**Bioinformation** by Biomedical Informatics Publishing Group *open access*

www.bioinformation.net **Hypothesis**

# Identification and comparative analysis of novel alternatively spliced transcripts of RhoGEF domain encoding gene in *C. elegans* and *C. briggsae*

**Luv Kashyap[1], Mohammad Tabish[1]*, Subramaniam Ganesh[2], Deepti Dubey[2]**

[1]Department of Biochemistry, Faculty of Life Sciences, Aligarh Muslim University, Aligarh, India; [2]Department of biological Sciences and Bioengineering, Indian Institute of Technology, Kanpur, India; *Mohammad Tabish - Email: tabish.biochem@gmail.com; Phone: 091 571 2700741; * Corresponding author

received July 27, 2007; revised August 16, 2007; accepted August 23, 2007; published online September 11, 2007

**Abstract:**
Y95B8A.12 gene of *C. elegans* encodes RhoGEF domain, which is a novel module in the Guanine nucleotide exchange factors (GEFs). Alternative splicing increases transcriptome and proteome diversification. Y95B8A.12 gene has two reported alternatively spliced transcripts by the *C. elegans* genome sequencing consortium. In the work presented here, we report the presence of four new spliced transcripts of Y95B8A.12 arising as a result of alternative splicing in the pre-mRNA encoded by Y95B8A.12 gene. Our methodology involved the use of various gene or exon finding programmes and several other bioinformatics tools followed by experimental validation. We have also studied alternative splicing pattern in RhoGEF domain encoding orthologues gene from *C. briggsae* and have obtained very similar results. These new unreported spliced transcripts, which were not detected through conventional approaches, not only point towards the extent of alternative splicing in *C. elegans* genes but also emphasize towards the need of analyzing genome data using a combinations of bioinformatics tools to delineate all possible gene products.

**Key words**: RhoGEF domain; computational analysis; exon prediction; alternative splicing; *C. elegans; C. briggsae*

**Background:**
The guanine nucleotide exchange factors for the Rho GTPases (RhoGEFs) are a large family of proteins that share a dual structural motif designated the DH/PH domain. RhoGEFs are regulators of the Rho proteins **[1]** that act as molecular switches; cycling between inactive (GDP-bound) and active (GTP-bound) states. **[2]** The interaction of Rho with residues within the DH domain enhances the exchange of GDP for GTP and thus converts Rho into the biologically active form. Thus, the intracellular ratio of the GTP/GDP-bound forms of Rho proteins determines the activation of signal transduction pathways regulating the spatial and temporal reorganization of cytoskeletal architecture. **[3]** Members of the Rho subfamily of Ras-like monomeric GTPases, including Rho, Rac and Cdc42, are involved in a broad range of functions including gene transcription, cell cycle progression, cell polarity and most notably regulation of the actin cytoskeleton and cell morphology. **[4]**

RhoGEF family is a widespread family found commonly in almost all organisms like Humans, Mouse, and Drosophila, *C. elegans* etc. In *C. elegans* this family consists of a large number of members and here the number of RhoGEF and RhoGAP regulators of Rho GTPases significantly exceeds the number of Rho family GTPases. These regulators likely provide the signaling specificity and spatial-temporal regulation required by the broadly expressed and functionally important Rho family GTPases. Alternative splicing has recently emerged as a major mechanism of generating protein diversity in higher eukaryotes including nematodes *Caenorhabditis elegans.* **[5, 6, 7]** Several cases of alternative RNA splicing have been found in various RhoGEF domain containing and Rho related proteins in various organisms **[8, 9]** and even in other RhoGEF domain containing genes of *C. elegans.* **[10]** Currently, most efficient methods use expressed sequence tags or microarray analysis for efficient detection of alternative splicing. **[11-14]** However, it is difficult to detect all alternative splice events with them because of their inherent limitations as discussed. **[15]**

Recently, we successfully demonstrated identification of novel transcripts using a bioinformatics methodology involving both computational and experimental analysis in *C. elegans.* **[15]** Genefinder prediction by the *C. elegans* sequencing consortium of genomic sequence of Y95B8A.12 has reported two spliced transcripts Y95B8A.12a and Y95B8A.12b that arise as a result of alternative splicing alternative splicing of pre-mRNA in the intronic region between exon 2 and exon 3 of Y95B8A.12. Detailed analysis of Y95B8A.12 gene using a wide array of bioinformatics tools and programmes like gene/exon/ORF finding tools, BLAST analysis, sequence alignment programmes and many more; we predicted the existence of atleast four new

alternatively spliced transcripts Y95B8A.12c, Y95B8A.12d, Y95B8A.12e and Y95B8A.12f. These were subsequently confirmed by the presence of different transcripts by RT-PCR using gene specific primers and RNA isolated from mixed population of *C. elegans*.

**Methodology:**

**Nematode**

The wild-type *C. elegans* strain, N2 (var. Bristol), and the attenuated *E. coli* strain was used as a food source in all experiments as described essentially in. **[16]** All nematodes were cultured at 20°C on standard NGM agar (0.3 percent NaCl, 0.25 percent peptone, 5 mg per ml cholesterol, 1 mM $CaCl_2$, 1 mM $MgSO_4$, 1.7 percent agar) seeded with live *Escherichia coli* (OP50).

**Chemicals**

RevertAid™ M-MuLV Reverse Transcriptase and oligo $(dT)_{18}$ primer were purchased from Fermentas, Hanover, (USA), *Taq* DNA Polymerase and PCR-Buffer were purchased from Bangalore Genie Pvt. Ltd., India. dNTP Mix (2.5 mM each) and 1kb DNA ladder was purchased from MBI Fermentas, USA. All other chemicals used in the experiments were of molecular biology grade.

**Primers used**

The following oligonucleotides primers were custom synthesized from MWG Biotech, Pvt. Ltd., India.
(1) F1: 5' GCAGTTGTTCACCCCGTTGATTAG 3'
(2) F2: 5' ACATACGAGGAAATCAATTGATTAAGCC 3'
(3) F3: 5' CGGCAAGACTTCAGGATCGATGGAG 3'
(4) R1: 5' TCCAGCTCATCATCCTCAATCTC 3'
(5) R1: 5' AATAATCTCCCTCCGTTTTGCTCTGCCC 3'

**Total RNA Isolation**

Total RNA was isolated from mixed-stage nematodes using the method described earlier. **[17]** Finally, total RNA was dissolved in diethyl pyrocarbonate-treated distilled water. The yield of RNA was determined spectrophotometrically, and the quality of the RNA was checked by gel electrophoresis. The resulting RNA was used for RT- PCR.

**Reverse transcriptase (RT)-PCR**

Computationally predicted transcripts encoding alternatively spliced exons were validated using RT-PCR. *C. elegans* total RNA (2 micro gram) was primed with oligo $(dT)_{18}$ and single-stranded cDNA was synthesized using the Revert Aid™ Reverse Transcriptase at 42 degree C for 60 min. (total vol. 20 micro liter). The target sequences were PCR amplified in 25 micro liter reaction volume. The program consisted of an initial denaturation step at 95 degree C for 5 min, followed by 30 cycles at 94 degree C for 1 min, 58 degree C for 1 min, and 72 degree C for 1 min. The program ended with a final elongation step at 72 degree C for 10 min. RT-PCR product (8 micro liter) obtained after 30 cycles were electrophoretically separated on a 2 percent agarose gel, stained with ethidium-bromide and photographed on a UV transilluminator.

**Semi nested PCR**

For further confirmation of results obtained after first round of PCR (for the presence of predicted spliced transcripts), semi nested PCR was performed. Here, after the first PCR amplification (as detailed above) the resulting RT-PCR product (2 micro liter) was used as a template for further amplification by PCR using the same forward but a new reverse primer (placed just internal to the reverse primer used in first PCR) specific for the same exon (as used in first PCR). The resulting semi nested PCR product (10 micro liter) was then subjected to electrophoresis on a 1 percent (weight by volume) agarose gel, stained with ethidium bromide and photographed on UV transilluminator.

**Tools used for Computational and Bioinformatic analysis**

Sequences of the RhoGEF domain encoding gene Y95B8A.12 was downloaded form the NCBI nucleotide database at http://www.ncbi.nlm.nih.gov/entrez. Various other gene/exon finding tools and other bioinformatics programmes used were same as described in **[15, 18]** and as below:

The **Genescan** program **[19]**, predicts complete, partial and multiple genes on both DNA strands. It can be used to identify introns, exons, promoter sites, polyA signals, etc. It is available at http://genes.mit.edu/GENSCAN.html

The **FEX** (Find Exon) **[20, 21]**, program initially predicts internal exons by linear discriminant function, evaluating open reading frames flanked by GT and AG base pairs (the 5' and 3' ends of typical introns). It is available at http://www.softberry.com/berry.phtml

**Twinscan [22]** makes predictions by combing the information from predicted coding regions and splice sites with conserservation measurements between the target sequence and sequences from a closely related genome. It is available at http://genes.cse.wustl.edu/

**FGENESH [23]** is also based on the Hidden Markov Model (HMM). It is available at http://sun1.softberry.com under section products.

**Splice Predictor** implements Bayesian models for splice site prediction **[24]** is available at http://bioinformatics.iastate.edu/cgi-bin/sp.cgi

The **NetGene2** server **[25]** is a service producing neural network predictions of splice sites in human, *C. elegans* and *A. thaliana* genomic sequences. The method is based on a hybrid of a Markov model and neural networks where parameters of the Markov model are learned by neural networks. It is located at http://www.cbs.dtu.dk:80/services/NetGene2

**Results and Discussion:**

**Computational prediction of new alternative splice transcripts of RhoGEF domain containing *C. elegans* gene**

Our analysis comprised use of a combination of various bioinformatics tools like gene finding programmes, exon predicting tools, ORF finders, blast searches and several other tools as detailed in our previous study. **[15]** Thus bioinformatics analysis of the approximately 19kb of untranslated region between the predicted initiation codon of the gene Y95B8A.2 (lying immediately upstream of Y95B8A.12a) and the initiation codon of Y95B8A.12a, predicted the possibility of 2 new potential spliced exons. These new exons were predicted to encode 15 to 21 amino acids and had methionine as the initiation codon and each potentially able to splice into exon 2 as shown (Figure 1 and Tables 1 and 2 (under supplementary material)). As there were two already reported spliced transcripts and computational analysis had predicted two new N-terminal exons, so by combining the results we were able to predict four new alternatively spliced transcripts of the gene Y95B8A.12 namely Y95B8A.12c, Y95B8A.12d, Y95B8A.12e and Y95B8A.12f. All these new spliced transcripts arise due to alternative splicing of Y95B8A.12 pre-mRNA in 5' untranslated region. In Y95B8A.12c a new N-terminal formed by 65 bp exon, splices with the second exon of the control transcript Y95B8A.12a and lies 7719 bp upstream of the first exon of the control Y95B8A.12a. Similarly, in Y95B8A.12d a new N-terminal exon of 47 bp, splices with the second exon of the control transcript and lies 7422 bp upstream from the first exon of the control Y95B8A.12a. (Figure 1 and Table 1 (under supplementary material)). So now we have a total of 6 spliced transcripts of *C. elegans* gene Y95B8A.12 (Y95B8A.12c, Y95B8A.12d, Y95B8A.12e and Y95B8A.12f) which splice together in the pattern as shown (Figure 1 and Tables 1 and 2 (in supplementary material)). Since all these transcripts arise because of alternative splicing of non-coding exons, it may add further complexity to Y95B8A.12 gene regulation and working mechanism.

Following the computational predictions of new spliced transcripts, we searched the Yuji Kohara's *C. elegans* EST database to look for putative EST or cDNA support for possible occurrence of these new exons or transcripts. A search of Yuji Kohara's *C. elegans* EST database for EST match didn't yield any fruitful matches, which is expected keeping in mind the problems and limitations of the EST database. Moreover, the likelihood of observing ESTs for alternative splice forms in a gene correlates with increasing number of ESTs for that gene; it is highest for highly expressed genes and virtually nil for low-abundance genes. This may be the most possible reason to explain why Yuji Khoara's *C. elegans* expressed sequence tag (EST) database search for putative EST or cDNA support for possible occurrence of these new exons/transcripts failed. NCBI BLAST search was done to look for homology of these new spliced transcripts but no significant information could be obtained about the prospective similarity with other polypeptides. Consensuses splice signals that are not normally used as splice sites (known as "cryptic" splice sites) occur frequently in a given pre-mRNA. Generally, intronic sequences at splice junctions are highly conserved (99.24 percent of introns have a "GT-AG" at their 5' and 3' ends, respectively) in almost all eukaryotic species. **[26, 27]** The presence of consensus sequences at the splice-donor /acceptor site is believed to be a structural feature that determines whether a specific splice site is used although these consensus sequences may not be always present. **[27]** Analysis of the exon-intron junction region of the newly predicted N-terminal exons of Y95B8A.12c, Y95B8A.12d, Y95B8A.12e and Y95B8A.12f indicates structural conservation, as depicted (Table 1 under supplementary material). The presence of consensus sequences in the splice-donor /acceptor site provides the further confirmatory evidence for the existence of new N-terminal exons of Y95B8A.12c, Y95B8A.12d, Y95B8A.12e and Y95B8A.12f. After the failure in search for supporting EST or cDNA matches for our newly predicted transcripts, the next way to confirm our findings was to validate them using RT-PCR.

**Experimental validation of computationally identified new spliced transcripts**

RT-PCR amplification was used to validate the possible existence of new transcripts of Y95B8A.12 gene as predicted on the basis of computational analysis. Here, our aim was only to prove that what predictions we got from our computational analysis were genuine enough. So, we selected only 1 of the 2 (Y95B8A.12a) *C. elegans* Sequencing Group reported alternatively spliced transcripts of Y95B8A.12 gene for practical validation. Moreover, we were also unable to verify some of our predicted transcripts mainly Y95B8A.12e, Y95B8A.12f (arising form combination with Sanger's prediction i.e. 2nd control Y95B8A.12b) as there was only around 20-30 bp difference in the transcripts, which was difficult to resolve on gel. For e.g. In Y95B8A.12a, the third exon (24 bp), is spliced out to give rise to another spliced transcript named as Y95B8A.12b (Figure 1). Briefly, for experimental confirmation of these spliced transcripts, total RNA, prepared from mixed-stage *C. elegans*, was reverse transcribed and the resulting single-stranded cDNA was PCR amplified (as detailed in materials and methods). Forward and Reverse primers used for the identification of splice transcripts were designed using a combination of bioinformatics tools and manual methods, so that they bind only to specific exon sequences. As a result, we would expect two reaction products, which are different in length if a splice event takes place. These exon specific primers were able to successfully validate the occurrence of the spliced transcripts by giving a band of anticipated size (Figure 2) when a particular primer pair specific to that particular exon was PCR amplified and products visualized on agarose gel, stained with ethidium bromide and photographed on a UV transilluminator. For transcripts of low abundance, for which the RT-PCR products could not be visualized after the procedure described above (as in case of predicted spliced transcripts Y95B8A.12d Figure 2, lane 5) and for further confirmation of results obtained in the first round of PCR

reaction, semi nested PCR was performed, in exactly same method as adopted above. So we were able to validate computational predictions for occurrence of new spliced transcripts of RhoGEF domain encoding Y95B8A.12 gene in *C. elegans*.

The goal of our work was to use a novel bioinformatics approach capable of complementing the conventional methods of identifying spliced transcripts by providing efficient delineation of all possible putative gene transcripts. Our findings emphasize towards the urgent need to analyze genome data using a combination of bioinformatics tools, programmes in order to delineate all possible gene products and to estimate the true extent of alternative splicing in *C. elegans* genes. It may also encourage other researchers to take up similar exhaustive studies in several other finished genomes especially of humans with whom *C. elegans* share a close gene homology. Moreover, further studies in this direction could be taken up which would enhance our knowledge about the biological and functional significance of these spliced transcripts and their possible role in RhoGEF gene working and regulation.

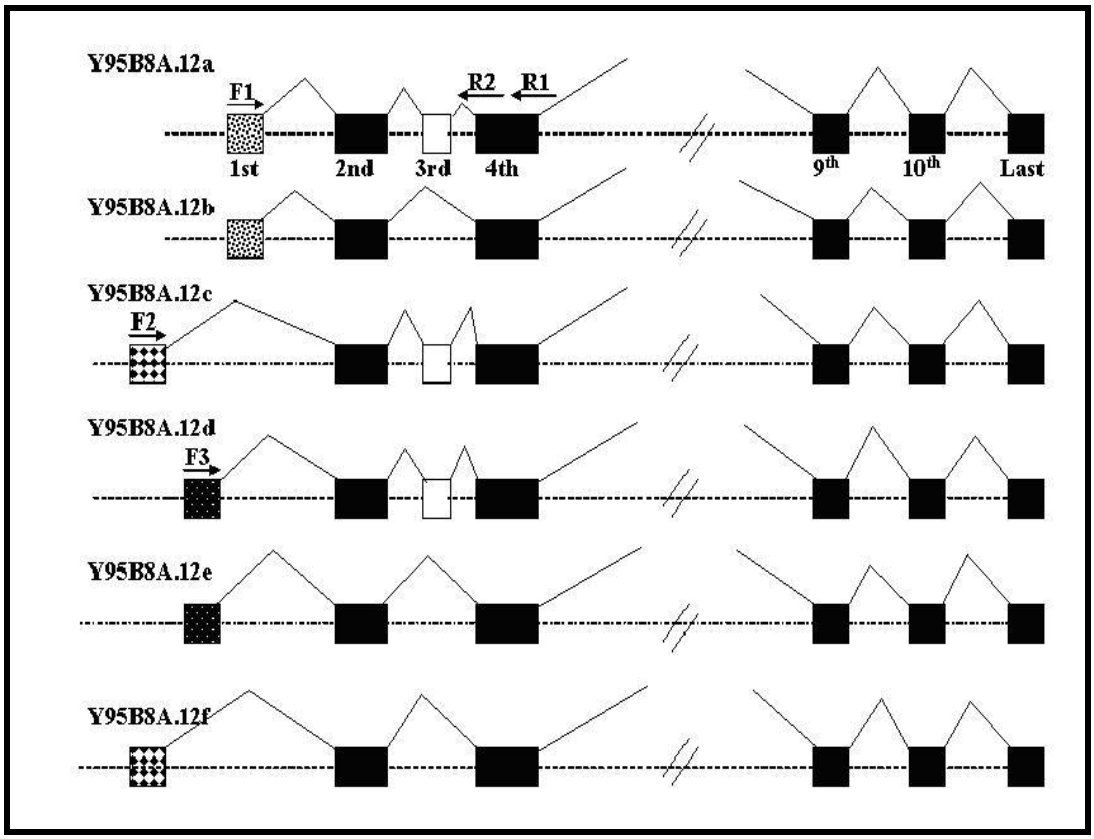


**Figure 1:** Showing the structure and organization of Y95B8A.12 gene with its predicted and already existing spliced transcripts: Organization of *C. elegans* RhoGEF domain containing gene Y95B8A.12 along with the predicted spliced transcripts: The exon, intron organization of the Y95B8A.12 gene, along with its existing spliced transcripts Y95B8A.12a, Y95B8A.12b and the newly predicted alternatively spliced transcripts Y95B8A.12c, Y95B8A.12d, Y95B8A.12e and Y95B8A.12f. Rectangular and square boxes indicate exons; dotted lines indicate the intronic and the untranslated regions, while solid joining lines show the splicing pattern of each spliced transcript. New exon of each Y95B8A.12a and Y95B8A.12b (control) and Y95B8A.12c, Y95B8A.12d, Y95B8A.12e and Y95B8A.12f (predicted alternative exons) is shown by a different pattern in the box. Arrows (F1, F2, F3, R1 and R2) indicate the Primers designed specific for each computationally predicted exon

**Bioinformation** by Biomedical Informatics Publishing Group *open access*

www.bioinformation.net **Hypothesis**

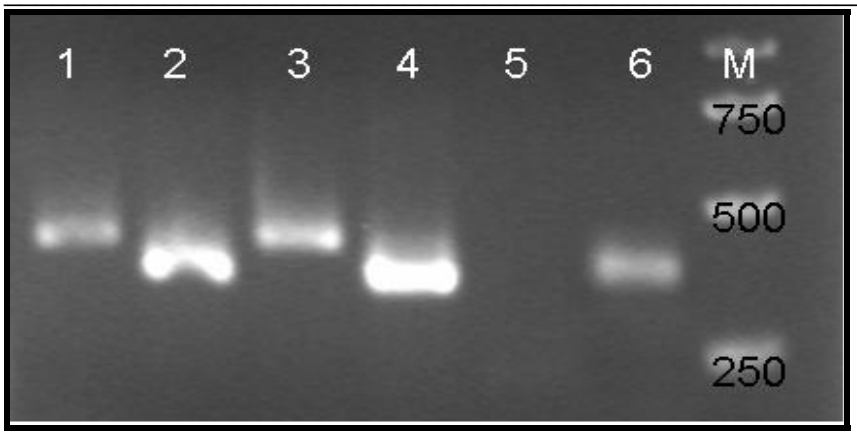


**Figure 2: RT-PCR analysis of predicted spliced transcript of the Y95B8A.12 gene:** RT-PCR analysis of predicted spliced transcript of the RhoGEF domain containing gene Y95B8A.12: RT-PCR amplification was used to determine the presence of transcripts, containing predicted exon, in total RNA prepared from mixed-stage *C. elegans* as described in the Materials and methods section. The migration of a series of size markers (M) is indicated on the right. RT-PCR products were obtained using a common reverse primer R1 (from exon 4) and exon specific forward primers representing each spliced transcripts. Lanes 1, 3, 5 represent the product size (bp) of 468, 463 and 460 obtained using common reverse primer (R1) in combination with forward primers F1, F2 and F3 respectively. While lanes 2, 4, 6 represent the product size (bp) 406, 401 and 404 obtained after semi nested PCR using common reverse primer(R2 designed internal to exon 4) in combination with forward primers F1, F2 and F3 respectively. Lane 5 is blank because no band was obtained in first PCR, while the corresponding transcript was validate by semi nested PCR (Lane 6)

**Comparative analysis of alternative splicing pattern of RhoGEF domain encoding genes in *C. elegans* and *C. briggsae***

With the availability of *C. briggsae* whole-genome shotgun assembly (cb25.agp8), with an estimated coverage of 98 percent of the whole genome it offers interesting research avenues to researchers engaged in parallel studies. As alternative splicing is a frequent and important aspect of gene regulation. It is of interest to compare the level and pattern of conservation of alternative splicing. These reasons prompted us to study alternative splicing pattern in *C. elegans* RhoGEF domain encoding orthologues gene in *C. briggsae.*

*C. elegans* and *C. briggsae* share a similar number of protein-coding genes (just under 20,000), with an estimate of 12,200 orthologous genes. **[28]** *C. elegans-C. briggsae* orthologs are annotated in Worm base (http://www.wormbase.org/) according to the criteria of Stein *et al.,* **[28]** *C. elegans* orthologue of RhoGEF domain gene was identified from inspection of Wormbase (http://www.wormbase.org/) and using Blastp function **[29]** at NCBI BLAST. Table 3 (supplementary material) shows the comparison between alternative splicing patterns obtained in *C. elegans-C. briggsae* orthologues genes encoding RhoGEF domain. From Table 3 (supplementary material), it is clear that both the genes not only had almost same size of unusually large 5' untranslated region (UTR) (approximately 19 to 20 kb) but also a similar pattern of splicing. Secondly, the results obtained on computational analysis of this unusually large UTR region to look for possible alternatively spliced transcripts also produced similar results. As shown in this paper in case of computational analysis of about 19 kb of *C. elegans* RhoGEF domain encoding gene Y95B8A.12, we successfully verified the existence of two new spliced transcripts arising due to alternative splicing of Y95B8A.12 pre-mRNA in 5' untranslated region (Y95B8A.12c & Y95B8A.12d). Similarly on computational analysis of about 20kb of untranslated region of CBG05122 (orthologue gene of RhoGEF domain gene in *C. elegans*), we predicted the existence of three novel spliced transcripts arising due to alternative splicing of pre-mRNA in 5' untranslated region of CBG05122 having new N-terminal exons (N'-1a, N'-1b, N'-1c). The splicing pattern obtained in the two organisms is remarkably similar, with almost same N-terminal exon size (Table 3 in supplementary material). Moreover, one of the predicted new N-terminal exons (N'-1a) has a marked similarity in the two organisms (1c & N'-1a) Figure 3(a) whereas this kind of amino acid similarity was not observed in the case of the other predicted spliced exons (1d & N'-1b) Figure 3(b). The possible reasons for which are still under investigation. Taking the above data and facts into consideration, we can easily say that N-terminal transcripts in both organisms arise due to alternative splicing of pre-mRNA in 5' untranslated region. The almost similar pattern of alternative splicing event indicates towards the evolutionarily conserved nature of alternative splicing in the two closely related genomes and its possible role in evolution of the two species. It may be that patterns of alternative splicing are conserved at similar levels to genes and gene structures. Thus, it can easily be hypothesized that many, and probably most, alternative splicing events are conserved between *C. elegans* and *C. briggsae* genomes.

**Bioinformation** by Biomedical Informatics Publishing Group *open access*

www.bioinformation.net **Hypothesis**

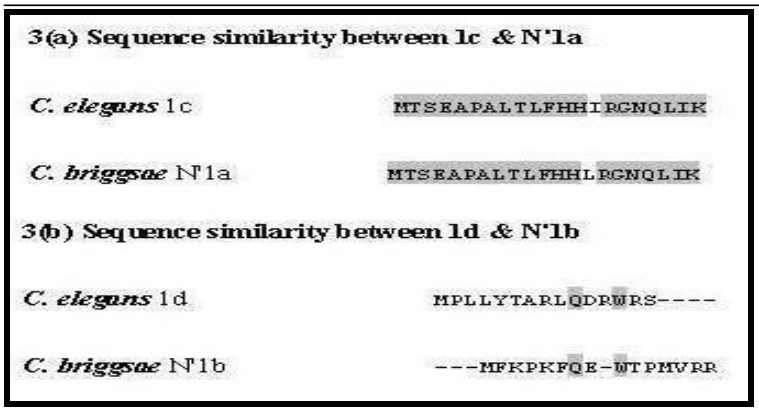


**Figure 3:** Comparative sequence alignment analysis between the polypeptide sequences encoded by *C. elegans* and *C. briggsae* orthologue genes (Y95B8A.12 and CBG05122): Sequence alignment picture of the polypeptide sequences encoded by alternative new N-terminal exons in *C. elegans* and *C. briggsae* orthologue genes (Y95B8A.12 and CBG05122) obtained using CLUSTAL W multiple sequence alignment. A grey highlight shows sequence similarity whereas no highlight means dissimilar residues

**Acknowledgement:**
The authors are thankful to the Council for Scientific and Industrial Research, University Grants Commission and Department of Science and Technology, New Delhi, India for providing special grants to the Department in the form of DRS and FIST for developing infrastructure facilities.

**References:**

**[01]** M. Schwartz, *J. Cell Sci.*, 117:5457 (2004) [PMID: 15509861]
**[02]** Y. Takai , *et al., Trends Biochem. Sci.*, 20:227 (1995) [PMID: 7543224]
**[03]** S. N. Prokopenko, *et al., Genes Dev.*, 13:2301 (1999) [PMID: 10485851]
**[04]** K. Burridge and K. Wennerberg, *Cell*, 116:167 (2004) [PMID: 14744429]
**[05]** A. A. Mironov, *et al., Genome Res.*, 9:1288 (1999) [PMID: 10613851]
**[06]** L. Croft, *et al., Nat. Genet.*, 24:340 (2000) [PMID: 10742092]
**[07]** D. Brett, *et al., Nat. Genet.*, 30:29 (2002) [PMID: 11743582]
**[08]** C. Pucharcos, *et al., Biochim Biophys Acta.*, 1521:1 (2001) [PMID: 11690630]
**[09]** L. Tsyba, *et al., Genomics*, 84:106 (2004) [PMID: 15203208]
**[10]** R. Steven, *et al., Genes Dev.*, 19:2016 (2005) [PMID: 16140983]
**[11]** Z. Kan, *et al., Genome Res.*, 11:889 (2001) [PMID: 11337482 ]
**[12]** B. Modrek, *et al., Nucleic Acids Res.*, 29:2850 (2001) [PMID: 11433032]
**[13]** J. .M. Johnson, *et al., Science*, 302:2141 (2003) [PMID: 14684825]
**[14]** M. R. Brent and R. Guigo, *Curr. Opin. Struct. Biol.*, 14:264 (2004) [PMID: 15193305]
**[15]** L. Kashyap, *et al., Bioinformation*, 2:17 (2007)
**[16]** S. Brenner, *Genetics*, 77:71(1974) [PMID: 4366476]
**[17]** M. Tabish, *et al., Biochem. J.*, 339:209 (1999) [PMID: 10085246]
**[18]** L. Kashyap and M. Tabish, *Bioinformation*, 1:203 (2006) [PMID: 17597889]
**[19]** C. Burge and S.Karlin, *J. Mol. Biol.*, 268:78 (1997) [PMID: 9149143]
**[20]** V. V. Solovyev, *et al., Nucleic Acids Res.*, 22:5156 (1994) [PMID: 7816600]
**[21]** C. M. Johnston, *et al., The Journal of Immunology*, 176:4221 (2006) [PMID: 16547259]
**[22]** I. Korf, *et al., Bioinformatics*, 17:S140 (2001) [PMID: 11473003]
**[23]** A. Salamov and V. Solovyev, *Genome Res.*, 10:516 (2000) [PMID: 10779491]
**[24]** V. Brendel, *et al., Bioinformatics*, 20:1157 (2004) [PMID: 14764557]
**[25]** S. M. Hebsgaard, *et al., Nucleic Acids Research*, 24:17 (1996)[PMID: 8811101]
**[26]** B. L. Robberson, *et al., Molecular Cell Biology*, 10:3439 (1990) [PMID: 2136768]
**[27]** S. M. Berget, *J. Biol. Chem.*, 270:2411 (1995) [PMID: 7852296]
**[28]** L. D. Stein, *et al., PLOS Biol.*, 1:E45 (2003) [PMID: 14624247]
**[29]** S. F. Altschul, *et al., Nucleic Acids Res.*, 25:3389 (1997) [PMID: 9254694]

**Edited by I. Roterman**
**Citation: Kashyap *et al.***, Bioinformation 2(2): 43-49 (2007)

**Bioinformation** by Biomedical Informatics Publishing Group *open access*

www.bioinformation.net **Hypothesis**

## Supplementary material

| Cosmid (Gene) | 5'-Exon-Intron boundary | 3'-Intron-Exon boundary | Exons organization in transcripts |
|---|---|---|---|
| Y95B8A.12a | CAGAG**gt**aaa | tcc**ag**CCAGA | 1a-2-3-4--Last (CN-1) |
| Y95B8A.12b | CAGAG**gt**aaa | tcc**ag**CCAGA | 1a-2-4----Last (CN-2) |
| Y95B8A.12c | AGCCG**gt**aag | tcc**ag**CCAGA | 1c-2-3-4---Last (PR-1) |
| Y95B8A.12d | CTGCG**gt**aag | tcc**ag**CCAGA | 1d-2-3-4---Last (PR-2) |
| Y95B8A.12e | CTGCG**gt**aag | tcc**ag**CCAGA | 1d-2-4------Last PR-3) |
| Y95B8A.12f | AGCCG**gt**aag | tcc**ag**CCAGA | 1c-2-4------Last (PR-4) |

**Table 1:** Details of Exon-Intron boundary of Y95B8A.12 gene with its predicted and already existing spliced transcripts: Y95B8A.12a and Y95B8A.12a are already existing spliced transcripts as per the *C. elegans* genome sequencing consortium marked as control (CN) while Y95B8A.12c, Y95B8A.12d, Y95B8A.12e and Y95B8A.12f represents the newly predicted spliced transcript marked as (PR). Consensus sequences at the splice-donor /acceptor site (exon-intron-exon boundaries) are shown in bold. Nucleotides indicated in lower case are part of introns and in upper case are parts of exons at the exon-intron-exon boundaries

| Exon | Amino acid sequence | N-terminal exon size | Exons organization in transcripts |
|---|---|---|---|
| 1a | MVDLQLFTPLIRYSEPEEGEPQTIHP | 15 | 1a-2-3-4--Last (CN-1) |
| 1b | MVDLQLFTPLIRYSEPEEGEPQTIHP | 15 | 1a-2-4----Last (CN-2) |
| 1c | MTSEAPALTLFHHIRGNQLIKPEEGEPQTIHP | 21 | 1c-2-3-4--Last (PR-1) |
| 1d | MPLLYTARLQDRWRSPEEGEPQTIHP | 15 | 1d-2-3-4--Last (PR-2) |
| 1e | MPLLYTARLQDRWRSPEEGEPQTIHP | 15 | 1d-2-4----Last(PR-3) |
| 1f | MTSEAPALTLFHHIRGNQLIKPEEGEPQTIHP | 21 | 1c-2-4----Last(PR-4) |

**Table 2:** Amino acid sequences encoded by the predicted alternatively spliced N-terminal exons: Deduced amino acid sequences encoded by alternative first exons (shown in regular font) and the first part of the amino acid sequence encoded by exon 2 (underlined) depending on the splicing pattern; number of amino acids encoded by the N-terminal exon is shown under N-terminal exon size; 1a, 1b, 1c, 1d, 1e and 1f are the first exons of Y95B8A.12a and Y95B8A.12b (control); Y95B8A.12c, Y95B8A.12d, Y95B8A.12e and Y95B8A.12f (predictions) respectively; splicing pattern shows the manner in which the splicing occurs in both new predicted spliced transcripts (PR) and controls (CN) predicted by the *C. elegans* genome sequencing consortium

| ***C. elegans* gene (5'UTR (untranslated) region approximately 19kb)** | | | |
|---|---|---|---|
| **Exon** | **Predicted Exon Size(amino acids)** | **Exons organization in transcripts** | **Amino acid Sequence encoded** |
| 1c | 21 | 1c-2-3-4----Last | MTSEAPALTLFHHIRGNQLIK |
| 1d | 15 | 1d-2-3-4----Last | MPLLYTARLQDRWRS |
| ***C. briggsae* gene (5'UTR (untranslated) region approximately 20kb)** | | | |
| **Exon** | **Predicted Exon Size(amino acids)** | **Exons organization in transcripts** | **Amino acid Sequence encoded** |
| N'-1a | 21 | N'1a-2-3-4----Last | MTSEAPALTLFHHLRGNQLIK |
| N'-b | 15 | N'1b-2-3-4----Last | MFKPKFQEWTPMVRR |
| N'-1c | 30 | N'1c--- 3-4----Last | MGNFRRQSAIIGNRRKPSAIIDQLSVTVHD |

**Table 3:** A comparative view of alternative splicing pattern obtained in *C. elegans* and *C. briggsae* orthologue genes: Alternative splicing pattern obtained after Gene/Exon finder analysis on genomic DNA sequences of RhoGEF domain encoding gene from *C. elegans* (Y95B8A.12) and the corresponding orthologue gene sequence from *C. briggsae* (CBG05122)