFYVQGI P SAVTGEHTCMVLRPLNSDDIEVNES SKCGFLDPFYQDFRQRFRRL LGTSFRH
LNFKLAMS V LASKIDFS DHEPSDYTNITSKILGDM LSPHDMKRLEAYSNNLVDYHLILD
LVPILAHQYFSEKLPVTLHGAQA AVLFCMGLQDKDIGATKEELGIEREQVLSNFIKTMKK
LYGYLHNIAGKEIEATL PRLKEIDTAPLKS LDEDLDEAAAREVKEQRR AIDEDD VDPKFLQ
KYAIDADDDEIEKALNGGKISASGVISVKSNTKADKQEKREKEMKSKRKGNDGEKSESK
KKRS

>A_trichopoda

MRKKVDERIRTLIENGVKSRRSMCVI VGDKSRDQIVNLYHILAKSVVKS RPKVLWCYKK

ELFLSSHKKKRMKQLKMMQRGLLDPEKDDPFLVSTSTEIRYCYSETENILGNTFGMC
ILQDFEALTPNLLARTIETVEGGGLIVLLLLSSLSLSSLYTMVMDVHERFRTEESHSEATA
RFNERFLLSLASCKACVVMDEELNIPVSSHRSINPVSVKEGSEGLSEGLDLKLNKEQ
LHDDFPVGPLVKGKRTLDQGKAVITFLDAILEKTLRSTVALTAARGRGKSAALGLAIAGA
IAAGYSNIFVTAPSPENLKTFLQFVFCNGFDALDYKEHLHYDLVQSTNPDFKKAIVRVNIY
KQHRQTIQYIQPDHEKLSQVELLVIDEAAAIPLPVVKSLGPLYLVFLSSTVNGYEGTGR
SLSLKLFQQLQEQQSAVNNESGVSVSGRAFRKIELSESIRYSSGDP IEAWLHGLLCDAT
NFIPNITGRLPHPNCELYYVNRDTLFSYHKDSETFLQRMALYVASHYKNSPNDLQLMA
DAPAHHLFVLLGVPVDESKNVLDPILCVIQVLEGQISRASVMKSLSEGYQPSGDQIPWKV
SEQFQDKVFPPLSGARIVRIAVHPSVLRAGYGSAAVELLIRYYEGQMADMAEIDSEQVEE
KTPVKITEAAEKVSLLEENIKPRANLPPLLMLRDERRSEKLHYISVSFGLTQQLHNFWSK
HKFVFPFYIGQIPSAITGEHTCMVLRPLNNDLEVELTGSGKAGFFDPFYQDFRRRFRLLGS
SFRTLNSALALSILDPKINFGEQAPSEEPSISGILGMKAMLTYPDMERLDAYANNLADYH
LILDLVPILARLYSEERIPVLSYSQAAVLLCMGLQHQDISYLQGLKQLGKQQILSLFN
KVIKKIHKYLHNAAAKEIESTLPRLKEVVMTPHKSLEDEDIDEAAREVIEKMKGESELGG
MLKPELLQOYAIVDREADFEDALLKGERKISSSGLLSVKSNRKNVDKSKLKEKEKGENR
GVAHSSSRPHKSKA

>N_colorata

MRKKVDERIRTLIENGVKCHRSMFVIVGDKHHYQIVNLHMMMRQATVKPMTSVLWCYKK
NLQLSSHQKKRMKQIKKMRQRGLLDVEKEDSYLLFLDSEKIRYCYSETENILGGTYGMC
VLQDFEGLTPNILARTIETVEGGGLVLLLLSSLSLTSLSLYTMVMDVHKRFRTEAHPCAAS
RFNERFLLSLALCEACIVMDEELNIPISSHMWSITPVPALGEGGLSESELELKNLKEQ
LHDDFPVGPLIGKCRITLDQGKAVISFLDAILEKSLRMTVALTAARGRGKSAALGLAIAGA
VAGGYSNIFVTAPSPENLKTLEFISKGLEALDYKEHTHYELIRNPDYKNATTRVNVFKQ
HRQTIQYIQPDHEKLSQVELLVIDEAAAIPLPVVTSLGPLYMIFLSSTVNGYEGTGRSL
SLKLFNSKLEAENHLLAKKDDKLSLGSGLFRKIELNESIRYAVKDP IESWLHGLLCDATNQ
IPRLNSSGLPPPSECDLYYVNRDTLFSYHKESEIFLQRMALYVSSHYKNSPNDLQLMAD
APAHHLFVMLGKVDKSNQPLDILCVIQVCLEGQISRESVIKSLSHGSQPSGDQIPWKFS
EQFQDNAPFSLSGARIVRIAVHPDIVGSRYGSAAVDLVTRYEGQLTNI SEAECDPKYD
EFENASLKVTEAAEKVSLLEDSEIKRRANLPPLLHVHQERRPEKLHYLGVSFGLTWGLYGF
WAKNKFAFPYISQIPNAVTEHTCMMLRPLNTDDVELNGANEWGFLDPFYQDFRQRFIRL
LGSSFRSLHPALALRILEPRVGAKKDIRAEAKEPTSTSNGLGLEATTLTPYDMKRLESY
TDNRMDFHLILDLPVLSHLFFGSKLPVELDYIHKIILLCIGLQNDFTYVKGVRIP LPC
REGHRFAKEVQLSGEQITASIFKVMTRIYRYLYGLATKEVESTLPRLREVVMPLPDSRVE
DDLNDAAKQVVEKMNKEEQDYLLNPDMLKQYVIVGQENEIEAALLKGQGKISSSGVVSVK
SSRRKEENGKVKFKNGKKRENDRGKFNPKKRKV

>S_moellendorffii

MRKKVDSRIRTLIENGKSRHRSFLVVVGDKNRDQVVNLHYMLSKAVVKSRSVSVLWCYKK
DLFLSSHKKKRMKQIKKMMKRGLLDAEKDDPFLVASTSIRYCYAETEKILGNTFGMC
VLQDFEALTPNLLARTIETVEGGGIILLSSLSLSSLSLTSLSLYTMMDVHSRFRTESHEDVVA
RFNERFILSLASCKACIFMDEELNIPVSSHVRSISSLPVLEGEGLTESEKDLKELKDS
LVDTQPAGPLVSKCKTLDQGKAVTFLDAISEKTLRSTVALTASRGRGKSAALGLAIAGA
VAFGYSNIFVTAPSPENLKTLEFVCKGFDALDYKEHIDYDLVQSTNPAFNKAVRVNIF
RQHRQTIQYIQPDHAKLAQAELLIIDEAAAIPLPVVKALLGPLYVFMCSSTVNGYEGTGR
SLSLKLIQQLRSQKSESAVSGRVREVELAEP IRYGAGDPIEGWLHELLCLNAADHIPS
IEGRLPPPSECRIFYVNRDTLFSYHKASELFLQRMALYVASHYKNTPNDLQLLSDAPAH
HLFVLLGVPVDETQNTFPDILCVIQVCLEGKISRKSALKSLSDGSSPSGDQIPWKVSQQFQ
DSEFPPLSGARIVRIAVHPNLKAGYGSAAVNLVRYEGQFTDLEEEDEDDVVHKHPV
KVAEAAAEISLLEEQIKPRANLPPLLVLHERNAEKVHYVGVSYGLTQQLHNFWKLLKLV
PVYVSQIPSAITGEHSCVLRPTEIGEVDTATAEQGSWVRPFNEDFSRRFISLLGSTFRS
FSPGLALSVLKPDIDFDEESQETISTEGILGGKNALLTPYDMKRLESYANNLADYHLIID
LVPTISRLYFMRTIPVTLASQAAILLCIGLQHDISFVESALDLPRNQVLALFNKSVRK
LHGQLHKAABAKEIETLPRLREVVMNPHAVSVDEDELEAAAKVKADQKEKVS SMLNPESL
QHFAIGGNDVDFEAALGGKVPKSGLVSVKSSSENKGEAKRPPQNKQLKNKSSKSR

>P_Patens

MRKKVDPRIIRMMIENGKSKQRTMFFVVVGDKGRDQVVNLHYMLSKAVVKARPSVSVLWCYKK
DLFLSSHKKKRMHQIKKMMARGLLDPEKDDPFFALFMASSTNIRYCYADTHKILGNTFGMC
VLQDFEALQPNLLARTIETVEGGGIILLSSLSLTSLSKLYTMTMDVHARFRTESHGDVVG
RFNERFILSLAACKSCLVMDDELNPLPLSSHTKSITPLPNLEGEEMLEADKDLKELKDS
LRDTQPAGLLVQKCRITLDQAKAVITFLDAISEKTLRSTVALTAARGRGKSAALGVAIAGA
VAFGYSNIFVTAPSPENLKTLEFIFKGFDAEMEYKEHIDYDLVESTNSAFNKAIVRVNIF

RQHRQTIQYIQPKDHEKLAQAELLVIDEAAAIPLPVVKALLGPYLVFLCSTVNGYEGTGR
SLSLKLIIQQLREQSNKLSNGSTDGGSVPGRTFRELELAEP IRYKAGDPIEAWLNELLCL
DAANHIPNIEGRLLPHPQECQLYVYVNRDALFSYHKESEMFLQRMALYVASHYKNTPNDLQ
LMSDAPAHHLFVLLGPVDDTQNRMPDILCVLQVCLEGEISKQSALRSLSEGNLPNGDMIP
WTISQQFQDPEFSPSLSGARIVRIAVHPSLIRAGYGSHTAVDILSRYYEGQLSDLTELEDDA
AEPKDSQKLVTEAAESEVSLMEETVKPREDLPSILTLVSRKKPEKLHLYLGVSFGITQALFN
FWQKRKFVVPVYVRQTPNEITGEHTCIVLKPLQSDDEVEGVAGVSGGWLSPFNQDFRRRFMS
LLGFQFRMFAPALALSILDPKITFTEQETMVGASTENIRGLTPLLTPYDLKRLESYASN
LVDYHLILDVPTVARLYFLERLPTSISYQAAIILLCLGLQYHQLEHVEAQLKLPQSQIL
ALFNKTIIRRLYNQLHKAASKDIEAALPRLREVTMPHETTVEDEDLISGAKKVEADIRSKM
EENVLDPEKLRFAITGADADFDEALGSGKIPKSGYVSVKSSKRDPGKDDQKQVKGAEKAK
RKKGADFKRPGGKSKH

>M_polymorpha

MAGLPSQQPQGAAGRGRGMTVPAWMKVTPAGSPPAASPTSGPPQAHIPGNTPPSGPLP
PPRMPGLGSPGAPPNYGVPQMQRFPQGGPPQPGQQFMPSSGGQQYRPAAGPPPGMGPPQHGM
PNMQGMGPQQGMPPSSQPSQYSPSPMQMLPRPPQQPQQSSQGPMSYPPQARPMAGG
PPPQQQPQPHGMPPSGPGNPGMQPPSSFGFNSSLYPYQAPNSYGMTPPHYQSPMQPQS I
QPMQSSQPPWPMIPGQSMPPALPQGLQAAAQPPAGASSLPGATALSAQSSVPGPSDWQE
HTAPDGRYYHNKRTKLSSEWEPVELMTPTERADASTVWKEFTTADGRKYYNKNKTKQSK
WTMPDEMKLAREQVDKSTATTTSGSSVAEVVPPSVTLGTTPTGTNPVSFLPPAASP KPSV
IVNPSPVFLTPPVALPIAPPVTTVPAFTSSTPSGIPPSVAALPAESVKEDSVETSAKEEA
EETSADLEEAKKAMPLTGKINITPFVVEEKPSVAAEPLTYANKTEAKNAFKELLESVHV
ESDWTWDQAMRVI INDKRYGALKTLGERKQAFNEYLAQRKKLETEEKRLKQKKAREDFLA
MLEESKELTSSMRWSKALTLFEEDPRLQAVEKDREREELFEDYLVDLERKEREKAREERK
KNLAEYRSFLESCDYIKANTQWRKVQDKLEDDERC SRLDKIDRLETFOEYIRDLEKEEEEE
EKKLQKEQLRRKERKRNDRFRNLMEEHKVAAGTLNARTQWRDYLVKIKDRVEYQAVANTS
GSTPKELFEDFAEELDKQYHEDKVRAKEAMKAGKISVTSTWNFDKFKAAIAEAGDLEAIA
EPNLKLVFDDLLERAKKDEKEAKKRRMLDDLTDVLRMTKQVTASSKWEECRPLLEETP
EFRAFTEESSRKKVFEDYVTHLQQRAGEKDRKREKEKEKVGPEVAQNGAKKEKEKEKEKER
ERRKEKDKDKDKDKDKDKDKKEREREKDKDKEKSKREKEVVDVSETGNHVVHKEEKKK
DKEREKDRDRERDKDKKYRKRHRSSSEDAGSDREDKDKEDSKRSRRHGIDHKKSSRKV

>C_braunii

MRKKVDSRIRMLVENGVKTRQRTMFVIVGDKGRDQVVNLHYMLSKAVVKARPTVLWCYKK
ELYLSSHRKKRLRQIRKMMQRGLIDPEKEDPFLFMASNTNIRYSYYSETHKILGNTFGMC
VLQDFEALTPNLLARTIETVEGGGAIVLLLSALSLSRLYTLAMDVHARFRTESSHSDVTG
RFNERFIMSLASCRTCVVMDDELNILPLSSHIRELQPVHQLEGEEGMSEAERELKELKES
LADTQPIGSLVAKCRTLDQAKAVITFLDAASEKTLRSTVALTAARGRGKSAALGVAIAGA
IALGYSNIFVTSPSPENLKTLEFVFKGFDAIEYKEHIDYDLVESTNPAFNKAIVRVNVF
RQHRQTIQYVQPQDNQKLSQAELVVIDEAAAIPLPVVKALLGPYLVFLSSTVNGYEGTGR
SLSLKLIIQQLREQSAKAGDPASATNGGSEAVVAGSGRVFKEVELSEPIRYAQGDPIEAWL
HELLCLDAANHIPRISGRLLPHPNECELYVYVNRDTLFSYHKASEVFLQRMALYVASHYKN
TPNDLLLMSDAPAHHLFVLLGPVDESKNALPDILCVLQVCLEGEISKKSAMKSLAEGNQF
TGDLPWTISQQYQDSEFPSPSLSGVRIVRIATHPDLVRAGYGSRAIELLARYYEAQMTDLG
EQEEDDEEEEEKEDLPTDKAKITEKAEAEVSLLETEQVKPRTGLPPLLVS LAERKPEKLHYM
GVSFGLTQPLLNFWRKMKHVPVYVRQTPNDITGEYTCIMIRPLQSDDELQQSDGEAWGSAM
WLSPFLEDFKKRFICLLGVSFQKFPALALSILDPKIVFSEQESSEGV RAPGVVSGMAL
LSPYDMKRLDSYSRNLVDYHLILDVPTIAQLYFLEKIPATVSYGQAAIILLSIGLQROEL
SEVEKSFATEFPALQQRPTIRW

>C_reinhardtii

MRKKVDARIRTLIENCAKTRQALLVLVGDGGRDQVVNLHYMLSKASVKARPSVLWCYKK
ELHLSHKKKRMKQIKKLAQRGLLDPEKEDPFALFVASTNIRYCYHETQNILGNTYGMA
VLQDFEALTPNLLARS IETVEGGGLVLLLSNLNSLTQLYSMTMDAHSRFRTESSHQDVVG
RFNERMVLSLASCANCLLLDDELNVLP TSSLI RYIEPLPTNP DGTVVDDPARSARQELGE
LSGSLADTQPAGALVGR CRTLDQARAVVTF L DAASEKTLRSTVALTASRGRGKSAALGLA
IAGALALGYSNIFVTAPSPENLR TLEFVFKGLDSL DYKEHIDYDLVESTNPSFGKAVVR
VNVFRNHRQT VQYIQPQHAKLAQAELLVIDEAAAIPLPVVKQLLGPYLVFLCSTVNGYE
GTGRSLSLKLIIQQLREQGAKISAGDNKAGSKAAGGADGTAAP IANGTSGRTFREVHLGEP
IRYAPGDPIESWLHELLCLDAGSHMPKPPRLPHPDCELFVVERDTLFSYHKASEKFLQ
QMVALFVASHYKNTPNDL LMSDAPAHQLFVLLAPVDQTANTMPN I LAVAQVALEGSISK
RSAVNSLAKGELPQGDLPWTVVSQQFQDADFPALSGGRVVR IAVHPELGRAGYGSRVLEQ
LRRYYQGELTSLDEEDDAGGRGGAGAGGSGSGEEDGDEARRRS GADAAANADGDAPLLSE

RLAPRKGLPPLLVLPLSDRKPDRHLHYLGVSYGLTQQLFTFWRKAGYQPLYLRQSASDTTGE
HTCVMVRPLEHPEVEGTAWLSPFVSDFKGRFTSLLGGAFREMAPALALSILDPQLNWGEE
EAQRATQAGVVVHRGDGSLDPYDLKRLQAYCSSLVVDYHLVLDLVPPLAAAYLRGRVPVS
LSYGQAAIMLVLGLQLREVSALLESGLPLSNQVLALFNKAMRKV

>G_Sulphuraria

MKKKIDARLKTLLLEVSCQNNHRALLVIIGDRGKNQVVNLHYMLSKISNIPKPSVLWCYKK
ELGFSSHRKKRMKQIKKLAQRGLLDTSRDDPFEIFVSNTHIRYCYAETHKILGNTFGMC
VLQDFEALTPNILARTVETVQGGGLVLLGTIQSLRQLYTMAMDVHDFRFTQAHKDVVA
RFNERFMLSLSRCSCLIVDDELNVLPSSPMKEGTFALSDKQEQIKKQQQEWSELKSN
LEETPILGELISRTKTLDAQALLTFIEAITEKTLRSTIALTAARGRGKSAALGLAVASA
IGHQYSNIFVTCSPENLKSFFDFVLIGFDALQFKEHVDYEVLQSTDASLQHCVVRINVF
RSHRQTIQYIDPRDASKALHQAELLVIDEAAAIPLPVKSMLGPYLIFMSSTITGYEGTG
RSLSLKLFQLRKGAHSDNRPSDFHLAEGGRTFREITLEAPIRYAIGDPIESWLYDVLCL
EAGKGPQPLIYGCPHPHDCQLYRVNRDALFSGHPLAESFLQTIMALFVSSHYNKSPNDLQ
LLSDAPAHRLFVLIAPTCKQLDSNGLPQVLCAIQVCLEGGQISKEYAQASMSRGRGAGDL
IPWTISQQFQDSSFASLSGVRVRIATHPDVQRMGYGSRALTLNDYYSKMI SLTEDGW
RGTLSDEPSTITNVDNSNDTLNHERLEPRKELPPLLEPLNVPDRVDYLGVSYGLTYELFC
FWKKQGMVPLYIRLVANEYTGHEHTCIMLKQLPEQSKMSWLSNFFHDFIRRFLALLPFAWS
SWNPALTLQIIDYGKEYYASHQVTMLNEHMHLLSTFDIQRLEAYAKNLDYHMIVDLL
PTIAKLYFSGQLDKEIHL SAAQSAILVSVGLQYMDLDRLEKVLNVPAQQLLALLNKMIRK
FAHYFKELEWSLMESHDASTIITETKTSMKDTNETLEPKQAPLVKKPKKNMEHKDKKEHR
KHRLE

>C_crispus

MKKKVDPRIRQCLQEGVNNHRSLIVLVGDKGRDQVVNLHYMLSKMTLQSRPSVLWCYKH
ELGFSSNKKRMKIKKRIKQGLVDPNRDDPFELFISSTSIRFAYYSESHKILGNTYGM
VLQDFEALTPNLLARTVETVQGGGLAVILINMMSLKKLYSITMDVHARLRTEAHRKVV
RFNERFILSLADCPAAIVADDELNVLPISSHVRAMSGEVTSTANRPTTENEPKELKES
LADTPPIGPLIAKAKTLDQAKALLRFEEAISEKTLRSTVALTASRGRGKSAALGLAVAAA
VAHDYANIFVTSPPENLRTFFAFLVFGFDAMDYQEHTDYEIIQSTD SGLNNAIVRVNVF
RSHRQTIQYIDPIDAPRSLQAELLVIDEAAAIPLPVIESMLGPYLVFMCSTITGYEGTG
RSLSLKLMTKLREGSARVPSQYKGVNGNSNRQSDRDAAAVSSRTGSGRVFREVQMEIP
IRYQKDPVEAWLYDLLCLEAGGVRQVLTGGAPHPEECELYYVERDALFSRHSASEAFLHRI
MALFVSSHYNKSPNDLQMLSDAPAHGLFVLLAPTRPDTERLPDVLCAIQVCLEGGKISSKS
SKTQLSRGQRASGDLIPWTISQQFQEPDFATLSGARVVRVATHPDVQKMGYGTALKLLL
EYYKGEHNI SLENNVERLSEQNPPAGQENGYTDSEQGDADQKKTLLEIIGPRKNIPP
LLRRLHERSAENLDYLGVSYGVTTQGLYSFWSRLGFMPYLRLTPNDLTGHEHTCIMATTSV
CSSEETKPSDWLSKFNQDFRRRFVHLLGYDFRSLPPGLALAILNTGAEFRESGDAEERS
SIDEIKGLFTPYDRRRLSRSRSLIDYHVILDMIPKLAELYGTGSLGLRDGLRLSPAQNAV
LMGIGLQRCVDDLEKRLGIQASQVLALLNKVIRKMSTFLRDMEETDIASRIGQTQGARG
MIPGVTIELPNSNKAFAALSTDDPSDEKGTGPKSLKAEEDGAGATEKGTMSNFKVGGS
KDEWKAALPKNALADGIPSTVSIKSGKKSKRKTISESKDVGTPRKRKKKQVK

>E_siliculosus

MVRKKVDARVRSLENGVKSNNHRSTFVLVDHGDKQVVNLHYILSKARVASRPSVLWCYK
KELGFSTHRKKRMRLKQKKSGLASAGGASEDPFELFVNSTNIRWCYKDTERVLGQTF
GVLVLQDFEALTPNLLARTIETVEGGGLVILLKTVTSLKQLYAMSMDVHSRFRTESSGD
VVGRFNERFMLTLGDCKSCLVCDDELNVLPSSHARDIVPLPPVEDAETTEAKELRELK
ESLEGTTPVGPVLAGAKTLDQARALLTFMEEAISEKTLRSTVALTASRGRGKSAAGLCLA
GAIAYGYSNIFVTAPSPENLSTVFEFVLKGFLLHFSEHLDYEVIIQSTNPAFNRAVVRIN
VFKDHRQTIQYIQPHDSEKLSHCELLAVDEAAAIPLPVKSLLGPYLVFLSSTVNGYEGT
GRALSLKLIDQLRKQGTAAQAARIAASEIHGTGRKKGKEKVHENRWKVAADAAANASK
HGGGGGGRTLREINLETPIRYGANDPVEAWLYRLLCLDAKVGHHRLVCGTPAPKDCELYY
VNKDALFSYHTLSEGFLQRVMSLYTSAHYKNQPNDLQMLSDAPAHHLFVLLGPKAMEQGA
KGNLPDVLVVQVALEGQISRASLEASLSRGRASGDLIPWSLSQQYQDQGFAGLSGAR
VRIATHTDVQKMGYGSRAIDLTKYYEGGLFSGTGGEEGSEEGSEESSEEESEEEEDDD
DDDDESSDEEEENGNTANKKKKASAKGGGLSSERLAPRKKLPPLLVSAGERPPERLHWL
GSSFGVTGQLRLWKRAGFRLVYLRQTTSDLTGHEHTGVMLRALDTEGMKEKPEDGWLASF
AADRCRIVSLLGHEFQGFDCCTTALSLEVAAPGGAAVEGGEAPAAGGARGAVGARELGY
CGPHDLKRLGLYSRNLVDYHLITDLLPMVSKLYFLGRMPSTRMSHLQEAILLAMGLQHR
VTALSKELQLPDNQLALFNKAVRKL CNCLTEVAEKEAEGEMETHLAKKKRLQTVKSAGE
MSMLPASMEDELEGGAKEAMKKMEGKQFAVMY

>P_tricornutum

MVRKRLDDRVRALLERSVVTGQRSMLVLVGDHKGDKQVPNLHQILTKCSVQARPKVLWCYK
KELGFSTHRKMKKLRDKSRGLVGGEADQADNFELFVSQTDITWCYYKDSHRVLGTTV
GVLVLQDFEALTPNLMARTIETVAGGGLVIFLLRTVKSLKQLYAMSMDVHARYRTE SAGD
LVPRFNERFILSLGKCPNCLVCDDELNVLPVSRKALNDLSPNAGWSKGDAGEVIVQDTPE
QRDLKEIQEALLDTPHVGVLVELTKTLDQAKALLVFLEACSEKTLKSTVAMTAARGRGS
AAMGLCLAGAISLGYSTICVTAPEPENLVSFDFLCRGLKALKYQEHMDYSVTYNSASGR
EQTKCITAINVHRSHRQVIQYVDP AETDKFTSAEIVAIDEAAAIPLPVVRALMSHPDRLT
FLSSTINGYEGTGRALRGRHAEMQAASSAANSIVGAKSKKGEAKVHEQRWAAAAAILEA
SEEIELLTPIRYAHGLDSVEAWLNKLLCLDCGSASNKLKNGGAPAPGDCELYSVDRDALFS
FHKLSEAFLLQKVMGLYTSAHYKNSPNDLQMLSDAPAHSLFVLLSPSAEQDANSLPDVLT
VQVALEGRISRKAVEAQLARGHRSAGDLIPWTISQQFGDSKFAQLSGARIVRVAVHPVQ
GMGYGSRAIELLYRFYNEEMVSLVNDENDDADSDAERNGEESDNDEPTTSGIGILGEN
LRPRKELPPLLLPLTEVDMPRLDWVGTSFGLTLQLHKFWSRSGMRMLYLRQTKNELTGEH
SSIMVRALPRRSGVDDSWLYAYLSDARRRFTTLFSGPFRHLDVRLALSVDNMDVPSNTT
EAKQRAGALAGTLTFKELDYFLTPYDLKRELELYGRNLCDHHLVMDLLPIIGRLYFTGRFG
SDFNLSVQAALFCGIGLQNKSV DILTRELGLPTNQVLAMFNKAVRKMSIALNSVVEEKE
KESLLTGEKRSRIEESAQMRHVSRTLDEDAEQAGQEA IATLRANEMANHLPELAHDT
MLKYVVKGSQKQWEKVLQDKDKREKLLTMTT

>T_thermophila

MRKKVDQRIQILMENCIKLRQRGFFIIVGDRGKEQVVIHYLLNKMNFKSKPNVLWCYK
ELGFSTNKQKRMKEIKKLQKGLYDNNVDDPFQLFVTSTQIRYCYKETHKILGQTYQML
ILQDFEAITPNVLCRTIETVEGGGFVIMLFKTMSSLKQLYTLTMDVHNRRTDQFQDVEP
RFNERFILSLSKSENTIVMDELNILPISNLIENIKSVSADDSLDSKMTKELEELKESLK
QNVLVGPFINI AKTLDQARAIMTFIDAITEKNTRATISLTAGRGRGKSAAVGVAIAAAIV
CGFSNIYVTAPSPENLKTLEFIFNGLEQNNYKENMHYDI IQSTNPDFNNAVIRVNVFKT
HKQIIQYVSPDFASIGHADLVVIDEAAAIPLVYVKQFLGPHLVFLCSTVNGYEGTGRSL
SLKLIQQLKQSKMNSSNSEKQTA VSSSGSRILRELTLNDPIRYSNDPIERWLNLL
CLEASQAVPLKSGLPHPNDCELFMVNRDTLFSYKksteQFLFNLMSLFIS SHYKNSPNDL
QLLSDAPSHAVFVLMGKIKEGEKKSEMPDILCAVQVALEGKISKETFNQONENRVRPAGD
LIPWIISEQFQDREFPSLSGVRVRIATHPAAQKMGYGSKSLELLQKFFEGELIDLSNET
LNRENYFLKAQRTADGEKLLKPLKRLTEIEPPQLHYLGVSFGLSKELFKFWKSGYYPL
YLRQAKNDITAEHSCIMLKSLLKQNDMENENVQENKYNDQWLKYYSDDFKKRFLQLLSY
EFKDVDVQLSLGVVDPFTVTDNDNSRLTNDLGLGDKDKNLGNYIMAPISNKEQLENYIN
TYDLKRLEAYSRLADFYLIIDLLPEISKMLFLNRLQSIKMSYSQCAIFLGLGLQHKTE
KLARELGIPVSQVMAIFNKGIRKVTQLFRSLFEQEIMKELNFGNSKEALEKMNP IKDNEN
EIQDAGKEYLKSLLKDNKKDEKEDKSDKLNKKRHARYNFGTEDIDVGNASINQNTGTI
TIARDSKVALNQDFTNYKSASAAQAAAVKNAKLLKKN

>P_tetraurelia

MKKKLDSDIQAALVQSGINEHIRSIFVMIGDRSKYQVATMHQLLSKLSMIQKPKVLWCYK
DLGFTTHKKKRETEIKNQOMKEETQLNEPFLNLFSSNEIRYCYAETQKVLGNTYNMLIL
QDFEALTPNILCRTVETIAGGGIILFLIKTMSSLKQLYSLTMDIHSRLRTPSFKEVQPRF
NERFIIISLRCQRCLFVDDDELNLLPVSSTALEKIEQGTPINSNQVADFAQTPNEIAQKL
IKNCKTLDQAKIILGLIDSLQEKVFRYTTISISAARGRGSAAIGMAAAAALKLGFSNIFV
TAPSPENLVTFEFLISSLETLDLFRQNQHFDVVIQSTNPDFNNAVIRVNAYKTHRQFVQYI
QPQDFVKVANNADLVFCDEAAAIPIQFVKSMFGSHLTVLSSSTINGYEGTGRSLSLKLLQK
LRETKQTGARLLKELQMNDPIRYSNKDIVEKWL YELLLL DADAPKLQKGLPHPNDC
ELYLCNRDTLFSSQKTSEKFLFNLMSLFIS SHYKNSPNDLQLLSDAPAHAVFVLLGPLPTDK
TMPDILCAVQVCFEGAIPKDLVDQVTNQQKPSGDLIPWTL SKYYLDNQFPQYSGVRVIRI
ATHPNAQKMGYGSKAMELLNRFCKGEILDPSAQLIHIDFREQATAEDEKDLIQPKKLLP
PLLKLLSEVYPPQIDYLGVSFGLTQELYKFWKSNQFSPAYLSLRKNDTTGEYTCIVLRPV
KDEGSGIDNLCKAFSNDFRNRFINLLGYDFKDLPEYLCIDILKANVTTANVDEDEIEIAKD
DNIITKEELRQQLSLSDYARLENYCKSMADFYLILDILPSICKIFFAQKCGKAIRMSRTQ
AALLLGMGLQYKNI ELASEQVDNAVNSLPLFNKAMRMRNF IKMVL DDEQHQRTPKPKQ
QQQPSKQSVIQKRGKVNYQIDGQIDQEEIQQSLQTGDVVTLGRSKDQAYINYKKKVKLG

>P_falci parum

MKKKVDNRKILIEENIAVQQRSMFFVVGDEGKNVVVNFYFLLNRLINRTHNILWCYK
LDFSTSKKKRYREMKKIKKGTFDQTI DNNFDMFLKNADIRFCFYKESKVLGKTY SICI
LQDFSYITPNILCRCL ETVIGGGLIIFL INKQNELKDLYNLTLNYHKKYTMNGICNVYNN
YIHRFFLSLNKCEAMFIDDEMNILPLNENYLKIKKITSKSGDNGIGNVTKLKHMEVQ
KEGEQSGSNINNNDDDDNDNDNYSNSNYSNSNYSNYSHTKSATLGGFLCPDKNLLQEK
LKFLESICEENEKKEEKKEEKFILYSSKFSRNDTCTKINKNGDIHSSDNKYTRDITYETYSV

TNTSTTANTYNNYI SKKTKKKKDVYSFLDRNIMNLLNICLSLDQLEVLNLMCKILRNDEEK
KKNIKEVLFNLLANRGRGKSATLGLLLSLSIYFNYSNIIICSGNNEGVTIYDFLTKGLH
ILGFNEFKDYEKIYDMGKIKEIVIFKNMKYLKQRIFFDIIEEDILNSELMIIDEAACIP
IDILKKKIKGEITILSTTLNGYEGTGKTFTFKLLKQKFKFITQLSYEEFKNMNPLYFDK
AYIDLTLTNPRIYSYNDQVEKWLNDFLCLNCNEPTNIKDNICSPSNCELYFVNKNIFKNY
NRTSEILLKKIMTLFITSHYKNTPNDIIMILDSQQHHLFILINKNNHYNNNNILEDQID
ELDIYGVLHCAIDGIINNGNVKKTVKLEDIHLVKNNNTQKENPTLDTKEINDSNDQKLE
KKENINQINSNDNHKELLINEYEGNLIPIIISDYFNYYFYNYIGIRIVRISIHPSMQNLNY
GSEFLKVKCNYYNEYNKNSKQRHDNNTKQKIKINNNSNNCGDSNNCGDSNNCDDSNNCDDN
VKYSENIILFNCCNGKNIYFDKLLKQIDYIGTCFGLTKGLLIFWQKNKFIIPVFLKQKNEI
TGEFSVLMKLDLNEKHLKHIANFYTDFVRSVTKLLPYSFKYMESEIVFNLLHNNNYSIPL
KQDDAPTIVKDHNTDHTYNHNGHIKNNLHNNHFYDNYEDQHLSFYNNNLINQQTIFYF
FHTNDICRLKRFVLEGRNFYELIYLLSTIANLILFKKVKIDLSFLEYTILYAVAFQKKT
KEISDEIKINVNTNAIFRKIIHRFYNYIKDIMEKQIEKEVNAELNEKLVKKKRSNIE
LPSEYKDELENSLIVKKKLLKKEKLSLMKEFNLEGDIKRKEINIKTNDNLEIKKKKKK
KKKKKQETN

>T_gondii

MRKRVDPRVRTLVENCIQLGQRSFFVIIGDRGSEQVNNLHYLLHRYFPAQKPSVLWCYKK
DLGFSSHRKRKQVKKKIQRGLYDPNIDDPFELFMSSTNVRFCYYKDSHTVLGMTTGMC
VLQDFEAITPNILCRTVETVQGGGIICLLLRSFASLQQLYDLSDIHGRFKTERFSQVKP
RFNERFILSLARCRNCLMVDELNVLPLTSFSQTLRSLKPGMQSEAAEGASAGLPPGLPP
DAAEALAEKKRTEELEALKRDFQETPPAGPLVNLCLSVDQAKAVLAFVDCLAQAASGKK
DAANCAAADTRGNLSALTHLGQVAHQTFALTAARGRKSAAALGLAVAAAAYDYTNIFV
SAPSPENVKTLFEFVVKGLTALGLKEHLDFSVSREDPGLAAAAHASHSSGVSSSENSSR
GGASRLITGVEVHRSHKQKVTFSVHEREGLGHADLLVIDEAASIPLPVKSIFGPYVVL
LSSTIHGYEGTGRALSKLIIQDLKQNKLSLSSNNPKGGFLARTLQTLTTPIRYSVGDG
VEAWLHDLCLDADPPKLSDAALPPASKCDLYLVDRLDALFSFHEASEAFHNLLSLFSV
SHYKNSPNDLMLMADAPAHLLFALLPPVDERTATVPDIYCAVQVSI EGHLSKQAIQQALG
RGLRPSGDLIPWTLAQQFADEDFGALTGARIVRIACHPSLQRLGYGTAAMQQLIDFFEMK
GVSLKTEGKNALSNSLLSSSQGGDEEDEDEDEDEDEDEDEDEDEGGTPAKGGKKGDKQAE
EEQVRDASSAAASASHASSPEAAKVWTKKAFKPRPAPPLSPCRSTRPDFSIDYFGVAFG
LSSSLFRFWTRRNFPVYVVRQTVTDVTGFEFSCILLRPAEVEGSLWPGESDKTGAADDRGG
DEDLRRKNRWQGIETGRAANELVGRGGIWL SAFASDFRARFLQLLGGAFRRLPASLALC
LVTCSKPRDNSEAPGSKETSIFPPLSLANLPNFFSSHDL SRLRKYAQQADLPLILDLL
PTLAALFFNDRFPPLAASSVTLSHLQOVVLLAVGLQRKTPDDVSTEFNIPPNTLALLNK
TVHKFVGFQKLQEAQVEAEVDSQWGVRSAA SRRLVVEGGGALLGELPSQSFTTEELADEA
KKVKRQKVELKNLLEALGGEEMKEHSMDAYTEEDFNEAVGAARPGSLQASTASRVFSL
KRKNPNKGQKNGGTGEKTDGEGEKKKQKRDFGFSFNKNKNKHKVYANR

>G_intestinalis

MAQSVRVNSRGEESLKAIDKRLTTLNLSGYESFQRSIIIVGDGGLYQVPLNLSILKMTRT
SQLSIMWCYDEMPKLSNKRKRNEFRQLKNRGELQEEELSVEAQFYANTNIRYLHYRDT
ELALGTTVDMLIQDFKALTPNIIGQTVCEVRGGGLIFLLNMMHDLSQLHAVVMNFHSR
FSNINNREVSNDLHIKARFNSRFIALLEYLPRFLLVNDTLEVIQITEGIQNLMTNLSAQD
GSVEKRVAAFTAYKNEHSSSLLAKTFISLCATLEQAKLVIQVLEMLDNVDRIRKLADKN
AMLEKEPVERDGGKPREPSRKRYFVREKETYTREDIAESSVDQDVDAASLEEQNAGHELLQ
TLTQNANALATATRFPICSLIAARGRKSASLGLIVALMIFQGYNHVLSAPNPINVQTV
FEFCIIIGLKALDYQEARDYTVRRAGPKATDPIIELRITRKVKQVIAYIMPSTLATYCYAN
ERASSSILNFIDAILLDECAAIPLKLVQRQIVDNVEANMLCFISSTCYGYEASGRALSCLK
LHDLERSGSKRGQGLVKSLENEPIRYAPNDSIEKWNQLLCLDQVNGLVTSLEEGPEIAA
SFPQPQSCTLYRVNKDALFSGLPQSELFLQTIWISIFSASHYKNSPNDLILLSDNPLHEIF
VLCPPITDSKRIPTVLCACQICYEGKLVKQIVKEALSQQKRLDGDMIAWLLSQFQDVVSF
SSLSGARVVRIATNASYMRGGYGSHTLELLSSFFKGELLNPTEIKRLEAKKKERIEKLAK
QKGISINDAQEIEALTPALLQNCCTVYPSDLTWLGVGFGITLDDLNFWRKNQYLPVYIR
QTRNETTGEHTCLMVRSLKENPFINNFAFSFRSRFLRLRYDFAELPIEIAFKVSWPTNT
VALAHDADQVAIRAALTKQDVVRLRGFVRHQLDMAVRDLVPIAELIFTEKLSPVSLSH
ASARIFLIGLMMNRQIDVAQQLGLVSDVAFSQVESIFTKTVEVLLRILEPLVGLGTELG
AGASK

>T_vaginalis

MATRIKKQIPKNLLDLSAGISSNHRSLIIVGPHGTDRIPTINTLLQKKMSTNSKVLWC
YHKELHISSNREKRNRLKQKTQNGTLNEDLATELDLFFANSKQVTFKAYSESHRVLGQT
FNILVLQDFSSLTPNILARTVETVSGGGLIVLLLDDMQSLNDLYTITMDYQRGMMNRDAFG

KVASRFNARFHNSLKSCKNCIAMTDEFDILDSVTPNFEPDSTISEDADLKEILAQYSKDI
ESVNERLEEITDEKQKKQVEHERDEKQQILSIVEKALTADQAKTVGKVIDSIKDRQLGKV
IGITAARGRGKSAAIGLSLAAALALGYSNIFVTAPSIANLTALFEFVIVGLTALGFREGE
SEDFEVVKSNTPTSSKERPYVTRVNVHLSHGRRQTVQYVRPNMTQYLGQCEILAIDEAAAI
PLPIVRQLMGPYTTLFSSTVDGYEGTGRALSCLKLFEKQRTMKSRFTQTTLTKPIRYAEG
DPIEAWLHNNLLLDVHPQECSLIPALSDTEFLINRDALFNRSRGSEELLRALVSLSVES
HYKNTPNDLTLMADSPNHRIFALINMKTVTGKGIPEIIAFVQVALEGKIGDEQFSNAENG
KGRRDGDLIPWNIARAFEEPEFAQLTGVRIVRIAVNPEMQKKGYGSVAIEKLVQYYSGHE
VVDTKIKPNAKTLMPYPLTAIAPEPVDYMGVSFGLTRDLAHFWRKGGFVPLYVVSRETNEVT
GEHSVIVRPIGETEWLPKFTTEFKSNVYRLLPTIFRDVTPFVIEEIMFPPPPSEKDSGD
RARLTDIDVKKLSRWSGDGIGAVKHLLLESVCDIFVDNPPIKLTRTEMFALLIMGKQMD
EEECARLEIQVNQVVALLSAIGKISRYLNKQORGETDSSMNSKIIIEKD

>T_brucei

MPPPTSVASTARSSSGATIKRKVDDRIRKVIENAVHKRHRALLLVGDRGKDQIVNLHQM
VSHANHNAKVNVLWCMKKEPDFGSTSLKQOEKARLEVKGGMSTDATKEAFQTFLSQTIIR
RFCQYKETHKILGQTFGMAVLQDFEAMTPNTLARTIETVKGGGIIVIMFRAMKSLKQLYT
IAMDVHSRYRTEALQDVVPRFNERFLLSLVDCDTSCLVDDDLNVLPITAKMCAVESARTT
SYDETLVLQGRQKHEADLAALKERLRASNEVGPLISLCQTLDQGKTILSLMQTVVEKSLN
TTCVVTAGRGRGKSAALGLATAGAVAQGYSNIFCTAPSPENLQTFEFVVRGLQELGYKE
RTDFEAMQSTNVEFSKCIIRINVFREHRQTVQYISPNDASKFAQAELAIIDEAAALPLPI
VKQMLGPYLVFLSSTISGYEGTGRSLSMKLVADMRRKHSNSTNASVAGDDRRLLRELSMSD
PIRYGPNDPVEFWLNLKLLCLDATVRPIAVKSSPHPNSELYYVNRDALFSYHPLAEQLLQ
AIVAMLVAHYKNQPNDLQLMSDAPGHHLFILCPSTVDAKSGIPDVFCVIHACEEGHVSS
EAIKSNLSRGLRPSGDLIPYTLQYYLEEGFAKLAGLRIVRIATNPQLRAGYGSRALEL
LQYYTGSVSLRPAASAAATENNELKAVDDADEVSSTISHTIKPRKRIPSLLVPLIERP
YEEIDYLGVSFQVTTPLFNFWKRGGFEPLYLRHAANELTGEHSCVMVRSFGFDLGLLRAE
FRRRFIPLLAMPFRHLPTLALSLKDVVDNDPQKLAATDLSHVAEHNVRVAGQPQCTW
KELQEIFISISDLKRLKLNATTFVEGGHVLDLVPALAKLYFESRLFHLPDGAEGVVLSHAQ
AAVLLGVGLQCQTIIEVGGQAVFSGVPTQQLRAFLQKAI SRIVDHFTRLQKLLTGSKEE
ETVKRNGGNDADENTHEIYDKEGRVVGLSVEKRVRSVINVDSTLLRDAKSATVGARND
GGTTGLQRAFKRSKKTRRG

>L_Major

MVKRKVDDRIKTLIDDIARHKHRGLILLVGDRAKDQVNLHLMISRANHNAKVNVLWCMR
EDPDFGSTGKKQOEKRRARLEVKGGLSTEASKEAFQTFLAQTNIRFCYKETHKILGQTFG
MAVLQDFEAINPNTLARTMETVKGGGLIVITFRAMRSLRQLYTIAMDVHARYRTETQRDV
VPRFNERFLLSLADCDTAMCVDDDLNVLPITHKMKSYGKSCKSDAYNADLAVQGRLOHEV
DLANVKETLRPSEDVGPLMQLCQTMQAKTVLSLMQTMKRLDSTCVVTAGRGRGKSA
LGMTIAGAIQGYSNIMCTAPTENVQTLFEFAIRGLKELGYRERTDFEALQGVSEEFK
CFIRINVFRGHRQTVQFVSATDTAKFAQAELCVIDEAAALPLTLVKRILGPYLVILSSTV
SGYEGTGRSLSMKLVADMRRGSSSGAADARHLKELSMSDP IRYGPGDPVEKWLNKLLSLD
ATMENTPLTTSPPHSACELFYVNRDALFSYHPLAEELLQRIQSLVAHYKNQPNDLQLL
SDAPGHHLFVLCADSVESAAAPTSTSGAARQVDPDIFCVIHACEEGQVSAQSIKSHL
SHGLRPSGDLIPYTLQYYLEEGFAKLAGLRIVRIATNPALPRAGYGSRALSLHQQYYSG
SISLTAEPKTAIATHRRGDSEADAREQAGKEHEGDDDESSAPTAMLAPRSHITNLLTPL
VDRPYELVDYLGVSFGLTTELLNFWKKASYIPLYVRQAPNELTGEHSCVMVVRPMGFDLRP
LQREFQHRLPLLAMPFRHLPTLALSLVGDFEAHDPKLSAATDLSEVDHHVVRVDGTV
QATDDIAATFTAGDVQRLRLVSTTFIEGGNVLDLVPVTLAKLYFGKRLFRCPDGTGTVL
SHAQAAVLLAVGLQCQTTEELAAQAFTGVSTQQLRGLFLKALSRLSEHLVALQKIPQRA
QRDGHNAGVVAGKRARSLSDGAVGDSAAEDVEDAEVYDSHGKGLKLVARKVKVQVC

Supplementary data S3

| | | | |
|---|-------------------|--|--|
| Figure 2A | sgRNA_U13_5-Fwd | CACCGAGGGAGGTGGGTCGTCGGGT | sgRNA_1 |
| | sgRNA_U13_5-Rev | AAACACCCGACGACCCACCTCCCTC | sgRNA_1 |
| | sgRNA_U13_7-Fwd | CACCGCCCGTCTGACCTGAACTTCA | sgRNA_2 |
| | sgRNA_U13_7-Rev | AAACTGAAGTTCAGGTCAGACGGGC | sgRNA_2 |
| | Del_U13_UP1 | GGATGTATCGCATTACGGTAGTCACC | Genotyping |
| | Del_U13_Down1 | GCAAACACCACTGGTTCCTACACC | Genotyping |
| Fig 2B | SNORD13-Rev | GGTCAGACGGGTAATGTGC | Northern |
| | SNORD3-Rev | GCCGGCTTCACGCTCA | Northern |
| Fig2C | | TAATGATCCTTCCGACGGTTCACCTACGGAAACCTTGTTACGACTTTTAC | Mung bean nuclease protection assay |
| | | | |
| Fig 2D, Fig3C, Fig5G, Fig6C; Supp Fig5 | h45 rev primer | TAATGATCCTTCCGACGGTTCACCTAC | Targeted ac4C sequencing |
| | h45 fwd primer | TGTACACACCGCCCGTCGC | Targeted ac4C sequencing |
| Fig 3A | sgRNA-zU13-6F | CACCGAATGTAACCACTGCGCAGAA | sgRNA_1 |
| | sgRNA-zU13-6R | AAACTTCTGCGCAGTGGTTACATTC | sgRNA_1 |
| | sgRNA-T7zU13-6F | GCGTAATACGACTCACTATAGAATGTAACCACTGCGCAGAA | In vitro transcription |
| | sgRNA-zU13-7F | CACCGGTAGGACGAGAAATAATAAG | sgRNA_2 |
| | sgRNA-zU13-7R | AAACCTTATTATTCTCGTCTACC | sgRNA_2 |
| | sgRNA-T7zU13-7F | GCGTAATACGACTCACTATAGGTAGGACGAGAAATAATAAG | In vitro transcription |
| | sgRNA-T7_Rev | AAAAGCACCGACTCGGTGCC | In vitro transcription |
| | Del zU13-up2 | CATTTTATCTATGAAAGAAAATCG | Genotyping |
| | Del zU13-down2 | GCCACAGAACTCAGACATAG | Genotyping |
| Fig 3B | zSNORD118-Rev | GTCAGACAGGATTCATAAAGGTG | Northern |
| | zSNORD13-Rev | CAGACGGTCAATGTGCCACG | Northern |
| Fig 5D; Fig5F | O_NB_CD1-droso | GTCGTAACATATTGGTTGGG | Northern , primer extension |
| | ORT_Lh3p | TCAGAGGAGTTAACCTGCTG | Northern |
| | ORT_Lhumile | CGAGTTAAACCCGTTTCATC | Northern |
| | ORT_Amellifera | GAGAACACGATATTCATC | Northern |
| Fig 6C | C.elegans h34 fwd | CACCATTAGGACTGACAGATTG | Targeted ac4C sequencing |
| | C.elegans h34 rev | CTAGAGTCTCGCTCGTTATCG | Targeted ac4C sequencing |
| | | | |
| Sup Fig 2 A/B | 5'ITS1-hs | CCTCGCCCTCCGGGCTCCGTTAATGATC | probe1; Northern |
| | ITS2b | CTGCGAGGGAACCCCGAGCCGCGCA | probe 2 (mix of ITS2b + ITS2d/e); Northern |

| | | | |
|--------------------|-------------------------|--------------------------------------|--|
| | ITS2d/e | GCGCGACGGCGGACGACACCGGGCGT | probe 2 (mix of ITS2b + ITS2d/e); Northern |
| Sup Fig 2 C | RNaseH-2 | TGTTACGACTTTTACTTCTCTAGATAGTC | RnaseH mapping |
| | 3'18S | GATCCTTCCGCAGGTTACCTACGG | RnaseH mapping |
| Sup Fig 3B | zCdh5-F | CGAGATTGCTGATGGAGGAACGCC | Rt-qPCR |
| | zCdh5-R | TGGCGAGGAGGGCACTGACA | Rt-qPCR |
| | z-fli1a-F | CCAAACATGACGACCAATGAGA | Rt-qPCR |
| | z-fli1a-R | GTGATCCGGAGACCACAGAGA | Rt-qPCR |
| | z-bcl2-F | TCACTCGTTCAGACCCTCAT | Rt-qPCR |
| | z-bcl2-R | ACGCTTTCACGCACAT | Rt-qPCR |
| | z-bax-F | GGCTATTTCAACCAGGGTTCC | Rt-qPCR |
| | z-bax-R | TGCGAATCACCAATGCTG | Rt-qPCR |
| | z-puma-F | GAACACACGGGTTACAAAAGAC | Rt-qPCR |
| | z-puma-R | GAAAATCCCAGAGTCTGTAAGTG | Rt-qPCR |
| | z-lhx1a-F | AGTCCGAGAAGAATGCGAAC | Rt-qPCR |
| | z-lhx1a-R | GGCCGTAGTACTCGCTTGA | Rt-qPCR |
| | z-pax6a-F | GCATGTTGCCCACTAGTCCT | Rt-qPCR |
| | z-pax6a-R | CGGGTACAGACACTCCAGGT | Rt-qPCR |
| | z-otx5-F | GATCAACCTGCCCGAGTCTA | Rt-qPCR |
| | z-otx5-R | CTTCTTCTGGGTGGTCTGG | Rt-qPCR |
| Sup Fig 5 | T. thermophila h34 fwd | AGAAGGGATTGACAGATTGAG | Targeted ac4C sequencing |
| | T. thermophila h34 rev | TTAAGGTCTCGTTCGTTAACG | Targeted ac4C sequencing |
| | P. polycephalum h34 fwd | GTATGAAAGTCAAGCTGAAAG | Targeted ac4C sequencing |
| | P. polycephalum h34 rev | CGCTCGTTATCGGAATAAAC | Targeted ac4C sequencing |