J Vector Borne Dis 52, March 2015, pp. 11–22

Research Articles

Computational approach for elucidating interactions of cross-species miRNAs and their targets in *Flaviviruses*

Santosh P. Shinde[1], Amit Kumar Banerjee[2], Neelima Arora[1], U.S.N. Murty[2], Venkateswara Rao Sripathi[1], Manika Pal-Bhadra[2] & Utpal Bhadra[1]

[1]*Functional Genomics and Gene Silencing Group, Centre for Cellular and Molecular Biology, Hyderabad;* [2]*Biology Division and Center for Chemical Biology, Indian Institute of Chemical Technology, Tarnaka, Hyderabad, India*

ABSTRACT

Background & objectives: Combating viral diseases has been a challenging task since time immemorial. Available molecular approaches are limited and not much effective for this daunting task. MicroRNA based therapies have shown promise in recent times. MicroRNAs are tiny non-coding RNAs that regulate translational repression of target mRNA in highly specific manner.

Methods: In this study, we have determined the target regions for human and viral microRNAs in the conserved genomic regions of selected viruses of *Flaviviridae* family using miRanda and performed a comparative target selectivity analysis among them.

Results: Specific target regions were determined and they were compared extensively among themselves by exploring their position to determine the vicinity. Based on the multiplicity and cooperativity analysis, interaction maps were developed manually to represent the interactions between top-ranking miRNAs and genomes of the viruses considered in this study. Self-organizing map (SOM) was used to cluster the best-ranked microRNAs based on the vital physicochemical properties.

Interpretation & conclusion: This study will provide deep insight into the interrelation of the viral and human microRNAs interactions with the selected *Flaviviridae* genomes and will help to identify cross-species microRNA targets on the viral genome.

Key words Bioinformatics; *Flaviviridae*; interaction map; miRanda; miRNA

INTRODUCTION

Diseases caused by the members of *Flaviviridae* cast a gloomy shadow by affecting millions across the globe and this spawn a vicious cycle of disease and poverty. Resurgence, reemergence and the spread of distribution of these viruses, which were once perceived to be confined to specific geographic locations, have raised much hue and cry. Diseases caused by *Flaviviruses* constitute a major share of mosquito-borne diseases[1]. Despite extensive research on viral biology, transmission dynamics and pathogenicity, we are miles away from success in the crusade against these deadly diseases. Among these mosquito-borne diseases, recurrent outbreaks of viral diseases such as Japanese encephalitis, dengue and dengue hemorrhagic fever (DHF) have been reported in Southeast Asian countries such as India, Pakistan and Thailand in the recent past[2]. Measures adopted to control these diseases such as spatio-temporal mapping of the effective zones[3], development of vaccines for an individual pathogen[4], analysis of apoptotic cell death in response to dengue virus infection[5] and etiological diagnosis[6] proved partially effective. Though, these methodologies are scientifically sound yet they are unable to provide an eternal solution. Moreover, these efforts are restricted to a specific pathogen; studies on group-specific pathogen targeting are very limited due to the complexity involved and considered cumbersome so far.

Family *Flaviviridae* is classified into three genera Flavivirus, Pestivirus, Hepacivirus and the unclassified group of GB virus C/hepatitis G viruses (GBV-C) comprising of 70 important pathogens of significant impact on global health. These viruses possess a single-stranded positive sense RNA (ss+RNA) genome of 9.6–12.3 kb comprising of one single long open reading frame flanked by 5′ and 3′ untranslated regions (UTR)[1–2, 7–8]. Keeping in mind, the fast evolution and increase in number of pathogens, there is a need for finding a common solution against these viruses. Irrespective of rapid evolution, the members of *Flaviviridae* share quite good amount of con-

servation and similarity in their genetic make-up[1, 8]. Different molecular, biochemical and pathophysiological aspects of these tiny culprits have been studied extensively in the past decades. The effect of dengue epidemic in relation to nutritional status was assessed extensively[9].

It is well established that the striving success of viruses relies on their ability to exploit cellular machinery of the host efficiently to their advantage for their survival and replication but this also makes them prone to gene regulatory mechanisms of the host often mediated by microRNAs. MicroRNAs are small non-coding RNAs involved in regulation of gene expression at the post-transcriptional level[10].

Recent emergence of data on microRNA mediated gene regulation in viral infections has brought cellular miRNAs as agents for targeting viruses to limelight[11]. While the members of *Flaviviridae* have been the objects of various studies, very little is known about microRNA mediated regulation in them. Ample theoretical and experimental evidences suggest that this reciprocity in host defense and viral offense has led to a co-evolutionary conflict in which microRNAs are the major players[12].

The knowledge that viruses encode miRNAs for autoregulation of their own gene expression as well as for defense from the host cell miRNAs has incited countless studies[13]. In the same fashion, viral genome can be the target of either specific miRNA or large number of miRNAs owing to fortuitous complementarities[14]. The renewed interest in identification of potential microRNA candidates using *in-silico* and experimental methods has witnessed an unparalleled progress for understanding viral epidemiology, once termed as imperfect science. High mutability confers evolutionary advantage to viruses and low homology makes the prediction of miRNAs in viruses a challenging task. Location of binding sites can unravel a lot about the mechanism of miRNA and their role in the regulatory pathways of viral diseases. With the unprecedented development of simple, rapid and accurate computer based methods and efficient algorithms for microRNA prediction; the count of such studies is all set to mount. Conventional biochemical profiling approaches suffer from transient and low level of microRNA expression, tissue specificity and complex interaction with targets[15]. Computational approaches had been used to identify potential miRNA targets for microRNA from Epstein-Barr virus (EBV)[16].

Screening microRNAs is a cumbersome task, which can be very specifically and quite accurately done using computational methodologies. Recently, computational approaches have been extensively developed and effectively utilized in searching target sites for microRNAs[10]. Targeting various regions of viral genomes with microRNAs can be an effective solution against these pathogens as microRNAs are very specific in selecting target regions. Recent success of RNAi technology in combating viral diseases has motivated us to explore the role of cellular and viral microRNA in host-parasite interactions in *Flaviviridae*.

In this study, we describe the common viral genomic target regions for viral and human microRNAs with significant sensitivity and specificity. Further, an interaction network interconnecting microRNA and genome was developed for human and viral microRNA targets considering the top ranked hits screened for all the 15 viral genomes. Self organizing map (SOM) is an unsupervised machine-learning method used to cluster and visualize high-dimensional data[17]. As physicochemical properties like composition and GC% provide an insight about the mechanism of binding and stability of duplex, they were computed for unique non-repetitive miRNAs top hits and the data were clustered using SOM to understand the inherent patterns.

MATERIAL & METHODS

This study aims at finding both viral and human microRNA targets in dengue virus serotypes and comparing the top ranking microRNAs with other miRNA hits for the selected genomic regions of members of *Flaviviridae*. The workflow is depicted in Fig. 1 in detail.

Genome sequences and extraction of target regions

Sequence retrieval: Complete genome sequences of 15 viruses of family *Flaviviridae* were collected in Fasta format from National Center for Biotechnology Information (NCBI). Dengue virus serotype 1, 2, 3, 4; West Nile virus (lineage I strain NY99), West Nile virus (lineage II strain 956), Japanese encephalitis virus, tick-borne encephalitis virus (TBE) and yellow fever virus belong to genus *Flavivirus* while Hepatitis C virus genotype 1, 2, 3, 4, 5, 6 belong to Hepacivirus. All the viruses selected from the *Flavivirus* genus are mosquito-borne viruses except TBE.

Identification of the conserved regions in dengue virus genomes: To identify the conserved regions in the viral genomes, a multiple sequence alignment (MSA) approach was adopted. It was performed in two steps (Fig. 1). Initially, conservation was analyzed in four dengue serotypes and later on, it was performed for the other viral genomes considered for the study. The sequence length ranged from 10,649 nt (Dengue serotype 4) to 10,735 nt (Dengue serotype 1). The sequences were sub-

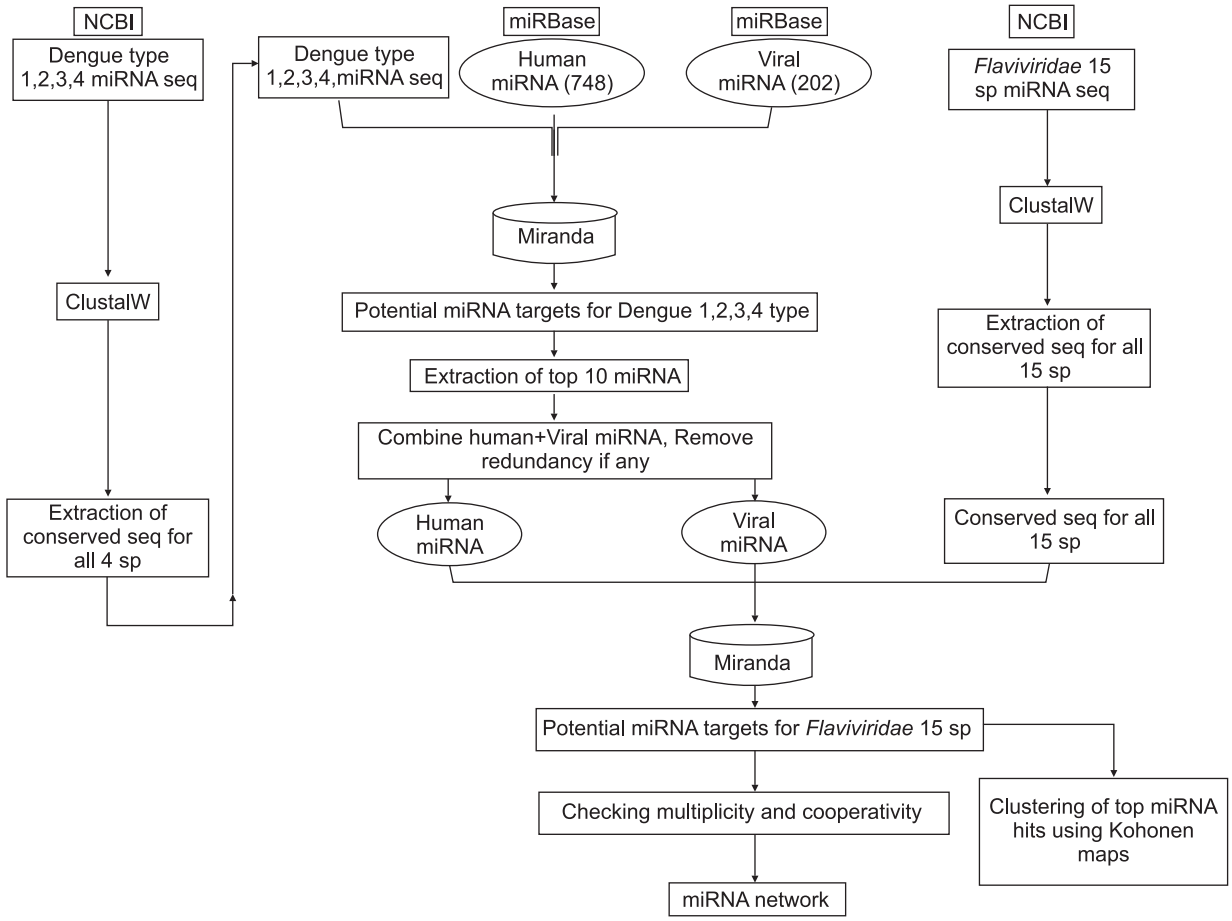

Fig. 1: Schematic representation of the workflow.

jected to multiple sequence alignment (MSA) using ClustalW with default gap opening and extension penalty and scoring matrix[18]. The ClustalW was used owing to its reliability in aligning sequences of similar length. The alignment exercise revealed the presence of two major conserved regions. Out of these two, the longer conserved region obtained was selected to carry out further analysis (data not shown). Apart from these four target genomes, 11 genomes, namely West Nile viruses (1 and 2), Japanese encephalitis virus, yellow fever virus, tick-borne encephalitis virus and hepatitis C viruses (1 to 6) were also considered for this study. All these viral genomes were aligned separately and the conserved region was identified and later on subjected to microRNA target prediction.

MicroRNA and their sequences: In microRNA biogenesis of mammals, the primary miRNA (pri-miRNA) is transcribed mainly by RNA polymerase II and sequentially processed by two RNase III enzymes—Drosha and Dicer to generate mature 21–23 nt double-stranded endogenous miRNAs[19–21]. The pathway of miRNA activity continues with the RISC formation followed by mRNA degradation, which includes many intermediate steps[19].

Several databases are available for cataloguing and depositing miRNA sequences. miRBase is an online repository which provides integrated interfaces to comprehensive microRNA sequence data, annotation and predicted gene targets[22]. Both viral (202) and human (747) mature microRNA sequences were downloaded from miRBase (*http://microrna.sanger.ac.uk/sequences*) in FASTA format and used as input query sequences for target prediction.

Selection of tool for miRNA target prediction: Although several tools are available for predicting miRNA target sites, miRanda offers efficiency and accuracy owing to high sensitivity and specificity, which results in elimination of false positive and true negative hits. In addition, this tool offers flexibility of applying user defined parameters[23–24]. Considering these advantages, Java based version of miRanda (v1.0b) [23–24] was selected for predicting miRNA targets in the conserved genomic regions of the selected 15 *Flaviviridae* genomes. For prediction of miRNA targets, miRanda (version 1.0b)

employs dynamic programming based on sequence complementarity, allotting higher weights to matches at the 5′ end of the mature miRNA while considering the free energy of the RNA-RNA duplex (calculated using RNA Fold)[25] and the extent of conservation of the miRNA target across related genomes. The parameters considered for this study were: Gap open penalty = –2, Gap extend penalty = –8, Score threshold = 140, Energy threshold = –20 kcal/mol, and Scaling parameter = 3, more or less similar to our previous studies[26–27].

Analysis of multiplicity and cooperativity: Multiplicity and cooperativity are of paramount importance for understanding the microRNA targets and patterns of miRNA based regulation of gene expression. miRNA can regulate target regions in two ways. One miRNA can control several genes or target regions, a phenomenon known as multiplicity or several microRNAs can collectively control the expression of a single target gene region, i.e. cooperativity[31]. Therefore, it is very important to look for such important issues for assessing the mode of gene expression regulation.

Construction of miRNA-genome interaction map: For all the 15 targeted genomic regions, top 10 ranking potent miRNAs were predicted and their pattern of interactions was analyzed. An interaction map of microRNA-genomic target region was constructed with the 10 best hits found for each target conserved genomic regions. Two separate networks were constructed for viral miRNA sequences and human sequences.

Conserved nucleotide pattern in top ranked miRNA: Conservation was checked for the top ranked viral and human microRNAs. Specific nucleotide conservation patterns were analyzed for the top viral and human miRNAs obtained which may aid in specific targeting of the considered viruses or probably for the entire *Flaviviridae* family.

Mining physicochemical properties for miRNA: It is important to understand the interrelation and interdependency of the parameters influencing the microRNA and target genome interaction. For this purpose, we employed advanced data mining technology. An assembled dataset of all top ten miRNAs found for each target genome was prepared and the repeating microRNAs were excluded to eliminate redundancy. Physicochemical properties of this unique set of top ranked microRNAs were calculated. Important physicochemical properties such as molecular weight (kilo Dalton), melting temperature (Tm), free energy (δG in kcal/mol) and composition of the nucleotides were calculated by using Oligo Calc: Oligonucleotide properties calculator. Molecular weight, melting temperature and free energy were calculated using the following formulae:

$$\textit{Molecular weight}\ (\mathrm{Mw}) = (\mathrm{A_n} \times 329.21) + (\mathrm{U_n} \times 306.17) + (\mathrm{C_n} \times 305.18) + (\mathrm{G_n} \times 329.21) + 159 \quad \ldots (1)$$

Where, A_n, U_n, C_n and G_n are the number of each respective nucleotide present in the RNA molecule under consideration. Additionally, weight 159 g/mole is added that accounts to 5′ triphosphate.

$$\textit{Melting temperature}\ (\mathrm{Tm}) = 64.9 + 41 \times (\mathrm{yG} + \mathrm{Zc} - 16.14)/(\mathrm{wA} + x\mathrm{U} + \mathrm{yG} + \mathrm{zC}) \quad \ldots(2)$$

Where, w, *x*, y and z are the number of the bases A, U, G, C in the sequence respectively.

$$\textit{Free energy}\ (\delta\mathrm{G}) = \mathrm{RT}\ \ln\ [(\mathrm{RNA.template})/(\mathrm{RNA})(\mathrm{template})] \quad \ldots(3)$$

Both equations 2 and 3 assume that the annealing occurs under the standard conditions of 50 nM primer, 50 mM Na+, and pH 7. Later on, detailed analysis was performed for gaining a comprehensive understanding on the nature of clustering based on these properties.

Self organizing maps (SOM) algorithm was used for clustering the properties calculated for the miRNAs. SOM has numerous applications in biological data clustering[28–29]. By applying SOM, we can obtain a map of input space where closeness of input vectors is reflected as proximity of units or clusters. Important feature that makes SOM a popular choice for analyzing biological data set is its ability to assign precise location to input vectors based on the association between them and also ensuring comparable neighborhood cell in lattice. Apart from this, ease of visualization of huge datasets makes the SOM map a useful tool.

Steps involved in the algorithm

(1) *Initialization:* A weight vector (Wi) is initialized for each neuron I randomly.

Wi = [wi1; wi2; . . . ; wi n]; where, *n* is the dimension of input data.

(2) *Sampling:* An input vector X = [*x1, x2, . . . , xn*] is selected.

(3) *Similarity matching:* A winning neuron is found whose weight vector best matches with the input vector

$j(t) = \arg\min \{\|X - Wi\|\}$

(4) *Updating:* Weight vector of winning neuron is updated, such that it becomes still closer to the input vector. Weight vectors of neighbouring neurons are also updated—the farther the neighbour, the lesser the degree of change.

$Wi\ (t+1) = Wi(t) + \alpha(t)\ X\ hij(t) X\ [X(t) - Wi(t)]$

Where, $\alpha\ (t)$: Learning rate that decreases with time t, $0 < \alpha\ (t) \leq 1$.

$hij\ (t) = \exp(-\| rj - ri \|\ 2/2 \times \sigma(t)2)$

Where, $\|rj\text{-}ri\|2$ = Distance between winning neuron and other neurons, $\sigma\ (t)$ = Neighbourhood radius that decreases with time *t*.

(5) *Continuation:* Steps 2–4 are repeated until there is no change in weight vectors or up to certain number of iterations. For each input vector, best matching weight vector is selected and the input vector is allotted to the corresponding neuron/cluster.

Data normalization: Different properties of microRNAs vary in terms of numerical values and units. This necessitates normalization of the data for achieving homogeneity for the heterogeneous dataset. Whole dataset was normalized and the obtained normalized input data set was used for clustering. Data were normalized linearly such that the data point values in each category ranged between 0 and 1. This ensured unbiased results while assigning equal importance to all parameters during clustering. We adopted the standard normalization formula used in our previous studies[30–31].

Following formula was applied for normalization:

$$N = (O - m)/(M - m)$$

Where, N is normalized data, O is the original data value, m is the minimum data value, and M is the maximum data value.

RESULTS & DISCUSSION

Extraction of conserved region in dengue virus genomes

All four dengue virus complete genomes (dengue serotype 1, 2, 3 and 4) were considered and the conserved sequence stretch was identified for further miRNA target prediction. In this alignment exercise, two major conserved regions were found and the longer stretch was considered for further study. The regions selected for four dengue serotypes are 1–2419, 96–2421, 1–2413, 1–2423 nucleotides respectively (data not shown).

Target identification for dengue virus genomes

The miRNA targets were identified in dengue virus serotypes 1, 2, 3 and 4 employing miRanda software for the conserved region. Target regions for 747 human and 202 viral miRNAs were predicted for both conserved regions. Top 10 hits were identified for all the four conserved dengue virus genome segments using both human and viral miRNAs. The viral miRNA top 10 target scores obtained are as follows in descending order, dengue serotype 1: 185 to 171, dengue serotype 2: 191 to 172, dengue serotype 3: 188 to 174 and dengue serotype 4: 189 to 170. Similarly, in case of human top 10 miRNA scores observed are dengue serotype 1: 196 to 177, dengue serotype 2: 184 to 175, dengue serotype 3: 196 to 180, and dengue serotype 4: 194 to 178.

Conservation pattern of viral and human miRNAs in dengue virus genomes

The alignment of the genomes revealed that dengue serotypes 1 and 3 were more conserved whereas dengue serotypes 2 and 4 showed more proximity to each other. Interestingly, similar conservation pattern was observed in the miRNA target regions also. We found good conservation in dengue serotypes 1 and 3 in miRNA binding regions where eight viral miRNAs showed same binding sites which may prove to be likely candidates for targeting both the genomes. These are hcmv-miR-UL36*, hiv1-miR-TAR-3p, kshv-miR-K12-10b, mdv1-miR-M6, mdv2-miR-M21, mcmv-miR-m01-4, mcmv-miR-m59-2 and mghv-miR-M1-7-5p miRNA (Table 1).

On the contrary, dengue serotypes 2 and 4 showed less conservation in binding sites as compared to type 1 and 3 serotypes, suggesting that they have more divergence in miRNA target regions. Similar results were obtained in case of human miRNA target analysis. Good conservation was found in dengue serotypes 1 and 3 in miRNA binding regions where 15 human miRNA showed same binding sites. These are hsa-miR-490-5p, hsa-miR-1296, hsa-miR-939, hsa-miR-608, hsa-miR-770-5p, hsa-miR-624*, hsa-miR-520d-3p, hsa-miR-593*, hsa-miR-431, hsa-miR-486-5p, hsa-miR-675, hsa-miR-188-5p, hsa-miR-433, hsa-miR-154 and hsa-miR-1234 miRNA

Table 1. Conserved viral miRNA binding regions in Dengue serotypes 1 and 3

S.No.	Viral miRNA	miRNA binding regions in dengue serotype 1	miRNA binding regions in dengue serotype 2
1.	hcmv-miR-UL36*	769–791	766–791
2.	hiv1-miR-TAR-3p	169–191	169–191
3.	kshv-miR-K12-10b	1361–1384	1361–1384
4.	mdv1-miR-M6	1714–1741	1708–1735
5.	mdv2-miR-M21	958–979	958–979
6.	mcmv-miR-m01-4	613–637	613–637
7.	mcmv-miR-m59-2	148–171	148–171
8.	mghv-miR-M1-7-5p	17–39	17–39

(Table 2). As compared to dengue serotypes 1 and 3 (Fig. 2b), dengue serotypes 2 and 4 do not demonstrate any conservation in binding sites.

Derived information from the obtained output revealed that majority of miRNAs target regions are spread over the NS5 encoding region in dengue virus genome serotype 1 (Fig. 2a). Multiple binding sites were observed for few human miRNAs in conserved region of dengue virus genomes. Two binding sites were identified for hsa-miR-187 in dengue serotype 1 and hsa-miR-1183 in dengue serotype 2. Two binding sites were found for hsa-miR-612 and hsa-miR-637 in NS5 region of dengue serotype 4. hsa-miR-612 targets NS5 region in dengue serotypes 1 and 2. Same miRNA also has target regions in dengue serotype 4 in NS4B as well as NS5 region. hsa-miR-187 targets NS5 region in dengue serotypes 1, 2 and 4. hsa-miR-637 targets NS5 region in the dengue serotypes 1 and 4 and NS4 gene region in dengue serotype 2.

Table 2. Conserved human miRNA binding regions in dengue serotypes 1 and 3

S. No.	Human miRNA	miRNA binding regions in dengue serotype 1	miRNA binding regions in dengue serotype 3
1.	hsa-miR-490-5p	2392 – 2412	2386 – 2406
2.	hsa-miR-1296	802 – 822	797 – 823
3.	hsa-miR-939	249 – 278	246 – 278
4.	hsa-miR-608	1977 – 2005	1971 – 1999
5.	hsa-miR-770-5p	984 – 1012	984 – 1015
6.	hsa-miR-624*	984 – 1010	1008 – 1032
7.	hsa-miR-520d-3p	1155 – 1180	1157 – 1180
8.	hsa-miR-593*	633 – 666	618 – 641
9.	hsa-miR-431	2167 – 2187	2160 – 2181
10.	hsa-miR-486-5p	987 – 1012	1000 – 1020
11.	hsa-miR-675	835 – 862	835 – 862
12.	hsa-miR-188-5p	921 – 947	919 – 943
13.	hsa-miR-433	446 – 475	451 – 475
14.	hsa-miR-154	646 – 667	647 – 667
15.	hsa-miR-1234	991 – 1019	991 – 1018

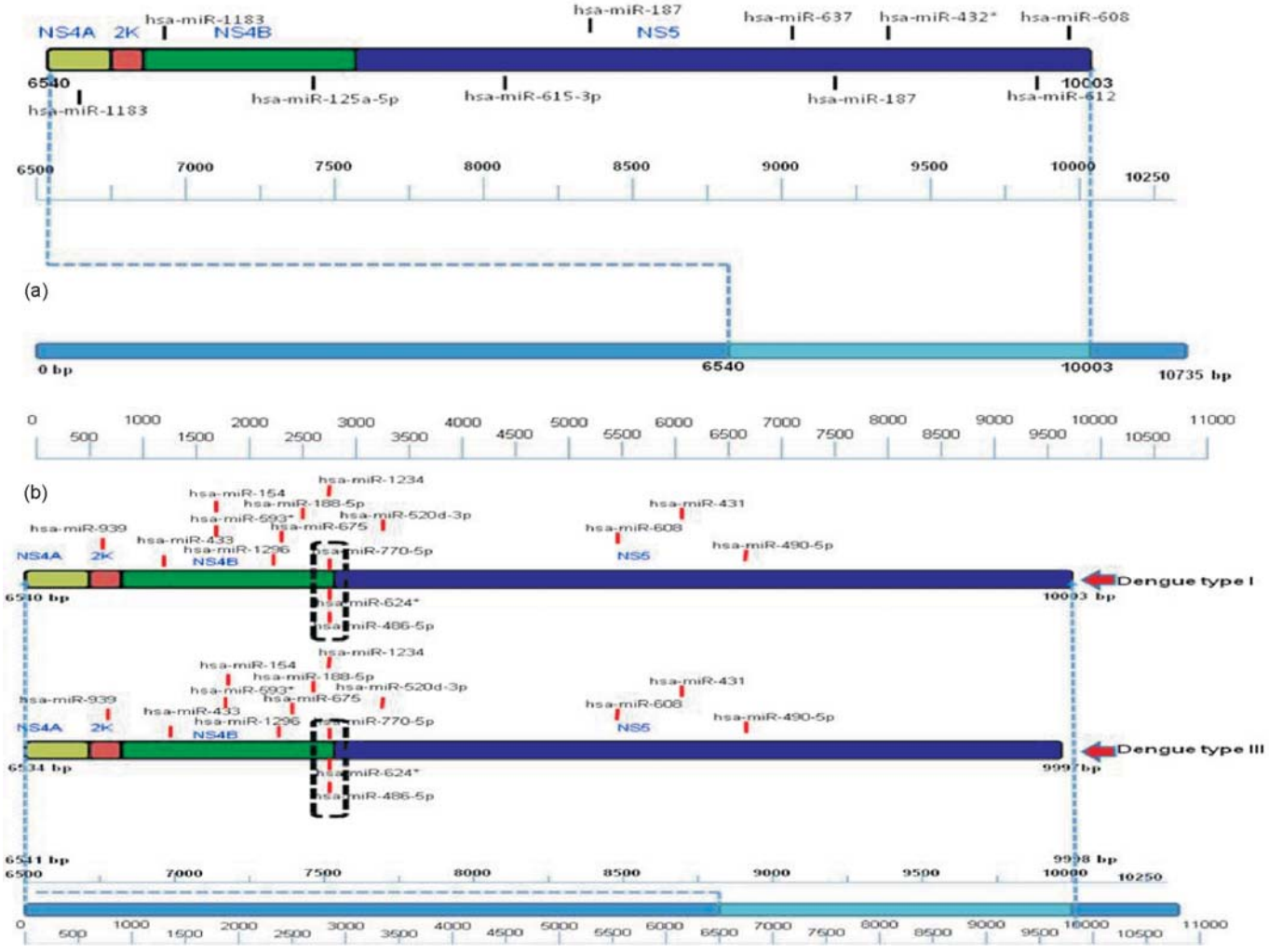

Fig. 2: Distribution and conservation of miRNA hits in the dengue genomes—(a) Hits showing microRNAs specific to different genomic regions in serotype 1 dengue genome; and (b) Target analysis of the obtained microRNA hits in dengue serotypes 1 and 3 genomes.

Binding sites for hsa-miR-1183 were predicted in NS4 region in dengue serotypes 1 and 3 and in NS5 region for dengue serotypes 2 and 4. Similarly, hsa-miR-619 targeted NS4B region in both dengue serotypes 2 and 3. NS5 region of dengue serotypes 2 and 3 was found to have target sites for hsa-miR-302d whereas same region (NS5) is targeted by hsa-miR-187 in dengue serotypes 1, 3 and 4. hsa-miR-302 showed binding affinity to NS4B and NS5 regions of dengue serotypes 4 and 3 respectively. It was observed that NS5 region of dengue serotype 4 was targeted by many human miRNAs suggesting a possible role for these miRNAs in host defense mechanism. Similarly, binding sites were predicted and investigated for all the members of *Flaviviridae* considered in this study (data not shown).

Interestingly, many miRNAs demonstrated specificity towards conserved genomic regions; this supports our hypothesis of searching the common miRNA target sites. Similar trend of target sites and targeting miRNA conservation was observed for the West Nile viruses as well as hepatitis C subtypes. In hepatitis C virus genotypes (1–6), similar miRNAs were found to target the region in different positions due to sequence similarity in six genomic regions. For instance, hsa-miR-608 target site is present in hepatitis C virus genotypes 1, 2, 4, 5 and 6 in five genomic regions but not in third genotype. In addition to this, hsa-miR-1183, hsa-miR-770-5p and hsa-miR-26b* demonstrated high frequency of target sites in genomic regions. Therefore, these results may pave a way for targeting all hepatitis virus genotypes with a single miRNA.

At region spanning from 7524 to 7552 (Fig. 2b) in dengue serotypes 1 and 3, we observed a cluster of three miRNA (hsa-miR-770-5p, hsa-miR-624*, hsa-miR-486-5p) targeting the same position (+10 or –10 nucleotides). Similar conserved clusters were not observed in dengue serotypes 2 and 4, which may be due to less similarity between the sequences.

All specific dengue virus genes and other genomic regions were observed and miRNA target specific to each gene and genomic regions were identified. For this analysis, NS4A, 2K, NS4B and NS5 genes were considered and positions of the top 10 miRNAs on these sites were observed. NS4A showed only one target of hsa-miR-1183 at 6626 to 6655 position. The 2K regions were not targeted by miRNA target, whereas NS4B showed two miRNA targets of hsa-miR-612 and hsa-miR-615-3p at position 9805 to 9837, and 8070 to 8097 respectively. NS5 possess seven miRNA targets of hsa-miR-187, hsa-miR-432*, hsa-miR-608, hsa-miR-637, hsa-miR-125a-5p, hsa-miR-187 and hsa-miR-1183 were observed at positions 9178 to 9203, 9330 to 9356, 9923 to 9955, 9022 to 9042, 7430 to 7461, 8342 to 8367, and 6884 to 6915 respectively.

Similar analysis and identification of microRNA target sites on specific genes and genomic regions were performed for all other considered genomes.

Identification of conserved region in 15 genomes

The conservation pattern found in the miRNA target regions tempted us to extend it to other viral genomes of the same *Flaviviridae* family. Keeping this in mind, 15 viral genomes of major concern were selected and collected from NCBI genome database. Similar exercise of genome alignment with ClustalW was carried out considering same parameters as taken previously for all four dengue virus genomes. As expected, all the selected 15 genomes showed good conservation of approximate length of 3500 nt. To find the conservation in miRNA target regions, top 10 microRNAs found in dengue virus genomes (serotypes 1–4) were taken and their target sites were explored in conserved genomic regions for all 15 viruses.

Target identification of 15 genomes

Total 40 microRNAs (top 10 from each dengue virus genome target) were considered initially and later on, redundant miRNAs were removed and unique 28 viral and 38 human miRNAs were taken as input for miRanda. MicroRNA target regions were identified for all 15 genomes. Top 10 miRNAs were identified based on the top score for all human and viral data sets. For viral miRNAs, top 10 scores are as follows: dengue serotype 1 (180 to 165), dengue serotype 2 (190 to 168), dengue serotype 3 (187 to 172), dengue serotype 4 (185 to 164), West Nile virus 1 (189 to 166), West Nile virus 2 (180 to 165), Japanese encephalitis virus (179 to 171), yellow fever virus (179 to 168), tick-borne encephalitis (185 to 170), hepatitis C virus 1 (184 to 165), hepatitis C virus 2 (185 to 170), hepatitis C virus 3 (176 to 167), hepatitis C virus 4 (183 to 168), hepatitis C virus 5 (182 to 169), and hepatitis C virus 6 (180 to 167).

For human miRNA targets, top scores obtained are dengue serotype 1 (179 to 167), dengue serotype 2 (201 to 173), dengue serotype 3 (184 to 168), dengue serotype 4 (184 to 165), West Nile virus1 (185 to 166), West Nile virus 2 (184 to 172), Japanese encephalitis virus (194 to 169), yellow fever virus (180 to 172), tick-borne encephalitis (180 to 171), hepatitis C virus 1 (184 to 172), hepatitis C virus 2 (183 to 164), hepatitis C virus 3 (193 to 172), hepatitis C virus 4 (187 to 173), hepatitis C virus 5 (194 to 167), and hepatitis C virus 6 (194 to 175).

Analysis of multiplicity

Multiplicity is an important criterion for understanding miRNA based regulation of gene expression and signifies a phenomenon in which a single microRNA targets several genes or genomic regions. Multiplicity was examined for the obtained top 11–12 miRNAs for all the genomes and ranked according to the maximum number of hits. Table 3 shows the obtained information on multiplicity of top 11 viral miRNAs. Among the viral miRNAs, mdv2-miR-M15 showed best results as it binds to all 15 genomes with maximum 20 hits, in some of genomes multiple hits were found (Table 3).

In case of human miRNAs, hsa-miR-1183 resulted in 17 hits in 12 genomes except yellow fever virus, hepatitis C virus 1, hepatitis C virus 2 and hepatitis C virus 6; hsa-miR-608 showed a maximum of 15 hits in 10 genomes while no hits were found in genomes of dengue virus serotype 2, dengue virus serotype 3, dengue virus serotype 4, tick-borne encephalitis virus genome and hepatitis C virus 3 (Table 4). Detailed analysis of multiplicity observed in other viruses is represented in Table 4. The viral microRNAs showed significant conservation and multiplicity in target prediction than human microRNAs.

Viral and human miRNA network

An interaction map was constructed based on miRNAs and their target genomic regions. Both human and viral top 10 miRNAs were considered (Fig. 3).

miRNA patterns

All top 10 miRNAs of viral and human were selected and multiple sequence alignment was carried out separately to analyze the conservation pattern of miRNAs. Alignment of the obtained top ranked viral miRNAs showed good conservation of uridine at 3rd, 8th, 13th, 19th and 21st positions. Adenine nucleotide conservation at miRNA binding sites was found in all 15 viral genomes. Alignment of obtained top human miRNA showed cytosine nucleotide conservation at the 13th position of the multiple sequence alignment (Fig. 4). Comparatively less conservation was observed for the human miRNAs than viral miRNAs.

Classification and clustering of the top 10 unique microRNAs

Top 10 miRNAs found in all 15 genomes were taken and redundant miRNAs were excluded from the list. After filtering, 30 human and 26 viral unique miRNAs were considered. Physicochemical parameters were calculated and 56 miRNAs were clustered. During clustering, the

Table 3. Analysis of multiplicity of viral miRNAs

Top 10 miRNA/ 15 species →	Dengue virus serotype 1	Dengue virus serotype 2	Dengue virus serotype 3	Dengue virus serotype 4	West Nile virus 1	West Nile virus 2	Japanese encephalitis virus	Yellow fever virus	Tick-borne encephalitis virus	Hepatitis C virus 1	Hepatitis C virus 2	Hepatitis C virus 3	Hepatitis C virus 4	Hepatitis C virus 5	Hepatitis C virus 6	Total No. of hits of miRNA
mdv2-miR-M15	1	3	1	2	1	2	1	1	1	1	1	2	1	1	1	20
mdv2-miR-M25-5p	1	0	0	0	1	2	0	1	0	2	1	1	1	1	1	12
mdv2-miR-M18-5p	2	0	0	1	3	1	1	1	1	0	1	0	0	1	0	12
ebv-miR-BART12	1	2	1	0	1	0	1	0	0	1	0	1	2	1	0	11
mdv2-miR-M16	0	2	0	1	1	1	1	1	1	0	0	0	1	1	1	11
mdv2-miR-M14-5p	0	0	0	0	0	0	0	0	0	1	2	2	0	1	1	7
rrv-miR-rR1-7-5p	0	0	0	0	1	1	0	1	0	1	0	1	0	0	1	6
rlcv-miR-rL1-15	1	1	0	0	0	0	1	0	0	1	1	0	0	1	0	6
mdv1-miR-M6	1	0	1	1	0	0	1	0	1	1	0	0	0	0	0	6
mdv1-miR-M5	0	0	2	0	0	0	0	0	1	0	1	0	2	0	0	6
hcmv-miR-UL36*	0	1	1	1	0	0	1	0	1	0	0	1	0	0	0	6
Total No. of hits	7	9	6	6	8	7	7	5	6	8	7	8	7	7	5	103

Table 4. Analysis of multiplicity of human miRNA

Top 10 miRNA/ 15 species →	Dengue virus serotype 1	Dengue virus serotype 2	Dengue virus serotype 3	Dengue virus serotype 4	West Nile virus 1	West Nile virus 2	Japanese encephalitis virus	Yellow fever virus	Tick-borne encephalitis virus	Hepatitis C virus 1	Hepatitis C virus 2	Hepatitis C virus 3	Hepatitis C virus 4	Hepatitis C virus 5	Hepatitis C virus 6	Total No. of hits of miRNA
hsa-miR-1183	2	2	2	1	2	1	1	0	1	0	0	2	2	1	0	17
hsa-miR-608	1	0	0	0	2	1	1	1	0	1	3	0	2	2	1	15
hsa-miR-593*	0	0	0	0	1	2	1	1	2	0	0	1	1	1	0	10
hsa-miR-612	1	0	1	1	1	1	1	0	1	0	0	1	0	1	0	9
hsa-miR-768-5p	0	0	0	1	0	0	0	1	1	1	1	1	0	1	1	8
hsa-miR-187	2	1	1	1	0	0	0	0	0	1	0	0	1	1	0	8
hsa-miR-637	1	0	1	1	1	1	0	0	0	1	1	0	1	0	0	8
hsa-miR-767-5p	0	1	1	1	1	1	0	1	1	0	0	0	0	0	0	7
hsa-miR-194*	0	0	0	0	0	1	1	1	1	1	1	0	0	1	0	7
hsa-miR-432*	1	0	0	0	1	1	1	0	0	0	0	0	0	0	1	5
hsa-miR-302a	0	1	1	1	0	0	1	0	0	0	0	0	0	0	1	5
hsa-miR-125a-5p	1	1	0	0	0	0	0	1	0	1	0	1	0	0	0	5
Total No. of hits	9	6	7	7	9	9	7	6	7	6	6	6	7	8	4	104

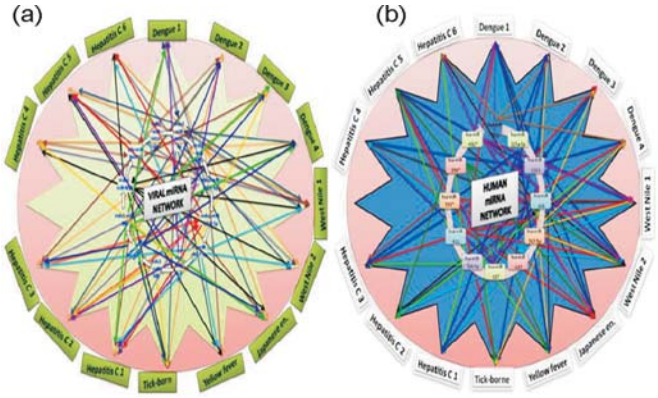

Fig. 3: Development of network for viral genomes and microRNA interactions—(a) Viral microRNA and target genome interaction network; and (b) Human microRNA and target genome interaction network.

dataset was normalized in a range of 0 to 1 and 2×2 grid was used to obtain the clusters with 100000 iterations and with a learning rate of 0.01 to 1.

Four clusters were generated successfully in a 2×2 grid (Fig. 5). Maximum number of sequences (26 sequences) constituted the 1st cluster (Cluster 1, 1) followed by 3rd (Cluster 2, 1), 2nd (Cluster 1, 2) and 4th (Cluster 2, 2) clusters, which contained 24, 5, and 1 sequences respectively. First and third clusters contain similar kind of viral and human miRNAs, whereas the second cluster comprises of four human miRNAs and a single viral miRNAs. The final cluster (Cluster 2, 2) of the grid contains a single human miRNA, i.e. hsa-miR-767-5p. These clusters of miRNAs are formed based on the computed physicochemical properties and further help us to understand and group the obtained top human and viral miRNAs. Maximum conservation in the whole data was observed in the second and third clusters (Figs. 5b and c) as evident from the figure with less colour difference. In the second cluster (cluster 1, 2), higher conservation was observed for GC content followed by melting temperature values whereas in 3rd cluster (cluster 2, 1) apart from these two parameters higher amount of conservation was observed for different energy values too. In comparison, the 1st cluster (cluster 1, 1) has shown divergence in almost all the parameters considered while grouping the miRNAs (Fig. 5a).

CONCLUSION

Flaviviruses cause complex viral diseases resulting in heavy socioeconomic burden. Understanding the viral genome structure and identifying the conserved regions and predicting the putative miRNA targets on such ge-

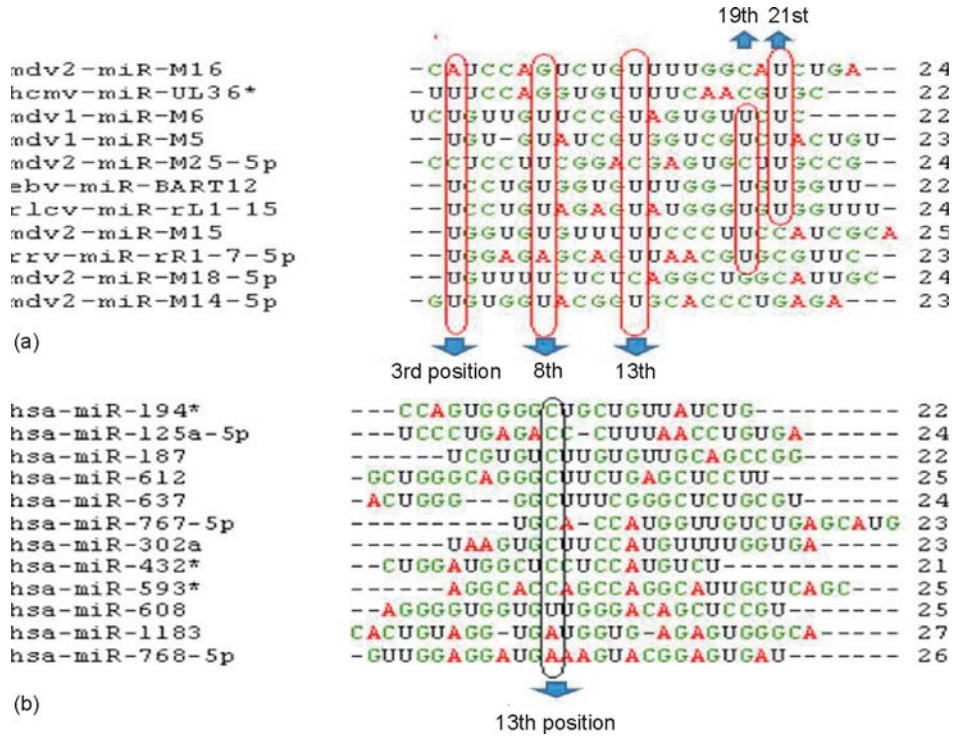

Fig. 4: Conservation pattern observed in the aligned top ranked microRNA sequences—(a) Observed conservation in the top ranked viral microRNA; and (b) Observed conservation in the top ranked human microRNA.

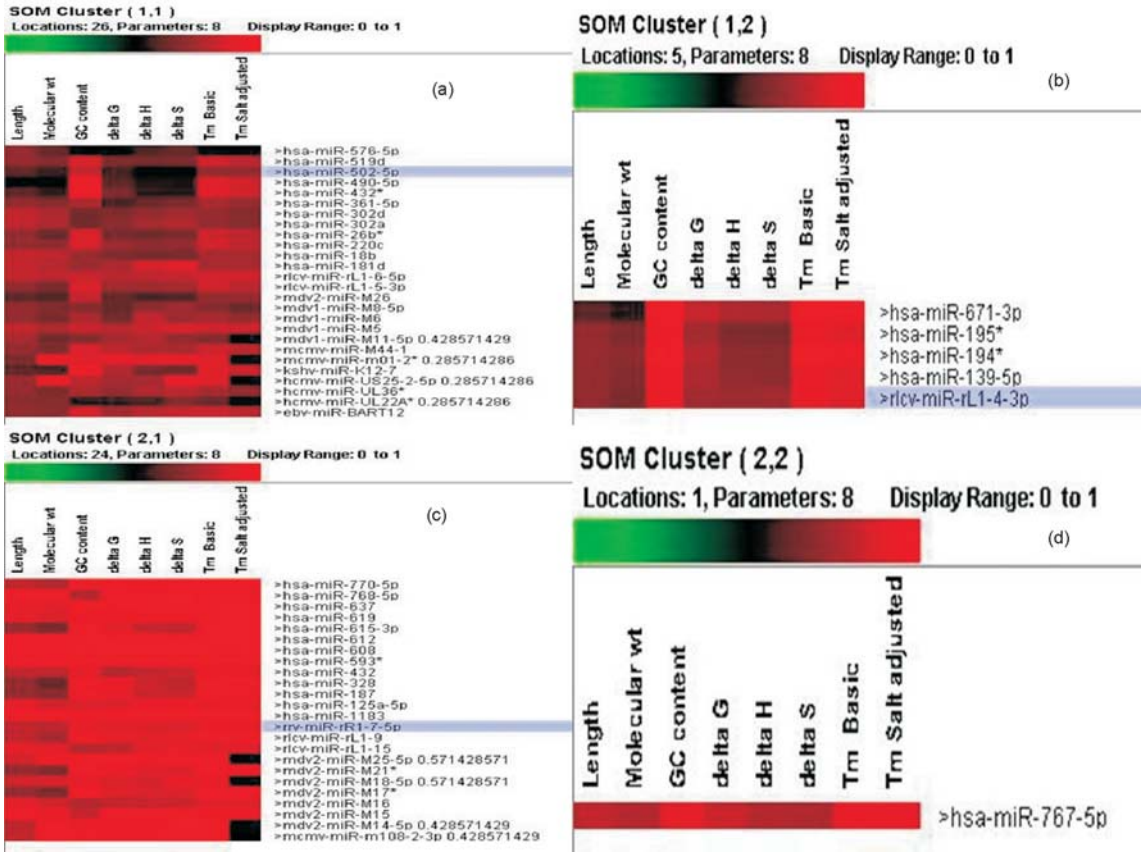

Fig. 5: Self organizing map (SOM) based clustering pattern for miRNAs based on their computed properties—(a) Cluster (1, 1); (b) Cluster (1, 2); (c) Cluster (2, 1); and (d) Cluster (2, 2).

nomic regions may prove to be an effective approach in controlling these viral diseases. Repression of candidate genes involved in pathogenesis or other associated activity by a single miRNA or a cluster of miRNAs may aid in efficiently controlling several diseases caused by different viruses of this family. Here, we propose a complex and comprehensive miRNA-target genome interaction map for viral and human miRNAs for the conserved regions of the 15 species of *Flaviviruses* considered in this study. Several physicochemical properties of top-ranking miRNAs were identified and analyzed specifically for this family.

Moreover, considerable similarity in network pattern between viral miRNA and human miRNA with viral genome implies a conserved pattern of target selection. The miRNAs from diverse species showed a close correlation of their target selection in viral genomes, which suggest that despite dynamic miRNA synthesis and function, target preference may be conserved.

The distribution of miRNA target sites on different regions of the viral genomes was also observed, which might aid in interpreting the mode of miRNA action at different levels including transcriptional, post-transcriptional or translational. The transient expression profiles of tiny miRNA markers can be used for easy diagnosis and can provide a platform for multi-dimensional treatment plan, which may aid in designing effective strategy against viral diseases. Despite complex relationship between miRNAs and viral genomes, emerging evidences suggest that the use of anti-miRs to target miRNA *in vivo* provides a novel and potential therapeutic approach for these viral diseases.

ACKNOWLEDGEMENTS

This work is funded by CSIR Network project to UB (BSC 0128) and CSIR Network project to MPB (SMILE CSC0111). AKB thanks CSIR for Senior Research Fellowship.

REFERENCES

1. Gould EA, Solomon T. Pathogenic flaviviruses. *Lancet* 2008; *371*(9611): 500–9.
2. Guzman MG, Halstead SB, Artsob H, Buchy P, Farrar J, Gubler DJ. *et al.* Dengue: A continuing global threat. *Nat Rev Microbiol* 2010; *8*(Suppl 12): S7–16.
3. Mammen MP, Pimgate C, Koenraadt CJ, Rothman AL, Aldstadt J, Nisalak A. *et al.* Spatial and temporal clustering of dengue virus transmission in Thai villages. *PLoS Med* 2008; *5*(11): e 205.
4. Lai CJ, Bray M, Men R, Cahour A, Chen W, Kawano H *et al.* Evaluation of molecular strategies to develop a live dengue vaccine. *Clin Diagn Virol* 1998; *10*(2–3): 173–9.
5. Marianneau P, Flamand M, Deubel V, Desprès P. Apoptotic cell death in response to dengue virus infection: The pathogenesis of dengue haemorrhagic fever revisited. *Clin Diagn Virol* 1998; *10*(2–3): 113–9.
6. Cardosa MJ, Baharudin F, Hamid S, Hooi TP, Nimmanitya S. A nitrocellulose membrane based IgM capture enzyme immunoassay for etiological diagnosis of dengue virus infections. *Clin Diagn Virol* 1995; *3*(4): 343–50.
7. Lindenbach BD, Thiel HJ, Rice CM. Flaviviridae: The viruses and their replication. In: Knipe DM, Howley PM, editors. *Fields Virology.* V edn. Philadelphia: Leppincott-Raven Publishers 2007; p. 1101–52.
8. Chambers TJ, Hahn CS, Galler R, Rice M. Flavivirus genome organization, expression, and replication. *Ann Rev Microbiol* 1990; *44:* 649–88.
9. Kalayanarooj S, Nimmannitya S. Is dengue severity related to nutritional status? *Southeast Asian J Trop Med Public Health* 2005; 36 (2): 378–84.
10. Lin HR, German D. Viral microRNA target allows insight into the role of translation in governing microRNA target accessibility. *Proc Natl Acad Sci USA* 2011; *108*(13): 5148–53.
11. Scaria V, Hariharan M, Maiti S, Pillai B, Brahmachari SK. Host-virus interaction: A new role for microRNAs. *Retrovirology* 2006; *3:* 68.
12. Obbard DJ, Gordon KH, Buck AH, Jiggins FM. The evolution of RNAi as a defence against viruses and transposable elements. *Philos Trans R Soc Lond B Biol Sci* 2009; *364*(1513): 99–115.
13. Sullivan CS, Ganem D. A virus-encoded inhibitor that blocks RNA interference in mammalian cells. *J Virol* 2005; *79*(12): 7371–9.
14. Simon-Mateo C, Garcia JA. MicroRNA-guided processing impairs Plum pox virus replication, but the virus readily evolves to escape this silencing mechanism. *J Virol* 2006; *80:* 2429–36.
15. Bentwich I. Prediction and validation of microRNAs and their targets. *FEBS Lett* 2005; *579*(26)*:* 5904–10.
16. Lagana A, Forte S, Russo F, Giugno R, Pulvirenti A, Ferro A. Prediction of human targets for viral-encoded microRNAs by thermodynamics and empirical constraints. *J RNAi Gene Silencing* 2010; *6*(1): 379–85.
17. Kohonen T. Self-organizing maps. *Springer Series in Information Sciences*, v 30. III edn. Berlin: Springer-Verlag 2001.
18. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994; *22* (22): 4673–80.
19. Singh SK, Pal Bhadra M, Girschick HJ, Bhadra U. MicroRNAs: Micro in size but macro in function. *FEBS J* 2008; *275*(20): 4929–44.
20. Winter J, Jung S, Keller S, Gregory RI, Diederichs S. Many roads to maturity: MicroRNA biogenesis pathways and their regulation. *Nat Cell Biol* 2009; *11*(3): 228–34.
21. Krol J, Loedige I, Filipowicz W. The widespread regulation of microRNA biogenesis, function and decay. *Nat Rev Genet* 2010; *11*(9): 597–610.
22. Griffiths-Jones S. The microRNA registry. *Nucleic Acids Res* 2004; *32* (Database Issue): D109–11.
23. Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. MicroRNA targets in *Drosophila*. *Genome Biol* 2003; *5*(1): R1.
24. John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS.

Human microRNA targets. *PLoS Biol* 2004; *2*(11): e363.
25. Wuchty S, Fontana W, Hofacker IL, Schuster P. Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers* 1999; *49*(2): 145–65.
26. Shinde PS, Arora N, Sarma P, Pal-Bhadra M, Bhadra U. Interaction map and selection of microRNA targets in Parkinson's disease-related genes. *J Biomed Biotechnol* 2009; *2009:* 363145.
27. Sripathi VR, Pal-Bhadra M, Arora N, Santosh S, Ray P, Banerjee AK, *et al*. Target sites for microRNA in pancreatic islets in type 2 *Diabetes mellitus* associated genes. *Onl J Bioinform* 2010; *11* (2): 224–43.
28. Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, *et al*. Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. *Proc Natl Acad Sci* 1999; *96:* 2907–12.
29. Murty US, Banerjee AK, Arora N. Application of Kohonen maps for solving the classification puzzle in AGC kinase protein sequences. *Interdiscip Sci* 2009; *1:* 173–8.
30. Banerjee AK, Arora N, Murty USN. Towards classifying organisms based on their protein physicochemical properties using comparative intelligent techniques. *Appl Arti Intell* 2011; *25*(5): 426–39.
31. Banerjee AK, Ravi V, Murty US, Sengupta N, Karuna B. Application of intelligent techniques for classification of bacteria using protein sequence-derived features. *Appl Biochem Biotechnol* 2013; *170*(6): 1263–81.

Correspondence to: Dr. Utpal Bhadra, Functional Genomics and Gene Silencing Group, Centre for Cellular and Molecular Biology, Hyderabad–500 007, India.
E-mail: utpal@ccmb.res.in

Received: 15 August 2013 *Accepted in revised form:* 5 December 2014