Preprints of the 12th IFAC Symposium on Computer Applications in Biotechnology
The International Federation of Automatic Control
16-18, 2013, December. Mumbai, India

miRNA Mediated Regulation of Rice (*Oryza sativa*) Genome ⋆

Chittabrata Mal * **Arindam Deb** * **Md. Aftabuddin** ** **Sudip Kundu** *

* *Department of Biophysics, Molecular Biology & Bioinformatics, University of Calcutta, 92, A.P.C. Road, Kolkata 700009, India (e-mail: skbmbg@caluniv.ac.in)*
** *West Bengal University of Technology, BF-142, Salt Lake, Sector I, Kolkata 700064, India*

Abstract: miRNAs play important roles in plant post-transcriptional gene regulation by targeting mRNAs through cleavage or repressing translation. Here, we have reconstructed genome scale level miRNA-miRNA co-targeting network of rice plant and identified several significant modules (high-density sub-graphs). Some of the modules are involved in multiple biological processes, while some are involved in a single biological process. The distribution of the out-going connectivities of co-target network can be approximated best by a power-law equation. We have also observed a wide variation in inter chromosomal regulation. The genes of chromosome 3 are highly targeted by the miRNAs synthesized from other chromosomes. The results presented here might provide a platform for testing the hypothetical role of co-targeting associations in rice post-transcriptional control.

Keywords: Rice *(Oryza sativa)*, miRNA co-target network, hub miRNAs, combinatorial regulation.

1. INTRODUCTION

Rice is a major staple food for more than half of the world's human population. There is an increasing demand for high yielding, stress tolerant rice cultivars. To meet this challenge we have to understand the cellular physiology of rice, especially the activities of their genes and their regulation. The regulation can occur at the level of transcription, RNA processing, mRNA lifetime and translation etc. The miRNAs are a large family of about 21-22 nucleotide endogenous non-protein-coding regulatory RNA sequences [Bartel and Bartel (2003)] which are the key players in post- transcriptional gene regulation and involve in cleavage, degradation or translational inhibition of its target mRNA with a resultant repression of gene expression in animals, plants, and fungi [Bartel and Bartel (2003); Hunter and Poethig (2003)]. Plant development, response to environmental stress, pathogen invasion and regulation of their own biogenesis etc. are finely tuned by miRNAs. Although miRNAs have been studied in recent years; to date no study has been performed on analysis of miRNA co-targeting network of rice. Here, we have exploited the available experimental and computational data of rice miRNA and their target genes; and have reconstructed and analyzed the rice miRNA co-targeting network.

Most miRNAs were identified by computational predictions [Adai et al. (2003); Bonnet et al. (2004); Rhoades et al. (2002)] or by direct cloning of small RNAs [Lai et al. (2003)]. Computational approaches provide a valuable method to predict miRNAs and their respective targets, and have been successfully applied in vertebrates [Lim et al. (2003); Lewis et al. (2003)], insects [Rajewsky and Socci (2004)], Arabidopsis and rice [Bonnet et al. (2004); Jones-Rhoades and Bartel (2004); Wang et al. (2004)]. According to the PMRD Database, 8433 miRNA genes have been collected from 121 plant species [Zhang et al. (2010)]. Out of those, 2540 miRNA genes have been identified in rice (*Oryza sativa*) of which 269 and 2271 miRNAs are identified experimentally and computationally, respectively. However, many miRNA genes and their targets, especially in plants, still to be discovered which in turn will help to understand their critical roles in diverse biological processes. Although thousands of miRNAs have been identified, the function of most miRNAs involved in biological networks remains unclear [Tsang et al. (2010)]. It has been hypothesized that miRNAs target a set of related genes to regulate a specific pathway or process [Xu (2011)]. But the incident of multiple miRNAs targeting same gene (co-targeting) is not investigated so far in case of rice.

Further, a relatively small number of miRNAs can set up remarkably complex spatial and temporal patterns of gene expression by means of combinatorial or differential gene regulation. Genome-scale combinatorial miRNA regulation has been studied mostly in case of human [Xu (2011); Balaga et al. (2012)] and a small number of plants [Je-Gun and Zhangjun (2009); Liu et al. (2012)] but not for rice. Here, we have constructed a genome scale level miRNA-miRNA co-targeting network of rice and also have

⋆ Financial support : UGC (RFSMS), Govt. of India.

identified some modules which are involved in different important biological processes.

2. METHODS

2.1 Data

The miRNA mature sequences are taken from PMRD database [Zhang et al. (2010)]. Only those miRNAs which have precursor sequences are considered. miRNA targets are extracted from psRNATarget database [Dai and Zhao (2011)] with default parameters (i.e., Maximum expectation: 3, Length for complementarity scoring (hspsize): 20, Target accessibility - allowed maximum energy to unpair the target site (UPE): 25, Flanking length around target site for target accessibility analysis: 17 bp in upstream / 13 bp in downstream, Range of central mismatch leading to translational inhibition: 9-11 nt). We want to mention that it has been suggested that the value of maximum expectation should be 4.0-5.0 to get a higher prediction coverage while that should be 0.0-2.0 for lowering the false positive prediction. Here, we have taken a conservative approach by taking the mid-value i.e., 3.0. GO annotations were downloaded from Biomart [Smedley et al. (2009)] release 12. Loci dataset was taken from *Oryza sativa* MSU Rice Genome Annotation (release 6.1).

2.2 Construction of post-transcriptional miRNA - miRNA co-targeting network (p-CNet)

miRNA target data sets are collected from database and preprocessed as described in the 'Data' section. In the p-Cnet, an edge is given between a pair of miRNAs if they target a common target gene. Thus,18045 pair-wise interactions between 2348 miRNAs have been identified. The edge-weight is denoted by the number of common target genes (TGs) targeted by the pair. Then this value is normalized by the ratio of the observed number of shared target genes by a pair of miRNA to the expected number of TGs shared by the same pair. The normalized value is termed as co-targeting coefficient (CC) value for each pair. The expected number is calculated by taking the average of shared TGs for a miRNA pair in 60,000 randomly generated post-transcriptional networks where degree distribution remains similar to that of the original network. The co-targeting network is then defined as the set of miRNAs and links between them that have $CC > 1$, i.e. higher than random co-targeting association. After randomization and considering O_{min} (minimum number of co-targeted loci)= 3, a p-Cnet with 16850 edges between 1370 miRNAs is obtained. The same threshold of CC value has been used in [Balaji et al. (2006)].

In the p-Cnet, the weight of the edge between miRNAs is a measure of the extent of co-targeting association between pairs of miRNAs over what is expected by chance. Bias arising due to chance sharing of TGs, especially seen in high-throughput datasets, is normalized by this procedure [Balaji et al. (2006)]. We have used this p-Cnet consisting of significant co-targeting interactions of miRNAs in rice (Figure 1) for further genome-scale analysis.

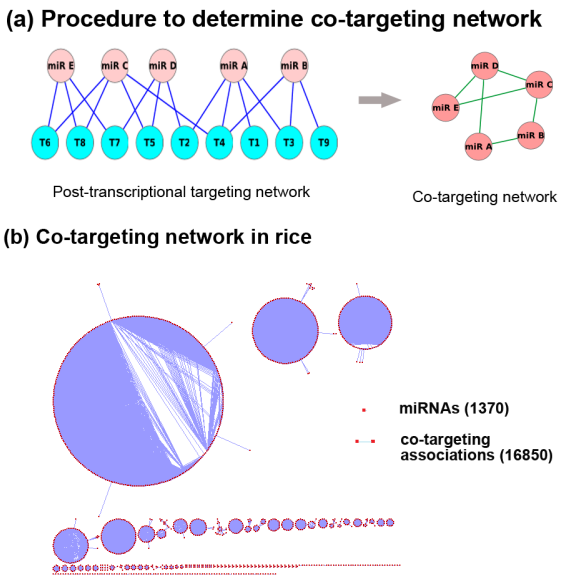

Fig. 1. (a) Procedure to determine a co-targeting network starting from a post-transcriptional target network. Pink circles represent miRNAs, cyan circles represent target genes. Targeting interactions are shown as blue lines connecting the miRNAs and the target gene in the post-transcriptional network. To create a network of miRNAs alone, we link two miRNAs (green lines) if they target a common target gene. (b) The co-targeting network in rice derived using the procedure described above. This network consists of 1370 miRNAs, with 16,850 co-targeting associations (i.e. pairs or edges with $CC > 1$)

3. RESULTS AND DISCUSSIONS

3.1 Analysis of combinatorial regulation in Oryza sativa genome-scale post-Transcriptional regulatory network

Computationally predicted and experimentally verified miRNAs and their target genes of rice are collected from PMRD database. We have used this data to reconstruct the largest possible post-transcriptional targeting network (p-Tnet) in *Oryza sativa*. We have obtained a network consisting of 2348 miRNAs and 6815 target genes (TGs) involving a total of 18045 targeting interactions. Here, both the miRNAs and their TGs are treated as nodes and interactions of miRNAs on TGs are considered as edges.

Next, we analysis the network architecture and observe that each miRNA targets ∼7.68 TGs; while on average, each TG is targeted by ∼2.65 miRNAs. The top 20% of miRNAs with high out-going connectivity (targeting the genes) are defined as hubs. These hubs, in total, are involved in more than half the number of interactions in the network (Figure 2). These hub miRNAs can act as master regulators. One of such hubs is osa-miR1439 which targets 25 genes.

The connectivity distributions in the post-transcriptional targeting network (p-Tnet) are shown in Figure 3. It indicates that several miRNAs can target a single TG. This may be happened due to two reasons - (i) due to the sequence similarities of two different miRNAs, they

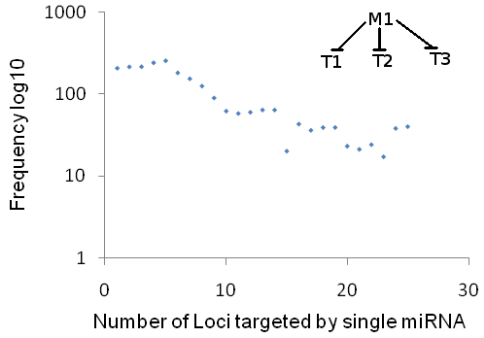

Fig. 2. Distribution of number of target loci per miRNA. x-axis represents the number of target loci, n, targeted by a miRNA and the y-axis (logarithmic scale) represents the number of miRNAs that target 'n' number of loci. M1 is miRNA and T1, T2, T3 are its targets.

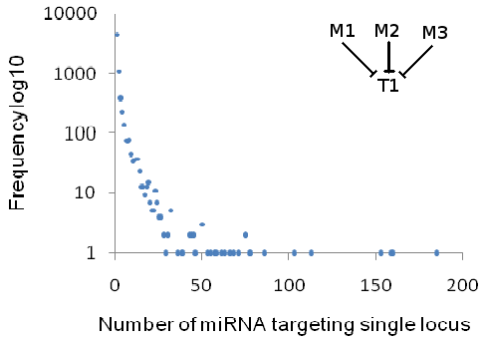

Fig. 3. Distribution of the number of miRNAs per target loci. The x-axis represents the number of miRNAs m targeting a target loci and the y-axis (logarithmic scale) represents the number of loci regulated by 'm' number of miRNAs. Three miRNAs M1, M2, M3 target one locus T1.

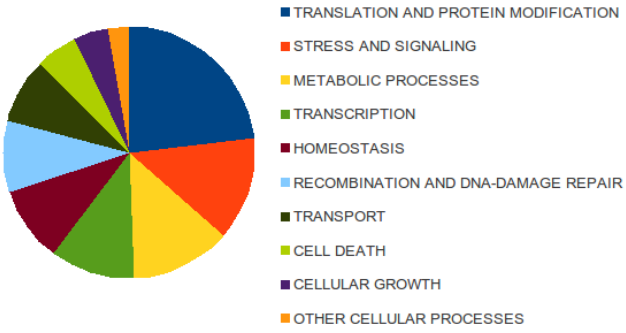

Fig. 4. pie-chart of miRNAs involved in different biological processes.

can target the same TG, and / or (ii) the different miRNAs have dissimilarities in their sequences and still bind to the same target gene at overlapping or non-overlapping sites. For example, the miRNAs osa-miR156a, osa-miR156b, osa-miR156c, osa-miR156d, osa-miR156e, osa-miR156f, osa-miR156g having identical sequences target a single gene (LOC_Os08g41940.1) at the same region. However, the precursor sequences of these miRNAs are different. On the other hand, some miRNAs of miR156 family (osa-miR156a, osa-miR156k, osa-miR156l, osa-miR156o) have different sequences but target a single locus (LOC_Os01g69830.1) at overlapping regions. Similarly, the miRNAs from two different families, namely, osa-miR1436 and osa-miR818d have different sequences and target a single gene (LOC_Os07g40450.2) at overlapping region. Thus, these multiple incoming connectivities of a TG indicate that there are miRNA mediated redundancies and regulations through binding either at the same site or at the overlapping sites of the TG. Previously it was hypothesized that plant miRNAs may act in a functionally redundant manner. In Arabidopsis, the MIR164 family (MIR164a, MIR164b and MIR164c) is important for shoot development. Sieber et al. has shown that a loss of entire miR164 activity leads to a severe disruption of shoot development, in contrast to the effect of mutation in any single MIR164 gene [Sieber et al. (2007)]. We have also observed a set of miRNAs which can co-target and bind to the same gene at non-overlapping sites of the TG. For example, the miRNAs, osa-miR168b and osa-miRf10096-akr have the potential binding sites at two non-overlapping positions (636-655 and 541-546, respectively) of the target gene LOC_Os08g04460.1. This indicates the possibility of redundancies as well as combinatorial regulation. We have also obtained a large set of this type of miRNA pairs which can bind at different sites of the same target genes. This encouraged us to find out general principles of co-targeting relations among different miRNAs in the post-transcriptional regulation of *Oryza sativa* (described in Sections 3.3-3.4).

3.2 GO classification of miRNA-targeting single and multiple loci

The miRNAs target a large number of loci and thus it is also expected that they have the potentiality to control a large number of biological processes. A pie-chart of miRNAs regulating different biological processes is described in the Figure 4.

Some biological processes are targeted by higher percentage of miRNAs. For example, 23.11% miRNAs are involved in translational and protein modification, 13.34% miRNAs are involved in stress and signalling, 12.97% miRNAs are involved in metabolic processes etc. Some biological processes are targeted by lower percentage of miRNAs. For example, 4.51% miRNAs are involved in cellular growth and 5.42% miRNAs are involved in cell death.

Further, the distributions of miRNAs and miRNA-targeted loci show that while a large number of miRNAs target a small number of loci, a few miRNAs target a large number of loci. For example, 75% miRNAs target small number of loci (1 to 10) while 25% miRNAs target large number of loci (11 to 25). The GO functional classification of those loci indicates that the genes targeted by the miRNAs are involved in different important functions of a cell. From the miRNA-target interactions (p-Tnet) data we have found that single miRNA targets maximum 25 loci. GO classification of the loci targeted by such 40 miRNAs, each having 25 target loci has shown that some of them have roles in various biological functions. These are osa-miR1439, osa-miR812e, osa-miR812a, osa-miR812b, osa-miR812c, osa-miR812d, osa-miR1884a, osa-miR414 along with some uncharacterized "-akr" group of miRNAs. For example,

osa-miR812a targets those loci which are involved in 17 biological functions like cytolysis (GO:0019835), protein ubiquitination (GO:0016567), modification-dependent protein catabolic process (GO:0019941), xylan catabolic process (GO:0045493), pathogenesis (GO:0009405) regulation of transcription (GO:0045449), response to bacterium (GO:0009617) etc. The loci targeted by osa-miR1884a are involved in 12 biological functions like diterpene phytoalexin precursor biosynthetic process pathway (GO:0051504), regulation of transcription, DNA-dependent (GO:0006355), oxidation reduction (GO:0055114), gibberellin metabolic process (GO:0009685), purine base metabolic process (GO:0006144) etc. Thus, we can conclude that there exists a large number of miRNAs which can regulate different cellular processes and thus indicates that the dysregulation of miRNAs can affect or alter the overall cellular process. These hub miRNAs can act as master regulators.

On the other hand, we have found that some of the loci are targeted by large number of miRNAs. For example, the LOC_Os05g15150.1 (leucyl-tRNA synthetase, cytoplasmic, putative, expressed) targeted by 61 miRNAs has role in two biological processes e.g.,translation (GO:0006412), tRNA aminoacylation for protein translation (GO:0006418). The LOC_Os03g22050.3 targeted by 45 miRNAs has role in two biological processes, namely, protein amino acid phosphorylation (GO:0006468)and signal transduction (GO:0007165). This probably indicate that depending on the specific relationship of combinatorial regulation and condition specific expression of miRNAs, the ultimate level of protein coded by that loci would vary.

Though most of the miRNA-targeted loci have been fallen in different GO category, some loci are not classified in any of them. For example, LOC_Os08g27180.1 targeted by 185 miRNAs is not classified under any GO category. The proper functional annotation of these uncharacterized genes, targeted by large number of miRNAs, would help the researchers to understand the regulation of cellular process.

Further, distributions of miRNA and their regulating GO categories also show that while a large number of miRNA regulate a small number of biological functions, a few miRNAs regulate a large number of biological functions (Figure 5). Among the 1678 miRNAs, which are involved in at least one GO category, the top 2.5% miRNAs regulate more than 13 biological processes. For example, osa-miRf10306-akr, osa-miRf11937-akr, osa-miR437 are involved in 20, 19, 18 biological processes, respectively. The hub miRNAs (top 2.5% miRNAs) including above three are involved in 126 different biological processes and among them, some biological processes are more frequently targeted by other miRNAs. Some of the biological functions targeted by large number of hub-miRNAs are metabolic process (GO:0008152), protein amino acid phosphorylation (GO:0006468), transport (GO:0006810) etc.

3.3 Characteristics of Co-targeting network

A network transformation procedure is applied here to build the post-transcriptional co-targeting network (p-Cnet) from the p-Tnet which finally helps us to understand

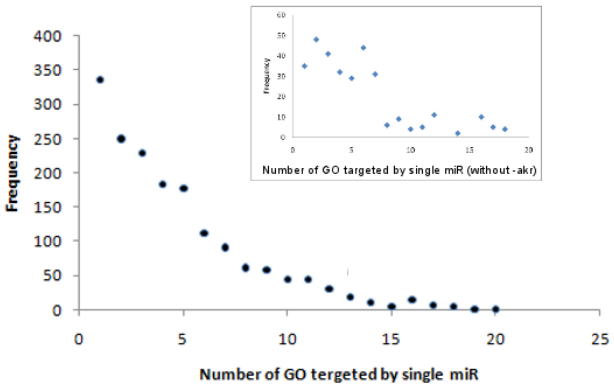

Fig. 5. Distribution of number of GO per miRNA. The x-axis represents the 'm' number of GO, targeted by a miRNA and the y-axis represents the number of miRNAs that target 'n' number of GO. Inset: Distribution of number of GO per miRNA (without -'akr').

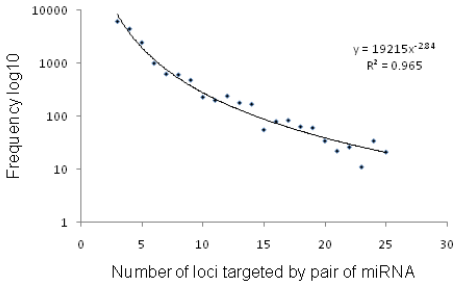

Fig. 6. Distribution of number of target loci per miRNA-pair. x-axis represents the number of target loci, n, targeted by a miRNA-pair and the y-axis (logarithmic scale) represents the number of miRNA-pairs that target n target loci. The outgoing degree distribution can be best approximated by a power-law equation $y = 19215x^{-2.84}, R^2 = 0.965$.

the co-targeting interactions [Balaji et al. (2006)]. Thus, a network is formed where all nodes are miRNAs and the edges are co-targeting associations between miRNAs.

We have observed that on average, each miRNA-pair targets ∼5.19 TGs.The distribution of the out-going connectivities of miRNA pair in p-CNet can be approximated best by a power-law equation $y = 19215x^{-2.84}, R^2 = 0.965$ (Figure 6).

The higher the CC-value indicates the higher association of the biological function (through their shared TGs) of a miRNA pair. We have identified that many miRNA pairs have a relatively low CC, but a small number of miRNA pairs have very high co-targeting coefficients (Figure 7). In several cases, pairs with above average co-targeting coefficients, for example, osa-miR812d and osa-miR812b (CC=24.1214), are involved in more number of biological processes which include twelve GOs such as GO:0044262 (carbohydrate metabolic process), GO:0016567 (protein ubiquitination), GO:0017148 (negative regulation of translation), GO:0006468 (protein amino acid phosphorylation), GO:0006810 (transport) etc. In other cases, we have also observed a large number of pairs with below average

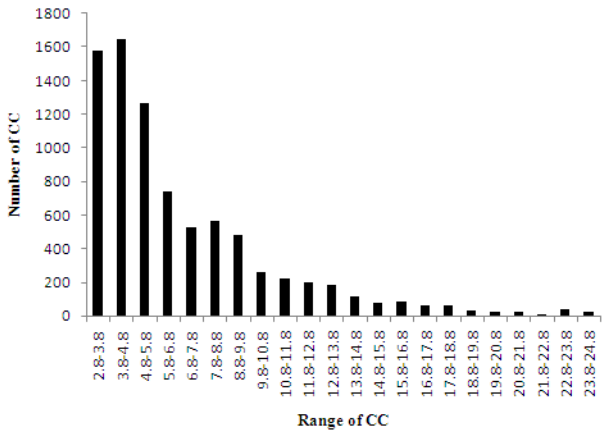

Fig. 7. Distribution of the co-targeting coefficient values in the co-targeting network. The x-axis represents the range of co-targeting coefficient, and the y-axis represents the number of pairs within that co-targeting coefficient range.

co-targeting coefficients which are involved in less number of biological processes. For example, osa-miRf10406 and osa-miR812b (CC=2.9001) are involved in only one GO; GO:0008152 (metabolic process).

3.4 Complexity at the modular level in the co-targeting network

In this work, we have used modular network concept to understand the effect of co-targeting miRNAs. We have defined a module in the co-targeting network as a k-clique, i.e. a highly-dense sub-graph with 'k' number of miRNAs with all miRNAs having co-targeting association with other miRNAs in the sub-graph. This procedure is performed using Cytoscape [Shannon et al. (2003)] AllegroMCODE Plug-in with k-core=3. We have identified 71 such modules, where each module has a unique composition of miRNAs and the same miRNA or the same pair do not occur in more than one module. The distribution of module size shows that modules with higher number of miRNAs are less in number whereas modules with lower number of miRNAs are more in number. The largest module has 101 miRNAs. Though most of the modules are comprised of "-akr" group of miRNAs, we have observed some mixed population of miRNAs - "-akr" along with non-"akr" families within some of the modules. For example, in Module 27, osa-miRf10201-akr is present with osa-miR809 family (osa-miR809h, osa-miR809e, osa-miR808, osa-miR809b, osa-miR809c, osa-miR809a, osa-miR809f, osa-miR809d, osa-miR809g). In Module 46, osa-miRf10978-akr is present with osa-miR399 family. This modularity may help in characterizing the possible biological effect of undefined miRNAs ("-akr").

3.5 Inter-chromosomal regulation of miRNAs

The miRNAs (excluding '-akr') synthesising from one chromosome target genes of the same chromosome as well as those of other chromosomes. The number of genes present in different chromosome vary. We have calculated the inter-chromosomal regulation, X_{ij} (regulation of j^{th} chromosomal genes by i^{th} chromosomal miRNAs) using the following expression: X_{ij} = (miRNAs of i^{th} chromosome targeting genes of j^{th} chromosome/Total number of genes present in the j^{th} chromosome) $\times$ 100%. Figure 8 represents the heat map of the distribution of target genes in different chromosomes. A wide variation of inter chromosomal regulation mediated by miRNA is clearly depicted from the Figure 8. We have also found that miRNAs synthesising from all the twelve chromosomes have greater percentage of target genes on chromosome 3.

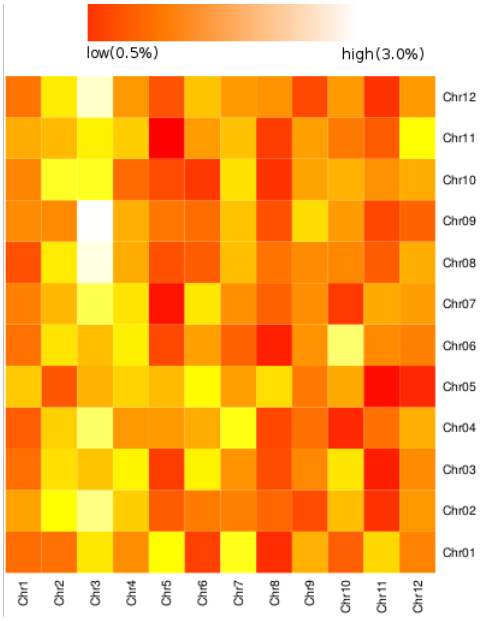

Fig. 8. Inter-chromosomal regulation of miRNAs. The miRNAs originated from chromosome 'i' in the y-axis regulate the genes of chromosome 'j' in the x-axis. The colour box shows the intensity of regulation.

ACKNOWLEDGEMENTS

CM and AD gratefully acknowledge UGC(RFSMS), Government of India for financial support and DIC, Calcutta University for computational facilities.

REFERENCES

Adai, A., Johnson, C., Mlotshwa, S., et al. (2003). Computational prediction of mirnas in arabidopsis thaliana. *Genome Res*, 15, 78–91.

Balaga, O., Friedman, Y., and Linial, M. (2012). Toward a combinatorial nature of microrna regulation in human cells. *Nucl. Acids Res.*, 40(19), 9404–16.

Balaji, S., Madan, B.M., Iyer, L.M., Luscombe, N.M., and Aravind, L. (2006). Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast. *Journal of Molecular Biology*, 360, 213–227.

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

Bonnet, E., Wuyts, J., Rouze, P., and Van de Peer, Y. (2004). Detection of 91 potential conserved plant micrornas in arabidopsis thaliana and oryza sativa identifies important target genes. *Proc Natl Acad Sci U S A*, 101, 11511–6.

Dai, X. and Zhao, P.X. (2011). psrnatarget: A plant small rna target analysis server. *Nucleic Acids Research*, 39, 155–9.

Hunter, C. and Poethig, R.S. (2003). Missing links: mirnas and plant development. *Curr Opin Genet Dev*, 13, 372–8.

Je-Gun, J. and Zhangjun, F. (2009). Identification of microrna regulatory modules in arabidopsis via a probabilistic graphical model. *Bioinformatics*, 25(3), 387–393.

Jones-Rhoades, M.W. and Bartel, D.P. (2004). Computational identification of plant micrornas and their targets, including a stress-induced mirna. *Mol Cell*, 14, 787–99.

Lai, E.C., Tomancak, P., Williams, R.W., and Rubin, G.M. (2003). Computational identification of drosophila microrna genes. *Genome Biol*, 4, R42.

Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microrna targets. *Cell*, 115, 787–8.

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

Liu, H., Jin, T., Liao, R., Wan, L., Xu, B., Zhou, S., and Guan, J. (2012). mirfans: an integrated database for arabidopsis thaliana microrna function annotations. *BMC Plant Biology*, 12, 68.

Rajewsky, N. and Socci, N.D. (2004). Computational identification of microrna targets. *Dev Biol*, 267, 529–35.

Rhoades, M.W., Reinhart, B.J., Lim, L., et al. (2002). Prediction of plant microrna targets. *Cell*, 110, 513–20.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, 13(11), 2498–504.

Sieber, P., Wellmer, F., Gheyselinck, J., Riechmann, J.L., and Meyerowitz, E.M. (2007). Redundancy and specialization among plant micrornas:role of the mir164 family in developmental robustness. *Development*, 134, 1051–1060.

Smedley, D., Haider, S., Ballester, B., Holland, R., London, D., Thorisson, G., and Kasprzyk, A. (2009). Biomart biological queries made easy. *BMC Genomics*, 10, 22.

Tsang, J.S., Ebert, M.S., and Oudenaarden, A. (2010). Genome-wide dissection of microrna functions and cotargeting networks using gene set signatures. *Molecular Cell*, 38, 140–153.

Wang, X.J., Reyes, J.L., Chua, N.H., and Gaasterland, T. (2004). Prediction and identification of arabidopsis thaliana micrornas and their mrna targets. *Genome Biology*, 5, R65.

Xu, J. (2011). Mirna - mirna synergistic network: construction via co-regulating functional modules and disease mirna topological features. *Nucl. Acids Res.*, 39(3), 825–836.

Zhang, Z., Yu, J., Li, D., Zhang, Z., Liu, F., Zhou, X., Wang, T., Ling, Y., and Su, Z. (2010). Pmrd: plant microrna database. *Nucleic Acids Research*, 38, D806–D813.