Genome Scale Pathway-Pathway Co-functional Synergistic Network (PcFSN) in Oryza Sativa

Chittabrata Mal[1](✉), Ayushman Kumar Banerjee[1], and Joyabrata Mal[2]

[1] Department of Bioinformatics, Maulana Abul Kalam Azad University of Technology, West Bengal, Nadia, India
chittabrata@gmail.com, chittabrata.mal@makautwb.ac.in
[2] Department of Biotechnoloygy, MNNIT Allahabad, Prayagraj, India

Abstract. Cohesive network modelling and systems biology have emerged as extremely potent tools which helps understanding the combinatorial effects of biomolecules. Synergistic modulation among biomolecules (e.g., enzymes, transcription factors, microRNAs, drugs, etc.) are significant in finding out complex regulatory mechanisms in biological networks and pathways. In some cases, although combinatorial interactions among some biomolecules in specific biological networks is available, our knowledge in that particular domain is very limited with context to a genomic scale. Here we explore the pathway-pathway network to identify and understand the network architecture of metabolic pathway mediated regulations at genomic and co-functional levels, in rice. Using network transformation methods, a genome scale pathway-pathway co-functional synergistic network (PcFSN) was constructed. Finally, the PcFSN modules are extracted. This in turn helps to identify the miRNAs and genes associated with the pathways, especially linked to the central metabolic network in rice.

Keywords: miRNA · synergistic network · modules · genome scale

1 Introduction

Rice is treated as a major staple food for more than half of the global population [1]. Rice-biotechnologists are continuously working towards finding novel methodologies to increase the production of this crop by generating high yielding, stress tolerant cultivars. This requires an understanding in the cellular physiology, which includes gene functions and their corresponding regulation at the transcriptional level, mRNA lifespan, processing of RNA, post-translational modifications, metabolic pathways, enzymatic activities, etc. Besides the genes, function of some important regulatory non-protein coding RNAs are also equally important to study.

The miRNAs are short sequences, 21-22 nucleotide long endogenous non-protein-coding regulatory RNA sequences which are believed to play significant roles in post-transcriptional gene regulation. miRNAs play a critical role in cleavage formation, degradation or translational inhibition of their respective mRNAs which can result in suppression of gene expression in animals, plants, and fungi. Such tiny molecules are the

V. Agarwal et al. (Eds.): BIOSANGAM 2022, ABSR 24, pp. 47–57, 2023.
https://doi.org/10.2991/978-94-6463-020-6_6

influencers of plant development, response to environmental stress, pathogen invasion and regulation of their own biogenesis i.e., the metabolic processes etc. [2]. These can be identified either with the help of computational methods, or by directly cloning small RNAs. The computationally predicted miRNAs and their targets are frequently considered for experimentation in several species of vertebrates, insects, Arabidopsis and rice to know their exact roles. However, several miRNAs and their target genes are yet to be discovered and whose discovery might help us to understand the critical role of miRNA mediated regulation in more detail.

Genome-wide miRNA regulation has been already studied in rice. miRNA mediated gene regulatory network in roots, signal transduction, leaf senescence, vegetative and reproductive regulation are few examples [3–6]. Recently pathway centric co-expression studies are also gaining importance so that gap between gene expression and pathway perturbation can be minimised [7, 8]. While these studies have shed light on condition and tissue specific regulatory role of miRNAs in other species, a comprehensive analysis of genome scale pathway-mediated co-target and co-functional networks in rice has not been done so far. We constructed a genome scale pathway-pathway synergistic co-functional network of rice and has subsequently identified some significant modules where the involved nodes can co-express and co-regulate different but related enzymes.

2 Materials and Methods

2.1 Data Retrieval

All the pathways, reactions, enzymes, and genes for Oryza sativa were downloaded from the release version 14 of the PMN (Plant Metabolic Network v14 [9] https://www.plantcyc.org/) database. The sequences of mature miRNAs were obtained from the freely available miRbase v22 (http://mirbase.org/) [10]. Rice genome data was collected from *Oryza sativa* MSU Rice Genome Annotation Project 7 [11]. MSU LOC to RAP-DB ID mapping file was used for the conversion of MSU LOC IDs of rice genes to updated RAP-DB IDs. (https://rapdb.dna.affrc.go.jp/download/archive/RAP-MSU_2017-04-14.txt.gz).

2.2 Software and Tools

The plant-small RNA Target analysis server, psRNATarget (release 2017 [12] http://plantgrn.noble.org/psRNATarget/) was used for the prediction of miRNA targets corresponding to identified miRNAs (with default parameters). Python 3 (https://www.python.org/) was used in this project for data pre-processing. Cytoscape v3.7.2 [13] was used for visualizing molecular interaction networks. A specific Cytoscape plugin called MCODE was used for module analysis.

2.3 Annotation of miRNAs

Mature miRNA sequences were obtained from miRbase. Some of the miRNAs that had identical sequences but were assigned different identifiers under different experiments. The unambiguous annotation of the miRNAs was achieved with the help of assigning a new identifier (beginning with “osmir”) to all the 738 unique miRNAs.

2.4 Target Prediction

The plant small RNA target analysis server (psRNATarget), an online tool, was used to predict miRNA targets corresponding to the miRNAs retrieved from miRbase. The prediction analysis was carried out using the default scoring schema V2 option [12] and user submitted small RNA transcripts. The target cut off was set at 3 to increase the normal stringency.

2.5 Network Construction

2.5.1 Construction of Gene – Metabolic Pathway Network (gmpNet)

Considering genes and their corresponding pathways as nodes, gmpNet was constructed. Interaction of each gene with its corresponding metabolic pathways were treated as an edge or link between each node. No filter was applied for the construction of this genome-scale network of pathways and genes for rice as depicted in Fig. 1.

2.5.2 Construction of Pathway-Pathway Preliminary Network (p2Net)

While constructing the p2Net, an edge was initially established between a pair of pathways only if they shared at least one common gene between them. From the constructed network, 161803 pairwise edges were identified among 571 pathways. The number of associated common genes was assigned to be the edge-weight for each of the pathway-pathway pairwise interaction edges. This was followed by calculation of a correlation coefficient (CC value) for each pair of connected pathways. And then the CC values were normalized by computing the average of common genes between each pathway pair for 80000 randomly generated pathway-pathway networks which were made to have the same distribution as the original pathway-pathway interaction network. This was followed by assigning CC value for each pathway pair.

2.5.3 Construction of Pathway-Pathway Co-interaction Network (PCIN)

From the previous network, the pathway pairs which were found to have a CC value less than 1 (i.e., observed co-functionality was lesser than randomly generated coefficient) were filtered out and only the pairs with values of greater than one was kept for further analysis. Next, Gmin $>= 3$ (minimum number of common genes per pair) filter was applied to the network to obtain the final p2Net for further genome-scale analysis in rice metabolism. The final p2Net was found to have 1018 edges between 449 nodes (pathways) after applying the CC > 1 filter and the Gmin $>= 3$ filters. The same CC value threshold was implemented as used in network construction methods prescribed by Balaji et. al. [14]. The common set of gene(s) ($p1 \cap p2$) were extracted for each pathway pair (p 1 & p2) identified from the previously obtained network. Those pathway pairs were used further, which shared at least 3 common genes (Gmin $>= 3$) [14], and also had a CC > 1. Only the common genes recognized after this step were used for further analysis. A hypergeometric distribution model was implemented for the calculation of

Fig. 1. Network graph showing the gene and metabolic pathway network, where • are genes and • are pathways.

the probability P for the given p1 ∩ p2 as per Eq. 1.

$$P = 1 - F(T, G, L) = 1 - \sum_{t=0}^{n} \frac{\binom{T}{t}\binom{s-T}{L-t}}{\binom{s}{L}} \tag{1}$$

Where T is the total number of target-genes, G is the number of genes that are targeted by their respective miRNA and associated with metabolic pathways, L is the size of p1 ∩ p2, n is the number of genes in p1 ∩ p2 linked to miRNAs targeting the genes. This hypergeometric distribution was applied to the network using the inhouse developed tool, and for FDR correction a cut-off value <0.05 was used with the Benjamini and Hochberg method. If at least one miRNA was significantly involved with a pair of pathways, they were annotated as co-dependent or co-functional. All the pathway pairs identified using the prescribed method in this section were then assembled to construct a pathway-pathway co-interactional network (PCIN). In this network, a single node

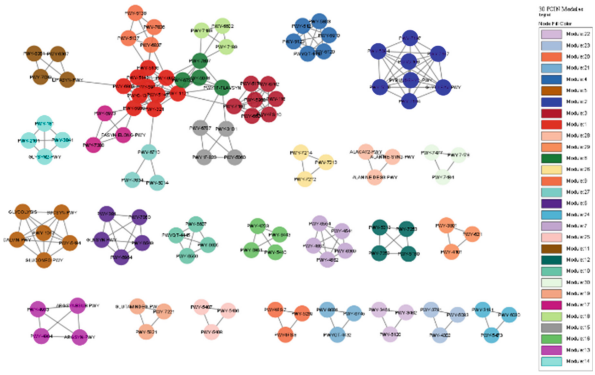

Fig. 2. Pathway-pathway co-interaction network (PCIN); the pathways in this network are considered to be nodes, and edges are drawn between a pair of pathways if there are at least 3 common associated genes between them, and the edge weight depends on the number of shared genes. If there are no common genes or less than 3, then no edge is considered to exist between them. Only pathways that belong to the 30 modules identified under Sect. 2.6 are labelled here.

represented a pathway, and two nodes were joined if the corresponding pair of pathways had a co-dependent relationship, otherwise no edge was considered.

2.6 Identification of Significant Modules from PCIN Using MCODE

All clusters from the PCIN network have been extracted using the MCODE plugin, in Cytoscape. MCODE defines a module as densely connected subgraphs in a given network based on topology. Modules from the PCIN, in this study have been defined as k-cliques, highly dense subgraphs with 'k' number of miRNAs where all miRNAs exhibit co-functional association with other miRNAs in the same subgraph. The MCODE plugin was run with the default parameters; (i) degree cut-off = 2, (ii) k-core value = 2, (iii) node score cut-off = 0.2, (iv) max. Depth = 100. Each of the modules that were obtained (Fig. 2) had a unique combination of pathways and none of the pathways or pathway pair were found to be present in more than one module. 125 pathways out of 449 pathways were found to be in the thirty modules extracted using MCODE, i.e., 127 miRNAs were found to be significantly co-functional. And the remaining 322 pathways were not found to have any associations with the corresponding gene sets targeted by miRNAs used in the present study.

3 Result and Discussion

3.1 Network Properties

3.1.1 Network Characteristics of the Gene – Metabolic Pathway Network (gmpNet)

With the interaction between genes and their corresponding metabolic pathways being considered as edges, gmpNet was constructed as described in Sect. 2.5.1. This network

had 4632 nodes (571 metabolic pathways and 4061 genes) and 7351 edges among them. Each gene had an average of ≈2 (1.81) associated metabolic pathways, while each metabolic pathway targeted an average of ≈13 (=12.87) genes. The frequency distribution of genes and their target metabolic pathways showed that a fairly large percentage (i.e., 94.6%) of genes (3843 out of 4061 genes) were involved in a comparatively lesser number of metabolic pathways (less than 4 pathways). While in the case of metabolic pathways, only about 4% of them targeted an excess of 50 genes in a single pathway (23 out of 571). Among all metabolic pathways in the network the one with maximum number of genes connected was, PWY-3781 (aerobic respiration I (cytochrome c)] with 310 genes associated with it.

3.1.2 Network Characteristics of the Preliminary Pathway-Pathway Co-interaction Network (p2Net)

At the beginning, the pathway-pathway network had 571 pathway nodes and 161803 edges between them. Then we considered edges will be drawn if three common genes exist between each pathway-pathway pair. Thus, the resultant p2Net contained 449 nodes and 1018 edges among them. In this network (p2Net), we found that each node (pathway) had an average of ≈5 (4.534) neighboring nodes (pathway) and an average clustering coefficient of 0.291.

3.1.3 Network Characteristics of the Pathway-Pathway Co-interaction Network (PCIN)

PCIN network has 203 nodes (pathways) from the previous network (p2Net) with 354 edges between them, after applying CC filter and application of a hypergeometric distribution. In this network too, it was found that each node (pathway) was connected to at least ≈4 (=3.487) neighboring nodes (pathways) and had an average clustering coefficient of 0.318 (Fig. 3).

3.2 PCIN Module Characteristics with Respect to Rice Metabolic Pathways

571 metabolic pathways of rice were retrieved from PMN, among which 125 pathways were found to be involved in the 30 modules extracted from the PCIN network using MCODE. The largest module comprises 10 pathways whereas the smallest module has only 3 pathways. The network further showed involvement of 119 reactions and 70 enzymes associated with those 125 metabolic pathways. From the thirty modules extracted from the PCIN network, the average pathways found for each module was ≈4 (=4.16), the average number of genes in each module was ≈3 (=2.73) and each module had an average of ≈4 (=3.93) miRNAs involved. The combined miRNA, gene, pathway, enzyme network for the 30 modules was found to have 460 nodes (68 miRNAs, 125 pathways, 77 genes, 119 reactions, and 70 enzymes) and 750 edges were found to exist among those nodes (Fig. 3).

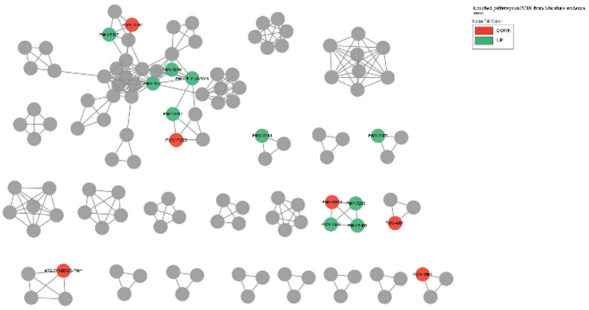

Fig. 3. Network graph representation of enriched pathways that overlap with PCIN pathways, based on DEGs identified for Indica vs Japonica seedling mRNA expression data. Pathways marked in green were found to be enriched for upregulated DEGs while pathways marked in red were enriched for downregulated DEGs.

3.3 PCIN Module Analysis

We have selected three modules randomly. The pathways, miRNAs and genes associated with these modules are represented in Table 1. From the figures (Fig. 4, 5, 6) it can be concluded that the pathways are highly interconnected through the genes which are regulated by the miRNAs.

3.4 Literature Validation

We used rice expression data available under accession GSE71925. The expression data in the study is extracted from 3 samples of Indica group and 3 samples of Japonica group of rice seedlings. To identify differentially expressed genes (DEGs), we used DESeq2 [15]. All genes that had log2FoldChange value more than 1 and less than −1 along with p-adjusted value of less than 0.05 were considered to be differentially expressed. Next, we performed an overrepresentation analysis with pathway annotation based on

Table 1. Modules and related pathways, miRNAs and genes

Modules	PATHWAYS	miRNAs	GENES
Module 3	PWY-5176, PWY-6762, PWY4FS-10, PWY-116, PWY-5968, PWY-6624, PWY-7186	osmir314, osmir598, osmir543, osmir683, osmir637,	Os12g0263800, Os05g0499800, Os01g0638600, Os01g0597800, Os09g0517900, Os01g0638000, Os03g0212000
Module 18	ALACAT2-PWY,ALANINE-DEG3-PWY, ALANINE-SYN2-PWY	osmir446, osmir505, osmir575, osmir517, osmir626, osmir126	Os07g0108300, Os12g0263800, Os10g0390600, Os10g0390500, Os07g0617800, Os03g0183600, Os09g0433900

(*continued*)

Table 1. (*continued*)

Modules	PATHWAYS	miRNAs	GENES
Module 24	PWY-7221, GLUTAMINDEG-PWY, PWY-5921	osmir554, osmir13, osmir483	Os12g0556600, Os11g0536800, Os03g0256400, Os06g0265000, Os04g0184500, Os02g0130100, Os05g0555600, Os03g0651000, Os11g0708500, Os03g0291500, Os01g0654100, Os02g0708100, Os04g0182900, Os11g0123033, Os10g0191600, Os05g0573300, Os01g0570700, Os04g0182875, Os04g0183300, Os06g0271300, Os11g0544800, Os01g0616900, Os05g0569700, Os07g0100200, Os12g0263800, Os04g0183500, Os04g0643200, Os01g0682001, Os10g0100700, Os05g0570400, Os10g0155400, Os07g0658400, Os04g0184100, Os08g0326600

the upregulated and downregulated DEGs using an online tool called WebGestalt [16] with custom Oryza sativa pathway database subset from the PMN database as described under Sect. 2.1.

Here we identified top 25 pathways each, sorted by p-values (less than 0.05) for upregulated and downregulated DEGs, respectively and overlapped it against the PCIN modules extracted in Sect. 2.6.

16 enriched pathways were found to overlap with the pathways identified in the PCIN modules, which again belonged to 10 different modules. Overlapping pathways are represented in Fig. 3.

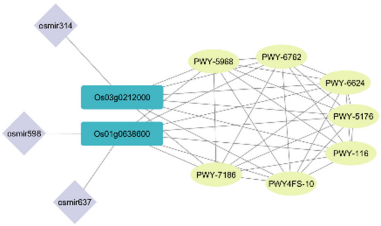

Fig. 4. Module 3.

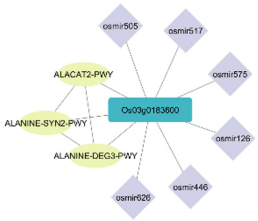

Fig. 5. Module 18.

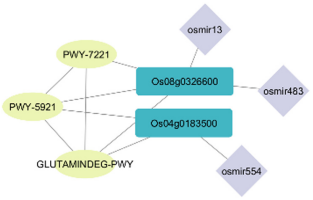

Fig. 6. Module 24.

Out of all the overlapping pathways 11 belong to secondary metabolite biosynthesis pathways of which 6 are flavonoid biosynthesis pathways. Given that the expression data is from rice seedlings, it makes sense that the flavonoid biosynthesis seem to be enriched. In plants, flavonoids have long been known to be synthesised in particular sites and are responsible for the colour and aroma of flowers, and in fruits to attract pollinators and consequently fruit dispersion to help in seed and spore germination, and the growth and development of seedlings [17].

Acknowledgments. CM gratefully acknowledges MAKAUT, WB for providing computational facilities.

Authors' Contributions. CM conceived the study, designed the experiments and edited the manuscript. AKB and CM performed the analysis using an in-house Python code. AKB and JM prepared the draft. All authors read and approved the final manuscript.

References

1. S. S. Gnanamanickam, An overview of progress in biological control. Biological Control of Rice Diseases, 2009, p. 43–51.
2. E. Bonnet, J. Wuyts, P. Rouzé, and Y. Van de Peer, Detection of 91 potential conserved plant microRNAs in Arabidopsis thaliana and *Oryza sativa* identifies important target genes. Proceedings of the National Academy of Sciences. 101(31), 2004, p. 11511–11516. DOI: https://doi.org/10.1073/pnas.0404025101
3. X. Xu, H. Bai, C. Liu, E. Chen, Q. Chen, J. Zhuang, and B. Shen, Genome-wide analysis of microRNAs and their target genes related to leaf senescence of rice. PLoS One. 9(12), 2014, p. e114313. DOI: https://doi.org/10.1371/journal.pone.0114313
4. S. Kansal, V. Panwar, R. D. Mutum, and S. Raghuvanshi, Investigations on Regulation of MicroRNAs in Rice Reveal [Ca2+] cyt Signal Transduction Regulated MicroRNAs. Frontiers in Plant Science, 2021, p. 2239. DOI: https://doi.org/10.3389/fpls.2021.720009
5. C. Mal, A. Deb, M. Aftabuddin, and S. Kundu, A network analysis of miRNA mediated gene regulation of rice: crosstalk among biological processes. Molecular BioSystems. 11(8), 2015, p. 2273–2280.
6. J. Hu, T. Zeng, Q. Xia, Q. Qian, C. Yang, Y. Ding, L. Chen, and W. Wang, Unravelling miRNA regulation in yield of rice (Oryza sativa) based on differential network model. Scientific reports. 8(1), 2018, p. 1–10. DOI: https://doi.org/10.1038/s41598-018-26438-w
7. Y. Pita-Juárez, G. Altschuler, S. Kariotis, W. Wei, K. Koler, C. Green, R. E. Tanzi, and W. Hide, The pathway coexpression network: revealing pathway relationships. PLoS computational biology. 14(3), 2018, p. e1006042. DOI: https://doi.org/10.1371/journal.pcbi.1006042
8. G. Minadakis, A. M.-P. Fuentes, G. Tsouloupas, I. Papatheodorou, and G. M. Spyrou, PathExNET: A tool for extracting pathway expression networks from gene expression statistics. Computational and Structural Biotechnology Journal. 19, 2021, p. 4336–4344. DOI: https://doi.org/10.1016/j.csbj.2021.07.033
9. C. Hawkins, D. Ginzburg, K. Zhao, W. Dwyer, B. Xue, A. Xu, S. Rice, B. Cole, S. Paley, and P. Karp, Plant Metabolic Network 15: A resource of genome-wide metabolism databases for 126 plants and algae. Journal of integrative plant biology. 63(11), 2021, p. 1888–1905. DOI: https://doi.org/10.1111/jipb.13163
10. A. Kozomara, M. Birgaoanu, and S. Griffiths-Jones, miRBase: from microRNA sequences to function. Nucleic acids research. 47(D1), 2019, p. D155-D162. DOI: https://doi.org/10.1093/nar/gky1141
11. Y. Kawahara, M. de la Bastide, J. P. Hamilton, H. Kanamori, W. R. McCombie, S. Ouyang, D. C. Schwartz, T. Tanaka, J. Wu, and S. Zhou, Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. Rice. 6(1), 2013, p. 1–10. DOI: https://doi.org/10.1186/1939-8433-6-4
12. X. Dai and P. X. Zhao, psRNATarget: a plant small RNA target analysis server. Nucleic acids research. 39(suppl_2), 2011, p. W155–W159. DOI: https://doi.org/10.1093/nar/gkr319
13. M. Kohl, S. Wiese, and B. Warscheid, Cytoscape: software for visualization and analysis of biological networks, in., Data mining in proteomics, Springer. 2011 p. 291–303. DOI: https://doi.org/10.1007/978-1-60761-987-1_18

14. S. Balaji, M. M. Babu, L. M. Iyer, N. M. Luscombe, and L. Aravind, Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast. Journal of molecular biology. 360(1), 2006, p. 213–227. DOI: https://doi.org/10.1016/j.jmb.2006.04.029
15. M. I. Love, W. Huber, and S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 2014, p. 550. DOI: https://doi.org/10.1186/s13059-014-0550-8
16. Y. Liao, J. Wang, E. J. Jaehnig, Z. Shi, and B. Zhang, WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic acids research. 47(W1), 2019, p. W199–W205. DOI: https://doi.org/10.1093/nar/gkz401
17. R. Griesbach, Biochemistry and genetics of flower color. Plant breeding reviews. 25, 2010, p. 89–114.