[Frontiers in Bioscience 19, 504-514, January 1, 2014]

Identification of *C. elegans* & *C. briggsae* miRNAs by modified miRsearch

Avirup Dutta[1], Raghunath Chatterjee[23], Keya Chaudhuri[1]

[1]Molecular and Human Genetics Division, Indian Institute of Chemical Biology, Kolkata-700032, India, [2]National Cancer Institute, NIH, Building 37 room 3128, Bethesda, MD 20892, [3]Human Genetics Unit, Biological Science Division, ISI, Kolkata 700108

TABLE OF CONTENTS

1. ABSTRACT

In this study, a modified miRsearch program was developed in C++ for the detection of miRNAs. All the mature miRNA sequences of *Caenorhabditis elegans, Caenorhabditis briggsae, Caenorhabditis remanei, Drosophila melanogaster*, *Homo sapiens* and *Rattus norvegicus* available in miRbase was searched by this program for homologous sequences with a maximum of 3-mismatches in the chromosomes of *C. elegans* excluding the miRNAs of *C. elegans*. The same strategy was repeated for *C. briggsae* excluding *C. briggsae* miRNAs. The probable pre-miRNA sequences with stem loop secondary structures were assessed by implementation of *longest-common-subsequence* (LCS) algorithm with appropriate scoring system. As miRNA genes could be on either strand, each sequence was searched in both forward and reverse strands. The putative miRNAs were viewed through Mapviewer to identify their intronic or intergenic location in *C. elegans* and *C. briggsae* genomes. Further, the quality of stem-loop formation of the remaining pre-miRNA sequences was assessed through RNAFOLD. This algorithm will be helpful in detection of potential miRNAs in future sequencing data, making this an invaluable tool for miRNA prediction.

2. INTRODUCTION

MicroRNAs (miRNAs) are an abundant class of endogenous, small, non-coding RNAs typically ~21nucleotide (nt) long expressed in plants, viruses and animals (1-4). These miRNAs are key regulators of gene expression at post transcriptional level and revolutionized our understanding of gene regulation (5, 6). Depending on the degree of complementarity between miRNA and its target transcript, miRNAs are known to cause degradation of target transcript leading to translational repression (7). miRNAs were first discovered in the nematode *Caenorhabditis elegans*, the founding members being *lin-4* and *let-7* gene products of *Caenorhabditis elegans* (8, 9), both of these act as repressors of their respective target genes *lin-14*, *lin-28* and *lin-41* (8, 10, 11). In all these cases repression was mediated by the presence of complementary miRNA sequences in the 3′ untranslated regions (UTRs) of the target mRNAs (11, 12). Since the discovery of the first miRNA, thousands of them have been identified in diverse organisms through random cloning and sequencing or computational prediction and analysis. Reports also suggest that many miRNAs are highly conserved across species (13-17).

Experimental data have shown that several miRNAs participate in essential biological processes and regulate functionally important pathways related to development (9), cell death (18), cell proliferation control (19, 20), hematopoiesis (19, 20), patterning of nervous system (19), pancreatic cell insulin secretion (20), adipocyte development (19, 20), cancer (21) and neurological diseases (22), angiogenesis and vascular integrity (23).

Direct cloning and sequencing of short RNA molecules has enabled the identification of many miRNAs; however, highly constrained tissue- and time-specific expression patterns, presence of degradation products from mRNAs, and other noncoding RNAs, has made it difficult and incomplete to clone miRNAs (24, 25). For finding those low-expression or tissue-specific miRNA genes, computational prediction provides an efficient strategy (13-17). The basic principle of the computational approaches is simple- they rely on the known characteristics of miRNAs and search those in other organisms. MiRNA detection relies on (i) conservation of miRNAs in the genomes of related species, (ii) formation of stable stem-loop structure by pre-miRNAs, (iii) the presence of mature miRNAs in the stem and not in the loop of pre-miRNAs and (iv) the presence of the mature miRNA in the intronic or the intergeneic regions.

The present study is a modification of a previously reported algorithm miRsearch for detecting miRNAs (26). In this study, the algorithm for detection of pre-miRNA sequence has been completely modified by (i) implementing *longest-common-subsequence* (LCS) algorithm (27) (ii) standardization of allowed mismatches during the search process and (iii) identification of possible new miRNAs in the *C. elegans and C. briggsae* genomes.

3. MATERIALS AND METHODS

3.1. Sources of nucleotide sequences

All 3410 mature miRNAs available from *C. elegans*, *C. briggsae*, *C. remanei, Drosophila melanogaster*, *Homo sapiens* and *Rattus norvegicus* were selected from the ftp site of Mirbase database (28, 29). The complete genome sequences of *C. elegans* and *C. briggsae* were downloaded from the ftp site at the NCBI (ftp://ftp.ncbi.nih.gov/genomes/).

3.2. Computational Screening of miRNA

The computational screening of miRNA was executed through the program modified miRsearch. Algorithm miRsearch, which was initially written in perl script (26), was modified and written in C++. The program started with a known mature miRNA sequence and searches in other organisms for their close homologs with mismatches using user given score where the S score is calculated as twice the difference of the length of the miRNA and twice the number of mismatches as defined earlier (26). Here mismatches are given more penalties to reduce the number of false positives. The mature miRNA sequences of *C. elegans* and *C. briggsae, C ramanei, D. melanogaster*, *H. sapiens* and *R. norvegicus* were used as probes by the modified miRsearch to search for their homologues in the chromosomes of *C. elegans* and *C. briggsae*. The whole dataset of mature miRNA sequences was divided into 3 sets- (i) *C. elegans* miRNAs (ii) *C. briggsae* miRNAs and (iii) Human, Rat and Drosophilla miRNAs. Datasets (ii) and (iii) were used for the *C. elegans* genome and datasets (i) and (iii) for *C. briggsae*. Both the forward and reverse complement sequences of complete genomes were searched for miRNAs. During the search, mismatch was allowed ranging from zero mismatches and was extended up to three mismatches.

For the detection of pre-miRNA sequence, the algorithm has been completely modified. To assess the stem loop secondary structure of the pre-miRNA sequences the *longest-common-subsequence* (LCS) algorithm (27) was implemented. The LCS problem is to find the maximum possible length of a common subsequence of two strings, 'a' of length |a| and 'b' of length |b|. Usually an actual LCS is also required. For example, using the alphabet A, C, G and T of genetic *bases*, an LCS of 'GCTAT' and 'CGATTA' is 'GTT' of length three. The LCS will determine the best possible match of the two input sequences to a LCS common-subsequence; in some cases this will result in multiple 'best possible' matches. Here we have implemented a scoring system based on the Watson:Crick pairing. Our algorithm will keep the match with best score. If multiple matches has similar score, then this phase of the algorithm will consider all pairs for further quality checking, e.g., free energy calculation, genomics locations of the miRNA. Here, reverse complement of the homolog of the query mature miRNA sequence was chosen as one sequence, and the other sequences were chosen from -80 to -1 from the first base of the target sequence and +1 to +80 from the last base of the target sequence. For each pair of sequences the LCS is calculated and correspondingly, a score (R) is given for each LCS as above (i.e., for A, T base pairing a score of 2 and for G, C base pairing a score of 3 is given). The scoring was done based on the fact that, A, T base-pairing involves a double hydrogen bond whereas the C, G base-pairing involves a stronger triple bond. More energy is required to break a C, G base-pair than an A, T base-pair and hence less likely to undergo a mutation than an A, T base pair. The sequence for which max score is obtained is chosen as the reverse strand of the pre-miRNA sequence. If the LCS is within the -80 to -1 then the other arm of the hairpin loop precursor miRNA is in the upstream of the target sequence else if it is within +1 to +80 then the other arm of the hairpin loop pre-miRNA is downstream of the target sequence. If the LCS is upstream, then pre-miRNA sequence was chosen as -10 nt of the start of the LCS to +10 nt of the end of the mature miRNA.

The results thus obtained were then segregated using a customized perl script depending on the mismatches. The pre-miRNA sequences thus obtained were passed through the RNA folding program RNAFOLD (30) selecting the cut-off values on the basis of training datasets (i.e., known pre-miRNA sequences). The observed characteristics from the RNAFOLD output were used to detect the new pre-miRNAs in an organism under study. A

structure is accepted as probable miRNA if they satisfy all the characteristics obtained from RNAFOLD output, i.e., (a) $\Delta G \leq$ -20.0K.cal/mole, (b) the longest helical arm contains at least (20-29) bp depending on miRNA sequence length and (c) the predicted miRNAs are within the long helical arm.

Using another perl script the co-ordinates of the probable miRNAs were finalized and checked through NCBI map viewer for their location if they were present in the intronic or intergenic regions or not as miRNAs are not supposed to be present in exonic region. The ones present in the exonic regions were rejected and only the ones present in the intronic or intergenic regions were selected and the corresponding sequences were used. The complete screening process starting from searching of the mature miRNA in the chromosome to final selection of probable new miRNA is explained in Figure 1.

3.3. Calculation true positive, true negative, false positive and false negative

To test the efficiency of modified miRsearch, the percentages of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) were calculated based on the datasets used for the search. To get an accurate estimate of the TP, TN, FP and FN, the number of miRNAs of each set, were carefully quantified and the identical miRNAs were excluded. For the *C. elegans* genome, the *C. briggsae* miRNAs which were identical to the *C. elegans* miRNAs and vice versa for *C. briggsae*, were not excluded during the search process, but they were excluded during the calculation of TP, TN, FP and FN as shown below.

3.3.1. True positive

Number of *C. elegans* and *C. briggsae* miRNAs identified with 0 mismatch from *C. elegans* and *C. briggsae* genomes respectively.

3.3.2. False negative

Number of *C. elegans* and *C. briggsae* miRNAs not identified with 0 mismatch from *C. elegans* and *C. briggsae* genomes respectively.

3.3.3. False positive

Number of miRNAs identified with 0 mismatch using *C. briggsae* dataset on *C. elegans* genome excluding identical *C. elegans* miRNAs and vice versa for *C. briggsae* genome.

3.3.4. True negative

Number of miRNAs not identified with 0 mismatch using *C. briggsae* dataset on *C. elegans* genome excluding identical *C. elegans* miRNA and vice versa for *C. briggsae* genome.

4. RESULTS

4.1. Computational Prediction of *C. elegans* and *C. briggsae* miRNA by Modified miRsearch

Evolutionary conservation of mature miRNA sequences and their characteristic of forming stem-loop secondary structure have been used in modified miRsearch to identify the miRNA sequences in the genomes of *C. elegans* and *C. briggsae*. Using all human, rat, drosophila, *C. elegans, C. ramanei* and *C. briggsae* mature miRNA sequences, available in miRbase database (28, 29) (November, 2011), homologous sequences with a maximum of 3 mismatches was searched for in the chromosomal sequences of *C. elegans* and *C. briggsae*. Homologous sequences, with 3 mismatches, may be present in many places in the genome, all of which may not have the capability of forming stem-loop precursor structure characteristics of pre-miRNAs. The LCS implementation of modified miRsearch could assess the stem-loop structure of probable miRNAs. Reverse complements of the homologue of miRNA sequences (reverse match) were searched at a position –80 to +80 from the matched sequence and their capability of stem-loop formation were assessed using the modified miRsearch. Modified miRsearch assigned a proper score value to the reverse matching sequence. The program was trained with all known miRNA sequences and we empirically set the minimum score value obtained from these miRNA sequences as the cut off score for *C. elegans* and *C. briggsae* miRNA. As miRNA genes can be located on either strand, we searched each sequence in both the forward strand as well as in its reverse complement. The putative miRNAs thus obtained was viewed through Mapviewer to identify their location in the *C. elegans* and *C. briggsae* genomes. miRNAs present in the exonic region were excluded from the set. Further evaluation of the quality of stem-loop formation of the remaining pre-miRNA sequences was assessed through the RNA folding program RNAFOLD and some selection procedure set empirically as mentioned in section on materials and methods.

In our previous report, we had reported 50 possible novel miRNAs in *A. gambiae.* Some of these novel miRNAs predicted by our algorithm miRsearch (26) has been used by other laboratory (31) and experimentally verified in Asian mosquito. In the present study out of the 368 *C. elegans* miRNAs reported in miRbase database, 364 miRNAs (i.e., ~98.91%) were detected in this method. In case of *C. briggsae*, of 131 miRbase reported miRNAs, 127 were detected (~96.94%). A cumulative query dataset of 3410 miRNAs of *C. elegans, C. briggsae, C. ramanei, D. melanogaster, H. sapiens* and *R. norvegicus* have been used for detecting miRNAs in *C. elegans* and *C. briggsae* genomes.

Table 1 represents the miRNAs which had passed through all level of screening after their prediction through modified miRsearch with 100% homology in *C. elegans* and *C. briggsae* genomes. The modified miRsearch managed to identify 24 and 71 probable new miRNAs in the genomes of *C. elegans* and *C. briggsae* respectively with 0 and 1 mismatch (Table 2a & b). The TP/FP/TP/TN were all calculated using the following datasets: 368 *C. elegans* and 131 *C. briggsae* known miRNAs for the calculation of TP and FN in the *C. elegans* and *C. briggsae* chromosomes respectively; whereas 333 *C. elegans* and 96 *C. briggsae* known miRNAs for the calculation of FP and

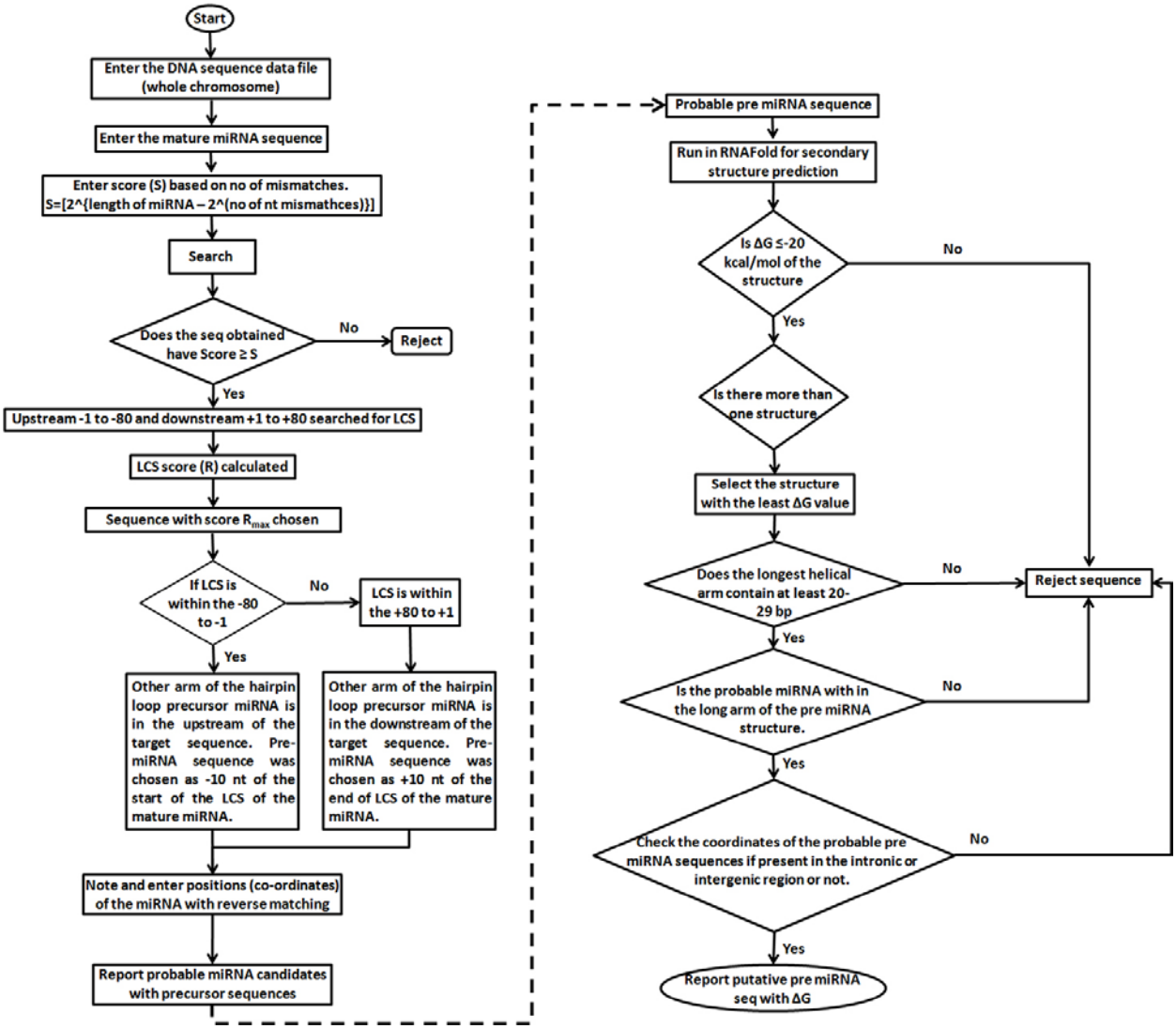


Figure 1. Algorithmic flowchart of modified miRsearch.

TN *C. briggsae* and *C. elegans* chromosomes respectively. The results for TP (364/368 for *C. elegans* and 127/131 for *C. briggsae* miRNAs, identified with 0 mismatch), TN (90/96 for *C. elegans* and 314/333 for *C. briggsae* miRNAs, not identified with 0 mismatch), FP (6/96 for *C. elegans* and 19/333 for *C. briggsae* miRNAs, identified with 0 mismatch) and FN (4/368 for *C. elegans* and 4/131 for *C. briggsae* miRNAs, not identified with 0 mismatch) are shown graphically in Figure 2. The results showed that ~98.91% and ~96.94% were true positive in case of *C. elegans* and *C. briggsae* respectively. The overall quality of the prediction was tested using precision, recall and F-score as shown in the summary statistics of Table 3.

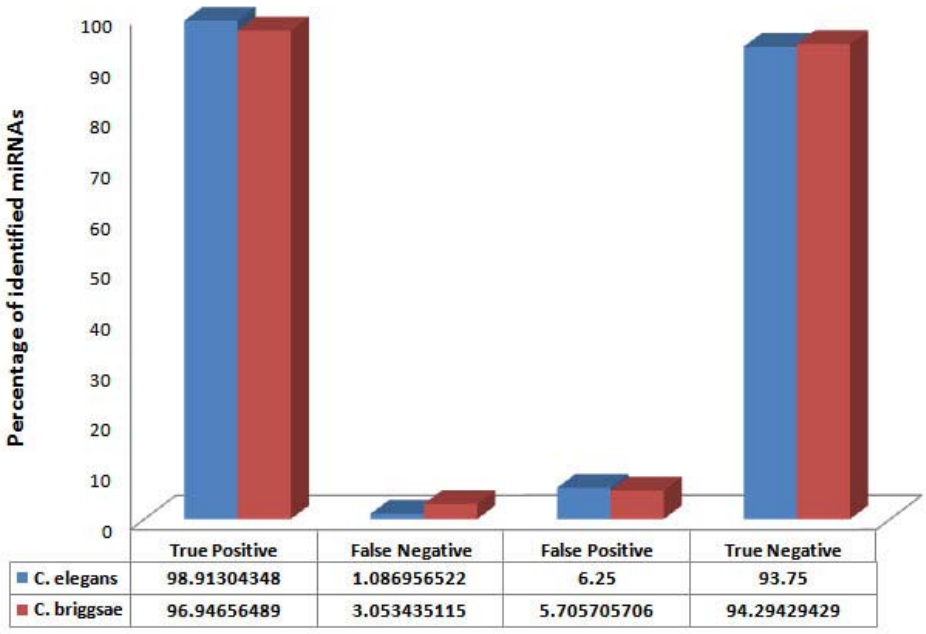
Four new miRNAs were found of *C. remanei* in the chromosome of *C. elegans* with zero mismatches, out of which three were common with *C. briggsae* miRNAs. Four and five new miRNAs of *C. remanei* and *C. briggsae* were identified in the *C. elegans* chromosomes with 1 mismatch of which two miRNAs were common to both and *C. briggsae*. Only one and seven new *C. remanei* miRNAs were found in *C. elegans* chromosomes with 2 or 3 mismatches respectively (Suppl Table 1a). Interestingly, all other newly identified miRNAs in *C. elegans* chromosomes except for one (Drophilla miRNA) were homologous to human miRNAs with a single mismatch (Table 2a).

In case of *C. briggsae*, 14 new *C. elegans* miRNAs and 4 new *C. remanei* miRNAs were identified with 100% homology out of which 7 miRNAs were common to both *C. elegans* and *C. remanei*. Even with a single mismatch, one new miRNA was identified homologous to *C. remanei* and 11 *C. elegans* miRNAs. Apart from these, all other miRNAs were homologous to human miRNAs with 1 mismatch except for one rat miRNA. When the mismatches were relaxed up to two and eventually to three, the number of miRNAs increased dramatically for both *C. elegans* and *C. briggsae* chromosomes with homologues mostly from human, drosophila and rat and a few from other species of

Table 1. *C. elegans* and *C. briggsae* miRNA predicted by modified miRsearch to and 100% similar to already reported miRNA

Sl No.	Position of miRNAs in *C. elegans*				100% homologous miRNAs in *C. elegans*	Position of miRNAs in *C. briggsae*				100% homologous miRNAs in *C. briggsae*
	Chr	Strand	Start	End		Chr	Strand	Start	End	
1	1	+	12608606	12608690	miR-1823	1	+	2688441	2688511	miR-244
2	1	+	13513942	13514021	miR-1824-5p	1	-	6757250	6757327	miR-245
3	1	+	13513934	13514031	miR-1824-3p	1	-	2601058	2601153	miR-353
4	1	+	10600427	10600523	miR-1828	1	-	8823489	8823585	miR-235b*
5	1	+	9447316	9447423	miR-2211-3p	1	-	8823504	8823578	miR-235b
6	1	+	13812570	13812638	miR-2218a-5p	2	+	319390	319475	miR-234
7	1	+	13812565	13812650	miR-2218a-3p	2	+	12875635	12875720	miR-2222
8	1	+	422899	422967	miR-4805-5p	2	+	1879130	1879215	miR-2227
9	1	+	1812669	1812764	miR-4805-3p	2	+	1875312	1875397	miR-2225
10	1	+	2921193	2921287	miR-4931	2	-	8058564	8058648	miR-44
11	1	+	9512309	9512375	miR-4932	2	-	8058564	8058648	miR-45
12	1	+	11885593	11885663	miR-5545-5p	2	-	7467799	7467884	miR-252
13	1	+	11885585	11885681	miR-5545-3p	2	-	6505552	6505637	miR-355
14	1	+	863958	864053	miR-5547-3p	3	+	13421059	13421154	miR-90
15	1	-	10869360	10869456	miR-1817	3	+	9825062	9825158	miR-231
16	1	-	13980835	13980931	miR-1818	3	+	6833085	6833180	miR-357
17	1	-	13834840	13834936	miR-2218b-5p	3	-	13422683	13422779	miR-90b*
18	1	-	13834850	13834938	miR-2218b-3p	3	-	13422693	13422775	miR-90b
19	1	-	1546628	1546724	miR-4805-5p	3	-	218516	218600	miR-65
20	1	-	2888450	2888546	miR-5546-5p	3	-	218514	218593	miR-65*
21	1	-	2888461	2888538	miR-5546-3p	3	-	218796	218868	miR-64b
22	1	-	454220	454308	miR-5547-3p	3	-	8971626	8971711	miR-2214
23	1	-	1861888	1861967	miR-5595-3p	3	-	219142	219238	miR-2226
24	2	+	6015486	6015581	miR-1822-3p	3	-	219152	219240	miR-2226*
25	2	+	4284430	4284525	miR-1832	4	+	149549	149653	miR-1834
26	2	+	11600195	11600285	miR-2207-5p	4	-	8555781	8555866	miR-228
27	2	+	11600193	11600299	miR-2207-3p	4	-	8548032	8548116	miR-790
28	2	+	4058282	4058378	miR-2217-3p	5	+	12272735	12272809	miR-241
29	2	+	6747094	6747189	miR-2953-3p	5	+	8908227	8908311	miR-255
30	2	+	238576	238666	miR-4805-5p	5	-	7813060	7813144	lsy-6
31	2	+	1061657	1061742	miR-4805-3p	5	-	9138944	9139029	miR-253
32	2	+	12238519	12238596	miR-5547-5p	5	-	1214378	1214455	miR-358
33	2	+	6065134	6065218	miR-5547-3p	X	+	10968465	10968550	miR-230
34	2	+	241486	241565	miR-5552-5p	X	+	1703779	1703875	miR-233
35	2	+	431481	431558	miR-5593-5p	X	+	4610292	4610388	miR-248
36	2	+	431469	431564	miR-5593-3p	X	+	11095215	11095300	miR-360
37	2	-	4371818	4371913	miR-1830-5p	X	+	15175325	15175404	miR-239a
38	2	-	4371823	4371913	miR-1830-3p	X	+	13573439	13573534	miR-240
39	2	-	13333652	13333736	miR-2215-5p	X	+	5320054	5320150	miR-254
40	2	-	13333651	13333728	miR-2215-3p	X	+	7891067	7891162	miR-785
41	2	-	10867011	10867096	miR-2216-5p	X	+	13573561	13573646	miR-786
42	2	-	10867019	10867093	miR-2216-3p	X	+	10981890	10981975	miR-791
43	2	-	1567906	1567991	miR-4805-5p	X	+	7890610	7890694	miR-359
44	2	-	12346196	12346281	miR-4806-5p	X	+	19202537	19202632	miR-392
45	2	-	12346207	12346281	miR-4806-3p	X	+	15333583	15333668	miR-35c
46	2	-	13461665	13461742	miR-4929	X	+	16201679	16201763	miR-237*
47	2	-	1804987	1805064	miR-5547-3p	X	+	15333872	15333968	miR-35b
48	2	-	2284823	2284893	miR-5548-3p	X	-	5851564	5851660	miR-251
49	2	-	287605	287712	miR-5552-5p	X	-	7892506	7892583	miR-249
50	2	-	270725	270809	miR-5593-5p	X	-	13820033	13820117	miR-784
51	2	-	11865020	11865104	miR-5594-5p	X	-	16579516	16579613	miR-788
52	2	-	11865026	11865100	miR-5594-3p	X	-	16837118	16837208	miR-54b
53	2	-	14399436	14399532	miR-5595-5p	X	-	4493702	4493787	miR-74b*
54	2	-	14399447	14399537	miR-5595-3p	X	-	4493707	4493781	miR-74b
55	3	+	13660035	13660112	miR-46-5p					
56	3	+	13660017	13660123	miR-46-3p					
57	3	+	11829103	11829198	miR-1832					
58	3	+	6968791	6968857	miR-2213-5p					
59	3	+	7278614	7278711	miR-2214-3p					
60	3	+	10867310	10867378	miR-4805-5p					
61	3	+	3249133	3249238	miR-4936					
62	3	+	421554	421649	miR-5547-3p					
63	3	-	12780308	12780404	miR-4805-5p					
64	3	-	13727938	13728022	miR-4814-5p					
65	3	-	13727936	13728013	miR-4814-3p					
66	3	-	6971916	6972000	miR-5549-5p					
67	3	-	6971920	6972004	miR-5549-3p					
68	4	+	7767456	7767543	miR-1820-5p					
69	4	+	7767446	7767541	miR-1820-3p					
70	4	+	1021762	1021830	miR-2208a-5p					

71	4	+	1021748	1021843	miR-2208a-3p					
72	4	+	1026570	1026638	miR-2208b-5p					
73	4	+	1026556	1026651	miR-2208b-3p					
74	4	+	1027080	1027175	miR-2209a-3p					
75	4	+	1027208	1027287	miR-2209c-5p					
76	4	+	1027196	1027292	miR-2209c-3p					
77	4	+	10914578	10914657	miR-2210-5p					
78	4	+	10914564	10914659	miR-2210-3p					
79	4	+	501804	501872	miR-4805-5p					
80	4	+	2398662	2398739	miR-4930					
81	4	+	409773	409850	miR-5550-5p					
82	4	+	409766	409861	miR-5550-3p					
83	4	+	9757494	9757573	miR-5551-5p					
84	4	+	9757485	9757581	miR-5551-3p					
85	4	-	15012299	15012389	miR-1833					
86	4	-	6660428	6660513	miR-1834-5p					
87	4	-	6660435	6660503	miR-1834-3p					
88	4	-	17092799	17092895	miR-1832b-5p					
89	4	-	17092808	17092887	miR-1832b-3p					
90	4	-	5280040	5280136	miR-2217-5p					
91	4	-	1021281	1021376	miR-2209b-5p					
92	4	-	1021293	1021370	miR-2209b-3p					
93	4	-	212538	212634	miR-4805-5p					
94	4	-	5668913	5668987	miR-4815					
95	4	-	1381717	1381802	miR-4937					
96	5	+	5440465	5440542	miR-255-5p					
97	5	+	10649801	10649896	lsy-6					
98	5	+	232362	232452	miR-4805-5p					
99	5	+	12921811	12921899	miR-4814-5p					
100	5	+	18565065	18565144	miR-4935					
101	5	+	17940320	17940416	miR-5552-5p					
102	5	+	18039303	18039399	miR-5552-3p					
103	5	-	14367009	14367094	miR-48-5p					
104	5	-	11772656	11772730	miR-61-3p					
105	5	-	12041269	12041365	miR-87-5p					
106	5	-	12041279	12041356	miR-87-3p					
107	5	-	14368763	14368868	miR-241-5p					
108	5	-	14368782	14368859	miR-241-3p					
109	5	-	11772515	11772599	miR-250-5p					
110	5	-	11772517	11772589	miR-250-3p					
111	5	-	10541605	10541700	miR-259-5p					
112	5	-	10541618	10541700	miR-259-3p					
113	5	-	2924397	2924493	miR-2219-5p					
114	5	-	2924402	2924490	miR-2219-3p					
115	5	-	15664887	15664972	miR-4805-5p					
116	5	-	1473030	1473107	miR-5547-3p					
117	5	-	18025474	18025559	miR-5552-5p					
118	X	+	16207868	16207952	miR-1819-3p					
119	X	+	14977026	14977094	miR-1829b					
120	X	+	16207730	16207814	miR-2212-3p					
121	X	+	1873811	1873877	miR-2220-5p					
122	X	+	1873805	1873889	miR-2220-3p					
123	X	+	1390850	1390945	miR-2221					
124	X	+	15016876	15016944	miR-4805-5p					
125	X	+	1872435	1872531	miR-4807					
126	X	+	1872936	1873010	miR-4808-5p					
127	X	+	1872928	1873013	miR-4808-3p					
128	X	+	1875715	1875799	miR-4810					
129	X	+	13759159	13759238	miR-5592-5p					
130	X	+	13759147	13759243	miR-5592-3p					
131	X	-	6435368	6435464	miR-1022-5p					
132	X	-	15236267	15236352	miR-1829c					
133	X	-	1947646	1947731	miR-4805-5p					
134	X	-	7185292	7185376	miR-4811-5p					
135	X	-	7185294	7185371	miR-4811-3p					
136	X	-	15242688	15242772	miR-4812-5p					
137	X	-	15242693	15242767	miR-4812-3p					
138	X	-	8077505	8077590	miR-5553-5p					
139	X	-	8077508	8077587	miR-5553-3p					
140	X	-	13759156	13759241	miR-5592-5p					
141	X	-	13759157	13759236	miR-5592-3p					



	True Positive	False Negative	False Positive	True Negative
C. elegans	98.91304348	1.086956522	6.25	93.75
C. briggsae	96.94656489	3.053435115	5.705705706	94.29429429

Figure 2. Sensitivity of modified miRsearch presented as True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN).

Caenorhabditis (Suppl Table 1a,b). It appears that human miRNAs are closer to *C. elegans* or *C. briggsae*.

The homologoue of human miRNA h-miR-574-5p was found in different chromosomes of both *C. elegans* and *C. briggsae* at 2 mismatches. Interestingly, they were present in large numbers in overlapping regions. In case of *C. elegans* this miRNA was predicted to be present in chromosome 1 (+strand positions between, 1251563 - 1251646, 4909091 - 4909185, 9382925 - 9383045), chromosome 2 (-strand, positions between 4577004 - 4576916, and +strand from 2195721 - 2195795, 10549388 - 10549458) in chromosome 3 (+strand, positions between 445576 - 445693) and chromosome 4 (+strand, positions between 560397 - 560489). In *C. briggsae*, the same miRNA was predicted to be located in chromosome 3 (-strand positions between 1868386 - 1868483), in chromosome 5 (-strand, positions between 1011996 - 1012097 and +strand from 3757814 - 3757906, 700032 - 700138) and in chromosome X (+strand, positions between 2978466 - 2978555, 4329179 - 4329288, 16563105 - 16563213). In either case homologues of h-mir-574-5p were present in an overlapping manner due to its repeat nature of the sequence.

5. DISCUSSION

Informatics approach to identify miRNA largely involves alignment of genomes of two closely related species to find conserved regions followed by identification of stem-loop precursor transcripts capable of processing and forming ~22 nt mature RNA (2, 24). Our earlier approach used sequence alignment of mature miRNAs, the structure conservation and assessing the position of the mature miRNAs in the pre-miRNA. Starting with the known mature miRNAs from an organism as query, homologues were searched in a related organism allowing few mismatches depending on their phylogenetic distance. Some of the novel predictions of our previous algorithm miRsearch (26) has been used by other laboratory (31) and experimentally verified in Asian mosquito. However, in this approach we may miss some of the miRNAs, which are exceptionally divergent and may not be homologous at all to the available miRNAs. The above program may be accommodated to identify miRNAs in not so related organism also (as many of the miRNAs are evolutionarily conserved) by increasing the number of mismatches in miRsearch, although the chances of getting a large number of false positives will be high. Further filtering techniques will be able to reduce the false positives.

In the more advanced method presented in this study, we have determined the miRNAs of *C. elegans* and *C. briggsae* starting from a cumulative dataset from reported miRNAs with a view to make our program more versatile and widely applicable. About 98.91% of the reported *C. elegans* miRNAs could be determined with this method and the scores for false positive were quite low. Since majority of miRNAs could be considered as evolutionarily conserved, this method detects most of the miRNAs with confidence and is faster than many of the existing methods. However, the limitation of this method lies in the fact that it only detects miRNAs that are homologous to the reported ones.

Over the past few years, the complex and subtle roles of miRNAs in gene regulation have been increasingly appreciated. Most miRNA prediction algorithms combine

Table 2. Newly identified miRNAs

Sl No.	Closest Homologue	Start	End	miRNA seq in *C. elegans*	Length	Chr	Strand	Mismatch
A. *C. elegans* predicted by modified miRsearch								
1	cbr-miR-46 / crm-miR-46	13660017	13660123	UGUCAUGGAGUCGCUCUCUUCA	22	3	+	0
2	crm-miR-46*	13660035	13660112	AAGAGAGCCGUCUAUUGACAGU	22	3	+	0
3	cbr-miR-87 / crm-miR-87	12041279	12041356	GUGAGCAAAGUUUCAGGUGUGC	22	5	-	0
4	cbr-miR-48 / crm-miR-48	14367009	14367094	UGAGGUAGGCUCAGUAGAUGCGA	23	5	-	0
5	h-miR-574-5p	4909023	4909130	UGAGUGUGUGUGUGUGUGUGU	22	1	+	1
6	h-miR-513a-5p	5918255	5918357	UUCUCAGGGAGGUGUCAU	17	1	+	1
7	h-miR-574-5p	9382894	9383001	UGAGUGUGUGUGUGUGUGUGU	22	1	+	1
8	h-miR-4328	3823886	3823987	CCAGUUUUCCCAGGAUG	16	2	+	1
9	cbr-miR-47 / crm-miR-47	13660017	13660123	UGUCAUGGAGUCGCUCUCUUCA	21	3	+	1
10	h-miR-3182	2916920	2917013	GCUUCUGUACUGUAGUC	16	5	-	1
11	h-miR-3182	2916976	2917048	GCUUCUGUACUGUAGUC	16	5	-	1
12	cbr-miR-259	10541605	10541700	AAAUCUCAUCCUAAUCUGGUA	20	5	-	1
13	cbr-lsy-6	10649801	10649896	UUUUGUAUGAGACGCAUUUCG	20	5	+	1
14	crm-miR-250*	11772515	11772599	CCUUCAGUUGCCUCGUGAUCCG	21	5	-	1
15	crm-miR-87*	12041269	12041365	CGCCUGAUACUUUCGUCUCAACCU	23	5	-	1
16	cbr-miR-241	14368763	14368868	UGAGGUAGGUGCGAGAAAUGA	20	5	-	1
17	cbr-miR-241 / crm-miR-241	14368827	14368901	UGAGGUAGGUGCGAGAAAUGA	20	5	-	1
18	d-miR-2491-3p	16780497	16780590	CAGCAACAGCAGCAGCAA	17	5	-	1
19	h-miR-574-5p	3837778	3837863	UGAGUGUGUGUGUGUGUGUGU	22	X	+	1
20	h-miR-3201	5099085	5099173	GGGAUAUGAAGGAAAAU	16	X	+	1
21	h-miR-1279	10167840	10167933	UCAUAUUGAUUCUUUCU	16	X	+	1
22	h-miR-4516	12179263	12179348	GGGAGAAGGGUCGGCGC	16	X	+	1
23	h-miR-1281	12440435	12440520	UCGUCUCCUCCUCUCCC	16	X	-	1
24	h-miR-574-5p	14494971	14495056	UGAGUGUGUGUGUGUGUGUGU	22	X	+	1
B. C. briggsae predicted by modified miRsearch								
1	cel-miR-244-5p / crm-miR-244	2688441	2688511	UCUUUGGUUGUACAAAGUGGUAUG	24	1	+	0
2	cel-miR-245-3p / crm-miR-245	6757250	6757327	AUUGGUCCCCUCCAAGUAGCUC	22	1	-	0
3	cel-miR-236-5p	5133850	5133946	CGUCUUACCUGUUCAAUAUUUAGA	24	2	-	0
4	crm-miR-355	6505553	6505637	UUUGUUUUAGCCUGAGCUAUGU	22	2	-	0
5	cel-miR-45-5p / crm-miR-44*/ crm-miR-45*	8058555	8058650	CUGGAUGUGCUCGUUAGUCAU	21	2	-	0
6	crm-miR-44 / crm-miR-45	8058564	8058648	UGACUAGAGACACAUUCAGCU	21	2	-	0
7	cel-miR-5552-5p	13592464	13592560	UGUAGUUUGUAGUCUAGCAGACC	23	2	-	0
8	cel-miR-5552-5p	13592475	13592560	UGUAGUUUGUAGUCUAGCAGACC	23	2	-	0
9	cel-miR-5552-5p	13639621	13639717	UGUAGUUUGUAGUCUAGCAGACC	23	2	-	0
10	cel-miR-5552-5p	13684530	13684637	UGUAGUUUGUAGUCUAGCAGACC	23	2	-	0
11	cel-miR-5552-5p	13684749	13684856	UGUAGUUUGUAGUCUAGCAGACC	23	2	-	0
12	cel-miR-5552-5p	13684968	13685075	UGUAGUUUGUAGUCUAGCAGACC	23	2	-	0
13	cel-miR-5552-5p	13685114	13685221	UGUAGUUUGUAGUCUAGCAGACC	23	2	-	0
14	cel-miR-5552-5p	13926293	13926400	UGUAGUUUGUAGUCUAGCAGACC	23	2	-	0
15	cel-miR-5552-5p	13933723	13933802	UGUAGUUUGUAGUCUAGCAGACC	23	2	+	0
16	cel-miR-5552-5p	1133919	1134026	UGUAGUUUGUAGUCUAGCAGACC	23	3	-	0
17	crm-miR-231	9825072	9825157	UAAGCUCGUGAACAACAGGCAGG	23	3	+	0
18	cel-miR-90-3p / crm-miR-90	13421059	13421155	UGAUAUGUUGUUUGAAUGCCCCU	23	3	+	0
19	cel-miR-790-5p / crm-miR-790	8548032	8548116	CUUGGCACUCGCGAACACCGCG	22	4	-	0
20	cel-miR-790-5p / crm-miR-790	8548642	8548737	CUUGGCACUCGCGAACACCGCG	22	4	-	0
21	cel-miR-228-5p / crm-miR-228	8555781	8555866	AAUGGCACUGCAUGAAUUCACGG	23	4	-	0
22	cel-miR-250-5p	6755122	6755206	CCUUCAGUUGCCUCGUGAUCCG	22	5	-	0
23	cel-miR-255-3p	8908227	8908311	AAACUGAAGAGAUUUUUUACAG	22	5	+	0
24	cel-miR-253-3p	9138951	9139039	UUAGUAGGCGUUGUGGGAAGGG	22	5	-	0
25	crm-miR-253	9138953	9139035	UUAGUAGGCGUUGUGGGAAG	20	5	-	0
26	cel-miR-233-3p / crm-miR-233	1703778	1703873	UUGAGCAAUGCGCAUGUGCGG	21	X	+	0
27	cel-miR-251	5851564	5851660	UUAAGUAGUGGUGCCGCUCUUAUU	24	X	-	0
28	cel-miR-360-3p	11095215	11095300	UGACCGUAAUCCCGUUCACAA	21	X	+	0
29	h-miR-4297	5007413	5007513	UGCCUUCCUUUCUGUG	15	1	+	1
30	h-miR-1281	2209676	2209745	UCGCCUCCUCCUCUCAC	16	2	+	1
31	h-miR-4279	3598845	3598945	CUCUCCUCACGGCUUC	15	2	-	1
32	h-miR-574-5p	5603535	5603620	UGUGUGUGUGUGUGUGAGUGUGU	22	2	-	1
33	h-miR-4297	5735906	5736006	UGCCUUCCUUUCUGUG	15	2	+	1
34	cel-miR-44-5p	8058555	8058650	CUGGAUGUGCUCGUUAGUCAUA	21	2	-	1
35	h-miR-4297	8699597	8699697	UGCCUUCCUUUCUGUG	15	2	+	1
36	h-miR-4468	11642237	11642330	AGAGCAGAAGAAUGAGAU	17	2	-	1
37	cel-miR-5552-5p	13445343	13445450	UGUAGUUUGUAGUCUAGCAGGCC	22	2	-	1
38	cel-miR-5552-5p	13596952	13597048	UGCAGUUUGUAGUCUAGCAGACC	22	2	-	1
39	cel-miR-5552-5p	13608035	13608142	UGUAGCUUGUAGUCUAGCAGACC	22	2	-	1
40	cel-miR-5552-5p	13620971	13621067	UGUAGUUUGUAGUCUAGCAUACC	22	2	-	1
41	cel-miR-5552-5p	13638477	13638584	UGUAGUUUGUUGUCUAGCAGACC	22	2	-	1
42	cel-miR-5552-5p	13679123	13679230	UGUAGUUUGUAGUCUAGCAGGCC	22	2	-	1
43	cel-miR-5552-5p	13679279	13679375	UGUAGUUUGUAGUCUAGCAGGCC	22	2	-	1
44	cel-miR-5552-5p	13679757	13679864	UGUAGUUUGUAGUCUAGCAGGCC	22	2	-	1
45	cel-miR-5552-5p	13693010	13693117	UGUAGUUUGUAGUCUAGCAGGCC	22	2	-	1
46	cel-miR-5552-5p	13693147	13693254	UGUAGUUUGUAGUCUAGCAGGCC	22	2	-	1
47	cel-miR-5552-5p	13920072	13920162	UGUAGUUUGUAGUUUAGCAGACC	22	2	+	1

48	cel-miR-5552-5p	13932463	13932570	UGUAGUUUGUAGUCUAGCAGGCC	22	2	-	1
49	cel-miR-5552-5p	13943413	13943503	UGUAGUUUGUAGUCUAGCAGAUC	22	2	+	1
50	cel-miR-5552-5p	1135561	1135668	CGUAGUUUGUAGUCUAGCAGACC	22	3	-	1
51	cel-miR-5552-5p	1141790	1141880	UGUAGUUUGUAGUCUAGCAGGCC	22	3	+	1
52	h-miR-4297	6114642	6114742	UGCCUUCCUUUCUGUG	15	3	+	1
53	cel-miR-2214-5p	8971626	8971711	ACUCGGUCCGGAGUCAAUGGG	20	3	-	1
54	cel-miR-2214-5p	9058999	9059084	ACUCGGUCCGGAGUCAAUGGG	20	3	-	1
55	cel-miR-356b-3p	10353552	10353648	UUUGUUCGCGUUGCUCAUCCACA	22	3	+	1
56	h-miR-4455	13365800	13365901	AGGGUGUGUGUGUGUUU	16	3	+	1
57	h-miR-4455	13374505	13374606	AGGGUGUGUGUGUGUUU	16	3	+	1
58	cel-miR-90-5p	13421070	13421149	CGGCUUUCAACGACUAUAUCAAC	22	3	+	1
59	h-miR-3201	4613951	4614031	GGGAUAUGAAAAAAAAU	16	4	-	1
60	h-miR-4481	9568536	9568629	GGAAUGGGCUGGUGGUU	16	4	+	1
61	h-miR-186-3p	14396975	14397081	GUCCAAAGGUGAAUUUUUUGGG	21	4	+	1
62	r-miR-186*	14396995	14397080	GUCCAAAGGUGAAUUUUUUGG	20	4	+	1
63	crm-miR-250*	6755122	6755206	CCUUCAGUUGCCUCGUGAUCCG	21	5	-	1
64	cel-lsy-6	7813060	7813144	UUUUGUAUGAGACGCAUUCCG	20	5	-	1
65	cel-miR-259-5p	7917249	7917337	AAAUCUCAUCCUAAUCUGGUUG	21	5	+	1
66	cel-miR-255-5p	8908232	8908309	GUAAGAAAUCUUUGUAGUUUUC	21	5	+	1
67	cel-miR-241-5p	12272735	12272809	UGAGGUAGGUGUGAGAAAUGA	20	5	+	1
68	h-miR-1281	1645451	1645552	UCGCCUCCUCCUGUCCC	16	X	-	1
69	cel-miR-230-3p	10968454	10968550	GUAUUAGUUGUGCGACCAGGAAA	22	X	+	1
70	h-miR-4499	11072267	11072360	AAGAGUGAGAGGAGGGA	16	X	+	1
71	cel-miR-239b-5p	15174132	15174227	UUUGUACUGCACAAAAGUACUG	21	X	-	1

Table 3. Summary Statistics for overall quality of prediction

Nematode	Precision	Recall	F-Score
C. elegans	0.983783784	0.989130435	0.986449864
C. briggsae	0.869863014	0.969465649	0.916967509

information on sequence, structure, and conservation and predict different numbers of candidate miRNA genes, few of which have been experimentally validated. Possible explanations could be that these represent false-positives or the gene is not simply expressed in the RNA sample examined. These algorithms so far have not been equipped with the predictions on the orientation of the transcript (plus or minus strand) with respect to genomic location, the position of the processing sites within the hairpin structure, and the determination of which of the paired segments of the hairpin will constitute the mature miRNA. Despite such uncertainties, *in silico* prediction methods for miRNAs have already become a valuable tool. Sensitive biological validation techniques are key factors in fine tuning informatics prediction algorithms. And yet, developing such biological techniques often depends on effective prediction algorithms. An integrated detection approach, which combines computational prediction together with high-throughput biological validation, has been most effective in discovery of miRNAs. The present algorithm of modified miRsearch along with the filtering processes makes this a very fast and efficient method for identifying miRNAs. The method not only accurately identified >95% of the annotated miRNAs of both *C. elegans* and *C. briggsae* in their respective chromosomes, but also provides the option of identifying distantly related miRNAs by allowing mismatches. The stringent filtering process is crucial in screening out the miRNAs and reducing the number of false positive predictions. The method identified a number of new miRNAs in the chromosomes *C. elegans* and *C. briggsae* having both *Caenorhabditis* and non *Caenorhabditis* origins. The identification of the locations of the newly identified miRNAs will help in the experimental validation of these miRNAs in a much faster and precise way. Moreover with the availability of Next Generation Sequencing data this algorithm can be used to detect potential miRNAs from them making this an invaluable tool for miRNA prediction. Now we know that the regulation of gene expression by miRNA is a widespread natural phenomenon regulating complex genetic pathways, and these miRNAs are modulated in many human diseases. Understanding the miRNA-guided network has enormous possibility of providing a new window for diagnostics and therapy of many human diseases. Many challenges remain in understanding miRNAs and dissecting the affected pathways. Currently we are working on the development of an algorithm for the prediction of the target for the miRNAs, which will be integrated with this present algorithm. Integrative approaches with crosstalk between *in silico* and experimental methods will continue to push forward future developments in this exciting field.

6. ACKNOWLEDGEMENTS

Avirup Dutta and Raghunath Chatterjee have equally contributed to this manuscript. The study was supported by the Council of Scientific and Industrial Research (CSIR), Govt. of India. A.D. is the recipient of the CSIR Senior Research Fellowship.

7. REFERENCES

1. B. Bartel and D. P. Bartel: MicroRNAs: at the root of plant development? *Plant Physiol*, 132(2), 709-17 (2003)

2. L. P. Lim, M. E. Glasner, S. Yekta, C. B. Burge and D. P. Bartel: Vertebrate microRNA genes. *Science*, 299(5612), 1540 (2003)

3. S. Pfeffer, M. Zavolan, F. A. Grasser, M. Chien, J. J. Russo, J. Ju, B. John, A. J. Enright, D. Marks, C. Sander and T. Tuschl: Identification of virus-encoded microRNAs. *Science*, 304(5671), 734-6 (2004)

4. S. Pfeffer, A. Sewer, M. Lagos-Quintana, R. Sheridan, C. Sander, F. A. Grasser, L. F. van Dyk, C. K. Ho, S.

Shuman, M. Chien, J. J. Russo, J. Ju, G. Randall, B. D. Lindenbach, C. M. Rice, V. Simon, D. D. Ho, M. Zavolan and T. Tuschl: Identification of microRNAs of the herpesvirus family. *Nat Methods*, 2(4), 269-76 (2005)

5. D. P. Bartel: MicroRNAs: target recognition and regulatory functions. *Cell*, 136(2), 215-33 (2009)

6. R. W. Carthew and E. J. Sontheimer: Origins and Mechanisms of miRNAs and siRNAs. *Cell*, 136(4), 642-55 (2009)

7. J. Krol, I. Loedige and W. Filipowicz: The widespread regulation of microRNA biogenesis, function and decay. *Nat Rev Genet*, 11(9), 597-610 (2010)

8. R. C. Lee, R. L. Feinbaum and V. Ambros: The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell, 75(5), 843-54 (1993)

9. B. J. Reinhart, F. J. Slack, M. Basson, A. E. Pasquinelli, J. C. Bettinger, A. E. Rougvie, H. R. Horvitz and G. Ruvkun: The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature, 403(6772), 901-6 (2000)

10. E. G. Moss, R. C. Lee and V. Ambros: The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA. Cell, 88(5), 637-46 (1997)

11. F. J. Slack, M. Basson, Z. Liu, V. Ambros, H. R. Horvitz and G. Ruvkun: The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. Mol Cell, 5(4), 659-69 (2000)

12. B. P. Lewis, I. H. Shih, M. W. Jones-Rhoades, D. P. Bartel and C. B. Burge: Prediction of mammalian microRNA targets. Cell, 115(7), 787-98 (2003)

13. M. Lagos-Quintana, R. Rauhut, W. Lendeckel and T. Tuschl: Identification of novel genes coding for small expressed RNAs. Science, 294(5543), 853-8 (2001)

14. R. C. Lee and V. Ambros: An extensive class of small RNAs in Caenorhabditis elegans. Science, 294(5543), 862-4 (2001)

15. N. C. Lau, L. P. Lim, E. G. Weinstein and D. P. Bartel: An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science, 294(5543), 858-62 (2001)

16. E. G. Moss and R. S. Poethig: MicroRNAs: something new under the sun. Curr Biol, 12(20), R688-90 (2002)

17. D. P. Bartel: MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, 116(2), 281-97 (2004)

18. J. Brennecke, D. R. Hipfner, A. Stark, R. B. Russell and S. M. Cohen: bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. *Cell*, 113(1), 25-36 (2003)

19. V. Ambros: The functions of animal microRNAs. *Nature*, 431(7006), 350-5 (2004)

20. B. D. Harfe: MicroRNAs in vertebrate development. *Curr Opin Genet Dev*, 15(4), 410-5 (2005)

21. G. A. Calin, C. Sevignani, C. D. Dumitru, T. Hyslop, E. Noch, S. Yendamuri, M. Shimizu, S. Rattan, F. Bullrich, M. Negrini and C. M. Croce: Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci U S A*, 101(9), 2999-3004 (2004)

22. J. Dostie, Z. Mourelatos, M. Yang, A. Sharma and G. Dreyfuss: Numerous microRNPs in neuronal cells containing novel microRNAs. *Rna*, 9(2), 180-6 (2003)

23. J. Meister and M. H. Schmidt: miR-126 and miR-126*: new players in cancer. *ScientificWorldJournal*, 10, 2090-100 (2010)

24. E. C. Lai, P. Tomancak, R. W. Williams and G. M. Rubin: Computational identification of Drosophila microRNA genes. *Genome Biol*, 4(7), R42 (2003)

25. L. P. Lim, N. C. Lau, E. G. Weinstein, A. Abdelhakim, S. Yekta, M. W. Rhoades, C. B. Burge and D. P. Bartel: The microRNAs of Caenorhabditis elegans. *Genes Dev*, 17(8), 991-1008 (2003)

26. R. Chatterjee and K. Chaudhuri: An approach for the identification of microRNA with an application to Anopheles gambiae. *Acta Biochim Pol*, 53(2), 303-9 (2006)

27. D. S. Hirschberg: A linear space algorithm for computing maximal common subsequences. *CACM*, 18(6), 431-343 (1975)

28. S. Griffiths-Jones: The microRNA Registry. *Nucleic Acids Res*, 32(Database issue), D109-11 (2004)

29. S. Griffiths-Jones, R. J. Grocock, S. van Dongen, A. Bateman and A. J. Enright: miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res*, 34(Database issue), D140-4 (2006)

30. I. L. Hofacker: Vienna RNA secondary structure server. *Nucleic Acids Res*, 31(13), 3429-31 (2003)

31. E. A. Mead and Z. Tu: Cloning, characterization, and expression of microRNAs from the Asian malaria mosquito, Anopheles stephensi. *BMC Genomics*, 9, 244 (2008)

Key Words : miRNA, *C. elegans and C. briggasae* miRNA, Modified Mirsearch

Send correspondence to: Keya Chaudhuri, Chief Scientist, Molecular and Human Genetics Division, CSIR-Indian Institute of Chemical Biology, 4, Raja S. C. Mullick Road, Kolkata-700 032, India,Tel: 91-33-2499-5762, Fax: 91-33-2473-5197, E-mail: keya.chaudhuri@gmail.com