

Alternatively spliced isoforms encoded by cadherin genes from *C. elegans* genome

Luv Kashyap and Mohammad Tabish*

Department of Biochemistry, Faculty of Life Sciences, Aligarh Muslim University, Aligarh, India;

*Mohammad Tabish - Email: tabish.biochem@gmail.com;

Phone: 091 571 2700741; * Corresponding author

received July 27, 2007; revised September 05, 2007; accepted September 11, 2007; published online September 24, 2007

Abstract:

Cadherins are calcium-dependent, homophilic, cell-cell adhesion receptors that regulate morphogenesis, pattern formation and cell migration. The *C. elegans* Genome Sequencing Consortium has reported 12 genes from *C. elegans* genome encoding members of the cadherin superfamily. Alternative splicing of eukaryotic pre-mRNAs is a mechanism for generating potentially many transcript isoforms from a single gene. Here, using a combination of various gene or exon finding programmes and several other bioinformatics tools followed by experimental validation using RT-PCR, we have studied alternative splicing pattern in the cadherin encoding genes from *C. elegans* genome. We have predicted that 7 of the 12 genes encoding the cadherin superfamily undergo extensive alternative splicing and encode for 12 new unreported alternatively spliced transcripts. Most of the alternatively spliced exons were found to be present at the 5' end of genes. These new previously un-detected spliced variants in *C. elegans* cadherin superfamily of genes could play vital roles in explaining the way cadherins act to control the processes like cell adhesion and morphogenesis.

Key words: cadherins; alternative splicing; computational analysis; exons; isoforms; *C. elegans*

Background:

Cadherins are calcium dependent cell-surface molecules that have been known to play a pivotal role in cell recognition, adhesion, proliferation and in multiple morphogenetic events in animal development, such as patterning of the central nervous system, and stable tissue formation. [1-3] These are known to occur both in vertebrates and invertebrates like humans, mouse, *C. elegans*, drosophila etc. The cadherin superfamily includes protocadherins, desmogleins and desmocollins, classical cadherins, CNRs (cadherin-related neuronal receptor), flamingo, fat like and many more. The absence of cadherins in yeast, bacteria genomes and other lower organisms suggest that they have evolved for better and much efficient cell-cell interactions in all metazoans where the genome is more complicated and organized.

The *C. elegans* Genome Sequencing Consortium has reported 12 genes in the cadherin superfamily from *C. elegans* genome. [4, 5] As discussed earlier, this superfamily has representatives of almost all the major subtypes of cadherins and related families. It includes, fat-like Cadherins (CDH-4, CDH-3) [6], Seven Helix Membrane Cadherins (CDH-6) [7], a Novel Conserved cadherin (CDH-11) [8], classical cadherins *hmr-1a*, *hmr-1b* [9] etc. The function of most of these cadherins are poorly understood, however classical cadherins are well studied in most of the organisms including *C. elegans*. *C. elegans* genome has a single classical cadherin gene W02B9.1 (*hmr-1*) reported to have two alternatively spliced isoforms W02B9.1a (*hmr-1a*) and W02B9.1b (*hmr-1b*). [9] Alternative splicing is a powerful means of regulating gene expression and enhancing protein diversity. [10] Alternative splicing has been reported in the cadherin superfamily of genes in various organisms [9, 11, 12, 13] and

recently in Protocadherins, a subclass of cadherin superfamily. [14-16] The prediction of alternatively spliced exons using purely bioinformatics means remains problematic till date. [17] This is mainly because the biological rules of splice site selection are still not well understood. It is currently not possible to create programs robust enough to recognize cryptic consensus splice sites that are not normally used by the splicing machinery and also recognize non-consensus splice sites that are actually used by the splicing machinery. Till date, no single approach has been complete in itself to delineate all possible repertoire of spliced transcripts of a gene. [17] So, in our novel approach, we have used an array of bioinformatics tools and programmes including gene/exon finders, ORF analysis tools, splice site prediction tools and many more instead of conventional just EST based or single gene finder based approaches.

Recently, we successfully demonstrated a unique method capable of delineating all possible spliced transcripts of a gene. [17] In the work presented here, we have studied the alternative splicing pattern in the cadherin or cadherin like genes encoded by whole genome of *C. elegans* and have identified new alternatively spliced transcripts of the cadherin encoding genes from *C. elegans* genome. Using bioinformatics tools, such as ORF predicting, exon predicting and various other gene finding programmes, we report the existence of 12 new alternatively spliced transcripts from 12 cadherin encoding genes from *C. elegans* genome. We further confirmed the existence of newly predicted transcript encoded by W02B9.1 gene using RT-PCR for the well-studied classical cadherin.

Methodology:

Animals

The Bristol N2 wild-type strains of *C. elegans*, grown on NGM plates at 20 degree C containing lawn of uracil requiring *Escherichia coli* (OP50) essentially as described in [18] were used in all experiments. Viability of nematodes was assessed using a light microscope.

Reagents

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

The following oligonucleotides primers were custom synthesized from MWG Biotech, Pvt. Ltd., India

1aF1: 5' CAGTGAGGATACGCCGGTTGGAACGG 3'

1bF2: 5' CACGCGCCACACATTCACGGAGCCAC 3'

1cF3: 5' GGTAATTAGGTGGATTTGTGAAGGAGAGG 3'

3R1: 5' CGCGGAAAATACATGGTCTCGGTACGG 3'

Preparation of RNA

Total RNA was isolated from mixed-stage nematodes using the method described earlier. [19] Finally, total RNA was dissolved in diethyl pyrocarbonate-treated distilled water. Purity of RNA was checked using denaturing agarose gel electrophoresis.

Synthesis of first strand cDNA

The cDNAs used for gene amplification by PCR were generated from total RNA isolated from mixed-stage nematodes. Reverse transcription (RT) reactions were performed by initially mixing 2 micro gram of total RNA, 0.2 micro gram oligo (dT)₁₈ primers in a total volume of 10 micro liter, incubating at 65 degree C for 10 minutes and at room temperature for 2 minutes. The following components were then added: 1 micro liter (10 U) ribonuclease inhibitor, 1 micro liter, 0.1 M DTT, 4 micro liter of 5X reaction buffer, 2 micro liter of 30 mM dNTP and 0.5 micro liter (5 U) of Reverse Transcriptase and 1 micro liter of sterile water. The reaction was incubated at 42 degree C for 1 hour. The reaction was stopped at 95 degree C for 2 mins. For RT PCR, 2 micro liter of first strand cDNA was utilized as template in standard PCR reactions.

Polymerase Chain Reaction (PCR)

Single strand cDNA was amplified by PCR in a total volume of 25 micro liter with a template of approximately 50 ng ss cDNA, 25 pmols of forward and reverse primers, 10 mM of each dNTP (Fermentas) and 1.5U of Taq Polymerase (Fermentas). Typical thermal cycling parameters consisted of an initial denaturation of 95 degree C for 5 mins, followed by 35 cycles of denaturation (95 degree C for 45 secs), annealing (60 degree C for 45 secs) and extension (72 degree C for 45 sec). 10 micro liter of PCR product was resolved on 2 percent agarose gel by electrophoresis and photographed on UV transilluminator.

Semi-nested RT-PCR

For further confirmation of results obtained after first round of PCR (for the presence of predicted spliced transcripts), nested PCR was performed. Here, after the first PCR amplification (as detailed above) the resulting RT-PCR product (1 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 nested PCR product (8 micro liter) was then subjected to electrophoresis on a 2 percent (weight/volume) agarose gel, stained with ethidium bromide and photographed on UV transilluminator.

Bioinformatics tools

Genomic sequences of the cadherin encoding genes from *C. elegans* genome were downloaded from the www.wormbase.org (release WS170 date February 12, 2007) and the NCBI nucleotide database (<http://www.ncbi.nlm.nih.gov/entrez>). Various bioinformatics tools, programmes and databases used were same as described. [17]

Result and Discussion:

Organization of W02B9.1 gene and predicted gene products

The classical cadherin encoding gene W02B9.1 (*hmr-1*) is known to have two isoforms W02B9.1a (*hmr-1a*) and W02B9.1b (*hmr-1b*), arising as a result of alternative splicing. [9] Genomic sequence was analyzed using a series of computational tools like gene/exon finding programmes, ORF finders and several other bioinformatics tools (as mentioned in materials and methods). The gene W02B9.1 spans between three clones namely ZK39, Y52B11B and W02B9 (accession numbers Z82093, AL032638 and Z82064 respectively) contains 32 exons. Of these, 31 exons are expressed together in a single transcript giving rise to a hypothetical protein of 2920 amino acid residues, called W02B9.1b (*hmr-1b*) isoform. W02B9.1 contains a large intron between exon 22 and exon 23 which codes for an exon capable of replacing first 22 exons by splicing with 23rd exon, making new exon as a first exon. Thus, splicing of new exon with 23rd exon (second exon in this transcript) makes the transcript highly truncated, containing only ten exons which encodes for a protein of 1223 amino acid residues, called W02B9.1a (*hmr-1a*). There is large untranslated region (approximately 7.5 kb) between upstream gene ZK39.4 and W02B9.1 at the 5' end (Figure 1). [20] We started with the detailed analysis of large 5' UTR and large intron (between exons 22nd and 23rd) in exactly same manner as detailed in our previous study. [17] Briefly, on extensive analysis of these unusually large UTR and intronic gap region through a pre-defined array of different bioinformatics programmes like gene/exon finders, ORF finders, Blast analysis tools, alignment programmes. These tools predicted several new exons possibly capable of replacing the existing exon(s) and thus creating alternatively spliced transcript of the gene. From the several exons predicted above, we selected only the "common exons" capable of replacing the existing exon(s) and thus creating spliced transcript of the gene without causing any change in the reading frame of the protein (Table 2 in supplementary material). Further, all the new exons for respective genes were

analysed using different other programmes to validate the predictions. Lastly, several other parameters like percent-amino acid replacement, codon usage, sense nature i.e. whether from positive or negative strand, the probability score for occurrence of that exon etc. were also checked to ensure accuracy of predicted spliced transcript of the gene.

After the detailed analysis(as described above), we identified one new exon that arise from the large intron between exon 22 and exon 23 of W02B9.1b (*hmr-1b*) capable of splicing with 23rd exon to form a new spliced transcript W02B9.1c (*hmr-1c*). It encodes a hypothetical protein of 1173 amino acids (Figure 1 and Table 1 (supplementary material)); in a similar fashion that generates W02B9.1a (*hmr-1a*) isoform.

First exon of W02B9.1c (*hmr-1c*) transcript encodes 57 amino acids residues (Table 2 in supplementary material). So, we predicted a third transcript of the W02B9.1 (*hmr-1*) that arise as a result of alternative splicing events making a total of three transcripts, namely W02B9.1a (*hmr-1a*), W02B9.1b (*hmr-1b*) and W02B9.1c (*hmr-1c*) (Table 2 in supplementary material). Identification of third spliced variants *hmr-1c* in *C. elegans* was very similar to the characterization of DN-cadherin in *D. melanogaster*, which was the consequence of alternative splicing with the exon from the large intron. [4, 21] We also searched for similarity and homologous sequences for this new, unreported exon, but not much information was available in the database.

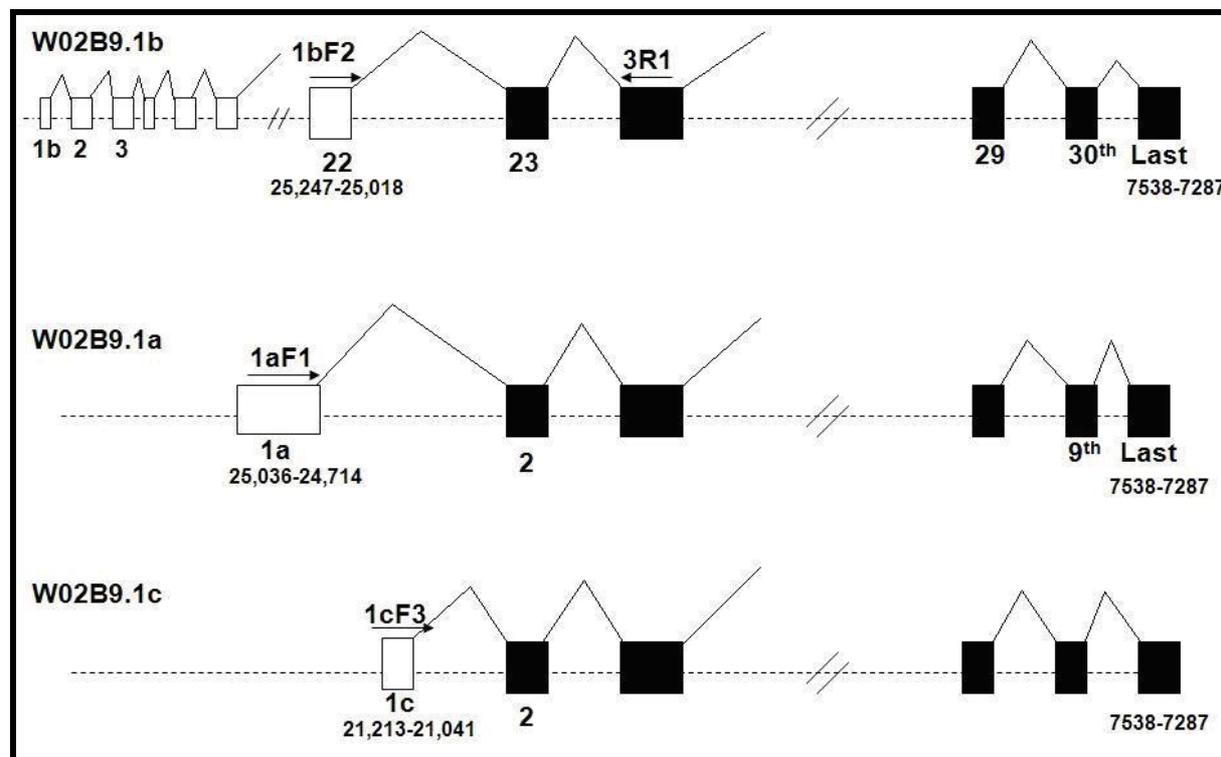


Figure 1: Organization of the W02B9.1 gene of *C. elegans* along with the predicted spliced variants: The Intron/Exon organization of the W02B9.1a (*hmr-1a*), W02B9.1b (*hmr-1b*) and the new predicted spliced variant W02B9.1c (*hmr-1c*). Exons are indicated by rectangular boxes, dotted lines indicate the intronic and the untranslated regions, while solid joining lines show the splicing pattern of each spliced variant. Arrows (1bF2, 1aF1, 1cF3 and 3R1) indicate the Primer designed specific for each exon

Experimental validation of computationally identified new transcripts of W02B9.1 gene

To experimentally validate our prediction, we used RT-PCR approach to confirm the presence of different transcripts encoded by W02B9.1 gene in RNA prepared from mixed stage nematodes. We selected the well studied and characterized cadherin gene *hmr-1*, known to have two spliced variants the W02B9.1a (*hmr-1a*) and W02B9.1b (*hmr-1b*) to be experimentally validated along with our new prediction W02B9.1c (*hmr-1c*). Total RNA was reverse transcribed and the resulting single-stranded cDNA was PCR amplified (as detailed in materials and methods). Exon specific primers used for the identification of splice variants were designed using a combination of bioinformatics tools

and manual methods. Using exon specific primers we calculated the RT-PCR products size which was different for different spliced variants due to presence and/or absence of some sequences. These exon specific primers were able to successfully amplify and confirm the occurrence of the predicted spliced variants by giving a band of anticipated size (Figure 2a and Figure 2b) when products were visualized on agarose gel. Further, to confirm the results obtained after the first PCR, semi-nested PCR was performed, in exactly same method as described above. Thus, we were able to validate computational predictions for existence of new spliced variant of *hmr-1*, W02B9.1c (*hmr-1c*). This paper confirmed the previously reported observation that *C. elegans* classical cadherin gene *hmr-1* encodes two isoforms *hmr-1a*, which is

involved in epidermal morphogenesis [5] and *hmr-1b*, which has a prospective role in neuronal development in [9]. In addition to these, we successfully identified a third transcript

W02B9.1c (*hmr-1c*) that arises as a result of alternative splicing, using computational predictions and RT-PCR combined approach.

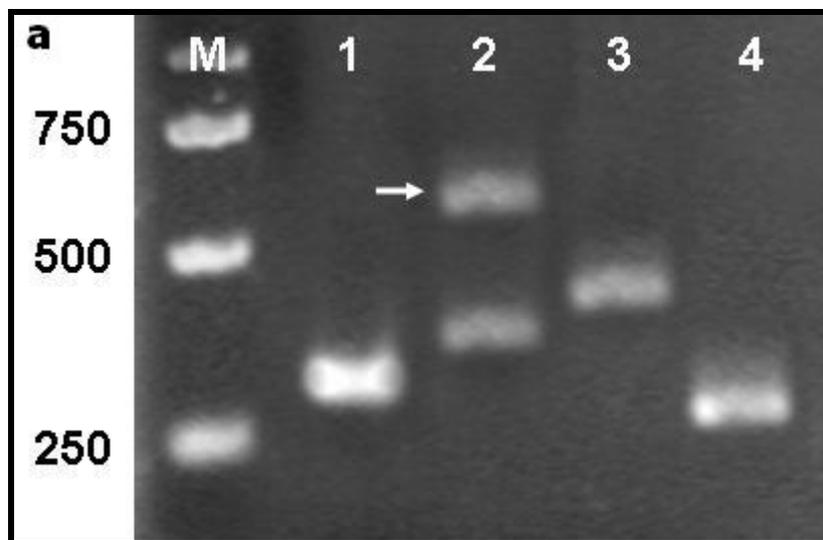


Figure 2: RT-PCR analysis of predicted spliced variant of the *hmr-1* encoding gene W02B9.1 and their product size: RT-PCR amplification was used to determine the presence of transcripts, containing predicted exon, in total RNA prepared from mixed-stage *C. elegans* as described in the Materials and methods. (a) RT-PCR analysis of predicted spliced variant of the *hmr-1* encoding gene W02B9.1: Top panel, the migration of a series of size markers (M) is indicated on the left. RT-PCR products were obtained using a common reverse primer 3R1 (from exon three) and exon specific forward primers (1bF2, 1aF1, 1cF3) representing each spliced variants. A product of and 650 bp (shown by arrow) was observed in lane 2. Primer 1bF2 can either recognize exon 22 of W02B9.1ab or 5' UTR of W02B9.1a transcript. When 1bF2 anneals with 5' UTR of W02B9.1a would result in the amplification product of 650 bp with reverse primer 3R1. Please see Table 3 (supplementary material) for RT-PCR product in each lane

Prediction of new alternatively spliced transcripts in cadherin genes from *C. elegans* genome

In light of previous work and studies, we were inquisitive to study and analyze the splicing pattern with special impetus to finding new undetected spliced variants in cadherin encoding genes from *C. elegans* genome. After downloading the full set of genomic sequences of the cadherin encoding genes from *C. elegans* genome, we analyzed the sequence as discussed above. In all there were only three genes reported to have multiple transcripts from a total of 12 cadherin genes, encoding 16 gene products from *C. elegans* genome (Table 1 in supplementary material). In addition to W02B9.1 (detailed above), F08B4.2 and B0034.3 genes are reported to have two (F08B4.2a and F08B4.2b) and three (B0034.3a, B0034.3b and B0034.3c) alternatively spliced isoforms respectively. Our computational analysis predicted alternative splicing in 7 of the total 12 cadherin genes giving rise to 12 new alternatively spliced transcripts (Table 1 in supplementary material) making a total of 28 gene products.

Following the computational predictions of new spliced isoforms, Yuji Kohara's *C. elegans* EST database was searched for putative EST/cDNA support for possible occurrence of these new exons/transcripts. A search of Yuji Kohara's *C. elegans* EST database (<http://nematode.lab.nig.ac.jp/dbest/keysrch.html>) did not yield any EST match for these new exons. The most probable reason for this being that the current available EST database for *C. elegans* is not adequately represented and so far at least

40% of the genes in the organism (~19,000) are not reflected in this database. The NCBI BLAST (blastp) search was performed for homology or prospective similarity with other polypeptides of these new spliced transcripts. It compares an amino acid query sequence against a protein sequence database. Here, on blastp, we have displayed (Figure 3) homology match of the computationally predicted newly exon with all other polypeptides from other genomes, no specific scaling down was done using E-value or Score (bits) criteria. Our aim here was to show that few of our computational predicted new alternative exons sequences had known polypeptide matches with that of cadherin sequences from other genomes. Out of the 7 genes in which we found alternative splicing, the computationally predicted spliced variant of F08B4.2 gene (namely F08B4.2c and F08B4.2d) showed significant similarity to "Protocadherins" and related proteins found in various other organisms like *Caenorhabditis briggsae* (Figure 3a), *Danio rerio* (Figure 3b, Figure 3d and Figure 3e), *Mus musculus* (Figure 3c). This new exon shares maximum homology with a hypothetical protein found in *C. briggsae*. The homology match between the newly predicted exon(s) to Protocadherins and related proteins found in *C. briggsae* and other organisms clearly indicates the evolutionary significance of the cadherin group in various species. Secondly, it also supports the hypothesis that cadherins have evolved for better and much efficient cell-cell interactions in almost all metazoans. The consensus sequences at the splice-donor or acceptor site are believed to be the structural features that determine whether a specific splice site

is used although these consensus sequences may not be always present. [19] Analysis of the exon-intron junction region of the newly predicted exons of various cadherin-encoding genes indicates structural conservation, as depicted (Table 1 in supplementary material). The presence of consensus sequences in the splice-donor /acceptor site provides further confirmatory evidence for the existence of

new exons of various cadherin encoding genes from *C. elegans* genome. Thus, we conclude that the strategy as detailed above for W02B9.1, we can very well experimentally validate all our computationally predicted new spliced transcripts of cadherin encoding genes from *C. elegans* genome.



Figure 3: Homology between the newly predicted first exon of F08B4.2 with that of Protocadherins found in various other organisms: Homology search using NCBI BLAST for newly predicted exon. The homologous amino acid sequence encoded by predicted first exon of F08B4.2 with protocadherins found in various organisms *C. briggsae* (3a), *Danio rerio* (3b, 3d and 3e) and *Mus musculus* (3c)

Determining the extent and importance of alternative splicing required the confluence of critical advances in data acquisition, improved understanding of biological processes and the development of fast and accurate computational analysis tools. In spite of a large pool of methods available, none of the approaches used till date, have been fully successful in delineating all possible alternative splice transcripts of a gene. So, there was a strong need to develop a method capable of delineating the full spectrum of all the possible spliced transcripts of a gene. Our results demonstrate that, we are still far from completely deciphering these hidden

spliced transcripts from the genome of sequenced organism and perhaps most of the studies have probably underestimated the extent of alternative splicing. This could be the reason why number of gene products is suspected to be underestimated. Our findings could assist other biologists in many ways: Firstly, they not only point towards the extent of occurrence of alternative splicing in the cadherin superfamily of genes in *C. elegans* genome but will also enhance the alternative splicing database of *C. elegans*, which could be further studied for better understanding of more complex genomes. It may motivate other researchers to take up similar

studies in other sequenced genomes as well. Secondly, our results not only point towards the need of developing more efficient algorithms and methods capable of identifying alternative spliced transcripts but also toward the need of analyzing genome data using a combination of gene/exon finders to delineate all possible gene products and to decipher the true extent of alternative splicing in *C. elegans* genes. Lastly, due to limited zone of our work, further studies using more advanced and sophisticated techniques like the RNA interference (RNAi) could be taken up aimed at assessing the biological and functional significance of these spliced transcripts and their possible hand in *C. elegans* gene working and regulation.

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] R. O. Hynes, *Trends Cell Biol.*, 9:33 (1999) [PMID: 10611678]
- [02] U. Tepass, *Curr. Opin. Cell Biol.*, 11:540 (1999) [PMID: 10508657]
- [03] K. Vleminckx & R. Kemler, *Bioessays*, 21:211 (1999) [PMID: 10333730]
- [04] E. Hill, *et al.*, *J. Mol. Biol.*, 305:1011 (2001) [PMID: 11162110]
- [05] E. A. Cox, *et al.*, *J. Cell Sci.*, 117:1867 (2004) [PMID: 15090591]
- [06] J. Pettitt, *et al.*, *Development.*, 122:4149 (1996) [PMID: 9012534]
- [07] T. Yagi & M. Takeichi, *Genes Dev.*, 14:1169 (2000) [PMID: 10817752]
- [08] L. Vogt, *et al.*, *Mol. Cell. Neurosci.*, 17:151 (2001) [PMID: 11161476]
- [09] I. D. Broadbent and J. Pettitt, *Curr. Biol.*, 12:59 (2002) [PMID: 11790304]
- [10] A. J. Lopez, *Annu. Rev. Genet.*, 32:279(1998) [PMID: 9928482]
- [11] R. Kawano, *et al.*, *J. Biol. Chem.*, 277:47679 (2002) [PMID: 12364338]
- [12] H. Makarenkova, *et al.*, *Biochim Biophys Acta.*, 1681:150 (2005) [PMID: 15627506]
- [13] X.Q. Wang, *et al.*, *Clin. Cancer Res.*, 11:483 (2005) [PMID: 15701831]
- [14] M. Goldberg, *et al.*, *Am. J. Physiol. Renal. Physiol.*, 283:454 (2002) [PMID: 12167596]
- [15] M. Frank and R. Kemler, *Curr. Opin. Cell Biol.*, 14:557 (2002) [PMID: 12231349]
- [16] K. Yoshida, *Cell. Mol. Biol. Lett.*, 8:735 (2003) [PMID: 12949613]
- [17] L. Kashyap, *et al.*, *Bioinformatics*, 2:17 (2007)
- [18] S. Brenner, *Genetics*, 77:71 (1974) [PMID: 4366476]
- [19] M. Tabish, *et al.*, *Biochem. J.*, 339:209 (1999) [PMID: 10085246]
- [20] *C. elegans* Sequencing Consortium, *Science*, 282:2012 (1998) [PMID: 9851916]
- [21] M. D. Adams, *et al.*, *Science*, 287:2185 (2000) [PMID: 10731132]

Edited by M. K. Sakharkar

Citation: Kashyap *et al.*, *Bioinformatics* 2(2): 50-56 (2007)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

Supplementary material

LG	Cosmid (Gene)	Spliced Isoform	5' Exon-Intron boundary	Intron-3' Exon boundary	Exons organization in transcripts	Amino acid
I	W02B9.1b	Yes*‡	(1b) TTTGgtaa	tcagGCAA (2)	1b-2----23---30-Last	2920
I	W02B9.1a	Yes*‡	(1a) TCTCgtaa	ccagATTC (2a)	1a-23 (2a)---30-Last	1223
I	W02B9.1c	Yes‡	(1c) AGAAgtac	ccagATTC (2a)	1c-23 (2a)---30-Last	1173
II	B0034.3a	Yes*	(1a) CGAGgtaa	tcagCGCC (2)	1a-2-3---9-10-11-Last	984
II	B0034.3b	Yes*	(1b) CGTGgtaa	ttagAACA (9)	1b-9-10-11-Last	167
II	B0034.3c	Yes*	(1c) CCTGgtaa	tcagGTGG (10)	1c-10-11-Last	160
II	R05H10.6	No*	(1) GGAGgtga	ccagACCT (2)	1-2-3---9-Last	1364
III	F25F2.2	No*	(1) TGTGgtaa	TcagGGTG (2)	1-2-3---9-Last	4307
III	Y71D11A.1a	Yes*	(1a) GTTGgtga	ccagAATC (2)	1a-2-3-4-5-6-7-Last	506
III	Y71D11A.1b	Yes	(1b) TGCGgtaa	tcagACCG (3)	1b-3-4-5-6-7-Last	444
III	ZK112.7a	Yes*	(22) AATGgtga	tcagAAAT (last)	1-2----20-21-22-Last	3343
III	ZK112.7b	Yes	(22) AATGgtga	tcagACGA (last')	1-2----20-21-22-Last'	3427
III	ZK112.7c	Yes	(21) TGAGgtga	tcagACGA (last')	1-2----20-21-Last'	3215
III	ZK112.7d	Yes	(20) TAAGgttg	tcagACGA (last')	1-2---19-20-Last'	2739
III	R10F2.1a	Yes*	(1) TAAGgtga	tcagTTCG (2)	1-2---18-19-Last	2779
III	R10F2.1b	Yes	(1b) TGGGgtga	tcagTTCG (2)	1b-2-3---18-19-Last	2762
III	R10F2.1c	Yes	(1b) TGGGgtga	ttagCTAT (3)	1b-3---18-19--Last	2702
III	R10F2.1d	Yes	(1b) TGGGgtga	tcagTTAG (4)	1b-4---18-19--Last	2612
IV	C45G7.5a	Yes*	(12) GTATgtga	tcagATTT (13a)	1-2---12-13a-14---20-Last	2302
IV	C45G7.5b	Yes	(12) GTATgtga	gcagTTTT (13b)	1-2---12-13b-15---20-Last	1854
IV	F08B4.2a	Yes*	(21) CGAAgtga	tcagTTCT (last)	1-2-3---21-Last	1544
IV	F08B4.2b	Yes*	(21) CGAAgtga	tcagACGT (last')	1-2-3---21-Last'	1306
IV	F08B4.2c	Yes	(1') AATAgtca	tcagAATC (5)	1'-5---20-21-Last	1377
IV	F08B4.2d	Yes	(1') AATAgtca	tcagAATC (5)	1'-5---20-21-Last'	1139
IV	F18F11.3	No*	(1) ACTGgtaa	GcagTTTT (2)	1-2-3---8-Last	427
V	F15B9.7	No*	(1) AGAAgtga	TcagATAT (2)	1-2-3---25-Last	2577
X	F59C12.1a	Yes*	(1a) CCAGgtat	gcagGAAC (2)	1a-2-3-4---13-Last	623
X	F59C12.1b	Yes	(1b) CACAgtaa	tcagGGAG (3)	1b-3-4---13-Last	576

Table 1: A summary of *C. elegans* genome encoding cadherin genes, Intron-Exon boundaries, alternatively spliced variants and computational predictions: A List of cadherin encoding genes from *C. elegans* chromosomes (LG I-V and X). Transcripts indicated with astricks (Yes*) were predicted by *C. elegans* sequencing consortium and the remaining were identified by our detailed analysis. Multiple transcripts of W02B9.1 indicated with '‡' were confirmed by RT-PCR analysis as well. Nucleotides indicated in lower case are part of introns and in upper case are parts of exons at the exon-intron-exon boundaries. Numbers mentioned under brackets before and after the sequences indicate the splicing exons. Spliced variant is indicated as a, b, c, d in their respective column.

Isoform (Gene)	First exon amino acid sequence	First exon size (a.a.)
W02B9.1a (<i>hmr-1a</i>)	MPETRQGRSSRGNCVIFGSSKRLWVTLGLGFCFVLSLIGGAEAFDLSLPGLE PSVAKSRFSSLVGGVRRARDIHVFVMKNISEDTPVGTVLETFKAHDPSNPMYNF <u>FRINRQSD</u>	107
W02B9.1b (<i>hmr-1b</i>)	MSWNILLILLISNLDEVL (plus sequences of exon 2 through exon 22) <u>FRINRQSD</u>	18
W02B9.1c (<i>hmr-1c</i>)	MANFWLFLSLHPKLTRWVIRWICEGEASPLLLILIYNGNNKCNKVKKKKK <u>KKKKKKFRINRQSD</u>	57

Table 2: Amino acid sequence encoded by the predicted alternatively spliced N-terminal exons: Deduced amino acid sequences encoded by alternative first exons (shown in regular font) and part of the amino acid sequence encoded by exon 2 (underlined). Number of amino acids encoded by first exons 1a, 1b and 1c are shown in the last column representing W02B9.1a (*hmr-1a*), W02B9.1b (*hmr-1b*) and W02B9.1c (*hmr-1c*) respectively.

Lane	PCR Product	Primer set (5'---3')	Predicted size
1	Exon 1a----exon 3	1aF1 Vs 3R1	311
2	Exon 22b---exon 3	1bF2 Vs3R1	346 and 650
3	Exon 1c----exon 3	1cF3 Vs 3R1	359
4	Exon 1c---exon 3(internal)	1cF3 Vs 3R2	295

Table 3: RT-PCR product for each lane in figure 2 is given.