
Operon and non-operon gene clusters in the *C. elegans* genome*

Thomas Blumenthal^{1§}, Paul Davis² and Alfonso Garrido-Lecca¹

¹Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, CO 80309 USA

²Wormbase, European Bioinformatics Institute, EMBL, Wellcome Trust Genome Campus, Hinxton, CB10 1SD, UK

Table of Contents

1. Trans-splicing and operons	2
2. SL2-type operons	2
3. Hybrid operons	3
4. SL2-type operons with long spacing	5
5. SL2-type operons with juxtaposed 3' end formation and trans-splice sites	6
6. SL2 trans-splicing at downstream exons	7
7. Alternative operons	8
8. SL1-type operons	9
9. Dicistronic mRNAs	11
10. Overlapping genes	11
11. Identification of operons	12
12. Regulation in and of operons	13
13. Genome architecture	14
14. Tables 1-8	14
15. References	19

Abstract

Nearly 15% of the ~20,000 *C. elegans* genes are contained in operons, multigene clusters controlled by a single promoter. The vast majority of these are of a type where the genes in the cluster are ~100 bp apart and the pre-mRNA is processed by 3' end formation accompanied by trans-splicing. A spliced leader, SL2, is

*Edited by Julie Ahringer Last revised September 2, 2014. Published April 28, 2015. This chapter should be cited as: Blumenthal T., Davis P., Garrido-Lecca A. Operon and non-operon gene clusters in the *C. elegans* genome (April 28, 2015), *WormBook*, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.175.1, <http://www.wormbook.org>.

Copyright: © 2015 Thomas Blumenthal, Paul Davis, Alfonso Garrido-Lecca. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

§To whom correspondence should be addressed. E-mail: tom.blumenthal@colorado.edu

specialized for operon processing. Here we summarize current knowledge on several variations on this theme including: (1) hybrid operons, which have additional promoters between genes; (2) operons with exceptionally long (> 1 kb) intercistronic regions; (3) operons with a second 3' end formation site close to the trans-splice site; (4) alternative operons, in which the exons are sometimes spliced as a single gene and sometimes as two genes; (5) SL1-type operons, which use SL1 instead of SL2 to trans-splice and in which there is no intercistronic space; (6) operons that make dicistronic mRNAs; and (7) non-operon gene clusters, in which either two genes use a single exon as the 3' end of one and the 5' end of the next, or the 3' UTR of one gene serves as the outtron of the next. Each of these variations is relatively infrequent, but together they show a remarkable variety of tight-linkage gene arrangements in the *C. elegans* genome.

1. Trans-splicing and operons

Operons are polycistronic clusters of genes transcribed from a promoter at the 5' end of the cluster. Although operons were considered to be absent from eukaryotic genomes, it is now clear that operons are present in the genomes of numerous eukaryotes (Spieth et al., 1993; Blumenthal et al., 2002; Guiliano and Blaxter, 2006). For example, the *Drosophila* genome contains >30 dicistronic clusters that make mRNAs that encode two different genes (Misra et al., 2002). Most operons in eukaryotes are of a different type in which the polycistronic pre-mRNA arising from the operon is co-transcriptionally processed by 3' end formation and spliced leader (SL) trans-splicing between the genes to make monocistronic mature mRNAs (Blumenthal, 2004). In SL trans-splicing, a short spliced leader RNA exon is spliced onto the 5' end of the pre-mRNA by conventional splicing mechanisms, providing a cap for the downstream mRNAs.

Nearly 15% of the ~20,000 protein-coding genes in the *C. elegans* genome are organized into ~1250 operons, tight clusters of two to eight genes (Allen et al., 2011). In most cases the genes are separated by an intercistronic region of ~100 bp from the polyA addition site of the upstream gene to the trans-splice site of the downstream gene. The polycistronic pre-mRNA is processed by coordinated 3' end formation of the upstream gene and trans-splicing at the 5' end of the downstream gene. This trans-splicing event involves a spliced leader, SL2, specialized for operon pre-mRNA processing.

However, SL1 trans-splicing is the more common kind of trans-splicing in *C. elegans*. Most SL1 trans-splicing occurs near the 5' ends of genes rather than downstream in operons. This removes the outtron, the RNA between the transcription start site and the first 3' splice site in the pre-mRNA. A rare type of operon, SL1-type, uses SL1 for trans-splicing a downstream gene. SL1-type operons are mechanistically quite interesting since they have no intercistronic sequence; polyadenylation of the upstream gene occurs right at the trans-splice site of the downstream gene (Williams et al., 1999). In these operons, 3' end formation may occur by SL1 trans-splicing of the downstream gene, resulting in a free 3' end upstream that may then be debranched and polyadenylated. Thus, in these cases, the same processing event at least sometimes may serve to create the 3' end of the upstream gene mRNA and the 5' end of the downstream gene mRNA.

Here we present an in-depth analysis of gene clusters in the *C. elegans* genome in an effort to determine whether they all represent true operons and whether there are alternative ways of processing operon pre-mRNAs. Several variations of SL2-type operon are considered. These include the relatively common hybrid operons, which have longer intercistronic regions that accommodate an extra promoter between the genes (Huang et al., 2007; Whittle et al., 2008), as well as several less common variations on the operon theme. For example, some operons have an extra polyadenylation signal (AAUAAA) near the trans-splice site that results in 3' end formation close to the site of trans-splicing. There are some operons with unusually long (> 1 KB) intercistronic regions (Morton and Blumenthal, 2011). Alternative operons are sometimes spliced as if they are a single gene (Jan et al., 2011; Morton and Blumenthal, 2011). The existence of a variation on the SL1-type operon, in which the upstream mRNA is discarded rather than being polyadenylated, is described. In this case, the expression of the upstream gene occurs from a dicistronic mRNA. The worm genome also contains occasional examples of operons that make dicistronic mRNAs, which are apparently translated in that form, much like the *Drosophila* operons. However, it is currently unknown how translation of the downstream cistron is initiated. Finally, two types of tightly linked non-operon gene clusters are identified. In one type, a single exon serves as the 3' end of an upstream gene as well as the 5' end of a downstream gene. In another type, the 3' UTR of an upstream gene serves as an outtron for SL1 trans-splicing of a downstream gene. These are not operons since the genes are not transcribed from the same promoter. In this paper, examples of each kind of gene cluster are described, and tables with lists of known examples of each type are included.

2. SL2-type operons

Allen and coworkers mapped SL1 and SL2 trans-splice sites from deep sequencing transcriptome data (Allen et al., 2011), and used this data to identify operons supported by SL2 trans-splicing (Supplemental File 3 in Allen et al., 2011). Almost all gene clusters in the *C. elegans* genome are of the type exemplified in Figure 1, in this case a four-gene operon. SL2-type operons range from two to eight genes expressed from a single promoter at the 5' end of the cluster. We have performed chromatin immunoprecipitation experiments with an Affymetrix tiling array (ChIP/chip) for H3K9 acetylation (H3K9ac), an epigenetic mark of active promoters. The data show that for most operons there is only a single H3K9ac peak, found at the 5' end of the cluster. An example is shown in Figure 1. The genes are very close together, usually ~100 bp separating the 3' end of the upstream gene (the site of polyA addition) and the trans-splice site of the downstream gene and virtually all trans-splicing is to SL2. Furthermore, Allen et al. demonstrated a strong relationship between the intergenic distance and the percent of trans-splicing to SL2: shorter distance correlates with increased SL2. There are around 1250 operons of this kind, a subset of which are hybrid operons considered in the next section.

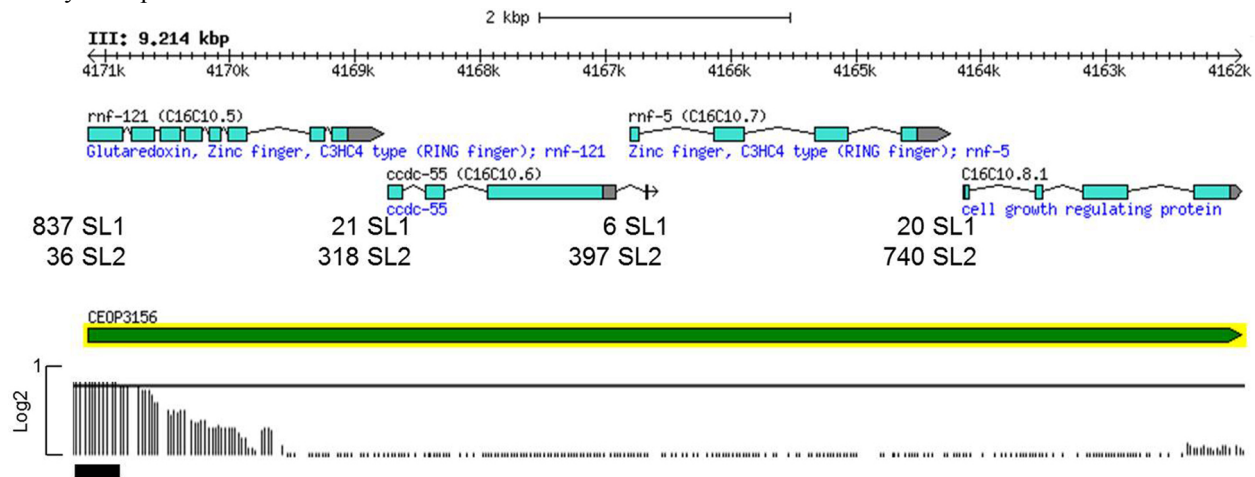


Figure 1. A typical SL2-type operon. The figure shows a four-gene operon with exons shown as colored boxes and introns as angled lines. The direction of transcription is from left to right, as indicated by the arrows on the 3' UTRs. The green bar denotes the extent of the operon. The pre-mRNA is processed to yield multiple mature mRNAs. The first gene in the operon *rnf-121* is trans-spliced mostly to SL1, whereas the three downstream genes are trans-spliced mostly to SL2 (data from Allen et al. 2011). In each case, the 3' cleavage sites are ~100 bp upstream of the SL2-specific trans-splice sites. This is a download from Wormbase. There are >1200 documented operons of this type in the *C. elegans* genome (Allen et al., 2011). Here and throughout, the number of SL1 and SL2 reads shown is data from the deep sequencing project reported in Allen et al. (2011). Sequencing was performed on RNA isolated from several different worm stages and the data pooled. Below the operon bar, H3K9ac ChIP/chip performed on young adult worms, which marks active promoters, is shown. The length of the vertical lines is proportional to the intensity of the hybridization signal corresponding to the indicated genomic position. The horizontal black bar marks a peak called by the data analysis program.

3. Hybrid operons

Many operons have the characteristics described above, but have another important feature: in addition to the upstream operon promoter, they contain a second promoter within the operon. This has been demonstrated in several ways. Transgenic constructs containing only the intergenic DNA fused to a reporter gene were found to be expressed in patterns characteristic of the endogenous genes (Huang et al., 2007). When analyzed by ChIP for proteins previously shown to peak at promoters—including the variant histone HTZ, the histone modification H3K4Me3, and RNA polymerase II—many operons were found to contain peaks between genes (Whittle et al., 2008; Baugh et al., 2009). Furthermore, some operons have a peak of *ser-5* phosphorylated RNA polymerase, which typically occurs near 5' ends of genes, between genes in operons (A. Garrido-Lecca and T. Blumenthal, unpublished observations). Finally, SL1 trans-splicing, generally associated with sites close to promoters, occurs preferentially at some downstream operon genes with long intergenic regions (Allen et al., 2011). Further validation of the relationship between an internal promoter and SL1 trans-splicing was provided by analysis of two deletion strains: a deletion within a hybrid operon intergenic region dramatically reduced SL1 usage at the downstream trans-splice site, while increasing the SL2 trans-splicing at this site. In contrast, a deletion at the 5' end of a different hybrid operon dramatically reduced the SL2 trans-splicing at a downstream gene, leaving the SL1 trans-splicing unchanged (Allen et al., 2011).

The intercistronic region from a typical hybrid operon is depicted in Figure 2. The downstream gene in this operon receives significant levels of SL1, presumably from the internal promoter, and of SL2, presumably from transcripts originating from the promoter at the 5' end of the operon. Note also the >500 bp intercistronic region. The ChIP/chip peak of H3K9ac between the genes indicates the location of an internal promoter (Figure 2).

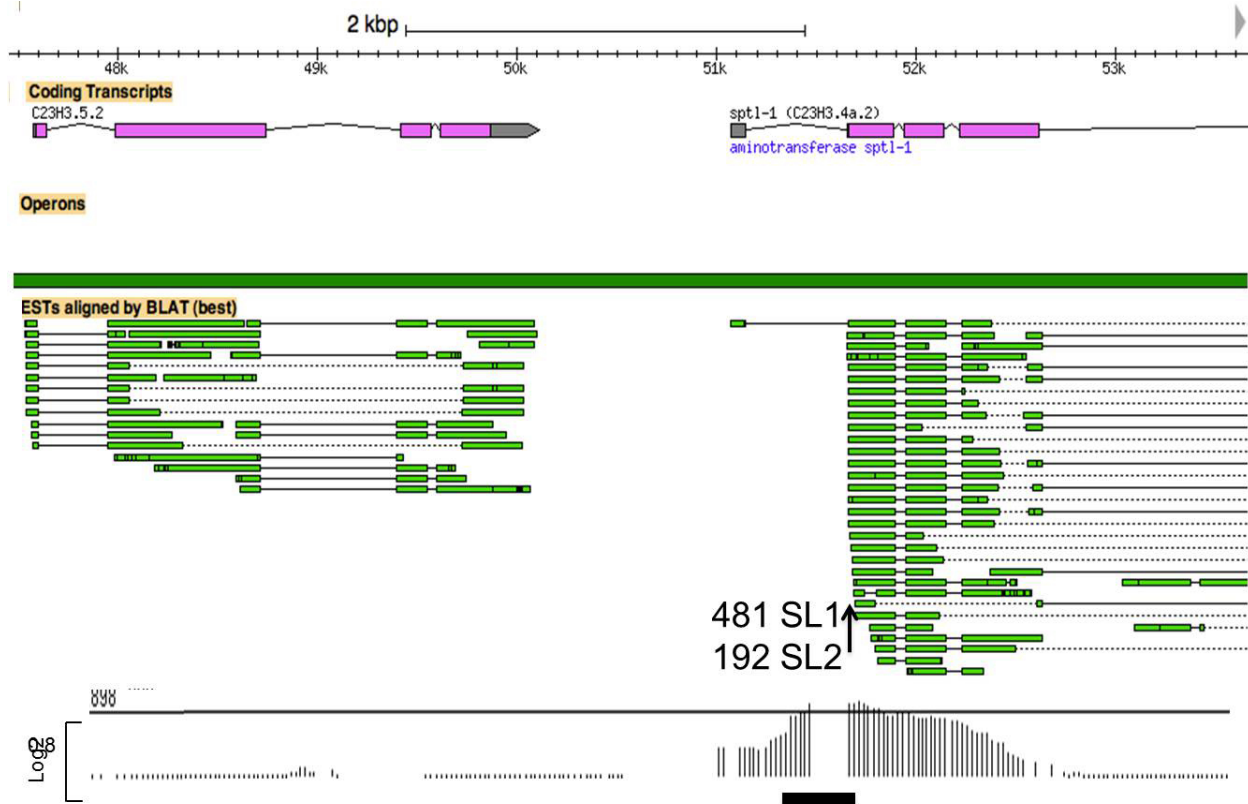


Figure 2. The intercistronic region from a typical hybrid operon. In this kind of operon, there is a promoter at the 5' end of the cluster (not shown) and an additional promoter between the genes as shown by H3K9ac ChIP/chip data (bottom). In the example shown here, the intercistronic distance is longer than in a typical operon (~500 bp). Transcripts from the promoter at the 5' end of the cluster undergo 3' end formation of C23H3.5.2, along with SL2 trans-splicing at the 5' end of *spt1-1*. The internal promoter is responsible for synthesis of an outtron that is trans-spliced by SL1 at the same site. In this example there are 192 SL2 and 481 SL1 reads at the site indicated by the vertical arrow for *spt1-1* (Allen et al., 2011). Below the operon bar H3K9ac ChIP/chip, which marks active promoters, is shown. The length of the vertical lines is proportional to the number of reads corresponding to the indicated genomic position. The horizontal black bar marks a peak called by the data analysis program.

Hybrid operons usually have intercistronic distances >500 bp and from 10-80% SL2 usage at the downstream trans-splice site. While not as common as standard SL2-type operons, the hybrid operon arrangement is not uncommon in the *C. elegans* genome. To what extent can intercistronic distance alone predict operon status? Intercistronic distance for gene pairs with the site of poly A addition of another gene less than 1000 bp upstream were calculated. They were divided into three groups based on their percent SL2 trans-splicing: >80%, 10-80%, and <10% (Figure 3). The vast majority of genes receiving >80% SL2 are 90-120 bp downstream of another gene, whereas the rest are generally much farther from genes in the same orientation. In fact, the difference is so dramatic that it is mostly accurate to conclude that genes in this close range can be reliably diagnosed as downstream in operons solely based on intercistronic length. Furthermore, the distribution of genes with very low SL2 levels (0-10%) is quite different from that of those receiving between 10 and 80% SL2. The former are presumably not in operons, while the latter likely represent hybrid operons.

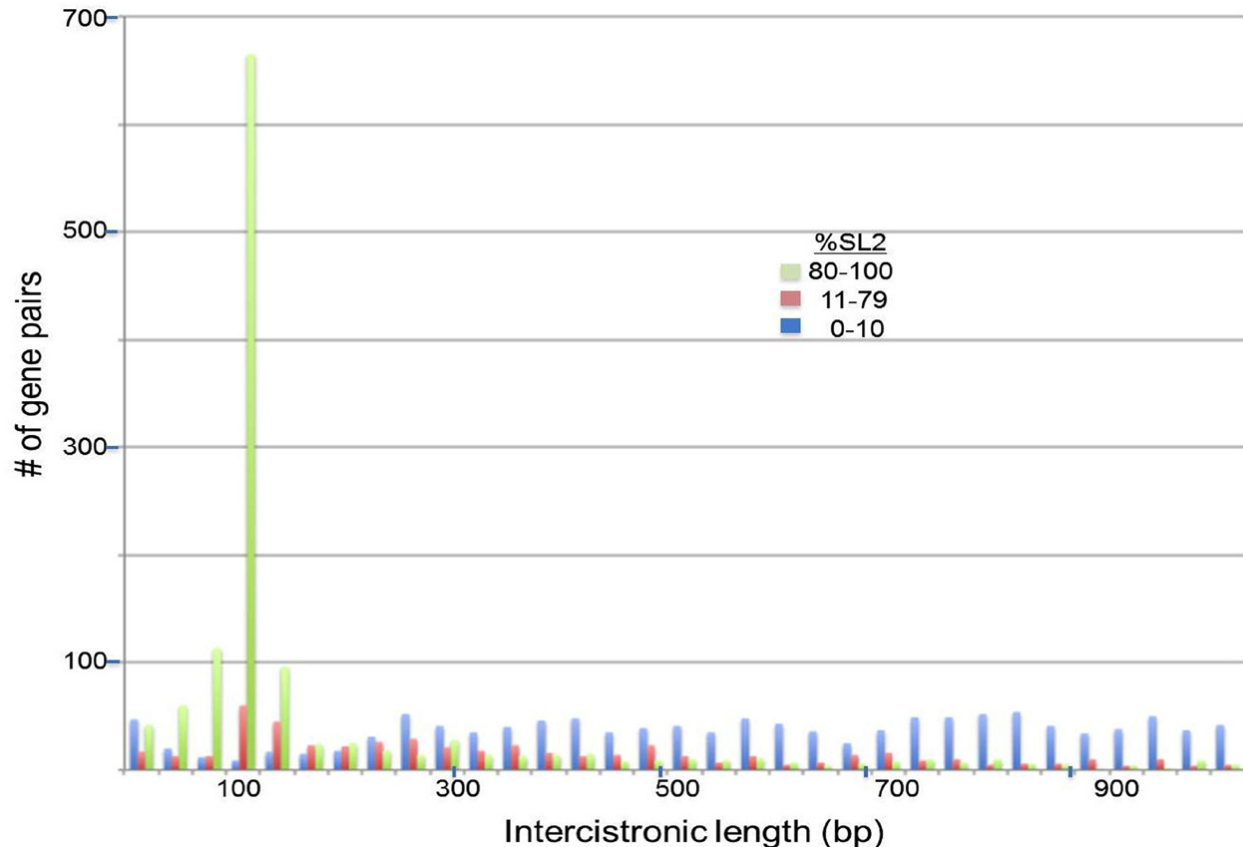


Figure 3. Intercistronic lengths of genes grouped by % SL2 trans-splicing. Data is from Allen et al (2011). All genes whose distances to the next upstream polyA addition site was 1000 bp or less and that were trans-spliced were analyzed for number of SL2 reads divided by total trans-spliced reads. The genes are grouped according to level of SL2 trans-splicing, and plotted (# of genes in each 10 bp bin of intergenic distance). The number of genes plotted is: 80-100% SL2: 1232; 11-79% SL2: 520; 0-10% SL2: 1272.

4. SL2-type operons with long spacing

Whereas hybrid operons have long spacing and a relatively low percent SL2 trans-splicing, there are also examples of genes that receive high levels of SL2, yet are spaced more than a thousand bp downstream of the nearest gene in the same orientation. It has not been clear until recently whether these represent operons with long intercistronic distances or whether these genes are trans-spliced to SL2 even though they are not downstream in operons. However, analysis of two examples with spacing >2000 bp demonstrated that in the two cases analyzed experimentally the entire region between the genes was transcribed. Thus, they appeared to be bona fide operons but with uncharacteristically long distances between genes (Morton and Blumenthal, 2011). A third example (Figure 4) provides some insight into this phenomenon. This operon is composed of *rab-18* and *uaf-1*. All of the non-coding regions of this operon, including the introns and the intercistronic region appear to be much longer than typical sizes for this kind of non-coding region in *C. elegans*. The intercistronic DNA is 3 kb long. Interestingly, this operon in *C. briggsae* appears to be typical, in that both the intercistronic region and the introns are short (Figure 4B). The expansion in *C. elegans* is due to insertion of large amounts of repetitive DNA throughout the non-coding regions of the operon, as shown in the lower part of Figure 4A. Furthermore, repetitive DNA is frequently present in *C. elegans* operons that have long intercistronic regions (J. Morton and T. Blumenthal, unpublished observations). It seems likely that most *C. elegans* operons have a short intercistronic region to facilitate processing of their pre-mRNA by coordinated 3' end formation and trans-splicing. The fact that operons like this one, with long intercistronic regions, are nevertheless trans-spliced mainly to SL2, suggests that the presence of extensive amounts of repetitive DNA may not interfere with this mechanism. Additional examples of SL2-type operons with long spacing are listed in Table 1 (Section 14). Interestingly, many of these examples have a gene (or in one case an entire operon) in the opposite orientation between the distantly spaced operon genes.

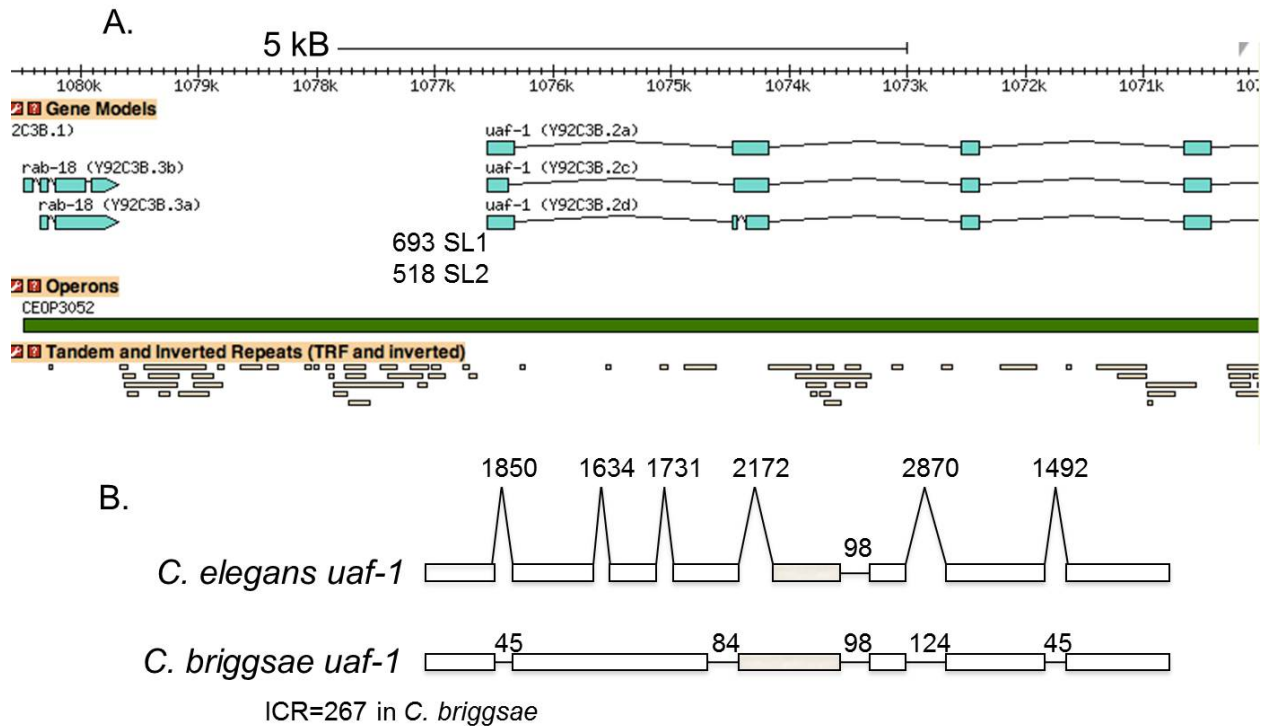
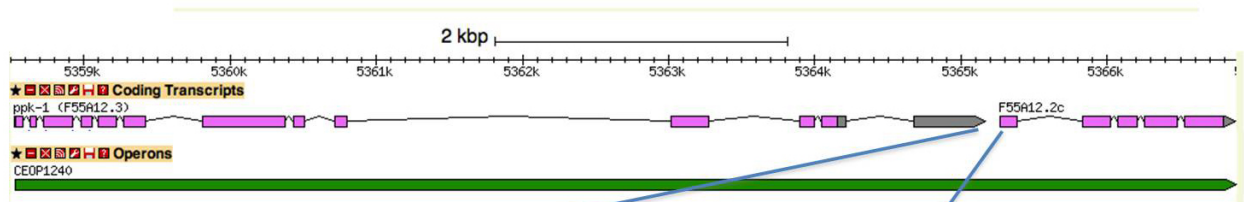


Figure 4. An SL2-type operon with exceptionally long spacing between the two genes. A. CEOP3052 is comprised of the *rab-18* and *uaf-1* genes. They are ~3 kb apart (Wormbase). The unusually long introns and ICR contain large amounts of repetitive DNA, shown in the lower track of panel A. Panel B shows a comparison of the gene structure of *uaf-1* in *C. elegans* with that of the related nematode, *C. briggsae*. Exons are shown as boxes and introns as horizontal or angled lines. Exons are to scale, but introns are not. An autoregulatory suicide exon is gray in panel B, but is not shown in panel A (MacMorris et al., 1999). In *C. briggsae* the introns are more typical of *Caenorhabditis* genes. Similarly, the ICR is only 267 bp in this species.

In sum, it appears there are two classes of operon with long spacing: (1) hybrid operons, which have substantial levels of SL1 trans-splicing emanating from transcripts initiating at an internal promoter; and (2) operons where virtually all of the downstream gene trans-splicing is to SL2 even though the intergenic region is long. These operons occur in regions of the chromosome expanded by insertion of repetitive DNA and even whole genes on the opposite strand. As in the case of the operon shown in Figure 4, these operons can sometimes be validated by the fact that the intergenic DNA in another *Caenorhabditis* species is of a more typical operon length.

5. SL2-type operons with juxtaposed 3' end formation and trans-splice sites

In some operons there is a 3' end formation site spaced a typical ~100 bp upstream of the trans-splice site, and there is a second 3' end formation site very close to the trans-splice site, as shown in Figure 5. Additional examples are listed in Table 2 (Section 14). Alternative 3' end formation is a well-established phenomenon, but in an operon it has an interesting and potentially significant extra dimension. If trans-splicing occurs before 3' end formation, then presumably only the 3' end formation site 100 bp upstream would be available because not enough RNA would be present to allow binding of the 3' end formation protein CstF, which binds downstream of the cleavage site. If 3' end formation occurs at the proximal site first, then trans-splicing would occur normally. However, if the distal 3' end formation site is used before trans-splicing occurs, the RNA upstream of the trans-splice site would be too short to allow trans-splicing. Thus it is possible that in these cases alternative 3' end formation is a regulatory phenomenon, as it often is. However, in this scenario the expression of the gene downstream in the operon is regulated by alternative 3' end formation of the gene upstream. Nothing is currently known about the order of processing events in these operons, so such regulation remains hypothetical. It is often possible to determine whether a phenomenon is important by determining whether it is conserved. In this case, many of the downstream polyA sites are present in *C. briggsae* (Table 2, Section 14) suggesting they may play important roles.



AATAAAAA GATTTTCCAACTT**AAAAC**
 TTTGATTTTCTTTGCTTTCTTTTGGT
 TTTAAGTCGCTTTTTCAATATTTGAAT
AAACATTTTCAGCATC**AGAACTTGA**
AAATATTTCAG/ATG

Figure 5. The intergenic region of an SL2-type operon with a second polyA site just upstream of the trans-splice site. This kind of operon has two, alternatively used polyA sites, a proximal site at ~100 bp upstream of the trans-splice site, and a distal site just upstream of the trans-splice site. When the distal site is used, subsequent trans-splicing of the downstream gene is impossible. The sequence shown is from the proximal polyA signal of F55A12.3 to the translation start site of F55A12.2. The polyA signals are underlined. The two sites of polyadenylation are underlined and boldface. Both sites are used based on known 3' end formation and both show ChIP-seq CstF-64 peaks (data not shown). The trans-splice site is underlined and italicized. F55A12.2 is trans-spliced 97% of the time to SL2 (196 reads in the transcriptome data (Allen et al., 2011)).

6. SL2 trans-splicing at downstream exons

There are several instances of SL2 trans-splicing occurring at a 3' splice site that is also used as a cis-splice site: that is, at the 3' end of an intron. A list of examples is given in Table 3 (Section 14), and the phenomenon is illustrated in Figure 6. Sometimes in these cases, the first exon is trans-spliced to SL2 as well, but sometimes there is no evidence for trans-splicing of the first exon. Furthermore, higher levels of SL2 trans-splicing are observed when the trans-splice sites are closer to the next upstream polyA site (Table 3, Section 14), consistent with the idea that they are spliced to SL2 because they are downstream in operons.

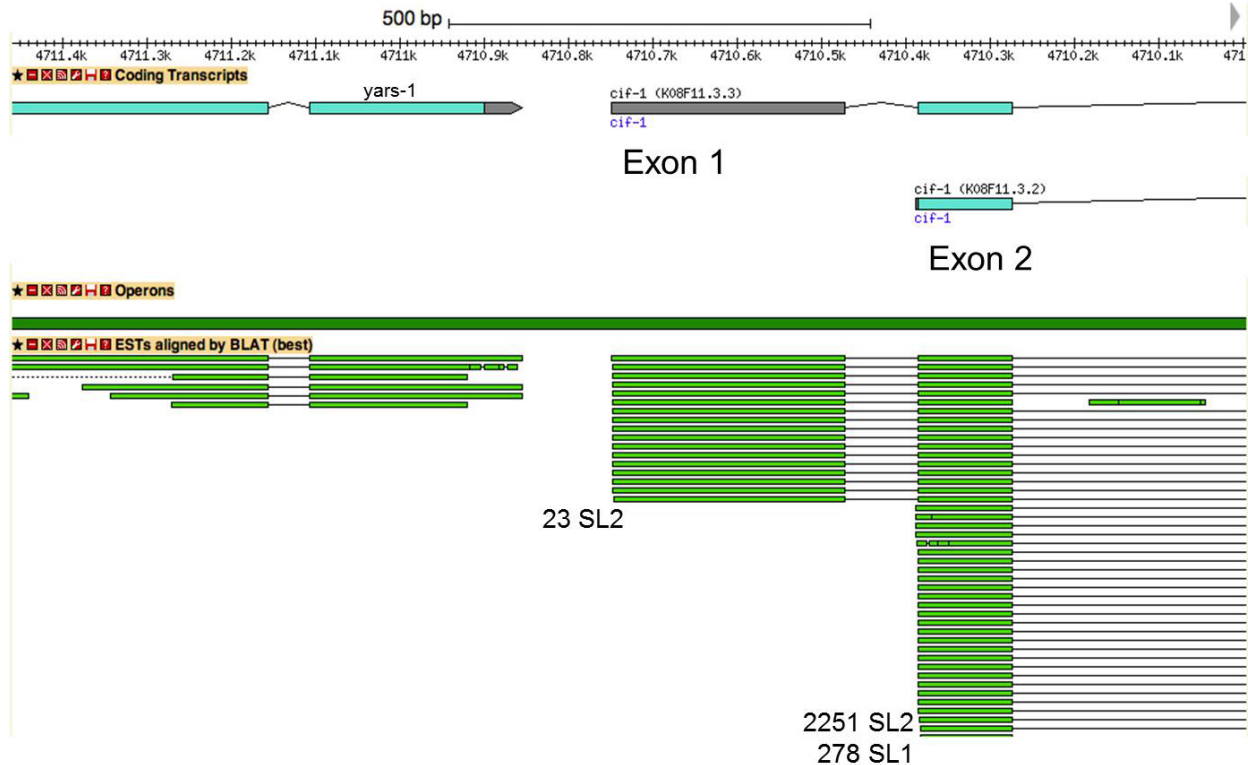


Figure 6. A gene with SL2 trans-splicing at the second exon. The picture shows portions of two operon genes, *yars-1* and *cif-1*, with the region between them. The numbers of SL1 and SL2 reads (Allen et al., 2011) are given at the sites of trans-splicing. Mapped ESTs are shown as green bars. Some contain exon 1, whereas others begin with trans-splicing at exon 2.

7. Alternative operons

The Blumenthal and Bartel groups independently reported several instances of SL2 trans-splicing accompanied by 3' end formation within annotated single genes, alternative operons (Jan et al., 2011; Morton and Blumenthal, 2011). Some interesting and important genes exhibit this phenomenon; e.g., *hdac-6* (Figure 7), *smg-6*, and *sma-9*. In these three cases there is good evidence for: (1) mRNAs encompassing the entire gene, (2) mRNAs comprising only the 5' part of the gene due to 3' end formation within the gene, and (3) mRNAs comprising only the 3' part of the gene resulting from SL2 trans-splicing ~100 nt downstream of the internal 3' end formation site. Mechanistically, the events that gave rise to the latter two types of mRNA are the same as any other operon. In the case of *sma-9*, this has been shown to be functionally significant (Yin et al., 2010). This phenomenon is also similar to other cases of alternative 3' end formation vs. cis-splicing. There is a choice between 3' end formation within an intron and removal of the intron by cis-splicing. Presumably the relative rates of 3' end formation and cis-splicing determine which process occurs. 3' end formation is accompanied by SL2 trans-splicing at the site that otherwise would have served as the 3' splice site for cis-splicing. These genes/operons, termed “alternative operons”, can potentially incorporate a level of regulation or specialization of function unique to this situation. A list of 17 clear cases and five additional possible cases are given in Table 4 (Section 14). In the latter cases, the internal SL2 trans-splicing is well documented, but the 3' end formation upstream is only predicted. However, appropriately located canonical 3' end formation signals make it likely that these are actual alternative operons, although it is possible that in these latter instances no stable upstream product accumulates.

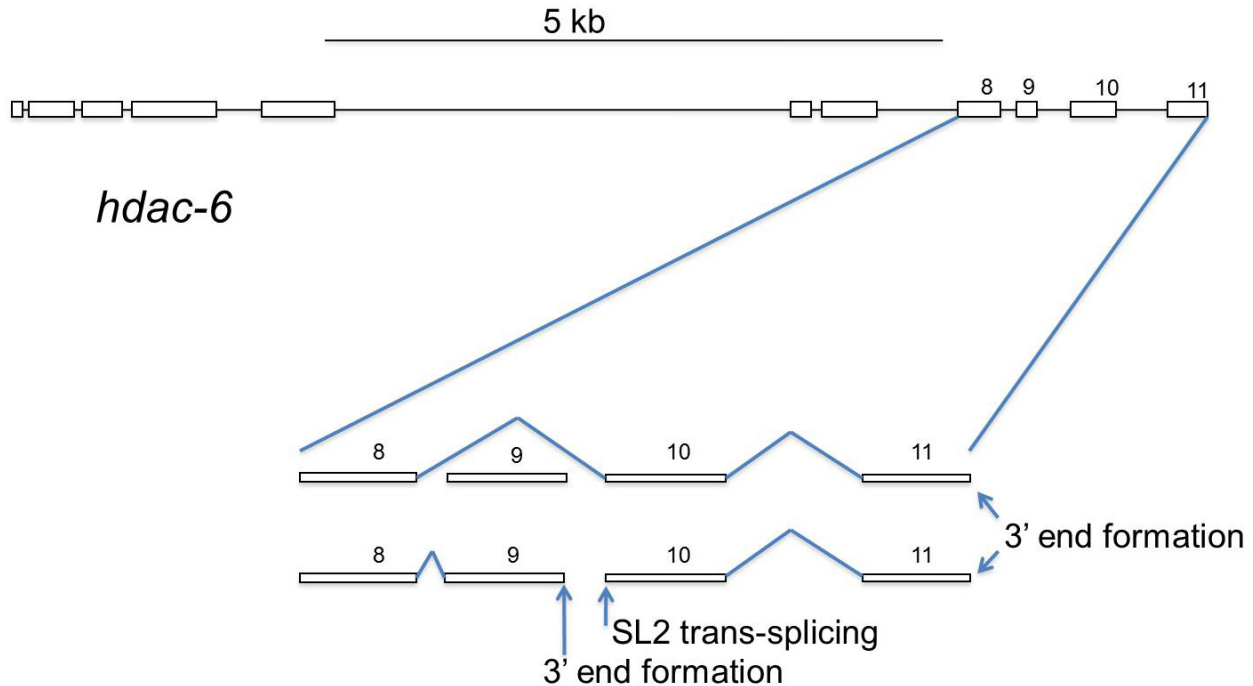


Figure 7. An alternative operon. *hdac-6*, which encodes a histone deacetylase, can be processed in either of two ways. When exon 9 is skipped, a full length *hdac-6* mRNA is synthesized by splicing exon 8 to exon 10. When exon 8 is spliced to exon 9, 3' end formation at the end of exon 9 occurs and is accompanied by SL2 trans-splicing just downstream, at the 5' end of exon 10, thereby creating a nine-exon upstream mRNA and a two-exon downstream mRNA.

Interestingly, several genes encoding proteins involved in 3' end formation/transcription termination are present on this list. These include *pcf-11*, *pfs-2*, and *T23H2.3*, which encodes the worm homolog of the transcription termination factor, TTF2. These genes may be autogenously regulated at the level of premature 3' end formation within the gene. This could occur when the levels of their encoded proteins are high. The formation of an SL2 trans-spliced downstream product could be an unintended consequence of premature 3' end formation or an integral part of an autoregulatory pathway.

8. SL1-type operons

In this type of operon there is no intercistronic DNA: the 3' end of the upstream gene immediately precedes the trans-splice site of the downstream gene, and all trans-splicing is to SL1, rather than SL2 (Figure 8A) (Williams et al., 1999). It was hypothesized that a single RNA processing event, SL1 trans-splicing, gives rise to both mature downstream mRNA, and an upstream mRNA that gets polyadenylated and debranched. An AAUAAA polyadenylation signal is present just upstream of the trans-splice site, making this quite a reasonable idea. However, when one such operon was studied experimentally in transgenic worms, 3' end formation of the upstream gene clearly competed with trans-splicing of the downstream gene, suggesting that a single pre-mRNA molecule does not give rise to both mRNAs.

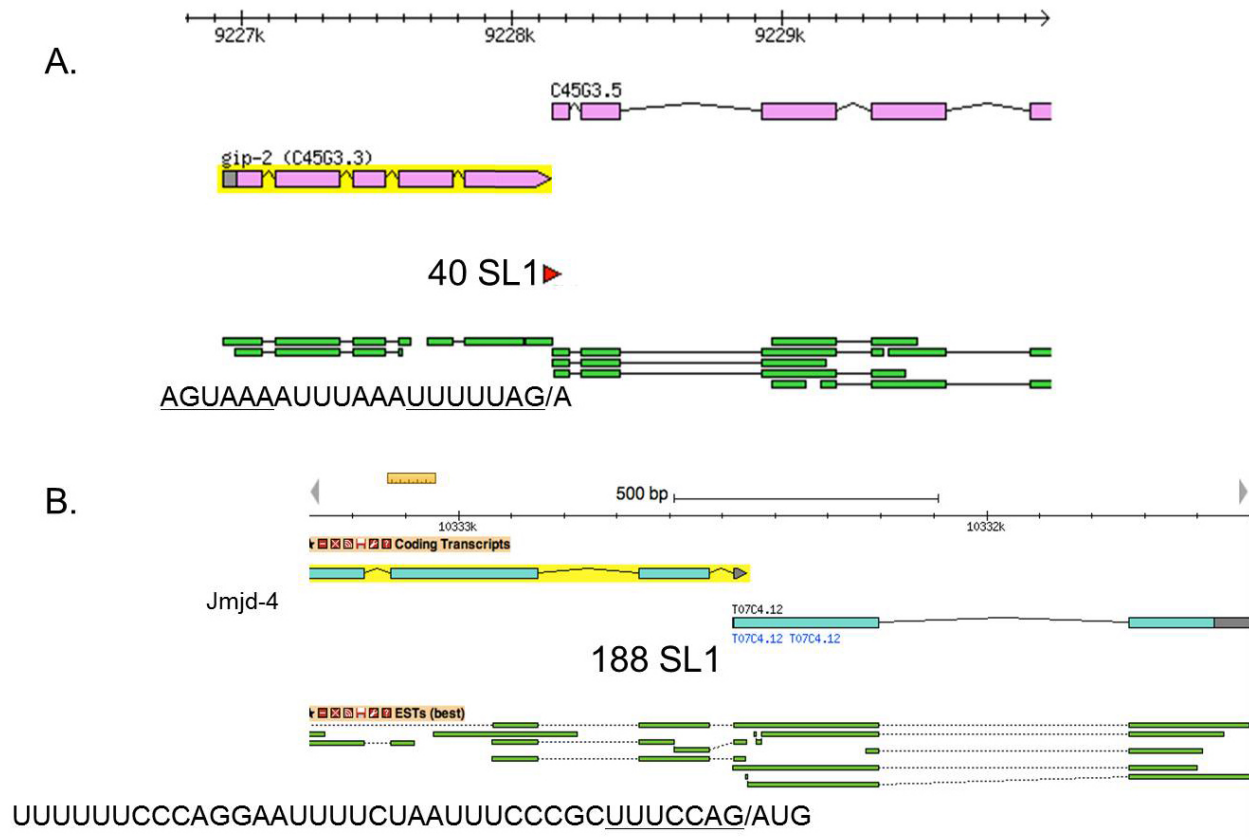


Figure 8. SL1-type operons. **A.** In this type of operon, 3' cleavage may occur by SL1 trans-splicing at least sometimes, resulting in a downstream mature SL1 trans-spliced mRNA and an upstream precursor that apparently can be polyadenylated at the site of trans-splicing-catalyzed cleavage. The transcriptome data has 40 SL1 reads and no SL2 reads (Allen et al., 2011). The green bars show extents of a few ESTs, where the *gip-2* EST ends precisely at the site where the C45G3.5 ESTs begin. The sequence of the genome at this site is shown below. The polyA addition signal, the variant sequence AGUAAA in this case, is underlined, as is the trans-splice site. **B.** An SL1-type operon with no monocistronic mRNA for the upstream gene. In this two-gene operon, 3' end formation occurs only at the 3' end of the cluster. There is no CstF-64 peak at the 3' end of *jmjd-4* in a ChIP-seq experiment, whereas there is a strong peak at the 3' end of T07C4.12 (A. Garrido-Lecca and T. Blumenthal, unpublished). Although there is no 3' end formation within the operon, there is SL1 trans-splicing that results in production of mature T07C4.12 mRNA, as indicated by 188 SL1 transcriptome reads at this site (Allen et al., 2011). The upstream gene, *jmjd-4*, must be expressed from the dicistronic mRNA where the T07C4.12 encoding exons are out of frame. The sequence shown at the bottom is of some of the intercistronic DNA. The trans-splice site is underlined.

Table 5 (Section 14) presents 23 examples of SL1-type operon in the genome, as defined both by SL1 trans-splicing and by an AAUAAA and 3' end formation at the trans-splice site. Each of these 23 operons also has the characteristic extended polypyrimidine tract reported in the first three SL1-type operons previously studied (Williams et al., 1999). Intriguingly, analysis of 3' end formation in the UTRome project (Mangone et al., 2010), as well as other determinations of 3' ends (our unpublished observations), demonstrates that for most of these genes, 3' ends occur directly on the AG of the trans-splice site, indicating that these 3' ends could be the result of polyadenylation at the free 3' end created by SL1 trans-splicing of the downstream gene. Analysis at the whole genome level of dinucleotides at which polyA is added indicates that AG is among the very rarest sites for polyadenylation in *C. elegans* (Mangone et al., 2010). It seems likely that this represents a new mechanism for 3' end formation—cleavage by trans-splicing, instead of by the usual mechanism (cleavage by 3' end formation machinery). Cleavage would then be followed by polyadenylation dependent on the AAUAAA near the free 3' end created by trans-splicing. The apparent discrepancy between the experimental results and the transcriptome data is unresolved. However, a reasonable explanation would be that both conventional and trans-splicing-dependent 3' end formation can occur at these 3' ends, perhaps depending on the circumstances.

Figure 8B shows an example of an unusual SL1-type operon where there is no AAUAAA upstream of the trans-splice site and no accumulation of an upstream mRNA. The 3' UTRome shows no 3' end formation at the 3' end of *jmjd-4*. Presumably in this case trans-splicing by SL1 is important for making downstream mature mRNA, but the upstream product is discarded. When trans-splicing does not occur, a dicistronic mRNA is presumably transported to the cytoplasm for translation of the upstream cistron.

9. Dicistronic mRNAs

Operons that make dicistronic mRNAs have been well documented in *Drosophila*, and they also are present in *C. elegans* (Figure 9; Table 6, Section 14). In these cases, there is no evidence for processing of the dicistronic mRNA, either by trans-splicing or 3' end formation between the two genes. These operons appear similar to those identified in *Drosophila*, where the mechanism for internal translation of the downstream open reading frame, perhaps an internal ribosome entry site (IRES), is not yet known.

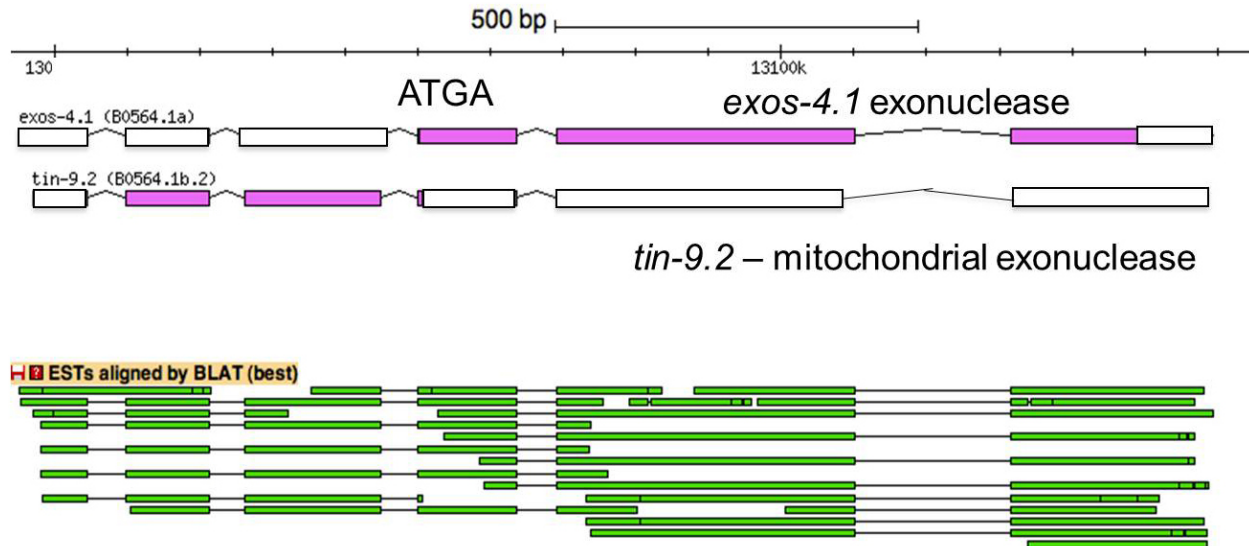


Figure 9. An operon producing only a dicistronic mRNA. The diagram shows the first two genes in CEOP4532, *tin-9.2* and *exos-4.1*, shown in their entirety, with filled boxes representing translated regions and open boxes untranslated. They specify two paralogous exonucleases. The ESTs shown below the operon bar demonstrate that only dicistronic mRNAs are produced. There is no evidence in the database for either 3' end formation or trans-splicing and no ChIP-seq peak of CstF-64 (data not shown) between the genes. The 3' UTRome data shows 3' end formation only at the 3' end of the two-gene cluster. Both proteins must be expressed from this single mRNA. It is not known how translation of the downstream protein is initiated. The *exos-4.1* ATG start codon overlaps the *tin-9.2* TGA stop codon: ATGA. In other operons making dicistronic mRNAs (listed in Table 6), the open reading frames of the two genes can have a greater overlap or they can be separated by non-translated sequences.

10. Overlapping genes

Overlapping genes of two kinds can be found. In one type, the 3' UTR of the upstream gene overlaps with the outtron of the downstream gene. In the example shown in Figure 10A, a single stretch of DNA serves two purposes, although the two resulting mature mRNAs have no sequences in common, just as with operons. A single stretch of DNA serves as the 3' UTR and 3' end formation signal of the upstream gene as well as the outtron, the region needed for trans-splicing of the downstream gene. These can easily be mistaken for SL1-type operons, since the genes are nearly touching and the trans-splicing is to SL1. However, the key difference is that the 3' ends do not occur at the exact sites of the 3' cleavage products created by trans-splicing. The polyA sites are a bit further upstream or can even occur downstream of the trans-splice site. The most likely explanation is that these genes are not in an operon, but there is a promoter for the downstream gene within the 3' UTR of the upstream gene. The H3K9ac track supports the existence of this hypothetical promoter (Figure 10A). Examples of this arrangement are listed in Table 7 (Section 14).

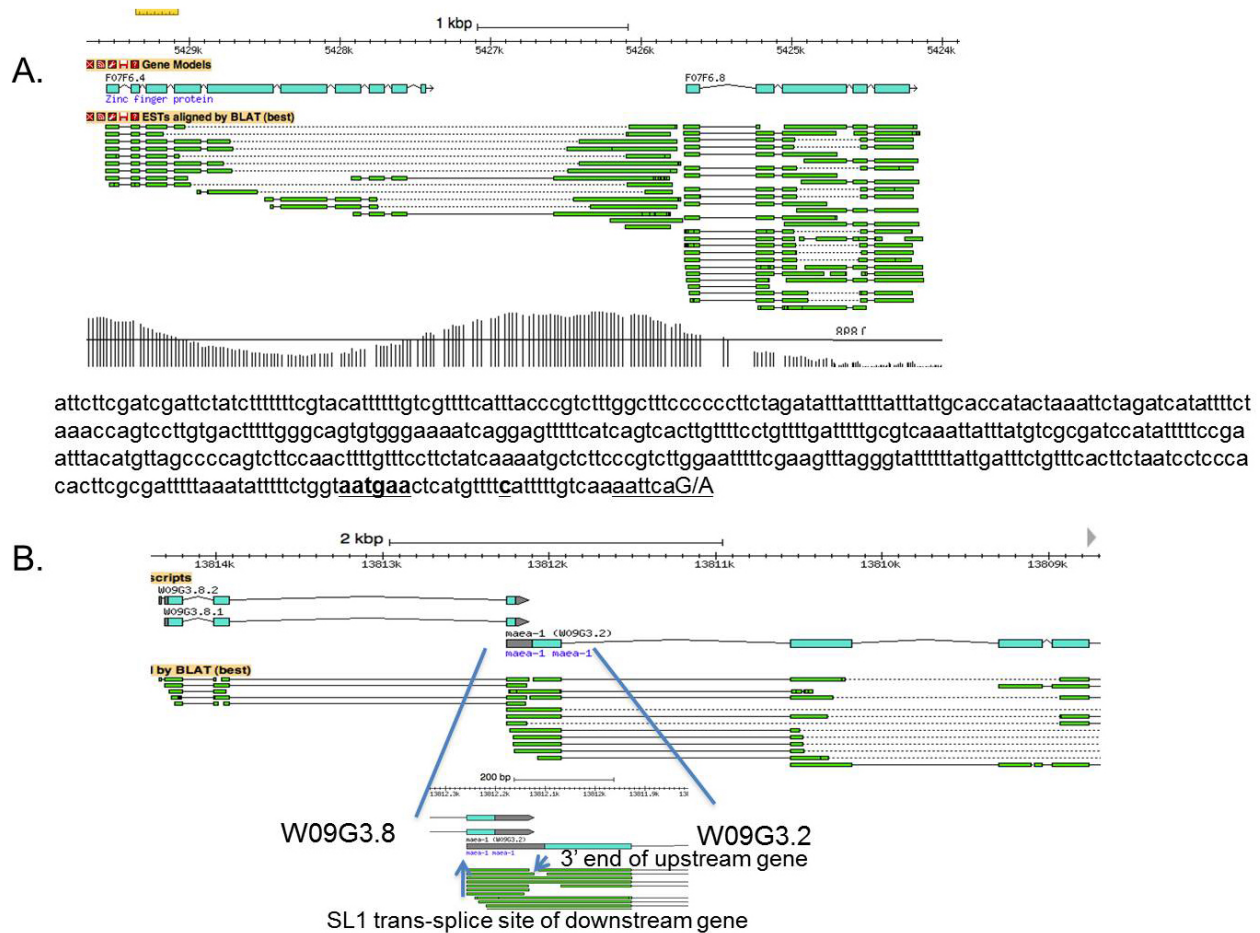


Figure 10. Overlapping genes. A. A two-gene cluster where the 3' UTR of one gene overlaps with the outtron of the next gene downstream. The sequence shows the 3' UTR of F07F6.4, with the site of polyA addition (dotted underline and boldface) and the polyA signal (underlined and in boldface). This site is only eighteen bp upstream of the SL1 trans-splice site of F07F6.8 (double underlined). Presumably a promoter within the F07F6.4 3' UTR results in the synthesis of the F07F6.8 outtron. The H3K9ac track shown provides evidence for this promoter. The length of the vertical lines is proportional to the number of reads corresponding to the indicated genomic position. **B.** Dual use of an exon. This panel shows an example of two genes that overlap. The upstream gene, W09G3.8, ends within the exon shown in the middle of the diagram. The downstream gene, W09G3.2, begins by SL1 trans-splicing at the 5' end of this same exon, as indicated by 555 SL1 transcriptome reads at this site. In this case, it is impossible for a single polycistronic pre-mRNA to give rise to both mRNAs.

The second kind of overlapping gene involves dual use of a single exon. Sometimes a single stretch of DNA can serve as parts of exons of two adjacent genes. Figure 10B illustrates an unusual arrangement in which an exon is sometimes used as the end of the coding region and 3' UTR of the upstream gene, or alternatively as the 5' UTR and beginning of the coding region of the downstream gene. Obviously, this cannot be an operon since the two mRNAs must be made from different pre-mRNA molecules. Table 8 (Section 14) lists additional examples.

11. Identification of operons

Originally, operons were identified based on both the presence of SL2 trans-splicing and close intergenic spacing. However, there are significant gray areas—closely spaced genes without much SL2 trans-splicing, or SL2 trans-spliced genes without genes closely spaced upstream. It is now clear that in *C. elegans*, SL2 trans-splicing is by far the best predictor of transcription from a promoter upstream of another gene. The only exceptions appear to be occasional SL2 trans-splicing at intron trans-splice sites, and many of these could be uncharacterized alternative operons. Therefore, operon designations for gene clusters are most reliably based solely on SL2 trans-splicing. To what extent do operons defined this way have different intergenic length distributions? The data in Figure 3 definitively answer this question: genes in the same orientation less than one kb apart can be divided into three groups according to their SL2 trans-splicing levels. The vast majority of genes receiving >80% SL2 are 90-120 bp downstream of another gene, whereas the rest have much longer spacing. In fact, the difference is so dramatic that genes in this close range can be diagnosed as very likely to be downstream in operons, solely based on intergenic

length. Furthermore, the distribution of genes with very low SL2 levels (0-10%) is quite different from that of those receiving between 10 and 80% SL2. The former are rarely in operons, whereas the latter are mostly in hybrid operons.

Genes spaced >1000 bp apart are very rarely in operons. However, the genes spaced between 130–1000 bp apart can not be accurately diagnosed based on intergenic length alone—some are in operons, some hybrid operons, and some not in operons. This highlights the danger of assigning genes in organisms that lack SL2 trans-splicing to operons based solely on intergenic distance.

12. Regulation in and of operons

The *C. elegans* genome has a surprisingly varied array of gene clusters. Because the vast majority of such clusters fit the general description of SL2-type operons: ~100 bp intergenic region containing a Ur element (Lasda et al., 2010) and with almost all trans-splicing to SL2 (Figure 1), it seems likely that this is an optimal arrangement for processing of the polycistronic RNA precursor to produce individual mature mRNAs. Because expression of downstream operon genes depends on successful 3' end formation and trans-splicing, it is not surprising that expression levels of genes more distant from the promoter drop off somewhat (Cutter et al., 2009; M. A. Allen and T. Blumenthal, unpublished observations). Although this does not constitute regulation *per se*, it does suggest that position within operons may have been selected based on need for higher or lower expression levels. In general, there is no obvious relationship between the functions of genes in the same operon, but there are interesting exceptions to this rule (Blumenthal and Gleason, 2003). For example, *ceop4596* contains three genes, all of which specify proteins involved in pre-mRNA splicing. The operon dataset as a whole (the operome) largely contains genes required for growth, including especially the genes for the basic machinery of gene expression and energy generation. The presence of genes in operons may allow the worm to respond rapidly to the need for growth during particular times in the life cycle, such as embryogenesis and relief from starvation (Zaslaver et al., 2011).

The other kinds of operons considered in this manuscript are quite rare compared with conventional SL2-type operons, and it seems likely that many of them have arisen in response to need for specific kinds of regulatory requirements. For example, hybrid operons contain a promoter at the 5' end of the cluster and a different promoter somewhere within the cluster. Presumably, the internal promoter results in expression of the gene or genes downstream of it at a time or place where the 5' end promoter is not expressed, or is expressed at an insufficiently high level. The internal promoter allows differential expression of the genes in the operon in response to particular signals, for example, but still allows the entire cluster to be expressed as a unit when needed. Both co-regulation and differential expression are thereby achieved.

SL2-type operons with an extra 3' end formation signal very close to the trans-splice site present a possibility for a very different sort of regulation. When the upstream 3' end formation signal is used, both upstream and downstream genes can be expressed as in a conventional SL2-type operon. However, when the 3' end formation site just upstream of the trans-splice site is used before trans-splicing has occurred, expression of the downstream gene from the same pre-mRNA would be eliminated because there would be insufficient RNA upstream of the trans-splice site to allow for trans-splicing. Hence the downstream RNA would lack a cap and would presumably be degraded. In contrast, if trans-splicing occurred before 3' end formation, then both genes could be expressed no matter which 3' end formation site was used. Thus, differences in the efficiency of 3' end formation at the two sites and of trans-splicing determines which genes are expressed from these polycistronic pre-mRNAs. Differential use of 3' end formation sites has been shown before to result in inclusion or exclusion of sites needed to bind regulatory RNAs and proteins, thereby regulating expression of the gene (Jan et al., 2011). With this type of operon, there should be an interesting extra effect of such regulation: expression levels of the gene downstream would depend on choice of 3' end formation site upstream.

The SL1-type operons present a similar kind of regulatory opportunity. If SL1 trans-splicing happens first, the free 3' end of the upstream mRNA can theoretically be polyadenylated, so both products could be made from a single pre-mRNA molecule. However, if 3' end formation happens first, then the downstream gene is inactivated, so only the upstream part of the pre-mRNA is produced. Thus, any change in the relative efficiencies of 3' end formation and SL1 trans-splicing could potentially result in changes in the relative expression of the two genes involved. Nonetheless, no SL1-type operon has yet been shown to be regulated this way.

Alternative operons present yet another regulatory possibility. In this case, when cis-splicing occurs, one gene product is made, but when 3' end formation accompanied by trans-splicing occurs, two gene products are potentially made from the same set of exons. Which processing event occurs would be determined by their relative efficiencies.

The fact that three of these alternative operons encode proteins required for 3' end formation or transcription termination strongly suggests that this arrangement is at least sometimes regulatory, apparently autoregulatory in these instances.

Two of the kinds of operon discussed here are presumably regulated at the level of translation. In the SL1-type operon that produces both a dicistronic mRNA and a trans-spliced mRNA, the expression of the downstream gene could be modulated by the efficiency of translation of the two mRNAs. It is currently not known how translation of a downstream gene in a dicistronic mRNA is initiated, so it is premature to speculate how such translation could be controlled. Second, the operons that make only a single dicistronic mRNA must also be subject to any regulation that occurs at the level of translation.

When two genes share exons, as in the case of the gene pairs, *cha-1/unc-17* or W09G3.8/2 (Figure 10B), they generally cannot be expressed from the same pre-mRNA. So for any given pre-mRNA only one of the two genes can be expressed. This can be regulated at the level of alternative splice site choice, as a decision between 3' end formation and splicing. In the case shown in Figure 10B, the exon can serve as the 3' end of an upstream gene or, alternatively, as the first exon of the downstream gene. How such choices are made is not currently known.

13. Genome architecture

The many ways of arranging genes with a minimum of wasted space discussed in this chapter suggest that *C. elegans* may have had to adapt by minimizing its genome size. The result is a surprising variety of gene arrangements where regulation can occur, but with a minimum of DNA dedicated to conventional regulatory elements. Although many of these interesting gene arrangements have been reported first in *C. elegans*, some have subsequently been found in other nematodes and in other phyla. It seems reasonable to predict, therefore, that some of the apparently unique gene arrangements discussed here will be found in other species.

14. Tables 1-8

Table 1. Operon gene pairs with > 1 kb spacing.

1	2	3	4	5	6
Gene pair	ceop	reads		spacing (kb)	intervening gene
		#	%SL2		
B0414.8 - .7	1003	6	67	2.6	B0414.1
F30A10.9 - .3	1005	701	33	10	F30A10.8
F49C12.8 - .11	4680	7374	89	1.2	F49C12.10
K01C8.9 - .6	2742	876	85	2.6	K01C8.8
T16H12.11 - .3	3652	1	100	4	ceop3648
Y51H7C.6 - .5	2048	208	80	3.5	Y51H7C.9
Y54E10BL.6 - .5	1943	2901	81	2.6	Y54E10BL.3
Y57A10A.18-16	2688	254	62	6.7	Y57A10A.35
ZK546.2 - .1	2690	812	52	1.9	ZK546.15

Column 1 indicates gene pairs, **2** operon designation, **3** number of trans-spliced reads (Allen et al., 2011), **4** % of these that are spliced to SL2 or one of its variants, **5** the spacing between the polyA site of the upstream gene and the trans-splice site of the downstream gene, **6** gene or genes, usually on the opposite strand, that occur between the operon genes.

Table 2. SL2-type operons with a polyA site near the trans-splice site.

1	2	3	4	5	6	7
Gene pair	polyA signal1	at	polyA signal2	at	%SL1	conserved?
C15H11.7/4	GATAAA	-141	AAAAAA	-29	18	-
C17E4.5/11	TAAAAA	-112	AGTAAA	-43	11	+/-
C24F3.1/4	AATAAA	-136	TATAAA	-36	6	-
C34E10.6/5	AATAAA	-121	AATAAA	-54	4	+
C43E11.4/11	AATAAT	-130	AATAAT	-23	5	-
C47B2.6/7	AATATA	-188	AAAAAA	-42	6	+
C56C10.8/7	AATAAA	-181	AATAAA	-25	22	+
D2096.8/7	AATACT	-170	AATAAA	-27	7	-
E02C12.13/4	CATAAA	-89	TATAAA	-47	23	+
F01G4.6/5	TATAAA	-171	AATAAA	-31	6	+
F21D5.7/6	AATTAA	-170	AATAAA	-16	1	-
F25H2.11/12	AATGAA	-115	AAAAAA	-25	5	-
F40G9.3/2	CATGAA	-121	TTTAAA	-40	5	+
F45H10.2/1	AGTGAA	-120	GAAAAA	-33	13	+/-
F46B6.5/4	GATGAA	-117	TATGAA	-31	3	-
F55A12.3/2	AATAAA	-115	AATAAA	-39	3	-
F59A2.3/4	AATAAA	-134	AATAAA	-19	14	+
K08E7.2/1	TATGAA	-130	GATAAA	-22	14	+/-
R119.4/3	AATAAA	-112	TATAAA	-30	0	+
T20H4.4/ R151.10	AATTAA	-116	AATAAA	-29	8	-
T05G5.6/5	ATATAT	-132	AAAAAA	-26	6	+
T05G5.9/8	AATGAA	-128	TATAAA	-33	6	-
Y32H12A.2/ZK121.1	AATATA	-124	AATAGA	-28	12	-
Y38A10A.7/6	TATAAT	-89	TATAAA	-29	6	-
Y47G6A.22/21	AATAAA	-136	AAAAAA	-45	53	+/-
Y62E10A.1/5	AATAAA	-109	AATAAA	-28	2	-
Y71H2AM.5/4	GATAAA	-121	AATAAA	-18	9	+
Y111B2A.2/3	GATGAA	-138	TTTAAA	-40	10	-
ZK856.13/10	TATAAA	-123	TATAAA	-29	3	-
ZK1010.1/2	AATAAA	-116	TATAAA	-12	8	-

These gene pairs (full designation for upstream gene/same designation for downstream gene up to the dot) all have two sites of polyA addition, a proximal site (**columns 2 and 3**), and a distal site (**columns 4 and 5**). However, they are not SL1-type operons based on the fact that most of their trans-splicing is to SL2 (**column 6**) (Allen et al., 2011) and the distal site is not identical with the site of trans-splicing. **Column 7** indicates whether the distal sites are also present in the related nematode, *C. briggsae* (+/-) indicates that a possible signal is present.

Table 3. SL2 trans-splicing to ds exons.

1	2	3		4		5		6		7		8	
		reads											
		ds exon				1 st exon				bp from upstream gene to:			
Gene	exon#	#	%SL2	#	%SL2	ds exon	1 st exon	ds exon	1 st exon	ds exon	1 st exon	ds exon	1 st exon
C10G11.5	2	453	81	903	5	260	75						
F57B10.3	2	773	24	0	-	1200	600						
K10C3.6	2	480	39	0	-	2100	600						
Y106G6H.8	2	5	100	2068	88	280	100						
Y87G2A.8	2	963	13	0	-	1100	600						
F43E2.6	2	16	88	0	-	900	400						
F42G9.5	2	83	47	89	0	1200	80						
R07E5.10	2	536	77	808	75	500	300						
ZK121.1	2	21	81	586	88	290	3,40,100,140						
F54E7.1	2	10	100	110	72	600	400						
R151.2	2	954	24	173	93	1600	110,190						
K08F11.3	2	2529	89	23	100	470	100						
F44E7.4	2	288	15	357	14	2200	1900						
K10B3.6	2	23	30	0	-	1500	800						
C30H7.2	2	390	39	9	100	1200	600						
Y48C3A.18	2	11	91	120	91	1100	800						
Y50D7A.8	2	8	88	310	50	2400	1100						
F01F1.5	2	84	68	310	85	500	200						
F01F1.1	2	211	51	2	50	500	90						
K08D10.2	2	18	94	456	86	800	100						
F28F8.6	2	707	47	25	72	700	140						
C34E10.4	2	101	93	18	0	600	44						
C34E10.4	3	11	82	18	0	700	44						
C34E10.4	5	17	88	18	0	1400	44						

Some SL2 trans-splicing events occur at downstream exons, rather than, or in addition to, the first exon in the gene. The numbers of reads (**columns 3 and 5**) and percent SL2 (**columns 4 and 6**) at the downstream (ds) exon (**columns 3 and 4**) and at the first exon (if trans-spliced) (**columns 5 and 6**) are indicated (Allen et al., 2011). Distances between the trans-splice site(s) and the nearest upstream polyA site are shown in **columns 7 and 8**. When multiple numbers appear in the rightmost column, the nearest upstream gene uses more than one polyA site.

Table 4. Alternative operons.

Gene	polyA signal	position (bp)	SL2 ts at exon #
C03H12.1	AATAAA	-147	11
C09E9.1	TATGAA	-181	3
<i>mks-5</i>	AATAAA	-119	10
C35E7.5	AATAAA	-135	4
C44B7.2/.1	AATAAA	-320	12
<i>hum-2</i>	TATAAA	-454	14
F40E10.6	ATTAAA	-158	7
<i>hdac-6</i>	AATAAA	-164	9
F55D12.5	TATAAA	-120	5
F58G1.2	AATAAA	-128	2
<i>pfs-2</i>	AATAAA	-200	4
<i>pcf-11</i>	AATAAA	-56	5
<i>sma-9</i>	AATAAA	-171	12
<i>hars-1</i>	AATAAA	-111	2
T23H2.3	AATAAA	-106	6
<i>smg-6</i>	AATAAT	-145	6
Y67D8C.3	AATAAA	-197	2
Possible alternative operons based on SL2 and AAUAAA			
C32D5.8	TATAAA	-87	2
<i>trt-1</i>	TATAAA	-96	2
<i>taf-5</i>	TATAAA	-119	2
<i>cdk-7</i>	AAAAAA	-107	2
<i>faah-4</i>	CATAAA	-109	2
<p>These are single genes when internal 3' end formation accompanied by SL2 trans-splicing does not occur, and operons when it does. The upper group represents confirmed alternative operons, and the lower group represents suspected alternative operons based on having similar properties to the first group. In the lower group, no mRNAs ending at internal polyA sites have been reported. In this table gene names are used instead of cosmid designations where available to illustrate the possible functional implications of these operons. The position of the polyA signal is given with respect to the alternative cis/trans splice site.</p>			

Table 5. SL1-type operons.

Gene pair	polyA signal	position	%SL1
<i>gna-1</i> – B0024.11	AATAAA	-12	100
B0432.9 – .13	AATAAA	-19	73
<i>rpb-6</i> – <i>dohh-1</i>	AATAAA	-13	100
C07H6.2 – <i>lig-4</i>	AATAAA	-19	98
<i>mes-6</i> – <i>cks-1</i>	AATAAA	-20	99
C43H8.1 – B0511.6	CATAAA	-23	98
C45G3.3 – <i>gip-2</i>	AGTAAA	-20	99

Gene pair	polyA signal	position	%SL1
F35D11.5 – <i>ipla-1</i>	AATAAA	-20	44
<i>cdd-2</i> – <i>pam-1</i>	AATAAA	-15	99
F52C12.2 – .6	AATAAA	-14	98
<i>marc-6</i> – C35E7.9	AATGAA	-14	none
<i>mff-2</i> – <i>tag-345</i>	AATAAA	-21	96
K04G7.11 – <i>rnp-7</i>	AATAAA	-18	99
<i>erv-46</i> – K09E9.3	GATAAA	-20	100
<i>mai-1</i> – <i>gpd-2</i>	none	NA	93
<i>hat-1</i> – <i>chl-1</i>	AATAAA	-27	33
M05D6.5 – <i>lact-4</i>	AATAAA	-28	97
<i>snfc-5</i> – <i>rnp-4</i>	AATAAA	-17	99
<i>mev-1</i> – <i>ced-9</i>	AATAAA	-14	98
T27E9.2 – <i>abcf-2</i>	AATAAA	-17	100
Y38C1AA.12 – <i>csn-3</i>	GATAAA	-22	99
<i>vha-11</i> – <i>vha-3</i>	AATAAA	-31	100
Y71H2AM.5 – .4	AATAAA	-18	9

Gene names are used when possible functional implications are evident. Trans-splicing data is from (Allen et al., 2011). The four operons with substantial levels of SL2 trans-splicing also have polyA addition sites further upstream, so these apparently can act as either SL2-type or SL1-type operons.

Table 6. Gene pairs that produce dicistronic mRNAs

Gene pair
W01A8.2 - .8
EEED8.18 - .13
B0334.15 - .5
Y39A1A.5 - .24
C30A5.2 – C02F5.7
Y62E10A.20 – .9
B0564.1
T07C4.11 - .12

Table 7. Non-operon gene clusters with a promoter of the downstream gene in the 3' UTR of the upstream gene.

Gene pair	polyA signal	position	%SL1
F07F6.4 – F07F6.8	AATGAA	-35	100
F08G5.1 – .2	AATAAA	-26	NA
F54A3.4 – Y59C2A.3	AATCAA	-22	99
F55A8.1 – F52C12.5	GATAAA	-32	100
M03F8.2 – M03F8.1	AATAAA	-34	100
T10E9.7	AATAAA	-41	100

Gene pair	polyA signal	position	%SL1
T24H10.7 – C16D2.1	AATAAA	-19	NA
Y17G7B.6 - .12	AATAAA	-51	100
Y32B12B.1 - .2	CATAAA	-19	100
Y56A3A.30 – .29	AATAAA	-44	99
Y57G11C.36 - .44	AATAAA	-21	NA
Y67D8C.4 - .3	AAAAAA	-24	100

The position of the polyA signal is given with respect to the trans splice site of the downstream gene in bp. Trans-splicing data is from (Allen et al., 2011).

Table 8. Use of a single exon by two genes

Gene pair
R06A10.1 (exon 5) - .5 (exon 1)
F28D9.4 (exon 4) – Y52B11C.1 (exon 1)
W09G3.8 (exon 3) - .2 (exon 1)
<i>cha-1</i> (exon 1) – <i>unc-17</i> (exon 1)
<i>unc-60a</i> (exon 1) – <i>unc-60b</i> (exon 1)

15. References

- Allen, M.A., Hillier, L., Waterston, R.H. and Blumenthal, T. (2011). A global analysis of trans-splicing in *C. elegans*. *Genome Res.* 21, 255-264. [Abstract Article](#)
- Baugh, L.R., Demodena, J., and Sternberg, P.W. (2009). RNA Pol II accumulates at promoters of growth genes during developmental arrest. *Science* 324, 92-94. [Abstract Article](#)
- Blumenthal, T. (2004). Operons in eukaryotes. *Brief. Funct. Genomic. Proteomics.* 3, 199-211. [Abstract Article](#)
- Blumenthal, T. and Gleason, K.S. (2003). *Caenorhabditis elegans* operons: form and function. *Nat. Rev. Genet.* 4, 110-118. [Abstract Article](#)
- Blumenthal, T., Evans, D., Link, C.D., Guffanti, A., Lawson, D., Thierry-Mieg, J., Thierry-Mieg, D., Chiu, W.L., Duke, K., Kiraly, M., and Kim, S.K.. (2002). A global analysis of the *Caenorhabditis elegans* operons. *Nature*: 417, 851-854. [Abstract Article](#)
- Cutter, A.D., Dey, A. and Murray, R.L. (2009). Evolution of the *Caenorhabditis elegans* genome. *Mol. Biol. Evol.* 26, 1199-1234. [Abstract Article](#)
- Guiliano, D.B. and Blaxter, M.L. (2006) Operon conservation and the evolution of *trans*-splicing in the Nematoda. *PLOS Genet.* 2, 1871- 1882. [Abstract Article](#)
- Huang P., Pleasance E.D., Maydan J.S., Hunt-Newbury R., O'Neil N.J., Mah A., Baillie D.L., Marra M.A., Moerman D.G., and Jones S.J. (2007). Identification and analysis of internal promoters in *Caenorhabditis elegans* operons. *Genome Res.* 17, 1478–1485. [Abstract Article](#)
- Jan, C.H., Friedman, R.C., Ruby, J.G. and Bartel, D.P. (2011). Formation, regulation and evolution of *Caenorhabditis elegans* 3' UTRs. *Nature* 469, 97-101. [Abstract Article](#)

- Lasda, E., Allen, M.A., and Blumenthal, T. (2010). Polycistronic pre-mRNA processing in vitro: snRNP and pre-mRNA role reversal in trans-splicing. *Genes Dev.* *24*, 1645-1658. [Abstract Article](#)
- MacMorris, M.A., Zorio, D.A.R., and Blumenthal, T. (1999). An exon that prevents transport of a mature mRNA. *Proc. Natl. Acad. Sci. U. S. A.* *96*, 3813-3818. [Abstract](#)
- Mangone, M., Manoharan, A.P., Thierry-Mieg, D., Thierry-Mieg, J., Han, T., Mackowiak, S.D., Mis, E., Zegar, C., Gutwein, M.R., Khivansara, V., *et al.* (2010). The landscape of *C. elegans* 3' UTRs. *Science* *329*, 432-435. [Abstract Article](#)
- Misra, S., Crosby, M.A., Mungall, C.J., Matthews, B.B., Campbell, K.S., Hradecky, P., Huang, Y., Kaminker, J.S., Milburn, G.H., Prochnik, S.E., *et al.* (2002) Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review. *Genome Biol.* *3* RESEARCH0083, 1-22. [Abstract Article](#)
- Morton, J.J., and Blumenthal, T. (2011). Identification of transcription start sites of trans-spliced genes: uncovering unusual operon arrangements. *RNA.* *17*, 327-337. [Abstract Article](#)
- Spieth, J., Brooke, G., Kuersten, S., Lea, K., and Blumenthal, T. (1993). Operons in *C. elegans*: Polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions. *Cell* *73*, 521-532. [Abstract Article](#)
- Whittle, C.M., Lazakovitch, E., Gronostajski, R.M., and Lieb, J.D. (2009). DNA-binding specificity and in vivo targets of *Caenorhabditis elegans* nuclear factor I. *Proc. Natl. Acad. Sci. U. S. A.* *106*, 12049-12054. [Abstract Article](#)
- Williams, C., Xu, L., and Blumenthal, T. (1999). SL1 trans-splicing and 3' end formation in a unique class of *Caenorhabditis elegans* operon. *Mol. Cell. Biol.* *19*, 376-383. [Abstract](#)
- Yin, J., Yu, L., and Savage-Dunn, C. (2010). Alternative trans-splicing of *Caenorhabditis elegans sma-9/schnurri* generates a short transcript that provides tissue-specific function in BMP signaling. *BMC Mol. Biol.* *11*, 46. [Abstract Article](#)
- Zaslaver, A., Baugh, L.R., and Sternberg, P.W. (2011). Metazoan operons accelerate recovery from growth-arrested states. *Cell* *145*, 981-992. [Abstract Article](#)

