

## **Supplementary data**

**Insights into the mechanism of pre-mRNA splicing of tiny introns from the genome of a giant ciliate *Stentor coeruleus***

**Jirayu Nuadthaisong, Tanaporn Phetruen<sup>#</sup>, Chanakan Techawisutthinan<sup>#</sup>,  
Sittinan Chanarat<sup>\*</sup>**

Laboratory of Molecular Cell Biology, Department of Biochemistry, Faculty of Science,  
Mahidol University, Bangkok, 10400, Thailand.

<sup>#</sup> These authors contributed equally to this work

<sup>\*</sup> Corresponding author: [sittinan.cha@mahidol.ac.th](mailto:sittinan.cha@mahidol.ac.th)

## Supplementary figure legends

### Figure S1. Functional enrichment analysis of genes harboring introns in *S. coeruleus*.

Gene ontology analysis was performed on g:Profiler using *Tetrahymena thermophila* (A) and *Paramecium tetraurelia* (B), two closely related ciliates in phylum Ciliophora, as organism input parameter. GO, Gene ontology; MF, molecular function; BP, biological process; CC, cellular component; KEGG, Kyoto Encyclopedia of Genes and Genomes pathway. Padj, a p-value adjusted for multiple testing using the Benjamini-Hochberg method [1].

### Figure S2. Annotated tiny introns of *S. coeruleus* are functional and indeed spliced out.

Comparison between sequences of four different genes harboring introns at the genomic DNA and mRNA level. Genomic DNA sequences were retrieved from the StentorDB [2], the mRNA sequences from RNAseq experiment [3]. Intronic regions are highlighted in red. Amino acid sequences translated corresponding mRNA are also shown below each alignment pair. Note that, for simplicity purpose, U at the mRNA level is shown as T. (A) SteCoe\_2110, (B) SteCoe\_13333, (C) SteCoe\_32708, and (D) SteCoe\_38208 genes were analyzed.

### Figure S3. Results from ‘cmsearch’ program in the Infernal package

(A) Alignment from cmsearch result represents matching between U1 snRNA candidate from *S. coeruleus* and covariance model (CM) of eukaryotic U1 snRNA (Rfam: RF00003). (B) – (E) Similar to Fig. S1A, but CMs of U2 (Rfam: RF00004), U4 (Rfam: RF00015), U5 (Rfam: RF00020), and U6 snRNAs (Rfam: RF00026) were used as queries. Contigs and genomic locations of the *S. coeruleus* U-snRNA candidates are also indicated.

### Figure S4. A simplified proposed model of RNA interaction network during pre-B complex

### Figure S5. Length of predicted spliceosomal proteins of *Stentor* versus human proteins

Comparison of the predicted proteins from *S. coeruleus* versus their human protein homologs. The line of identity (1:1) is shown as dashed line. aa: amino acids.

### Figure S6. Certain parts of the branching factors that are adjacent to the spliceosome active site are noticeably non-conserved.

Sequence alignments of the three branching factors Yju2/CCDC94 (A), Cwc25/CCDC49 (B), and Ntc30/ISY1 (C) are shown. N-terminal regions of proteins that are projected to the spliceosomal active site are enclosed in red dashed boxes.

### Table S1. Functional enrichment analysis of genes harboring introns in *S. coeruleus*.

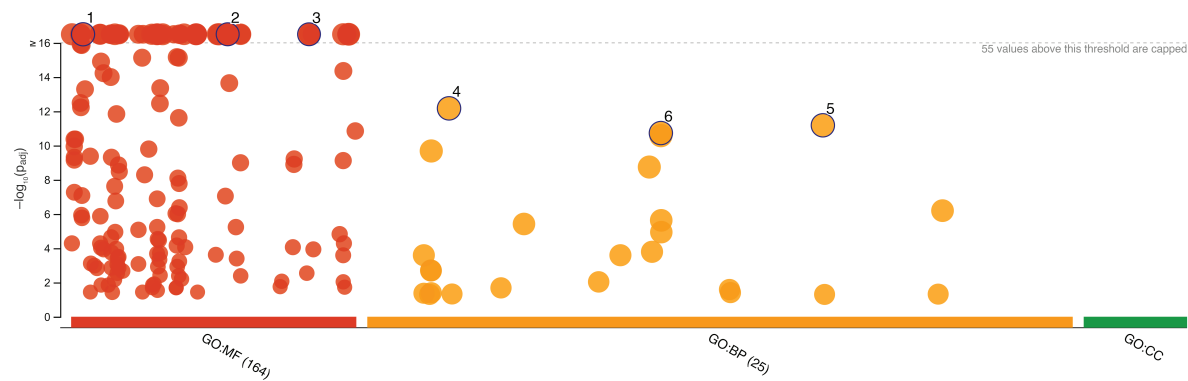
Similar to Figure S1, but the full set of results were shown in table.

## References

1. Raudvere, U.; Kolberg, L.; Kuzmin, I.; Arak, T.; Adler, P.; Peterson, H.; Vilo, J. G:Profiler: A Web Server for Functional Enrichment Analysis and Conversions of Gene Lists (2019 Update). *Nucleic Acids Research* **2019**, *47*, W191–W198, doi:10.1093/nar/gkz369.
2. StentorDB | Stentor Genome Database Wiki Available online: <http://stentor.ciliate.org/index.php/home/welcome>.
3. Sood, P.; Lin, A.; Yan, C.; McGillivray, R.; Diaz, U.; Makushok, T.; Nadkarni, A.V.; Tang, S.K.; Marshall, W.F. Modular, Cascade-like Transcriptional Program of Regeneration in Stentor. *eLife* **2022**, *11*, e80778, doi:10.7554/eLife.80778.

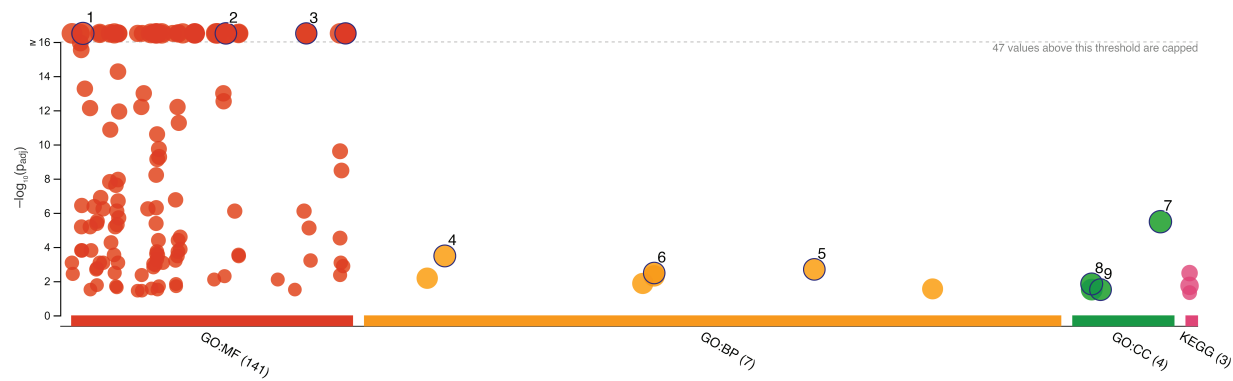
Figure S1

A



ID	Source	Term ID	Term Name	P <sub>adj</sub> (query_1)
1	GO:MF	GO:0003824	catalytic activity	$4.941 \times 10^{-324}$
2	GO:MF	GO:0043167	ion binding	$4.941 \times 10^{-324}$
3	GO:MF	GO:0097159	organic cyclic compound binding	$2.283 \times 10^{-303}$
4	GO:BP	GO:0008152	metabolic process	$6.724 \times 10^{-13}$
5	GO:BP	GO:0071704	organic substance metabolic process	$6.572 \times 10^{-12}$
6	GO:BP	GO:0044237	cellular metabolic process	$1.863 \times 10^{-11}$

B



ID	Source	Term ID	Term Name	P <sub>adj</sub> (query_1)
1	GO:MF	GO:0003824	catalytic activity	$4.941 \times 10^{-324}$
2	GO:MF	GO:0043167	ion binding	$3.324 \times 10^{-292}$
3	GO:MF	GO:0097159	organic cyclic compound binding	$1.395 \times 10^{-253}$
4	GO:BP	GO:0008152	metabolic process	$3.328 \times 10^{-4}$
5	GO:BP	GO:0071704	organic substance metabolic process	$2.037 \times 10^{-3}$
6	GO:BP	GO:0044238	primary metabolic process	$3.316 \times 10^{-3}$
7	GO:CC	GO:0110165	cellular anatomical entity	$3.305 \times 10^{-6}$
8	GO:CC	GO:0016020	membrane	$1.449 \times 10^{-2}$
9	GO:CC	GO:0031224	intrinsic component of membrane	$3.115 \times 10^{-2}$

## Figure S2

A

```
SteCoe_2110_gDNA -----ATGCTGAGCAAGTCAAGGTGGGGCTTGCTAAAAACCCAGTAACTC 50
SteCoe_2110_mRNA ATATCCATAATATCTTTTGAAGAAATTAAGCTGTATGACCCCTCTGATAATATGGCATCATTGTAATAATGTCAGACACAGTCAAGTGGGGCTTGCTAAAAACCCAGTAACTC 120
SteCoe_2110 -----M S E H K S R W G L L K T P V T 16

SteCoe_2110_gDNA AGCCAGCTTTGAAATCTTGAAGTCTCTGATCTCAAGGCAAGGTTTTCATCAAAACCCCTTTAGAAATGATCGTAAGCTATATTAGAAAAATGTTGTTAAGCAATTTATTCG 170
SteCoe_2110_mRNA AGCCAGCTTTGAAATCTTGAAGTCTCTGATCTCAAGGCAAGGTTTTCATCAAAACCCCTTTAGAAATGATCGTAAGCTATATTAGAAAAATGTTGTTAAGCAATTTATTCG 225
SteCoe_2110 -----Q P R F E N L E V P D F Y K A K F S S N N P L E I D Q N T L L S N L F 51

SteCoe_2110_gDNA AGATATGTCAATGTCTACAAATACAGGCCACCTAGCTACGCTGATTTAAATAGAGTTTGGCTTCCAAATCAAAATGATTGTGTGGCTTTATTTGTTAAGCAAGGTGATA 290
SteCoe_2110_mRNA AGATATGTCAATGTCTACAAATACAGGCCACCTAGCTACGCTGATTTAAATAGAGTTTGGCTTCCAAATCAAAATGATTGTGTGGCTTTATTTGTTAAGCAAGGTGATA 330
SteCoe_2110 -----E D M S M S T N T G P P S Y D S F L L P I P N D C V G L I I G K Q G D 86

SteCoe_2110_gDNA TGATAAGCATCTTCGAGTAACCTCAATGCTAGTGTGATACAAAGTAGTAAGATTGTGCACCTGGATGTGCTTACAGAAATGATTTATTGAAGGATCCCTAGACAGCTTTGCCTTG 410
SteCoe_2110_mRNA TGATAAGCATCTTCGAGTAACCTCAATGCTAGTGTGATACAAAGTAGTAAGATTGTGCACCTGGATGTGCTTACAGAAATGATTTATTGAAGGATCCCTAGACAGCTTTGCCTTG 450
SteCoe_2110 -----N I K H L R V T S N A S V I Q V A K D C A P G C A Y I R N V F I E G S L E S Y C L 126

SteCoe_2110_gDNA CAAAAAATGATATACGAATCATGATGAGAACATAAAGCGAAGAGACTTCCAAACATAAATTTGACCCACACACTCTCTCTGAAACATATGTAGAAATTTATGCTCCCTGAGCCCT 530
SteCoe_2110_mRNA CAAAAAATGATATACGAATCATGATGAGAACATAAAGCGAAGAGACTTCCAAACATAAATTTGACCCACACACTCTCTCTGAAACATATGTAGAAATTTATGCTCCCTGAGCCCT 570
SteCoe_2110 -----A K K M I Y E I I D E N N K A K R L P T Y N L T H T L S S E P Y V E I I V P E P 166

SteCoe_2110_gDNA ATATAGAGCTCATGTAGGAGCTGATTTATTAATCTCAAAAATCATCTGAGTCTGGAGCCAGCTTCTAGTCCCAAGAGTGTGCCAGGTACAAATATCTGTTATTTGACCA 650
SteCoe_2110_mRNA ATATAGAGCTCATGTAGGAGCTGATTTATTAATCTCAAAAATCATCTGAGTCTGGAGCCAGCTTCTAGTCCCAAGAGTGTGCCAGGTACAAATATCTGTTATTTGACCA 690
SteCoe_2110 -----Y I E L M L G S D L L L L K K I H A E S G A Q V L V P T K C A P G T T N N R Y L T 206

SteCoe_2110_gDNA TACTGGCCAGCAGAAATGATAGATTAGCAAAAAGATGATTATGAATCTTGAACACATATTTGATATATTAGCATATAGAGAACATATAGAGAACATATAGAGAAC 770
SteCoe_2110_mRNA TACTGGCCAGCAGAAATGATAGATTAGCAAAAAGATGATTATGAATCTTGAACACATATTTGATATATTAGCATATAGAGAACATATAGAGAACATATAGAGAAC 795
SteCoe_2110 -----I T G Q P E C I R L A K R M I Y E I L E P A Y R E H A V A N G V G V I 241

SteCoe_2110_gDNA NCAAGAGCAAAATGATCAAGAGTCTTCTACGAGCTTTTATGCGTGTAGTGTGCAAGATCAATGATTACTAGCTGCTTGTGAGCAACTCTCAACCCAGCACTGCAGCAATGA 890
SteCoe_2110_mRNA NCAAGAGCAAAATGATCAAGAGTCTTCTACGAGCTTTTATGCGTGTAGTGTGCAAGATCAATGATTACTAGCTGCTTGTGAGCAACTCTCAACCCAGCACTGCAGCAATGA 915
SteCoe_2110 -----N H E Q I D P E F P Y C S F Y G L V L K V Q M I T M T A C E A T S N P A T A A L 281

SteCoe_2110_gDNA AAGCTATAAACAACATTTTCAATA----- 916
SteCoe_2110_mRNA AAGCTATAAACAACATTTTCAATAATTTTATAGAGTTTATAATTTTCAATAAATAATATCA 977
SteCoe_2110 -----K S I N K H F Q * 289
```

B

```
SteCoe_13333_gDNA --ATGTATCGACAAAGAAAAAGTCTCTTAAAAAACAATCATATCTCAACAACTCAAAGTTCTGGAAAAAACAATACTCTAATGAACCTGTATAAAAAATCAAAATTTACGTTA 118
SteCoe_13333_mRNA GAATGTATCGACAAAGAAAAAGTCTCTTAAAAAACAATCATATCTCAACAACTCAAAGTTCTGGAAAAAACAATACTCTAATGAACCTGTATAAAAAATCAAAATTTACGTTA 120
SteCoe_13333 -----M Y R Q R K K F L K T K A H I L T I Q R F W K K T Y P L M K K I Y K K I F Y V 39

SteCoe_13333_gDNA AAAACATGAGCAAAATATGCTGAGAAATGTATAAAAAAACAAGACAGGATTTCTATACAAAGCTTTTATAAGAAATATAAAGTTAGGAAGAGTATAGGATAAATAATAGCAAA 238
SteCoe_13333_mRNA AAAACATGAGCAAAATATGCTGAGAAATGTATAAAAAAACAAGACAGGATTTCTATACAAAGCTTTTATAAGAAATATAAAGTTAGGAAGAGTATAGGATAAATAATAGCAAA 240
SteCoe_13333 -----K K L G K Y A E E M Y K K K T K A A I C I Q K F Y K K Y K V R K E Y K D K I Y E 79

SteCoe_13333_gDNA AAGTAAATTAACATGATGAAAGAACTTTTGAAGGACAAACCAAGTATCTCAAGTAGGAACAGGAAAAACAAGCTGTTAAGGTTATTGAAAGGTAAATTTATTAGAGATGCT 358
SteCoe_13333_mRNA AAGTAAATTAACATGATGAAAGAACTTTTGAAGGACAAACCAAGTATCTCAAGTAGGAACAGGAAAAACAAGCTGTTAAGGTTATTGAAAGGTAAATTTATTAGAGATGCT 345
SteCoe_13333 -----K V K I T M H E K L L R R Q S K L V Q D R N R K R Q A V K V I E R E W 114

SteCoe_13333_gDNA TGAAAGAGATTAAGAGCAAGAAATGAGAAATTTAAGGAAGTATTTGGCCACATCTTATGAGTGCAGGATTTAGTTTAAATATAAGCAAGCTTAAAGAGAACTGATGTTTAC 478
SteCoe_13333_mRNA TGAAAGAGATTAAGAGCAAGAAATGAGAAATTTAAGGAAGTATTTGGCCACATCTTATGAGTGCAGGATTTAGTTTAAATATAAGCAAGCTTAAAGAGAACTGATGTTTAC 465
SteCoe_13333 -----L K R M K R Q R M R S L R K Y L A T I P Y E C R V L C L K Y K Q A K R E T D V L 154

SteCoe_13333_gDNA AAGCTGATTGTAAGATTTATAAAAAACAAGTGGATTGGGTGTTATTAG----- 528
SteCoe_13333_mRNA AAGCTGATTGTAAGATTTATAAAAAACAAGTGGATTGGGTGTTATTAGGTTAAGTTATGCTT 530
SteCoe_13333 -----Q A D L N K V I K T S G F G C Y * 170
```

C

```
SteCoe_32708_gDNA -----ATGAAAGAAAGCTCTCAGATCAAGCTACGATATAAAAG 40
SteCoe_32708_mRNA SCGTATAGCTGTATTTTATCAGTAGAGCGCTTTCGAGATTATCTTTGTAAGGATTGATAATAGCTATCCATATTACAATGAAGAAAGCTCTCAGATCAAGCTACGATATAAAAG 119
SteCoe_32708 -----M K E K L S D P R L P I K 13

SteCoe_32708_gDNA TAACTTTTATTAGAGACAGCTCAAAATCAAGGAGATGATACAAAGAAACCAAGTCAAGCTGTAAGTGTGAAAGAAATTAAGCAATTCGATGCTTTATATCAATCATCC 160
SteCoe_32708_mRNA TAACTTTTATTAGAGACAGCTCAAAATCAAGGAGATGATACAAAGAAACCAAGTCAAGCTGTAAGTGTGAAAGAAATTAAGCAATTCGATGCTTTATATCAATCATCC 225
SteCoe_32708 -----R A R S N I K E D D S K K K R P V T A V E L E K I K K F D A F I D Q I 48

SteCoe_32708_gDNA ACAAAATAGACCCCTTCCAGTTCAGCGCTTTGACACTTCAGTCTCGACATTATAGCAAGGCTTCAGAAATCCCAATATGACTGAATTCAGAAATTTATACCGAGAGTATATCAAG 280
SteCoe_32708_mRNA ACAAAATAGACCCCTTCCAGTTCAGCGCTTTGACACTTCAGTCTCGACATTATAGCAAGGCTTCAGAAATCCCAATATGACTGAATTCAGAAATTTATACCGAGAGTATATCAAG 345
SteCoe_32708 -----H K I D P F T V D A F D T S V S I S I G K L S E F F N D T E F Q D L Y R E Y I K 88

SteCoe_32708_gDNA AGCCTCAAAAGCAAAAGAAAGTAAATCTTATTCAGAAATTTAGAGCAAGAAAGCTTTTCTGTTGTTATGAAAGCAAGCTGAGCAAGATTCCTATATCTCAAGCTTTTAC 400
SteCoe_32708_mRNA AGCCTCAAAAGCAAAAGAAAGTAAATCTTATTCAGAAATTTAGAGCAAGAAAGCTTTTCTGTTGTTATGAAAGCAAGCTGAGCAAGATTCCTATATCTCAAGCTTTTAC 465
SteCoe_32708 -----E P Q K Q K K S N S Y S E F T E C K K R F D L V M N K E D P E Q D S L S Q R F 128

SteCoe_32708_gDNA AGATCTACAAAGCAAGCAATACCAATAAATCAAGTTCGAATATGCTTGAAGTTCAGAAATCAAGTTCAGAAATCAAGTTCAGAAATCAAGTTCAGAAATCAAGTTCAG 520
SteCoe_32708_mRNA AGATCTACAAAGCAAGCAATACCAATAAATCAAGTTCGAATATGCTTGAAGTTCAGAAATCAAGTTCAGAAATCAAGTTCAGAAATCAAGTTCAGAAATCAAGTTCAG 585
SteCoe_32708 -----E I Y N K A I P L N N D V Q N M L E F L K D L P R T N A A I E R K S I Q R H A L 168

SteCoe_32708_gDNA AGCTTAAAGCAAGAAATTTAGTCTGAGTCTAAGAGATAGTCTGATGATAAAAAAATGGCGCTAAGTTCAGAAATCAAGCTTCAGAAATCAAGTTCAGAAATCAAGTTCAG 640
SteCoe_32708_mRNA AGCTTAAAGCAAGAAATTTAGTCTGAGTCTAAGAGATAGTCTGATGATAAAAAAATGGCGCTAAGTTCAGAAATCAAGCTTCAGAAATCAAGTTCAGAAATCAAGTTCAG 705
SteCoe_32708 -----E L K R K K F T E S K S I S P D K K I A A K V L E T D P P P S G P L S I Q E K L 208

SteCoe_32708_gDNA TCTGCTGAGCGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCT 760
SteCoe_32708_mRNA TCTGCTGAGCGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCT 825
SteCoe_32708 -----V L L R A L A K H M G D W E K I S K E F K D K P V S G D L L K K I W R C I K V T 248

SteCoe_32708_gDNA TGAAGAGAGAGCTTTGATATCAGGAAAGAGCGCGCTCAATATCAATTTTAAAGGCTTGGGCTGCTGTAGAAAGTTGAGCAATCAAGTTCAGAAATCAAGTTCAGAAATCAAG 880
SteCoe_32708_mRNA TGAAGAGAGAGCTTTGATATCAGGAAAGAGCGCGCTCAATATCAATTTTAAAGGCTTGGGCTGCTGTAGAAAGTTGAGCAATCAAGTTCAGAAATCAAGTTCAGAAATCAAG 945
SteCoe_32708 -----M K E E V I D I R K K A P Q Y H Y F K W L R A A V R K F E Q N K G K K S K S K A 288

SteCoe_32708_gDNA TTTCTATTTAAAGCTTGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1000
SteCoe_32708_mRNA TTTCTATTTAAAGCTTGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1065
SteCoe_32708 -----F T S I N P K P K E Q R F D I L S I M A E A E N I K C S G I E S S T H L N I T S 328

SteCoe_32708_gDNA CAAGTCTTTTAAAGCATACACAAATTCAGGAACCTGCAAGAACTGTCAACTAA----- 1056
SteCoe_32708_mRNA CAAGTCTTTTAAAGCATACACAAATTCAGGAACCTGCAAGAACTGTCAACTAAATTAAGTTAGATTGGCACTTTGATCTAA 1149
SteCoe_32708 -----S S S F K A Y T N S G T L Q E L C N * 346
```

D

```
SteCoe_38208_gDNA ATGAAGAGCTTCAGGATGATCTGGAGAGTTATTGATTTAGAAATACAGGGATATTTTGGCAATATATGATTGCTTTCACATTTTGCATGCTTAAGTCAATCCGCTGTATCAG 120
SteCoe_38208_mRNA ATGAAGAGCTTCAGGATGATCTGGAGAGTTATTGATTTAGAAATACAGGGATATTTTGGCAATATATGATTGCTTTCACATTTTGCATGCTTAAGTCAATCCGCTGTATCAG 119
SteCoe_38208 -----M K E L Q D E S G E L F D L E I T G I F L A I Y G L L S L F A C L K S I R C I R 40

SteCoe_38208_gDNA AATATGAAAGCAGATATAATATAGAAATCTTGTGTTTACTTTATCATTAACAGCTGCACATTTTGAATCTTATTAGTAAGATAGTTTAAATCTACTTTGGCAATATAC 240
SteCoe_38208_mRNA AATATGAAAGCAGATATAATATAGAAATCTTGTGTTTACTTTATCATTAACAGCTGCACATTTTGAATCTTATTAGTAAGATAGTTTAAATCTACTTTGGCAATATAC 224
SteCoe_38208 -----N M K A E Y N Y R I L W F Y F I I N S V T F L R I V W C L N Y F G I Y 75

SteCoe_38208_gDNA TCAGAAATATCTATAATTTTATGCTTATCTTCAGTACATTTGAGTATCTGATATACATTTTATAGTATTTTGGCTGTATATATCATAGATATAGATCTTTTATCTCAAAA 360
SteCoe_38208_mRNA TCAGAAATATCTATAATTTTATGCTTATCTTCAGTACATTTGAGTATCTGATATACATTTTATAGTATTTTGGCTGTATATATCATAGATATAGATCTTTTATCTCAAAA 344
SteCoe_38208 -----S E N T Y N L L D C L S S A T L E L L G Y T F I V F W L D I Y I D I D L F I S K 115

SteCoe_38208_gDNA GAAAGAGCCAGTATATAGTATCACTTTCGAGTGTCTGATGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCT 480
SteCoe_38208_mRNA GAAAGAGCCAGTATATAGTATCACTTTCGAGTGTCTGATGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCT 464
SteCoe_38208 -----E K A Q L Y S I T L R C A Y A G L T L S V F T Y A A H I Y L Y F Q I F S I N S I C T D 155

SteCoe_38208_gDNA TGTGTTATTTACTTCAAGCTTTTATCTCTATATCTGATCAATAGCTTCTATCTATCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
SteCoe_38208_mRNA TGTGTTATTTACTTCAAGCTTTTATCTCTATATCTGATCAATAGCTTCTATCTATCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 525
SteCoe_38208 -----S V I Y F S A P I S I F V S I A L I I S G R L L V K K I N L V Y S Q S I G E K V 195

SteCoe_38208_gDNA AATAAAGAGCTTAAGCTAATAATAATAGCAAGTAGCTTTATCTATATTTCAAAAAGCTGTTGTTTATACAGCTGAATCTCACTATAAAACATAAAATTTGGTTTGTATCGCATTT 720
SteCoe_38208_mRNA N K R V K L I I I A S S F I Y I S K S C L F I T A E F T H K T K N F G F D I A F 235

SteCoe_38208_gDNA ATGTTTATTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCT 840
SteCoe_38208_mRNA M F I Y V I G T E I F L P L C S V L Y F L K P N L L E N S S L N S S W I P S G L R 275

SteCoe_38208_gDNA AGTGAAGAAAGTTTCAATATCAAAAGATCAAGAATTTTAAACCTTTTATGAAAGCAAGATTTTGAAGATTTTATGATATTTAGAATCTGATCATAGGCTTTGAGTTTACCT 960
SteCoe_38208_mRNA S E K S S L S K V S R I F K T F M K S K V F E N T L D Y L E S D H R S L S L P 315

SteCoe_38208_gDNA CTACTAATTTAAATACAGAGAGTTTGAAGACAGAAAGTATAGCTAGCATTTAA 1017
SteCoe_38208_mRNA A T K L N T R E F V K Q K S I A S I * 333
```

### Figure S3

[illegible][illegible]

**C**

```

v NC
U4 1 .....<<<<<<<<<_____>>>>>>>->,,,<<<_____>>>>,,<< CS
AUCCUUUGUC .AGggcGAaUuccccugcCAGUGAGGAUucgCcAGgcggggUauUgCUAgUugAAACUAAuuuccaAcaccGcc 88
C CUUUGUC AGggGCAUA: C: C+G: AGUGA G UU+ C+GA: GG: G: U AUUGCA: UU: AAAACUAU CC: AA: CG+
SteCoe_contig_1964 3831 AGCCUUGUCuAGggGCAUAUUGcGACGAUGAAGCUUA -GCUGAUGCGCGAUAUGUcUAGUUGAAACUAU- CCAAUU- -CGU 3749
*****989999...688. PP

v NC
U4 89 <<<<<<<<<_____>>>>>>>->,,,<<<_____>>>>,,<< CS
uaggccuccuucagaggccgaagcCAAUUUUUGgAAuccccuuaagggg 140
: : : : : C : : : : : CAAUUU UGG A : : : : C +AAG: :
SteCoe_contig_1964 3748 GGUCGUUU- -ACCC- -AAUAGUCU- -CAAUUUCUGGCA- GGUCGAAAGACC 3705
77888877. 8888..77888877...9*****99. ***** PP
```

**D**

```

NC
CS
U5 1 auctccuagggccacuaucuaucGAACAagucUcUcGCCUUUUUACuAgAgacuuCCGugugagucccauuuaaggggguuuuaaauC AU 89
+ :C:C GG ::CA:U:CA+A+CGA +A:U:C:UCGCCUUUUACUA:AGA:U:CCGUG:A:UG::C ++G:G:UU AA C A U
SteCoe_contig_1638 8263 UAUCGACGGGUUCAGUUCAAAAACGAUUUAUUCUUCGCCUUUUUACUAAAGAUUACGUGAGCUGGGUCUUAUGUUGCAUUGAAC-C AUU 8350
*****99.*** PP

v NC
-----<-<<<<<____>>>>> CS
U5 90 UUUUGgaagccuuuuuuuuuaagggcu 116
UUUUG ::CUU AAG::
SteCoe_contig_1638 8351 UUUUGCAGCGGCUUA----AAGCGC 8372
*****8888885.....8888888 PP
```

[illegible]

Figure S4

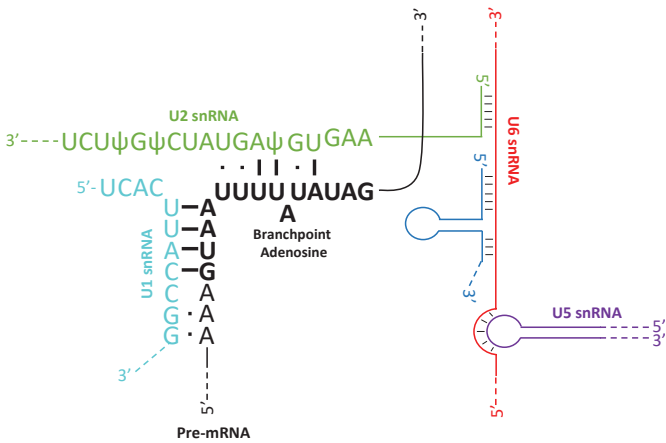
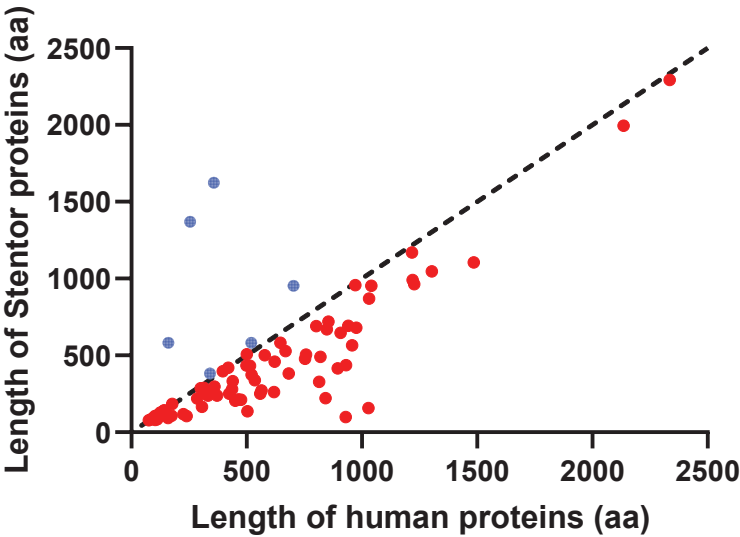


Figure S5





### Figure S6

[illegible][illegible][illegible]

<i>Stentor_ceruleus</i>	-----	
<i>Arabidopsis_thaliana</i>	GSDEDED---	331
<i>Mus_musculus</i>	RDESDSDN--	313
<i>Homo_sapiens</i>	LDSDSDSGSN	323
<i>Danio_rerio</i>	SDSDSDSSD-	345
<i>Caenorhabditis_elegans</i>	GSGSDDSD--	323
<i>Cryptococcus_neoformans</i>	KKAKA----	321
<i>Schizosaccharomyces_pombe</i>	-----	
<i>Saccharomyces_cerevisiae</i>	-----	

**B**

Species	Sequence	Position
<i>Cryptococcus neoformans</i>	VGSDGNDLKKSWHFLITLIDQENVWKAECANPEKQMLACVLRGEETEEQLLEHHR	111
<i>Schizosaccharomyces pombe</i>	VGSDGNDLKKSWHFLITLIDQENVWKAECANPEKQMLACVLRGEETEEQLLEHHR	108
<i>Homo sapiens</i>	VGSDGNDLKKSWHFLITLIDQENVWKAECANPEKQMLACVLRGEETEEQLLEHHR	108
<i>Mus musculus</i>	VGSDGNDLKKSWHFLITLIDQENVWKAECANPEKQMLACVLRGEETEEQLLEHHR	102
<i>Danio rerio</i>	VGSDGNDLKKSWHFLITLIDQENVWKAECANPEKQMLACVLRGEETEEQLLEHHR	102
<i>Stentor coeruleus</i>	VGSDGNDLKKSWHFLITLIDQENVWKAECANPEKQMLACVLRGEETEEQLLEHHR	102
<i>Saccharomyces cerevisiae</i>	VGSDGNDLKKSWHFLITLIDQENVWKAECANPEKQMLACVLRGEETEEQLLEHHR	105

```

Cryptococcus neoformans      KNIG-----AKREFIALQNTATRTAATREDEFLATKQKQAAALAAQVQDRIKQIPAAKQKATVGTNGQGETKKEERKARQKATQERKRRKRRKRRSSPPSSQDYDDYDRRRRRD 129
Schizosaccharomyces pombe  QNWS-----LAKTEFIALQNSLQTAQMRULDELAATQKQVQLQGTMEK-----RKYSLDSDFQERRRHRDRHNRNRRSRRESRSDND 225
Homo sapiens                KEAGCSSETGLLPGSIFAPSGNSLLMASTREDEFLPIKQKQEEPKRVLNLRVQMKKQKELLQMSLEK-----KEKKKQKQKQKQKQKQKRRSSSSRRSSRDEHSQRS- QK 213
Mus musculus               REAGCSSETGLLPGSIFAPSGNSLLMASTREDEFLPIKQKQEEPKRVLNLRVQMKKQKELLQMSLEK-----KEKKKQKQKQKQKQKQKRRSSSSRRSSRDEHSQRS- QK 213
Danio rerio                QESGSAETGLLPGSIFAPSGNSLLMASTREDEFLPIKQKQEEPKRVLNLRVQMKQKQMLQMLEK-----KQKKKQKQKQKQKQKQKRRSSSSRRSSRDEHSQRS- QK 215
Stentor coeruleus          -----EMRPTLCQKQANEIFVHDEDTFAAQKQKQRRKEVTSN-----QKRVNKKRETEKQKQK-----KQKKQKQKQKQKQKQKQKRRSSRDEHSQRS- QK 215
Saccharomyces cerevisiae   QETTPVRAATTSISAGSAAATSSISQKQKSLKQD-----MSFSKVTAKKQKQKQKQKQKQKRRSSRDEHSQRS- QK 175

```

<i>Cryptococcus neoformans</i>	SYNSDRNDNRRTYDNRSDRSRSTRSESPKRGKGEKDYSSK-----DLDYRRDDTTGSGFSELEKDEGNRKRSDHRDRFRERQDDYDIPPPYFSHNYSYTVVPPFSPPPPFAAALPVT	339
<i>Schizosaccharomyces pombe</i>	HSKDKHRSRYSYRNDNRNDNRKERTHNR-----YRRDKDYGSGYKGYDDDMRFQGHQFD-----ERDLKKYVRTSRQYSRSPSDPRTNHQHPHSRSDRSPQT	288
<i>Homo sapiens</i>	MANSSFVLSKVPFGYGLQVRNSDRNGLQLGPITAEQGRHGMSQKHSRSRSSSHSPRRHASKSKTKEAGSDRRRSRLGRKSRSPRPSKLHNSKNRRRTGQTRSPSPKKEVYQRRHAGVT	338
<i>Mus musculus</i>	MANSSFFVLSKVPFGYGLQVRNSDRNRLQSGSLGEQRAIKN-----NSRSRSSSPRRHASKSKTKEAGSDRRRSRSPSRSSRSPRPSKPHTSKNVNRK---DSPSPKKEAYQRRHAGVT	324
<i>Danio rerio</i>	SDVTTNNGHRSYGLQLPAHRSTHSSASNHTDHR-----SDRSRSRSLSHKADRNHSSREREPKAKPKAP-----SPPRHRDRYRQRPQTNTS	300
<i>Stentor coeruleus</i>	KNRYQKEASCDSQLYNLNDNKRSL-----	224
<i>Saccharomyces cerevisiae</i>	DLDY-----	179

```

Cryptococcus_neoformans NTSTLEDQRAAKLAAMSADAILYSSRSKSLAARAEERERQDEQKMKQYKQKEQASANFFSQ----QSQGLGLGALQKGGKLLKDI--- 425
Schizosaccharomyces_pombe RHTDYS----RLQKMDNAKLEDESRKKLTLELKKERDEEQFLEKERRDTARKWGQDFTIRMKRYSYSGDSVLADRVNSRHNMLRP- 376
Homo_sapiens RKLAAEELERKQEMENAKWKEERLINLTKRHADERE-QRLKLELDGRDGKFLHRMKLESASTSSLEDVRKNIHSLQTSVALEKNFMKR 425
Mus_musculus RKLAAEELERKQEMENAKWKEERLINLTKRHADERE-HRLERLDGRSGKFLHRMKLESASTSSLEDVRKNIHSLQTSVALEKNFMKR 416
Danio_rerio RKLAAEELKKGREMDFAREREVERENVQYRKQKEQEKAREDAKQDRHAGFITHMKLESASTSSLEDVRKNIHSLQTPAALKENFMKR 393
Stentor_coeruleus -----

```

[illegible]

<i>Cryptococcus neoformans</i>	-----RQVFGQKTPGKALPGVRLLESTPGQATESASFSFQTHQPGVQGE	ELKGLDSDVQAREGVEDARKSASALGISDITLRPPVYSISKLKPD	224
<i>Mus musculus</i>	-----KQVGNCRTPGKALPGVRLLESTPGQATESASFSFQTHQPGVQGE	ELKGLDSDVQAREGVEDARKSASALGISDITLRPPVYSISKLKPD	224
<i>Homo Sapiens</i>	-----KQVGNCRTPGKALPGVRLLESTPGQATESASFSFQTHQPGVQGE	ELKGLDSDVQAREGVEDARKSASALGISDITLRPPVYSISKLKPD	224
<i>Danio rerio</i>	-----KQVGNCRTPGKALPGVRLLESTPGQATESASFSFQTHQPGVQGE	ELKGLDSDVQAREGVEDARKSASALGISDITLRPPVYSISKLKPD	224
<i>Arabidopsis thaliana</i>	NIIDVFNPSGRGRGKTPGKALPGVRLLESTPGQATESASFSFQTHQPGVQGE	ELKGLDSDVQAREGVEDARKSASALGISDITLRPPVYSISKLKPD	224
<i>Caenorhabditis elegans</i>	-----RQVGNCRTPGKALPGVRLLESTPGQATESASFSFQTHQPGVQGE	ELKGLDSDVQAREGVEDARKSASALGISDITLRPPVYSISKLKPD	224
<i>Stentor coeruleus</i>	-----KQVGNCRTPGKALPGVRLLESTPGQATESASFSFQTHQPGVQGE	ELKGLDSDVQAREGVEDARKSASALGISDITLRPPVYSISKLKPD	224
<i>Saccharomyces cerevisiae</i>	MN-----SQKGLNCRTPGKALPGVRLLESTPGQATESASFSFQTHQPGVQGE	ELKGLDSDVQAREGVEDARKSASALGISDITLRPPVYSISKLKPD	224

<i>Cryptococcus neoformans</i>	SVGLGTVQVNE	344
<i>Mus musculus</i>	SGKLGSGEALKTGNSKKRKSGLNDLGGGQKEQEDANSTKDTVSAINNVVGGVNSAAVAATQAQAAFLVDLSLGIPTMPSIDGKAAVLVLDVVRQAALKE	345
<i>Homo sapiens</i>	EENIYNAVYTESDEEGNQEKAGGGGQKFTANPVSQGLDEAFVRRRSGEQVQKQYASSTQQAQSEEARRLLYG	289
<i>Danio rerio</i>	EENIYNAVYTESDEEGNQEKGGDSQKFTANPVSQGLDEAFVRRRSGEQVQKQYASSTQQAQSEEARRLLYG	289
<i>Arabidopsis thaliana</i>	EETIYAVRDSRSDDETQKGGEGDGAQSTFANPVSQGLDEAFVRRRSGEQVQKQYASSTQQAQSEEARLLGL	289
<i>Caenorhabditis elegans</i>	AREVLFEEDVVEEERMERKEEKEEREFVVRPLDQKQVLEKLNKMDLSKYSALDELVGQTEAKSMLNIHR	307
<i>Stentor coeruleus</i>	ENIYKVEEDDDDLDETSTVIGEDGRMPTFRLLPLDQKQVLEKLNKMDLSKYSALDELVGQTEAKSMLNIHR	266
<i>Saccharomyces cerevisiae</i>	---KVRVLVEDDTLRFTEKANDKYLKLGEG	235
	LTEFENAWTSILKARYNVFVNDDEEMSRQTEIHVGLADNCHVVRKMKMDLNL	235

Cryptococcus_neoformans	YGVY	348
Mus_musculus	----	
Homo_Sapiens	----	
Danio_rexio	----	
Arabidopsis_thaliana	----	
Caenorhabditis_elegans	----	
Stentor_coeruleus	----	
Saccharomyces_cerevisiae	----	