Published in final edited form as:

Int J Parasitol. 2010 January ; 40(1): 63. doi:10.1016/j.ijpara.2009.07.001.

Functional analysis of putative operons in *Brugia malayi*

Canhui Liu^a, Ana Oliveira^b, Chitra Chauhan^a, Elodie Ghedin^c, and Thomas R. Unnasch^{a,*}

^aGlobal Health Infectious Disease Research Program, Department of Global Health, University of South Florida, Tampa, FL USA

^bGeographic Medicine, University of Alabama at Birmingham, Birmingham, AL USA

^cDivision of Infectious Diseases, Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania USA

Abstract

Operons are a common mode of gene organization in *Caenorhabditis elegans*. Similar gene arrangements suggest that functional operons may exist in *Brugia malayi*. To definitively test this hypothesis, a bicistronic reporter vector consisting of an upstream firefly luciferase gene and a downstream renilla luciferase gene was constructed. The genome was then surveyed to identify 16 gene pairs that were likely to represent operons. Two of four domains upstream of the 5' gene from these clusters exhibited promoter activity. When constructs replicating the promoter and intergenic arrangement found in the native putative operon were transfected into embryos, both firefly and renilla activities were detected, while constructs with the promoter alone or intergenic region alone produced no activity from the downstream reporter. These data confirm that functional operons exist in *B. malayi*. Mutation of three U-rich element homologues present in one of the operons resulted in a decrease in downstream renilla reporter activity, suggesting that these were important in mRNA maturation. Hemi-nested reverse transcriptase-PCR assays demonstrated that while the mRNA encoding the native downstream open reading frame of one operon contained an SL1 spliced leader at its 5' end, the renilla gene mRNA produced from the corresponding transgenic construct did not.

Keywords

Filariasis; Trans-splicing; Transfection; Biolistics; Operon

1. Introduction

Filarial parasites represent a significant public health problem throughout much of the developing world (Michael et al., 1996). There is currently a lack of effective drugs to treat these infections, making studies to identify new chemotherapeutic candidates a priority. As a result, studies of the basic biology of these organisms, and in particular studies of biological processes that do not appear to be shared between the parasite and its human host, are of particular importance.

© 2009 Australian Society for Parasitology. Published by Elsevier Ltd. All rights reserved.

*Corresponding author. Thomas R. Unnasch, Global Health Infectious Disease Program, 3720 Spectrum Blvd., Suite 304, Tampa, FL 33612 USA., Tel.: +1-813-974-0507; fax: 1-813-974-0992., tunnasch@health.usf.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

During the past decade, it has become clear that many of the genes of the free-living nematode *Caenorhabditis elegans* are organized into operons (Zorio et al., 1994; Blumenthal et al., 2002). This type of gene arrangement, while common among prokaryotes, appears to be confined to certain groups of eukaryotes. Most operons in *C. elegans* contain a small number of genes (mean 2.6) (Blumenthal et al., 2002), although the maximum number of genes found to be arranged in an operon is eight (Blumenthal et al., 2002). In *C. elegans* operons, transcription proceeds from a single promoter which is located upstream of the first gene in the operon (Spieth et al., 1993). Transcription proceeds through the upstream gene and then through a short intergenic spacer domain into the downstream genes, resulting in a polycistronic transcript. The resulting polycistronic pre-mRNA is then resolved into mono-cistronic mature mRNAs. This occurs through the process of trans-splicing. In *C. elegans* operons, the mature mRNA of the gene located at the 5' end of these clusters receives a spliced leader (SL) known as SL1, while the genes located downstream in the polycistronic pre-RNA are usually matured by the addition of SL2 sequences (Spieth et al., 1993). However, the SL1 can substitute for the SL2 in the maturation of some downstream mRNAs, particularly in operons where the intergenic region is large (Graber et al., 2007). Thus, identification of bona fide operons in *C. elegans* is facilitated by experiments that look for the presence of SL2 sequences at the 5' end of the mature mRNAs of genes whose function and genomic arrangement suggest that they might be part of an operon (Blumenthal et al., 2002).

The maturation of the polycistronic mRNAs produced from operons is somewhat complicated by the fact that this process would be expected to result in the production of an uncapped 5' end on the mRNA intermediate containing the downstream gene, something that normally would mark it for rapid degradation in the cell. This difficulty is overcome by the interaction of a special small nuclear ribonucleoprotein (SnRNP) containing the SL2 RNA with sequences encoded in the intergenic region, resulting in the addition of an SL2 to the downstream intermediate and the production of a mature SL2 containing mRNA (MacMorris et al., 2007). This process is mediated by a U-rich (Ur) domain encoded in the intergenic region (Huang et al., 2001). The Ur domain consists of a motif (with a consensus sequence of either UAUUUU or UUUUAU), that is generally located approximately 27 nucleotides (nt) downstream from the poly A addition signal of the upstream gene in the operon (Graber et al., 2007).

Analysis of the genomic sequence of the human nematode filarial parasite *Brugia malayi* has suggested that, like *C. elegans*, many of the genes in this parasite might also be arranged into operons (Guiliano and Blaxter, 2006; Ghedin et al., 2007). However, the presence of functional operons in *B. malayi* has been difficult to confirm. This is because in these parasites only the SL1 sequence is found on all mRNAs. Thus, any downstream genes in a putative operon would probably also receive an SL1 sequence at their 5' ends, making them experimentally indistinguishable from genes which are transcribed from their own promoter. One study has reported detecting an mRNA species spanning the two open reading frames (ORFs) of genes in a putative operon of *B. malayi* using reverse transcriptase-PCR (RT-PCR) (Guiliano and Blaxter, 2006). However, as pointed out by Blumenthal, (2004) such RNAs encompassing two adjacent genes might represent dead-end products and not physiologically relevant pre-mRNAs. Thus, proving that putative operon structures represent bona fide operons requires that one demonstrate the intergenic region does not contain a promoter, and that all transcription originates from the region upstream of the 5' gene of the cluster.

A transient transfection system based upon biolistic bombardment of isolated *B. malayi* embryos has been used to study promoter structure and trans-splicing in this human filarial parasite (Higazi et al., 2002, 2005; Shu et al., 2003; Higazi and Unnasch, 2004; Liu et al., 2007, 2009; Oliveira et al., 2008). We hypothesized that this method might also be used to test the hypothesis that operons exist in *B. malayi*. To accomplish this, a bicistronic reporter vector was constructed that included two reporter genes. This vector could then be used to prepare

constructs that recapitulated the structure of a putative operon but in which the upstream and downstream native ORFs were replaced by the separately assayable reporters. Here we report the results obtained in using this reporter vector to test the hypothesis that operons exist in *B. malayi* and to begin to examine the mechanisms involved in processing of nascent mRNAs generated from one such synthetic operon construct.

2. Materials and methods

2.1. Identification of putative operons for testing

A list of potential operons in the *B. malayi* genome had been previously developed by looking for genes in *B. malayi* which were in close apposition (within 4 kb), had common a transcriptional orientation and whose homologues were arranged in operons in *C. elegans* (Ghedini et al., 2007). Using this list as a starting point, we further limited our selection of putative operons by identifying gene pairs which: i) were not characterized by gene duplications or pseudogenes; ii) were comprised of genes separated by an intergenic distance below 500 bp; and iii) had a strong homology to characterized syntenic *C. elegans* operons. Taking into consideration the relative lack of short-range synteny observed between the caenorhabditid and *B. malayi* genomes, the presence of a syntenic cluster across both genomes was hypothesized to be a particularly powerful indication of potential operon conservation. Based upon these criteria, it was possible to narrow down the list of the over 3,500 operon candidates to 15 gene clusters which met these specific selection parameters (Table 1). These were then analyzed to deduce potential functional relationships among the predicted coding domains encoded within each cluster. To accomplish this, the information provided for the gene ontologies, protein family and domain databases available at the Uniprot website (<http://www.uniprot.org/uniprot/>) was analyzed to determine whether the gene pairs in each putative operon were functionally related. Gene pairs for which there was not enough annotation data available (e.g. hypothetical proteins) were characterized as “insufficient data”. Proteins encoded by the genes localized to the same organelle or interacting with the same type of molecules, were characterized as potentially related.

2.2. Preparation of a synthetic bicistronic reporter vector

The strategy used in the construction of the bicistronic reporter vector is summarized in Fig. 1. Construction began with pGL3 basic, a promoterless reporter vector containing a firefly luciferase reporter gene (Promega, Madison, WI, USA). The parental plasmid was modified by in vitro mutagenesis to introduce a *Bss*H2 restriction site at position 1,753 of pGL3 Basic, using the GeneTailor in vitro mutagenesis system (Invitrogen, Carlsbad, CA, USA), following the manufacturer's instructions. The *Bss*H2 site was introduced 14 bp downstream from the stop codon of the firefly luciferase reporter gene. The renilla luciferase gene and the downstream SV40 poly A addition signal from the reporter vector pRL null (Promega) was then amplified by PCR using a thermostable polymerase with error proofing capability (Pfu Turbo gold, Stratagene, La Jolla, CA, USA) employing a primer set containing a synthetic *Bss*H2 site at the 5' end of the 5' (coding) primer and a synthetic *Bam*H1 site at the 5' end of the 3' (non-coding) primer. The resulting amplicon was cloned into the traditional TA cloning vector (Invitrogen), individual clones isolated and the DNA sequence of their inserts confirmed. Plasmid DNA was prepared from one clone whose sequence matched that of the starting template DNA and digested with *Bss*H2 and *Bam*H1 to liberate the insert containing the renilla reporter and the SV40 poly A addition signal. The insert was then directionally cloned into the modified pGL3 basic plasmid digested with *Bss*H2 and *Bam*H1. This resulted in the production of a dual reporter plasmid containing the firefly luciferase gene and the renilla luciferase gene separated by a unique *Bss*H2 site (Fig. 1).

2.3. Preparation of synthetic operon constructs

The genomic sequence located upstream from the initiating codon of the upstream gene in each putative operon was extracted from the *B. malayi* genomic sequence and analyzed to develop primers that could be used to amplify an amplicon extending from position -1 relative to the start of the upstream ORF to a point approximately 1 kbp upstream of the start of the upstream ORF. This size fragment was chosen as previous studies have suggested that the core promoters of *B. malayi* genes reside in a region located within 1,000 bp of the initiation codon of the ORF (Higazi et al., 2005; Oliveira et al., 2008; Liu et al., 2009). The primers were designed with synthetic *Bgl*III restriction sites on their 5' ends to facilitate cloning of the promoter fragment into the bicistronic vector described above. Following amplification, the resulting amplicons were cloned into the classic TA cloning vector (Invitrogen) and the DNA sequence of individual clones were determined. Representative clones containing an insert with the predicted sequence were chosen and plasmid DNA was prepared. The inserted DNA encoding the putative promoter was then cloned into the unique *Bgl*III site which was located in the polylinker situated upstream of the firefly luciferase gene in the bicistronic vector. The DNA sequences of subclones containing the promoter fragments were then experimentally confirmed.

The intergenic regions of the putative operons (consisting of the entire sequence located downstream from the stop codon of the upstream ORF to the initiating codon of the downstream ORF) were similarly amplified from *B. malayi* genomic DNA employing primers designed with synthetic *Bss*H2 restriction sites on their 5' ends. The resulting amplicons were cloned and the DNA sequence of selected clones confirmed as described above. Inserts from clones whose DNA sequence were confirmed were excised with *Bss*H2 and cloned into the *Bss*H2 site separating the firefly and renilla luciferase reporter genes in the bicistronic vector. The DNA sequences of selected clones were experimentally confirmed. For each putative operon, three constructs were prepared, consisting of the promoter alone, the intergenic sequence alone, and both the promoter and intergenic sequence.

The putative Ur domains present in the intergenic region of the 14,993 gene cluster were mutated using the GeneTailor site directed mutagenesis system (Invitrogen). The primers were designed according to the instructions provided in the kit manual and the native sequences were replaced by the most distantly related nucleotides (A was replaced by C and G with T, and vice versa).

2.4. Transient transfection and analysis of reporter gene activity

Isolated *B. malayi* embryos were transiently transfected as previously described (Shu et al., 2003). In brief, 0.6 μ m gold beads coated with plasmid DNA were introduced by biolistic bombardment into embryos isolated by dissection from gravid adult female parasites. The embryos were maintained in culture for 48 h after bombardment. The embryos were then isolated by centrifugation, frozen and ground in passive lysis buffer (Promega, Madison WS, USA). Firefly and renilla activity in the homogenates was measured using Stop and Glo™ dual luciferase assay reagent (Promega) following the manufacturer's instructions. Protein content in the homogenates was determined using the Bradford method (Bradford, 1976). Each experiment consisted of triplicate transfections with each construct, and all constructs were assayed in at least two independent experiments.

2.5. PCR analysis of cis and trans-splicing

Total RNA prepared from 48 h post-transfection embryos was used as the template in hemi-nested RT-PCRs to detect evidence of trans-splicing in the downstream gene of the putative 14,933 operon, using a modification of a previously described method (Liu et al., 2007). Total RNA was purified from each batch of transfected embryos using Trizol (Invitrogen) following the manufacturer's instructions. A total of 1 μ g of this purified RNA was used as a template

in a single tube RT-PCR amplification reaction, employing Qiagen's One Step reagent (Qiagen, Valencia, CA, USA), following the manufacturer's protocol. The one step reaction contained primers derived from the SL1 sequence (5' GGTTTAATTACCCAAGTTTGAG 3') and the renilla gene encoded in the dual reporter vector (pRLNULL nc567: 5' ATTTGCCTGATTTGCCATA 3'). A total of 5 µl of this amplification reaction was used as a template in a 50 µl hemi-nested PCR reaction, employing Pfu Turbo DNA polymerase (Invitrogen), the buffer and amplification conditions recommended by the manufacturer. The hemi-nested PCR reaction employed the SL primer and a nested primer derived from the renilla gene (pRLNULL nc494: 5' ATGTCGCCATAAATAAGAAGAG 3').

5' rapid amplification of cDNA ends (5' RACE) was carried out to map the 5' ends of the mRNAs derived from the renilla gene in embryos transfected with the bicistronic construct representing the putative 14,933 operon. 5' RACE were carried out using a commercially available kit (Invitrogen) following the manufacturer's instructions. cDNA for the 5' RACE was primed using primer pRLNULL nc567, and the tailed cDNA amplified using the 5' adaptor primer provided by the manufacturer and pRLNULL nc494. The resulting amplicons were cloned into the classic TA cloning vector, and the DNA sequence of 20 randomly selected clones determined.

To detect transgenic mRNA encoding the renilla gene, total RNA purified from transgenic embryos as described above was extensively treated with RNase free DNase to eliminate contaminating DNA sequences, as previously described (Liu et al., 2007). The DNase treated RNA was then utilized as a template in a single tube RT-PCR amplification reaction, employing Qiagen's One Step reagent (Qiagen, Valencia, CA, USA), following the manufacturer's protocol. The primers used in the RT-PCR to detect the transgenic mRNA were pRLNULL c336 (5' CAAAGGAAACGGATGATAACT) 3' and pRLNULL nc567. A total of 5 µl of this amplification reaction was used as a template in a 50 µl hemi-nested PCR reaction, employing Pfu Turbo DNA polymerase (Invitrogen), the buffer and amplification conditions recommended by the manufacturer. The hemi-nested PCR reaction employed pRLNULL c336 and pRLNULL nc494. The amplification reactions were analyzed by agarose gel electrophoresis. The identity of amplicons produced in positive reactions was confirmed by DNA sequencing.

Trans-splicing in the native downstream gene in the 14,933 operon was assayed using a similar hemi-nested protocol. In these reactions, the SL-based primer was used in conjunction with a primer derived from the predicted ORF of the downstream gene (14,933 3'CD nc395: 5' CAGAACTGTCTTTTCATTTGT 3'), followed by a hemi-nested reaction employing the SL primer and a nested primer from the ORF (14933 3'CD nc209: 5' CTCGCCATAAATATCAGCTTC 3'). Products were analyzed by agarose gel electrophoresis and DNA sequencing as described above.

3. Results

3.1. Testing the hypothesis that operon-like gene clusters in *B. malayi* represent functional operons

Putative operons for this study were identified using a stringent set of selection criteria, as described in Materials and methods. A total of 15 gene clusters met all of these criteria (Table 1). Analysis of the predicted proteins encoded in these 15 gene clusters suggested that 3/15 clusters contained genes encoding proteins which were functionally related and 2/15 contained genes encoding proteins which exhibited some evidence for a functional relationship. Of the remaining gene clusters, 6/15 contained genes that encoded proteins for which no evidence for a functional relationship was found, while the annotation available for the remaining clusters was not sufficient to deduce a functional relationship.

Of the 15 putative operons identified in Table 1, four were chosen at random for further study. As a first step in characterizing these, the putative promoter domains located upstream of the 5' gene in each cluster were cloned into the bicistronic vector containing the firefly and renilla reporter genes. The resulting constructs were transiently transfected into *B. malayi* embryos and assayed for reporter gene activity. Of the four putative promoters assayed in this manner, two (14,933 and 14,406) produced consistently detectable levels of firefly luciferase activity (Fig. 2), indicating that they were active as promoters in the transient transfection system. These two putative operons were chosen for further study.

Constructs were then prepared in which the intergenic region of the 14,933 and 14,406 gene clusters were inserted into the *Bss*H2 site located between the firefly and renilla luciferase reporter genes. The intergenic regions were inserted into constructs which contained and lacked the upstream promoter domain. This process resulted in a set of three complementary constructs for each gene cluster; one containing the promoter alone, one the intergenic region alone and one containing both the promoter and the intergenic domain. The constructs were then transiently transfected into *B. malayi* embryos and assayed for both firefly and renilla luciferase activity. In both cases, the construct containing the promoter alone produced firefly luciferase activity, but no detectable renilla luciferase activity (Fig. 3). In contrast, constructs containing the intergenic region alone did not produce detectable levels of either firefly or renilla luciferase (Fig. 3), suggesting that the intergenic region was not a functional promoter. However, the constructs containing both the promoter and intergenic regions produced detectable levels of both firefly and renilla luciferase activity (Fig. 3). These results, when taken together, are consistent with the hypothesis that the 14,933 and 14,406 gene clusters represent functional operons in *B. malayi*.

3.2. Processing of native and transgenic pre-mRNAs from *B. malayi* operons

In *C. elegans*, polycistronic mRNA maturation is usually coupled to addition of an SL2 to the downstream mature mRNA (Spieth et al., 1993). However, as described above, SL2 sequences are not present in any *B. malayi* mRNA, suggesting that SL1 may be involved in polycistronic mRNA maturation. To test this hypothesis, primers derived from the predicted downstream gene in the 14,933 operon (encoding gene model 14933 m.00221, a predicted *B. malayi* homologue of pre-mRNA splicing protein prp5) were employed in a hemi-nested RT-PCR together with a 5' primer derived from the SL1 sequence, as described in Materials and methods. This produced an amplicon suggestive of a trans-spliced product when *B. malayi* RNA was used as a template (Fig. 4). DNA sequence of this amplicon confirmed that it represented a trans-spliced product derived from the predicted *B. malayi* mRNA.

In *C. elegans*, resolution of the polycistronic pre-mRNA into mature mRNA molecules involves cleavage associated with the addition of the SL2 to the downstream mRNA (Spieth et al., 1993). This process appears to involve recognition of a Ur element, a motif that is located approximately 30 bp downstream from the polyadenylation signal of the mature upstream mRNA (Huang et al., 2001; Graber et al., 2007). To determine whether sequences similar to the Ur element were important in polycistronic mRNA maturation in *B. malayi*, the intergenic region of the 14,933 gene cluster was examined for the presence of sequences similar to the *C. elegans* Ur element. Three Ur-like sequences were found in the intergenic region of the 14,933 cluster (Fig. 5A). To determine whether these Ur homologues were playing a role in trans-splicing, the three putative Ur elements in the 14,933 operon construct were individually mutated as described in Materials and methods. The mutant constructs were then transfected into embryos and the level of firefly and renilla luciferase activity resulting from the constructs assayed. As expected, mutation of the putative Ur elements had no effect on the amount of firefly luciferase activity seen (Fig. 5B). However the amount of renilla activity was reduced in all mutant constructs relative to that seen in the wild type construct. Mutation of the first

element resulted in the loss of approximately 90% of the renilla luciferase activity, while mutation of the second and third elements resulted in the loss of about 50–70% of the renilla luciferase activity (Fig. 5B).

Previous studies have shown that trans-splicing of the transgenic mRNAs derived from constructs transfected into *B. malayi* embryos required the presence of sequences encoded downstream of the start of the ORF (Shu et al., 2003). For example, the transgenic mRNA derived from constructs containing just the promoter derived from the gene encoding the *B. malayi* homologue of the 70 kDa heat shock protein (BmHSP70) is not trans-spliced (Shu et al., 2003). However, inclusion of sequences located downstream from the start of the BmHSP70 ORF upstream of the firefly luciferase reporter gene (including the first exon, first intron and part of the second exon) resulted in transgenic mRNA which was trans-spliced (Shu et al., 2003). In this case, trans-splicing appeared to require a 7 nt trans-splicing motif (TSM) encoded in intron 1, a motif which was both necessary and sufficient to direct trans-splicing (Liu et al., 2007). A TSM is not present in the predicted first intron of the *prp5* gene, the downstream gene of the 14,933 operon. However, similar to what has been seen in *C. elegans*, it was conceivable that SL addition to the downstream gene in the operons of *B. malayi* might proceed by a different mechanism than that used to trans-splice mRNAs derived from free-standing genes and the upstream-most gene in operons. If this was the case, a downstream TSM might not be necessary to obtain trans-splicing of the transgenic downstream gene in our synthetic operon constructs. To test this hypothesis, a hemi-nested RT-PCR employing primers derived from the SL and the renilla ORF was used to look for evidence of trans-splicing in transgenic mRNA derived from the downstream renilla reporter gene in embryos transfected with 14,933 operon construct. The hemi-nested RT-PCR employing the SL primer produced no product (Fig. 6A), suggesting that the renilla gene containing mRNA was not trans-spliced. In support of this observation, none of the 20 cloned 5' RACE products targeting this transgenic renilla mRNA, whose DNA sequence was determined, contained a trans-spliced leader sequence of any sort (data not shown). In contrast, a similar hemi-nested RT-PCR employing primers derived entirely from within the renilla ORF did produce a detectable product (Fig. 6B), demonstrating that renilla transgenic mRNA was present in the RNA preparation, and that this mRNA was detectable using a similar hemi-nested RT PCR assay.

4. Discussion

As mentioned in the Introduction, analogy by gene synteny has suggested that operons exist in *B. malayi*. However, this supposition has been difficult to prove experimentally. Here, we have developed a unique bicistronic reporter vector containing two individually assayable luciferase genes separated by a unique restriction site. Using this vector, it was possible to recapitulate the structure of putative *B. malayi* operons, replacing the upstream ORF in the operon with the firefly luciferase reporter gene and the downstream ORF with the renilla luciferase reporter gene. When transfected into *B. malayi* embryos, it was possible to determine whether these transgenic constructs indeed functioned as operons. In both putative operons tested using this system, renilla reporter gene activity was dependent upon the presence of both the putative promoter of the operon and the intergenic region. As this was the result expected for a functional operon, these data therefore provide experimental confirmation for the existence of functional operons in *B. malayi*.

It is clear from the data presented above that the intergenic regions of both the 14,933 and 14,406 gene clusters exhibit no intrinsic promoter activity, since constructs containing these sequences alone produced no renilla reporter gene activity. Thus, the presence of the promoter of the gene cluster was necessary for both reporters to be active. Interestingly, however, the promoter alone was not sufficient to produce renilla reporter gene activity, as constructs

containing the promoter alone also did not produce any renilla activity. These data suggest that renilla activity was not just a consequence of run-through transcription continuing through the firefly luciferase gene and into the renilla gene, but that activity of the downstream renilla reporter gene was also dependent upon sequences encoded in the intergenic region. Presumably, the role of the intergenic region is to ensure that the pre-mRNA is processed to mature mRNA molecules. Furthermore, these data suggest that translation of downstream ORFs in a polycistronic mRNA molecule (e.g. via an internal ribosome entry site) does not occur in *B. malayi*, or that this process is very inefficient.

In contrast to the expression pattern seen with the downstream renilla reporter gene, levels of expression of the upstream firefly luciferase reporter gene were not affected by the presence or absence of the intergenic regions. This is despite the fact that the sequences necessary for proper 3' end maturation and polyadenylation of the native upstream ORF of the gene clusters were encoded in the intergenic region. In previous studies which have employed the single reporter vector pGL3 basic to study promoter function, robust expression of firefly luciferase has been seen in the absence of the native 3' untranslated regions (UTRs) (Shu et al., 2003; Liu et al., 2007; Oliveira et al., 2008), and inclusion of a native 3' UTR into such constructs seems to have little or no effect on reporter gene activity (C. Liu and T.R. Unnasch, unpublished data). Presumably, the SV40 poly A addition signal present just downstream of the firefly luciferase ORF in the pGL3 basic vector is sufficient to support maturation of the transgenic mRNAs in this case. However, in the case of the bicistronic vector used in these studies, the vector-encoded poly A addition signal was separated from the firefly luciferase ORF by the renilla ORF. The fact that inclusion of the native intergenic region downstream from the firefly luciferase reporter (which encoded the native sequences necessary for 3' end maturation) had no effect on the amount of luciferase activity obtained suggests that the 3' UTR plays only a small role in determining either the amount of mature mRNA produced or its stability.

In *C. elegans*, most of the downstream genes in operons contain an SL2 at the 5' ends of their messages (Spieth et al., 1993). To date, no evidence of an SL2 sequence exists in *B. malayi*. In keeping with this observation, the native gene encoded in the 14,933 operon was found to contain an SL1 sequence at the 5' end of its message. SL1-capped mRNAs are found in some downstream genes in operons in *C. elegans*, generally occurring in those operons in which the intergenic region is large (Williams et al., 1999). Thus, it is possible that SL1 substitutes completely for SL2 in *B. malayi*, and that it is trans-spliced to upstream genes in introns, free standing genes and downstream genes in operons alike.

Polycistronic mRNA maturation also depends upon the presence of a Ur domain in *C. elegans*, which is located in the intergenic region, usually within 30 nt of the poly A addition signal for the upstream gene. It appears that homologues of the *C. elegans* Ur domain play a similar role in polycistronic mRNA maturation in *B. malayi*. There were three putative Ur domains located downstream of the putative poly A addition signal of the upstream ORF in the 14,933 operon, and mutation of each decreased the amount of downstream renilla reporter gene activity seen. Furthermore, the putative Ur domain that appeared to contribute the most to the expression of the downstream reporter was located 27 nt downstream from the putative poly A addition signal of the upstream gene (Fig. 5). This position is identical to the position where the Ur motif is generally located in *C. elegans*. These data suggest that the mechanisms involved in polycistronic mRNA processing may be substantially similar in the two nematodes.

Previous studies have demonstrated that sequences encoded downstream of the start of the ORF are necessary to permit trans-splicing in *B. malayi* monocistronic genes. For example, in the BmHSP70 gene a motif necessary for trans-splicing is encoded in the first intron of the gene (Higazi and Unnasch, 2004), while in the gene encoding a homologue of a RNA binding protein, sequences located in the second exon appear to be necessary to support trans-splicing

(Liu et al., 2007). The data presented here suggest that a similar situation might apply to the downstream genes in the operons of *B. malayi*. Thus, although the mature mRNA derived from the downstream native gene in the 14,933 operon was trans-spliced, the transgenic mRNA encoding the downstream renilla reporter gene was not trans-spliced. Further studies will be required to determine whether sequences encoded downstream of the start of the ORF are indeed necessary to support trans-splicing of such downstream genes, and if so how they might relate to the BmHSP70 TSM.

Although these studies provide evidence that the transgenic mRNA encoding the downstream renilla reporter gene is not trans-spliced, it is not clear what the composition of the 5' end of this mRNA is. Because expression of the downstream reporter requires the presence of the intergenic region and the Ur elements encoded therein, it is tempting to speculate that although these mRNAs are not trans-spliced, they may be partially processed, perhaps through a mechanism involving the Ur elements. Further studies are needed to resolve this question.

Acknowledgments

We would like to thank Dr. Naomi Lang-Unnasch for critically reading the manuscript. This research was supported by a grant from the United States National Institute of Allergy and Infectious Diseases (Project #R01AI048562) to TRU.

References

- Blumenthal T, Evans D, Link CD, Guffanti A, Lawson D, Thierry-Mieg J, Thierry-Mieg D, Chiu WL, Duke K, Kiraly M, Kim SK. A global analysis of *Caenorhabditis elegans* operons. *Nature* 2002;417:851–854. [PubMed: 12075352]
- Blumenthal T. Operons in eukaryotes. *Brief Funct Genomic Proteomic* 2004;3:199–211. [PubMed: 15642184]
- Bradford MM. A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem* 1976;72:248–254. [PubMed: 942051]
- Ghedini E, Wang S, Spiro D, Caler E, Zhao Q, Crabtree J, Allen JE, Delcher AL, Guiliano DB, Miranda-Saavedra D, Angiuoli SV, Creasy T, Amedeo P, Haas B, El-Sayed NM, Wortman JR, Feldblyum T, Tallon L, Schatz M, Shumway M, Koo H, Salzberg SL, Schobel S, Pertea M, Pop M, White O, Barton GJ, Carlow CK, Crawford MJ, Daub J, Dimmic MW, Estes CF, Foster JM, Ganatra M, Gregory WF, Johnson NM, Jin J, Komuniecki R, Korf I, Kumar S, Laney S, Li BW, Li W, Lindblom TH, Lustigman S, Ma D, Maina CV, Martin DM, McCarter JP, McReynolds L, Mitreva M, Nutman TB, Parkinson J, Peregrin-Alvarez JM, Poole C, Ren Q, Saunders L, Sluder AE, Smith K, Stanke M, Unnasch TR, Ware J, Wei AD, Weil G, Williams DJ, Zhang Y, Williams SA, Fraser-Liggett C, Slatko B, Blaxter ML, Scott AL. Draft genome of the filarial nematode parasite *Brugia malayi*. *Science* 2007;317:1756–1760. [PubMed: 17885136]
- Graber JH, Salisbury J, Hutchins LN, Blumenthal T. *C. elegans* sequences that control trans-splicing and operon pre-mRNA processing. *RNA* 2007;13:1409–1426. [PubMed: 17630324]
- Guiliano DB, Blaxter ML. Operon conservation and the evolution of trans-splicing in the phylum Nematoda. *PLoS Genet* 2006;2:e198. [PubMed: 17121468]
- Higazi TB, Merriweather A, Shu L, Davis R, Unnasch TR. *Brugia malayi*: Transient transfection by microinjection and particle bombardment. *Exp. Parasitol* 2002;100:95–102. [PubMed: 12054699]
- Higazi TB, Unnasch TR. Intron encoded sequences necessary for trans splicing in transiently transfected *Brugia malayi*. *Mol. Biochem. Parasitol* 2004;137:181–184. [PubMed: 15279965]
- Higazi TB, DeOliveira A, Katholi CR, Shu L, Barchue J, Lisanby M, Unnasch TR. Identification of elements essential for transcription in *Brugia malayi* promoters. *J. Mol. Biol* 2005;353:1–13. [PubMed: 16154590]
- Huang T, Kuersten S, Deshpande AM, Spieth J, MacMorris M, Blumenthal T. Intercistronic region required for polycistronic pre-mRNA processing in *Caenorhabditis elegans*. *Mol. Cell. Biol* 2001;21:1111–1120. [PubMed: 11158298]

- Liu C, de Oliveira A, Higazi TB, Ghedin E, DePasse J, Unnasch TR. Sequences necessary for trans-splicing in transiently transfected *Brugia malayi*. Mol. Biochem. Parasitol 2007;156:62–73. [PubMed: 17727976]
- Liu C, Chauhan C, Katholi CR, Unnasch TR. The splice leader addition domain represents an essential conserved motif for heterologous gene expression in *B. malayi*. Mol. Biochem. Parasitol. 2009 in press.
- MacMorris M, Kumar M, Lasda E, Larsen A, Kraemer B, Blumenthal T. A novel family of *C. elegans* snRNPs contains proteins associated with trans-splicing. RNA 2007;13:511–520. [PubMed: 17283210]
- Michael E, Bundy DAP, Grenfall BT. Re-assessing the global prevalence and distribution of lymphatic filariasis. Parasitology 1996;112:409–428. [PubMed: 8935952]
- Oliveira AD, Katholi CR, Unnasch TR. Characterization of the promoter of the *Brugia malayi* 12kDa small subunit ribosomal protein (RPS12) gene. Int. J. Parasitol 2008;38:1111–1119. [PubMed: 18364245]
- Shu L, Katholi C, Higazi T, Unnasch TR. Analysis of the *Brugia malayi* HSP70 promoter using a homologous transient transfection system. Mol. Biochem. Parasitol 2003;128:67–75. [PubMed: 12706798]
- Spieth J, Brooke G, Kuersten S, Lea K, Blumenthal T. Operons in *C. elegans*: polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions. Cell 1993;73:521–532. [PubMed: 8098272]
- Williams C, Xu L, Blumenthal T. SL1 trans splicing and 3'-end formation in a novel class of *Caenorhabditis elegans* operon. Mol. Cell. Biol 1999;19:376–383. [PubMed: 9858561]
- Zorio DA, Cheng NN, Blumenthal T, Spieth J. Operons as a common form of chromosomal organization in *C. elegans*. Nature 1994;372:270–272. [PubMed: 7969472]

Fig. 1.
Strategy for production of a bicistronic reporter vector.

Fig. 2. Activity of putative promoters from putative operons of *Brugia malayi*. Activities are expressed in as relative light units (RLU) (raw light units minus background) per milligram of total protein in the extract.

Fig. 3.

Firefly and renilla luciferase reporter gene activities in *Brugia malayi* embryos transfected with putative operon constructs. Constructs were prepared in the bicistronic vector and transfected into *B. malayi* embryos as described in Materials and methods. A) Constructs containing the promoter and intergenic regions from the putative 14933 operon. B) Constructs containing the promoter and intergenic regions from the putative 14406 operon. In each panel, paired columns represent firefly (F) and renilla (R) luciferase activities obtained from the same construct. Columns represent the mean, and error bars the S.D. of data obtained from six independent transfections. The tables below each graph indicate whether the promoter and/or intergenic domain was present or absent in each construct. Activities are expressed in as relative light units (RLU) (raw light units minus background) per milligram of total protein in the extract.

Fig. 4.

Trans-splicing of the native mRNA derived from the downstream gene in the 14933 operon (the *Brugia malayi* pre-mRNA splicing protein prp5 homologue). Trans-splicing was assayed using the hemi-nested reverse transcriptase-PCR (RT-PCR) assay described in Materials and methods. Lane 1 = hemi-nested spliced leader (SL)-mediated PCR positive control (hemi-nested SL mediated RT-PCR targeting the BmHSP70 gene employing mRNA prepared from embryos transfected with the construct BmHSP70/luc). Lane 2 = hemi-nested SL-mediated PCR targeting the *B. malayi* pre-mRNA splicing protein prp5 homologue.

A

TCTTAAGGAGTTTTTGGTCATGTTTGAAGTCTGGAATTTGCGTGCCATTTTCAAACTTGATGACTGATAATGGGTGCGGA
 TAATGGTCAAAATATCCTTAATTGTAATGTTAAGTAAGGATTTTCTGCTGCTGGAAAAATTAATTACTAATTTCCACAAAAGTTGTTT
 CCATTTTTTGTATTGTTTGTACGTAATCTTGCATAAAGCATTAAAAATTCCTCATCATTAG**TATTTT**GTTGTTCTGGCAG**ATT**
TATATATTTCAGAT**TATTTT**GTGAGGTGGTTAACTCATTTTTTAAAGTGTGCTTCTTACATGTAATGTAATACTATGAAATGAGGA
 TTTATTGAAGTTTTAACTTAATTAGTGAAGGTTAATTACCGTGGGTCGCCGTTTGCTTTAAGTAGGATATGCATTAAATTTTGGTA
 AATTTTGCTCTAATTTTTAATCGTTGTTGAGGATAATGGAACAAAT**TGCTG**CCCCGGATCTGA

↑

B

Fig. 5.

Analysis of the role of putative U-rich (Ur) domains in expression of the downstream renilla reporter gene in the 14933 operon. A) Sequence of the 14933 intergenic region. The putative Ur elements are indicated by bold type. Italic type indicates the sequence corresponding to the 5' end of the downstream open reading frame (ORF) of the *Brugia malayi* pre-mRNA splicing protein prp5 homologue. The putative poly A addition signal for the upstream gene is indicated by underlining. The arrow indicates the addition site for spliced leader 1 (SL1) on the native the *B. malayi* pre-mRNA splicing protein prp5 homologue mRNA, determined by DNA sequence analysis of the amplicon shown in Fig. 4. B) Firefly and renilla luciferase activity in embryos transfected with bicistronic constructs containing mutations in the putative Ur

domains. Ur domains are numbered by their relationship to the putative upstream poly A addition site, with Ur#1 representing the element closest to the poly A addition site. Columns represent the mean, and error bars the S.D. of six independent transfections. Activities are expressed in as relative light units (RLU) (raw light units minus background) per milligram of total protein in the extract.

Fig. 6.

Hemi-nested spliced leader (SL)-mediated reverse transcriptase-PCR (RT-PCR) analysis of transgenic mRNA in embryos transfected with the complete 14933 operon construct. mRNA isolated from embryos transfected with the bicistronic 14933 operon construct containing both the promoter and intergenic regions was assayed for trans-splicing in the resulting transgenic mRNA encoding the renilla open reading frame (ORF), as described in Materials and methods. A) Hemi-nested SL-mediated RT-PCR to detect trans-splicing in the renilla ORF-containing mRNA. B) Hemi-nested RT-PCR to detect renilla ORF-containing mRNA. In each panel, the lane labeled “pos” represents a positive control, the lane labeled “+RT” represents a hemi-nested RT-PCR carried out in the presence of reverse transcriptase and the lane labeled “-RT” represents a sham hemi-nested RT-PCR carried out in the absence of reverse transcriptase.

Table 1

Putative Operons in *Brugia malayi*.

Gene cluster	Functionally related	5' gene model	Predicted protein	<i>Caenorhabditis elegans</i> ortholog	Intergenic distance (bp)	3' gene model	Predicted protein	<i>C. elegans</i> ortholog
14479	None determined	14479.m00135	PSP family protein	ZK632.11	103	14479.m00134	Pre-SET motif family protein	R05D3.11
14975a	Yes	14975.m04524	Zinc finger motif, C2HC5-type family protein	Y53F4B.15	166	14975.m04523	Hypothetical protein	Y53F4B.15
14992	Insufficient data	14992.m10996	Hypothetical protein	C48B6.3	246	14992.m10995	TGF beta-inducible nuclear protein 1, putative	W09C5.1
14931	Insufficient data	14931.m00330	WW domain containing protein	T21D12.3	328	14931.m00329	Bifunctional aminoacyl-tRNA synthetase, putative	ZC434.5
14975b	None determined	14975.m04450	Translocon-associated protein gamma, putative	Y38F2AR.2	331	14975.m04449	Tudor domain containing protein	Y38F2AR.1
14406	Yes	14406.m00032	60S ribosomal protein L29, putative	B0513.3	336	14406.m00031	IL5 receptor binding protein, putative	B0513.2a
14975c	Potential	14975.m04541	MGC69179 protein-related	Y53C12B.1	351	14975.m04540	MGC89796 protein, putative	Y53C12B.2
15304	Insufficient data	15304.m00119	Probable NADH-ubiquinone oxidoreductase B18 subunit, putative	D2030.4	370	15304.m00118	Hypothetical protein	D2030.3
14424	Insufficient data	14424.m00388	Choline/ Carnitine o-acyl-transferase family protein	R07H5.2	418	14424.m00387	Conserved hypothetical protein	R07H5.3a
14367	Potential	14367.m00311	TM2 domain containing protein	Y66D12A.21	455	14367.m00310	Mitochondrial import inner membrane translocase subunit TIM10, putative	Y66D12A.22
14799	None determined	14799.m00205	Ubiquinone biosynthesis methyl-transferase ZK652.9, putative	ZK652.9	461	14799.m00204	Hypothetical protein	ZK652.10
14977	None determined	14977.m04905	Hypothetical protein	F56D1.2	469	14977.m04904	Zinc finger, C2H2 type family protein	F56D1.1
14250	None determined	14250.m00298	60S ribosomal protein L36, putative	F37C12.4	470	14250.m00297	Phospho-pantetheine attachment site family protein	F37C12.3
14933	Yes	14933.m00222	Pre-mRNA splicing factor ATP-dependent RNA helicase, putative	T05E8.3	482	14933.m00221	Pre-mRNA splicing protein pp5, putative	D1054.15
14990	None determined	14990.m08018	Small nuclear ribonucleo-protein-associated protein homolog F9F13.90 - <i>Arabidopsis thaliana</i> , putative	W08E3.1	488	14990.m08017	GTP-binding protein W08E3.3, putative	W08E3.3