SUPPLEMENTARY FIGURE LEGENDS

Figure S1. Intron prediction pipeline.

Figure S2. Visualization of RNA-seq read coverage and splice site annotations using IGV: (A) *A. motanka*: visualization of intron65437 (*contig08337:1450-1484*), intron65439 (*contig08337:1703-1748*), intron79365 (*contig048:5204-5239*), and intron86326 (*contig08012:2472-2507*). (B) *N. karyoxenos*: visualization of intron12964 (*contig050515:701-763*), intron15375 (*contig020064:1119-1199*), intron04754 (*contig07549:464-520*), and intron20183 (*contig015374:1065-1120*). The Integrated Genome Viewer (IGV) displays RNA-seq read coverage (top track) mapped to the genome alongside corresponding splice site annotations (bottom track). Exons are represented as thick blocks, while introns are depicted as thin connecting lines or gaps between the blocks in the annotation track. The coverage track highlights the read depth across exons, while intronic regions exhibit low or absent read coverage, consistent with splicing events.

Figure S3. (A) Length distribution of introns in six diplonemid species: *A. motanka*, *N. karyoxenos*, *S. specki*, *L. lanifica*, *D. japonicum*, *and R. humris*. Panels are faceted by species, with the x-axis representing intron length (bp) and the y-axis representing the intron count. (B) *P. papillatum* splice site sequence logos from recently sequenced genome (Valach, M et al. 2023) and annotated introns using our pipeline.

Figure S4. (A) Logos of a branch-site in canonical introns. (B) Identified sequence motifs in diplonemid introns.

Figure S5. Splicing of selected *N. karyoxenos* genes in *T. brucei* and *P. papillatum*. Three genes from *N. karyoxenos* (*MRT4, POC1*, and *GK*) containing canonical and non-canonical introns and one gene from *P. papillatum* (*MRT4)* containing a canonical intron were inserted into *T. brucei* genome and their expression induced by 24 hour treatment with doxycycline, followed by extraction of total RNA and RT-PCR. (A) Representative gels shown for untreated (0 h) and treated (24 h) cells. Description of the genes is the same as in Figure 2. (B) Quantification of spliced/unspliced products in three independent experiments. Ratios were calculated using band intensities corresponding to spliced and unspliced products adjusted for background. Mean and standard deviation are shown. (C) Sequence of the PCR products corresponding to the spliced RNA of *GK* gene, intron 1 (*N. karyoxenos*) isolated

from the gel. (D) Results of gene expression in *P. papillatum* showing successful expression of the *POC1* gene. Representative gel images are shown on the left with corresponding exon (e) - intron (i) structures indicating splice site boundaries. The expected spliced/unspliced RT-PCR products are indicated on the right. Exons are depicted as boxes and introns as lines.

Figure S6. Secondary structure prediction for U1, U2, U4, U5, and U6 snRNAs from *A. motanka.* 5'ss and branch point (BP) binding regions and the conserved U6 AGAGA box are underlined, Sm-binding sites are indicated by black rectangles.

Figure S7. Secondary structure prediction for U1, U2, U4, U5, and U6 snRNAs from *N. karyoxeno*s. 5'ss and branch-point binding regions and the conserved U6 AGAGA box are underlined, Sm-binding sites are indicated by black rectangles.

Figure S8. Secondary structure prediction for canonical U1, non-canonical U1, U2, U4, U5, and U6 from *S. specki*. 5'ss and BP binding regions and the conserved U6 AGAGA box are underlined, Sm-binding sites are indicated by black rectangles.

Figure S9. Secondary structure prediction for U1, U2, U4, U5, and U6 snRNAs from *L. lanifica*. 5'ss and BP binding regions and the conserved U6 AGAGA box are underlined, Sm-binding sites are indicated by black rectangles.

Figure S10. Secondary structure prediction for U1, U2, U4, U5, and U6 snRNAs from *D. japonicum*. 5'ss and BP binding regions and the conserved U6 AGAGA box are underlined, Sm-binding sites are indicated by black rectangles.

Figure S11. Secondary structure prediction for U2, U4, U5, and U6 for *R. humris*. The 5'ss and BP binding regions and the conserved U6 AGAGA box are underlined, Sm-binding sites are indicated by black rectangles.

Figure S12. Secondary structure of selected introns from *A. motanka*.

Figure S13. Secondary structure of selected introns from *N. karyoxenos*.

Figure S1

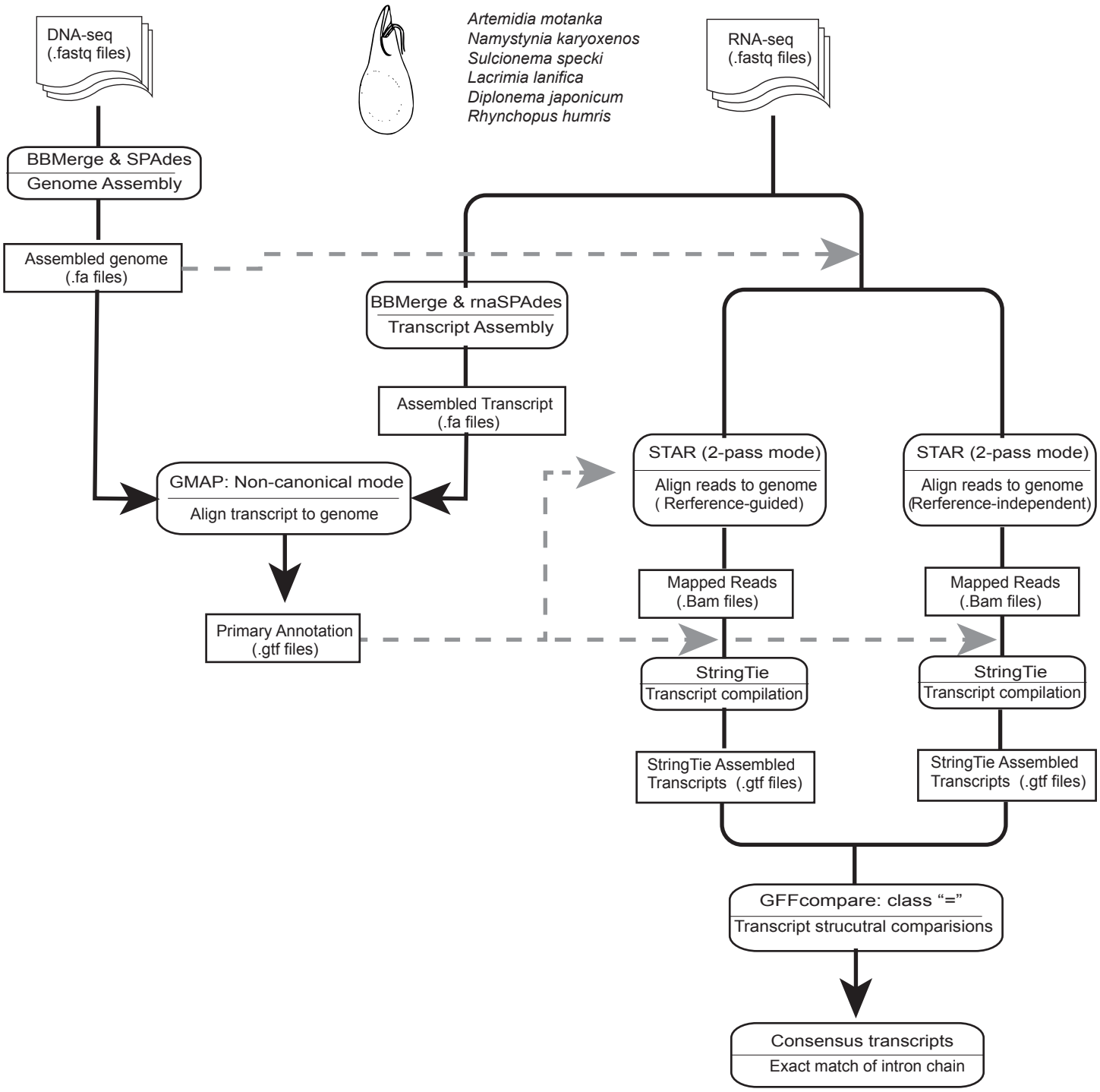

DNA-seq
(.fastq files)
BBMerge & SPAdes
Genome Assembly
Assembled genome
(.fa files)
Artemidia motanka
Namystynia karyoxenos
Sulcionema specki
Lacrimia lanifica
Diplonema japonicum
Rhynchopus humris
RNA-seq
(.fastq files)
BBMerge & rnaSPAdes
Transcript Assembly
Assembled Transcript
(.fa files)
GMAP: Non-canonical mode
Align transcript to genome
Primary Annotation
(.gtf files)
STAR (2-pass mode)
Align reads to genome
(Rerference-guided)
STAR (2-pass mode)
Align reads to genome
(Rerference-independent)
Mapped Reads
(.Bam files)
Mapped Reads
(.Bam files)
StringTie
Transcript compilation
StringTie
Transcript compilation
StringTie Assembled
Transcripts (.gtf files)
StringTie Assembled
Transcripts (.gtf files)
GFFcompare: class "="
Transcript strucutral comparisions
Consensus transcripts
Exact match of intron chain

Figure S2

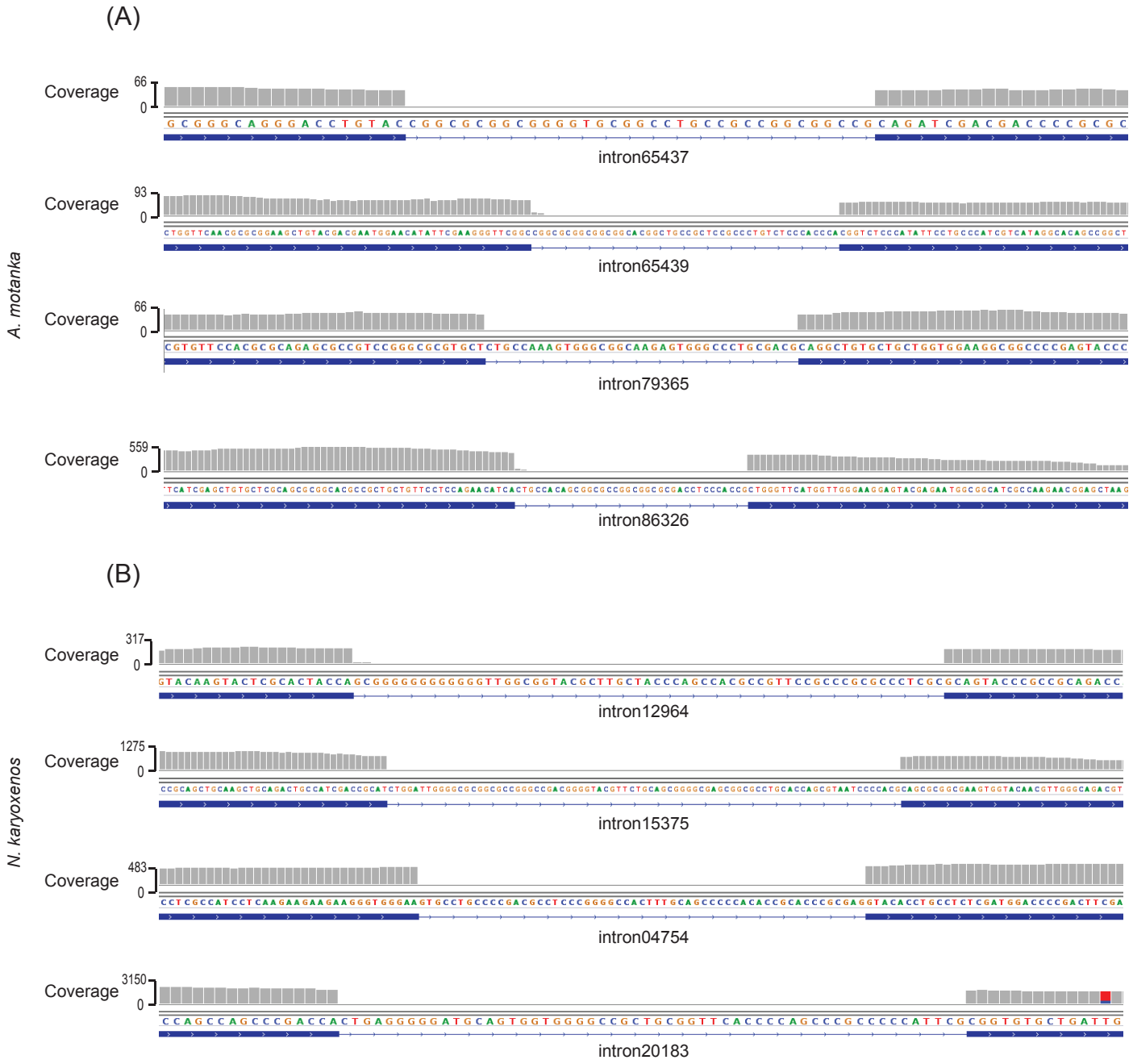
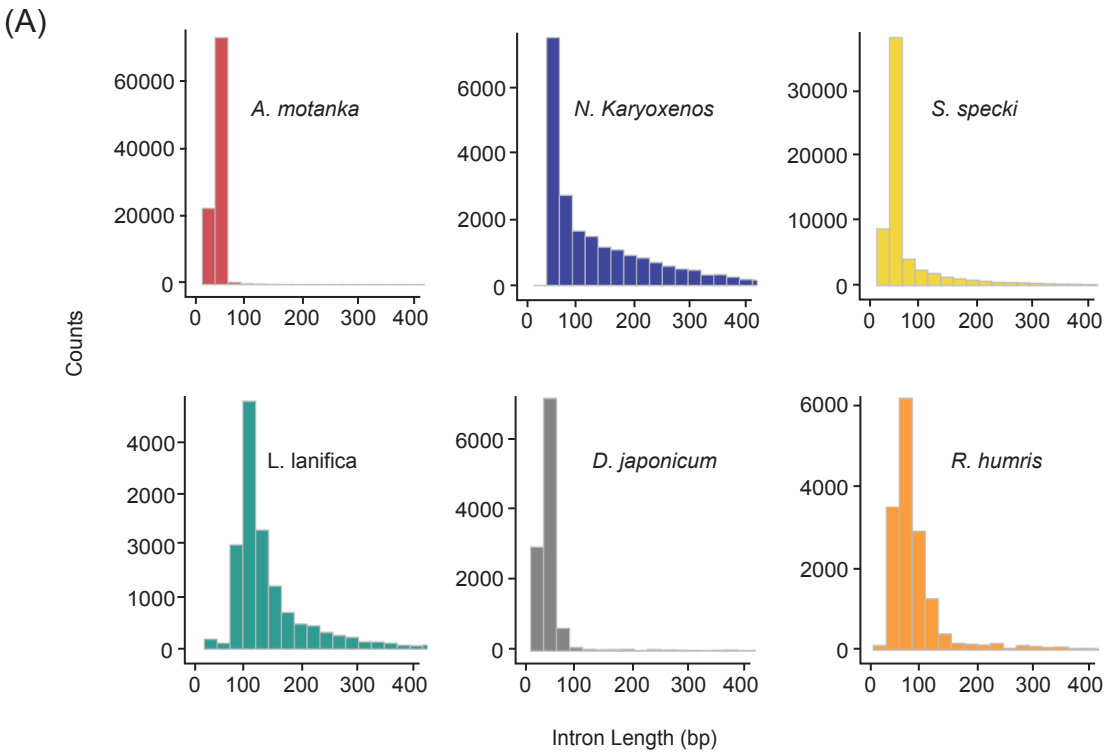

(A)
A. motanka
Coverage
66
0
intron65437
Coverage
93
0
intron65439
Coverage
66
0
intron79365
Coverage
559
0
intron86326
(B)
N. karyoxenos
Coverage
317
0
intron12964
Coverage
1275
0
intron15375
Coverage
483
0
intron04754
Coverage
3150
0
intron20183

Figure S3


(A)
A. motanka
N. Karyoxenos
S. specki
L. lanifica
D. japonicum
R. humris
Counts
Intron Length (bp)

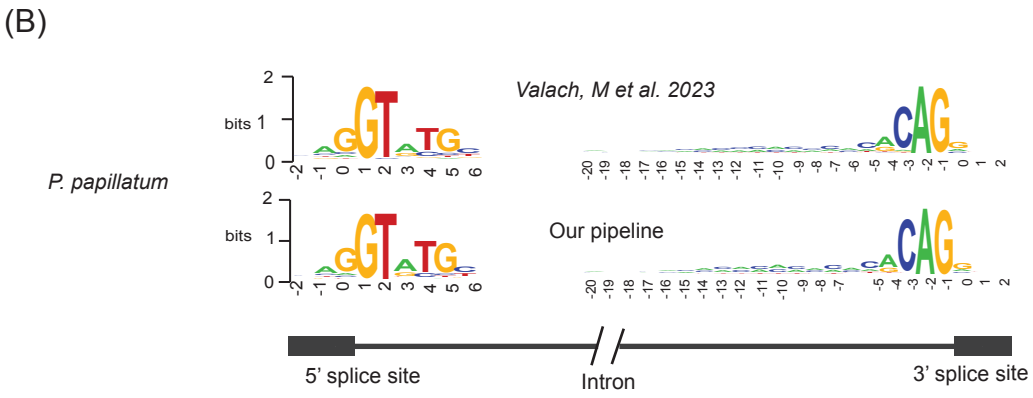
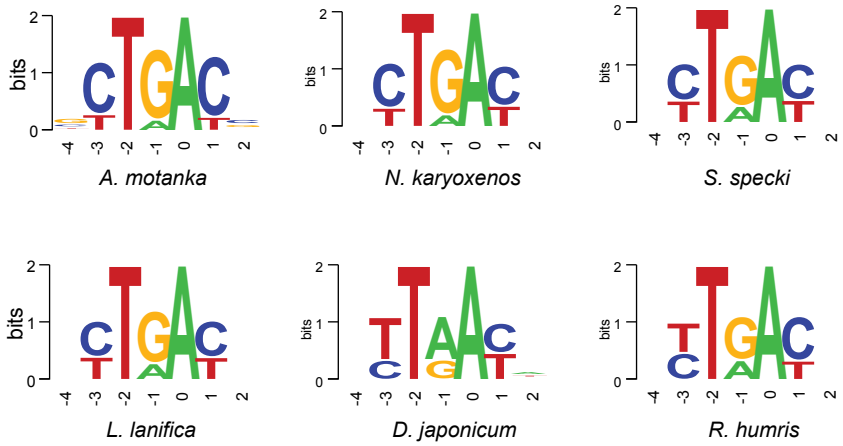

(B)
P. papillatum
Valach, M et al. 2023
Our pipeline
bits
5' splice site
Intron
3' splice site

Figure S4

(A)


bits
A. motanka
N. karyoxenos
S. specki
L. lanifica
D. japonicum
R. humris

(B)

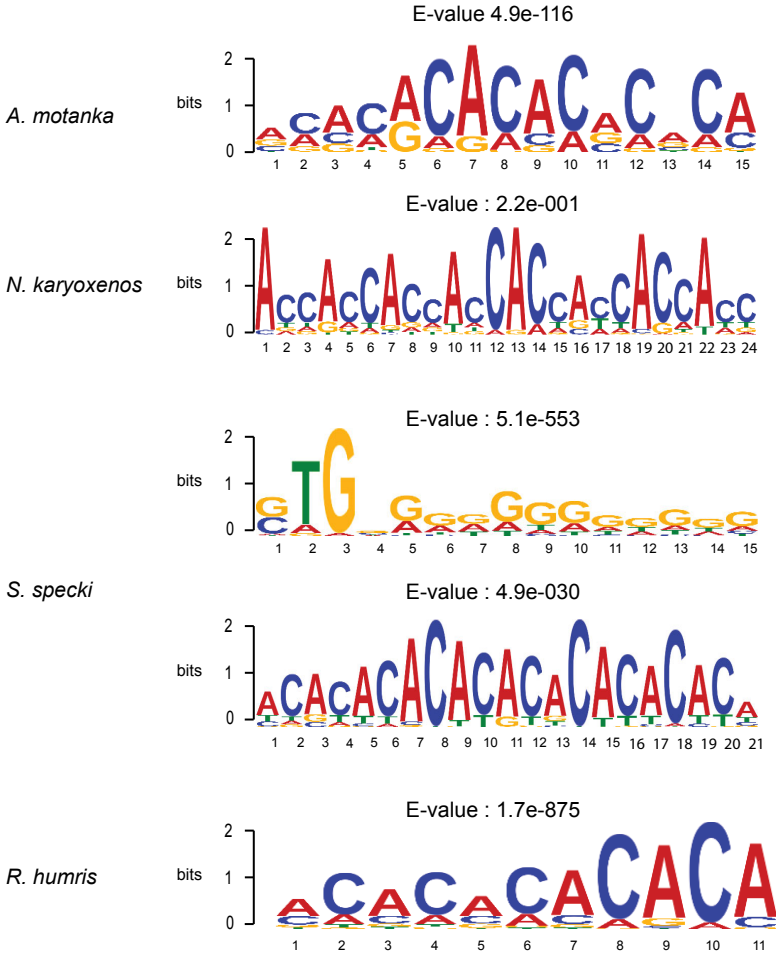
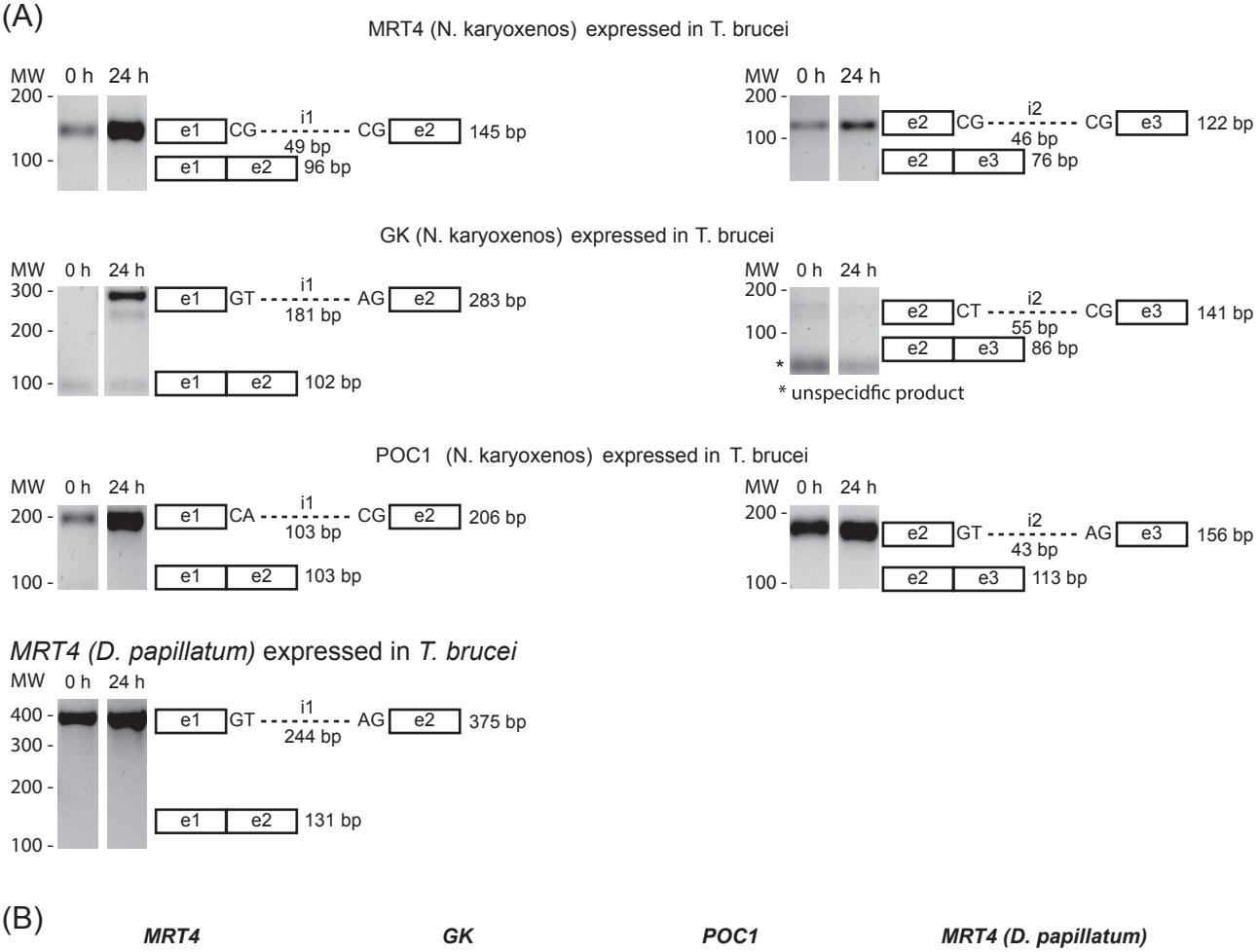
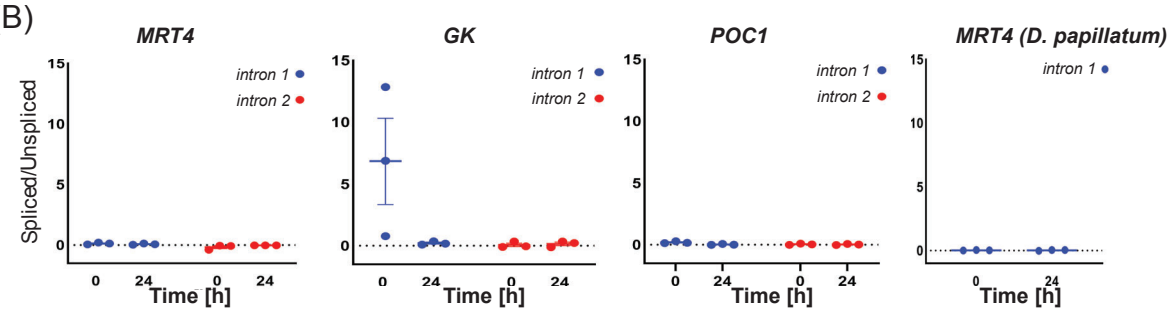

E-value 4.9e-116
A. motanka
E-value : 2.2e-001
N. karyoxenos
E-value : 5.1e-553
S. specki
E-value : 4.9e-030
E-value : 1.7e-875
R. humris
bits

Figure S5

(A)

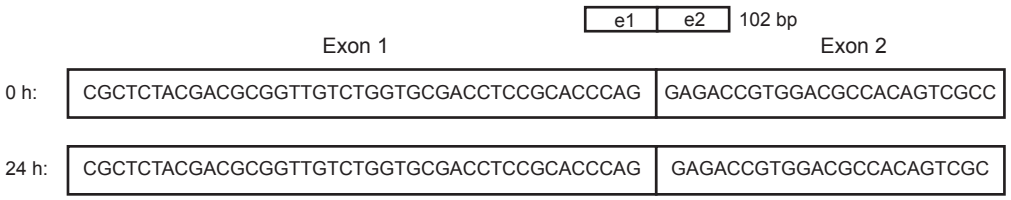


(B)



(C)

Sequence data for *GK (N. karyoxenos)* expressed in *T. brucei* intron 1:



(D)

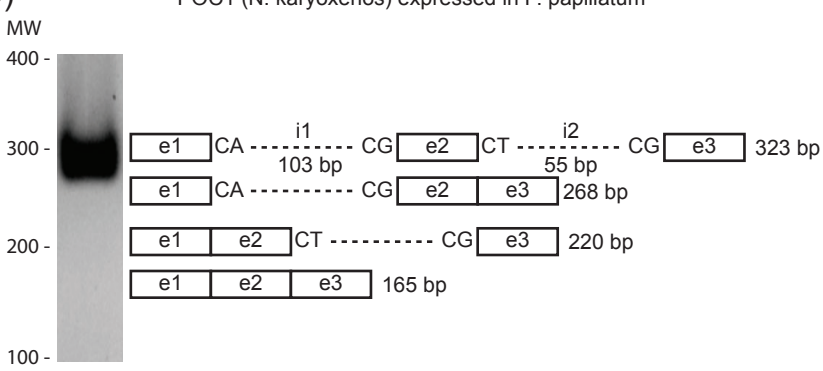


Figure S6

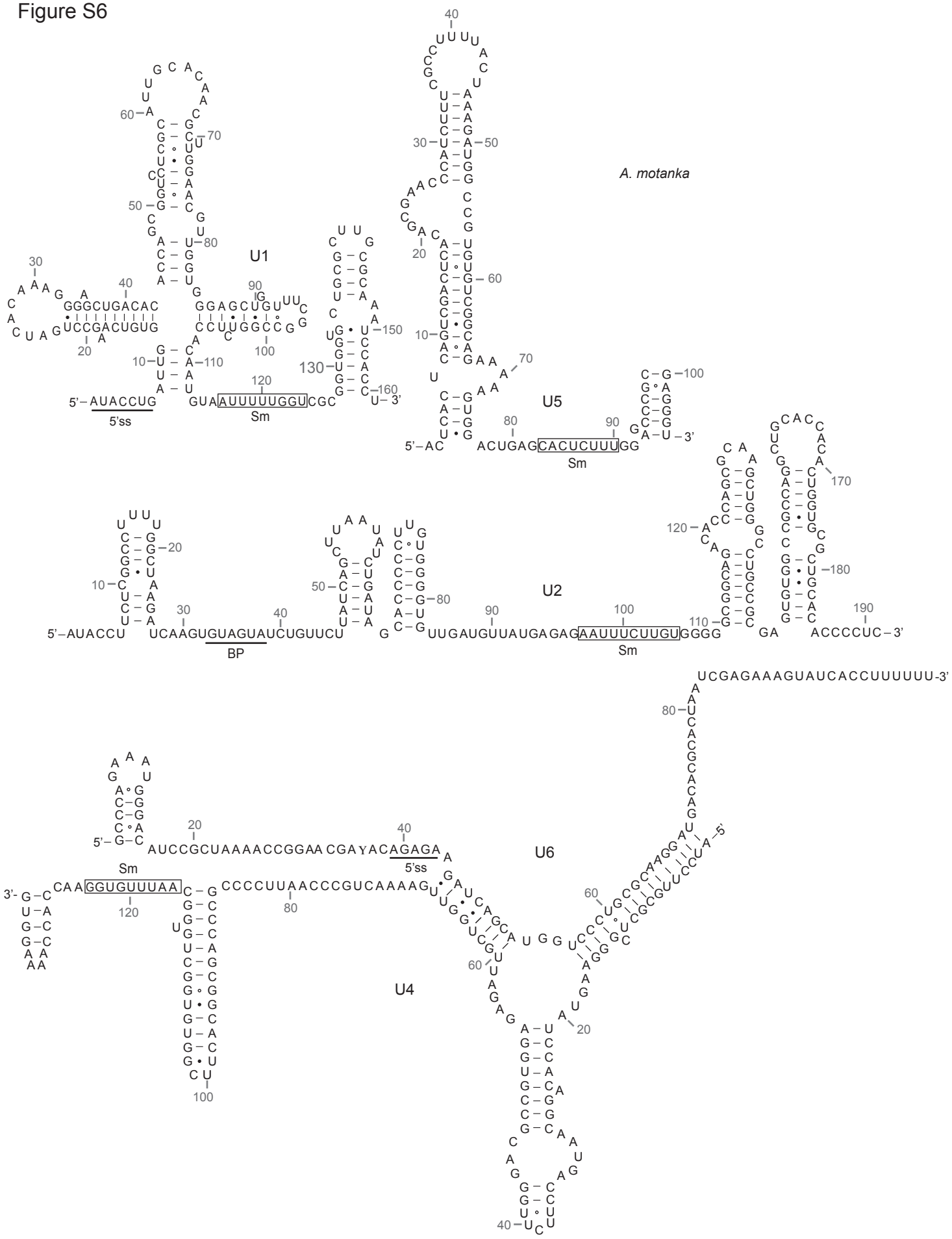

A. motanka
U1
U5
U2
U4
U6
5'ss
Sm
BP
5'-AUACCUG
GUAAUUUUUGGUCGC
5'-AUACCU
UCAAGUGUAGUAUCUGUUCU
UUGAUGUUAUGAGAG
AAUUUCUUGUGGGG
GA
ACCCCUC-3'
UCGAGAAAGUAUCACCUUUUUU-3'
AUCCGCUAAAACCGGAACGAYACAGAGAA
CAAGGUGUUUAA
CCCCUUAACCCGUCAAAAG

Figure S7

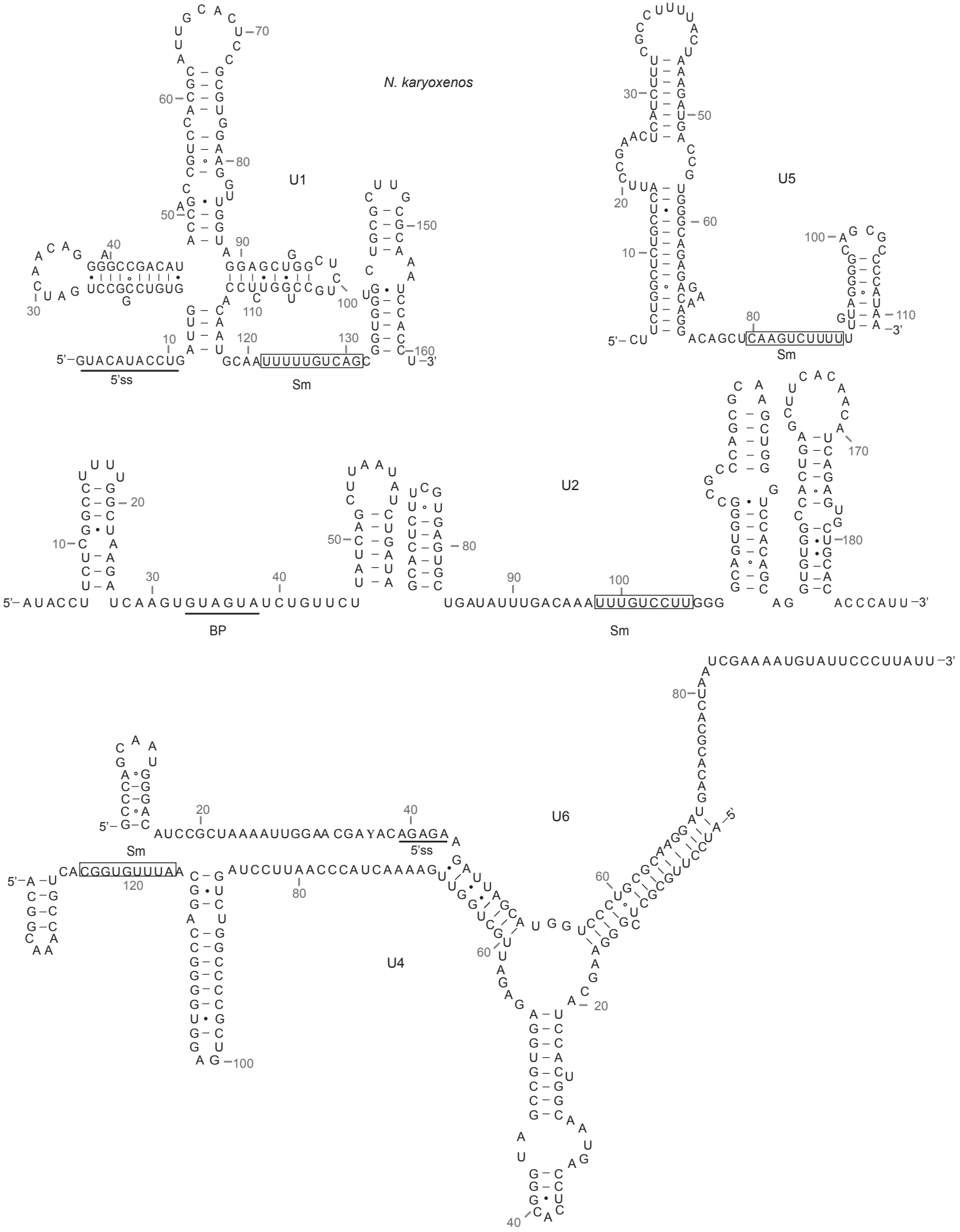

N. karyoxenos
U1
5'–GUACAUACCUG
5'ss
GCAAUUUUUGUCAGC
Sm
U5
5'– CU
ACAGCUCAAGUCUUUUU
Sm
U2
5'–AUACCU
UCAAGUGUAGUAUCUGUUCU
BP
UGAUAUUUGACAAAUUUGUCCUUGGG
Sm
U6
UCGAAAAUGUAUUCCCUUAUU–3'
5'–A–U
CACGGUGUUUAA
Sm
AUCCGCUAAAAUUGGAA CGAYAC AGAGA A
5'ss
AUCCUUAACCCAUCAAAAG
U4

Figure S8

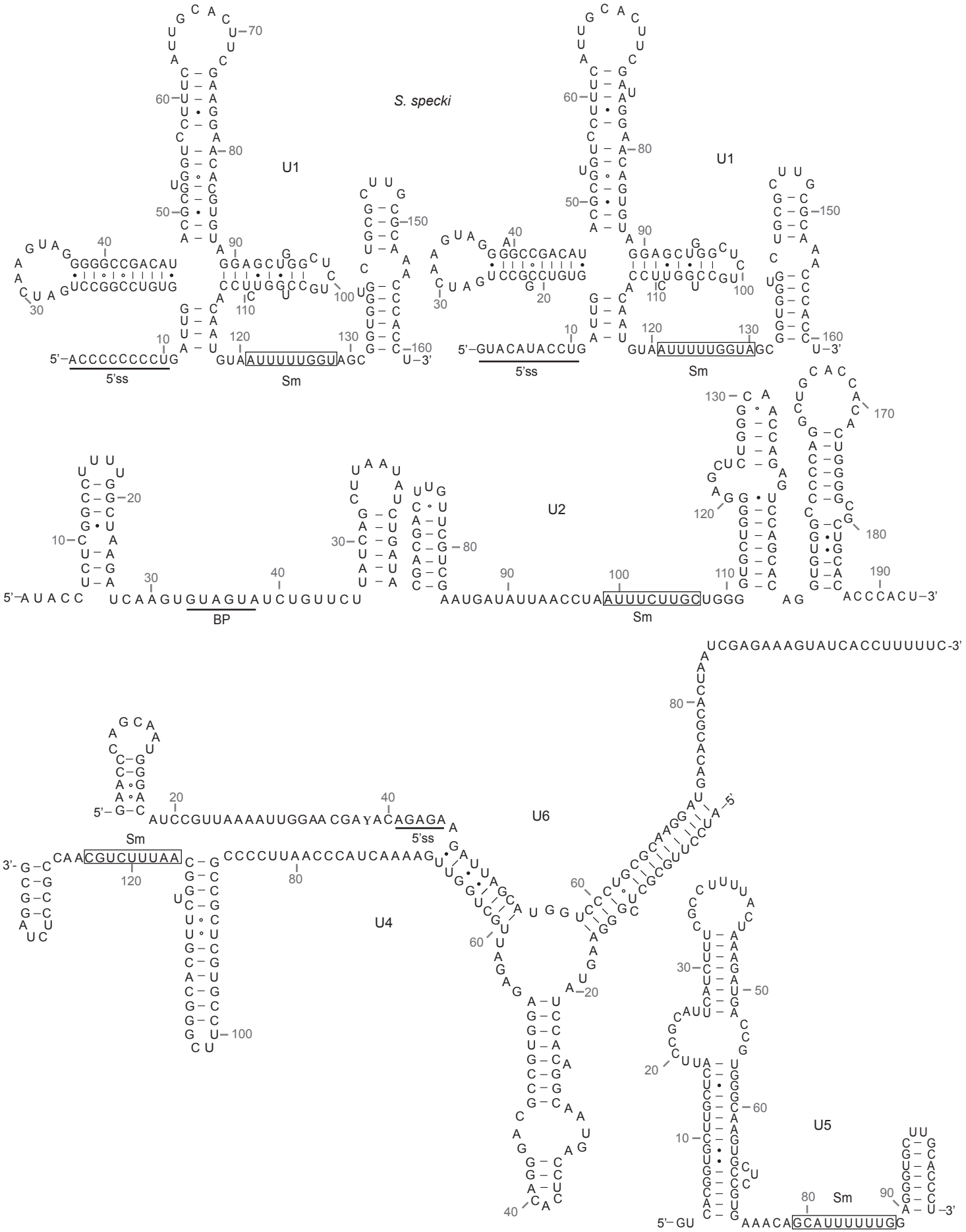

S. specki
U1
5'–ACCCCCCCCUG
5'ss
Sm
U1
5'–GUACAUACCUG
5'ss
Sm
U2
5'–AUACC UCAAGUGUAGUAUCUGUUCU
BP
AAUGAUAUUAACCUAAUUUCUUGCUGGG AG ACCCACU–3'
Sm
U6
AUCCGUUAAAAUUGGAACGAYACAGAGA A
5'ss
CAACGUCUUUAA
Sm
CCCCUUAACCCAUCAAAAG
U4
UCGAGAAAGUAUCACCUUUUUC-3'
U5
5'–GU AAACAGCAUUUUUUUG
Sm

Figure S9

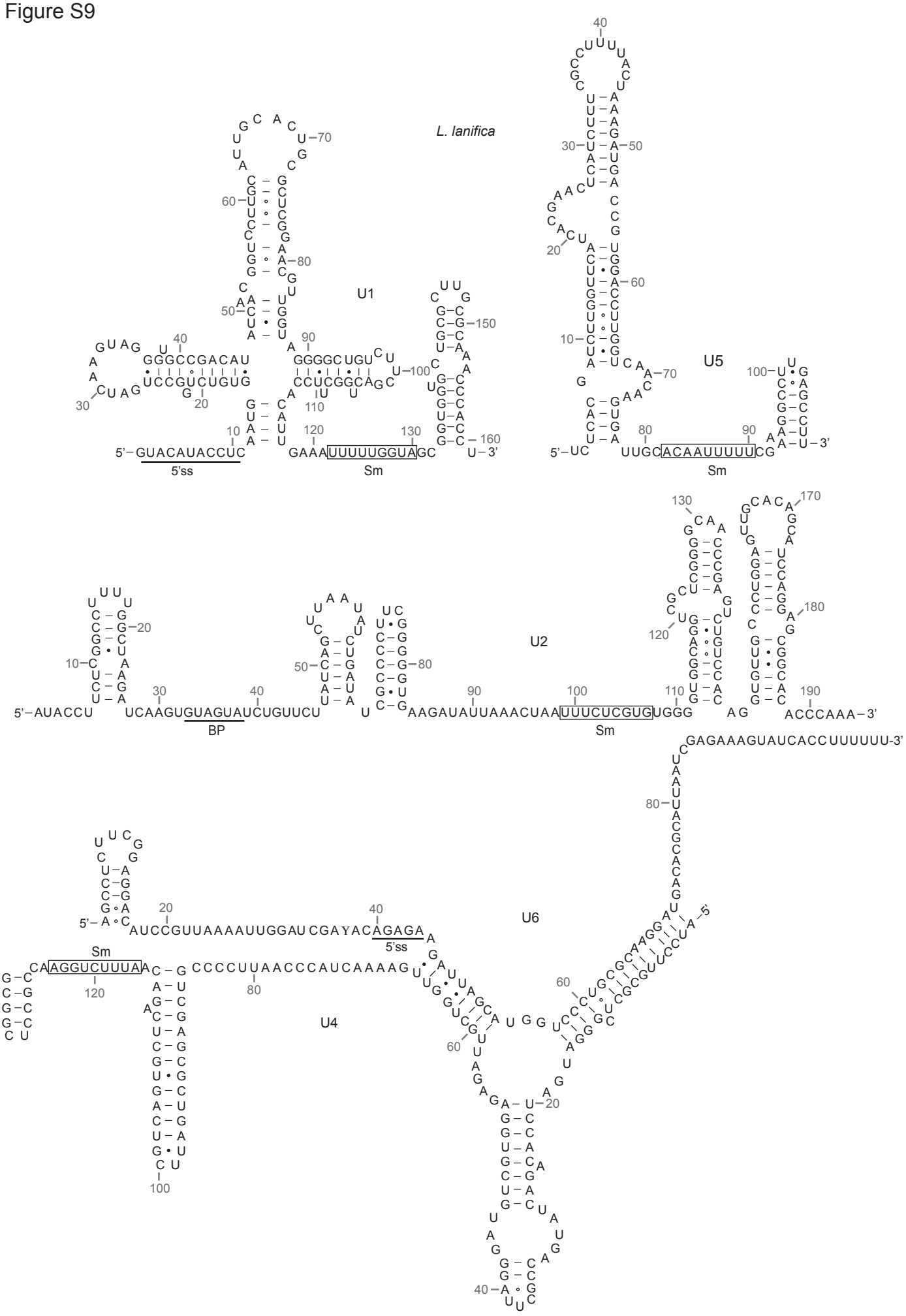


Figure S10

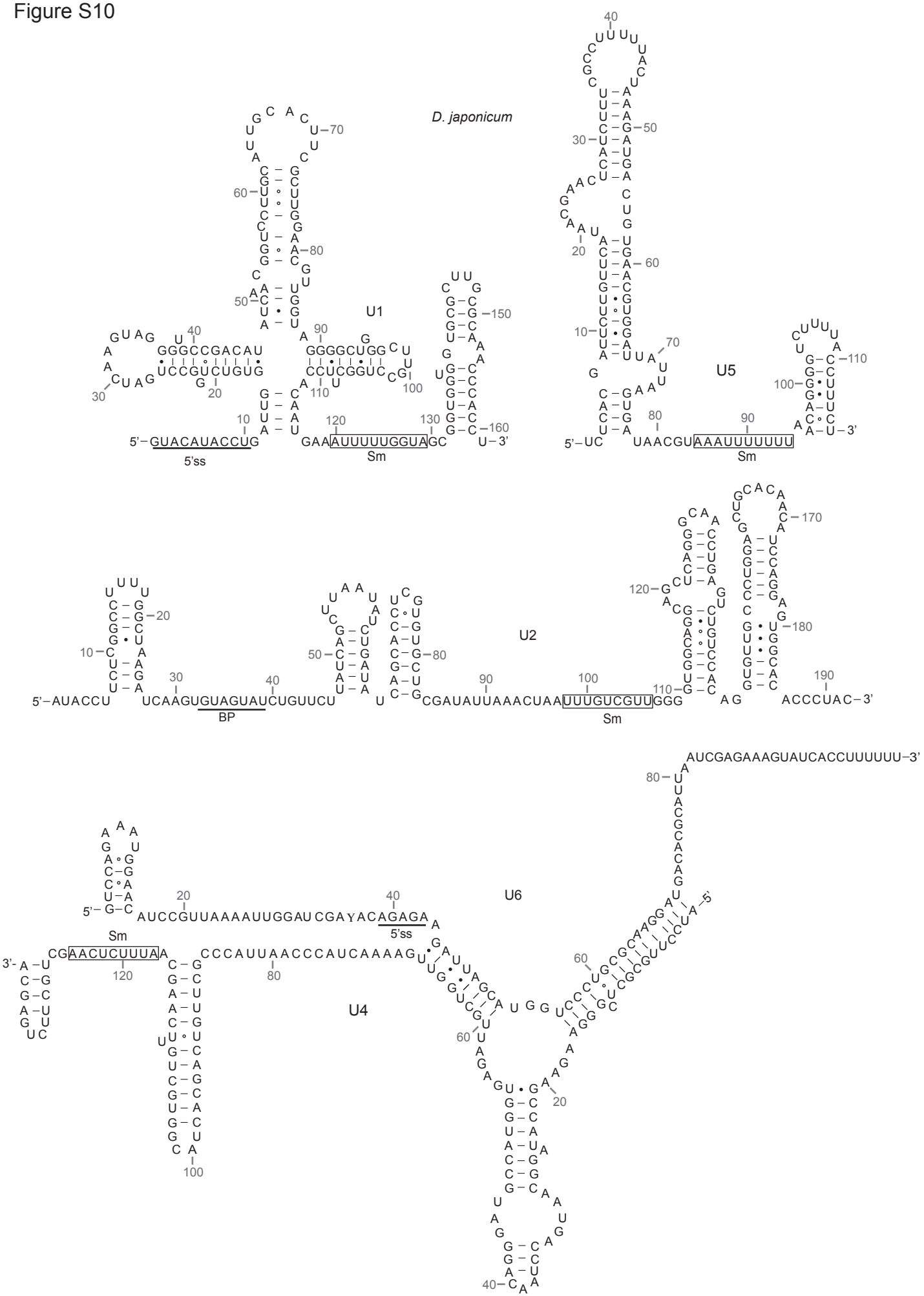

D. japonicum
U1
U5
U2
U4
U6
5'ss
Sm
BP
5'-GUACAUACCUG
GAAAUUUUUGGUAGC
5'-UC
UAACGUAAAUUUUUUU
5'-AUACCU
UCAAGUGUAGUAUCUGUUCU
CGAUAUUAAACUAAUUUGUCGUUGGG
ACCCUAC-3'
AUCGAGAAAGUAUCACCUUUUUU-3'
5'-G-C
AUCCGUUAAAAUUGGAUCGAYACAGAGAA
CGAACUCUUUAA
CCCAUUAACCCAUCAAAAG

Figure S11

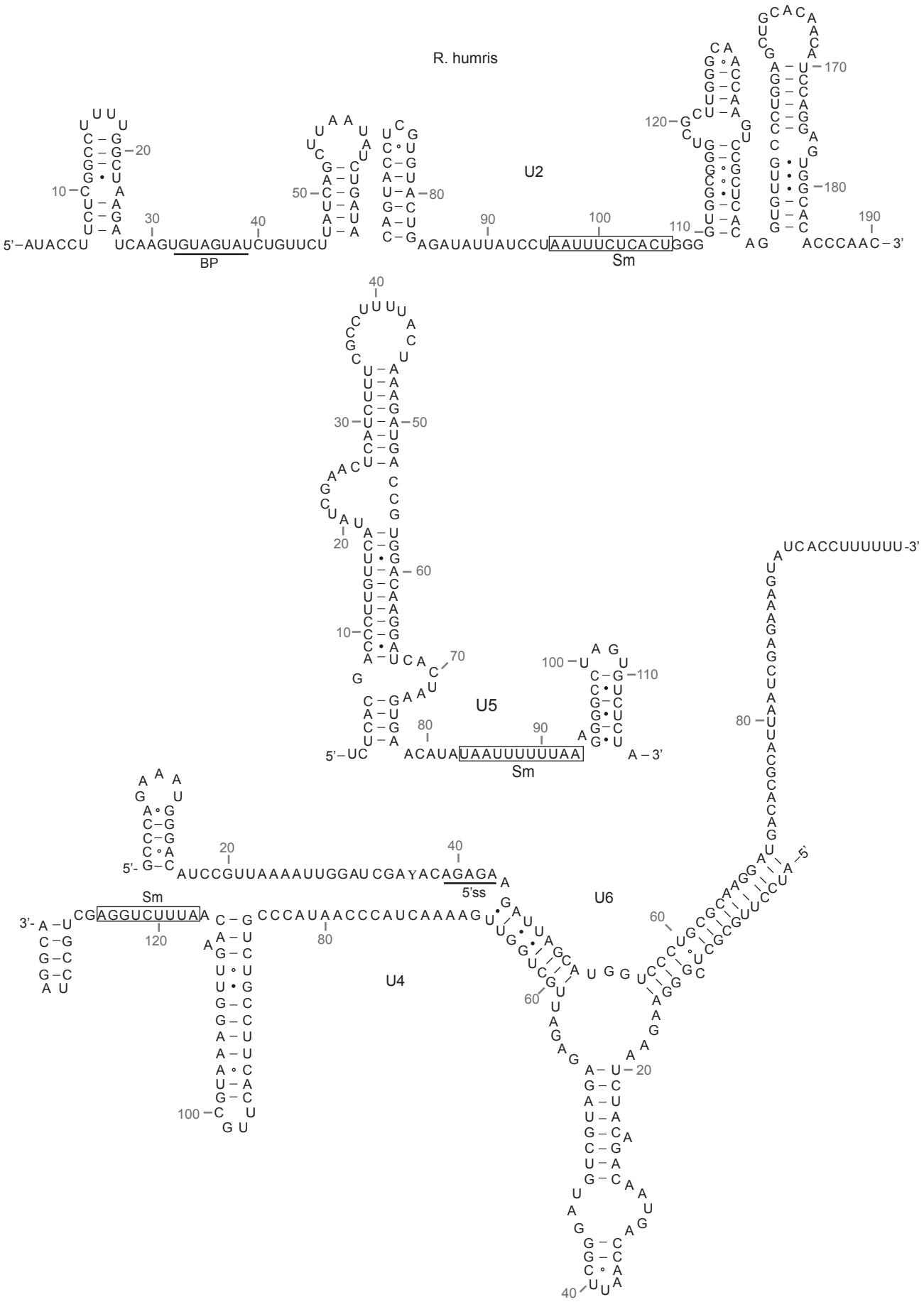
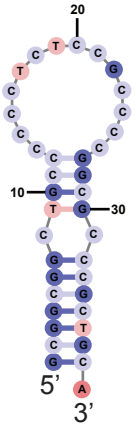
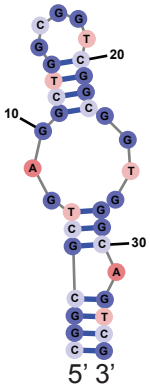


Figure S12

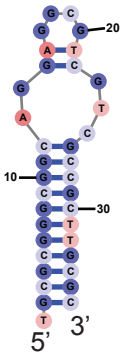
A. motanka



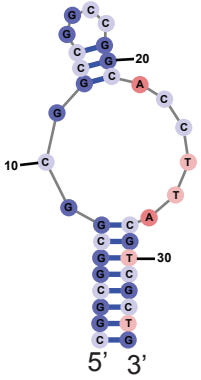
intron00004



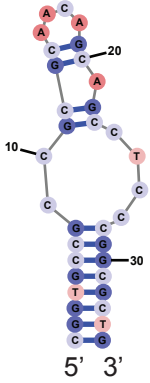
intron00048



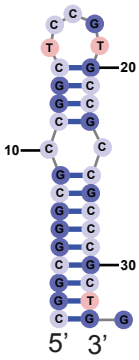
intron00109



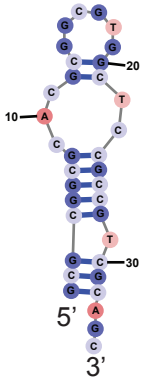
intron00259



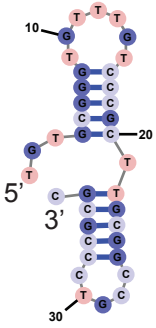
intron00302



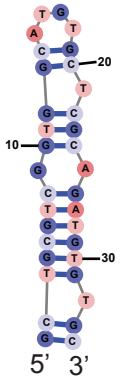
intron00303



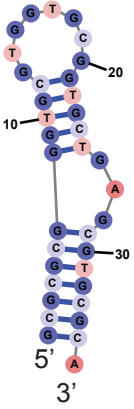
intron00339



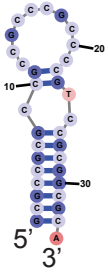
intron00483



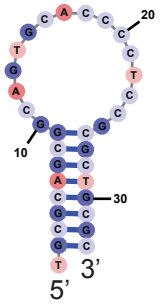
intron00561



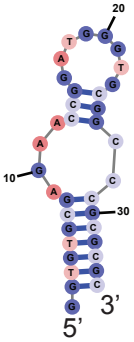
intron00616



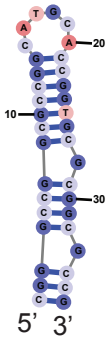
intron00819



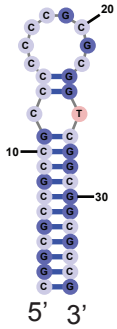
intron01021



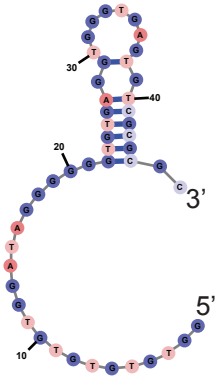
intron01046



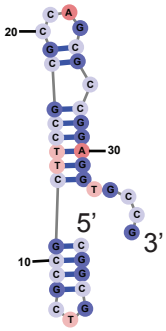
intron01107



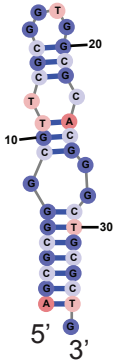
intron01129



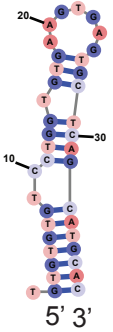
intron01200



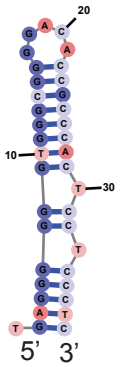
intron01210



intron01327



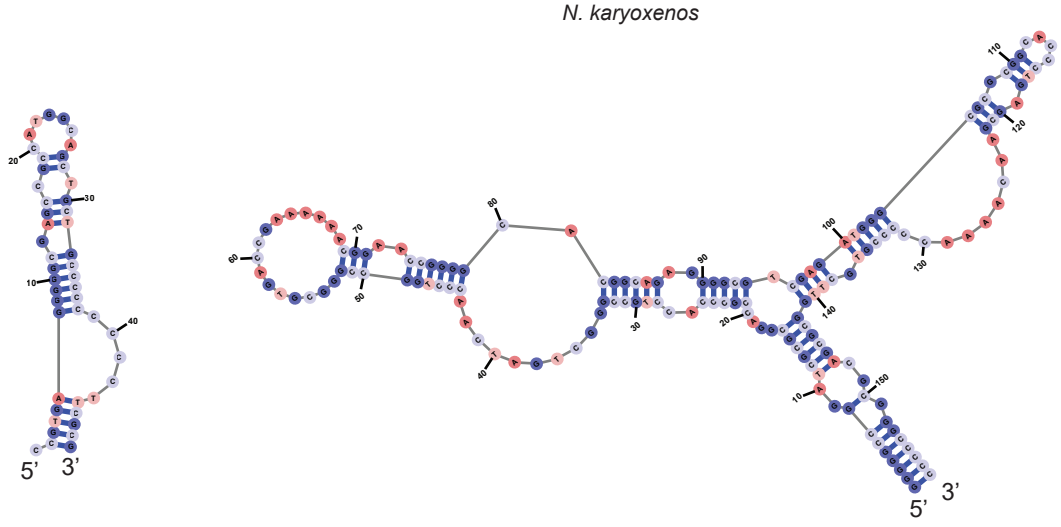
intron01624



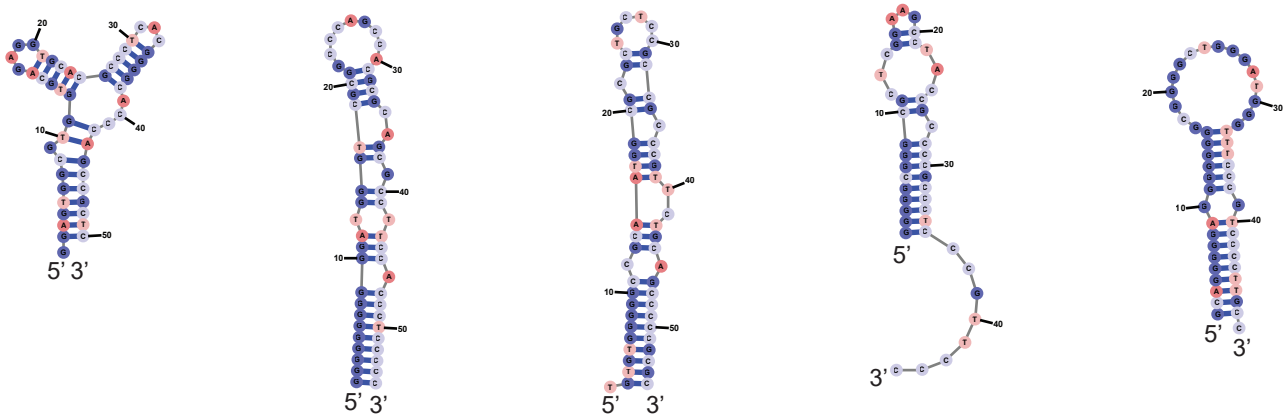
intron01637

Figure S13

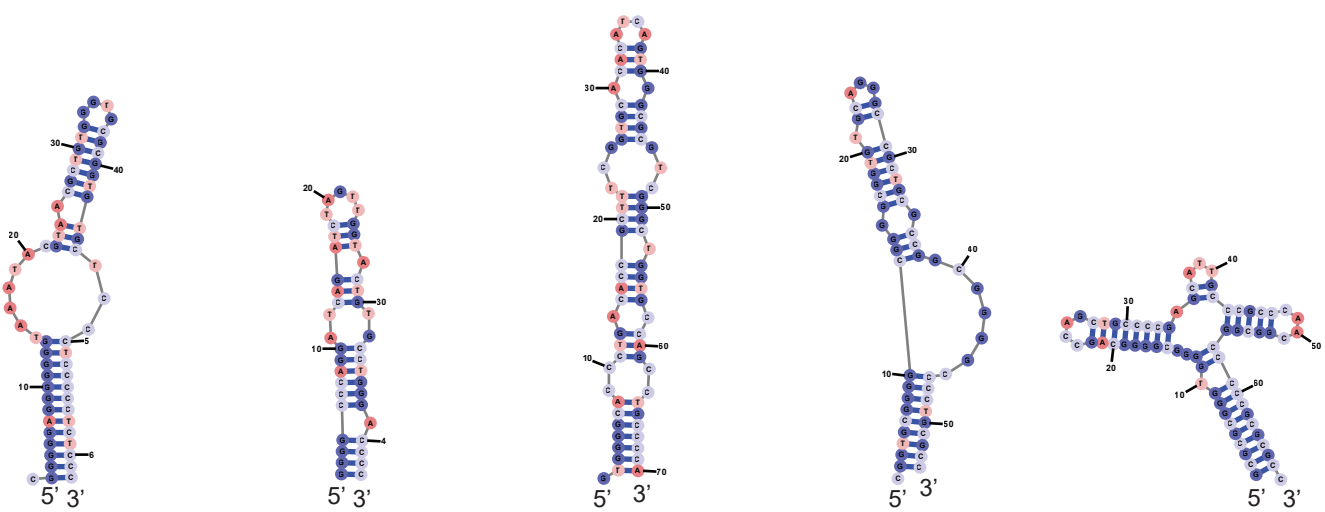
N. karyoxenos



intron00171 intron00242 intron00251



intron00258 intron00752 intron00757 intron00759 intron00912



intron00412 intron00531 intron00617 intron00647 intron00741

SUPPLEMENTARY TABLES

Table S1. Genome assembly statistics.

Table S2. Assessment of genome assembly

Table S3. List of used primers.

Table S1. Statistics of the diplonemids genome assembly

	No. of contigs	Total length, bp	Max length, bp	N_{50}	N_{90}	N_{80}
A. motanka	91695	81481355	34661	2658	231	957
N. karyoxenos	102844	118343072	124812	1160	770	834
S. specki	35642	34908406	19484	1175	757	833
L. lanifica	15958	73316053	83793	11390	2873	4798
D. japonicum	160587	84236011	293136	2515	144	205
R. humris	189683	87108244	57896	2226	141	184

Table S2. Assessment of genome assembly completeness in six diplonemid species using KAT(Mapleson et al. 2017) and BUSCO(Waterhouse et al. 2019).

Species	KAT Completeness (%)	BUSCO Completeness (%)
A. motanka	75.56	21.1
N. karyoxenos	56.57	5.8
S. specki	59.26	11.1
L. lanifica	40.00	70.2
D. japonicum	45.83	70.8
R. humris	50.00	73.7

Mapleson D, Garcia Accinelli G, Kettleborough G, Wright J, Clavijo BJ. 2017. KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics* **33**: 574-576.

Waterhouse RM, Seppey M, Simao FA, Zdobnov EM. 2019. Using BUSCO to Assess Insect Genomic Resources. *Methods Mol Biol* **1858**: 59-74.

Table S3: List of used primers

Name	5'-3' Sequence
FwY1-mRNA-XbaI	GCAGTCTAGACGGGAGGCCGCCGATGAG
RevY1-mRNA-HindIII	GCAGAAGCTTGGGCGAAGTCATCGCGCTCG
1621_miniG1_F1	TACGAGAACTTCCGCACCAC
1621_miniG1_F2	ACAAAGTGATGCAGATC
1621_miniG1_R1	GGCGATCTGCATCACTTTGT
1621_miniG1_R2	CGCTTTGACAGCTTCCACAT
FwY2-kinase-XbaI	GCAGTCTAGACAGCGCGATCCGGAGGCCG
RevY2-kinase-HindIII	AGCGAAGCTTAGAAGTACGGGCTGCAGGGC
1621_miniG2_F1	GAGACGGTGGTGATCTGGGA
1621_miniG2_F2.2	AGACCGTGGACGCCACAGTC
1621_miniG2_R1	GGCGACTGTGGCGTCC
1621_miniG2_R2	GAAGTACGGGCTGCAGGG
FwY3-centriole-XbaI	GCAGTCTAGACAGCGGGCACACACATTGG
RevY3-centriole-HindIII	AGCGAAGCTTGCTGCCCTCCTGGTGCCATC
1621_miniG3_F1	AGCGGGCACACACATTGG
1621_miniG3_F2	GGGGCGATGATAAGCAAGTG
1621_miniG3_R1	GGTCGAGGTCCCACAATTTCA
1621_miniG3_R2	GGTGCCATCTGGGTGGAAA
FwY4-centriole-BamH1	ATTCGGATCCGCCGTCGACGAGTATGAAAAC
RevY4-centriole-HindIII	ATCCAAGCTTTGTGGCTGTCTTCTTTGCC
D.P._miniG_F1	GCTTCACGTTCGAGTTGCAG
D.P._miniG_R1	GTCTTCTTTGCCCCTTCCGA
1604_G1_F1	GAAAAGGGATCCGCCCACAA
1604_G1_R1	AAGCCCTTCTCGCACTTTCC
1604_G1_F2	CTGCGCTAGCAAGACATCAC
1604_G1_R2	CGTTGGCCTTCAAAGCACC
1604_G1_F3	CACCAGGCTTTGCGTAGACA
1604_G1_R3	CGTTCGAGTTGGACGGGTAG
1604_G2_F1	AGGGTACTACTTCCAGGGGT
1604_G2_R1	CACCTCAGAAGACGGGTGTG
1604_G2_F2	GTCTTCTGAGGTGTGCGTCC
1604_G2_R2	AGGCTTTCCTCTCAACCGTC
1604_G3_F1	GGTGATCCACTGTGAGGGTG
1604_G3_R1	ATCTCACGAACCTGGGCAAG
1604_G3_F2	AACGCGCCTGCTCTTATCTT
1604_G3_R2	GCCTTCATCCCGTCCATGAG
1604_G3_F3	CAGGTGATCGTGCTTGCTG

1604_G3_R3	CGATCTCATCGGGCACACC
1608_G1_F1	CTTCAGGGGAAAGCGTGTCG
1608_G1_R1	CTGCGTCGACGTGGCAAT
1608_G1_F2	GTGTGGACCCCGCGTA
1608_G1_R2	TCCTCACCCTTCTTCGTCGT
1608_G2_F1	ATCGGTGTCCTGACGGTTCT
1608_G2_R1	CAACGGCGAGGAGGATACC
1608_G2_F2	CGTTGGCTTCATCCACCAGT
1608_G2_R2	TTTAGCCACACACCTCACGC
1608_G3_F1	GAAGTCCTCGCACACGCAAA
1608_G3_R1	AACCGGAGAGCAGAATTCCC
1608_G3_F2	ATTCTGCTCTCCGGTTCCAG
1608_G3_R2	TCAGGAGGGCGATGATGAGA
1608_G3_F3	TTCTGCGAGGCTGTTGGTAT
1608_G3_R3	AATACCGCAGGCGAAGTTGC
1621_G1_F1	TACGAGAACTTCCGCACCAC
1621_G1_R1	GCATCACTTTGTTGTTCCCCA
1621_G1_F2	TGGGGAACAACAAAGTGATGC
1621_G1_R2	TAAAGAGCAGCCCGCAAGAA
1621_G1_F3	GCTCTTTACCAACAGGCCCAA
1621_G1_R3	ATCTCGTGCCCGAACTGCTC
1621_G1_F4	CGAGATGATGCACCAGCTTAC
1621_G1_R4	CCCTCCTGACAGACCTTGG
1621_G2_F1	AAGCCCATGCTTACTGTCCC
1621_G2_R1	TACTGGTCAGACCCTTCGCC
1621_G2_F2	CTGGGTTCCCCACTTGTCG
1621_G2_R2	CCTTGGCGTCCTGAGCGTA
1621_G3_F1	CCAGATGAACAGCATCTCGCT
1621_G3_R1	CCTCATCACCTCCTGGCAGT
1621_G3_F2	CCGAGATCACCGGGAAGATG
1621_G3_R2	TCGATGGCTTCGTCCATTGT