Molecular & Biochemical Parasitology 125 (2002) 59 71

Abundant larval transcript-1 and -2 genes from Brugia malayi: diversity of genomic environments but conservation of 5 promoter sequences functional in Caenorhabditis elegans

Natalia Gomez-Escobar [a], William F. Gregory [a], Collette Britton [b], Linda Murray [b], Craig Corton [c], Neil Hall [c], Jen Daub [a], Mark L. Blaxter [a], Rick M Maizels [a,*]

[a] Institute of Cell, Animal and Population Biology, Uni ersity of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK
[b] Wellcome Centre for Molecular Parasitology, Uni ersity of Glasgow, Glasgow, UK
[c] Pathogen Sequencing Unit, Wellcome Trust Sanger Institute Genome Campus, Hinxton, Cambridge, UK

Received 15 July 2002; received in revised form 2 September 2002; accepted 2 September 2002

Abstract

The genomic organisation of two abundant lar al transcript (alt) genes from the filarial nematode Brugia malayi has been defined. The products of these genes are 78% identical in amino acid sequence, and are highly expressed in a stage-specific manner by mosquito-borne infective larvae. alt-1 is present as two near-identical copies organised in an inverted repeat of 7.6 kb, occupying a total of 16 kb of the genome. alt-2 is a single-copy gene at a different locus to alt-1. The two alt-1 genes (alt-1.1 and -1.2) are 99.7% identical in coding sequence and 99.5% in intronic sequences. Both alt-1 and -2 contain 3 introns, and the third intron of alt-2 exhibits a size polymorphism evident in different individual parasites from the laboratory-maintained strain. Genomic sequence up- and down-stream from alt-1.1/1.2 (26 and 6 kb, respectively) and alt-2 (6 and 4 kb, respectively) show that neither gene is in a multiple array or an operon. Most notably, the neighbouring genes of alt-1 and -2 show no similarity to each other, or to the genes flanking the distant alt homologue in Caenorhabditis elegans. Despite this diversity in flanking genes, the 5 UTR tracts extending some 800 bp upstream of each B. malayi alt gene show a high degree of similarity (overall 59% identity with tracts of 77 86% identity). Surmising that this region may contain conserved promoter elements, constructs containing the B. malayi alt 5 UTR with or without coding sequence were made fused to -galactosidase reporter protein. These constructs were injected into the syncytical gonad of C. elegans and progeny stained for -gal expression. Our results show relatively strong expression in the gut cells of C. elegans for both alt-1 and -2 constructs, commencing in larval worms and continuing into adulthood. Moreover, expression was enhanced when constructs contained segments of alt-1 coding and intronic sequence in addition to the 5 UTR. We conclude that the high level of alt transcription in filarial L3s is not due to expression from a multi-copy gene family but to a set of strong promoter elements shared between the two alt genes.

Keywords: Filarial nematode; Promoter; Transfection; Vaccine antigen

1. Introduction

The abundant lar al transcript (alt)-1 and -2 genes from Brugia malayi are among the most highly-expressed genes from the mosquito-borne infective larvae [1], and represent a gene family present in all filarial parasites [2 7]. The products, which have only low levels of similarity to known proteins from nonfilarial organisms, are of major interest on three accounts. First, they are expressed in a largely stage-specific manner: for example, alt-2 represents 3% of all ESTs from B. malayi L3, but not one transcript has been found among 18000 ESTs from all other stages in the life cycle [8]. Second, they have no known function, although the conservation of ALT products across the filarial taxon

Note: Nucleotide sequence data reported in this paper have been submitted to the GenBank™ data base with the accession numbers AY141874 and AY141875.

* Corresponding author. Tel.: 44-131-650-5511; fax: 44-131-650-5450
E-mail address: rick.maizels@ed.ac.uk (R.M. Maizels).

implies that these proteins play an essential role in establishment of infection [9]. Third, the ALT proteins are the antigens most strongly implicated in protective immunity [10], a proposition currently being tested in a number of experimental models.

Eukaryotic parasites are distinguished by their complex life histories which enable them to exploit multiple niches during their transmission cycle. The filarial nematodes exemplify this feature, because early larval stages (L1–L3) grow in arthropod vectors, readying themselves for entry into the vertebrate host [11]. The ALT proteins are synthesized in the later phase of growth in the mosquito, and in the infective L3 larva are found packaged in the oesophageal gland (Wu, Bianco, Maizels, Gregory, unpublished). This close association between *alt* expression and the L3 stage makes this protein family particularly attractive as a vaccine target [12].

We are interested in how filarial nematodes devote large proportions of their synthetic capacity to one or two specialised products, and how this synthesis is so closely co-ordinated with parasite development. For example, the *alt* genes may be present in multiple copies or arrays, a finding which would imply capacity for variation and escape from vaccine-induced immunity. Gene expression could be tightly linked in an operon as occurs for numerous *Caenorhabditis elegans* loci [13–15], and the identity of neighbouring genes might thereby provide a clue to the function of ALT proteins. Further, the *alt* genes must require powerful promoter motifs, subject to stringent stage- and tissue-specific regulation.

Little is currently known about promoter sequences in filarial parasites, due in no small part to the difficulty of achieving even transient transfection for functional testing in homologous systems [16,17]. However, success has been recorded for exploiting the free-living nematode *C. elegans* as tested for parasite nematode promoters. Transfection is relatively straightforward with well-developed vectors [18,19] and *C. elegans* has been shown to recognise promoter elements from nematode parasites such as *Haemonchus contortus* and *Ostertagia circumcincta* [20] as well as *Globodera rostochiensis* [21]. There are precedents also for filarial nematode promoters in work showing that *Brugia* heat-shock elements induce reporter gene expression in *C. elegans* at permissive temperatures [22], and that the 5′ UTR of *Onchocerca volvulus* glutathione S-transferase drives reporter gene expression in *C. elegans* [23]. If the *alt* promoters are highly active in *Brugia*, it may be surmised that filarial promoter sequences might be functional in *C. elegans*, and we test that proposition in this report.

Equally, little is currently known about the genomic organisation of filarial gene loci and only a few segments large enough to encode >1 gene (>5–10 kb) have so far been sequenced ([24]; J. Hirzmann, personal communication). We therefore determined the flanking sequence (5–26 kb) on either side of *alt-1* and *-2*, compared the upstream region presumed to contain regulatory elements, and identified other open reading frames. Our analysis is presented below.

2. Methods

2.1. Parasites and genomic DNA

Male adult jirds (*Meriones unguiculatus*) infected intraperitoneally with *B. malayi* organisms were purchased from TRS Labs (Athens, GA) and used as a source of adult parasites. *B. malayi* genomic DNA was prepared from 10 individual male worms as described previously [25].

2.2. Genomic PCR

Genomic DNA was used as the template for PCR reactions. 2 μl of DNA were mixed with the following *alt-2* specific primers: sense primer (nt 70–96 of cDNA, where the first ATG is 1–3, spanning intron 1) 5′-GAC GAA GAG TTC GAT GAC TCC GCA GCC-3′, antisense primer (nt 1525–1499 of genomic) 5′-GTA GTA TCA AAG ACT GAT TCA TTC CTA-3′. The PCR was cycled between 94, 55, and 72 °C (1 min each) for 35 rounds, followed by 1 round of 72 °C for 10 min.

2.3. Long-range PCR

The Boehringer Mannheim (Roche) Expand Long Range kit (Cat. No, 1-681-834) was used for long-range PCR, according to the manufacturer's instructions. Briefly, Buffer system 3 was combined with 15 pmol of a single primer, 1 μg of lambda DNA or of *B. malayi* genomic DNA, 10 mM of each dNTP in a final volume of 50 μl. To amplify the *alt-2* insert from lambda bacteriophage, T7 and T3 primers were used. To confirm the arrangement of tail-to-tail copies of *alt-1*, 3 pairs of sense and antisense primers were designed to the same sequence, namely nt 51–78 and reverse (in exon 1); 417–444 and reverse (in exon 2); and 909–937 and reverse (in intron 3). The primer sequences (sense strands only) were GCC ATG TAT GTC ACA ATC AGA TGA CGA A; CTC AGG AAT CCT CAG GTG CTG ATG AGG G; and CCG TTG TTT GCG AGT CAG TAA ATG CGA AC. PCR was performed on an MJ Research PT-200 DNA Engine machine as follows: 94 °C for 2 min, then cycles between 94 °C 10 s, 55 °C 1 min, 68 °C 10 min for 10 rounds and then between 94 °C 10 s, 55 °C 1 min, 68 °C 10 min 20 s plus a further 20 s every cycle for 15 rounds before a final incubation at 68 °C for 7 min.

2.4. Isolation of bacterial artificial chromosomes containing alt-1

The Filarial Genome Project 18 000-clone *B. malayi* large-insert *Hin*dIII partial digest BAC library (BMBAC1&2, estimated eightfold genome coverage, [26,27]) was screened for *alt-1* and *-2* genes by PCR and hybridization. Pooled samples of the gridded library were tested by PCR using primers specific to each gene. Only *alt-1* products were observed, using the following primers (sense primer, nt 431–451, 5′-GGT GCT GAT GAG GGA GGA GAT-3′, antisense prime, nt 712–686, 5′-GCA CAC AAT TTA CAA GTG TAG CAA TAT-3′). The row pools corresponding to the positive plates were then amplified with the same primers. This was followed by amplification of each of the 12 clones in the positive rows to determine the final coordinates of the positive BAC clones. This identified a 42-kb BAC, BMBAC07G12, containing the *alt-1* gene which was analysed by shotgun sequencing.

A second BAC, BMBAC47K06, was isolated by hybridization of the BAC library with an *alt-1*-specific probe as previously described [26]. The probe was the PCR product amplified using sense primer (nt 69–89, relative to ATG in the cDNA) 5′-GAT GAC GAA TTC GAC GAC GAA-3′, antisense primer (complementary to nt 1101–1078 of genomic sequence) 5′-TTG TTT TGC TTG CTT TGT AAG CAT-3′, which was labelled through the incorporation of biotinylated nucleotides by random priming (NEBlot Photo-tope™ kit; New England Biolabs, USA) and subsequently hybridized to the BAC library. Positive hybridizations were identified by chemiluminescent detection [26].

End sequencing of BMBAC07G12 (C. Whitton, J. Daub, N. Hall, M Quail and M. Blaxter, unpublished observations) showed that its SP6 end corresponded to the small subunit ribosomal RNA gene of *B. malayi* [28], while 47K06 was found to encode part of *alt-1* at the SP6 end. A third clone, BMBAC368C11, was identified from a *B. malayi* library constructed using partial *Sau*3A fragments (BMBAC3; C. Whitton, M. Quail and M. Blaxter, unpublished). This clone was found to overlap with BMBAC47K06 by comparison of end sequence information.

2.5. BAC sequencing

BMBAC07G12 was sequenced using a standard two-stage strategy involving random sequencing of subcloned DNA followed by directed sequencing to resolve problem areas. In the first stage, DNA prepared from BAC clones was shattered by sonification and fragments of 1.4–2 kb cloned into pUC18. DNA from randomly selected clones was sequenced with dye-terminator chemistry and analysed on automated sequencers. Each BAC was sequenced to a depth of sevenfold coverage. Contig assembly was performed using phrap (Phil Green, Washington University Genome Sequencing Center, unpublished). Manual base calling and finishing was carried out using Gap4 (http://www.mrc-lmb.cam.ac.uk/pubseq/manual/gap4_unix_1.html). Gaps and low quality regions were resolved by techniques such as primer walking, PCR and resequencing clones under conditions giving increased read lengths.

BMBAC47K06 was sequenced by primer walking from each end. Primers designed at 1 kb intervals from the 07G12 sequence were also used where the presence of a repeat was indicated. BAC DNA was prepared from 200-ml overnight culture broth using 4-ml volumes of Qiagen P1, P2 and N3 buffers in accordance with the manufacturer's instructions (Qiagen, UK), then transferring the resultant supernatant onto a Qiagen Midi-Prep column, elution with pre-warmed QF buffer, and isopropanol precipitation, ethanol wash, and resolublisation of the pellet in 200 μl of 0.1 × TE. BMBAC368C11 DNA was prepared by a modified mini prep procedure in which following alkaline lysis and neutralisation of a 10 ml overnight culture pellet with Qiagen buffers P1, P2 and N3, DNA was precipitated with ice-cold absolute ethanol, washed in 70% ethanol and air-dried before uptake in TE.

2.6. Isolation of lambda bacteriophage containing alt-2

Screening of the BMBAC1&2 library failed to isolate a clone containing *alt-2*. Therefore, a lambda bacteriophage genomic library (kindly provided by Dr Jörg Hirzmann, University of Gießen, Germany) was hybridised under high stringency conditions with a ^{32}P-labelled probe generated by PCR using primers spanning the second intron (sense primer, nt 476–502, 5′-CTG AGG AAT CGG TTC CCT TTC CAT CTC-3′, antisense primer, nt 865–839, 5′-GTA GCA AGC TTC GTG CGA AGA GCA CTC-3′). Plaque 2.1.1 positive for *alt-2* was amplified using long-range PCR (see above). The PCR product was subcloned into TOPO XL (Invitrogen) and fully sequenced by primer walking. Plasmid DNA was prepared using Qiagen mini prep spin columns and sequenced on an Applied Biosystems 377 automated sequencer using oligonucleotides 24 bp in length.

2.7. Preparation of lac Z reporter constructs

To investigate the expression pattern of the promoter region the 5′-end of the *alt* genes was inserted into the multiple cloning site of the *C. elegans lac Z* and *gfp* reporter vector pPD96.04 (kindly provided by the Fire Laboratory [18,19]). This vector encodes a nuclear localisation signal, targeting β-galactosidase to the nucleus allowing the detection of the expressing cell types.

BAC07G12 was used as the template for a set of PCR reactions as follows. Construct NG-033 was a transcriptional construct of *alt-1* using sense primer, 24246–24266 of 07G12, −1936 to −1916 relative to the ATG of *alt-1*, 5′-ACA TGC ATG CTT ATG TCT GAT AAG ACA TAT G-3′, *Sph*I site underlined; and antisense primer −1 to −22 relative to the ATG, 5′-ACG CGT CGA CTT TTT CCT TCT GGA AAA CAA CC-3′, *Sal*I site underlined. A 1939-bp fragment upstream of the *alt-1* ATG was isolated and subcloned into the *Sph*I/*Sal*I site of pPD96.04. Similarly, NG-035 was a transcriptional construct of the 2524-bp segment upstream of the *alt-2* ATG. This was amplified from plaque 2.1.1 by PCR (sense primer, 3512–3529 of lambda clone, −2523 to −2508 relative to the ATG of *alt-2*, 5′-ACG CGT CGA CTT ATT ATG TCT AAT AAA TTA A-3′, *Sal*I site underlined; and antisense primer, 5933–5912 of lambda clone, −1 to −20 relative to the ATG, 5′-CGC CGA TCC TTT CTC TTC CTG CAA AAA CAA C-3′, *BamH*I site underlined) and subcloned into the same vector.

In order to find out whether the presence of introns changes the pattern of expression, a translational fusion was also constructed using the same vector. NG-034 was a translational construct of *alt-1*, amplified from BAC 07G12 with the sense primer (−1936 to −1916) given above, and antisense primer (26965–26942 of 07G12, 784–761 relative to ATG, 5′-ACG CGT CGA CCT GCA AAT ATT TCA GAT TGG TGA A-3′, *Sal*I site underlined, so that the *alt-1* promoter, the first and second exons and the first and second introns are cloned in frame with the reporter vector (see Section 3.6).

2.8. DNA transformation of C. elegans

C. elegans was transformed by microinjection of plasmid DNA into the distal arm of the N2 hermaphrodite gonad as described previously [18]. Plasmid DNA for microinjection was prepared using Qiagen mini prep spin columns. *lac Z* and *gfp* reporter constructs at a final concentration of 50 μg/ml were co-injected with plasmid pRF4 [18,29] at a final concentration of 100 μg/ml. Plasmid pRF4 contains a dominant mutant allele of the *rol-6* gene and allows transformants to be identified by their right roller phenotype. Lines in which F2 and subsequent generations showed the roller phenotype were stained for β-galactosidase expression.

2.9. Fixation and staining for β-galactosidase activity

The staining procedure was carried out as described by Fire et al. [19]. Briefly, nematodes were fixed in 1.25% (v/v) glutaraldehyde, washed and dried onto glass slides. They were then fixed for 4 min in acetone at −20 °C before incubating in stain solution containing 0.01% (w/v) X-gal. Staining was carried out overnight at room temperature in a humid chamber. Co-staining with DAPI (4′,6-di-amidino-2-phenylindole) at a final concentration of 0.1% (w/v) aided in the identification of stained nuclei.

2.10. Bioinformatics

The BMBAC07G12 sequence was analysed using the Artemis workbench [30]. All sequences were used for searches at NCBI GenBank™ (BLASTN and TBLASTX on NR databases) and at http://nema.cap.ed.ac.uk/ncbi_blast.html for comparisons to known *B. malayi* sequences, including the Filarial Genome Project's 22 000 EST sequences [31,32]. Sequences were aligned with ASSEMBLYLIGN and analysed with MACVECTOR 7.0. Promoter motifs were examined on the MatInspector website, at http://www.gsf.de/cgi-bin/matsearch.pl.

3. Results and discussion

3.1. alt-2 is encoded by a single, polymorphic locus

The high level expression of ALT proteins could be generated by either a multiplicity of *alt* gene loci, or an unusually powerful promoter, or both. We first analysed the genomic organisation of *alt-2*, the most abundant transcript of the filarial L3 stage [1]. In our initial characterisation of the *alt-2* gene structure [10], it was noted that two forms were observed with respect to intron 3. Genomic copies of the *alt-2* gene contained either 468- or 341-nt introns, representing 6 or 9 copies, respectively of a ∼46-bp repeat. To investigate whether each form is derived from an independent gene, or if these are alleles from the same locus, we performed genomic PCR on individual adult male worms. The results, presented in Fig. 1, indicate that individual parasites encode different allelic forms of *alt-2*, with homozygotes and heterozygotes being evident for 1075- and 1212-bp products. Further studies have identified less common variants in the laboratory strain of *B. malayi* which differ in size of intron 3, and also appear to be allelic forms (Gregory and Callister, unpublished observations). Southern blot analysis, using genomic DNA from a large batch of adult worms, was consistent with a single gene locus (data not shown).

The genomic environment of *alt-2* was then studied, to ascertain whether this locus is arraigned in a likely operon, to investigate whether neighbouring genes could provide further indications of the function of the *alt* genes, and to identify potential promoter elements which may be responsible for the tightly-regulated high-level expression pattern.

Screening a *B. malayi* BAC genomic DNA library for *alt-2* proved unsuccessful, although we were able to

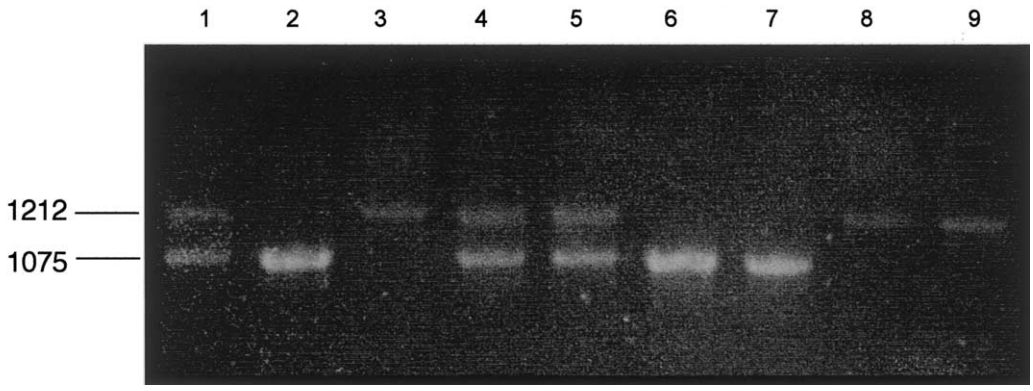

Fig. 1. Polymorphism of the *alt-2* gene in *B.malayi*. DNA samples from 9 individual adult male worms were used as the template with *alt-2*-specific primers in PCR. The 2 major bands correspond to known variants in the length of intron 3 [10].

isolate an *alt-1* clone from the same library (see below). We therefore screened a lambda bacteriophage genomic library, with smaller inserts, and identified a single positive clone. The insert of 11 794 bp was completely sequenced. Three additional genes were identified (Fig. 2). At the extreme 5′ end relative to *alt-2*, is a homologue of a highly conserved enzyme, 3-keto-acyl-CoA thiolase. Comparison to *C. elegans* identifies 4 out of the 5 exons of this gene within the insert; the 3′ end of the last exon is 4.8 kb distant from the ATG of *alt-2*. Within this intergenic region are two potential genes identified only by the presence of ESTs from *B. malayi*. One (at 3270–3643 of the clone) is 98% identical to a single EST (AI105513) with no open reading frame or poly-A tail, and which is chimaerised to a *Wolbachia* fragment. This we consider to be a genomic sequence which has been inadvertently cloned. The evidence that the second gene (at 3973–5142) is transcribed is more consistent, as there are 4 ESTs each of which appear to be spliced from 3 exons in the genomic sequence. Nevertheless, this gene, which pending further information we name *upt-1* (UP from alt-Two-1) is of uncertain status because of the presence of multiple stop codons in all frames. The *upt-1* transcripts begin only 758 bp 5′ of *alt-2*, and in the reverse direction. At the extreme 3′ end, 4.0 kb from the stop codon of *alt-2*, is the first exon of a steroid carrier termed Nuclear Factor I-A.

Neither 3-keto-acyl-CoA thiolase nor Nuclear Factor I-A are represented by ESTs, and there is no obvious association between either product and the L3 stage in which *alt-2* is so highly represented. Both 3-keto-acyl-CoA thiolase and Nuclear Factor I-A are located on Chromosome II of *C. elegans*, on which they are separated by nearly 5 Mb rather than the 10 kb observed in *B. malayi*. Information on these genes is summarised in Table 1.

3.2. The alt-1 genomic locus is dissimilar to that of alt-2

The *alt-1* genomic locus was characterised by analysis of BAC clones, from a library constructed by the Filarial Genome Project [8] with approximately eight-fold coverage. PCR amplification from pooled samples of the gridded library identified a clear positive for *alt-1*, and analysis of individual members of the positive pool identified an *alt-1*-encoding BAC (BMBAC07G12), which was ~42 kb in length. The SP6 end comprises 2 copies of the ribosomal RNA repeat spanning ~10

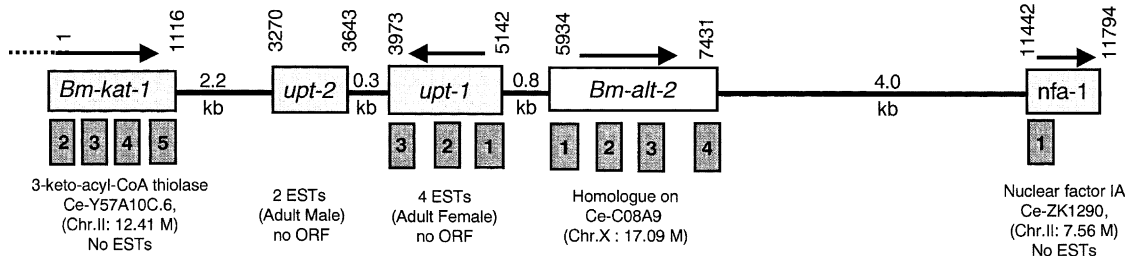

Fig. 2. Genomic map of *alt-2* based on a lambda bacteriophage clone. Larger boxes represent genes from initiating to terminating codons (including intronic sequences); smaller numbered boxes denote individual predicted exons. Arrows show direction of transcription, and intergenic intervals are indicated. Positions of ESTs from the Filarial Genome Project are indicated, and homologies with *C. elegans* genes provided. Note that neither of the flanking genes are completely encoded within the phage clone. Gene names are: *kat,* 3-keto-acyl-CoA thiolase; *upt*, UP from alt-Two; and *nfa*, Nuclear Factor IA. The sequence of this genomic clone has been deposited with the Accession Number AY141874.

Table 1

Position relative to *alt-1.1*	Gene abbreviation	Gene name	BM cluster number	*C. elegans* homologue	Other homologues	*B.malayi* ESTs (stages)
(A) *alt-1*						
−6100	*Bm-upo-1*	UP from One-1	None	D1069.3	*Meloidogyne*	None
−14276	*Bm-upo-2*	UP from One-2	05497	None	None	H52910 (L3)
−15004	*Bm-mog-4*	RNA helicase	None	C04H5.6	None	None
−20473	*Bm-prl-1*	Pleiotropic regulator	07552	D1054.15	*Heterodera, Meloidogyne, Strongyloides, Toxocara, Trichinella*	AA627027 (AF)
−25526	*Bm-upo-3*	UP from One-3	00224	None	None	H21290 (L3) N43490 (d10 L3) AA728016 (L4)
Position relative to *alt-2*	Gene abbreviation	Gene name	BM cluster number	*C. elegans* homologue	Other homologues	*B.malayi* ESTs (stages)
(B) *alt-2*						
+4011	*Bm-nfa-1*	Nuclear Factor I-A	None	ZK1290	None	None
−792	*Bm-upt-1*	UP from Two-1	01750	None	None	AA257460 (AF), AA509142 (AF), AA257592 (AF), AA509132 (AF)
−2289	*Bm-upt-2*	UP from Two-2	None	None	None	AI105513 (AM), W23369 (AM)
−4818	*Bm-kat-1*	3-keto-acyl-CoA thiolase	None	Y57A10C.6	None	None

kb. This region of the BAC has not been fully sequenced, in part because of problems with multiple copies of ~50 bp degenerate repeats in the extragenic spacer region of the ribosomal RNA cistron. Since the *alt-1* gene was ~26 kb distant from the rRNA locus, no attempt was made to complete the latter sequence, and the analysis presented here summarises the remaining 32442 nt of genomic information encoded on 07G12 and submitted to the GenBank™ data base.

Five candidate genes were identified between the ribosomal RNA repeat and *alt-1* (Fig. 3, Table 1); two of these have named homologues in *C. elegans* and the *B. malayi* genes have been named correspondingly. The remaining 3 have been given *upo* (UP from alt-One) gene names. The closest to *alt-1* is *upo-1*, a ~3-kb gene starting 6.7 kb upstream from the ATG of *alt-1*. *upo-1* is similar to a novel gene of no known function from *C. elegans* (*Ce*-D1069.3). After a further interval of 6.6 kb, a short sequence matches one *B. malayi* EST. This locus (*upo-2*) may not be a functional gene, as the single EST contains only a short ORF and no poly-A tail, and gives >400 nt uninterrupted identity to genomic DNA.

Two more genes are found in close succession which have homologues in *C. elegans*. *Bm-mog-4* encodes a homologue of *Ce*-C04H5.6 (*mog-4*), a DEAH-box RNA helicase of >1000 aa. *Bm-mog-4* closely matches *C. elegans* in the seven 3′ exons, but the 5′ end is less well-conserved and with no EST as yet from *B. malayi,* the open reading frame cannot currently be resolved. In *C. elegans*, *mog-4* is proposed to be a post-transcriptional regulator of the gametogenesis switch from sperm to oocyte production [33].

Just 392 bp upstream, and in the same orientation, is a homologue of *Ce*-D1054.15, a member of a cytosolic/nuclear protein family which includes *Arabidopsis* pleiotropic regulatory (*prl*) protein [34] and animal β-transducins. *Bm-prl-1* has 10 predicted exons, spanning just over 3 kb and encoding a 484-aa protein. Comparison of the *prl-1* sequences between *C. elegans* and *B. malayi* shows a matching intron–exon structure supported by the single adult EST in *B. malayi*. Similar sequences have been described from a range of nematode species (Table 1).

Adjacent to the first 50-bp repeat segment of the ribosomal RNA repeat, a novel gene encodes a predicted protein of 97 aa in two exons. This gene, which we have named *upo-3*, corresponds to the BMC00224 cluster containing 3 ESTs from larval *B. malayi*, but with no *C. elegans* homologue.

Comparison to the *C. elegans* genome reveals distinctly nonsyntenous evolution. Strikingly, the *B. malayi* homologues of *Ce*-D1054.15 (Chromosome V) and *Ce*-C04H5.6 (Chromosome II) are encoded sufficiently close together (392 bp apart) to suggest they have formed an operon. Moreover, in the *C. elegans* genome, D1054.15 is the first gene in a three-gene operon [15], but the other two genes are not represented in this segment of *B. malayi* sequence. A further example of disrupted synteny is seen in two adjacent *B. malayi* genes which have homologues in *C. elegans* in widely

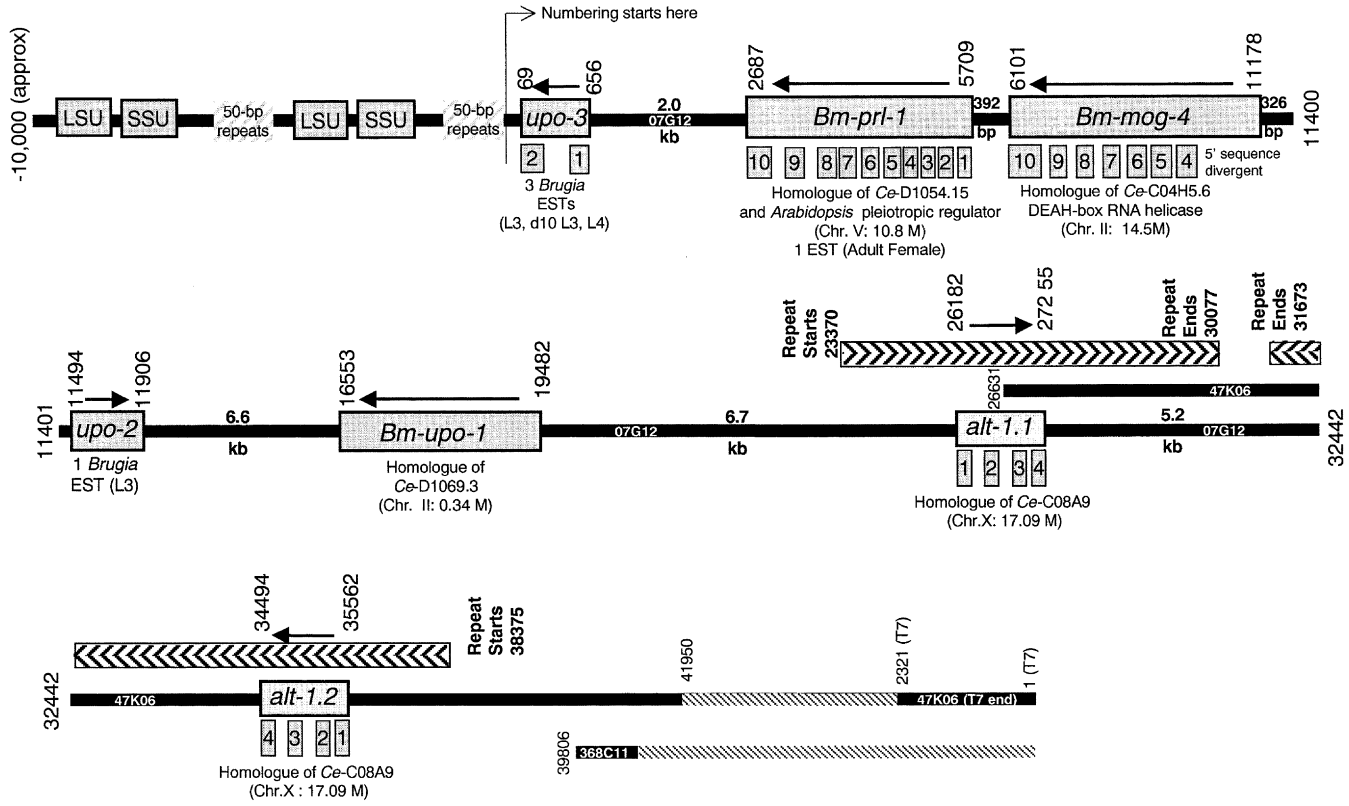

Fig. 3. Genomic map of the ~50 kb segment containing *alt-1.1* and *-1.2*. BMBAC07G12 comprises *alt-1.1*, 5 additional predicted genes, and two ribosomal RNA repeat units. BMBAC47K06 overlaps by 5.8 kb, includes the 3′ end of *alt-1.1* and the whole of *alt-1.2*. Solid lines represent sequenced portions of BACs, hatched lines unsequenced. Larger boxes represent genes from initiating to terminating codons (including intronic sequences); smaller numbered boxes denote individual predicted exons. The flecked boxes indicate the positions of the inverted repeat which includes the duplicated *alt-1* genes. The numbering commences as indicated, immediately 3′ of a 50-bp repeat segment which could not be unambiguously sequenced. Arrows show direction of transcription, and intergenic intervals are indicated. Gene names are: *upo*, UP from alt-One; *prl*, Pleiotropic ReguLator; *mog*, Masculinization Of the Germ line; *alt*, Abundant Larval Transcript. Names of *C. elegans* homologues and their loci in the genome are included. ESTs from the Filarial Genome Project are indicated where appropriate. The accession numbers for each sequence segment are as follows: BMBAC07G12 1–32442, AJ508355; BMBAC47K06, 34494-41950, AF540001; BMBAC47K06 T7 end, 1–2321, AY141875; BMBAC368C11, T7 end, 1–641, BH769870.

differing chromosomal locations: *alt-1* (C08A9, Chromosome X) and *upo-1*/D1069.3 (Chromosome II, 14 Mbp distant from C04H5.6).

In functional terms as well as in patterns of expression, none of the genes upstream of *alt-1* share any common features. The homologue of *Ce*-C04H5.6 has no EST as yet from *B. malayi*, and neither *upo-1*/*Ce*-D1069.3 nor *upo-3* have any known function in either *C. elegans* or *B. malayi*.

3.3. Duplication of the alt-1 gene locus

The final 765 nt at the 3′ end of 07G12 (31673–32442) forms an inverted repeat which closely matches 30077–29312 of the same BAC. To investigate this further, and to identify the neighbouring gene to the 3′ end of *alt-1*, we turned our attention to a second BAC, BMBAC47K06, which overlaps BMBAC07G12 by 5812 bp. End-sequencing showed that BMBAC47K06 encodes exons 3 and 4 of *alt-1* at its SP6 end. Interestingly, primer walking on 47K06 revealed the inverted repeat to be ~6.7 kb in length, and to contain an entire second copy of the *alt-1* gene. The interval between the two *alt-1* genes, which are in a tail-to-tail (3′ to 3′) orientation is 7.24 kb. This interval was confirmed by long-range PCR using SP6 and an *alt-1* exon 1-specific primer on BMBAC47K06 DNA (data not shown). The two genes were named *alt-1.1* (on BMBAC07G12 and in part on BMBAC47K06) and *alt-1.2* (on BMBAC47K06), as summarised in Fig. 3.

Primer walking on 47K06 extended for 5 kb upstream from the 5′ end of *alt-1.2* without identifying another putative gene. However, an identity was found with the T7 end of BMBAC368C11 (Fig. 3). This BAC was found not to contain additional copies of *alt-1* by PCR using two sets of *alt-1*-specific primers (data not shown).

To confirm that the inverted duplication of *alt-1* is an authentic representation of the *Brugia* genome, we carried out long-range PCR with genomic DNA, using a set of primers specific for Exon 1, Exon 2 and Intron 3

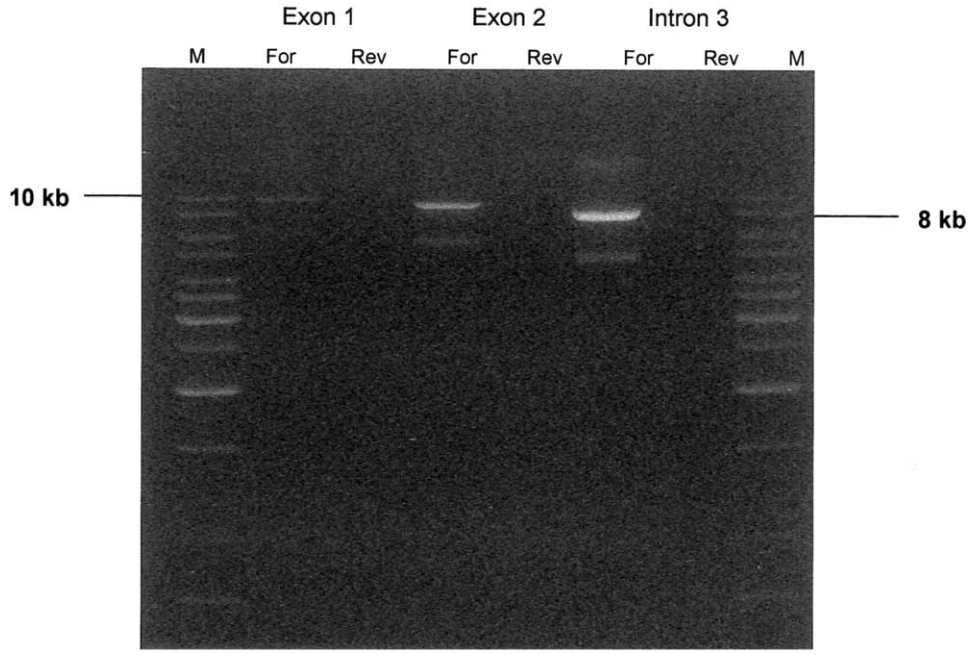

Fig. 4. Long-range PCR of *B. malayi* genomic DNA. Positive amplification was observed when single primers corresponding to the sense strand for *alt-1* were used (F, forward), but not when primers corresponded to the antisense strand (R, reverse). Three sets of primers were designed to nt 51–78 (in exon 1), 417–444 (in exon 2), and 909–937 (in intron 3). Lanes on either side marked M contain 1-kb DNA markers. The major products observed were ~10, 9 and 8 kb, respectively. Additional products about 3 kb smaller were also observed, which may be due to minor variation in genomic sequence in the nonclonal parasite population used to derive DNA.

of *alt-1*. As shown in Fig. 4, PCR using single primers in the forward direction gave products of the predicted size, but primers to the same sequence in the reverse sense gave none. As the different primers anneal at sites ~500 bp apart, and the genes are 3′ to 3′, their products differ, as expected, by 1000 bp.

The *alt-1.1* open reading frame (on BMBAC07G12, closer to the SSU repeats) is identical to the cDNA determined previously [1]. The *alt-1.2* gene shows a single amino acid difference (Lys-55 in place of Glu-55); interestingly this substitution is present in one of a family of related genes which are expressed at lower levels in *B. malayi* (Gregory et al. in preparation), but none of 36 ESTs examined show this substitution, which raises doubt whether alt-1.2 is expressed as a functional gene. This one difference among 375 coding nucleotides between *alt-1.1* and *-1.2* gives an identity of 99.7% at the nucleotide level. The two sets of intronic sequences differ by only 4/798 nt (0.5%).

3.4. Similarities of alt-1 and -2 introns, exons and noncoding regions

Both *alt-1* and *-2* share a similar gene structure. Previously we reported 2 exons in each gene, using a procedure which omitted the 5′ ends of the genes [10]. We can now provide a full description (Fig. 5). The second and third introns of *alt-2* are composed of repeat motifs absent from *alt-1*, and *-2* also contains more overall intron sequence than *alt-1* (1113 nt vs 699 nt),

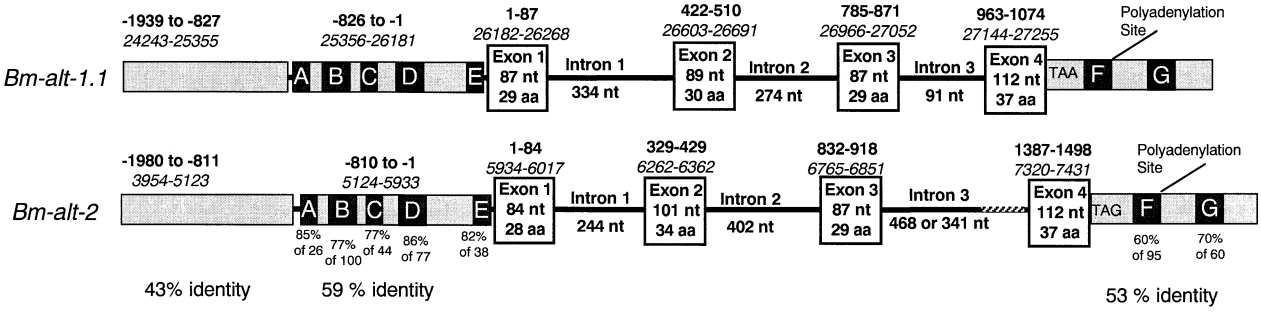

Fig. 5. Gene structures for *alt-1.1* and *-2*. Exons are in larger boxes, connected by introns shown as thin lines. Potential 5′ and 3′ regulatory regions are shown as thinner boxes. Motifs showing higher levels of similarity are lettered on dark boxes. The sequences of the 5′ motifs A–E are shown in Fig. 6. Bold numbering is relative to the ATG of each gene; light italic numbering is that of the genomic clones deposited in GenBank™ (Accession numbers given in legends to Figs. 2 and 3).

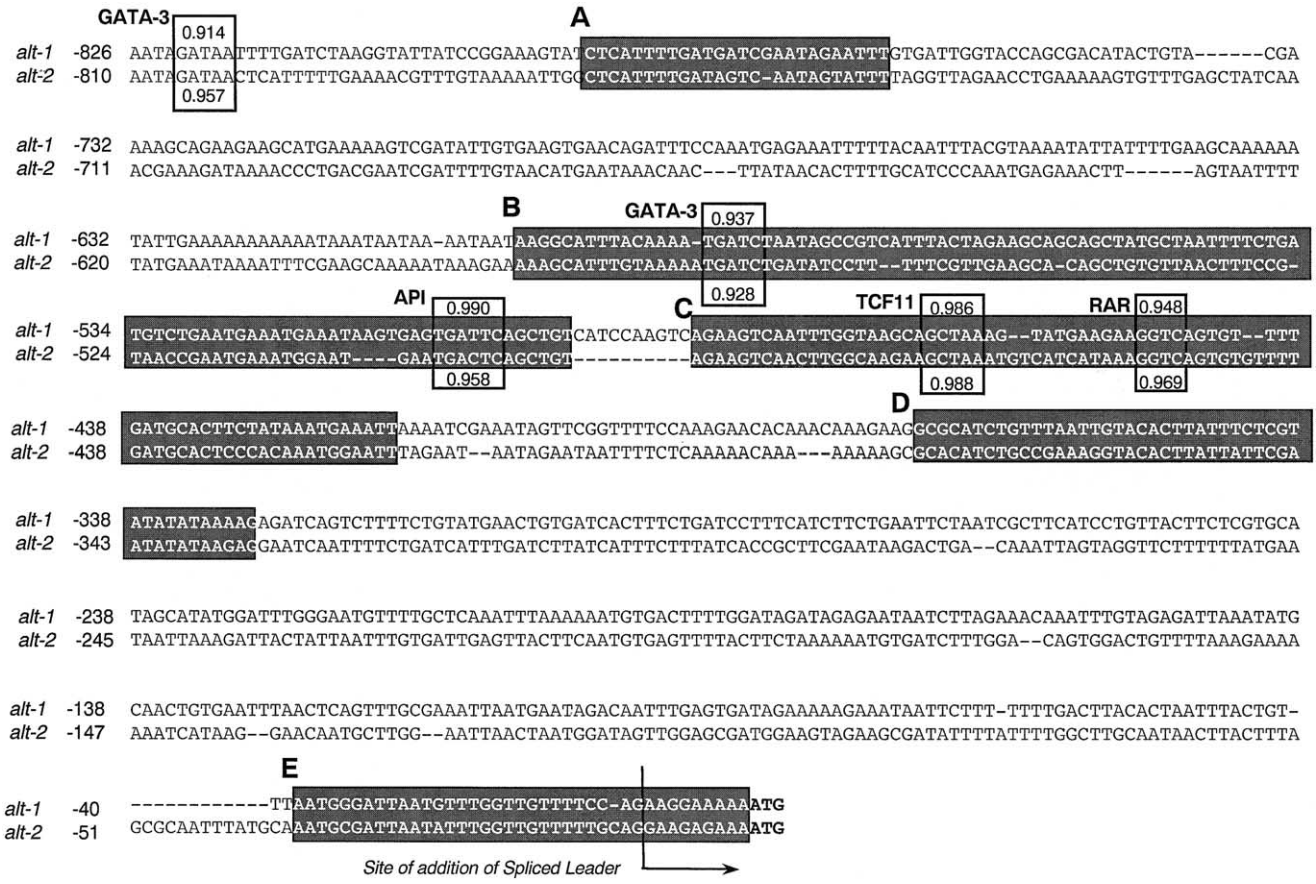

Fig. 6. Alignment of the 5′ noncoding regions of *alt-1* and *-2*. The dark boxes represent the 5 motifs with highest similarity between the two genes, notably focussed between 330 and 600 nt upstream from the ATG initiation codons. The open boxes signify promoter motifs which exceeded the criterion for matrix similarity of 0.9 for both *alt-1* and *-2*; the individual scores for each gene are given above and below the sequence respectively, together with the relevant transcription factor. The site of *trans*-splicing for mature mRNA formation is indicated close to the ATG.

which may be attributed to expansion of these repeat motifs.

We hypothesised that since both *alt-1* and *-2* are extremely highly transcribed in a stage-regulated manner, common regulatory motifs might be discerned in 5′ noncoding regions and other relevant sequences. We did indeed find extensive similarities (Fig. 5), with the 826 nt immediately upstream of the ATG of *alt-1* showing 59% identity with the corresponding segment of *alt-2*. Within this region are 5 shorter stretches of extremely high similarity, ranging from 77/100 to 22/26 (Figs. 5 and 6), containing consensus motifs for conserved transcription factors such as GATA-3, API, TCF11 and RAR (Fig. 6). Moreover, there are further homologous tracts in the 3′ region both before and after the polyadenylation site (Fig. 5). No significant similarities were found between the *C. elegans* genome and these 5 motifs and, nor with the 5′ UTR segments as a whole.

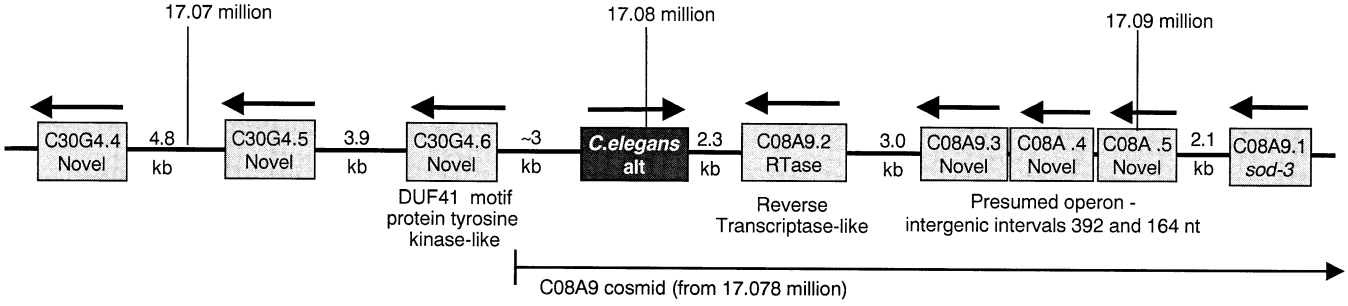

Fig. 7. Genomic environment of the *C. elegans alt* homologue on cosmids C30G4 and C08A9. Numbering represents Chromosome V positions as retrieved from wormbase (http://www.wormbase.org/). Arrows show direction of transcription, and intergenic intervals are indicated. Note that *C. elegans alt* does not have a gene number (unlike C08A9.2 for reverse transcriptase) as it was not identified by GeneFinder in the original annotation.

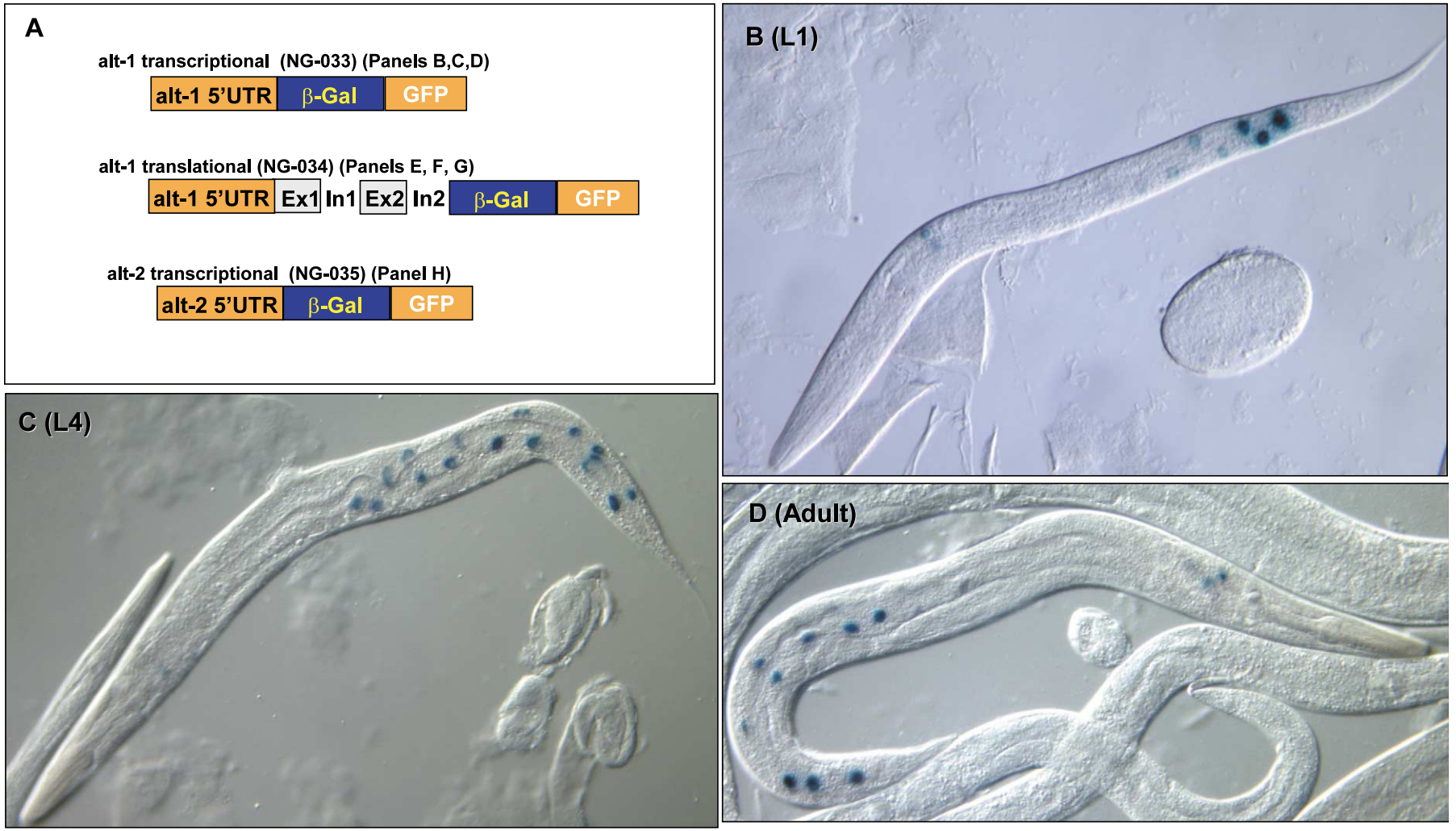

Fig. 8. Promoter analysis in *C. elegans*. (A) Schematic of 3 constructs in pPD96.04; (B) *alt-1* transcriptional construct (NG-033) in L1; (C) *alt-1* transcriptional construct (NG-033) in L4; (D) *alt-1* transcriptional construct (NG-033) in young Adults; (E) *alt-1* translational construct (NG-034) in L2; (F) *alt-1* translational construct (NG-034) in Adults; (G) *alt-1* translational construct (NG-034) under higher power; (H) *alt-2* transcriptional construct (NG-035) in L1.

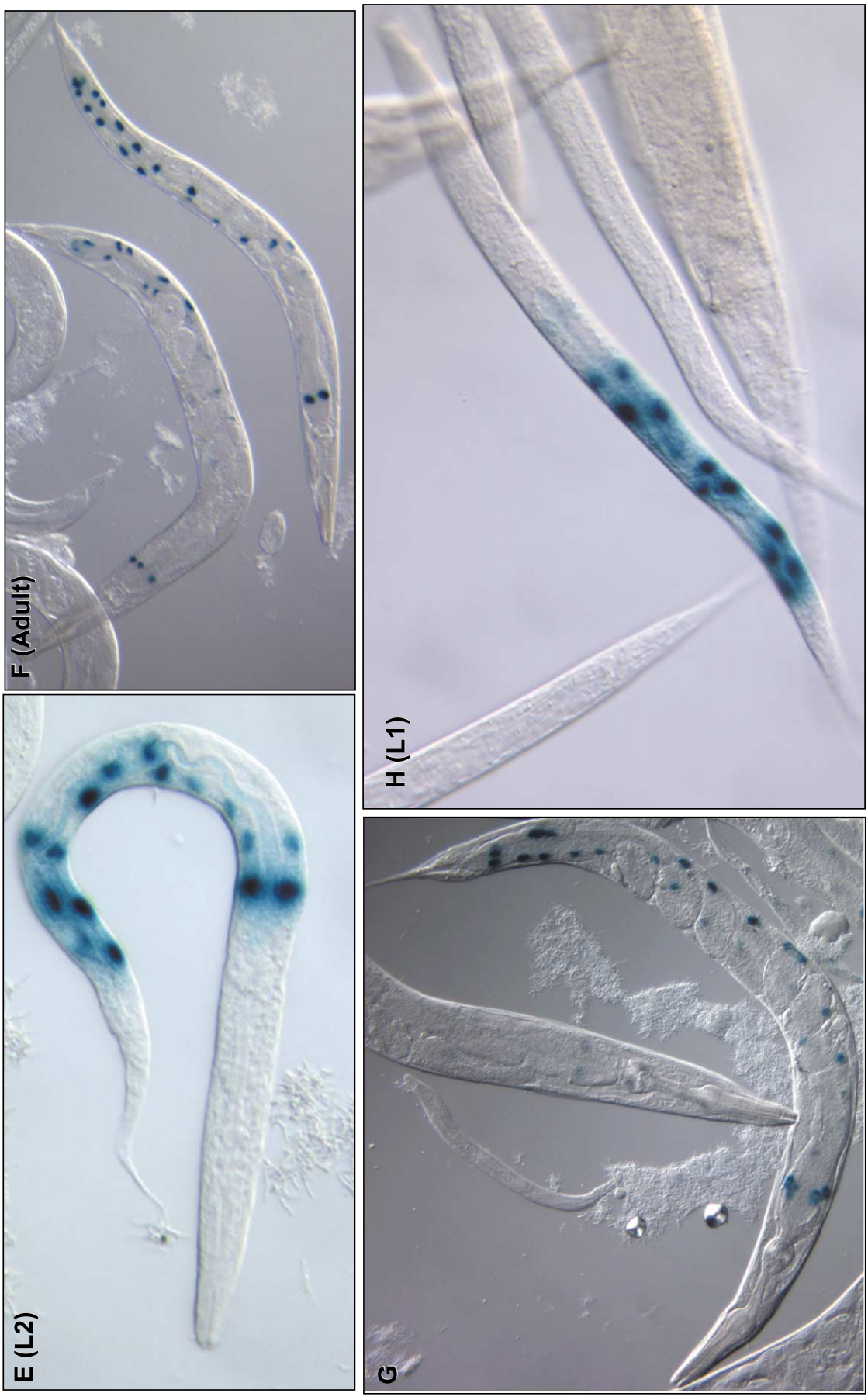

Fig. 8 (*Continued*)

3.5. Comparison with C. elegans locus on chromosome X

A significant similarity exists between filarial *alt* genes and a predicted ORF in *C. elegans*, on cosmid C08A9 from chromosome X [10]. *C. elegans alt* gene has not been identified by GeneFinder, and our earlier studies could not resolve which of several potential reading frames might be transcribed. In studies to be published elsewhere (Gregory et al. manuscript in preparation), 5′RACE was used to define two upstream exons of *Ce-alt* in addition to the 2 exons previously described [10]. Our analysis shows that is located in a ~7 kb interval between the novel gene C30G4.6 (which has a protein tyrosine kinase-like domain) and C08A9.2, a reverse transcriptase homologue (Fig. 7). There are no flanking genes for some distance up- and downstream with any similarity to those found neighbouring *Bm-alt-1* and *-2*.

3.6. B. malayi alt promoter analysis in C. elegans

The sequence similarities in the 5′ noncoding regions of *alt-1* and *-2* suggested a common promoter, and the high level of expression in filarial larvae implies that the *alt* gene promoter is a powerful one. To test these hypotheses, we transfected *C. elegans* with constructs containing *alt* upstream sequence with and without downstream exons and introns (Fig. 8A). The results obtained with these constructs are presented in Fig. 8B–H. All constructs elicited some expression, demonstrating successful recognition of *Brugia* regulatory sequences in *C. elegans*. The expression patterns of *alt-1* and *-2* transcriptional constructs were very similar, with gut-specific expression occurring most frequently in early larvae and in adults. Expression was predominantly in the posterior gut cells, with slightly less frequent expression in the anterior cells of the gut. A few worms also showed expression in midgut cells. It is notable that although *C. elegans* can recognise the alt promoters, expression does not display the same degree of stage- and tissue-specific regulation that is observed in the homologous filarial environment.

The translational construct of *alt-1* gave stronger expression, in more cells and in many more worms than observed with the transcriptional construct (Fig. 8E–G), suggesting that the introns include enhancer elements. Strongly staining cells in the anterior of the *alt-1* translational lines are the most anterior gut cells. At higher magnification, their position within the gut, just behind the pharyngeal bulb, can be seen (Fig. 8G). Expression in these cells occurs with the transcriptional constructs but at a lower level and in far fewer worms and is not so apparent (not shown).

4. Conclusion

The genomic information provided here reveals a number of important features of the filarial *alt* genes. First, our data indicates that *alt-2* is a single gene locus, which can nevertheless provide up to 3.2% of cDNA in the L3 stage. The *alt-1* locus presents a highly conserved inverted repeat, indicating a recent genetic duplication event. We show that the *alt-1* and *-2* loci are not closely linked to each other, and that neither is organised as an operon with any other gene.

Analysis of neighbouring genes reveals no conformity between the *B. malayi alt* genes, or indeed with the locus most similar in *C. elegans*. Moreover, where genes flanking *Bm-alt-1* and *-2* do have homologues in *C. elegans,* they are found in diverse parts of the genome, suggesting extensive disruption of synteny. These major rearrangements may be associated with the adaptation of *alt* structure and expression to the needs of a parasitic mode of life.

We have also inferred putative promoter sequences in the 5′ noncoding regions of these *B. malayi* genes. Although transfection protocols for *Brugia* are in their infancy [17], the ability to test filarial nematode promoter structure and function in a heterologous system [22,23] is invaluable, and in this light we have transfected *C. elegans* with *Bm-alt* reporter constructs. The complete gene sequencing of *B. malayi* within the next two years will also allow us to search for additional genes which may bear similarities either to the *alt* coding sequences or to the motifs associated with gene promoter activity.

Acknowledgements

We thank the Wellcome Trust for Programme Grant Support (NG-E and RMM), and the MRC for Fellowship Support to WFG, and for support to the Filarial Genome Project (MLB). We thank Jill Lovell of the ICAPB Sequencing Service for individual plasmid sequencing. We thank David Dresser for advice, Cecilia Fernandez for review of the manuscript, and David Guiliano for informatics expertise. We are also indebted to Yvonne Harcus and Janice Murray for providing *B. malayi* parasite material for this study and for the preparation of *B. malayi* DNA.

References

[1] Gregory WF, Blaxter ML, Maizels RM. Differentially expressed, abundant *trans*-spliced cDNAs from larval *Brugia malayi*. Mol Biochem Parasitol 1997;87:85–95.

[2] Bianco AE, Wu Y, Jenkins RE. *Onchocerca* spp: a 'family' of secreted acidic proteins expressed by infective larvae in blackflies. Exp Parasitol 1995;81:344–54.

[3] Frank GR, Tripp CA, Grieve RB. Molecular cloning of a developmentally regulated protein isolated from excretory–secretory products of larval *Dirofilaria immitis*. Mol Biochem Parasitol 1995;75:231–40.
[4] Joseph GT, Huima T, Lustigman S. Characterization of an *Onchocerca volvulus* L3-specific larval antigen, Ov-ALT-1. Mol Biochem Parasitol 1998;96:177–83.
[5] Pogonka T, Oberlander U, Marti T, Lucius R. *Acanthocheilonema viteae*: characterization of a molt-associated excretory/secretory 18-kDa protein. Exp Parasitol 1999;93:73–81.
[6] Allen JE, Daub J, Guilliano D, et al. Analysis of genes expressed at the infective larval stage validate the utility of *Litomosoides sigmodontis* as a murine model for filarial vaccine development. Infect Immun 2000;68:5454–8.
[7] Maizels RM, Gomez-Escobar N, Gregory WF, Murray J, Zang X. Immune evasion genes from filarial nematodes. Int J Parasitol 2001;31:889–98.
[8] Williams SA, Lizotte-Waniewski MR, Foster J, et al. The filarial genome project: analysis of the nuclear, mitochondrial and endosymbiont genomes of *Brugia malayi*. Int J Parasitol 2000;30:411–9.
[9] Maizels RM, Blaxter ML, Scott AL. Immunological genomics of *Brugia malayi*: filarial genes implicated in immune evasion and protective immunity. Parasite Immunol 2001;23:327–44.
[10] Gregory WF, Atmadja AK, Allen JE, Maizels RM. The *abundant larval transcript 1/2* genes of *Brugia malayi* encode stage-specific candidate vaccine antigens for filariasis. Infect Immun 2000;68:4174–9.
[11] Scott AL. Lymphatic-dwelling filariae. In: Nutman TB, Pasvol G, editors. Filariasis, 1999:5–39.
[12] Maizels RM, Holland M, Falcone FH, Zang XX, Yazdanbakhsh M. Vaccination against helminth parasites: the ultimate challenge for immunologists? Immunol Rev 1999;171:125–48.
[13] Zorio DAR, Cheng NN, Blumenthal T, Spieth J. Operons as a common form of chromosomal organization in *C. elegans*. Nature 1994;372:270–2.
[14] Evans D, Zorio D, MacMorris M, Winter CE, Lea K, Blumenthal T. Operons and SL2 *trans*-splicing exist in nematodes outside the genus *Caenorhabditis*. Proc Natl Acad Sci USA 1997;94:9751–6.
[15] Blumenthal T, Evans D, Link CD, et al. A global analysis of *Caenorhabditis elegans* operons. Nature 2002;417:851–4.
[16] Jackstadt P, Wilm TP, Zahner H, Hobom G. Transformation of nematodes via ballistic DNA transfer. Mol Biochem Parasitol 1999;103:261–6.
[17] 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.
[18] Mello CC, Kramer JM, Stinchcomb D, Ambros V. Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. EMBO J 1991;10:3959–70.
[19] Fire A, Harrison SW, Dixon D. A modular set of *lacZ* fusion vectors for studying gene expression in *Caenorhabditis elegans*. Gene 1990;93:189–98.
[20] Britton C, Redmond DL, Knox DP, McKerrow JH, Barry JD. Identification of promoter elements of parasite nematode genes in transgenic *Caenorhabditis elegans*. Mol Biochem Parasitol 1999;103:171–82.
[21] Qin L, Smant G, Stokkermans J, Bakker J, Schots A, Helder J. Cloning of a *trans*-spliced glyceraldehyde-3-phosphate-dehydrogenase gene from the potato cyst nematode *Globodera rostochiensis* and expression of its putative promoter region in *Caenorhabditis elegans*. Mol Biochem Parasitol 1998;96:59–67.
[22] Thompson FJ, Cockroft AC, Wheatley I, Britton C, Devaney E. Heat shock and developmental expression of hsp83 in the filarial nematode *Brugia pahangi*. Eur J Biochem 2001;268:5808–15.
[23] Krause S, Sommer A, Fischer P, Brophy PM, Walter RD, Liebau E. Gene structure of the extracellular glutathione S-transferase from *Onchocerca volvulus* and its overexpression and promoter analysis in transgenic *Caenorhabditis elegans*. Mol Biochem Parasitol 2001;117:145–54.
[24] Guiliano DB, Hall N, Jones SJM, et al. Conservation of long-range and micro-synteny between the genomes of two distantly related nematodes. Genome Biol 2002;3:57.1–57.14.
[25] Maizels RM, Robertson BD, Blaxter ML, Selkirk ME. Parasite antigens, parasite genes. A laboratory manual for molecular parasitology. Cambridge University Press, 1991.
[26] Foster JM, Kamal IH, Daub J, et al. Hybridization to high-density filter arrays of a *Brugia malayi* BAC library with biotinylated oligonucleotides and PCR products. Biotechniques 2001;30:1216–8, 1220, 1222, passim.
[27] Guiliano D, Ganatra M, Ware J, et al. Chemiluminescent detection of sequential DNA hybridizations to high-density, filter-arrayed cDNA libraries: a subtraction method for novel gene discovery. Biotechniques 1999;27:146–52.
[28] Blaxter ML, De Ley P, Garey JR, et al. A molecular evolutionary framework for the phylum Nematoda. Nature 1998;392:71–5.
[29] Kramer JM, French RP, Park E-C, Johnson JJ. The *Caenorhabditis elegans* rol-6 gene, which interacts with the sqt-1 collagen gene to determine organismal morphology, encodes a collagen. Mol Cell Biol 1990;10:2081–9.
[30] Rutherford K, Parkhill J, Crook J, et al. Artemis: sequence visualization and annotation. Bioinformatics 2000;16:944–5.
[31] Parkinson J, Whitton C, Guiliano D, Daub J, Blaxter M. 200000 nematode expressed sequence tags on the net. Trends Parasitol 2001;17:394–7.
[32] Blaxter M, Daub J, Guiliano D, Parkinson J, Whitton C. The *Brugia malayi* genome project: expressed sequence tags and gene discovery. Trans R Soc Trop Med Hyg 2002;96:7–17.
[33] Puoti A, Kimble J. The hermaphrodite sperm/oocyte switch requires the *Caenorhabditis elegans* homologs of PRP2 and PRP22. Proc Natl Acad Sci USA 2000;97:3276–81.
[34] Németh K, Salchert K, Putnoky P, et al. Pleiotropic control of glucose and hormone responses by PRL1, a nuclear WD protein, in *Arabidopsis*. Genes Dev 1998;12:3059–73.