Operons Are a Conserved Feature of Nematode Genomes

Jonathan Pettitt,[1] Lucas Philippe, Debjani Sarkar, Christopher Johnston, Henrike Johanna Gothe, Diane Massie, Bernadette Connolly, and Berndt Müller

School of Medical Sciences, University of Aberdeen, Institute of Medical Sciences, Foresterhill, Aberdeen AB25 2ZD, United Kingdom

ORCID IDs: 0000-0001-7988-6833 (J.P.); 0000-0002-0672-4371 (C.J.); 0000-0003-1279-5421 (B.M.)

ABSTRACT The organization of genes into operons, clusters of genes that are co-transcribed to produce polycistronic pre-mRNAs, is a trait found in a wide range of eukaryotic groups, including multiple animal phyla. Operons are present in the class Chromadorea, one of the two main nematode classes, but their distribution in the other class, the Enoplea, is not known. We have surveyed the genomes of *Trichinella spiralis*, *Trichuris muris*, and *Romanomermis culicivorax* and identified the first putative operons in members of the Enoplea. Consistent with the mechanism of polycistronic RNA resolution in other nematodes, the mRNAs produced by genes downstream of the first gene in the *T. spiralis* and *T. muris* operons are *trans*-spliced to spliced leader RNAs, and we are able to detect polycistronic RNAs derived from these operons. Importantly, a putative intercistronic region from one of these potential enoplean operons confers polycistronic processing activity when expressed as part of a chimeric operon in *Caenorhabditis elegans*. We find that *T. spiralis* genes located in operons have an increased likelihood of having operonic *C. elegans* homologs. However, operon structure in terms of synteny and gene content is not tightly conserved between the two taxa, consistent with models of operon evolution. We have nevertheless identified putative operons conserved between Enoplea and Chromadorea. Our data suggest that operons and "spliced leader" (SL) *trans*-splicing predate the radiation of the nematode phylum, an inference which is supported by the phylogenetic profile of proteins known to be involved in nematode SL *trans*-splicing.

THE organization of open-reading frames into operons, such that multiple, distinct gene products are produced from a single, polycistronic transcript, is commonplace in prokaryote genomes (Jacob *et al.* 1960). Operons are also found in eukaryotes, although their distribution is sporadic and it does not seem likely that they represent an ancestral eukaryotic trait (Lawrence 1999; Hastings 2005). In prokaryotes, translation of multiple open reading frames in a polycistronic RNA occurs through multiple independent translation initiations. In eukaryotes, the polycistronic RNAs must first be processed into individual mRNAs before being translated. This creates a problem in that the processed, downstream mRNAs would lack a cap structure necessary for RNA stability and translation. A number of eukaryotes are able to circumvent this problem through the *trans*-splicing of a short "spliced leader" (SL) RNA onto the 5′ end of the mRNA. Because the precursor SL RNAs that donate the SL are trimethylguanosine-capped, the *trans*-splicing event provides the cap structure for the mRNA. Thus by providing a mechanism that allows the formation of monocistronic, capped mRNAs from polycistronic RNA, SL *trans*-splicing enables the organization of eukaryotic genes into operons. It is striking that, at least to date, all eukaryotes in which operon usage is widespread also undergo SL *trans*-splicing (Johnson *et al.* 1987; Spieth *et al.* 1993; Davis and Hodgson 1997; Blumenthal *et al.* 2002; Ganot *et al.* 2004; Guiliano and Blaxter 2006; Satou *et al.* 2008; Marlétaz *et al.* 2008; Dana *et al.* 2012; Protasio *et al.* 2012; Tsai *et al.* 2013), suggesting that the resolution of polycistronic RNA is dependent upon SL *trans*-splicing.

Although operon organization is widespread in numerous eukaryotic taxa, the evolutionary mechanisms that have resulted in this form of gene organization are not well

doi: 10.1534/genetics.114.162875
Manuscript received February 12, 2014; accepted for publication June 6, 2014; published Early Online June 13, 2014.
Available freely online through the author-supported open access option.
Supporting information is available online at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.162875/-/DC1.
The *T. spiralis* and *T. muris* mRNA 5′ ends with/without SL sequences are deposited in GenBank (accession nos. KF442418–KF442435, KF511776, KF511777, and KF768019).
[1]Corresponding author: School of Medical Sciences, Institute of Medical Sciences, University of Aberdeen, Aberdeen AB25 2ZD, UK. E-mail: j.pettitt@abdn.ac.uk

understood. The most detailed analysis of the problem has come from studies in *Caenorhabditis elegans*, which led to the hypothesis that operon organization allows the marshalling of multiple genes under the control of a single promoter. This makes cells better able to cope with situations when transcription factors are present in limiting concentrations, such as recovery from growth arrest (Zaslaver *et al.* 2011). However, it is far from clear whether this is the only mechanism responsible for the evolution of operon organization, and the general applicability of this hypothesis to other members of the nematode phylum is not known.

To better understand the relationship between operon evolution and SL *trans*-splicing, it is necessary to determine the distribution of operon organization across the nematode phylum. Nematodes can be divided into two major classes: Enoplea and Chromadorea (Holterman *et al.* 2006; Meldal *et al.* 2007), with the latter class being much better characterized in terms of gene expression mechanisms, largely because it contains *C. elegans*. Both SL *trans*-splicing and operons have been identified in multiple nematodes within the Chromadorea (Evans *et al.* 1997; Lee and Sommer 2003; Guiliano and Blaxter 2006). However, the presence of operons has not been reported in nematodes from the other taxon.

We have previously identified SL *trans*-splicing in the enoplean nematodes *Trichinella spiralis* and *Prionchulus punctatus* (Pettitt *et al.* 2008; Harrison *et al.* 2010), suggesting that they may also possess operons. The draft genome of *T. spiralis* should be a useful resource for identifying operons in this nematode (Mitreva *et al.* 2011). However, identification of operons is not straightforward. The original discovery of operons in *C. elegans* was dependent upon the discovery of a specific spliced leader, SL2, which is *trans*-spliced to most mRNAs derived from genes downstream of the first gene in operons (Spieth *et al.* 1993), but not all nematodes use a specialized SL RNA to resolve polycistronic RNAs (Guiliano and Blaxter 2006). Thus, this feature cannot be considered diagnostic for mRNAs derived from nematode operons. The other feature common to operonic genes is that, at least in *C. elegans*, the distance between genes in an operon (the intercistronic region, ICR) is unusually short, with a mean ICR size of 126 bp (Blumenthal *et al.* 2002). Again, this trait is not definitive: the ICR size can be considerably larger in the operons of other nematodes (Guiliano and Blaxter 2006; Ghedin *et al.* 2007) and even in *C. elegans* operons exist with large ICR distances (Morton and Blumenthal 2011).

Previous approaches to identify operons in *T. spiralis* (Mitreva *et al.* 2011) looked for pairs of *T. spiralis* genes whose homologs were in the same operon in *C. elegans*. This resulted in a limited set of 16 neighboring pairs of genes that potentially correspond to *T. spiralis* operons; however, further characterization of these candidate operons was not undertaken. We have used conserved synteny, coupled with the fact that mRNAs derived from downstream genes in operons are dependent on SL *trans*-splicing to elucidate a set of putative *T. spiralis* operons. Detailed analysis of two of these putative operons indicates that they display all the molecular characteristics expected of loci that generate polycistronic RNA. Taken together our data indicate that the organization of genes into operons was present in the last common ancestor of the Chromadorea and Enoplea.

Materials and Methods

Bioinformatic identification of T. muris SL RNA genes

Trichuris muris SL RNA genes *Tmu*-SL1, *Tmu*-SL2, *Tmu*-SL3, and *Tmu*-SL9 were identified by searching the *T. muris* genome dataset with *T. spiralis* SL sequences using the BLASTN tool. Hits were considered if two of the three following criteria were met: a candidate Sm protein binding site was detected (AATTTTTG), the 5′ splice site sequence was conserved (AGGT), and a run of at least three Ts was found located ~100 bp from the end of the putative SL sequence. *T. muris* SL RNA genes *Tmu*-SL4, *Tmu*-SL5, *Tmu*-SL6, *Tmu*-SL7, *Tmu*-SL8, and *Tmu*-SL10 were identified by searching the *T. muris* genome dataset with the *Tmu*-SL1, *Tmu*-SL2, *Tmu*-SL3, or *Tmu*-SL9 sequences using the BLASTN tool and fulfilling the same criteria as above. Genes for *Tmu*-SL1, *Tmu*-SL2, *Tmu*-SL6, *Tmu*-SL8, *Tmu*-SL9, *Tmu*-SL10, and *Tmu*-SL11 were also identified with a PERL script (Pettitt *et al.* 2008) used previously to identify *T. spiralis* SL genes, except that the parameters for the Sm binding site were changed to AATTTTTG/TG.

Analysis of T. spiralis SL containing ESTs and Identification of Putative Conserved Operons

The ESTs were identified earlier (Pettitt *et al.* 2008). To identify the corresponding gene from which each EST was derived, the EST sequences were mapped onto the *T. spiralis* draft genome sequencing using BLASTN. The corresponding gene was annotated as being in a putative operon if its upstream or downstream neighbor genes were on the same DNA strand with an intercistronic distance of ≤1 kb. If the neighbors were on the same strand, but between 1 kb and 5 kb away, they were recorded as ambiguous. Otherwise the genes were annotated as nonoperonic. A minority of the ESTs matched to more than one predicted *T. spiralis* gene.

To identify the operonic status of the *C. elegans* homologs of each SL *trans*-spliced *T. spiralis* EST, BLASTX searches were carried out. We used an *E*-value cutoff of 10^{-5} to determine homology. In addition, if we obtained similar *E*-values for multiple *C. elegans* genes, we excluded that EST from the analysis.

T. spiralis and *Romanomermis culicivorax* homologs of operonic gene pairs conserved between *C. elegans* and *Brugia malayi* (Ghedin *et al.* 2007) were identified using BLASTP searches with the *C. elegans* upstream homologs from each pair as a query in searches against the respective gene predictions. The predicted coding region of the *T. spiralis/R. culicivorax* gene immediately downstream of the gene identified by this search was then used as a query sequence in a "reciprocal" BLASTP search against the *C. elegans* gene predictions. Since it was apparent that *T. spiralis* genes, which are separated by unusually short intergenic distance (such as might be expected in genes organized into operons), are prone to misannotation

and conflation into a single gene prediction, we also carried out an additional step using both *C. elegans* gene pairs as query sequences in BLASTP searches against the *T. spiralis* gene predictions, looking for cases where both *C. elegans* genes return matches to the same *T. spiralis* gene. Manual examination of such putative gene prediction errors, guided by the results of the sequence similarity searches, was then used to identify the intercistronic regions in each case. In all BLASTP searches, we used an *E*-value cutoff of 10^{-5} for the establishment of homology.

Phylogenetic profiling of SL trans-splicing snRNPs

Homologs of *C. elegans* *sna-1*, *sut-1*, and *sna-2* were identified by carrying out BLAST searches against the National Center for Biotechnology Information (NCBI) nonredundant database, except in the case of *R. culicivorax* and *T. muris*, where BLAST searches were carried out against datasets downloaded from http://www.nematodes.org/genomes/romanomermis_culicivorax/ and http://www.sanger.ac.uk/resources/downloads/helminths/trichuris-muris.html, respectively. Phylogenetic tree construction was carried out with the online implementation of PhyML (Dereeper *et al.* 2008) using default settings.

Nematode isolation and RNA preparation

T. spiralis RNA was produced as described (Pettitt *et al.* 2008). *T. muris* RNA was a generous gift from Allison Bancroft and Richard Grencis (University of Manchester).

Analysis of RNA 5′ ends

The 5′ ends of cDNAs were obtained through 5′ RACE using the GeneRacer kit (Invitrogen), according to the manufacturer's instructions. Gene-specific primers used are given in Supporting Information, Table S3, and the cDNAs, amplified by PCR, using either GoTaq polymerase (Promega) or Expand High Fidelity polymerase (Roche), were cloned into pGEM T-Easy (Promega). The resulting plasmid inserts were sequenced by the University of Dundee Sequencing Service.

Detection of processing intermediates of polycistronic transcripts

RNA was reverse transcribed using SuperScript III Reverse Transcriptase (Invitrogen) and random primers according to the instructions of the manufacturer. In control reactions ("−RT") all reagents were included except the reverse transcriptase. Processing intermediates were normally amplified by two rounds of PCR with nested primer pairs (Table S4) and either GoTaq polymerase (Promega) or Expand High Fidelity polymerase (Roche) and visualized by agarose gel electrophoresis. The identity of the PCR products was determined by cloning into pGEM-T Easy (Promega) and sequencing of plasmid inserts.

Identification of SL RNA 3′ ends

SL RNA 3′ ends were determined essentially as described previously (Pettitt *et al.* 2008). *T. muris* total RNA (~5 μg) was poly(A) tailed using yeast poly(A) polymerase, reverse transcribed using an oligo-dT-anchor primer (GCGAGCTCCGCGG CCGCGTTTTTTTTTTTTTTT) and then PCR amplified using an SL-specific primers (GGTTAATTACCCAATTTAAAAG) and an anchor primer (GCGAGCTCCGCGGCCGCG). PCR fragments were inserted into pGEM-T Easy (Promega), and inserts were sequenced at the University of Dundee DNA Sequencing Facility.

SL RNA secondary structure prediction: Secondary structure prediction of *T. muris* SL RNA was performed using MFOLD Version 2.3 (Zuker 2003) using the default folding conditions (1 M NaCl, 37°) and with the constraint that the Sm-binding site (5′-AAUUUUUUG-3′) was required to be single stranded.

Generation and analysis of synthetic operon constructs

The GFP coding region was amplified from pTG96 using the primers 5′-CAATACAGACTTCCCGGGATTGGCCAAAGGACC CAAA-3′ and 5′-GCTCACCATGCTAGCCTATTTGTATAGTTC ATCCATGC-3′. The mCherry coding region, coupled to the *unc-54* 3′-UTR, was amplified from pPD95.75Cherry (a derivative of pPD95.75 in which the GFP coding region was replaced by mCherry) using the primers 5′-ACAAATAGGCT AGCATGGTGAGCAAGGGCGAG-3′ and 5′-CGCGCGAGACG AAAGGGCCCAGGAAACAGTTATGTTTGGTAT-3′. The primers were designed so that they had overlapping complementary 5′ extensions that introduced an *Nhe*I restriction site. The two amplicons were purified and fused using a PCR fusion strategy (Hobert 2002). The resulting amplicon consisting of the GFP and mCherry coding regions flanking an *Nhe*I site was cloned into *Sma*I–*Apa*I cut pTG96 using In-Fusion HD (Clontech Laboratories) to generate pTG96-Op. The ICRs were cloned from PCR products amplified from genomic DNA. The Tsp-*cpt-2*∼*nuaf-3* ICR was amplified from *T. spiralis* genomic DNA using primers 5′-ATACAAATAGGCTAG CACGAATTATCACTTTTATAAC-3′ and 5′-TGCTCACCATGC TAGCTTACGCCAAACTAGGAAATTATTGA-3′, and the *Cel-cpt-2*∼*prx-14* ICR was amplified from *C. elegans* genomic DNA using primers 5′-ATACAAATAGGCTAGCTTGTTTGAT GACATTTATGTATTTAT-3′ and 5′-TGCTCACCATGCTAGCTTT CAACCTGAAGCTTTAAAAT-3′. The resulting PCR products were cloned into *Nhe*I cut pTG96-Op using In-Fusion HD (Clontech Laboratories) and the resultant plasmids, pPE#LP1 (*Tsp-cpt-2*∼*nuaf-3* ICR clone) and pPE#LP2 (*Cel-cpt-2*∼*prx-14* ICR clone) were sequenced to confirm the integrity of the cloning process. To generate transgenic *C. elegans* strains, the plasmids were co-injected (100 ng/μl) with $P_{myo\text{-}2}$::*dTomato* (10 ng/μl) into Bristol (N2) wild-type hermaphrodites. For each construct, several lines were obtained, each of which gave identical expression patterns. Single lines for each construct were selected for the experiments reported here: PE612, *feEx304* [*sur-5*::*gfp*::$ICR^{Tsp\text{-}cpt\text{-}2\sim nuaf\text{-}3}$::*mCherry* $P_{myo\text{-}2}$::*dTomato*] and PE613, *feEx305* [*sur-5*::*gfp*::$ICR^{Cel\text{-}cpt2\sim prx\text{-}14}$::*mCherry* $P_{myo\text{-}2}$::*dTomato*]. *Trans*-splicing of reporter gene transcripts was analyzed as described previously (Harrison *et al.* 2010). Briefly, total RNA was reverse transcribed and *trans*-spliced

transcripts were PCR amplified using *C. elegans* SL2-specific (5′-GGTTTTAACCCAGTTACTCAAG-3′) and mCherry-specific (5′-CCGTCCTCGAAGTTCATCAC-3′) primers. Primers derived from *gpd-1* (5′-CCAACTGTCTGGCACCACT-3′ and 5′-GTCTT CTGGGTTGCGGTTAC-3′) were used to normalize the reactions. cDNA fragments were cloned into pGEM-T Easy (Promega), and inserts were sequenced at the University of Dundee DNA Sequencing Facility.

Results

A putative enoplean operon

As part of the analysis of the transcriptome of the free-living enoplean, *Prionchulus punctatus*, we identified an EST corresponding to an SL *trans*-spliced mRNA. Sequence similarity searches using this sequence identified a single predicted *T. spiralis* gene, Tsp_06075. However, subsequent sequence analysis of Tsp_06075 showed that it corresponds to an erroneous gene prediction, which conflates three genes that are the orthologs of the *C. elegans* genes *zgpa-1* (C33H5.17), *dif-1*, and *aph-1*, respectively. That Tsp_06075 is actually three separate genes was confirmed by sequence analysis of 5′ RACE products. It seems likely that the unusually short intergenic regions that exist between these three *T. spiralis* genes caused the gene annotation error (Figure 1). Such short intergenic distances are characteristic of nematode genes that are arranged into operons (Blumenthal *et al.* 2002; Guiliano and Blaxter 2006; Ghedin *et al.* 2007), and we thus decided to investigate the possibility that *Tsp-zgpa-1*, *Tsp-dif-1*, and *Tsp-aph-1* constitute an operon. In parallel, we also analyzed the homologs of these three genes in the closely related enoplean, *T. muris*, which show the same syntenic arrangement, although the intergenic distance between *Tmu-dif-1* and *Tmu-aph-1* is much larger than expected for an ICR. The three *C. elegans* homologs although organized into operons, are not found in the same operon. However, in a close relative of *C. elegans*, *Pristionchus pacificus*, *zgpa-1* and *dif-1* could potentially constitute a single operon, but again the intergenic space between the two genes is also relatively large compared to the average size of ICRs in *C. elegans* operons.

We determined the overall pattern of SL *trans*-splicing of the mRNAs derived from the putative operons in both *T. spiralis* and *T. muris* by analyzing the 5′ ends of *zgpa-1*, *dif-1*, and *aph-1* mRNAs using 5′ RACE (Figure 1A; Table S2). The analysis of *Tsp-zgpa-1* and *Tmu-zgpa-1* transcripts mapped the mRNA 5′ ends to a region 200–250 bp upstream of the start codon, and we failed to detect any SL *trans*-spliced transcripts derived from this gene in either nematode. In contrast, all *dif-1* transcripts analyzed were subject to SL *trans*-splicing in both organisms (Table S2). *Tsp-dif-1* transcripts were *trans*-spliced to *Tsp*-SL10 [note this SL was previously given the designation TSL-10 (Pettitt *et al.* 2008), but we have renamed it to conform to accepted nematode gene nomenclature rules, which employ a species-specific prefix (Beech *et al.* 2010)] and *Tmu-dif-1* transcripts were *trans*-spliced to the newly identified *Tmu*-SL1, *Tmu*-SL4, and *Tmu*-SL12 (Table S2). Analysis of *aph-1* transcripts showed that in some cases the transcripts are SL *trans*-spliced, but we were also able to detect transcripts that initiated ~200–300 bp upstream of the start codon, indicating that they were not subject to SL *trans*-splicing (Figure 1A; Table S2). It is notable that the distance between *Tmu-aph-1* and *Tmu-dif-1* is relatively large, suggesting the possibility that there are promoter elements immediately upstream of *Tmu-aph-1* that would allow the production of transcripts without the need for SL *trans*-splicing. Such "hybrid operons" have been described in *C. elegans* (Huang *et al.* 2007).

As part of this analysis, we identified spliced leaders in *T. muris*, leading to the discovery of 13 *Tmu*-SLs (Figure 2; Table S2). Previous studies have shown that the primary sequences of spliced leaders in *T. spiralis* are much more variable than those found in the Chromadorea (Pettitt *et al.* 2008), and many lack the conserved motifs that characterize spliced leaders from these latter nematodes. In contrast, those of *P. punctatus* do not show the same diversity and display a greater degree of sequence similarity to the Chromadorid spliced leaders (Harrison *et al.* 2010). Analysis of the 13 distinct *T. muris* spliced leaders, designated *Tmu*-SL1–13, support this view, since the *T. muris* spliced leaders possess the same 5′ GGUWW and central CCC motifs that are highly conserved in the *P. punctatus* spliced leaders and Chromadorid SL1 and SL2 families, but missing in most of the *T. spiralis* spliced leaders. The presence of canonical nematode spliced leaders in *T. muris* and *P. punctatus*, despite the fact that the former nematode is more closely related to *T. spiralis*, supports the inference that the *T. spiralis* spliced leaders are derived features.

If *zgpa-1*, *dif-1*, and *aph-1* are components of a *bona fide* operon in the two enoplean nematodes, we would expect to be able to detect the polycistronic RNA from which their mRNAs are derived. Although not a definitive criterion, the presence of polycistronic, partially processed pre-mRNAs is a predicted property of operon usage. We tested for the presence of such RNA molecules in both *T. spiralis* and *T. muris* (Figure 1B; Figure S1) by reverse transcription of total RNA followed by PCR with gene-specific primers. PCR products were then analyzed by agarose gel electrophoresis. In *T. spiralis*, we detected RNA species connecting the open reading frames of *Tsp-zgpa-1* with *Tsp-dif-1* (*Tsp-zgpa-1*~*dif-1*) and *Tsp-dif-1* with *Tsp-aph-1* (*Tsp-dif-1*~*aph-1*) (Figure 1B). As we failed to amplify any products in control reactions performed in parallel with RNA subjected to mock reactions without reverse transcriptase (Figure S1), these products represent processing intermediates of polycistronic transcripts.

The *Tsp-zgpa-1*~*dif-1* intermediates contained the intercistronic region, and two of the intermediates lacked introns. The *Tsp-dif-1*~*aph-1* processing intermediates detected were all subject to *cis*-splicing of *dif-1* introns, but we failed to detect an intermediate containing the complete ICR. Instead, the ICR was removed by *cis*-splicing of a cryptic splice donor site located in exon 7 of the *dif-1* gene to the SL splice acceptor site of *aph-1* (Figure 1B). Such *cis*-splicing events have

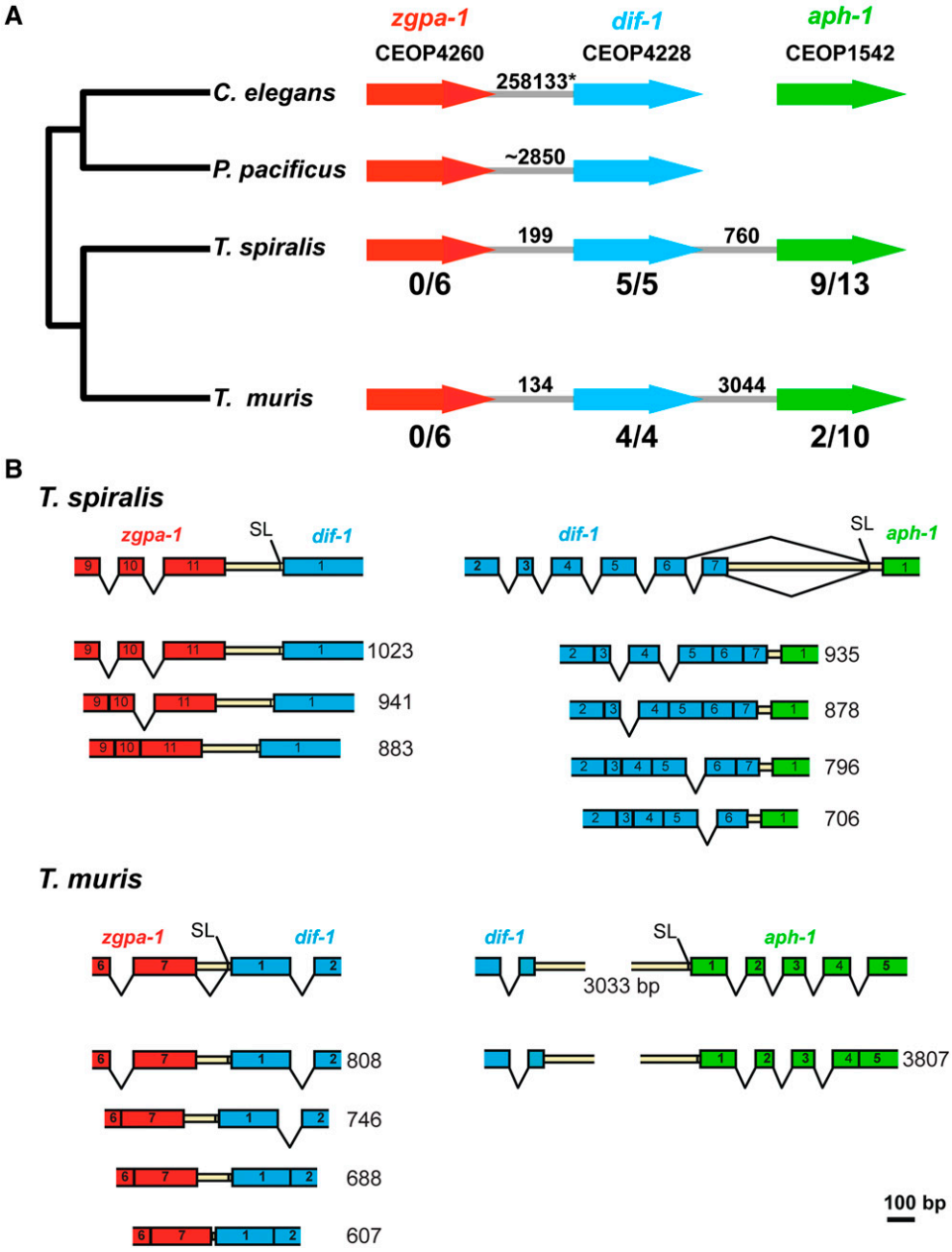


Figure 1 Evidence for the existence of an enoplean operon. (A) Schematic showing the genomic organization of *zgpa-1*, *dif-1*, and *aph-1* in selected nematodes mapped onto their phylogenetic relationships. Arrows represent genes, and the gray lines represent the intercistronic regions (ICRs). Numbers above the ICRs represent the distances, in base pairs, between the stop and start codons of the upstream and downstream genes, respectively. The *C. elegans* operon numbers are given where appropriate. Fractions below the *T. spiralis* and *T. muris* genes represent the proportion of cDNAs derived from those genes that begins with a spliced leader sequence (see also Table S2). In *C. elegans*, the three genes are part of different operons. * indicates the distance between genes on chromosome IV. (B) Detecting polycistronic RNAs derived from the *zgpa-1*∼*dif-1*∼*aph-1* operon in enoplean nematodes. The exon–intron structures of the amplicons used to identify polycistronic RNAs are shown, with exons represented by boxes (shaded to identify the genes from which they are derived using the same color coding that was used in A. The intercistronic regions are represented by cream-colored boxes. The positions of the SL *trans*-spliced 3′ splice sites are indicated. The length of each cDNA is indicated.

also been detected in putative polycistronic RNAs discovered in tapeworm genomes (Tsai *et al.* 2013). Moreover, this demonstrates that *Tsp-dif-1* and *aph-1* are transcribed as a single transcript. In *T. muris* we also detected processing intermediates corresponding to *Tmu-zgpa-1*∼*dif-1* and *Tmu-dif-1*∼*aph-1* transcripts (Figure 1B; Table S2). The latter observation is significant, since the ICR between *Tmu-dif-1* and *Tmu-aph-1* is predicted to be 3033 nt long, a distance substantially longer than the length of an average ICR in *C. elegans*, although ICRs of similar length are also present in some *C. elegans* operons (Morton and Blumenthal 2011).

Identification of additional putative enoplean operons

To more systematically identify enoplean operons, we adopted two approaches. First, we used a set of EST sequences derived from SL *trans*-spliced *T. spiralis* mRNAs (Pettitt *et al.* 2008) to identify their corresponding genes via sequence similarity searches (we also identified two *T. muris* mRNAs via the same approach: FF145866 and CB277782). For each gene, we then looked for neighboring genes predicted to be transcribed in the same orientation and that lay within 1 kb. Using this approach, we were able to identify multiple potential operons in the *T. spiralis* genome (Table 1; Table S1). We further analyzed this set of genes by identifying the *C. elegans* orthologs of each *T. spiralis* gene and determining whether these correspond to genes within operons. Our analysis revealed that at least 75 of the *C. elegans* orthologs are arranged in operons. This represents 44% of the *C. elegans* genes identified as orthologs of our *T. spiralis* SL *trans*-spliced EST set. Since only 15% of *C. elegans* genes are organized into operons

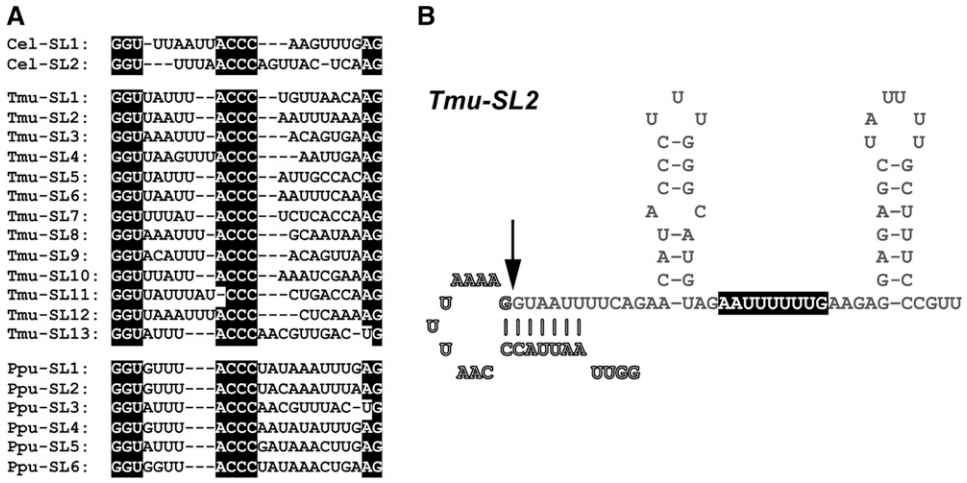


Figure 2 *T. muris* SL sequences and SL RNA structure. (A) *Tmu*-SL1–13 genes were identified using a combination of cDNA sequencing and bioinformatics tools as described in *Materials and Methods*. *Tmu*-SL12 was found by 5′ RACE *trans*-spliced to *nuaf-3* mRNA, and *Tmu*-SL13 was found *trans*-spliced to *aph-1* mRNA. In the alignment, only the SL sequences are shown. *T. muris* SL sequences were manually aligned and conserved groups are countershaded. *C. elegans* SL1 and SL2 and the previously identified *P. punctatus* SL sequences were included for comparison. (B) The intron of *Tmu*-SL2 was experimentally identified and also found in the genome sequence. The proposed secondary structure was produced using M-fold (Zuker 2003). The SL sequence is shown in outline font and the putative Sm sequence motif is countershaded.

(Allen *et al.* 2011), we would expect only 15% of *T. spiralis* genes in our dataset to match *C. elegans* operonic homologs if they were selected at random. It is difficult to determine the reason for the increased likelihood of matches to *C. elegans* operonic genes among the *T. spiralis* SL *trans*-spliced ESTs; it may be that the corresponding *T. spiralis* gene set is biased for highly expressed genes, for instance. However, it is consistent with the possibility that this dataset is enriched for transcripts derived from operonic genes.

As an alternative approach, which would potentially identify operons that have been conserved since the separation of the Enoplea and Chromadorea, we looked for *T. spiralis* homologs of a set of putative operons conserved between *C. elegans* and *B. malayi* (Ghedin *et al.* 2007). Of the 107 operonic gene pairs screened, we identified 12 *T. spiralis* gene pairs that displayed conserved synteny, and whose component genes were separated by an average intergenic distance of 607 bp (Table 2). To determine whether any of these operons are also conserved in another enoplean species, we examined the organization of the corresponding homologous genes in the genome of *R. culicivorax* (Schiffer *et al.* 2013). This analysis revealed that 4 of the 12 gene pairs were arranged in putative operons (assuming a maximum ICR distance of up to 1 kb) in this nematode (Table 2).

Taken together, our analysis indicates that there are multiple *T. spiralis* gene pairs whose genomic arrangement is consistent with their corresponding to operons. Moreover, it is possible to identify gene pairs conserved between *T. spiralis*, *R. culicivorax*, *B. malayi*, and *C. elegans*, suggesting that these represent operons that were present in the last common ancestor of the three species.

Characterization of a conserved nematode operon

The analysis of one of the two SL *trans*-spliced *T. muris* mRNAs (GenBank accession no. FF145866) led to the identification of a putative operon conserved between multiple nematodes (Figure 3A). The two genes contained in these putative operons have *C. elegans* homologs, *cpt-2* and *nuaf-3*, respectively, which are in the same operon (CEOP4424) (Figure 3A), although there is an additional gene, *prx-14*, located between these genes in CEOP4424 that is not present in the putative homologous *T. spiralis* operon. Examination of the genomic organization of the homologous genes in a selection of nematode species confirmed the evolutionary conservation of the synteny of the *cpt-2* and *nuaf-3* homologs (Figure 3A). This analysis also showed that insertion of *prx-14* into the *cpt-2*∼*nuaf-3* operon was a relatively recent event, since it is present only in *C. elegans* and other closely related *Caenorhabditis* species. We also find, based on the head-to-tail organization and spacing between coding regions, that there is variation in the composition of both operons in different species, and these genes in *R. culicivorax* are not in operons.

We further focused on the *cpt-2*∼*nuaf-3* operon in *T. spiralis* and *T. muris* to determine the pattern of SL *trans*-splicing exhibited by the mRNAs derived from this operon and to verify that we were able to detect cDNAs consistent with the production of polycistronic RNAs.

Determination of the 5′ ends of *cpt-2* and *nuaf-3* transcripts by 5′ RACE revealed that *nuaf*-3 mRNA is subject to

Table 1 Operonic status of genes that match ESTs derived from *T. spiralis* SL *trans*-spliced transcripts

	Location in operon			
Species	Upstream	Downstream	Ambiguous	Nonoperonic
T. spiralis	30	35	40	54
C. elegans	29	46	5[a]	89

The status of each *T. spiralis* gene was determined using criteria given in *Materials and Methods*. The status of *C. elegans* genes was obtained from WormBase (Release WS237). Eleven EST matches were absent from the *T. spiralis* data relative to the *C. elegans* data as the corresponding *T. spiralis* gene could not be identified (see *Materials and Methods*). Full details of the individual EST sequence matches are given in Table S1.

[a] Genes not annotated as operons, but having intergenic spacing with respect to their neighbors, that suggests they may be organized in an operon.

Table 2 Putative conserved nematode operons

T. spiralis		*R. culicivorax*		*C. elegans*			*B. malayi*	
Upstream gene	Downstream gene	Upstream gene	Downstream gene	Upstream gene	Downstream gene	Operon no.	Upstream gene	Downstream gene
Tsp_00685	[a]	Nonoperonic		*mrps-17*	*C05D11.9*	3372	Bm1_13520	Bm1_13525
Tsp_03140	Tsp_03139	t32947	t32944	*T26E3.4*	*par-6*	1672	Bm1_48785	Bm1_48780
Tsp_05540	Tsp_05541	Nonoperonic		*K11B4.1*	*K11B4.2*	1764	Bm1_55805	Bm1_55810
Tsp_06077	Tsp_06076	Nonoperonic		*Y62E10A.2*	*Y62E10A.6*	4540	Bm1_54855	Bm1_54850
Tsp_06996	[a]	t05598/9	t05596	*sel-1*	*mrps-5*	5365	Bm1_45745	Bm1_45750
Tsp_09103	Tsp_09102	t35569	t35568	*snu-23*	*ZK686.3*	3452	Bm1_13735	Bm1_13740
Tsp_09506	[a]	Nonoperonic		*H20J04.6*	*mog-2*	2124	Bm1_15855	Bm1_15860
Tsp_09539	[a]	Nonoperonic		*E02H1.5*	*E02H1.6*	2436	Bm1_24720	Bm1_24715
Tsp_10673	Tsp_10674	t34344.1	t34344.2	*B0491.1*	*B0491.7*	2532	Bm1_10780	Bm1_10785
Tsp_10698	Tsp_10702	Nonoperonic		*trpp-8*	*vha-10*	1264	Bm1_12140	Bm1_12135
Tsp_10959	[a]	Nonoperonic		*ubxn-2*	*Y94H6A.8*	4665	Bm1_36515	Bm1_36520
Tsp_11898	[a]	Nonoperonic		*lst-6*	*sqv-7*	2276	Bm1_24075	Bm1_24080

[a] Single gene annotation matches to both *C. elegans* genes in the operon pair consistent with annotation error caused by short intergenic spacing.

SL *trans*-splicing (Table S2) in both nematodes. However, we failed to detect any SL *trans*-spliced *cpt-2* transcripts, similar to the situation with *zgpa-1* transcripts in the *zgpa-1∼dif-1∼aph-1* operon.

We were also able to detect processing intermediates derived from the putative polycistronic *T. spiralis cpt-2∼nuaf-3* transcripts. As for the *zgpa-1∼dif-1∼aph-1* operon, in addition to unprocessed, polycistronic transcripts, we detected *cis*-spliced intermediates lacking *cpt-2* and *nuaf-3* introns and a transcript in which the ICR was removed by splicing from a cryptic 3′ splice site in *cpt-2* to the *nuaf-3* SL splice acceptor site (Figure 3B).

The T. spiralis cpt-2∼nuaf-3 ICR can mediate polycistronic RNA processing in C. elegans

Analysis of the intercistronic region between *Tsp-cpt-2* and *nuaf-3* downstream of the polyadenylation signal of *Tsp-cpt-2* revealed a clear Ur element and there are several U-rich regions, characteristics of the ICRs in *C. elegans* operons (Graber *et al.* 2007). To investigate the possibility that this region is able to function in polycistronic RNA processing, we determined whether the ICR from it could be recognized and processed if heterologously expressed in *C. elegans*. We generated an artificial operon consisting of *sur-5*::*gfp* (Gu *et al.* 1998) and mCherry genes flanking the ICR from *Tsp-cpt-2∼nuaf-3*. Transgenic animals carrying this construct expressed nuclear GFP and cytoplasmic mCherry, consistent with the processing of the two coding regions under the direction of the *Tsp-cpt-2∼nuaf-3* ICR. We confirmed that this involved *trans*-splicing to SL2, as expected for polycistronic RNA processing in *C. elegans*, by showing that we could detect SL2 *trans*-splicing to the mCherry mRNA in RNA derived from transgenic animals (Figure 4). Thus, the predicted ICR between *Tsp-cpt-2* and *nuaf-3* is recognized and used as a substrate for polycistronic RNA processing in *C. elegans*.

Conservation of SL trans-splicing snRNPs between Enoplea and Chromadorea

Our data show that both SL *trans*-splicing and operons were likely present in the last common ancestor of the nematode phylum. This suggests that the processing machinery necessary for the coordination of these processes was already in place prior to radiation of the nematode phylum. To address this, we sought to determine the conservation of known protein components that are specifically involved in SL *trans*-splicing. Previous studies have shown the existence of two interacting proteins, conserved between *Ascaris suum* and *C. elegans*, that are components of the SL small nuclear ribonucleic particle (snRNP) (Denker *et al.* 2002; MacMorris *et al.* 2007). The two proteins, termed SNA-1 and SNA-2 in *C. elegans*, form a complex with SL RNA. In addition, a paralog of SNA-1, SUT-1, forms novel snRNPs containing SNA-2 and a family of nematode-specific RNAs, designated Sm Y (MacMorris *et al.* 2007; Jones *et al.* 2009). The function of Sm Y RNAs is not known, but they are associated with SL *trans*-splicing (Maroney *et al.* 1996), and a role in the recycling of Sm proteins following SL *trans*-splicing has been proposed (MacMorris *et al.* 2007).

We were able to identify credible homologs of *sna-2* in the genomes of all nematodes in which searches were carried out, including *T. spiralis* and *R. culicovorax*, indicating that this gene encodes a nematode-wide SL *trans*-splicing component (Figure S2). Searches of the same datasets identified clear *sna-1* and *sut-1* orthologs in the chromadorean nematodes (Figure 5), extending the previously reported phylogenetic distribution of *sut-1* (MacMorris *et al.* 2007). In contrast, the only enoplean genome in which we were able to identify a *sut-1* homolog was *R. culicivorax* (Figure 5). We could not detect *sna-1* homologs in any of the enoplean genomes that we assayed. These data suggest that the gene duplication event that gave rise to *sna-1* and *sut-1* occurred after the separation of the two major nematode taxa; however, we cannot rule out that the possibility that our failure to detect *sna-1* homologs is due to the incomplete state of the enoplean genome drafts.

Discussion

The incidence of operons as a means to coordinate gene expression has been investigated in only one of the two main nematode taxa, leaving unanswered the question about when

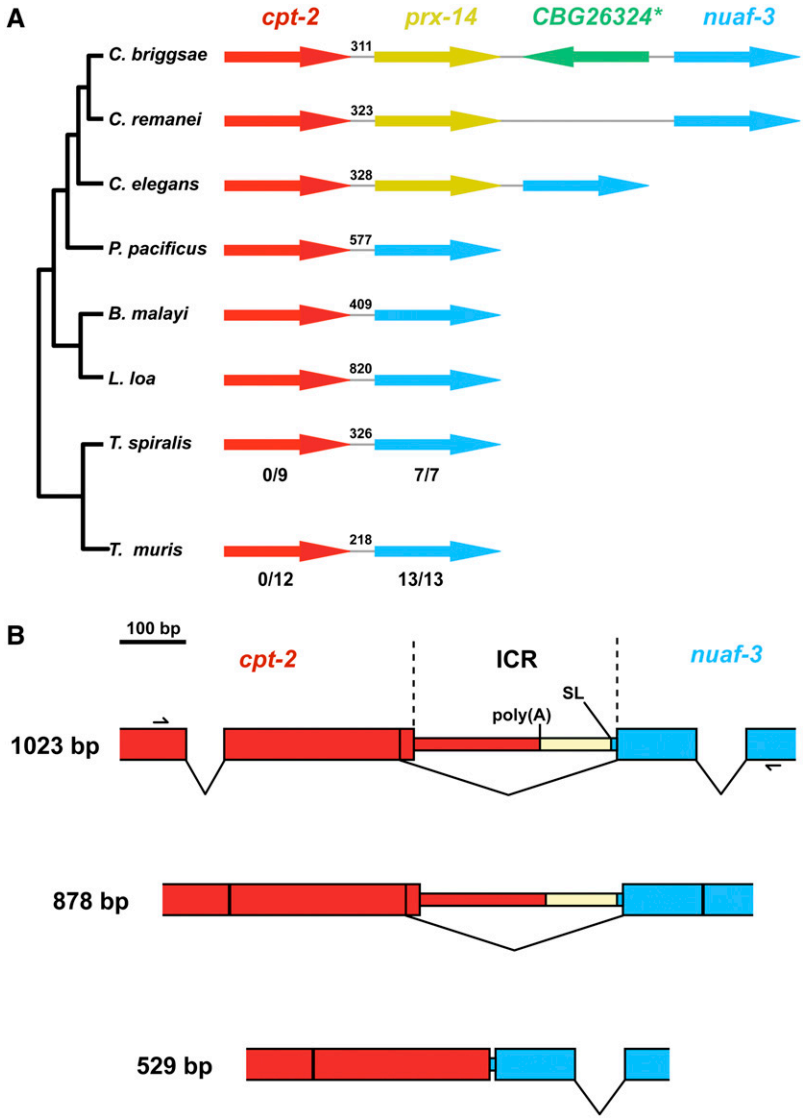


Figure 3 Evidence for an evolutionarily conserved nematode operon. (A) The structure of the *cpt-2-nuaf-3* genomic regions from a range of nematode species mapped onto the nematode phylogeny. Genes are represented by arrows, and the gray lines represent the intercistronic regions (ICRs). Numbers above the ICRs represent the distances, in base pairs, between the stop and start codons of the upstream and downstream genes, respectively. The *C. elegans* operon numbers are given where appropriate. Fractions below the *T. spiralis* and *T. muris* genes represent the proportion of cDNAs derived from those genes that begins with a spliced leader sequence. (B) Detecting polycistronic RNAs derived from the *cpt-2*∼*nuaf-3* operon in *T. spiralis*. The exon–intron structures of the amplicons used to identify polycistronic RNAs are shown, with exons represented by boxes (shaded to identify the genes from which they are derived using the same color coding that was used in A. The region removed during operon processing is represented by cream-colored boxes. The positions of the SL *trans*-splice 3′ spliced sites are indicated.

this mechanism first occurred during nematode evolution (Guiliano and Blaxter 2006; Ghedin *et al.* 2007; Liu *et al.* 2010). The work presented here provides strong support for the existence of operons in the Enoplea, and that operons were likely present in the ancestor of the Chromadorea and the Enoplea. The unequivocal identification of operons is not straightforward in those nematodes that do not utilize a specialized spliced leader to resolve polycistronic RNAs; the use of SL2 in *C. elegans* and its close relatives has greatly facilitated the identification of operons in these species. In contrast, in enoplean species we can only infer the presence of operons through multiple lines of evidence.

Our work demonstrates the existence of clusters of genes ordered in a head-to-tail arrangement with short, less than 1 kb, intergenic distances, consistent with their being organized into operons. The fact that we can identify homologous pairs of closely spaced genes conserved between enoplean and chromadorean nematodes suggests that at least in these cases there has been selective pressure to retain short intergenic distances, consistent with what would be expected if they were part of operons. Analysis of the transcripts produced by these putative operons shows that, as expected, genes predicted to be downstream in operons produce mRNAs that are SL *trans*-spliced. Further supporting evidence comes from the fact that we can detect unprocessed, polycistronic pre-mRNAs derived from these putative operons. Thus these genes exhibit the molecular properties expected of operons. Most significantly, the intergenic region between *Tsp-cpt-2* and *nuaf-3* acts as a substrate for the polycistronic RNA processing machinery of *C. elegans*, providing the strongest evidence that it is part of an operon in *T. spiralis*.

The identification of SLs from *T. muris* has extended our understanding of nematode SL evolution. Previous studies have shown that *T. spiralis* and *T. pseudospiralis* have unusually diverse SLs that do not readily correspond to those SLs found within Chromadorea (Pettitt *et al.* 2008, 2010). Another enoplean, *P. punctatus*, possesses SLs that resemble the specialized SL2 associated with *trans*-splicing and polycistronic RNA

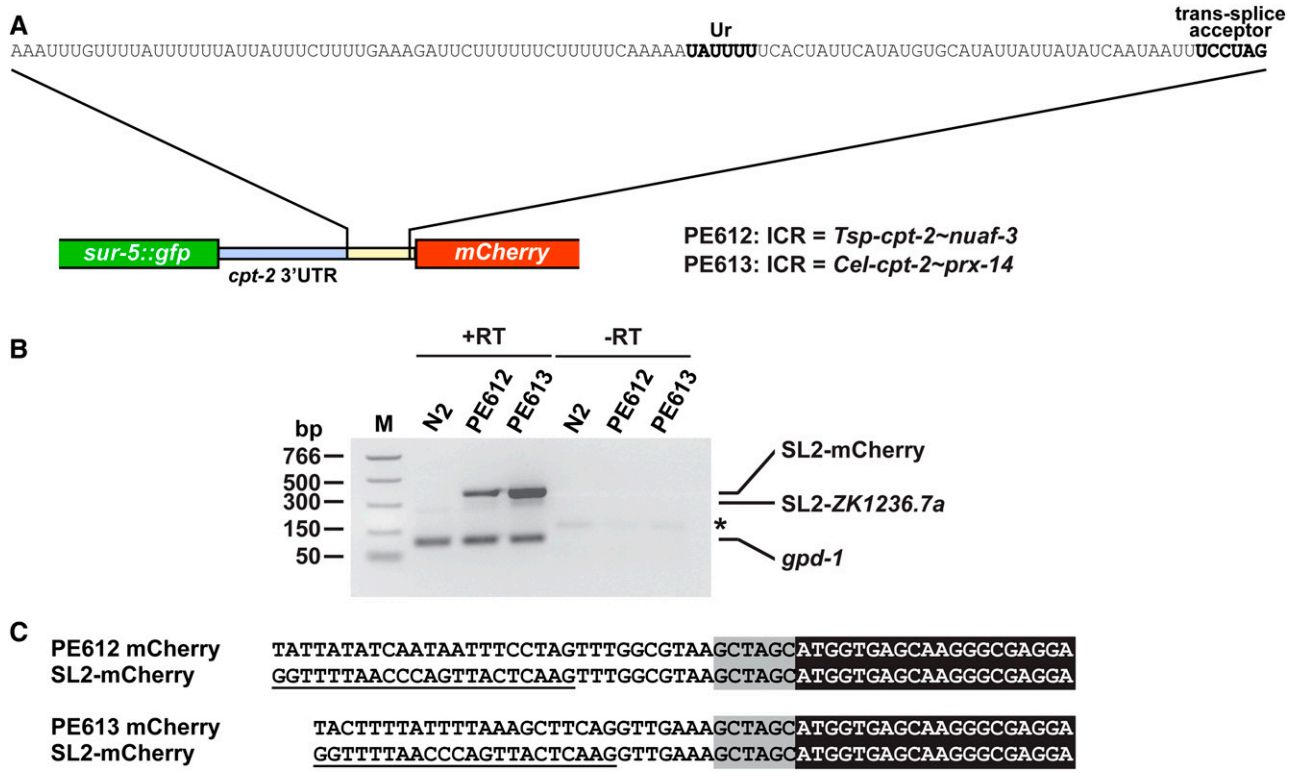


Figure 4 Processing of a synthetic operon containing the *T. spiralis cpt-2~nuaf-3* intercistronic region in *C. elegans*. (A) Schematic of the structure of *sur-5::gfp~mCherry* synthetic operon construct containing the *Tsp-cpt-2~nuaf-3* intercistronic region. The sequence immediately downstream from the *Tsp-cpt-2* 3′-UTR is shown, illustrating the presence of the Ur motif and the *trans*-splice acceptor site. Strain PE613 contains an identical construct, but with the ICR replaced with that from between *Cel-cpt-2* and *prx-14*. (B) Detection of operon transcript processing by SL2 *trans*-splicing. mCherry transcripts *trans*-spliced to SL2 were detected by reverse transcription of RNA prepared from either PE612 or PE613 animals followed by PCR with an SL2 primer and a primer located in the mCherry coding region (+RT). Primers amplifying *gpd-1* were included to control for sample variation. *gpd-1* genomic DNA was detected in the −RT control reaction (*) and SL2-ZK1236.7a is a minor product detected in the +RT reactions. Reactions with RNA isolated from N2 wild-type animals were included as control. M is a DNA size standard. (C) Alignment of transgene sequences and SL2-mCherry transcripts confirming correct splice site usage. The beginning of the mCherry open reading frame is countershaded black, the *Nhe*I cloning site is countershaded gray, and the SL2 sequences are underlined.

resolution found in *C. elegans* and other closely related nematodes. At least one of these is specifically recognized by the SL2 *trans*-splicing machinery in *C. elegans* when expressed heterogeneously (Harrison *et al.* 2010). Rather surprisingly, since it is more closely related to *Trichinella* species than *P. punctatus*, the *T. muris* SLs are more similar in terms of sequence composition to those of the latter nematode (the most extreme example of this being *Ppu*-SL3 and *Tmu*-SL13, which possess >95% sequence identity). Thus, the diverse SL complement of the *Trichinella* species is likely to be a derived rather than ancestral trait.

In addition to elucidating the extent of nematode SL RNA conservation, we have determined the phylogenetic distribution of the protein components of the SL snRNP. This reveals that one of these components, SNA-2, is conserved throughout the phylum, consistent with its essential role in SL *trans*-splicing in both *A. suum* and *C. elegans* (Denker *et al.* 2002; MacMorris *et al.* 2007). Our analysis of SNA-1/SUT-1 has revealed that these two paralogs are distributed throughout the Chromadorea. These data, together with the presence of a single clear SUT-1 homolog in *R. culicivorax* (and absence of a SNA-1 homolog) is consistent with an evolutionary scenario whereby the ancestral *sut-1* gene was duplicated in the ancestor of the Chromadorea, with one paralog, *sut-1*, retaining the ancestral function, while the other, *sna-1*, evolved a derived function. However, the fact that *sna-1* and *sut-1* show a synthetic loss of function phenotype in *C. elegans* (MacMorris *et al.* 2007) indicates that their functions are related and possibly overlapping. The puzzling absence of SUT-1 homologs in *T. spiralis* and *T. muris* means that such an interpretation must be provisional, pending broader sampling of genome sequences from other enopleans. If these nematodes actually lack these proteins this would be surprising, given that loss of both *sna-1* and *sut-1* function in *C. elegans* results in embryonic lethality (MacMorris *et al.* 2007) and SNA-1 is required for SL *trans*-splicing in *A. suum* splicing extracts (Denker *et al.* 2002).

An important question that remains to be answered is whether there is any evidence for functional specialization of the enoplean SLs with regard to polycistronic RNA resolution. In *C. elegans*, SL2-type SLs are clearly specific for this latter process (Blumenthal *et al.* 2002), with SL1 *trans*-splicing almost exclusively employed in mRNAs derived from nonoperonic

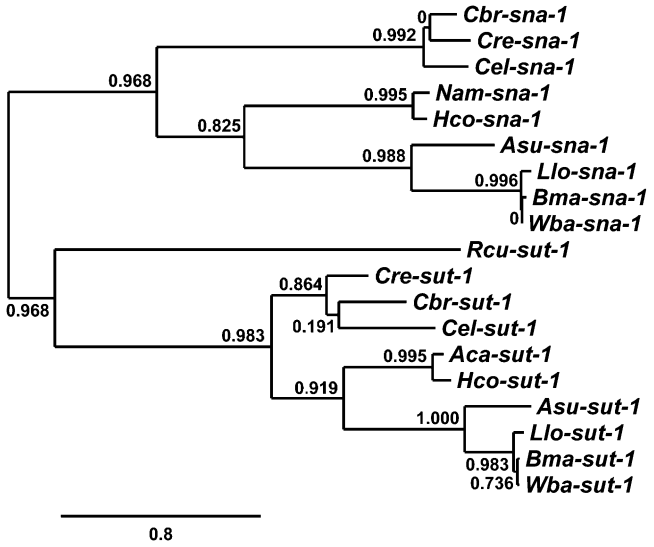


Figure 5 Evolutionary relationship between snRNPs associated with SL *trans*-splicing. Unrooted PhyML tree showing the relationship between *sna-1* and *sut-1* homologs identified in selected nematodes. Genes were named on the basis of their *C. elegans* homologs. The following species-specific prefixes were used: Aca, *Angiostrongylus cantonensis*; Asu, *Ascaris suum*; Bma, *Brugia malayi*; Cel, *C. elegans*; Cbr, *C. briggsae*; Cre, *C. remanei*; Hco, *Haemochus contortus*; Llo, *Loa loa*; Nam, *Necator americanus*; Rcu, *Romanomermis culicivorax*; and Wba, *Wucheria bancrofti*. The numbers at each node are approximate likelihood ratio test statistics.

genes, or the first gene in an operon. Other nematodes within the Rhabditina also show the same specialization (Evans *et al.* 1997; Lee and Sommer 2003), but other taxa within the Chromadorea appear to use the same set of SLs for all SL *trans*-spliced mRNAs (Guiliano and Blaxter 2006). Within the Enoplea we clearly see multiple different SLs used in the *trans*-splicing of mRNAs derived from downstream genes in operons, but our data are not comprehensive enough to determine whether some SLs are preferentially used to process the transcripts arising from such genes.

We have identified multiple examples of gene pair synteny conserved between *T. spiralis*, *R. culicivorax*, *B. malayi*, and *C. elegans*, suggesting these correspond to conserved operons present in the last common ancestor of the Enoplea and Chromadorea. Nevertheless, the majority of the putative operons that we have identified in *T. spiralis* are not syntenic with *C. elegans* operons, although in many cases we see that the component genes of these putative *T. spiralis* operons have *C. elegans* orthologs that are found in operons. An example of such an operon is that comprising *Tsp-zgpa-1*, *Tsp-dif-1*, and *Tsp-aph-1*, whose *C. elegans* orthologs are located in different operons. We also observed putative *T. spiralis* operon genes whose *C. elegans* orthologs are not part of operons. The changes in the operon complements between the two nematodes could arise from the lineage-specific rearrangements of an ancestral set of nematode operons, but could also be accounted for by differential *de novo* operon generation in the two clades, or a mixture of the two processes. A key question is whether the synteny shown between the putative enoplean operons and their chromadorean orthologs is significant, *i.e.*, are the genes that comprise these operons more constrained to be located in the same operon than other operonic genes, or is the conservation of synteny merely random chance? The availability of tools to engineer the *C. elegans* genome (Golic 2013) might allow this question to be addressed by assaying the function of selected operons compared with their individual component genes each expressed under their own promoters.

Finally, it is clear from this work and previous studies that SL *trans*-splicing and operon organization arose prior to the divergence of the Enoplea and Chromadorea. An important question is whether "nematode" SL *trans*-splicing and operons predate the foundation of the phylum. It will thus be important to establish whether they are also present in the other, so far uncharacterized, phyla that are closely related to the nematodes.

Acknowledgments

We thank Allison Bancroft and Richard Grencis for providing *T. muris* RNA. The Bristol (N2) *C. elegans* N2 strain was provided by the *Caenorhabditis* Genetics Center, which is funded by National Institutes of Health Office of Research Infrastructure Programs (P40 OD010440). This work was funded by a Biotechnology and Biological Sciences Research Council project grant BB/J007137/1.

Literature Cited

Allen, M. A., L. W. Hillier, R. H. Waterston, and T. Blumenthal, 2011 A global analysis of *C. elegans trans*-splicing. Genome Res. 21: 255–264.

Beech, R. N., A. J. Wolstenholme, C. Neveu, and J. A. Dent, 2010 Nematode parasite genes: what's in a name? Trends Parasitol. 26: 334–340.

Blumenthal, T., D. Evans, C. D. Link, A. Guffanti, D. Lawson *et al.*, 2002 A global analysis of *Caenorhabditis elegans* operons. Nature 417: 851–854.

Dana, C. E., K. M. Glauber, T. A. Chan, D. M. Bridge, and R. E. Steele, 2012 Incorporation of a horizontally transferred gene into an operon during cnidarian evolution. PLoS ONE 7: e31643.

Davis, R. E., and S. Hodgson, 1997 Gene linkage and steady state RNAs suggest trans-splicing may be associated with a polycistronic transcript in *Schistosoma mansoni*. Mol. Biochem. Parasitol. 89: 25–39.

Denker, J. A., D. M. Zuckerman, P. A. Maroney, and T. W. Nilsen, 2002 New components of the spliced leader RNP required for nematode *trans*-splicing. Nature 417: 667–670.

Dereeper, A., V. Guignon, G. Blanc, S. Audic, S. Buffet *et al.*, 2008 Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 36: W465–W469.

Evans, D., D. Zorio, M. MacMorris, C. E. Winter, K. Lea *et al.*, 1997 Operons and SL2 *trans*-splicing exist in nematodes outside the genus Caenorhabditis. Proc. Natl. Acad. Sci. USA 94: 9751–9756.

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

Ghedin, E., S. Wang, D. Spiro, E. Caler, Q. Zhao *et al.*, 2007 Draft genome of the filarial nematode parasite *Brugia malayi*. Science 317: 1756–1760.

Golic, K. G., 2013 RNA-guided nucleases: a new era for engineering the genomes of model and nonmodel organisms. Genetics 195: 303–308.

Graber, J. H., J. Salisbury, L. N. Hutchins, and T. Blumenthal, 2007 *C. elegans* sequences that control trans-splicing and operon pre-mRNA processing. RNA 13: 1409–1426.

Gu, T., S. Orita, M. Han, 1998 *Caenorhabditis elegans* SUR-5, a novel but conserved protein, negatively regulates LET-60 Ras activity during vulval induction. Mol. Cell. Biol. 18: 4556–4564.

Guiliano, D. B., and M. L. Blaxter, 2006 Operon conservation and the evolution of *trans*-splicing in the phylum Nematoda. PLoS Genet. 2: e198.

Harrison, N., A. Kalbfleisch, B. Connolly, J. Pettitt, and B. Müller, 2010 SL2-like spliced leader RNAs in the basal nematode *Prionchulus punctatus*: New insight into the evolution of nematode SL2 RNAs. RNA 16: 1500–1507.

Hastings, K. E. M., 2005 SL *trans*-splicing: easy come or easy go? Trends Genet. 21: 240–247.

Hobert, O., 2002 PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic *C. elegans*. Biotechniques 32: 728–730.

Holterman, M., A. van der Wurff, S. van den Elsen, H. van Megen, T. Bongers *et al.*, 2006 Phylum-wide analysis of SSU rDNA reveals deep phylogenetic relationships among nematodes and accelerated evolution toward crown Clades. Mol. Biol. Evol. 23: 1792–1800.

Huang, P., E. D. Pleasance, J. S. Maydan, R. Hunt-Newbury, N. J. O'Neil *et al.*, 2007 Identification and analysis of internal promoters in *Caenorhabditis elegans* operons. Genome Res. 17: 1478–1485.

Jacob, F., D. Perrin, C. Sánchez, J. Monod, and S. Edelstein, 1960 The operon: a group of genes with expression coordinated by an operator. C. R. Acad. Sci. Paris 250: 1727–1729.

Johnson, P. J., J. M. Kooter, and P. Borst, 1987 Inactivation of transcription by UV irradiation of *T. brucei* provides evidence for a multicistronic transcription unit including a VSG gene. Cell 51: 273–281.

Jones, T., W. Otto, M. Marz, S. Eddy, and P. Stadler, 2009 A survey of nematode SmY RNAs. RNA Biol. 6: 5–8.

Lawrence, J., 1999 Selfish operons: the evolutionary impact of gene clustering in prokaryotes and eukaryotes. Curr. Opin. Genet. Dev. 9: 642–648.

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

Liu, C., A. Oliveira, C. Chauhan, E. Ghedin, and T. R. Unnasch, 2010 Functional analysis of putative operons in *Brugia malayi*. Int. J. Parasitol. 40: 63–71.

MacMorris, M., M. Kumar, E. Lasda, A. Larsen, B. Kraemer *et al.*, 2007 A novel family of *C. elegans* snRNPs contains proteins associated with trans-splicing. RNA 13: 511–520.

Marlétaz, F., A. Gilles, X. Caubit, Y. Perez, C. Dossat *et al.*, 2008 Chaetognath transcriptome reveals ancestral and unique features among bilaterians. Genome Biol. 9: R94.

Maroney, P. A., Y. T. Yu, M. Jankowska, and T. W. Nilsen, 1996 Direct analysis of nematode *cis*- and *trans*-spliceosomes: a functional role for U5 snRNA in spliced leader addition *trans*-splicing and the identification of novel Sm snRNPs. RNA 2: 735–745.

Meldal, B. H. M., N. J. Debenham, P. De Ley, I. T. De Ley, J. R. Vanfleteren *et al.*, 2007 An improved molecular phylogeny of the Nematoda with special emphasis on marine taxa. Mol. Phylogenet. Evol. 42: 622–636.

Mitreva, M., D. P. Jasmer, D. S. Zarlenga, Z. Wang, S. Abubucker *et al.*, 2011 The draft genome of the parasitic nematode *Trichinella spiralis*. Nat. Genet. 43: 228–235.

Morton, J. J., and T. Blumenthal, 2011 Identification of transcription start sites of *trans*-spliced genes: uncovering unusual operon arrangements. RNA 17: 327–337.

Pettitt, J., N. Harrison, I. Stansfield, B. Connolly, and B. Müller, 2010 The evolution of spliced leader *trans*-splicing in nematodes. Biochem. Soc. Trans. 38: 1125–1130.

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

Protasio, A. V., I. J. Tsai, A. Babbage, S. Nichol, M. Hunt *et al.*, 2012 A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. PLoS Negl. Trop. Dis. 6: e1455.

Satou, Y., K. Mineta, M. Ogasawara, Y. Sasakura, E. Shoguchi *et al.*, 2008 Improved genome assembly and evidence-based global gene model set for the chordate *Ciona intestinalis*: new insight into intron and operon populations. Genome Biol. 9: R152.

Schiffer, P. H., M. Kroiher, C. Kraus, G. D. Koutsovoulos, S. Kumar *et al.*, 2013 The genome of *Romanomermis culicivorax*: revealing fundamental changes in the core developmental genetic toolkit in Nematoda. BMC Genomics 14: 923.

Spieth, J., G. Brooke, S. Kuersten, K. Lea, and T. Blumenthal, 1993 Operons in *C. elegans*: polycistronic mRNA precursors are processed by *trans*-splicing of SL2 to downstream coding regions. Cell 73: 521–532.

Tsai, I. J., M. Zarowiecki, N. Holroyd, A. Garciarrubio, A. Sanchez-Flores *et al.*, 2013 The genomes of four tapeworm species reveal adaptations to parasitism. Nature 496: 57–63.

Zaslaver, A., L. R. Baugh, and P. W. Sternberg, 2011 Metazoan operons accelerate recovery from growth-arrested states. Cell 145: 981–992.

Zuker, M., 2003 Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31: 3406–3415.

Communicating editor: M. P. Colaiácovo

GENETICS

Supporting Information
http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.162875/-/DC1

Operons Are a Conserved Feature of Nematode Genomes

Jonathan Pettitt, Lucas Philippe, Debjani Sarkar, Christopher Johnston, Henrike Johanna Gothe, Diane Massie, Bernadette Connolly, and Berndt Müller

DOI: 10.1534/genetics.114.162875

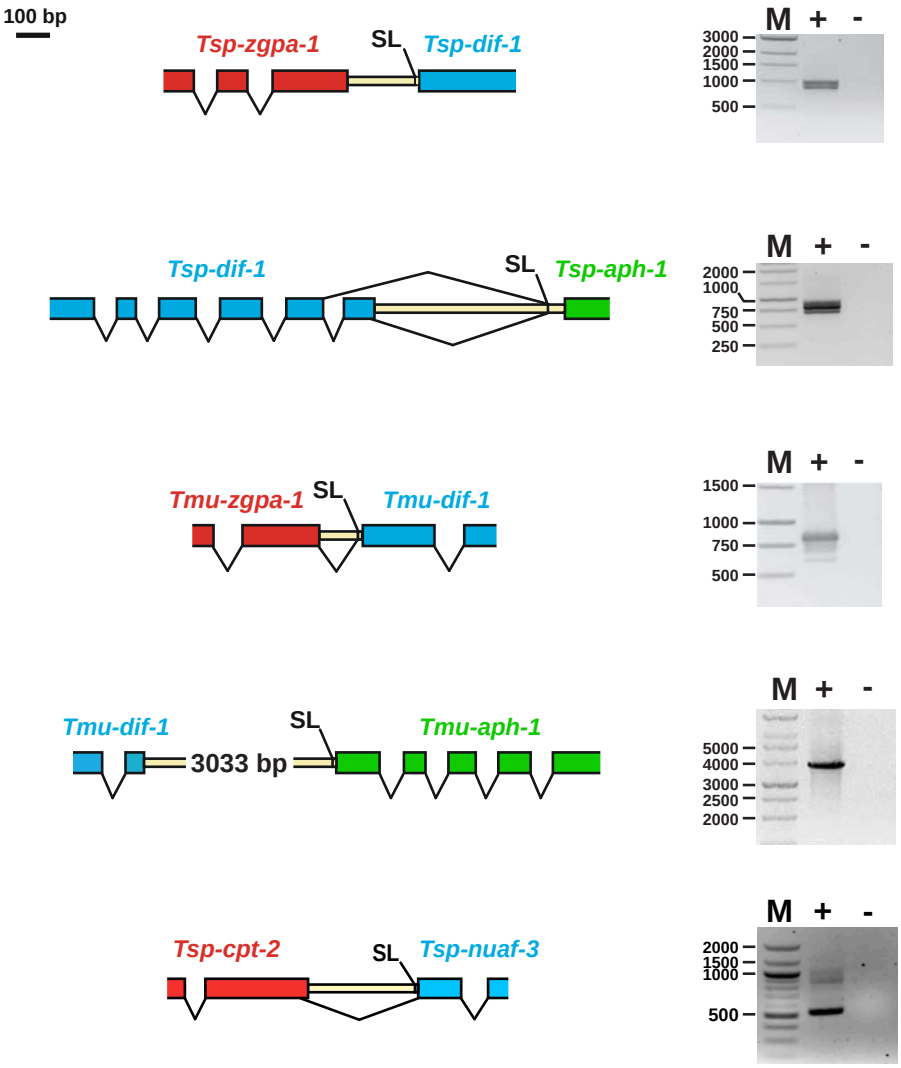
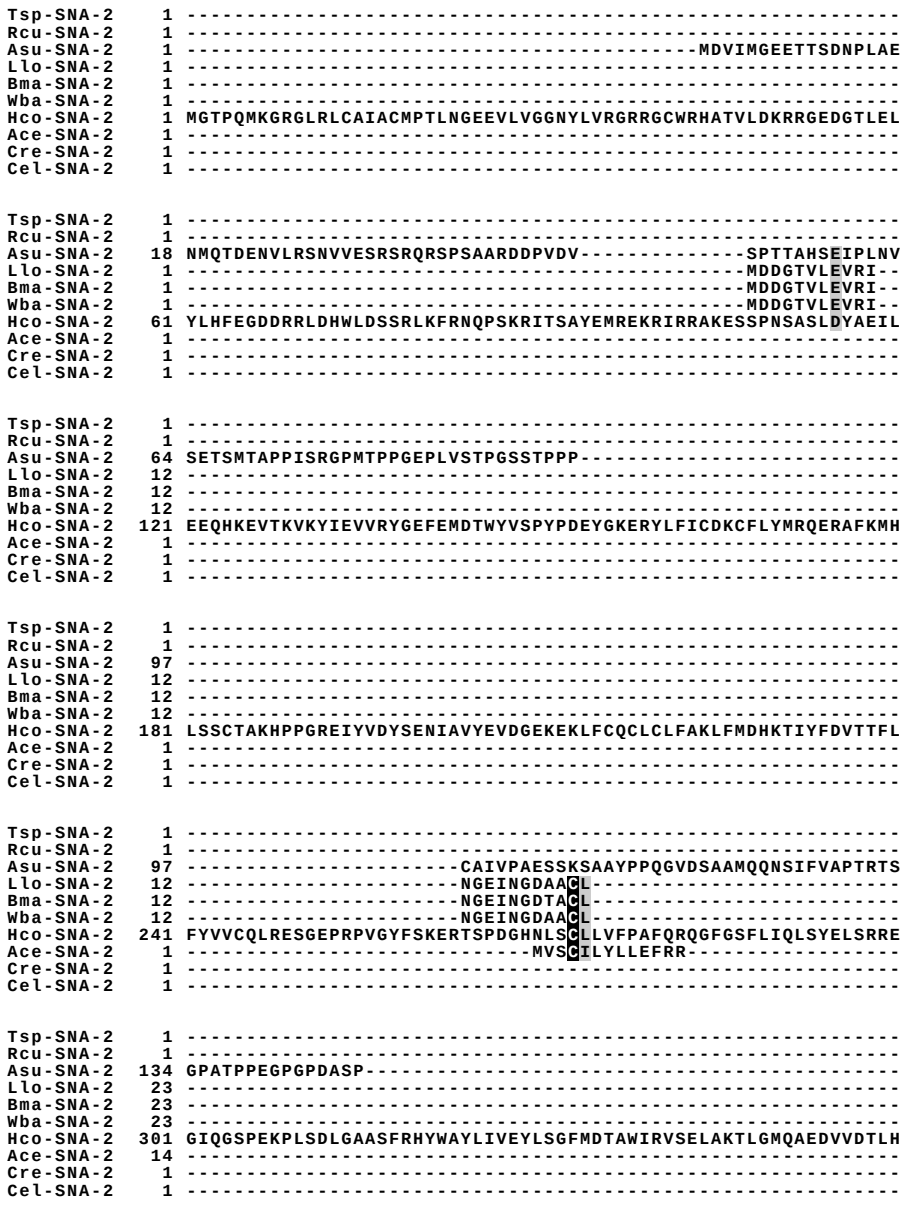
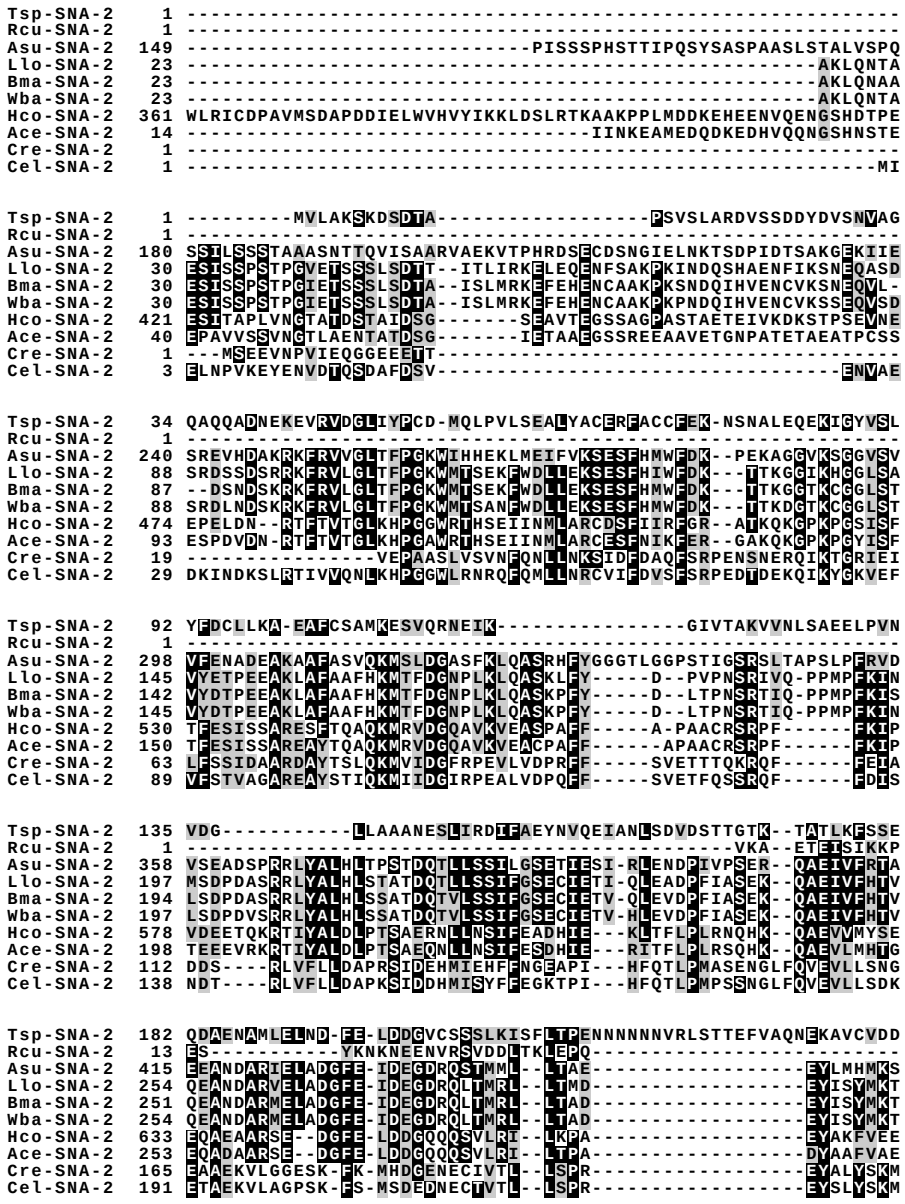
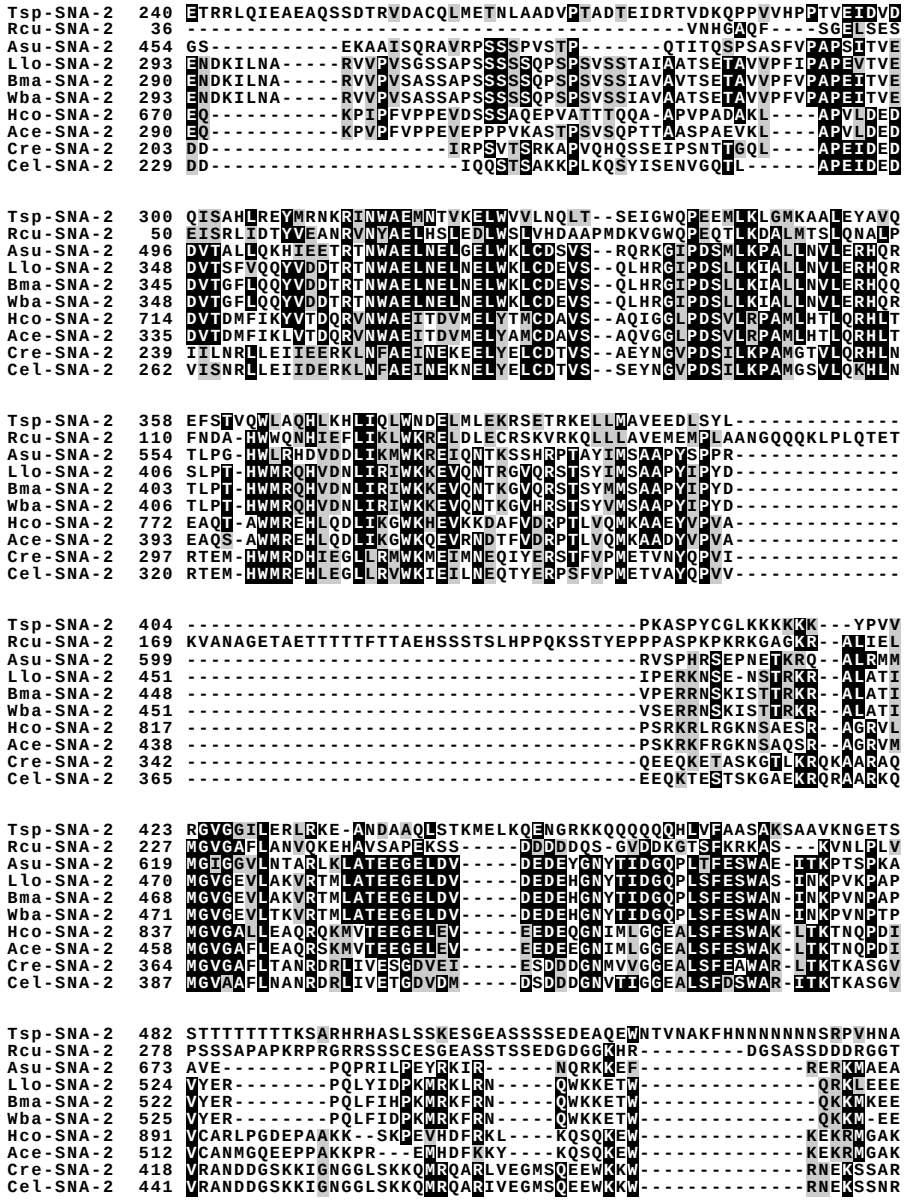


Figure S1 RT-PCR detection of polycistronic RNA. The exon-intron structures of the amplicons used to identify polycistronic RNAs are shown, with exons represented by boxes (shaded to identify the genes from which they are derived using the same colour coding that was used in Figures 1A and 3A). Cream coloured boxes represent the intercistronic regions. The positions of the SL trans-splice acceptor sites are indicated. Images of representative RT-PCR products produced from each pair of primers are shown to the right of each amplicon cartoon. Reactions were carried out in the presence or absence of reverse transcriptase, + or –, respectively. M indicates DNA size markers.


Tsp-SNA-2 1 ----------
Rcu-SNA-2 1 ----------
Asu-SNA-2 1 ----------MDVIMGEETTSDNPLAE
Llo-SNA-2 1 ----------
Bma-SNA-2 1 ----------
Wba-SNA-2 1 ----------
Hco-SNA-2 1 MGTPQMKGRGLRLCAIACMPTLNGEEVLVGGNYLVRGRRGCWRHATVLDKRRGEDGTLEL
Ace-SNA-2 1 ----------
Cre-SNA-2 1 ----------
Cel-SNA-2 1 ----------
Tsp-SNA-2 1 ----------
Rcu-SNA-2 1 ----------
Asu-SNA-2 18 NMQTDENVLRSNVVESRSRQRSPSAARDDPVDV----------SPTTAHSEIPLNV
Llo-SNA-2 1 ----------MDDGTVLEVRI--
Bma-SNA-2 1 ----------MDDGTVLEVRI--
Wba-SNA-2 1 ----------MDDGTVLEVRI--
Hco-SNA-2 61 YLHFEGDDRRLDHWLDSSRLKFRNQPSKRITSAYEMREKRIRRAKESSPNSASLDYAEIL
Ace-SNA-2 1 ----------
Cre-SNA-2 1 ----------
Cel-SNA-2 1 ----------
Tsp-SNA-2 1 ----------
Rcu-SNA-2 1 ----------
Asu-SNA-2 64 SETSMTAPPISRGPMTPPGEPLVSTPGSSTPPP----------
Llo-SNA-2 12 ----------
Bma-SNA-2 12 ----------
Wba-SNA-2 12 ----------
Hco-SNA-2 121 EEQHKEVTKVKYIEVVRYGEFEMDTWYVSPYPDEYGKERYLFICDKCFLYMRQERAFKMH
Ace-SNA-2 1 ----------
Cre-SNA-2 1 ----------
Cel-SNA-2 1 ----------
Tsp-SNA-2 1 ----------
Rcu-SNA-2 1 ----------
Asu-SNA-2 97 ----------
Llo-SNA-2 12 ----------
Bma-SNA-2 12 ----------
Wba-SNA-2 12 ----------
Hco-SNA-2 181 LSSCTAKHPPGREIYVDYSENIAVYEVDGEKEKLFCQCLCLFAKLFMDHKTIYFDVTTFL
Ace-SNA-2 1 ----------
Cre-SNA-2 1 ----------
Cel-SNA-2 1 ----------
Tsp-SNA-2 1 ----------
Rcu-SNA-2 1 ----------
Asu-SNA-2 97 ----------CAIVPAESSKSAAYPPQGVDSAAMQQNSIFVAPTRTS
Llo-SNA-2 12 ----------NGEINGDAACL----------
Bma-SNA-2 12 ----------NGEINGDTACL----------
Wba-SNA-2 12 ----------NGEINGDAACL----------
Hco-SNA-2 241 FYVVCQLRESGEPRPVGYFSKERTSPDGHNLSCLLVFPAFQRQGFGSFLIQLSYELSRRE
Ace-SNA-2 1 ----------MVSCILYLLEFRR----------
Cre-SNA-2 1 ----------
Cel-SNA-2 1 ----------
Tsp-SNA-2 1 ----------
Rcu-SNA-2 1 ----------
Asu-SNA-2 134 GPATPPEGPGPDASP----------
Llo-SNA-2 23 ----------
Bma-SNA-2 23 ----------
Wba-SNA-2 23 ----------
Hco-SNA-2 301 GIQGSPEKPLSDLGAASFRHYWAYLIVEYLSGFMDTAWIRVSELAKTLGMQAEDVVDTLH
Ace-SNA-2 14 ----------
Cre-SNA-2 1 ----------
Cel-SNA-2 1 ----------


Tsp-SNA-2 1
Rcu-SNA-2 1
Asu-SNA-2 149 PISSSPHSTTIPQSYSASPAASLSTALVSPQ
Llo-SNA-2 23 AKLQNTA
Bma-SNA-2 23 AKLQNAA
Wba-SNA-2 23 AKLQNTA
Hco-SNA-2 361 WLRICDPAVMSDAPDDIELWVHVYIKKLDSLRTKAAKPPLMDDKEHEENVQENGSHDTPE
Ace-SNA-2 14 IINKEAMEDQDKEDHVQQNGSHNSTE
Cre-SNA-2 1
Cel-SNA-2 1 MI
Tsp-SNA-2 1 MVLAKSKDSDTA PSVSLARDVSSDDYDVSNVAG
Rcu-SNA-2 1
Asu-SNA-2 180 SSILSSSTAAASNTTQVISAARVAEKVTPHRDSECDSNGIELNKTSDPIDTSAKGEKIIE
Llo-SNA-2 30 ESISSPSTPGVETSSSLSDTT--ITLIRKELEQENFSAKPKINDQSHAENFIKSNEQASD
Bma-SNA-2 30 ESISSPSTPGIETSSSLSDTA--ISLMRKEFEHENCAAKPKSNDQIHVENCVKSNEQVL-
Wba-SNA-2 30 ESISSPSTPGIETSSSLSDTA--ISLMRKEFEHENCAAKPKPNDQIHVENCVKSSEQVSD
Hco-SNA-2 421 ESITAPLVNGTAIDSTAIDSG------SEAVTEGSSAGPASTAETEIVKDKSTPSEVNE
Ace-SNA-2 40 EPAVVSSVNGTLAENTATDSG-------IETAAEGSSREEAAVETGNPATETAEATPCSS
Cre-SNA-2 1 ---MSEEVNPVIEQGGEEETT
Cel-SNA-2 3 ELNPVKEYENVDTQSDAFDSV ENVAE
Tsp-SNA-2 34 QAQQADNEKEVRVDGLIYPCD-MQLPVLSEALYACERFACCFEK-NSNALEQEKIGYVSL
Rcu-SNA-2 1
Asu-SNA-2 240 SREVHDAKRKFRVVGLTFPGKWIHHEKLMEIFVKSESFHMWFDK--PEKAGGVKSGGVSV
Llo-SNA-2 88 SRDSSDSRRKFRVLGLTFPGKWMTSEKFWDLLEKSESFHIWFDK---TTKGGIKHGGLSA
Bma-SNA-2 87 --DSNDSKRKFRVLGLTFPGKWMTSEKFWDLLEKSESFHMWFDK---TTKGGTKCGGLST
Wba-SNA-2 88 SRDLNDSKRKFRVLGLTFPGKWMTSANFWDLLEKSESFHMWFDK---TTKDGTKCGGLST
Hco-SNA-2 474 EPELDN--RTFTVTGLKHPGGWRTHSEIINMLARCDSFIIRFGR--ALKQKGPKPGSISF
Ace-SNA-2 93 ESPDVDN-RTFTVTGLKHPGAWRTHSEIINMLARCESFNIKFER--GAKQKGPKPGYISF
Cre-SNA-2 19 ----------VEPAASLVSVNFQNLLNKSIDFDAQFSRPENSNERQIKTGRIEI
Cel-SNA-2 29 DKINDKSLRTIVVQNLKHPGGWLRNRQFQMLLNRCVIFDVSFSRPEDTDEKQIKYGKVEF
Tsp-SNA-2 92 YFDCLLKA-EAFCSAMKESVQRNEIK GIVTAKVVNLSAEELPVN
Rcu-SNA-2 1
Asu-SNA-2 298 VFENADEAKAAFASVQKMSLDGASFKLQASRHFYGGGTLGGPSTIGSRSLTAPSLPFRVD
Llo-SNA-2 145 VYETPEEAKLAFAAFHKMTFDGNPLKLQASKLFY-----D--PVPNSRIVQ-PPMPFKIN
Bma-SNA-2 142 VYDTPEEAKLAFAAFHKMTFDGNPLKLQASKPFY-----D--LTPNSRTIQ-PPMPFKIS
Wba-SNA-2 145 VYDTPEEAKLAFAAFHKMTFDGNPLKLQASKPFY-----D--LTPNSRTIQ-PPMPFKIN
Hco-SNA-2 530 TFESISSARESFTQAQKMRVDGQAVKVEASPAFF-----A-PAACRSRPF------FKIP
Ace-SNA-2 150 TFESISSAREAYTQAQKMRVDGQAVKVEACPAFF------APAACRSRPF------FKIP
Cre-SNA-2 63 LFSSIDAARDAYTSLQKMVIDGFRPEVLVDPRFF-----SVETTTQKRQF------FEIA
Cel-SNA-2 89 VFSTVAGAREAYSTIQKMIIDGIRPEALVDPQFF-----SVETFQSSRQF------FDIS
Tsp-SNA-2 135 VDG----------LLAAANESLIRDIFAEYNVQEIANLSDVDSTTGTK--TATLKFSSE
Rcu-SNA-2 1 VKA--ETETISIKKP
Asu-SNA-2 358 VSEADSPRRLYALHLTPSTDQTLLSSILGSETIESI-RLENDPIVPSER--QAEIVFRTA
Llo-SNA-2 197 MSDPDASRRLYALHLSTATDQTLLSSIFGSECIETI-QLEADPFIASEK--QAEIVFHTV
Bma-SNA-2 194 LSDPDASRRLYALHLSSATDQTVLSSIFGSECIETV-QLEVDPFIASEK--QAEIVFHTV
Wba-SNA-2 197 LSDPDVSRRLYALHLSSATDQTVLSSIFGSECIETV-HLEVDPFIASEK--QAEIVFHTV
Hco-SNA-2 578 VDEETQKRTIYALDLPTSAERNLLNSIFEADHIE---KLTFLPLRNQHK--QAEVVMYSE
Ace-SNA-2 198 TEEEVRKRTIYALDLPTSAEQNLLNSIFESDHIE---RITFLPLRSQHK--QAEVLMHTG
Cre-SNA-2 112 DDS----RLVFLLDAPRSIDEHMIEHFFNGEAPI---HFQTLPMASENGLFQVEVLLSNG
Cel-SNA-2 138 NDT----RLVFLLDAPKSIDDHMISYFFEGKTPI---HFQTLPMPSSNGLFQVEVLLSDK
Tsp-SNA-2 182 QDAENAMLELND-FE-LDDGVCSSSLKISFLTPENNNNNNVRLSTTEFVAQNEKAVCVDD
Rcu-SNA-2 13 ES----------YKNKNEENVRSVDDLTKLEPQ
Asu-SNA-2 415 EEANDARIELADGFE-IDEGDRQSTMML--LTAE----------EYLMHMKS
Llo-SNA-2 254 QEANDARVELADGFE-IDEGDRQLTMRL--LTMD----------EYISYMKT
Bma-SNA-2 251 QEANDARMELADGFE-IDEGDRQLTMRL--LTAD----------EYISYMKT
Wba-SNA-2 254 QEANDARMELADGFE-IDEGDRQLTMRL--LTAD----------EYISYMKT
Hco-SNA-2 633 EQAEAARSE--DGFE-LDDGQQQSVLRI--LKPA----------EYAKFVEE
Ace-SNA-2 253 EQADAARSE--DGFE-LDDGQQQSVLRI--LTPA----------DYAAFVAE
Cre-SNA-2 165 EAAEKVLGGESK-FK-MHDGENECIVTL--LSPR----------EYALYSKM
Cel-SNA-2 191 ETAEKVLAGPSK-FS-MSDEDNECTVTL--LSPR----------EYSLYSKM


Tsp-SNA-2 240 ETRRLQIEAEAQSSDTRVDACQLMETNLAADVPTADTEIDRTVDKQPPVVHPPTVEIDVD
Rcu-SNA-2 36 ----------VNHGAQF----SGELSES
Asu-SNA-2 454 GS----------EKAAISQRAVRPSSSPVSTP-------QTITQSPSASFVPAPSITVE
Llo-SNA-2 293 ENDKILNA-----RVVPVSGSSAPSSSSSQPSPSVSSTAIAATSETAVVPFIPAPEVTVE
Bma-SNA-2 290 ENDKILNA-----RVVPVSASSAPSSSSSQPSPSVSSIAVAVTSETAVVPFVPAPEITVE
Wba-SNA-2 293 ENDKILNA-----RVVPVSASSAPSSSSSQPSPSVSSIAVAATSETAVVPFVPAPEITVE
Hco-SNA-2 670 EQ----------KPIPFVPPEVDSSSAQEPVATTTQQA-APVPADAKL----APVLDED
Ace-SNA-2 290 EQ----------KPVPFVPPEVEPPPVKASTPSVSQPTTAASPAEVKL----APVLDED
Cre-SNA-2 203 DD----------IRPSVTSRKAPVQHQSSEIPSNTTGQL----APEIDED
Cel-SNA-2 229 DD----------IQQSTSAKKPLKQSYISENVGQTL------APEIDED
Tsp-SNA-2 300 QISAHLREYMRNKRINWAEMNTVKELWVVLNQLT--SEIGWQPEEMLKLGMKAALEYAVQ
Rcu-SNA-2 50 EISRLIDTYVEANRVNYAELHSLEDLWSLVHDAAPMDKVGWQPEQTLKDALMTSLQNALP
Asu-SNA-2 496 DVTALLQKHIEETRTNWAELNELGELWKLCDSVS--RQRKGIPDSMLKPALLNVLERHQR
Llo-SNA-2 348 DVTSFVQQYVDDTRTNWAELNELNELWKLCDEVS--QLHRGIPDSLLKIALLNVLERHQR
Bma-SNA-2 345 DVTGFLQQYVDDTRTNWAELNELNELWKLCDEVS--QLHRGIPDSLLKIALLNVLERHQQ
Wba-SNA-2 348 DVTGFLQQYVDDTRTNWAELNELNELWKLCDEVS--QLHRGIPDSLLKIALLNVLERHQR
Hco-SNA-2 714 DVTDMFIKYVTDQRVNWAEITDVMELYTMCDAVS--AQIGGLPDSVLRPAMLHTLQRHLT
Ace-SNA-2 335 DVTDMFIKLVTDQRVNWAEITDVMELYAMCDAVS--AQVGGLPDSVLRPAMLHTLQRHLT
Cre-SNA-2 239 IILNRLLEIIEERKLNFAEINEKEELYELCDTVS--AEYNGVPDSILKPAMGTVLQRHLN
Cel-SNA-2 262 VISNRLLEIIDERKLNFAEINEKNELYELCDTVS--SEYNGVPDSILKPAMGSVLQKHLN
Tsp-SNA-2 358 EFSTVQWLAQHLKHLIQLWNDELMLEKRSETRKELLMAVEEDLSYL--------------
Rcu-SNA-2 110 FNDA-HWWQNHIEFLIKLWKRELDLECRSKVRKQLLLAVEMEMPLAANGQQQKLPLQTET
Asu-SNA-2 554 TLPG-HWLRHDVDDLIKMWKREIQNTKSSHRPTAYIMSAAPYSPPR--------------
Llo-SNA-2 406 SLPT-HWMRQHVDNLIRIWKKEVQNTRGVQRSTSYIMSAAPYIPYD--------------
Bma-SNA-2 403 TLPT-HWMRQHVDNLIRIWKKEVQNTKGVQRSTSYMMSAAPYIPYD--------------
Wba-SNA-2 406 TLPT-HWMRQHVDNLIRIWKKEVQNTKGVHRSTSYVMSAAPYIPYD--------------
Hco-SNA-2 772 EAQT-AWMREHLQDLIKGWKHEVKKDAFVDRPTLVQMKAAEYVPVA--------------
Ace-SNA-2 393 EAQS-AWMREHLQDLIKGWKQEVRNDTFVDRPTLVQMKAADYVPVA--------------
Cre-SNA-2 297 RTEM-HWMRDHIEGLLRMWKMEIMNEQIYERSTFVPMETVNYQPVI--------------
Cel-SNA-2 320 RTEM-HWMREHLEGLLRVWKIEILNEQTYERPSFVPMETVAYQPVV--------------
Tsp-SNA-2 404 ----------PKASPYCGLKKKKKKK---YPVV
Rcu-SNA-2 169 KVANAGETAETTTTTFTTAEHSSSTSLHPPQKSSTYEPPPASPKPKRKGAGKR--ALIEL
Asu-SNA-2 599 ----------RVSPHRSEPNETKRQ--ALRMM
Llo-SNA-2 451 ----------IPERKNSE-NSTRKR--ALATI
Bma-SNA-2 448 ----------VPERRNSKISTTRKR--ALATI
Wba-SNA-2 451 ----------VSERRNSKISTTRKR--ALATI
Hco-SNA-2 817 ----------PSRKRLRGKNSAESR--AGRVL
Ace-SNA-2 438 ----------PSKRKFRGKNSAQSR--AGRVM
Cre-SNA-2 342 ----------QEEQKETASKGTLKRQKAARAQ
Cel-SNA-2 365 ----------EEQKTESTSKGAEKRQRAARKQ
Tsp-SNA-2 423 RGVGGILERLRKE-ANDAAQLSTKMELKQENGRKKQQQQQQHLVFAASAKSAAVKNGETS
Rcu-SNA-2 227 MGVGAFLANVQKEHAVSAPEKSS-----DDDDDQS-GVDDKGTSFKRKAS---KVNLPLV
Asu-SNA-2 619 MGIGGVLNTARLKLATEEGELDV-----DEDEYGNYTIDGQPLTFESWAE-ITKPTSPKA
Llo-SNA-2 470 MGVGEVLAKVRTMLATEEGELDV-----DEDEHGNYTIDGQPLSFESWAS-INKPVKPAP
Bma-SNA-2 468 MGVGEVLAKVRTMLATEEGELDV-----DEDEHGNYTIDGQPLSFESWAN-INKPVNPAP
Wba-SNA-2 471 MGVGEVLTKVRTMLATEEGELDV-----DEDEHGNYTIDGQPLSFESWAN-INKPVNPTP
Hco-SNA-2 837 MGVGALLEAQRQKMVTEEGELEV-----EEDEQGNIMLGGEALSFESWAK-LTKTNQPDI
Ace-SNA-2 458 MGVGAFLEAQRSKMVTEEGELEV-----EEDEEGNIMLGGEALSFESWAK-LTKTNQPDI
Cre-SNA-2 364 MGVGAFLTANRDRLIVESGDVEI-----ESDDDGNMVVGGEALSFEAWAR-LTKTKASGV
Cel-SNA-2 387 MGVAAFLNANRDRLIVETGDVDM-----DSDDDGNVTIGGEALSFDSWAR-ITKTKASGV
Tsp-SNA-2 482 STTTTTTTTKSARHRHASLSSKESGEASSSSEDEAQEWNTVNAKFHNNNNNNNSRPVHNA
Rcu-SNA-2 278 PSSSAPAPKRPRGRRSSSCESGEASSTSSEDGDGGKHR---------DGSASSDDDRGGT
Asu-SNA-2 673 AVE---------PQPRILPEYRKIR------NQRKKEF--------------RERKMAEA
Llo-SNA-2 524 VYER--------PQLYIDPKMRKLRN-----QWKKETW--------------QRKLEEE
Bma-SNA-2 522 VYER--------PQLFIHPKMRKFRN-----QWKKETW--------------QKKMKEE
Wba-SNA-2 525 VYER--------PQLFIDPKMRKFRN-----QWKKETW--------------QKKM-EE
Hco-SNA-2 891 VCARLPGDEPAAKK--SKPEVHDFRKL-----KQSQKEW--------------KEKRMGAK
Ace-SNA-2 512 VCANMGQEEPPAKKPR---ENHDFKKY----KQSQKEW--------------KEKRMGAK
Cre-SNA-2 418 VRANDDGSKKIGNGGLSKKQMRQARLVEGMSQEEWKKW--------------RNEKSSAR
Cel-SNA-2 441 VRANDDGSKKIGNGGLSKKQMRQARIVEGMSQEEWKKW--------------RNEKSSNR

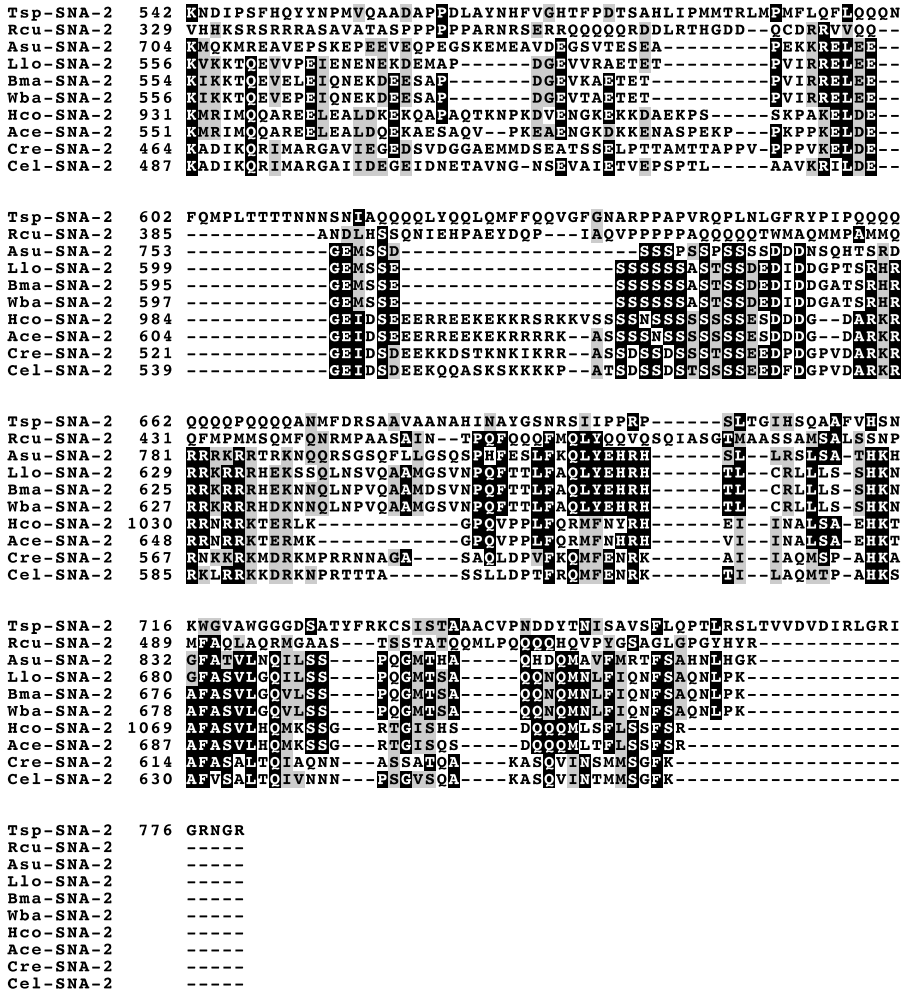


Figure S2 Multiple sequence alignment of SNA-2 homologues. Residues conserved in 50% or more of the homologues are shaded black (identities) or grey (similarities). Species prefixes used as follows: Tsp - *T. spiralis*; Rcu - *R. culicivorax*; Asu - *A. suum*; Llo - *L. loa*; Bma - *B. malayi*; Wba - *W. bancrofti*; Hco - *H. contortus*; Ace - *A. ceylanicum*; Cre - *C. remanei*; Cel - *C. elegans*.

Table S1 Operonic status of *T. spiralis* genes that produce SL *trans*-spliced transcripts.

T. spiralis Gene[1]	*Tsp*-SL Number	EST	*T. spiralis* Gene Operon Status[2]	*C. elegans* homologue	*C. elegans* Gene Operon Status	*C. elegans* Operon Number
Tsp_04323	3	ES566285	Upstream	*ZK829.7*	Upstream	4484
Tsp_04801	1	ES565008	Upstream	*prmt-1*	Upstream	5508
Tsp_07643	3	ES272890	Upstream	*Y57G11C.22*	Upstream	4661
Tsp_09747	3	ES565435	Upstream	*fkb-2*	Upstream	1903
Tsp_10562	3	ES566995	Upstream	*K10C3.4*	Upstream	1566
Tsp_10928	3	ES570931	Upstream	*C17E4.6*	Upstream	1544
Tsp_08576	2	ES568840	Upstream	*ung-1*	Upstream	3746
Tsp_00718	5	ES570675	Upstream	*ufd-1*	Downstream	4512
Tsp_01411	2	ES560816	Upstream	*C16A3.4*	Downstream	3364
Tsp_02079	3	ES570654	Upstream	*idi-1*	Downstream	3480
Tsp_06176	12	ES565259	Upstream	*T04G9.4*	Downstream	X147
Tsp_06763	4	BG353375	Upstream	*C18E3.5*	Downstream	1152
Tsp_08492	4	ES567181	Upstream	*wwp-1*	Downstream	1934
Tsp_08574	3	BQ692442	Upstream	*vps-28*	Downstream	1708
Tsp_08643	3	ES565814	Upstream	*pdha-1*	Downstream	2336
Tsp_11146	3	ES566810	Upstream	*hpo-11*	Downstream	1368
Tsp_09859	3	ES570398	Upstream	*gpdh-2*	Downstream	3795
Tsp_12645	12	ES567036	Upstream	*acs-16*	Non-Operonic	
Tsp_00181	2	ES569343	Upstream	*Y53F4B.42*	Non-Operonic	
Tsp_01518	2	ES563585	Upstream	*tomm-1*	Non-Operonic	
Tsp_02629	2	ES569407	Upstream	*tsp-11*	Non-Operonic	
Tsp_04773	3	ES565760	Upstream	*apd-3*	Non-Operonic	
Tsp_05351	1	ES565863	Upstream	*smn-1*	Non-Operonic	
Tsp_05770	2	ES570548	Upstream	*jnk-1*	Non-Operonic	
Tsp_09338	1	ES567564	Upstream	*rpl-32*	Non-Operonic	
Tsp_09639	2	ES568996	Upstream	*aldo-1*	Non-Operonic	
Tsp_10534	3	ES570566	Upstream	*skr-1*	Non-Operonic	
Tsp_11514	3	ES568803	Upstream	*rps-9*	Non-Operonic	
Tsp_12763	3	ES565626	Upstream	*cyn-7*	Non-Operonic	
Tsp_13841	2	ES273211	Upstream	*Y37D8A.17*	Non-Operonic	
Tsp_00182	4	ES570621	Downstream	*swd-2.2*	Upstream	4248
Tsp_00331	8	ES569101	Downstream	*ubl-1*	Upstream	3088
Tsp_00685	2	ES568435	Downstream	*mrps-17*	Upstream	3372
Tsp_00837	12	ES568088	Downstream	*rrbs-1*	Upstream	5416
Tsp_03905	2	ES563877	Downstream	*F53F4.10*	Upstream	5392
Tsp_04143	10	ES565871	Downstream	*C06A8.2*	Upstream	2731
Tsp_08528	2	ES569156	Downstream	*Y55F3AM.9*	Upstream	4642
Tsp_11545	9	ES564614	Downstream	*atp-5*	Upstream	5272
Tsp_00424	1	ES562658	Downstream	*F44E7.4*	Downstream	5565
Tsp_00734	2	ES560968	Downstream	*F54C8.7*	Downstream	3584
Tsp_02509	2	ES569909	Downstream	*ard-1*	Downstream	4412

Tsp_02873	9	ES569583	Downstream	*eif-3F*	Downstream	2424
Tsp_03622	12	ES565661	Downstream	*clpf-1*	Downstream	3108
Tsp_05308	2	ES562011	Downstream	*T09B4.9*	Downstream	1304
Tsp_06507	2	ES561596	Downstream	*raga-1*	Downstream	2528
Tsp_06818	12	ES563737	Downstream	*Y62E10A.10*	Downstream	4544
Tsp_08736	12	ES567003	Downstream	*lars-2*	Downstream	1396
Tsp_11177	3	ES563835	Downstream	*rlbp-1*	Downstream	1416
Tsp_08003	4	ES566741	Downstream	*ins-1*	Non-Operonic	
Tsp_00064	11	ES564912	Downstream	*mlp-1*	Non-Operonic	
Tsp_00608	1	ES568215	Downstream	*ppt-1*	Non-Operonic	
Tsp_00609	1	ES564896	Downstream	*mdt-19*	Non-Operonic	
Tsp_00958	2	ES564918	Downstream	*vha-8*	Non-Operonic	
Tsp_01274	2	ES564040	Downstream	*asg-1*	Non-Operonic	
Tsp_03114	3	BG353703	Downstream	*pdhb-1*	Non-Operonic	
Tsp_03857	3	ES570432	Downstream	*exc-7*	Non-Operonic	
Tsp_06181	3	ES569208	Downstream	*idh-2*	Non-Operonic	
Tsp_06812	1	ES564176	Downstream	*B0035.1*	Non-Operonic	
Tsp_06870	2	ES273592	Downstream	*rpl-33*	Non-Operonic	
Tsp_07223	4	ES566475	Downstream	*amph-1*	Non-Operonic	
Tsp_07289	3	ES273487	Downstream	*M153.2*	Non-Operonic	
Tsp_08373	3	ES565720	Downstream	*dyn-1*	Non-Operonic	
Tsp_11596	4	ES565698	Downstream	*cpl-1*	Non-Operonic	
Tsp_12613	2	ES565903	Downstream	*lip-1*	Non-Operonic	
Tsp_14129	10	ES564415	Downstream	*ace-1*	Non-Operonic	
Tsp_01343	3	ES273458	Ambiguous	*fnta-1*	Upstream	4008
Tsp_07410	3	ES569154	Ambiguous	*ent-7*	Upstream	1542
Tsp_11644	1	BG732210	Ambiguous	*rpb-7*	Upstream	1924
Tsp_00266	2	ES570691	Ambiguous	*C10H11.8*	Downstream	1780
Tsp_00680	2	ES561947	Ambiguous	*arl-5*	Downstream	3636
Tsp_01365	3	ES569529	Ambiguous	*hpl-2*	Downstream	3701
Tsp_02416	4	BG353715	Ambiguous	*B0205.6*	Downstream	1644
Tsp_04477	3	ES565704	Ambiguous	*kin-10*	Downstream	1456
Tsp_05422	3	ES569822	Ambiguous	*C50D2.7*	Downstream	2012
Tsp_05915	2	ES565057	Ambiguous	*D2085.3*	Downstream	2380
Tsp_07955	3	ES563974	Ambiguous	*prmt-7*	Downstream	1528
Tsp_08270	3	ES570232	Ambiguous	*Y32H12A.4*	Downstream	3802
Tsp_09368	2	ES568561	Ambiguous	*lsm-8*	Downstream	4176
Tsp_10835	3	ES565721	Ambiguous	*C09G9.1*	Downstream	4310
Tsp_11447	12	ES569795	Ambiguous	*C05C10.3*	Downstream	2751
Tsp_10765	2	ES567336	Ambiguous	*T25G3.1*	Ambiguous	
Tsp_00026	2	ES563422	Ambiguous	*F52E4.5*	Non-Operonic	
Tsp_00262	2	ES567711	Ambiguous	*pitp-1*	Non-Operonic	
Tsp_00717	4	ES273452	Ambiguous	*T07D1.2*	Non-Operonic	
Tsp_01078	2	ES564803	Ambiguous	*F21F3.7*	Non-Operonic	
Tsp_01228	4	ES570063	Ambiguous	*cco-2*	Non-Operonic	
Tsp_01287	4	ES562009	Ambiguous	*mtrr-1*	Non-Operonic	

Tsp_01573	2	ES565382	Ambiguous	*F37A8.5*	Non-Operonic	
Tsp_01930	1	ES570094	Ambiguous	*ifa-1*	Non-Operonic	
Tsp_03184	7	BG353788	Ambiguous	*F53F4.3*	Non-Operonic	
Tsp_03217	13	BQ541428	Ambiguous	*cmd-1*	Non-Operonic	
Tsp_03724	2	ES569761	Ambiguous	*nrfl-1*	Non-Operonic	
Tsp_04029	2	ES273134	Ambiguous	*BE0003N10.1*	Non-Operonic	
Tsp_04084	1	ES568229	Ambiguous	*F52D10.2*	Non-Operonic	
Tsp_04141	7	ES565838	Ambiguous	*T18D3.9*	Non-Operonic	
Tsp_04152	3	ES566044	Ambiguous	*C15H9.5*	Non-Operonic	
Tsp_05204	3	ES567104	Ambiguous	*C32E8.5*	Non-Operonic	
Tsp_07239	1	ES564125	Ambiguous	*swt-1*	Non-Operonic	
Tsp_08116	5	BG354896	Ambiguous	*F35G12.7*	Non-Operonic	
Tsp_09247	3	ES563796	Ambiguous	*F31D4.2*	Non-Operonic	
Tsp_09527	4	ES566597	Ambiguous	*mtch-1*	Non-Operonic	
Tsp_10184	2	BG353445	Ambiguous	*ck-1*	Non-Operonic	
Tsp_10493	10	ES563996	Ambiguous	*cpz-1*	Non-Operonic	
Tsp_10907	2	ES570466	Ambiguous	*kin-19*	Non-Operonic	
Tsp_11214	4	ES563030	Ambiguous	*C32F10.8*	Non-Operonic	
Tsp_01291	3	ES568398	Non-Operonic	*rpl-6*	Upstream	3836
Tsp_01684	13	BG353107	Non-Operonic	*rsp-6*	Upstream	4252
Tsp_01730	2	ES563676	Non-Operonic	*ubxn-4*	Upstream	3520
Tsp_05024	3	ES566087	Non-Operonic	*C01G5.6*	Upstream	4180
Tsp_05180	2	ES568248	Non-Operonic	*lin-53*	Upstream	1552
Tsp_06564	4	ES566240	Non-Operonic	*npp-1*	Upstream	4340
Tsp_06972	2	ES564495	Non-Operonic	*rla-0*	Upstream	1624
Tsp_07066	2	ES569769	Non-Operonic	*ubh-4*	Upstream	2328
Tsp_08234	4	ES570299	Non-Operonic	*C01A2.5*	Upstream	1696
Tsp_08309	4	ES564762	Non-Operonic	*mnat-1*	Upstream	2070
Tsp_00048	3	ES570469	Non-Operonic	*exos-4.1*	Downstream	4532
Tsp_00176	7	BG732141	Non-Operonic	*ced-9*	Downstream	3666
Tsp_00183	3	ES560678	Non-Operonic	*cua-1*	Downstream	3792
Tsp_01361	4	ES272901	Non-Operonic	*mms-19*	Downstream	5533
Tsp_02307	3	ES562862	Non-Operonic	*Y50D4C.3*	Downstream	5559
Tsp_02829	1	BG322177	Non-Operonic	*gdi-1*	Downstream	4580
Tsp_04483	7	ES273272	Non-Operonic	*cts-1*	Downstream	3660
Tsp_06831	2	ES566483	Non-Operonic	*dpy-11*	Downstream	5120
Tsp_09589	2	ES566654	Non-Operonic	*F45F2.9*	Downstream	5196
Tsp_11375	4	ES561664	Non-Operonic	*rba-1*	Downstream	1552
Tsp_08545	7	ES273001	Non-Operonic	*drh-3*	Downstream	1979
Tsp_00870	3	ES569690	Non-Operonic	*smr-1*	Ambiguous	
Tsp_03751	3	ES569638	Non-Operonic	*rps-4*	Ambiguous	
Tsp_08048	3	ES563226	Non-Operonic	*B0491.5*	Ambiguous	
Tsp_09957	2	ES567293	Non-Operonic	*rpl-7*	Ambiguous	
Tsp_05009	3	ES569586	Non-Operonic	*ncs-7*	Non-Operonic	
Tsp_10058	3	ES569899	Non-Operonic	*cec-1*	Non-Operonic	
Tsp_00219	8	ES563383	Non-Operonic	*hcf-1*	Non-Operonic	

Tsp_00352	2	ES569781	Non-Operonic	*C16E9.2*	Non-Operonic	
Tsp_01044	1	ES273428	Non-Operonic	*kin-2*	Non-Operonic	
Tsp_01088	4	BG322014	Non-Operonic	*sli-1*	Non-Operonic	
Tsp_02150	1	ES273162	Non-Operonic	*flp-14*	Non-Operonic	
Tsp_03394	2	ES570427	Non-Operonic	*T03F1.12*	Non-Operonic	
Tsp_04026	1	ES561895	Non-Operonic	*pcyt-1*	Non-Operonic	
Tsp_05440	4	ES568782	Non-Operonic	*sulp-4*	Non-Operonic	
Tsp_05646	3	ES569812	Non-Operonic	*K09E9.1*	Non-Operonic	
Tsp_05819	4	ES570798	Non-Operonic	*C25F6.7*	Non-Operonic	
Tsp_05837	3	ES561164	Non-Operonic	*ZK795.1*	Non-Operonic	
Tsp_05990	1	ES273061	Non-Operonic	*ppm-1*	Non-Operonic	
Tsp_06158	12	ES567235	Non-Operonic	*B0272.3*	Non-Operonic	
Tsp_06305	11	ES566685	Non-Operonic	*ife-3*	Non-Operonic	
Tsp_06353	1	ES569025	Non-Operonic	*ppk-2*	Non-Operonic	
Tsp_07030	3	ES569609	Non-Operonic	*mecr-1*	Non-Operonic	
Tsp_07118	2	ES563420	Non-Operonic	*Y82E9BR.3*	Non-Operonic	
Tsp_07593	7	ES567967	Non-Operonic	*ZK809.3*	Non-Operonic	
Tsp_07622	12	ES563060	Non-Operonic	*F17A9.2*	Non-Operonic	
Tsp_07825	2	ES569035	Non-Operonic	*lipl-6*	Non-Operonic	
Tsp_07989	6	BG520790	Non-Operonic	*pck-1*	Non-Operonic	
Tsp_08023	1	ES563253	Non-Operonic	*Y54E10BR.2*	Non-Operonic	
Tsp_09739	3	ES273513	Non-Operonic	*rab-10*	Non-Operonic	
Tsp_10501	12	ES564714	Non-Operonic	*F58F12.1*	Non-Operonic	
Tsp_10907	2	ES561739	Non-Operonic	*lin-19*	Non-Operonic	
Tsp_11867	7	BG302133	Non-Operonic	*hmg-1.2*	Non-Operonic	
Tsp_12507	2	ES568668	Non-Operonic	*rps-14*	Non-Operonic	
No match	2	BG354854	Unknown	*dcn-1*	Upstream	3172
Ambiguous	2	BQ693036	Unknown	*icln-1*	Downstream	4324
Ambiguous	2	ES561812	Unknown	*T02C12.3*	Downstream	3148
Ambiguous	3	ES563178	Unknown	*pkc-1*	Downstream	5312
No match	2	ES568194	Unknown	*gst-1*	Downstream	3572
Ambiguous	3	ES566071	Unknown	*F53B1.8*	Non-Operonic	
Tsp_00916	1	ES564849	Unknown	*clc-3*	Non-Operonic	
Tsp_09335	4	ES566864	Unknown	*plp-1*	Non-Operonic	
No match	12	ES561871	Unknown	*glr-1*	Non-Operonic	
No match	3	BQ693326	Unknown	*K09E9.4*	Non-Operonic	
No match	2	ES567398	Unknown	*cdc-42*	Non-Operonic	

[1] In some cases we found the EST sequence matched to multiple genes (Ambiguous), or we failed to find a match in *T. spiralis* draft genome (No match) and we were thus unable to determine the operonic status.
[2] A number of genes are present on short contigs, so we were unable to determine its operonic status.

Table S2 The 5' ends of *T. spiralis and T. muris* mRNAs identified by 5' RACE. Initiation codons are underlined and identical residues present in all sequences are indicated (*). For the ends of non *trans*-spliced mRNAs the distances from the initiation codon position in the cDNA (with A in ATG set as 0) are indicated. 'n' indicates how many times the sequence was found.

T. spiralis mRNA 5' ends				
			n	GenBank
zgpa-1	genome	AACTGTACTACAGTTTGACAGTAATCCTATTCTGTGATGCAGTCGTGTAATTATTTGCTTCACGCGTATTTTTTGAAGATTAATTGAACTGTGTCACATTAATGTG		
	-205	----------GATGCAGTCGTGTAATTATTTGCTTCACGCGTATTTTTTGAAGATTAATTGAACTGTGTCACATTAATGTG	1	
	-214	----------CTATTCTGTGATGCAGTCGTGTAATTATTTGCTTCACGCGTATTTTTTGAAGATTAATTGAACTGTGTCACATTAATGTG	1	
	-179	----------ACGCGTATTTTTTGAAGATTAATTGAACTGTGTCACATTAATGTG	1	
	-192	----------AATTATTTGCTTCACGCGTATTTTTTGAAGATTAATTGAACTGTGTCACATTAATGTG	1	
	-229	----------AGTTTGACAGTAATCCTATTCTGTGATGCAGTCGTGTAATTATTTGCTTCACGCGTATTTTTTGAAGATTAATTGAACTGTGTCACATTAATGTG	1	
	-228	----------GTTTGACAGTAATCCTATTCTGTGATGCAGTCGTGTAATTATTTGCTTCACGCGTATTTTTTGAAGATTAATTGAACTGTGTCACATTAATGTG	1	KF442420
		**********		
dif-1	genome	TTGTTCGGCGAATTGTATACAGATAAATATGGTGGAAAATGATTCGAAATCGATCGATTTGGTGAAACCACGTCAAGATACGGATCCACTGAGAAATTTCTTAGC		
	Tsp-SL10/*dif-1*	GGTAATATTTACTGAATTCAAGATAAATATGGTGGAAAATGATTCGAAATCGATCGATTTGGTGAAACCACGTCAAGATACGGATCCACTGAGAAATTTCTTAGC	5	KF442421
		* * * **********		
aph-1	genome	ATCAATATGTATATTCTTTTTAGGATTATATTAATAATCGCCTTGAGAAGCATTCCGTCCGAGTACACGTTTCAACTATGGGATTTCAAGAATTCACAGGATAT		
	Tsp-SL11/*aph-1*	--TACCTTTGAACCCACTTCAAGGATTATATTAATAATCGCCTTGAGAAGCATTCCGTCCGAGTACACGTTTCAACTATGGGATTTCAAGAATTCACAGGATAT	1	KF442422
trans-spliced	*Tsp*-SL1/*aph-1*	-AGGTATTTACCAGATCTAAAAGGATTATATTAATAATCGCCTTGAGAAGCATTCCGTCCGAGTACACGTTTCAACTATGGGATTTCAAGAATTCACAGGATAT	2	KF442423
	Tsp-SL7/*aph-1*	-AACCTGCACGACTTGTTCGAAGGATTATATTAATAATCGCCTTGAGAAGCATTCCGTCCGAGTACACGTTTCAACTATGGGATTTCAAGAATTCACAGGATAT	2	KF442423
	Tsp-SL12/*aph-1*	ACGAATTTACCGTATTTGTCAAGGATTATATTAATAATCGCCTTGAGAAGCATTCCGTCCGAGTACACGTTTCAACTATGGGATTTCAAGAATTCACAGGATAT	2	KF442425
	Tsp-SL14/*aph-1*	ATACCGTTCAATTAATTTTGAAGGATTATATTAATAATCGCCTTGAGAAGCATTCCGTCCGAGTACACGTTTCAACTATGGGATTTCAAGAATTCACAGGATAT	1	KF442426
	Tsp-SL9/*aph-1*	--AGACGTGGTTATTTATTGAAGGATTATATTAATAATCGCCTTGAGAAGCATTCCGTCCGAGTACACGTTTCAACTATGGGATTTCAAGAATTCACAGGATAT	1	KF442427
		********** **********		
non-*trans*-spliced	genome	ATCTTACTGTAGCTTTATTTGAACGAAAACTGTTGAATGTTATTGTTGCTTTCTTTTTAGCTTTTTACTTTTTATGTAAAATCTGTGTTTGTGGTAAAAAGTCA		
	-201	----------GAACGAAAACTGTTGAATGTTATTGTTGCTTTCTTTTTAGCTTTTTACTTTTTATGTAAAATCTGTGTTTGTGGTAAAAAGTCA	2	
	-205	----------ATTTGAACGAAAACTGTTGAATGTTATTGTTGCTTTCTTTTTAGCTTTTTACTTTTTATGTAAAATCTGTGTTTGTGGTAAAAAGTCA	1	
	-211	----------GCTTTATTTGAACGAAAACTGTTGAATGTTATTGTTGCTTTCTTTTTAGCTTTTTACTTTTTATGTAAAATCTGTGTTTGTGGTAAAAAGTCA	1	KF442428
		**********		
cpt-2	genome	TCATTTTACTGTCTGCACATTATTGTTTACATAGTGTAAGAGCTGTTGTATTTTGAGAAACAAATTTGTGTGCTTGAATTGCTTGCATTTCTTTTAATTCGAACACATT		
	-204	----------ATTATTGTTTACATAGTGTAAGAGCTGTTGTATTTTGAGAAACAAATTTGTGTGCTTGAATTGCTTGCATTTCTTTTAATTCGAACACATT	9	KF442419
		**********		
nuaf-3	genome	TTATTATATCAATAATTTCCTAGTTTGGCGTAAATGTTGGCTATTAGACGAGCTTTACTTAAAAATCGCTGCTTGTTTAATCAAACTAATGCAATCAGTGGAAAG		
	Tsp-SL12/*nuaf-3*	ACGAATTTACCGTATTTGTCAAGTTTGGCGTAAATGTTGGCTATTAGACGAGCTTTACTTAAAAATCGCTGCTTGTTTAATCAAACTAATGCAATCAGTGGAAAG	7	KF442418
		* * ** ** * **********		

T. muris mRNA 5′ ends

Table S3 Primers used for the amplification of 5'RACE products.

mRNA 5' ends	Primers
Tsp-cpt-2	GGCAAGGCAGCTCTGTATCGGAA and Gene Racer 5' primer (CGACTGGAGCACGAGGACACTGA).
Tsp-nuaf-3	1st PCR: CACTTTCCATGATAATGCAGCCTGTGG and Gene Racer 5' primer. 2nd PCR: GCCAGTGATTCATTCCAAGCC and Gene Racer 5' nested primer (GGACACTGACATGGACTGAAGGAGTA).
Tsp-zgpa-1	1st PCR: AGTTCTGTCGCTTCCAACAACGCAT and Gene Racer 5' nested primer. 2nd PCR: CGCTTCCAACAACGCATCTATACTAC and Gene Racer 5' nested primer.
Tsp-dif-1	1st PCR: TGCTGCAGCCGAAGAAGTACAACGC and Gene Racer 5' nested primer. 2nd PCR: CCGAAGAAGTACAACGCAAACAGC and Gene Racer 5' nested primer.
Tsp-aph-1	1st PCR: ACTTCTTGTAAGAAGACAAACAAGAAAACGG and Gene Racer 5' nested primer. 2nd PCR: GATCGTGCATGATCACGCAGAGA and Gene Racer 5' nested primer.
Tmu-nuaf-3	1st PCR: GCCATGATGTCAACACTCAATC and Gene Racer 5' nested primer. 2nd PCR: CCGACGACAAATACGCCATTGG and Gene Racer 5' nested primer.
Tmu-zgpa-1	1st PCR: GGAATAGCTTAACCAACAGACCCTTT and Gene Racer 5' nested primer. 2nd PCR: AGTAAACAAGATACAAGCCACTCACC and Gene Racer 5' nested primer.
Tmu-dif-1	1st PCR: TTCGTCTGGGTGCCTTTGTTGGAG and Gene Racer 5' nested primer. 2nd PCR: GAAAATACAAGGCGAATAGCGGAGC and GR 5' nested primer.
Tmu-aph-1	1st PCR: ATGAGTGCGGTAGGCCAATAGTTCC and Gene Racer 5' nested primer. 2nd PCR: AGAAGGTTTAGCAGTGCAACTACGCC and Gene Racer 5' nested primer.

Table S4 Primers used for the detection of processing intermediates of polycistronic transcripts by PCR.

Transcripts	Primers
Tsp-cpt-2~nuaf-3	GAAGCATGTTCCAAGCAGCATTC
	GCCAGTGATTCATTCCAAGCC
Tsp-zgpa-1~dif-1	AAGCAAAATTGGCGAAAAGGACCGAAG
	TGCTGCAGCCGAAGAAGTACAACGC
Tsp-dif-1~aph-1	TGCTGGAGCTTTGTCAGGTATGATG
(1st PCR)	ACTTCTTGTAAGAAGACAAACAAGAAAACGG
Tsp-dif-1~aph-1 (2nd PCR)	CGGGCGAGAGGATCAAATGC
	GATCGTGCATGATCACGCAGAGA
Tmu-zgpa-1~dif-1 (1st PCR)	TTGTTCGCGGGCAAGGCTTAG
	TTCGTCTGGGTGCCTTTGTTGGAG
Tmu-zgpa-1~dif-1 ((2nd PCR)	TGATCGAAAGTGTAGAGCGGATACGG
	GAAAATACAAGGCGAATAGCGGAGC
Tmu-dif-1~aph-1 (1st PCR)	TGCTGGAGCTTTGTCAGGTATGATG
	ATGAGTGCGGTAGGCCAATAGTTCC
Tmu-dif-1~aph-1 (2nd PCR)	CGGGCGAGAGGATCAAATGC
	AGAAGGTTTAGCAGTGCAACTACGCC