Note

Algae 2011, 26(3): 229-235
http://dx.doi.org/10.4490/algae.2011.26.3.229
Open Access

Spliced leader sequences detected in EST data of the dinoflagellates *Cochlodinium polykrikoides* and *Prorocentrum minimum*

Ruoyu Guo[1] and Jang-Seu Ki[1,*]

[1]*Department of Green Life Science, Sangmyung University, Seoul 110-743, Korea*

Spliced leader (SL) *trans*-splicing is a mRNA processing mechanism in dinoflagellate nuclear genes. Although studies have identified a short, conserved dinoflagellate SL (dinoSL) sequence (22-nt) in their nuclear-encoded transcripts, whether the majority of nuclear-coded transcripts in dinoflagellates have the dinoSL sequence remains doubtful. In this study, we investigated dinoSL-containing gene transcripts using 454 pyrosequencing data (*Cochlodinium polykrikoides*, 93 K sequence reads, 31 Mb; *Prorocentrum minimum*, 773 K sequence reads, 291 Mb). After making comparisons and performing local BLAST searches, we identified dinoSL for one *C. polykrikoides* gene transcript and eight *P. minimum* gene transcripts. This showed transcripts containing the dinoSL sequence were markedly fewer in number than the total expressed sequence tag (EST) transcripts. In addition, we found no direct evidence to prove that most dinoflagellate nuclear-coded transcripts have this dinoSL sequence.

Key Words: *Cochlodinium polykrikoides*; expressed sequence tag; *Prorocentrum minimum*; spliced leader sequence; *trans*-splicing

INTRODUCTION

The dinoflagellates are an interesting model for eukaryotic evolutionary studies, due to their extraordinary genomic characteristics. Dinoflagellate chromosomes remain permanently condensed during the entire cell life cycle, their nuclear membranes remain intact during mitosis, and they lack nucleosomes and typical histones (Dodge 1966, Hackett et al. 2004, Moreno Díaz de la Espina et al. 2005, Lin et al. 2010). Moreover, dinoflagellates contain modified nuclear DNA; for example, 5-hydroxymethyluraci replaces 12-70% of the nuclear DNA's thymine, while 5-methylcytosine replaces some cytosine (Lin 2011). Dinoflagellates possess a sizable quantity of DNA, ranging from 1.5 to 225 pg per cell (LaJeunesse et al. 2005). In addition, dinoflagellates' gene regulation mechanisms, such as alternative splicing and post-transcriptional regulation, differ substantially from those of typical eukaryotes (Brunelle and Van Dolah 2011, Zhang et al. 2011). In particular, studies have shown spliced leader (SL) *trans*-splicing to be a common mRNA processing mechanism in the dinoflagellate nuclear genes (Lidie and Van Dolah 2007, Zhang and Lin 2008, 2009, Zhang et al. 2009, 2011, Lin et al. 2010), whereas most eukaryotic mRNA editing employs *cis*-splicing in processing. In general, this mRNA processing using SL *trans*-splicing appends a short RNA fragment, such as a SL RNA, to the 5′-untranslated region (UTR) of transcribed pre-mRNA (Pouchkina-Stantcheva and Tunnacliffe 2005, Zhang et al. 2007). Researchers have identified SL *trans*-splicing in other eukaryotic organisms, including trypanosomes, euglena, nematodes, platyhelminthes, rotifers, a tunicate

Received June 21, 2011, **Accepted** August 2, 2011
***Corresponding Author**
E-mail: kijs@smu.ac.kr
Tel: +82-2-2287-5449, Fax: +82-2-2287-0070

(*Ciona intestinalis*), and so on (Murphy et al. 1986, Krause and Hirsh 1987, Tessier et al. 1991, Davis et al. 1994, Pouchkina-Stantcheva and Tunnacliffe 2005).

Recently, Lin and colleagues (Zhang et al. 2007, Zhang and Lin 2009) have studied dinoflagellate SL (dinoSL) *trans*-splicing extensively, and identified a short, conserved dinoSL sequence, 5′-DCC GTA GCC ATT TTG GCT CAA G-3′ (D = U, A, or G), comparing 5′-UTR sequences from cDNA libraries (Zhang et al. 2007). They pointed out that dinoflagellate nuclear encoded transcripts mostly have dinoSL sequences at the 5′-UTR end (Zhang et al. 2007, 2009). With this distinct characteristic, the authors can isolate dinoflagellate genes from environmental samples by using the dinoSL sequence as a marker, or dinoflagellate-specific primer, on the SL (Zhang and Lin 2008). However, Bachvaroff and Place (2008), after determining 47 genes of dinoflagellate *Amphidinium carterae*, examined those having cDNAs for dinoSL sequences and detected this dinoSL in only about two-thirds of all examined transcripts (i.e., approximately one-third failed to show *trans*-splicing). Following this study, Zhang and Lin (2009) re-investigated the genes lacking dinoSL, which Bachvaroff and Place (2008) had suggested might be "not *trans*-spliced," successfully detected the dinoSL at the 5′-ends of their transcripts, and reinstated the postulate that dinoSL is widespread among dinoflagellate nuclear-encoded transcripts. Taking these previous findings into consideration, the presence of dinoSL in the majority of dinoflagellate transcripts remains controversial. To determine the expressed sequence tag (EST) sequencings of other dinoflagellates and other strains from different geographical regions requires further studies.

In the present study, we investigated the dinoSL sequence using our EST databases that comprised a naked dinoflagellate, *Cochlodinium polykrikoides*, and an armored dinoflagellate, *Prorocentrum minimum*. To determine these EST sequences, we employed 454 sequencing (a pyrosequencing system of 454 Life Sciences, Roche, Branford, CT, USA). Researchers consider *C. polykrikoides* and *P. minimum* to be harmful algal bloom species. In particular, *P. minimum* can produce the potent diarrhetic shellfish poisoning, which is one of the major types of illness that result from harmful algal blooms.

MATERIALS AND METHODS

Cochlodinium polykrikoides and *Prorocentrum minimum* cultures

We obtained the two dinoflagellate strains, *C. polykrikoides* (CP-1) and *P. minimum* (D-127), from the National Fisheries Research and Development Institute (NFRDI) and the Korea Marine Microalgae Culture Center (KMMCC), respectively. Cultures of both strains were grown in f/2 medium, at 20°C, following a 12 : 12 h light : dark cycle. We harvested the cells at various growth phases, using exponential growth phase cultures for various stress treatments, including heat shock, cold shock, exposure to metals, and UV. The *Cochlodinium* and *Prorocentrum* cells were then harvested via centrifugation at 3,000 rpm for 10 min. We immediately diluted all harvested cells with ten volumes of TRIzol (Invitrogen, Carlsbad, CA, USA), froze them in liquid nitrogen, and stored them at -80°C until we extracted their RNA.

Total RNA extraction

To isolate the total RNA from these harvested cells, we used the TRIzol method (Invitrogen), according to the manufacturer's instructions. After physically breaking the cells via freeze-thawing in liquid nitrogen, we homogenized them using zirconium beads (diameter 0.1 mm) and a Mini-Beadbeater (BioSpec Products Inc., Bartlesville, OK, USA). We measured each RNA sample's concentration and purity using a DU730 life science UV-Vis spectrophotometer (Beckman Coulter, Fullerton, CA, USA) and verified the RNA's integrity via electrophoresis on agarose gels.

EST sequencing and annotations

First, we pooled a variety of total RNAs from various conditions (e.g., heat shock, cold shock, toxic chemical exposure, and different life stages) into a single tube, which we then subjected to EST sequencing via a GS-FLX Titanium instrument (454 Life Sciences, Roche), assembling the EST sequences having 95% similarity levels with one another. Next, we separately characterized contigs and singletons of each EST data set by means of BLAST-X comparisons, using public domain databases. This process allowed us to treat EST sequences with E-values over 1.0E-05 as "No Hit," as they probably belonged to UTRs.

DinoSL sequence searches

Finally, we investigated EST sequences having more than 100 bp of 5′-UTR for the SL sequence. In addition, we constructed local nucleotide databases of our *Cochlodinium* and *Prorocentrum* EST data, using BioEdit

Table 1. Summary of EST data constructed from *Cochlodinium polykrikoides* and *Prorocentrum minimum*

Species	Reads	Sequences	Contig		Singleton	
			Annotated EST	No hit	Annotated EST	No hit
C. polykrikoides	93,124	31,406,675	1,909	1,264	6,647	14,874
P. minimum	773,409	291,217,368	9,147	11,973	25,404	100,136

Contigs and singletons were annotated by BLAST-X searches.
EST, expressed sequence tag.

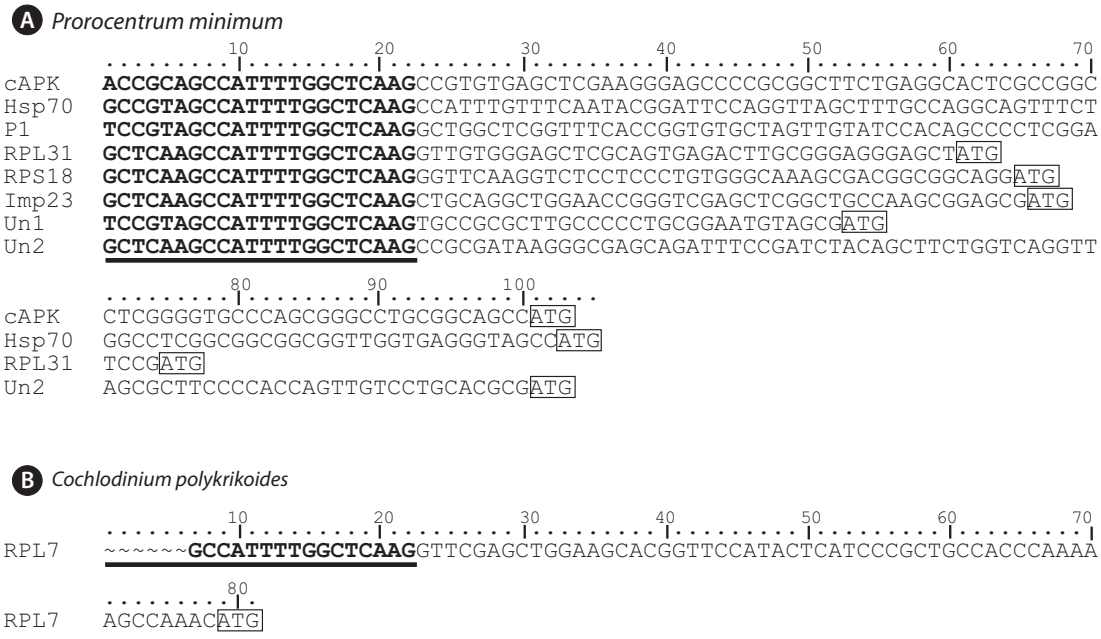

Fig. 1. Spliced leader and adjacent sequences detected from (A) *Prorocentrum minimum* and (B) *Cochlodinium polykrikoides* expressed sequence tags (ESTs). Nucleotides in boxes and under lines represent the start codons and dinoflagellate spliced leader sequences, respectively. cAPK, cAMP-dependent protein kinase; Hsp 70, heat shock protein 70; P1, acidic ribosomal protein P1; RPL31, 60S ribosomal protein L31; RPS18, 40S ribosomal protein S18; Imp23, imm downregulated protein 23; Un, unknown protein; RPL7, 60S ribosomal protein L7.

version 5.0.6 (Hall 1999), and used them for local BLAST searches. To analyze nuclear encoded transcripts (or EST sequences) that had the dinoSL sequence, we used Genetyx version 7.0 software (Genetyx Corp., Tokyo, Japan).

RESULTS AND DISCUSSION

In the present study, we determined the large-scale EST sequences of two dinoflagellates, *C. polykrikoides* (93 K sequence reads, 31 Mb) and *P. minimum* (773 K sequence reads, 291 Mb). From our GS-FLX sequencing, we identified 3,173 contigs and 21,521 singletons in *Cochlodinium* cDNA and 21,120 contigs and 125,540 singletons in *Prorocentrum* cDNA (Table 1). BLAST-X searches showed many cDNA sequences contained 5′-UTR sequences. For example, we identified *P. minimum* ESTs more than 100 bp of 5′-UTR in sequence length at 1,491 contigs and 414 singletons. Upon comparing these 5′-UTR sequences of *P. minimum* ESTs and a conserved dinoSL sequence (5′-DCC GTA GCC ATT TTG GCT CAA G-3′), we identified eight dinoSL sequences belonging to ribosomal protein, 40S ribosomal protein, 60S ribosomal protein, cAMP-dependent protein kinase regulatory subunits, and acidic ribosomal protein (Table 2, Fig. 1). On the other hand, we only detected one dinoSL sequence belonging to the 60S ribosomal protein in the *C. polykrikoides* ESTs.

In addition, we used BLAST searches to investigate dinoSL-containing transcripts in our local nucleotide database. Through this analysis, we detected 55 dinoSL sequence-containing ESTs (17 contigs, 38 singletons)

from the *P. minimum* EST data. Using BLAST searches, we analyzed all sequences in the GenBank database, listing the closest matched genes in Table 2. Of these, we could annotate 3 out of 38 singleton-ESTs (8%) and 9 out of 17 contig-ESTs (53%). The identified genes included ribosomal protein, acidic ribosomal protein, cAMP-dependent protein kinase, conserved hypothetical protein, heat shock protein 70, imm downregulate 23 protein, and unknown proteins in *P. minimum*, and 60S ribosomal protein in *C. polykrikoides*. Our data showed that transcript numbers containing dinoSL sequence were lower than the total reads of EST data. In particular, we detected only one dinoSL sequence from *C. polykrikoides* ESTs. These results resembled those in the study by Bachvaroff and Place (2008), in which they detected the dinoSL sequence in *Amphidinium carterae* EST data, but it was not universal. These findings are incompatible with those of Zhang and Lin (2009), which showed that the dinoSL sequence in the 5′-UTR has a wide distribution among dinoflagellate nuclear-encoded transcripts. With the present and previous data, we could not conclude that dinoflagellates' nuclear-gene transcripts most commonly include the dinoSL sequence, because we identified relatively few gene transcripts containing the dinoSL sequence from large-scale ESTs of *C. polykrikoides* and *P. minimum*.

To our knowledge, the dinoSL sequence is added to the 5′-end of dinoflagellate gene transcripts. For investigating whether all or parts of dinoflagellate nuclear gene transcripts contain dinoSL sequence, researchers should retain intact 5′-ends of the genes. In addition, to detect the dinoSL-containing transcripts, studies should amplify transcripts entirely by means of reverse transcriptase. However, many dinoflagellates contain inhibitors that will strongly inhibit either reverse transcriptase or Taq DNA polymerase activity (Zhang and Lin 2009). Problems such as these may explain why few nuclear gene transcripts contain the dinoSL sequence, in both the previous data (Zhang et al. 2007, Bachvaroff and Place 2008) and in our EST data. On the other hand, Bachvaroff and Place (2008) showed that the SL *trans*-splicing of dinoflagellates nuclear genes correlates with expression level, suggesting that the high-expression-level gene is more likely to be SL *trans*-spliced. By surveying the dinoflagellate gene transcripts that contain the dinoSL sequence, researchers have identified numerous genes involved in the various cell functions (Table 3). Interestingly, all of the annotated genes in Table 3 play important roles in cells' biological processes and have high expression levels within these cells. In view of the summarized data, we consider that finding genes containing the dinoSL sequence might be much easier using high-expression-level genes rather than using low-expression genes. For example, studies have found ribosomal protein genes containing the dinoSL sequence in almost all dinoflagellates (except *Noctiluca scintillans*). Reportedly, proliferating cell nuclear antigen (*PCNA*) contains the dinoSL sequence throughout the phylum (Zhang et al. 2007). Both ribosomal protein and *PCNA* are expressed throughout the cell cycle, and at high expression levels, as well.

This study investigated the dinoSL sequence location by surveying reported dinoSL-containing gene transcripts and our EST data (Table 3). Having detected the dinoSL sequence in *C. polykrikoides* and *P. minimum* nuclear gene transcripts (Fig. 1), we found the dinoSL sequence location ranged from 52 to102 bp upstream of

Table 2. *Cochlodinium* and *Prorocentrum* ESTs containing dinoSL RNA sequences at the 5′-end and their closest matches from GenBank data

Species	Gene	Code	Closest match in the GenBank		
			Accession No.	Species	E-value
C. polykrikoides	60S ribosomal protein L7	RPL7	XP_002775844	*Perkinsus marinus*	1.00E-15
P. minimum	40S ribosomal protein S18	RPS18	XP_002141581	*Cryptosporidium muris*	3.00E-23
	60S ribosomal protein L31	RPL31	XP_002368706	*Toxoplasma gondii*	2.00E-05
	Acidic ribosomal protein P1	P1	XP_001699629	*Chlamydomonas reinhardtii*	3.00E-23
	cAMP-dependent protein kinase	cAPK	XP_002764691	*Perkinsus marinus*	6.00E-06
	Conserved hypothetical protein	Un2	ACU44930	*Prorocentrum minimum*	9.00E-62
	Heat shock protein 70	Hsp70	ABI14407	*Prorocentrum minimum*	1.00E-43
	Imm downregulated protein 23	Imp23	EFA78150	*Polysphondylium pallidum*	5.00E-10
	Unknown protein	Un1	ABX80193	*Prorocentrum minimum*	8.00E-13

Here, we used our *Cochlodinium* and *Prorocentrum* EST data for dinoSL detection.
EST, expressed sequence tag; DinoSL, dinoflagellate spliced leader.

Table 3. Genes, GenBank accession numbers, and locations of the dinoSL sequence upstream of ATG, summarized from available dinoflagellates' *trans*-spliced genes

Species	Gene	Accession No.	Location No.	Reference
Alexandrium fundyense	Actin	EF133869	99	Zhang et al. 2007
	40S ribosomal protein S8	EF133862	103	
	Alpha tubulin	EU742865	84	Bachvaroff and Place 2008
	Basic nuclear protein	EU742858	86	
	Calmodulin	EF133873	141	Zhang et al. 2007
	DNA damage checkpoint protein	EF133874-EF133878	89-104	
	GAPDH	EU742866	90	Bachvaroff and Place 2008
	Glutamate semialdehyde synthase	EU742797	37	
	Heat shock protein 70 / 90	EU742860 / EU742821	81/76	
	PCNA	EF133957	77	Zhang et al. 2007
	Protein disulfide-isomerase	EF133860-EF133861	75-94	
	Small nuclear ribonucleoprotein E	EU742814	120	Bachvaroff and Place 2008
	Unknown protein	EF133894	225	Zhang et al. 2007
Amphidinium carterae	60S ribosomal protein L26	EF133964-EF133967	90	Zhang et al. 2007
	Solute carrier family 35	EF133962-EF133963	195	
	Unknown protein	EF133998-EF134002	315	
Cochlodinium polykrikoides	60S ribosomal protein L7	JN560151	69	This study
Karenia brevis	Beta-tubulin	EU078557	106	Lidie and Van Dolah 2007
	Glyceraldehyde-3-phosphate dehydrogenase	EU078558	69	
	PCNA	EU078559	91	
Karlodinium micrum	Heat shock protein 90	DQ884433	103	Zhang et al. 2007
	60S acidic ribosomal protein P2	EF134133	81	
	Isomerase	DQ884436-DQ884439	96-101	
	Tubulin alpha chain	EF134085	141	
	40S ribosomal protein S6 / S10	DQ884434 / EF134134	80	
	Beta-tubulin	EF134086	93	
	PCNA	EF134029	83	
	Major basic nuclear protein	EF134102- EF134104	81-87	
	Thioredoxin-like	EF134124	302	
Noctiluca scintillans	Actin	EF134225	70	Zhang et al. 2007
	RNA-binding protein	EF134232	79	
	Ubiquitin-conjugating enzyme E2-like	EF134212-EF134215	70	
	Unknown protein	EF134243-EF134246	179	
	Unknown protein	EF134250-EF134252	279	

Table 3. Continued

Species	Gene	Accession No.	Location No.	Reference
Oxyrrhis marina	Rhodopsin	EF134312-EF134318	44-131	Zhang et al. 2007
	40S ribosomal protein S16	EF134326-EF134327	61-71	
	60S ribosomal protein L31 / L26	EF134323 / EF134325	74/68	
	Ribosomal protein L27a	EF134332-EF134333	50-58	
Pfiesteria piscicida	Unknown protein	DQ864904 / DQ864911	208/171	Zhang et al. 2007
	40S ribosomal protein S24	DQ864830	64	
	60S ribosomal protein I21 / L32	DQ864827 / DQ864828	80/99	
	ADP-ribosylation factor	DQ864770-DQ864785	68-85	
	Calmodulin	DQ864822-Q864826	73-80	
	Cytochrome c	DQ864762-DQ864769	66-76	
	PCNA	DQ239831	78	
	Major basic nuclear protein	DQ864798-DQ864819	70-80	
	Serine / Threonine protein kinases	DQ864872 / DQ864873	65/119	
	Unknown protein	DQ864915-DQ864927	58-93	Zhang et al. 2007
Prorocentrum minimum	Beta-tubulin	EF134371	74	
	Centrin	DQ884415-DQ884418	107-116	
	Calmodulin	DQ884428-DQ884431	94-97	
	40S ribosomal S14	DQ884419	94	
	Heat shock protein 70	DQ884421	86	
	Major basic nuclear protein	DQ884423-Q884427	88-101	
	PCNA	EF134017	93	
	Unknown protein	EF134374-EF134387	48-137	
	Lipopolysaccharide-induced NF factor-like	EF134372	101	
	Uridine kinase	EF134373	78	
	Form II Rubisco	DQ884420	104	
	Nitrite transporter	DQ884413	120	
	Single domain cyclophilin Type peptidyl-prolyl *cis-trans*	DQ884414	70	
	40S ribosomal protein S18	JN560149	64	This study
	60S ribosomal protein L31	JN560144	60	
	Acidic ribosomal protein P1	JN560148	74	
	cAMP-dependent protein kinase	JN560147	100	
	Conserved hypothetical protein	JN560143	100	
	Heat shock protein 70	JN560150	102	
	Imm downregulated protein 23	JN560146	65	
	Unknown protein	JN560145	52	

DinoSL, dinoflagellate spliced leader; Location No., location of dinoSL sequence upstream of the start codon (ATG); *GAPDH*, glyceraldehyde 3-phosphate dehydrogenase; *PCNA*, proliferating cell nuclear antigen.

the start codon (ATG). Moreover, additional summarized data (Table 3) showed that the dinoSL sequence's major locations ranged from 40 to 160 bp upstream of ATG. Only in a few genes, and particularly in unknown function genes, did the dinoSL sequence occur more than 170 bp upstream of ATG. Perhaps the SL *trans*-splicing process mostly tends to append the dinoSL sequence to the short nucleotides (< 170 bp) upstream of the start codon.

ACKNOWLEDGEMENTS

This work was supported by both the Marine and Extreme Genome Research Center Program of the Ministry of Land, Transportation and Maritime Affairs, Republic of Korea, and by a National Research Foundation of Korea (NRF) grant, funded by the Korea government (MEST; No. 2010-0009669).

REFERENCES

Bachvaroff, T. R. & Place, A. R. 2008. From stop to start: tandem gene arrangement, copy number and *trans*-splicing sites in the dinoflagellate *Amphidinium carterae*. PloS One 3:e2929.

Brunelle, S. A. & Van Dolah, F. M. 2011. Post-transcriptional regulation of S-phase genes in the dinoflagellate, *Karenia brevis*. J. Eukaryot. Microbiol. 58:373-382.

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

Dodge, J. D. 1966. The dinophyceae. *In* Godward, M. B. E. (Ed.) *The Chromosomes of the Algae*. St. Martin's Press, New York, pp. 96-115.

Hackett, J. D., Anderson, D. M., Erdner, D. L. & Bhattacharya, D. 2004. Dinoflagellates: a remarkable evolutionary experiment. Am. J. Bot. 91:1523-1534.

Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41:95-98.

Krause, M. & Hirsh, D. 1987. A *trans*-spliced leader sequence on actin mRNA in *C. elegans*. Cell 49:753-761.

LaJeunesse, T. C., Lambert, G., Andersen, R. A., Coffroth, M. A. & Galbraith, D. W. 2005. *Symbiodinium* (Pyrrhophyta) genome sizes (DNA content) are smallest among dinoflagellates. J. Phycol. 41:880-886.

Lidie, K. B. & Van Dolah, F. M. 2007. Spliced leader RNA-mediated *trans*-splicing in a dinoflagellate, *Karenia brevis*. J. Eukaryot. Microbiol. 54:427-435.

Lin, S. 2011. Genomic understanding of dinoflagellates. Res. Microbiol. 162:551-569.

Lin, S., Zhang, H., Zhuang, Y., Tran, B. & Gill, J. 2010. Spliced leader-based metatranscriptomic analyses lead to recognition of hidden genomic features in dinoflagellates. Proc. Natl. Acad. Sci. U. S. A. 107:20033-20038.

Moreno Díaz de la Espina, S., Alverca, E., Cuadrado, A. & Franca, S. 2005. Organization of the genome and gene expression in a nuclear environment lacking histones and nucleosomes: the amazing dinoflagellates. Eur. J. Cell. Biol. 84:137-149.

Murphy, W. J., Watkins, K. P. & Agabian, N. 1986. Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for *trans*-splicing. Cell 47:517-525.

Pouchkina-Stantcheva, N. N. & Tunnacliffe, A. 2005. Spliced leader RNA-mediated *trans*-splicing in phylum Rotifera. Mol. Biol. Evol. 22:1482-1489.

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

Zhang, H., Campbell, D. A., Sturm, N. R. & Lin, S. 2009. Dinoflagellate spliced leader RNA genes display a variety of sequences and genomic arrangements. Mol. Biol. Evol. 26:1757-1771.

Zhang, H., Dungan, C. F. & Lin, S. 2011. Introns, alternative splicing, spliced leader *trans*-splicing and differential expression of *pcna* and *cyclin* in *Perkinsus marinus*. Protist 162:154-167.

Zhang, H., Hou, Y., Miranda, L., Campbell, D. A., Sturm, N. R., Gaasterland, T. & Lin, S. 2007. Spliced leader RNA *trans*-splicing in dinoflagellates. Proc. Natl. Acad. Sci. U. S. A. 104:4618-4623.

Zhang, H. & Lin, S. 2008. mRNA editing and spliced-leader RNA *trans*-splicing groups *Oxyrrhis*, *Noctiluca*, *Heterocapsa*, and *Amphidinium* as basal lineages of dinoflagellates. J. Phycol. 44:703-711.

Zhang, H. & Lin, S. 2009. Retrieval of missing spliced leader in dinoflagellates. PLoS One 4:e4129.