Research

Conservation of long-range synteny and microsynteny between the genomes of two distantly related nematodes

DB Guiliano*, N Hall†, SJM Jones‡, LN Clark†, CH Corton†, BG Barrell† and ML Blaxter*

Addresses: *Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, UK. †Pathogen Sequencing Unit, The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. ‡Genome Sequence Centre, British Columbia Cancer Research Centre, Vancouver V5Z 4E6, Canada.

Correspondence: ML Blaxter. E-mail: mark.blaxter@ed.ac.uk

Published: 26 September 2002

Genome ***Biology*** 2002, **3(10)**:research0057.1–0057.14

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2002/3/10/research/0057

(Print ISSN 1465-6906; Online ISSN 1465-6914)

Received: 22 May 2002
Revised: 19 July 2002
Accepted: 22 August 2002

Abstract

Background: Comparisons between the genomes of the closely related nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae* reveal high rates of rearrangement, with a bias towards within-chromosome events. To assess whether this pattern is true of nematodes in general, we have used genome sequence to compare two nematode species that last shared a common ancestor approximately 300 million years ago: the model *C. elegans* and the filarial parasite *Brugia malayi*.

Results: An 83 kb region flanking the gene for *Bm-mif-1* (macrophage migration inhibitory factor, a *B. malayi* homolog of a human cytokine) was sequenced. When compared to the complete genome of *C. elegans*, evidence for conservation of long-range synteny and microsynteny was found. Potential *C. elegans* orthologs for 11 of the 12 protein-coding genes predicted in the *B. malayi* sequence were identified. Ten of these orthologs were located on chromosome I, with eight clustered in a 2.3 Mb region. While several, relatively local, intrachromosomal rearrangements have occurred, the order, composition, and configuration of two gene clusters, each containing three genes, was conserved. Comparison of *B. malayi* BAC-end genome survey sequence to *C. elegans* also revealed a bias towards intrachromosome rearrangements.

Conclusions: We suggest that intrachromosomal rearrangement is a major force driving chromosomal organization in nematodes, but is constrained by the interdigitation of functional elements of neighboring genes.

Background

All genomes encode conserved genes. The arrangement of these genes on chromosomal elements is determined by a balance between stochastic rearrangements and functional constraints. The level of conservation of gene order (synteny) and linkage between two genomes will depend on the relative contributions of inter- and intrachromosomal rearrangements. Whereas shared ancestry and functional constraints will increase conservation of linkage and synteny between taxa, rearrangement events will tend to randomize gene order over time. In the Metazoa, several gene clusters have been identified that remain linked because of functional constraints. These include the histone genes [1], the Hox gene clusters [2], the immunoglobulin cluster [3], and

the major histocompatibility complex (MHC) [4], but most genes are believed to be free to move within the genome. The tempo of gene rearrangement varies between taxa [5,6]. Vertebrate chromosomes are mosaic structures containing large conserved segments that can reside in different linkage groups in different species. There is a surprising conservation of synteny between distantly related species (approximately 450 million years (Myr) divergence) [7]. However, some lineages, such as rodents, show more extensive rearrangement than others, such as teleosts.

In protostomes, comparative studies of the genomes of closely related dipterans (*Drosophila* sp. and *Aedes aegypti* [5,8]) and nematodes (*Caenorhabditis elegans* and *C. briggsae* [6,9]) revealed a high rate of rearrangement. Chromosome rearrangements between closely related *Drosophila* species are mainly large pericentric inversions that may be facilitated by flanking transposon sequences [10,11]. *C. elegans* and *C. briggsae* are closely related, with estimates of 25-120 Myr divergence based on sequence comparisons [6,12]. Two groups have attempted to assess genome rearrangement rates and modes in comparisons between these two species. Kent and Zahler [9] compared 8.1 megabases (Mb) of fragmentary *C. briggsae* sequence derived from sequenced cosmid clones to *C. elegans* and derived a mean syntenic fragment length of 8.6 klobases (kb), or approximately 1.8 genes (there is one gene per 5 kb in *C. elegans*) [13]. In contrast, Coghlan and Wolfe [6], comparing 12.9 Mb of *C. briggsae* cosmid-derived sequence, found a mean syntenic fragment length of 53 kb. The difference appears to be purely methodological, as Kent and Zahler analyzed a subset of the data of Coghlan and Wolfe, and probably derives from a more relaxed definition of matching genes and use of cosmid fingerprinting physical map information by the latter study [6]. Estimation of rates of intrachromosomal to between-chromosome rearrangements showed that both were very frequent (approximately fourfold greater than that observed in *D. melanogaster*). Again, repeat sequences were associated with rearrangement boundaries [6]. It remains to be established whether this high rate of rearrangement is peculiar to the *Caenorhabditis* lineage, or is a general feature of nematode genomes.

To address this question we have begun analysis of a third nematode genome, that of the human filarial parasite *Brugia malayi*, which is estimated to have last shared a common ancestor with *C. elegans* 300-500 Myr ago [14]. *B. malayi* has a genome size of 100 Mb [15] and a gene complement estimated to be similar to *C. elegans* [16], and is the subject of a mature, expressed sequence tag (EST)-based genome project [16,17]. Unlike *C. elegans*, which has five autosomes and an XX/X0 sex-determination system [18], *B. malayi* has four autosomes and an XX/XY system [19]. The small size of condensed nematode chromosomes has precluded accurate *in situ* analysis of conservation of gene order. We have therefore taken a sequence-based approach, and here compare an 83 kb region surrounding the *B. malayi* macrophage-migration-inhibitory factor 1 locus (*Bm-mif-1*), a *B. malayi* homolog of a vertebrate cytokine [20], to the *C. elegans* genome and have found evidence for conservation of linkage and microsynteny between these two distantly related nematodes. The general features of this comparison were confirmed using a survey of genome sequences from *B. malayi*.

Results

General sequence features of an 83 kb segment of the *B. malayi* genome

Two overlapping bacterial artificial chromosome clones (BACs) were isolated that spanned the *Bm-mif-1* locus. The inserts of BMBAC01L03 and BMBAC01P19 were 28,757 base pairs (bp) and 64,685 bp, respectively, with 10,637 bp of overlap, yielding a contiguated sequence of 82,805 bp (Figure 1). AT content overall was 68.0%; exonic DNA had an AT content of 59.9% and intergenic and intronic DNA had AT contents of 69.3% and 70.4% respectively. The average predicted gene size was 4.7 kb (range 0.6-20 kb). The average distance between genes was 3.1 kb (range 0.3-10.5 kb), giving an average gene density of one gene per 6.9 kb. There was an average of 9.3 introns per gene, with an average intron length of 316 bp (range 48-2,767 bp). The *C. elegans* orthologs of the *B. malayi* genes (see below) had a mean length of 3.2 kb, with an average of 5.5 introns per gene (mean size of 142 bp). The *B. malayi* genes were longer as a result of increased mean length and number of introns. Comparison to *C. elegans* presumed orthologs (see below) showed that only 50% of *C. elegans* introns were conserved in *B. malayi* (29 of 56 introns), and 25% of *B. malayi* introns (29 of 107) were conserved in *C. elegans* (Table 1). Of the 12 predicted *B. malayi* genes, seven were tested and confirmed by cDNA-PCR, and alternatively spliced transcripts were identified for four. Five of the 12 genes had corresponding ESTs (Table 1).

Comparison of predicted genes to *C. elegans*

All 12 predicted genes had *C. elegans* homologs, but putative orthology could only be assigned to 11 pairs (Figure 1, Table 1). Orthology definition is possibly problematic, as the complete genome sequence of *B. malayi* is not known, and it is thus possible that genes more similar to these *C. elegans* comparators could be present. We note, however, that no *B. malayi* EST-defined genes (23,000 ESTs defining approximately 8,300 genes [16]) have better matches to these *C. elegans* proteins (data not shown), and that orthology definition included coextension of the proteins, and conservation of intron position and phase (Table 1). The exception, BMBAC01L03.3, contained two domains, an amino-terminal LON ATP-dependent serine protease domain (domain PF02190) and an anonymous carboxy-terminal domain (PFB022940). Proteins predicted from the *Arabidopsis thaliana* (AAC42255.1), *Mus musculus* (NP_067424), and *Homo sapiens* (XP_0421219) genomes

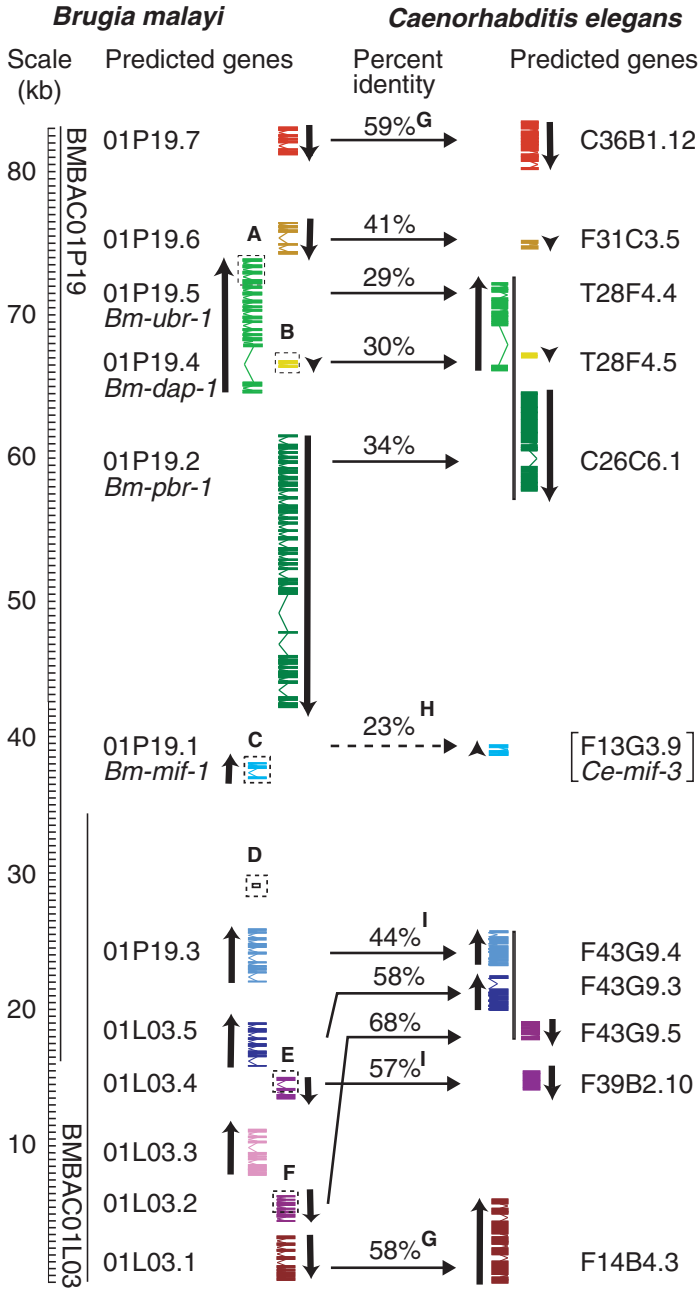

Figure 1
The BMBAC01L03/BMBAC01P19 contig compared to the *C. elegans* genome. Genes are indicated by exon (box) and intron (bracket) structures. For each species, the direction of transcription of the genes is indicated by an arrow. The *C. elegans* gene structures are drawn to the same scale as the *B. malayi* contig. A, Match to *B. malayi* EST cluster BMC03169 [16]. *Brugia* EST (BMC) and *Onchocerca volvulus* (OVC) clusters are viewable in NemBase [39,60]. B, Highly similar to *O. volvulus* EST cluster OVC02481 [61]. C, Match to *B. malayi* EST cluster BMC00238. D, Match to *B. malayi* EST clusters BMC02055 and BMC01932. However, no ORF was identified, and it may not represent protein-coding sequence (see text for discussion). E, Match to *B. malayi* EST cluster BMC06334. F, Match to *B. malayi* EST cluster BMC00400. G, BMBAC01L03.1 and BMBAC01P19.7 are gene fragments. Percent identity was calculated on the alignable portion of the *C. elegans* ortholog. H, F13G3.9 (*Ce-mif-3*) is on *C. elegans* chromosome I. However, F13G3.9 is not the predicted ortholog of *Bm-mif-1* and thus the relationship is indicated by a dashed arrow (see text). I, Percent identity was calculated for BMBAC01P19.3 and BMBAC01L03.4 only within the PWWP or dnaJ domains respectively. Homolog pairs are indicated by the coloring of the gene models.

share this architecture, but there are no *C. elegans* proteins that have both domains.

Some genes were similar to hypothetical, functionally uncharacterized genes from *C. elegans*. BMBAC01P19.7a/b

Table 1

Genes predicted on the BMBAC01L03/BMBAC01P19 contig

B. malayi open reading frame	Predicted cDNA length (bp)	Predicted peptide length	Number of introns	*C. elegans* ortholog	Percent identity with *C. elegans* ortholog	Number of introns in *C. elegans* ortholog	Number of shared intron positions with *C. elegans* ortholog	Putative identity
BMBAC01L03.1	1340*	446*	7*	*Ce*F14B4.3	58†	3‡	3	Amino-terminal fragment of the β subunit of RNA polymerase I
BMBAC01L03.2	693	230	6	*Ce*F43G9.5	68	3	1	Pre-mRNA cleavage factor
BMBAC01L03.3	1239	412	8	-	-	-	-	Contains LON-ATP-dependent serine protease domain
BMBAC01L03.4	630	209	2	*Ce*F39B2.10	57§	3	1	Contains dnaJ domain
BMBAC01L03.5	918	305	6	*Ce*F43G9.3	58	6	2	Mitochondrial carrier protein
BMBAC01P19.1 (*Bm-mif-1*)	535	115	2	*Ce*Y56A3A.3	41	2	2	Macrophage-migration-inhibitory factor homolog
BMBAC01P19.2a/b (*Bm-pbr-1*)	5955/5748	1934/1865	37/35	*Ce*C26C6.1	34	14	9	Polybromo domain protein, BAF180 homolog
BMBAC01P19.3 a/b	1182/919	367/283	9/7	*Ce*F43G9.4	44¶	8	2	Contains PWWP domain
BMBAC01P19.4 (*Bm-dap-1*)	446	111	1	*Ce*T28F4.5	30	1	1	Homolog of mammalian death-associated protein DAP-1
BMBAC01P19.5a/b (*Bm-ubr-1*)	2679/2602	847/821	18/17	*Ce*T28F4.4	27	12	5	Unknown
BMBAC01P19.6	804	190	4	*Ce*F31C3.5	41	1	1	Conserved protein of unknown function
BMBAC01P19.7a/b	1039/932*	274/298*	6/7*	*Ce*C36B1.12	60#	3‡	2	Carboxy-terminal fragment of a novel transmembrane protein

*Gene fragments (see text). †BMBAC01L03.1 gene fragment aligned with the amino-terminal 450 amino acids of *Ce*F14B4.3. ‡Number of introns in the aligned portion of the *C. elegans* ortholog. §Percent identity over the dnaJ domains of BMBAC01L03.4 and *Ce*F39B2.10. ¶Percent identity over the PWWP domains of BMBAC01P19.3 and *Ce*F43G9.4. #The gene fragment of BMBAC01P19.7 aligned with the carboxy-terminal 380 amino acids of *Ce*C36B1.12.

had multiple predicted transmembrane segments also found in a number of peptides from other species (PFB002843) and were most similar to C36B1.12 (60% identity). There is only one homolog of BMBAC01P19.3a in any organism - F43G9.4 from *C. elegans*. The amino termini of both BMBAC01P19.3a and F43G9.4 contained PWWP domains (PF00855). PWWP domains are found in proteins with nuclear location and roles in cell growth and differentiation [21,22]. PSORT profiling indicated that BMBAC01P19.3 and F43G9.4 were likely to have nuclear localizations. The amino terminus of BMBAC01L03.4 contains a dnaJ-like domain (PF00684). The dnaJ domain is found in 41 *C. elegans* proteins, but BMBAC01L03.4 showed highest identity (57%) to F39B2.10. Both proteins had the dnaJ domain at their amino terminus and shared a common position of the first intron in this region. The remainder of the protein was not conserved.

BMBAC01P19.1 encodes *Bm-mif-1* (Figure 2) [20]. Mammalian MIF is a cytokine involved in inflammation, growth, and differentiation of immune cells [23]: *B. malayi* MIF-1 may have a role in immunomodulation of the host [20,24]. *C. elegans* has four MIF-like genes: *Ce-mif-1* (Y56A3A.3), *Ce-mif-2* (C52E4.2), *Ce-mif-3* (F13G3.9), and *Ce-mif-4* (Y73B6BL.13). Transgenic reporter and immunolocalization studies suggest that *C. elegans* MIFs may have roles in development and the dauer stage [13,25]. *Bm*-MIF-1 has highest pairwise similarity to *Ce*-MIF-1 (41% compared to 23-29% for the other three paralogues; Figure 2) [20], and phylogenetic analysis of over seventy MIF-like proteins from eukaryotes confirms this assignment (D.B.G. and M.L.B., manuscript in preparation). Comparison of *Bm*-MIF-1 to the *C. elegans* MIFs, a second *B. malayi* MIF (*Bm*-MIF-2), and human MIF-1 (Figure 2) revealed that *Bm-mif-1* and *Ce-mif-1* shared

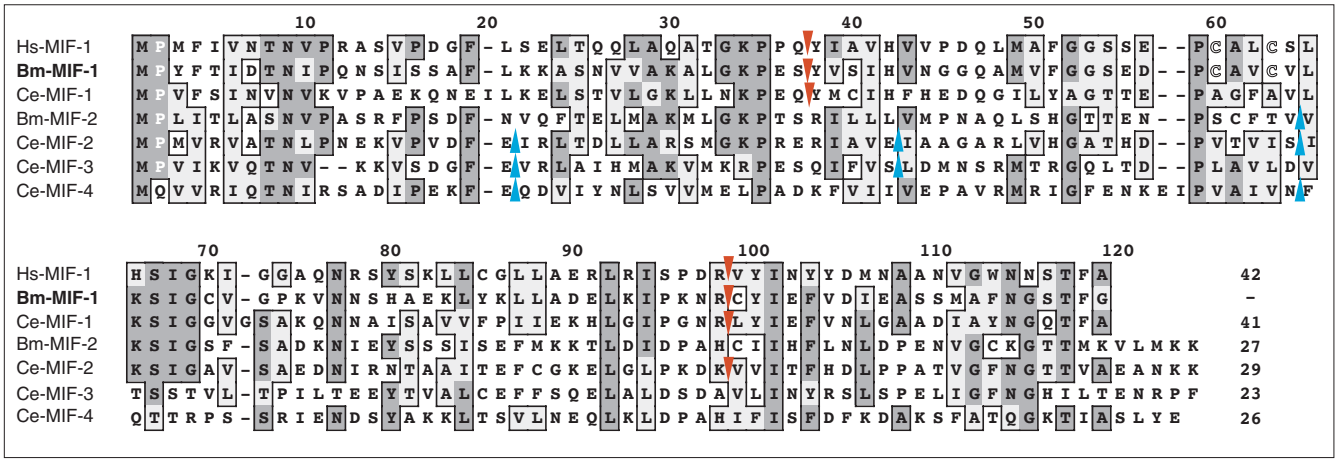

Figure 2
Comparison of *B. malayi* and *C. elegans* MIF proteins. *Bm*-MIF-1 (accession AAC82502) was aligned with human *Hs*-MIF-1(AAA21814), *C. elegans* MIF homologs *Ce*-MIF-1 (CAB60512), *Ce*-MIF-2 (CAB01412), *Ce*-MIF-3 (CAA95795), *Ce*-MIF-4 (AAG23475), and *Bm*-MIF-2b (AAF91074). Intron positions are marked by triangles (red, conserved with *Hs*-MIF-1; blue, *Ce*-MIF-2, -3 and -4 specific). The proline at position 2 (white) is important for immune function, and the CXXC motif at positions 60-63 is essential for thiol-oxidoreductase activity in mammalian MIF. The percent identity of each protein to *Bm*-MIF-1 is given at the end of the alignment.

two intron/exon boundaries also found in vertebrate MIFs. One of these introns was also present in *Ce-mif-3*, but *Ce-mif-3* and the other two *C. elegans mif* genes shared a set of introns not present in the *mif-1* genes. *Bm*-MIF-1 and other filarial MIF-1 homologs contain a CXXC motif (single-letter amino-acid code) critical for the thiol-oxidoreductase activities of vertebrate MIF [26]. None of the *C. elegans* MIF homologs contained this motif.

Conserved gene clusters

Two clusters of three genes in close proximity are conserved. The first involves BMBAC01L03.2, .3 and .5. The *C. elegans* orthologs of these genes are F43G9.5, F43G9.4, and F43G9.3 respectively. F43G9.5 and F43G9.3 are divergently transcribed from a 631 bp intergenic region. F43G9.3 is followed by F43G9.4 in the same transcriptional orientation with 501 bp separating the genes. In *B. malayi* this local synteny is conserved, except that two additional genes - BMBAC01L03.3 and .4 - are found between BMBAC01L03.2 and .5.

The second cluster also involves three genes. Proteins predicted from both alternative transcripts of BMBAC01P19.2 were found to be homologous to large proteins from *Homo sapiens* (BAF180, AAG34760 [27]), *Gallus gallus* (JC5056 [28]), *D. melanogaster* (CG11375, AAF56339), and *C. elegans* (C26C6.1) (Figure 3). These proteins shared six bromodomains (PF00439), two BAH domains (bromo-adjacent homology, PF01426), a HMG box (high mobility group, PF00505), and an anonymous carboxy-terminal domain (PFB007669). The *B. malayi*, *C. elegans*, and *D. melanogaster* polybromodomain (PBR) proteins also contain two C2H2 zinc fingers. PBR proteins may be involved in chromatin-remodeling complexes. Bromodomains interact with acetylated lysine in histone complexes, while HMG boxes are found in chromatin proteins that bind to single-stranded DNA and unwind double-stranded DNA. Human BAF180 has been shown to localize to the kinetochores of mitotic chromosomes [27]. None of the vertebrate PBR homologs contains zinc fingers, which may indicate additional functions for the nematode and fly proteins.

Two conserved genes were identified immediately upstream from *pbr-1* (Figure 3). BMBAC01P19.5 (named *Bm-ubr-1* (upstream of *pbr-1*)) showed significant similarity only to T28F4.4 from *C. elegans* (27% identity). The protein encoded by BMBAC01P19.4 is homologous to *C. elegans* T28F4.5 (30% identity). Iterative searches of GenBank using PSI-BLAST [29] indicated that BMBAC01P19.4 and T28F4.5 belong to a group of small peptides that include human DAP-1 (death-associated protein). DAP-1 is a nuclear protein and positive regulator of interferon gamma-induced apoptosis in HeLa cells [30]. PSORT profiling indicated that both nematode proteins may have a nuclear localization. BMBAC01P19.2 (*Bm-pbr-1*) and BMBAC01P19.5 (*Bm-ubr-1*) are divergently transcribed and BMABAC01P19.4 (*Bm-dap-1*) is found in the large third intron of BMBAC01P19.5 in the same transcriptional orientation as BMBAC01P19.2 (Figure 3). In the *C. elegans* instance of the PBR cluster, C26C6.1 (*Ce-pbr-1*) and T28F4.4 (*Ce-ubr-1*) are also divergently transcribed from a 1,233 bp intergenic region. The third gene, T28F4.5 (*Ce-dap-1*) is found in the large third intron of T28F4.4 on the same strand as C26C6.1.

Comparison of the intergenic and upstream regions of both clusters, and of the orthologous gene pairs, did not reveal any clear motifs that might be involved in transcriptional

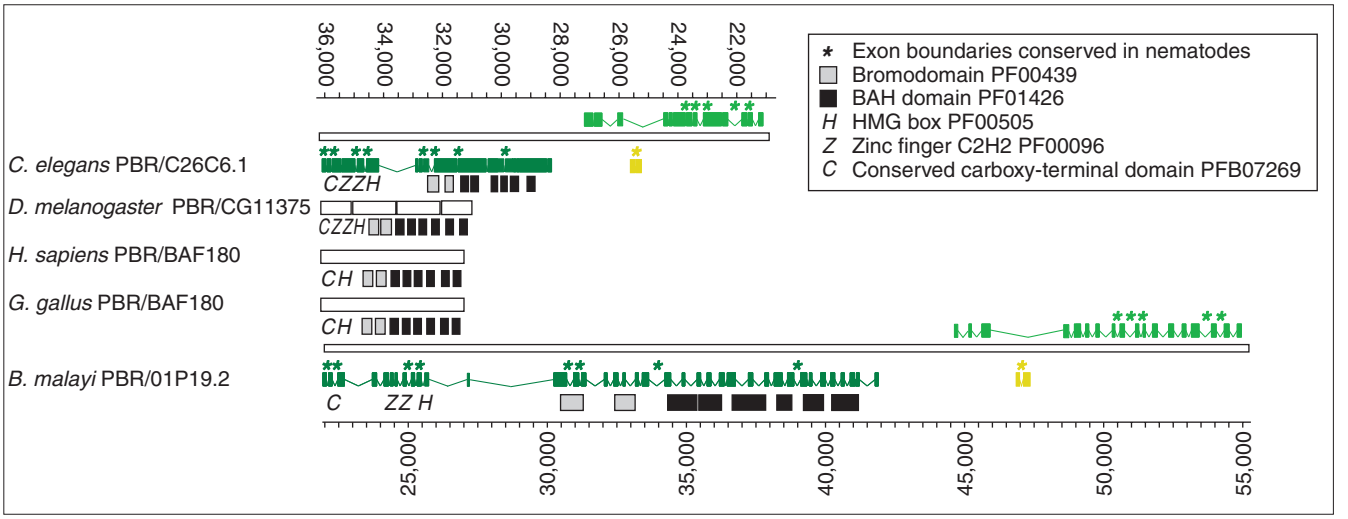

Figure 3
The *pbr* synteny cluster and *pbr* homologs in other species. The genomic organization of the *pbr* synteny cluster in *C. elegans* and *B. malayi*, and the domain structure of the PBR homologs in *Drosophila melanogaster*, *Gallus gallus*, and *Homo sapiens* are illustrated. Intron/exon boundaries that are conserved between the nematodes are indicated by asterisks. White boxes represent the contiguous DNA underlying the gene models.

regulation. In particular, the intergenic DNA between *pbr-1* and *ubr-1*, and the first intron of *ubr-1*, had less than 30% pairwise identity throughout, and there were no stretches of greater identity. The AT richness of the *B. malayi* genome compared to *C. elegans* may obscure any conserved elements. No RNA-coding genes were found. Two *B. malayi* ESTs matched at > 99.5% identity to two regions of BMBAC01P19 separated by 200 bp that were not predicted to be part of a transcript (see Figure 1). These regions are downstream of gene BMBAC01P19.3, and may derive from alternative 3′ untranslated regions: the furthest downstream match includes a good polyadenylation site. The 3′ end of the cDNA determined for this gene may have derived from internal priming from an A-rich segment of the 3′ untranslated region.

Fractured synteny between the genomes of *B. malayi* and *C. elegans*

All of the *C. elegans* orthologs, except for Y56A3A.3 (*Ce-mif-1*, 41% identity to *Bm-mif-1*, on chromosome III), are located on chromosome I (Figure 4). F13G3.9 (*Ce-mif-3*, 23% identity to *Bm-mif-1*) is found on *C. elegans* chromosome I in close proximity to the orthologs of *B. malayi* genes BMBAC01P19.2, .4, and .5. This could suggest that our orthology assignment is wrong. As described above, however, *Ce-mif-1* and *Bm-mif-1* share two intron positions and are more similar to each other than either is to *Ce-mif-3*, which has one concordant intron position, and one discordant intron position. The conflict between location and structure could be due to a gene-conversion event in either lineage, or an event of directed movement or insertion.

Eight of the 10 remaining *C. elegans* orthologs lay within a 2.3 Mb region in the center of chromosome I (6.7-9 Mb) (Figure 4). The orthologs of the other two genes (BMBAC0L03.4 and BMBAC01P19.6) are found at the distal tip of chromosome I. While there has been extensive rearrangement of gene order, when compared to the *C. elegans* orthologs, 10 of the *B. malayi* genes were in the same relative transcriptional orientation. Examination of the boundaries of the *C. elegans* cluster and individual gene regions did not show any association with repeat-sequence classes, including those shown to be commonly associated with rearrangements between *C. elegans* and *C. briggsae* [6].

Genome survey sequence comparison and synteny

To ascertain whether the segment sequenced was representative of the relationship between the *B. malayi* genome and that of *C. elegans*, we surveyed the *B. malayi* BAC-end derived genome survey sequences (GSSs; J. Daub, C. Whitton, N.H., M. Quail and M.L.B., unpublished observations). There are over 18,000 GSSs from *B. malayi*, derived from three independent libraries. Each BAC-end sequence was compared to the *C. elegans* proteome (Wormpep [31]) and significant similarities recorded (BLASTX probabilities < e-8). The chromosomal position of each matching *C. elegans* protein was derived from Wormbase [32]. One hundred and sixty-four BACs had matches at both ends to *C. elegans* proteins under these conditions (summarized in Table 2, details in Table 3). We note that these matches are not necessarily to orthologs, as we have not carried out intensive analysis of each one, but random selection of genes should not yield greater linkage estimation despite the problem of gene families and domain matches. While much of the *C. elegans* proteome consists of protein families, very few of these have a chromosomally restricted distribution [33,34].

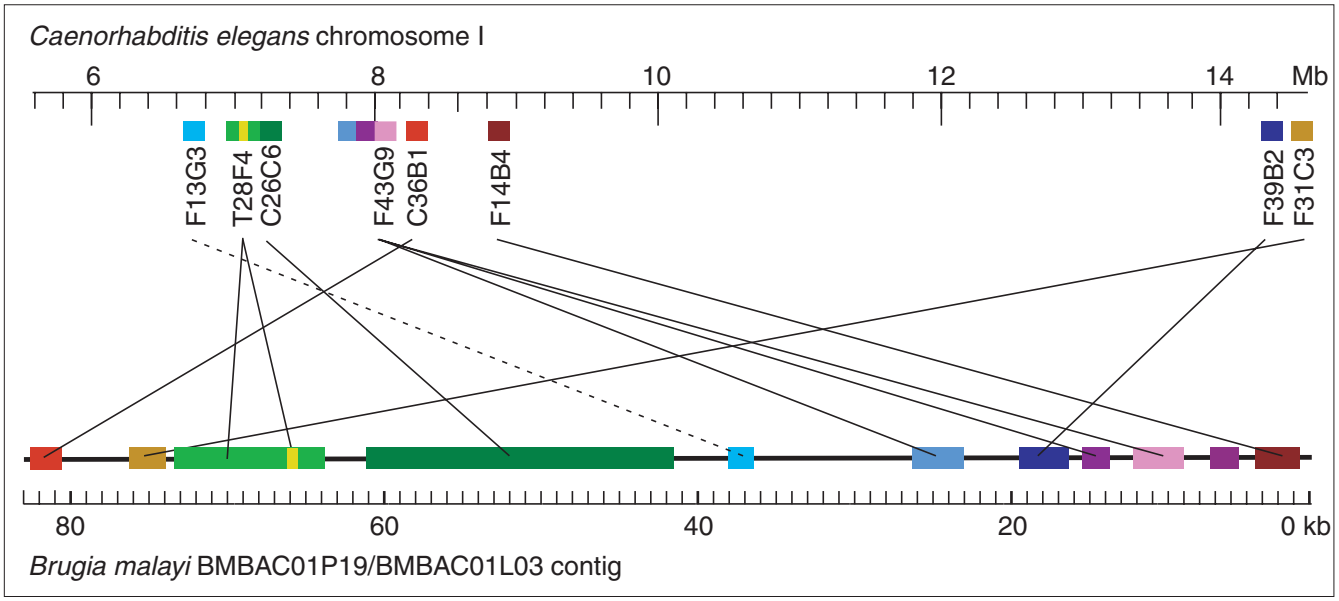

Figure 4
Comparison of linkage and synteny with *C. elegans*. The *B. malayi* contig is compared to an approximately 9 Mb segment of *C. elegans* chromosome I. The relative positions of the ortholog pairs, colored as in Figure 1, are indicated. The link between *Bm-mif-1* and *Ce-mif-3* (F13G3.9) is dashed to indicate that these two genes are paralogs rather than orthologs (see text for details).

C. elegans has six chromosomes. Under a minimal model, if a genome rearrangement were equally likely to involve a between-chromosome as a within-chromosome event, and was only dependent on the length of DNA in the within-chromosome versus not-within-chromosome classes, we would expect approximately five of every six rearrangements to involve between-chromosome events and one-sixth to involve within-chromosome events. This model ignores the fact that *B. malayi* has only five chromosome pairs: four autosomes and one XY pair. The derivation of the two karyotypes is unknown, and cannot be deduced from phylogenetic comparisons (see [35]). While most nematodes of clade V have six chromosomes like *C. elegans*, other taxa in the Secernentea have from one to >100 [36]. If we assume that the *C. elegans* complement derives from splitting of an ancestral chromosome retained in *B. malayi*, the expectation would be that 20% of rearrangements would be within-chromosome.

Many more BACs had significantly more ends mapping to the same chromosome than would be expected under these models (approximately 55%, χ^2 test $p < 0.01$ for all comparisons in Table 2 under the above model). The mean distance between the *C. elegans* matches was 4.4 Mb, which may be compared to an expected approximately 45 kb for the separation between the *B. malayi* BAC ends.

Discussion

B. malayi is a human parasite only distantly related to the model nematode *C. elegans* [14,37]; therefore, genome

Table 2

Synteny conservation between *B. malayi* BAC-end genome survey sequences and *C. elegans* genome sequence

Maximal probability of either of blast matches	Number of BACs with both ends matching *C. elegans* proteins	Number of BACs with both ends matching *C. elegans* proteins on the same chromosome	Distance between *C. elegans* proteins (megabases)	Percentage of matches on same chromosome
<e-8	164	90	4.4	54.88
<e-10	138	78	4.6	56.52
<e-15	51	29	4.7	56.86
<e-20	17	10	5.3	58.82

B. malayi BAC end sequences were compared to the *C. elegans* proteome using BLASTX. Matches with a probability <e-8 were noted, and chromosomal positions determined from WormBase. Of 2,200 BACs with matches, 164 had matches to both ends.

Table 3

B. malayi* BAC end comparisons to *C. elegans

	T7 end				SP6 end				
Brugia malayi BAC clone	*C. elegans* match	*C. elegans* chromosome	Position on chromosome	Exponent of probability in BLAST search	*C. elegans* match	*C. elegans* chromosome	Position on chromosome	Exponent of probability in BLAST search	Distance between matches
BMBAC01M03	CE27661	IV	7844080	18	CE03144	II	11592445	30	NA
BMBAC01I11	CE12826	II	2125627	24	CE27131	X	12434210	12	NA
BMBAC01O12	CE12384	IV	11536217	21	CE24899	X	10540216	13	NA
BMBAC01I15	CE07931	X	1138520	11	CE00450	III	7926986	9	NA
BMBAC01F17	CE06551	V	11711418	18	CE00946	III	4668338	18	NA
BMBAC01F18	CE06551	V	11711418	18	CE00946	III	4668338	18	NA
BMBAC03I06	CE04396	X	4702371	12	CE01008	III	3436926	17	NA
BMBAC03F12	CE22809	IV	10961452	9	CE28366	V	2688438	15	NA
BMBAC03O15	CE29604	I	10151104	24	CE08947	V	11984722	10	NA
BMBAC03F17	CE00316	III	9821062	22	CE14390	V	6500809	8	NA
BMBAC03J17	CE20445	IV	3562754	23	CE15856	II	13201660	9	NA
BMBAC04M12	CE14750	I	4619453	9	CE17599	V	14520036	11	NA
BMBAC04B14	CE26776	IV	2800306	36	CE03447	X	10583738	36	NA
BMBAC04B18	CE01099	III	9303554	45	CE16711	V	18449410	43	NA
BMBAC06B01	CE15044	V	4304442	23	CE26600	I	1494247	10	NA
BMBAC07G03	CE07756	II	3032776	9	CE13435	I	6135032	20	NA
BMBAC08D11	CE22116	II	14151234	16	CE17662	I	9188977	19	NA
BMBAC08E17	CE18356	I	3663582	17	CE24671	X	1800708	13	NA
BMBAC09F11	CE08682	I	4162592	13	CE08947	V	11984722	10	NA
BMBAC09K18	CE26381	IV	7210081	26	CE27040	III	1491791	12	NA
BMBAC09A22	CE24671	X	1800708	38	CE14734	II	1143941	33	NA
BMBAC10N08	CE14734	II	1143941	29	CE11078	X	14666566	18	NA
BMBAC11P11	CE18826	I	12580986	63	CE01074	III	4761237	39	NA
BMBAC301H09	CE00436	III	8966904	15	CE03397	II	10033351	10	NA
BMBAC303G12	CE25661	X	10088725	12	CE28910	IV	12096051	37	NA
BMBAC305D10	CE05811	IV	12222786	10	CE26022	I	13790068	13	NA
BMBAC306C12	CE19038	II	12001566	10	CE17716	V	5828441	25	NA
BMBAC307F09	CE26106	III	11214188	14	CE24397	I	398952	12	NA
BMBAC308B07	CE01495	III	4243241	9	CE23997	I	4301621	52	NA
BMBAC308E07	CE10254	V	8596497	16	CE22541	IV	1058851	39	NA
BMBAC309G05	CE19946	V	13652193	10	CE20405	I	10121170	21	NA
BMBAC310G03	CE00169	III	8560276	16	CE03487	IV	11101538	20	NA
BMBAC310F07	CE26106	III	11214188	20	CE17565	I	12897554	9	NA
BMBAC311D10	CE05492	IV	9045220	11	CE09323	I	8357446	11	NA
BMBAC312B12	CE03263	X	12785597	11	CE20461	II	11358344	28	NA
BMBAC314G02	CE04726	X	7500571	15	CE16564	III	10779508	14	NA
BMBAC314G05	CE04726	X	7500571	15	CE16564	III	10779508	14	NA
BMBAC314C06	CE14448	V	8303220	13	CE25695	III	7697801	23	NA
BMBAC321E09	CE00901	III	3777196	28	CE04196	IV	7171873	27	NA
BMBAC324A05	CE23883	X	10640426	10	CE24718	IV	9708837	15	NA
BMBAC325E11	CE24076	IV	16170439	27	CE20681	III	3942090	19	NA
BMBAC327E05	CE29377	II	14249402	21	CE05190	I	7147729	9	NA
BMBAC328H12	CE11268	I	6056251	27	CE20346	IV	359584	33	NA
BMBAC331C11	CE00639	III	10524644	12	CE07306	V	8110632	39	NA
BMBAC332H10	CE03812	X	11374102	41	CE03398	II	10030927	18	NA

Table 3 *(continued)*

	T7 end				SP6 end				
Brugia malayi BAC clone	*C. elegans* match	*C. elegans* chromosome	Position on chromosome	Exponent of probability in BLAST search	*C. elegans* match	*C. elegans* chromosome	Position on chromosome	Exponent of probability in BLAST search	Distance between matches
BMBAC335D03	CE00713	III	6820989	37	CE26022	I	13790068	10	NA
BMBAC335B06	CE03657	X	12880838	36	CE28110	II	12072195	15	NA
BMBAC335H06	CE04374	III	7093735	29	CE27906	II	7218339	12	NA
BMBAC335B11	CE28095	II	6474361	11	CE12664	IV	10627077	9	NA
BMBAC335G11	CE14211	II	526662	12	CE00644	III	4417152	9	NA
BMBAC338H04	CE21000	I	3609203	39	CE01643	II	8066397	57	NA
BMBAC340C01	CE08947	V	11984722	12	CE06100	I	7963691	10	NA
BMBAC340H10	CE24671	X	1800708	28	CE28961	II	8518425	11	NA
BMBAC341A06	CE24422	II	15153828	10	CE26560	IV	2637785	14	NA
BMBAC341H09	CE20297	I	10962692	19	CE07462	X	16821321	18	NA
BMBAC342D11	CE27040	III	1491791	14	CE11074	X	14645097	11	NA
BMBAC352C10	CE00713	III	6820989	44	CE26022	I	13790068	18	NA
BMBAC353A03	CE00949	III	4694946	10	CE06100	I	7963691	18	NA
BMBAC353E06	CE09682	IV	17269732	48	CE02716	II	4609014	10	NA
BMBAC354G08	CE24000	X	13899761	16	CE21401	I	12747957	34	NA
BMBAC354C09	CE16562	III	10808992	23	CE17579	IV	1178045	9	NA
BMBAC355C03	CE04838	IV	7225306	12	CE21023	I	2496034	24	NA
BMBAC356B08	CE06116	V	10355247	11	CE26971	I	311402	14	NA
BMBAC357C02	CE14754	I	4624187	24	CE19593	III	867498	12	NA
BMBAC360E07	CE06034	IV	11733052	15	CE02044	II	6736839	11	NA
BMBAC362E03	CE05492	IV	9045220	11	CE28001	III	6020770	16	NA
BMBAC365D07	CE15463	IV	12871709	16	CE01508	II	11384821	12	NA
BMBAC365F09	CE15612	V	10250527	10	CE05747	IV	12401915	20	NA
BMBAC365D11	CE15892	I	13091093	13	CE28340	III	13328281	9	NA
BMBAC368B08	CE21026	X	8125574	15	CE09880	I	8898846	16	NA
BMBAC374G02	CE24292	II	12681620	11	CE06704	IV	5987165	18	NA
BMBAC375H10	CE01537	II	9588260	15	CE04726	X	7500571	15	NA
BMBAC376D04	CE02705	II	5918674	9	CE29504	IV	4212960	17	NA
BMBAC377D05	CE03061	X	12966730	14	CE15044	V	4304442	10	NA
BMBAC01G04	CE12942	II	163142	12	CE15754	II	13443071	19	13279929
BMBAC01J11	CE17559	III	3729721	13	CE27691	III	6439903	14	2710182
BMBAC01N16	CE19942	II	6157856	20	CE01090	II	7858336	15	1700480
BMBAC01A23	CE27862	I	4952222	8	CE16340	I	13239686	8	8287464
BMBAC01M24	CE02307	II	10222779	15	CE04813	II	4902586	25	5320193
BMBAC02F03	CE01008	III	3436926	30	CE02018	III	5268852	17	1831926
BMBAC02M10	CE18369	IV	14965985	26	CE27782	IV	32953	13	14933032
BMBAC03D10	CE01563	II	10146750	11	CE18563	II	14006392	10	3859642
BMBAC03L15	CE27488	IV	2976034	13	CE20122	IV	12679870	30	9703836
BMBAC03O17	CE27311	III	1616853	13	CE00946	III	4668338	20	3051485
BMBAC03J24	CE21971	II	12727192	17	CE22157	II	13670692	12	943500
BMBAC04P08	CE17474	IV	8034836	13	CE06702	IV	5987165	37	2047671
BMBAC04J10	CE17474	IV	8034836	12	CE06702	IV	5987165	35	2047671
BMBAC04G15	CE03492	III	10465212	14	CE01161	III	5016428	29	5448784
BMBAC04L18	CE26381	IV	7210081	23	CE06302	IV	10375062	23	3164981
BMBAC06H01	CE16413	V	11222884	11	CE08630	V	4818967	13	6403917
BMBAC07C02	CE13736	I	5616992	9	CE18454	I	7384257	27	1767265
BMBAC07C06	CE28324	X	4830732	21	CE23711	X	14708595	24	9877863

Table 3 *(continued)*

	T7 end				SP6 end				
Brugia malayi BAC clone	*C. elegans* match	*C. elegans* chromosome	Position on chromosome	Exponent of probability in BLAST search	*C. elegans* match	*C. elegans* chromosome	Position on chromosome	Exponent of probability in BLAST search	Distance between matches
BMBAC07E21	CE16194	III	10818631	33	CE26632	III	12724299	46	1905668
BMBAC07C22	CE16565	III	10784128	12	CE17401	III	3778796	16	7005332
BMBAC08P03	CE27215	X	6626852	21	CE09403	X	4445872	16	2180980
BMBAC09E01	CE25011	III	7031238	12	CE24009	III	4722390	10	2308848
BMBAC09B17	CE25196	II	2913916	15	CE18730	II	11578670	11	8664754
BMBAC09E19	CE22045	III	11317051	12	CE20681	III	3942090	23	7374961
BMBAC09J20	CE27601	IV	3694675	13	CE17308	IV	3625656	9	69019
BMBAC09A24	CE04504	IV	8315582	37	CE29005	IV	6345808	12	1969774
BMBAC10M23	CE01473	II	8023190	13	CE28454	II	5867637	17	2155553
BMBAC11H08	CE11494	II	10872814	20	CE03412	II	11545610	17	672796
BMBAC11K08	CE28485	IV	16845895	21	CE06705	IV	5987165	45	10858730
BMBAC11C09	CE21401	I	12747957	12	CE08532	I	3704246	13	9043711
BMBAC11H20	CE08377	I	10117043	20	CE17566	I	12903800	16	2786757
BMBAC13A23	CE29235	II	6995861	10	CE23659	II	13251947	19	6256086
BMBAC301F09	CE28770	V	7400098	9	CE06116	V	10355247	13	2955149
BMBAC303H10	CE18123	X	10768551	12	CE04392	X	5627431	15	5141120
BMBAC303E12	CE01105	III	3992607	28	CE05066	III	6081444	39	2088837
BMBAC306B02	CE19437	IV	1935178	12	CE06634	IV	11985224	30	10050046
BMBAC306F02	CE21208	V	11417176	9	CE15044	V	4304442	46	7112734
BMBAC306B09	CE05594	IV	11574005	10	CE18268	IV	262009	12	11311996
BMBAC309A07	CE27186	II	1490131	20	CE20311	II	14794131	18	13304000
BMBAC309H07	CE15235	I	6586100	12	CE15751	I	8715988	13	2129888
BMBAC311C01	CE26713	X	10830672	11	CE05839	X	14719098	16	3888426
BMBAC312B02	CE23530	II	9886945	21	CE05732	II	9892692	9	5747
BMBAC318E08	CE03335	II	9006072	32	CE01731	II	10094778	33	1088706
BMBAC320B05	CE24687	I	13505250	32	CE10608	I	5535918	18	7969332
BMBAC321D05	CE03487	IV	11101538	12	CE06601	IV	12361570	12	1260032
BMBAC323H11	CE28173	II	7045967	12	CE03349	II	8811285	12	1765318
BMBAC326G05	CE18454	I	7384257	16	CE19979	I	14650506	17	7266249
BMBAC327F03	CE05372	I	8656418	20	CE17767	I	14934285	15	6277867
BMBAC327E08	CE22135	II	13280025	18	CE03397	II	10033351	14	3246674
BMBAC329E10	CE26424	III	7268168	10	CE26172	III	2584463	13	4683705
BMBAC333B09	CE06291	III	9853062	21	CE00018	III	9542362	18	310700
BMBAC335G07	CE23108	V	18907704	25	CE08145	V	7207535	24	11700169
BMBAC336F09	CE23823	V	904798	12	CE08939	V	10165831	10	9261033
BMBAC338B09	CE19930	IV	11494578	12	CE27358	IV	12787708	12	1293130
BMBAC339A05	CE03536	X	11156821	19	CE29169	X	15581083	10	4424262
BMBAC340B12	CE28433	V	12591291	11	CE06114	V	10352190	21	2239101
BMBAC341B01	CE06362	IV	11131027	28	CE17284	IV	507058	10	10623969
BMBAC344B10	CE27691	III	6439903	9	CE18868	III	13830997	16	7391094
BMBAC345G11	CE16052	III	13507164	17	CE01319	III	7408460	22	6098704
BMBAC346C07	CE20899	III	9066807	42	CE06204	III	10983239	9	1916432
BMBAC348D09	CE27859	X	4671617	13	CE03447	X	10583738	14	5912121
BMBAC349D02	CE28454	II	5867637	30	CE01473	II	8023190	15	2155553
BMBAC349A03	CE01694	II	9647841	22	CE01697	II	9649394	30	1553
BMBAC350E01	CE05839	X	14719098	31	CE28227	X	10830175	11	3888923
BMBAC350F06	CE07421	IV	7521267	17	CE17427	IV	606450	11	6914817

Table 3 *(continued)*

	T7 end				SP6 end				
Brugia malayi BAC clone	*C. elegans* match	*C. elegans* chromosome	Position on chromosome	Exponent of probability in BLAST search	*C. elegans* match	*C. elegans* chromosome	Position on chromosome	Exponent of probability in BLAST search	Distance between matches
BMBAC351D02	CE17559	III	3729721	37	CE29455	III	7295192	14	3565471
BMBAC351E11	CE04813	II	4902586	15	CE01843	II	6318128	9	1415542
BMBAC352A02	CE16057	X	9835707	16	CE23711	X	14708595	13	4872888
BMBAC352H11	CE27011	III	2386289	22	CE23035	III	12323434	40	9937145
BMBAC354D02	CE09506	V	13767614	15	CE26193	V	7064763	11	6702851
BMBAC354C06	CE28000	V	5680455	19	CE18785	V	14010680	10	8330225
BMBAC357F01	CE07705	I	5160615	15	CE17689	I	7197186	12	2036571
BMBAC357D06	CE05481	V	9909478	10	CE18731	V	12911699	11	3002221
BMBAC360G09	CE26686	V	19889249	20	CE12204	V	12228547	13	7660702
BMBAC361F02	CE21971	II	12727192	15	CE24422	II	15153828	14	2426636
BMBAC364D04	CE19878	IV	13020730	47	CE12664	IV	10627077	24	2393653
BMBAC364D12	CE05066	III	6081444	36	CE01648	III	10372177	11	4290733
BMBAC365G04	CE27551	I	1567610	19	CE09340	I	9956916	15	8389306
BMBAC367E09	CE29511	III	7551909	19	CE28049	III	10232451	11	2680542
BMBAC369F08	CE20121	IV	12674275	12	CE06362	IV	11131027	44	1543248
BMBAC370D08	CE09762	IV	3867451	19	CE17122	IV	7994133	17	4126682
BMBAC372C01	CE06239	I	8521548	15	CE16055	I	10289506	13	1767958
BMBAC372A05	CE23035	III	12323434	18	CE27402	III	5842056	15	6481378
BMBAC372F06	CE29472	III	5231760	15	CE00100	III	8521627	9	3289867
BMBAC372A09	CE00872	III	4156692	10	CE20934	III	3133584	18	1023108
BMBAC373F04	CE09880	I	8898846	12	CE06511	I	7477616	19	1421230
BMBAC374E02	CE00946	III	4668338	10	CE03076	III	3936413	20	731925
BMBAC374F12	CE21847	IV	1757609	21	CE06364	IV	11128632	11	9371023
BMBAC375A04	CE25585	IV	6754827	12	CE04562	IV	7326282	11	571455
BMBAC375F12	CE22210	V	14353297	17	CE21224	V	7077544	16	7275753

Clones with significant matches at both ends. NA, not applicable.

comparisons between these species will yield data concerning longer-term changes in structure and function that cannot be derived from within-genus comparisons. In the 83 kb of genomic DNA flanking the *B. malayi mif-1* locus we found a fractured conservation of microsynteny between the two nematode genomes, and conservation of linkage. Twelve protein-coding genes were predicted, and 11 of these had putative orthologs in the *C. elegans* genome. Ten of these orthologs were on *C. elegans* chromosome I, with eight in a 2.3 Mb segment in the center of the chromosome and two at the distal tip of chromosome I. Some of these genes have remained tightly linked in the same or slightly modified relative transcriptional orientations in both species.

This pattern, of conservation of linkage with disruption of precise synteny, was confirmed using BAC-end sequences. Of the 171 clones with matches at both ends to *C. elegans* genes, over 55% were localized to the same chromosome in *C. elegans*. While the mean distance separating the *B. malayi* genes is 45 kb (the length of the BAC clones; [38] and C. Whitton and M.L.B., unpublished work), the mean distance between the matching *C. elegans* genes is approximately 4.4 Mb.

The 83 kb fragment of *B. malayi* genomic DNA is the largest contiguated portion of sequenced genomic DNA from a non-rhabditid nematode described to date. A large proportion (around 60%) of genes identified in the *B. malayi* EST dataset (23,000 ESTs corresponding to around 8,300 unique transcripts [39]) have no close *C. elegans* homologue [16]. In this study, however, *C. elegans* orthologs were identified for 11 of the 12 identified *B. malayi* genes. Some of these orthologous pairs were confirmed by congruence in length of open reading frame and shared intron positions, despite low pairwise identity. Global searches with ESTs would not have detected these pairs (BLAST probability values of approximately e-4), and thus the true proportion of *B. malayi* unique genes is likely to be less than 60%. *B. malayi* genes were found to have larger and more numerous introns than *C. elegans* genes (2.2 times longer and 1.7

times more frequent), in keeping with previous estimates made using data from several highly expressed genes [40]. If the contig is representative and gene complement is equivalent to *C. elegans,* the *B. malayi* genome may be larger (120-140 Mb) than estimated previously (100 Mb [41]). Four of seven genes confirmed by reverse transcriptase PCR had alternative transcripts, a figure consistent with *C. elegans* EST and cDNA projects [42]. Additionally, five genes had *B. malayi* EST matches, a proportion congruent with the estimate that the EST program has identified around 40% of the expected 20,000 *B. malayi* genes [16].

Conserved linkage between the genomes of closely related eukaryotic organisms has been shown in several taxa. But it is only recently, with the sequencing of discrete segments or whole genomes, that examples of conservation of microsynteny between the genomes of distantly related species (not involving functionally related genes) have been described [43,44]. The microsyntenic gene clusters retained between *C. elegans* and *B. malayi* do not fall into any clear functional categories. However, all genes contained in the second cluster (BMBAC01P19.2, .4, and .5) are predicted to have nuclear localization signals and could be co-regulated. Alternatively, promoters or *cis*-acting regulatory elements required for their proper function could be embedded within other cluster members. Interdigitation of these regulatory elements could be constraining the movement of genes away from this cluster. No conserved motifs were found, however, and this possibility can thus only be tested by transgenesis experiments. This phenomenon has been observed in other systems such as fungal genomes, where gene pairs predicted to have overlapping regulatory elements are more likely to be conserved between species [45].

Many genes in *C. elegans* are co-transcribed in operons [46,47] and this could constrain synteny breakage. The *C. elegans* orthologs of BMBAC01L03.5 and BMBAC01P19.3 are separated by 501 bp, an intergenic distance found in other *C. elegans* operons, and the downstream gene (*Ce*-F43G9.4) was shown to be *trans*-spliced to the SL2 spliced leader, a feature of downstream genes in *C. elegans* operons [47]. However, in *B. malayi*, BMBAC01L03.5 and BMBAC01P19.3 are separated by 2.8 kb, which is outside the range of operon intergenic spacing. The functions of *C. elegans* genes on chromosome I have been investigated by RNA-mediated interference and a phenotype was identified for one gene in each cluster: embryonic lethality (F39G4.5 [48]) and altered adult morphology (C26C6.1 [49]). Therefore, it is possible that the clusters are conserved because removing other members would interfere with functions of these essential genes. The one exception to the conservation of linkage is the *Bm-mif-1*/*Ce-mif-1* ortholog pair. Another *C. elegans* MIF homolog, *Ce-mif-3*, is found in close proximity to the genes in the *pbr-1* synteny cluster, raising the possibility that a gene-conversion event may have obscured orthology assignment for this gene.

In the Metazoa, long-range synteny between the genomes of distantly related species (>300 Myr divergence) has only been identified previously in vertebrates (teleost fish and humans [50,51]). In vertebrates, interchromosomal exchanges seem to be rare events, and some linkage groups, such as human chromosomes 6 and X, are conserved across most eutherian mammals [7]. From the analyses presented here we can suggest some general patterns of gene rearrangement in nematodes. Most of the *C. elegans* orthologs were located in a small segment of chromosome I (nine of eleven genes in 2.3 Mb or 16% of the chromosome), suggesting that local intrachromosomal inversions or rearrangements have occurred more frequently than long-range intrachromosomal, or interchromosomal rearrangements. This is consistent with patterns observed in closely related dipterans, where the composition of linkage groups is conserved but not the order within the chromosome. Mechanistically this may occur because intrachromosomal rearrangements require fewer DNA breaks than interchromosomal translocations, and the nuclear scaffold may hold local chromosomal regions in closer association. The high rate of rearrangement of genes within the nematode chromosomes makes it unlikely that the positional information of genes in the *Caenorhabditis* genomes will be useful in finding orthologous genes in the genomes of distantly related nematodes such as *B. malayi*.

Materials and methods

Identification of candidate genomic clones for sequencing

A probe for *Bm-mif-1* was synthesized by labeling full-length cDNA (GenBank accession U88035) with biotin (Phototope; New England Biolabs), hybridized to high-density arrays of 18,000 BAC clones containing *B. malayi* genomic DNA [52], and detected with the Phototope detection kit (New England Biolabs). Hybridization-positive BACs were PCR verified using gene-specific primers Bm-MIF-1.F1a (ATGCCATATTTTAC-GATTGATAC) and Bm-MIF-1.R1a (GAACACCATCGCTTGTC-CACC) using standard reaction and cycling conditions (0.2 mM dNTPs, 1.5 mM MgCl, 0.5 pM primer; 1 cycle of 94°C for 3 min; 35 cycles of 94°C for 15 sec, 55°C for 20 sec, 72°C for 3 min; 1 cycle of 72°C for 10 min). BMBAC01P19 was selected for sequencing. Sequence from the T7 end of the insert was used to design specific primers 01P19.T7.F1 (GCAGCAAAT-GCTTATTTGTCTTG) and 01P19.T7.R1 (GTTTGGTGATTCAT-GTCCATGAGC). Primers 01P19.T7.R1 and 2BiotinBACF3 (designed to the BAC vector; $(biotinU)_2$GAGTCGACCT-GCAGGCATGC; New England BioLabs Organic Synthesis Unit) were used to synthesize a biotin-labeled end probe. The probe was hybridized to the BAC library filter using a modified hybridization and detection protocol [38]. Positive BACs were PCR verified with primers 01P19.T7.R1 and 01P19.T7.F1, and insert DNA prepared using a kit (Qiagen). BAC ends were end-sequenced using the Sanger Institute protocol [53]. BMBAC01L03 showed minimal overlap with

BMBAC01P19 compared to other clones and was selected for sequencing.

Preparation, subcloning, and sequencing of BACs

The BACs were 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 analyzed on automated sequencers. Each BAC was sequenced to a depth of sevenfold coverage. Contigs were assembled using phrap (Phil Green, Washington University Genome Sequencing Center, unpublished). Manual base calling and finishing was carried out using Gap4 [54]. Gaps and low-quality regions were resolved by techniques such as primer walking, PCR and resequencing clones under conditions that give increased read lengths.

Sequence analysis

The finished sequences of BMBAC01P19 and BMBAC01L03 were compared to the GenBank nonredundant (nucleic acid and protein) EST database (dbEST), the *C. elegans* genome and protein and the custom *B. malayi* clustered EST [16] databases using BLAST [55,56]. GeneFinder (P. Green and L. Hillier, Washington University Genome Sequencing Center, unpublished) was trained with 162 publicly available *B. malayi* gene sequences and used to analyze the contiguated sequence. The sequence was annotated on the Artemis workbench [57]. Predicted protein sequences were compared to Pfam [58] and cellular localization examined using PSORTII [59]. The annotated sequence is available in GenBank (accession AL606837).

Verification of gene predictions

To confirm gene predictions from BMBAC01P19, primers were designed and PCR was carried out on oligo(dT)-primed *B. malayi* mixed adult first-strand cDNA with gene-specific primers. To isolate cDNA ends, the GeneRacer 3′ RACE primer (Invitrogen) (GCTGTCAACGATACGCTACGTAACG-GCATGACAGTG), or the nematode SL1 sequence (GGTT-TAATTACCCAAGTTTGAG) were used with specific primers. Secondary PCRs were carried out using nested primers and 2% of the primary PCR product. Positive PCR products were cloned and sequenced.

BAC-end sequence analysis

The *B. malayi* BAC-end sequence dataset was compared to the *C. elegans* proteome in Wormpep. Significant matches were filtered, and BAC clones having matches on both ends retained. The chromosomal position of the *C. elegans* genes was determined from [32].

Acknowledgements

We thank the Filarial Genome Project for the *B. malayi* BAC library and clones, Yvonne Harcus, Janice Murray, William Gregory, and Rick Maizels for *B. malayi* materials, Jen Daub and Claire Whitton for BAC-end analysis, Dan Lawson for help with *C. elegans* genome queries, and New England BioLabs for reagents. Funding for this work was provided by the Medical Research Council. We acknowledge the support and hard work of sequencing team 14 at the Sanger Institute.

References

1. Hentschel CC, Birnstiel ML: **The organization and expression of histone gene families.** *Cell* 1981, **25:**301-313.
2. Ferrier DE, Holland PW: **Ancient origin of the Hox gene cluster.** *Nat Rev Genet* 2001, **2:**33-38.
3. Litman GW, Rast JP, Shamblott MJ, Haire RN, Hulst M, Roess W, Litman RT, Hinds-Frey KR, Zilch A, Amemiya CT: **Phylogenetic diversification of immunoglobulin genes and the antibody repertoire.** *Mol Biol Evol* 1993, **10:**60-72.
4. Ohta Y, Okamura K, McKinney EC, Bartl S, Hashimoto K, Flajnik MF**: Primitive synteny of vertebrate major histocompatibility complex class I and class II genes.** *Proc Natl Acad Sci USA* 2000, **97:**4712-4717.
5. Ranz JM, Casals F, Ruiz A: **How malleable is the eukaryotic genome? Extreme rate of chromosomal rearrangement in the genus *Drosophila*.** *Genome Res* 2001, **11:**230-239.
6. Coghlan,A, Wolfe KH: **Fourfold faster rate of genome rearrangement in nematodes than in *Drosophila*.** *Genome Res* 2002, **12:**857-867.
7. O'Brien SJ, Menotti-Raymond M, Murphy WJ, Nash WG, Wienberg J, Stanyon R, Copeland NG, Jenkins NA, Womack JE, Marshall Graves JA: **The promise of comparative genomics in mammals.** *Science* 1999, **286:**458-462, 479-481.
8. Fulton RE, Salasek ML, DuTeau NM, Black WCT: **SSCP analysis of cDNA markers provides a dense linkage map of the *Aedes aegypti* genome.** *Genetics* 2001, **158:**715-726.
9. Kent WJ, Zahler AM: **Conservation, regulation, synteny, and introns in a large-scale *C. briggsae*-*C. elegans* genomic alignment.** *Genome Res* 2000, **10:**1115-1125.
10. Caceres M, Ranz JM, Barbadilla A, Long M, Ruiz A: **Generation of a widespread *Drosophila* inversion by a transposable element.** *Science* 1999, **285:**415-418.
11. Caceres M, Puig M, Ruiz A: **Molecular characterization of two natural hotspots in the *Drosophila buzzatii* genome induced by transposon insertions.** *Genome Res* 2001, **11:**1353-1364.
12. Thomas WK, Wilson AC: **Mode and tempo of molecular evolution in the nematode caenorhabditis: cytochrome oxidase II and calmodulin sequences.** *Genetics* 1991, **128:**269-279.
13. The *C. elegans* Sequencing Consortium: **Genome sequence of the nematode *C. elegans*: a platform for investigating biology.** *Science* 1998, **282:**2012-2018.
14. Vanfleteren JR, Van de Peer Y, Blaxter ML, Tweedie SA, Trotman C, Lu L, Van Hauwaert ML, Moens L: **Molecular genealogy of some nematode taxa as based on cytochrome c and globin amino acid sequences.** *Mol Phylogenet Evol* 1994, **3:**92-101.
15. McReynolds LA, DeSimone SM, Williams SA: **Cloning and comparison of repeated DNA sequences from the human filarial parasite *Brugia malayi* and the animal parasite *Brugia pahangi*.** *Proc Natl Acad Sci USA* 1986, **83:**797-801.
16. Blaxter M, Daub J, Guiliano DB, Parkinson J, Whitton C, Project FG: **The *Brugia malayi* genome project: expressed sequence tags and gene discovery.** *Trans Roy Soc Trop Med Hyg* 2001, in press.
17. Williams SA, Lizotte-Waniewski MR, Foster J, Guiliano D, Daub J, Scott AL, Slatko B, Blaxter ML: **The filarial genome project: analysis of the nuclear, mitochondrial and endosymbiont genomes of *Brugia malayi*.** *Int J Parasitol* 2000, **30:**411-419.
18. Meyer BJ: **Sex determination and X chromosome dosage compensation.** In *C.elegans II*. Edited by Riddle DL, Blumenthal T, Meyer BJ, Priess JR. Plainview, New York: Cold Spring Harbor Laboratory Press; 1997: 209-240.
19. Sakaguchi Y, Tada I, Ash LR, Aoki Y: **Karyotypes of *Brugia pahangi* and *Brugia malayi* (Nematoda: Filaroidea).** *J Parasitol* 1983, **69:**1090-1093.
20. Pastrana DV, Raghavan N, FitzGerald P, Eisinger SW, Metz C, Bucala R, Schleimer RP, Bickel C, Scott AL: **Filarial nematode parasites secrete a homologue of the human cytokine macrophage migration inhibitory factor.** *Infect Immun* 1998, **66:**5955-5963.
21. Stec I, Wright TJ, van Ommen GJ, de Boer PA, van Haeringen A, Moorman AF, Altherr MR, den Dunnen JT: **WHSC1, a 90 kb SET**

domain-containing gene, expressed in early development and homologous to a *Drosophila* dysmorphy gene maps in the Wolf-Hirschhorn syndrome critical region and is fused to IgH in t(4;14) multiple myeloma. *Hum Mol Genet* 1998, **7:**1071-1082.

22. Stec I, Nagl SB, van Ommen GJ, den Dunnen JT: **The PWWP domain: a potential protein-protein interaction domain in nuclear proteins influencing differentiation?** *FEBS Lett* 2000, **473:**1-5.
23. Nishihira J: **Macrophage migration inhibitory factor (MIF): its essential role in the immune system and cell growth.** *J Interferon Cytokine Res* 2000, **20:**751-762.
24. Falcone FH, Loke P, Zang X, MacDonald AS, Maizels RM, Allen JE: **A *Brugia malayi* homolog of macrophage migration inhibitory factor reveals an important link between macrophages and eosinophil recruitment during nematode infection.** *J Immunol* 2001, **167:**5348-5354.
25. Marson AL, Tarr DEK, Scott AL: **Macrophage migration inhibitory (*mif*) transcription is significantly elevated in *Caenorhabditis elegans* dauer larvae.** *Gene* 2001, **278:**53-62.
26. Kleemann R, Rorsman H, Rosengren E, Mischke R, Mai NT, Bernhagen J: **Dissection of the enzymatic and immunologic functions of macrophage migration inhibitory factor. Full immunologic activity of N-terminally truncated mutants.** *Eur J Biochem* 2000, **267:**7183-7193.
27. Xue Y, Canman JC, Lee CS, Nie Z, Yang D, Moreno GT, Young MK, Salmon ED, Wang W: **The human SWI/SNF-B chromatin-remodeling complex is related to yeast *rsc* and localizes at kinetochores of mitotic chromosomes.** *Proc Natl Acad Sci USA* 2000, **97:**13015-13020.
28. Nicolas RH, Goodwin GH: **Molecular cloning of polybromo, a nuclear protein containing multiple domains including five bromodomains, a truncated HMG-box, and two repeats of a novel domain.** *Gene* 1996, **175:**233-240.
29. Altschul SF, Koonin EV: **Iterated profile searches with PSI-BLAST-a tool for discovery in protein databases.** *Trends Biochem Sci* 1998, **23:**444-447.
30. Deiss LP, Feinstein E, Berissi H, Cohen O, Kimchi A: **Identification of a novel serine/threonine kinase and a novel 15-kD protein as potential mediators of the gamma interferon-induced cell death.** *Genes Dev* 1995, **9:**15-30.
31. **Wormpep** [http://www.sanger.ac.uk/Projects/C_elegans/wormpep]
32. **WormBase** [http://www.wormbase.org/]
33. Friedman R, Hughes AL**: Gene duplication and the structure of eukaryotic genomes.** *Genome Res* 2001, **11:**373-381.
34. Lespinet O, Wolf YI, Koonin EV, Aravind L: **The role of lineage-specific gene family expansion in the evolution of eukaryotes.** *Genome Res* 2002, **12:**1048-1059.
35. Blaxter ML: **Genes and genomes of *Necator americanus* and related hookworms.** *Int J Parasitol* 2000, **30:**347-355.
36. Walton AC: **Some parasites and their chromosomes.** *J Parasitol* 1959, **45:**1-20.
37. Blaxter ML, De Ley P, Garey JR, Liu LX, Scheldeman P, Vierstraete A, Vanfleteren JR, Mackey LY, Dorris M, Frisse LM, *et al.*: **A molecular evolutionary framework for the phylum *Nematoda*.** *Nature* 1998, **392:**71-75.
38. Foster JM, Kamal IH, Daub J, Swan MC, Ingram JR, Ganatra M, Ware J, Guiliano D, Aboobaker A, Moran L, *et al.*: **Hybridization to high-density filter arrays of a *Brugia malayi* BAC library with biotinylated oligonucleotides and PCR products.** *Biotechniques* 2001, **30:**1216-1218, 1220, 1222 passim.
39. Parkinson J, Whitton C, Guiliano D, Daub J, Blaxter ML: **200,000 nematode ESTs on the net.** *Trends Parasitol* 2001, **17:**394-396.
40. Zang X, Yazdanbakhsh M, Jiang H, Kanost MR, Maizels RM: **A novel serpin expressed by blood-borne microfilariae of the parasitic nematode *Brugia malayi* inhibits human neutrophil serine proteinases.** *Blood* 1999, **94:**1418-1428.
41. Maina CV, Grandea III AG, Tuyen LTK, Asikin N, Williams SA, McReynolds LA: ***Dirofilaria immitis*: Genomic complexity and characterisation of a structural gene.** In *Molecular Paradigms for Eradicating Helminthic Parasites*. Edited by MacInnis AJ. Alan Liss: New York; 1987: 193-204.
42. Reboul J, Vaglio P, Tzellas N, Thierry-Mieg N, Moore T, Jackson C, Shin IT, Kohara Y, Thierry-Mieg D, Thierry-Mieg J, *et al.*: **Open-reading-frame sequence tags (OSTs) support the existence of at least 17,300 genes in *C. elegans*.** *Nat Genet* 2001, **27:**332-336.
43. Brunner B, Todt T, Lenzner S, Stout K, Schulz U, Ropers HH, Kalscheuer VM: **Genomic structure and comparative analysis of nine *Fugu* genes: conservation of synteny with human chromosome Xp22.2-p22.1.** *Genome Res* 1999, **9:**437-448.
44. Hamer L, Pan H, Adachi K, Orbach MJ, Page A, Ramamurthy L, Woessner JP: **Regions of microsynteny in *Magnaporthe grisea* and *Neurospora crassa*.** *Fungal Genet Biol* 2001, **33:**137-143.
45. Huynen MA, Snel B, Bork, P: **Inversions and the dynamics of eukaryotic gene order.** *Trends Genet* 2001, **17:**304-306.
46. Blumenthal T, Steward K: **RNA processing and gene structure.** In C.elegans *II*. Edited by Riddle DL, Blumenthal T, Meyer BJ, Priess JR. Plainview, New York: Cold Spring Harbor Laboratory Press; 1997: 117-146.
47. 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.
48. Maeda I, Kohara Y, Yamamoto M, Sugimoto A: **Large-scale analysis of gene function in *Caenorhabditis elegans* by high-throughput RNAi.** *Curr Biol* 2001, **11:**171-176.
49. Fraser AG, Kamath RS, Zipperlen P, Martinez-Campos M, Sohrmann M, Ahringer J: **Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference.** *Nature* 2000, **408:**325-330.
50. Grant D, Cregan P, Shoemaker RC: **Genome organization in dicots: genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*.** *Proc Natl Acad Sci USA* 2000, **97:**4168-4173.
51. Ku HM, Vision T, Liu J, Tanksley SD: **Comparing sequenced segments of the tomato and *Arabidopsis* genomes: large-scale duplication followed by selective gene loss creates a network of synteny.** *Proc Natl Acad Sci USA* 2000, **97:**9121-9126.
52. Guiliano D, Ganatra M, Ware J, Parrot J, Daub J, Moran L, Brennecke H, Foster JM, Supali T, Blaxter M, *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:**146-152.
53. **End sequencing protocol** [http://www.sanger.ac.uk/Teams/Team51/PACBACPrep.shtml]
54. **Organization of the Gap4 manual** [http://www.mrc-lmb.cam.ac.uk/pubseq/manual/gap4_unix_1.html]
55. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403-410.
56. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25:**3389-3402.
57. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B: **Artemis: sequence visualization and annotation.** *Bioinformatics* 2000, **16:**944-945.
58. Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL: **The Pfam protein families database.** *Nucleic Acids Res* 2000, **28:**263-266.
59. Nakai K, Kanehisa M: **A knowledge base for predicting protein localization sites in eukaryotic cells.** *Genomics* 1992, **14:**897-911.
60. **NemBase** [http://www.nematodes.org/]
61. Lizotte-Waniewski M, Tawe W, Guiliano DB, Lu W, Liu J, Williams SA, Lustigman S: **Identification of potential vaccine and drug target candidates by expressed sequence tag analysis and immunoscreening of *Onchocerca volvulus* larval cDNA libraries.** *Infect Immun* 2000, **68:**3491-3501.