Carnival of SL RNAs:
Structural variants and the possibility of a common origin

Manja Marz[a], Nathalie Vanzo[f], Peter F. Stadler[a,b,c,d,e]

[a]*Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany*
[b]*Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, D-04103 Leipzig, Germany*
[c]*Fraunhofer Institut für Zelltherapie und Immunologie – IZI, Perlickstraße 1, D-04103 Leipzig, Germany*
[d]*Department of Theoretical Chemistry University of Vienna, Währingerstraße 17, A-1090 Wien, Austria*
[e]*Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA*
[f]*Centre de Biologie du Développement, UMR 5547 C.N.R.S. Université Paul Sabatier, F-31062 Toulouse Cedex, France*

Abstract

SL RNA *trans*-splicing so far has been described in seven eukaryotic phyla. A careful re-evaluation of predicted SL RNA secondary structures reveils striking similarities among Euglenida, Kinetoplastida, Dinophyceae, cnidarians, rotifers, nematods, platyhelminths and tunicates. Together with conserved sequence motifs, our analysis suggests an early common origin of spliced leader *trans*-splicing.

Key words:

Introduction

Many eukaryotes have two fundamental modes of spliceosomal splicing. *cis*-splicing is the excision of introns. In *trans*-splicing, on the other hand, a short *leader* sequence is transfered to the 5' end of a (typically protein coding) mRNA, which is usually processed from a polycistronic transcript. This leader contains the 5' hypermodified cap structure necessary for translational initiation [1]. In all cases so far, the leader sequence is derived from small non-coding RNA, the spliced-leader (SL) RNA, which has a common organization, Fig. 1.
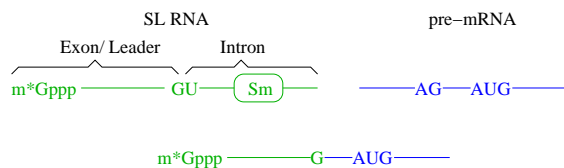

Figure 1: Schematic drawing of a typical SL RNA

The first SL RNAs were discovered in kinetoplastids a quarter of a century ago [2, 3]. A few years later, related RNAs were found in *Euglena gracilis* [4]. The first metazoan examples were the nematode *Caenorhabditis elegans* [5] and in platyhelminth *Schistosoma mansoni* [6]. Many more examples were soon found in related species see Tab. 2, but it took until the turn of the millenium before SL RNAs were discovered in additional metazoan phyla (cnidaria [7], tunicates [8, 9], rotifera [10]), and in dinoflagellates [11, 12]. In some species, multiple divergent copies of the SL RNA have been reported [13, 4, 14, 10], and some groups of species harbour two or more clearly distinct types of SL RNAs [15, 7, 14]. Of these SL1 and SL2 are distinguished also in the Rfam [16]. On the other hand, several model organism do not seem to utilize *trans*-splicing: despite substantial efforts, no evidence for *trans*-splicing could be gathered for *Drosophila melanogaster*, *Saccharomyces cerevisae*, and *Arabidopsis thaliana* [17, 18], and no evidence for *trans*-splicing has been reported in vertebrates despite the availability of extensive transcriptomics data.

The unexpectedly scattered distribution of SL *trans*-splicing across the phylogeny of Eukarya has prompted interest in the evolution of *trans*-splicing already two decades ago. To-date the two major competing hypotheses (reviewed in [7, 17, 18]) still remain unresolved:

1. SL RNAs have a common origin (Fig. 2a). Based

Email addresses: manja@bioinf.uni-leipzig.de (Manja Marz), vanzo@cict.fr (Nathalie Vanzo), studla@bioinf.uni-leipzig.de (Peter F. Stadler)
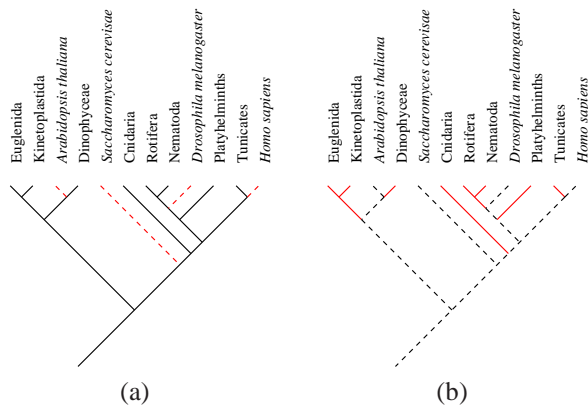
Figure 2: Evolution of SL RNAs ([7, 17]). (a) SL RNAs have a common eukaryotic ancestor. (b) SL RNAs derived independently in seven clades, which we here show to be improbable.

on the fact, that all SL RNAs have the same function of resolving Pol II transcripts to monocistronic mRNAs.

2. SL RNAs arose on multiple occasions (Fig. 2b)

The first hypothesis is supported primarily by the functional and mechanistic similarities of SL-*trans*-splicing, while the failure to detect sequence homology between SL RNAs from different phyla and the apparent disparity of SL RNA secondary structure has lead to the second hypothesis, see e.g. [19].

Here we re-evaluate the secondary structures from the previous studies. We find that much of the apparent differences between SL RNA secondary structures can be attributed to methodological problems. We find that SL RNAs in fact can fold into fairly similar secondary structure that even admit similar structural alternatives even though the sequences themselves show very few similarities across phyla. Although we cannot provide as definite proof, our evidence strongly leans towards a common origin of the SL-*trans*-splicing early in the evolution of Eukarya.

Results

Re-evaluation of secondary structures

Tab. 2 summarizes the published secondary structures together with selected structures that computed with parameters adjusted to each organism's optimal ambient temperature. Additional structural alternatives can be found in the supplemental material. Many of the published structures were obtained using `mfold` with standard parameters (i.e., a temperature of $T = 37°C$),

which is lethal for most of the organisms in question, in particular the unicellular ones [20]. In several cases, which we will discuss in detail in the next paragraphs, our analysis deviates drastically from the published data. Small corrections and differences are briefly mentionend in the Methods section.

Dinoflagellata. The *K. brevis* SL RNA was reported in [11] with a donor splice site that is not consistent with the EST data from the same work. Moreover, the reported structure is energetically unfavourable at all temperatures, with and without an additional constraints that forces the Sm binding site to be unpaired. The structural analysis makes it likely, furthermore, that this SL RNA is 24nt shorter than the published sequence, consistent with proposed A_5 termination signal [11]. Completely different models of dinoflagellate SL RNAs are proposed in [12]. For both *K. micrum* and *P. piscicida* the Sm binding site is shown upstream of the splice site instead of downstream. If correct, this would indicate major differences in the organization of the *trans*-splicing machinery. However, the Sm proteins and U1 snRNA seem to be conserved in dinoflagelates [21]. The *P. minimum* is reported without Sm binding site. A simple sequence alignment of this sequence with the SL RNA of *K. brevis* shows that the SL RNA is clearly conserved among these alveolates. Thus, either the published *K. brevis* is much too long, or the dinoflagelate SL RNA sequences from the work of [12] are truncated. Adding 4'nt on the 5' side and 79nt on the 3' side to the published sequence from the genomic DNA available from GenBank (**EF143079.1**, **EF143082.1**, **EF143084.1**), we easily obtain structure models that conform to the common organization of SL RNAs of other phyla. Detailed alignments can be found in the Electronic Supplement at www.bioinf.uni-leipzig.de/Publications/SUPPLEMENTS/09-009.

Euglenida. Stem IV of the published *Entosiphon* SL RNA [22] has a positive folding energy under all parameter settings, strongly suggesting that this substructure is not formed. For *E. sulcatum* we identify a possible alternative Sm binding site, which would suggest that this SL RNA would be 16 nt shorter, see Tab. 2. The extended sequence then folds similarly to the known secondary structures in other phyla.

Tunicata. Within tunicates, the sequence similarity between *C. intestinalis* and *O. dioica* is much higher than previously reported [9]. We therefore expanded the *C. intestinalis* sequence 20nt to both sides, covering the well-conserved part of the genomic locus.
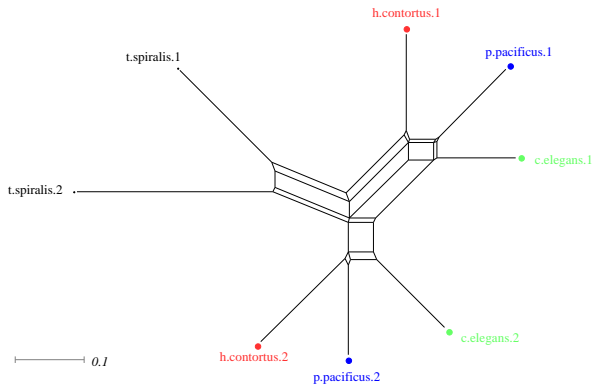
Figure 3: Neighbor-net analysis confirms that nematode SL1 and SL2 RNAs form distinct paralog groups, with the possible exception of *Trichinella*.

Sm Binding Motif. In most of the previous publications, structures were computed with the constraint of an externally unpaired Sm-binding site. However, most of the sequences can fold in a haipin structure in which (most of) the Sm-binding site is located in an accessible loop, as reported e.g. for *Euglena* [4] and *Hydra* [7]. Therefore we report structures with and without a constraint on the Sm-binding site in Tab. 2.

Phylum specific alignments

The SL RNAs are fairly well conserved as sequence level within each of the 7 phyla. Alignments can be found in the Electronic Supplement. An exception is the 3' part of the *T. brucei* SL RNA, which has a long insert upstream of the Sm-binding site.

The two paralogous SL RNA in *Hydra* are probably recent duplicates, sharing a highly conserved 17nt long block in the 3'region. Neither paralogs has a recognizable homolog in *Nematostella vectensis*.

Most nematodes have multiple SL RNAs. Within Rhabditina, there are clearly dicernible paralog groups SL1 and SL2 [23]. The SL RNAs of *Trichinella* (clade Dorylaimia) [14] do not fit well into this scheme, however (Fig. 3).

A sequence-based alignment between two adjacent phyla shows no conserved regions, as expected. However, aligning just exonic parts reveals some similarities between rotifers and nematods, Fig. 4.

Ubiquitous Sequence Features

Not surprisingly, the donor splice-site, `G|GU`, and the U-rich Sm-binding site are well conserved within each phylum and can be detected easily using `Meme` [24] in

```
# STOCKHOLM 1.0

c.elegans.1    GGTTT.....AA.TTACCCAAGTTT....GAG
p.pacificus1   GGTTT.....AA.TTACCCAAGTTT....GAG
h.contortus1   GGTTT.....AA.TTACCCAAGTTT....GAG
c.elegans2     GGTTT..TA.....ACCCA.GTTACT.CAAG
h.contortus2   GGTTT..TA.....ACCCA.GTATCT.CAAG
p.pacificus2   GGTTTAT........ACCCA.GTATCT.CAAG
a.ricciae1     GGCTTATTACAACTTA.CCAAG........AG
philodina      GGCTTATTACAACTTA.CCAAG........AG
t.spiralis     GG..TAT......TTA.CCA.G.ATCTAAAAG
//
```

Figure 4: A sequence feature shared between rotifer and nematod SL RNAs.
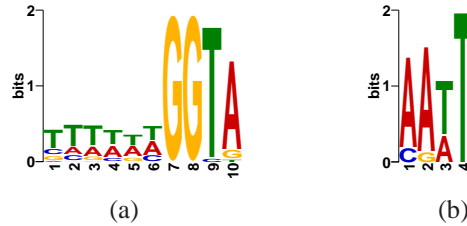
Figure 5: Consensus of (a) Donor splice site and (b) Sm binding site of SL RNA for all seven clades.

the complete data set, Fig. 5). No other sequence similarities have been reported previously, and a sequence alignment does not pick up any additional motifs. However, a comparison of the IUPAC-code consensus sequences of the 7 phyla identifies several weak sequence features, Fig. 6. Details can be found in the Electronic Supplement.

- Upstream of SL RNA stem I, there is an A/C-rich region containing an occasional `T`. We denote this region by H^* since `H`= {A,C,T}. This `G`-poor sequence conform to the proposed
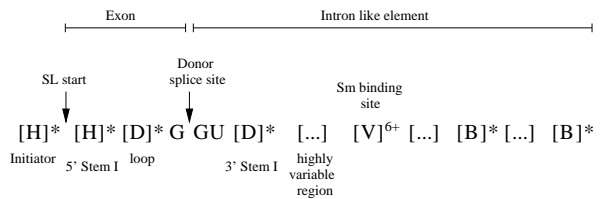

Figure 6: Common sequence features of SL-RNAs: Initiator lacking of `G`, a variable length of Stem I, which 5' part consists mainly of `A,C,T` followed by a loop, which for SL RNA-α contains mainly `T` and for SL RNA-β `A,G`. The Donor Splice site is located downstream of the loop in the hairpin with the highly conserved `GGU`-motif. Stem I continous nearly without any `C`. After a highly variable region, the known Sm binding site with the common motif `AATTTTTGG` and a possible unpaired region a last part often structured as a hairpin shows no `A` to be paired.

initiator sequences `UNCU` in euglenoids [19], `YA`$^{+1}$`NU/AYYY` generally observed in metazoans [25], `YYHBYA`$^{+1}$`ACU` described for trypanosomes [26] and `CA`$^{+1}$`AUCUC` in *K. brevis* [11].

- The loop and 3' part of the first hairpin are depleted in `C`. This may be associated with constraints associated with the splice-site and/or the binding affinity between SL RNA to mRNA.

- The 5' part of the first hairpin (most of the exon) shows a lack of `G`, explained by the pairing with the `D= {A, G, T}` region mentioned above. This H-block is less well-conserved than the 3' D-block due to the possibility of G-U pairing. A succession of `A-U` pairs followed by `U-A` pairs in the outer part of stem I (just "below" the splice site) was reported in [19] for euglenoids. In several other clades such a stringent pattern is not visible, however.

- The loop of stem I shows a clear prevalence of `W= {A,U}`. Depending on the A/U-ratio, two subtypes can be distinguished (see below).

- The donor-splicing site is highly conserved with the sequence `G|GU`, Fig. 5. It is always located downstream of loop I.

- The region between stem I and the Sm-binding site is highly variable not only between but also within each phylum.

- The SM binding site consists of a highly conserved D-region. The common pattern is `AAUUUUUGG`, Fig. 5b, with the sole exception of *Oikopleura dioica*.

- Stems downstream of the Sm binding site show a highly conserved `C`,`G`-rich `B= {C,G,U}` stem. The loop in contrast is A-rich.

Taken together, we find recognizable sequence constraints covering almost the entire SL RNA gene.

The loop region of stem I clearly distinguished two sub-types of SL RNAs. In class α, the loop consists mostly of `U`s, while the loop in class β is essentially free of `U`s. In most metazoans with more than SL RNA (e.g. *C. elegans*, *H. contortus*, *P. pacificus*, *T. spiralis*, and *S. mediterannea* both types are present. The two rotifers *A. ricciae* and *Rhilodina sp.*, as well as *C. intestinalis* have only type β, while otherwise type α appears to be prevalent. In the cnidarian *Hydra* the classification remains ambiguous.

Secondary Structure Analysis

The secondary structures obtained by folding SL RNAs with and without constraints, and those reported in the literature, can be classified into 10 groups. Using each of these structures as constraints, we computed the folding energies of all known SL RNAs. Tab. 1 lists the results, which can be summarized as follows:

1. Most euglenid SL RNA folds into 4 hairpins. However, other phyla such as cnidaria, rotifers, or tunicates (due to the shorter sequence) never fold into such a structure.
2. The SL RNAs of all species can fold into a single hairpin including both the donor splice site and a Sm binding site. Except for *Oikopleura*, however, this structure is never energetically preferred.
3. Stem I upstream of the Sm-binding site is highly conserved. All SL RNAs except that of *Oikopleura* can form this structure. In most cases, this structure is also thermodynamically highly favoured, (see Tab. 1, 3rd and 4th column).
4. For the Sm-binding site either a completely unbound external structure or a mostly unbound location within a hairpin loop were discussed. Both structural models are plausible, the hairpin variant is always energetically favorable.
5. In most SL RNAs, stem I folds can attain two different hairpins, see below (Fig. 7).
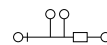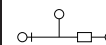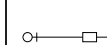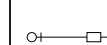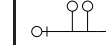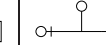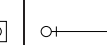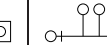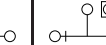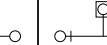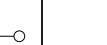
Interestingly, the sequence underlying stem I can form two alternative structures, Fig 7. Since the sequences are highly divergent between phyla, it is very unlikely to observe the same structural alternative throughout the entire dataset. We therefore conclude that the conformational change between the two alternatives is required for SL RNA function.

Secondary structure alignments show common features within each phylum, as well as for rotifers and nematods together. Weak signals fuer conserved structure elements were obtained by aligning all prostomia together, Fig. 8. The corresponding alignments are compiled in the Electronic Supplement.

Discussion

SL RNAs share a common function in *trans*-splicing [18, 17]. They provide a short exonic leader sequence with an unusual trimethyl-guanosine cap and they play an active role in the spliceosome-catalyzed processing by virtue of binding to the Sm protein. SL RNAs are found in wide range of eukaryotic phyla, they are conspicuosly absent in many major clades suggesting a complex evolutionary history.

Table 1: Secondary structures, their MFE_{min} for all constraint folded MFE_{cons} and their ratio of contraint folding to minimum MFE (MFE_{cons}/MFE_{min}).

Organism	T	MFE_{min}	S1		S2		S3		S4		S5		S6		S7		S8		S9		S10	
E. gracilis	29	-35.02	-35.02	1.00	-23.38	0.67	-20.51	0.59	-6.22	0.18	-22.52	0.64	-27.15	0.78	-21.54	0.62	-31.50	0.90	-29.27	0.84	-21.94	0.63
P. curvicauda	22	-41.51	-41.51	1.00	-34.19	0.82	-24.35	0.59	-11.65	0.28	-32.71	0.79	-32.86	0.79	-23.18	0.56	-34.98	0.84	-35.13	0.85	-25.45	0.61
C. acus	22	-54.10	-54.10	1.00	-40.10	0.74	-27.01	0.50	-13.84	0.26	-46.80	0.87	-42.92	0.79	-33.50	0.62	-49.63	0.92	-40.73	0.75	-28.97	0.54
R. costata	23	-42.69	-42.69	1.00	-32.29	0.76	-17.57	0.41	-6.61	0.15	-27.23	0.64	-24.78	0.58	-14.38	0.34	-32.01	0.75	-30.09	0.70	-19.69	0.46
M. pellucidum	23	-35.73	-35.73	1.00	-28.32	0.79	-15.28	0.43	-7.66	0.21	–	0	-22.26	0.62	-20.05	0.56	-23.67	0.66	-27.55	0.77	-17.75	0.50
E. sulcatum-long	25	-35.98	–	0	–	0	–	0	-7.13	0.20	-35.98	1.00	-33.35	0.93	-14.13	0.39	–	0	–	0	–	0
*E. sulcatum-short	25	-35.07	-16.47	0.47	-19.29	0.55	-11.84	0.34	-7.15	0.20	–	0	-35.07	1.00	-33.86	0.97	–	0	-33.66	0.96	-27.13	0.77
T. cruzi	23	-44.05	-34.93	0.79	-38.23	0.87	-32.84	0.75	-24.96	0.57	-33.17	0.75	-38.58	0.88	-42.24	0.96	-34.75	0.79	-44.05	1.00	-43.74	0.99
T. vivax	23	-53.64	-46.90	0.87	-53.64	1.00	-43.94	0.82	-22.16	0.41	–	0	–	0	-42.79	0.80	-48.25	0.90	-51.40	0.96	-48.61	0.91
T. brucei	23	-75.05	-60.45	0.81	-75.05	1.00	-34.05	0.45	-15.95	0.21	-48.89	0.65	-45.54	0.61	-44.64	0.59	-55.73	0.74	-71.99	0.96	-43.42	0.58
**T. brucei*5'SM	23	-80.31																				
L. collosoma	28	-27.43	–	0	-27.43	1.00	-21.07	0.77	-15.60	0.57	–	0	-23.54	0.86	-22.69	0.83	–	0	-21.85	0.80	-22.53	0.82
C. fasciculata	20	-30.83	-25.40	0.82	-30.83	1.00	-29.69	0.96	-21.44	0.70	-21.20	0.69	-23.13	0.75	-30.10	0.98	–	0	-30.76	1.00	-29.62	0.96
L. enriettii	36	-24.72	-12.68	0.51	-17.26	0.70	-11.00	0.44	-10.36	0.42	–	0	–	0	-24.72	1.00	–	0	–	0	–	0
K. brevis	20	-45.34	-38.06	0.84	-45.34	1.00	-40.31	0.89	-22.47	0.50	–	0	-33.03	0.73	–	0	–	0	-21.04	0.46	-22.77	0.50
**K. brevis*3'SM	20	-50.23	-41.21	0.82	-39.88	0.79	-22.74	0.45	-19.30	0.38	-45.41	0.90	-50.23	1.00	-39.31	0.78	-28.14	0.56	-26.02	0.52	-16.17	0.32
K. micrum-long	20	-54.66	–	0	-53.29	0.97	-53.56	0.98	-26.06	0.48	–	0	-23.93	0.44	-30.62	0.56	–	0	-47.97	0.88	-54.66	1.00
Hydra-A	18	-18.06	–	0	-10.96	0.61	-9.26	0.51	-6.60	0.37	–	0	–	0	-18.06	1.00	–	0	-12.63	0.70	-14.35	0.79
Hydra-B1	18	-18.80	–	0	–	0	-18.80	1.00	-7.63	0.41	–	0	-12.19	0.65	–	0	–	0	–	0	-18.03	0.96
Hydra-B2	18	-18.03	–	0	-18.03	1.00	-11.79	0.65	-8.20	0.45	–	0	-15.86	0.88	-17.23	0.96	–	0	-14.70	0.82	-10.85	0.60
Hydra-B3	18	-24.27	–	0	-11.62	0.48	-8.82	0.36	-8.20	0.34	–	0	-22.17	0.91	-20.88	0.86	–	0	-24.27	1.00	-14.93	0.62
A. ricciae	24	-46.48	–	0	–	0	-36.14	0.78	-6.76	0.15	–	0	–	0	–	0	–	0	–	0	-46.48	1.00
Philodina sp.	20	-49.82	–	0	–	0	-42.28	0.85	-7.77	0.16	–	0	–	0	–	0	–	0	–	0	-49.82	1.00
C. elegans 1	20	-44.73	-30.04	0.67	-43.54	0.97	-28.75	0.64	-12.27	0.27	-28.25	0.63	-33.80	0.76	-12.27	0.27	–	0	-44.73	1.00	-31.67	0.71
C. elegans 2	20	-58.38	-43.52	0.75	-57.14	0.98	-47.88	0.82	-23.35	0.40	-33.79	0.58	-32.85	0.56	-34.06	0.58	-52.20	0.89	-58.38	1.00	-49.20	0.84
T. spiralis 1	36	-30.15	-20.92	0.69	-27.80	0.92	-21.08	0.70	-8.34	0.28	–	0	-17.41	0.58	-19.06	0.63	–	0	-30.15	1.00	-20.71	0.69
T. spiralis 2	36	-19.51	–	0	-19.51	1.00	-8.40	0.43	-7.91	0.41	–	0	-17.78	0.91	-11.74	0.60	–	0	-18.27	0.94	-11.81	0.61
P. pacificus 1	20	-54.23	-19.16	0.35	-50.76	0.94	-36.91	0.68	-13.79	0.25	-38.17	0.70	-36.34	0.67	-37.31	0.69	-54.03	1.00	-54.23	1.00	-37.30	0.69
P. pacificus 2	20	-51.68	-48.85	0.95	-48.98	0.95	-39.48	0.76	-19.94	0.39	-29.28	0.57	-32.14	0.62	-38.51	0.75	-42.26	0.82	-51.68	1.00	-37.19	0.72
H. contortus 2	25	-41.15	-33.91	0.82	-39.91	0.97	-35.12	0.85	-17.74	0.43	-17.52	0.43	-17.81	0.43	-41.15	1.00	-34.69	0.84	-39.26	0.95	-40.79	0.99
S. mansoni	28	-31.43	–	0	-23.53	0.75	-26.65	0.85	-21.63	0.69	–	0	-26.41	0.84	-21.63	0.69	–	0	–	0	-31.43	1.00
F. hepatica	16	-57.02	-32.80	0.58	-53.75	0.94	-32.89	0.58			–	0	-41.96	0.74	-41.80	0.73	–	0	-57.02	1.00	-52.15	0.91
S. mediterranea-1	22	-44.16	-32.07	0.73	-43.87	0.99	-37.88	0.86	-21.96	0.50	-15.99	0.36	-28.24	0.64	-26.57	0.60	-31.91	0.72	-44.16	1.00	-41.19	0.93
E. multilocularis	35	-42.20	-33.35	0.79	-36.34	0.86	-26.78	0.63	-19.78	0.47	–	0	-42.20	1.00	-37.35	0.89	–	0	-37.37	0.89	-28.69	0.68
C. intestinalis	21	-14.98	–	0	–	0	-5.60	0.37	-5.71	0.38	–	0	–	0	-14.98	1.00	–	0	–	0	–	0
*O. dioica*SM1	20	-14.24	–	0	–	0	–	0	-14.24	1.00	–	0	–	0	–	0	–	0	–	0	–	0
*O. dioica*SM2+9	20	-16.35	–	0	-0.82	0.05	-16.35	1.00	-14.24	0.87	–	0	–	0	–	0	–	0	–	0	–	0

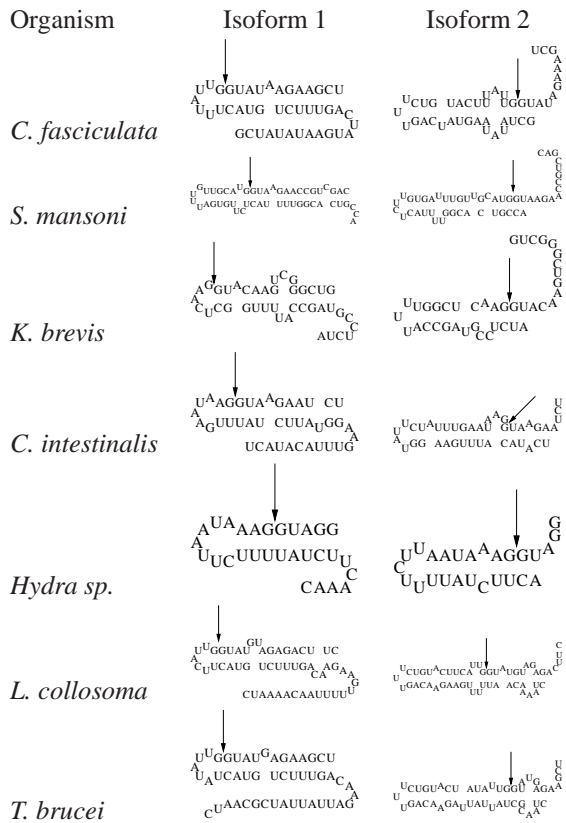Legend: 0 | 0.6-0.7 | 0.8-0.9 | 0.9-1.0

Figure 7: Two alternative secondary structures of stem I can be formed in all phyla. In case of *C. intestinalis* 16 nucleotides were added to the 5' end of the published sequence.
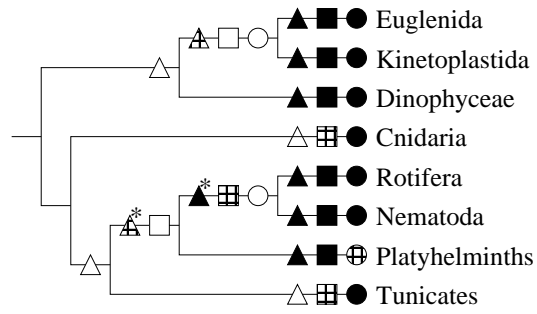

Figure 8: Sequence and structure similarities obtained by standard alignment programs `ClustalW` (● – full sequence, ■ – exon only) and the structural alignment tool `locarnate` (▲ – full sequence). Filled symbols indicate similarities, empty symbols indicate that there are no obvious common features; * – alignable with nematods SL1 only.

A re-evaluation of the available evidence on SL RNAs across Eukarya shows that there a several sequence and structure similarities among SL RNAs that together strongly suggest not only a common function but also a common mechanism. SL RNAs share:

1. The relative positioning of the splice-donor site and the Sm-binding site is the same.

2. There is a weak but recognizable shared sequence pattern, suggesting common descent or common selection pressures.

3. Structural similarties between SL RNA are much greater than recognized in previous work when natural ambient temperatures are taken into account.

4. All SL RNAs share the possibility of two alternative conformations of stem I, suggesting that the structural transition between the two states is involved in SL RNA function.

While logically possible, it seems quite unlikely that *trans*-splicing arose *de novo* several times to give rise to SL RNAs that always share the same sequence and structure constraints, since these similarities suggest that they interacting with the same proteins. Our analysis thus favours a common orgin of *trans*-splicing, with frequent losses througout the eukaryotic tree [8].

The inability to find SL RNA genes in genomic sequence data does make a good argument against the existence of *trans*-splicing since mutation studies in kinetoplasts and nematodes showed that much of the sequence and structure can be disrupted without consequence to function in *trans*-splicing [17, 27], indicating that SL RNAs are evolutionary very plastic. This property is shared with several other functionally crucial ncRNAs such as telomerase RNA [28, 29] and 7SK RNA [30, 31], which so far also have been found only in a rather scattered collection of clades. Is is quite conceivable, therefore, that SL RNA *trans*-splicing will be discovered in multiple additional phyla as sufficiently large EST libraries become available. The reason is that SL *trans*-splicing might well be a relatively rarely used processing mechanism in some clades. For example, the basal eukaryotes *Giardia lamblia* [32, 33] and *Trichinella spiralis* [34] have functional spliceosomes but there are only a handful of mRNAs with spliceosomal introns in these species. Similarly, introns are quite rare in *Saccharomyces cerevisae* [35]. We argue, therefore, that the absence of SL *trans*-splicing is plausibly established only in a few organisms such as mammals, fruitflies, yeast, or *Arabidopsis*. For these, the transcriptome is known sufficiently well to rule out with near certainty that there unrelated mRNAs with a common leader sequence of unclear origin.

If one accepts a common origin of SL RNAs, their structural evolution must have followed one of the three scenaria outline in Fig. 9: (1) Most ancestral SL RNAs
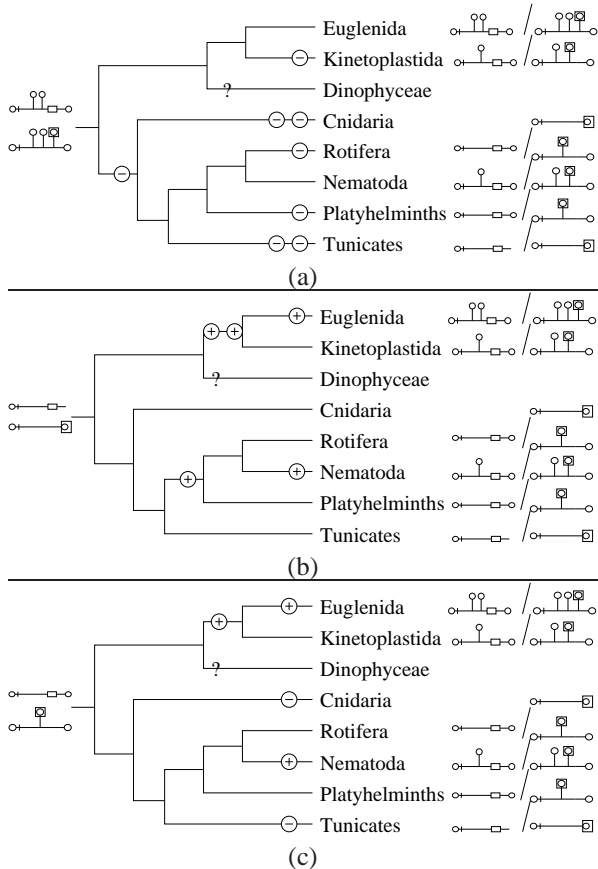
Figure 9: Three alternative scenaria for the evolution of SL RNA secondary structure. (a) Ancestral state with 4 hairpins, (b) ancestral state containing only features that are still contained in all present-day SL RNAs, (c) most parsimonious scenario minimizing the number of hairping gain/loss events.

contain 4 hairpins close to the Sm binding site, Fig. 9a, with subsequent simplifications in some clades. (2) Alternatively, the structural complexity may have increased, Fig. 9b. (3) A maximum parsimony analysis in which we interpret the hairpins as characters, also points a structurally fairly simple ancestor, Fig. 9c. The inferred ancestral states in these scenaria should help with constructing descriptors for homology search of SL RNAs.

In summary, the re-evaluation of SL RNA sequences and secondary structures does not definitively resolve the questions regarding the origin of SL transsplicing, but it provides several lines of circumstantial evidence that consistently favour an ancient common origin in Eukarya.

Methods

Sequences. SL RNA sequences were taken from the following publications: Chordata: *Ciona intestinalis* [8], *Oikopleura dioica* [9]; Nematoda: *Caenorhabditis elegans* [5], *Ascaris* [36], *Wucheria bancrofti* [37], *Haemonchus contortus* [38], *Pristionchus pacificus* [39], *Trichinella spiralis* [14]; Platyhelminthes: *Schistosoma mansoni* [6], *Fasciola hepatica* [40], *Echinococcus multilocularis* [41], *Schmidtea mediterranea* [42]; Rotifera: *Philodina sp.*, *Adineta ricciae* [10]; Cnidaria: *Hydra sp.* [7]; Euglenida: *Euglena gracilis* [4], *Entosiphon sulcatum* [22], *Cyclidiopsis acus*, *Phacus curvicauda*, *Rhabdomonas castata*, *Menoidium pellucidum* [19] Kinetoplastida: *Trypanosoma cruzi*, *T. vivax*, *T. brucei*, *Leptomonas collosoma* [2, 3], *Leishmania enriettii* [43], *Crithidia fasciculata* [44], *Bodo caudatus* [45]. Dinoflagellata: *Karenia brevis* [11], *Karlodinium micrum*, *Pfiesteria piscicida*, *Prorocentrum minimum* [12].

A `blastn` search in GenBank returns additional full length SL RNA sequences, which were used in subsequent analyses: Nematoda: *Ascaris* **AB022045.1**, *Loa* **U31638.1**, *Mansonella* **AJ279033.1**, *Acanthocheilonema* **U31646.1**, *Onchocerca* **M37737.1**, *Foleyella* **AJ250988.1**, *Setaria* **AF282181.1**, *Toxocara* **U65503.1**, *Enterobius* **AY234784.1**, *Nippostrongylus* **EB185208.1**, *Meloidogyne* **CN443291.1**, *Haemonchus* **CA994732.1**, *Teladorsagia* **CB043522.1**; Platyhelminthes: *Echinostoma* **U85825.1**; Rotifera: *Bdelloidea* **AY823993.1**; Kinetoplastida: *Herpetomonas* **AY547489.1**, *Phytomonas* **AF243335.1**, *Wallaceina* **AY547488.1**.

Alignments. Sequences based alignments were calculated by `ClustalW` 2.0.10 with standard parameters [46]. Structure-based alignments were calculated by `Locarnate` [47].

Secondary Structures. Secondary Structures were computed using the `Vienna RNA Package` 1.8.2 [48]. We used `RNAalifold` for consensus structure prediction, `RNAsubopt` for assessing structural alternatives, and `RNAeval` to determine the folding energy of published RNA secondary models. Whenever individual sequences were folded, the temperature parameter was set to temperature under which the organisms are handled, see Tab. 2 for references.

The *S. mansoni* structure proposed in [40] based the sequence published in [6] contained small differences in sequences. We used the original sequence folded into the published structure.

Other Software. We used the pattern search program `Meme` [24] with default parameters to search for conserved sequence motifs. Phylogenetic analysis was performed with the `SplitsTree` package [49].
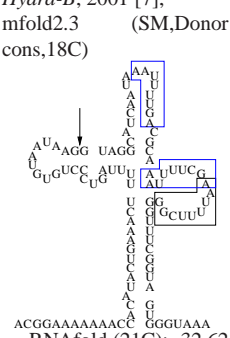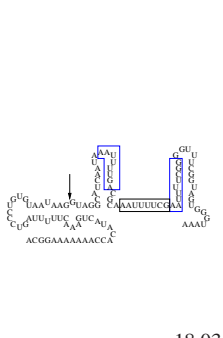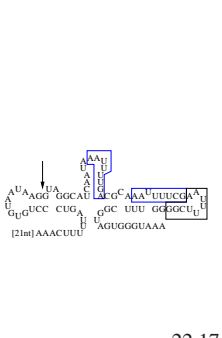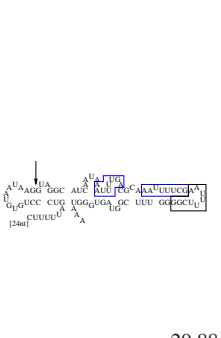
Acknowledgements
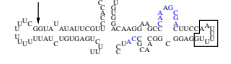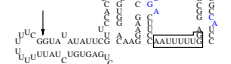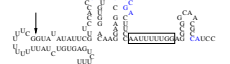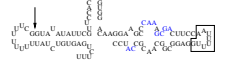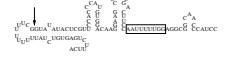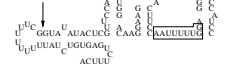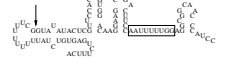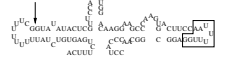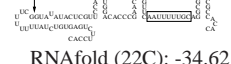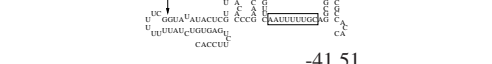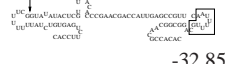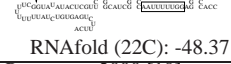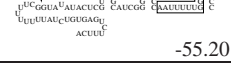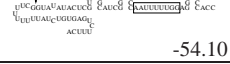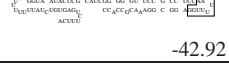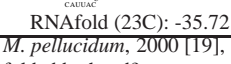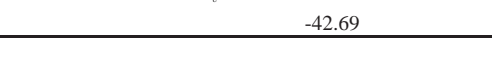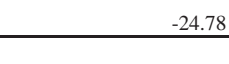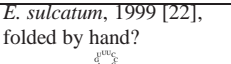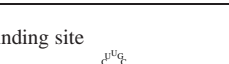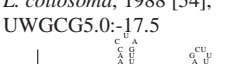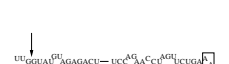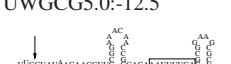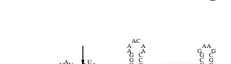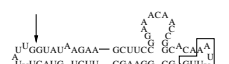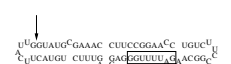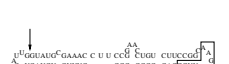
References
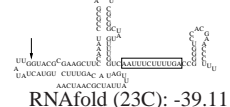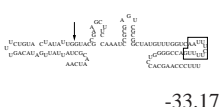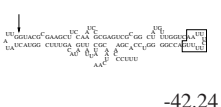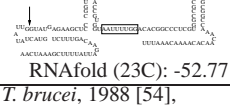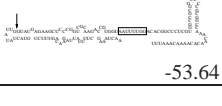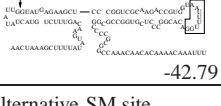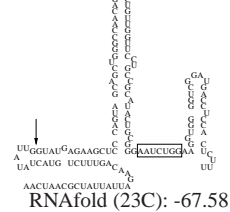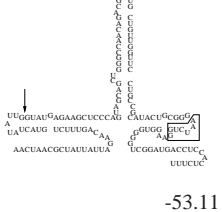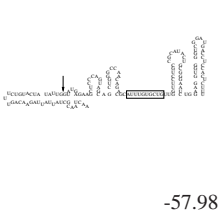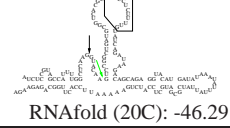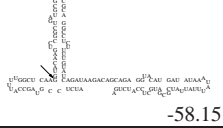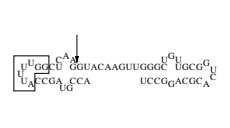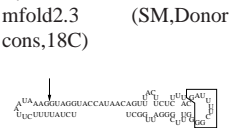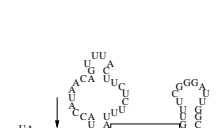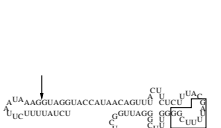
[1] W J Murphy, K P Watkins, and N Agabian. Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for trans splicing. *Cell*, 47(4):517–525, Nov 1986.

[2] M Milhausen, R G Nelson, S Sather, M Selkirk, and N Agabian. Identification of a small RNA containing the trypanosome spliced leader: a donor of shared 5' sequences of trypanosomatid mRNAs? *Cell*, 38(3):721–729, Oct 1984.

[3] T De Lange, T M Berkvens, H J Veerman, A C Frasch, J D Barry, and P Borst. Comparison of the genes coding for the common 5' terminal sequence of messenger RNAs in three trypanosome species. *Nucleic Acids Res*, 12(11):4431–4443, Jun 1984.

[4] L H Tessier, M Keller, R L Chan, R Fournier, J H Weil, and P Imbault. Short leader sequences may be transferred from small RNAs to pre-mature mRNAs by trans-splicing in Euglena. *EMBO J*, 10(9):2621–2625, Sep 1991.

[5] M Krause and D Hirsh. A trans-spliced leader sequence on actin mRNA in C. elegans. *Cell*, 49(6):753–761, Jun 1987.

[6] A Rajkovic, R E Davis, J N Simonsen, and F M Rottman. A spliced leader is present on a subset of mRNAs from the human parasite Schistosoma mansoni. *Proc Natl Acad Sci U S A*, 87(22):8879–8883, Nov 1990.

[7] N A Stover and R E Steele. Trans-spliced leader addition to mRNAs in a cnidarian. *Proc Natl Acad Sci U S A*, 98(10):5693–5698, May 2001.

[8] A E Vandenberghe, T H Meedel, and K E Hastings. mRNA 5'-leader trans-splicing in the chordates. *Genes Dev*, 15(3):294–303, Feb 2001.

[9] P Ganot, T Kallesoe, R Reinhardt, D Chourrout, and E M Thompson. Spliced-leader RNA trans splicing in a chordate, Oikopleura dioica, with a compact genome. *Mol Cell Biol*, 24(17):7795–7805, Sep 2004.

[10] N N Pouchkina-Stantcheva and A Tunnacliffe. Spliced leader RNA-mediated trans-splicing in phylum Rotifera. *Mol Biol Evol*, 22(6):1482–1489, Jun 2005.

[11] K B Lidie and F M van Dolah. Spliced leader RNA-mediated trans-splicing in a dinoflagellate, Karenia brevis. *J Eukaryot Microbiol*, 54(5):427–435, Sep-Oct 2007.

[12] H Zhang, Y Hou, L Miranda, D A Campbell, N R Sturm, T Gaasterland, and S Lin. Spliced leader RNA trans-splicing in dinoflagellates. *Proc Natl Acad Sci U S A*, 104(11):4618–4623, Mar 2007.

[13] M Blaxter and L Liu. Nematode spliced leaders–ubiquity, evolution and utility. *Int J Parasitol*, 26(10):1025–1033, Oct 1996.

[14] J Pettitt, B Müller, I Stansfield, and B Connolly. Spliced leader trans-splicing in the nematode Trichinella spiralis uses highly polymorphic, noncanonical spliced leaders. *RNA*, 14(4):760–770, Apr 2008.

[15] D Evans and T Blumenthal. trans splicing of polycistronic caenorhabditis elegans pre-mRNAs: analysis of the SL2 RNA. *Mol Cell Biol*, 20(18):6659–6667, Sep 2000.

[16] S Griffiths-Jones, S Moxon, M Marshall, A Khanna, S R Eddy, and A Bateman. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res*, 33(Database issue):121–124, Jan 2005.

[17] T W Nilsen. Evolutionary origin of SL-addition trans-splicing: still an enigma. *Trends Genet*, 17(12):678–680, Dec 2001.

[18] K E Hastings. SL trans-splicing: easy come or easy go? *Trends Genet*, 21(4):240–247, Apr 2005.

[19] C Frantz, C Ebel, F Paulus, and P Imbault. Characterization of trans-splicing in Euglenoids. *Curr Genet*, 37(6):349–355, Jun 2000.

[20] CS MCDANIEL. Microorganism coating components, coatings, and coated surfaces. *WIPO*, 2005. URL {http://www.wipo.int/pctdb/en/wo.jsp?IA= WO2005026269&wo=2005026269&DISP%LAY=DESC}.

[21] E Alverca, S Franca, and S M Díaz de la Espina. Topology of splicing and snrnp biogenesis in dinoflagellate nuclei. *Biol Cell*, 98(12):709–720, Dec 2006.

[22] C Ebel, C Frantz, F Paulus, and P Imbault. Trans-splicing and cis-splicing in the colourless Euglenoid, Entosiphon sulcatum. *Curr Genet*, 35(5):542–550, Jun 1999.

[23] D B Guiliano and M L Blaxter. Operon conservation and the evolution of trans-splicing in the phylum Nematoda. *PLoS Genet*, 2(11), Nov 2006.

[24] TL Bailey and C Elkan. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, pages 28–36, 1994.

[25] H Zhang, D A Campbell, N R Sturm, and S Lin. Dinoflagellate spliced leader RNA genes display a variety of sequences and genomic arrangements. *Mol Biol Evol*, 2009.

[26] H Luo, G Gilinger, D Mukherjee, and V Bellofatto. Transcription initiation at the TATA-less spliced leader RNA gene promoter requires at least two dna-binding proteins and a tripartite architecture that includes an initiator element. *J Biol Chem*, 274(45):31947–31954, Nov 1999.

[27] P A Maroney, G J Hannon, J D Shambaugh, and T W Nilsen. Intramolecular base pairing between the nematode spliced leader and its 5' splice site is not essential for trans-splicing in vitro. *EMBO J*, 10(12):3869–3875, Dec 1991.

[28] J. L. Chen, M. A. Blasco, and C. W. Greider. Secondary structure of vertebrate telomerase RNA. *Cell*, 100:503–514, 2000.

[29] M Xie, A Mosig, X Qi, Y Li, P F Stadler, and J J Chen. Structure and function of the smallest vertebrate telomerase RNA from teleost fish. *J Biol Chem*, 283(4):2049–2059, Jan 2008.

[30] Andreas R. Gruber, Dorota Koper-Emde, Manja Marz, Hakim Tafer, Stephan Bernhart, Gregor Obernosterer, Axel Mosig, Ivo L. Hofacker, Peter F. Stadler, and Bernd-Joachim Benecke. Invertebrate 7SK snRNAs. *J. Mol. Evol.*, 107-115:66, 2008.

[31] Andreas Gruber, Carsten Kilgus, Axel Mosig, Ivo L. Hofacker, Wolfgang Hennig, and Peter F. Stadler. Arthropod 7SK RNA. *Mol. Biol. Evol.*, 1923-1930:25, 2008.

[32] G. B. Simpson, Alastair, Erin K. MacQuarrie, and Andrew J. Roger. Eukaryotic evolution: Early origin of canonical introns. *Nature*, 419:270, 2002.

[33] X S Chen, W T White, L J Collins, and D. Penny. Computational identification of four spliceosomal snRNAs from the deep-branching eukaryote *Giardia intestinalis*. *PLoS ONE*, 3:
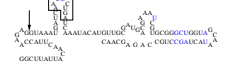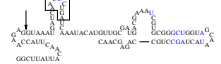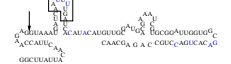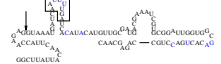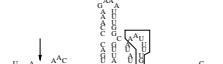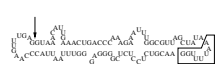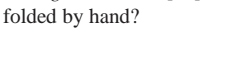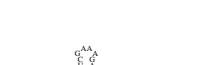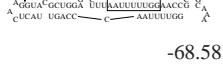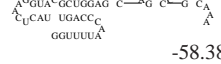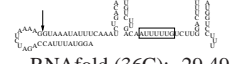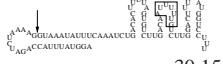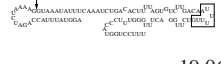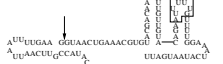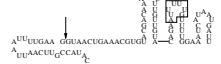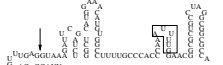
e3106, 2008.

[34] A Simoes-Barbosa, D Meloni, J A Wohlschlegel, M M Konarska, and P J. Johnson. Spliceosomal snRNAs in the unicellular eukaryote *Trichomonas vaginalis* are structurally conserved but lack a 5'-cap structure. *RNA*, 14:1617–1631, 2008.

[35] E Bon, S Casaregola, G Blandin, B Llorente, C Neuvéglise, M Munsterkotter, U. Guldener, H W Mewes, J Van Helden, B Dujon, and C. Gaillardin. Molecular evolution of eukaryotic genomes: hemiascomycetous yeast spliceosomal introns. *Nucleic Acids Res.*, 31:1121–1135, 2003.
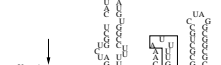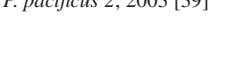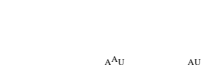
[36] P A Maroney, J A Denker, E Darzynkiewicz, R Laneve, and T W Nilsen. Most mRNAs in the nematode ascaris lumbricoides are trans-spliced: a role for spliced leader addition in translational efficiency. *RNA*, 1(7):714–723, Sep 1995.

[37] R S Dassanayake, N V Chandrasekharan, and E H Karunanayake. Trans-spliced leader RNA, 5s-rRNA genes and novel variant orphan spliced-leader of the lymphatic filarial nematode wuchereria bancrofti, and a sensitive polymerase chain reaction based detection assay. *Gene*, 269(1-2):185–193, May 2001.

[38] D L Redmond and D P Knox. Haemonchus contortus SL2 transspliced RNA leader sequence. *Mol Biochem Parasitol*, 117(1): 107–110, Sep 2001.

[39] K Z Lee and R J Sommer. Operon structure and trans-splicing in the nematode Pristionchus pacificus. *Mol Biol Evol*, 20(12): 2097–2103, Dec 2003.

[40] R E Davis, H Singh, C Botka, C Hardwick, M Ashraf el Meanawy, and J Villanueva. RNA trans-splicing in Fasciola hepatica. Identification of a spliced leader (SL) RNA and sl sequences on mRNAs. *J Biol Chem*, 269(31):20026–20030, Aug 1994.

[41] K Brehm, K Jensen, and M Frosch. mRNA trans-splicing in the human parasitic cestode Echinococcus multilocularis. *J Biol Chem*, 275(49):38311–38318, Dec 2000.

[42] R M Zayas, T D Bold, and P A Newmark. Spliced-leader transsplicing in freshwater planarians. *Mol Biol Evol*, 22(10):2048–2054, Oct 2005.

[43] S I Miller, S M Landfear, and D F Wirth. Cloning and characterization of a Leishmania gene encoding a RNA spliced leader sequence. *Nucleic Acids Res*, 14(18):7341–7360, Sep 1986.

[44] M L Muhich, D E Hughes, A M Simpson, and L Simpson. The monogenetic kinetoplastid protozoan, Crithidia fasciculata, contains a transcriptionally active, multicopy mini-exon sequence. *Nucleic Acids Res*, 15(7):3141–3153, Apr 1987.

[45] D A Campbell. Bodo caudatus medRNA and 5s rRNA genes: tandem arrangement and phylogenetic analyses. *Biochem Biophys Res Commun*, 182(3):1053–1058, Feb 1992.

[46] M A Larkin, G Blackshields, N P Brown, R Chenna, P A McGettigan, H McWilliam, F Valentin, I M Wallace, A Wilm, R Lopez, J D Thompson, T J Gibson, and D G Higgins. Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21):2947–2948, Nov 2007.

[47] Wolfgang Otto, Sebastian Will, and Rolf Backofen. Structure local multiple alignment of RNA. In *Proceedings of German Conference on Bioinformatics (GCB'2008)*, volume P-136 of *Lecture Notes in Informatics (LNI)*, pages 178–188. Gesellschaft für Informatik (GI), 2008. ISBN 987-3-88579-230-7.

[48] I L Hofacker. Vienna RNA secondary structure server. *Nucleic Acids Res*, 31(13):3429–3431, Jul 2003.

[49] Daniel H. Huson and D. Bryant. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.*, 23:254–267, 2006.

[50] JR Cook. Adaptations to temperature in two closely related strains of euglena gracilis. *Biological Bulletin*, 1966. URL http://www.biolbull.org/cgi/reprint/ 131/1/83.pdf.

[51] WH Kusber. A study on *Phacus smulkowskianus (Euglenophyceae)* - a rarely reported taxon found in waters of the botanic garden berlin-dahlem. *Wlldenowia*, 28:239–247, 1998.

[52] T L Webb and D Francis. Mating types in Stentor coeruleus. *J Protozool*, 16(4):758–763, Nov 1969.

[53] J Kielhorn and G Rosner. Morpholine. first draft. *INTERNATIONAL PROGRAMME ON CHEMICAL SAFETY*, 1996. URL http://www.inchem.org/documents/ehc/ ehc/ehc179.htm.

[54] J P Bruzik, K Van Doren, D Hirsh, and J A Steitz. Trans splicing involves a novel form of small nuclear ribonucleoprotein particles. *Nature*, 335(6190):559–562, Oct 1988.

[55] D L LEHMANN. Culture forms of *Trypanosoma ranarum* (lankester, 1871). II. effect of temperature upon reproduction and cyclic development. *J Protozool*, 9:325–326, Aug 1962.

[56] ADC Manaia, MDCM de Souza, EDS Lustosa, and Roitman I. Leptomonas lactosovorans n.sp., a lactose-utilizing trypanosomatid: Description and nutritional requirements. *J.Protozool*, 28(1):124–126, 1981.

[57] JBT da Silva and Roitman I. Effect of temperature and osmolarity on growth of crithidia fasciculata, c. hutneri, c. luciliae thermophila, and herpetomonas samuelpessoai. *J.Protozool*, 29 (2):269–272, 1982.

[58] J D Berman and F A Neva. Effect of temperature on multiplication of Leishmania amastigotes within human monocyte-derived macrophages in vitro. *Am J Trop Med Hyg*, 30(2):318–321, Mar 1981.

[59] M Gray, B Wawrik, J Paul, and E Casper. Molecular detection and quantitation of the red tide dinoflagellate Karenia brevis in the marine environment. *Appl Environ Microbiol*, 69(9):5726–5730, Sep 2003.

[60] H D Park, N E Sharpless, and A B Ortmeyer. Growth and differentiation in Hydra. I. the effect of temperature on sexual differentiation in Hydra littoralis. *J Exp Zool*, 160(3):247–254, Dec 1965.

[61] C Ricci, M Caprioli, and D Fontaneto. Stress and fitness in parthenogens: is dormancy a key feature for bdelloid rotifers? *BMC Evol Biol*, 7 Suppl 2, 2007.

[62] LI Lebedeva and TN Gerasimova. Peculiarities of *Philodina roseola* growth and reproduction under various temperature conditions. *Int. Revue ges. Hydrobiol.*, 70(4):509–525, 1985.

[63] W A Van Voorhies and S Ward. Genetic and environmental conditions that increase longevity in Caenorhabditis elegans decrease metabolic rate. *Proc Natl Acad Sci U S A*, 96(20):11399–11403, Sep 1999.

[64] J Martinez, J Perez-Serrano, W E Bernadina, and F Rodriguez-Caabeiro. Oxidative, heat and anthelminthic stress responses in four species of trichinella: comparative study. *J Exp Zool*, 293 (7):664–674, Dec 2002.

[65] A Pires da Silva. Pristionchus pacificus genetic protocols. *WormBook*, pages 1–8, 2005.

[66] L F Lejambre and J H Whitlock. Optimum temperature for egg development of phenotypes in haemonchus contortus cayugensis as determined by arrhenius diagrams and sacher's entropy function. *Int J Parasitol*, 3(3):299–310, May 1973.

[67] N J Lwambo, E S Upatham, M Kruatrachue, and V Viyanant. The host-parasite relationship between the saudi arabian schistosoma mansoni and its intermediate and definitive hosts. 2. effects of temperature, salinity and ph on the infection of mice by s. mansoni cercariae. *Southeast Asian J Trop Med Public Health*, 18(2):166–170, Jun 1987.

[68] J R Claxton, J Sutherst, P Ortiz, and M J Clarkson. The effect of cyclic temperatures on the growth of Fasciola hepatica and Lymnaea viatrix. *Vet J*, 157(2):166–171, Mar 1999.

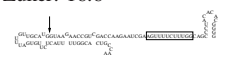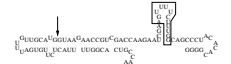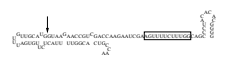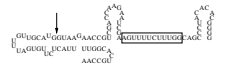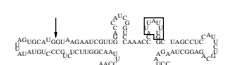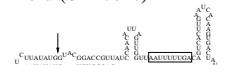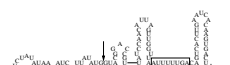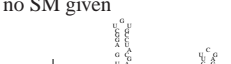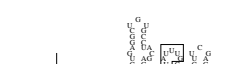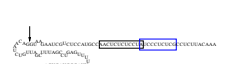[69] D Palakodeti, M Smielewska, and B R Graveley. MicroRNAs from the planarian schmidtea mediterranea: a model system for stem cell biology. *RNA*, 12(9):1640–1649, Sep 2006.

[70] M Novak. Growth of echinococcus multilocularis in gerbils exposed to different environmental temperature. *Experientia*, 39 (4):414–414, Apr 1983.

[71] JK Petersen and HU Riisgard. Filtration capacity of the ascidian *Ciona intestinalis* and its grazing impact in a shallow fjord. *Mar.Ecol.Prog.Ser*, 88:9–17, 1992. URL `http://www. int-res.com/articles/meps/88/m088p009.pdf`.

[72] F Broms and P Tiselius. Effects of temperature and body size on the clearance rate of oikopleura dioica. *J.Plan.Res.*, 25(5): 573–577, 2003.

Table 2: Sequences, secondary structures, and folding energies ΔG(kcal/mol) of known SL RNAs. Donor splice site (arrow) and Sm-binding site (box) are marked. **Left column:** Structures proposed in the literature. **Right column:** Alternative structural models proposed in this work. Abbreviations: UWGCG – University of Wusconsin Genetics Computer Group; * – Recalculated; *T* – natural ambient temperatur of organism; Blue nucleotides (*Euglena*, Rotifera) indicate mutations within known SL RNA alignments; Blue Box (*Hydra*) – alternative SM-binding sites; Green Arrow (*K. brevis*) indicates erroneous splice site from the literature; Sequences, constraints and drawings are available at www.bioinf.unileipzig.de/Publications/SUPPLEMENTS/09-009

Published SL RNA	T	Alternative possible structures at organisms temperature		
E. gracilis, 1991 [4], folded by hand	29C [50]	Additional Constraint: `locarna`-output		
RNAfold (29C): -26.52		-35.02	-32.30	-28.04
E. gracilis, 1999 [22], folded by hand?	29C [50]	Additional Constraint: `locarna`-output		
RNAfold (29C): -23.49		-32.59	-33.69	-28.54
P. curvicauda, 2000 [19], folded by hand?	22C [51]			
RNAfold (22C): -34.62			-41.51	-32.85
C. acus, 2000 [19], folded by hand?	22C			
RNAfold (22C): -48.37		-55.20	-54.10	-42.92
R. costata, 2000 [19], folded by hand?	23C [52]			
RNAfold (23C): -35.72			-42.69	-24.78
M. pellucidum, 2000 [19], folded by hand?	23C			
RNAfold (23C): -26.58			-35.73	-22.94
E. sulcatum, 1999 [22], folded by hand?	25C [53]		16 nt shorter / different SM-binding site	
RNAfold (25C): -33.00		-33.55	-34.10	-33.55
L. collosoma, 1988 [54], UWGCG5.0:-17.5	28C [56]	suboptimal stem I		
RNAfold (28C): -24.00		-27.43	-24.00	-20.04
C. fasciculata, 1988 [54], UWGCG5.0:-12.5	20C [57]	suboptimal stem I		
RNAfold (20C): -30.83		-30.10	-30.83	-31.40
L. enriettii, 1988 [54], UWGCG5.0:-6.4	36C [58]			
RNAfold (36C): -16.26		-20.40	-24.72	-20.87

11

Published SL RNA	T	Alternative possible structures at organisms temperature		
T. cruzi, 1988 [54], UWGCG5.0:-24.1 RNAfold (23C): -39.11	23C [55]	-34.93	-33.17	-42.24
T. vivax, 1988 [54], UWGCG5.0:-26.5 RNAfold (23C): -52.77	23C [55]	-46.90	-53.64	-42.79
T. brucei, 1988 [54], UWGCG5.0:-42.3 RNAfold (23C): -67.58	23C [55]	-60.45	-53.11	alternative SM site -57.98
K. brevis, 2007 [11], mfold (SM cons) RNAfold (20C): -46.29	20C [59]	24 nt 3' removed -58.15	-39.64	-46.43
K. micrum, 2007 [12], mfold3.1.2 (SM cons,20C) RNAfold (20C): -27.68	20C	79 nt 3' added -66.47	-50.85	-63.97
Hydra-A, 2001 [7], mfold2.3 (SM,Donor cons,18C) RNAfold (21C): -14.37	21C [60]	-10.96	-9.26	-18.06
Hydra-B, 2001 [7], mfold2.3 (SM,Donor cons,18C) RNAfold (21C): -32.62	21C [60]	-18.03	-22.17	-20.88

Published SL RNA	T	Alternative possible structures at organisms temperature		
A. ricciae, 2005 [10], mfold RNAfold (24C): -45.31	24C [61]	-36.14	0	-46.28
Philodina sp., 2005 [10], mfold RNAfold (20C): -51.41	20C [62]	-42.28	-42.21	-49.82
C. elegans 1, 1988 [54], UWGCG5.0:-16.0 RNAfold (20C): -38.14	20C [63]	-44.73	-43.54	-33.80
C. elegans 2, 2000 [15], folded by hand? RNAfold (20C): -54.79	20C [63]	-68.58	-57.14	-58.38
T. spiralis 1, 2008 [14], mfold (SM cons) RNAfold (36C): -29.49	36C [64]	-30.15	-27.80	-19.06
T. spiralis 2, 2008 [14], mfold (SM cons) RNAfold (36C): -19.51	36C [64]	-17.78	-11.74	-18.27
P. pacificus 1, 2003 [39] RNAfold (20C): -45.98	20C [65]	-57.87	-50.76	-54.23
P. pacificus 2, 2003 [39] RNAfold (20C): -41.68	20C [65]	-48.85	-48.98	-51.68
H. contortus 2, 2001 [38] RNAfold (25C): -37.63	25C [66]	-47.38	-39.91	-41.15

Published SL RNA	T	Alternative possible structures at organisms temperature		
S. mansoni, 1990 [6], Zuker:-16.6 RNAfold (28C): -24.84	28C [67]	-31.05	-26.65	-23.53
F. hepatica, 1994 [40], Zuker RNAfold (16C): -46.72	16C [68]	-57.02	-47.25	-52.15
S. mediterranea-1, 2005 [42], mfold (SM cons) RNAfold (22C): -43.84	22C [69]	-44.16	-32.07	-41.19
E. multilocularis, 2000 [41], mfold no SM given RNAfold (35C): -36.34	35C [70]	-42.20	-33.35	-37.37
C. intestinalis, 2001 [8], mfold3.0 RNAfold (21C): -5.71	21C [71]	-14.98	-5.71	4nt 5' added -17.06
O. dioica, 2004 [9], folded by hand? RNAfold (20C): -14.19	20C [72]	-14.24		