International Journal of Computational Bioinformatics and In Silico Modeling
Vol. 2, No. 3 (2013): 123-127
Research Article
Open Access

ISSN: 2320-0634

Regulatory Motifs in precursors of miR156 family in plants

Sayak Ganguli*, Pritha Mukhopadhyay, Sohini Gupta and Abhijit Datta

DBT Centre for Bioinformatics, Presidency University, Kolkata

*Corresponding author: Sayak Ganguli, e-mail: sayakbif@yahoo.com

Received: 09 April 2013 Accepted: 23 April 2013 Online: 01 May 2013

ABSTRACT

Understanding the basic mechanisms of micro RNA control in the cell has been one of the open ended problems in this field of study. Though next generation sequencing has helped us unearth a large number of candidate microRNAs and their possible targets, we still do not know the exact mechanisms by which the production of these microRNAs is actually controlled. MicroRNAs in plants have been well documented to be involved in developmental processes and their expression levels have also been sketched throughout the life cycle of a plant. The upregulation and downregulation of these micro RNA entities provides clear indication that a regulatory network exists for the control of their differential expression. This work focuses on the identification of the regulatory motifs in the precursor sequences using position specific weight matrix based pattern classification algorithm. The candidate microRNA family used for this study is mir156 family which has been reported in a large number of plant species in monocotyledons as well as dicotyledons. Results indicate that regulatory motifs are present in large numbers at the DNA as well as RNA levels. Few motifs were found to be overrepresented.

Keywords: Position Specific weight matrix, regulatory motifs, microRNA, mir156 family

INTRODUCTION

Plant development can be distinctly divided in phases such as vegetative growth, a reproductive phase and finally seed set and senescence. The shifts between these phases are under coordinated control systems where multiple genes and proteins form distinct circuits that integrate endogenous and environmental signals. In recent years, however, it has become evident that the genetic networks that underlie these phase transitions share some common factors. Non coding RNAs have been found to be one of the most important regulators of plant development [1] and recent advances in the field of plant phase transitions, have highlighted the role of two microRNAs – miR156 and miR172 – and their respective targets during these transitions. miR156 represents one of the most conserved microRNA families in plants and sequencing experiments have identified their presence in the following plants:

Arachis hypogea, Boechera stricta, Brassica napus, Brassica oleracea, Brassica rapa, Bruguiera gymnorhiza, Citrus paradisi, Poncirus trifoliate, Euphorbia esula, Fragaria vesca, Gossypium hirsutum, Gossypium raimondii, Glycine max, Helianthus annuus, Ipomoea nil, Lycopersicon esculentum, Lactuca sativa, Lotus japonicus, Malus domestica, Medicago truncatula, Nicotiana tabacum, Oryza australiensis, Oryza brachyanth, Oryza punctata, Oryza ridleyi, Oryza rufipogon, Oryza sativa, Populus trichocarpa, Poncirus trifoliata, Prunus persica, Solanum tuberosum, Sorghum bicolor, Triticum aestivum, and Zea mays. Three important strategies have been utilized for the identification of plant miRNAs – direct cloning from small RNA libraries, computational prediction by genome analyses and mutational screening for altered phenotypes. Recent miRNA databases such as mirBase harbor over 5000 sequences of microRNAs. Plant miRNAs have large and varied stem loop structures and interact with their target sites with near perfect complementarity. Many miRNA genes have been identified to be conserved in rice and *Arabidopsis thaliana*, indicative of the fact that they had originated prior to the divergence of monocots and dicots. This is supported by the fact class III homeodomain-leucine zipper (HD-ZIP III) binding sites on microRNA genes are conserved from bryophytes to seed plants. The

main interacting partners that are of importance in microRNA biogenesis pathways are Dicer and Argonaute proteins which have been reported to have specific interaction sites [2, 3]. miR156 is one of the most abundant miRNAs in plants, which is at its highest expression during the seedling stage [4,5]. Schwab et.al. (2005) [6] has reported a moderate delay in flowering along with the production of juvenile features in leaves when miR156 has been constitutively overexpressed [7]. Since miR156 has been reported to be ubiquitously expressed, it is important to understand the regulatory modules that are present in its sequences so that they can be correlated with the expression levels of the transcription factors that form an important lineage of genes which are also ubiquitously expressed in most plant developmental cascades. Recent deep sequencing experiments have identified that in plants the amount of expression of non-conserved small RNAs are greater than in other species and thus a proper understanding of the DNA and RNA regulatory modules should be important steps in the analyses of the developmental cascades of plants in greater detail.

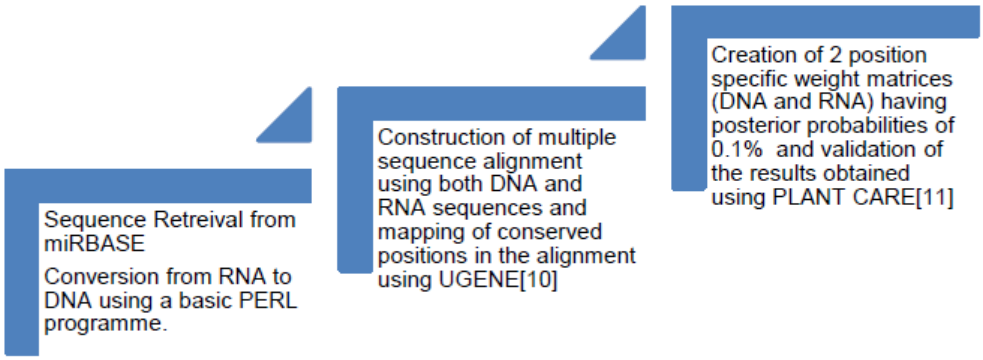

Figure 1. Workflow followed for the analyses

MATERIALS AND METHODS

Available mir156 sequences both mature and precursor sequences were retrieved from mirBASE and initially the precursor sequences were converted to their complimentary DNA sequences using a basic PERL program. Thus now there were two sets of sequences one in the DNA form and the other in the database format with ribonucleotides. A multiple sequence alignment was constructed using UGENE for both the sequence files and the conserved positions were identified. Following the construction of multiple sequence alignment, the DNA and RNA sequences were subjected to two position specific weight matrices, constructed according to the methods previously described by Sandelin (2004) [8] and Wasserman (2004) [9] one for DNA and the other for RNA having a posterior probability of 0.1% so that the false positive results could be minimized. The results that were obtained through conserved sequence mapping were then correlated with the weight matrix results. Finally the results were validated using the Plant Care database which is a repository for conserved plant regulatory motifs. The outline of the workflow is described in Figure 1.

RESULTS AND DISCUSSION

The results of the analyses clearly indicate the presence of regulatory elements which are cis acting in nature in the DNA sequences of the precursors and when RNA sequence regulatory elements were analyzed a large number of such elements were identified in those precursor sequences as well. This leads us to comment that our understanding of the myriad mechanisms controlling the biogenesis of microRNAs is still in its nascent stage as each microRNA family may possess specific modular regulatory elements in both the DNA and RNA levels indicative of a failsafe loop which makes the biogenesis specific and reduces the error rate in the production of the mature functional microRNAs.

The phylogenetic analyses performed revealed 24 highly conserved positions in the alignment that was generated: Highly conserved indicates that out of the 165 sequences under study the positions are conserved in almost each of the sequences except the ones mentioned in the comment column.

Table 1. Phylogenetically conserved positions in the sequences under study (The rows highlighted in the table represent positions which are part of the regulatory sites in the precursor sequences)

Serial	Position(s)	Number	Residue/motif	Comment
1	321	1	U	Absent in aqcmir156b and aqcmir529 substituted by 'C' in athmir156g
2	323 - 331	8	GACAGAAG	HIGLY CONSERVED
3	346 - 352	7	GAGCACA	HIGLY CONSERVED
4	490	1	G	HIGLY CONSERVED
5	491	1	U	HIGLY CONSERVED
6	492	1	G	Substituted by 'U' in mtrmir156
7	493	1	C	Substituted by 'U' in mtrmir156
8	494	1	U	HIGHLY CONSERVED
9	510	1	U	Substituted by 'A' in mir156a and b and by 'C' in mtrmir156
10	511 - 512	2	CA	'A' substituted by U in mtrmir156 and hvumir156

Phylogenetically conserved positions identified through multiple sequence alignment and the fact that these conserved positions form intricate part of the regulatory elements further signify the assumption that the regulatory modules are indeed functional and hence have retained their position specificity throughout the events of divergence and occurrence of this particular family of micro RNA in plants.

SERIAL NUMBER	REGULATORY MOTIF	FUNCTION	FREQUENCY OF OCCURRENCE
1	TATA BOX	CORE PROMOTER; ELEMENT FOUND FROM -50 TO -20 OF THE TSS	219
2	UNNAMED 4	UNKNOWN BUT REPORTED TO ASSOCIATE WITH ANTISENSE SEQUENCE ACTIVITY	142
3	5'UTR PY RICH STRETCH	REGION IN THE 5'UTR CONFERRING HIGH TRANSCRIPTION LEVELS WITHOUT THE NEED FOR OTHER UPSTREAM CIS ELEMENTS EXCEPT FOR A TATA-BOX	94
4	G BOX	CIS-ACTING ELEMENT INVOLVED IN LIGHT RESPONSIVENESS	41
5	CAAT BOX	PROMOTER ELEMENT FOUND IN THE UPSTREAM REGION AND IS REPORTED TO BE ASSOCIATED WITH ENHANCER FUNCTION	41
6	MBS	MYB BINDING SITE INVOLVED IN DROUGHT INDUCIBILITY	36
7	GAG MOTIF	PART OF A LIGHT RESPONSIVE ELEMENT	12

Figure 1. Overrepresented DNA regulatory motifs in the sequences and their descriptions

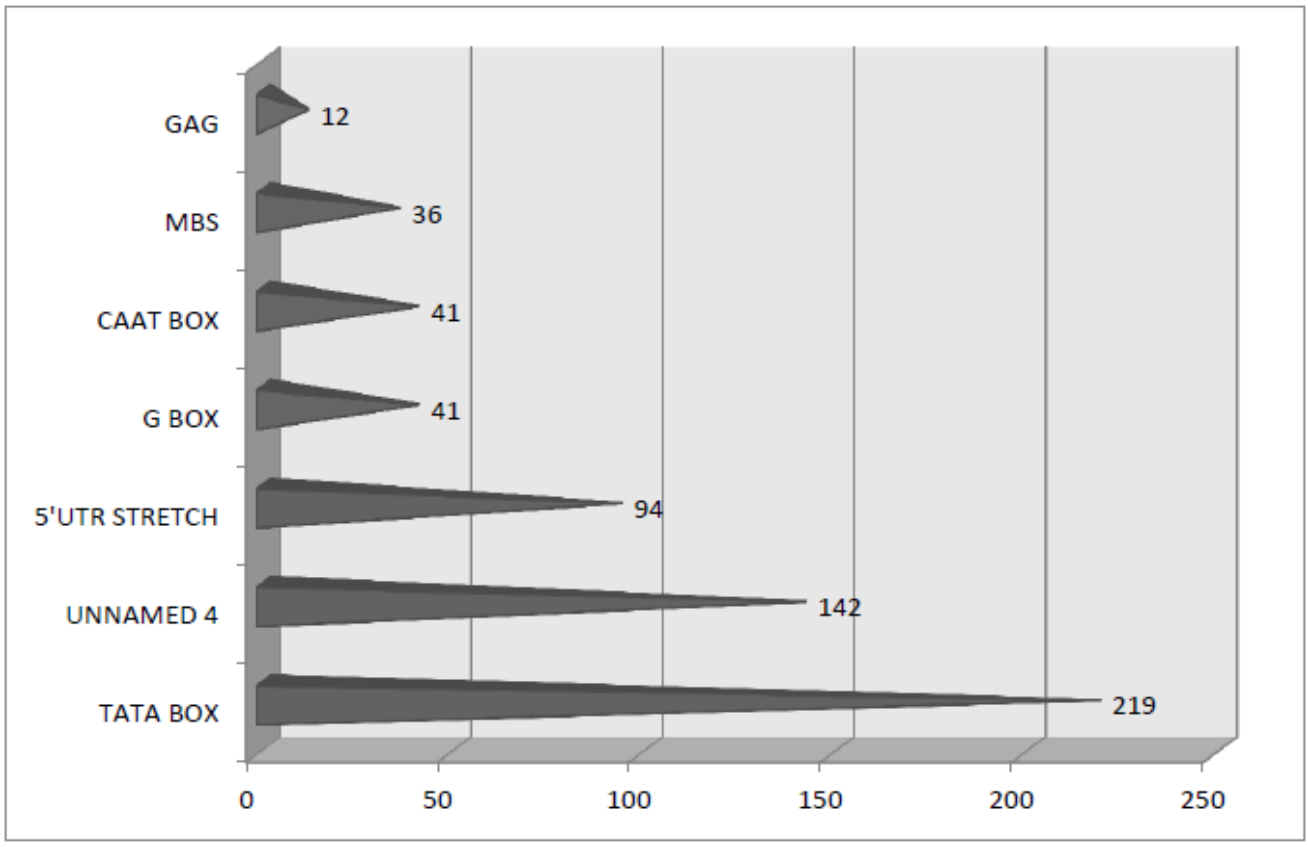

Figure 2. Graphical Representation of the overrepresented DNA regulatory motifs

REGULATORY MOTIF	DESCRIPTION
Terminal Oligopyrimidine Tract (TOP)	This motif begins with a C and is followed by a G residue (consensus sequence C(Py)nG where n=3-14 pyrimidines); otherwise, no other requirements appear to be necessary except for consecutive pyrimidines.
Upstream Open Reading Frame (uORF)	Among the cis-elements that play a role in translation regulation are upstream open reading frames (uORFs) located in the 5'UTR of mRNA.
Cytoplasmic polyadenylation element (CPE)	Cytoplasmic polyadenylation
K-Box (KB)	The K box (cTGTGATa) is present in one or more copies in many of these 3'UTRs and mediates negative post-transcriptional regulation.
Exonic Splicing Enhancer	WCWWC motif - involved in regulation of exonic splicing by interacting with an enhancer protein
Exonic Splicing Enhancer	GRYYMCYR motif - involved in regulation of exonic splicing by interacting with an enhancer protein
Exonic Splicing Enhancer	WGGACRA motif - involved in regulation of exonic splicing by interacting with an enhancer protein
Exonic Splicing Enhancer	CRSMSGW- involved in regulation of exonic splicing by interacting with an enhancer protein
Exonic Splicing Enhancer	YRCRKM -- involved in regulation of exonic splicing by interacting with an enhancer protein
Exon silencer	TTAG motif - involved in the binding of regulatory proteins such as hnRNPA1 to silence the exons.
Exonic Splicing Silencer	GAAGAAGA motif - involved in the binding of regulatory proteins such as hnRNPA1 to silence the exons.
Exon silencer	CAAGG motif - involved in the binding of regulatory proteins such as hnRNPA1 to silence the exons.
Exon silencer	TGGT motif – binds multiple regulatory proteins and prevents assembly of splicing apparatus
Intron enhancer	TGCATG motif – involved in regulation of the alternative splicing pathway
Intron enhancer	YCAY – motif involved in regulation of the alternative splicing pathway
EGFR (epidermal growth factor receptor)	TCCGCT motif – involved in binding of human epidermal growth factor
CRY-delta1	CACCTA motif – involved in regulation of delta 1 crystallin receptor

Figure 3. Overrepresented RNA regulatory elements in the sequences and their descriptions

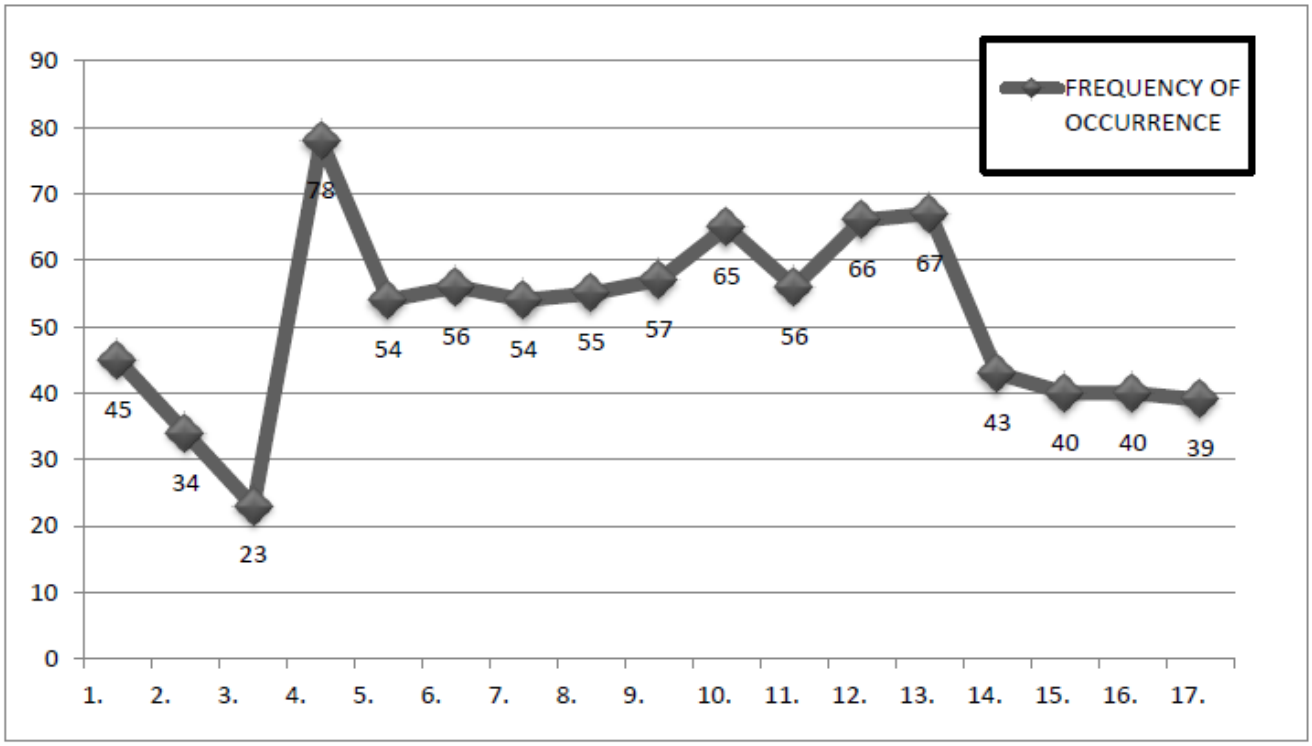

Figure 4. Graphical Representation of the overrepresented RNA regulatory motifs

Several regulatory controls are evident if we take a look at the differential expression of the microRNAs of this family. Kozomara and Griffiths-Jones, (2011) [12] have shown that many miR156 isoforms such as *MIR156a-f*, *MIR156g-h* and *MIR157a-d* can be encoded by multiple loci in *Arabidopsis.* When miR156 mimicry targets (*MIM156*) have been overexpressed to perturb endogenous miR156 function, those transgenics exhibit adult features in organs after flowering. This clearly points out to the fact that miR156 is a necessity for the expression of the juvenile phase. Low temperatures may positively influence mildly a few *MIR156* loci but miR156 accumulation has not been observed to be affected by any of the known ambient temperature pathway components [13]. Yang et al (2011) [11] have shown that leaves may often be a source of mir156 repressing factor using organ ablation; but still the nature of the regulatory factors which control the expression of miR156 remains a mystery.

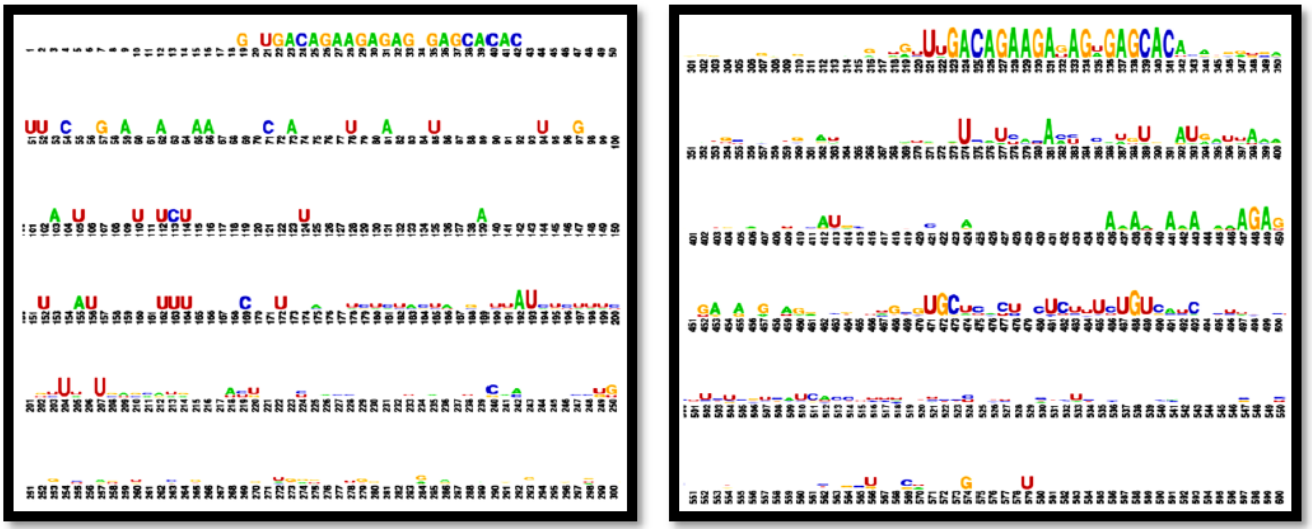

Figure 5. Positional conservedness of nucleotides in the mir156 family sequences.

CONCLUSIONS

A large number of regulatory sequence motifs were identified at the DNA and the RNA levels and were shown to be phylogenetically conserved. These data indicate towards the fact that mir156 expression can be under the influence of a complex regulatory control circuit which may involve both transcriptional and post transcriptional regulatory events.

Acknowledgements

The authors acknowledge the contribution of the BTBI BTIS net initiative of the Department of Biotechnology, Government of India.

REFERENCES

1. Chen X. MicroRNA biogenesis and function in plants. FEBS Lett 2005; 579: 5923-31.
2. Ganguli S, De M and Datta A. 2011a. Analyses of Argonaute–MicroRNA interactions in *Zea mays* International Journal of Computational Biology, (2)1 32 – 34
3. Ganguli. S. and Datta. A. 2011b RNAi – Interactomics and Therapeutics In: RNAi Technology Edited By: V.K. Gupta et.al. CRC Press (Taylor & Francis Group), PAGES 347 – 356 (Chapter 19)
4. Fahlgren N, Howell MD, Kasschau KD, *et al.* High-Throughput Sequencing of Arabidopsis microRNAs: Evidence for Frequent Birth and Death of MIRNA Genes. PLoS ONE 2007; 2: e219.
5. Wang, J. W., Czech, B. and Weigel, D. (2009). miR156-regulated SPL transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana*. *Cell* 138, 738-749.
6. Schwab R, Ossowski S, Riester M, Warthmann N, Weigel D. Highly specific gene silencing by artificial microRNAs in Arabidopsis. Plant Cell 2006; 18: 1121-33.
7. Wu G, Poethig RS. Temporal regulation of shoot development in Arabidopsis thaliana by miR156 and its target SPL3. Development 2006; 133: 3539-47.
8. Sandelin A, Wasserman W.W. 2004. Constrained binding site diversity within families of transcription factors enhances pattern discovery bioinformatics J Mol Biol. 338(2):207-15.
9. Wasserman W.W. and Sandelin A 2004 "Applied Bioinformatics for the Identification of Regulatory Elements" Nat. Rev. Gen (5) 276 – 287
10. Okonechnikov K, Golosova O, Fursov M and the UGENE team 2012. Unipro UGENE: a unified bioinformatics toolkit doi:10.1093/bioinformatics/bts091
11. Lescot M, Dhais P, Thijs G, Marchal K, Moreau Y, Peer Y, Rouz P and Rombauts S 2002 PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences *Nucleic Acids Res. 2002 Jan 1;30(1):325-327.*
12. Kozomara, A. and Griffiths-Jones, S. (2011). miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* 39, D152-D157.
13. Lee, H., Yoo, S. J., Lee, J. H., Kim, W., Yoo, S. K., Fitzgerald, H., Carrington, J. C. and Ahn, J. H. (2010). Genetic framework for flowering-time regulation by ambient temperature-responsive miRNAs in *Arabidopsis*. *Nucleic Acids Res.* 38, 3081-3093.
14. Yang, L., Conway, S. R. and Poethig, R. S. (2011). Vegetative phase change is mediated by a leaf-derived signal that represses the transcription of miR156. *Development* 138, 245-249.

*****