Cover image:

Edith Hunsberger, *“Corny Attitude: Every field of corn hides millions of amazing sculptural views.”* Painting - Acrylic On Board, 2012.
http://fineartamerica.com/featured/corny-attitude-edith-hunsberger.html

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

DOCTORATE SCHOOL OF CROP SCIENCE

CURRICULUM AGROBIOTECHNOLOGY – CYCLE XXVII

DEPARTMENT OF AGRONOMY, FOOD, NATURAL RESOURCES,
ANIMALS AND ENVIRONMENT

EXAMINING THE INFLUENCE OF THE ENDOGENOUS SMALL RNAS ON GENE EXPRESSION AND GENOME STABILITY IN THE MAIZE LEAF

Direttore della Scuola : Ch.mo Prof. Antonio Berti

Supervisore : Ch.mo Prof.ssa Serena Varotto

Dottoranda : Alice Lunardon

DATA CONSEGNA TESI

2 febbraio 2015

Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Alice Lunardon February, 2nd 2015

A copy of the thesis will be available at
http://paduaresearch.cab.unipd.it/dottorato/

Dichiarazione

Con la presente affermo che questa tesi è frutto del mio lavoro e che, per quanto io ne sia a conoscenza, non contiene materiale precedentemente pubblicato o scritto da un'altra persona né materiale che è stato utilizzato per l'ottenimento di qualunque altro titolo o diploma dell'università o altro istituto di apprendimento, a eccezione del caso in cui ciò venga riconosciuto nel testo.

Alice Lunardon 2 febbraio 2015

Una copia della tesi sarà disponibile presso
http://paduaresearch.cab.unipd.it/dottorato/

“Transposable elements and the transposases they encode underlie the evolvability of higher eukaryotes’ massive, messy genomes.”

Nina V. Fedoroff

Acknowledgments

I would like to express a huge “thank you” to my advisor Professor Serena Varotto (Department of Agronomy Animal Food Natural Resources and Environment, University of Padova), for guiding me over these years. The door of Prof. Varotto was always open whenever I had questions about my research and I needed emotional support. She allowed me realize my ideas and helped me develop critical thinking in my research. I must thank her for the time she spent with me during journal clubs: I have learnt so much discussing with her about science. I would like to thank her for encouraging me participating to many conferences and workshops that have been important for my scientific training and where I have met inspiring scientists. I will always keep in my mind all your precious suggestions and advises.

I would like to thank Professor Michael Axtell (Department of Biology and Huck Institutes of the Life Sciences, Pennsylvania State University), for allowing me working for six months in his lab. He provided me fundamental help for my research. He had time for answering to all my questions and he taught me more than what he thinks. In his lab I have experienced a collaborative scientific method and I have learnt to be more efficient and productive in my research. I would like to thank Ceyda, Saima, Qikun, Feng and Seth for their scientific help and their friendship.

I’d like to thank all the people that helped me in my work: Doctor Emanuele De Paoli (Dipartimento di Scienze Agrarie Ed Ambientali, University of Udine) and Doctor Cristian Del Fabbro (Department of Math and Computer Science, University of Udine), for introducing me to the bioinformatics world and giving me important directions at the beginning of my research. Emanuela Aleo (Istituto di Genomica Applicata, Udine) for her willingness to always answering to my e-mails. Doctor Thomas Hardcastle (Department of Plant Sciences, University of Cambridge) for his collaboration at the beginning of my research.

I would like to thank all the people working at the Department of Agronomy Animal Food Natural Resources and Environment of the University of Padova, especially the guys I share the office with: Giulio Galla, Alessandro Vannozzi, Mirko Volpato, for being so nice with me, helping me with my PhD experience and making me smiling a lot! Thanks to my colleagues for making me a stronger person. Thanks to Sara Balzan, for our American skype conversations, we both experienced the cold of a freezy American winter.

A special thanks to my family: Mum, Dad, my sister Erika and my Grandma Gianna for their enormous help, patience, love, support and inspiration.
Thanks to the new entry in my life Alessio.
Thanks to Fry, I have never been alone during the writing of this thesis ☺

Index

Abstract

Small non-coding RNAs are widespread in all kingdoms of life (Michaux et al. 2014) where they participate in RNA-mediated silencing pathways to regulate and fine-tune gene expression, through transcriptional gene silencing (TGS) and post-transcriptional gene silencing (PTGS) mechanisms. Not all mechanisms of RNA interference (RNAi) are conserved among organisms, which is true for example for the TGS pathway termed RNA-directed DNA methylation (RdDM). RdDM occurs in the nucleus to repress target genes at the transcriptional level, it is an epigenetic pathway because it does not alter the DNA sequence but instead causes gene expression variation by small RNA-guided modifications of chromatin, for example cytosine methylation and histone modifications. In plants RdDM is unique among small RNA-mediated chromatin modifications because it depends on two plant-specific RNA polymerase enzymes called Pol IV and Pol V (Matzke and Mosher 2014). This increases the complexity of RNAi mechanisms in plants, which have been investigated for a large amount of studies in the model species *Arabidopsis thaliana* (hereafter referred to as *Arabidopsis*). Small RNAs (sRNAs) and RNAi mechanisms play fundamental roles in many biological processes; in particular, their observed participation in the phenomena of hybrid vigor, stress-response and formation of epialleles makes them an important source of growth in crop production. *Arabidopsis* shows many differences in genome size, structure and dynamics compared to crops, therefore it is necessary, and challenging, to transfer the knowledge acquired in this model plant to crop species (Mirouze and Vitte 2014). Maize is one of the most important food and feed crops in the world and has a wide range of industrial applications as well. The maize genome has unique characteristics, such as the unusual number of well-characterized active transposable elements (Lisch D 2012), which are the main targets of RdDM. For these reasons it is of particular importance the research aimed to expand our knowledge on how sRNAs control genome activity in maize.

This is the general background to this PhD project, whose aim was to characterize the endogenous sRNA population of maize leaf in terms of genomic annotation and abundance, to further examine its influence on gene expression and its response to abiotic stresses. To analyse the sRNA control of gene expression, in addition to wild type maize plants, the *rmr6-1* mutant was also studied: impaired in Pol IV function this mutant is characterized by the absence of siRNAs participating in RdDM that require Pol IV for their biogenesis (Erhard et al. 2009). The absence of Pol IV-dependent siRNAs allowed testing what was their impact on genome stability. The sRNA population was characterized through the analysis of sRNA-seq data obtained from our samples. Gene annotation and expression level in wt and mutant plants was retrieved from the analysis of total RNA-seq data obtained by our laboratory from the same samples. To assess the role of sRNAs in stress response we examined the sRNA population of wild type and *rmr6-1* mutant plants subjected to abiotic stresses. The abiotic stresses studied were field-mimicked conditions of drought, salinity and the combination of the two, drought plus salinity, because these are the most crucial abiotic stresses that limit the production of the world crops (Munns R 2011). In particular, salinization constitutes a problem also in Mediterranean coastal areas (Flowers TJ 2004) and, considering the region of Veneto, in the coastal soils of the Venice Lagoon (Carbognin and Tosi 2003).

The PhD started with the collaborative project between the laboratories of Prof.ssa S. Varotto, Prof. F. Morari and Dr. F. Meggio. The aim of the project was to set up a reproducible protocol for the application of drought and salinity conditions to maize plants that was agronomically realistic and representative of field stress conditions. To mimic field progressive stress conditions, drought, salinity and the combination of the two, drought plus salinity, were applied to plants progressively for ten days and the stress response was evaluated at different time points during the stress application. In field conditions after a period of stress, environmental conditions usually turn more favourable, therefore after ten days of treatment the stresses were removed and plants were grown in optimal conditions to test their recovery capacity. Two different lines were studied: the stress-sensitive inbred line B73 and a stress-resistant F1 commercial hybrid. At the time points of stress application and recovery from the stress, plants

responses were analysed with agronomic, physiological and genetic parameters. Agronomic parameters were evaluated by the laboratory of Prof. F. Morari and physiological parameters by Dr. F. Meggio. Our collaboration consisted in the study of the genetic responses of plants. In particular, literature was investigated to identify a set of genes known to be differentially expressed (DE) by stress or belonging to the main pathways involved in abiotic stress response and their transcript level was analysed in our experiment using real time quantitative PCR (qRT-PCR). All the analysed parameters confirmed that the applied treatments were effective in inducing a stress condition in plants. Therefore our stress protocol represents a valid tool for further studies concerning the stress response in maize, which retain their value under field conditions, thus increasing the result translatability for crop improvement. The combination of the examined agronomic, physiological and genetic parameters allowed gaining insights into the mechanisms regulating the different tolerance to the stress of the stress-sensitive and stress-resistant lines.

The main work of the PhD project was dedicated to the analysis of sRNA-seq data obtained from wt and *rmr6-1* mutant plants, to characterize the endogenous sRNA population of the maize leaf and investigate its effect on gene expression and its stress response. 48 sRNA-seq libraries were sequenced from leaf samples of wt and mutant plants, in control conditions or subjected to abiotic stresses and after the recovery from the stresses. Reads from each library were pre-processed and the quality of the clean reads was verified. Reads were then mapped to the reference maize B73 genome, revealing the typical maize sRNA population profile with the highest abundance of 24-nt sRNAs, followed by the 22-nt and the 21-nt sRNAs. The bioinformatics tools ShortStack was used to *de novo* identify the maize genome loci responsible for a significant production of sRNAs in the leaf, starting from the merged set of sRNAs of the 48 samples. The identified *MIRNA* loci were examined first. We found differences between our microRNA annotation and that reported in miRBase that might reflect inaccurate annotation in miRBase or leaf-specific differences in *MIRNA* processing patterns. The prediction of the microRNA targets was performed on the transcripts annotated in the transcriptome assembly reconstructed from RNA-seq. This allowed identifying a newly annotated transcript as target of a conserved microRNA, helping

elucidating the role of this microRNA in maize. Putative novel microRNAs were identified: a number of them had characteristics of bona fide microRNAs while others appeared to be new 'proto-miRNAs' or instead siRNAs. The other identified sRNA loci categories were analysed in terms of co-occupancy with protein-coding genes, transposon and long non-coding RNA (lncRNA) transcripts. A significant enrichment of the loci predominated by the production of 24-nt sRNAs was found in the flanking regions of all the analysed set of genes. In particular, expressed genes were flanked by sRNA loci of 24-nt size class with higher frequencies compared to the non-expressed genes. In the *rmr6-1* mutant, despite the dramatic loss of siRNAs observed mainly in gene flanking regions, the number of DE genes compared to wt was 1013 and the downregulation of an sRNA locus was not generally sufficient not even necessary to predict the up or downregulation of its close gene. Therefore, the absence of siRNAs had little impact on the genome stability of the maize leaf, indeed leaves of mutant plants did not have morphological defects and were identical to those of wt plants. The mechanisms that maintain gene silencing when siRNAs are lost and thus RdDM control of gene expression is impaired still remains to be elucidated. Literature data show evidences that the RdDM pathway might be essential to ensure the transgenerational transmission of the epigenetic information. In this hypothesis, to elucidate the role of siRNAs in the control of gene expression it would be helpful to study the activity of siRNAs and the effects of RdDM mutations in other cell types such as the gametes. Alternatively, it would be helpful to study epigenetic changes of gene expression in multiple generations of plants. The absence of siRNAs, although it was not found to compromise the genome stability in the leaf, did have some effects on gene expression that appeared to be secondary effects of the mutation. In particular, in the *rmr6-1* mutant it was registered the upregulation of stress-responsive genes and cytochromes and the downregulation of genes involved in the regulation of cell cycle and genes encoding core histone proteins. Finally, the sRNA stress response was examined. We applied the stress protocol previously set up and found a few numbers of miRNAs and sRNA loci of the other categories that were DE in stress conditions. Although the DE sRNAs were less numerous compared to previous works assessing the sRNA stress

response in crops, they might be better candidates for stress-tolerance studies because they were found to be DE during stresses mimicking field conditions.

Published works cited here are reported in the ‘References’ section of Chapter 2.

Riassunto

I piccoli RNA non codificanti sono stati riscontrati in tutti i regni della vita (Michaux et al. 2014). Essi partecipano ai meccanismi di regolazione genica di silenziamento del DNA mediato da RNA, che si distinguono in meccanismi di silenziamento genico trascrizionale (TGS) e post-trascrizionale (PTGS). Non tutti questi pathway sono conservati negli organismi, come ad esempio il meccanismo chiamato di metilazione del DNA RNA-dipendente (in inglese 'RNA-directed DNA methylation', RdDM). Esso avviene nel nucleo, dove induce la repressione delle sequenze target a livello trascrizionale. Il pathway RdDM è un esempio di meccanismo epigenetico di controllo dell'espressione genica, in quanto la variazione di espressione viene indotta senza alterazioni di sequenza del DNA, attraverso modificazioni della cromatina guidate dall'azione dei piccoli RNA, come ad esempio la metilazione delle citosine o le modifiche istoniche. Nelle piante il pathway RdDM prevede l'azione di due RNA polimerasi specifiche del regno vegetale, l'RNA polimerasi IV (Pol IV) e l'RNA polimerasi V (Pol V) (Matzke and Mosher 2014). La specificità di questi enzimi riservata al regno vegetale è indice che le piante hanno evoluto un livello aggiuntivo di complessità dei meccanismi di silenziamento del DNA RNA-dipendenti, che sono stati studiati soprattutto nella pianta modello *Arabidopsis thaliana* (abbreviata d'ora in poi con il nome *Arabidopsis*). I piccoli RNA e i meccanismi di silenziamento del DNA RNA-dipendenti ricoprono ruoli fondamentali in diversi processi biologici. In particolare, il loro coinvolgimento nei fenomeni quali il vigore dell'ibrido, la risposta allo stress e la formazione di epialleli li rende un'importante fonte di studio al fine del miglioramento delle piante da coltivazione. Il genoma della pianta *Arabidopsis* presenta molteplici differenze in termini di dimensione, struttura ed organizzazione dinamica rispetto ai genomi delle piante da coltivazione. Queste differenze sostanziali rendono necessario, ma anche difficoltoso, il trasferimento delle conoscenze acquisite in *Arabidopsis* da questa pianta modello alle piante da coltivazione (Mirouze and Vitte 2014). Il mais è una delle più importanti

coltivazioni a livello mondiale per la produzione di alimenti e mangimi e viene utilizzato in diverse catene industriali. Il suo genoma possiede caratteristiche uniche, come ad esempio la presenza di un inusuale elevato numero di elementi trasponibili attivi (Lisch D 2012), che sono i principali target del pathway RdDM. Per queste ragioni è di particolare importanza la ricerca scientifica volta ad aumentare la conoscenza dei meccanismi di controllo dell'attività del genoma di mais guidati dai piccoli RNA.

L'attività del progetto di Dottorato si inserisce all'interno di questo quadro di ricerca. Il principale scopo del progetto è stato la caratterizzazione della popolazione di piccoli RNA endogeni in foglia di mais, in termini di annotazione genomica e abbondanza, che ha permesso poi di valutare gli effetti dei piccoli RNA sull'espressione genica e la loro risposta a stress abiotici. Al fine di indagare il controllo esercitato dai piccoli RNA sull'espressione genica sono state studiate due linee di mais, la linea inbred B73 e il mutante *rmr6-1*. Questo mutante presenta una forma non funzionale della Pol IV che provoca la mancata produzione dei piccoli RNA che partecipano al pathway RdDM e che dipendono dalla Pol IV per la loro biogenesi, i quali sono chiamati siRNA, dall'inglese 'small interfering RNA' (Erhard et al. 2009). L'assenza dei siRNA ha permesso di valutarne gli effetti sulla stabilità del genoma. La popolazione dei piccoli RNA è stata caratterizzata attraverso l'analisi di dati di sequenziamento di piccoli RNA ottenuti dai nostri campioni. L'annotazione dei geni e i loro livelli di espressione sono stati ottenuti nel nostro laboratorio attraverso l'analisi di dati di sequenziamento di RNA totale proveniente dagli stessi campioni. Al fine di valutare il ruolo dei piccoli RNA nella risposta allo stress la loro espressione è stata analizzata in piante wild type e mutanti sottoposte a stress abiotici. I protocolli di stress utilizzati sono stati trattamenti che mimano gli episodi di stress idrico, salino e la combinazione dei due, idrico più salino, che si verificano in condizioni di campo. Sono stati scelti questi stress abiotici in quanto sono le tipologie di stress più frequenti che abbassano le rese della produzione delle piante da coltivazione a livello mondiale (Munns R 2011). In particolare, la salinizzazione costituisce un problema anche nelle zone costiere del Mediterraneo (Flowers TJ 2004) e, a livello della regione Veneto, nei suoli costieri della laguna di Venezia (Carbognin and Tosi 2003).

Il lavoro del Dottorato è iniziato con la partecipazione ad un progetto di collaborazione tra i laboratori della Prof.ssa S. Varotto, del Prof. F. Morari e del Dr. F. Meggio. Lo scopo del progetto è stato la messa a punto di un protocollo riproducibile per l'applicazione di stress idrico, salino e idrico più salino in combinazione a piante di mais, che fosse realistico a livello agronomico e quindi simile alle condizioni di stress che avvengono in campo. Al fine di mimare gli episodi stress progressivo che si verificano in campo, gli stress idrico, salino e la loro combinazione sono stati applicati alle piante in modo progressivo per dieci giorni e la risposta delle piante allo stress è stata esaminata in diversi momenti durante la sua applicazione. In condizioni di campo solitamente accade che dopo un episodio di stress le condizioni ambientali tornino favorevoli, quindi dopo i dieci giorni di applicazione di stress quest'ultimo è stato rimosso e le piante sono state cresciute in condizioni ottimali per valutarne la capacità di recupero dallo stress. Due diverse linee di mais sono state esaminate: la linea inbred B73 sensibile agli stress e un ibrido commerciale F1 selezionato per la sua resistenza agli stress. In diversi momenti durante l'applicazione dello stress e poi durante la fase di recupero dallo stress la risposta delle piante è stata valutata attraverso l'analisi di parametri agronomici, fisiologici e genetici. I parametri agronomici sono stati studiati dal laboratorio del Prof. F. Morari e i parametri fisiologici dal Dr. F. Meggio. L'attività svolta nel lavoro di Dottorato ha riguardato lo studio della risposta delle piante a livello genetico. In particolare, sono stati cercati in letteratura geni per i quali fosse nota l'espressione differenziale in seguito agli stress studiati o l'appartenenza alle principali vie molecolari di risposta a stress abiotici. Il loro livello di espressione è stato studiato nei nostri campioni attraverso la tecnica della PCR quantitativa in real-time. Tutti i parametri analizzati hanno confermato che i trattamenti sono stati efficaci nell'indurre la condizione di stress nelle piante. Di conseguenza, il protocollo messo a punto costituisce un valido strumento per studi successivi riguardanti la risposta di piante di mais a questi stress, i cui risultati mantengano validità in caso di applicazione in campo agronomico. Lo studio combinato dei parametri agronomici, fisiologici e genetici ha permesso di approfondire i meccanismi che regolano la diversa tolleranza allo stress delle due linee di mais studiate.

Il lavoro principale del Dottorato ha riguardato l'analisi bioinformatica di dati di sequenziamento di piccoli RNA, ottenuti da piante wild type e mutanti *rmr6-1*, con lo scopo di caratterizzare la popolazione dei piccoli RNA endogeni della foglia di mais, esaminarne gli effetti sull'espressione genica e la risposta a stress abiotici. 48 librerie di piccoli RNA sono state sequenziate da campioni di foglia di piante wild type e mutanti, cresciute in condizioni di controllo, in condizioni di stress abiotici e di recupero dallo stress. Le sequenze ottenute sono state pre-processate e la loro qualità è stata inizialmente verificata. Dopodiché esse sono state allineate al genoma di B73 e le sequenze allineate hanno mostrato il tipico profilo dei piccoli RNA di mais: i più abbondanti con lunghezza di 24-nt, seguiti da quelli con lunghezza di 22-nt e poi di 21-nt. Il progamma bioinformatico ShortStack è stato utilizzato per identificare *de novo* i loci genomici responsabili di una produzione significativa di piccoli RNA in foglia di mais, partendo dall'insieme di tutte le sequenze prodotte dai 48 campioni sequenziati. I loci *MIRNA* identificati sono stati i primi a essere analizzati. Sono state riscontrate delle differenze tra la nostra annotazione prodotta dei microRNA e quella riportata nel database miRBase, le quali potrebbero riflettere un'inaccurata annotazione presente in miRBase o differenze specifiche della foglia nel processamento dei precursori dei microRNA. La predizione dei target dei microRNA è stata eseguita sui trascritti annotati nel trascrittoma di mais ricostruito dai dati di sequenziamento di RNA totale. Un trascritto nuovo annotato è stato predetto come target di un microRNA di mais che è conservato in diverse specie vegetali, aiutando a capire la funzione di questo microRNA in mais. Nuovi putativi microRNA sono stati identificati: una parte di essi ha presentato le caratteristiche per essere considerati microRNA bona fide, invece altri hanno presentato caratteristiche tipiche dei 'proto-miRNA' o dei siRNA. Le altre categorie identificate di loci di piccoli RNA sono state analizzate in termini di co-occupancy con i geni codificanti proteine, con i trascritti di trasposoni e con i lunghi RNA non codificanti (lncRNA). I loci con produzione primaria di piccoli RNA di 24-nt di lunghezza sono stati trovati significativamente arricchiti nelle regioni fiancheggianti di tutte e tre le tipologie di geni considerate. In particolare, i geni espressi hanno mostrato una maggiore probabilità di essere fiancheggiati da loci di piccoli RNA di lunghezza di 24-nt rispetto ai geni non espressi. Nel mutante *rmr6-1*, nonostante la perdita sostanziale dei siRNA

osservata soprattutto nelle regioni fiancheggianti dei geni, un totale di 1013 geni sono stati trovati differenzialmente espressi (DE) rispetto al wild type e la downregolazione di un locus di piccoli RNA non è risultato in generale un criterio sufficiente e nemmeno necessario per predire la up o downregolazione del suo gene vicino. Di conseguenza, l'assenza dei siRNA non ha mostrato avere un grosso impatto nella stabilità del genoma in foglia di mais, infatti, le foglie del mutante non hanno evidenziato difetti morfologici e sono state osservate essere identiche a quelle delle piante wild type. I meccanismi coinvolti nel mantenimento del silenziamento genico quando i siRNA non sono presenti e il pathway RdDM è alterato nella sua funzione rimangono ancora sconosciuti. Dati di letteratura evidenziano la possibilità che il pathway RdDM sia essenziale per garantire la trasmissione transgenerazionale dell'informazione epigenetica. In questa ipotesi, al fine di approfondire il ruolo dei siRNA nel controllo dell'espressione genica, risulterebbe informativo lo studio dell'attività dei siRNA e delle mutazioni del pathway RdDM in altri tipi cellulari, ad esempio i gameti. Risulterebbe informativo anche lo studio delle variazioni epigenetiche di espressione genica in generazioni successive di piante. La mancanza dei siRNA, nonostante sia stato verificato non compromettere la stabilità del genoma nella foglia, è stato osservato indurre cambiamenti di espressione genica che sono apparsi come effetti secondari della mutazione. In particolare, nel mutante *rmr6-1* è stata registrata l'upregolazione di geni di risposta allo stress e di geni codificanti citocromi e la downregolazione di geni coinvolti nella regolazione del ciclo cellulare e di geni codificanti proteine istoniche. Infine è stata esaminata la risposta allo stress dei piccoli RNA. Sono stati applicati i trattamenti di stress precedentemente messi a punto ed è stato identificato un piccolo numero di microRNA e loci di piccoli RNA delle altre categorie differenzialmente espressi in condizioni di stress. Nonostante questo numero sia risultato inferiore rispetto a quello trovato in precedenti lavori che hanno analizzato la risposta dei piccoli RNA allo stress, i piccoli RNA DE identificati potrebbero essere candidati migliori per studi di tolleranza allo stress, in quanto la loro espressione differenziale è stata indotta da condizioni di stress simili a quelle che si verificano in campo.

I lavori qui citati sono riportati nella bibliografia del secondo capitolo di questa tesi.

Abbreviations

AGO	ARGONAUTE
ARF	AUXIN RESPONSE FACTOR
bp	base pair
C	control non-stressful conditions
ChIP	chromatin immunoprecipitation
CLSY1	CLASSY 1
D	drought stress alone
D+S	drought and salinity stress combined
DCL	DICER-LIKE
DDL	DAWDLE
DDM1	DECREASE IN DNA METHYLATION 1
DE	differentially expressed
DME	DEMETER
DMS3	DEFECTIVE IN MERISTEM SILENCING 3
DRD1	DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1
DRM2	DOMAINS REARRANGED METHYLTRANSFERASE 2
dsRNA	double-stranded RNA
DTF1/SHH1	DNA-BINDING TRANSCRIPTION FACTOR 1/SAWADEE HOMEODOMAIN HOMOLOG 1
EVD	*Evadé*
FLC	*FLOWERING LOCUS C*
FWA	*FLOWERING WAGENIGEN*
GO	Gene Ontology
GRF	GROWTH-REGULATING FACTOR
H2B	histone 2B
H3	histone 3
H3K4	unmethylated lysine 4
H3K9me	methylated lysine 9

H3K9me2	dimethylation of lysine 9
hc-siRNA	heterochromatic siRNA
HDA6	HISTONE DEACETYLASE 6
HEN1	HUA ENHANCER 1
HESO1	HEN1 SUPPRESSOR 1
HP loci	hairpin loci
hpRNA	hairpin RNA
HST	HASTY
HYL1	HYPONASTIC LEAVES1
INV	invariant method
IPS1	INDUCED BY PHOSPHATE STARVATION 1
JMJ14	JUMONJI 14
kb	kilobase
KTF1	KOW DOMAIN-CONTAINING TRANSCRIPTION FACTOR 1
LTR	long-terminal repeat
MCM	MINICHROMOSOME MAINTENANCE
MET1	METHYLTRANSFERASE 1
MIR	*MIRNA*
miRNA	microRNA
MITE	Miniature Inverted–Repeat Transposable Element
mop1-1	*Mediator of paramutation1-1*
mRNA	messenger RNA
Mu	*Mutator*
MULE	*Mutator*-like element
NAM	NO APICAL MERISTEM
NAT-siRNA	natural antisense transcript siRNA
NGS	Next Generation Sequencing
non-HP loci	non-hairpin loci
NRPD1	NUCLEAR RNA POLYMERASE D1
NRPD2/NRPE2	NUCLEAR RNA POLYMERASE D2/NUCLEAR RNA POLYMERASE E2
NRPD2a	NUCLEAR RNA POLYMERASE D2a

NRPE1	NUCLEAR RNA POLYMERASE E1
nt	nucleotide
phasiRNA	phased secondary siRNA
PHB	PHABULOSA
Pol II	RNA polymerase II
Pol IV	RNA polymerase IV
Pol V	RNA polymerase V
pre-miRNA	precursor miRNA
PTGS	post-transcriptional gene silencing
RdDM	RNA-directed DNA methylation
RDM1	RNA-DIRECTED DNA METHYLATION 1
RDR2	RNA-DEPENDENT POLYMERASE 2
RDR6	RNA-DEPENDENT RNA POLYMERASE 6
RISC	RNA-induced silencing complex
rmr1	*Required to maintain repression1*
rmr2	*Required to maintain repression2*
rmr6	*Required to maintain repression6*
rmr7	*Required to maintain repression7*
RNAi	RNA interference
ROS	reactive oxygen species
ROS1/ DML1	REPRESSOR OF SILENCING 1/DEMETER-LIKE1
RPM	Reads Per Million
S	salinity stress alone
SA	salicylic acid
SAM	significance analysis of microarrays
SAM	shoot apical meristem
SDN	SMALL-RNA-DEGRADING NUCLEASE
SE	SERRATE
SINE	short interspersed nuclear element
siRNA	small interfering RNA
SOD	SUPEROXIDE DISMUTASES
SPCH	SPEECHLESS

SPL	SQUAMOSA PROMOTER BINDING PROTEIN-LIKE
sRNA	small RNA
ssRNA	single-stranded RNA
tasiRNA	*trans*-acting siRNA
TCP	TEOSINTE BRANCHED1/CYCLOIDEA/PROLIFERATING CELL FACTOR1
TE	transposable element
TF	transcription factor
TGS	transcriptional gene silencing
TIR	Terminal Inverted Repeat
TMM	trimmed mean of M value
TSS	transcription start site
UBP26	UBIQUITIN-SPECIFIC PROTEASE 26
UTR	untranslated region
VSN	variance stabilization
wt	wild type

Chapter 1

Temporal progression of agronomic, physiological and genetic responses to field-mimicked conditions of drought, salinity and recovery in stress-sensitive and stress-tolerant maize lines

Scudiero E.[1]*, Meggio F.[2]*, Lunardon A.[2]*, Forestan C.[2], Farinati S.[2], Morari F.[2], Varotto S.[2]

*these authors contributed equally to this work.

[1]United States Department of Agriculture, Agricultural Research Service, U.S. Salinity Lab., 450 West Big Springs Rd., Riverside California, 92507-4617

[2]Department of Agronomy Animal Food Natural Resources and Environment (DAFNAE), University of Padova, Agripolis, Viale dell'università, 16, 35020 Legnaro, Italy

Abstract

Drought and salinity are abiotic stresses that reduce plant growth and have a strong impact on crop yield. These stresses will have a high future impact on crop productivity, due to both the increase competition for land, water, energy and climate changes. The response to drought (D), salinity (S) and the combined stress (D+S) was monitored in time course of stress applications in two maize genotypes: the inbred line B73 and a F1 commercial hybrid selected for its tolerance to stress. To mimic field progressive stress conditions, a stress protocol was developed and the stress conditions analyzed in terms of effect on plant growth at different time points, indicating that all the applied stresses were effective in limiting growth in the hybrid and arresting it in the inbred line. When subjected to salt stress conditions, the two genotypes had different ion accumulation and translocation capacity, particularly for Na^+ and Cl^-. The response of the two genotypes to stresses was physiologically different: the hybrid rapidly reduced all the analyzed physiological parameters and kept them reduced until the recovery, while the B73 decreased more gradually all physiological parameters, being mainly affected by S stress. Both genotypes recovered better from the D stress compare to the other stresses. Expression analysis of stress marker genes indicated that gene expression was modulated in response to the applied stresses in the two genotypes. Gene expression patterns were not coincident and reflected the different capacity of the two genotypes to cope with D, S and D+S treatments. Combining agronomic and physiological data with gene expression analyses yielded insight into the mechanisms regulating the different tolerance to the stress of the two genotypes.

Introduction

Drought and salinity are abiotic stresses that adversely affect plant growth and productivity, because they negatively influence both photosynthesis and plant reproduction. In the future these stresses will have a high impact on crop yield, due to both the increase competition for land, water, energy and climate changes (FAO 2002, Ahuja et al. 2010). In particular, competition for water resources among different social and economic sectors is growing, with agriculture being progressively forced to use lower quality water (Araus J-L 2004). For example, the problem of salinity is becoming increasingly serious particularly near coastal areas. The exploitation of groundwater involves the increase of saline intrusion with implications in salt accumulation and soil degradation. On the other hand, irrigation-induced salinity represents a main constraint limiting productivity for many crops. Selecting more drought and salt-tolerant genotypes is a desirable way of improving crops (Tester and Langridge, 2010). Maize, one of the most important food, feeding and industrial crops, has a pronounced susceptibility to drought and salinity (Bänziger and Araus 2007): improving the stress resistance of this crop is thus of strategic significance.

A fair amount of studies has been focusing on the comparison of the differential responses of crops to water and salt stresses (eg. Hu et al. 2007, Munns R 2002, Elmetwalli et al. 2012) as they both lower soil water potential, normally leading to similar physiological responses. The effects of water deficiency stress on plants are well known: reduction of the photochemical activity of the chlorophyll (Souza et al. 2004), reduced activity of the roots in the adsorption of nutrients from the soil, and slacken roots to shoots nutrient transport (Kramer and Boyer 1995). Even at high moisture content, soil salinity induces disequilibrium in the ionic ratios in the plant (Grattan and Grieve 1999), resulting in physiological drought with the abovementioned effects on plants (Corwin DL 2005). On the other hand, soil salinity can also cause specific ion toxicity (Rhoades et al. 1999), and compromise the repartition of macro- and micronutrients within leaves (Hu et al. 2007, Neves-Piestun and Bernstein 2005).

In many plant species, genetic studies have shown that drought and salinity stress tolerance is a complex trait. However, its understanding can be facilitated by the adoption of expression analysis approaches, which help elucidating the molecular basis of stress adaptation and identifying the numerous pathways important for the plant growth under limiting water or in saline soil (Shinozaki and Yamaguchi-Shinozaki 2007, Bartels and Sukar 2005, Deinlein et al. 2014). These pathways tend to be conserved among plant species, indeed one of the most obvious features of the adaptation to drought and salinity is the change in transcript profiles for genes involved in many biochemical, cellular and physiological processes, from transcription regulation to signal transduction, protein

biosynthesis and decay, membrane trafficking and photosynthesis (Vinocur and Altman 2005). For example, from genetic studies it is evident that plant adaptation to drought is a complex biological processes, which includes up or downregulation of specific genes, transient increase in ABA levels, build-up of compatible solutes and protective enzymes, increasing levels of antioxidants and inhibition of energy-consuming pathways (Salekdeh et al. 2009). However, the conservation of pathways and genes is not sufficient to translate results from one species and even genotype to another because the high conservation of the core gene machinery between plants may not correlate with the expression timing of the stress-induced genes. A diverse stress tolerance between two genotypes may reflect differences in the timing of up or/and downregulation of specific gene sets (Skirycz et al. 2011).

Another important aspect of abiotic stress studies in plants is the need to apply stress conditions that retain their value under field conditions, thus improving translational research from model plants to crops, for agronomical solutions. In many experimental works dealing with stress response, tolerance is assessed predominantly in severe conditions in which plant survival would be compromised in the case of prolonged treatment application. However, in field conditions, limited resource availability rarely causes plant death and after a period of stress, environmental conditions usually turn more favourable, determining reduced crop yields but without compromising the survival of plants (Skirycz et al. 2011, Deikman et al. 2012).

In this work, we analysed the stress response to drought, salinity and the combined stress in two maize genotypes: the reference inbred line B73 for which genomic tools are available and a F1 hybrid selected for its tolerance to stress. We developed a protocol with the aim to mimic field progressive stress conditions and evaluate the stress response of the two genotypes in time course of stress application and after four days of recovery. The strategies adopted by the two diverse genotypes to cope with stresses were evaluated using agronomic, physiological and genetic parameters.

Materials and Methods

Experimental set-up

The experiment was carried out at the experimental farm of the University of Padova, Italy (45°21' N, 11°58' E, 6 m a.s.l.) in the period May-July 2012. The response to drought and salinity was tested on two varieties of maize (*Zea mays* L.): the hybrid PR32P26 (hereafter simply called P26, Pioneer Hi-Bred Italia, Gadesco Pieve Delmona, Italy) and the inbred line B73. In a field provided with an automatic mobile roof avoiding rainfall input, pots (diameter 23 cm, height 23 cm, volume 9500 cm^3) were filled with a 50%-weight mixture of native sandy loam and silica sand. The resulting substrate (66% sand, 27.5% silt, and 6.5% clay) was sub-alkaline (pH 7.8), had an organic carbon content of 0.40%, and was non-saline (saturated paste electrical conductivity, EC_e= 0.8 dS m^{-1}). The substrate was packed in the pots in order to obtain a bulk density of 1.42 ± 3.6*10^{-3} g cm^{-3}. Pot water capacity and wilting point were 0.154 ± 1.94*10^{-3} cm^3 cm^{-3} and 0.072 ± 0.9*10^{-4} cm^3 cm^{-3}, respectively. Before sowing, 0.50 g N, 0.22 g P_2O_5 and 0.15 g K_2O were added to each pot. Maize seeds were pre-germinated for 2 days in wet, rolled paper towels at 25 °C, after which 3 germinating seeds were transferred in each pot. The seedlings were thinned to one per pot after 7 days.

The two varieties of maize were tested under the factorial combinations of two water regimes and two salt concentrations in the soil, in four treatments: C (unstressed plants, the control), D (drought stress), S (salt stress) and D+S (drought and salt stress combined). The experimental design consisted in a randomized block with 3 replications. Since destructive plant samplings were performed on 5 dates, a total of 120 pots were prepared.

During the experiment pots were weighted daily. Water-unstressed plants were grown at a water content of 100% available water capacity, replenishing every day the water lost by evapotranspiration. On the contrary, water-stressed plants were watered replenishing only the 60% of daily evapotranspiration to a minimum water content threshold of 0.10 cm^3 cm^{-3} (i.e. 40% of the available water capacity). The saline water (electrical conductivity=20 dS m^{-1}) consisted in a controlled mix of ions (Cristal Sea Marinemix®: 54.92% Cl^-; 30.82% Na^+; 7.68% SO_4^{2-}; 3.81% Mg^{2+}; 1.21% Ca^{2+}; 1.12% K^+; 0.44% $NaHCO_4$) reproducing saline groundwater typically found in the coastal soils of the southern margin of the Venice Lagoon, Italy (Scudiero et al. 2012). D+S plants were watered replenishing only the 60% of daily evapotranspiration, like in D, with saline water, like in S. The use of this protocol implied that the quantity of ions mix was lower in the pots of D+S treatments compared to S.

Stress treatments started on June 18th and were applied to V6 plants. Until that day, water content was maintained at the pot water capacity in each pot. Plants were sampled at the beginning of the treatments (T0), on June 20th (T2), on June

22nd (T4), on June 28th (T10), and on July 2nd (T14). To verify the plant recovery capacity from water and salt stress conditions, from June 28th to July 2nd all plants were watered twice a day with non-saline water, up to a water content of 0.30 cm^3 cm^{-3}, in order to promote salt leaching and optimal soil moisture status.

Physiological analyses

Single-leaf gas exchange measurements were performed with a LI-6400 portable photosynthesis system (Li-Cor Inc. Lincoln, Nebraska, USA). Analyses were performed using the circular 2 cm^2 leaf cuvette equipped with the 6400-40 fluorometer as the light source. Measurements were subjected to at least 10-min acclimation at a constant saturating photosynthetic photon flux density (PPFD) of 1500 µmol of photons m^{-2} s^{-1}, a CO_2 concentration of 390 µmol mol^{-1} and relative humidity (RH) between 60 and 70% allowing ~ 1.7 vapor pressure deficit (VPD) inside the chamber. Block temperature was maintained at 27°C allowing leaf temperature to range between 29 and 36°C. In addition to net assimilation rate (A_n, µmol CO_2 $m^{-2}s^{-1}$) and stomatal conductance (g_s, mmol H_2O $m^{-2}s^{-1}$) the incorporated fluorometer allowed determination of the actual photochemical efficiency of photosystem II (φPSII). This was determined by measuring steady-state fluorescence (F_s) and maximum fluorescence during a light-saturating pulse of *c*. 8000 µmol m^{-2} s^{-1} (F'_m), following the procedures of Genty et al. (1989): $\varphi PSII=[(F'_m- F_s)/ F'_m]$. Measurements were performed on fully expanded leaves per plant comprising at least three leaves per treatment at regular times during the experimental period, between 11.00 am and 2.00 pm solar time.

Chemical analyses on plants and soil

Once physiological analyses were performed, plants were weighted and analyzed for ions composition and soil was sampled for salinity assessment. Roots and shoots were dried at 60°C for 48 hours and dry weights were measured. Powered biomass was analyzed for cations (Na^+, K^+, Mg^{2+}, Ca^{2+} and NH_4^+) and anions (Cl^-, SO_4^{2-}, and PO_4^{3-}) by ion chromatography (ICS 900, Dionex, Sunnyvale, CA, USA) according to Nicoletto et al. (2013). The soil in the pots was air dried and sieved at 0.5 cm and then analyzed for saturated paste electrical conductivity (EC_e) (Rhoades et al. 1999). The osmotic potential of the saturated extract was then analyzed with the WP4-T Dewpoint PotentiaMeter (Decagon Devises Inc., Pullman, WA,USA).

Real time quantitative PCR (qRT-PCR)

The last expanded leaf was collected for RNA extraction. Three biological replicates were used for the two time points June 28th T10 and July 2nd T14 of each treatment: C, D, S, and D+S. Biological replicates were pooled together and total RNA was extracted from maize leaves using the RNeasy Plant Mini Kit (QiAgen) and subjected to on-column DNase treatment (QiAgen). cDNA synthesis

was performed with the SuperScript III reverse transcriptase kit (Invitrogen), according to the manufacturer's instructions. One microgram of total RNA was used as a template together with 1µl of oligo(dT)$_{12–18}$ (0.5µg/µl – Invitrogen). Quantitative Real-Time PCR expression analysis was performed using a StepOnePlus™ Real-Time PCR Systems (AppliedBiosystems) and the FAST SYBR® GREEN PCR MasterMix (Life Technologies), following the manufacturer's guidelines. Real-time conditions were: 20 s at 95 ˚C, 40 cycles of: 3 s at 95 ˚C and 30 s at 60 ˚C. For each reaction, we observed product melting curves by heating from 60 to 95 ˚C at 0.2 ˚C/s. For all transcripts, this procedure allowed identification of a single product, which we confirmed by analysis on 2% agarose gel. Three technical replicates were carried out for each primer combination. The constitutively expressed *GAPC2* gene was used as housekeeping internal control of the cDNA/RNA quantity. Relative quantification of gene expression (normalized to *GAPC2* transcript quantities) was performed with the Pfaffl method (Pfaffl 2001) using previously determined amplification efficiencies for each gene. Specific primers were designed using Primer BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) or were selected from published papers (Supp.Table 1).

Statistical Analyses

A tree-way ANOVA (mixed model with repeated measures) by maize variety, salinity level and water regime was used to analyze agronomic and physiological parameters. Comparison between means was performed by adjusted Tukey's test. In order to estimate a possible linear relationship between parameters the Pearson correlation coefficient was calculated. The general structure of the interdependences existing between physiological response, plant growth, chemical composition, and gene expression was finally evaluated performing a correlation-based principal component analysis (PCA) on 12 variables measured before (T10) and after the recovery (T14): leaf dry matter, leaf and root Na^+, leaf Cl^-, ratio K^+/Na^+ in root, net assimilation (A_n), expression patterns for *PMP3-4*, *HSP70*, *CAT1*, *CoAred* and *SUS*. Variables were selected according to Kaiser's measure of sampling adequacy (MSA). The overall MSA was 0.74 indicating that PCA was suitable (Kaiser, 1974). Rotated orthogonal components (varimax normalized method of rotation) with eigenvalues >1 were extracted (Kaiser, 1960) and the relative scores were determined. Statistical analyses were performed with STATISTICA 7.0 (Statsoft Inc., Tulsa, OK, USA) and SAS 9.3 (Cary, NC, USA).

Supplemental Table 1 List of qRT-PCR primers

Gene (Genbank Acc. No. or MaizeGDB Acc. No.)*		Sequence (5'->3')
ZmGAPC2 (GRMZM2G180625)	sense	AATGGCAAGCTCACTGGC
	antisense	CTGTCACCGGTGAAGTCG
ZmLEA3 (NM_001153473) [1]	sense	GTCCGTGACCCTGTTTGC
	antisense	CCGCCCGACTCGTTTA
ZmPMP3-4 (EU954642.1)	sense	TTCTGGATCGACCTCTTGCT
	antisense	TCCTCCTCTTCGCACAACTT
ZmHSP70 (CA404511) [2]	sense	GATCCCCTCAAGCTCCTTCAT
	antisense	AGATCGAAGATGCCGTTGACA
ZmCAT1 (NM_001111945.1)	sense	CCTGTGGTACAAACCCTGCT
	antisense	ATCCTTGCTGCATCTGTCCG
ZmPP2C (EF195257.1)	sense	CTGATGATATCAGTGTCGTGATCGTGCAG
	antisense	CGCCAGCGAAGTAACATATCATGTCTACC
putative B2/DP1 HVA22 (GRMZM2G154735)	sense	ATCCTCACTCACCTCCACTCCCTAGC
	antisense	GAGCTCGTACCAGATGGGGATCCAGTAT
putative calcium-binding EF-hand (GRMZM5G827398)	sense	TGTCCGCTTGGAGTTCAGTCACTACG
	antisense	GAGCTCAGGTTACCATCGCAGTTAGC
putative hydroxymethylglutaryl-coenzymeA reductase (CO440726) [2]	sense	AGACAAACGTACAGGCTCTCG
	antisense	GCTGCCACAATGTTACTTGC
ZmSUS (X02382) [3]	sense	CCCTTCAATGCCTCCTTTCCTC
	antisense	TCAACATCATCGTCGTGCCC
ZmIVR1 (U16123.1) [4]	sense	GCTGCCTTCCTTATCCTTCTTGTG
	antisense	CCTGCTCCCTGCTCCTCTTATC
ZmGLN1 (NM_001254779.1) [2]	sense	GGCGGGTTTGAAGAGATCAA
	antisense	CCAGTCAGTCTTCTTTCATTTCCTT
putative Rab GTPase (GRMZM2G018619)	sense	ACTAGTGCGTATTACCGAGGCGCTGT
	antisense	CGGTAGATCTGAGCTAGGACTTCTGC
Zmβ-EXP7 (AF332180) [5]	sense	CAACCTTGTCCTCCACAGTAG
	antisense	GTGAGGTCGGAGGCGTTAAA
ZmNHX4 and ZmNHX5 (NM_001112473.1 and NM_001111753.1)	sense	AATCTCTCTCGGCGCAATAG
	antisense	CACAGAATCCGTTGCAGAAA
ZmRMR6 (NM_001195895.1) [6]	sense	GAGGGTTTGAATCCATTGGAATGTC
	antisense	GGAGTCCTCTAAACCATTGACCG
ZmHDA108 (GRMZM2G136067)	sense	AGACTACTACTACGGGCAAG
	antisense	CACGCCTGTGGAACTTGAGGAGCTCG
putative Really Interesting New Gene Zn-finger (GRMZM2G148908)	sense	GCTCGGCCTCCTCAAGGTTATGCTATAC
	antisense	GTTCTCCCTAGTCAAGGTATCCGTGTCC
putative RNA-binding KH domain-containing protein (AC218972.3_FG007)	sense	GAGTTGAAGCTACTACAGGTTGCCGTGT
	antisense	GGTTTCAGCAATCCTCCAGTATCTC

*Gene bank numbers according to http://www.ncbi.nlm.nih.gov/. Maize GDB numbers according to http://www.maizegdb.org.

[1] Liu Y et al. 2013.

[2] under MTA contract with Biogemma, 8 rue des frères Lumière, 63100 Clermont-Ferrand, France.

[3] Wang et al. 2003.

[4] Kakumanu et al. 2012.

[5] Geilfus et al. 2010.

[6] primers provided by Dr. V.Rossi, Consiglio per la Ricerca e la Sperimentazione in Agricoltura, Unità di Ricerca per la Maiscoltura, Via Stezzano 24, I-24126 Bergamo, Italy

Results

Plant development in response to stress

To analyze the effect of the stress on plant growth we measured both shoot and root dry weight of control and stressed plants of the two genotypes, during stress applications (at T2, T4 and T10) and after the recovery from the stresses (T14). Genotypes were different in their growing capacity being the hybrid more productive than the B73 inbred for both shoots and roots ($P<0.01$; Table1).

Shoot dry weight accumulation indicated that the P26 hybrid coped better with the stress conditions with respect to the B73 inbred line (Fig.1a,b). In hybrid, compared with control treatment (C) both drought (D) and salinity (S) reduced shoot growth that was, however, stopped in D+S (Fig.1a). B73 shoots were affected in their growth by D+S, than D and S (Fig.1b). D influenced the growth of plant roots with a reduction of almost 50% compared with control plants both in the hybrid and inbred line (Fig.1c,d). In the time course of stress applications, in P26 root growth was less reduced in D and D+S whose effects were similar if compared to C. In this genotype, S blocked the root growth (Fig.1c). In B73, both S and D+S arrested root growth, whereas root growth was reduced in D (Fig.1d).

The two genotypes showed a different capacity to recover from the stresses. The shoots of hybrid plants increased their growth soon after the D and S were removed, whereas the removal of the D+S did not promote shoot growth (Fig.1a). The shoot growing capacity of the B73 plants did not change after the stresses removal and even decreased in D (Fig.1b). In the case of root, D removal affected the growth capacity of P26 hybrid plants that accelerated their growth after the recovery (Fig.1c). Conversely, root d.w. of the hybrid decreased after D+S removal and in S recovery root d.w. did not varied at all (Fig.1c). No increase in root d.w was observed in B73 plants after recovery from any of the different stresses (Fig.1d). These results indicated that D and D+S reduced the growth of hybrid shoot and root compared to C, whereas S completely inhibited the growth of this genotype that showed a reduced recovery capability in terms of d.w. at T14. B73 plant shoots and roots did not grow during S and D+S time course and recovery. Relatively more tolerance to D was showed by the inbred line that, anyway, was not able to recover at T14 as the hybrid did.

	ECs	Soil ψπ	*Shoot* dw	*Root* dw	*Leaf* ψ_t	*Leaf* ψπ	A_n	g_s	Φ_{PSII}
	µS cm^{-1}	MPa	g	g	MPa	MPa	µmol CO_2 $m^{-2}s^{-1}$	mmol H_2O $m^{-2}s^{-1}$	efficiency
Variety									
wild	4585 *ns*	-0.31 *ns*	1.48 *b*	1.37 *b*	-2.30 *ns*	-2.41 *ns*	14.06 *b*	0.10 *b*	0.09 *ns*
hybrid	5675 *ns*	-0.35 *ns*	2.90 *a*	2.65 *a*	-2.35 *ns*	-2.06 *ns*	17.55 *a*	0.12 *a*	0.10 *ns*
Water salinity									
no salt	1736 *b*	-0.24 *a*	2.66 *a*	2.57 *a*	-2.27 *ns*	-1.91 *a*	21.64 *a*	0.14 *a*	0.12 *a*
salt	8524 *a*	-0.43 *b*	1.73 *b*	1.46 *b*	-2.37 *ns*	-2.55 *b*	9.97 *b*	0.07 *b*	0.06 *b*
Soil water content									
60 %	4050 *ns*	-0.30 *ns*	1.79 *b*	1.75 *ns*	-2.29 *ns*	-2.34 *ns*	13.06 *b*	0.09 *b*	0.08 *b*
100 %	6209 *ns*	-0.36 *ns*	2.60 *a*	2.27 *ns*	-2.35 *ns*	-2.12 *ns*	18.55 *a*	0.13 *a*	0.11 *a*
Day									
T2	4133 *b*	-0.28 *b*	1.07 *c*	1.19 *c*	-1.70 *a*	-2.09 *ab*	19.57 *ns*	0.13 *a*	0.11 *ns*
T4	5942 *b*	-0.31 *b*	1.33 *c*	1.23 *c*	-2.90 *b*	-2.25 *ab*	15.26 *ns*	0.10 *ab*	0.09 *ns*
T10	9593 *a*	-0.55 *c*	2.46 *b*	2.46 *b*	-2.66 *b*	-2.83 *b*	13.93 *ns*	0.09 *b*	0.09 *ns*
T14	851 *c*	-0.20 *a*	3.91 *a*	3.17 *a*	-2.03 *a*	-1.75 *a*	14.45 *ns*	0.11 *a*	0.09 *ns*

Table 1 Factor analysis of electrical conductivity (ECs), soil osmotic potential (ψπ), shoot and root dry weight (d.w.), leaf turgor potential (ψ_t), leaf osmotic potential (ψπ), net CO_2 assimilation (A_n), stomatal conductance (g_s) and quantum efficiency of photosystem II (Φ_{PSII}).

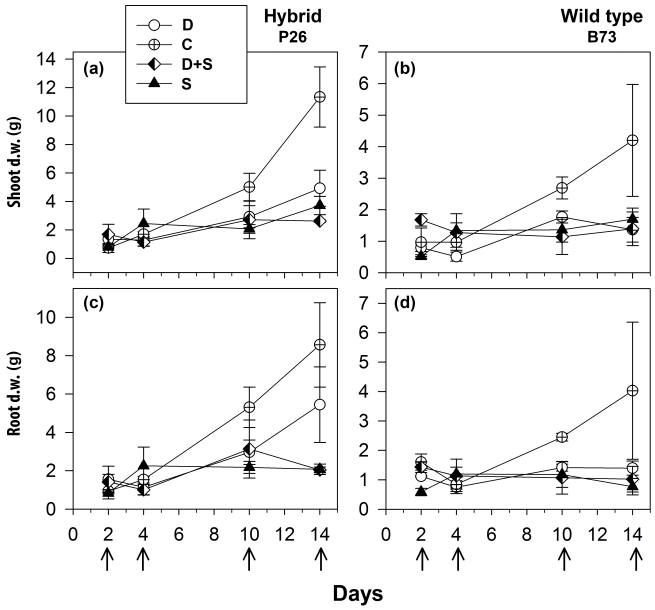

Figure 1 Dry weight (d.w.) of: **(a)** P26 shoots, **(b)** B73 shoots, **(c)** P26 roots and **(d)** B73 roots of plants grown for 2, 4 and 10 days under control (C), drought (D), salinity (S), or the combination of drought and salinity (D+S) and after 4 days of recovery from the stresses. Values represent means (±SE) of two independent replicates.

Ion contents

To verify the mechanisms of uptake and translocation of ions in the two genotypes, we measured the ion contents in both leaves and roots. Ion concentrations were measured during stress application at T2, T4 and T10 and after 4 days of recovery (T14) in B73 and hybrid plants (Fig.2, Fig.3 and Table 2).

Na^+ concentration was significantly higher in both leaves and roots of hybrid and B73 plants grown under S and D+S compared with D and C treatments (Fig.2 and Table 2). At T10, Na^+ concentration in roots of hybrid plants grown under S and D+S treatments was about three and two times higher, respectively, than those found in plants grown under C and D (Fig.3c). At the same time point, in B73 Na^+ root concentration in D+S and S was about three and four times higher respectively than in D and C (Fig.2d). Considering Na^+ accumulation, the response of P26 hybrid plants to S and D+S was very rapid in roots, being the increase of Na^+ concentrations observed already at T2, whereas in B73 roots the increase was began to be evident at T4. In leaves of S treated plants Na^+ concentration increase was detected at T4 in both genotypes (Fig. 2a,b).

Considering the effect of recovery in the hybrid, it is interesting to note that the Na^+ concentration in roots dropped to the same value of C under S, while recovery had no effects under D+S (Fig.2c). An opposite Na^+ concentration trend was observed in hybrid leaves (Fig.2a). The recovery had no effect on plant leaves grown under S and a decrease in Na^+ concentration was instead observed in shoots grown under D+S. In B73 plant roots grown under S and D+S, Na^+ concentration dropped to the concentration level of non-treated and D treated plants after recovery application while in leaves grown both under S and D+S a reduced concentration of Na^+ was observed after the recovery (Fig. 2b). However, Na^+ concentration remained four times (S) and two times (D+S) higher than the concentration measured in C and D plant leaves. Factor analysis revealed that the ratio between leaf Na^+ and root Na^+ was significantly different between the hybrid and B73, 0.57 and 1.29 respectively ($P<0.01$), and interestingly, that this ratio significantly increased from T10 to the recovery at T14, 0.78 to 1.80 respectively ($P<0.01$) (Table 2).

In plants grown under C and D, Cl^- concentrations were very similar for the two genotypes, and no significant variations were found in the time course of 10 days of stress application in leaves and roots (Fig.3 and Table 2). However, when plants were grown in S and D+S a significant increase in Cl^- concentration was found in shoot of the hybrid and B73 plants starting at T2 in roots and at T4 in leaves. An evident difference in concentration values of Cl^- between the leaves of the hybrid compared with those of the inbred line was observed (Fig.3a,b). B73 leaves accumulated up to 50mg/g of Cl^- after ten days of salt stress while hybrid leaves reached the maximum concentration of 14mg/g. Conversely, Cl^- concentration values in the roots of the two genotypes were quite similar (Fig.3c,d).

The effect of recovery in the hybrid leaves was different after the S and D+S. After 4 days of recovery from the D+S, the concentration of Cl^- was reduced of about 50% in the leaves of the hybrid whereas it continued to increase during the recovery from the S (Fig.3a). In the B73 leaves, Cl^- concentrations decreased during recovery both from S and D+S, but the ion amount remained higher in S compared to the other treatments (Fig.3b). Also in the root of the two genotypes, the effect of recovery from the stresses determined a reduction in Cl^- concentrations that reached the values of the C and D, with the exception of the hybrid roots in D+S where the Cl^- concentration was only partially reduced (Fig.3c,d). As observed for Na^+, also the repartition of the Cl^- between leaf and root was significantly different between the hybrid and B73, with a ratio of 1.43 and 5.94 respectively (Table 2).

Potassium (K^+) concentration resulted unaffected by treatments in both leaves and roots with the exception of S that decreased K^+ concentration in roots, from 5.75 to 4.38 mg g^{-1}(Table 2).

S increased the concentrations of the other analyzed cations, NH_4^+, Mg^{2+} and Ca^{2+}, in the leaves of the two genotypes while no significant effects were observed for roots (Table 2). Moreover, leaf Mg^{2+} and Ca^{2+} concentration were both affected by the variety, with higher values in B73 than hybrid; an opposite behavior was observed for Ca^{2+} in the roots (Table 2). No significant difference in leaf and root content of PO_4^{3-} was observed between genotypes or due to the stress treatments. Leaf SO^{4-} concentration was significantly higher in B73 compare to the hybrid. Root SO^{4-} concentration increased subsequently to D. Considering the recovery, its effect was significant for the concentration of K^+, Mg^+, Ca^{2+} in the leaf and Mg^{2+} Ca^{2+} PO^{4-} and SO^{4-} in the roots (Table 2).

Taken together these data showed that the two genotypes have different ion accumulation and translocation capacity when subjected to stress conditions. This is particularly evident in the case of Na^+ and Cl^- accumulation in roots and in leaves of the two genotypes grown under S.

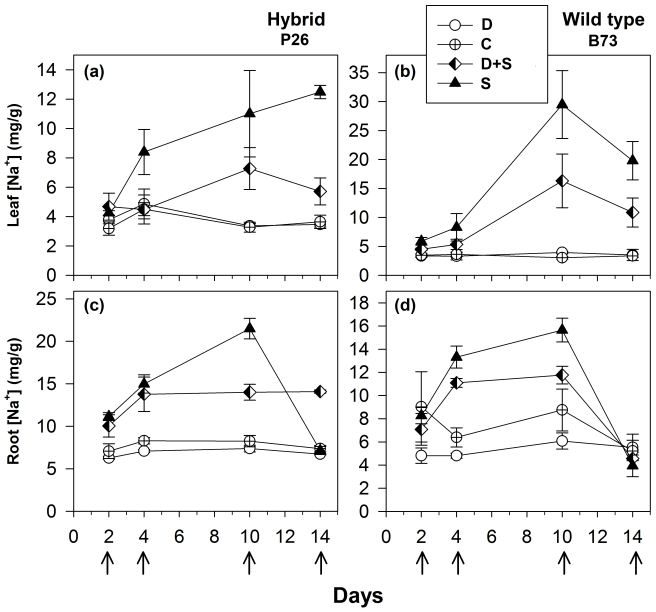

Figure 2 Na^+ concentration of: **(a)** P26 leaves, **(b)** B73 leaves, **(c)** P26 roots and **(d)** B73 roots of plants grown for 2, 4 and 10 days under control (C), drought (D), salinity (S), or the combination of drought and salinity (D+S) and after 4 days of recovery from the stresses. Values represent means (±SE) of two independent replicates.

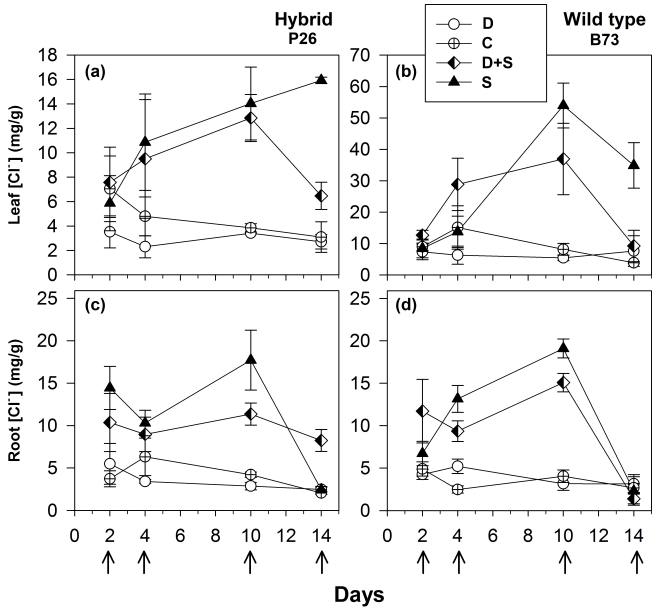

Figure 3 Cl^- concentration of: **(a)** P26 leaves, **(b)** B73 leaves, **(c)** P26 roots and **(d)** B73 roots of plants grown for 2, 4 and 10 days under control (C), drought (D), salinity (S), or the combination of drought and salinity (D+S) and after 4 days of recovery from the stresses. Values represent means (±SE) of two independent replicates.

	Leaf (L)								*Root (R)*								Na^+ *L/R*	Cl^- *L/R*
	Na^+	K^+	NH_4^+	Mg^{2+}	Ca^{2+}	Cl^-	PO_4^{3-}	SO_4^{2-}	Na^+	K^+	NH_4^+	Mg^{2+}	Ca^{2+}	Cl^-	PO_4^{3-}	SO_4^{2-}		
	mg g^{-1}			mg g^{-1}	mg g^{-1}	mg g^{-1}			mg g^{-1}			mg g^{-1}	mg g^{-1}	mg g^{-1}				
Variety																		
wild	8.02 *a*	12.73 *ns*	1.65 *ns*	4.96 *a*	9.23 *a*	16.36 *a*	1.00 *ns*	0.70 *a*	7.89 *b*	5.01 *ns*	0.88 *ns*	2.53 *ns*	6.93 *ns*	6.80 *ns*	0.61 *ns*	3.27 *ns*	1.29 *a*	5.94 *a*
hybrid	5.53 *b*	13.40 *ns*	1.73 *ns*	3.65 *b*	6.62 *b*	7.11 *b*	0.74 *ns*	0.31 *b*	10.32 *a*	5.12 *ns*	0.85 *ns*	2.32 *ns*	7.54 *ns*	7.15 *ns*	0.49 *ns*	3.54 *ns*	0.57 *b*	1.43 *b*
Water salinity																		
no salt	3.62 *b*	12.32 *ns*	1.31 *b*	4.06 *b*	7.59 *b*	5.85 *b*	0.89 *ns*	0.40 *ns*	6.82 *b*	5.75 *a*	0.95 *ns*	2.41 *ns*	7.31 *ns*	3.78 *b*	0.55 *ns*	3.20 *ns*	0.57 *b*	1.75 *b*
salt	9.93 *a*	13.82 *ns*	2.07 *a*	4.55 *a*	8.26 *a*	17.62 *a*	0.84 *ns*	0.61 *ns*	11.40 *a*	4.38 *b*	0.78 *ns*	2.43 *ns*	7.16 *ns*	10.17 *a*	0.55 *ns*	3.61 *ns*	1.29 *a*	5.61 *a*
Soil water content																		
60%	5.56 *b*	11.81 *b*	1.60 *ns*	4.12 *ns*	7.51 *b*	10.17 *b*	0.78 *ns*	0.45 *ns*	8.44 *b*	4.84 *ns*	0.86 *ns*	2.27 *b*	7.20 *ns*	6.66 *ns*	0.48 *ns*	2.97 *b*	0.88 *ns*	2.80 *ns*
100%	7.99 *a*	14.33 *a*	1.78 *ns*	4.49 *ns*	8.35 *a*	13.30 *a*	0.96 *ns*	0.57 *ns*	9.77 *a*	5.29 *ns*	0.87 *ns*	2.57 *a*	7.27 *ns*	7.29 *ns*	0.62 *ns*	3.83 *a*	0.98 *ns*	4.57 *ns*
Day																		
T2	4.15 *b*	15.55 *a*	1.78 *ns*	4.40 *b*	8.50 *b*	7.69 *b*	1.12 *ab*	0.66 *a*	7.97 *c*	6.53 *a*	1.04 *ns*	2.58 *a*	7.08 *b*	7.70 *ab*	0.80 *a*	4.60 *a*	0.57 *b*	1.44 *b*
T4	5.36 *b*	15.58 *a*	1.83 *ns*	2.83 *c*	4.03 *c*	11.44 *b*	0.55 *b*	0.27 *b*	9.97 *b*	5.87 *a*	0.96 *ns*	2.70 *a*	8.01 *ab*	7.41 *b*	0.46 *ab*	3.75 *a*	0.57 *b*	1.93 *b*
T10	9.72 *a*	12.01 *b*	1.60 *ns*	5.69 *a*	10.19 *a*	17.33 *a*	1.22 *ab*	0.77 *a*	11.67 *a*	4.51 *b*	0.72 *ns*	2.75 *a*	8.43 *a*	9.69 *a*	0.60 *ab*	3.60 *a*	0.78 *b*	1.70 *b*
T14	7.87 *a*	9.12 *b*	1.53 *ns*	4.29 *b*	9.00 *b*	10.49 *b*	0.59 *ab*	0.33 *b*	6.82 *c*	3.36 *b*	0.74 *ns*	1.65 *b*	5.42 *c*	3.10 *c*	0.33 *b*	1.66 *b*	1.80 *a*	9.67 *a*

Table 2 Factor analysis of leaf and root cations and anions and leaf/root ratio of Na^+ and Cl^-.

Photosynthetic parameters

To determine the physiological response of plants to the stresses, net assimilation, stomatal conductance and quantum efficiency of photosystem II were studied (Fig.4 and 5). Net assimilation (A_n) measured in the control condition (C) was 19.48 ± 5.85 and 23.33 ± 2.43 µmol CO_2 /(m^2*s) for P26 and B73 genotypes, respectively. At the same time, stomatal conductance (g_s) and quantum efficiency of photosystem II (Φ_{PSII}) were 133.99 ± 27.95 mol H_2O/(m^2*s) and 0.10 ± 0.03 for P26 and 148.10 ± 19.68 mol H_2O/(m^2*s) and 0.13 ± 0.02 for B73. As a consequence of D, S and their combination D+S, A_n, g_s and Φ_{PSII} decreased in both genotypes, as shown on a percentage of control basis in Fig.4. The stress effect became evident already at early stages (T4) in P26 with reductions of ~ 60 % for all parameters measured compared to C. On the contrary, after the same time, in B73 only a small reduction (~20 %) was measured for D and S treatments, for D+S the effect was higher leading to a halving of all three parameters measured. When stress conditions became more severe (T10) their effect was progressively higher in B73 than in P26, becoming evident and statistically significant between genotype and treatment. After 10 days, no significant differences were measured among genotypes for D and D+S treatments. Under S, while in P26 values similar to those for D were measured, an almost complete inhibition of photosynthetic apparatus (A_n, Φ_{PSII}) and quasi-complete stomatal closure (g_s) were detected in B73. At T14, a recovery capacity upon re-watering up to values of 50-70 % compared to C was measured for both genotypes under D. Under S and D+S, while B73 demonstrated, although small, a

recover capability to values of 30-40 %, P26 showed no significant differences from the previous time point (T10) for both treatments, leading to values of about 20-30 % compared to C. These results indicated that the response of the two genotypes to the applied stresses is physiologically different: at T4 the hybrid perceived the stress, reduced all the analyzed physiological parameters, particularly in D+S, and kept them reduced until the recovery, where it reacted better to the D compare to the other stresses. B73 decreased more gradually all physiological parameters until T10, being mainly affected by S, and recovered immediately after the stress removal, especially from D.

The dependence of A_n on g_s (i.e. their ratio or leaf intrinsic water use efficiency, $iWUE_{leaf}$) as well as of g_s on soil osmotic potential (soil Y_0) was analyzed (Fig.5). Data comprising C, D, S and D+S for both genotypes are presented, and the best-fitting regression curves are shown. When g_s is plotted against A_n (Fig.5a,b) and against soil Y_0 (Fig.5c,d) a linear and exponential growth function, respectively, satisfactorily fitted data from both genotypes. The evaluation of these regressions enabled the detection of three distinct phases, which were characterised by a 'mild or no stress', a 'moderate stress' and a 'severe stress' phase, respectively (Fig.5). The results revealed a similar pattern of photosynthetic response for both D and S stress and their combination D+S, but with different ranges between the two genotypes. In the early stages of the mild or no stress phase, A_n values for P26 were higher than those detected for B73 (Fig.5a,b). After an early stress effect resulting in partial stomatal closure (phase 2, Fig.5a,b, moderate stress), further reduction of g_s was evident as stress gradually proceeded leading to severe conditions (T10), with a simultaneous dramatic reduction of g_s (phase 3) and an almost complete inhibition of A_n for P26 under D+S treatment (Fig.5a). In contrast, for B73 an even higher stomatal closure leading to a complete inhibition of A_n was measured instead under S conditions (Fig.5b). When g_s is plotted against soil Y_0 (Fig.5c,d) only plants for T2 and T10 were used and the results revealed a similar pattern of soil Y_0 response for both D, S stress and their combination (D+S) following the same g_s threshold observed for A_n/g_s relationship. These results underlined how under severe stress (T10) plants of both genotypes under S and D+S experienced lowest soil Y_0 values up to values of ~-1 MPa on average.

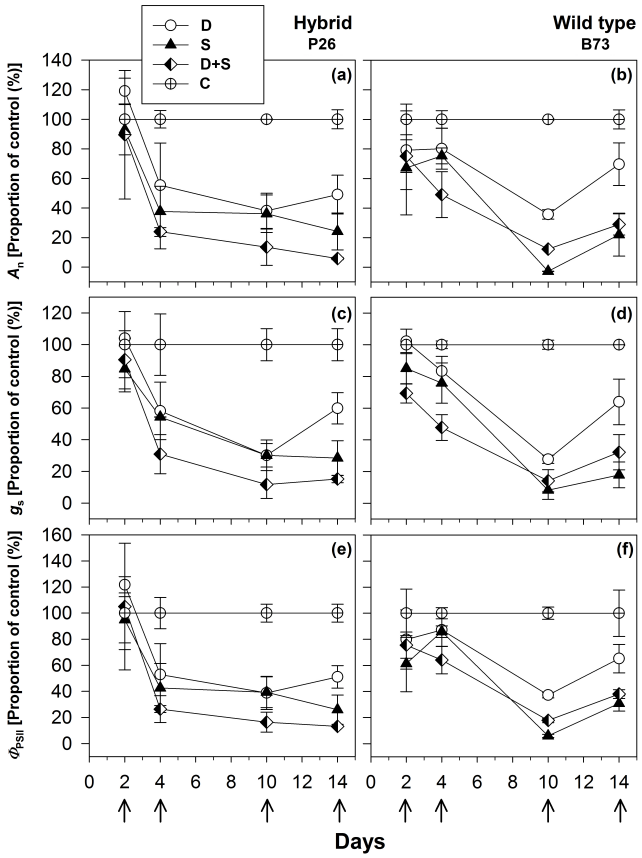

Figure 4 Effects of water stress (D), salt stress (S) and their combination (D+S) on the **(a,b)** net CO_2 assimilation (A_n), **(c,d)** stomatal conductance (g_s) and **(e,f)** quantum efficiency of photosystem II (Φ_{PSII}), for P26 **(left)** and B73 **(right)** genotype plants. Average ±SE values of A_n, g_s and Φ_{PSII} are expressed as a proportion of the control. Arrows show the times of sampling during stress treatments (2,4,10) and recovery (14).

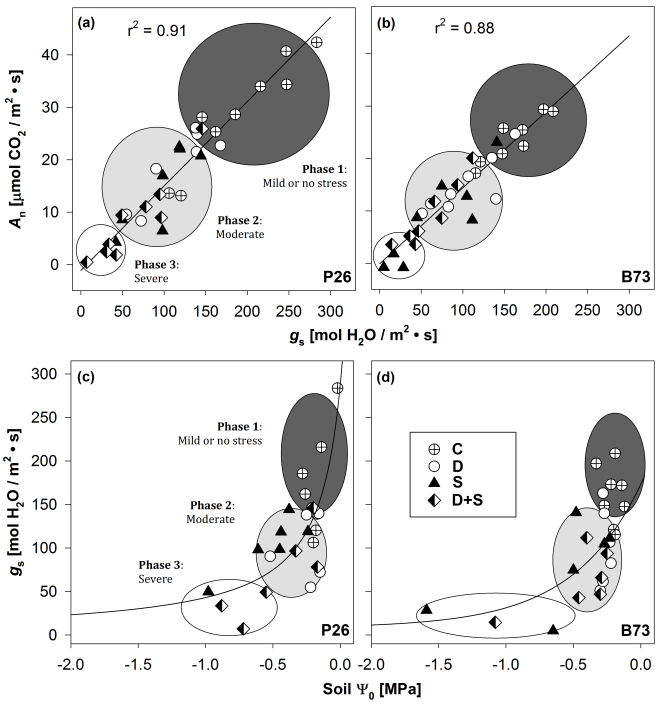

Figure 5 Stomatal conductance (g_s) in **(a)** P26 [y = -2.75+(265.79·x)/(1367.33+x), R^2 = 0.91] and in **(b)** B73 [y = -0.89+(220.5·x)/(1312.94+x), R^2 = 0.88] as a function of net CO_2 assimilation rate (A_n), and of soil osmotic potential (Y_0) in **(c)** P26 [y = 287.83·-0.18/(-0.18+x), R^2 = 0.54] and in **(d)** B73 [y = 299.73·-0.16/(-0.16+x), R^2 = 0.37] in well watered (⊕), water stressed (○) salt stressed (▲) and their combination (◐) plants of the two genotypes, P26 and B73. Each point corresponds to measurements on different sampling days (2, 4, 10 and 14). The curve of best fit for **(a,b)** and **(c,d)** plots was a single rectangular hyperbola and a hyperbola decay function, respectively. Three main regions were distinguished along the curves using g_s as a reference parameter: mild or no stress (Phase 1), moderate (Phase 2) or severe stress (Phase 3).

Gene expression analyses

To assess whether the diverse stress tolerance between the B73 inbred line and the commercial hybrid P26 was related to a difference in the type and timing of specific gene up or/and downregulation, a gene expression analysis was performed. The transcript level of genes known to be differentially expressed by stress or belonging to the main pathways involved in abiotic stress response was analysed using real time quantitative PCR (qRT-PCR). The expression analysis was performed at two time points: after ten days of stress application (T10) and after four days of recovery (T14) from the stress. For each genotype, gene expression was normalized to *GAPC2* transcript quantity and then expressed as the fold change relative to the expression level of the control non-stressed sample at T10 (Supp. Table 2; see Materials and methods for details). To highlight the expression pattern of the genes and for a better understanding of the results obtained, the $\log_2$ of these values were reported in a heatmap as coloured cells: from red, corresponding to negative $\log_2$ fold change values, to blue, corresponding to positive values (Fig.6).

Late embryogenesis abundant (LEA) proteins are major hydrophilic proteins, which can reduce the damage caused by adverse environmental conditions (Liu Y et al. 2013). Both genotypes showed the highest induction of *LEA3* expression in response to S and, at less extent, to the combined D+S stress. B73 upregulated *LEA3* transcript during the stress treatments and the transcript decreased after the recovery. Conversely, hybrid P26 upregulation of *LEA3* transcript was observed during the recovery time. To compare directly the two genotypes, *LEA3* expression of all samples of both B73 and P26 was normalized to the control sample of B73 at T10 (data not shown): *LEA3* transcript level in P26 C sample was one third of the B73 C sample level and the stress-dependent upregulation of the transcript was higher in the hybrid compared to the B73. However, the final transcript levels reached during S in B73 and after the recovery from salt stress in hybrid P26 were similar. Interestingly, the final transcript level reached after the recovery from D+S in hybrid was higher than the final level reached in B73 sample in D+S.

Plasma membrane protein 3 (PMP3) is class of small molecular weight hydrophobic proteins, its members were observed to respond to abiotic stresses in maize (Fu et al. 2012) and one member was reported to participate in maintaining intracellular ion homeostasis in *Arabidopsis* (Mitsuya et al. 2005). *PMP3-4* was a second gene characterized by a diverse expression pattern in B73 in comparison with P26. In control conditions, both genotypes showed an increase in *PMP3-4* transcript level when considering the timing of leaf collection, indicating that the gene expression might be influenced by the developmental stage. At T10 in S and D+S, B73 significantly upregulated *PMP3-4* transcription while in D the transcript increased only slightly. At T10, the P26 hybrid showed the same upregulation trend of B73 but with a considerably lower fold change. After the

recovery from D, both genotypes upregulated *PMP3-4* abundance. B73 slightly downregulated *PMP3-4* abundance after the recovery from S, conversely, the hybrid significantly upregulated it. Both genotypes upregulated *PMP3-4* abundance after the recovery from D+S, but the hybrid had a considerably greater increase.

The HSP70 heat shock protein family encompasses many chaperones, which have an important role in the folding and assembly of proteins during synthesis and in the removal and disposal of non-functional and degraded proteins; they are usually induced by environmental stress and developmental stimuli (Bartels and Sunkar 2005). B73 and hybrid P26 genotypes had a similar trend of *HSP70* transcription after stress applications. Expression induction after the recovery from S and D+S was higher in the hybrid compared to the B73, on the contrary after the recovery from D, B73 showed a greater fold change compared to the hybrid.

Catalases (CAT) eliminate hydrogen peroxide that is produced in plant cells under biotic or abiotic stresses. The expression of maize *CAT1* was shown to be upregulated more in drought-sensitive maize lines than in drought-tolerant lines (Zheng et al 2010). *CAT1* was induced by both S and D+S in B73 and the transcript was maintained at high levels after the recovery. A similar trend but with lower fold changes were observed in the hybrid.

Protein phosphatases 2Cs are serine/threonine phosphatases and their involvement in stress is well known, in particular PP2C action was studied in relation to ABA signalling (Bartels and Sunkar 2005). In both genotypes, D caused an induction of *PP2C* transcription and after the recovery the transcript level was partially decreased. In B73 S caused an upregulation of *PP2C* that remained high after the recovery, while in P26 S did not affect *PP2C* transcript levels compared to control samples. In D+S samples, the transcript level increased only after the recovery in B73, while it increased at T10 in P26.

HVA22 is an early ABA-inducible gene, which is thought to encode for a highly conserved stress-inducible protein playing an important role in protecting cells from damage under stress conditions (Shen et al. 2001). A putative maize *HVA22* gene was analysed (GRMZM2G154735, simply called here *HVA22*). Control samples of both genotypes showed a transcript abundance increase of about three times at the second considered time point. B73 upregulated *HVA22* transcript in all stress conditions, particularly during D and D+S and in samples that experienced the salt application high fold changes were detected after recovery. In the hybrid the upregulation observed was higher compared to B73, especially for D+S, but with similar patterns at both T10 and T14.

The involvement of Ca^{2+} signalling in response to osmotic and ionic stress is well documented (Bartels and Sunkar 2005) and the EF-hand motif is the most common calcium-binding motif found in proteins. In B73 the considered abiotic stresses did not alter the transcription of a gene encoding a putative calcium-binding EF-hand protein (GRMZM5G827398, simply called here *EF-hand*), except

of S that only slightly increased it. In the hybrid, all stress treatments caused an upregulation of the *EF-hand* transcription that was maintained high at T14.

3-Hydroxy-3-methylglutaryl Coenzyme A Reductase (CoAred) is a protein involved in plants isoprenoid metabolism that regulates the synthesis of mevalonic acid (Stermer et al. 1994). A gene encoding a putative CoAred (CO440726, simply called here *CoAred*) was analysed. Following stress treatments, the two genotypes differed in their response: B73 upregulated *CoAred* transcript in S and D+S both at T10 and T14, whereas in the hybrid same treatments slightly affected its expression only at T14.

Sucrose synthase (SUS) is one of the key enzymes involved in sucrose synthesis metabolism. In *Arabidopsis* mature leaves it was reported to be very low expressed under normal physiological conditions, while its expression was stimulated during stress condition (Déjardin et al. 1999). In maize its upregulation was reported under salt stress in roots (Wang et al. 2003). S treatment induced *SUS* expression with different timing in B73 (at T10) and P26 hybrid (T14). D and D+S had a slight effect on both genotypes at T10, while only D+S induced *SUS* at T14 in P26.

IVR1 is a soluble invertase that was previously reported to show drought-mediated increased transcript abundance in the basal leaf meristem (Kakumanu et al. 2012). B73 upregulated *IVR1* transcription mainly during S and the D+S and high fold changes were maintained after the recovery. P26, instead, upregulated *IVR1* transcription at T10 during all kind of stresses, but only in D and S single stresses at T14.

Under water stress, total Glutamine Synthetase activity was observed to be significantly decreased in roots and leaves in wheat and rice (Nagy et al. 2013, Singh and Ghosh 2013). The stress-dependent decrease in maize Glutamine Synthetase1 (*GLN1*) expression was delayed in P26 compared to B73, with the inbred line responding already at T10 and the hybrid responding at T10 in D+S and only at T14 in D and S.

The Rab protein family is the largest member of the Ras superfamily of monomeric G proteins, also referred to as small ATPases. Along with their essential function in intracellular vesicular trafficking, they have also been implicated in defence and stress signalling pathways (Hong et al. 2013). The applied abiotic stresses only slightly decreased the expression of a putative Rab GTPase ecoding gene (GRMZM2G018619, simply called here *Rab GTPase*) in P26 after S and D+S application.

It has been proposed that regulation of expansin mRNA pools likely contributes to fast adjustment of cell wall-loosening in maize under water deficit conditions (Geilfus et al. 2010). In our study, both genotypes significantly upregulated *ß-EXP7* abundance only in D at T10. B73 strongly reduced its expression in all the three treatments at T14, while only weak variation characterized the hybrid.

Tonoplast-associated Na^+/H^+-antiporters are responsible for detoxifying the cytoplasm by pumping Na^+ into the vacuole. Efficient Na^+ exclusion significantly improves the salt tolerance in maize. Under salt stress a drought-sensitive maize line was reported to induce the expression of these Na^+/H^+-antiporters only in roots and not in shoots, while no changes were reported in a drought-tolerant line (Zörb et al. 2005). In all stress treatments B73 upregulated *NHX4* and *NHX5* (*NHX4-5*) expression at both T10 and T14, especially in S. Conversely, hybrid P26 did not upregulated *NHX4-5* at T10, but at recovery and exclusively in S and D+S.

Finally, the expression patterns of four genes involved in gene expression regulation and protein-protein interaction were analysed in response to stresses. Two were maize epiregulators: *RMR6,* coding for a subunit of Pol IV (Erhard et al. 2009), and *HDA108,* coding for a histone deacetylases (Forestan et al. submitted). A putative RING Zn-finger coding gene (GRMZM2G148908, simply called here *RING fing*) was analysed because the overexpression of another maize gene of the same family was observed to be involved in drought tolerance (Liu J et al. 2013) and a putative RNA-binding KH domain-containing protein coding gene (AC218972.3_FG007, simply called here *RNA-binding KH*) was analysed because a gene with the same domain was reported to participate to stress response in *Arabidopsis* (Guan et al. 2013). With a few exceptions, these genes were not differentially expressed in both genotypes, both at T10 and T14.

Taken together our results indicated that gene expression was modulated in response to the applied stresses in the two genotypes. However, gene expression patterns were not coincident and reflected the different capacity of the two genotypes to cope with D, S and D+S and to differently respond at the recovery.

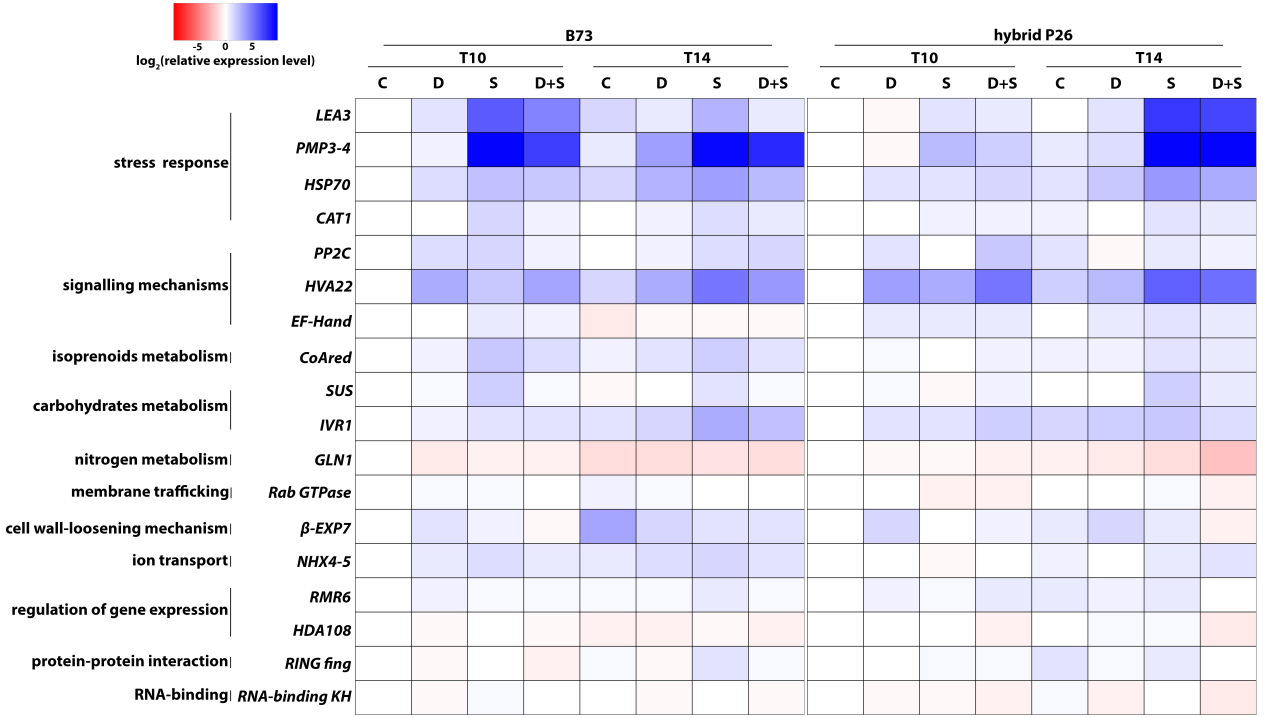

Figure 6 Heat map representing the relative quantification of gene expression in maize leaves of B73 inbred line and hybrid P26 at two time points, after 10 days of stress (T10) and 4 days of recovery (T14) following the application of drought (D), salinity (S) and drought+salinity (D+S). The maize *GAPC2* gene was selected as internal control. Each experiment was run in triplicate. Data from qRT-PCR experiments were analyzed according to the Pfaffl method and gene expression was calculated as the fold change (FC) relative to the expression level of the control non-stressed sample (C) of the same genotype at T10 (±SE). The $\log_2$(FC) values were reported as coloured cells: blue colour represents higher relative expression and red colour represents lower relative expression.

Principal Component Analysis (PCA)

The PCA was done to establish the general structure of the interdependences existing between the changes in the levels of genetic stress markers and the fluctuations in the selected environmental parameters associated with D, S and D+S. The PCA referred to those markers related to ion homeostasis and to the maintenance of cellular osmotic balance: water content in plants (estimated as dry weights in shoots), inorganic ions related to stress applications (leaf and root Cl^-, leaf and root Na^+, root K^+/Na^+ ratios) and A_n. We also included in this analysis a set of genes markers of stress and belonging to different stress responsive pathways. As determined by qRT-PCR, *PMP3-4*, *CoAred* and *SUS* presented a dissimilar expression patterns in the two genotypes, B73 and the hybrid, in response to stress; *HSP70* had upregulation levels mainly related to the kind of applied stress and *CAT1* shared the same expression pattern in the two genotypes.

Application of PCA to data allowed extracting 3 components explaining more than 80 % of the total variability. The first component (PC1), which accounted for the 56 % of the variance, was highly correlated (factor loadings ≥ 0.78) with leaf characteristics: Na^+ and Cl^- contents and upregulated stress-responsive genes (*CAT1* and *CoAred*). The second (PC2) and third (PC3) components explained the 19 % and 8.8 % of the variance, respectively, and were correlated with Na^+ and Cl^- contents in roots and with *PMP3-4* and *HSP70* expression.

Plotting data according to PC1 and PC2 (Fig.7) allowed identifying a cluster in quadrant 3, including mainly the plants not subjected to S irrespectively of the recovery application. They were associated to high *An* and leaf d.w. values. In the opposite quadrant 1, were grouped B73 plants under S and D+S treatments before the recovery. Salt concentration in leaf (Na^+ and Cl^-) and expression of *PMP3-4*, *CAT1* and *SUS* were the primary clustering factors. After the recovery, D+S B73 was shifted toward the group of unstressed plants of quadrant 3, while S B73 was positioned in quadrant 2 driven by the reduction of salt concentration in roots (PC2<0) and persisting of high Na^+ and Cl^- concentration in leaves (PC1>1.5). Finally, hybrid under S and D+S was clustered in quadrant 4 by both higher and lower concentration of Na^+ and Cl^- in root and leaf, respectively. The effect of recovery was depicted by the shift of the hybrid under S treatment in quadrant 3 whereas hybrid under D+S treatment remains unaffected.

The analysis confirmed that the inbred line B73 was very sensitive to S and more sensitive to S than the combined D+S in our condition. The recovery from the D+S condition showed a positive effect on this genotype, while less evident was the effect of recovery after the S application. The analysis also indicated that the hybrid recovered very well from the S and was only slightly affected by the D and D+S.

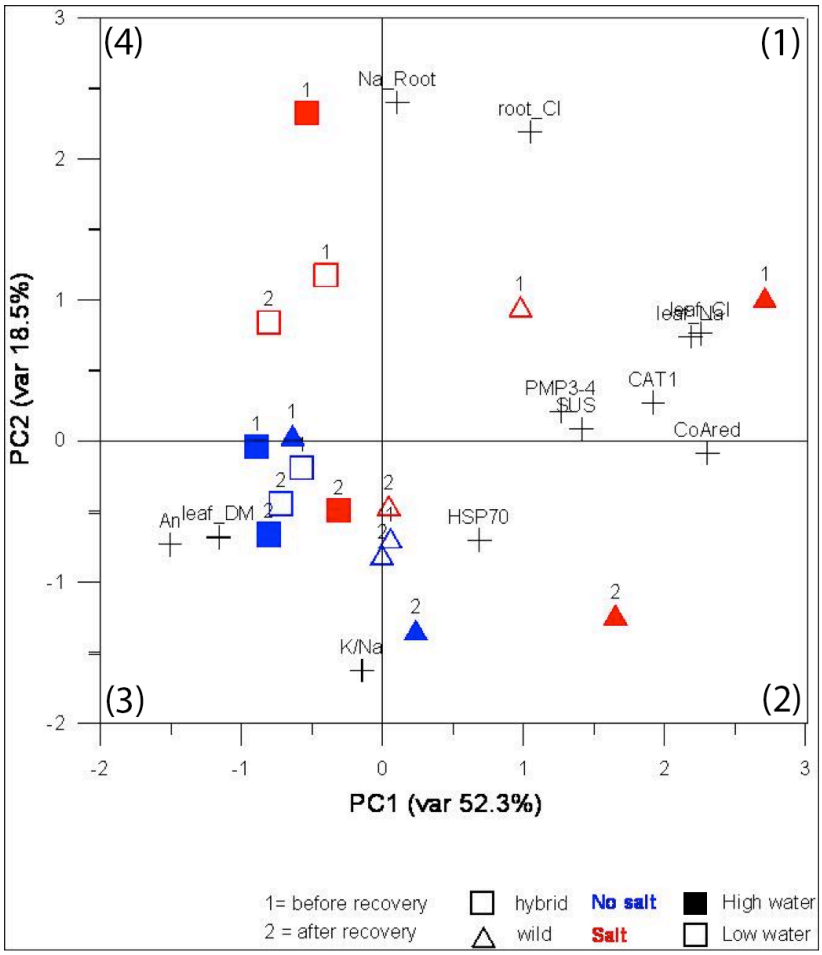

Figure 7 Site score plot of the studied variables on the two principal components (PC1, PC2). PCAs included, as the analysed variables, those related to osmotic adjustment or those related to gene expression. Plotted points belong to the genotypes (squares and triangles), time points during stress application (1 and 2) and stress types (blue and red, background and border) variables. **(1)-(2)-(3)-(4)** are quadrants 1-2-3-4.

Gene	B73								hybrid P26							
	T10				T14				T10				T14			
	C	D	S	D+S	C	D	S	D+S	C	D	S	D+S	C	D	S	D+S
LEA3	1 (±0.27)	1.93 (±0.37)	83.12 (±14.79)	25.99 (±4.1)	3.01(±0.5)	1.58 (±0.27)	7.77 (±1.28)	1.72 (±0.28)	1 (±0.07)	0.85 (±0.08)	1.98 (±0.46)	1.62 (±0.15)	1 (±0.2)	2.09 (±0.3)	192.16 (±18.2)	126.1 (±9.18)
PMP3-4	1 (±0.21)	1.5 (±0.2)	871.9 (±116.01)	160.48 (±23.15)	1.74 (±0.35)	12.31 (±1.6)	653.37 (±89.15)	255.53 (±34.21)	1 (±0.17)	0.86 (±0.13)	6.16 (±0.69)	3.63 (±0.64)	1.63 (±0.2)	2.58 (±0.49)	773.81 (±82.08)	857.69 (±133.06)
HSP70	1 (±0.12)	2.58 (±0.28)	5.07 (±0.55)	4.23 (±0.42)	3.08 (±0.34)	7.91 (±0.82)	11.99 (±1.48)	6.44 (±0.85)	1 (±0.1)	2 (±0.2)	1.9 (±0.17)	3.08 (±0.24)	2.05 (±0.21)	4.22 (±0.42)	16.01 (±1.57)	8.12 (±0.74)
CAT1	1 (±0.09)	0.98 (±0.08)	3.01 (±0.19)	1.49 (±0.1)	1.05 (±0.08)	1.4 (±0.1)	2.63 (±0.16)	1.8 (±0.24)	1 (±0.06)	1 (±0.08)	1.46 (±0.08)	1.38 (±0.2)	1.55 (±0.1)	0.93 (±0.08)	1.99 (±0.28)	1.67 (±0.18)
PP2C	1 (±0.11)	2.55 (±0.35)	2.96 (±0.33)	1.39 (±0.12)	1.07 (±0.1)	1.35 (±0.12)	2.54 (±0.29)	2.98 (±0.33)	1 (±0.1)	1.88 (±0.17)	0.93 (±0.07)	4.17 (±0.35)	2 (±0.2)	0.83 (±0.08)	1.69 (±0.12)	1.47 (±0.15)
HVA22	1 (±0.11)	9.41 (±0.85)	4.04 (±0.35)	10.16 (±0.92)	3.04 (±0.24)	8.01 (±0.63)	37.83 (±2.44)	15.25 (±0.98)	1 (±0.44)	12.74 (±3.18)	9.43 (±2.43)	35.24 (±8.35)	3.75 (±0.95)	5.74 (±1.55)	58.46 (±14.19)	43.4 (±10.61)
EF-Hand	1 (±0.27)	1 (±0.17)	1.59 (±0.27)	1.38 (±0.28)	0.6 (±0.11)	0.87 (±0.15)	0.9 (±0.14)	0.88 (±0.18)	1 (±0.13)	1.77 (±0.18)	1.73 (±0.17)	1.74 (±0.29)	0.95 (±0.11)	1.7 (±0.21)	2.18 (±0.31)	1.65 (±0.17)
CoAred	1 (±0.11)	1.48 (±0.21)	4.63 (±0.56)	2.45 (±0.3)	1.57 (±0.23)	1.94 (±0.23)	3.65 (±0.46)	2.22 (±0.3)	1 (±0.15)	1.13 (±0.13)	1(±0.1)	1.54 (±0.15)	1.48 (±0.22)	1.45 (±0.21)	2.22 (±0.22)	1.76 (±0.23)
SUS	1 (±0.09)	1.31 (±0.09)	3.39 (±0.24)	1.26 (±0.12)	0.81 (±0.08)	1.03 (±0.07)	2.21 (±0.15)	1.17 (±0.07)	1 (±0.03)	1.27 (±0.06)	0.79 (±0.04)	1.36 (±0.05)	1.01 (±0.06)	0.95 (±0.06)	3.44 (±0.24)	1.8 (±0.1)
IVR1	1 (±0.17)	1.34 (±0.24)	2.09 (±0.32)	2.15 (±0.32)	1.96 (±0.25)	2.96 (±0.47)	8.01 (±0.96)	5.16 (±0.65)	1 (±0.07)	2.14 (±0.23)	2.1 (±0.17)	3.52 (±0.33)	2.76 (±0.24)	3.52 (±0.3)	4.61 (±0.33)	2.35 (±0.28)
GLN1	1 (±0.08)	0.57 (±0.04)	0.64 (±0.07)	0.76 (±0.07)	0.44 (±0.02)	0.42 (±0.05)	0.51 (±0.03)	0.41 (±0.03)	1 (±0.04)	0.89 (±0.03)	0.86 (±0.03)	0.67 (±0.04)	0.76 (±0.04)	0.58 (±0.03)	0.39 (±0.02)	0.21 (±0.02)
Rab GTPase	1 (±0.15)	1.28 (±0.16)	1.22 (±0.16)	1.05 (±0.12)	1.33 (±0.18)	1.25 (±0.12)	0.98 (±0.11)	0.94 (±0.14)	1 (±0.05)	0.97 (±0.05)	0.75 (±0.04)	0.76 (±0.05)	0.97 (±0.03)	1.02 (±0.06)	1.21 (±0.1)	0.65 (±0.06)
β-EXP7	1 (±0.21)	1.91 (±0.37)	1.32 (±0.21)	0.79 (±0.11)	11.36 (±1.59)	2.74 (±0.4)	1.98 (±0.44)	2.1 (±0.52)	1 (±0.37)	2.85 (±0.74)	1 (±0.27)	1.44 (±0.47)	1.86 (±0.54)	2.79 (±0.6)	1.76 (±0.4)	0.72 (±0.2)
NHX4-5	1 (±0.12)	1.59 (±0.34)	2.39 (±0.38)	1.77 (±0.19)	1.74 (±0.19)	2.62 (±0.31)	2.97 (±0.29)	2.12 (±0.34)	1 (±0.16)	0.92 (±0.15)	0.84 (±0.1)	0.98 (±0.1)	1.33 (±0.13)	0.98 (±0.09)	1.68 (±0.17)	2.05 (±0.28)
RMR6	1 (±0.2)	1.54 (±0.21)	1.25 (±0.24)	1.13 (±0.21)	1.2 (±0.17)	1.29 (±0.2)	1.62 (±0.27)	1.22 (±0.17)	1 (±0.09)	1.31 (±0.13)	1.1 (±0.07)	1.81 (±0.19)	1.62 (±0.16)	1.47 (±0.12)	1.71 (±0.26)	0.92 (±0.13)
HDA108	1 (±0.11)	0.91 (±0.13)	1.03 (±0.09)	0.88 (±0.07)	0.72 (±0.08)	0.75 (±0.05)	0.91 (±0.06)	0.69 (±0.07)	1 (±0.16)	1.02 (±0.12)	1.01 (±0.15)	0.66 (±0.12)	1.06 (±0.11)	1.11 (±0.14)	1.27 (±0.19)	0.55 (±0.13)
RING fing	1 (±0.18)	0.85 (±0.11)	1.09 (±0.17)	0.67 (±0.1)	1.26 (±0.19)	0.88 (±0.19)	1.93 (±1.23)	1.27 (±0.15)	1 (±0.03)	0.97 (±0.09)	1.29 (±0.17)	1.16 (±0.08)	2.16 (±0.19)	1.25 (±0.04)	1.57 (±0.26)	1.04 (±0.13)
RNA-binding KH	1 (±0.16)	0.88 (±0.12)	1.11 (±0.13)	0.93 (±0.14)	0.96 (±0.14)	0.77 (±0.09)	0.96 (±0.17)	0.78 (±0.08)	1 (±0.15)	0.9 (±0.13)	0.78 (±0.13)	0.72 (±0.15)	1.16 (±0.22)	0.72 (±0.12)	0.99 (±0.13)	0.6 (±0.06)

Supplemental Table 2 Relative quantification of gene expression in maize leaves of B73 inbred line and hybrid P26 at two time points, after 10 days of stress (T10) and 4 days of recovery (T14) following the application of drought (D), salinity (S) and drought+salinity (D+S). The maize *GAPC2* gene was selected as internal control. Each experiment was run in triplicate. Data from qRT-PCR experiments were analyzed according to the Pfaffl method and gene expression was calculated as the fold change relative to the expression level of the control non-stressed sample (C) of the same genotype at T10 (±SE).

Discussion

This study was conceived to compare the response of two different maize genotypes (the B73 inbred line for which genomic tools are largely available and the P26 commercial hybrid) to a progressive time-limited (10 days) application of drought, salinity and a combination of both stresses. These genotypes were already known having different ability to cope with stress, although the genetic basis of P26 tolerance to stress was not known. The idea was to apply the stresses reproducing the real stressful field conditions experienced by maize plants during growth in our region and assess whether, after 4 days from the removal of the stresses, plants could recover to complete their life cycle. Both drought and salinity are major abiotic stresses that limit growth and affect crop productivity in many areas of the world. They are due to the reduced availability of water and increasing use of poor quality of water for irrigation and soil salinization (Trembort et al. 2014, Rozema and Flowers 2008). With the aim to investigate the effect of these two abiotic stresses, this work compared a realistic stress protocol (for salinity alone S, drought alone D and combined drought plus salinity stress D+S) simulating a field environment, in which combined salinity plus drought is achieved watering with a reduced quantity of saline water. The agronomic data demonstrated that the combined stress D+S represented a less severe salinity stress condition for the plants, due to the lower ECs values reached with this treatment than S. As outlined in previous works, apply realistic protocols, standardizing the measure and description of plant stresses makes findings in crops more valuable for data comparisons or for translating the findings to crop breeding (Skirycz et al. 2011, Nelson et al. 2007, Talame et al. 2007).

To achieve our primary objective we monitored the stress response using a combination of agronomic, physiological and genetic parameters and elaborated the retrieved data sets to depict a complete picture of stress response and recovery ability of the two genotypes. Firstly, the stress conditions were analyzed in terms of effect on plant growth, indicating that all the applied stresses were effective in limiting both shoot and root growth in the hybrid and arresting the growth in the inbred line. After four days from the removal of the stress conditions, B73 leaf d.w. slightly increased only for S and D+S while the hybrid showed a better recovery capability in both D and S, but not in D+S. These results indicated that a longer recovery time is needed to the inbred line shoots to start growing again. Even more complex was the recovery capability at root level, since no effect on growth after stress removal was observed in both genotypes, with the exception of hybrid following D. These observations on growth inhibition are consistent with the physiological data on net assimilation, stomatal conductance and quantum efficiency of photosystem II. Furthermore, these data indicated that the tolerance to stress is not necessarily associated to a prompt recovery

capability of a genotype (Efeoglu et al. 2009, Nayyar and Gupta, 2006). However, it would be important to breed for maize varieties with a high recovery capability, especially in those regions where drought and salinity stresses can have limited time duration in the growing season because water availability is naturally restored after a period of drought (Nelson et al. 2007).

When uptake and translocation of ions were considered, it was evident that the two genotypes had significantly different concentrations of Na^+ and Cl^- in their tissues, both during S and D+S, and also after recovery from these two stress conditions. Interestingly, we observed that B73 and the hybrid accumulated similar concentration of Na^+ at root level; however Na^+ concentration was significantly different in the leaves of the two genotypes, suggesting that B73 accumulated higher level of Na^+ in the leaf through translocation from the roots, during S and D+S. As expected, after the recovery from S and D+S, the Na^+ concentration in B73 root dropped to C level and clearly decreased in the leaf, remaining, however, at high levels when compared with both the C and hybrid. A very similar trend was observed when considering the Cl^- accumulation in the roots and leaves in the two genotypes that had equal concentration of Cl^- in their roots but a drastically different concentration of this ion in their leaves during both S and D+S. Also in the case of Cl^- the recovery determined an evident reduction in this ion concentration in B73 root and leaf, where the Cl^- concentration remained very high after the removal of S. The data on ions uptake and translocation clearly indicated that the different ability to cope with stress, particularly with S and D+S, of the two genotypes is somehow associated to a different dynamic in Na^+ and Cl^- translocation in the shoot. Indeed the hybrid accumulated both Na^+ and Cl^- in roots and might not (or only partially) translocated them to the shoot in S and D+S compared to the D and C conditions while, under the same stress conditions, B73 increased the amount of Na^+ and Cl^- in roots and particularly in leaves, where Cl^- reached a very high concentration. It is well known that an important mechanism of salinity tolerance is the ability to limit the quantity of Na^+ entering the plant through the roots (Laurie et al. 2002, Tester and Davenport 2003, Munns and Tester 2008). In particular, the control of Na^+ transport by secreting and sequestering it in cellular compartments such as tissues, cells or organelles where Na^+ is less toxic, is also critical to cope better with salinity (Munns and James 2003; Parida and Das 2005). Indeed, salinity stress is due to the accumulation of high concentrations of Na^+ in the leaf cell cytoplasm (Jha et al. 2010). However, in some species Cl^- is the main stressful ion (Prior et al. 2007) because these species are better at excluding Na^+ than Cl^- (Munns and Tester 2008). When both Na^+ and Cl^- are taken up in large amount by the roots, they negatively affect plant growth by impairing metabolic processes and decreasing photosynthetic efficiency (Deinlein et al. 2014). Interestingly, in our study a clear relationship between Na^+ and Cl^- exclusion and salinity tolerance in P26 hybrid does exist. Further investigations are needed for the understanding of the mechanisms involved in the uptake and movement of Na^+ and Cl^- throughout the plant of P26 hybrid as

well as the identifications of genes involved in Na^+ and Cl^- homoeostasis in this genotype, to elucidate the mechanisms that mediated its salinity tolerance.
To assess the water and salt stresses actually endured by plants, net assimilation, stomatal conductance and quantum efficiency of photosystem II were recorded. These parameters provided precise information on the drought and salinity stress intensity occurring in the plant, and allowed the definition of three phases (mild or no stresses, moderate and severe) during the progressive application of D, S and D+S. The physiologically parameters confirmed that P26 was less tolerant to D+S and B73 very sensitive to S and enabled to establish a more accurate correlation between gene expression variation and stress progression. It has been observed that the kinetics of stress treatments are particularly important and should be carefully considered in experimental designs, especially when genes expression analyses are performed to identify stress responsive genes (Deyholos 2010). In our work, the genetic analysis was performed determining the transcript levels of genes previously showed to be markers of drought and salinity stresses and belonging to different stress response pathways. It is worth noting that in many out of the previous works the expression of these and others stress-marker genes was monitored on samples collected from plants subjected to high-intensity stress treatments frequently developed in a very short time after the application of the stress (Kawasaki et al. 2001, Kreps et al. 2002, Ozturk et al. 2002, Seki et al. 2002, Rabbani et al, 2003, Rensink and Buell 2005), whereas we monitored the transcript level at the end of the progressive stress application (T10) corresponding to the severe phase of stress and after four days of recovery from the stresses (T14). Therefore, due to the particular design of this experiment, gene expression was specifically affected both by the stress duration and severity and it cannot be excluded that some drought and/or tolerance-related genes could be activated earlier, to prepare the plant to a developing water and salt stress. This expression analysis permitted both to confirm the stress-marker nature of some transcripts for a specific type of stress and highlight possible differences between the expressions of these marker genes in the two genotypes, having a high correlation with the stress condition at physiological level. The transcript level variations observed at the two considered time points in our experiments were broad and depended upon both the applied stress and the genotype. In our conditions, some genes were confirmed to be good markers of stress: *HVA22* was upregulated at T10 and T14 in D, S and D+S in both genotypes, confirming previous observation in other plant species (Shen et al. 2001). *EF-hand* was a good marker of the three stresses in P26 at both time points: it is well known that most of the Ca^{2+} sensors bind Ca^{2+} using a helix-loop-helix motif termed as the 'EF hand' or the elongation factor, which binds a single Ca^{2+} molecule with high affinity (Tutejaa and Mahajan 2007) and it is also well known that Ca^{2+} signalling play a pivotal role in stress response (Knight H 2000). Also *LEA3* and *HSP70* represented good markers of S and D+S, but with a distinction between the two genotypes: both these genes were upregulated in B73 at T10 while their transcript

increase was more evident at T14 in P26, indicating that the two genotypes might regulate the expression of these genes, commonly expressed in different stress conditions (Liu et al. 2013; Wang et al. 2003), with different timing. Previous observations reported different transcript levels of *CAT1* in stress-susceptible and tolerant maize inbred lines (Zheng et al. 2010), the upregulation of maize *SUS* a few hours after salt stress application (Wang et al. 2003) and the increased transcript abundance of *IVR1* in the leaf meristem following drought stress (Kakumanu et al. 2012). These genes appeared to be good markers of salinity stress at least in B73: *CAT1*, *SUS*, *IVR1* and also *CoA-red* transcripts were all upregulated in S and less in D+S at both T10 and T14, suggesting that four days of recovery were not a sufficient time to resume the transcript levels observed in the control condition. In P26 the transcripts of these genes had more variable trends, depending on the stress but also on the time point and the overall results indicated that they are not good markers of stress for this genotype. Some genes were downregulated during stress applications: this was the case of *GLN1*, whose trend confirmed reported data (Singh and Ghosh, 2013), and *Rab GTPase* that was slightly downregulated in S and D+S in P26 but not in B73. Being part of a large protein family, it is possible that other Rab-GTPases could be differently regulated in B73 as observed in other crop species (Hong et al. 2013). Finally, *ß-EXP7* was upregulated by the D in both genotypes, but strongly downregulated in B73 after the recovery from all the three stresses. Also in this case the presence of many paralog genes in the maize genome could result in different expression timing and levels of the different genes, as previously reported in different maize genotypes (Geilfus et al. 2010).

PCA, that was used to combine some selected and correlated parameters, clearly showed the exiting difference in stress tolerance between the two genotypes: it associated the tolerance of the hybrid to leaf d.w. and *An*. It also correlated B73 low tolerance to the Cl^- and Na^+ concentration in leaf and root and to the expression of genes that are good marker of stress for the inbred line. Interestingly, it highlighted the effect of recovery that was evident for the hybrid under S, whereas it had no effect under D+S.

The ultimate aim of this work was to set reproducible D, S and D+S protocols in which these three time-limited stress conditions were verified at agronomic, physiological and genetic levels, allowing reproducing these stress protocols in following experiments and analyze their effect at epigenetic and genetic genome-wide level. It would be interesting to better dissect the characteristic of the recovery response in both tolerant e susceptible genotypes, to evaluate the effect of these transitory stresses on plant productivity and investigate whether a transitory stress can cause a sort of “memory” for subsequent stressful events of the same kind. Indeed the stress protocols described in this work were set and reproduced for a genome-wide analysis that was performed using the B73 inbred line coupling data of RNA-seq, sRNA-seq and ChIP–seq that are currently being processed with the aim to continue our investigations.

References

Ahuja I, de Vos RC, Bones AM, Hall RD: Plant molecular stress responses face climate change. *Trends Plant Sci* 2010, 15:664–74

Araus J-L: The problem of sustainable water use in the Mediterranean and research requirements for agriculture. *Ann appl Biol* 2004, 144:259–272

Bänziger M, Araus J-L: Recent advances in breeding maize for drought and salinity stress tolerance. In: Jenks MA, Hasegawa PM, Jain SM, editors. Advances in molecular breeding toward drought and salt tolerant crops. *Springer Netherlands* 2007, pp. 587-601

Bartels D, Sunkar R: Drought and salt tolerance in plants. *Crit Rev Plant Sci* 2005, 24, 23–58

Corwin DL: Soil salinity. In: Lehr JH, Keeley J, editors. Water Encyclopedia: Surface and Agricultural Water. *John Wiley & Sons* 2005

Deikman J, Petracek M, Heard JE: Drought tolerance through biotechnology: improving translation from the laboratory to farmers' fields. *Curr Opin Biotechnol* 2012, 23: 243–250

Deinlein U, Stephan AB, Horie T, Luo W, Xu G, Schroeder JI: Plant salt-tolerance mechanisms. *Trends Plant Sci* 2014, 19(6):371-9

Déjardin A, Sokolov LN, Kleczkowski LA: Sugar/osmoticum levels modulate differential abscisic acid-independent expression of two stress-responsive sucrose synthase genes in *Arabidopsis*. *Biochem J* 1999, 1;344 Pt 2:503-9

Deyholos MK: Making the most of drought and salinity transcriptomics. *Plant Cell Environ* 2010, 33(4):648-54

Efeoglu B, Ekmekçi Y, Çiçek N: Physiological responses of three maize cultivars to drought stress and recovery. *S Afr J Bot* 2009, 75:34–42

Elmetwalli AMH, Tyler AN, Hunter PD, Salt CA: Detecting and distinguishing moisture-and salinity-induced stress in wheat and maize through in situ spectroradiometry measurements. *Remote Sens Lett* 2012, 3,363-372

Erhard KF Jr, Stonaker JL, Parkinson SE, Lim JP, Hale CJ, Hollick JB: RNA polymerase IV functions in paramutation in *Zea mays*. *Science* 2009, 323: 1201–5

FAO. Crops and drops making the best use of water for agriculture. World Food Day. Rome, Italy; 2002. ftp://ftp.fao.org/agl/aglw/docs/cropsdrops_e.pdf.

Forestan C, Farinati S, Rouster J, Lassagne H, Lauria M, Dal Ferro N, Varotto S: ZmHDA108 has an active role both in controlling maize plant vegetative and reproductive development and setting the histone code. Submitted

Fu J, Zhang DF, Liu YH, Ying S, Shi YS, Song YC, Li Y, Wang TY: Isolation and characterization of maize PMP3 genes involved in salt stress tolerance. *PLoS One* 2012, 7(2):e31101

Gee G, Campbell M, Campbell G, Campbell J: Rapid measurement of low soil water potentials using a water activity meter. *Soil Sci Soc Am J* 1992, 56, 1068-1070

Geilfus CM, Zörb C, Mühling KH: Salt stress differentially affects growth-mediating β-expansins in resistant and sensitive maize (*Zea mays* L.). *Plant Physiol Biochem* 2010, 48(12):993-8

Genty B, Briantais JM, Baker NR: The relationship between the quantum yield of photosynthetic electron transport and quenching of chlorophyll fluorescence. *Biochim Biophys Acta* 1989, 990:87–92

Grattan SR, Grieve CM: Salinity–mineral nutrient relations in horticultural crops. *Sci Hortic-Amsterdam* 1999, 78, 127-157

Guan Q, Wen C, Zeng H, Zhu J: A KH domain-containing putative RNA-binding protein is critical for heat stress-responsive gene regulation and thermotolerance in *Arabidopsis*. *Mol Plant* 2013, 6(2):386-95

Hong MJ, Lee Ym, Son YS, Im CH, Yi YB, Rim YG, Bahk JD, Heo JB: Rice Rab11 is required for JA-mediated defense signaling. *Biochem Biophys Res Commun* 2013, 17;434(4):797-802

Hu Y, Burucs Z, von Tucher S, Schmidhalter U: Short-term effects of drought and salinity on mineral nutrient distribution along growing leaves of maize seedlings. *Environ Exp Bot* 2007, 60, 268-275

Jha D, Shirley N, Tester M, Roy SJ: Variation in salinity tolerance and shoot sodium accumulation in *Arabidopsis* ecotypes linked to differences in the natural

expression levels of transporters involved in sodium transport. *Plant Cell Environ* 2010, 33(5):793-804

Kaiser HF: The application of electronic computers to factor analysis. *Educ Psychol Meas* 1960, 20, 141-151

Kaiser HF: An index of factorial simplicity. *Psychometrika* 1974, 39, 31-36

Kakumanu A, Ambavaram MM, Klumas C, Krishnan A, Batlang U, Myers E, Grene R, Pereira A: Effects of drought on gene expression in maize reproductive and leaf meristem tissue revealed by RNA-Seq. *Plant Physiol* 2012, 160(2):846-67

Kawasaki S, Borchert C, Deyholos M, Wang H, Brazille S, Kawai K, Galbraith D, Bohnert HJ: Gene expression profiles during the initial phase of salt stress in rice. *Plant Cell* 2001, 13(4):889-905

Knight H: Calcium signaling during abiotic stress in plants. *Int Rev Cytol* 2000, 195:269-324

Kramer PPJ, Boyer JS: Water relations of plants and soilds. *AP* 1995, San Diego, CA, USA

Kreps JA, Wu Y, Chang HS, Zhu T, Wang X, Harper JF: Transcriptome changes for *Arabidopsis* in response to salt, osmotic, and cold stress. *Plant Physiol* 2002, 130: 2129–2141

Laurie S, Feeney KA, Maathuis FJM, Heard PJ, Brown SJ, Leigh RA: A role for HKT1 in sodium uptake by wheat roots. *Plant J* 2002, 32:139–149

Liu Y, Wang L, Xing X, Sun L, Pan J, Kong X, Zhang M, Li D: ZmLEA3, a multifunctional group 3 LEA protein from maize (*Zea mays* L.), is involved in biotic and abiotic stresses. *Plant Cell Physiol* 2013, 54(6):944-59

Liu J, Xia Z, Wang M, Zhang X, Yang T, Wu J: Overexpression of a maize E3 ubiquitin ligase gene enhances drought tolerance through regulating stomatal aperture and antioxidant system in transgenic tobacco. *Plant Physiol Biochem* 2013, 73:114-20

Mitsuya S, Taniguchi M, Miyake H, Takabe T: Disruption of RCI2A leads to over-accumulation of Na+ and increased salt sensitivity in *Arabidopsis thaliana* plants. *Planta* 2005, 222(6):1001-9

Munns R: Comparative physiology of salt and water stress. *Plant Cell Environ* 2002, 25, 239-250

Munns R, James RA: Screening methods for salinity tolerance: a case study with tetraploid wheat. *Plant and Soil* 2003, 253,201–218

Munns R, Tester M: Mechanisms of salinity tolerance. *Annu Rev Plant Biol* 2008, 59:651-81

Nagy Z, Németh E, Guóth A, Bona L, Wodala B, Pécsváradi A: Metabolic indicators of drought stress tolerance in wheat: glutamine synthetase isoenzymes and Rubisco. *Plant Physiol Biochem* 2013, 67:48-54

Nayyar H, Gupta D: Differential sensitivity of C3 and C4 plants to water deficit stress: Association with oxidative stress and antioxidants. *Environ Exp Bot* 2006, 58 106–113

Nelson DE, Repetti PP, Adams TR, Creelman RA, Wu J, Warner DC, Anstrom DC, Bensen RJ, Castiglioni PP, Donnarummo MG, Hinchey BS, Kumimoto RW, Maszle DR, Canales RD, Krolikowski KA, Dotson SB, Gutterson N, Ratcliffe OJ, Heard JE: Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance and lead to improved corn yields on water-limited acres. *Proc Natl Acad Sci USA* 2007, 16;104(42):16450-5

Neves-Piestun BG, Bernstein N: Salinity-induced changes in the nutritional status of expanding cells may impact leaf growth inhibition in maize. *Funct Plant Biol* 2005, 32, 141-152

Nicoletto C, Santagata S, Bona S, Sambo S: Influence of cut number on qualitative traits in different cultivars of sweet basil. *Ind Crop Prod* 2013, 44, 465-472

Ozturk ZN, Talamé V, Deyholos M, Michalowski CB, Galbraith DW, Gozukirmizi N, Tuberosa R, Bohnert HJ: Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley. *Plant Mol Biol* 2002, 48:551-573

Parida AK, Das AB: Salt tolerance and salinity effects on plants: a review. Ecotox Environ Safe 2005, 60, 324-349

Pfaffl MW: A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 2001, 29(9):e45

Prior LD, Grieve AM, Bevington KB, Slavich PG: Long-term effects of saline irrigation water on 'Valencia' orange trees: relationships between growth and yield, and salt levels in soil and leaves. *Aust J Agr Res* 2007, 58:349-358

Rabbani MA, Maruyama K, Abe H, Khan MA, Katsura K, Ito Y, Yoshiwara K, Seki M, Shinozaki K, Yamaguchi-Shinozaki K: Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. *Plant Physiol* 2003, 133: 1755–1767

Rensink WA, Buell CR: Microarray expression profiling resources for plant genomics. *Trends Plant Sci* 2005, 10(12):603-9

Rhoades JD, Chanduvi F, Lesch SM: Soil salinity assessment: Methods and interpretation of electrical conductivity measurements. FAO 1999, Rome, Italy

Rozema J, Flowers T: Crops for a salinized world. *Science* 2008, 322: 1478–1480

Salekdeh GH, Reynolds M, Bennett J, Boyer J: Conceptual framework for drought phenotyping during molecular breeding. *Trends Plant Sci* 2009, 14(9):488-96

Scudiero E, Berti A, Teatini P, Morari F: Simultaneous monitoring of soil water content and salinity with a low-cost capacitance-resistance probe. *Sensors* 2012, 12, 17588-17607

Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T, Satou M, Akiyama K, Taji T, Yamaguchi-Shinozaki K, Carninci P, Kawai J, Hayashizaki Y, Shinozaki K: Monitoring the expression profiles of ca 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant J* 2002, 31: 279–292

Shen Q, Chen CN, Brands A, Pan SM, Ho TH: The stress- and abscisic acid-induced barley gene HVA22: developmental regulation and homologues in diverse organisms. *Plant Mol Biol* 2001, 45(3):327-40

Shinozaki K, Yamaguchi-Shinozaki K: Gene networks involved in drought stress response and tolerance. *J Exp Bot* 2007, 58: 221– 66

Singh KK, Ghosh S: Regulation of glutamine synthetase isoforms in two differentially drought-tolerant rice (*Oryza sativa* L.) cultivars under water deficit conditions. *Plant Cell Rep* 2013, 32(2):183-93

Skirycz A, Vandenbroucke K, Clauw P, Maleux K, De Meyer B, Dhondt S, Pucci A, Gonzalez N, Hoeberichts F, Tognetti VB, Galbiati M, Tonelli C, Van Breusegem F,

Vuylsteke M, Inzé D: Survival and growth of *Arabidopsis* plants given limited water are not equal. *Nat Biotechnol* 2011, 29(3):212-4

Souza RP, Machado EC, Silva JAB, Lagôa AMMA, Silveira JAG: Photosynthetic gas exchange, chlorophyll fluorescence and some associated metabolic changes in cowpea (Vigna unguiculata) during water stress and recovery. *Environ Exp Bot* 2004, 51, 45-56

Stermer BA, Bianchini GM, Korth KL: Regulation of HMG-CoA reductase activity in plants. *J Lipid Res* 1994, 35(7):1133-40

Talamè V, Ozturk NZ, Bohnert HJ, Tuberosa R: Barley transcript profiles under dehydration shock and drought stress treatments: a comparative analysis. *J Exp Bot* 2007, 58(2):229-40

Tester M, Langridge P: Breeding technologies to increase crop production in a changing world. *Science* 2010, 327:818–22

Trenberth KE, Dai A, van der Schrier G, Jones PD, Barichivich J, Briffa KR, Sheffield J: Global warming and changes in drought. *Nat Climate Change* 2014, 4: 17–22

Tuteja N, Mahajan S: Calcium signaling network in plants: an overview. *Plant Signal Behav* 2007, 2(2):79-85

Vinocur B, Altman A: Recent advances in engineering plant tolerance to abiotic stress: achievements and limitations. *Curr Opin Biotechnol* 2005, 16(2):123-32

Wang H, Miyazaki S, Kawai K, Deyholos M, Galbraith DW, Bohnert HJ: Temporal progression of gene expression responses to salt shock in maize roots. *Plant Mol Biol* 2003, 52(4):873-91

Zheng J, Fu J, Gou M, Huai J, Liu Y, Jian M, Huang Q, Guo X, Dong Z, Wang H, Wang G: Genome-wide transcriptome analysis of two maize inbred lines under drought stress. *Plant Mol Biol* 2010, 72(4-5):407-21

Zörb C, Noll A, Karl S, Leib K, Yan F, Schubert S: Molecular characterization of Na+/H+ antiporters (ZmNHX) of maize (*Zea mays* L.) and their expression under salt stress. *J Plant Physiol* 2005, 162(1):55-66

Chapter 2

1 Introduction

1.1 Plant small RNAs

Plant small RNAs (sRNAs) are a pool of 20-nucleotides (nt) to 24-nt non-coding RNAs that participate in a set of pathways termed RNA-mediated silencing, or RNA interference (RNAi), controlling the expression of genes, the quiescence of viruses and the movement of transposable elements (TEs). sRNAs exert RNAi through different mechanisms: the post-transcriptional gene silencing (PTGS) or the transcriptional gene silencing (TGS). In PTGS, occurring in the cytoplasm, sRNAs target complementary messenger RNAs (mRNAs), inducing their degradation or translational repression. In TGS, occurring in the nucleus, sRNAs direct repressive epigenetic modifications, such as DNA cytosine methylation and histone methylation, to homologous regions of the genome (Matzke and Mosher 2014).

There are exogenous sRNAs, produced from transgene-derived or virus-derived transcripts, and endogenous sRNAs, produced from endogenous transcripts. Exogenous sRNAs were first discovered in 1999 in plants by Hamilton and Baulcombe (1999): they uncovered the presence of sRNAs corresponding to transgenes, only in plants undergoing PTGS of the transgenes, and sRNAs corresponding to viral sequences, in plants infected with viruses. Endogenous sRNAs in plants were found later with the cloning of the first-discovered *Arabidopsis* microRNAs (Llave et al. 2002, Park et al. 2002, Reinhart et al. 2002).

Since their discovery, sRNAs have been the focus of extensive studies that led to the comprehensive appreciation of their biogenesis, modes of actions and biological functions. It became clear that sRNAs, as regulators molecules, influence almost all aspects of plant biology, playing important roles in genome stability maintenance, plant growth and development, adaptation to abiotic stresses and responses to biotic pathogens.

1.2 Classification of plant endogenous small RNAs

Plant endogenous sRNAs share common features in their biogenesis and function mechanisms. They are produced from the processing of helical RNA precursors into small double-stranded duplexes, varying in size from 20-nt to 24-nt, by the endonuclease activities of DICER-LIKE (DCL) proteins, which are RNAse III enzymes. An ARGONAUTE (AGO) protein, contained in the RNA-induced silencing complex (RISC), binds to one strand of the initial duplex, which then becomes available to match target RNAs, by sequence complementarity, and subsequently direct their repression. Longer sRNAs, from 30-nt to 40-nt, sharing many common features with known sRNAs, have been identified in *Arabidopsis* upon pathogen infection or under specific growth conditions (Katiyar-Agarwal et al. 2007). Longer sRNAs, as well as exogenous sRNAs, are not described here because they have not been the focus of this study.

The categorization of endogenous sRNAs, based on differences in biogenesis and function, is here reported following the classification system of Axtell MJ (2013a) (Figure 1). sRNAs can be primarily divided into two main categories that differ in the structure of the helical RNA precursor. One group is composed by the hairpin RNAs (hpRNAs), which are produced from single-stranded RNA (ssRNA) precursors that have intramolecular nucleotide sequence complementarity resulting in a hairpin loop structure. The other group is composed by the small interfering RNAs (siRNAs), which are produced from double-stranded RNA (dsRNA) precursors that are formed by the intermolecular hybridization of two complementary RNA strands. hpRNAs and siRNAs can be further subdivided in different child categories. A hairpin RNA precursor can be processed in a precise way, producing one or a few functional sRNAs called microRNAs (miRNAs), or in an imprecise way, producing sRNAs from diverse regions of the hairpin. miRNAs can be conserved in different species or can be specifically detected only in one species or a few closely related species. miRNAs are usually 20-nt to 22-nt long but longer 23-nt and 24-nt forms of miRNAs have been found that function similarly to siRNAs. The majority of siRNAs are heterochromatic siRNAs (hc-siRNAs), which are 23-nt or 24-nt long and are produced mainly from intergenic and/or repetitive regions where they direct the deposition of repressive epigenetic marks. Less numerous categories of siRNAs

are the secondary siRNAs and the natural antisense transcript siRNAs (NAT-siRNAs). Secondary siRNAs biogenesis requires the initial cleavage of an RNA transcript directed by other sRNAs and its subsequent conversion into a dsRNA that is then processed by DCL proteins. Secondary siRNAs can be processed in phase (phasiRNAs), for example the *trans*-acting siRNAs (tasiRNAs), or not. NAT-siRNAs are a less described category of sRNAs derived from two distinct, homologous, and interacting mRNAs that are transcribed from overlapping or nonoverlapping genes.

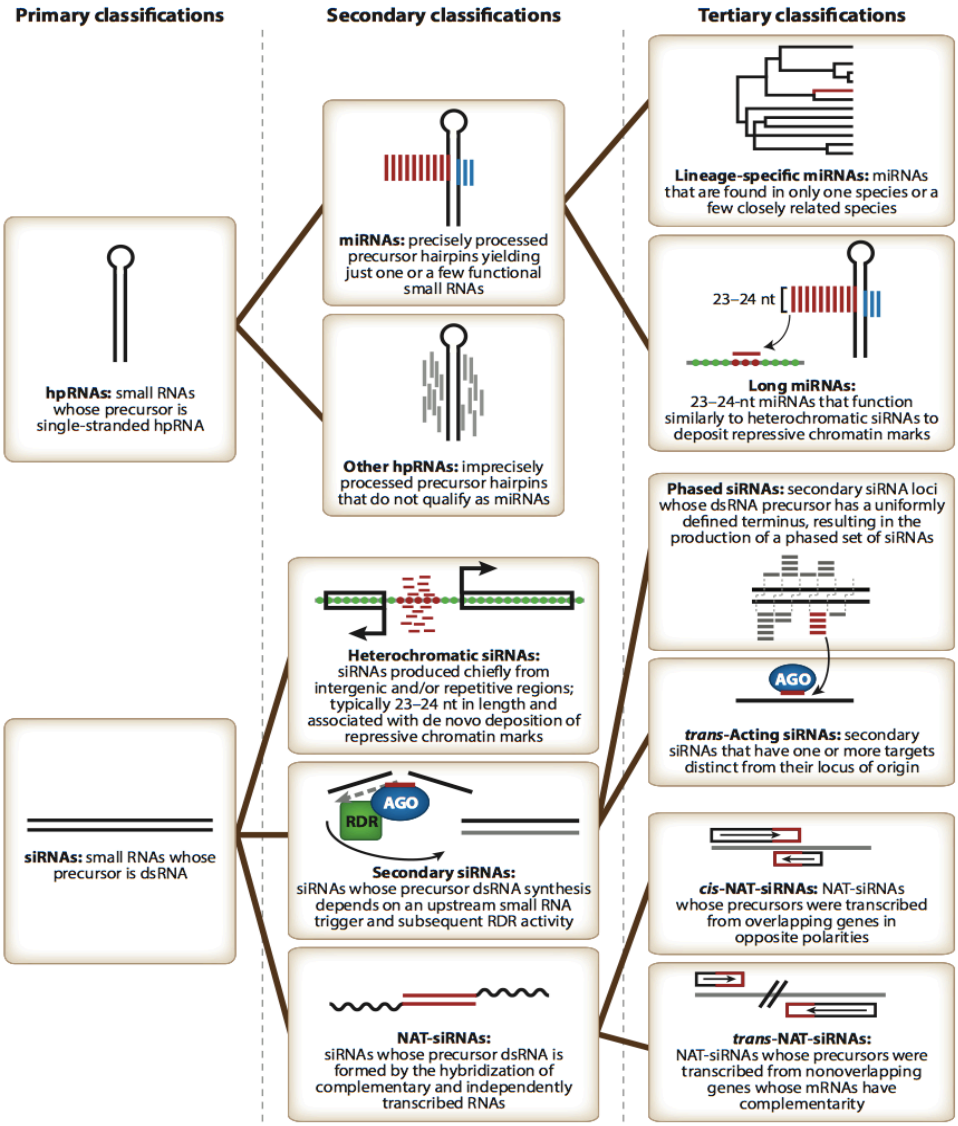

Figure 1: Plant endogenous sRNAs classification (from Axtell MJ 2013a).

1.3 MicroRNAs

1.3.1 MicroRNA biogenesis

Mature plant miRNAs range in size from 20-nt to 24-nt but most of them are 21-nt long, their biogenesis is summarized in Figure 2. miRNAs are encoded by endogenous *MIRNA* (*MIR*) genes that are located in intergenic or genic regions and can be found both in exons and introns of their host genes. *MIR* genes are transcribed by the RNA polymerase II (Pol II; Xie et al. 2005) into capped and polyadenylated pri-miRNAs that range from approximately 70 to thousands of bases and contain imperfect, self-complementary foldback regions. Pri-miRNAs are presumably stabilized by the RNA-binding protein DAWDLE (DDL; Yu et al. 2008). In the nuclear processing centres called D-bodies, pri-miRNAs are processed to precursor miRNAs (pre-miRNAs) by DCL1 protein, necessitating the activity of the dsRNA-binding protein HYPONASTIC LEAVES1 (HYL1) and the C2H2-zinc finger protein SERRATE (SE; Fang and Spector 2007). Pre-miRNAs are stem-loop structures subjected to subsequent cleavages by DCL1 to form miRNA/miRNA* duplexes with 3' overhangs. The 3' ends of the miRNA/miRNA* duplex are 2'-*O*-methylated by the nuclear S-adenosyl methionine-dependent methyltransferase HUA ENHANCER 1 (HEN1) protein (Yu et al. 2005), which blocks uridylation by HEN1 SUPPRESSOR 1 (HESO1; Zhao et al. 2012) and decay of miRNAs by 3'–5' exoribonucleases SMALL-RNA-DEGRADING NUCLEASE 1 (SDN1), SDN2 and SDN3 (Ramachandran and Chen 2008). Most miRNAs exit the nucleus and enter the cytoplasm with the assistance of the plant homolog of Exportin-5, HASTY (HST; Park et al. 2005); an additional export pathway seems to be involved but it remains unknown. In the cytoplasm, the miRNA/miRNA* duplex is loaded onto AGO1 protein: the miRNA* passenger strand is removed and only the miRNA guide strand is retained, to carry out the silencing reactions. The miRNA* is usually rapidly degraded but there are documented cases in which it has similar or higher abundance levels than the corresponding canonical miRNA and it appears to regulate specific targets (Zhang et al. 2011, Manavella et al. 2013).

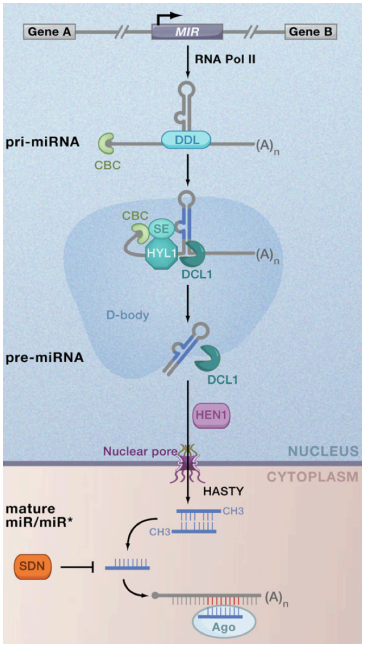

Figure 2: Biogenesis of plant microRNAs (from Voinnet O 2009).

1.3.2 MicroRNA mechanisms of action

The miRNA loaded onto the AGO1 protein guides the RISC complex to target mRNAs through base pairing. In plants, targets are recognized when the base pairing is extensive: there are only few examples with more than five mismatches between the miRNA and the target. The critical region of the base pairing is between positions 2-13 from the 5'-end of the miRNA: here a single mismatch is tolerated but rare (Axtell MJ 2013a). miRNAs with high levels of sequence similarity belong to the same 'miRNA family' and are assigned the same number. Most plant miRNA families have zero to ten known targets (usually from the same gene family) in a single genome (Jones-Rhoades MW 2012). Multiple different mechanisms of miRNA target repression have been reported and are described below.

The recognized mRNA, target of a miRNA, can be cleaved by the endonucleolytic activity of AGO1 between positions 10 and 11 of the alignment and then followed by RNA degradation. Alternatively, the miRNA can induce mRNA translational inhibition. For many targets both the mRNA cleavage and translational repression are known to co-occur (Voinnet O 2009). Some specific

miRNAs target and cleave *TAS* transcripts, which are consequently made double-stranded and processed in a 21-nt phasing manner from either 5′ or 3′ of the miRNA-cleaved fragments; the resulting 21-nt tasiRNAs act in *trans* to regulate target mRNAs as well as miRNAs (Yoshikawa et al. 2013).

In addition to post-transcriptionally mechanisms of miRNAs target repression, some cases have been described in which miRNAs cause transcriptional gene repression: in *Arabidopsis*, the miR165/166 is though to bind to the newly synthetized and processed *PHABULOSA (PHB)* mRNA to direct methylation of the corresponding *PHB* template DNA (Bao et al. 2004). In rice a number of long miRNAs of 24-nt, in some cases produced by dual-coding precursors that give rise to both a canonical miRNA and the 24-nt species, direct DNA methylation both in *cis*, at loci where they are originated, and in *trans*, at target genes (Wu et al. 2010, Hu et al. 2014).

In plants, there is only one documented case in which the pairing between the miRNA and the target is interrupted by a central mismatched loop that prevents the slicing of the target and instead causes the sequestration of the miRNA by target. This is the case of the *Arabidopsis* transcript encoded by the non-coding gene *INDUCED BY PHOSPHATE STARVATION1 (IPS1)* that sequesters the phosphate (P_i) starvation–induced miRNA miR399 (Franco-Zorrilla et al. 2007). By sequestering the miR399, *IPS1* transcript modulates the activity of the miRNA by competing with its canonical target gene *PHO2* encoding an ubiquitin-conjugating E2 enzyme, which is a major component for the maintenance of Pi homeostasis (Bari et al. 2006).

1.3.3 *MIRNA* gene evolution

Two different models of *MIR* gene evolutionary emergence have been proposed (Axtell et al. 2011). One mode in which *MIR* genes are thought to evolve is through the duplications of intragenomic regions. This hypothesis is based on the observation that the young *MIR* genes share extensive sequence complementarity with their targets (Allen et al. 2004), suggesting that inverted duplication of genes formed the young 'proto-*MIRs'*. The proto-*MIR* transcripts are initially imprecisely processed by one or more DCL enzymes to produce heterogeneous sRNAs of multiple size classes (Dunoyer et al. 2007). The proto-

MIRs further accumulate mutations that can make them inert. Contrarily, when their target regulation is beneficial for the host, they are positively selected by evolution to become young *MIR* genes: they acquire both mutations that cause fold-back mispairing and DCL1 dependence that leads to the precise processing of the precursors into the mature miRNA and miRNA* sequences. The intragenomic duplications giving rise to *MIR* genes involve protein-coding genes and also TEs (Li et al. 2011). Pre-existing intragenomic duplications characteristics of the DNA-type nonautonomous elements Miniature Inverted–Repeat Transposable Elements (MITEs) also appear to be a source of *MIR* gene genesis. MITEs transcripts form small imperfect hairpins typical of miRNA precursors (Piriyapongsa and Jordan 2008) and several young *MIR* genes have been found to map to MITEs. A recent accurate analysis of the rice TE-derived *MIR* genes revealed that at least some of them are bona fide miRNAs (Li et al. 2011). Another mode in which *MIR* genes are thought to evolve is through the 'spontaneous evolution' from random fold-backs sequences found in the genomes (Voinnet O 2009). When the occasional regulation mediated by the emerging miRNA confers benefits to the plant, the gene can be selected and gains competence for miRNA biogenesis, accumulating mutations that improve the hairpin cleavage by DCL1 and the gene transcriptional capacity (Axtell et al. 2011).

MIR genes appear to have high rates of birth and death because the majority of miRNAs in any given plant species are not conserved and only found in one species or closely related species. The lineage-specific miRNAs tend to show characteristics of the young miRNAs: they are often expressed at low levels, processed in a heterogeneous way from their precursors and lack targets. These observations indicate that many of the lineage-specific miRNAs are likely to be proto-miRNAs, transient, nonfunctional entities (Axtell MJ 2013a).

1.3.4 MicroRNA roles in drought and salinity stress response and tolerance

Drought and salinity are among the major environmental stresses worldwide, which adversely affect plant growth and productivity. To tolerate stresses in their sessile lifestyle, plants have evolved networks of molecular events that confer them developmental plasticity. Plant stress responses involve transcriptional and post-transcriptional gene regulation: several genes and miRNAs have been observed to be up or downregulated in many species under abiotic stress conditions. Stress tolerance is a complex genetic trait, for this reason breeding for stress tolerance and the creation of stress-tolerant transgenic plants is challenging (Bartels and Sunkar 2005, Jewell et al. 2010). Similarly, the fact that a miRNA is differentially regulated in response to an environmental stress does not necessarily mean that the miRNA is involved in stress adaptation responses (Khraiwesh et al. 2012), but presumably as the understanding of the roles of miRNAs during stress deepens, the possibilities for using miRNA-mediated gene regulation to enhance plant stress tolerance would significantly increase (Sunkar et al. 2012).

To study the role of miRNAs in drought and salinity stress response and tolerance many works have been done subjecting plants to stress conditions and detecting the expression of miRNAs, in some works also of their targets, in both control and stressed samples. Drought and/or salinity-responsive miRNAs that have been detected in maize are summarized below (Table 1). These are the results of many works where the stress response has been studied focusing on different aspects: the comparison between stress sensitive and tolerant varieties (Ding et al. 2009, Wang et al. 2014a) or between inbred lines and hybrids (Kong et al. 2010), the study of the time-course of the stress response (Ding et al. 2009, Wei et al. 2009, Wang et al. 2014a, Luan et al. 2015) or the study of the miRNA precursor expression instead of that of miRNAs (Zhang et al. 2014). Diverse responses of an individual miRNA family to a stress condition can be due either to different behaviors of its distinct members or to different behaviors observed in diverse genetic lines, time points of stress application or stress protocols. For example, maize roots of plants grown with salty water show different expressions of miR164 and miR167 between salt-tolerant and salt-sensitive lines that could

contribute to their diverse level of stress tolerance (Ding et al. 2009). In this experiment miR164 is initially upregulated in both lines after the stress application but while it remains upregulated in the salt-sensitive line, it decreases its expression in the salt-tolerant line after 24 hours of treatment. miR167 is salt-repressed only in the salt-tolerant line and remains unaffected in the salt-sensitive line. The specific downregulation of miR164 and miR167 families in the salt-tolerant line could lead to a higher accumulation of their predicted targets, respectively transcripts of *NO APICAL MERISTEM* (*NAM*) and *AUXIN RESPONSE FACTOR* (*ARF*) genes (Zhang et al. 2009). The higher accumulation of these targets could enhance the auxin response and thus enhance shoot and root development, accumulating more biomass to counteract the wastage brought on by the salt shock. Furthermore, these effects could contribute to the adaptive advantage of the salt-tolerant line (Ding et al. 2009).

	drought stress	salinity stress
miR156	(+) [3] (-) [2] (+/-) [5]	(+) [3] (+/-) [1]
miR159	(+/-) [5]	(-) [1]
miR160	(-) [2] (+/-) [5]	(+) [1]
miR162	(+) [5]	(+/-) [1]
miR164	(+/-) [5]	(-) [3] (+/-) [1]
miR166	(+) [3] (-) [2] (+/-) [5]	(+) [3] (+/-) [1]
miR167	(+) [4] (+/-) [2;5]	(-) [1]
miR168	(-) [2;3] (+/-) [5]	(+/-) [1]
miR169	(+/-) [4;6]	(+/-) [6]
miR171	(+/-) [5]	(-) [3] (+/-) [1]
miR172	(+) [4]	
miR319	(+) [5] (+/-) [3]	(-) [3] (+/-) [1]
miR390	(+/-) [5]	
miR365		(+) [1]
miR396	(+/-) [2;5]	(-) [1]
miR397	(-) [5]	
miR398	(-) [2] (+/-) [5]	
miR399	(-) [4] (+/-) [5]	(-) [1]
miR408	(+) [5] (-) [2]	
miR528	(-) [2] (+/-) [5]	
miR529	(-) [5]	
miR827	(-) [4] (+/-) [5]	
miR1432	(+/-) [5]	

Table 1: Summary of drought- and/or salinity-responsive miRNAs or miRNA precursors in maize. Stress responsive miRNAs: [1] Ding et al. 2009 [2] Wei et al. 2009 [3] Kong et al. 2010 [5] Wang et al. 2014a [6] Luan et al. 2015. Stress responsive miRNA precursors: [4] Zhang et al. 2014. +=upregulated. -=downregulated. +/-=some members were upregulated, some were downregulated or different miRNA trends were found in diverse genetic lines, time points of stress application or stress protocols.

The involvement of miRNAs in drought and salinity stress response and tolerance has been suggested for many miRNA families, in maize and other species, as described in the following examples. In *Arabidopsis,* miR156 targets *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE* (*SPL*) genes, as well as in maize (Chuck et al. 2007). SPLs play a role in the regulation of leaf cell number and size (Usami et al. 2009), thus miR156 could contribute to the modulation of leaf and shoot development under stress conditions (Ding et al. 2009). A similar role could be played by miR319 and miR396 families. In *Arabidopsis,* miR319 targets *TEOSINTE BRANCHED1/CYCLOIDEA/PROLIFERATING CELL FACTOR1* (*TCP*) genes, as also predicted in maize (Zhang et al. 2009), thus playing a role in the regulation of leaf cell proliferation (Palatnik et al. 2003, Martin-Trillo and Cubas 2010); miR396 in *Arabidopsis* targets *GROWTH-REGULATING FACTOR* (*GRF*) genes, as also predicted in maize (Zhang et al. 2009), thus contributing to the regulation of cell expansion in leaf (Wang et al. 2011). Finally, in *Arabidopsis* miR398 targets Cu/Zn *SUPEROXIDE DISMUTASES* (*SOD*) genes, which are directly involved in stress responses because they are important for the scavenging of reactive oxygen species (ROS; Sunkar et al. 2006) that are produced in excess under drought and salinity conditions. Similarly, in maize miR528 is predicted to target Cu/Zn SODs (Zhang et al. 2009).

1.3.5 MicroRNA annotation and expression profiling through massive parallel sequencing of small RNAs

The first step when studying miRNAs is the cloning of the miRNA sequence. Once the presence of the miRNA has been demonstrated, downstream and upstream analyses are used to complete its functional characterization. Downstream analyses include the expression profiling of the miRNA, the validation of its predicted targets and the study of its activity regulation. Upstream analyses are aimed to understand the miRNA expression modulation that can be exerted at the chromatin level or at the RNA level (Chen et al. 2010).

Several strategies are employed to clone miRNAs: bioinformatics prediction based on criteria for plant miRNA definition (Meyers et al. 2008), mutant screening, genetic cloning, microarrays and massive parallel sequencing of sRNAs. In each case the expression of a certain miRNA should be detected

through Northern hybridization or qRT-PCR (Chen et al. 2010). Microarrays, widely employed in the past, suffer from background and cross-hybridization problems and measure only the relative abundances of known microRNAs. They have been gradually replaced by the more accessible technique of massive parallel sequencing, also called high-throughput sequencing or Next Generation Sequencing (NGS) that allows measuring the absolute abundance of miRNAs in a wider range than microarrays and permits the discovery of new miRNAs (Creighton et al. 2009).

NGS technology is therefore employed not only to clone known miRNAs and to annotate new *MIRNA* loci but also to annotate other sRNA species and define their expression pattern, with a cost that is significantly decreasing over time with increasing performances.

1.3.5.1 NGS: annotation of MIRNA loci and detection of miRNA variants

sRNA sequencing experiments produce a huge number of reads (millions) for each analyzed library that must be pre-processed before to be analyzed. The pre-processing includes a filter to select reads with a minimum quality score, the trimming of the adapter and a second filter to remove low complexity reads; variants of the pre-processing step exist, such as performing the mapping of the reads against the *Escherichia coli* genome to filter out reads coming from potential contaminants. After the reads have been cleaned they are aligned against the genome of the studied species.

In order to identify and annotate *MIRNA* loci the regions of the genome that can be potentially folded into single-stranded, stem-loop hairpin structures are first selected. To classify a predicted hairpin as a *MIRNA* locus the pattern of aligned reads within its sequence must satisfy some strict criteria that have been defined to distinguish *bona fide* miRNAs from other sRNAs or RNA degradation products (Figure 3). Hairpins with single reads, heterogeneously processed reads or reads without the 3' overhangs do not show evidence of a precise DCL-dependent processing typical of miRNAs and thus are not considered putative *MIRNA* loci (Figure 3a). Hairpins that show preferential mapping of reads in the candidate miRNA/miRNA* duplex region but lack precise 5'/3' ends of reads or the presence of the miRNAs* are considered as *MIRNA* loci 'candidates' and require

further studies to be confirmed (Figure 3b). A high-confident *MIRNA* locus generates relatively precise miRNA/miRNA* duplexes with 3' overhangs (Figure 3c).

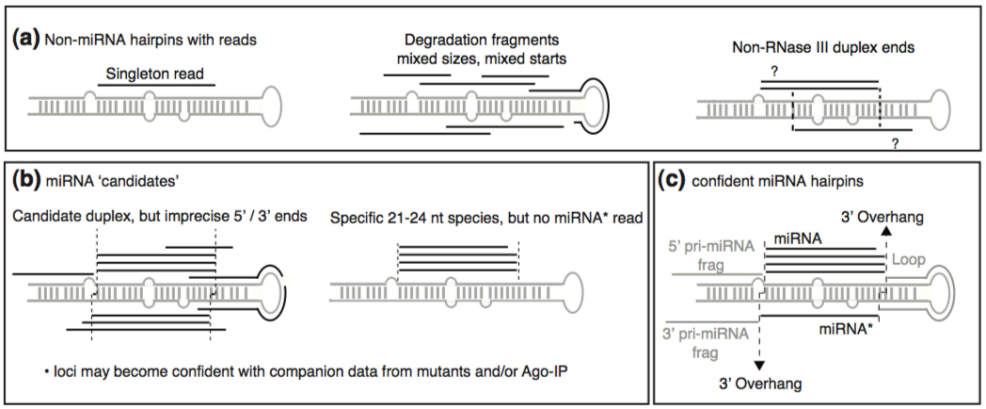

Figure 3: Possible patterns of aligned reads to predicted hairpins. **(a)** Examples of loci that should not be annotated as *MIRNA* loci. **(b)** Examples of loci that have low evidence for miRNA biogenesis. **(c)** Reads alignment pattern of high-confident loci for miRNA biogenesis (from Axtell et al. 2011).

Annotations of *MIRNA* loci and miRNA/miRNA* mature sequences are reported in the database miRBase (Kozomara and Griffiths-Jones 2011). A recent study in *Arabidopsis* by Coruh et al. (2014) has shown that the mature sequences annotated in miRBase are not always identical to the most abundant sRNA species mapping within a *MIRNA* locus, detected from sRNA sequencing experiments. This reflects the presence of inaccurate annotations in miRBase and/or the fact that nearly all known *MIRNA* loci produce more than a single product. Indeed, two previous works in *Arabidopsis* and rice demonstrated that the sRNA sequencing of samples from different tissues, of plants with diverse genetic backgrounds, wild type (wt) and mutants with impaired sRNA biogenesis pathways, subjected to different environmental and nutrient stresses, allows the detection of products resulting from the alternative processing of the *MIRNA* hairpins (Jeong D-H et al. 2013, Jeong D-H et al. 2011). The alternative processing of a *MIRNA* precursor can give rise to sequences that are length and/or sequence variants of the annotated mature miRNA, named isomiRs, or to sequences that originates from a different, nonoverlapping region of the hairpin.

isomiRs are categorized into three main classes: 5' isomiRs, 3' isomiRs and polymorphic isomiRs. 5' and 3' isomiRs show differences compared to mature annotated miRNA respectively in the 5'- and in the 3'-end of the sequence, while polymorphic isomiRs harbor different internal nucleotide sequences. 5' and polymorphic isomiRs are rare while 3' isomiRs are observed frequently. isomiRs can derive from the activity of exoribonucleases, nucleotidyl transferases and in animals also the RNA editing process is though to modify miRNA sequences (Neilsen et al. 2012). It is still unclear if isomiRs are functionally significant but there are some evidences: 5' isomiRs can influence miRNA target selection in *Arabidopsis* (Jeong D-H et al. 2013) and 3' isomiRs can influence the stability of miRNAs in *Arabidopsis*, rice and maize (Zhai et al. 2013).

Another alternative to the canonical *MIRNA* precursor processing is the formation of a more abundant sequence nonoverlapping with the annotated mature miRNA. In *Arabidopsis* it has been observed that such variants can influence the preferential AGO loading (Jeong D-H et al. 2013). A specific case falling under this category of alternative hairpin processing is when the miRNA* accumulates at higher levels compared to the miRNA. Also miRNAs* can be loaded by AGO proteins and several miRNA*s have known functions (Zhang et al. 2011, Manavella et al. 2013).

1.3.5.2 NGS: expression profiling of miRNAs

sRNA sequencing experiments return for each analyzed library the number of sequenced reads corresponding to each unique detected sRNA sequence, which is used to examine the absolute abundance of miRNAs in each individual sample or compare the expression of miRNAs between distinct samples. To perform differential expression analysis of miRNAs between two or more samples it is first necessary to normalize their expression in each sequenced library, to reduce the impact of nonbiological sources of variation that can add noise to sRNA sequencing experiments. Many normalization methods have been developed to normalize the abundance of sRNAs, which are classified into two categories, according to the application of linear scaling or not.

Linear scaling methods include scaling, upper quartile, global, Lowess and trimmed mean of M value (TMM). For miRNA normalization the most frequently

used method is total count scaling (Smyth et al. 2003), which divides gene counts of a sample by its normalization baseline and multiplies by a fixed number, such as the mean total count across all the samples of the dataset. The normalization baseline can be the total number of reads sequenced or the total number of aligned reads to the genome or, in experiments focusing on a specific RNA type, such as miRNAs, it can be the total number of these sequences. Upper quartile (Bullard et al. 2010) is similar to the total count scaling, instead that the normalization baseline is the upper quartile of total counts. In sRNA sequencing the 75th-percentile sRNA are found at only one or two copies per library, furthermore this method needs to be modified to be applied to sRNA data (McCormick et al. 2011). Less used are global (Smyth et al. 2003), Lowess (Smyth et al. 2003) and TMM (Robinson and Oshlack 2010) methods. TMM assumes that most genes are not differentially expressed (DE) between samples and thus that their true relative expression levels should be pretty similar: it calculates, for each baseline element, the log expression ratio of the experimental sample to a control sample (or the mean or median of all samples) and uses their trimmed mean as a linear scaling factor. TMM is good for dataset including tens of thousands RNA species, furthermore its use is discouraged for studies limited to the smaller datasets of miRNAs (McCormick et al. 2011, Garmire and Subramaniam 2012).

Nonlinear scaling methods include quantile (Bolstad et al. 2003), variance stabilization (VSN) (Huber et al. 2002), invariant method (INV) (Pradervand et al. 2009) and two-step nonlinear regression (Taslim et al. 2009). For miRNA normalization the most frequently used is quantile, which assumes that most genes are not DE between samples and that the true expression distribution is similar between different samples: the highest values of each sample take the values of the average of the all the highest values and the procedure is repeated for every set of next highest values.

There are several tools to perform differential expression analysis available at Bioconductor (www.bioconductor.org). The most frequently used for miRNAs are: edgeR (Robinson et al. 2010) and baySeq (Hardcastle and Kelly 2010), which use a model based on negative binomial distribution to estimate differential expression, and DESeq (Anders and Huber 2010), which assumes that

the mean is a good predictor of the variance and tests for differences between the base means of two conditions. SAM-seq method (Fahlgren et al. 2009) adapts the significance analysis of microarrays (SAM) to sequencing data. It is not frequently used because to guarantee a good power of DE genes detection it requires a number of replicates that is usually not reached by sequencing experiments.

1.4 Small interfering RNAs

Following the classification of the endogenous sRNA described above (Axtell MJ 2013a), siRNAs include secondary siRNAs, NAT-siRNAs and hc-siRNAs: only the latter are described here, because the others were not the focus of this study. Heterochromatic-siRNAs are so called because they derive mainly from intergenic and/or repetitive genomic regions where they direct the *de novo* deposition of repressive chromatin marks through an epigenetics process named RNA-directed DNA methylation (RdDM). RdDM is involved in the transcriptional silencing of these regions (Matzke et al. 2009) and it is defined as an epigenetic pathway because it does not affect the DNA sequence of its target but it influences their regulation by modifying the chemical properties of DNA and chromatin, such as inducing DNA methylation and post-translational modifications of histone tails.

In maize, a recent work has demonstrated that the genomic loci undergoing RdDM, defined by their production of siRNAs, are characterized by a different chromatin environment compared to that of heterochromatin, traditionally defined as chromatin regions that remain condensed throughout the cell cycle (Gent et al. 2014) (Figure 4). Briefly, regions that are not targeted by siRNAs and RdDM are characterized by inaccessible, transcriptionally inactive heterochromatin and are enriched in DNA symmetric methylation (CG and CHG contexts, where H = A, C, or T) and dimethylation of lysine 9 (H3K9me2). In contrast, loci targeted by siRNAs, thus by RdDM, are characterized by accessible, transcriptionally active chromatin and are enriched in asymmetric DNA methylation (CHH context) and show relatively low H3K9me2. Here the production of siRNAs ensures the silencing of these regions in a transcriptionally active environment. Unlike heterochromatic regions, RdDM loci are preferentially located next to genes, which are characterized by accessible, active euchromatin and relatively low levels of DNA methylation, allowing for mRNA production.

Considering this result, siRNAs associated with RdDM in the silencing of DNA, before named as hc-siRNAs, will be hereafter named with the general name of siRNAs.

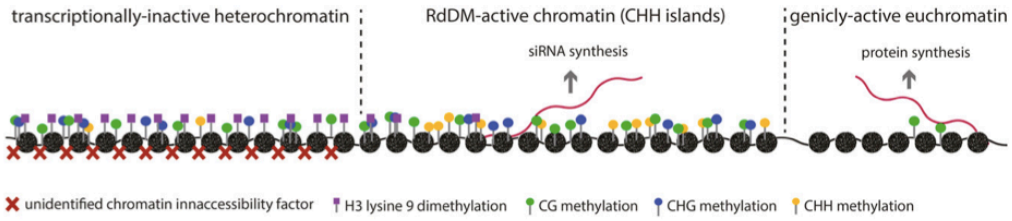

Figure 4: Three major chromatin environments in maize (from Gent et al. 2014).

1.4.1 Small interfering RNA biogenesis and function in the canonical RNA-directed DNA methylation pathway

The canonical RdDM pathway is summarized in Figure 5. The production of siRNAs participating in RdDM requires the transcription of template DNA by the plant-specific RNA polymerase IV (Pol IV) (Zhang et al. 2007), which is assumed to transcribe ssRNAs. The mechanisms by which Pol IV selects its targets are not completely clear: up to now it has been demonstrated that for a large subset of the most active sites of siRNA production Pol IV is directed to DNA by the interaction with the DNA-BINDING TRANSCRIPTION FACTOR 1/SAWADEE HOMEODOMAIN HOMOLOG 1 (DTF1/SHH1), which interacts directly with the chromatin remodeling protein CLASSY 1 (CLSY1) and binds to the methylated lysine 9 (H3K9me) and unmethylated lysine 4 (H3K4) (Law et al. 2013, Zhang et al. 2013). RNA-DEPENDENT POLYMERASE 2 (RDR2) physically associates with Pol IV and generates dsRNAs using Pol IV transcripts as templates, with the assistance of CLSY1 (Law et al. 2011, Smith et al. 2007). The dsRNAs are cleaved by DCL3 into 24-nt siRNA duplexes (Kasschau et al. 2007), which are stabilized by methylation at their 3'-OH groups by HEN1 (Ji and Chen 2012) and transported to the cytoplasm for AGO4 loading. Loaded with the guide strand AGO4 is imported back to the nucleus where it associates with non-coding transcripts produced by a second plant-specific RNA polymerase V (Pol V) through complementarity to the siRNAs (Wierzbicki et al. 2009). The slicer activity of AGO4 is required for the cleavage of the passenger strand of the initial siRNA

duplex: it is therefore necessary for the siRNA loading (Ye et al. 2012) but it is still unknown whether it is necessary for the mechanism of target recognition. The scaffolding transcripts produced by Pol V thus form the set of siRNA targets (Axtell MJ 2013a). Perfect complementarities between siRNAs and Pol V transcripts are functional for target selection but it is still unknown if other base-pairing patterns are also functional. As for Pol IV, the mechanisms of Pol V selection of its targets remain incompletely understood although some insights into its binding site preferences have been revealed. Pol V-occupied loci, detected by experiments of chromatin immunoprecipitation followed by sequencing (ChIP–seq), are associated with 24-nt siRNAs and CHH methylation. However, a subset of loci does not show these characteristics, suggesting that Pol V occupancy alone is not sufficient for RdDM (Wierzbicki et al. 2012). The Pol V-occupied loci are preferentially found in euchromatic regions, in the immediate 5′ proximal region next to known Pol II promoters, especially where "young" TEs are located (Zhong et al. 2012). Pol V transcription and association with chromatin depends critically on the DDR complex (Zhong et al. 2012), which comprises the putative chromatin remodeller DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (DRD1), the hinge-domain protein DEFECTIVE IN MERISTEM SILENCING 3 (DMS3) and the single-stranded DNA-binding protein RNA-DIRECTED DNA METHYLATION 1 (RDM1) (Law et al. 2010). The recruitment of Pol V to some targets is helped by three members of the SU(VAR)3-9 histone methyltransferase family, SUVH2, SUVH9, which bind methylated DNA, and SUVR2 (Johnson et al. 2014, Liu et al. 2014). The KOW DOMAIN-CONTAINING TRANSCRIPTION FACTOR 1 (KTF1) is associated with Pol V and is supposed to act as an organizer by interacting with AGO4 and methylated DNA (Bies-Etheve et al. 2009, He et al. 2009). The association between siRNA-loaded AGO4 and Pol V transcripts leads to the recruitment of the DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) which catalyses the *de novo* DNA methylation in all cytosine context, including CHH, at the homologous genomic sites of the Pol V transcripts (Pélissier et al. 1999, Matzke and Mosher 2014). A subset of siRNAs requires the AGO4 slicer activity for their accumulation (Qi et al. 2006). Furthermore, it has been suggested that the association between siRNA-loaded AGO4 and Pol V transcripts might cause the AGO4-mediated cleavage of

a subset of Pol V transcripts (Axtell MJ 2013a). *De novo* DNA methylation by DRM2 consequently directs chromatin modifications that transcriptionally silence the target loci: the nucleosome positioning, adjusted by the SWI/SNF complex (Zhu et al. 2013), and the deposition of repressive histone marks, such as H3K9me by SUVH4, SUVH5 and SUVH6 (Enke et al. 2011), which is facilitated by the removal of active marks by HISTONE DEACETYLASE 6 (HDA6) (To et al. 2011), JUMONJI 14 (JMJ14) (Searle et al. 2010) and UBIQUITIN-SPECIFIC PROTEASE 26 (UBP26) (Sridhar et al. 2007).

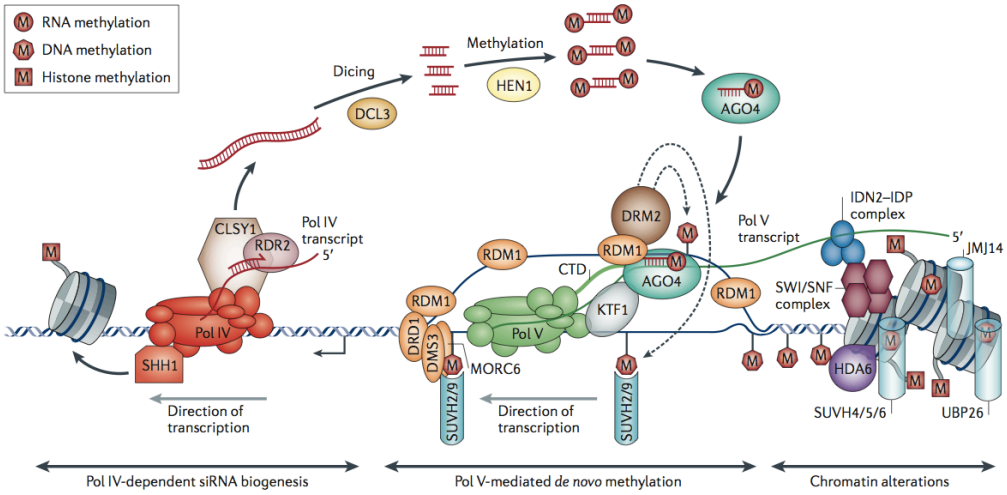

Figure 5: Canonical RdDM pathway (from Matzke and Mosher 2014).

There are evidences that RdDM is a self-reinforcing system. Both Pol IV and Pol V are preferentially associated with methylated DNA *in vivo* (Wierzbicki et al. 2012, Zhong et al. 2012, Law et al. 2013). Three proteins facilitating Pol V association with chromatin all bind to methylated DNA: RDM1, which is part of the DDR complex and also interacts with both AGO4 and DRM2 (Gao et al. 2010), and SUVH2 and SUVH9, which aid the Pol V recruitment to a subset of targets (Johnson et al. 2014, Liu et al. 2014). Pol V is required for the accumulation of siRNAs at some but not all loci (Mosher et al. 2008), indicating that DNA methylation promotes Pol IV activity: indeed mutants with impaired DNA methylatransferase activity show reduced siRNA accumulation (Lister et al. 2008, Stroud et al. 2014).

1.4.2 Control of transposon silencing by canonical and non-canonical RNA-directed DNA methylation pathways

The vast majority of siRNAs is transcribed from TEs and other repeats, which are the major targets of RdDM: siRNAs play crucial roles in the control of TE transcriptional silencing and inhibition of transposition.

1.4.2.1 RdDM: establishment and stabilization of transposon transcriptional silencing

Different mechanisms have been suggested for the three steps in TE transcriptional silencing: i) the establishment of TE silencing through diverse non-canonical RdDM pathways ii) the stabilization of silencing through the canonical RdDM pathway and iii) the maintenance of silencing through RdDM-independent pathways (Kim and Zilberman 2014).

The recognition of a transcriptionally active TE and the subsequent initiation of its silencing is achieved through different mechanisms upon the presence or not of a homologous sequence in the host genome. In the case of cross-hybridization within a single species or closely related species it is likely that the host genome can contain a homologous TE copy to the newly entered active TE. If the homologous TE copy has been previously silenced and its silencing has been stabilized through the canonical RdDM pathway, the 24-nt siRNAs matching the TE can recognize the active TE and quickly target it, resulting in homology-dependent *trans*-silencing of the active TE through RdDM (Nuthikattu et al. 2013, Panda and Slotkin 2013).

In the case of horizontal transfer it is likely that the incoming TE is unique to the genome it enters, so the cell is not able to silence it based on homology. In this case a non-canonical RdDM pathway seems to act to initiate TE silencing. An active TE is transcribed by Pol II and its transcripts are recognized as being somehow aberrant and copied by RNA-DEPENDENT RNA POLYMERASE 6 (RDR6) to produce dsRNAs. It is not known how active TEs are recognized, but there are evidences suggesting that the mobilization of active TEs often produces natural rearrangements or TE tandem or inverted duplications that drive the production of dsRNAs, triggering siRNA production and TGS (Slotkin et al. 2005). dsRNAs are processed by DCL2 and DCL4 into 21-nt and 22-nt siRNAs. These

siRNAs are loaded onto AGO1 and guide the cleavage of TE transcripts in a typical PTGS pathway, representing the first layer of defense of the cell against the new TE. Some of these siRNAs are loaded onto other AGO proteins, which might be AGO2 or AGO6, and act in the nucleus targeting Pol V scaffolding transcripts to initiate DRM2-dependent *de novo* DNA methylation at the active TE site. It is not known how Pol V is initially recruited at these loci, if particular targets are selected or if all regions are targeted by low-level or transient Pol V transcription (Nuthikattu et al. 2013, Panda and Slotkin 2013). At some low-copy number loci it has been observed that Pol II transcription or transcripts can function to recruit Pol IV and Pol V (Zheng et al. 2009). If the TE is methylated at the promoter or other regulatory sequences, Pol II transcription is therefore attenuated or shut off (Inagaki and Kakutani 2013). The established low-level DNA methylation at the TE is then reinforced and stabilized by canonical RdDM that consequently ensures TGS at the TE, independently from Pol II transcription.

An alternative mechanism of TE silencing initiation and transition from PTGS to TGS has been described in *Arabidopsis* for a member of the evolutionarily young, low-copy *ATCOPIA93* family of long-terminal repeat (LTR) retrotransposons, *Evadé (EVD).* Mutants impaired in DNA methylation maintenance mechanisms activate *EVD* transcription, as well as other endogenous TEs in the cell. Pol II *EVD* transcripts are copied by RDR6 into dsRNAs that are processed by DCL2 and DCL4 into 21-nt and 22-nt siRNAs, which are loaded onto AGO1-AGO2 to partly degrade *EVD* RNA through PTGS. Over generations of mutants *EVD* increases the number of new inserted copies in the genome, when it reaches the threshold of ~40 copies dsRNAs levels saturate DCL2 and DCL4 and become available for processing by DCL3 into 24-nt siRNAs, which are loaded onto AGO4 and in the nucleus direct *de novo EVD* methylation. Over a few subsequent generations *EVD* TGS is eventually achieved to through canonical RdDM (Marí-Ordóñez et al. 2013).

Once DNA methylation is established at a certain TE sequence, through the mechanisms described above, it is commonly stabilized through canonical RdDM pathway. Over time, depending on its size, chromatin environment and likely intrinsic sequence features, a TE can exit the RdDM cycle and proceed to a deeply silenced status, in which CHH methylation is lost or reduced and the TE

silencing is maintained and propagated through generations by CG and CHG methylation (Panda and Slotkin 2013, Kim and Zilberman 2014). According to this model, TEs and other repeats can be silenced through different mechanisms, depending or not on RdDM. Short TEs that reside near genes depend on RdDM for constant DNA methylation reinforcement to achieve TGS. In these regions RdDM allows the silencing of TEs in a transcriptionally compatible chromatin environment required by close genes (Kim and Zilberman 2014, Gent et al. 2014). Longer TEs distant from genes only depend on symmetrical DNA methylation for silencing, they can be not transcribed at all, in intergenic inaccessible heterochromatin regions, or they can still produce 24-nt siRNAs required to initiate RdDM homology-dependent silencing of any incoming active TEs with sequence similarity (Nuthikattu et al. 2013, Kim and Zilberman 2014, Gent et al. 2014). In confirmation of this model, in *Arabidopsis* most TEs produce 24-nt siRNAs (Mosher et al. 2008) but mutants impaired in symmetrical DNA methylation maintenance show the reactivation of a greater number of transposons compared to mutants impaired in RdDM pathway, indicating that many TEs are still targeted by RdDM but do not depend on it for silencing (Zemach et al. 2013).

1.4.2.2 RdDM: repression of transposon mobility

There are evidences in *Arabidopsis* that RdDM control of transposon silencing is involved in avoiding the mobilization of activated TEs. *EVD* is activated when METHYLTRANSFERASE 1 (MET1), which propagates CG methylation, is mutated but its transposition is only observed during inbreeding of hybrid epigenomes consisting of *met1*- and wt-derived chromosomes. When combining MET1 mutation with a mutated version of NUCLEAR RNA POLYMERASE D2/NUCLEAR RNA POLYMERASE E2 (NRPD2/NRPE2), encoding the common subunit of Pol IV and Pol V, or a mutated version of SUVH4, the transposition is activated instantaneously and inbreeding is not required (Mirouze et al. 2009). Another *Copia*-type retrotransposon, *ONSEN*, is activated by heat stress to synthetize extrachromosomal DNA that can potentially transpose. The level of *ONSEN* transcripts or extrachromosomal DNA is higher in mutants impaired in siRNA biogenesis, indicating that siRNAs play a role in the regulation of *ONSEN* expression. The TE transposition is not observed in vegetative tissues of the wt

and the mutant of NUCLEAR RNA POLYMERASE D1 (NRPD1), encoding the largest subunit of Pol IV, but only in the progeny of stressed *nrpd1* plants (Ito et al. 2011). These findings demonstrate that RdDM functions not only to suppress TE transcription but also to suppress TE transposition but it is still unknown if siRNAs are able to avoid transposition by targeting extrachromosomal DNA for degradation or by inhibiting their integration into the host genome or through a combination of the two mechanisms (Ito 2012). Moreover, observations on *ONSEN* transposition events indicate that its mobilization happens before gametogenesis and that the siRNA control of TE transposition occurs in the somatic cells that produce the gametes, so in a developmental or tissue-specific manner (Ito et al. 2011).

1.4.3 Biological roles of RNA-directed DNA methylation pathways

Canonical and non-canonical RdDM pathways control TE transcriptional silencing and mobility inhibition, preventing potentially deleterious effects caused by TE movements and participating in the maintenance of genome stability (Ito 2012). In addition to this general role, RdDM is involved in many biological processes of plant development, morphogenesis and reproduction, revealing its great biological importance (Matzke and Mosher 2014).

1.4.3.1 Reinforcement of TE silencing in gametes and seed

The proposed models for reinforcement of TE silencing in gametes and seed in *Arabidopsis* are reviewed in (Feng et al. 2010). In the female gametophyte the central cell is actively demethylated by DEMETER (DME), leading to the activation of TEs and upregulation of RdDM. The TE-derived siRNAs direct *de novo* DNA methylation of TEs in the central cell and might move to the egg cell where they enhance TE silencing. In the male gametophyte several key RdDM proteins are downregulated and DECREASE IN DNA METHYLATION 1 (DDM1), a chromatin remodeler required for DNA and histone methylation and transposon silencing, is only expressed in the sperm cells and not in the vegetative nucleus, leading to the activation of TEs and downregulation of RdDM in the vegetative cell. As for the female gametophyte, TE-derived siRNAs might travel from the vegetative nucleus to the sperm cells where they reinforce TE silencing. Similarly in the seed,

maternal TEs stay activated in the endosperm and produce siRNAs, which might move to the embryo to reinforce silencing. In all cases a similar mechanism seems to occur: the production of TE-derived siRNAs happens in cells not contributing to the genetic information of the next generation. Furthermore, the massive activation of TEs in these cells is not deleterious for the next generation and allows the reinforcement of the TE silencing in the germ line and embryo, which is important to avoid TE transposition being transmitted to the next generation.

1.4.3.2 Genomic imprinting

Genomic imprinting is the phenomenon by which different epigenetic marks are deposited in maternal and paternal alleles resulting in a parent-of-origin-specific expression of genes. There are evidences that genomic imprinting is associated with RdDM. In *Arabidopsis* and rice seed endosperm, Pol IV-dependent siRNAs, derived in part from TEs and repetitive elements, specifically originates from maternal chromosomes (Mosher et al. 2009, Rodrigues et al. 2013). Moreover, all known imprinted genes in *Arabidopsis* are either proximal to or overlapping with siRNA loci (Gehring et al. 2009).

1.4.3.3 Genome interaction

The crossing between two different varieties of the same species or two distinct but closely related species produces intraspecific or interspecific hybrids, respectively. There are evidences that RdDM mediates the epigenetic interactions between maternal and paternal genomes during hybridization and that could contribute to hybrid vigor (Matzke and Mosher 2014). Both in *Arabidopsis* and maize intraspecific hybrids show non-additive levels of 24-nt siRNAs and DNA methylation relative to their parental species (Groszmann et al. 2011, Barber et al. 2012, Greaves et al. 2012). It has been suggested that the non-additive methylation and siRNA expression is probably due to interallelic RdDM: if one allele produces high levels of siRNAs they could target in *trans* the sister non-expressing allele which becomes subjected to RdDM and produces additional siRNAs and becomes methylated; if one allele produces low levels of siRNAs they could be insufficient to target the sister non-expressing allele and not even to

maintain RdDM to the original allele (Greaves et al. 2012). In *Arabidopsis*, both interspecific hybrids and allopolyploids show alteration of siRNA production and TE expression, indicating that siRNAs serve as a buffer against the genomic shock occurring in F1. These changes are stably maintained through generations, suggesting that stable inheritance of transposon-associated siRNA maintains chromatin and genome stability (Ha et al. 2009).

1.4.3.4 Stress responses

The possible influence of environmental stresses on epigenetic silencing mechanisms controlling TEs and the consequent effects in the host are reviewed in (Mirouze and Vitte 2014) (Figure 6). Briefly, silencing pathways like RdDM can be destabilized by biotic or abiotic stresses, inducing DNA hypomethylation. TEs hypomethylation can cause their activation or the activation of nearby genes. An activated TE can produce siRNAs that target a gene in trans and lead to its decrease in expression. Alternatively, an activated TE can transpose and its insertion into a new genomic position can lead to *cis* effects on close regions.

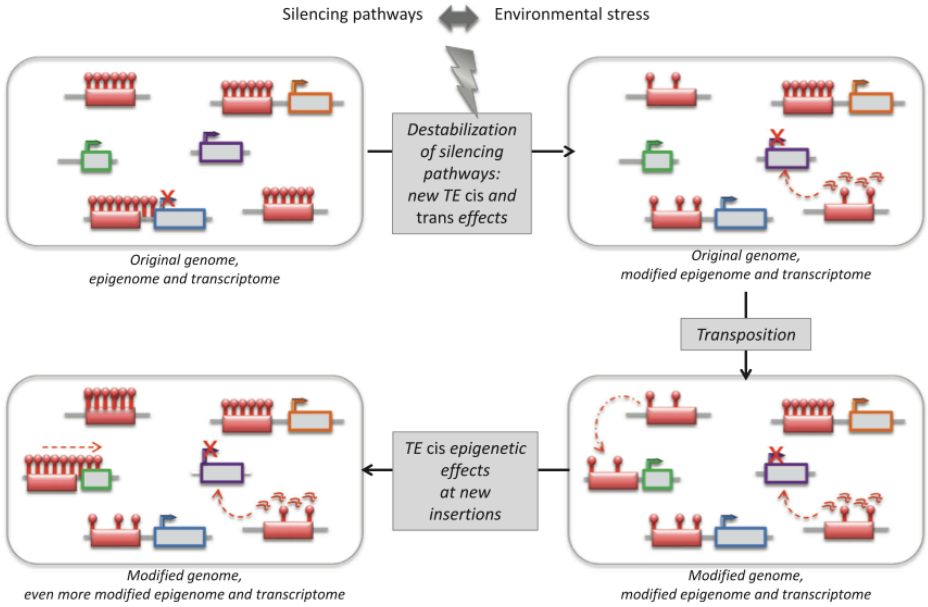

Figure 6: Possible influence of environmental stresses on TE epigenetic silencing pathways and consequences on genome, epigenome and transcriptome. red blocks: TEs. grey blocks: genes. lollipops: DNA methylation (from Mirouze and Vitte 2014).

The influence of abiotic stresses on the methylome has been observed in several cases. For example, low growth temperatures activate the TE Tam3 to transpose in *Antirrhinum majus* and its activity is strictly suppressed at high growth temperatures (Carpenter et al. 1987). DNA methylation of the activated Tam3 is markedly lower than that of the silent TE, suggesting that at low temperatures the siRNA-mediated methylation might decrease, leading to the expression of the previously silenced TE (Hashida et al. 2006). DNA methylation is also altered in plants by biotic stresses. For example, in *Arabidopsis* the infection with the bacterial pathogen *Pseudomonas syringae* causes active demethylation and the impairment of RdDM, which transcriptionally activate TEs (Yu et al. 2013).

Stress-induced demethylation and activation of TEs can cause the simultaneous activation of nearby genes. For example, the exposure of *Arabidopsis* plants to the salicylic acid (SA) hormone causes the differential methylation status of a number of TE-associated regions, which show upregulation of 21-nt siRNAs and are often coupled with differential expression of the TE and/or the proximal gene (Dowen et al. 2012).

An example of a TE-derived siRNA produced following the TE activation and regulating in *trans* a gene involved in stress response has been described in *Arabidopsis*. When retrotransposons of the *Athila* family are epigenetically activated the siRNA854 is produced and regulates in *trans* at post-transcriptional and translational levels the *UBP1b* mRNA, which encodes an RNA-binding protein involved in stress granule formation. The siRNA854 repression of *UBP1b* mRNA results in a phenocopy of the stress-sensitive *ubp1b* mutant phenotype. This demonstrates that the epigenetic activity of TEs can modulate the host organism's stress response (McCue et al. 2012a).

The *Arabidopsis* TE *ONSEN* is an exemplar case in which an abiotic stress transcriptionally activates the TE and the RdDM pathway plays a fundamental role to impede its transgenerational transposition and avoid potentially consequent deleterious effects on the progeny (Ito et al. 2011). However, in the second generation of mutants plants impaired in RdDM treated with heat stress the retrotransposition of *ONSEN* has an impact on the transcriptional regulation of endogenous loci harbouring new *ONSEN* insertions,

which become heat responsive (Ito et al. 2011). The acquired new regulation of these loci could be advantageous for stress adaptation, indicating that TE movements are not always deleterious and can contribute to new phenotypic variation important for evolution.

Stress conditions can alternatively cause DNA methylation of TEs with consequent repression of TE nearby genes. In *Arabidopsis,* low relative humidity induces methylation at a TE sequence and upregulation of TE-derived siRNAs. TE methylation spreads into regulatory and genic regions of the close locus *SPEECHLESS* (*SPCH*), which becomes methylated and decreases in expression. In the same stress conditions another gene, *FAMA*, without close TEs, becomes methylated, siRNAs from its genic sequences are upregulated and it decreases in expression. Both stress-altered genes are involved in stomatal development and their repression is correlated with the reduction in stomatal index that follows stress application. This example indicates that gene expression alteration resulting from environmental stress-induced epigenetic modifications has a measurable biological effect on the plant development and can contribute to the plant stress response. (Tricker et al. 2012)

Environmental stresses have also been documented to cause genome-wide alterations of the siRNA profile: in foxtail millet PEG-simulated drought conditions alter the expression of thousands of sRNA loci (Qi et al. 2013) and in *Brachypodium* cold, heat and salinity stresses provoke the differential expression of hundreds of siRNA sequences with predicted effects on gene expression regulation (Wang et al. 2014b).

1.4.3.5 Formation of epialleles

Epialleles are alleles with identical DNA sequence but different expression levels due to different epigenetic regulation, frequently changes in DNA methylation. They are classified into three groups based on their dependence on genotype: i) obligate epialleles completely depend on a genetic variant, ii) facilitated epialleles depend on a genetic variant only their formation but not for their maintenance and iii) pure alleles, which are independent of any genetic variation. An example of obligate epiallele is that described in *Arabidopsis* for the *FLOWERING LOCUS C* (*FLC*) locus, a central repressor of flowering. The natural *Arabidopsis thaliana*

accession Columbia lacks an insertion in the first intron of *FLC* and the gene express normally. Another natural accession, Landsberg *erecta,* has an insertion of a *Mutator*-like element (MULE) TE into the first intron of *FLC*. The resulting transcript in L*er*, containing the MULE sequence, is targeted by RdDM and produces siRNAs that induces the deposition of repressive chromatin modification at the *FLC* locus. This results in reduced *FLC*-L*er* expression and vernalization-independent early flowering of L*er* (Liu et al. 2004). An example of facilitated epiallele, recovered from a mutagenesis experiment, is that described for the *FLOWERING WAGENIGEN* (*FWA*) locus in *Arabidopsis*. *FWA* is an imprinted gene specifically expressed in the endosperm but silent in vegetative tissues. The tissue-specific imprinted expression of *FWA* depends on siRNA-targeting and DNA methylation of its promoter, comprised of two direct repeats homologous to a short interspersed nuclear element (SINE). The heterochromatic spreading of the TE silencing influences the expression of the nearby *FWA* gene. Mutants that alter siRNA processing or DNA methylation can result in ectopic expression of *FWA*, resulting a late flowering phenotype (Kinoshita et al. 2007). Pure epialleles have been observed in *Arabidopsis* through the study of its DNA methylome at single base pair (bp) resolution. These studies uncovered a rate of base level spontaneous variation in DNA methylation that in some cases significantly influenced the transcription level of the affected locus (Becker et al. 2011, Schmitz et al. 2011).

TEs and their control by RdDM are a source of epialleles formation and thus of genome evolution. This system can act in *cis* or in *trans*, depending if TE polymorphisms influence the expression of nearby or distant genes, respectively. An example of *cis* effects of RdDM TE regulation on gene expression is the case of *FWA*: the heterochromatic spreading of the close TE silencing influences the expression of the *FWA* gene. At the genome-wide level TE methylation spreading to flanking regions does not exist in *Arabidopsis* (Cokus et al. 2008, Ahmed et al. 2011) and it is restricted to particular TE families in maize (Eichten et al. 2012). In *Arabidopsis* also *trans* effects of RdDM TE regulation on gene expression have been documented: in two reported cases, TE-derived siRNAs can regulate the expression of an endogenous gene in *trans* (McCue and Slotkin 2012b, McCue et al. 2013).

There are several cases in which epialleles determine phenotypic consequences for the organism, for example the *peloric* epiallele in *Linaria vulgaris* (Cubas et al. 1999), the *colorless non-ripening* epiallele in *Solanum lycopersicum* (Manning et al. 2006) and the *B'* epiallele in maize (Stam et al. 2002). Epialleles could theoretically be positively selected by evolution in the case in which the phenotypic variation they determine is advantageous for the organism, but whether natural selection operates on epialleles is still not known (Hirsch et al. 2012). The importance of epialleles in crop phenotypic variation and domestication remains unknown, but it has been suggested (Mirouze and Vitte 2014). As reviewed by Springer NM (2013), the generation of epiRILs can create substantial variation for several quantitative traits. epiRILs are individuals of a population of recombinant inbred lines that differ primarily in epigenetic information. They are generated by exposing the genome to a mutation able to remove DNA methylation and then segregating away the mutation and allowing for segregation of genomic segments with altered DNA methylation patterns. It might be possible to use a genetic approach similar to the epiRILs to generate variation in crop plants, but there are several difficulties arising from the nature of the crop plant genomes that would require more studies for the development of these strategies.

1.4.3.6 Genome evolvability

The hypothesis for which epigenetic mechanisms, such as RdDM, have evolved to control invading, parasitic TEs and minimize their deleterious effects on host genomes (Yoder et al. 1997) has been interestingly discussed by Fedoroff NV (2012). The suggested thesis is that epigenetic silencing mechanisms of TEs have evolved to control their activity not simply with the aim to reduce their deleterious effects but also to allow at the same time their accumulation in the host. For example, epigenetic mechanisms also control homology-dependent recombination, without these mechanisms, ectopic, homology-dependent recombination among dispersed TEs would rapidly eliminate them. The maintenance and accumulation of TEs in host genomes is hypothesized to function as a source of genome evolvability: TE activity induces genetic and epigenetic variability and if the TE-induced variation has an adaptive advantage

for the host it could be positively selected by evolution. Because of their sessile lifestyle, plants have no recourse to behavioral responses in coping with stressful environments, so probably they developed a more complex and redundant array of epigenetic silencing mechanisms than animals to keep TEs in their genome as a source of adaptation. Moreover, TEs activity can cause rapid genome restructuring, which is at the heart of eukaryotic evolvability.

1.4.4 Mutations on RNA-directed DNA methylation pathways much greatly affect the phenotype of crops than *Arabidopsis*

RdDM pathways are involved in many important biological processes of plants and both *Arabidopsis* and crops RdDM mutants show many TEs and genes with altered expression levels. Surprisingly, mutants of single RdDM components show little or no phenotype in *Arabidopsis*, while crop plants show more severe phenotypes when the same components are mutated. For example, rice DMR2 mutants are sterile (Moritoh et al. 2012) and DCL3 mutants show significantly reduced plant height at heading stage, increased bending angle of the lamina joint and smaller panicles (Wei et al. 2014), while in *Arabidopsis* these mutations have no such phenotypes (Cao and Jacobsen 2002). Maize mutants for the orthologs of *Arabidopsis RDR2* and *NRPD1*, have striking, albeit stochastic or not fully penetrant pleiotropic, developmental phenotypes, including altered leaf morphogenesis, stunting and flowering defects like feminized tassels (Dorweiler et al. 2000, Parkinson et al. 2007), while in *Arabidopsis* these mutations have no such dramatic phenotypes (Pikaard et al. 2008). Mutations on *Arabidopsis NRPD1*, *NUCLEAR RNA POLYMERASE E1* (*NRPE1*, encoding the largest subunit of Pol V), *RDR2*, *DCL3*, *AGO4* and *DRM* loci, although non-essential in terms of viability, nonetheless play roles in development: under short-day conditions mutants flowering is significantly delayed, as an effect of altered DNA methylation status at the *FWA* locus that affects its expression (Pontier et al. 2005, Chan et al. 2004). These data indicate that RdDM disruption affects multiple plant developmental processes, in particular plant reproduction systems, in a more severe way in crops compared to *Arabidopsis*, suggesting that the epigenetic control of genome stability in crops might be more essential for proper plant development than it is for *Arabidopsis* (Mirouze and Vitte 2014). RdDM

might be more important in crops compared to *Arabidopsis* because of their different content and genomic distribution of TEs and repeats, which are under the RdDM silencing control. *Arabidopsis* has a low amount of TEs that have not been observed to be active in wt plants (Bucher et al. 2012), while TEs are highly abundant in crops and some TEs are active in wt rice and maize plants (Nakazaki et al. 2003, Lisch D 2012). Moreover, *Arabidopsis* TEs have a clear tendency toward clustering in the gene-poor pericentromeric regions (The *Arabidopsis* Genome Initiative 2000) while maize TEs are widely distributed throughout the genome (Meyers et al. 2001), more interspersed with genes and more frequently inserted into gene introns (Haberer et al. 2005).

1.4.4.1 maize RdDM mutants characterized by loss of siRNAs

Maize mutants of RdDM components characterized by the loss of siRNAs that have been identified so far are described below. All of them have been identified in mutant screens for plants unable to maintain paramutation at the *pl1*, *b1*, and *r1* alleles.

Mediator of paramutation1-1 (*mop1-1*) is the mutated allele of *MOP1*, the ortholog of *Arabidopsis RDR2*. The *mop1-1* mutation, identified in the K55 genetic background, causes deleterious pleiotropic phenotypes when compared with the wt, including delayed flowering, shorter stature, spindly and barren stalks, and aberrant development resulting in feminized tassels. Differences in flowering time are reproducible, whereas other abnormalities are variably penetrant and expressive and seem to be influenced by environmental factors (Dorweiler et al. 2000). The *mop1-1* mutation has been initially shown to affect TE methylation: it reduces the cytosine methylation of some elements of the DNA transposon super-family *Mutator* (*Mu*) (Lisch et al. 2002); in particular, the Terminal Inverted Repeat (TIR) regions of the *Mu* gene *mudrA* become hypomethylated immediately in the *mop1-1* background, while the gene is progressively and stochastically reactivated after several generations in the *mop1-1* background (Woodhouse et al. 2006). In further studies, *mop1-1* genome-wide profiles of sRNAs and genes have been analyzed. *mop1-1* shows a dramatic reduction of the 24-nt siRNAs but the retention of the highly abundant TE-derived ~22-nt class of sRNAs (Nobuta et al.

2008). In shoot apical meristems (SAMs) of *mop1-1* mutants, most DE DNA TEs are upregulated, while most DE retrotransposons are downregulated, suggesting that distinct silencing mechanisms are applied to different silencing templates. In addition, more than 6000 genes are DE, including nearly 80% of genes in chromatin modification pathways and key regulators of SAM development, consistently with the different SAM morphology between *mop1-1* and wt plants (Jia et al. 2009). In ear shoots of *mop1-1* mutants, introgressed in the B73 genetic background, cell nuclei show increased chromatin accessibility at chromosome arms. In the same mutants 349 genes are upregulated and 413 are downregulated, suggesting a role for *MOP1* in regulation of higher-order chromatin organization where loss of MOP1 activity at a subset of loci triggers a broader cascade of transcriptional consequences and genome-wide changes in chromatin structure. A subset of the DE genes have been identified as direct targets of the MOP1-mediated RdDM activity, based on multiple signals that include accumulation of 24-nt siRNAs and the presence of specific classes of gene-proximal transposons, but neither of these attributes alone has been found to be sufficient to predict transcriptional misregulation in *mop1-1* homozygous mutants (Madzima et al. 2014). The role of MOP1 in the phenomenon of hybrid vigor has been investigated: despite *mop1-1* mutation significantly reduces plant height and cob weight, delays flowering and, at molecular level, causes a dramatic loss of 24-nt siRNAs, it has little impact on the degree of hybrid vigor displayed by B73xMo17 (Barber et al. 2012). Recently it has been demonstrated that RdDM loci, defined as genomic loci showing loss of 24-nt siRNAs in *mop1-1,* are characterized by relatively high CHH methylation, while non-RdDM loci has low CHH methylation; high CG and CHG methylation are present at all genomic loci except genes (Gent et al. 2014).

Required to maintain repression1 (rmr1) is the mutated allele of *RMR1*, identified in the B73 genetic background. RMR1 belongs to a subfamily of Snf2 proteins defined by Rad54, an ATPase involved in homologous recombination via interactions with single-stranded and double-stranded DNA, as well as CLSY1 and *Arabidopsis* DRD1 (Hale et al. 2007). *rmr1* mutation is not associated with any obvious perturbation of genome homeostasis: the mutants do not show any

gross morphological or sterility phenotype, are not affected in plant height or flowering time, do not show obvious pollen sterility and do not have any large scale cytological defects. Instead, the loss of RMR1 appears to dampen the phenotypic variances typical of inbreeding depression. RMR1 is necessary for the accumulation of the majority of 24-nt siRNAs and the accumulation of the non-polyadenylated RNA transcripts of two families of LTR retrotransposons, as well as RDR2, in a manner that is distinct from the role of Pol IV, which is necessary for the repression of polyadenylated transcripts from the same sampling of elements that are targeted by RMR1 and RDR2. Furthermore, it has been suggested a model in which Pol IV functions independently of the sRNA accumulation facilitated by RMR1 and RDR2 and support that a loss of Pol IV leads to RNA Polymerase II–based transcription (Hale et al. 2009).

Required to maintain repression2 (rmr2) is the mutated allele of *RMR2*, encoding the founding member of a small clade of plant-specific proteins whose molecular function is not obvious. RMR2 affects paramutation at *pl1* allele but not at *r1* allele, is required for the accumulation of 24-nt siRNAs from both repetitive and unique genomic regions, which are not absolutely required to promote paramutation at either *pl1* or *r1*. RMR2 is required for the maintenance of a 5-methylcytosine pattern distinct from that maintained by RNA polymerase IV. These data indicate that RMR2 plays a role in the establishment of paramutation specifically at *pl1* and that it has both Pol IV–overlapping functions and functions distinct from Pol IV, representing a novel component of the increasingly diverse set of nuclear systems available to generate and maintain heritable epigenetic variation in maize (Barbour et al. 2012).

Required to maintain repression6 (rmr6) is the mutated allele of *RMR6*, the ortholog of *Arabidopsis NRPD1*, identified in a genetic background different from B73 (Hollick et al. 2005). Compared to wt, plants homozygous for *rmr6* show pleiotropic phenotypes, including delayed flowering, reduced stature, shorter vegetative internodes, delayed juvenile-to-adult transition and compressed apical inflorescence architecture due primarily to decreased tassel internode length. *rmr6* mutant plants show also other phenotypes: altered abaxial leaf fates, defects on

lateral meristem repression and feminized tassel, which appear for the first time among *rmr6* mutant plants after one or two generations of homozygous sibling crosses, that is after the genome had been exposed to a meiotic division in the absence of RMR6 function. This behavior suggests that RMR6 acts to maintain epigenetic marks at its target loci through meiosis. *rmr6* mutants rarely produce any seed past the S3 generation or any morphologically normal plants past the S2 generation (Parkinson et al. 2007). RMR6 has been demonstrated to be required for the accumulation of the vast majority of 24-nt siRNAs (Erhard et al. 2009).

Required to maintain repression7 (rmr7) is the mutated allele of *RMR7*, the ortholog of *Arabidopsis NUCLEAR RNA POLYMERASE D2a* (*NRPD2a)*, identified in a genetic background different from B73. *NRPD2a* encodes the sole second largest subunit shared between *Arabidopsis* Pol IV and Pol V. *RMR7* is one of three maize loci predicted to encode a protein similar to AtNRPD2a, which appear to express RNA more or less constitutively throughout growth and development. All these three proteins are predicted to be functional. Like RMR6, RMR7 affects paramutation at *pl1* allele and is required for the vast majority of all 24-nt siRNA accumulation, consistent with a Pol IV-type function. In contrast to *rmr6*, *rmr7* mutants do not show any obvious developmental abnormalities. Therefore, the loss of RMR7 function does not completely mimic the loss of RMR6, as *rmr7* mutants have unique molecular, genetic, and morphological phenotypes. These contrasting results suggest that the individual RMR7-type subunits overlap only for certain RNA polymerase functions and that RMR7 is required for only a subset of presumed Pol IV functions, supporting the hypothesis that maize utilizes functionally distinct Pol IV-type RNA polymerases defined by a shared RMR6 together with one or the other RMR7-type subunits (Stonaker et al. 2009).

1.4.5 Small interfering RNA annotation and expression profiling through massive parallel sequencing of small RNAs

The experimental methods that can be used to study siRNAs are the same of those used for miRNAs. Usually siRNA sequences are identified at genome-wide level by massive parallel sequencing of sRNAs and their effective participation in RdDM is confirmed by their absence in RdDM mutants known to be impaired in siRNA production. As for miRNAs, the expression of a certain siRNA can be confirmed through Northern hybridization or qRT-PCR. Downstream and upstream analyses are also performed on siRNAs to characterize their expression profiles, validate their targets, examine their consequences on transcriptional and post-transcriptional silencing of targets and understand their expression modulation.

1.4.5.1 NGS: annotation of siRNA loci

sRNA reads obtained by massive parallel sequencing experiments are first preprocessed and aligned to the reference genome, as describe for miRNA analysis. sRNA sequences can be analyzed as unique individual entities but more frequently they are clustered to identify significant genome loci of sRNA production. Several approaches have been used to identify sRNA loci, in all cases the effective participation of sRNAs in RdDM must be experimentally verified, in order to distinguish siRNA loci from non-siRNA loci.

To identify sRNA loci a simple method is to split the genome sequence into nonoverlapping loci of identical length and select those with sRNA reads overlapping with them for a minimum fraction of their length. For example, in (Gent et al. 2014) among all the maize sRNA loci identified with this approach, those found in intergenic regions with at least 3-fold expression decrease in the mutant of the maize homologous to *Arabidopsis RDR2* are defined as participating in the RdDM pathway. Possible problems of this method are that sRNAs that arise from different transcripts might frequently be inappropriately assigned to the same group or, similarly, sRNAs that arise from the same transcript might be assigned to different groups. To limit these problems an alternative approach is to define a cluster as a group of sRNAs in which each sRNA is separated from its next nearest sRNA by at maximum a set number of nucleotides, and select those

containing a minimum set number of sRNA reads (Moxon et al. 2008, Johnson et al. 2009). There are a number of bioinformatics tools that identify sRNA loci from sRNA-seq data applying different methods and statistics. segmentSeq (Hardcastle et al. 2012) looks for regions of the genome with high densities of sRNA matches, inferring a segmentation of the genome into regions of biological significance. The segmentation is performed simultaneously from multiple samples, taking into account replicate data, in order to create a consensus segmentation of the genome, by an empirical Bayesian method. ShortStack (Axtell MJ 2013b) utilizes a diverse two-step procedure to identify sRNA loci: first, islands with a minimum set coverage of sRNA are identified, second, the initial islands are "padded" up and downstream by a set number of nucleotides and are merged to next overlapping islands to form a cluster. Padding is important to smooth the data when accumulation of sRNAs varies substantially from different regions of the same precursor, which is expected for sRNA loci as the result of differential stabilization of the initial sRNAs based upon AGO loading preferences and strand selection from initial sRNA duplexes (Axtell MJ 2013b). With the same command, in addition to identify sRNA loci, ShortStack also annotates hairpin-associated loci and *MIRNA* loci, tests for the phasing of aligned sRNAs and analyzes loci based on sRNA size composition, strandedness, and repetitiveness.

1.4.5.2 NGS: expression profiling of siRNAs

The same statistics described for miRNAs can be applied to perform differential expression analysis of siRNAs between different samples. First, the abundance of the individual siRNAs or the identified siRNA loci must be normalized with linear o non-linear scaling methods. In contrast to miRNAs, the number of obtained unique siRNAs or siRNA loci is sufficiently high to apply the TMM method for their normalization (Robinson and Oshlack 2010). Differential expression analysis can then be performed, as for miRNAs, with many tools as edgeR (Robinson et al. 2010), baySeq (Hardcastle and Kelly 2010) and DESeq (Anders and Huber 2010).

2 Materials and Methods

2.1 Plant materials

Maize (*Zea mays* L.) stocks used had the following genetic backgrounds:

- inbred line B73;
- *rmr6-1* homozygous mutant introgressed in the B73 reference genome: the *rmr6-1* allele is a loss of function allele resulting from a point mutation in its 8th exon that creates a premature nonsense codon (Erhard et al. 2009).

rmr6-1 seeds were obtained by hand pollination, applying pollen from *rmr6-1* heterozygous plants to the emerging silks of heterozygous *rmr6-1* plants. To select homozygous plants among the segregating F1 population, each plant was genotyped to reveal the presence of the mutation in the *RMR6* alleles.

2.2 Phenol/chloroform extraction and ethanol precipitation of genomic DNA

Genomic DNA was extracted from approximately 100 mg of leaf sample stored at -80°C. The leaf sample was ground to obtain powder and 500µL of Extraction Buffer were added to the tube. The solution was resuspended by vortexing for 2 minutes and then incubated at 65°C for 5 minutes. The resuspension and incubation steps were repeated. 500µL of phenol:chloroform:isoamyl alcohol mix (25:24:1) were added to the tube. The solution was resuspended by vortexing for 2 minutes and than centrifuged at 12,000 *g* for 10 min at room temperature. The pellet was discarded and 400µL of supernatant were collected and transferred into a new tube. 400µL of isopropanol were added to the tube and the solution was gently resuspended by inverting the tube multiple times. The solution was centrifuged at 12,000 *g* for 10 min at room temperature. The supernatant was discarded. 190µL of 70% cold ethanol were added to the tube and the solution was centrifuged at 12,000 *g* for 10 min at room temperature. The supernatant was

accurately discarded and the pellet was dried at 37°C. DNA was finally resuspended in 50µL of sterile H_2O.

Extraction Buffer composition:

-NaCl 0.2M

-EDTA 25mM

-Tris ph7.5 50mM

-SDS 0.5%

2.3 Polymerase Chain Reaction (PCR)

PCR was performed to amplify a region of the *RMR6* allele. The reaction was performed with the *Taq* DNA Polymerase recombinant (Invitrogen™) in a 25µl volume as follows: 0.5µl of genomic DNA extracted with the phenol/chloroform protocol, dNTPs (Invitrogen™) 0.2mM, $MgCl_2$ 3mM, 1x 10x-PCR buffer minus Mg^{++}, forward and reverse oligonucleotide primers 0.4µM each, *Taq* DNA Polymerase recombinant (Invitrogen™) 1 unit, sterile H_2O to reach final volume. Thermal cycling consisted of 5 minutes at 96°C (1 cycle); 1 minute at 95°C, 30 seconds at 57°C, 50 seconds at 72°C (45 cycles); 12 minutes at 72 °C (1 cycle). The length of the amplified *RMR6* region is 283bp.

forward primer: 5'-GAGGGTTTGAATCCATTGGAATGTC-3'

reverse primer: 5'-GGAGTCCTCTAAACCATTGACCG-3'

The primers were provided by Dr. V.Rossi (Consiglio per la Ricerca e la Sperimentazione in Agricoltura, Unità di Ricerca per la Maiscoltura, Via Stezzano 24, I-24126 Bergamo, Italy).

2.4 Restriction enzyme digestion

The amplified region of the *RMR6* allele was digested with the *MwoI* (Fermentas) restriction enzyme. The reaction was performed directly on the PCR product solution in a 25µL volume as follows: 20µL PCR product solution, *MwoI* (Fermentas) restriction enzyme 10 units, 1x 10x-Buffer Tango™, sterile H_2O to reach final volume. The digestion was performed over night at 37 °C. Digestion products (10µl) were analyzed by electrophoresis on 1X TAE agarose gels (2% w/v) and visualized by SYBR® Safe (Life Technologies) staining.

In the amplified region, the wt *RMR6* allele has one *MwoI* restriction site, which is lost in the mutated *rmr6-1* allele. The digestion of the wt allele released three bands: 38bp, 97bp, 148bp, only the 97bp and 148bp bands were visualized in the gel. The digestion of the *rmr6-1* mutant allele released two bands: 38bp and 245bp, only the 245bp band was visualized in the gel.
Examples of obtained products (Figure 1): 1) 1kb Plus DNA Ladder (Invitrogen™), 2) non-digested PCR product, 3)-4)-6)-8) heterozygous *RMR6/rmr6-1* plants, 5)-7) homozygous *rmr6-1/rmr6-1* plants, 9) wt plant.

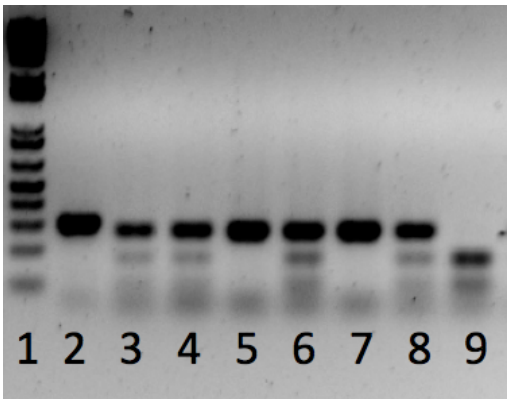

Figure 1: examples of *MwoI* digestion products.

2.5 Stress protocols and tissue collection

Plants from inbred B73 and *rmr6-1* stocks were grown in pots in a greenhouse at the "Lucio Toniolo" experimental Farm of the University of Padova (Legnaro, PD, Italy), with temperatures between 28°C to 30°C at day and 20°C to 22°C at night and relative humidity between 60% to 80%. Plants were watered till pot saturation until the V5/V6 developmental stage, when stress treatments were applied as described in detail in Chapter 1, with some changes compared to the original protocol. Briefly, control plants were watered with 75% of disposable water at 0.1dS/m salt concentration (C); drought-stressed plants with 25% of disposable water at 0.1dS/m salt concentration (D); salinity-stressed plants with 75% of disposable water at 15dS/m salt concentration (S); drought plus salinity-stressed plant with 25% of disposable water at 15dS/m salt concentration (D+S). To mimic the composition of highly saline soils, a complex mixture of salts (Cristal Sea

Marinemix®) was added to water to reach the defined electrical conductivity values. Treatments were applied daily for 10 days and in the 10th day of treatment the youngest wrapped leaf was harvested from each of a subset of plants that were after eliminated. Subsequently, the remaining plants were watered to pot capacity for 7 days to recover from stress; in the 7th day of recovery (+7) the youngest wrapped leaf was harvested from each plant. Leaf samples of same genotype, treatment and sampling time point were pooled together, flash-frozen in liquid nitrogen and stored at -80°C. The complete experiment was replicated three times (R1, R2, R3).

2.6 RNA extraction and sRNA sequencing

Total RNA was extracted from frozen tissue using the Spectrum Plant Total RNA Kit (SIGMA), using "Protocol A" with 750µL of Binding Solution, to recover more of the small-sized RNA, and subjected to On-Column DNase Digestion (SIGMA). Total RNA was quantified spectrophotometrically using a Nanodrop™ 1000 Spectrophotometer (Wilmington, USA) and integrity checked by agarose gel electrophoresis. A total of 48 sRNA libraries (two genotypes, four treatments, two time points, three biological replicates) were produced using the TruSeQ® small RNA Sample Preparation Kit (Illumina) and sequenced on a Illumina Hiseq2000 platform at the Istituto di Genomica Applicata (Udine, Italy). Samples of the R1 biological replicate were sequenced with a multiplexing level of 8, while those of R2 and R3 biological replicates were sequenced with a multiplexing level of 16.

2.7 sRNA data handling

3' and 5' adapters were removed from sequences using cutadapt (Martin M 2011), with default parameters except for the following: "-m 15", to remove reads shorter than 15-nt. Low quality sequences, containing only two different nucleobases, were removed through a customized Perl script. FastQC (Andrews S.) was used to evaluate the libraries quality. ShortStack version 1.2.3 (Axtell, 2013) with default parameters was used in "Mode 2" to map the reads to the maize reference genome (RefGen ZmB73 Assembly AGPv3); ShortStack aligns the reads using bowtie (Langmead et al. 2009), allowing up to one mismatch and randomly selecting one valid alignment per read. The .bam files of the 48 libraries were

merged together and used as input to ShortStack in “Mode 3”, setting the “plant” parameters for the *MIRNA* loci identification and the “--inv_file” produced by the "invert_it.pl" script included in the ShortStack TUTORIAL. Phased loci were identified with ShortStack, the p-values were corrected for multiple testing and a Benjamini-Hochberg adjusted significance level of 0.05 was used.

2.8 Gene and transcript annotation and classification

The same samples used to perform sRNA-seq were sequenced for total RNA at the Istituto di Genomica Applicata and the RNA-seq data were analyzed by my colleagues: from the analysis of these data we recovered the reannotation of the maize transcriptome. Genes and transcripts annotated in this transcriptome assembly were used in our analyses and were distinguished in: ‘protein-coding genes’, ‘TE transcripts’, ‘lncRNA transcripts’. Protein-coding genes were identified from the set of annotated genes following the “protein-coding” classification of the RefGen ZmB73 Annotation AGPv3.20 (gene biotypes and descriptions reported in tables of the ‘Results’ section were recovered from the same source). lncRNA and TE transcripts were instead identified from the set of annotated transcripts. Potential lncRNA transcripts were recovered from the analysis performed by our lab with the collaboration of Sequentia Biotech (Barcelona, Spain). Transcripts were classified as TEs when their sequence overlapped for their entire length, on the same strand, with TEs or repetitive regions annotated in the RefGen ZmB73 repeat-masked Assembly AGPv3, using the superfamilies classification reported in the assembly. TEs of unknown classification were called ‘TXX’ and repetitive regions of unknown classification were called ‘XXX’. The annotation of the complete set of TEs and repetitive regions recovered from the RefGen ZmB73 RepeatMasked Assembly AGPv3 was named ‘repeats’. Gene names were obtained from the Maize Genetics and Genomics Database (http://alpha.maizegdb.org), both ‘classical genes’ and ‘MaizeGDB curated genes’. The 484 chromatin-associated transcripts reported in the Chromatin Database (Gendler et al. 2008) were mapped to the transcripts sequences using criteria of 85% identity and 95% coverage, to identify their correspondent transcripts in the annotation employed. Best *Arabidopsis* and rice BLASTP hits (Altschul et al. 1990) of translated genes were obtained from the Phytozome v10.0 Annotation

v6a. sRNA loci were considered masked by repeats when they overlapped for at least 50% of their length with repeats. Gene Ontology (GO) annotation of genes was obtained by integrating the public Phytozome v10.0 GO Annotation v6a with that produced by our lab with the collaboration of Sequentia Biotech.

2.9 MicroRNA analysis

Mature sequences of known *MIRNA* loci confirmed by our data were manually compared to those annotated in miRBase 20. Data used to make S-plots were obtained with a customized Perl script from the ShortStack "MIRNA detail files". New *MIRNA* loci mature and hairpin sequences were aligned against those reported in miRBase with BLASTN (Altschul et al. 1990), setting "-strand plus": a new *MIRNA* locus was considered a member of a known family when its mature sequences had at most three mismatches with the known miRNAs and miRNA*s. With the same criterion, new *MIRNA* loci mature and hairpin sequences were then aligned against themselves to find new miRNA families. The miRNA targets were predicted using TargetFinder (Fahlgren et al. 2007), the analysis was performed twice: setting the miRNA:alignment penalty score cut-off to the stringent value of 2.5 and to the more permissive value of 3.5. Targets were predicted among the transcripts annotated in the reconstructed assembly performed by my colleagues. Blast2GO (Conesa et al. 2005) was used to perform the GO term enrichment analysis of targets, with the one-tailed "Fisher's Exact Test" function, setting the FDR<5%. The Blast2GO function "GO Distribution by Level" was used to obtain the number of targets associated to each GO term for "Biological Process" and "Molecular Function". InterProScan 5 (Jones et al. 2014) was used to find structural domains in the putative proteins encoded by the genes GRMZM2G381709 and GRMZM2G149108.

2.10 Genomic distributions of sRNA loci and co-occupancy analysis

The length fractions of the chromosome 1-Mb domains covered by each class of sRNA loci were calculated with the BEDTools (Quinlan and Hall 2010) function "coverageBed", with default parameters. For the co-occupancy analysis the count of the observed non-redundant overlapping nucleotides between a sRNA loci class and a genomic feature was obtained with a customized Perl script provided

by Dr. Axtell MJ. The expected number of non-redundant overlapping nucleotides was calculated as follows: ((reference total non-redundant nt/genome size)*(query total non-redundant nt/genome size))*genome size. Enrichment/depletion was calculated as follow: $\log_2$ (observed overlapping nt/expected overlapping nt). The genomic features studied were: protein-coding genes, their exons and introns, TE transcripts, lncRNA transcripts, the 2-kilobases (kb) flanking regions of genes and transcripts, and the repeats. The sRNA loci categories studied were the hairpin and non-hairpin loci with size class from 20-nt to 24-nt.

2.11 Distribution of 23-nt and 24-nt size class sRNA loci in gene and transcript flanking regions

RPKM values of genes and transcripts for wt control samples collected after ten days of experiment were obtained from the analysis of RNA-seq data performed by my colleagues. Protein-coding genes were divided into four equivalent quartiles, from lowest to highest RPKM value. TE and lncRNA transcripts were divided into five groups: one group contained all non-expressed transcripts, corresponding to 72.9% and 51.3% of the total TE and lncRNA transcripts, respectively; the other four groups contained all expressed transcripts, divided into four equivalent quartiles, from lowest to highest RPKM value, each including 6.77% and 12.19% of the total TE and lncRNA transcripts, respectively. The presence or absence of 23-nt and 24-nt size class sRNA loci in each of the gene and transcript upstream and downstream region was obtained with the BEDTools (Quinlan and Hall 2010) function “coverageBed” with the parameter “-d”. A customized Perl script was used to calculate the fraction of genes and transcripts having a close sRNA locus at each position of the flanking regions. The number of overlaps between the flanking regions of genes and transcripts and the sRNA loci with size class of 23-nt and 24-nt, reporting the strand polarity of the two features when defined, was obtained with the BEDTools (Quinlan and Hall 2010) function “intersectBed” with the parameter “-wo”.

2.12 Differential expression analysis

The counts of the miRNA and miRNA* sequences were extracted from each library using a customized Perl script and subjected to pairwise differential

expression analysis using edgeR, setting the FDR<1%. The counts of all identified sRNA loci were obtained from the ShortStack "Results.txt" files of each library and subjected to pairwise differential expression analysis using edgeR, applying the TMM normalization method and setting the FDR<1%. Gene differential expression results, obtained with the tool Cuffdiff (http://cole-trapnell-lab.github.io/cufflinks/), were recovered from the analysis of RNA-seq data performed by my colleagues; for the comparison of genes between the wt and *rmr6-1* mutant, the control samples collected after ten days of experiment were used. Genes with at least one spliced transcript classified as lncRNA were categorized as 'lncRNAs', genes with at least one spliced transcript classified as TE were categorized as 'TEs'. Blast2GO was used to perform the GO term enrichment analysis of up and downregulated genes, with the one-tailed "Fisher's Exact Test" function, setting the FDR<5%. GO terms of up and downregulated DE genes were also scored with the Blast2GO function "Distribution by Level". In the comparison between *rmr6-1* and wt control samples, overlaps between the total and DE sRNA loci with size class from 20-nt to 24-nt and the flanking regions of genes and transcripts were calculated with the BEDTools function "intersectBed" with the parameter "-wo". Similarly, overlaps between the non-DE and DE sRNA loci with size class from 20-nt to 24-nt and gene body and flanking regions of DE genes were calculated with the BEDTools function "intersectBed" with the parameter "-wo".

3 Results

3.1 *De novo* identification of maize leaf sRNA loci by high-throughput sequencing

In order to characterize the sRNA population of maize leaf sRNA-seq experiments were performed on *rmr6-1* homozygous mutants and wt leaf samples of V5/V6 plants grown for ten days under control conditions (C), drought (D), salinity (S) and drought plus salinity (D+S) stresses and subsequently for seven days of watering to pot capacity (+7), to recover form the stress. Three biological replicates were made for each of these 16 conditions (R1, R2, R3). The Illumina sequencing of the sRNA-seq libraries of the 48 samples yielded a total of 4.88E8 raw reads. After removing the low quality sequences and trimming the adapters we obtained 3.59E8 clean reads, which had Q scores ≥ 28 across all bases. A total of 3.36E8 of these reads could be aligned to the maize genome (ZmB73 AGPv3) allowing up to one mismatch: the average value of mapped reads over total reads was 71.8% in wt samples and 63.9% in *rmr6-1* samples.

The length distribution of the majority of aligned reads was observed in the range of 17-nt to 30-nt; a smaller but still considerable number of aligned reads was detected in the range of 32-nt to 37-nt. Focusing on the control samples collected after ten days (Figure 1), wt samples had two major peaks at 24-nt (22%) and 22-nt (7.8%) and five minor peaks at 30-nt (4.8%), 23-nt (4.5%), 21-nt (4.3%), 17-nt (3.4%) and 20-nt (2.1%). In *rmr6-1* mutant samples we observed a reduction of 23-nt and 24-nt sRNAs (respectively 4% and 4.7%), as described in (Erhard et al. 2009). In contrast, the other sRNA size classes were increased in *rmr6-1* mutants compared to wt (by 0.4% to 3.5%).

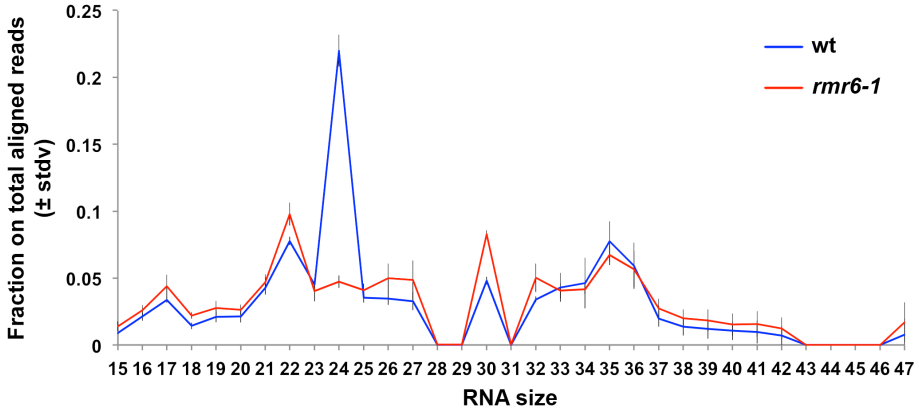

Figure 1 Length distribution and abundance of aligned reads in wt and *rmr6-1* mutant control samples (C). The abundance is reported as fraction of reads with a specific length on the total aligned reads; values are averages of the three biological replicates of control samples collected after ten days (± standard deviation).

The length distribution was consistent across samples with same genotype (Figure 2): neither the stress treatments nor the different developmental stage of plants (+7 samples) caused substantial alterations in the size distribution of the sRNA population. (The *rmr6-1* samples D+S,R1 and S,+7,R1 had higher proportions of 24-nt reads compared to the other *rmr6-1* samples because from a genotype screening, performed on the pools of leaves that were sequenced, they resulted to be contaminated by wt samples, so their read abundances were not further considered in the analyses).

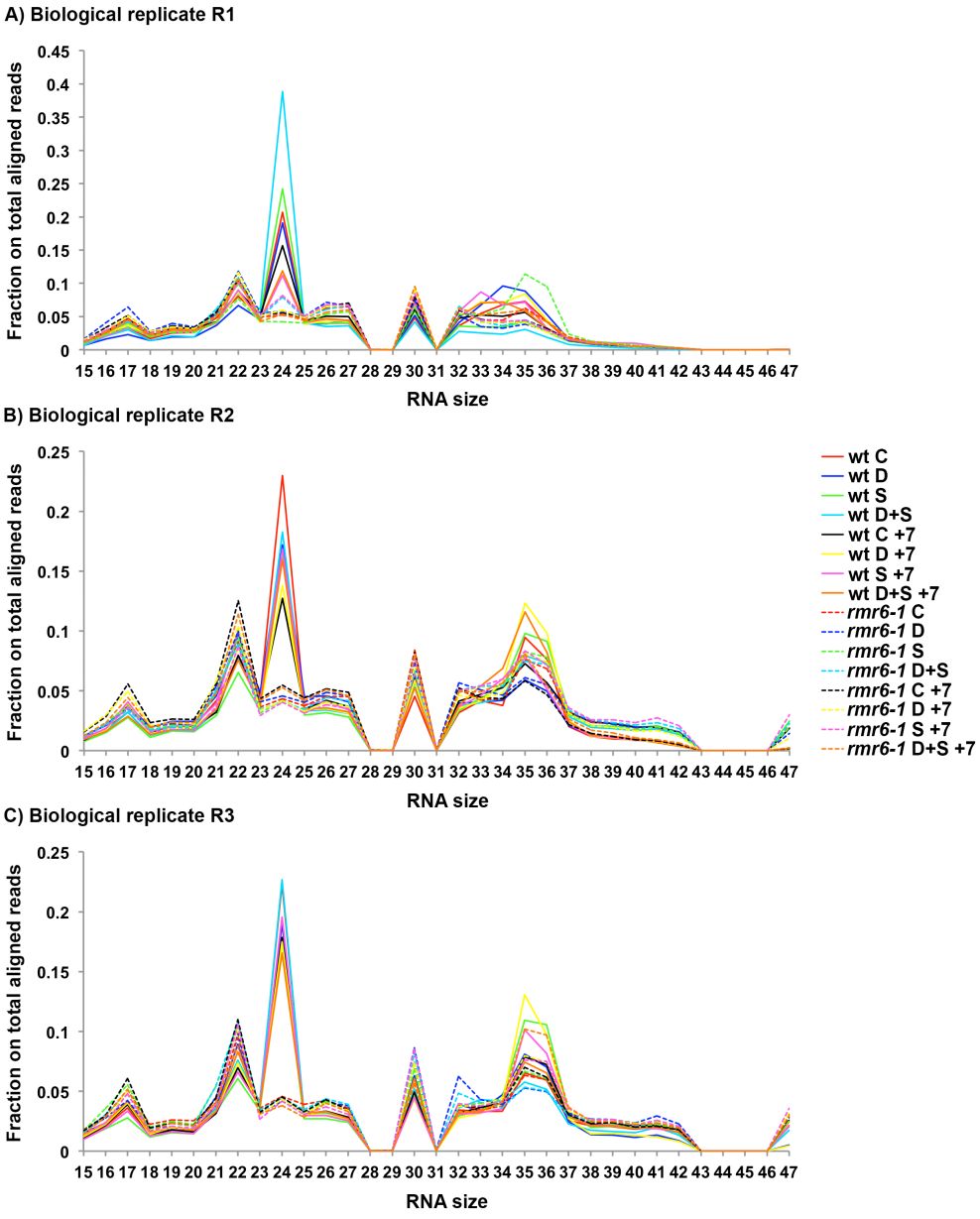

Figure 2 Length distribution and abundance of aligned reads in the 48 samples. The abundance is reported as fraction of reads with a specific length on the total aligned reads. **A) B) C)** plots refer to samples of biological replicates 1, 2, 3, respectively.

To obtain a comprehensive *de novo* annotation of sRNA loci we used ShortStack (Axtell MJ 2013) to predict sRNA clusters from the merged set of all sRNA reads. A total of 188,938 clusters were identified (Table 1), differentiated in *MIRNA* loci, hairpin loci (HP) and non-hairpin loci (non-HP). A size class indicating the most abundant sRNA size observed at the locus was assigned to each cluster (20-nt, 21-nt, 22-nt, 23-nt, 24-nt and N, indicating sizes out of the range from 20-nt to 24-nt). To estimate the consistency of this value across the individual libraries we calculated, for every library separately, for each expressed locus, the fraction of mapping reads with length equal to the size class assigned to the locus. Average and median values of these fractions were calculated for each of the 15 sRNA loci categories considered (*MIRNA*, HP, non-HP loci, of size class from 20-nt to 24-nt) (Appendix A): in wt, 11 categories showed values greater than 0.5 in all libraries, four categories (20-nt HP, 21-nt non-HP, 23-nt HP and 23-nt non-HP, which were among the less numerous), showed a few number of libraries with values smaller than 0.5. In *rmr6-1* mutant, in addition to these four categories, also those with size class of 24-nt exhibited values smaller than 0.5 in the quite totality of libraries, which is explained with the specific loss of 24-nt sRNAs observed in the mutant. These data indicate that for the vast majority of libraries the size class assigned to the loci from the merged set of all sRNA reads still represent the most abundant sRNA size when analyzing the alignments individually for each library.

The majority of reads mapped within loci with size class <20-nt or >24-nt, therefore they were not examined because they were not likely generated by the catalytic activity of DCL proteins. The majority of sRNA loci of 22-nt and 24-nt size class were classified as non-HP precursors of sRNAs, which also accounted for the majority of sRNA alignments within the loci with these size classes. The majority of sRNA loci with 20-nt and 21-nt size class were classified as non-HP precursors of sRNAs, but the greatest numbers of reads mapping within the loci with these size classes were produced from *MIRNA* loci (Table 1).

About 10% of the total sRNA loci were predicted to have a hairpin secondary structure (HP loci) but did not meet the criteria for *MIRNAs*. We analyzed the maximum delta G/stem length (∆G/sl) values of both the HP and the *MIRNA* loci. Values for the HP loci were distributed in the range from -0.5 (the

maximum value accepted to call an hairpin) and -2.4, with a median value of -0.7. Values for the *MIRNA* loci were distributed in the range from -0.6 and -2, with a median of -1.1. While 50% of the HP loci had a ΔG/sl value ≥ -0.7, only 2.8% of the identified *MIRNA* loci had a ΔG/sl value ≥ -0.7. *MIRNA* loci are considered high confidence hairpins, therefore the differences observed in the frequency distributions of the ΔG/sl parameter between *MIRNA* and HP loci indicate that a subset of the predicted HP loci might likely be false positives and lack an effective hairpin structure.

Table 1 Summary of total identified sRNA loci

Size class	Non-HP loci	HP loci	*MIRNA* loci	Non-HP alignments[a]	HP alignments[a]	miRNA alignments[a]
<20-nt or >24-nt	41210	93	0	285406093	4628	0
20-nt	32	26	12	16317	4881	21630
21-nt	520	215	91	904194	757480	2387852
22-nt	17678	1181	12	4109758	623932	105208
23-nt	311	121	2	175044	15739	72513
24-nt	109729	17679	26	19983173	2353610	6064

[a]alignments=total number of reads that mapped within the sRNA loci categories.

Among the loci with 21-nt size class, 19 were predicted to have a phasing pattern of sRNA production (Appendix B). Four mapped within known maize *TAS* genes, *TAS3a-TAS3d* (Nogueira et al. 2007). Six others overlapped with known maize *MIRNA* loci (*MIR159b*, *MIR159f*, *MIR160b*, *MIR167h*, *MIR390b*, *MIR399e*). The other nine phased clusters were not previously annotated, one of these was a novel *MIRNA* locus identified in this study, which overlapped with a protein-coding gene as well as other five loci. Over these nine, four were masked by repeats (RefGen ZmB73 RepeatMasked Assembly AGPv3) of the following super-families: MITE*, Gypsy* and *CACTA*. We did not find any potential phase-initiating miRNA for the identified phased-loci. In some species it has been demonstrated that *TAS3* transcripts are targeted and cleaved by miR390 to direct the synthesis of *trans*-acting siRNAs (Fei et al. 2013) but in our analysis we did not find any potentially cleavable site of miR390 in *TAS3* transcripts because the miR390:*TAS3* alignments did not satisfy the criteria set by TargetFinder to predict a canonical miRNA target site.

3.2 Annotation of conserved maize microRNAs can be refined specifically for the young leaf tissue

143 *MIRNA* loci were identified with the applied *de novo* method (Appendix C). These included 70 out of the 159 maize *MIRNA* loci and 25 out of the 29 maize miRNA families annotated in miRBase 20 (Kozomara and Griffiths-Jones 2011). To evaluate the precision of the existing miRBase miRNA annotations when applied to our young leaf samples, we analyzed the 70 loci that mapped within known maize *MIRNA* loci and compared their mature sequences with those previously annotated. 36 loci had identical miRNA and miRNA* annotations and the mature sequences of all members of the miR390, miR394, miR398, miR528 and miR529 families were exactly confirmed in our data. 34 loci showed different precursor processing that generated mature miRNAs being isomiRs of the annotated sequences (members of the miR156, miR162, miR164, miR166, miR167, miR168, miR171, miR172, miR319, miR393, miR395, miR396, miR397 and miR399 families), or that resulted in a higher expression of the annotated miRNA* compared to the miRNA (members of the miR167, miR169, miR171, miR172, miR393, miR396 and miR399 families), or in a higher expression of unrelated sequences, nonoverlapping with their precursor's annotated miRNA and miRNA* (two members of the miR169 family). We observed that for all these 34 cases the identified mature sequences had higher expression than those previously annotated. These patterns were consistent across all of our libraries. Six representative examples are shown in S-plots (Figure 3), where the abundance of sequences from all 48 libraries that matched to any point within the miRNA precursors were plotted against the 5' positions of the same sequences within the precursors (Jeong et al. 2011). It is possible that the discrepancies found with miRBase annotation reflect leaf-specific differences in *MIRNA* processing patterns, or they may also reflect inaccurate annotations in miRBase.

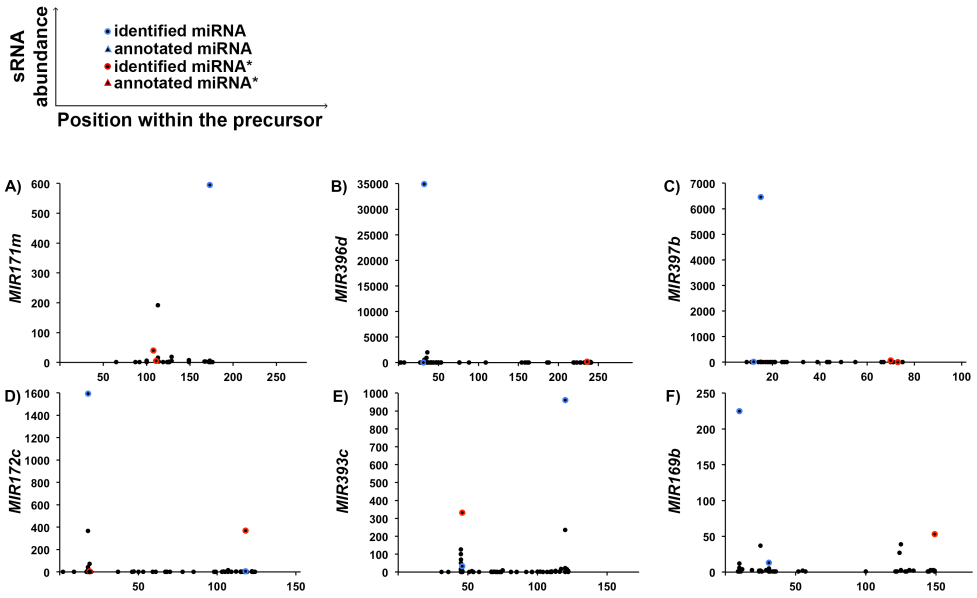

Figure 3 S-plots of *MIRNA* loci producing different mature sequences compared those previously annotated. The abundance of sequences from the merged set of all 48 libraries that matched to any point within the miRNA precursors is plotted against their 5' positions within the precursors. Blue circle, miRNA identified in this study; blue triangle, miRNA reported in miRBase; red circle, miRNA* identified in this study; red triangle, miRNA* reported in miRBase. **A)** *MIR171m*, **B)** *MIR396d* and **C)** *MIR397b* loci: examples of *MIRNA* loci producing mature sequences being isomiRs of those annotated in miRBase. **D)** *MIR172c* and **E)** *MIR393c* loci: examples of *MIRNA* loci producing mature miRNAs corresponding to the miRBase annotated miRNA* sequences **F)** *MIR169b* locus: example of *MIRNA* loci producing mature sequences that do not overlap with those annotated in miRBase.

Not all of the 159 annotated maize *MIRNA* loci in miRBase 20 were confirmed: over the 89 that our *de novo* analysis did not find to have strong *MIRNA* evidence, 25 simply had little or no sRNA reads in our libraries (members of the miR159, miR160, miR169, miR171, miR2118, miR2275, miR395, miR397, miR399 and miR482 families). The other 64 overlapped with regions of significant sRNA production from our samples. Over these 64 loci, one locus, the *MIR396h*, was not confirmed because sRNAs were produced only from the opposite strand where the *MIR396b* locus is located. Over the total 64 loci, 33 were classified as

hairpins: six of these were not predicted to have a possible miRNA/miRNA* duplex within the precursor (members of the miR159, miR160 and miR319 families) and 27 lacked required evidence for miRNA* expression (members of the miR1432, miR159, miR160, miR164, miR166, miR167, miR169, miR171, miR319, miR395 and miR399 families). Over the total 64 loci, 30 did not have a size class between 20-nt to 24-nt or were not classified as hairpins (members of the miR156, miR159, miR164, miR166, miR167, miR168, miR169, miR171, miR172, miR2118, miR395, miR396 and miR827 families). We found that a subset of the loci lacking a valid hairpin structure according to our methods had an identified length much larger than those previously annotated, because of proximal mapping reads that extended their extremities, likely leading the program to fail to fold the entire sequences as hairpins. This was evident for the following *MIRNA* loci that are located in tandem in the genome: *MIR166k, MIR166m* and *MIR2118b, MIR2118d*. In two other cases the known loci were found to be extended and included a previously unannotated, highly expressed sRNA. In the case of the *MIR169j* locus, the new abundant sRNA mapped on the same strand of the miR169j and showed homology with the miR169 family. In the case of the *miR827* locus, the new abundant sRNA mapped on the opposite strand compared to the miR827 and did not show significant homology with any of the miRBase annotated miRNAs. The *MIRNA* methods used were set to minimize false positives, and as consequence, we expected some false negatives (Axtell MJ 2013). Because of this reason we believe that a number of the non-confirmed loci might still be bona fide *MIRNA* loci.

Curiously, none of the members of the most numerous maize miRNA family, miR169 with 17 members, was exactly confirmed by our analysis: three loci were confirmed as *MIRNA* loci but the mature sequences were inverted miRNA/miRNA* or unrelated sequences; three loci did not show the expression of the miRNA*; five loci expressed mature sequences of 19-nt (out of the 20-nt to 24-nt range considered for a likely Dicer-like activity biogenesis); three loci did not show a hairpin structure and three loci were not expressed in our samples.

3.3 Novel *MIRNA* loci are enriched in class II DNA transposable elements

Due to the essential role of miRNAs in the negative regulation of gene expression at the posttranscriptional level it is important to search for new uncharacterized miRNAs. The putative novel *MIRNA* loci were predicted based on the merged set of sRNA sequencing data of the 48 samples. We applied an abundance threshold, keeping only those *MIRNA* loci whose predicted mature sequence showed at least five reads in at least one library: 15 loci did not pass the filter and were not included in further analyses, while 58 loci passed the filter and were considered as new *MIRNA* loci (Appendix C). Over the total 58 loci, five were new members of the known miRNA families miR156, miR166 and miR167: four of them (*MIR-NEW156m, MIR-NEW166o, MIR-NEW166p, MIR-NEW167k*) showed 99% identity with *MIRNA* loci reported in miRBase but lacking a genome annotation (respectively, *MIR156c, MIR166g, MIR166b, MIR167i*). Over the total 58 loci, 53 were new loci with novel mature sequences, belonging to 46 new miRNA families. They had low expression levels on the individual samples, which is typical of the less-conserved miRNAs (Ma et al. 2010). Compared to the conserved *MIRNA* loci that are located mainly in intergenic (53%) and exon (34%) regions, the predicted novel *MIRNA* loci were found mainly in intergenic regions (40%) and in exon and intron regions with the same probability (~23%) (Appendix C). 37 out of the 53 new loci with novel mature sequences were masked by repeats, mainly TIR TEs and MITEs super-families: 19 miRNAs were 24-nt long and 15 out of 19 had the 5' terminal A, which are typical characteristics of siRNAs. These results suggest that of the 37 loci masked by repeats some of them migth be new 'proto-miRNAs', as it has been shown for a number of TE-derived miRNAs (Li et al. 2011), or siRNAs actively transcribed from TE rearrangements, involved in the establishment of transcriptional silencing (Lisch D 2012).

3.4 Target prediction of conserved miRNAs can be improved including assembled transcripts from total RNA-seq experiments

In order to assess the function of miRNAs, the potential targets of the conserved and the putative novel miRNAs were predicted with the tool TargetFinder (Fahlgren et al. 2007). We decided to use the set of transcripts that were reconstructed from total RNA-seq data obtained from the same samples used in this study to perform sRNA-seq. This allowed testing if previously unannotated transcripts found to be expressed in our samples could be the targets of miRNAs. Among the total transcripts, TEs and potential lncRNAs were identified as described in the chapter Materials and Methods.

We first predicted the potential targets of the conserved miRNAs. The penalty score cutoff applied to the identified miRNA:target alignments was stringent (≤2.5) (Appendix D) but still allowed us to capture most of the miRNAs targets that were conserved across different plant species, including *Arabidopsis* (Adai et al. 2005) and rice (Sunkar et al. 2005). Analysis of target enrichment in GO molecular function and biological process categories showed that targets of conserved miRNA were enriched in activities related to the DNA-dependent regulation of transcription, confirming that the majority of them were transcription factors (TFs) (targets of miR156, miR159, miR160, miR164, miR171, miR172, miR319, miR396 and miR529 families) (Zhang et al. 2009). We confirmed target prediction also for the following miRNA families: miR390, miR393, miR394, miR395, miR397, miR408 (Zhang et al. 2009), miR162 (Zhang Z et al. 2008). Known targets of miR166, miR2275 and miR528 families (Liu et al. 2014, Zhang et al. 2009) were confirmed by our analysis only when using a more permissive miRNA:target alignment penalty score cut-off (≤3.5) (not reported). For the miR166 family we also predicted other targets with better scores than the canonical targets: two uncharacterized transcripts, one of which was classified as a potential lncRNA, and a TE transcript.

Even with more permissive scores, known targets of four miRNA families were not detected in our study: miR2118 was predicted to target an uncharacterized transcript, for miR168 and miR398 we failed to predict any targets, while miR167 was predicted to target homologous proteins of *Arabidopsis* and rice pumilio-family RNA binding proteins instead of the previously reported

ARF TFs (Zhang et al. 2009). The miR169 known targets, NFYA TFs (Zhang et al. 2009), were also not detected because only 3 out of the 17 annotated *MIR169* loci were confirmed by our study: all were found to produce mature sequences unrelated to those previously annotated, one was also found to produce a second miRNA/miRNA* duplex where the mature miRNA corresponded to the annotated miRNA*, but none of the identified mature sequences was predicted to have targets.

We obtained an interesting result for the miR399 family. In *Arabidopsis* miR399 targets the *PHO2* gene (*UBC24*, encodes an ubiquitin-conjugating E2 enzyme), which is a major component for the maintenance of Pi homeostasis (Bari et al. 2006). The miR399 is upregulated by Pi starvation and its target is downregulated, through transcript cleavage (Allen et al. 2005) and probably also through translational repression (Bari et al. 2006). The miR399 contributes to the regulation of the Pi homeostasis and it was hypothesized to act as a long-distance Pi starvation signal (Pant et al. 2008). In *Arabidopsis,* miR399 has been experimentally verified to cleave the 5'-untranslated region (UTR) of the *PHO2* gene at five target sites distributed in a range of 300-bp (Allen et al. 2005). In maize, miR399 was previously described to target genes encoding inorganic phosphate transporters, a number of genes with unknown function (Zhang et al. 2009) and, more recently, the GRMZM2G149108 gene encoding a putative ubiquitin-like 1-activating enzyme E1A (Wang et al. 2014a). While the *Arabidopsis PHO2* gene possess the structural domains ubiquitin-conjugating enzyme/RWD-like (IPR016135) and ubiquitin-conjugating enzyme, E2 (IPR000608), the GRMZM2G149108 gene was found to have three different domains: NAD(P)-binding domain (IPR016040), molybdenum cofactor biosynthesis, MoeB (IPR009036) and UBA/THIF-type NAD/FAD binding fold (IPR000594), so the previously reported miR399 maize targets are not homologous to the *Arabidopsis PHO2* gene. We predicted three targets for the miR399 family: an uncharacterized transcript, previously reported (Zhang et al. 2009), that resulted to be a potential lncRNA, a TE transcript and a new transcript detected in our samples. The new transcript, named TCONS_00124738, was found to harbor up to six putative miR399 target sites, distributed in a range of 417-bp: six sites for the miR399a, miR399c, miR399e and miR399j and five for the miR399f, with alignment penalty

scores between 1.5 and 3.5 (Figure 4A). This miRNA:target binding pattern was very similar to that described in *Arabidopsis* between miR399 and the *PHO2* transcript. The TCONS_00124738 transcript 3'-end was located 93 bp upstream of the gene GRMZM2G381709 (Figure 4B) that has a short annotated 5'-UTR of 52-nt and encodes a putative ortholog of the *Arabidopsis* PHO2 (Calderón-Vázquez et al. 2011). We compared the amino acid sequences of the two proteins and found that 97% of the maize sequence was covered, with 45% identity, by the *Arabidopsis* sequence. Moreover, the structure of the coding sequence was conserved in these two species, except that one of the exons was split into two in maize. We suggest that the previously unannotated transcript, identified through total RNA-seq, may constitute the complete 5'-UTR of the downstream gene, which could therefore be the target of the miR399.

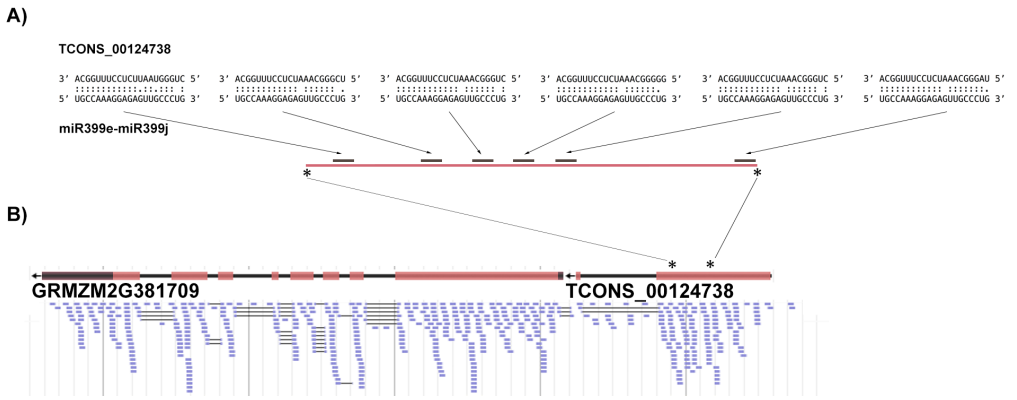

Figure 4 miR399 predicted target sites on the newly annotated transcript TCONS_00124738. A) miR399:TCONS_00124738 alignments for the miR399e-miR399j mature sequence. ":", ordinary Watson-Crick base pair; ".", G:U base pair; " ", mismatch. The arrows indicate the position of the alignments on the TCONS_00124738 zoomed region. **B)** RNA-seq reads from the wt control sample (R2) mapping to the GRMZM2G381709 gene and the newly identified TCONS_00124738 transcript. Gene exons are represented as red blocks, introns as black lines, 5'-UTR and 3'-UTR as grey blocks, arrows indicates the genes are located in the negative strand of chromosome 6.

3.5 Predicted targets of maize-specific miRNAs have different characteristics compared to those of conserved miRNAs

The potential targets for 28 out of the 53 unique putative novel miRNAs were successfully predicted (Appendix D). Compared to the predicted targets of the conserved miRNAs, those of the novel, maize-specific miRNAs showed a number of differences: i) in most cases the targets were uncharacterized transcripts, without any sequence homology with both *Arabidopsis* and rice proteins, or the matched sequences were not annotated in these reference species; ii) none of the GO molecular function and biological process terms assigned to the targets were enriched: the targets of novel miRNAs had diverse biological functions and, unlike conserved miRNAs, we also found many TE transcripts among them; iii) some putative novel miRNAs resulted to target multiple genes belonging to different cellular pathways.

Even if not significantly enriched, the GO terms related with the DNA-dependent regulation of transcription were the most represented, for example miR-NEW12 was predicted to target a putative MYB TF and miR-NEW19 a putative WRKY TF. Two miRNAs, miR-NEW18 and miR-NEW20, both of which overlapped with MITEs, resulted to have the same putative target gene encoding a putative ABC transporter, while two potassium transporters were predicted to be the target of miR-NEW1. Zinc finger C3HC4 type domain containing proteins were a common predicted target family for miR-NEW10, miR-NEW21, miR-NEW22 and miR-NEW28.

The observed differences between the predicted targets of the conserved and the putative novel miRNAs may confirm our hypothesis: the putative novel miRNAs might be transient pre-evolved miRNAs, only a subset of them with an effective selective value (Axtell MJ 2008), or they migth be siRNAs and not miRNAs.

3.6 Most abundant miRNAs are conserved miRNA/miRNA* sequences

The expression of the mature miRNAs irrespective of their genomic origins were analyzed. The conserved miRNAs were evaluated first and they varied significantly in expression levels. The miR166 family showed the highest abundance in all libraries (from 2000 to 16000 Reads Per Million, RPM=(number of miRNA reads/total number of reads aligned to the genome)*10^6); in maize, miR166 pattern accumulation in the leaf establishes organ adaxial/abaxial polarity (Juarez et al. 2004). The other miRNA involved in the same control mechanism of the leaf dorsoventral polarity, miR390 (Nogueira et al. 2007), showed lower abundances (from 30 to 270 RPM). The miR168 was highly expressed in all samples (from 190 to 1600 RPM): by the targeting of AGO1, miR168 maintains the steady-state balance of the RNA silencing machinery (Vaucheret et. al 2004). Also the miR168* showed high expression levels (from 57 to 900 RPM) in all libraries. Among the most abundant miRNAs there was also the miR396 family, which is involved in the regulation of cell expansion in leaf (Wang et al. 2011); the two members miR396c and miR396d showed significantly higher expression (from 47 to 550 RPM) than the other members (< 22 RPM), like it was previously reported for juvenile tissues (Zhang et al. 2009). This result suggests that the regulatory role of a miRNA family can be exerted by a restricted number of its members in the young leaf tissue. This behaviour was observed also for other miRNA families highly abundant in all leaf samples that preferentially expressed the following members: miR399e and miR399j; miR160b and miR160g; miR156a, miR156b, miR156d, miR156f, miR156g, miR156h, miR156l and miR156m; miR167e, miR167f, miR167g, miR167k, miR167d (identical to the miR167d* annotated in miRBase) and miR167k*. The following miRNA families were expressed at low levels in all libraries (< 25 RPM): miR171, miR172, miR2118, miR2275, miR169, miR393, miR394, miR395 and miR529. The miR172 and miR156 families are involved in the vegetative phase change and they are characterized by anti-correlated expression levels in the juvenile and adult phases of development (Chuck et al. 2009). Indeed we observed that in our samples the miR156 was highly expressed while the miR172 had low abundances. Other miRNAs were very low expressed possibly because they were active in different tissues, as it was previously observed for the miR529, that shares the same

targets, SBP-box TFs, with the miR156, but has a tassel-specific expression (Zhang et al. 2009), and for the miR2118 and miR2275 families that in rice are mainly or exclusively expressed in the stamens (Song et al. 2012). A number of *MIRNA* loci were not detected in our dataset (members of the miR159, miR160, miR169, miR171, miR2118, miR2275, miR395, miR397, miR399 and miR482 families). The miR482 family consists of only one member and was absent in our samples, indicating that its expression may be developmental and/or tissue-specific.

The vast majority of the putative novel miRNAs had very low abundances (< 10 RPM) in all samples. The most expressed miRNA family was the miR-NEW10, homologous to *Mu* repeat elements, with four members, each predicted to give rise to two miRNA/miRNA* duplexes. The mature sequence produced by one of the two duplexes, named miR-NEW10.1, was identical between the four miR-NEW10 family members and had higher abundances (from 45 to 300 RPM) compared to the sequences produced by the other duplex, named miR-NEW10.2. Since the four members of the miR-NEW10 family shared the same miR-NEW10.1 mature sequence, which only mapped to these four genomic locations, we could not assess if all the loci were effectively responsible for the production of the miRNAs. The putative novel miRNA miR-NEW58 had high expression levels in all samples (from 35 to 250 RPM), its locus was not masked by repeats and its predicted target, a low confidence gene, was not detected in our samples through RNA-seq. Other four putative novel miRNAs showed an abundance > 10 RPM: miR-NEW4, miR-NEW24, miR-NEW26 and miR-NEW34, whose loci were all masked by repeats.

3.7 Long-term abiotic stresses and plant development affect the expression of a few numbers of miRNAs

The unique mature miRNA sequences were tested for differential expression (log_2FC >1 or <-1) with the tool edgeR (Robinson et al. 2010) (Table 2). We also included the sequences of the conserved miRNAs whose precursors were not confirmed by our analysis or that lacked a genome annotation in miRBase, to see if there were DE sequences among them. The 25 miRNAs that we earlier found to be not expressed in our samples were excluded from the analysis. The effects of the abiotic stresses applied to the plants for the long interval of ten days, of the recovery from the stresses and of the plant development (samples in control conditions at +7) were evaluated: none of them caused strong effects on miRNA expression. Only the DE miRNAs, more numerous than the DE miRNA*'s, are discussed below.

The conserved miRNAs were evaluated first. Two mature sequences of the conserved miR156 family (one encoded by the *MIR156a*, *MIR156c*, *MIR156e*, *MIR156f*, *MIR156g*, *MIR156h*, *MIR156i*, *MIR156l* and *MIR156m* precursors and the other encoded by the *MIR156b* and *MIR156d* precursors) were upregulated in the wt after ten days of drought stress. In *rmr6-1* mutant only the miR156 sequence encoded by the miR156b and miR156d precursors was upregulated following both the drought and the salinity stresses. We found that the miR156d precursor gave rise to the canonical miRNA/miRNA* duplex annotated in miRBase (here renamed 156d.1) and to a second putative duplex (named miR156d.2), without sequence homology with the miR156 family. The miR156d.2 mature sequence was upregulated in wt following the drought stress treatment and its expression decreased after the recovery time; it was altered also in the *rmr6-1* mutant, showing an upregulation after all the three applied stresses, but the expression remained high after the recovery. We failed to predict any target for the miR156d.2, so its biological role remains to be elucidated. Four miRNAs were DE only in the wt following the ten days of drought stress: miR397b and the miR398 family were upregulated, while one mature sequence of the miR166 family (encoded by the *MIR166b*, *MIR166c*, *MIR166e, MIR166f*, *MIR166g* and *MIR166h* precursors) and one mature sequence of the miR396 family (encoded by the *MIR396e* and *MIR396f* precursors) were downregulated. The miR166 was

the only DE miRNA encoded by miRNA precursors not confirmed by our analysis and showed differential expression also in wt samples in control conditions between the two time points of sample collection. Three miRNAs were DE only in the *rmr6-1* mutant following the ten days of drought plus salinity stress: the miR319c was upregulated and the miR399b and miR528 family were downregulated. Except for the miR156d.2, all the other DE miRNAs remained up or downregulated after the recovery, suggesting that the pathways regulated by the miRNAs responsive to the long-term abiotic stresses might continue to be altered even when the stress has been removed.

The differential expression analysis of the novel miRNAs revealed that miR-NEW46 oppositely responded to the drought plus salinity stress in the two genotypes. The mature sequences of the two miRNA/miRNA* duplexes produced by the miR-NEW46 precursor were named miR-NEW46.1 and miR-NEW46.2: miR-NEW46.2 was downregulated in the *rmr6-1* mutant after the ten days of treatment, while both miR-NEW46.1 and miR-NEW46.2 were upregulated in wt after the recovery from the stress. miR-NEW46.1 was predicted to target a putative AP2 TF and both miR-NEW46.1 and miR-NEW46.2 were predicted to target a TE transcript; both transcripts were not detected in our samples in these conditions through RNA-seq. Three other putative novel miRNAs were DE only in wt: miR-NEW24 was upregulated following the ten days of drought stress, and miR-NEW34 (encoded by the *MIR-NEW34a* and *MIR-NEW34c* precursors) and miR-NEW58 were upregulated in control conditions at +7.

None of the predicted targets of the DE miRNAs were found to be DE in our samples, according to the differential expression analysis performed on gene counts obtained through RNA-seq.

Table 2 DE miRNAs in stress conditions and/or at the developmental stage of plants at +7

miRNA[a]	Comparison[b]	log_2FC[c]
miR156a-miR156f-miR156g-miR156h-miR156l-miR-NEW156m-*miR156c-miR156e-miR156i*	wt, D vs C	1.2
miR156b(isoMIR)-miR156d.1(isoMIR)	wt, D vs C	1.86
	rmr6-1, D vs C	1.1
	rmr6-1, S vs C	1.1
miR156d.2(unrelated)	wt, D vs C	2.5
	wt, D, +7 vs +0	-2
	rmr6-1, D vs C	1.5
	rmr6-1, S vs C	1.5
	rmr6-1, D+S vs C	1.3
miR166b-miR166c-miR166e-miR166f-miR166g-miR166h	wt, D vs C	-1.23
	wt, C, +7 vs +0	-1.35
miR319c(isoMIR)	*rmr6-1*, D+S vs C	3
miR396f*(miRNA)-*miR396e*	wt, D vs C	-1.86
miR397b(isoMIR)	wt, D vs C	1.2
miR398a-miR398b	wt, D vs C	1.5
miR399b	*rmr6-1*, D+S vs C	-2.6
miR528a-miR528b	*rmr6-1*, D+S vs C	-1.8
miR-NEW24	wt, D vs C	1.5
miR-NEW34a-miR-NEW34c	wt, C, +7 vs +0	1.65
miR-NEW46.1	wt, D+S, +7 vs +0	2.7
miR-NEW46.2	wt, D+S, +7 vs +0	3.7
	rmr6-1, D+S vs C	-3.7
miR-NEW58	wt, C, +7 vs +0	1.8

[a]()=when the identified miRNA sequence was not identical to that reported in miRBase, their relationship is indicated: isoMIR=isoMIR of the miRNA annotated in miRBase; unrelated=nonoverlapping with miRBase annotated sequences. miRNAs in italic are sequences whose precursor was not confirmed by our analysis or that lacked a genomic annotation in miRBase.

[b]C=control; D=drought stress; S=salinity stress; D+S=drought+salinity stress. +0=ten days of treatment; +7=seven days of recovery.

[c]FDR<1%.

3.8 Among the putative novel miRNAs homologous to repeat elements only the 24-nt species are Pol IV-dependent

Some differential expression of miRNAs between wt and *rmr6-1* were also observed (Table 3). Three mature miRNAs were upregulated in the *rmr6-1* mutant compared to the wt: miR156d.2, miR-NEW10a.2-miR-NEW10b.2 and miR-NEW24. The following mature miRNAs were instead downregulated in the *rmr6-1* mutant compared to the wt: miR-NEW1, mir-NEW4, miR-NEW5, miR-NEW15, miR-NEW26 (encoded by the *MIR-NEW26a* and *MIR-NEW26b* precursors), miR-NEW29 and miR-NEW30. When performing the differential expression analysis on *MIRNA* loci rather than individual mature miRNAs, we confirmed the

downregulation of the reported miRNAs and also found six other loci downregulated in the mutant: *miR-NEW8*, *miR-NEW22*, *miR-NEW27*, *mir-NEW28*, *miR-NEW31* and *miR-NEW32*. The 14 downregulated miRNAs (mature sequence or locus) were all 24-nt long and masked by repeats. Five other putative novel miRNAs with the same characteristics were not downregulated in *rmr6-1*. None of the 21-nt and 22-nt putative novel miRNAs homologous to repeat elements showed a downregulation in the *rmr6-1* mutant. These results suggest that only the 24-nt putative novel miRNAs homologous to repeat elements depend for their transcription on the activity of the Pol IV enzyme and therefore are the most likely to be siRNAs instead of bona fide miRNAs.

Table 3 DE miRNAs and *MIRNA* loci in *rmr6-1* mutant compared to wt (C)

Differentially expressed	**miRNA[a]**	**log_2FC[b]**
both miRNA and *MIRNA* locus	miR156d.2(unrelated)	1.3
	miR-NEW1	-6.4
	miR-NEW4	-9.3
	miR-NEW5	-6.8
	miR-NEW10a.2-miR-NEW10b.2	1.9
	miR-NEW15	-7.3
	miR-NEW24	1.4
	miR-NEW26a-miR-NEW26b	-6.2
	miR-NEW29	-7
	miR-NEW30	-6.4
only *MIRNA* locus	miR-NEW8	-5.4
	miR-NEW22	-5.3
	miR-NEW27	-3.4
	miR-NEW28	-6.9
	miR-NEW31	-4.9
	miR-NEW32	-5.5

[a]()=when the identified miRNA sequence was not identical to that reported in miRBase, their relationship is indicated: unrelated=nonoverlapping with miRBase annotated sequences.

[b]FDR<1%. For those miRNAs for which both the mature sequence and the precursor were DE, the log_2FC of the mature miRNA is reported. For those miRNAs for which only the precursor was DE, the log_2FC of the *MIRNA* locus is reported.

3.9 Gene flanking regions tend to be enriched in sRNA loci of 21-nt, 23-nt and 24-nt size class and depleted in sRNA loci of 22-nt size class

The vast majority of the identified sRNA loci were non-*MIRNA* loci, so to complete the characterization of the sRNA population of maize leaf we analyzed all the sRNA loci excluding the *MIRNA* loci, both those identified in our work and those previously annotated in miRBase but not confirmed by our analysis. These sRNA loci, divided by size class, from 20-nt to 24-nt, and by precursor structure, HP and non-HP, were first examined for their genomic and genic distributions.

We evaluated the genomic distribution of the sRNA loci locations (not their abundances) by plotting for all the chromosomes, for each of their 1-Mb domains, the fraction of the domain length covered by the sRNA loci. The sRNA loci with a size class of 22-nt and 24-nt, the most numerous size classes, were analyzed first (Figure 5). Both the HP and non-HP loci with a size class of 22-nt showed uniform low levels of chromosome domain occupancy across the chromosomes length, with a number of spikes observed in both chromosome arms and centromere regions. The non-HP loci with a size class of 24-nt, the most numerous category of sRNA loci, occupied mainly the chromosomes arms, with highest fractions of chromosome domains covered bases towards the telomeres and lowest fractions towards the centromeres. The HP loci with a size class of 24-nt showed the same trend of the non-HP loci with same size class but with less marked differences between the centromere and the chromosome arms.

The sRNA loci with a size class of 20-nt, 21-nt and 23-nt were analyzed separately from the other loci because they were less numerous (Figure 6). Each of these categories was composed by a small numbers of loci, therefore their observed fractions of chromosome domains covered bases were zero across all the chromosomes length, with a few number of spikes mainly concentrated in the chromosomes arms. The observed chromosomal distribution of the sRNA loci confirmed the previously described chromosomal distributions of the sRNA abundances (Gent et al. 2012) and the sRNA loci (He et al. 2013), respectively obtained in maize root tips and seedlings shoots and roots.

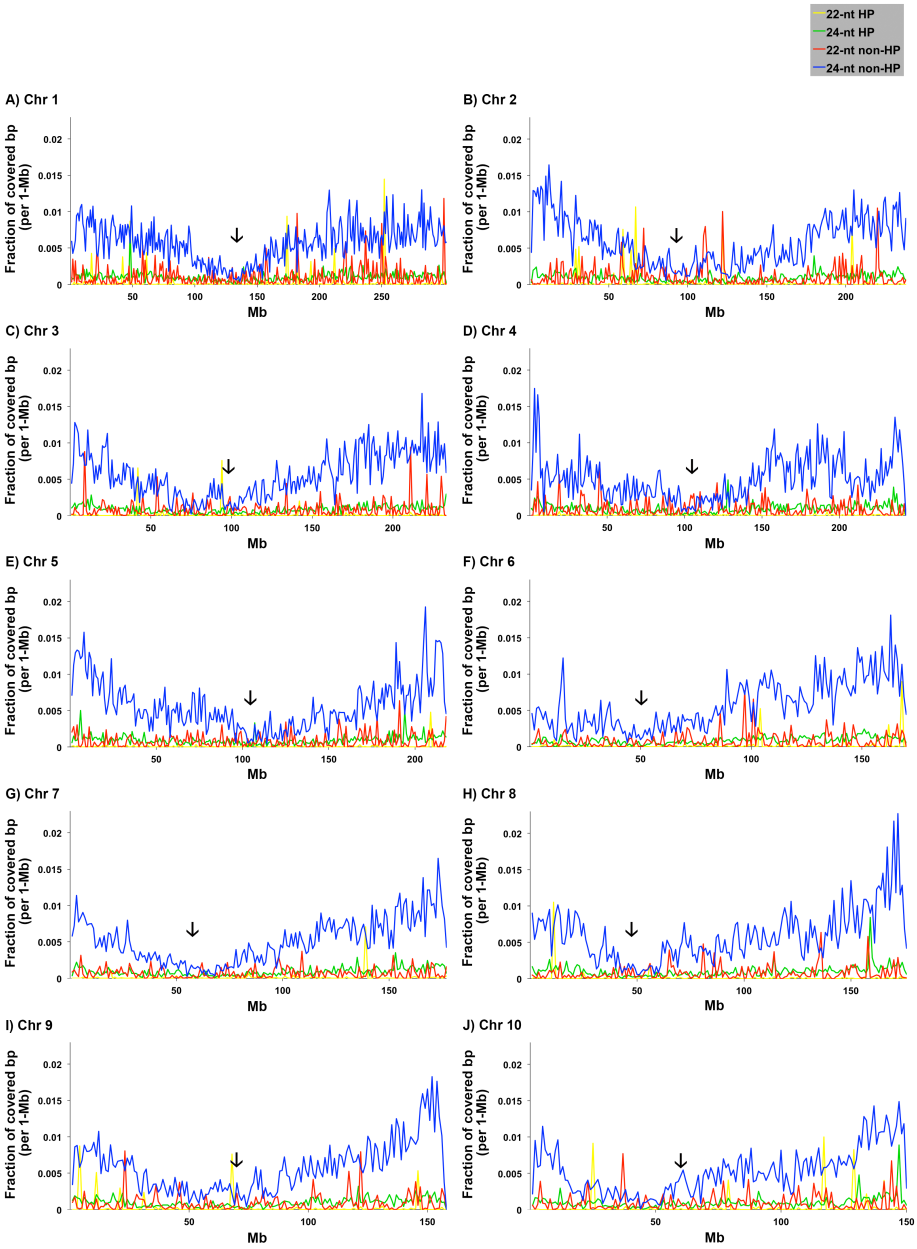

Figure 5 Chromosomal distribution of sRNA loci with size class of 22-nt and 24-nt. For each chromosome, for each of its 1-Mb domain is plotted the fraction of the domain length covered by the sRNA loci. **A) - J)** chromosomes 1 - 10. Centromere positions are indicated by black arrows.

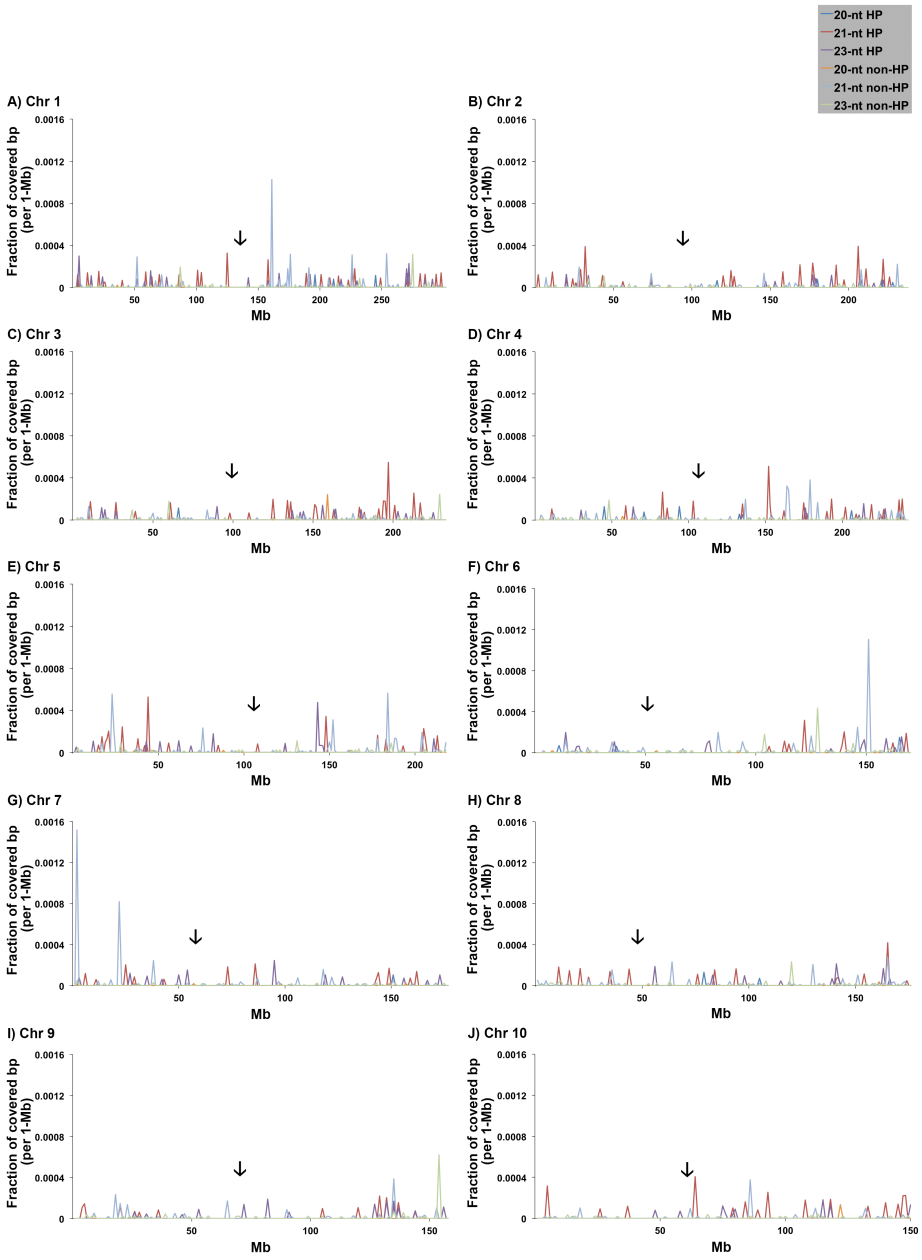

Figure 6 Chromosomal distribution of sRNA loci with size class of 20-nt, 21-nt and 23-nt. For each chromosome, for each of its 1-Mb domain is plotted the fraction of the domain length covered by the sRNA loci. **A) - J)** chromosomes 1 - 10. Centromere positions are indicated by black arrows.

Previous results showed that the sRNA loci of all size classes, except of the 22-nt, covered with the highest densities the chromosomes arms, which are genes-enriched regions (Schnable et al. 2009). Therefore, we decided to study in detail the relative position of sRNA loci and genes. Among the total genes, and their transcripts, annotated in the reconstructed transcriptome assembly, we analyzed the following three sets of loci: protein-coding genes (total number=39252), TE transcripts (total number=33132) and potential lncRNA transcripts (total number=16730), identified as described in the chapter Materials and Methods. The following genomic features were also considered: the exons and introns of protein-coding genes, the immediate flanking regions of three sets of loci and the repeats (the complete set of TEs and repetitive regions). Between the three sets of loci there was redundancy: 234 protein-coding genes had at least one of their spliced transcript classified as TE; 3294 protein-coding genes had at least one of their spliced transcript classified as lncRNA and 2406 TE transcripts were also classified as lncRNAs. Theoretically, the same locus cannot be at the same time coding and non-coding, however a gene could have only one of its spliced transcript classified as lncRNA and the identified lncRNAs were potential lncRNAs, not experimentally verified. For these reasons and because the redundancy between the three sets of loci involved a minor percentage of the total units, we decided to use hereafter the redundant classification of these sets of loci.

To assess the level of enrichment/depletion of each of the sRNA loci categories in each of the above-mentioned genomic features, the analysis of co-occupancy was performed to compare the number of expected and observed non-redundant overlapping nucleotides between them (Figure 7; Appendix E). For each size class, the HP and non-HP sRNA loci showed very similar trends of enrichment/depletion in the different genomic features, indicating that the length of the sRNAs is more influential than their precursor secondary structure in explaining their genomic locations relative to the genomic features studied.

The sRNA loci categories were divided in three main groups depending on their trends of enrichment/depletion. The first group comprised the sRNA loci of 23-nt and 24-nt size class (both HP and non-HP), which were depleted in body regions of TE and lncRNA transcripts and in exons of protein-coding genes but

were enriched in introns of protein-coding genes and in flanking regions of all the analyzed sets of loci. While the sRNA loci of 23-nt size class showed a higher enrichment in downstream regions of genes/transcripts, those of 24-nt size class showed the opposite trend with higher enrichment values in upstream regions of genes/transcripts. In total, 51.8% of the 2-kb upstream regions of protein-coding genes and ~29% of those of TE and lncRNA transcripts exhibited an overlap with at least one non-HP sRNA locus of 24-nt size class, while 9.1% of the 2-kb upstream regions of protein-coding genes and ~5% of those of TE and lncRNA transcripts showed an overlap with at least one HP sRNA locus of 24-nt size class. On the other hand, 19.8% and 23.1% of the HP and non-HP sRNA loci of 24-nt size class, respectively showed overlap with the flanking regions of protein-coding genes and ~9% and ~5% of the total sRNA loci of 24-nt size class exhibited an overlap with the flanking regions of TE and lncRNA transcripts respectively. The percentages of both the overlapping genes/transcripts and sRNA loci were only slightly lower when considering the downstream regions. The second group included the sRNA loci of 21-nt size class (both HP and non-HP), which were depleted in repeats and in TE transcripts but enriched in protein-coding genes, in lncRNA transcripts and in flanking regions of all three sets of loci. The third group comprised the sRNA loci of 20-nt and 22-nt size class (both HP and non-HP), which were enriched in protein-coding genes and lncRNA transcripts but were mainly depleted in their flanking regions.

The genomic features analyzed were also divided in groups depending on the sRNA loci categories they were enriched/depleted in. The first group comprised protein-coding genes, their exons and lncRNA transcripts, indicating that the non-coding feature of lncRNAs had co-occupancy results most similar to the coding feature of protein-coding genes. These genomic features were characterized by a high enrichment in sRNA loci of 20-nt and 21-nt size class. The protein-coding genes also exhibited a low level of enrichment in the other sRNA loci size classes, which was due to a significant overlap of these sRNA loci with the protein-coding genes in their introns and not in their exons, where the sRNA loci were clearly depleted. The second group was made by repeats and TE transcripts, both depleted or only slightly enriched in sRNA loci of all size classes, even though the latter overlapped with these elements with high percentages. The

exception were the sRNA loci with 22-nt size class and the non-HP loci with 20-nt size class, which showed a considerable enrichment (log_2[observed/expected] > 0.85) in TE transcripts. The third group included the flanking regions of TE transcripts and the introns of protein-coding genes, characterized by the enrichment, at variable levels, in sRNA loci of all size classes, with the only exceptions of the non-HP loci of 20-nt size class and partially of the non-HP loci of 23-nt size class. Finally, the forth group comprised the flanking regions of protein-coding genes and lncRNA transcripts, confirming the similarity of these two genomic features in the spatial association with different sRNA loci categories. These regions were clearly enriched in sRNA loci of 21-nt, 23-nt and 24-nt size class and depleted in those of 22-nt size class, while the trend of the 20-nt class varied upon the precursor structure.

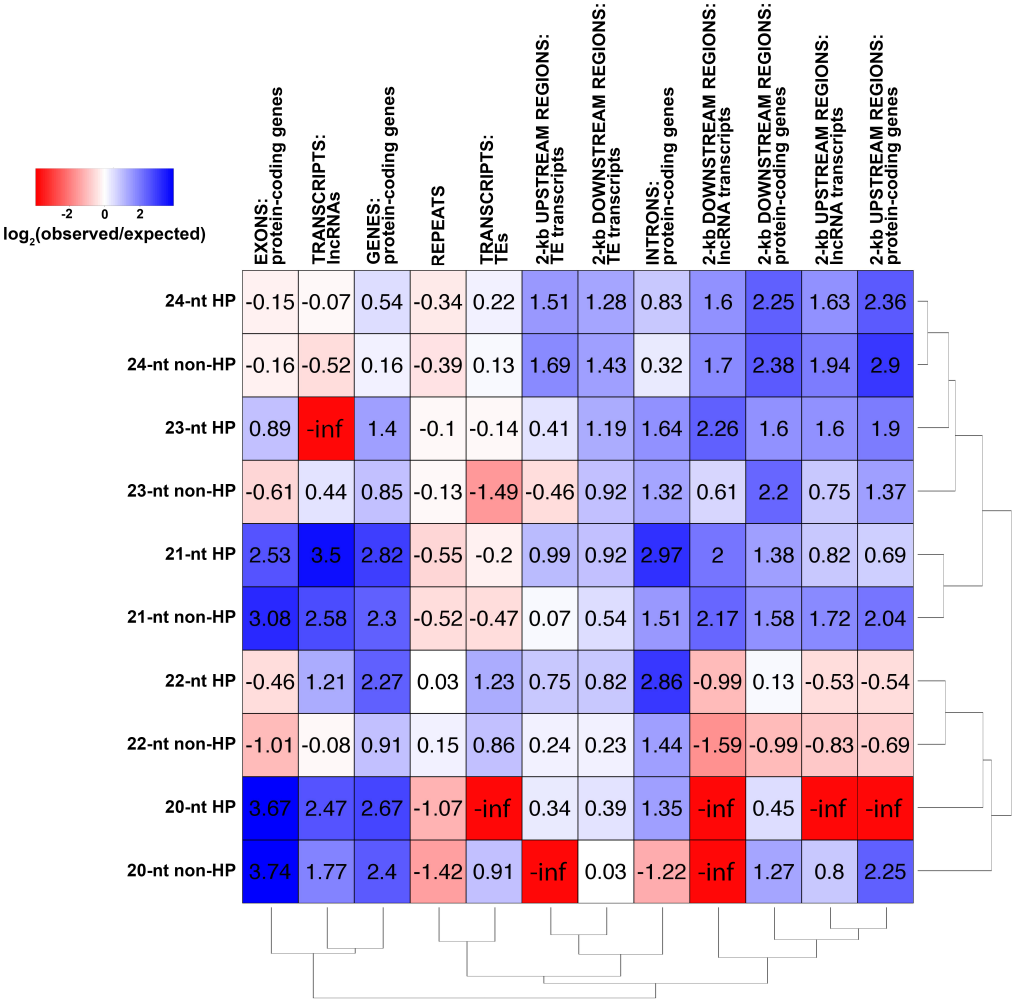

Figure 7 Co-occupancy analysis results. For each sRNA loci category-genomic feature combination the level of enrichment/depletion is reported as the $\log_2$ (observed overlapping nt/expected overlapping nt) between them. R (R Development Core Team 2013) function "heatmap.2" with the default parameters "Rowv=TRUE" and "Colv=TRUE" was used to perform the clustering of both the sRNA loci categories and the genomic features.

The sRNA loci with size class of 22-nt and 24-nt were the most numerous sRNA loci categories, for which it was therefore easier to observe genome-wide trends. Both the genomic and genic distributions evidenced considerable differences between these categories. In summary, the sRNA loci of 22-nt size class: i) had a homogenous distribution across the chromosomes length, ii) the vast majority of them (>93%) mapped to repeats, iii) they were enriched in body regions of TE transcripts but less in their flanking regions, iv) they were enriched in introns of protein-coding genes and v) clearly depleted in the flanking regions of protein-coding genes and lncRNA transcripts. In contrast, the sRNA loci of the 24-nt size class: i) were found preferentially in the chromosome arms where genes are more densely arranged, ii) they were associated with repeats to a lesser extent (~76% of the total loci), iii) they showed little enrichment in body regions of TE transcripts but considerable higher enrichment in their flanking regions and in those of protein-coding genes and lncRNA transcripts, and iv) they exhibited the lowest levels of enrichment observed in introns (together with the non-HP loci of 20-nt size class). These data suggest that the sRNA loci with a size class of 22-nt are more directly correlated to the position of repeats compared with those with a size class of 24-nt. It was previously demonstrated that maize 24-nt sRNAs tend to be concentrated very close to the ends of full-length cDNA genes (Wang et al. 2009), protein-coding genes (Gent et al. 2013), pseudogenes and TEs (Xin et al. 2014). We confirmed this trend providing statistical evidence and extended it for the lncRNA transcripts. Therefore, sRNA loci of 24-nt size class, together with those of 23-nt size class showing similar co-occupancy results, might play a role in the control of gene transcription, in a way independent from the coding or non-coding nature of the gene they are close to.

3.10 Expressed genes are flanked by upstream sRNA loci of 23-nt or 24-nt size class with higher probabilities compared to non-expressed genes

In order to examine the relationship between gene expression and the occupancy of flanking sRNA loci we plotted the distribution of the sRNA loci with size class of 23-nt and 24-nt along these regions, separately for genes with different expression levels (Figure 8). We analyzed these three sets of loci: protein-coding genes, TE and lncRNA transcripts. Only sRNA loci with size class of 23-nt and 24-

nt were included because they were enriched in the flanking regions of these sets of loci and had similar co-occupancy results. We used the gene/transcript expression levels measured through RNA-seq in the wt samples in control condition after ten days of experiment. Protein-coding genes were divided into four groups of equivalent number of elements based on expression level. TE and lncRNA transcripts were divided into five groups because the non-expressed transcripts where more than 25% of the total elements: one group contained only non-expressed transcripts (that accounted for the 72.9% of the TE transcripts and 51.3% of the lncRNA transcripts) and the other four groups consisted of the expressed transcripts divided into four numerically equivalent sets, based on their expression level. The plots (Figure 8) show the fraction of genes having a close sRNA locus, for each gene/transcript expression level and in each position of the 2-kb flanking regions. All the analyzed sets of loci had a higher probability to be flanked by a 23-nt or 24-nt size class sRNA locus in their upstream region than in their downstream region. Protein-coding genes showed the highest values of fraction of genes with flanking sRNA loci, followed by lncRNA and TE transcripts. Protein-coding genes and lncRNA transcripts exhibited the highest peak of sRNA loci occupancy in the interval between 300 and 400 bp upstream of the transcription start site (TSS), while TE transcripts in the interval between 150 and 200 bp upstream of the TSS. A positive correlation between the expression level of genes and the occupancy of upstream sRNA loci was evident for the protein-coding genes. Similar results were obtained by Gent et al. (2013) for the expression level of genes in the filtered gene set (version 5b) and the abundance of their 2-kb flanking 24-nt sRNA sequences. Non-expressed TE and lncRNA transcripts were characterized by a lower probability to have upstream sRNA loci compared to the expressed transcripts but none evident correlation was observed when considering only the transcripts that were expressed at different levels. All the sets of loci did not show any correlation in their downstream regions. A drop of sRNA loci occupancy at the level of the TSS was observed for all the sets of loci: protein-coding genes and TE transcripts showed low but still appreciable values of sRNA loci occupancy at the TSS, while lncRNA transcripts at the same position had zero or close to zero sRNA loci occupancy values.

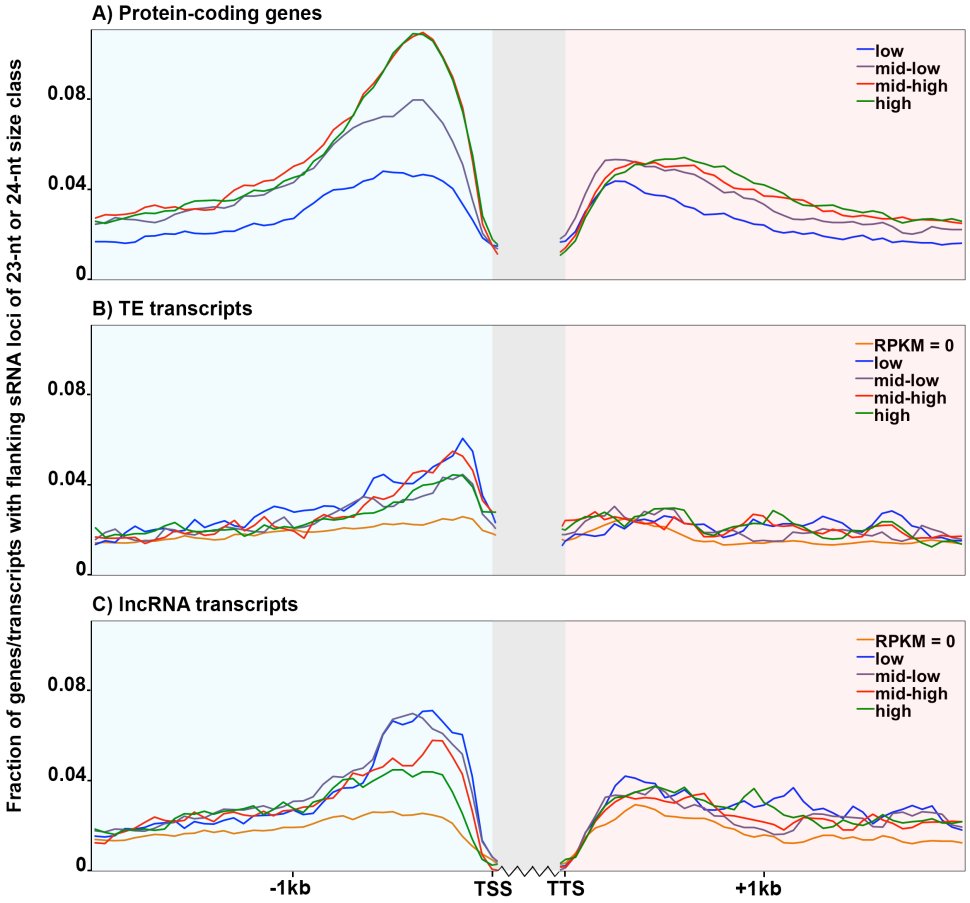

Figure 8 Distribution of sRNA loci with size class of 23-nt and 24-nt in gene flanking regions. The fraction of genes/transcripts overlapping in their 2-kb flanking regions with sRNA loci with size class of 23-nt and 24-nt is reported for each 50-bp interval of the flanking regions, separately for genes/transcripts with different expression levels. **A)** protein-coding genes, total number=39252. They were divided into four equivalent groups from low to high expression. **B)** TE transcripts, total number=33132 **C)** lncRNA transcripts, total number=16730. Both TE and lncRNA transcripts were divided into five groups: one group of non-expressed transcripts, "RPKM=0", and four equivalent groups of expressed transcripts from low to high expression. TSS, transcription start site. TTS, transcription termination site.

In maize, it has been demonstrated that protein-coding genes, pseudogenes and TEs have higher densities of 24-nt sRNAs on their upstream antisense strand than on their upstream sense strand (Xin et al. 2014). We examined the overlaps between sRNA loci with size class of 23-nt and 24-nt and the flanking regions of protein-coding genes, TE and lncRNA transcripts (Table 4). The results obtained did not confirm the trend demonstrated for the 24-nt sRNAs densities: from 44.5% to 49% of cases the overlaps involved sRNA loci with no strand polarity, in the other cases the overlaps involved sRNA loci with strand polarity but we did not observe a significant bias towards the antisense strand compared to the sense strand, both for the upstream and the downstream regions.

Table 4 Overlaps between sRNA loci with size class of 23-nt and 24-nt and gene flanking regions

Gene/transcript flanking regions	Overlaps with sRNA loci of 23-nt and 24-nt size class			
	total overlaps	% overlaps with sRNA loci without strand polarity	% overlaps with sRNA loci with strand polarity: in the same strand	in the opposite strand
protein-coding genes 2-kb upstream regions	30150	49.07	27.25	23.68
protein-coding genes 2-kb downstream regions	24439	46.10	25.75	28.15
TE transcripts 2-kb upstream regions	13648	45.80	23.46	30.74
TE transcripts 2-kb downstream regions	11769	45.81	26.81	27.38
lncRNA transcripts 2-kb upstream regions	7281	46.98	28.07	24.95
lncRNA transcripts 2-kb downstream regions	6686	44.48	27.61	27.91

3.11 Long-term abiotic stresses and plant development affect the expression of a few numbers of sRNA loci

Environmental stresses can influence the expression of different categories of sRNAs. In addition to miRNAs, siRNAs are altered in their expression to consequently modulate target genes as part of the plant response to the stress (Tricker et al. 2012) or to defend the genome from the potentially deleterious effects caused by the stress-induced movements of TEs (Ito et al. 2011). To test the effects of the stress treatments and the different developmental stage of plants (+7 samples) on non-*MIRNA* sRNA loci, we analyzed their counts for differential expression with the tool edgeR, applying the TMM normalization method in addition to the scaling on the total number of reads mapping within the sRNA loci per library ($log_2FC>1$ or <-1, FDR<1%). For each pairwise comparison, the obtained DE sRNA loci with size class from 20-nt to 24-nt, which in all the three

biological replicates were expressed with at least one read in at least one of the two libraries subjected to the comparison were selected (Appendix F). Neither the long-term abiotic stresses applied to the plants nor the different plant developmental stage (+7 samples) caused strong effects on siRNA loci expression: in total 19 loci resulted DE. Hundreds of sRNA sequences were reported to be DE after cold, heat and salinity stresses in *Brachypodium* (Wang et al. 2014b) and thousands of sRNA loci showed differential expression after PEG-simulated drought conditions in foxtail millet (Qi et al. 2013). Compared to these works, we detected a lower number of DE sRNA loci, maybe because we analyzed the effects of the treatments after ten days of stress application, not in the immediate hours after the stress application.

Over the total 19 DE sRNA loci, 12 showed differential expression only in the wt during drought stress and three only in the *rmr6-1* mutant in drought or drought plus salinity stresses; four of these DE sRNA loci, all with 22-nt size class, showed a similar level of differential expression in the non-stressed samples at +7. Four DE sRNA loci showed differential expression only in the wt after the recovery from the drought stress. Salinity stress alone had no significant effects on sRNA loci expression, consistent with the lower effects observed on miRNA expression compared to the other stresses. In general the DE sRNA loci showed no bias towards a particular size class, precursor structure, repeats masking, genomic location or differential expression trend. 11 over 19 DE sRNA loci were DE also in the comparison between *rmr6-1* mutant and wt in control conditions, ten of these showed a similar trend of differential expression (up or downregulation) in the stress or developmental stage comparisons and in the genotype comparison.

The majority of DE sRNA loci were located in genic regions: 12 over the total 19. Two sRNA loci, both upregulated in the wt in drought conditions, overlapped with exons of genes homologous to *Arabidopsis* zinc transporter precursors; four were located in introns of genes homologous to *Arabidopsis* genes with diverse functions; one was found in antisense to a gene encoding a (S)-beta-macrocarpene synthase and the other five sRNA loci overlapped with uncharacterized protein-coding genes or low confidence genes. The expression of these genes and of the genes located in the 10-kb flanking regions of each DE

sRNA locus was retrieved from the RNA-seq data. None of the genes found in the 10-kb flanking regions was DE in the same conditions where the close sRNA locus resulted DE, while in two cases both the sRNA locus and its overlapping gene were upregulated in the same stressed samples and their expression did not significantly decrease after the recovery.

In one of these two cases, the drought stress caused the upregulation, only in the wt, of the sRNA locus Cluster_63380 (log_2FC=1.6) and its overlapping gene GRMZM2G093276, which was one of the two genes homologous *Arabidopsis* zinc transporter precursors (log_2FC=1.26, q-value=0.041). The gene comprises three exons that were all covered by an sRNA locus with size class of 21-nt. The Cluster_63380 was one these loci, the most abundant among them and the only one DE as the gene. It showed a preferential processing to one individual abundant sRNA sequence. The discrepancy in the expression of the three sRNA loci might be due to the wrong mapping position assigned to the abundant sRNA sequence included in the Cluster_63380: the sequence had two possible genome positions, the other one inside the Cluster_63370, which was the other DE sRNA locus overlapping with a gene homologous to an *Arabidopsis* zinc transporter precursor. Alternatively, this might be explained in the hypothesis that the Cluster_63380 was the only one among the three sRNA loci to generate a functionally active sRNA that had a role in the response to drought stress and thus increased its abundance during the stress treatment.

In the other case, drought and drought plus salinity stresses caused the upregulation, in the *rmr6-1* mutant, of the sRNA locus of 24-nt size class Cluster_99151 (log_2FC=2.28 and log_2FC=2.73 respectively) but only in drought conditions the overlapping gene AC216891.3_FG004_X was upregulated (log_2FC=1.92, q-value=0.04). The wt showed upregulation of the gene following drought treatment (log_2FC=1.8, q-value=0.023), but not of the sRNA locus. These data indicate that in drought-stressed samples the upregulation of the sRNA locus observed in the *rmr6-1* mutant did not influence the expression of the overlapping gene, which increased in the two genotypes at a similar extent. The final gene expression level reached in the drought-stressed *rmr6-1* mutant samples was higher (Reads Per Kilobase per Million (RPKM)=16.63) compared to that reached in the stressed wt samples (RPKM=4.11), because its basal expression in control

condition was significantly higher in the *rmr6-1* mutant compared to the wt (log_2FC=2.24, q-value 0.004). The Cluster_99151 showed lower abundances in the *rmr6-1* mutant compared to the wt in control conditions but the difference was not significant. The AC216891.3_FG004_X gene was previously unannotated and newly detected in our samples, overlapping on the opposite strand with the AC216891.3_FG004 gene, which is not characterized and does not have homologs in Arabidopsis or rice. The sequence of the newly detected AC216891.3_FG004_X gene was analyzed and resulted to be part of a LTR retrotransposon. The LTR TE appears to be a TE-relic because it does not possess complete LTR regions and all the protein domains required for its retrotransposition. We visualized this locus in the MaizeGDB Genome Browser (Sen et al. 2010) and found that the LTR regions of this TE-relic are methylated in B73 according to Eichten et al. (2011). Taken together these observations suggest that the gene might be a TE-relic targeted by 24-nt sRNAs that are at least in part independent of the Pol IV activity, that do not function to repress the TE-relic expression and that increase in abundance following the drought stress only in the absence of a functional Pol IV enzyme.

Our data suggest that the plant response to the stress treatments applied for ten days did not involve the action of sRNAs (all sRNA categories except miRNAs) as a general strategy to modulate gene expression; indeed the *rmr6-1* mutants did not show more severe phenotypes in stress conditions compared to the wt.

3.12 The majority of sRNA loci located in gene flanking regions are of Pol IV-dependent

RMR6 gene encodes for the largest subunit of RNA Pol IV and its loss of function allele, *rmr6-1*, is responsible for the reduction of the vast majority of ~24-nt RNA species (Erhard et al. 2009). To confirm this result the abundances of the sRNA loci were compared between the wt and *rmr6-1* samples collected after ten days of experiment in control growing conditions, with the tool edgeR. To normalize the expression values the TMM method was used in addition to the scaling on the total number of reads mapping within the sRNA loci per library. Since TMM assumes that most elements are not differentially expressed between samples we

included in the analysis all the identified sRNA loci, also those with size class <20-nt or >24-nt, which accounted for the majority of mapping reads and are not supposed to change in the mutant. Among the sRNA loci with size class in the range from 20-nt to 24-nt, a total of 147,492 (51.5%) were DE (log_2FC>1 or <-1, FDR<1%) between the two genotypes (Table 5). To assess the influence of the normalization method used on the number of DE sRNA loci, the same comparison was performed without the TMM normalization and only 0.83% more DE sRNA loci with size class in the range from 20-nt to 24-nt were obtained: given the small difference, to retrieve stringent results the comparison with the TMM normalization was finally considered. Among the sRNA loci with size class <20-nt or >24-nt, a total of 6.9% were DE, confirming that these loci did not undergo big changes in the mutant. Results for the sRNA loci with size class in the range from 20-nt to 24-nt are summarized in Table 5. Only 0.3% of the total sRNA loci were upregulated in the *rmr6-1* mutant, while 51.2% were downregulated, consisting primarily of loci with size class of 24-nt. While the DE sRNA loci with size class of 23-nt and 24-nt were for the most part downregulated in the *rmr6-1* mutant, those with size class of 20-nt to 22-nt were in most cases upregulated, both HP and non-HP loci. sRNA loci with size class from 20-nt to 24-nt downregulated in the mutant were therefore dependent on Pol IV for their biogenesis, so they were called siRNA loci because they showed evidence for participation in the RdDM pathway.

Table 5 DE sRNA loci between *rmr6-1* mutant and wt

sRNA loci[a]	total	% DE[b]	% upregulated[b]	% downregulated[b]
all	147492	51.48	0.28	51.20
20-nt HP	26	11.54	11.54	0.00
21-nt HP	215	6.05	5.58	0.47
22-nt HP	1181	5.76	5.00	0.76
23-nt HP	121	26.45	4.96	21.49
24-nt HP	17679	48.49	0.06	48.43
20-nt non-HP	32	3.13	3.13	0.00
21-nt non-HP	520	8.27	6.92	1.35
22-nt non-HP	17678	1.98	1.52	0.46
23-nt non-HP	311	18.97	2.89	16.08
24-nt non-HP	109729	60.87	0.01	60.86

[a]sRNA loci with size class in the range from 20-nt to 24-nt, except *MIRNA* loci.

[b]log_2FC>1 or <-1, FDR<1%.

In the wt, the majority of the Pol IV-dependent siRNA loci were expressed between 1 and 10 RPM (average of the three biological replicate values, calculated on the total number of reads aligned to the genome for each library), while in the *rmr6-1* mutant the 90.9% of them were not expressed (Figure 9). In the wt, the total Pol IV-dependent siRNA loci accounted for the 75% of the total number of reads mapping within sRNA loci with size class in the range from 20-nt to 24-nt and the subset of Pol IV-dependent siRNA loci with 0 RPM in the mutant accounted for 62.3% of these reads. Together these results confirmed the previously described dramatic loss of 23-nt and 24-nt sRNAs occurring in the *rmr6-1* mutant (Erhard et al. 2009).

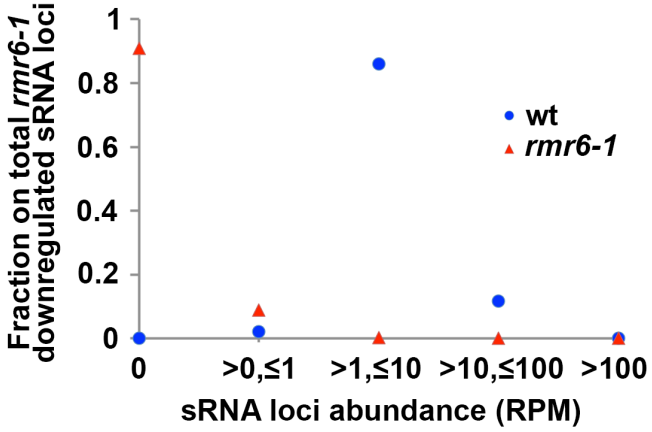

Figure 9 Abundance distribution of Pol IV-dependent sRNA loci. The abundance of the sRNA loci with size class between 20-nt and 24-nt downregulated in *rmr6-1* mutant compared to wt is reported as average of the three biological replicates. RPM are calculated on the total number of reads aligned to the genome for each library.

Considering all genes annotated in the transcriptome assembly used, over the total number of Pol IV-dependent siRNA loci, the 65.1% overlapped with their 2-kb flanking regions, of those the 86.2% overlapped with the 2-kb flanking regions of protein-coding genes, TE and lncRNA transcripts. These results are consistent with the previous finding that RdDM loci, defined in the mutant *mop1-1* by their loss of 24-nt sRNAs, are primarily limited to regions close to genes, that are accessible chromatin environments where siRNAs are thought to repress transposons (Gent et al. 2014). A substantial proportion of Pol IV-dependent

siRNA loci were not located in the immediate proximity of genes, suggesting that RdDM control of transcriptional silencing might not be restricted to these regions.

In the co-occupancy analysis, the partial redundancy between the analysed genomic features and the fact that the same sRNA locus could theoretically overlap with different genomic features and vice versa, made impossible to sum the fraction values to obtain general results for a single sRNA loci category or genomic feature. Therefore, the total non-redundant number of sRNA loci with size class from 20-nt to 24-nt overlapping with gene flanking regions was calculated. Considering all genes annotated in the transcriptome assembly used, over the total number of sRNA loci with size class from 20-nt to 24-nt, the 54.3% overlapped with their 2-kb up or downstream regions, of those the 61.3% were Pol IV-dependent. Considering only protein-coding genes and TE and lncRNA transcripts, over the total number of sRNA loci with size class from 20-nt to 24-nt, the 46.4% overlapped with their 2-kb up or downstream regions, of those the 61.9% were Pol IV-dependent. We then analyzed separately: i) the protein-coding genes from the TE transcripts and from the lncRNA transcripts, ii) their different expression levels and iii) the upstream regions from the downstream regions (Table 6). In each case, over the total overlapping sRNA loci with size class from 20-nt to 24-nt in gene/transcript flanking regions, the number of upregulated loci was never >1% and the number of downregulated loci was pretty constant around ~61%. For protein-coding genes and lncRNA transcripts the percentage of downregulated loci was lower for the non-expressed or lowly expressed units compared to the expressed units, both for up and downstream regions. These results indicate that over the total loci located in the immediate gene flanking regions, the Pol IV-dependent siRNA loci are always the majority, with similar proportions, around 61%, when considering different coding and non-coding units, with different expression levels, in their up and downstream regions.

Table 6 Total and DE sRNA loci overlapping with gene/transcript flanking regions

Genes/transcripts	Expression level	sRNA loci[a] overlapping with 2-kb upstream regions			sRNA loci[a] overlapping with 2-kb downstream regions		
		total	% upregulated	%downregulated	total	% upregulated	%downregulated
protein-coding genes	low	5312	0.09	62.22	4857	0.12	60.68
	mid-low	7363	0.14	66.33	6131	0.10	62.47
	mid-high	8778	0.01	68.43	6594	0.00	62.22
	high	8660	0.02	68.71	6718	0.06	61.04
TE transcripts	RPKM = 0	8960	0.15	59.19	8026	0.30	57.53
	low	1072	0.09	62.50	808	0.87	57.55
	mid-low	1057	0.66	61.49	901	0.22	55.60
	mid-high	1030	0.49	58.74	897	0.56	55.96
	high	915	0.33	57.16	855	0.23	57.89
lncRNA transcripts	RPKM = 0	3063	0.13	61.51	2888	0.14	60.39
	low	970	0.00	66.39	870	0.00	64.60
	mid-low	1051	0.00	66.13	904	0.33	61.17
	mid-high	967	0.10	66.80	858	0.12	60.49
	high	969	0.21	65.53	948	0.00	63.29

[a]sRNA loci with size class in the range from 20-nt to 24-nt, except *MIRNA* loci and miRBase annotated *MIRNA* loci not confirmed by our analysis.

About 40% of the total sRNA loci in gene flanking regions were not significantly differentially expressed in *rmr6-1* mutant. The 96% of them decreased in expression in *rmr6-1* compared to wt but not at significant levels and at a lower extent (average log_2FC=-3) compared to Pol IV-dependent loci (average log_2FC=-5.8). In the wt these loci had already on average four-fold lower expression levels compared to the Pol IV-dependent loci. However, in the mutant the 90% of them were not expressed at all, similarly to the Pol IV-dependent loci. These data suggest that these loci were either not targeted by RdDM or were undergoing RdDM at a much lower extent compared to the Pol IV-dependent loci siRNA loci.

3.13 Pol IV mutation induces gene expression changes in leaves of *rmr6-1* mutants without altering their morphology

To evaluate the effects of siRNAs loss and subsequent RdDM impairment on gene expression, gene counts of wt and *rmr6-1* mutant samples were subjected to pairwise differential expression analysis with the tool Cuffdiff (log_2FC>1 or <-1, FDR<5%). As performed for sRNA loci differential expression analysis, only samples collected after ten days of experiment in control conditions were used to evaluate the effects of the *rmr6-1* mutation on genes. Despite the global loss of siRNAs occurring in the mutant and their enriched position in gene flanking regions, Pol IV-mutation did not cause dramatic changes in leaf gene expression: a total of 1013 genes were DE between wt and *rmr6-1* (Appendix G). Results

were similar to the number of DE genes detected in *mop1-1* mutant immature ears (~762 DE genes; Madzima et al. 2014), while in *mop1-1* SAMs the number of DE genes detected was higher (~6000 DE genes; Jia et al. 2009).

At the time point of sample collection, *rmr6-1* mutant plants did not show any gross morphological defects. However, *rmr6-1* plants were delayed in development compared to wt and we noted that when grown in field they showed even more delay in development and flowering time and reduced fertility. The developmental delay was also observed in heterozygous *RMR6/rmr6-1* plants, but attenuated. Since the *rmr6-1* mutation was not induced in B73 but was introgressed in this background, these results might be due to the not 100% identical genetic background between wt and mutant plants. At the time point of sample collection, leaves were identical between mutant and wt plants. Altered abaxial leaf fates were described in *rmr6-1* mutants but this phenotype was observed only after one or two generations of homozygous sibling crosses (Parkinson et al. 2007), while we employed homozygous plants derived by the crossing of heterozygous individuals. Reduced fertility was only observed in homozygous *rmr6-1/rmr6-1* plants. Both *rmr6-1* and *mop1-1* mutants have been described to be most strongly affected in the development of floral organs and reproduction processes, although also these phenotypes are not fully penetrant or are visible only after one or two generations of homozygous sibling crosses (Dorweiler et al. 2000, Parkinson et al. 2007).

These data indicate that the effects of loss of siRNAs on genome homeostasis primarily negatively affect reproductive organ development and that phenotypes on vegetative organs become more severe after generations of homozygous crossing. Indeed we did not observe morphological defects on leaves, despite the observed gene expression changes.

3.14 Gene expression changes induced by the loss of siRNAs are not predictable upon the relative position of siRNAs and genes

Differential expression was performed on genes, not on transcripts, therefore we decided to classify a gene as TE and/or lncRNA when it had at least one of its spliced transcripts classified as TE and/or lncRNA. Therefore, there was redundancy between the following categories: one protein-coding gene was also

classified as TE, 31 protein-coding genes were also classified as lncRNAs and 14 TEs were also classified as lncRNAs (Appendix G). Theoretically, the same locus cannot be at the same time coding and non-coding but because a gene could have only one of its spliced transcript classified as lncRNA and because the identified potential lncRNAs were not verified experimentally, we decided to keep both the coding and non-coding classification for the 31 redundant genes. This redundancy involved a few number of loci, therefore it could not have significant impact on the following analysis. Genes that could not be classified neither as protein-coding genes, nor as TEs and lncRNAs were named 'genes not classified'.

Over the total 1013 DE genes, 777 were upregulated and 236 were downregulated in *rmr6-1* compared to wt (Table 7). About half of the total DE genes were protein-coding genes, about one third were genes that could not be assigned to one of the classification used, 154 were lncRNAs and 191 were TEs. As expected based on the mechanism associated with the RdDM silencing pathway, for each category and especially for TEs, lncRNAs and unclassified genes, the number of upregulated genes was greater than the number of downregulated genes. Among the 167 upregulated TEs, 109 belonged to the class I TE order of LTR, mainly of *Copia* and *Gypsy* superfamilies; 13 upregulated TEs belonged to the class II TE order of TIR; the other 45 upregulated TEs lacked a specific classification (Appendix G).

Over the total DE genes, 489 were expressed only in *rmr6-1* mutant, of these 142 were TEs and 116 were lncRNAs, and 45 were expressed only in wt, of these 11 were TEs and 14 were lncRNAs. Therefore, the majority of DE TEs and lncRNAs were expressed specifically in only one genotype and most of them were expressed only in the mutant. Over the total DE genes, 248 were not previously annotated, of these 206 were expressed only in *rmr6-1*, about half were TEs, indicating that the loss of siRNAs caused the activation of TEs and other genes that were not previously detected. Among the genes that were not classified, 196 were expressed only in the mutant and of these 81 were new loci (Appendix G).

Table 7 DE genes between *rmr6-1* mutant and wt

Genes[a]	total DE[b]	upregulated[b]	% upregulated[b]	downregulated[b]	% downregulated[b]
all	1013	777	76.70	236	23.30
protein-coding genes	418	267	63.88	151	36.12
TEs	191	167	87.43	24	12.57
lncRNAs	154	132	85.71	22	14.29
genes not classified	296	249	84.12	47	15.88

[a]the classification of genes into protein-coding, TE and lncRNA was redundant for a total of 46 genes.

[b]$\log_2$FC>1 or <-1, FDR<5%.

In *rmr6-1* mutant compared to wt, thousands of sRNA loci were DE, while 1013 genes were DE, indicating that the loss of siRNAs is not itself sufficient to predict the differential expression of a close gene. In order to verify if the presence of a DE sRNA locus were a necessary requisite for the differential expression of a close gene or was increasing significantly its probability to be DE, we calculated the fraction of DE genes overlapping with DE and non-DE sRNA loci in their gene body or flanking regions (Figure 10). The presence of only one DE gene associated with an upregulated sRNA locus, found across its gene body and downstream region, was likely due to the much smaller number of upregulated sRNA loci detected in the mutant compared to the number of downregulated loci. For each of the analysed subset of DE genes, up and downregulated protein-coding genes, TEs and lncRNAs, the fraction of genes with inner or flanking DE sRNA loci was never >50%, with very similar behaviour between up and downregulated genes of the same category. These results suggest that the downregulation of an sRNA locus is not generally sufficient not even necessary to predict the up or downregulation of its close gene. However, we cannot exclude that in single cases the differential expression of a gene required the DE expression of its inner or proximal sRNA loci.

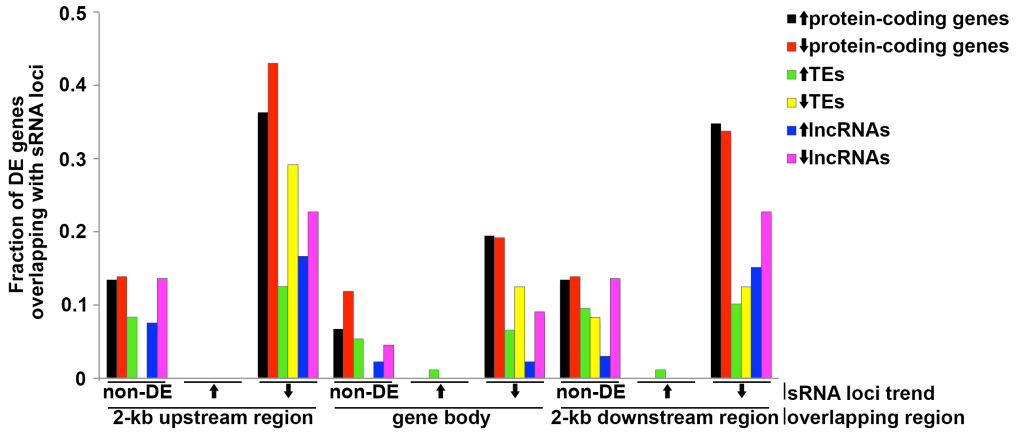

Figure 10 Frequencies of DE genes overlapping with DE and non-DE sRNA loci. For each category of up and downregulated protein-coding genes, TEs and lncRNAs in *rmr6-1* compared to wt, the fraction of genes overlapping with at least one sRNA locus in their gene body or 2-kb flanking regions are reported. DE and non-DE sRNA loci refer to the same genotype comparison: *rmr6-1* versus wt. ↑=upregulated gene or sRNA locus. ↓=downregulated gene or sRNA locus.

3.15 Gene expression changes occurring in *rmr6-1* are indicative of a secondary response directed by the mutant against its loss of Pol IV-dependent siRNAs and RdDM impairment

To understand how *rmr6-1* plants responded to the absence of siRNAs and consequent RdDM impairment in term of modulation of gene expression in the leaf, the GO enrichment analysis was performed on the DE genes in *rmr6-1* compared to wt (Table 8). All the GO categories were analysed: molecular function, biological process and cellular component. Among the total 777 upregulated genes, 251 were assigned to at least one GO term and 152 genes resulted having enriched GO terms. Among the total 236 downregulated genes, 132 were assigned to at least one GO term and 59 genes resulted having enriched GO terms. Best *Arabidopsis* and rice BLASTP hits of translated genes and correspondent transcripts in the Chromatin Database were also analyzed (Appendix G).

Among the upregulated genes, 16 were assigned to the enriched GO term of heme binding (and its parent term tetrapyrrole binding) and a subset of them

also to the terms monooxygenase activity and iron ion binding. Among these genes, 12 had homology with *Arabidopsis* and rice cytochrome P450 proteins, two with peroxidases and one with oxygenases proteins. Cytochromes P450 are heme-dependent oxidase enzymes that generally catalyze the insertion of one oxygen atom in a substrate after activation of molecular oxygen. They catalyze different kind of reactions and are involved in many different metabolic processes: the synthesis of secondary metabolites, the biosynthesis and catabolism of phytohormones and the synthesis of many compounds which are essential for the normal growth and development of plant cells, like sterol and xanthophylls, or that are important structural components, UV protectants, antioxidants or antimicrobials (Mizutani M 2012).

All upregulated genes with enriched GO terms were assigned to the GO term of catalytic activity, which is too generic and thus not informative. To examine more in detail the upregulated genes, we looked at their highest scored GO terms (score>10) of the biological process category. In fact even if they were not enriched they were useful to understand what were the mostly affected processes in the mutant. The most scored GO terms were: oxidation-reduction process (38 genes), response to cadmium ion (14 genes), response to oxidative stress (20 genes), proteolysis (19 genes), response to cold (12 genes) and response to salt stress (13 genes). Other GO terms related to response to abiotic stresses had lower scores but equally represented in terms of number of genes. Examples of upregulated genes encoding proteins typically involved in stress responses were: heat shock proteins, glutathione S-transferases and chaperonins. The maize gene encoding cystatin2, a cysteine proteinase inhibitor, was upregulated: in *Arabidopsis* the overexpression of two cystatin proteins have been demonstrated to increase tolerance to salt, drought, oxidation and cold stresses (Zhang X et al. 2008). Two genes encoding homologs of *Arabidopsis* HVA22 proteins were also upregulated: in *Arabidopsis* the levels of *HVA22* mRNA has been observed to increase following cold, salt, dehydration stresses and ABA treatment (Chen et al. 2002).

Among the upregulated genes, another GO term had a high score: the regulation of transcription, DNA-templated (21 genes). Indeed, many upregulated genes were found to encode putative o characterized TFs, belonging to the

following families: WRKY, GRAS, MYB, MADS-box and homeobox. The most numerous family was the homeobox, which plays a variety of important roles in plant development: three genes were predicted and three genes were annotated as encoding homeobox TFs. Two genes encoded MADS-box TFs, one of them was homologous to *Arabidopsis* AGAMOUS-like 8. MADS-box genes regulate reproductive organ identity during floral development and function through interactions with chromatin-associated proteins and other transcriptional regulators (Ng et al. 2009). The upregulation of TF encoding genes has also been observed in *mop1-1* mutant (Madzima et al. 2014). As suggested for *mop1-1*, the loss of silencing observed at TFs might trigger secondary responses in *rmr6-1*.

Downregulated genes in the mutant were enriched in GO terms related to the regulation of development. Among them, many genes were specifically related to the regulation of cell cycle. For example, one gene had homology to *Arabidopsis* and rice cyclin-dependent kinases, involved in regulation of the G2/M transition of the mitotic cell cycle, and another had homology with cyclin-dependent kinase inhibitors. Two genes were homologous to RPA proteins that in *Arabidopsis* have been suggested to play a role in DNA damage repair (Aklilu et al. 2014). Another gene had homology with the *Arabidopsis* gene encoding the RETINOBLASTOMA-RELATED protein, a cell cycle regulator that controls cell proliferation, differentiation, and regulation of a subset of Polycomb Repressive Complex 2 genes and MET1 in the male and female gametophytes, as well as cell fate establishment in the male ad female gametophytes (Johnston et al. 2010). One gene had homology with an *Arabidopsis* gene encoding a MINICHROMOSOME MAINTENANCE (MCM) protein. MCMs form heterohexameric complex to serve as licensing factor for DNA replication to make sure that genomic DNA is replicated completely and accurately once during S phase in a single cell cycle (Tuteja et al. 2011). The downregulation of this gene has been observed also in *mop1-1* (Madzima et al. 2014)

Among downregulated genes involved in the regulation of development, a subset was enriched in GO terms related to the chromatin organization. For some of them their transcripts had a correspondent in the Chromatin Database of chromatin-associated genes. Two genes encoded the core histone 3 (H3) (ChromDB names: HTR103, HTR106), three genes encoded the core histone 2B

(H2B) (ChromDB names: HTB104, HTB105, HTB107) and one gene encoded the component subunit D of the condensin complex (ChromDB name: CPD101), homolog of the *Schizosaccharomyces pombe* Cnd1 protein, essential for mitotic condensation (Sutani et al. 1999). The downregulation of chromatin-associated genes has been observed also in *mop1-1* mutant SAMs: numerous chromatin-related genes have been detected to change expression in *mop1-1* and the vast majority of them were downregulated (Jia et al. 2009). In contrast to these results obtained in *mop1-1*, we did not observe significant changes in the expression of RdDM-related genes.

The homolog of *Arabidopsis REPRESSOR OF SILENCING 1* (*ROS1*)/ *DEMETER-LIKE1* (*DML1*), encoding a DNA glycosylase protein that actively demethylates DNA (Morales-Ruiz et al. 2006, Agius et al. 2006), was significantly downregulated in *rmr6-1* compared to wt. Its downregulation has also been observed in *mop1-1* (Jia et al. 2009, Madzima et al. 2014) and in several *Arabidopsis* RdDM mutants, such as *pol IVa*, *IVb*, *rdr2,* and *drd1* (Huettel et al. 2006, Li et al. 2012). DNA demethylation is thought to regulate epigenome dynamics in opposition to the RdDM pathway. The downregulation of DNA demethylation activity might be a strategy adopted by the cell to counteract the effects provoked by the perturbation of RdDM gene silencing control.

Together these results suggest that many DE genes might be part of an orchestrated network activated in the mutant to respond to its dramatic loss of siRNAs and consequent RdDM impairment. This would mean that many DE genes were not direct targets of the mutation but instead secondary targets triggered to buffer the effects provoked by the misregulation of the gene silencing mechanism controlled by RdDM. The upregulation of stress-related genes might indicate that mutant plants were sensing as a stress condition the alteration of cell homeostasis likely provoked by the RdDM impairment. An increased activity of cytochromes might be important to ensure the proper development of plants. The downregulation of genes involved in the regulation of cell cycle suggests that mutant plants experienced a misregulation of cell proliferation mechanisms. Decreased amount of core histone proteins may cause a slowdown of the cell cycle and increased accessibility of chromatin. Indeed, *mop1-1* mutation has been demonstrated to alter the chromatin accessibility by increasing it at the

chromosome arms and decreasing it at pericentromeric regions (Madzima et al. 2014). The activation of TFs also suggests the involvement of a secondary response occurring in the mutant. Finally, the downregulation of the homolog of *ROS1*/*DML1* strongly suggests that mutant cells respond to the alteration of the RdDM epigenetic pathway modulating the activity of another distinct epigenetic mechanism of gene expression control.

Table 8 Enriched GO terms among DE genes

Enriched GO term				DE genes			All genes	
ID	name	category[a]	FDR[b]	*rmr6-1*/wt	total	% on DE genes[c]	total	% on all genes[c]
GO:0003824	catalytic activity	F	1.69E-02	upregulated	152	60.56	16111	45.82
GO:0020037	heme binding	F	3.48E-02	upregulated	16	6.37	594	1.69
GO:0046906	tetrapyrrole binding	F	4.17E-02	upregulated	16	6.37	629	1.79
GO:0004497	monooxygenase activity	F	4.17E-02	upregulated	12	4.78	370	1.05
GO:0005506	iron ion binding	F	4.37E-02	upregulated	15	5.98	582	1.66
GO:0006996	organelle organization	P	1.63E-02	downregulated	35	26.52	4086	11.58
GO:0050793	regulation of developmental process	P	2.14E-02	downregulated	19	14.39	1606	4.55
GO:0044772	mitotic cell cycle phase transition	P	2.14E-02	downregulated	6	4.55	126	0.36
GO:0044770	cell cycle phase transition	P	2.14E-02	downregulated	6	4.55	126	0.36
GO:0048833	specification of floral organ number	P	2.37E-02	downregulated	3	2.27	11	0.03
GO:0032993	protein-DNA complex	C	2.37E-02	downregulated	6	4.55	143	0.41
GO:1902749	regulation of cell cycle G2/M phase transition	P	2.37E-02	downregulated	5	3.79	94	0.27
GO:0010389	regulation of G2/M transition of mitotic cell cycle	P	2.37E-02	downregulated	5	3.79	94	0.27
GO:0000086	G2/M transition of mitotic cell cycle	P	2.37E-02	downregulated	5	3.79	95	0.27
GO:0044839	cell cycle G2/M phase transition	P	2.37E-02	downregulated	5	3.79	95	0.27
GO:0048832	specification of organ number	P	2.37E-02	downregulated	3	2.27	15	0.04
GO:1902589	single-organism organelle organization	P	2.37E-02	downregulated	22	16.67	2274	6.45
GO:0006260	DNA replication	P	2.37E-02	downregulated	11	8.33	652	1.85
GO:0006325	chromatin organization	P	2.37E-02	downregulated	14	10.61	1040	2.95
GO:0007010	cytoskeleton organization	P	2.66E-02	downregulated	14	10.61	1058	3.00
GO:0006259	DNA metabolic process	P	2.86E-02	downregulated	21	15.91	2158	6.12
GO:0006275	regulation of DNA replication	P	4.00E-02	downregulated	6	4.55	185	0.52
GO:0006323	DNA packaging	P	4.01E-02	downregulated	5	3.79	116	0.33
GO:0048731	system development	P	4.01E-02	downregulated	35	26.52	4900	13.89
GO:1901990	regulation of mitotic cell cycle phase transition	P	4.01E-02	downregulated	5	3.79	119	0.34
GO:1901987	regulation of cell cycle phase transition	P	4.01E-02	downregulated	5	3.79	119	0.34
GO:0016043	cellular component organization	P	4.18E-02	downregulated	41	31.06	6194	17.56
GO:0051567	histone H3-K9 methylation	P	4.38E-02	downregulated	7	5.30	288	0.82
GO:0005662	DNA replication factor A complex	C	4.57E-02	downregulated	2	1.52	3	0.01
GO:0006261	DNA-dependent DNA replication	P	4.99E-02	downregulated	7	5.30	300	0.85
GO:0008283	cell proliferation	P	4.99E-02	downregulated	8	6.06	403	1.14

[a]F=Molecular Function. P=Biological process. C= Cellular component.

[b]FDR<5%.

[c]percentages are calculated on the total DE genes with an assigned GO term and on the total annotated genes in the assembly with an assigned GO term.

4 Discussion

The main goals of this PhD project were: i) the identification and characterization of the maize genome loci responsible for a significant production of sRNAs in the leaf, ii) the analysis of the sRNA effects on gene expression and genome stability through the study of the *rmr6-1* mutant line lacking the Pol IV-dependent small interfering RNAs (siRNAs) and iii) the evaluation of the sRNAs response to field-mimicked stress conditions. To achieve these goals the Next Generation Sequencing technique was employed to collect data, sRNA-seq and RNA-seq. Many studies have been published where NGS of plant biological samples was used to obtain a picture of gene and small RNA expression of wild type and mutant plants impaired in the RNA-directed DNA methylation (RdDM) pathway of gene expression regulation, but very few were conducted in maize (Nobuta et al. 2008, Jia et al. 2009). So far, no sRNA-seq or RNA-seq experiment has been published for the maize mutant of Pol IV, *rmr6-1*. Moreover, while several NGS studies have been performed to study the effects of abiotic stresses on maize miRNAs, there are currently no NGS studies investigating these effects on the maize siRNAs.

4.1 Small RNA sequencing: data processing and identification of sRNA loci

For each of the sequenced 48 leaf samples, the profile of length distribution and abundance of the reads aligned to the maize genome was analysed. Wild type leaf samples showed a predominance of 24-nt sRNAs, followed by the 22-nt sRNAs. This observation confirmed previous data obtained in immature ears and tassels (Nobuta et al. 2008) indicating that maize is an exception even within monocots: it possesses a more abundant group of 22-nt sRNAs compared to the group of 21-nt sRNAs. High abundances of reads out of the range from 20-nt to 24-nt were obtained, in particular at 17-nt, 30-nt and from 32-nt to 42-nt. The analysis of the most abundant reads with these sizes revealed they were mainly

part of tRNAs, rRNAs, signal recognition particles (SRPs) and Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) subunit mRNAs; in particular a sequence of 30-nt homologous to rRNAs, was conserved in all the analyzed libraries where it made up to 1% of the total reads. These data indicated that at least part of the reads with size class out of the range from 20-nt to 24-nt were not likely derived by the processing of DCL proteins and might be degradome, therefore, they were not further analyzed. Mutant leaf samples showed the expected dramatic loss of 24-nt sRNAs (Erhard et al. 2009) and a slight increase in abundance of all other sRNA sizes compared to wt, which was expected as a result of the loss of siRNAs (Nobuta et al. 2008). The sRNA loci differential expression analysis between *rmr6-1* and wt revealed that a significant upregulation affected only 1% of the total identified loci and 0.3% of the total loci with size class from 20-nt to 24-nt, indicating that the upregulation of sRNA loci was not a widespread effect in the mutant and interested a limited number of loci. The consistent profile of read length distribution and abundance between samples of the same genotype and with previous published works confirmed the quality of the sequenced data that was a necessary pre-requisite to verify for proceeding with the analyses.

The first aim of this work was the annotation of the sRNA loci expressed in the maize leaf based upon the sRNA-seq data. Clustering the sRNA reads into sRNA loci was preferable to analyse the individual sequences, because *MIRNA* loci usually give rise to one or few functionally sequences but siRNA loci can show heterogeneous processing in different samples, resulting in distinct sequences originating from the same functional locus, especially in the case of low abundant loci. Considering also that less is known about siRNA sequence requirements for target selection, the approach to identify and study the sRNA loci was preferred. Starting from a set of 48 sRNA-seq libraries, it was necessary to find an appropriate method to work with this large amount of data that was both accurate and computationally achievable. The tool segmentSeq was initially tested, which has the advantage of taking into account replicate data in its process of *de novo* identification of sRNA loci (Hardcastle et al. 2012): it was computationally too intensive and we were not able to apply it to our complete set of data. The tool ShortStack is characterized by good performances in terms of sensitivity and

specificity of *de novo* sRNA loci identification, providing detailed description of the found loci (Axtell MJ 2013b): it was successfully applied to our data and allowed to identify a total of 188,938 clusters in the time of few days. This software has now been updated including a tool producing small RNA-seq alignments where multimapped sRNAs tend to be placed near regions of confidently high density (Axtell MJ 2014), increasing the balancing between precision and sensitivity. The alignment of multimapping reads is a critical step in the identification of genomic loci effectively producing sRNAs, especially in maize due to its highly repetitive genome: in our further analysis of sRNA-seq data we will test this updated ShortStack version with the aim to improve the sRNA loci annotation step.

As expected from the profiles of read length distribution and abundance, the most numerous categories of sRNA obtained with the size class in the range of interest (from 20-nt to 24-nt) were the hairpin (HP) and non-hairpin (non-HP) loci with size class of 24-nt and 22-nt. A fraction of sRNA loci was predicted to have an HP secondary structure but did not meet the criteria for *MIRNAs*: these loci comprised the 10% of the total identified loci and the 13% of the total loci with size class between 20-nt and 24-nt. The analysis of the maximum delta G/stem length (ΔG/sl) values suggested that a fraction of the predicted HP loci might be false positives. It was observed that HP and non-HP loci with the same size class (between 20-nt and 24-nt) showed very similar enrichment/depletion patterns relative to the investigated genomic features in the co-occupancy analysis. We calculated the fraction of each of these sRNA category masked by each of the transposable element (TE) superfamilies annotated in RefGen ZmB73 RepeatMasked Assembly AGPv3 (data not shown in the 'Results' section) and found no differences among HP and non-HP with the same size class. Moreover, HP and non-HP loci with the same size class were similarly affected in the mutant: the 22-nt HP and non-HP loci were the most numerous upregulated categories and the 24-nt HP and non-HP loci were the most numerous downregulated categories. In summary, HP and non-HP loci with the same size class were not differentiated in terms of co-occupancy with genes, TE superfamily association and expression trend in the mutant, confirming the hypothesis that a fraction of the predicted HP loci might not be real hairpins. We will further investigate this hypothesis examining publically available *mop1-1* sRNA data sets (Nobuta et al.

2008). These data were obtained in different tissues than the leaf, nevertheless they could help confirming our hypothesis because real HP loci should be unaffected by *mop1-1* mutation for the generation of sRNAs.

4.2 Analysis of *MIRNA* loci and microRNA mature sequences

Over the total 159 maize *MIRNA* loci annotated in miRBase, 70 were confirmed by our data, including 25 over the total 29 annotated miRNA families. About half of the confirmed loci produced mature sequences that were variants of those reported in miRBase. The discrepancies found in miRBase might reflect inaccurate annotation in miRBase or leaf-specific differences in *MIRNA* processing patterns, confirming that many known *MIRNA* hairpins produce more than a single product (Coruh et al. 2014, Jeong D-H et al. 2013, Jeong D-H et al. 2011). The same targets for the known mature miRNAs and their detected variants were predicted, suggesting that the observed variations in *MIRNA* loci processing did not probably alter miRNA target selection. In a number of known miRNA families, some members were expressed at higher levels compared to others. As suggested for the miRNA variants, also the preferential expression within a miRNA family of a subset of its *MIRNA* loci members might reflect a tissue-specific regulation (Zhang et al. 2009). The *de novo* identification of MIRNA loci did not confirm 64 miRBase loci, while 25 were simply not expressed. Among the non-confirmed loci, we believe that at least some of them might be instead real *MIRNA* loci, because the *MIRNA* method used was set to minimize false positives, and as consequence, we expected some false negatives (Axtell MJ 2013). For example, 27 were rejected because they lacked evidence of the miRNA* expression but the absence of the miRNA* might be due to the very low stability of this sequence. Furthermore, known *MIRNA* loci that are located in tandem in the genome were rejected because the program identified a single locus comprising all of them failing to identify distinct hairpins inside the same locus. Although the approach followed might have led to the loss of real *MIRNA* loci we preferred to retrieve more reliable results at the cost of some false negatives, especially for the other classes of sRNA loci where less is known about their annotation.

Previously reported targets of the maize known miRNAs were confirmed by the prediction performed with TargetFinder applying low penalty score cutoff values, except for miR167, miR168, miR169, miR2118 and miR398. miR167 is a conserved miRNA among different species, previously predicted to target ARF TFs by Zhang et al. (2009). In our analysis miR167 was predicted to target two putative homologs of pumilio proteins, which in human are general regulators of miRNA accessibility to targets (Incarnato et al. 2013) and are themselves targeted by miRNAs (Fiore et al. 2014). miR168 is another conserved miRNA, previously predicted to target AGO1 (Zhang et al. 2009): this prediction was not confirmed by our analysis and none target was predicted for miR168 probably because of the stringent parameters used. miR168* showed lower abundances compared to miR168 but was the only miRNA* with high expression levels only registered for the mature miRNAs. This indicate that for miR168 also the star sequence might have a functional role: however none target was predicted for miR168* so its potential role remains unknown. We failed to predict any target also for miR398. The function of the sole miR398 target gene in maize is unknown (Xu et al. 2011). A previous work predicted miR398 to target *SOD* genes (Pei et al. 2013), which are the validated targets for miR398 in *Arabidopsis* (Sunkar et al. 2006), instead other works predicted only one unknown gene as miR398 target (Zhang et al. 2009, Xu et al. 2011). These results suggest that the function of miR398 might not be conserved in maize. Interestingly, in our analysis the miR399 was predicted to target an uncharacterized transcript, previously reported (Zhang et al. 2009), and also a TE transcript and a new transcript detected in our samples through the total RNA sequencing. Evidences that this new transcript could be the 5'-UTR of its downstream gene encoding a putative ortholog of the *Arabidopsis PHO2* (Calderón-Vázquez et al. 2011) were showed. Moreover, in some preliminary PCR results (data not shown in the 'Results' section) primers designed spanning the gap between the two genes allowed detecting a single transcript of the expected length, in the hypothesis the two genes were not separate but were effectively the same gene. The miR399 targeting of PHO2 was showed to be a conserved regulatory mechanism across a number of species including *Arabidopsis*, rice, poplar and *Medicago* (Bari et al. 2006) but never reported before in maize. Therefore, the experimental validation of the targeting prediction

and the function of the predicted target would be important to verify if it is conserved also in maize.

Plant *MIRNA* loci are transcribed by Pol II (Xie et al. 2005), therefore the 14 putative novel miRNAs that were dependent for their transcription on the activity of the Pol IV enzyme were thought to be most likely siRNAs instead of bona fide miRNAs. These miRNAs were all 24-nt long and homologous to repeat elements and 12 of them had the 5' terminal A. Among the other putative novel miRNAs homologous to repeats, five that were 24-nt long were not downregulated in *rmr6-1* and none of the 21-nt and 22-nt was Pol IV-dependent. Moreover, none of the putative novel miRNAs without homology to repeat elements was downregulated in the mutant. These results indicated that the simultaneous presence of all these characteristics: i) homology to repeat elements, ii) length of 24-nt and iii) 5' terminal A, could be a good predictor for the Pol IV-dependence of a miRNA. We believe that the Pol IV-dependent putative novel miRNAs were instead siRNAs transcribed from TEs with precursors lacking the predicted hairpin structure, or actively transcribed from TE rearrangements, involved in the establishment of transcriptional silencing (Lisch D 2012). The non-Pol IV-dependent putative novel miRNAs were the best candidate to be bona fide miRNAs, including those homologous to repeat elements (Li et al. 2011). Among these, some were not DE between wt and mutant in control conditions, were 21-nt long and had a predicted target being a putative protein-coding gene: miR-NEW12 and miR-NEW19 were predicted to target TFs, miR-NEW18 and miR-NEW20 were predicted to target a gene encoding a putative ABC transporter and miR-NEW21 was predicted to target a putative Zinc finger C3HC4 type domain containing protein. Even if expressed at low levels, these miRNAs would be the best candidate of being true new miRNAs because they showed the typical characteristics of known miRNAs: absence of TE-homology, 21-nt length, presence of target with a putative function and most of them had the 5' terminal U. The identified putative novel miRNA family miR-NEW10, had four members encoding 21-nt and 22-nt mature sequences homologous to *Mutator* elements: their high expression, their upregulation observed in *rmr6-1* mutant and the presence of predicted targets, low confidence genes and a putative Zinc finger

C3HC4 type domain containing protein, made them another interesting case for experimental validation.

4.3 Analysis of sRNA loci and their effects on gene expression and genome stability

Later, all the other classes of sRNA loci except *MIRNA* were analyzed. Maize is unique even among monocots in possessing a more abundant 22-nt sRNA population than that of 21-nt, suggesting an additional layer of sRNA complexity existing in maize (Nobuta et al. 2008). A previous work demonstrated that in *mop1-1* mutant 24-nt sRNAs were in general downregulated while 22-nt sRNAs were instead retained (Nobuta et al. 2008). Our analysis performed on the sRNA loci confirmed this data in *rmr6-1* mutant: <1% of the sRNA loci with size class of 22-nt was downregulated, about 6% and 2% of the HP and non-HP loci respectively were upregulated and the vast majority was not DE. The sRNA loci with size class of 22-nt, both HP and non-HP, were found to be strongly associated to the presence of repetitive elements: the majority of them (>93%) mapped to repeats and they were enriched in body regions of TE transcripts and in introns of protein-coding genes, where TEs are commonly inserted in maize (Schnable et al. 2009). On the contrary, they were less enriched in TE transcript flanking regions and clearly depleted in the flanking regions of protein-coding genes and lncRNA transcripts. Together these data suggested that most of the 22-nt sRNAs did not participate to canonical RdDM and were produced by a different pathway preferentially targeting repetitive elements. This might be the RDR6/DCL4 pathway (Nobuta et al. 2008), which is involved in the silencing initiation of Pol II-actively transcribed TEs (Slotkin et al. 2005). Indeed the maize genome has an unusual number of well-characterized active transposable elements compared to other plants (Lisch D 2012) and this might explain its unusual high abundance of 22-nt sRNAs. Therefore, to confirm this hypothesis we will further investigate in the wt the association between the sRNA loci with size class of 22-nt and the actively transcribed repetitive elements, identified through RNA-seq expression data. Data obtained in the *rmr6-1* mutant already confirm this hypothesis: only four genes were found to be upregulated in *rmr6-1* compared to wt and at the same time overlapping with an upregulated sRNA locus and in all

cases the sRNA locus was processed predominantly into 22-nt sRNAs. Over these four genes, two were unclassified genes expressed only in the mutant and sharing the same overlapping 22-nt size class sRNA locus that was also expressed in the wt but at lower levels. The other two genes were TEs: in both cases, both the TE and the overlapping 22-nt size class sRNA locus were specifically expressed only in the mutant. The reason why only two over the total 167 upregulated TEs in the mutant were targeted by 22-nt sRNAs remains unknown: it might be possible that only these two were recognized as aberrant transcripts and targeted by non-canonical RdDM pathways for silencing. It would be interesting to examine their CHH methylation levels to confirm this hypothesis.

The sRNA loci with size class of 24-nt, both HP and non-HP, which were the most abundant, were preferentially found in the chromosome arms where genes are more densely arranged and they showed high enrichment in the flanking regions of protein-coding genes, TEs and lncRNAs. sRNA loci with 24-nt size class included almost all of the Pol IV-dependent siRNA loci, which were the majority among the loci located in the flanking regions of genes. In a previous work, it was demonstrated that near-gene transposons induced *de novo* CHH methylation independent of transposon sequence or identity (Gent et al. 2013). Together these observations indicated that in gene flanking regions the major cause for RdDM loci was not the nature of the gene or the type of TE found in its flanking regions but instead the proximity to the gene itself.

All the analyzed sets of genes showed a higher probability to be flanked by sRNA loci with size class of 24-nt in their upstream region than in their downstream region. In gene-upstream regions, the sRNA loci with size class of 24-nt were correlated to the presence of actively transcribed genes while this correlation of not detected for the gene downstream regions. Only for protein-coding genes a similar positive correlation was also found when considering only the expressed genes, between the expression level of genes and the occupancy of upstream sRNA loci. These results suggested a possible influence of sRNAs on the expression of downstream genes, more evident for protein-coding genes. We preferred to investigate this correlation plotting the fraction of genes with flanking sRNA loci occupancy regardless of the sRNA abundance because plotting the average sRNA coverage in gene flanking regions, the approach followed by Gent

et al. (2013), was not a good method to show the general trend: the median values were zero, thus the average sRNA coverage was too much influenced by the high expression of a few sRNA loci.

Considering the enrichment of RdDM loci next to genes, the role of RdDM in DNA silencing and the positive correlation between 24-nt sRNA loci occupancy and the expression of the downstream genes, the observed consequences of the absence of siRNAs on gene expression in the mutant were not expected. Indeed the number of DE genes in *rmr6-1* was 1013 in contrast to the thousands siRNA loci that were lost in the mutant and the downregulation of an sRNA locus was not generally sufficient not even necessary to predict the up or downregulation of its close gene. Without excluding that in some cases the differential expression a gene could require the DE expression of its inner or proximal sRNA loci, our data indicated that this was not a general trend. These results together with the absence of morphological defects in the leaves indicated that the lack of siRNAs did not have a great impact on the genome stability of the leaf. It might be possible that different mechanisms maintained DNA silencing in gene proximal regions, where TEs are usually found in maize, in the absence of siRNAs. To help discussing this open question we primarily referred to results previously obtained in *Arabidopsis* RdDM mutants and in the maize mutant *mop1-1*, which is the only maize RdDM mutant for which NGS data have been analysed. A possible mechanism was proposed in *Arabidopsis* observing that the demethylase *ROS1*/*DML1* was significantly downregulated in *rdr2* mutants: the decrease in production of the ROS1 might lead to hypermethylation at CG sites and additional protection against the activation of transposons (Penterman et al. 2007). The maize homolog of *ROS1* was found to be downregulated in maize *mop1-1* mutant (Jia et al. 2009, Madzima et al. 2014) and in *rmr6-1* mutant according to our RNA-seq data, therefore the same mechanism proposed in *Arabidopsis* might be active in maize. However, a recent work demonstrated that in *mop1-1* DNA methylation in all C contexts was decreased at genomic loci targeted by RdDM and it was not significantly altered in other genomic loci (Gent et al. 2014). In *mop1-1* CHH methylation was decreased at loci targeted by RdDM but not completely removed and it was suggested that the residual CHH methylation could be triggered by MOP1-independent siRNAs or by siRNA-independent DNA methylase activity at

these loci (Gent et al. 2014). Lower levels of residual CHH methylation might still be sufficient to ensure DNA silencing. In this hypothesis, however, it still remains unclear why do plants engage Pol IV-dependent production of siRNAs in gene flanking regions when the silencing of these regions can be maintained even in the absence of siRNAs.

A possible answer to the open question regarding the roles of siRNAs in the control of gene expression could be the hypothesis in which siRNAs would be essential for the maintenance of genome stability ensuring the transgenerational transmission of the epigenetic information. Indeed it has been demonstrated in *Arabidopsis* that siRNAs prevented the transposition of stress-activated TEs and that this control happened in the somatic cells that produce the gametes (Ito et al. 2011). Moreover, in *Arabidopsis* it has been demonstrated that *dcl2* and *dcl3* deficiency mutants were partially impaired in the establishment of transgenerational changes in homologous recombination frequency and DNA methylation in the progeny of heat-stressed plants (Boyko et al. 2010). Finally, there are evidences that RdDM could have fundamental roles during meiosis. For example, it has been previously described that the effects on development appeared among *rmr6* mutant plants only after the genome had been exposed to a meiotic division in the absence of RMR6 function (Parkinson et al. 2007).

TEs distant from genes can only depend on symmetrical DNA methylation for silencing or can still produce 24-nt siRNAs required to initiate RdDM homology-dependent silencing of any incoming active TEs with sequence similarity (Nuthikattu et al. 2013, Kim and Zilberman 2014), many TEs are still targeted by RdDM but do not depend on it for silencing (Zemach et al. 2013). This could contribute explaining why the loss of siRNAs did not cause a widespread activation of TEs in the mutant. Indeed, about 35% of the total identified Pol IV-dependent siRNA loci were not located in the 2-kb flanking regions of genes annotated in our reconstructed transcriptome assembly.

The absence of siRNAs, although it was not found to compromise the genome stability in the leaf, did have some effects on gene expression. In *rmr6-1* the DE genes associated with Pol IV-dependent siRNA loci constituted <50% of the total DE genes, indicating that many DE genes were likely not direct targets of RdDM but instead secondary targets of the mutation. Considering the protein-

coding genes, altering their expression as a secondary effect of the absence of siRNAs might imply that these genes had specific roles involved in the response of the cell to the misregulation of the gene silencing mechanism controlled by RdDM and might explain why only a specific set of genes was DE in the mutant. The first evidence was the activation of TFs, because their function is the expression regulation of other precise sets of genes carrying specific sequence characteristics. Many upregulated genes with associated stress-response functions were found, which might be an indicator of a stress-like condition experience by cells in which RdDM is impaired. In this hypothesis, the upregulation of genes encoding cytochromes might be important to ensure the proper development of plants. Another indicator of altered cell homeostasis occurring in mutants was the downregulation of genes involved in the regulation of cell cycle, suggesting a misregulation of cell proliferation mechanisms provoked by the absence of a functional RdDM pathway of gene expression regulation. A decrease in the expression of genes encoding core histone proteins was detected in the mutant and this might contribute to cause a slowdown of the cell cycle. In the youngest wrapped leaf collected in our experiment, in addition to actively dividing cells a population of cells undergoing expansion was present: here a decrease amount of core histone proteins might cause the alteration of the chromatin organization with consequences on chromatin accessibility. Indeed, *mop1-1* mutation has been demonstrated to alter the chromatin accessibility by increasing it at the chromosome arms and decreasing it at pericentromeric regions (Madzima et al. 2014). It might be possible that an altered chromatin organization was induced in *rmr6-1* in the leaf cells and also in other cell types where it would have greater impact on gene expression. Indeed, the *mop1-1* mutation was observed to induce the differential expression of a substantial greater number of genes in the SAM (Jia et al. 2009) compared to the ear shoots (Madzima et al. 2014).

TEs are the main targets of siRNAs, so their upregulation observed in the Pol IV mutant *rmr6-1* was expected to be more directly linked to the loss of siRNAs. Surprisingly, only ~10% of the upregulated TEs in *rmr6-1* was associated with a downregulated sRNA locus in their body regions or flanking regions and the fraction of downregulated TEs associated with downregulated sRNA loci was

higher. This result suggested that for the majority of DE TEs their higher expression level in the mutant compared to the wt could be RdDM-independent. Over the total 167 upregulated TEs in the mutant, 142 were not expresses at all in the wt, so they were expressed specifically in the mutant. The non-100% identical genetic background between the wt and the *rmr6-1* mutant might explain part of this apparent RdDM-independent differential expression of TEs, especially for those detected only in the mutant. We have sequenced the wt sibling of the *rmr6-1* mutant and the analysis of these data would help distinguishing the genotype effects from the mutation effects on the expression of genes and in particular TEs.

4.4 Small RNA stress response evaluation

Stress response of both the miRNAs and the other sRNA categories was determined in wt and *rmr6-1* plants. Known miR156 mature sequences were upregulated after ten days of drought stress both in wt and mutants, while salinity stress affected their expression only in mutant plants. In *Arabidopsis*, the miR156 targeting of a subset of SPL proteins plays a role in the regulation of leaf cell number and size (Usami et al. 2009): this mechanism might be altered in the plant response to stress, indeed the stress treatments, in particular the drought and the drought plus salinity combined stress, caused visible alteration of the shape of the expanded leaves. The collected youngest wrapped did not show such damage but could anyway differentially express the miR156 as a result of the stress primarily sensed by the oldest leaves, especially in drought conditions. The newly identified miRNA called 156d.2 mapped within the same precursor of the conserved known miR156d sequence (here called miR156d.1) but didn't share homology with any miRNA annotate in miRBase. It was upregulated in the wt following drought stress and in the *rmr6-1* mutant following all the applied stresses and showed downregulation after the recovery from drought only in wt. In control conditions its expression was very low but in stress conditions it was >10RPM, suggesting a possible functional role for this new miRNA, which remains unknown because the miRNA was not predicted to have any targets. Four miRNAs were differentially expressed (DE) only in the wt following the ten days of drought stress. The upregulated were: the miR397b, previously reported to be upregulated in *Arabidopsis* in drought conditions (Sunkar and Zhu 2004) and downregulated in

rice in drought conditions (Zhou et al. 2010), and the miR398 family that plays a role in the oxidative stress response in *Arabidopsis* (Sunkar et al. 2006) but whose target function still remains unknown in maize. The downregulated were: one mature sequence of the miR166 family, which was previously reported to be instead upregulated following drought stress in a drought-sensitive maize line (Wang et a. 2014a), and one mature sequence of the miR396 family, previously reported to show different regulations caused by drought stress depending upon the species (Sunkar et al. 2012). Three miRNAs were DE only in the *rmr6-1* mutant following the ten days of drought plus salinity combined stress: the miR319c was upregulated and the miR399b and miR528 family were downregulated. The miR319 family has been previously demonstrated to be upregulated in leaves of both drought-sensitive and drought-tolerant maize varieties during PEG-induced drought stress (Wang et al. 2014a), and its role in the regulation of leaf cell proliferation by the targeting of the TCP TFs has been demonstrated in *Arabidopsis* (Palatnik et al. 2003, Martin-Trillo and Cubas 2010). The miR399 has been previously reported to be involved in different stress conditions in *Arabidopsis* (Pant et al. 2008). The miR528 family predicted targets are multicopper-oxidase and laccase genes involved in energy metabolism and scavenging of the oxidative species produced during stress (Zhang et al. 2009); we observed a down-regulation of the miR528 family after the combined stress, accordingly to its down-regulation observed in *Triticum dicoccoides* in shock drought stress (Kantar et al. 2011) but opposite to previous data in sugarcane (Ferreira et al. 2012) and *Brachypodium* (Budak and Akpinar 2011, Bertolini et al. 2013), where it was found to be upregulated after drought stress. In summary, only the miR156 family was DE in both wt and mutant plants, all the other miRNAs were DE only in one specific genotype, suggesting a possible influence of the *rmr6-1* mutation in the stress response of plants. Salt stress, alone and in combination with drought, influenced the expression of known miRNAs only in the *rmr6-1* mutant, which might was more susceptible to salt than the wt. Except for the miR156d.2, all the other DE miRNAs in stress conditions did not significantly change their expression after the recovery, suggesting that the pathways regulated by the long-term abiotic stress-responsive miRNAs might continue to be altered even when the stress has been removed. In total, a few numbers of

miRNAs were DE following the stresses. One putative novel miRNA family, miR-NEW46, showed an opposite response to the drought plus salinity combined stress in the two genotypes: it was downregulated in the *rmr6-1* mutant after ten days of treatment compared to control conditions and it was upregulated in wt after the recovery from the stress compared to before the recovery during the stress. One of the two mature sequences encoded by the miR-NEW46 precursor was predicted to target a putative AP2 TF, as for miR172 family, and both the two mature sequences were predicted to target a TE transcript. Although this miRNA showed low expression levels in all samples, its differential expression could be experimentally tested to validate its different behavior in the two genotypes. In addition to miRNAs, 19 sRNA loci of the other categories also responded to the applied stresses: 12 showed differential expression only in the wt during drought stress and three only in the *rmr6-1* mutant in drought or drought plus salinity stresses. The majority of the DE sRNA loci were located in genic regions but only in two cases both the sRNA locus and the overlapping genes were DE. In these cases the stress caused the upregulation of both the sRNA locus and the gene, suggesting that the gene might be the precursors of the sRNAs. We predicted the potential targets of the most abundant sRNA species within the DE sRNA loci applying the same method for miRNAs: none of the predicted target was found to be DE in stress conditions (data not shown in the 'Results' section). The function of the DE sRNAs remains unclear because we didn't find for them any evidence of having an influence on the expression of genes in *cis* and in *trans*. Our data suggested that the plant response to the applied stress treatments did not involve the action of siRNAs as a general strategy to modulate gene expression; indeed the *rmr6-1* mutants did not show more severe phenotypes in stress conditions compared to the wt.

Considering the literature assessing the miRNA stress response to drought and salinity in maize, one work found a similar number of DE miRNAs compared to our analysis (Kong et al. 2010), whereas tens to hundreds of miRNAs were found DE following these stresses in other works (Ding et al. 2009, Wei et al. 2009, Wang et al. 2014a). A difference in terms of the extent of differential expression induced by abiotic stress was also noted for the other classes of sRNAs, for example hundreds of sRNA sequences were reported to be

DE after cold, heat and salinity stresses in *Brachypodium* (Wang et al. 2014b) and thousands of sRNA loci showed differential expression following drought stress in foxtail millet (Qi et al. 2013). This diversity might de due to the different applied stress protocols: previous works applied more severe stress conditions, such as PEG-simulated drought conditions (Qi et al. 2013, Wang et al. 2014a) or applied the stress to younger plants at the seedling-stage (Ding et al. 2009) or detected the stress effects after a shorter period of treatment application, at most of three days (Wei et al. 2009, Wang et al. 2014b). We applied realistic stress conditions and examined the stress effects in the leaf of adult plants, after ten days of treatment application. Therefore, we cannot exclude that a greater number of sRNAs could responded to the stress in the earlier stages of its application, returning to basal levels after ten days of stress, or in different tissues than the youngest wrapped leaf. The followed approach to identify stress responsive sRNAs that could be involved in stress tolerance mechanisms, although allowed retrieving less numerous DE sRNAs, could have a greater translatability for crop improvement because we applied stress episodes mimicking the field conditions: we applied agronomically realistic drought and salinity conditions that were reached gradually (see Chapter 1 for the set up of the stress protocols).

5 References

Adai A, Johnson C, Mlotshwa S, Archer-Evans S, Manocha V, Archer-Evans S, Vance V, Sundaresan V: Computational prediction of miRNAs in *Arabidopsis thaliana*. *Genome Res* 2005, 15: 78-91

Agius F, Kapoor A, Zhu JK: Role of the *Arabidopsis* DNA glycosylase/lyase ROS1 in active DNA demethylation. *Proc Natl Acad Sci USA* 2006, 103:11796-801

Ahmed I, Sarazin A, Bowler C, Colot V, Quesneville H: Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in *Arabidopsis*. *Nucleic Acids Res* 2011, 39, 6919–6931

Aklilu BB, Soderquist RS, Culligan KM: Genetic analysis of the Replication Protein A large subunit family in *Arabidopsis* reveals unique and overlapping roles in DNA repair, meiosis and DNA replication. *Nucleic Acids Res* 2014, 42(5):3104-18

Allen E, Xie Z, Gustafson AM, Sung GH, Spatafora JW, Carrington JC: Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. *Nat Genet* 2004, 36:1282-1290

Allen E, Xie Z, Gustafson AM, Carrington JC: MicroRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* 2005, 121, 207–221

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. *J Mol Biol* 1990, 215:403-410

Anders S, Huber W: Differential expression analysis for sequence count data. *Genome Biol* 2010, 11: R106

Andrews S: FastQC A quality control tool for high throughput sequence data *http://www.bioinformatics.babraham.ac.uk/projects/fastqc/*

Axtell MJ: Evolution of microRNAs and their targets: Are all microRNAs biologically relevant? *Biochim Biophys Acta* 2008, 1779:725–34

Axtell MJ, Westholm JO, Lai EC: Vive la différence: biogenesis and evolution of microRNAs in plants and animals. *Genome Biol* 2011, 12(4):221

Axtell MJ: Classification and comparison of small RNAs from plants. *Annu Rev Plant Biol* 2013a, 64:137-59

Axtell MJ: ShortStack: Comprehensive annotation and quantification of small RNA genes. *RNA* 2013b, 19: 740–751

Axtell MJ: Butter: High-precision genomic alignment of small RNA-seq data. 2014 preprint: bioRxiv doi: http://dx.doi.org/10.1101/007427

Bao N, Lye K-W, Barton MK: MicroRNA binding sites in *Arabidopsis* class III HD-ZIP mRNAs are required for methylation of the template chromosome. *Dev Cell* 2004, 7:653–62

Barber WT, Zhang W, Win H, Varala KK, Dorweiler JE, Hudson ME, Moose SP: Repeat associated small RNAs vary among parents and following hybridization in maize. *Proc Natl Acad Sci USA* 2012, 109, 10444–10449

Barbour JE, Liao IT, Stonaker JL, Lim JP, Lee CC, Parkinson SE, Kermicle J, Simon SA, Meyers BC, Williams-Carrier R, Barkan A, Hollick JB: required to maintain repression2 is a novel protein that facilitates locus-specific paramutation in maize. *Plant Cell* 2012, 24(5):1761-75

Bari R, Pant BD, Stitt M, Scheible WR: PHO2, microRNA399, and PHR1 define a phosphate-signaling pathway in plants. *Plant Physiol* 2006, 141, 988–999

Bartels D, Sunkar R: Drought and salt tolerance in plants. *Crit Rev Plant Sci* 2005, 24, 23–58

Becker C, Hagmann J, Müller J, Koenig D, Stegle O, Borgwardt K, Weigel D: Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature* 2011, 20;480(7376):245-9

Bertolini E, Verelst W, Horner DS, Gianfranceschi L, Piccolo V, Inze D, Pe ME, Mica E: Addressing the role of microRNAs in reprogramming leaf growth during drought stress in *Brachypodium distachyon*. *Mol Plant* 2013, 6:423–443

Bies-Etheve N, Pontier D, Lahmy, S, Picart C, Vega D, Cooke R, Lagrange T: RNA-directed DNA methylation requires an AGO4-interacting member of the SPT5 elongation factor family. *EMBO Rep* 2009, 10, 649–654

Bolstad BM, Irizarry RA, Astrand M, Speed TP: A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 2003, 19: 185– 193

Boyko A, Blevins T, Yao Y, Golubov A, Bilichak A, Ilnytskyy Y, Hollunder J, Meins F Jr, Kovalchuk I: Transgenerational adaptation of *Arabidopsis* to stress requires DNA methylation and the function of Dicer-like proteins. *PLoS One* 2010, 5,e9514

Bucher E, Reinders J, Mirouze M: Epigenetic control of transposon transcription and mobility in *Arabidopsis*. *Curr Opin Plant Biol* 2012, 15, 503–510

Budak H, Akpinar A: Dehydration stress-responsive miRNA in *Brachypodium distachyon*: evident by genome-wide screening of microRNAs expression. *OMICS* 2011, 15:791-799

Bullard JH, Purdom E, Hansen KD, Dudoit S: Evaluation of statistical methods for normalization and differential expression in mRNA-seq experiments. *BMC Bioinformatics* 2010, 11:94

Calderón-Vázquez C, Sawers RJH, Herrera-Estrella L: Phosphate deprivation in maize: genetics and genomics. *Plant Physiol* 2011, 156(3):1067-1077

Cao X, Jacobsen SE: Role of the *Arabidopsis* DRM methyltransferases in *de novo* DNA methylation and gene silencing. *Curr Biol* 2002, 12, 1138–1144

Carbognin L, Tosi L: Il progetto ISES per l'analisi dei processi di intrusione salina e subsidenza nei territori meridionali delle provice di padova e venezia. Istituto per lo studio della dinamica delle grandi masse. C.N.R. Venezia. 2003

Carpenter R, Martin C, Coen ES: Comparison of genetic behavior of the transposable element Tam3 at 2 unlinked pigment loci in *Antirrhinum-Majus*. *Mol Gen Genet* 1987, 207, 82–89

Chan SW, Zilberman D, Xie Z, Johansen LK, Carrington JC, Jacobsen SE: RNA silencing genes control *de novo* DNA methylation. *Science* 2004, 27;303(5662):1336

Chen CN, Chu CC, Zentella R, Pan SM, Ho TH: AtHVA22 gene family in *Arabidopsis*: phylogenetic relationship, ABA and stress regulation, and tissue-specific expression. *Plant Mol Biol* 2002, 49(6):633-44

Chen M, Meng Y, Mao C, Chen D, Wu P: Methodological framework for functional characterization of plant microRNAs. *J Exp Bot* 2010, 61(9):2271-80

Chuck G, Cigan AM, Saeteurn K, Hake S: The heterochronic maize mutant Corngrass1 results from overexpression of a tandem microRNA. *Nat Genet* 2007, 39(4):544-9

Chuck G, Candela H, Hake S: Big impacts by small RNAs in plant development. *Curr Opin Plant Biol* 2009, 12: 81–86

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE: Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* 2008, 452, 215–219

Conesa A, Götz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 2005, Vol. 21, pp. 3674-3676

Coruh C, Shahid S, Axtell MJ: Seeing the forest for the trees: annotating small RNA producing genes in plants. *Curr Opin Plant Biol* 2014,18:87-95

Creighton CJ, Reid JG, Gunaratne PH: Expression profiling of microRNAs by deep sequencing. *Brief Bioinform* 2009, 10(5):490-497

Cubas P, Vincent C, Coen E: An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* 1999, 9;401(6749):157-61

Ding D, Zhang L, Wang H, Liu Z, Zhang Z, Zheng Y: Differential expression of miRNAs in response to salt stress in maize roots. *Ann Bot* 2009, 103(1):29-38

Dorweiler JE, Carey CC, Kubo KM, Hollick JB, Kermicle JL, Chandler VL: *mediator of paramutation1* is required for establishment and maintenance of paramutation at multiple maize loci. *Plant Cell* 2000, 12(11):2101-18

Dowen RH, Pelizzola M, Schmitz RJ, Lister R, Dowen JM, Nery JR, Dixon JE, Ecker JR: Widespread dynamic DNA methylation in response to biotic stress. *Proc Natl Acad Sci USA* 2012, 7;109(32):E2183-91

Dunoyer P, Himber C, Ruiz-Ferrer V, Alioua A, Voinnet O: Intra- and intercellular RNA interference in *Arabidopsis* thaliana requires components of the microRNA and heterochromatic silencing pathways. *Nat Genet* 2007, *39*, 848–856

Eichten SR, Swanson-Wagner RA, Schnable JC, Waters AJ, Hermanson PJ, Liu S, Yeh C-T, Jia Y, Gendler K, Freeling M, Schnable PS, Vaughn MW, Springer NM: Heritable epigenetic variation among maize inbreds. *PLoS Genet* 2011, 7(11):e1002372

Eichten SR, Ellis NA, Makarevitch I, Yeh C-T, Gent JI, Guo L, McGinnis KM, Zhang X, Schnable PS, Vaughn MW, Dawe RK, Springer NM: Spreading of heterochromatin is limited to specific families of maize retrotransposons. *PLoS Genet* 2012, 8, e1003127

Enke RA, Dong Z, Bender J: Small RNAs prevent transcription-coupled loss of histone H3 lysine 9 metylation in *Arabidopsis thaliana. PLoS Genet* 2011, 7, e1002350

Erhard KF Jr, Stonaker JL, Parkinson SE, Lim JP, Hale CJ, Hollick JB: RNA polymerase IV functions in paramutation in *Zea mays*. *Science* 2009, 323: 1201–5

Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Law TF, Grant SR, Dangl JL, Carrington JC: High-throughput sequencing of *Arabidopsis* microRNAs: evidence for frequent birth and death of MIRNA genes. *PLoS ONE* 2007, 2, e219

Fahlgren N, Sullivan CM, Kasschau KD, Chapman EJ, Cumbie JS, Montgomery TA, Gilbert SD, Dasenko M, Backman TWH, Givan SA, Carrington JC: Computational and analytical framework for small RNA profiling by high-throughput sequencing. *RNA* 2009, 15:992-1002

Fang Y, Spector DL: Identification of nuclear dicing bodies containing proteins for microRNA biogenesis in living *Arabidopsis* plants. *Curr Biol* 2007, *17*, 818–823

Fedoroff NV: Presidential address. Transposable elements, epigenetics, and genome evolution. *Science* 2012, 9;338(6108):758-67

Fei Q, Xia R, Meyers BC: Phased, Secondary, Small Interfering RNAs in Posttranscriptional Regulatory Networks. *Plant Cell* 2013, 25(7):2400-2415

Feng S, Jacobsen SE, Reik W: Epigenetic reprogramming in plant and animal development. *Science* 2010, 330, 622–627

Ferreira TH, Gentile A, Vilela RD, Costa GGL, Dias LI, Endres L, Menossi M: MicroRNAs associated with drought response in the bioenergy crop sugarcane (*Saccharum* spp.). *PLoS ONE* 2012, 7:e46703

Fiore R, Rajman M, Schwale C, Bicker S, Antoniou A, Bruehl C, Draguhn A, Schratt G: MiR-134-dependent regulation of Pumilio-2 is necessary for homeostatic synaptic depression. *EMBO J* 2014, 1;33(19):2231-46

Flowers TJ: Improving crop salt tolerance. *J Exp Bot* 2004, 55:307-319

Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-Somoza I, Leyva A, Weigel D, García JA, Paz-Ares J: Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat Genet* 2007, 39:1033–37

Gao ZH, Liu HL, Daxinger L, Pontes O, He X, Qian W, Lin H, Xie M, Lorkovic ZJ, Zhang S, Miki D, Zhan X, Pontier D, Lagrange T, Jin H, Matzke AJ, Matzke M, Pikaard CS, Zhu JK: An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation. *Nature* 2010, 465, 106–109

Garmire LX, Subramaniam S: Evaluation of normalization methods in mammalian microRNA-Seq data. *RNA* 2012, 18(6):1279-1288

Gehring M, Bubb KL, Henikoff S: Extensive demethylation of repetitive elements

during seed development underlies gene imprinting. *Science* 2009, 324, 1447–1451

Gendler K, Paulsen T, Napoli C: ChromDB: the chromatin database. *Nucleic Acids Res* 2008, 36: D298–D302

Gent JI, Dong Y, Jiang J, Dawe RK: Strong epigenetic similarity between maize centromeric and pericentromeric regions at the level of small RNAs, DNA methylation and H3 chromatin modifications. *Nucleic Acids Res* 2012, 40(4): 1550–1560

Gent JI, Ellis NA, Guo L, Harkess AE, Yao Y, Zhang X, Dawe RK: CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize. *Genome Res* 2013, 23 628–637

Gent JI, Madzima TF, Bader R, Kent MR, Zhang X, Stam M, McGinnis KM, Dawe RK: Accessible DNA and relative depletion of H3K9me2 at maize loci undergoing RNA-directed DNA methylation. *Plant Cell* 2014

Greaves IK, Groszmann M, Ying H, Taylor JM, Peacock WJ, Dennis ES: *Trans* chromosomal methylation in *Arabidopsis* hybrids. *Proc Natl Acad Sci USA* 2012, 109, 3570–3575

Groszmann M, Greaves IK, Albertyn ZI, Scofield GN, Peacock WJ, Dennis ES: Changes in 24-nt siRNA levels in *Arabidopsis* hybrids suggest an epigenetic contribution to hybrid vigor. *Proc Natl Acad Sci USA* 2011, 108, 2617–2622

Ha M, Lu J, Tian L, Ramachandran V, Kasschau KD, Chapman EJ, Carrington JC, Chen X, Wang XJ, Chen ZJ: Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids. *Proc Natl Acad Sci USA* 2009, 106, 17835–17840

Haberer G, Young S, Bharti AK, Gundlach H, Raymond C, Fuks G, Butler E, Wing

RA, Rounsley S, Birren B, Nusbaum C, Mayer KF, Messing J: Structure and architecture of the maize genome. *Plant Physiol* 2005, 139(4):1612-24

Hale CJ, Stonaker JL, Gross SM, Hollick JB: A novel Snf2 protein maintains *trans*-generational regulatory states established by paramutation in maize. *PLoS Biol* 2007, 16;5(10):e275

Hale CJ, Erhard KF Jr, Lisch D, Hollick JB: Production and processing of siRNA precursor transcripts from the highly repetitive maize genome. *PLoS Genet* 2009, 5(8):e1000598

Hamilton AJ, Baulcombe DC: A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* 1999, 286:950–952

Hardcastle TJ, Kelly KA: baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics* 2010, 11: 422

Hardcastle TJ, Kelly KA, Baulcombe D: Identifying small interfering RNA loci from high-throughput sequencing data. *Bioinformatics* 2012, 15;28(4):457-63

Hashida SN, Uchiyama T, Martin C, Kishima Y, Sano Y, Mikami T: The temperature-dependent change in methylation of the *Antirrhinum* transposon Tam3 is controlled by the activity of its transposase. *Plant Cell* 2006, 18, 104–118

He XJ, Hsu YF, Zhu S, Wierzbicki AT, Pontes O, Pikaard CS, Liu HL, Wang CS, Jin H, Zhu, JK: An effector of RNA-directed DNA methylation in *Arabidopsis* is an ARGONAUTE 4- and RNA-binding protein. *Cell* 2009, 137, 498–508

He G, Chen B, Wang X, Li X, Li J, He H, Yang M, Lu L, Qi Y, Wang X, Deng XW: Conservation and divergence of transcriptomic and epigenomic variation in maize hybrids. *Genome Biol* 2013, 14: R57

Hirsch S, Baumberger R, Grossniklaus U: Epigenetic variation, inheritance, and

selection in plant populations. *Cold Spring Harb Sym* 2012, 77, 97–104

Hollick JB, Kermicle JL, Parkinson SE: *Rmr6* maintains meiotic inheritance of paramutant states in *Zea mays*. *Genetics* 2005, 171(2):725-40

Hu W, Wang T, Xu J, Li H: MicroRNA mediates DNA methylation of target genes. *Biochem Biophys Res Commun* 2014, 21;444(4):676-81

Huber W, von Heydebreck A, Sültmann H, Poustka A, Vingron M: Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* (Suppl1) 2002, 18: S96–S104

Huettel B, Kanno T, Daxinger L, Aufsatz W, Matzke AJ, Matzke M: Endogenous targets of RNA-directed DNA methylation and Pol IV in *Arabidopsis*. *EMBO J* 2006, 25:2828-36

Inagaki S, Kakutani T: What triggers differential DNA methylation of genes and TEs: contribution of body methylation? *Cold Spring Harb Sym* 2013, 77, 155–160

Incarnato D, Neri F, Diamanti D, Oliviero S: MREdictor: a two-step dynamic interaction model that accounts for mRNA accessibility and Pumilio binding accurately predicts microRNA targets. *Nucleic Acids Res* 2013, 41(18):8421-33

Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski J: An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature* 2011, 472, 115–119

Ito H: Small RNAs and transposon silencing in plants. *Dev Growth Differ* 2012, 54(1):100-7

Jeong D-H, Park S, Zhai J, Gurazada SG, De Paoli E, Meyers BC, Green PJ: Massive analysis of rice small RNAs: mechanistic implications of regulated microRNAs and variants for differential target RNA cleavage. *Plant Cell* 2011, 23:

4185–4207

Jeong D-H, Thatcher SR, Brown RSH, Zhai J, Park S, Rymarquis LA, Meyers BC, Green PJ: Comprehensive investigation of microRNAs enhanced by analysis of sequence variants, expression patterns, argonaute loading, and target cleavage. *Plant Physiol* 2013, 162(3):1225-1245

Jewell MC, Campbell BC, Godwin ID: Transgenic plants for abiotic stress resistance. In: Kole K, Michler CH, Abbott AG, Hall TC, editors. Transgenic crop plants. *Springer-Verlag Berlin Heidelberg* 2010, p. 67-132

Ji LJ, Chen XM: Regulation of small RNA stability: methylation and beyond. *Cell Res* 2012, 22, 624–636

Jia Y, Lisch DR, Ohtsu K, Scanlon MJ, Nettleton D, Schnable PS: Loss of RNA-dependent RNA polymerase 2 (RDR2) function causes widespread and unexpected changes in the expression of transposons, genes, and 24-nt small RNAs. *PLoS Genet* 2009, 5(11):e1000737

Johnson C, Kasprzewska A, Tennessen K, Fernandes J, Nan GL, Walbot V, Sundaresan V, Vance V, Bowman LH: Clusters and superclusters of phased small RNAs in the developing inflorescence of rice. *Genome Res* 2009, 19(8):1429-40

Johnson LM, Du J, Hale CJ, Bischof S, Feng S, Chodavarapu RK, Zhong X, Marson G, Pellegrini M, Segal DJ, Patel DJ, Jacobsen SE: SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature* 2014, 507, 124–128

Johnston AJ, Kirioukhova O, Barrell PJ, Rutten T, Moore JM, Baskar R, Grossniklaus U, Gruissem W: Dosage-sensitive function of retinoblastoma related and convergent epigenetic control are required during the *Arabidopsis* life cycle. *PLoS Genet* 2010, 17;6(6):e1000988

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S: InterProScan 5: genome-scale protein function classification. *Bioinformatics* 2014, 30: 1236–1240

Jones-Rhoades MW: Conservation and divergence in plant microRNAs. *Plant Mol Biol* 2012, 80:3–16

Juarez M, Kui J, Thomas J, Heller B, Timmermans M: microRNA-mediated repression of *rolled leaf1* specifies maize leaf polarity. *Nature* 2004, **428:** 84–88

Kantar M, Lucas S, Budak H: MiRNA expression patterns of Triticum dicoccoides in response to shock drought stress. *Planta* 2011, 233: 471–484

Kasschau KD, Fahlgren N, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Carrington JC: Genome-wide profiling and analysis of *Arabidopsis* siRNAs. *PLoS Biol* 2007, 5, e57

Katiyar-Agarwal S, Gao S, Vivian-Smith A, Jin H: A novel class of bacteria-induced small RNAs in *Arabidopsis*. *Genes Dev* 2007, 21:3123–34

Khraiwesh B, Zhu J-K, Zhu J: Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochim Biophys Acta* 2012, 1819(2):137-148

Kim MY, Zilberman D: DNA methylation as a system of plant genomic immunity. *Trends Plant Sci* 2014, 19(5):320-6

Kinoshita Y, Saze H, Kinoshita T, Miura A, Soppe WJ, Koornneef M, Kakutani T: Control of FWA gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. *Plant J* 2007, 49, 38–45

Kong YM, Elling AA, Chen B, Deng XW: Differential expression of microRNAs in maize inbred and hybrid lines during salt and drought stress. *AJPS* 2010, 1, 69-76

Kozomara A, Griffiths-Jones S: miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* 2011, 39: D152-D157

Langmead B, Trapnell C, Pop M, Salzberg SL: Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009, 10: R25

Law JA, Ausin I, Johnson LM, Vashisht AA, Zhu J-K, Wohlschlegel JA, Jacobsen SE: A protein complex required for polymerase V transcripts and RNA-directed DNA methylation in *Arabidopsis*. *Curr Biol* 2010, 20: 951–56

Law JA, Vashisht AA, Wohlschlegel JA, Jacobsen SE: SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA polymerase IV. *PLoS Genet* 2011, 7, e1002195

Law JA, Du J, Hale CJ, Feng S, Krajewski K, Palanca AMS, Strahl BD, Patel DJ, Jacobsen SE: polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* 2013, 498, 385–389

Li Y, Li C, Xia J, Jin Y: Domestication of transposable elements into MicroRNA genes in plants. *PLoS One* 2011, 6:e19212

Li X, Qian W, Zhao Y, Wang C, Shen J, Zhu JK, Gong Z: Antisilencing role of the RNA-directed DNA methylation pathway and a histone acetyltransferase in *Arabidopsis*. *Proc Natl Acad Sci USA* 2012, 109:11425-30

Lisch D, Carey CC, Dorweiler JE, Chandler VL: A mutation that prevents paramutation in maize also reverses *Mutator* transposon methylation and silencing. *Proc Natl Acad Sci USA* 2002, 30;99(9):6130-5

Lisch D: Regulation of transposable elements in maize. *Curr Opin Plant Biol* 2012,

13: 1–6

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR: Highly integrated single-base resolution maps of the *Arabidopsis* genome. *Cell* 2008, 133, 523–536

Liu J, He Y, Amasino R, Chen X: siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in *Arabidopsis*. *Genes Dev* 2004, 1;18(23):2873-8

Liu H, Qin C, Chen Z, Zuo T, Yang X, Zhou H, Xu M, Cao S, Shen Y, Lin H, He X, Zhang Y, Li L, Ding H, Lübberstedt T, Zhang Z, Pan G: Identification of miRNAs and their target genes in developing maize ears by combined small RNA and degradome sequencing. *BMC Genomics* 2014, 15:25

Liu ZW, Shao CR, Zhang CJ, Zhou JX, Zhang SW, Li L, Chen S, Huang HW, Cai T, He XJ: The SET domain proteins SUVH2 and SUVH9 are required for Pol V occupancy at RNA-directed DNA methylation loci. *PLoS Genet* 2014, 10, e1003948

Llave C, Kasschau KD, Rector MA, Carrington JC: Endogenous and silencing-associated small RNAs in plants. *Plant Cell* 2002, 14:1605–1619

Luan M, Xu M, Lu Y, Zhang L, Fan Y, Wang L: Expression of zma-miR169 miRNAs and their target ZmNF-YA genes in response to abiotic stress in maize leaves. *Gene* 2015, 555 178–185

Ma Z, Coruh C, Axtell MJ: *Arabidopsis lyrata* small RNAs: transient *MIRNA* and small interfering RNA loci within the *Arabidopsis* genus. *Plant Cell* 2010, 22: 1090–103

Madzima TF, Huang J, McGinnis KM: Chromatin structure and gene expression changes associated with loss of MOP1 activity in *Zea mays*. *Epigenetics* 2014, 9(7):1047-59

Manavella PA, Koenig D, Rubio-Somoza I, Burbano HA, Becker C, Detlef W: Tissue-specific silencing of *Arabidopsis* SU(VAR)3-9 HOMOLOG8 by miR171a*. *Plant Physiol* 2013, 161: 805–812

Manning K, Tör M, Poole M, Hong Y, Thompson AJ, King GJ, Giovannoni JJ, Seymour GB: A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nat Genet* 2006, 38(8):948-52

Marí-Ordóñez A, Marchais A, Etcheverry M, Martin A, Colot V, Voinnet O: Reconstructing *de novo* silencing of an active plant retrotransposon. *Nat Genet* 2013, 45(9):1029-39

Martin M: Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 2011, Vol. 17 http://journal.embnet.org/index.php/embnetjournal/article/view/200/479

Martin-Trillo M, Cubas P: TCP genes: a family snapshot ten years later. *Trends Plant Sci* 2010, 15:31-39

Matzke M, Kanno T, Daxinger L, Huettel B, Matzke AJM: RNA-mediated chromatin-based silencing in plants. *Curr Opin Cell Biol* 2009, 21(3):367-76

Matzke MA, Mosher RA: RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nature Rev Genet* 2014, 15, 394–408

McCormick KP, Willmann MR, Meyers BC: Experimental design, preprocessing, normalization and differential expression analysis of small RNA sequencing experiments. *Silence* 2011, 2:2

McCue AD, Nuthikattu S, Reeder SH, Slotkin RK: Gene expression and stress response mediated by the epigenetic regulation of a transposable element small

RNA. *PLoS Genet* 2012a, 8(2):e1002474

McCue AD, Slotkin RK: Transposable element small RNAs as regulators of gene expression. *Trends Genet* 2012b, 28, 616–23

McCue AD, Nuthikattu S, Slotkin RK: Genome-wide identification of genes regulated *in trans* by transposable element small interfering RNAs. *RNA biol* 2013, 10, 1379–1395

Meyers BC, Tingey SV, Morgante M: Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Res* 2001, 11(10):1660-76

Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, Bowman JL, Cao X, Carrington JC, Chen X, Green PJ, Griffiths-Jones S, Jacobsen SE, Mallory AC, Martienssen RA, Poethig RS, Qi Y, Vaucheret H, Voinnet O, Watanabe Y, Weigel D, Zhu J-K: Criteria for annotation of plant microRNAs. *Plant Cell* 2008, 20, 3186–3190

Michaux C, Verneuil N, Hartke A, Giard JC: Physiological roles of small RNA molecules. *Microbiology* 2014, 160(Pt 6):1007-19

Mirouze M, Reinders J, Bucher E, Nishimura T, Schneeberger K, Ossowski S, Cao J, Weigel D, Paszkowski J, Mathieu O: Selective epigenetic control of retrotransposition in *Arabidopsis*. *Nature* 2009, 461, 427–430

Mirouze M, Vitte C: Transposable elements, a treasure trove to decipher epigenetic variation: insights from *Arabidopsis* and crop epigenomes. *J Exp Bot* 2014, 65(10):2801-12

Mizutani M: Impacts of diversification of cytochrome P450 on plant metabolism. *Biol Pharm Bull* 2012, 35(6):824-32

Morales-Ruiz T, Ortega-Galisteo AP, Ponferrada-Marín MI, Martínez-Macías MI, Ariza RR, Roldán-Arjona T: DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases. *Proc Natl Acad Sci USA* 2006, 103:6853-8

Moritoh S, Eun C-H, Ono A, Asao H, Okano Y, Yamaguchi K, Shimatani Z, Koizumi A, Terada R: Targeted disruption of an orthologue of DOMAINS REARRANGED METHYLASE 2 (OsDRM2) impairs the growth of rice plants by abnormal DNA methylation. *Plant J* 2012, 71, 85–98

Mosher RA, Schwach F, Studholme D, Baulcombe DC: Pol IVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. *Proc Natl Acad Sci USA* 2008, 105:3145–50

Mosher RA, Melnyk CW, Kelly KA, Dunn RM, Studholme DJ, Baulcombe DC: Uniparental expression of Pol IV-dependent siRNAs in developing endosperm of Arabidopsis. *Nature* 2009, 460, 283–286

Moxon S, Schwach F, Dalmay T, Maclean D, Studholme DJ, Moulton V: A toolkit for analysing large-scale plant small RNA datasets. *Bioinformatics* 2008, 1;24(19):2252-3

Munns R: Plant adaptations to salt and water stress: differences and commonalities. *Adv Bot Res* 2011, 57:1–32

Nakazaki T, Okumoto Y, Horibata A, Yamahira S, Teraishi M, Nishida H, Inoue H, Tanisaka T: Mobilization of a transposon in the rice genome. *Nature* 2003, 421, 170–172

Neilsen CT, Goodall GJ, Bracken CP: IsomiRs – the overlooked repertoire in the dynamic microRNAome. *Trends Genet* 2012, 28(11):544-9

Ng KH, Yu H, Ito T: AGAMOUS controls GIANT KILLER, a multifunctional chromatin

modifier in reproductive organ patterning and differentiation. *PLoS Biol* 2009, 7:e1000251

Nobuta K, Lu C, Shrivastava R, Pillay M, De Paoli E, Accerbi M, Arteaga-Vazquez M, Sidorenko L, Jeong DH, Yen Y, Green PJ, Chandler VL, Meyers BC: Distinct size distribution of endogeneous siRNAs in maize: Evidence from deep sequencing in the *mop1-1* mutant. *Proc Natl Acad Sci USA* 2008, 30;105(39):14958-63

Nogueira FT, Madi S, Chitwood DH, Juarez MT, Timmermans MC: Two small regulatory RNAs establish opposing fates of a developmental axis. *Genes Dev* 2007, 21 750–755

Nuthikattu S, McCue AD, Panda K, Fultz D, DeFraia C, Thomas EN, Slotkin RK: The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs. *Plant Physiol* 2013, *162*(1), 116–131

Palatnik JF, Allen E, Wu X, Schommer C, Schwab R, Carrington JC, Weigel D: Control of leaf morphogenesis by microRNAs. *Nature* 2003, 425:257-263

Panda K, Slotkin RK: Proposed mechanism for the initiation of transposable element silencing by the RDR6-directed DNA methylation pathway. *Plant Signal Behav* 2013, 8(8):e25206

Pant BD, Buhtz A, Kehr J, Scheible WR: MicroRNA399 is a long-distance signal for the regulation of plant phosphate homeostasis. *Plant J* 2008, 53, 731–738 10

Park W, Li J, Song R, Messing J, Chen X: CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr Biol* 2002, 12:1484–1495

Park MY, Wu G, Gonzalez-Sulser A, Vaucheret H, Poethig RS: Nuclear processing and export of microRNAs in *Arabidopsis*. *Proc Natl Acad Sci USA* 2005, 102:3691–3696

Parkinson SE, Gross SM, Hollick JB: Maize sex determination and abaxial leaf fates are canalized by a factor that maintains repressed epigenetic states. *Dev Biol* 2007, 15;308(2):462-73

Pei L, Jin Z, Li K, Yin H, Wang J, Yang A: Identification and comparative analysis of low phosphate tolerance-associated microRNAs in two maize genotypes. *Plant Physiol Biochem* 2013, 70:221-34

Pélissier T, Thalmeir S, Kempe D, Sänger HL, Wassenegger M: Heavy *de novo* methylation at symmetrical and non-symmetrical sites is a hallmark of RNA-directed DNA methylation. *Nucleic Acids Res* 1999, 27, 1625–1634

Penterman J, Uzawa R, Fischer RL: Genetic interactions between DNA demethylation and methylation in *Arabidopsis*. *Plant Physiol* 2007, 145:1549 – 1557

Pikaard CS, Haag JR, Ream T, Wierzbicki AT: Roles of RNA polymerase IV in gene silencing. *Trends Plant Sci* 2008, 13(7):390-7

Piriyapongsa J, Jordan IK: Dual coding of siRNAs and miRNAs by plant transposable elements. *RNA* 2008, 14:814-821

Pontier D, Yahubyan G, Vega D, Bulski A, Saez-Vasquez J, Hakimi MA, Lerbs-Mache S, Colot V, Lagrange T: Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. *Genes Dev* 2005, 1;19(17):2030-40

Pradervand S, Weber J, Thomas J, Bueno M, Wirapati P, Lefort K, Dotto GP, Harshman K: Impact of normalization on miRNA microarray expression profiling. *RNA* 2009, 15: 493–501

Qi Y, He X, Wang X-J, Kohany O, Jurka J, Hannon GJ: Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* 2006, 443:1008–12

Qi X, Xie S, Liu Y, Yi F, Yu J: Genome-wide annotation of genes and noncoding RNAs of foxtail millet in response to simulated drought stress by deep sequencing. *Plant Mol Biol* 2013, 83:459–473

Quinlan AR, Hall IM: BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010, 26: 841–842

R Development Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013, ISBN 3-900051-07-0, URL http://www.R-project.org/

Ramachandran V, Chen X: Degradation of microRNAs by a family of exoribonucleases in *Arabidopsis*. *Science* 2008, 321(5895): 1490–1492

Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP: MicroRNAs in plants. *Genes Dev* 2002, 16:1616–1626

Robinson MD, McCarthy DJ, Smyth GK: edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010, 26: 139–140

Robinson MD, Oshlack A: A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 2010, 11:R25

Rodrigues JA, Ruan R, Nishimura T, Sharma MK, Sharma R, Ronald PC, Fischer RL, Zilberman D: Imprinted expression of genes and small RNA is associated with localized hypomethylation of the maternal genome in rice endosperm. *Proc Natl Acad Sci USA* 2013, 110(19):7934-7939

Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, Urich MA, Libiger O, Schork NJ, Ecker JR: Transgenerational epigenetic instability is a source of novel methylation variants. *Science* 2011, 21;334(6054):369-73

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B *et* al: The B73 maize genome: complexity, diversity, and dynamics. *Science* 2009, 20;326(5956):1112-5

Searle IR, Pontes O, Melnyk CW, Smith LM, Baulcombe DC: JMJ14, a JmjC domain protein, is required for RNA silencing and cell-to-cell movement of an RNA silencing signal in *Arabidopsis. Genes Dev* 2010, 24, 986–991

Sen TZ, Harper LC, Schaeffer ML, Andorf CM, Seigfried T, Campbell DA, Lawrence CJ: Choosing a genome browser for a Model Organism Database: surveying the Maize community. *Database* 2010, 2010:baq007

Slotkin RK, Freeling M, Lisch D: Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. *Nat Genet* 2005, 37: 641–644

Smith LM, Pontes O, Searle I, Yelina N, Yousafzai FK, Herr AJ, Pikaard CS, Baulcombe DC: An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in *Arabidopsis. Plant Cell* 2007, 19, 1507–1521

Smyth GK, Yang YH, Speed TP: Statistical issues in microarray data analysis. *Methods Mol Biol* 2003, 224: 111–136

Song X, Li P, Zhai J, Zhou M, Ma L, Liu B, Jeong DH, Nakano M, Cao S, Liu C, Chu C, Wang XJ, Green PJ, Meyers BC, Cao X: Roles of DCL4 and DCL3b in rice phased small RNA biogenesis. *Plant J* 2012, 69 (3), 462-474

Springer NM: Epigenetics and crop improvement. *Trends Genet* 2013, 29(4):241-7

Sridhar VV, Kapoor A, Zhang K, Zhu J, Zhou T, Hasegawa PM, Bressan RA, Zhu JK: Control of DNA methylation and heterochromatic silencing by histone H2B deubiquitination. *Nature* 2007, 447, 735–738

Stam M, Belele C, Dorweiler JE, Chandler VL: Differential chromatin structure within a tandem array 100 kb upstream of the maize *b1* locus is associated with paramutation. *Genes Dev* 2002, 1;16(15):1906-18

Stonaker JL, Lim JP, Erhard KF Jr, Hollick JB: Diversity of Pol IV function is defined by mutations at the maize *rmr7* locus. *PLoS Genet* 2009, 5(11):e1000706

Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, Patel DJ, Jacobsen SE: Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. *Nat Struct Mol Biol* 2014, 21, 64–72

Sunkar R, Zhu JK: Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. *Plant Cell* 2004, 16, 2001–2019

Sunkar R, Girke T, Jain PK, Zhu J: Cloning and characterization of microRNAs from rice. *Plant Cell* 2005, 17(5): 1397–1411

Sunkar R, Kapoor A, Zhu J: Posttranscriptional induction of two Cu/Zn superoxide dismutase genes in *Arabidopsis* is mediated by downregulation of miR398 and important for oxidative stress tolerance. *Plant Cell* 2006, 18(8):2051–2065

Sunkar R, Li YF, Jagadeeswaran G: Functions of microRNAs in plant stress responses. *Trends Plant Sci* 2012, (4), 196-203

Sutani T, Yuasa T, Tomonaga T, Dohmae N, Takio K, Yanagida M: Fission yeast condensin complex: essential roles of non-SMC subunits for condensation and Cdc2 phosphorylation of Cut3/SMC4. *Genes Dev* 1999, 1;13(17):2271-83

Taslim C, Wu J, Yan P, Singer G, Parvin J, Huang T, Lin S, Huang K: Comparative study on ChIP-seq data: normalization and binding pattern characterization. *Bioinformatic*s 2009, 25:2334-2340

The *Arabidopsis* Genome Initiative: Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 2000, 408**:** 796–815

To TK, Kim JM, Matsui A, Kurihara Y, Morosawa T, Ishida J, Tanaka M, Endo T, Kakutani T, Toyoda T, Kimura H, Yokoyama S, Shinozaki K, Seki M: *Arabidopsis* HDA6 regulates locus-directed heterochromatin silencing in cooperation with MET1. *PLoS Genet* 2011, 7, e1002055

Tricker PJ, Gibbings JG, Rodríguez López CM, Hadley P, Wilkinson MJ: Low relative humidity triggers RNA-directed *de novo* DNA methylation and suppression of genes controlling stomatal development. *J Exp Bot* 2012, 63(10):3799-3813

Tuteja N, Tran NQ, Dang HQ, Tuteja R: Plant MCM proteins: role in DNA replication and beyond. *Plant Mol Biol* 2011, 77(6):537-45

Usami T, Horiguchi G, Yano S, Tsukaya H: The more and smaller cells mutants of *Arabidopsis* thaliana identify novel roles for SQUAMOSA PROMOTER BINDING PROTEIN-LIKE genes in the control of heteroblasty. *Development* 2009, 136:955-964

Vaucheret H, Vazquez F, Crété P, Bartel DP: The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. *Genes & Development* 2004, 18: 1187–1197

Voinnet O: Origin, biogenesis, and activity of plant microRNAs. *Cell* 2009, 136(4):669-87

Wang X, Elling AA, Li X, Li N, Peng Z, He G, Sun H, Qi Y, Liu XS, Deng XW: Genome-wide and organ-specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in maize. *Plant Cell* 2009, 21(4):1053–1069

Wang L, Gu X, Xu D, Wang W, Wang H, Zeng M, Chang Z, Huang H, Cui X: miR396-targeted AtGRF transcription factors are required for coordination of cell division and differentiation during leaf development in *Arabidopsis*. *J Exp Bot* 2011, 62(2):761–773

Wang YG, An M, Zhou SF, She YH, Li WC, Fu FL: Expression profile of maize microRNAs corresponding to their target genes under drought stress. *Biochem Genet* 2014a, 52:474–493

Wang H-LV, Dinwiddie BL, Lee H, Chekanova JA: Stress-induced endogenous siRNAs targeting regulatory intron sequences in *Brachypodium. RNA* 2014b

Wei L, Zhang D, Xiang F, Zhang Z: Differentially expressed miRNAs potentially involved in the regulation of defense mechanism to drought stress in maize seedlings. *Int J Plant Sci* 2009, 170(8):979–989

Wei L, Gu L, Song X, Cui X, Lu Z, Zhou M, Wang L, Hu F, Zhai J, Meyers BC, Cao X: Dicer-like 3 produces transposable element-associated 24-nt siRNAs that control agricultural traits in rice. *Proc Natl Acad Sci USA* 2014, 11;111(10):3877-82

Wierzbicki AT, Ream TS, Haag JR, Pikaard CS: RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nature Genet* 2009, 41, 630–634

Wierzbicki AT, Cocklin R, Mayampurath A, Lister R, Rowley MJ, Gregory BD, Ecker JR, Tang H, Pikaard CS: Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the *Arabidopsis* epigenome. *Genes Dev* 2012, 26:1825–36

Woodhouse MR, Freeling M, Lisch D: The *mop1 (mediator of paramutation1)* mutant progressively reactivates one of the two genes encoded by the *MuDR* transposon in maize. Genetics 2006, 172(1):579-92

Wu L, Zhou H, Zhang Q, Zhang J, Ni F, Liu C, Qi Y: DNA methylation mediated by a microRNA pathway. *Mol Cell* 2010, 38:465–75

Xie Z, Allen E, Fahlgren N, Calamar A, Givan SA, Carrington JC: Expression of *Arabidopsis MIRNA* genes. *Plant Physiol* 2005, 138:2145-2154

Xin M, Yang R, Yao Y, Ma C, Peng H, Sun Q, Wang X, Ni Z: Dynamic parent-of-origin effects on small interfering RNA expression in the developing maize endosperm. *BMC Plant Biology* 2014, 14:192

Xu Z, Zhong S, Li X, Li W, Rothstein SJ, Zhang S, Bi Y, Xie C: Genome-wide identification of microRNAs in response to low nitrate availability in maize leaves and roots. *PLoS One* 2011, 6(11):e28009

Ye R, Wang W, Iki T, Liu C, Wu Y, Ishikawa M, Zhou X, Qi Y: Cytoplasmic assembly and selective nuclear import of *Arabidopsis* ARGONAUTE4/siRNA complexes. *Mol Cell* 2012, 46:859–70

Yoder JA, Walsh CP, Bestor TH: Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* 1997, 13(8):335-40

Yoshikawa M: Biogenesis of *trans*-acting siRNAs, endogenous secondary siRNAs in plants. *Genes Genet Syst* 2013, 88(2):77-84

Yu B, Yang Z, Li J, Minakhina S, Yang M, Padgett RW, Steward R, Chen X: Methylation as a crucial step in plant microRNA biogenesis. *Science* 2005, 307:932–5

Yu B, Bi L, Zheng B, Ji L, Chevalier D, Agarwal M, Ramachandran V, Li W, Lagrange T, Walker JC, Chen X: The FHA domain proteins DAWDLE in *Arabidopsis* and SNIP1 in humans act in small RNA biogenesis. *Proc Natl Acad Sci USA* 2008, 105, 10073–10078

Yu A, Lepère G, Jay F, Wang J, Bapaume L, Wang Y, Abraham A-L, Penterman J, Fischer RL, Voinnet O, Navarro L: Dynamics and biological relevance of DNA demethylation in *Arabidopsis* antibacterial defense. *Proc Natl Acad Sci USA* 2013, 110, 2389–94

Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D: The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* 2013, 153:193-205

Zhai J, Zhao Y, Simon SA, Huang S, Petsch K, Arikit S, Pillay M, Ji L, Xie M, Cao X, Yu B, Timmermans M, Yang B, Chen X, Meyers BC: Plant microRNAs display differential 3' truncation and tailing modifications that are ARGONAUTE1 dependent and conserved across species. *Plant Cell* 2013, 25(7):2417-2428

Zhang X, Henderson IR, Lu C, Green PJ, Jacobsen SE: Role of RNA polymerase IV in plant small RNA metabolism. *Proc Natl Acad Sci USA* 2007, 104, 4536–4541

Zhang X, Liu S, Takano T: Two cysteine proteinase inhibitors from *Arabidopsis thaliana*, AtCYSa and AtCYSb, increasing the salt, drought, oxidation and cold tolerance. *Plant Mol Biol* 2008, 68(1-2):131-43

Zhang Z, Wei L, Zou X, Tao Y, Liu Z, Zheng Y: Submergence-responsive microRNAs are potentially involved in the regulation of morphological and metabolic adaptations in maize root cells. *Ann Bot* 2008, 102(4):509-519

Zhang L, Chia J-M, Kumari S, Stein JC, Liu Z, Narechania A, Maher CA, Guill K, McMullen MD, Ware D: A genome-wide characterization of microRNA genes in maize. *PLoS Genet* 2009, 5:e1000716

Zhang X, Zhao H, Gao S, Wang WC, Katiyar-Agarwal S, Huang H-D, Raikhel N, Jin H: *Arabidopsis* Argonaute 2 regulates innate immunity via miRNA393*-mediated silencing of a Golgi-localized SNARE gene, MEMB12. *Mol Cell* 2011, 42: 356–366

Zhang H, Ma Z-Y, Zeng L, Tanaka K, Zhang CJ, Ma J, Bai G, Wang P, Zhang S-W, Liu Z-W, Cai T, Tang K, Liu R, Shi X, He X-J, Zhu J-K: DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV. *Proc Natl Acad Sci U.S.A.* 2013, 110, 8290–8295

Zhang W, Han Z, Guo Q, Liu Y, Zheng Y, Wu F, Jin W: Identification of maize long non-coding RNAs responsive to drought stress. *PLoS ONE* 2014, 9(6):e98958

Zhao Y, Yu Y, Zhai J, Ramachandran V, Dinh TT, Meyers BC, Mo B, Chen X: The *Arabidopsis* nucleotidyl transferase HESO1 uridylates unmethylated small RNAs to trigger their degradation. *Curr Biol* 2012, 22: 689–694

Zheng B, Wang Z, Li S, Yu B, Liu JY, Chen X: Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in *Arabidopsis. Genes Dev* 2009, 23, 2850–2860

Zhong X, Hale CJ, Law JA, Johnson LM, Feng S, Andy Tu, Jacobsen SE: DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. *Nat Struct Mol Biol* 2012, 19:870–75

Zhou L, Liu Y, Liu Z, Kong D, Duan M, Luo L: Genome-wide identification and analysis of drought-responsive microRNAs in *Oryza sativa*. *J Exp Bot* 2010, 61, 4157–4168

Zhu Y, Rowley MJ, Bohmdorfer G, Wierzbicki AT: A SWI/SNF chromatin-remodeling complex acts in noncoding RNA-mediated transcriptional silencing. *Mol Cell* 2013, 49, 298–309

Appendix A

Evaluation of sRNA loci size class consistency across individual libraries

C = control; D = drought stress; S = salinity stress; D+S = drought+salinity stress. +7=seven days of recovery.
R1, R2, R3 = biological replicate 1,2,3.

sRNA loci category	parameter*	wt C R1	wt C R2	wt C R3	wt D R1	wt D R2	wt D R3	wt S R1	wt S R2	wt S R3	wt D+S R1	wt D+S R2	wt D+S R3
20-nt MIRNA	average	0.75	0.82	0.70	0.83	0.91	0.83	0.75	0.86	0.78	0.87	0.87	0.83
	stdev of average	0.27	0.22	0.31	0.19	0.15	0.16	0.32	0.18	0.22	0.17	0.17	0.17
	1° percentile	0.55	0.83	0.50	0.82	0.91	0.80	0.74	0.83	0.80	0.85	0.89	0.80
	2° percentile	0.90	0.89	0.89	0.91	0.94	0.85	0.88	0.89	0.86	0.91	0.93	0.84
	3° percentile	0.92	0.95	0.91	0.93	1.00	0.94	0.92	1.00	0.89	0.98	0.94	0.90
20-nt HP	average	0.55	0.51	0.51	0.64	0.47	0.34	0.60	0.52	0.64	0.48	0.63	0.61
	stdev of average	0.36	0.35	0.39	0.39	0.40	0.35	0.36	0.44	0.39	0.44	0.41	0.39
	1° percentile	0.32	0.38	0.33	0.31	0.00	0.00	0.38	0.10	0.51	0.00	0.38	0.43
	2° percentile	0.50	0.50	0.50	0.75	0.54	0.40	0.50	0.40	0.71	0.34	0.75	0.53
	3° percentile	0.83	0.70	1.00	1.00	0.69	0.56	1.00	1.00	1.00	1.00	1.00	1.00
20-nt non-HP	average	0.57	0.57	0.68	0.69	0.55	0.75	0.69	0.62	0.63	0.69	0.65	0.65
	stdev of average	0.36	0.33	0.38	0.37	0.45	0.31	0.40	0.38	0.39	0.34	0.41	0.40
	1° percentile	0.33	0.48	0.45	0.45	0.00	0.50	0.55	0.50	0.46	0.50	0.43	0.45
	2° percentile	0.63	0.56	0.86	0.86	0.72	0.87	0.87	0.67	0.67	0.78	0.90	0.78
	3° percentile	0.92	0.82	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
21-nt MIRNA	average	0.79	0.80	0.80	0.77	0.77	0.81	0.84	0.77	0.77	0.83	0.74	0.75
	stdev of average	0.23	0.25	0.21	0.26	0.28	0.21	0.19	0.28	0.23	0.20	0.31	0.28
	1° percentile	0.68	0.72	0.67	0.60	0.65	0.67	0.73	0.67	0.67	0.73	0.58	0.60
	2° percentile	0.85	0.87	0.83	0.83	0.85	0.87	0.90	0.87	0.82	0.91	0.82	0.82
	3° percentile	0.96	0.99	0.98	0.98	0.99	1.00	0.99	1.00	0.94	0.99	0.99	0.99
21-nt HP	average	0.67	0.65	0.60	0.66	0.61	0.67	0.65	0.64	0.63	0.63	0.66	0.65
	stdev of average	0.38	0.40	0.39	0.36	0.42	0.40	0.37	0.40	0.39	0.36	0.39	0.39
	1° percentile	0.43	0.33	0.22	0.41	0.00	0.40	0.43	0.30	0.33	0.34	0.33	0.38
	2° percentile	0.85	0.83	0.72	0.79	0.82	0.83	0.84	0.81	0.75	0.73	0.84	0.78
	3° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
21-nt non-HP	average	0.54	0.52	0.50	0.49	0.49	0.53	0.51	0.51	0.50	0.50	0.47	0.49
	stdev of average	0.39	0.40	0.39	0.39	0.41	0.42	0.39	0.39	0.40	0.36	0.40	0.40
	1° percentile	0.14	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.21	0.00	0.00
	2° percentile	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.44	0.50
	3° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.83	1.00	1.00
22-nt MIRNA	average	0.59	0.67	0.55	0.57	0.65	0.73	0.58	0.61	0.62	0.65	0.62	0.60
	stdev of average	0.27	0.22	0.19	0.20	0.29	0.19	0.23	0.26	0.27	0.15	0.28	0.25
	1° percentile	0.45	0.52	0.47	0.48	0.55	0.55	0.52	0.49	0.50	0.55	0.49	0.53
	2° percentile	0.62	0.63	0.60	0.54	0.66	0.69	0.59	0.66	0.63	0.65	0.62	0.59
	3° percentile	0.69	0.79	0.64	0.65	0.81	0.91	0.64	0.71	0.77	0.69	0.75	0.67
22-nt HP	average	0.69	0.69	0.70	0.70	0.67	0.69	0.69	0.67	0.69	0.68	0.69	0.68
	stdev of average	0.35	0.36	0.35	0.36	0.40	0.38	0.36	0.38	0.37	0.32	0.37	0.38
	1° percentile	0.50	0.50	0.50	0.50	0.44	0.50	0.50	0.50	0.50	0.50	0.50	0.50
	2° percentile	0.79	0.83	0.82	0.85	1.00	0.88	0.83	0.83	0.90	0.75	0.85	0.83
	3° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
22-nt non-HP	average	0.78	0.77	0.77	0.77	0.77	0.77	0.77	0.76	0.76	0.76	0.77	0.77
	stdev of average	0.31	0.33	0.33	0.32	0.35	0.34	0.32	0.34	0.35	0.27	0.34	0.34
	1° percentile	0.67	0.62	0.62	0.67	0.63	0.62	0.64	0.56	0.58	0.67	0.60	0.62
	2° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.83	1.00	1.00
	3° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

sRNA loci category	parameter*	wt C +7 R1	wt C +7 R2	wt C +7 R3	wt D +7 R1	wt D +7 R2	wt D +7 R3	wt S +7 R1	wt S +7 R2	wt S +7 R3	wt D+S +7 R1	wt D+S +7 R2	wt D+S +7 R3
20-nt MIRNA	average	0.83	0.70	0.79	0.86	0.87	0.83	0.88	0.91	0.68	0.79	0.82	0.79
	stdev of average	0.21	0.32	0.21	0.20	0.19	0.19	0.17	0.18	0.31	0.30	0.20	0.36
	1° percentile	0.71	0.67	0.62	0.87	0.84	0.81	0.90	0.93	0.57	0.78	0.74	0.80
	2° percentile	0.93	0.78	0.90	0.93	0.93	0.88	0.92	0.96	0.80	0.90	0.80	0.97
	3° percentile	0.97	0.88	0.93	1.00	1.00	0.95	1.00	1.00	0.84	0.96	1.00	1.00
20-nt HP	average	0.57	0.36	0.49	0.70	0.56	0.58	0.64	0.70	0.67	0.52	0.34	0.58
	stdev of average	0.42	0.44	0.44	0.35	0.45	0.42	0.42	0.35	0.37	0.38	0.40	0.41
	1° percentile	0.08	0.00	0.00	0.33	0.00	0.22	0.31	0.50	0.51	0.29	0.00	0.33
	2° percentile	0.67	0.17	0.45	1.00	0.60	0.60	1.00	0.67	0.68	0.58	0.27	0.58
	3° percentile	1.00	0.73	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.75	0.50	1.00
20-nt non-HP	average	0.75	0.59	0.75	0.66	0.76	0.60	0.61	0.67	0.56	0.61	0.63	0.64
	stdev of average	0.34	0.41	0.31	0.37	0.32	0.34	0.37	0.40	0.44	0.42	0.38	0.37
	1° percentile	0.49	0.32	0.52	0.42	0.66	0.45	0.35	0.55	0.00	0.15	0.42	0.49
	2° percentile	1.00	0.57	0.83	0.80	0.93	0.65	0.51	0.80	0.75	0.77	0.73	0.67
	3° percentile	1.00	1.00	1.00	1.00	1.00	0.83	1.00	1.00	0.98	1.00	1.00	1.00
21-nt MIRNA	average	0.82	0.78	0.77	0.78	0.79	0.79	0.81	0.82	0.77	0.80	0.76	0.79
	stdev of average	0.20	0.28	0.28	0.25	0.23	0.22	0.22	0.22	0.26	0.24	0.26	0.24
	1° percentile	0.71	0.65	0.65	0.65	0.67	0.62	0.67	0.69	0.62	0.66	0.66	0.66
	2° percentile	0.89	0.90	0.84	0.83	0.86	0.83	0.90	0.92	0.83	0.89	0.84	0.85
	3° percentile	1.00	1.00	0.99	0.99	0.99	0.99	1.00	1.00	0.99	1.00	0.97	0.99
21-nt HP	average	0.63	0.67	0.62	0.67	0.65	0.66	0.61	0.63	0.60	0.70	0.66	0.66
	stdev of average	0.41	0.40	0.40	0.38	0.40	0.40	0.41	0.39	0.41	0.36	0.38	0.41
	1° percentile	0.20	0.38	0.33	0.36	0.33	0.33	0.14	0.33	0.09	0.50	0.42	0.33
	2° percentile	0.80	0.93	0.79	0.85	0.85	0.85	0.80	0.83	0.74	0.87	0.82	0.90
	3° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
21-nt non-HP	average	0.51	0.47	0.52	0.53	0.53	0.52	0.50	0.47	0.47	0.54	0.50	0.49
	stdev of average	0.40	0.43	0.40	0.41	0.41	0.41	0.41	0.41	0.40	0.41	0.40	0.42
	1° percentile	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2° percentile	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
	3° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
22-nt MIRNA	average	0.54	0.52	0.59	0.56	0.67	0.66	0.53	0.63	0.57	0.56	0.67	0.65
	stdev of average	0.34	0.27	0.25	0.20	0.17	0.25	0.24	0.24	0.25	0.21	0.21	0.12
	1° percentile	0.39	0.52	0.56	0.40	0.54	0.55	0.46	0.56	0.50	0.50	0.53	0.57
	2° percentile	0.63	0.58	0.68	0.68	0.64	0.63	0.50	0.66	0.60	0.56	0.69	0.62
	3° percentile	0.70	0.68	0.70	0.70	0.79	0.82	0.70	0.72	0.66	0.62	0.77	0.73
22-nt HP	average	0.67	0.71	0.70	0.71	0.69	0.68	0.69	0.68	0.72	0.68	0.68	0.68
	stdev of average	0.38	0.39	0.37	0.37	0.38	0.38	0.40	0.38	0.35	0.39	0.39	0.39
	1° percentile	0.48	0.50	0.50	0.50	0.50	0.50	0.37	0.50	0.50	0.44	0.50	0.48
	2° percentile	0.80	1.00	0.89	1.00	0.91	0.86	1.00	0.85	0.96	1.00	0.86	1.00
	3° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
22-nt non-HP	average	0.77	0.76	0.77	0.77	0.77	0.77	0.77	0.76	0.76	0.77	0.77	0.76
	stdev of average	0.33	0.36	0.34	0.35	0.35	0.35	0.36	0.35	0.34	0.36	0.34	0.35
	1° percentile	0.64	0.54	0.60	0.60	0.60	0.60	0.60	0.57	0.57	0.60	0.60	0.58
	2° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	3° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

sRNA loci category	parameter*	*rmr6-1* C R1	*rmr6-1* C R2	*rmr6-1* C R3	*rmr6-1* D R1	*rmr6-1* D R2	*rmr6-1* D R3	*rmr6-1* S R1	*rmr6-1* S R2	*rmr6-1* S R3	*rmr6-1* D+S R2	*rmr6-1* D+S R3
20-nt MIRNA	average	0.87	0.80	0.83	0.80	0.85	0.75	0.85	0.85	0.75	0.79	0.82
	stdev of average	0.18	0.20	0.18	0.22	0.19	0.30	0.20	0.16	0.31	0.20	0.20
	1° percentile	0.86	0.77	0.82	0.82	0.81	0.72	0.86	0.86	0.83	0.79	0.79
	2° percentile	0.93	0.85	0.90	0.85	0.87	0.86	0.92	0.89	0.85	0.85	0.91
	3° percentile	0.96	0.92	0.93	0.92	0.99	0.93	0.97	0.92	0.89	0.91	0.92
20-nt HP	average	0.58	0.59	0.54	0.43	0.49	0.56	0.78	0.68	0.54	0.60	0.46
	stdev of average	0.33	0.38	0.46	0.40	0.42	0.37	0.33	0.36	0.38	0.38	0.43
	1° percentile	0.36	0.34	0.13	0.00	0.00	0.35	0.43	0.49	0.34	0.38	0.00
	2° percentile	0.56	0.53	0.44	0.44	0.50	0.50	1.00	0.67	0.50	0.54	0.33
	3° percentile	0.88	1.00	1.00	0.73	1.00	1.00	1.00	1.00	0.94	1.00	1.00
20-nt non-HP	average	0.63	0.69	0.46	0.66	0.53	0.65	0.75	0.62	0.78	0.65	0.77
	stdev of average	0.43	0.34	0.36	0.43	0.42	0.43	0.34	0.42	0.31	0.43	0.30
	1° percentile	0.31	0.48	0.00	0.17	0.00	0.31	0.53	0.27	0.62	0.31	0.59
	2° percentile	0.84	0.75	0.50	0.89	0.50	0.84	0.96	0.79	0.93	0.86	0.91
	3° percentile	1.00	1.00	0.67	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
21-nt MIRNA	average	0.78	0.80	0.78	0.80	0.80	0.81	0.80	0.79	0.77	0.79	0.83
	stdev of average	0.27	0.22	0.27	0.26	0.24	0.22	0.26	0.27	0.24	0.26	0.21
	1° percentile	0.69	0.70	0.61	0.73	0.69	0.69	0.68	0.67	0.66	0.69	0.70
	2° percentile	0.86	0.86	0.88	0.89	0.89	0.88	0.90	0.89	0.82	0.88	0.89
	3° percentile	0.98	0.99	1.00	0.99	1.00	1.00	1.00	1.00	0.97	0.99	0.99
21-nt HP	average	0.73	0.68	0.73	0.72	0.76	0.73	0.72	0.70	0.72	0.71	0.72
	stdev of average	0.36	0.37	0.36	0.36	0.32	0.36	0.36	0.36	0.34	0.35	0.34
	1° percentile	0.50	0.46	0.50	0.50	0.51	0.50	0.50	0.42	0.50	0.50	0.50
	2° percentile	0.91	0.83	0.98	0.92	0.95	0.97	0.95	0.89	0.86	0.89	0.86
	3° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
21-nt non-HP	average	0.59	0.59	0.55	0.58	0.56	0.56	0.57	0.56	0.56	0.57	0.56
	stdev of average	0.37	0.38	0.39	0.38	0.39	0.40	0.38	0.40	0.39	0.38	0.38
	1° percentile	0.33	0.28	0.20	0.25	0.22	0.14	0.25	0.20	0.22	0.27	0.25
	2° percentile	0.56	0.60	0.50	0.56	0.50	0.56	0.55	0.54	0.51	0.56	0.50
	3° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
22-nt MIRNA	average	0.72	0.65	0.68	0.62	0.58	0.79	0.69	0.62	0.69	0.59	0.58
	stdev of average	0.16	0.26	0.14	0.31	0.22	0.16	0.15	0.33	0.17	0.32	0.34
	1° percentile	0.63	0.63	0.59	0.64	0.54	0.66	0.60	0.59	0.58	0.59	0.48
	2° percentile	0.72	0.68	0.66	0.67	0.62	0.68	0.62	0.65	0.60	0.63	0.62
	3° percentile	0.78	0.78	0.72	0.73	0.68	1.00	0.75	0.84	0.75	0.80	0.72
22-nt HP	average	0.72	0.71	0.71	0.72	0.72	0.69	0.73	0.73	0.72	0.72	0.73
	stdev of average	0.34	0.35	0.38	0.35	0.35	0.39	0.34	0.36	0.35	0.35	0.34
	1° percentile	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
	2° percentile	0.86	0.88	1.00	0.92	0.96	1.00	0.93	1.00	0.89	0.93	0.89
	3° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
22-nt non-HP	average	0.78	0.78	0.78	0.78	0.78	0.77	0.77	0.79	0.77	0.77	0.77
	stdev of average	0.31	0.32	0.35	0.32	0.32	0.34	0.32	0.32	0.34	0.32	0.32
	1° percentile	0.67	0.67	0.65	0.67	0.67	0.60	0.67	0.67	0.61	0.67	0.64
	2° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	3° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

sRNA loci category	parameter*	*rmr6-1* C +7 R1	*rmr6-1* C +7 R2	*rmr6-1* C +7 R3	*rmr6-1* D +7 R1	*rmr6-1* D +7 R2	*rmr6-1* D +7 R3	*rmr6-1* S +7 R2	*rmr6-1* S +7 R3	*rmr6-1* D+S +7 R1	*rmr6-1* D+S +7 R2	*rmr6-1* D+S +7 R3
20-nt MIRNA	average	0.71	0.84	0.84	0.68	0.79	0.84	0.82	0.86	0.73	0.85	0.85
	stdev of average	0.38	0.18	0.24	0.35	0.20	0.18	0.19	0.20	0.29	0.17	0.22
	1° percentile	0.61	0.85	0.78	0.48	0.76	0.85	0.83	0.85	0.72	0.84	0.89
	2° percentile	0.88	0.89	0.96	0.85	0.83	0.89	0.87	0.92	0.83	0.88	0.92
	3° percentile	0.96	0.91	1.00	0.89	0.88	0.94	0.89	0.98	0.86	0.95	1.00
20-nt HP	average	0.53	0.63	0.54	0.26	0.59	0.53	0.68	0.37	0.53	0.77	0.48
	stdev of average	0.35	0.42	0.40	0.33	0.38	0.36	0.37	0.40	0.41	0.34	0.41
	1° percentile	0.33	0.35	0.33	0.00	0.38	0.37	0.40	0.00	0.25	0.50	0.00
	2° percentile	0.50	0.83	0.50	0.27	0.58	0.48	0.83	0.29	0.54	1.00	0.48
	3° percentile	0.77	1.00	1.00	0.33	1.00	0.88	1.00	0.63	1.00	1.00	1.00
20-nt non-HP	average	0.55	0.53	0.56	0.66	0.80	0.69	0.53	0.62	0.67	0.61	0.57
	stdev of average	0.43	0.40	0.45	0.39	0.31	0.39	0.32	0.44	0.37	0.38	0.43
	1° percentile	0.06	0.06	0.00	0.48	0.72	0.48	0.45	0.19	0.40	0.36	0.04
	2° percentile	0.71	0.63	0.67	0.83	0.99	0.88	0.50	0.86	0.79	0.69	0.73
	3° percentile	0.99	0.91	1.00	1.00	1.00	1.00	0.71	1.00	1.00	1.00	0.98
21-nt MIRNA	average	0.82	0.80	0.79	0.80	0.82	0.79	0.81	0.76	0.80	0.81	0.76
	stdev of average	0.23	0.23	0.26	0.25	0.22	0.24	0.23	0.25	0.25	0.23	0.27
	1° percentile	0.71	0.68	0.68	0.67	0.71	0.67	0.67	0.62	0.74	0.71	0.66
	2° percentile	0.91	0.88	0.85	0.90	0.89	0.82	0.90	0.83	0.89	0.86	0.86
	3° percentile	0.98	0.99	1.00	1.00	1.00	0.99	1.00	0.98	1.00	0.99	0.98
21-nt HP	average	0.71	0.76	0.67	0.70	0.73	0.71	0.74	0.68	0.70	0.67	0.72
	stdev of average	0.34	0.32	0.38	0.36	0.34	0.38	0.34	0.38	0.36	0.38	0.35
	1° percentile	0.50	0.60	0.50	0.43	0.50	0.50	0.50	0.33	0.35	0.35	0.50
	2° percentile	0.83	0.97	0.86	0.92	0.93	0.96	0.93	0.88	0.87	0.84	0.92
	3° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
21-nt non-HP	average	0.61	0.56	0.55	0.60	0.59	0.51	0.55	0.49	0.61	0.53	0.59
	stdev of average	0.35	0.37	0.41	0.37	0.38	0.40	0.40	0.39	0.38	0.39	0.38
	1° percentile	0.39	0.27	0.03	0.32	0.25	0.00	0.15	0.00	0.33	0.14	0.30
	2° percentile	0.59	0.53	0.54	0.65	0.66	0.50	0.50	0.50	0.67	0.50	0.60
	3° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96	1.00	1.00	1.00
22-nt MIRNA	average	0.58	0.67	0.58	0.51	0.60	0.69	0.53	0.53	0.60	0.75	0.63
	stdev of average	0.28	0.20	0.14	0.30	0.30	0.20	0.33	0.25	0.11	0.17	0.17
	1° percentile	0.48	0.54	0.48	0.58	0.51	0.63	0.37	0.51	0.56	0.65	0.54
	2° percentile	0.68	0.65	0.61	0.63	0.68	0.65	0.59	0.56	0.63	0.71	0.63
	3° percentile	0.71	0.77	0.66	0.67	0.69	0.76	0.78	0.66	0.67	0.88	0.70
22-nt HP	average	0.72	0.72	0.74	0.70	0.74	0.71	0.71	0.70	0.70	0.70	0.71
	stdev of average	0.34	0.35	0.36	0.38	0.35	0.36	0.37	0.37	0.37	0.36	0.37
	1° percentile	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
	2° percentile	0.83	0.92	1.00	1.00	1.00	0.95	1.00	1.00	0.97	0.88	1.00
	3° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
22-nt non-HP	average	0.78	0.78	0.78	0.77	0.78	0.78	0.77	0.78	0.78	0.77	0.77
	stdev of average	0.30	0.33	0.35	0.35	0.33	0.34	0.34	0.35	0.33	0.33	0.34
	1° percentile	0.67	0.67	0.67	0.62	0.67	0.66	0.61	0.65	0.67	0.67	0.60
	2° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	3° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

sRNA loci category	parameter*	wt C R1	wt C R2	wt C R3	wt D R1	wt D R2	wt D R3	wt S R1	wt S R2	wt S R3	wt D+S R1	wt D+S R2	wt D+S R3
23-nt MIRNA	average	0.70	0.73	0.73	0.72	0.72	0.78	0.74	0.71	0.66	0.78	0.73	0.76
	stdev of average	0.20	0.16	0.17	0.22	0.12	0.23	0.16	0.17	0.13	0.16	0.17	0.21
	1° percentile	0.63	0.68	0.67	0.64	0.68	0.70	0.69	0.65	0.62	0.73	0.67	0.68
	2° percentile	0.70	0.73	0.73	0.72	0.72	0.78	0.74	0.71	0.66	0.78	0.73	0.76
	3° percentile	0.77	0.79	0.79	0.80	0.77	0.86	0.80	0.77	0.71	0.84	0.78	0.83
23-nt HP	average	0.52	0.51	0.51	0.54	0.48	0.49	0.47	0.56	0.57	0.52	0.54	0.48
	stdev of average	0.33	0.36	0.38	0.35	0.39	0.37	0.39	0.38	0.41	0.30	0.39	0.40
	1° percentile	0.36	0.25	0.18	0.33	0.00	0.21	0.00	0.25	0.13	0.33	0.12	0.00
	2° percentile	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.60	0.60	0.52	0.57	0.45
	3° percentile	0.67	0.84	1.00	1.00	1.00	0.95	0.81	1.00	1.00	0.69	1.00	1.00
23-nt non-HP	average	0.54	0.51	0.55	0.56	0.54	0.48	0.52	0.44	0.54	0.56	0.55	0.50
	stdev of average	0.35	0.41	0.40	0.38	0.44	0.42	0.40	0.40	0.39	0.30	0.41	0.40
	1° percentile	0.33	0.00	0.17	0.25	0.00	0.00	0.00	0.00	0.00	0.40	0.00	0.00
	2° percentile	0.57	0.50	0.50	0.57	0.50	0.50	0.50	0.50	0.50	0.56	0.50	0.50
	3° percentile	0.80	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.75	1.00	1.00
24-nt MIRNA	average	0.86	0.81	0.80	0.87	0.85	0.80	0.83	0.83	0.85	0.81	0.74	0.82
	stdev of average	0.19	0.29	0.20	0.23	0.29	0.30	0.23	0.22	0.21	0.26	0.32	0.20
	1° percentile	0.71	0.60	0.67	0.84	0.81	0.71	0.68	0.76	0.83	0.79	0.50	0.71
	2° percentile	0.95	1.00	0.80	0.94	0.98	0.88	0.93	0.91	0.89	0.90	0.88	0.87
	3° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.93	1.00	1.00
24-nt HP	average	0.85	0.85	0.85	0.86	0.86	0.86	0.85	0.86	0.85	0.86	0.86	0.87
	stdev of average	0.23	0.25	0.24	0.23	0.27	0.26	0.24	0.25	0.27	0.17	0.25	0.24
	1° percentile	0.75	0.77	0.76	0.80	0.83	0.83	0.78	0.80	0.75	0.79	0.81	0.80
	2° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.91	1.00	1.00
	3° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
24-nt non-HP	average	0.83	0.84	0.84	0.84	0.85	0.85	0.84	0.85	0.83	0.85	0.85	0.85
	stdev of average	0.21	0.23	0.22	0.21	0.25	0.24	0.22	0.23	0.24	0.16	0.24	0.22
	1° percentile	0.75	0.75	0.75	0.76	0.79	0.79	0.75	0.76	0.75	0.79	0.77	0.78
	2° percentile	0.89	0.94	0.91	0.91	1.00	1.00	0.91	1.00	0.94	0.88	1.00	1.00
	3° percentile	0.70	0.73	0.73	0.72	0.72	0.78	0.74	0.71	0.66	0.78	0.73	0.76

sRNA loci category	parameter*	wt C +7 R1	wt C +7 R2	wt C +7 R3	wt D +7 R1	wt D +7 R2	wt D +7 R3	wt S +7 R1	wt S +7 R2	wt S +7 R3	wt D+S +7 R1	wt D+S +7 R2	wt D+S +7 R3
23-nt MIRNA	average	0.70	0.61	0.72	0.72	0.73	0.71	0.72	0.75	0.69	0.76	0.72	0.76
	stdev of average	0.07	0.01	0.18	0.20	0.16	0.15	0.12	0.13	0.14	0.26	0.16	0.18
	1° percentile	0.67	0.60	0.66	0.65	0.67	0.66	0.68	0.70	0.64	0.67	0.66	0.70
	2° percentile	0.70	0.61	0.72	0.72	0.73	0.71	0.72	0.75	0.69	0.76	0.72	0.76
	3° percentile	0.72	0.61	0.79	0.79	0.79	0.76	0.77	0.79	0.74	0.86	0.78	0.83
23-nt HP	average	0.52	0.59	0.58	0.47	0.47	0.52	0.46	0.54	0.50	0.55	0.47	0.60
	stdev of average	0.40	0.38	0.39	0.39	0.39	0.41	0.43	0.40	0.37	0.40	0.37	0.39
	1° percentile	0.00	0.33	0.33	0.00	0.00	0.00	0.00	0.08	0.22	0.23	0.00	0.33
	2° percentile	0.50	0.56	0.59	0.44	0.47	0.53	0.44	0.55	0.50	0.50	0.50	0.53
	3° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.80	1.00
23-nt non-HP	average	0.55	0.56	0.52	0.56	0.54	0.50	0.55	0.51	0.49	0.58	0.50	0.55
	stdev of average	0.41	0.44	0.43	0.41	0.42	0.41	0.41	0.40	0.41	0.41	0.42	0.42
	1° percentile	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2° percentile	0.58	0.58	0.50	0.50	0.50	0.50	0.60	0.50	0.50	0.60	0.50	0.50
	3° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
24-nt MIRNA	average	0.81	0.89	0.91	0.88	0.78	0.72	0.84	0.78	0.88	0.83	0.93	0.86
	stdev of average	0.21	0.23	0.17	0.19	0.35	0.33	0.23	0.28	0.15	0.30	0.13	0.22
	1° percentile	0.75	0.84	0.87	0.83	0.75	0.60	0.72	0.67	0.80	0.78	0.90	0.76
	2° percentile	0.86	1.00	1.00	1.00	0.90	0.80	0.86	0.86	0.90	1.00	1.00	1.00
	3° percentile	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
24-nt HP	average	0.85	0.85	0.85	0.84	0.85	0.86	0.85	0.86	0.85	0.84	0.85	0.86
	stdev of average	0.27	0.29	0.26	0.28	0.27	0.27	0.29	0.25	0.25	0.29	0.27	0.27
	1° percentile	0.80	0.83	0.80	0.79	0.80	0.80	0.81	0.80	0.78	0.80	0.80	0.80
	2° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	3° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
24-nt non-HP	average	0.84	0.84	0.84	0.83	0.84	0.84	0.84	0.85	0.84	0.83	0.84	0.84
	stdev of average	0.24	0.28	0.25	0.27	0.25	0.25	0.27	0.24	0.23	0.27	0.25	0.26
	1° percentile	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.78	0.75	0.75	0.75	0.75
	2° percentile	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.93	1.00	1.00	1.00
	3° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

sRNA loci category	parameter*	*rmr6-1* C R1	*rmr6-1* C R2	*rmr6-1* C R3	*rmr6-1* D R1	*rmr6-1* D R2	*rmr6-1* D R3	*rmr6-1* S R1	*rmr6-1* S R2	*rmr6-1* S R3	*rmr6-1* D+S R2	*rmr6-1* D +S R3
23-nt MIRNA	average	0.75	0.72	0.76	0.76	0.70	0.74	0.70	0.76	0.75	0.72	0.73
	stdev of average	0.19	0.17	0.19	0.14	0.12	0.17	0.15	0.19	0.16	0.17	0.16
	1° percentile	0.69	0.66	0.69	0.71	0.66	0.68	0.65	0.69	0.69	0.66	0.68
	2° percentile	0.75	0.72	0.76	0.76	0.70	0.74	0.70	0.76	0.75	0.72	0.73
	3° percentile	0.82	0.78	0.83	0.81	0.74	0.80	0.76	0.82	0.81	0.77	0.79
23-nt HP	average	0.52	0.48	0.55	0.45	0.52	0.50	0.55	0.52	0.54	0.54	0.60
	stdev of average	0.35	0.40	0.41	0.40	0.40	0.41	0.39	0.39	0.39	0.41	0.39
	1° percentile	0.33	0.00	0.16	0.00	0.08	0.00	0.25	0.06	0.29	0.00	0.25
	2° percentile	0.50	0.49	0.50	0.44	0.50	0.50	0.50	0.50	0.50	0.50	0.67
	3° percentile	0.83	1.00	1.00	0.89	1.00	1.00	1.00	1.00	1.00	1.00	1.00
23-nt non-HP	average	0.59	0.50	0.49	0.55	0.52	0.48	0.45	0.45	0.50	0.52	0.57
	stdev of average	0.39	0.43	0.42	0.38	0.42	0.42	0.41	0.42	0.42	0.43	0.42
	1° percentile	0.33	0.00	0.00	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2° percentile	0.67	0.50	0.50	0.56	0.50	0.50	0.46	0.45	0.50	0.50	0.54
	3° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
24-nt MIRNA	average	0.44	0.41	0.30	0.77	0.70	0.26	0.69	0.24	0.46	0.32	0.80
	stdev of average	0.50	0.45	0.45	0.33	0.33	0.43	0.25	0.38	0.44	0.31	0.30
	1° percentile	0.00	0.00	0.00	0.44	0.41	0.00	0.47	0.00	0.25	0.00	0.67
	2° percentile	0.47	0.43	0.30	0.77	0.70	0.26	0.68	0.17	0.44	0.32	0.80
	3° percentile	1.00	0.71	0.50	1.00	1.00	0.30	0.78	0.33	0.86	0.50	1.00
24-nt HP	average	0.35	0.30	0.26	0.32	0.31	0.32	0.28	0.32	0.31	0.29	0.34
	stdev of average	0.44	0.42	0.40	0.44	0.43	0.44	0.42	0.43	0.43	0.42	0.44
	1° percentile	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2° percentile	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	3° percentile	1.00	0.75	0.50	1.00	0.79	1.00	0.67	0.80	0.78	0.67	1.00
24-nt non-HP	average	0.23	0.20	0.17	0.21	0.25	0.21	0.21	0.22	0.20	0.23	0.20
	stdev of average	0.40	0.39	0.36	0.39	0.41	0.40	0.39	0.40	0.38	0.41	0.38
	1° percentile	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2° percentile	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	3° percentile	0.33	0.00	0.00	0.01	0.50	0.00	0.00	0.20	0.00	0.33	0.00

sRNA loci category	parameter*	*rmr6-1* C +7 R1	*rmr6-1* C +7 R2	*rmr6-1* C +7 R3	*rmr6-1* D +7 R1	*rmr6-1* D +7 R2	*rmr6-1* D +7 R3	*rmr6-1* S +7 R2	*rmr6-1* S +7 R3	*rmr6-1* D+S +7 R1	*rmr6-1* D+S +7 R2	*rmr6-1* D+S +7 R3
23-nt MIRNA	average	0.75	0.70	0.74	0.68	0.71	0.74	0.71	0.77	0.76	0.78	0.74
	stdev of average	0.18	0.06	0.19	0.11	0.20	0.20	0.16	0.25	0.20	0.15	0.22
	1° percentile	0.68	0.68	0.67	0.64	0.64	0.66	0.65	0.68	0.69	0.72	0.66
	2° percentile	0.75	0.70	0.74	0.68	0.71	0.74	0.71	0.77	0.76	0.78	0.74
	3° percentile	0.81	0.72	0.81	0.71	0.78	0.81	0.76	0.86	0.83	0.83	0.82
23-nt HP	average	0.57	0.48	0.54	0.60	0.58	0.52	0.57	0.62	0.50	0.47	0.49
	stdev of average	0.38	0.41	0.44	0.41	0.38	0.45	0.40	0.42	0.39	0.39	0.42
	1° percentile	0.28	0.00	0.00	0.26	0.33	0.00	0.13	0.22	0.10	0.00	0.00
	2° percentile	0.50	0.50	0.67	0.69	0.51	0.50	0.67	0.89	0.50	0.50	0.50
	3° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.94	1.00
23-nt non-HP	average	0.55	0.54	0.43	0.48	0.53	0.48	0.51	0.47	0.56	0.53	0.48
	stdev of average	0.39	0.42	0.44	0.41	0.42	0.43	0.43	0.42	0.41	0.42	0.42
	1° percentile	0.16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2° percentile	0.50	0.58	0.38	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
	3° percentile	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
24-nt MIRNA	average	0.19	0.40	0.50	0.71	0.50	0.54	0.69	0.25	0.26	0.31	0.64
	stdev of average	0.27	0.55	0.53	0.48	0.55	0.42	0.46	0.29	0.37	0.47	0.37
	1° percentile	0.00	0.00	0.00	0.54	0.00	0.44	0.44	0.06	0.00	0.00	0.33
	2° percentile	0.19	0.40	0.52	0.79	0.52	0.52	0.69	0.27	0.27	0.20	0.64
	3° percentile	0.33	1.00	1.00	1.00	1.00	0.75	1.00	0.50	0.32	0.60	1.00
24-nt HP	average	0.32	0.35	0.28	0.36	0.28	0.31	0.30	0.29	0.40	0.36	0.22
	stdev of average	0.43	0.45	0.41	0.46	0.42	0.44	0.43	0.43	0.46	0.45	0.38
	1° percentile	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2° percentile	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	3° percentile	0.86	1.00	0.62	1.00	0.67	0.84	0.75	0.75	1.00	1.00	0.45
24-nt non-HP	average	0.25	0.25	0.18	0.22	0.19	0.19	0.19	0.20	0.25	0.23	0.17
	stdev of average	0.41	0.42	0.37	0.40	0.37	0.38	0.38	0.39	0.41	0.41	0.36
	1° percentile	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2° percentile	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	3° percentile	0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.33	0.00

*parameter = the statistical parameter calculated on the total values of fraction (fraction of mapping reads with length equal to the size class assigned to a locus) of a sRNA loci category in a library:
"average" = indicates the mean of the fraction values;
"stdev of average" = indicates the standard deviation of the mean;
"1° percentile" = indicates the first percentile, 0.25, of the fraction values;
"2° percentile" = indicates the second percentile, 0.5, the median, of the fraction values;
"3° percentile" = indicates the third percentile, 0.75, of the fraction values.

Appendix B

List of phased loci and related overlapping genes

phased locus											
phased locus	**chromosome**	**start**	**end**	**strand**	**phasing p-value***	**phasing FDR****	**nt size class**	**precursor**	**repeats masking**		**name*****
Cluster_7064	10	85326130	85326482	-	0.000305926	0.012695798	21				*TAS3d* tasiRNA locus
Cluster_17857	9	35933176	35933260	+	0.002227397	0.043527031	21	HP	RLG		
Cluster_25484	9	134543069	134543433	.	0.000373791	0.013622614	21				*TAS3b* tasiRNA locus
Cluster_35583	6	93489286	93489393	-	0.002228688	0.043527031	21				*zma-MIR167h*
Cluster_46019	8	10589182	10589365	+	0.002545566	0.046385862	21	HP			*zma-MIR159b*
Cluster_55047	8	129743949	129744157	+	4.49E-06	0.000736867	21				*TAS3a* tasiRNA locus
Cluster_76570	5	2764087	2764309	+	0.002226523	0.043527031	21	*MIRNA*			*zma-MIR390b*
Cluster_79385	5	22325565	22326073	+	0.000274126	0.012695798	21				
Cluster_81441	5	43049564	43049727	-	0.002804241	0.04841005	21	*MIRNA*			*zma-MIR160b*
Cluster_91725	5	183551265	183551479	.	1.74E-07	5.70E-05	21				
Cluster_91727	5	183567337	183567684	.	3.42E-05	0.003077335	21				
Cluster_94495	5	205836953	205837078	-	0.001107017	0.036310146	21	HP	MITE		
Cluster_97736	3	9947847	9947974	.	0.001304655	0.038902454	21		MITE		
Cluster_99405	3	25492207	25492486	-	0.001617051	0.043527031	21	*MIRNA*			*zma-MIR159f*
Cluster_119657	2	16353049	16353216	-	0.000276555	0.012695798	21	*MIRNA*			*zma-MIR399e*
Cluster_137435	2	221277550	221277817	+	0.002255974	0.043527031	21	HP	DTC		
Cluster_138945	2	232386803	232387014	-	0.000309654	0.012695798	21	*MIRNA*			
Cluster_149370	4	136069834	136070031	.	3.75E-05	0.003077335	21				
Cluster_166266	1	51291875	51292165	.	0.001823115	0.043527031	21				*TAS3c* tasiRNA locus

phased locus	overlapping gene overlapping gene ID****	AGPv3.20 biotype	AGPv3.20 description	*Arabidopsis* homolog	*Arabidopsis* annotation	rice homolog	rice annotation
Cluster_7064	GRMZM2G124744	protein_coding	Uncharacterized protein				
Cluster_17857	GRMZM2G455687	protein_coding	Uncharacterized protein	AT5G388401	SMAD/FHA domain-containing protein	LOC_Os06g164301	FHA domain containing protein, putative, expressed
Cluster_25484	GRMZM2G020468;GRMZM2G323061	protein_coding;low_confidence	Uncharacterized protein; /				
Cluster_35583	*zma-MIR167h*	miRNA					
Cluster_46019	GRMZM2G534485; *zma-MIR159b*	protein_coding;miRNA	Uncharacterized protein; /				
Cluster_55047	GRMZM2G178686	protein_coding	Uncharacterized protein				
Cluster_76570	*zma-MIR390b*	miRNA					
Cluster_79385	GRMZM5G863735	low_confidence					
Cluster_81441	*zma-MIR160b*	miRNA					
Cluster_91725	GRMZM2G082055	low_confidence					
Cluster_91727	GRMZM2G512113	protein_coding					
Cluster_94495	GRMZM2G140150	protein_coding	Pyruvate dehydrogenase E1 component alpha subunit	AT1G599001	pyruvate dehydrogenase complex E1 alpha subunit	LOC_Os02g506201	dehydrogenase E1 component domain containing protein, expressed
Cluster_97736	GRMZM2G100349	protein_coding	60S ribosomal protein L18A	AT1G170801	Ribosomal protein L18ae family	LOC_Os01g140701	60S ribosomal protein L18a-1, putative, expressed
Cluster_99405	*zma-MIR159f*	miRNA					
Cluster_119657	*zma-MIR399e*	miRNA					
Cluster_137435	GRMZM2G558405	transposable_element					
Cluster_138945	AC2172933_FG007	protein_coding	Uncharacterized protein	AT1G212701	wall-associated kinase 2	LOC_Os10g060901	OsWAK104 - OsWAK receptor-like protein kinase, expressed
Cluster_149370	GRMZM5G806469	protein_coding	Uncharacterized protein				
Cluster_166266	GRMZM2G084821;GRMZM5G833991	protein_coding;protein_coding	Putative uncharacterized protein; Uncharacterized protein				

*phasing p-value = calculated by ShortStack program

**phasing FDR = p-values were corrected for multiple testing and a Benjamini-Hochberg adjusted significance level of 0.05 was used

***name = indicates the name of the known maize *MIRNA* loci, as reported in miRBase 20, and the name of the known *TAS* loci

****gene annotation = transcriptome assembly reconstructed from our RNA-seq experiment

Appendix C

List of *MIRNA* loci

MIRNA loci in grey = loci whose predicted mature sequence did not pass the abundance filter of at least five reads in at least one library.

*UI = Uniqueness Index of *MIRNA* loci as defined by ShortStack program.
**location = genomic location of *MIRNA* loci based on the transcriptome reannotation obtained from our RNA-seq experiment:
"exon" or "intron" = indicates that the locus is located within an exon or an intron for its entire length;
"exon-intron" = indicates that the locus overlaps with an intron and an exon of a gene;
"antisense" = indicates that the locus is antisense to a gene;
"intergenic" = indicates that the locus is located between genes for its entire length.
***total reads = reads abundance in the merged set of 48 sRNA-Seq libraries as calculated by ShortStack program.
****name = indicates the name of the known maize *MIRNA* loci, as reported in miRBase 20, and the name of the putative novel *MIRNA* loci identified in this study, as "*zma-MIR-NEW…"*

locus	coordinates	strand	length	size class	UI*	location**	repeats masking	miRNA sequence	total reads***	miRNA* sequence	total reads***	name****
Cluster_146717	4:83913921-83914117	-	197	20	0.1032	exon		UGACAGAAGAGAGUGAGCAC	1506	GCUCACUUCUCUCUCUGUCAGU	1	*zma-MIR156a*
Cluster_97438	3:7775485-7775625	-	141	21	0.4524	exon		UUGACAGAAGAGAGUGAGCAC	12430	GCUCACCCUCUAUCUGUCAGUC	2	*zma-MIR156b*
Cluster_85431	5:92406162-92406425	-	264	21	0.6538	antisense		UUGACAGAAGAGAGUGAGCAC (1) AUGCGGAGAGGCUCUCGAGAGA (2)	12365 (1) 7097 (2)	GCUCACUUCUCUUUCUGUCAGCC (1) UCUCGAUGGCUUAUCUCCCGUCU (2)	54 (1) 1 (2)	*zma-MIR156d*
Cluster_132409	2:180843029-180843319	-	291	20	0.1219	intergenic		UGACAGAAGAGAGUGAGCAC	1550	GCUCACUUCUCUUUCUGUCAGC	69	*zma-MIR156f*
Cluster_68754	7:115997073-115997193	+	121	20	0.1255	exon		UGACAGAAGAGAGUGAGCAC	1556	GCUCACUUCUCUUUCUGUCAGC	85	*zma-MIR156g*
Cluster_11028	10:128549994-128550106	-	113	20	0.1043	intron		UGACAGAAGAGAGUGAGCAC	1476	GCUCACUGCUCUUUCUGUCAUC	2	*zma-MIR156h*
Cluster_70307	7:130685434-130685602	-	169	21	0.4093	intergenic		AUGACAGAAGAGAGAGAGCAC (1) GCUCUCUGCUCUCACUGUCAUC (2)	18 (1) 12 (2)	GCUCUCUGCUCUCACUGUCAUCG (1) UGACAGAAGAGAGAGCAC (2)	1 (1) 12 (2)	*zma-MIR156j*
Cluster_35825	6:96290867-96291070	-	204	20	0.9901	exon		UGACAGAAGAGAGUGAGCAC (1) UUGACAGAAGAGAGCGAGCAC (2)	353 (1) 3135 (2)	GCUCGCUUCUCUUUCUGUCAGC (1) GCUCGCUUCUCUUUCUGUCAGCC (2)	197 (1) 42 (2)	*zma-MIR156k*
Cluster_91762	5:183907808-183908030	-	223	20	0.104	intergenic		UGACAGAAGAGAGUGAGCAC	1568	GCUCACUGCUCUAUCUGUCACC	4	*zma-MIR156l*
Cluster_45988	8:10392807-10393091	+	285	21	0.3381	exon		UUUGGAUUGAAGGGAGCUCUG	5263	GAGCUCCUAUCAUUCCAAUGA	407	*zma-MIR159a*
Cluster_99405	3:25492207-25492486	-	280	21	0.3394	exon		UUUGGAUUGAAGGGAGCUCUG	5197	GAGCUCCUCUCAUUCCAAUGA	16	*zma-MIR159f*
Cluster_18968	9:57908485-57908644	+	160	21	0.1847	exon		UGCCUGGCUCCCUGUAUGCCA	4011	GCGUGCAAGGAGCCAAGCAUG	98	*zma-MIR160b*
Cluster_81441	5:43049564-43049727	-	164	21	0.186	exon		UGCCUGGCUCCCUGUAUGCCA	3902	GCGUGCAAGGAGCCAAGCAUG	128	*zma-MIR160b*
Cluster_139588	2:237695395-237695577	-	183	21	0.7729	intergenic		UGCCUGGCUCCCUGUAUGCCG	186	GCGUGCGAGGUGCCAGGCAUG	68	*zma-MIR160f*
Cluster_35402	6:91612990-91613083	+	94	21	0.1792	intergenic		UGCCUGGCUCCCUGUAUGCCA	3925	GCGUGCAAGGAGCCAAGCAUG	95	*zma-MIR160g*
Cluster_91567	5:182091011-182091161	-	151	21	1	exon		UCGAUAAACCUCUGCAUCCAG	7395	GGGCGCAGUGGUUUAUCGAUC	21	*zma-MIR162*
Cluster_40762	6:141801278-141801485	+	208	21	0.2483	exon		UGGAGAAGCAGGGCACGUGCA	862	GAUGUGCCCAUCUUCUCCACC	13	*zma-MIR164b*
Cluster_42413	6:153550670-153550962	-	293	21	0.5137	intergenic		UGGAGAAGCAGGGCACGUGCA	857	CAUGUGCCCUUCUUCUCCAUC	305	*zma-MIR164c*
Cluster_75500	7:172764726-172764939	-	214	21	0.3431	intergenic		UGGAGAAGCAGGGCACGUGCA	814	CACGUGGUCUCCUUCUCCAUC	137	*zma-MIR164d*
Cluster_34795	6:84390192-84390456	-	265	21	0.1259	intergenic		UCGGACCAGGCUUCAUUCCCC	267227	GGAAUGUUGUCUGGCUCGGGG	308	*zma-MIR166a*
Cluster_79341	5:21954949-21955111	-	163	21	0.1141	exon		UCGGACCAGGCUUCAUUCCCC	267046	GGAAUGUUGUCUGGUUCAAGG	981	*zma-MIR166d*
Cluster_181949	1:241357057-241357151	+	95	21	0.1122	intergenic		UCGGACCAGGCUUCAUUCCCC	266897	GGAAUGUCGUCUGGCGCGAGA	90	*zma-MIR166i*
Cluster_76840	5:4304401-4304532	+	132	21	0.3572	intergenic		UGAAGCUGCCAGCAUGAUCUA	289	GAUCAUGCUGUGGCAGCCUCACU	42	*zma-MIR167c*
Cluster_187284	1:287799562-287799694	-	133	23	0.958	intergenic		GGUCAUGCUGCUGCAGCCUCACU	5566	UGAAGCUGCCAGCAUGAUCUA	287	*zma-MIR167d*
Cluster_62341	7:9835391-9835482	-	92	21	0.3222	intergenic		UGAAGCUGCCAGCAUGAUCUG	1217	GAUCAUGCUGUGCAGUUUCAUC	73	*zma-MIR167e*
Cluster_112061	3:192151748-192151870	+	123	21	0.2304	intergenic		UGAAGCUGCCAGCAUGAUCUG	1243	GAUCGUGCUGCGCAGUUUCACC	51	*zma-MIR167f*
Cluster_105138	3:119196650-119196937	+	288	22	0.3563	antisense		UGAAGCUGCCAGCAUGAUCUGG	5890	AGGUCAUGCUGUAGUUUCAUC	359	*zma-MIR167g*
Cluster_159937	4:239639188-239639288	-	101	21	0.689	exon		UCGCUUGGUGCAGAUCGGGAC	116042	CCCGCCUUGCAUCAAGUGAAU	93166	*zma-MIR168b*
Cluster_100484	3:37610441-37610611	+	171	21	0.8571	intergenic		CAUCGGCAAGUUGUUCUUGGC (1) UUAGGCUCGGGGACUAUGGUG (2)	179 (1) 205 (2)	CAAGGAUGACUUGCCGAUCGAU (1) CCGUGGCUCCUGCUCCUGAUG (2)	2 (1) 59 (2)	*zma-MIR169a*
Cluster_45345	8:4792705-4792880	+	176	21	0.8519	intergenic		UUAGGCUCGGGGACUACGGUG	225	CCGUGGCUCCUGCUCCUGAUG	53	*zma-MIR169b*
Cluster_22718	9:109158850-109159049	-	200	24	0.9363	intergenic		UGAAGCGAGAAAGAGAGUGGGAUG	144	UCUUGUUCUUUUUCUCUCACG	1	*zma-MIR169r*
Cluster_24552	9:126503120-126503250	-	131	21	0.4869	exon		UGUUGGCUCGGCUCACUCAGA	396	UGAUUGAGCCGUGCCAAUAUC	160	*zma-MIR171d*
Cluster_161380	1:7950201-7950318	+	118	21	0.5521	intergenic		UGUUGGCUCGGCUCACUCAGA	404	UGAUUGAGCCGUGCCAAUAUC	171	*zma-MIR171e*
Cluster_79182	5:20331918-20332202	-	285	21	0.661	exon		UUGAGCCGCGUCAAUAUCUCC	595	AGGUAUUGGCGCGCCUCAAUC	40	*zma-MIR171m*
Cluster_21633	9:97675321-97675507	-	187	21	0.3579	intergenic		UUGAGCCGCGCCAAUAUCUCU	96	CGAUAUUGGUGAGGUUCAAUC	4	*zma-MIR171n*
Cluster_96207	5:215901559-215901714	-	156	21	0.4009	intergenic		AGAAUCUUGAUGAUGCUGCAU	364	GCAGCACCAUCAAGAUUCACA	197	*zma-MIR172b*
Cluster_152653	4:171155871-171156023	+	153	21	0.8973	exon		GCAGCACCACCAAGAUUCACA	1594	AGAAUCUUGAUGAUGCUGCAU	369	*zma-MIR172c*
Cluster_123329	2:49985277-49985542	+	266	22	0.9916	exon		UUCCUGAUGCCUCCUAUUCCUA	129	GGAAUAGGAACAUGAAGGAAAG	4	*zma-MIR2118g*
Cluster_142785	4:30231213-30231321	+	109	21	0.9809	intergenic		AGGAUUAGAGGCAACUGAACC	120	UUCAGUUUCCUCUAAUAUCUCA	4	*zma-MIR2275b*
Cluster_56976	8:147815424-147815606	-	183	21	0.2981	antisense		UUGGACUGAAGGGUGCUCCCU	854	AGAGCUCUCUUCAGUCCACUC	11	*zma-MIR319c*
Cluster_187880	1:292655914-292656125	+	212	21	0.4997	intergenic		AAGCUCAGGAGGGAUAGCGCC	20518	CGCUAUCUAUCCUGAGCUCCA	24	*zma-MIR390a*
Cluster_76570	5:2764087-2764309	+	223	21	0.5013	intergenic		AAGCUCAGGAGGGAUAGCGCC	20695	CGCUAUCUAUCCUGAGCUCCA	32	*zma-MIR390b*
Cluster_117147	2:754233-754354	+	122	21	0.5057	intergenic		UCAGUGCAAUCCCUUUGGAAU	957	UCCAAAGGGAUCGCAUUGAUC	322	*zma-MIR393a*
Cluster_98675	3:18647076-18647298	+	223	21	0.3392	intergenic		UCCAAAGGGAUCGCAUUGAUC	304	UCAAUGCGAUCCUUUUGGAGG	2	*zma-MIR393b*
Cluster_13920	10:148152620-148152792	-	173	21	0.4928	intergenic		UCAGUGCAAUCCCUUUGGAAU	961	UCCAAAGGGAUCGCAUUGAUC	332	*zma-MIR393c*
Cluster_92981	5:193871970-193872173	+	204	20	0.3488	exon		UUGGCAUUCUGUCCACCUCC	334	AGGUGGGCAUACUGCCAAUG	1	*zma-MIR394a*
Cluster_150922	4:154193671-154193866	+	196	20	0.2878	exon		UUGGCAUUCUGUCCACCUCC	361	AGGUGGGCAUACUGCCAAUG	2	*zma-MIR394b*
Cluster_118144	2:6334647-6334786	-	140	20	0.5686	intergenic		UGAAGUGUUUGGGGGAACUC	23	GUUCUCCUCAAACCACUUCAGU	16	*zma-MIR395a*
Cluster_118142	2:6332999-6333108	-	110	20	0.1333	intergenic		UGAAGUGUUUGGGGGAACUC	18	GUUCCCUUCAAGCACUUCACA	2	*zma-MIR395e*
Cluster_118143	2:6333795-6333943	-	149	20	0.1442	intergenic		UGAAGUGUUUGGGGGAACUC	20	GUUCCCUUCAAGCACUUCACA	3	*zma-MIR395h*
Cluster_152938	4:173324506-173324638	+	133	21	0.3517	exon		UUCCACAGCUUUCUUGAACUG	800	GUUCAAUAAAGCUGUGGGAAA	220	*zma-MIR396a*
Cluster_15881	9:15367436-15367590	-	155	21	0.3253	exon-intron		UUCCACAGCUUUCUUGAACUG	815	GUUCAAUAAAGCUGUGGGAAA	187	*zma-MIR396b*
Cluster_117326	2:1962352-1962645	-	294	21	0.5	intron		UCCACAGGCUUUCUUGAACUG	34972	GUUCAAGAAAGCCCAUGGAAA	202	*zma-MIR396c*
Cluster_13801	10:147337694-147337986	+	293	21	0.5002	exon		UCCACAGGCUUUCUUGAACUG	34902	GUUCAAGAAAGCCCAUGGAAA	182	*zma-MIR396d*

locus	coordinates	strand	length	size class	UI*	location**	repeats masking	miRNA sequence	total reads***	miRNA* sequence	total reads***	name****
Cluster_95911	5:214389058-214389216	+	159	21	0.8858	intergenic		GGUCAAGAAAGCUGUGGGAAG	984	UUCCACAGCUUUCUUGAACUU	796	*zma-MIR396f*
Cluster_95914	5:214399551-214399699	-	149	21	0.3271	intergenic		UUCCACAGCUUUCUUGAACUG	882	GUUCAAUAAAGCUGUGGGAAA	191	*zma-MIR396g*
Cluster_84686	5:82459688-82459789	-	102	21	0.9989	intergenic		UUGAGCGCAGCGUUGAUGAGC	6461	UCACCAGCGCUGCACUCAAUU	65	*zma-MIR397b*
Cluster_131352	2:170163043-170163244	+	202	21	0.6996	exon-intron	TXX	UGUGUUCUCAGGUCGCCCCCG	2457	GGGGCGAACUGAGAACACAUG	809	*zma-MIR398a*
Cluster_64542	7:38552546-38552814	+	269	21	0.8489	exon		UGUGUUCUCAGGUCGCCCCCG	2441	GGGGCGGACUGGGAACACAUG	2210	*zma-MIR398b*
Cluster_158504	4:229179889-229179984	-	96	21	0.3448	intergenic		UGCCAAAGGAGAAUUGCCCUG	542	GUGCGGUUCUCCUCUGGCACG	3	*zma-MIR399a*
Cluster_150969	4:154743139-154743251	-	113	21	0.9983	exon		UGCCAAAGGAGAGCUGUCCUG	896	GUGCAGCUCUCCUCUGGCAUG	34	*zma-MIR399b*
Cluster_43356	6:159890121-159890406	+	286	21	0.54	exon-intron	TXX	UGCCAAAGGAGAAUUGCCCUG	586	GGGUACGUCUCCUUUGGCACA	253	*zma-MIR399c*
Cluster_111727	3:189262800-189262890	+	91	21	0.8124	exon		GUGUGGCUCUCCUCUGGCAUG	311	UGCCAAAGGAGAGCUGCCCUG	108	*zma-MIR399d*
Cluster_119657	2:16353049-16353216	-	168	21	0.3624	intergenic		UGCCAAAGGAGAGUUGCCCUG	7948	GGGCUUCUCUUUCUUGGCAGG	73	*zma-MIR399e*
Cluster_43351	6:159882385-159882480	-	96	21	0.9986	intergenic		UGCCAAAGGAAAUUUGCCCCG	783	GGGCAACUUCUCCUUUGGCAGA	45	*zma-MIR399f*
Cluster_16950	9:24378446-24378650	-	205	21	0.4608	intergenic		UGCCAAAGGAGAGUUGCCCUG	7833	GGCAGCUCUCCUCUGGCAGG	1162	*zma-MIR399j*
Cluster_48429	8:38494396-38494659	-	264	21	0.9951	exon		CUGCACUGCCUCUUCCCUGGC (1) UGCACUGCCUCUUCCCUGGC (2)	5347 (1) 2399 (2)	CAGGGACGAGGCAGAGCAUGG (1) CAGGGACGAGGCAGAGCAUG (2)	201 (1) 97 (2)	*zma-MIR408b*
Cluster_161127	1:6409230-6409385	+	156	21	0.5478	overlap		UGGAAGGGGCAUGCAGAGGAG	2412	CCUGUGCCUGCCUCUUCCAUU	448	*zma-MIR528a*
Cluster_28456	9:153752325-153752434	-	110	21	0.5121	intergenic		UGGAAGGGGCAUGCAGAGGAG	2552	CCUGUGCCUGCCUCUUCCAUU	458	*zma-MIR528b*
Cluster_89370	5:160612950-160613131	+	182	21	1	intergenic		AGAAGAGAGAGAGUACAGCCU	953	GCUGUACCCUCUCUCUUCUUC	1	*zma-MIR529*
Cluster_97439	3:7775768-7775922	-	155	20	0.1045	exon		UGACAGAAGAGAGUGAGCAC	1444	GCUCACUUCUCUCUCUGUCAGC	4	*zma-MIR-NEW156m*
Cluster_162215	1:13558787-13558940	+	154	21	0.2018	intergenic		UCGGACCAGGCUUCAUUCCCC	267547	GGAAUGUUGUCUGGUUGGAGA	530	*zma-MIR-NEW166o*
Cluster_179753	1:223279170-223279319	-	150	21	0.1139	exon		UCGGACCAGGCUUCAUUCCCC	266876	GGAAUGUUGUCUGGUUCAAGG	975	*zma-MIR-NEW166p*
Cluster_179764	1:223371637-223371836	-	200	21	0.1139	antisense		UCGGACCAGGCUUCAUUCCCC	266150	GAAAUGUUGUCUGGUUCAAGG	340	*zma-MIR-NEW166q*
Cluster_22845	9:110562488-110562617	+	130	21	0.3971	antisense		UGAAGCUGCCAGCAUGAUCUGG	5936	AGAUCAUGUUGCAGCUUCACU	5896	*zma-MIR-NEW167k*
Cluster_110225	3:178015654-178015748	-	95	24	0.3207	intergenic	DTA;MITE	AUCCAGCGACAAGUAAAAAGAAAC	97	UUCGUUUUAGUUGUCGCUGGAUAG	1	*zma-MIR-NEW1*
Cluster_97510	3:8229464-8229514	-	51	24	0.5373	intergenic	DTA;RLG	UCGUGGAAUGGACGAGCCGAGCCU	21	GCCGAGCUCGUUCAGGCACCGAGC	2	*zma-MIR-NEW2*
Cluster_12750	10:141026541-141026611	-	71	24	1	intergenic	DTC	AUGGGCUCGGCCGUCGAAAAUAGA	10	UAUUUUCGGCGGCUGAACCUUGGC	1	*zma-MIR-NEW3*
Cluster_125850	2:83264783-83264916	+	134	24	0.304	intergenic	DTC	AUCCGGAACAACCGAACAGGGCCU	46	GGCCUGUUCGGUUGUAUCGGAAUA	3	*zma-MIR-NEW4*
Cluster_161285	1:7352075-7352237	-	163	24	0.7948	intergenic	DTC	ACCGAACAGGGCUUAAAAGGAACC	242	UUCCAUAAGGGCCUGUUCGGUUG	1	*zma-MIR-NEW5*
Cluster_142500	4:27097053-27097158	-	106	24	0.3589	intergenic	DTC;RLG	AGGGCCAUAGCACUCGACAGAGAA	33	CUUUGCCGAGUGUUGUGGCCCCGA	2	*zma-MIR-NEW6*
Cluster_43700	6:162062485-162062596	+	112	24	0.9527	intergenic	DTC;RLG	GAAGUGACACUCGGCAAAGAGAGC	25	AGAUUUUGCCGAGUGUUAUUAGGU	1	*zma-MIR-NEW7*
Cluster_21182	9:91340852-91340978	+	127	24	0.951	intergenic	DTH;MITE	AGAAUAAAAUGUAGGGACCCUAGA	103	UUUAGUCCCUUCAUUUUAUUCCAU	1	*zma-MIR-NEW8*
Cluster_77897	5:11065006-11065190	+	185	21	0.7246	intron	DTM	AGCCUCCUAGACCUUUAGAUC	17	UCUAAAGGUCCAGGAGGCUCG	2	*zma-MIR-NEW9*
Cluster_31840	6:33172016-33172147	+	132	21	0.3496	intergenic	DTM	CCCGCCGGCGAGCGCUUUCCU (1) GCGAGGGCGCCGCCGUCGGUGA (2)	12532 (1) 9335 (2)	GAGGGCGCCGCCGUCGGUGA (1) CCCGCCGGCGAGCGCUUUCCUGC (2)	461 (1) 59 (2)	*zma-MIR-NEW10a*
Cluster_31862	6:33433180-33433311	+	132	21	0.3512	exon	DTM	CCCGCCGGCGAGCGCUUUCCU (1) GCGAGGGCGCCGCCGUCGGUGA (2)	12240 (1) 9361 (2)	GAGGGCGCCGCCGUCGGUGA (1) CCCGCCGGCGAGCGCUUUCCUGC (2)	445 (1) 51 (2)	*zma-MIR-NEW10b*
Cluster_171565	1:116730685-116730816	+	132	22	0.3857	intergenic	DTM	CCCGCCGGCGAGCGCUUUCCU (1) GUGAGGGCGCCGCCGUCGGUGA (2)	12311 (1) 20164 (2)	GAGGGCGCCGCCGUCGGUGA (1) CCCGCCGGCGAGCGCUUUCCUGU (2)	484 (1) 1093 (2)	*zma-MIR-NEW10c*
Cluster_171567	1:116752581-116752712	+	132	22	0.3856	intergenic	DTM	CCCGCCGGCGAGCGCUUUCCU (1) GUGAGGGCGCCGCCGUCGGUGA (2)	12483 (1) 20416 (2)	GAGGGCGCCGCCGUCGGUGA (1) CCCGCCGGCGAGCGCUUUCCUGU (2)	488 (1) 1096 (2)	*zma-MIR-NEW10d*
Cluster_35154	6:88754463-88754749	+	287	21	0.9817	intron	DTM	UUAAAUCUGGACCCUUCAUCU	115	ACGAAGGGUCUAGAUUUAAUA	3	*zma-MIR-NEW11*
Cluster_111912	3:191058499-191058611	+	113	21	0.9792	exon	DTM	UUAAGGACCCGAAAUGUAUUC	59	AUACAUUUUCGGUCCUUAAAC	1	*zma-MIR-NEW12*
Cluster_146957	4:89124445-89124607	+	163	21	1	intron	DTM	UCUAACGGACCAGAAUGCUCA	138	AGCAUUCUGUUUCGUUAGAUC	1	*zma-MIR-NEW13*
Cluster_184161	1:261285253-261285497	-	245	24	0.872	intron	DTM	AUCUAACGGACCAAGAGAGCUGCU	24	CAGGUUUUUUAGUCCGUUAGGUCU	3	*zma-MIR-NEW14*
Cluster_75722	7:173701161-173701450	-	290	24	0.3454	intergenic	DTM	UACCUCGGCGCCAAUGAUGAUGAC	48	CAUCAUUAUUGGCGUCGAGGUAAC	4	*zma-MIR-NEW15*
Cluster_71880	7:145769374-145769505	-	132	21	0.6897	exon	MITE	AGCCCCUCCAAUUCUCUCCAU	14	GAAGGGGAUUGGAGGGGCUAA	1	*zma-MIR-NEW16*
Cluster_160752	1:4056630-4056761	-	132	21	0.799	intron	MITE	AGCAAGAGGAUUGGAGGGGCU	21	CCCCUCCAAUCCCUUUGCUCC	3	*zma-MIR-NEW17*
Cluster_175514	1:180855798-180855932	-	135	21	0.7842	intron	MITE	ACCGGAGGGGAUUGGAGAGAU	90	CCCUCCAAUCCCCUCCGGUAU	2	*zma-MIR-NEW18*
Cluster_68786	7:116404987-116405092	+	106	21	0.5276	intron	MITE	GAUUGAGAGGGAUUGGAGGGG	67	CCUCCAAUCCCCCUCAAUCCA	1	*zma-MIR-NEW19*
Cluster_14894	9:7034835-7034978	-	144	21	0.3524	intergenic	MITE	UCCUCCAAUCCCCUCCGGUAU	22	ACCGGAGGGGAUUGGAGGCGC	6	*zma-MIR-NEW20a*
Cluster_33566	6:65665008-65665151	-	144	21	0.3848	intergenic	MITE	UCCUCCAAUCCCCUCCGGUAU	17	ACCGGAGGGGAUUGGAGGCGC	3	*zma-MIR-NEW20b*
Cluster_168441	1:73390290-73390441	-	152	21	0.6893	intron	MITE	UUGGGAGCCACAAAACUGAAG	121	UCGGUUUUGUGGCUUCCAAAC	1	*zma-MIR-NEW21*
Cluster_99658	3:28217205-28217322	+	118	24	0.9586	intergenic	MITE	GAGGAUUGGAGGGGUUAGAAUCCC	59	GAUUAUAGCCCCCUCAAUCUCCUU	3	*zma-MIR-NEW22*
Cluster_131350	2:170153030-170153209	-	180	21	0.9579	intergenic	MITE	ACCGGAGGGGUUGAAGGUGCU	50	CUCCUCCAAUCUCCUCCGGUAU	12	*zma-MIR-NEW23*
Cluster_155415	4:195967261-195967394	-	134	21	0.998	exon	MITE	ACCGGAGGAGGUUAGAGGAGC	4060	CCCUCUAAUCUCCUCCGAUUU	85	*zma-MIR-NEW24*
Cluster_97151	3:5671567-5671664	+	98	21	0.9167	exon	MITE	UUUUCUGGCUGCCAAACUAGC	26	UAGUUUAGGAGCCAGAAAAUC	19	*zma-MIR-NEW25*

locus	coordinates	strand	length	size class	UI*	location**	repeats masking	miRNA sequence	total reads***	miRNA* sequence	total reads***	name****
Cluster_18416	9:45982538-45982699	-	162	24	0.3713	intergenic	MITE	AGUGAAUUGAGGGGGCUAGAAUCC	265	AUUAUAGCCCUCUCAAUCCCCUCC	4	*zma-MIR-NEW26a*
Cluster_52595	8:104285091-104285235	-	145	24	0.2171	intergenic	MITE;RLG	AGUGAAUUGAGGGGGCUAGAAUCC	281	AUUAUAGCCCCCUCAAUCCCCUCC	9	*zma-MIR-NEW26b*
Cluster_27139	9:146368401-146368561	-	161	24	0.9463	intergenic	MITE	AAUCCCCUCCGGGUUUUAGGCUCC	148	AGUCUAAAACCCGAAGAGGAUUAA	1	*zma-MIR-NEW27*
Cluster_45173	8:3160876-3161040	+	165	24	0.9888	intergenic	MITE	ACCGGAGAGAAUUAGAUGGGCUAA	100	AACCCCUCUAAUUUUCUCUGAUUU	1	*zma-MIR-NEW28*
Cluster_92071	5:186880366-186880547	+	182	24	0.5045	antisense	MITE	AAUCCCCUCCGGUUUUGUGGCUCC	22	AACCACAAAACCGGAGAGGAUUGA	1	*zma-MIR-NEW29*
Cluster_135011	2:202972078-202972221	-	144	24	0.93	intergenic	MITE	AUUCCCUUCGGGUUUGAGGCUCCC	238	AAGUCCAAAACCCGGAGGGGAUUC	1	*zma-MIR-NEW30*
Cluster_3717	10:33024642-33024833	-	192	24	0.9882	intergenic	RLC	AUCGGACCGGUCUGGAGCACCAGC	37	UGGCGCAUCAGACCGGUCUGGAGC	5	*zma-MIR-NEW31*
Cluster_143581	4:38092116-38092268	+	153	24	0.889	intergenic	RLC;RLG	ACGGAUUGAGAAAGAUUUUAACUU	77	GAUUAAAACUCCCUCAAUCCCCGU	5	*zma-MIR-NEW32*
Cluster_24240	9:123683632-123683828	-	197	21	0.9526	exon	RLC	UCUUCCUCCGCAUAUAAGCCU	206	GAUUACAUGCGGAGGAAGAUA	2	*zma-MIR-NEW33*
Cluster_78483	5:15055919-15056013	+	95	22	0.44	intergenic	RLC	CGCCCCGCAUCCCCAUCCCGCC	2308	CGGAUGGGUGCUGCGCGGUGGU	1343	*zma-MIR-NEW34a*
Cluster_158900	4:232600920-232601023	+	104	22	0.6059	intergenic	RLC	CGGAUGGGUGCUGCGCGGUGGU	1354	CGCCUCGCAUCCCCAUCCCGCC	996	*zma-MIR-NEW34b*
Cluster_185180	1:270127169-270127263	+	95	22	0.4414	antisense	RLC	CGCCCCGCAUCCCCAUCCCGCC	2269	CGGAUGGGUGCUGCGCGGUGGU	1278	*zma-MIR-NEW34c*
Cluster_148452	4:121038943-121039074	+	132	24	0.1526	intergenic	RLC	AAUCCCCCUCAAUCCCUAUGAAUU	35	ACCAUAGGGAUUGAGGGGGAUUGG	1	*zma-MIR-NEW35*
Cluster_59999	8:168663931-168664077	-	147	24	0.34	intergenic	RLG	GUCUGCCCCUGUGCAGAAGGCACC	49	AGCCUCGCACCAGGCCGCGGACCG	3	*zma-MIR-NEW36*
Cluster_109670	3:172519393-172519680	-	288	21	0.8962	intron	RLX	UCAAUCCGGACGCCGAUGUGG	128	ACGUCGGUAUCCGGUUUGAGC	11	*zma-MIR-NEW37*
Cluster_114045	3:209147786-209147927	-	142	24	0.9088	exon	RLX	CUGGAUAACUGUCACACUUUGCCU	17	AUAAAGUAUGACAAGUUAGACAGCG	1	*zma-MIR-NEW38*
Cluster_155777	4:199407377-199407588	-	212	22	0.8685	intron	TXX	CCUCCCCUAUGGAUCCCCUCUA	31	GAGGGAUCCCUAGGGAAGGAA	1	*zma-MIR-NEW39*
Cluster_32516	6:44544337-44544596	-	260	21	1	antisense		AAUACACAUGGGUUGAGGAGG	15	ACCUCAACUCAUGUGUAUUGG	3	*zma-MIR-NEW40*
Cluster_45252	8:3888146-3888291	+	146	21	0.9949	intron		AAGAAUGGAGAUGUGAGGAUC	273	UCCUCCCACCUCCAUUCUUGG	11	*zma-MIR-NEW41*
Cluster_52994	8:108537520-108537706	-	187	21	0.5286	exon		AACACAUGAAAAACACAGGAUU	12	UCUGCGUUUUUCCUGUGUUUU	7	*zma-MIR-NEW42*
Cluster_108273	3:158392524-158392662	+	139	21	0.9338	intron		AUUCUGGCCCCCUCAAACCCA	56	GGUUUGAGGGGGCUAGAAUCC	2	*zma-MIR-NEW43*
Cluster_125477	2:78236567-78236731	+	165	21	0.3933	antisense		ACCCCUUCAAUACCCUCCAUU	10	UGAAGGGGAUUGAGGGGGCUA	4	*zma-MIR-NEW44*
Cluster_188789	1:299826850-299826972	-	123	21	1	intron		AAGAAUGGAGGCGGGAGGACC	78	UUCUCCCACCUUCAUUCUUGU	2	*zma-MIR-NEW45*
Cluster_137914	2:224808203-224808319	+	117	21	0.9994	exon		CUUCAGGAGGGGGGGGGGAUC (1) CUUCAGGAGGGGGGGGGGAUCC (2)	885 (1) 523 (2)	UCCCUCUCUCCCCUUGAAGGC (1) GUCCCUCUCUCCCCUUGAAGGC (2)	115 (1) 1 (2)	*zma-MIR-NEW46*
Cluster_118688	2:9991046-9991274	-	229	21	0.925	intron		GCUUUAAGAUUGAUGGGGUGG	24	ACCCCAUCAAUCUUAAAGUAA	1	*zma-MIR-NEW47*
Cluster_121051	2:28003389-28003635	-	247	21	0.996	exon		GCGGCGGCGGGGGUGGCGAGG	83	UCGUCGUCAACCGCCGCCGCCC	54	*zma-MIR-NEW48*
Cluster_69572	7:123915118-123915226	-	109	21	1	intergenic		UUUGAGAUUCGUAGCUUUUAC	22	AAAACCUAUGAAUCUGAAAAA	1	*zma-MIR-NEW49*
Cluster_77047	5:5324010-5324143	+	134	21	0.8714	exon		UCCUCUCCCCUGCUGCCACGA	62	GCGGCGGCGGGGCGGAGGAGG	27	*zma-MIR-NEW50*
Cluster_81571	5:45156199-45156441	-	243	21	0.929	intron		UUCAAUGCCAUCGCCAGCGGC	37	CACUGGCAAUGACAUGGAAGG	1	*zma-MIR-NEW51*
Cluster_104499	3:108883303-108883475	+	173	21	0.9608	intron		UUCAUCCGAGAAAAGUGGGGG (1) UCAUCCGAGAAAAGUGGGGGC (2)	12 (1) 23 (2)	CUCACUUUUUUCGGAUGAAGU (1) CCUCACUUUUUUCGGAUGAAG (2)	1 (1) 2 (2)	*zma-MIR-NEW52*
Cluster_138945	2:232386803-232387014	-	212	21	0.9925	intron		UUUCAUGCUAUCGUCAGUGGC (1) UUCAUGCUAUCGUCAGUGGCA (2)	245 (1) 194 (2)	CACUGACGAGGGCAUGAAAAA (1) CCACUGACGAGGGCAUGAAAA (2)	79 (1) 5 (2)	*zma-MIR-NEW53*
Cluster_64050	7:30173821-30176997	+	3177	22	0.9818	exon		AGGAAGUAGAGGGUGACGAGGA	1706	CUCGUCACCAUCUACUUCCUCU	1	*zma-MIR-NEW54*
Cluster_9125	10:109602805-109603005	-	201	22	1	exon		CAGGAGAUGGCAGAGGAGGGGC	80	UCCUCCUCAUCCGUCUCGGCC	4	*zma-MIR-NEW55*
Cluster_50300	8:73653055-73653337	+	283	22	0.9272	intergenic		CUCGGAGGGGAUUGAAGGGGCU	79	CCUCUCCAAUCCCCUCUGGAUU	11	*zma-MIR-NEW56*
Cluster_72242	7:148921359-148921551	-	193	22	1	intron		CAAGCAUAAAGGUGAAAGGACC	12	UCCUCCCACUUCUAUGCUUGCG	4	*zma-MIR-NEW57*
Cluster_16743	9:22668218-22668304	+	87	23	0.9988	exon		CGAGCUUUUUAAAACUGAUGCCC	39859	GCAGGUGAUGAUGAUGUUUGGA	68	*zma-MIR-NEW58*
Cluster_16020	9:16689804-16689911	-	108	24	0.9538	exon		AUCUCCUCGUCAUCCCUGGCCACC	27	UGACCAGAGAUCACGAGAGGAUCC	2	*zma-MIR-NEW59*
Cluster_127306	2:110118172-110118326	+	155	24	0.9535	intergenic		AGAGAAUUGAGCGGACUAGAAUCC	17	AUUCUAGCCCCCUCAACCCUCUUC	1	*zma-MIR-NEW60*
Cluster_172623	1:141281285-141281651	-	367	24	0.8502	intron		AGGCCCCUGAACGAUCGACACUGA	56	AGUGUCGAUCGUUCAUGGGGCUCC	3	*zma-MIR-NEW61*

Appendix D

List of predicted targets

miRNA:
() = when the identified miRNA or miRNA* sequences are not identical to those reported in miRBase, their relationship is indicated:
isoMIR = isoMIR of the miRNA annotated in miRBase; miRNA* = miRNA* annotated in miRBase; miRNA* isoMIR = isoMIR of the miRNA* annotated in miRBase; unrelated = nonoverlapping with miRBase annotated sequences.

miRNAs and miRNA's in italic* = sequences whose precursor was not confirmed by our analysis or that lacked a genome annotation in miRBase.

miRNA	**target transcript ID***	***Arabidopsis* homolog**	***Arabidopsis* annotation**
miR156a-miR156f-miR156g-miR156h-miR156l-miR-NEW156m-*miR156c-miR156e-miR156i*	GRMZM2G101511_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G163813_T01	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G126018_T01	AT2G42200.1	squamosa promoter binding protein-like 9
	AC233751.1_FGT002	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G097275_T04	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G163813_T02	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G106798_T02	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM5G878561_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G065451_T02 GRMZM2G097275_T04_j_1	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G065451_T01 GRMZM5G806833_T01_j_1	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G460544_T01 GRMZM2G163813_T04	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G061734_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G097275_T01	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G097275_T03 GRMZM2G148467_T02	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G414805_T05	AT1G27370.1	squamosa promoter binding protein-like 10
	GRMZM2G160917_T03 GRMZM2G052921_T01	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G126018_T02 GRMZM2G414805_T07 GRMZM2G371033_T01_j_1	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G101511_T02	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G414805_T03	AT1G27370.1	squamosa promoter binding protein-like 10
	GRMZM2G160917_T02	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G160917_T01 GRMZM5G806833_T01	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G106798_T03	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G371033_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G414805_T04 GRMZM2G414805_T01	AT1G27370.1	squamosa promoter binding protein-like 10
	GRMZM2G307588_T01	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G106798_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G414805_T02	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G148467_T01	AT1G27370.1	squamosa promoter binding protein-like 10

	GRMZM2G097275_T02	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G414805_T03_j_1		
	GRMZM2G067624_T02	AT1G53160.1	squamosa promoter binding protein-like 4
	GRMZM2G126827_T01	AT1G27370.1	squamosa promoter binding protein-like 10
	GRMZM2G113779_T01	AT3G15270.1	squamosa promoter binding protein-like 5
	GRMZM2G067624_T01	AT1G53160.1	squamosa promoter binding protein-like 4
	GRMZM2G156621_T01	AT1G27370.1	squamosa promoter binding protein-like 10
miR156b(isoMIR)-miR156d.1(isoMIR)	GRMZM2G065451_T02	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G065451_T01	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G097275_T04_j_1		
	GRMZM2G097275_T02	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G097275_T03	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G160917_T01	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G097275_T01	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G097275_T04	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G460544_T01	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G160917_T02	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G160917_T03	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G307588_T01	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G126018_T01	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G126018_T02	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G414805_T03_j_1		
	GRMZM5G878561_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G414805_T01		
	GRMZM2G414805_T03	AT1G27370.1	squamosa promoter binding protein-like 10
	GRMZM2G414805_T05	AT1G27370.1	squamosa promoter binding protein-like 10
	GRMZM2G414805_T02	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G414805_T07		
	GRMZM2G163813_T01	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G101511_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G163813_T02	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G414805_T04	AT1G27370.1	squamosa promoter binding protein-like 10
	AC233751.1_FGT002	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G101511_T02	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G371033_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G163813_T04		
	GRMZM2G371033_T01_j_1		
	GRMZM2G052921_T01		
	GRMZM2G061734_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G148467_T02		
	GRMZM2G106798_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G148467_T01	AT1G27370.1	squamosa promoter binding protein-like 10
	GRMZM2G106798_T02	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G106798_T03	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM5G806833_T01		
	GRMZM5G806833_T01_j_1		
	GRMZM2G126827_T01	AT1G27370.1	squamosa promoter binding protein-like 10
	GRMZM2G156621_T01	AT1G27370.1	squamosa promoter binding protein-like 10
miR156j.1(isoMIR)	GRMZM2G061734_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G148467_T02		
	GRMZM2G106798_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G148467_T01	AT1G27370.1	squamosa promoter binding protein-like 10
	GRMZM2G106798_T02	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G106798_T03	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM5G806833_T01		
	GRMZM5G806833_T01_j_1		
	GRMZM2G414805_T03_j_1		
	GRMZM5G878561_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein

	GRMZM2G414805_T01		
	GRMZM2G065451_T02	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G414805_T03	AT1G27370.1	squamosa promoter binding protein-like 10
	GRMZM2G414805_T05	AT1G27370.1	squamosa promoter binding protein-like 10
	GRMZM2G065451_T01	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G097275_T04_j_1		
	GRMZM2G097275_T02	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G097275_T03	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G160917_T01	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G414805_T02	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G414805_T07		
	GRMZM2G163813_T01	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G097275_T01	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G101511_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G163813_T02	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G097275_T04	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G460544_T01	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G414805_T04	AT1G27370.1	squamosa promoter binding protein-like 10
	AC233751.1_FGT002	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G101511_T02	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G160917_T02	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G160917_T03	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G307588_T01	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G126018_T01	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G067624_T01	AT1G53160.1	squamosa promoter binding protein-like 4
	GRMZM2G126018_T02	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G067624_T02	AT1G53160.1	squamosa promoter binding protein-like 4
	GRMZM2G371033_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G163813_T04		
	GRMZM2G371033_T01_j_1		
	GRMZM2G052921_T01		
	GRMZM2G113779_T01	AT3G15270.1	squamosa promoter binding protein-like 5
	GRMZM2G126827_T01	AT1G27370.1	squamosa promoter binding protein-like 10
	GRMZM2G156621_T01	AT1G27370.1	squamosa promoter binding protein-like 10
miR156k.1	AC233751.1_FGT002	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G414805_T01		
	GRMZM5G806833_T01_j_1		
	GRMZM2G097275_T03	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G148467_T02		
	GRMZM2G052921_T01		
	GRMZM5G806833_T01		
	GRMZM2G414805_T03_j_1		
	GRMZM2G097275_T02	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G414805_T07		
	GRMZM2G371033_T01_j_1		
	GRMZM2G414805_T03	AT1G27370.1	squamosa promoter binding protein-like 10
	GRMZM2G160917_T01	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G106798_T02	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G160917_T02	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM5G878561_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G097275_T04	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G460544_T01	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G106798_T03	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G097275_T01	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G371033_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G106798_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G126018_T02	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G414805_T05	AT1G27370.1	squamosa promoter binding protein-like 10
	GRMZM2G160917_T03	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G126018_T01	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G101511_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G101511_T02	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein

	GRMZM2G163813_T01	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G061734_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G163813_T04		
	GRMZM2G148467_T01	AT1G27370.1	squamosa promoter binding protein-like 10
	GRMZM2G414805_T02	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G414805_T04	AT1G27370.1	squamosa promoter binding protein-like 10
	GRMZM2G307588_T01	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G163813_T02	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G065451_T01	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G065451_T02	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G097275_T04_j_1		
	GRMZM2G126827_T01	AT1G27370.1	squamosa promoter binding protein-like 10
	GRMZM2G113779_T01	AT3G15270.1	squamosa promoter binding protein-like 5
	GRMZM2G067624_T01	AT1G53160.1	squamosa promoter binding protein-like 4
	GRMZM2G156621_T01	AT1G27370.1	squamosa promoter binding protein-like 10
	GRMZM2G067624_T02	AT1G53160.1	squamosa promoter binding protein-like 4
miR156k.2(isoMIR)	GRMZM2G065451_T02	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G097275_T04_j_1		
	GRMZM2G065451_T01	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G097275_T02	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G097275_T03	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G160917_T01	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G097275_T01	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G097275_T04	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G460544_T01	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G160917_T02	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G160917_T03	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G307588_T01	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G126018_T01	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G126018_T02	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G414805_T03_j_1		
	GRMZM5G878561_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G414805_T01		
	GRMZM2G414805_T03	AT1G27370.1	squamosa promoter binding protein-like 10
	GRMZM2G414805_T05	AT1G27370.1	squamosa promoter binding protein-like 10
	GRMZM2G414805_T02	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G414805_T07		
	GRMZM2G163813_T01	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G101511_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G163813_T02	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G414805_T04	AT1G27370.1	squamosa promoter binding protein-like 10
	AC233751.1_FGT002	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G101511_T02	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G371033_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G163813_T04		
	GRMZM2G371033_T01_j_1		
	GRMZM2G052921_T01		
	GRMZM2G061734_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G148467_T02		
	GRMZM2G106798_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G148467_T01	AT1G27370.1	squamosa promoter binding protein-like 10
	GRMZM2G106798_T02	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G106798_T03	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM5G806833_T01		
	GRMZM5G806833_T01_j_1		
	GRMZM2G126827_T01	AT1G27370.1	squamosa promoter binding protein-like 10
	GRMZM2G156621_T01	AT1G27370.1	squamosa promoter binding protein-like 10
miR159a-miR159f-*miR159b-miR159j-miR159k*	GRMZM2G167088_T01	AT2G32460.1	myb domain protein 101
	GRMZM2G416652_T02	AT2G32460.1	myb domain protein 101
	GRMZM2G416652_T01	AT2G32460.1	myb domain protein 101
	GRMZM2G423833_T01_j_1		
	GRMZM2G423833_T01_j_1		
	GRMZM2G093789_T02_j_1		
	GRMZM2G093789_T01	AT2G32460.1	myb domain protein 101
	GRMZM2G004090_T01	AT2G32460.1	myb domain protein 101

	GRMZM2G075064_T01	AT3G11440.1	myb domain protein 65
	AC209015.3_FGT004		
miR160b-miR160b-miR160g-*miR160a-miR160c-miR160d-miR160e*	GRMZM2G390641_T01_j_1		
	GRMZM2G159399_T01	AT4G30080.1	auxin response factor 16
	GRMZM2G153233_T01	AT4G30080.1	auxin response factor 16
	GRMZM2G390641_T01	AT4G30080.1	auxin response factor 16
	GRMZM2G390641_T02	AT4G30080.1	auxin response factor 16
	GRMZM2G005284_T01	AT4G30080.1	auxin response factor 16
	AC207656.3_FGT002	AT4G30080.1	auxin response factor 16
	GRMZM5G808366_T01	AT1G77850.1	auxin response factor 17
	GRMZM2G081406_T01	AT4G30080.1	auxin response factor 16
	GRMZM2G081406_T01_j_1		
miR160f	GRMZM2G390641_T01_j_1		
	GRMZM2G159399_T01	AT4G30080.1	auxin response factor 16
	GRMZM2G153233_T01	AT4G30080.1	auxin response factor 16
	GRMZM2G390641_T01	AT4G30080.1	auxin response factor 16
	GRMZM2G390641_T02	AT4G30080.1	auxin response factor 16
	GRMZM2G005284_T01	AT4G30080.1	auxin response factor 16
	AC207656.3_FGT002	AT4G30080.1	auxin response factor 16
	GRMZM5G808366_T01	AT1G77850.1	auxin response factor 17
	GRMZM2G081406_T01	AT4G30080.1	auxin response factor 16
	GRMZM2G081406_T01_j_1		
miR162(isoMIR)	GRMZM2G040762_T01_j_1		
	GRMZM2G040762_T01	AT1G01040.1	dicer-like 1
miR164b-miR164c-miR164d-*miR164a-miR164g*	GRMZM2G114850_T01	AT1G56010.2	NAC domain containing protein 1
	GRMZM2G063522_T01	AT1G56010.2	NAC domain containing protein 1
	GRMZM2G393433_T01_j_1		
	GRMZM2G393433_T01_j_1		
	GRMZM2G393433_T01	AT5G53950.1	NAC (No Apical Meristem) domain transcriptional regulator superfamily protein
	GRMZM2G139700_T01	AT5G53950.1	NAC (No Apical Meristem) domain transcriptional regulator superfamily protein
miR166a-miR166d(isoMIR)-miR166i(isoMIR)-miR-NEW166o-miR-NEW166p-miR-NEW166q	GRMZM5G845891_T01		
	GRMZM2G499154_T01		
	GRMZM5G897556_T01		
	GRMZM5G845891_T01		
miR167e-miR167f-*miR167h-miR167i-miR167j*	GRMZM2G042623_T01	AT2G29200.1	pumilio 1
	GRMZM2G042623_T02	AT2G29200.1	pumilio 1
	GRMZM2G112769_T01	AT2G29200.1	pumilio 1
miR167g(isoMIR)-miR-NEW167k	GRMZM2G042623_T01	AT2G29200.1	pumilio 1
	GRMZM2G042623_T02	AT2G29200.1	pumilio 1
	GRMZM2G112769_T01	AT2G29200.1	pumilio 1
miR171m(isoMIR)	GRMZM5G825321_T02	AT4G00150.1	GRAS family transcription factor
	GRMZM2G098800_T02	AT4G00150.1	GRAS family transcription factor
	GRMZM2G037792_T01	AT4G00150.1	GRAS family transcription factor
	GRMZM2G098800_T01	AT4G00150.1	GRAS family transcription factor
	GRMZM2G079470_T01	AT4G00150.1	GRAS family transcription factor
	GRMZM2G051785_T01	AT4G00150.1	GRAS family transcription factor
	GRMZM5G825321_T01	AT4G00150.1	GRAS family transcription factor
	GRMZM2G176124_T01	AT4G00150.1	GRAS family transcription factor
	GRMZM2G418899_T02	AT4G00150.1	GRAS family transcription factor
	GRMZM2G418899_T01	AT4G00150.1	GRAS family transcription factor
	GRMZM2G011947_T01	AT4G00150.1	GRAS family transcription factor
	GRMZM2G118913_T01		
	AC187788.3_FGT008	AT4G00150.1	GRAS family transcription factor
	GRMZM5G825321_T02	AT4G00150.1	GRAS family transcription factor
	GRMZM2G098800_T02	AT4G00150.1	GRAS family transcription factor
	GRMZM2G098800_T01	AT4G00150.1	GRAS family transcription factor
	GRMZM5G825321_T01	AT4G00150.1	GRAS family transcription factor
	GRMZM2G037792_T01	AT4G00150.1	GRAS family transcription factor
	GRMZM2G079470_T01	AT4G00150.1	GRAS family transcription factor
	GRMZM2G051785_T01	AT4G00150.1	GRAS family transcription factor
	GRMZM2G176124_T01	AT4G00150.1	GRAS family transcription factor
	GRMZM2G418899_T02	AT4G00150.1	GRAS family transcription factor
	GRMZM2G418899_T01	AT4G00150.1	GRAS family transcription factor
	GRMZM2G011947_T01	AT4G00150.1	GRAS family transcription factor
	AC187788.3_FGT008	AT4G00150.1	GRAS family transcription factor
	GRMZM2G110579_T01	AT4G00150.1	GRAS family transcription factor
	GRMZM2G317338_T01		
	GRMZM2G060265_T01	AT4G00150.1	GRAS family transcription factor
miR172b(isoMIR)-miR172c*(isoMIR)	GRMZM2G416701_T01_j_1		
	GRMZM5G862109_T02	AT2G28550.3	related to AP2.7
	GRMZM5G862109_T03	AT4G36920.1	Integrase-type DNA-binding superfamily protein
	GRMZM2G700665_T01	AT5G60120.2	target of early activation tagged (EAT) 2
	GRMZM2G700665_T01_j_1		
	GRMZM2G700665_T03	AT2G28550.3	related to AP2.7
	GRMZM2G700665_T02	AT2G28550.1	related to AP2.7
	GRMZM2G176175_T02	AT2G28550.3	related to AP2.7
	GRMZM2G176175_T01	AT2G28550.3	related to AP2.7

	GRMZM5G862109_T01	AT2G28550.3	related to AP2.7
	GRMZM2G416725_T01		
	siRNA_Z27kG1_07922		
	GRMZM5G878615_T03_j_1		
	GRMZM5G878615_T04		
	GRMZM5G878615_T07		
	GRMZM5G878615_T03		
	GRMZM5G878615_T02		
	GRMZM5G878615_T06	AT3G53750.1	actin 3
	GRMZM5G878615_T05		
	GRMZM2G017847_T01	AT3G12110.1	actin-11
	GRMZM2G383472_T01		
miR319c(isoMIR)	GRMZM2G115516_T01	AT4G18390.1	TEOSINTE BRANCHED 1, cycloidea and PCF transcription factor 2
	GRMZM2G089361_T01_O_1		
	GRMZM2G089361_T01	AT4G18390.1	TEOSINTE BRANCHED 1, cycloidea and PCF transcription factor 2
	GRMZM2G115516_T02		
	GRMZM2G020805_T01	AT4G18390.1	TEOSINTE BRANCHED 1, cycloidea and PCF transcription factor 2
	GRMZM2G020805_T01_j_1		
	GRMZM2G028054_T02_j_1		
	GRMZM2G028054_T01	AT3G11440.1	myb domain protein 65
	GRMZM2G028054_T02	AT3G11440.1	myb domain protein 65
	GRMZM2G028054_T03	AT3G11440.1	myb domain protein 65
miR390a-miR390b	GRMZM2G304745_T01_j_1		
	GRMZM2G304745_T01	AT1G63430.1	Leucine-rich repeat protein kinase family protein
miR393a*(isoMIR)-miR393b(isoMIR)-miR393c*(isoMIR)	GRMZM5G848945_T02_j_1		
	GRMZM5G848945_T02	AT3G26810.1	auxin signaling F-box 2
	GRMZM2G135978_T01	AT3G62980.1	F-box/RNI-like superfamily protein
miR394a-miR394b	GRMZM2G119650_T01	AT1G27340.1	Galactose oxidase/kelch repeat superfamily protein
	GRMZM2G064954_T01	AT1G27340.1	Galactose oxidase/kelch repeat superfamily protein
miR395a(isoMIR)-miR395e(isoMIR)-miR395h(isoMIR)	GRMZM2G042171_T01_j_1		
	GRMZM2G042171_T01	AT5G10180.1	sulphate transporter 2;1
	GRMZM2G149952_T01	AT4G14680.1	Pseudouridine synthase/archaeosine transglycosylase-like family protein
	GRMZM2G051270_T01	AT4G14680.1	Pseudouridine synthase/archaeosine transglycosylase-like family protein
	GRMZM2G051270_T03	AT4G14680.1	Pseudouridine synthase/archaeosine transglycosylase-like family protein
	GRMZM2G051270_T02_j_1		
	GRMZM2G051270_T02	AT3G22890.1	ATP sulfurylase 1
miR396c(isoMIR)-miR396d(isoMIR)	GRMZM2G041223_T01_j_1		
	GRMZM2G041223_T01	AT3G13960.1	growth-regulating factor 5
	GRMZM2G119359_T01_j_1		
	GRMZM2G098594_T03_j_1		
	GRMZM2G033612_T02_j_1		
	GRMZM2G098594_T04		
	GRMZM2G034876_T01	AT3G13960.1	growth-regulating factor 5
	GRMZM5G850129_T04	AT3G13960.1	growth-regulating factor 5
	GRMZM2G129147_T02		
	GRMZM2G018414_T01	AT4G37740.1	growth-regulating factor 2
	GRMZM2G119359_T01		
	GRMZM2G067743_T01	AT4G37740.1	growth-regulating factor 2
	GRMZM5G850129_T02	AT3G13960.1	growth-regulating factor 5
	GRMZM2G105335_T01	AT3G13960.1	growth-regulating factor 5
	GRMZM2G105335_T02	AT3G13960.1	growth-regulating factor 5
	GRMZM2G098594_T01	AT3G13960.1	growth-regulating factor 5
	GRMZM2G098594_T02	AT3G13960.1	growth-regulating factor 5
	GRMZM2G034876_T02	AT3G13960.1	growth-regulating factor 5
	GRMZM5G850129_T03	AT3G13960.1	growth-regulating factor 5
	GRMZM2G067743_T02	AT3G13960.1	growth-regulating factor 5
	GRMZM2G033612_T02	AT4G37740.1	growth-regulating factor 2
	GRMZM5G850129_T01	AT3G13960.1	growth-regulating factor 5
	GRMZM2G129147_T01	AT3G13960.1	growth-regulating factor 5
	GRMZM2G018414_T02	AT4G37740.1	growth-regulating factor 2
	GRMZM2G098594_T06	AT3G13960.1	growth-regulating factor 5
	GRMZM2G034876_T03	AT3G13960.1	growth-regulating factor 5
	GRMZM2G098594_T03		
	GRMZM5G893117_T01	AT3G13960.1	growth-regulating factor 5
	GRMZM2G045977_T01	AT2G36400.1	growth-regulating factor 3
	GRMZM2G124566_T01	AT2G36400.1	growth-regulating factor 3
	GRMZM2G124566_T02	AT2G36400.1	growth-regulating factor 3
	lncRNA_Z27kG1_06923		
miR397b(isoMIR)	GRMZM2G072808_T01	AT5G60020.1	laccase 17
	GRMZM2G400390_T01	AT3G09220.1	laccase 7
	GRMZM2G132169_T02	AT5G05390.1	laccase 12
	GRMZM2G336337_T01	AT5G05390.1	laccase 12
	GRMZM2G132169_T01	AT5G05390.1	laccase 12
	GRMZM2G132169_T01_j_1		
miR399a-miR399c-*miR399h*	TCONS_00124738		

	TCONS_00124738		
	GRMZM2G165734_T01		
miR399b	GRMZM2G165734_T01		
miR399d(miRNA*)	GRMZM2G165734_T01		
	GRMZM2G125378_T01	AT1G60170.1	pre-mRNA processing ribonucleoprotein binding region-containing protein
miR399e-miR399j-*miR399i*	GRMZM2G165734_T01		
	TCONS_00124738		
	TCONS_00124738		
miR399f	GRMZM2G068186_T01		
miR408b.1-*miR408a*	GRMZM5G866053_T01	AT2G02850.1	plantacyanin
	GRMZM2G004012_T01	AT2G02850.1	plantacyanin
miR408b.2(isoMIR)	GRMZM2G004012_T01	AT2G02850.1	plantacyanin
	GRMZM2G352678_T01	AT2G02850.1	plantacyanin
	GRMZM5G866053_T01	AT2G02850.1	plantacyanin
	GRMZM2G097851_T01	AT2G32300.1	uclacyanin 1
miR529	GRMZM2G101511_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	AC233751.1_FGT002	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G101511_T02	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G414805_T03_j_1		
	GRMZM2G414805_T01		
	GRMZM2G414805_T03	AT1G27370.1	squamosa promoter binding protein-like 10
	GRMZM2G414805_T05	AT1G27370.1	squamosa promoter binding protein-like 10
	GRMZM2G414805_T02	AT5G43270.1	squamosa promoter binding protein-like 2
	GRMZM2G414805_T07		
	GRMZM2G414805_T04	AT1G27370.1	squamosa promoter binding protein-like 10
	GRMZM2G160917_T01	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G061734_T01	AT5G50670.1	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
	GRMZM2G460544_T01	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G160917_T02	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G160917_T03	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G307588_T01	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G126018_T01	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM2G126018_T02	AT2G42200.1	squamosa promoter binding protein-like 9
	GRMZM5G806833_T01		
	GRMZM5G806833_T01_j_1		
	GRMZM2G148467_T02		
	GRMZM2G148467_T01	AT1G27370.1	squamosa promoter binding protein-like 10
miR-NEW1	GRMZM2G171279_T01_j_1		
	GRMZM2G020766_T01	AT3G02050.1	K+ uptake transporter 3
	GRMZM2G114704_T01_j_1		
	GRMZM2G046909_T01	AT3G50910.1	
	GRMZM2G076468_T01	AT5G61650.1	CYCLIN P4;2
	GRMZM2G150674_T01		
	GRMZM2G314692_T04	AT4G13400.1	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
	GRMZM2G099297_T01	AT4G35160.1	O-methyltransferase family protein
	GRMZM2G314692_T03	AT4G13400.1	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
	GRMZM2G314692_T03_j_1		
	GRMZM2G123703_T01		
	GRMZM2G349651_T01	AT5G59970.1	Histone superfamily protein
	GRMZM2G151997_T01	AT5G01750.2	Protein of unknown function (DUF567)
	GRMZM2G314692_T01	AT4G13400.1	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
	GRMZM2G054481_T01	AT5G58375.1	Methyltransferase-related protein
	GRMZM2G449257_T01	AT4G31730.1	glutamine dumper 1
	GRMZM2G314692_T02	AT4G13400.1	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
	GRMZM2G349651_T01_j_1		
	GRMZM2G149375_T01	AT2G19090.1	Protein of unknown function (DUF630 and DUF632)
	GRMZM2G510905_T01		
	GRMZM2G484444_T01		
	GRMZM2G157243_T01		
	GRMZM2G541399_T01		
	GRMZM2G112986_T01		
	GRMZM2G432757_T01		
	GRMZM2G100253_T01		
	GRMZM5G854179_T01		
	GRMZM2G532329_T01		
	GRMZM2G022792_T01		
	GRMZM2G009136_T01	AT1G68090.1	annexin 5
	GRMZM2G559355_T01		
	GRMZM2G582910_T01		
	GRMZM2G061728_T01_j_1		
	GRMZM2G122767_T01	AT3G18190.1	TCP-1/cpn60 chaperonin family protein
	GRMZM2G171022_T02_j_1		

	GRMZM2G061728_T01	AT5G26770.1	
	GRMZM2G038384_T01	AT3G19460.1	Reticulon family protein
	GRMZM2G074423_T01	AT3G10250.1	Plant protein 1589 of unknown function
	GRMZM2G038384_T01_j_1		
	GRMZM2G171022_T02	AT5G39250.1	F-box family protein
	GRMZM2G074423_T01_j_1		
	GRMZM2G113347_T01		
	GRMZM2G171022_T01	AT5G39250.1	F-box family protein
	GRMZM2G171236_T02		
	GRMZM2G137239_T01	AT2G43040.1	tetratricopeptide repeat (TPR)-containing protein
	GRMZM2G171236_T03	AT3G03070.1	NADH-ubiquinone oxidoreductase-related
	GRMZM2G120084_T01		
	GRMZM2G578161_T01		
miR-NEW2	GRMZM2G474537_T01	AT5G20610.1	
	GRMZM2G137541_T01	AT1G68920.1	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
miR-NEW4	GRMZM2G148773_T01	AT1G34320.1	Protein of unknown function (DUF668)
	GRMZM2G011731_T02	AT4G10080.1	
	GRMZM2G011731_T01	AT4G10080.1	
	siRNA_Z27kG1_23954		
	GRMZM2G146490_T01	AT1G80420.1	BRCT domain-containing DNA repair protein
	GRMZM2G146490_T02	AT1G80420.1	BRCT domain-containing DNA repair protein
miR-NEW6	GRMZM2G028228_T01		
miR-NEW10a.2 - miR-NEW10b.2	GRMZM2G478553_T01	AT1G72310.1	RING/U-box superfamily protein
	GRMZM2G412850_T01		
	GRMZM2G167151_T01		
	GRMZM5G884800_T01		
	GRMZM2G040592_T01		
miR-NEW10c.2 - miR-NEW10d.2	GRMZM2G384536_T01		
miR-NEW11	GRMZM2G338785_T01		
miR-NEW12	GRMZM2G100709_T01	AT2G40260.1	Homeodomain-like superfamily protein
miR-NEW13	GRMZM2G088349_T01_j_1		
	siRNA_Z27kG1_20468		
	GRMZM2G088349_T01_j_1		
miR-NEW14	GRMZM2G119322_T01		
miR-NEW15	GRMZM2G062567_T01		
	GRMZM2G168909_T06_j_1		
	GRMZM2G168909_T05	AT2G26310.2	Chalcone-flavanone isomerase family protein
	GRMZM2G168909_T04	AT5G14105.1	
	GRMZM2G168909_T06	AT2G26310.2	Chalcone-flavanone isomerase family protein
miR-NEW18	GRMZM2G484653_T01		
	GRMZM2G081541_T01	AT5G60790.1	ABC transporter family protein
miR-NEW19	GRMZM2G134753_T01_j_1		
	GRMZM2G134753_T01	AT2G39340.1	SAC3/GANP/Nin1/mts3/eIF-3 p25 family
	GRMZM2G134753_T02	AT2G39340.1	SAC3/GANP/Nin1/mts3/eIF-3 p25 family
	GRMZM2G068255_T02	AT3G04500.1	RNA-binding (RRM/RBD/RNP motifs) family protein
	GRMZM2G068255_T01	AT3G04500.1	RNA-binding (RRM/RBD/RNP motifs) family protein
	GRMZM2G032348_T02_j_1		
	GRMZM2G163418_T01	AT4G23810.1	WRKY family transcription factor
	GRMZM2G163418_T02	AT2G46400.1	WRKY DNA-binding protein 46
	GRMZM2G383240_T02_j_1		
	GRMZM2G383240_T07	AT4G24400.2	CBL-interacting protein kinase 8
	GRMZM2G020450_T01	AT2G24960.1	
	GRMZM2G455687_T01_j_1		
	GRMZM2G455687_T01	AT5G38840.1	SMAD/FHA domain-containing protein
	shRNA_Z27kG1_15518		
	GRMZM2G154900_T01		
	GRMZM2G020450_T01_j_1		
	GRMZM2G521946_T01		
miR-NEW20a -miR-NEW20b	GRMZM5G837999_T01		
	GRMZM2G081541_T01	AT5G60790.1	ABC transporter family protein
	GRMZM2G347056_T01	AT2G44160.1	methylenetetrahydrofolate reductase 2
miR-NEW21	GRMZM2G151223_T01	AT2G01830.2	CHASE domain containing histidine kinase protein
	GRMZM2G380668_T01	AT1G65720.1	
	GRMZM2G056645_T01_j_1		
	GRMZM2G056645_T01	AT1G03060.1	Beige/BEACH domain ;WD domain, G-beta repeat protein
	GRMZM2G055116_T01		
	GRMZM2G144841_T01_j_1		
	GRMZM2G144841_T01	AT1G32370.2	tobamovirus multiplication 2B
	GRMZM2G115658_T03	AT5G13750.2	zinc induced facilitator-like 1
miR-NEW22	GRMZM2G364703_T01	AT1G14130.1	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
	GRMZM2G109464_T01	AT2G02960.1	RING/FYVE/PHD zinc finger superfamily protein
	GRMZM2G109464_T02	AT2G02960.1	RING/FYVE/PHD zinc finger superfamily protein
	GRMZM5G837999_T01		
	GRMZM2G392003_T02	AT4G31490.1	Coatomer, beta subunit
	GRMZM2G094699_T02	AT4G31490.1	Coatomer, beta subunit
miR-NEW28	GRMZM2G055116_T01		
miR-NEW29	GRMZM5G837999_T01		

	GRMZM2G174537_T01	AT4G00170.1	Plant VAMP (vesicle-associated membrane protein) family protein
	siRNA_Z27kG1_18385		
	GRMZM2G036123_T01	AT2G37690.1	phosphoribosylaminoimidazole carboxylase, putative / AIR carboxylase, putative
	GRMZM2G152925_T01	AT1G22450.1	cytochrome C oxidase 6B
	GRMZM2G151223_T01	AT2G01830.2	CHASE domain containing histidine kinase protein
miR-NEW30	GRMZM2G034551_T01	AT2G30080.1	ZIP metal ion transporter family
miR-NEW33	TCONS_00122037		
	GRMZM5G890787_T01		
	GRMZM2G504151_T01		
miR-NEW35	GRMZM2G151807_T03_j_1		
	GRMZM2G561630_T01_X_1		
	GRMZM2G056829_T03_j_1		
	GRMZM2G056829_T04		
	GRMZM2G169899_T04		
	GRMZM2G000741_T01	AT1G72820.1	Mitochondrial substrate carrier family protein
	GRMZM2G066755_T01	AT3G55960.1	Haloacid dehalogenase-like hydrolase (HAD) superfamily protein
	shRNA_Z27kG1_24208		
	shRNA_Z27kG1_14549		
	GRMZM2G102616_T01		
miR-NEW41	GRMZM2G144362_T03_j_1		
	GRMZM2G144362_T04	AT2G44420.1	protein N-terminal asparagine amidohydrolase family protein
	GRMZM2G144362_T03	AT2G44420.1	protein N-terminal asparagine amidohydrolase family protein
	GRMZM2G144362_T02	AT2G44420.1	protein N-terminal asparagine amidohydrolase family protein
	GRMZM2G144362_T05	AT2G44420.1	protein N-terminal asparagine amidohydrolase family protein
	GRMZM2G144362_T01	AT2G44420.1	protein N-terminal asparagine amidohydrolase family protein
miR-NEW45	GRMZM2G056526_T01		
miR-NEW46.1	GRMZM2G156006_T01	AT5G25190.1	Integrase-type DNA-binding superfamily protein
	AC186377.3_FGT006		
miR-NEW46.2	AC186377.3_FGT006		
miR-NEW48	GRMZM2G134329_T02	AT5G67610.2	Uncharacterized conserved protein (DUF2215)
	GRMZM2G134329_T01_j_1		
	GRMZM2G134329_T03	AT5G67610.2	Uncharacterized conserved protein (DUF2215)
	GRMZM2G134329_T04	AT5G67610.2	Uncharacterized conserved protein (DUF2215)
	GRMZM2G134329_T01	AT5G67610.2	Uncharacterized conserved protein (DUF2215)
	GRMZM2G389462_T01	AT5G42340.1	Plant U-Box 15
	GRMZM5G878732_T01	AT1G49850.1	RING/U-box superfamily protein
	GRMZM5G878732_T02		
	AC225718.2_FGT005	AT3G28920.1	homeobox protein 34
	GRMZM5G878732_T01	AT1G49850.1	RING/U-box superfamily protein
	GRMZM2G098420_T03	AT5G17290.1	autophagy protein Apg5 family
	GRMZM2G098420_T02	AT5G17290.1	autophagy protein Apg5 family
	GRMZM2G098420_T01	AT5G17290.1	autophagy protein Apg5 family
	GRMZM2G098420_T01_j_1		
	AC207358.3_FGT003		
	GRMZM2G318689_T01	AT3G04580.1	Signal transduction histidine kinase, hybrid-type, ethylene sensor
	GRMZM2G068688_T01	AT1G05820.1	SIGNAL PEPTIDE PEPTIDASE-LIKE 5
	GRMZM2G068688_T02	AT1G05820.1	SIGNAL PEPTIDE PEPTIDASE-LIKE 5
	GRMZM2G542515_T01		
	GRMZM2G082823_T01	AT1G12040.1	leucine-rich repeat/extensin 1
	GRMZM2G475170_T01	AT2G28670.1	Disease resistance-responsive (dirigent-like protein) family protein
	GRMZM2G130358_T02	AT1G75350.1	Ribosomal protein L31
	GRMZM2G130358_T01	AT1G75350.1	Ribosomal protein L31
	GRMZM2G026556_T02	AT2G41370.1	Ankyrin repeat family protein / BTB/POZ domain-containing protein
	GRMZM2G358619_T01	AT5G47910.1	respiratory burst oxidase homologue D
	AC204619.3_FGT003		
	GRMZM5G894582_T01		
	GRMZM2G093716_T03	AT5G22370.1	P-loop containing nucleoside triphosphate hydrolases superfamily protein
	GRMZM2G093716_T02	AT5G22370.1	P-loop containing nucleoside triphosphate hydrolases superfamily protein
	GRMZM2G093716_T01	AT5G22370.1	P-loop containing nucleoside triphosphate hydrolases superfamily protein
	GRMZM2G093716_T05		
	GRMZM2G467466_T01	AT1G02020.1	nitroreductase family protein
	GRMZM2G485880_T01		
	GRMZM2G122108_T02	AT5G15330.1	SPX domain gene 4
	GRMZM2G122108_T03	AT5G15330.1	SPX domain gene 4
	GRMZM2G122108_T01	AT5G15330.1	SPX domain gene 4
	GRMZM2G122108_T01_j_1		

GRMZM5G815881_T01		
lncRNA_Z27kG1_02332		
GRMZM5G896805_T01		
lncRNA_Z27kG1_02332		
GRMZM2G031917_T01_j_1		
GRMZM2G171518_T03	AT2G43970.1	RNA-binding protein
GRMZM2G031917_T01		
GRMZM2G031917_T02		
GRMZM2G171518_T02	AT2G43970.1	RNA-binding protein
GRMZM2G171518_T01	AT2G43970.1	RNA-binding protein
GRMZM2G485880_T01		
GRMZM2G325580_T01	AT3G06390.1	Uncharacterised protein family (UPF0497)
GRMZM2G082508_T01	AT3G11810.1	
GRMZM2G386209_T01	AT5G47530.1	Auxin-responsive family protein
GRMZM2G386209_T02	AT5G35735.1	Auxin-responsive family protein
GRMZM2G386209_T03	AT5G35735.1	Auxin-responsive family protein
GRMZM2G386209_T05	AT5G35735.1	Auxin-responsive family protein
GRMZM2G090779_T01	AT1G51760.1	peptidase M20/M25/M40 family protein
GRMZM2G386209_T04	AT5G35735.1	Auxin-responsive family protein
GRMZM2G448687_T03	AT1G58250.1	Golgi-body localisation protein domain ;RNA pol II promoter Fmp27 protein domain
GRMZM2G448687_T01	AT1G58250.1	Golgi-body localisation protein domain ;RNA pol II promoter Fmp27 protein domain
GRMZM2G034206_T01	AT4G24972.1	tapetum determinant 1
GRMZM2G369839_T01		
GRMZM2G107162_T01		
GRMZM5G827174_T01	AT1G10020.1	Protein of unknown function (DUF1005)
GRMZM2G054020_T02		
GRMZM2G024838_T01_O_1		
GRMZM2G487776_T01	AT2G16190.1	
GRMZM2G024838_T01	AT5G13510.1	Ribosomal protein L10 family protein
GRMZM2G092797_T01	AT3G56990.1	embryo sac development arrest 7
GRMZM2G054020_T01	AT2G03510.1	SPFH/Band 7/PHB domain-containing membrane-associated protein family
GRMZM2G173724_T01	AT5G05350.1	PLAC8 family protein
GRMZM2G399333_T01	AT5G11550.1	ARM repeat superfamily protein
GRMZM2G371159_T01	AT1G02900.1	rapid alkalinization factor 1
GRMZM2G074974_T02		
GRMZM2G074974_T01		
GRMZM2G092165_T02	AT2G40780.1	Nucleic acid-binding, OB-fold-like protein
GRMZM5G816314_T01	AT2G44940.1	Integrase-type DNA-binding superfamily protein
GRMZM2G318530_T01		
GRMZM2G360023_T01		
GRMZM5G822928_T01		
GRMZM5G873917_T01	AT3G52710.1	
GRMZM2G092165_T01	AT2G40780.1	Nucleic acid-binding, OB-fold-like protein
GRMZM2G306105_T01		
GRMZM2G306105_T02		
GRMZM5G868875_T02	AT1G11510.1	DNA-binding storekeeper protein-related transcriptional regulator
GRMZM2G454189_T01		
GRMZM2G367459_T01		
GRMZM2G382591_T01		
GRMZM2G306105_T01		
GRMZM2G127853_T01	AT1G69295.1	plasmodesmata callose-binding protein 4
GRMZM5G868875_T01	AT1G11510.1	DNA-binding storekeeper protein-related transcriptional regulator
GRMZM2G028007_T01	AT1G23270.1	
GRMZM2G382591_T01		
AC214448.3_FGT007_O_1		
GRMZM2G118515_T01_j_1		
GRMZM2G073943_T01	AT1G33800.1	Protein of unknown function (DUF579)
GRMZM2G382591_T02		
GRMZM2G010505_T01_j_1		
GRMZM2G118515_T02	AT5G48970.1	Mitochondrial substrate carrier family protein
GRMZM2G139031_T01	AT1G21710.1	8-oxoguanine-DNA glycosylase 1
GRMZM2G118515_T01	AT5G48970.1	Mitochondrial substrate carrier family protein
GRMZM2G073943_T01_O_1		
GRMZM2G166430_T01	AT1G79060.1	
GRMZM2G012999_T01	AT5G07900.1	Mitochondrial transcription termination factor family protein
GRMZM2G010505_T01	AT3G06910.1	UB-like protease 1A
GRMZM5G877941_T02	AT5G06580.1	FAD-linked oxidases family protein
GRMZM2G428470_T01	AT2G44730.1	Alcohol dehydrogenase transcription factor Myb/SANT-like family protein
GRMZM5G877941_T01	AT5G06580.1	FAD-linked oxidases family protein
GRMZM5G877941_T03	AT5G06580.1	FAD-linked oxidases family protein
GRMZM2G034430_T01	AT5G11350.1	DNAse I-like superfamily protein
GRMZM5G877941_T03_j_1		
GRMZM2G000739_T01	AT5G40850.1	urophorphyrin methylase 1
AC214448.3_FGT007_O_1		
GRMZM2G033829_T02_j_1		
GRMZM2G033829_T01	AT5G60640.1	PDI-like 1-4
GRMZM2G035601_T01_O_1		
GRMZM2G035601_T01	AT3G58030.1	RING/U-box superfamily protein

GRMZM2G152419_T01	AT2G32480.1	ARABIDOPSIS SERIN PROTEASE
GRMZM2G357804_T03	AT3G11540.1	Tetratricopeptide repeat (TPR)-like superfamily protein
GRMZM2G357804_T02	AT3G11540.1	Tetratricopeptide repeat (TPR)-like superfamily protein
GRMZM2G000739_T02	AT5G40850.1	urophorphyrin methylase 1
GRMZM2G357804_T01_j_1		
GRMZM2G357804_T01	AT3G11540.1	Tetratricopeptide repeat (TPR)-like superfamily protein
GRMZM2G000739_T01_j_1		
GRMZM2G116086_T01	AT3G03940.1	Protein kinase family protein
GRMZM2G026639_T01	AT5G47400.1	
GRMZM2G021483_T01_j_1		
GRMZM2G479717_T01	AT4G37700.1	
GRMZM2G024264_T06	AT3G12490.2	cystatin B
GRMZM2G386095_T01	AT3G13850.1	LOB domain-containing protein 22
GRMZM2G024264_T05	AT3G12490.2	cystatin B
GRMZM2G024264_T01	AT3G12490.2	cystatin B
GRMZM2G024264_T04	AT3G12490.2	cystatin B
GRMZM2G024264_T03	AT3G12490.2	cystatin B
GRMZM2G133620_T02	AT3G12490.2	cystatin B
GRMZM2G174558_T01	AT1G28280.1	VQ motif-containing protein
GRMZM2G009892_T02	AT5G61430.1	NAC domain containing protein 100
GRMZM2G072274_T01	AT2G45430.1	AT-hook motif nuclear-localized protein 22
GRMZM2G133620_T01	AT3G12490.2	cystatin B
GRMZM2G009892_T01	AT5G61430.1	NAC domain containing protein 100
GRMZM2G359038_T03	AT1G80300.1	nucleotide transporter 1
GRMZM2G009892_T03	AT5G61430.1	NAC domain containing protein 100
GRMZM2G359038_T01_j_1		
GRMZM5G894569_T01		
GRMZM2G005939_T05	AT1G27660.1	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
GRMZM2G027043_T01	AT2G45150.3	cytidinediphosphate diacylglycerol synthase 4
GRMZM2G005939_T02	AT1G05710.1	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
GRMZM2G005939_T01	AT1G05710.1	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
GRMZM2G005939_T06	AT1G05710.1	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
GRMZM2G005939_T01_j_1		
GRMZM2G101080_T03	AT2G38800.1	Plant calmodulin-binding protein-related
GRMZM2G005939_T04	AT1G27660.1	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
GRMZM2G101080_T01_j_1		
GRMZM2G005939_T03	AT1G27660.1	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
GRMZM2G024264_T02	AT3G12490.2	cystatin B
GRMZM2G158831_T01	AT2G20815.1	Family of unknown function (DUF566)
GRMZM2G130868_T02	AT1G71110.1	
GRMZM2G323309_T01	AT2G15630.1	Pentatricopeptide repeat (PPR) superfamily protein
GRMZM2G359038_T01	AT1G80300.1	nucleotide transporter 1
GRMZM2G130868_T01	AT1G71110.1	
GRMZM2G144843_T01	AT1G59720.1	Tetratricopeptide repeat (TPR)-like superfamily protein
GRMZM2G359038_T02	AT1G80300.1	nucleotide transporter 1
GRMZM2G358219_T01		
GRMZM2G028763_T02	AT3G13062.2	Polyketide cyclase/dehydrase and lipid transport superfamily protein
GRMZM2G028763_T01	AT3G13062.2	Polyketide cyclase/dehydrase and lipid transport superfamily protein
GRMZM2G403636_T01	AT2G42520.1	P-loop containing nucleoside triphosphate hydrolases superfamily protein
GRMZM2G091293_T01		
GRMZM2G158520_T01	AT1G56020.1	
AC226227.2_FGT003	AT4G06744.1	Leucine-rich repeat (LRR) family protein
GRMZM2G102790_T01_O_1		
GRMZM2G134756_T02	AT2G33840.1	Tyrosyl-tRNA synthetase, class Ib, bacterial/mitochondrial
GRMZM2G134756_T01	AT2G33840.1	Tyrosyl-tRNA synthetase, class Ib, bacterial/mitochondrial
GRMZM2G102790_T01	AT5G16770.1	myb domain protein 9
GRMZM2G424491_T01_j_1		
GRMZM5G836475_T01		
GRMZM2G057247_T01_j_1		
GRMZM2G057247_T01	AT5G67470.1	formin homolog 6
GRMZM5G871489_T01		
GRMZM2G378217_T01		
GRMZM2G134759_T01	AT3G02790.1	zinc finger (C2H2 type) family protein
GRMZM2G086628_T01	AT2G32300.1	uclacyanin 1
GRMZM2G134759_T02	AT3G02790.1	zinc finger (C2H2 type) family protein
GRMZM2G436688_T01		
GRMZM2G047255_T01	AT4G18570.1	Tetratricopeptide repeat (TPR)-like superfamily protein
GRMZM2G158034_T04	AT4G00100.1	ribosomal protein S13A

	GRMZM2G158034_T01	AT4G00100.1	ribosomal protein S13A
	GRMZM2G452695_T01_O_1		
	GRMZM2G452695_T01	AT5G05550.2	sequence-specific DNA binding transcription factors
	GRMZM2G407513_T01		
	GRMZM2G121878_T01	AT5G14740.2	carbonic anhydrase 2
	GRMZM2G121878_T02	AT5G14740.2	carbonic anhydrase 2
	GRMZM2G121878_T03	AT1G70410.1	beta carbonic anhydrase 4
	GRMZM2G121878_T05	AT5G14740.1	carbonic anhydrase 2
	GRMZM2G121878_T01_j_1		
	GRMZM2G121878_T06	AT5G14740.1	carbonic anhydrase 2
miR-NEW56	GRMZM2G098331_T01	AT1G20080.1	Calcium-dependent lipid-binding (CaLB domain) family protein
	GRMZM2G098331_T02	AT1G20080.1	Calcium-dependent lipid-binding (CaLB domain) family protein
	GRMZM2G034551_T01	AT2G30080.1	ZIP metal ion transporter family
	GRMZM5G837999_T01		
miR-NEW58	GRMZM5G872943_T01		

*transcript annotation = transcriptome assembly reconstructed from our RNA-seq experiment

Appendix E

Co-occupancy analysis results

	sRNA loci category	genomic feature													
		protein-coding genes				TE transcripts			lncRNA transcripts				REPEATS
		GENES	EXONS	INTRONS	2-kb UPSTREAM REGIONS	2-kb DOWNSTREAM REGIONS	TRANSCRIPTS	2-kb UPSTREAM REGIONS	2-kb DOWNSTREAM REGIONS	TRANSCRIPTS	2-kb UPSTREAM REGIONS	2-kb DOWNSTREAM REGIONS	
log_2(observed/ expected)*	20-nt HP	2.67	3.67	1.35	−inf	0.45	−inf	0.34	0.39	2.47	−inf	−inf	-1.07
	21-nt HP	2.82	2.53	2.97	0.69	1.38	-0.20	0.99	0.92	3.50	0.82	2.00	-0.55
	22-nt HP	2.27	-0.46	2.86	-0.54	0.13	1.23	0.75	0.82	1.21	-0.53	-0.99	0.03
	23-nt HP	1.40	0.89	1.64	1.90	1.60	-0.14	0.41	1.19	−inf	1.60	2.26	-0.10
	24-nt HP	0.54	-0.15	0.83	2.36	2.25	0.22	1.51	1.28	-0.07	1.63	1.60	-0.34
	20-nt non-HP	2.40	3.74	-1.22	2.25	1.27	0.91	−inf	0.03	1.77	0.80	−inf	-1.42
	21-nt non-HP	2.30	3.08	1.51	2.04	1.58	-0.47	0.07	0.54	2.58	1.72	2.17	-0.52
	22-nt non-HP	0.91	-1.01	1.44	-0.69	-0.99	0.86	0.24	0.23	-0.08	-0.83	-1.59	0.15
	23-nt non-HP	0.85	-0.61	1.32	1.37	2.20	-1.49	-0.46	0.92	0.44	0.75	0.61	-0.13
	24-nt non-HP	0.16	-0.16	0.32	2.90	2.38	0.13	1.69	1.43	-0.52	1.94	1.70	-0.39
% of sRNA loci overlapping with genomic features	20-nt HP	56.00	44.00	28.00	0.00	4.00	0.00	4.00	4.00	4.00	0.00	0.00	44.00
	21-nt HP	51.87	22.99	37.43	6.95	8.56	3.21	5.35	5.35	6.42	3.21	5.88	65.78
	22-nt HP	20.36	3.56	17.90	3.39	3.48	4.75	5.00	4.58	0.85	1.87	0.85	93.55
	23-nt HP	22.31	6.61	17.36	13.22	11.57	2.48	4.13	5.79	0.00	4.13	6.61	83.47
	24-nt HP	11.71	3.73	8.76	19.79	17.69	2.98	8.59	7.40	0.70	4.96	4.82	75.19
	20-nt non-HP	28.57	25.00	3.57	7.14	10.71	7.14	0.00	7.14	3.57	3.57	0.00	53.57
	21-nt non-HP	24.17	9.82	15.72	7.66	5.89	2.75	5.30	3.54	1.77	2.16	2.95	82.71
	22-nt non-HP	9.96	0.69	9.53	1.27	1.42	3.41	2.98	2.86	0.42	0.57	0.48	98.38
	23-nt non-HP	14.79	3.22	11.90	10.61	8.04	1.29	3.22	4.50	1.29	2.25	2.57	82.64
	24-nt non-HP	10.43	3.23	7.75	23.06	17.95	2.99	9.31	8.26	0.55	5.08	4.69	76.67
% of genomic features overlapping with sRNA loci	20-nt HP	0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
	21-nt HP	0.26	0.02	0.05	0.04	0.05	0.02	0.04	0.04	0.08	0.04	0.07	0.02
	22-nt HP	0.56	0.03	0.14	0.11	0.12	0.19	0.21	0.20	0.07	0.14	0.07	0.19
	23-nt HP	0.07	0.00	0.01	0.04	0.04	0.01	0.02	0.03	0.00	0.03	0.05	0.02
	24-nt HP	5.01	0.35	0.96	9.10	8.56	1.68	5.32	4.66	0.79	5.58	5.49	2.28
	20-nt non-HP	0.02	0.00	0.00	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.00	0.00
	21-nt non-HP	0.30	0.03	0.05	0.10	0.09	0.05	0.10	0.07	0.06	0.07	0.10	0.07
	22-nt non-HP	1.70	0.09	0.45	0.49	0.55	1.31	1.40	1.33	0.24	0.58	0.43	1.43
	23-nt non-HP	0.12	0.01	0.02	0.09	0.07	0.01	0.04	0.05	0.03	0.05	0.05	0.04
	24-nt non-HP	20.61	1.84	4.29	51.75	42.86	8.89	28.90	26.10	3.07	29.29	27.44	13.48

*the observed number of non-redundant overlapping nucleotides between a sRNA loci category and a genomic feature was obtained with a customized Perl script provided by Dr. Axtell MJ;
the expected number of non-redundant overlapping nucleotides was calculated as follows: ((total non-redundant nt of genomic feature/genome size)*(total non-redundant nt of sRNA loci category/genome size))* genome size

Appendix F

Differentially expressed sRNA loci (not *MIRNA* loci) in stress conditions and/or at the developmental stage of plants at +7

C = control; D = drought stress; S = salinity stress; D+S = drought+salinity stress.
+0=ten days of treatment; +7=seven days of recovery.

*log2 fold change = only values with FDR <1% are reported.
**UI = Uniqueness Index of sRNA loci as defined by ShortStack program.
***location = genomic location of sRNA loci based on the transcriptome reannotation obtained from our RNA-seq experiment:
"exon" or "intron" = indicates that the locus is located within an exon or an intron for its entire length;
"exon-intron" = indicates that the locus overlaps with an intron and an exon of a gene;
"antisense" = indicates that the locus is antisense to a gene;
"intergenic" = indicates that the locus is located between genes for its entire length;
"genic-intergenic" = indicates that the locus partially overlaps with a gene.
****gene annotation = transcriptome assembly reconstructed from our RNA-seq experiment.

DE locus	differential expression						nt size class	UI**	precursor	repeats masking	phasing	name	location***
	stress conditions and/or development effect				genotype effect								
	genotype	stress vs control at +0:	+7 vs +0 condition:	log_2 fold change*	*rmr6-1* vs wt C at +0	log_2 fold change*							
Cluster_4704	wt		D	-5.63			24	0.9527	HP	DTC;RLC			antisense
Cluster_11047	*rmr6-1*	D+S		-3.71			22	0.7609	HP	RLC			intron
	rmr6-1		C	-3.00									
Cluster_19729	wt	D		1.31	downregulated	-6.20	24	0.5075					intergenic
Cluster_22470	wt	D		-5.67	downregulated	-5.65	24	1		DTC			intergenic
Cluster_45368	wt	D		1.63	upregulated	2.19	22	0.0762		45S rDNA			intron
	wt		C	2.00									
Cluster_57354	wt		D	1.82			21	0.1261					intron
Cluster_63370	wt	D		2.02	upregulated	2.6	22	0.8159					genic-intergenic
	wt		C	1.69									
Cluster_63380	wt	D		1.60	upregulated	2.48	21	0.5391					exon
Cluster_69288	wt	D		-2.11	downregulated	-2.78	24	0.1593					intergenic
Cluster_83609	wt	D		-6.25			22	0.1281		RLC			intergenic
Cluster_91727	wt	D		-1.53			21	0.9987			yes		exon
Cluster_99151	*rmr6-1*	D		2.28			24	0.7041		RLC			exon-intron
	rmr6-1	D+S		2.73									
Cluster_108377	wt		D	-3.20	downregulated	-6.10	24	1					intron
Cluster_145222	wt	D		-3.24	downregulated	-6.26	24	0.9869					antisense
Cluster_147517	*rmr6-1*	D		2.49	upregulated	2.9	22	0.8721	HP	RLG			intergenic
	rmr6-1		C	2.60									
Cluster_166266	wt	D		-1.47			21	0.9024			yes	*TAS3c* tasiRNA locus	exon-intron
Cluster_178758	wt	D		1.23			24	0.9992	HP				exon-intron
Cluster_180696	wt	D		-1.83	downregulated	-8.23	24	1		MITE			intergenic
Cluster_182776	wt		D	-1.82	downregulated	-5.43	24	0.9989	HP				intergenic

DE locus	overlapping gene				
	overlapping gene ID****	AGPv3.20 biotype	AGPv3.20 description	*Arabidopsis* homolog	*Arabidopsis* annotation
Cluster_4704	GRMZM2G127087	protein_coding	(S)-beta-macrocarpene synthase	AT1G70080.1	Terpenoid cyclases/Protein prenyltransferases superfamily protein
Cluster_11047	GRMZM2G071264	protein_coding	Uncharacterized protein	AT5G11040.1	TRS120
Cluster_19729					
Cluster_22470					
Cluster_45368	GRMZM2G475017	protein_coding	Uncharacterized protein	AT2G47850.1	Zinc finger C-x8-C-x5-C-x3-H type family protein
Cluster_57354	GRMZM2G169116	protein_coding	Uncharacterized protein	AT1G79270.1	evolutionarily conserved C-terminal region 8
Cluster_63370	AC205008.4_FG001	low_confidence		AT3G12750.1	zinc transporter 1 precursor
Cluster_63380	GRMZM2G093276	protein_coding	ZIP zinc/iron transport family protein	AT1G05300.1	zinc transporter 5 precursor
Cluster_69288					
Cluster_83609					
Cluster_91727	GRMZM2G512113	protein_coding			
Cluster_99151	AC216891.3_FG004;AC216891.3_FG004_X	protein_coding; /			
Cluster_108377	GRMZM2G171650;AC205520.3_FG004	protein_coding;low_confidence	Uncharacterized protein;Uncharacterized protein	AT2G03710.3	K-box region and MADS-box transcription factor family protein
Cluster_145222	GRMZM2G492315	low_confidence			
Cluster_147517					
Cluster_166266	GRMZM2G084821;GRMZM5G833991	protein_coding;protein_codin	Putative uncharacterized protein;Uncharacterized protein		
Cluster_178758	GRMZM2G536825	protein_coding			
Cluster_180696					
Cluster_182776					

Appendix G

Differentially expressed genes in *rmr6-1* mutant compared to wt (control conditions, after ten days of experiment)

*gene annotation = transcriptome assembly reconstructed from our RNA-seq experiment. Genes named as "Cluster…" and "XLOC…" were new loci, previously not annotated.
**log2 fold change = only values with FDR <5% are reported. "inf" indicates that genes were expressed only in *rmr6-1* mutant. "-inf" indicates that genes were expressed only in wt.

DE gene ID*	*rmr6-1*/wt	log2 fold change**	AGPv3.20 biotype	TE	lncRNA	Classical genes	MaizeGDB curated genes	ChromDB name	*Arabidopsis* homolog	*Arabidopsis* annotation
AC149829.2_FG002	upregulated	2.16	protein_coding							
AC177911.4_FG005	upregulated	1.57	protein_coding							
AC184127.3_FG003	upregulated	inf	protein_coding							
AC184133.3_FG001	upregulated	inf	protein_coding						AT5G64130.1	cAMP-regulated phosphoprotein 19-related protein
AC185468.3_FG002	upregulated	inf	low_confidence							
AC189077.3_FG002	upregulated	inf	low_confidence		lncRNA					
AC190516.1_FG001	upregulated	inf	low_confidence							
AC190933.3_FG004	upregulated	3.16	protein_coding						AT2G47710.1	Adenine nucleotide alpha hydrolases-like superfamily protein
AC191255.2_FG003	upregulated	inf	low_confidence							
AC192381.3_FG004	upregulated	inf	protein_coding							
AC193414.4_FG002	upregulated	inf	low_confidence		lncRNA					
AC194091.2_FG001	upregulated	inf	low_confidence						AT1G80680.1	SUPPRESSOR OF AUXIN RESISTANCE 3
AC194121.3_FG001	upregulated	inf	low_confidence							
AC197164.4_FG002	upregulated	inf	protein_coding						AT5G08020.1	RPA70-kDa subunit B
AC197442.3_FG002	upregulated	inf	low_confidence		lncRNA					
AC197699.3_FG004	upregulated	inf	low_confidence						AT3G47950.1	H(+)-ATPase 4
AC197705.4_FG003	upregulated	2.34	protein_coding						AT5G01300.1	PEBP (phosphatidylethanolamine-binding protein) family protein
AC198787.3_FG009	upregulated	inf	protein_coding							
AC199065.2_FG001	upregulated	inf	low_confidence		lncRNA					
AC199638.4_FG005_O_1	upregulated	inf			lncRNA					
AC200725.4_FG002	upregulated	inf	transposable_element	RLG;TXX						
AC201780.3_FG003	upregulated	inf	low_confidence							
AC202129.3_FG001	upregulated	inf	protein_coding		lncRNA					
AC202878.3_FG002	upregulated	inf	low_confidence		lncRNA					
AC203435.3_FG001	upregulated	inf	low_confidence		lncRNA					
AC205035.3_FG004	upregulated	inf	protein_coding		lncRNA				AT1G13270.2	methionine aminopeptidase 1B
AC205376.4_FG005	upregulated	inf	transposable_element	TXX	lncRNA					
AC205886.3_FG001	upregulated	2.37	protein_coding						AT5G58420.1	Ribosomal protein S4 (RPS4A) family protein
AC206637.3_FG001	upregulated	inf	low_confidence		lncRNA					
AC207655.2_FG003_O_1	upregulated	inf								
AC208040.3_FG002	upregulated	inf	transposable_element	TXX						
AC208213.3_FG003	upregulated	inf	low_confidence							
AC208614.3_FG001	upregulated	inf	low_confidence							
AC210173.4_FG005	upregulated	1.52	protein_coding						AT4G36220.1	ferulic acid 5-hydroxylase 1
AC213398.3_FG006	upregulated	inf	low_confidence						AT2G37250.1	adenosine kinase
AC213398.3_FG008	upregulated	inf	low_confidence							
AC213612.3_FG001	upregulated	1.60	protein_coding							
AC213629.3_FG009	upregulated	inf	transposable_element	TXX						
AC213651.3_FG005	upregulated	inf	protein_coding		lncRNA					
AC214634.3_FG003	upregulated	inf	low_confidence							
AC216891.3_FG004_X_1	upregulated	2.24								
AC217052.2_FG009	upregulated	inf	low_confidence		lncRNA					
AC217401.3_FG004	upregulated	inf	low_confidence						AT5G12400.1	DNA binding;zinc ion binding;DNA binding
AC217811.3_FG003	upregulated	1.79	protein_coding							
AC217840.3_FG005	upregulated	inf	low_confidence							
AC218964.2_FG003	upregulated	inf	protein_coding							
AC218998.2_FG008	upregulated	inf	low_confidence						AT4G11470.1	cysteine-rich RLK (RECEPTOR-like protein kinase) 31
AC218999.2_FG001_X_1	upregulated	inf			lncRNA					
AC225338.3_FG001	upregulated	inf	low_confidence							

DE gene ID*	*rmr6-1*/wt	log2 fold change**	AGPv3.20 biotype	TE	lncRNA	Classical genes	MaizeGDB curated genes	ChromDB name	*Arabidopsis* homolog	*Arabidopsis* annotation
AC231409.1_FG001	upregulated	inf	protein_coding							
AC235534.1_FG007	upregulated	3.92	protein_coding			ocl2			AT3G61150.1	homeodomain GLABROUS 1
AC235546.1_FG002	upregulated	inf	protein_coding						AT3G52960.1	Thioredoxin superfamily protein
AF466202.2_FG001	upregulated	2.81	low_confidence						AT2G19900.1	NADP-malic enzyme 1
ClusterV2_1	upregulated	1.96								
ClusterV2_18	upregulated	inf		RLC						
ClusterV2_34	upregulated	1.52								
ClusterV2_46	upregulated	2.58								
ClusterV2_49	upregulated	inf								
ClusterV2_8	upregulated	inf		RLC						
Cluster_11	upregulated	inf		RLX						
Cluster_139	upregulated	inf			lncRNA					
Cluster_146	upregulated	inf		DTA						
Cluster_147	upregulated	2.88		RLX						
Cluster_15	upregulated	inf								
Cluster_17	upregulated	5.46								
Cluster_175	upregulated	inf								
Cluster_181	upregulated	inf								
Cluster_209	upregulated	inf								
Cluster_226	upregulated	1.91		RLC						
Cluster_236	upregulated	inf								
Cluster_237	upregulated	inf		RLG						
Cluster_246	upregulated	1.52								
Cluster_249	upregulated	inf		RLX						
Cluster_25	upregulated	inf								
Cluster_26	upregulated	3.61								
Cluster_281	upregulated	inf		RLX						
Cluster_282	upregulated	inf		RLX						
Cluster_29	upregulated	1.35								
Cluster_298	upregulated	5.34								
Cluster_315	upregulated	inf		RLX						
Cluster_317	upregulated	inf								
Cluster_337	upregulated	inf								
Cluster_34	upregulated	inf		DTH						
Cluster_350	upregulated	inf		RLC;RLG						
Cluster_351	upregulated	inf								
Cluster_356	upregulated	inf								
Cluster_359	upregulated	inf		RLX						
Cluster_362	upregulated	inf		RLC						
Cluster_364	upregulated	inf			lncRNA					
Cluster_366	upregulated	inf								
Cluster_37	upregulated	inf								
Cluster_46	upregulated	inf		RLX						
Cluster_60	upregulated	inf		RLC						
Cluster_8	upregulated	inf		RLC						
Cluster_97	upregulated	inf								
GRMZM2G001375	upregulated	2.01	low_confidence		lncRNA					
GRMZM2G001508	upregulated	3.19	protein_coding						AT4G29260.1	HAD superfamily, subfamily IIIB acid phosphatase
GRMZM2G001908	upregulated	inf	transposable_element	TXX						

DE gene ID*	*rmr6-1*/wt	log2 fold change**	AGPv3.20 biotype	TE	lncRNA	Classical genes	MaizeGDB curated genes	ChromDB name	*Arabidopsis* homolog	*Arabidopsis* annotation
GRMZM2G001915	upregulated	1.99	protein_coding						AT5G15640.1	Mitochondrial substrate carrier family protein
GRMZM2G002656	upregulated	3.15	protein_coding							
GRMZM2G002678	upregulated	2.58	low_confidence						AT3G23560.1	MATE efflux family protein
GRMZM2G003426	upregulated	2.09	protein_coding							
GRMZM2G004036	upregulated	2.48	protein_coding						AT1G75290.1	NAD(P)-binding Rossmann-fold superfamily protein
GRMZM2G004160	upregulated	1.70	protein_coding						AT2G02850.1	plantacyanin
GRMZM2G004385	upregulated	inf	low_confidence						AT2G47850.2	Zinc finger C-x8-C-x5-C-x3-H type family protein
GRMZM2G004909	upregulated	2.23	protein_coding						AT3G18280.1	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
GRMZM2G005700	upregulated	inf	low_confidence						AT5G13640.1	phospholipid:diacylglycerol acyltransferase
GRMZM2G006884	upregulated	inf	protein_coding							
GRMZM2G007426	upregulated	inf	protein_coding		lncRNA					
GRMZM2G009987	upregulated	inf	protein_coding						AT5G58290.1	regulatory particle triple-A ATPase 3
GRMZM2G010636	upregulated	2.57	protein_coding						AT5G08020.1	RPA70-kDa subunit B
GRMZM2G010740	upregulated	2.30	protein_coding						AT3G54780.1	Zinc finger (C3HC4-type RING finger) family protein
GRMZM2G010791	upregulated	3.77	low_confidence						AT3G14940.1	phosphoenolpyruvate carboxylase 3
GRMZM2G012160	upregulated	2.47	protein_coding			psei2			AT3G12490.2	cystatin B
GRMZM2G012865	upregulated	1.52	protein_coding						AT5G25140.1	cytochrome P450, family 71, subfamily B, polypeptide 13
GRMZM2G013530	upregulated	2.69	protein_coding						AT5G48930.1	hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase
GRMZM2G014622	upregulated	inf	transposable_element	TXX						
GRMZM2G017418	upregulated	inf	low_confidence							
GRMZM2G018793	upregulated	3.08	protein_coding				AY111333		AT3G51000.1	alpha/beta-Hydrolases superfamily protein
GRMZM2G019095	upregulated	inf	low_confidence							
GRMZM2G019574	upregulated	inf	low_confidence						AT1G73370.1	sucrose synthase 6
GRMZM2G019708	upregulated	inf	low_confidence		lncRNA					
GRMZM2G020987	upregulated	inf	low_confidence						AT5G23320.1	homolog of yeast STE14 A
GRMZM2G022972	upregulated	4.22	protein_coding						AT1G73805.1	Calmodulin binding protein-like
GRMZM2G023113	upregulated	inf	low_confidence							
GRMZM2G023152	upregulated	1.84	protein_coding						AT4G35160.1	O-methyltransferase family protein
GRMZM2G023827	upregulated	3.54	protein_coding						AT4G11000.1	Ankyrin repeat family protein
GRMZM2G024411	upregulated	4.04	low_confidence						AT4G30790.1	
GRMZM2G024668	upregulated	1.99	protein_coding				pza03531		AT5G56030.1	heat shock protein 81-2
GRMZM2G024904	upregulated	inf	low_confidence							
GRMZM2G025322	upregulated	1.40	protein_coding						AT1G53670.1	methionine sulfoxide reductase B 1
GRMZM2G025761	upregulated	inf	low_confidence							
GRMZM2G025832	upregulated	1.87	protein_coding				pr1		AT5G07990.1	Cytochrome P450 superfamily protein
GRMZM2G025833	upregulated	2.37	protein_coding						AT3G23920.1	beta-amylase 1
GRMZM2G026143	upregulated	2.50	protein_coding		lncRNA					
GRMZM2G027478	upregulated	1.66	protein_coding		lncRNA				AT3G56290.1	
GRMZM2G027851	upregulated	1.37	protein_coding						AT3G05030.1	sodium hydrogen exchanger 2
GRMZM2G028535	upregulated	1.69	protein_coding						AT3G55610.1	delta 1-pyrroline-5-carboxylate synthase 2
GRMZM2G028640	upregulated	4.53	protein_coding						AT4G10300.1	RmlC-like cupins superfamily protein
GRMZM2G030583	upregulated	3.52	protein_coding				tps26		AT5G23960.1	terpene synthase 21
GRMZM2G030995	upregulated	1.68	transposable_element	TXX						
GRMZM2G031504	upregulated	1.49	protein_coding	DTA						
GRMZM2G031782	upregulated	1.75	low_confidence							
GRMZM2G032339	upregulated	3.42	protein_coding				mads4		AT1G69120.1	K-box region and MADS-box transcription factor family protein
GRMZM2G033967	upregulated	inf	low_confidence							

DE gene ID*	*rmr6-1*/wt	log2 fold change**	AGPv3.20 biotype	TE	lncRNA	Classical genes	MaizeGDB curated genes	ChromDB name	*Arabidopsis* homolog	*Arabidopsis* annotation
GRMZM2G034471	upregulated	1.53	protein_coding						AT2G46660.1	cytochrome P450, family 78, subfamily A, polypeptide 6
GRMZM2G035040	upregulated	inf	low_confidence		lncRNA				AT5G11790.1	N-MYC downregulated-like 2
GRMZM2G036288	upregulated	3.08	protein_coding				zim14		AT5G20900.1	jasmonate-zim-domain protein 12
GRMZM2G036290	upregulated	2.04	protein_coding						AT5G62790.1	1-deoxy-D-xylulose 5-phosphate reductoisomerase
GRMZM2G036980	upregulated	1.36	protein_coding							
GRMZM2G038825	upregulated	inf	transposable_element	TXX						
GRMZM2G038827	upregulated	inf	protein_coding		lncRNA					
GRMZM2G039268	upregulated	inf	low_confidence							
GRMZM2G039993	upregulated	1.50	protein_coding						AT5G66430.1	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
GRMZM2G040669	upregulated	inf	low_confidence							
GRMZM2G041619	upregulated	inf	transposable_element	TXX						
GRMZM2G043112	upregulated	inf	low_confidence							
GRMZM2G043369	upregulated	inf	low_confidence		lncRNA					
GRMZM2G044383	upregulated	2.23	protein_coding						AT1G10370.1	Glutathione S-transferase family protein
GRMZM2G044627	upregulated	inf	protein_coding						AT1G05510.1	Protein of unknown function (DUF1264)
GRMZM2G044851	upregulated	1.48	protein_coding						AT1G32450.1	nitrate transporter 1.5
GRMZM2G044992	upregulated	1.94	protein_coding							
GRMZM2G045473	upregulated	1.93	protein_coding						AT2G32830.1	phosphate transporter 1;5
GRMZM2G045638	upregulated	1.36	protein_coding		lncRNA					
GRMZM2G045916	upregulated	inf	transposable_element	TXX						
GRMZM2G046918	upregulated	inf	low_confidence							
GRMZM2G047385	upregulated	inf	low_confidence							
GRMZM2G048129	upregulated	1.92	protein_coding						AT5G61820.1	
GRMZM2G049047	upregulated	inf	low_confidence		lncRNA					
GRMZM2G049349	upregulated	2.05	protein_coding						AT1G17870.1	ethylene-dependent gravitropism-deficient and yellow-green-like 3
GRMZM2G050262	upregulated	inf	protein_coding							
GRMZM2G050674	upregulated	inf	low_confidence						AT4G22305.1	alpha/beta-Hydrolases superfamily protein
GRMZM2G050748	upregulated	3.77	protein_coding						AT4G15480.1	UDP-Glycosyltransferase superfamily protein
GRMZM2G051952	upregulated	1.99	protein_coding						AT4G28740.1	
GRMZM2G052344	upregulated	1.93	protein_coding						AT5G66070.2	RING/U-box superfamily protein
GRMZM2G053669	upregulated	2.14	protein_coding						AT3G47340.1	glutamine-dependent asparagine synthase 1
GRMZM2G054115	upregulated	1.71	protein_coding						AT1G34060.1	Pyridoxal phosphate (PLP)-dependent transferases superfamily protein
GRMZM2G054371	upregulated	2.75	low_confidence						AT1G80070.1	Pre-mRNA-processing-splicing factor
GRMZM2G054803	upregulated	1.58	protein_coding				pza01936		AT5G58770.1	Undecaprenyl pyrophosphate synthetase family protein
GRMZM2G054938_O_1	upregulated	2.14								
GRMZM2G055908	upregulated	inf	low_confidence							
GRMZM2G055908_X_1	upregulated	inf								
GRMZM2G056500	upregulated	1.93	protein_coding							
GRMZM2G056743	upregulated	inf	protein_coding							
GRMZM2G058173	upregulated	1.64	protein_coding						AT5G58770.1	Undecaprenyl pyrophosphate synthetase family protein
GRMZM2G058402	upregulated	3.35	protein_coding						AT5G04750.1	F1F0-ATPase inhibitor protein, putative
GRMZM2G058491	upregulated	2.27	protein_coding						AT1G64110.2	P-loop containing nucleoside triphosphate hydrolases superfamily protein
GRMZM2G060718	upregulated	inf	low_confidence						AT2G38550.1	Transmembrane proteins 14C
GRMZM2G061016	upregulated	inf	protein_coding						AT1G61680.1	terpene synthase 14
GRMZM2G063244	upregulated	2.38	protein_coding						AT3G62030.3	rotamase CYP 4

DE gene ID*	*rmr6-1*/wt	log2 fold change**	AGPv3.20 biotype	TE	lncRNA	Classical genes	MaizeGDB curated genes	ChromDB name	*Arabidopsis* homolog	*Arabidopsis* annotation
GRMZM2G063431	upregulated	1.90	protein_coding		lncRNA					
GRMZM2G063455	upregulated	inf	low_confidence							
GRMZM2G064805	upregulated	inf	low_confidence		lncRNA				AT1G07360.1	CCCH-type zinc fingerfamily protein with RNA-binding domain
GRMZM2G064810	upregulated	inf	protein_coding							
GRMZM2G064862	upregulated	2.42	transposable_element	TXX						
GRMZM2G065023	upregulated	inf	low_confidence						AT1G55490.1	chaperonin 60 beta
GRMZM2G065088	upregulated	2.28	protein_coding						AT1G09560.1	germin-like protein 5
GRMZM2G066441	upregulated	1.75	protein_coding						AT2G45510.1	cytochrome P450, family 704, subfamily A, polypeptide 2
GRMZM2G068169	upregulated	inf	low_confidence							
GRMZM2G068763	upregulated	2.08	protein_coding							
GRMZM2G070475	upregulated	1.45	protein_coding						AT3G14200.1	Chaperone DnaJ-domain superfamily protein
GRMZM2G071145	upregulated	2.05	low_confidence						AT5G23450.3	long-chain base (LCB) kinase 1
GRMZM2G071771	upregulated	3.41	protein_coding		lncRNA				AT5G48485.1	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
GRMZM2G073281	upregulated	inf	low_confidence						AT1G27760.3	interferon-related developmental regulator family protein / IFRD protein family
GRMZM2G073969	upregulated	1.54	protein_coding						AT3G16660.1	Pollen Ole e 1 allergen and extensin family protein
GRMZM2G074648	upregulated	inf	protein_coding		lncRNA				AT1G08450.1	calreticulin 3
GRMZM2G074759	upregulated	1.95	protein_coding				o7		AT3G48990.1	AMP-dependent synthetase and ligase family protein
GRMZM2G076031	upregulated	inf	low_confidence						AT1G17420.1	lipoxygenase 3
GRMZM2G076972	upregulated	2.26	protein_coding							
GRMZM2G077055	upregulated	inf	transposable_element	TXX	lncRNA					
GRMZM2G077256	upregulated	2.37	protein_coding						AT2G34480.1	Ribosomal protein L18ae/LX family protein
GRMZM2G077482	upregulated	inf	low_confidence							
GRMZM2G078189	upregulated	inf	low_confidence							
GRMZM2G078465	upregulated	2.25	protein_coding						AT2G43840.1	UDP-glycosyltransferase 74 F1
GRMZM2G078806	upregulated	2.28	protein_coding						AT1G36070.1	Transducin/WD40 repeat-like superfamily protein
GRMZM2G079774	upregulated	1.93	protein_coding						AT5G51070.1	Clp ATPase
GRMZM2G080103	upregulated	1.62	protein_coding						AT2G26530.1	Protein of unknown function (DUF1645)
GRMZM2G080293	upregulated	inf	low_confidence						AT3G13990.1	Kinase-related protein of unknown function (DUF1296)
GRMZM2G080438	upregulated	inf	transposable_element	DTM	lncRNA					
GRMZM2G082966	upregulated	inf	transposable_element	TXX						
GRMZM2G083907	upregulated	inf	low_confidence							
GRMZM2G084583	upregulated	1.60	protein_coding				myb38		AT4G38620.1	myb domain protein 4
GRMZM2G084947	upregulated	2.20	protein_coding						AT4G29110.1	
GRMZM2G086340	upregulated	inf	low_confidence						AT3G23560.1	MATE efflux family protein
GRMZM2G086450	upregulated	inf	protein_coding							
GRMZM2G086532	upregulated	1.84	low_confidence							
GRMZM2G086841	upregulated	1.74	protein_coding		lncRNA				AT1G80920.1	Chaperone DnaJ-domain superfamily protein
GRMZM2G087801	upregulated	inf	transposable_element	TXX						
GRMZM2G088501	upregulated	1.92	protein_coding						AT4G08690.1	Sec14p-like phosphatidylinositol transfer family protein
GRMZM2G088951	upregulated	1.64	protein_coding						AT3G57260.1	beta-1,3-glucanase 2
GRMZM2G088964	upregulated	2.62	protein_coding						AT2G30070.1	potassium transporter 1
GRMZM2G089921	upregulated	inf	low_confidence						AT1G05230.1	homeodomain GLABROUS 2
GRMZM2G090872	upregulated	3.48	protein_coding						AT3G60210.1	GroES-like family protein
GRMZM2G091855	upregulated	inf	transposable_element	DTM						
GRMZM2G091855_O_1	upregulated	inf		DTM						
GRMZM2G091980	upregulated	2.86	protein_coding							
GRMZM2G092200	upregulated	inf	transposable_element	DTM						

DE gene ID*	*rmr6-1*/wt	log2 fold change**	AGPv3.20 biotype	TE	lncRNA	Classical genes	MaizeGDB curated genes	ChromDB name	*Arabidopsis* homolog	*Arabidopsis* annotation
GRMZM2G092474	upregulated	1.81	protein_coding						AT4G11650.1	osmotin 34
GRMZM2G092938	upregulated	1.57	transposable_element	RLC;RLG; RLX						
GRMZM2G094028	upregulated	3.22	protein_coding						AT1G67000.1	Protein kinase superfamily protein
GRMZM2G094286	upregulated	inf	low_confidence		lncRNA				AT5G53120.6	spermidine synthase 3
GRMZM2G095873	upregulated	1.94	protein_coding						AT5G15820.1	RING/U-box superfamily protein
GRMZM2G096412	upregulated	1.98	protein_coding						AT3G16520.3	UDP-glucosyl transferase 88A1
GRMZM2G098066	upregulated	inf	transposable_element	TXX						
GRMZM2G098739	upregulated	3.60	protein_coding							
GRMZM2G099049	upregulated	3.00	protein_coding						AT1G56220.4	Dormancy/auxin associated family protein
GRMZM2G099454	upregulated	1.53	protein_coding						AT3G12500.1	basic chitinase
GRMZM2G100665	upregulated	inf	transposable_element	TXX	lncRNA					
GRMZM2G101928	upregulated	3.25	protein_coding						AT5G13750.1	zinc induced facilitator-like 1
GRMZM2G102473	upregulated	inf	low_confidence							
GRMZM2G103273	upregulated	2.15	low_confidence						AT1G45130.1	beta-galactosidase 5
GRMZM2G103801	upregulated	2.84	low_confidence							
GRMZM2G103945	upregulated	1.64	protein_coding						AT2G25810.1	tonoplast intrinsic protein 4;1
GRMZM2G104231	upregulated	1.82	protein_coding						AT1G14590.1	Nucleotide-diphospho-sugar transferase family protein
GRMZM2G105449	upregulated	inf	low_confidence						AT3G57320.1	
GRMZM2G105564	upregulated	inf	transposable_element	TXX						
GRMZM2G106344	upregulated	1.47	protein_coding						AT1G60420.1	DC1 domain-containing protein
GRMZM2G106393	upregulated	2.11	protein_coding						AT4G10265.1	Wound-responsive family protein
GRMZM2G106413	upregulated	1.81	protein_coding		lncRNA				AT4G10270.1	Wound-responsive family protein
GRMZM2G106950	upregulated	1.66	protein_coding						AT5G48220.1	Aldolase-type TIM barrel family protein
GRMZM2G108003	upregulated	2.23	protein_coding							
GRMZM2G108692	upregulated	inf	low_confidence						AT1G72330.3	alanine aminotransferase 2
GRMZM2G108699	upregulated	inf	low_confidence		lncRNA				AT5G57870.1	MIF4G domain-containing protein / MA3 domain-containing protein
GRMZM2G108807	upregulated	inf	protein_coding						AT2G31200.1	actin depolymerizing factor 6
GRMZM2G110840	upregulated	inf	low_confidence		lncRNA					
GRMZM2G111570	upregulated	inf	low_confidence							
GRMZM2G111954	upregulated	1.71	protein_coding						AT1G06690.1	NAD(P)-linked oxidoreductase superfamily protein
GRMZM2G113633	upregulated	2.13	protein_coding							
GRMZM2G113761	upregulated	1.87	protein_coding						AT2G36870.1	xyloglucan endotransglucosylase/hydrolase 32
GRMZM2G115035	upregulated	inf	transposable_element	TXX						
GRMZM2G115453_X_1	upregulated	inf			lncRNA					
GRMZM2G115476	upregulated	inf	protein_coding						AT5G19855.1	Chaperonin-like RbcX protein
GRMZM2G116288	upregulated	inf	low_confidence							
GRMZM2G116487	upregulated	inf	low_confidence							
GRMZM2G116494_O_1	upregulated	inf								
GRMZM2G116896	upregulated	3.61	low_confidence							
GRMZM2G116966	upregulated	2.44	protein_coding						AT5G04370.2	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
GRMZM2G117084	upregulated	inf	transposable_element	TXX	lncRNA					
GRMZM2G117164	upregulated	1.99	protein_coding				hb41		AT5G65310.1	homeobox protein 5
GRMZM2G117319	upregulated	2.55	protein_coding				tps4		AT3G14490.1	Terpenoid cyclases/Protein prenyltransferases superfamily protein
GRMZM2G117878	upregulated	1.95	protein_coding						AT5G05860.1	UDP-glucosyl transferase 76C2
GRMZM2G119744	upregulated	inf	transposable_element	TXX						

DE gene ID*	*rmr6-1*/wt	log2 fold change**	AGPv3.20 biotype	TE	lncRNA	Classical genes	MaizeGDB curated genes	ChromDB name	*Arabidopsis* homolog	*Arabidopsis* annotation
GRMZM2G120038	upregulated	inf	low_confidence						AT3G54630.1	
GRMZM2G120084	upregulated	inf	low_confidence							
GRMZM2G120159	upregulated	inf	transposable_element	TXX						
GRMZM2G120166	upregulated	inf	transposable_element	TXX						
GRMZM2G120587	upregulated	2.43	protein_coding						AT2G27920.1	serine carboxypeptidase-like 51
GRMZM2G121928	upregulated	3.21	protein_coding							
GRMZM2G124927	upregulated	4.79	low_confidence							
GRMZM2G125032	upregulated	4.98	protein_coding						AT4G16260.1	Glycosyl hydrolase superfamily protein
GRMZM2G125482	upregulated	1.95	protein_coding						AT5G07990.1	Cytochrome P450 superfamily protein
GRMZM2G125571	upregulated	3.83	protein_coding							
GRMZM2G126878	upregulated	inf	low_confidence						AT2G33210.1	heat shock protein 60-2
GRMZM2G127336	upregulated	2.29	protein_coding				tps23		AT5G23960.1	terpene synthase 21
GRMZM2G127397	upregulated	inf	low_confidence		lncRNA					
GRMZM2G127521	upregulated	3.62	protein_coding						AT1G67430.1	Ribosomal protein L22p/L17e family protein
GRMZM2G127924	upregulated	3.49	protein_coding						AT1G67900.1	Phototropic-responsive NPH3 family protein
GRMZM2G129815	upregulated	1.90	protein_coding						AT3G56710.1	sigma factor binding protein 1
GRMZM2G131001	upregulated	inf	low_confidence							
GRMZM2G133721	upregulated	2.65	low_confidence						AT2G36530.1	Enolase
GRMZM2G133841	upregulated	inf	low_confidence							
GRMZM2G134711	upregulated	2.31	protein_coding						AT1G23140.1	Calcium-dependent lipid-binding (CaLB domain) family protein
GRMZM2G135165	upregulated	1.91	protein_coding						AT5G17210.1	Protein of unknown function (DUF1218)
GRMZM2G135240	upregulated	inf	low_confidence						AT1G32050.1	SCAMP family protein
GRMZM2G135722	upregulated	2.17	protein_coding						AT5G49690.1	UDP-Glycosyltransferase superfamily protein
GRMZM2G136364	upregulated	1.76	protein_coding		lncRNA				AT5G05960.1	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
GRMZM2G136453	upregulated	2.35	protein_coding		lncRNA				AT5G50400.1	purple acid phosphatase 27
GRMZM2G136960	upregulated	1.55	protein_coding						AT2G21970.1	stress enhanced protein 2
GRMZM2G138710	upregulated	2.46	protein_coding						AT4G24220.1	NAD(P)-binding Rossmann-fold superfamily protein
GRMZM2G139300	upregulated	1.30	protein_coding			incw1			AT3G52600.1	cell wall invertase 2
GRMZM2G140107	upregulated	4.35	protein_coding				phm9914		AT1G04920.1	sucrose phosphate synthase 3F
GRMZM2G140160	upregulated	1.69	protein_coding						AT5G20910.1	RING/U-box superfamily protein
GRMZM2G141252	upregulated	inf	protein_coding						AT4G22220.1	SufE/NifU family protein
GRMZM2G143139	upregulated	1.96	protein_coding						AT1G15520.1	pleiotropic drug resistance 12
GRMZM2G143332	upregulated	inf	protein_coding		lncRNA				AT2G31200.1	actin depolymerizing factor 6
GRMZM2G144870	upregulated	inf	low_confidence							
GRMZM2G145313	upregulated	2.17	protein_coding						AT3G23160.1	Protein of unknown function (DUF668)
GRMZM2G145427	upregulated	inf	low_confidence		lncRNA					
GRMZM2G145440	upregulated	inf	protein_coding						AT1G71830.1	somatic embryogenesis receptor-like kinase 1
GRMZM2G145783	upregulated	inf	low_confidence		lncRNA				AT1G76150.1	enoyl-CoA hydratase 2
GRMZM2G145941	upregulated	inf	low_confidence						AT2G31200.1	actin depolymerizing factor 6
GRMZM2G146627	upregulated	2.22	protein_coding						AT2G25810.1	tonoplast intrinsic protein 4;1
GRMZM2G146821	upregulated	inf	low_confidence						AT5G45550.1	Mob1/phocein family protein
GRMZM2G147210	upregulated	inf	protein_coding						AT1G77510.1	PDI-like 1-2
GRMZM2G147210_O_1	upregulated	inf			lncRNA					
GRMZM2G147319	upregulated	1.56	protein_coding						AT3G54780.1	Zinc finger (C3HC4-type RING finger) family protein
GRMZM2G147716	upregulated	1.30	protein_coding				mads67	FLCP113	AT5G60910.1	AGAMOUS-like 8
GRMZM2G148052	upregulated	1.68	protein_coding						AT3G26310.1	cytochrome P450, family 71, subfamily B, polypeptide 35
GRMZM2G149122	upregulated	inf	low_confidence						AT1G03370.1	C2 calcium/lipid-binding and GRAM domain containing protein
GRMZM2G149273	upregulated	2.35	protein_coding							

DE gene ID*	*rmr6-1*/wt	log2 fold change**	AGPv3.20 biotype	TE	lncRNA	Classical genes	MaizeGDB curated genes	ChromDB name	*Arabidopsis* homolog	*Arabidopsis* annotation
GRMZM2G150688	upregulated	1.32	protein_coding		lncRNA				AT1G10588.1	Gibberellin-regulated family protein
GRMZM2G150906	upregulated	1.35	protein_coding						AT5G40390.1	Raffinose synthase family protein
GRMZM2G151208	upregulated	inf	low_confidence							
GRMZM2G151896	upregulated	inf	low_confidence		lncRNA				AT2G32430.1	Galactosyltransferase family protein
GRMZM2G152054	upregulated	inf	protein_coding							
GRMZM2G152079	upregulated	3.13	protein_coding		lncRNA				AT3G53235.1	
GRMZM2G153178	upregulated	3.22	protein_coding						AT4G17960.1	
GRMZM2G154735	upregulated	1.94	protein_coding						AT5G50720.1	HVA22 homologue E
GRMZM2G156129	upregulated	inf	protein_coding				phd21		AT2G27980.1	Acyl-CoA N-acyltransferase with RING/FYVE/PHD-type zinc finger domain
GRMZM2G157102	upregulated	1.79	protein_coding						AT5G25050.1	Major facilitator superfamily protein
GRMZM2G157437	upregulated	inf	low_confidence						AT3G10350.1	P-loop containing nucleoside triphosphate hydrolases superfamily protein
GRMZM2G157505	upregulated	3.50	protein_coding							
GRMZM2G157586	upregulated	inf	low_confidence						AT4G34570.1	thymidylate synthase 2
GRMZM2G157772	upregulated	inf	protein_coding						AT2G28760.1	UDP-XYL synthase 6
GRMZM2G157822	upregulated	inf	protein_coding						AT1G17580.1	myosin 1
GRMZM2G158742	upregulated	3.08	low_confidence						AT1G52150.2	Homeobox-leucine zipper family protein / lipid-binding START domain-containing protein
GRMZM2G158793	upregulated	inf	low_confidence							
GRMZM2G159811	upregulated	2.36	protein_coding						AT5G51760.1	Protein phosphatase 2C family protein
GRMZM2G160541	upregulated	2.52	protein_coding						AT3G10340.1	phenylalanine ammonia-lyase 4
GRMZM2G161255	upregulated	2.12	protein_coding						AT5G57520.1	zinc finger protein 2
GRMZM2G161326	upregulated	inf	low_confidence		lncRNA					
GRMZM2G162056	upregulated	3.53	protein_coding				hagtf13		AT2G32030.1	Acyl-CoA N-acyltransferases (NAT) superfamily protein
GRMZM2G163182	upregulated	inf	low_confidence	DTC						
GRMZM2G163462	upregulated	inf	low_confidence						AT1G66950.1	pleiotropic drug resistance 11
GRMZM2G165919	upregulated	1.34	protein_coding						AT2G47180.1	galactinol synthase 1
GRMZM2G167841	upregulated	2.10	transposable_element	TXX						
GRMZM2G168115	upregulated	1.75	protein_coding						AT4G35880.1	Eukaryotic aspartyl protease family protein
GRMZM2G168956	upregulated	1.73	protein_coding						AT1G25450.1	3-ketoacyl-CoA synthase 5
GRMZM2G169149	upregulated	1.64	protein_coding				wrky104		AT4G31800.1	WRKY DNA-binding protein 18
GRMZM2G169504	upregulated	inf	low_confidence							
GRMZM2G170613	upregulated	inf	low_confidence						AT3G54630.1	
GRMZM2G170799	upregulated	inf	transposable_element	TXX	lncRNA					
GRMZM2G171539	upregulated	inf	low_confidence						AT3G03070.1	NADH-ubiquinone oxidoreductase-related
GRMZM2G172338	upregulated	inf	low_confidence							
GRMZM2G173379	upregulated	inf	low_confidence		lncRNA				AT1G56080.1	
GRMZM2G173715	upregulated	inf	low_confidence						AT5G49610.1	F-box family protein
GRMZM2G174147	upregulated	2.42	protein_coding							
GRMZM2G174449	upregulated	3.06	protein_coding						AT3G50150.1	Plant protein of unknown function (DUF247)
GRMZM2G176085	upregulated	1.40	protein_coding						AT5G67400.1	root hair specific 19
GRMZM2G176206	upregulated	1.32	protein_coding						AT5G25930.1	Protein kinase family protein with leucine-rich repeat domain
GRMZM2G177982	upregulated	1.82	protein_coding						AT1G19400.1	Erythronate-4-phosphate dehydrogenase family protein
GRMZM2G178769	upregulated	2.33	protein_coding						AT5G48930.1	hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase
GRMZM2G178780	upregulated	inf	low_confidence							
GRMZM2G179292	upregulated	1.55	protein_coding						AT4G19390.1	Uncharacterised protein family (UPF0114)
GRMZM2G181000	upregulated	2.24	protein_coding						AT1G73480.1	alpha/beta-Hydrolases superfamily protein

DE gene ID*	*rmr6-1*/wt	log2 fold change**	AGPv3.20 biotype	TE	lncRNA	Classical genes	MaizeGDB curated genes	ChromDB name	*Arabidopsis* homolog	*Arabidopsis* annotation
GRMZM2G181168	upregulated	inf	low_confidence						AT5G44480.1	NAD(P)-binding Rossmann-fold superfamily protein
GRMZM2G302873_X_1	upregulated	inf								
GRMZM2G303010	upregulated	inf	protein_coding							
GRMZM2G303121	upregulated	3.59	transposable_element	RLC;RLX						
GRMZM2G304372	upregulated	inf	protein_coding		lncRNA					
GRMZM2G305362	upregulated	2.90	protein_coding						AT3G10040.1	sequence-specific DNA binding transcription factors
GRMZM2G305685	upregulated	2.57	low_confidence							
GRMZM2G306859	upregulated	inf	protein_coding						AT2G37520.1	Acyl-CoA N-acyltransferase with RING/FYVE/PHD-type zinc finger domain
GRMZM2G309071	upregulated	1.39	protein_coding						AT5G20610.1	
GRMZM2G309145	upregulated	5.19	low_confidence						AT2G26670.1	Plant haem oxygenase (decyclizing) family protein
GRMZM2G313750	upregulated	3.43	protein_coding						AT5G07990.1	Cytochrome P450 superfamily protein
GRMZM2G314769	upregulated	3.00	protein_coding						AT1G19715.1	Mannose-binding lectin superfamily protein
GRMZM2G315767	upregulated	2.91	protein_coding						AT4G13840.1	HXXXD-type acyl-transferase family protein
GRMZM2G317687	upregulated	inf	low_confidence							
GRMZM2G320366	upregulated	inf	transposable_element	TXX						
GRMZM2G322717	upregulated	inf	protein_coding							
GRMZM2G325857	upregulated	2.54	low_confidence						AT5G25880.1	NADP-malic enzyme 3
GRMZM2G326339	upregulated	inf	transposable_element	TXX						
GRMZM2G326489	upregulated	inf	transposable_element	TXX						
GRMZM2G329962	upregulated	3.23	protein_coding							
GRMZM2G330012	upregulated	1.50	protein_coding						AT3G20220.1	SAUR-like auxin-responsive protein family
GRMZM2G332752	upregulated	inf	low_confidence						AT3G44620.1	protein tyrosine phosphatases;protein tyrosine phosphatases
GRMZM2G333095	upregulated	2.31	protein_coding						AT2G40475.1	
GRMZM2G335635	upregulated	2.58	low_confidence							
GRMZM2G339327	upregulated	2.84	protein_coding		lncRNA					
GRMZM2G339645	upregulated	2.14	protein_coding						AT1G02730.1	cellulose synthase-like D5
GRMZM2G349749	upregulated	3.54	protein_coding						AT2G26560.1	phospholipase A 2A
GRMZM2G352773	upregulated	inf	transposable_element	TXX	lncRNA					
GRMZM2G354172	upregulated	inf	low_confidence							
GRMZM2G355342	upregulated	inf	protein_coding						AT1G03360.1	ribosomal RNA processing 4
GRMZM2G355358	upregulated	2.92	low_confidence							
GRMZM2G355752	upregulated	1.84	protein_coding				elip1		AT3G22840.1	Chlorophyll A-B binding family protein
GRMZM2G359559	upregulated	4.21	protein_coding						AT2G13360.1	alanine:glyoxylate aminotransferase
GRMZM2G360104	upregulated	inf	low_confidence							
GRMZM2G360541	upregulated	2.58	protein_coding						AT5G19770.1	tubulin alpha-3
GRMZM2G364685	upregulated	inf	low_confidence		lncRNA					
GRMZM2G365544	upregulated	3.52	protein_coding						AT3G18660.2	plant glycogenin-like starch initiation protein 1
GRMZM2G372681	upregulated	inf	protein_coding							
GRMZM2G374302	upregulated	1.31	protein_coding				AY110562		AT2G16500.1	arginine decarboxylase 1
GRMZM2G374971	upregulated	1.83	protein_coding						AT4G11650.1	osmotin 34
GRMZM2G375833	upregulated	inf	low_confidence						AT5G51820.1	phosphoglucomutase
GRMZM2G379295	upregulated	inf	low_confidence		lncRNA				AT1G60670.2	Protein of unknown function (DUF3755)
GRMZM2G386407	upregulated	inf	low_confidence						AT4G18750.1	Pentatricopeptide repeat (PPR) superfamily protein
GRMZM2G388070	upregulated	inf	low_confidence						AT4G16800.1	ATP-dependent caseinolytic (Clp) protease/crotonase family protein
GRMZM2G389065	upregulated	inf	low_confidence						AT4G19130.1	Replication factor-A protein 1-related
GRMZM2G391559	upregulated	inf	low_confidence		lncRNA				AT5G49910.1	chloroplast heat shock protein 70-2
GRMZM2G392010	upregulated	inf	transposable_element	TXX						

DE gene ID*	*rmr6-1*/wt	log2 fold change**	AGPv3.20 biotype	TE	lncRNA	Classical genes	MaizeGDB curated genes	ChromDB name	*Arabidopsis* homolog	*Arabidopsis* annotation
GRMZM2G392076	upregulated	inf	protein_coding						AT3G27110.1	Peptidase family M48 family protein
GRMZM2G392513	upregulated	3.49	protein_coding						AT2G22240.1	myo-inositol-1-phosphate synthase 2
GRMZM2G392791	upregulated	1.47	protein_coding						AT3G51000.1	alpha/beta-Hydrolases superfamily protein
GRMZM2G395274	upregulated	inf	transposable_element	TXX						
GRMZM2G395526	upregulated	1.77	low_confidence							
GRMZM2G395569	upregulated	inf	protein_coding						AT1G12680.1	phosphoenolpyruvate carboxylase-related kinase 2
GRMZM2G397950	upregulated	2.81	transposable_element	TXX						
GRMZM2G398996	upregulated	3.74	protein_coding						AT5G06570.1	alpha/beta-Hydrolases superfamily protein
GRMZM2G400655	upregulated	9.50	protein_coding						AT3G56310.1	Melibiase family protein
GRMZM2G407495	upregulated	2.92	protein_coding		lncRNA					
GRMZM2G409771	upregulated	1.34	protein_coding						AT4G13420.1	high affinity K+ transporter 5
GRMZM2G410162	upregulated	inf	low_confidence						AT5G26760.2	
GRMZM2G410809	upregulated	inf	transposable_element	TXX						
GRMZM2G410916	upregulated	1.98	protein_coding						AT3G18170.1	Glycosyltransferase family 61 protein
GRMZM2G411216	upregulated	3.98	protein_coding							
GRMZM2G412986	upregulated	1.86	protein_coding						AT3G18170.1	Glycosyltransferase family 61 protein
GRMZM2G416998	upregulated	inf	low_confidence		lncRNA					
GRMZM2G421449	upregulated	inf	low_confidence		lncRNA					
GRMZM2G422187	upregulated	inf	protein_coding							
GRMZM2G423324_X_1	upregulated	inf			lncRNA					
GRMZM2G425136	upregulated	inf	protein_coding						AT1G77090.1	Mog1/PsbP/DUF1795-like photosystem II reaction center PsbP family protein
GRMZM2G426140	upregulated	inf	protein_coding		lncRNA		hb27		AT3G09770.2	RING/U-box superfamily protein
GRMZM2G427073	upregulated	inf	protein_coding						AT4G23690.1	Disease resistance-responsive (dirigent-like protein) family protein
GRMZM2G427836	upregulated	4.81	low_confidence						AT4G13930.1	serine hydroxymethyltransferase 4
GRMZM2G428740	upregulated	inf	protein_coding						AT5G31412.1	hAT transposon superfamily protein
GRMZM2G429982	upregulated	2.92	protein_coding						AT2G28790.1	Pathogenesis-related thaumatin superfamily protein
GRMZM2G430455	upregulated	inf	protein_coding						AT3G59100.1	glucan synthase-like 11
GRMZM2G430983	upregulated	inf	low_confidence						AT5G57870.1	MIF4G domain-containing protein / MA3 domain-containing protein
GRMZM2G431039	upregulated	4.07	protein_coding						AT5G55180.2	O-Glycosyl hydrolases family 17 protein
GRMZM2G431288	upregulated	1.79	protein_coding						AT5G07990.1	Cytochrome P450 superfamily protein
GRMZM2G432615	upregulated	inf	low_confidence						AT4G34610.1	BEL1-like homeodomain 6
GRMZM2G432653	upregulated	inf	protein_coding						AT2G02730.1	Protein of unknown function (DUF1664)
GRMZM2G434509	upregulated	inf	low_confidence							
GRMZM2G437134	upregulated	inf	transposable_element	DTA						
GRMZM2G437627	upregulated	inf	low_confidence						AT1G10940.1	Protein kinase superfamily protein
GRMZM2G443549	upregulated	inf	transposable_element	DTM						
GRMZM2G444683	upregulated	inf	low_confidence							
GRMZM2G447806	upregulated	inf	protein_coding							
GRMZM2G450233	upregulated	1.52	protein_coding						AT1G05260.1	Peroxidase superfamily protein
GRMZM2G452475	upregulated	inf	low_confidence						AT5G65110.1	acyl-CoA oxidase 2
GRMZM2G452896	upregulated	1.64	protein_coding						AT2G24280.1	alpha/beta-Hydrolases superfamily protein
GRMZM2G453805	upregulated	1.90	protein_coding						AT5G24090.1	chitinase A
GRMZM2G455774	upregulated	inf	transposable_element	TXX						
GRMZM2G455817	upregulated	2.06	protein_coding						AT4G13420.1	high affinity K+ transporter 5
GRMZM2G455909	upregulated	1.98	protein_coding							

DE gene ID*	*rmr6-1*/wt	log2 fold change**	AGPv3.20 biotype	TE	lncRNA	Classical genes	MaizeGDB curated genes	ChromDB name	*Arabidopsis* homolog	*Arabidopsis* annotation
GRMZM2G455959	upregulated	inf	transposable_element	RLC;RLX; TXX						
GRMZM2G456241	upregulated	2.30	protein_coding						AT3G62240.1	RING/U-box superfamily protein
GRMZM2G456997	upregulated	4.54	protein_coding						AT4G33720.1	CAP (Cysteine-rich secretory proteins, Antigen 5, and Pathogenesis-related 1 protein) superfamily protein
GRMZM2G459535	upregulated	inf	low_confidence						AT1G58230.1	binding
GRMZM2G459563	upregulated	2.06	protein_coding						AT2G46950.1	cytochrome P450, family 709, subfamily B, polypeptide 2
GRMZM2G460497	upregulated	inf	low_confidence		lncRNA					
GRMZM2G460669	upregulated	inf	transposable_element	TXX						
GRMZM2G470981	upregulated	inf	protein_coding						AT5G59240.1	Ribosomal protein S8e family protein
GRMZM2G476076	upregulated	4.31	low_confidence							
GRMZM2G476515	upregulated	inf	protein_coding							
GRMZM2G476515_X_1	upregulated	inf								
GRMZM2G477366	upregulated	inf	protein_coding						AT3G15640.1	Rubredoxin-like superfamily protein
GRMZM2G478160	upregulated	1.63	protein_coding						AT2G44310.1	Calcium-binding EF-hand family protein
GRMZM2G479038	upregulated	1.40	protein_coding						AT2G36800.1	don-glucosyltransferase 1
GRMZM2G485630_X_1	upregulated	3.34			lncRNA					
GRMZM2G488126	upregulated	inf	low_confidence							
GRMZM2G488333	upregulated	inf	low_confidence							
GRMZM2G488337	upregulated	inf	transposable_element	TXX						
GRMZM2G490920	upregulated	inf	low_confidence							
GRMZM2G491083_X_1	upregulated	inf								
GRMZM2G492670	upregulated	1.56	transposable_element	RLX						
GRMZM2G492768	upregulated	inf	protein_coding							
GRMZM2G494294	upregulated	inf	low_confidence							
GRMZM2G494504	upregulated	inf	low_confidence							
GRMZM2G497801	upregulated	inf	low_confidence	DTC						
GRMZM2G499758	upregulated	inf	low_confidence		lncRNA					
GRMZM2G500737	upregulated	inf	low_confidence		lncRNA					
GRMZM2G500739	upregulated	inf	low_confidence		lncRNA					
GRMZM2G504992_X_1	upregulated	inf			lncRNA					
GRMZM2G507885	upregulated	inf	low_confidence							
GRMZM2G508838	upregulated	inf	transposable_element	RLG;RLX	lncRNA					
GRMZM2G509699	upregulated	inf	protein_coding							
GRMZM2G514005	upregulated	2.66	transposable_element	RLC;RLG; RLX						
GRMZM2G517515	upregulated	inf	transposable_element	RLC;RLX						
GRMZM2G517529	upregulated	inf	low_confidence							
GRMZM2G518053	upregulated	inf	low_confidence		lncRNA					
GRMZM2G518751	upregulated	inf	protein_coding							
GRMZM2G521213_O_1	upregulated	inf								
GRMZM2G523120	upregulated	inf	low_confidence							
GRMZM2G523549	upregulated	inf	low_confidence		lncRNA					
GRMZM2G525708_X_1	upregulated	inf								
GRMZM2G527064	upregulated	inf	low_confidence							
GRMZM2G528190	upregulated	inf	low_confidence		lncRNA					
GRMZM2G529020	upregulated	inf	transposable_element	TXX						
GRMZM2G529294	upregulated	inf	low_confidence		lncRNA					
GRMZM2G532031	upregulated	inf	low_confidence							

DE gene ID*	*rmr6-1*/wt	log2 fold change**	AGPv3.20 biotype	TE	lncRNA	Classical genes	MaizeGDB curated genes	ChromDB name	*Arabidopsis* homolog	*Arabidopsis* annotation
GRMZM2G536584	upregulated	2.32	protein_coding						AT1G71800.1	cleavage stimulating factor 64
GRMZM2G540655	upregulated	2.39	transposable_element	TXX						
GRMZM2G542380	upregulated	inf	transposable_element	TXX						
GRMZM2G543660	upregulated	inf	low_confidence							
GRMZM2G544521	upregulated	3.56	transposable_element	RLX						
GRMZM2G544705	upregulated	6.64	transposable_element	RLG;TXX						
GRMZM2G545361	upregulated	2.18	transposable_element	RLC;RLG; RLX						
GRMZM2G545782	upregulated	inf	low_confidence		lncRNA					
GRMZM2G546227	upregulated	3.12	transposable_element	RLX						
GRMZM2G549080	upregulated	inf	transposable_element	DTC;RLG; RLX						
GRMZM2G552162_X_1	upregulated	inf			lncRNA					
GRMZM2G552982	upregulated	2.39	transposable_element	RLC;RLG; RLX						
GRMZM2G559161	upregulated	inf	low_confidence							
GRMZM2G561369	upregulated	inf	low_confidence		lncRNA					
GRMZM2G569575	upregulated	inf	low_confidence		lncRNA					
GRMZM2G569688	upregulated	inf	low_confidence		lncRNA					
GRMZM2G571415	upregulated	inf	low_confidence		lncRNA					
GRMZM2G572949_X_1	upregulated	inf								
GRMZM2G573715	upregulated	2.78	transposable_element	RLX;TXX						
GRMZM2G575328	upregulated	6.68	low_confidence							
GRMZM2G576564	upregulated	inf	low_confidence		lncRNA					
GRMZM2G578653_X_1	upregulated	inf			lncRNA					
GRMZM2G578656	upregulated	inf	low_confidence		lncRNA					
GRMZM2G579335	upregulated	5.51	low_confidence							
GRMZM2G580648	upregulated	inf	low_confidence		lncRNA					
GRMZM2G584110	upregulated	6.68	low_confidence							
GRMZM2G585314	upregulated	1.95	transposable_element	RLX						
GRMZM2G588812	upregulated	inf	low_confidence							
GRMZM2G590958	upregulated	5.44	low_confidence							
GRMZM2G700785	upregulated	2.43	transposable_element	TXX						
GRMZM2G701063	upregulated	2.41	protein_coding			pl1				
GRMZM2G702673	upregulated	inf	transposable_element	DTM						
GRMZM2G703582	upregulated	2.89	protein_coding						AT3G21420.1	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
GRMZM2G703602	upregulated	inf	protein_coding		lncRNA					
GRMZM2G703625	upregulated	3.73	protein_coding							
GRMZM5G800159	upregulated	inf	low_confidence						AT1G07530.1	SCARECROW-like 14
GRMZM5G801627	upregulated	1.39	protein_coding				c3h39		AT4G29190.1	Zinc finger C-x8-C-x5-C-x3-H type family protein
GRMZM5G803981	upregulated	2.00	protein_coding						AT4G15210.1	beta-amylase 5
GRMZM5G807269	upregulated	1.89	low_confidence							
GRMZM5G809218	upregulated	1.54	low_confidence							
GRMZM5G810393	upregulated	2.13	protein_coding						AT3G56630.1	cytochrome P450, family 94, subfamily D, polypeptide 2
GRMZM5G812246	upregulated	inf	protein_coding							
GRMZM5G812494	upregulated	inf	low_confidence							
GRMZM5G812525	upregulated	inf	low_confidence						AT1G52360.2	Coatomer, beta\' subunit
GRMZM5G818655	upregulated	3.43	protein_coding						AT3G10040.1	sequence-specific DNA binding transcription factors

DE gene ID*	*rmr6-1*/wt	log2 fold change**	AGPv3.20 biotype	TE	lncRNA	Classical genes	MaizeGDB curated genes	ChromDB name	*Arabidopsis* homolog	*Arabidopsis* annotation
GRMZM5G821047	upregulated	3.16	low_confidence							
GRMZM5G830269	upregulated	2.13	protein_coding							
GRMZM5G834747	upregulated	inf	low_confidence		lncRNA					
GRMZM5G837563	upregulated	2.43	low_confidence							
GRMZM5G842115	upregulated	inf	low_confidence		lncRNA					
GRMZM5G847687	upregulated	2.22	low_confidence							
GRMZM5G851493	upregulated	inf	low_confidence		lncRNA				AT5G49930.1	zinc knuckle (CCHC-type) family protein
GRMZM5G852116	upregulated	2.99	transposable_element	TXX						
GRMZM5G854490	upregulated	inf	protein_coding						AT2G37980.1	O-fucosyltransferase family protein
GRMZM5G857585	upregulated	inf	transposable_element	RLX						
GRMZM5G859334_O_1	upregulated	2.69								
GRMZM5G859350	upregulated	3.40	low_confidence							
GRMZM5G860235	upregulated	2.44	protein_coding						AT1G05590.1	beta-hexosaminidase 2
GRMZM5G862967	upregulated	inf	transposable_element	TXX						
GRMZM5G864536	upregulated	inf	transposable_element	RLC						
GRMZM5G864903	upregulated	2.50	protein_coding						AT2G46950.1	cytochrome P450, family 709, subfamily B, polypeptide 2
GRMZM5G865543	upregulated	inf	protein_coding						AT3G16250.1	NDH-dependent cyclic electron flow 1
GRMZM5G865804	upregulated	3.42	low_confidence							
GRMZM5G866269	upregulated	2.74	protein_coding							
GRMZM5G870900	upregulated	inf	transposable_element	TXX						
GRMZM5G871182	upregulated	inf	low_confidence		lncRNA					
GRMZM5G871264	upregulated	4.09	low_confidence							
GRMZM5G872680	upregulated	1.86	low_confidence		lncRNA					
GRMZM5G875581	upregulated	inf	low_confidence						AT4G12030.1	bile acid transporter 5
GRMZM5G877552	upregulated	inf	low_confidence		lncRNA					
GRMZM5G878607	upregulated	2.09	protein_coding						AT1G78570.1	rhamnose biosynthesis 1
GRMZM5G881062	upregulated	inf	transposable_element	TXX						
GRMZM5G882114	upregulated	inf	transposable_element	TXX						
GRMZM5G884151	upregulated	3.92	protein_coding							
GRMZM5G888387	upregulated	inf	low_confidence		lncRNA					
GRMZM5G889705	upregulated	inf	protein_coding						AT2G20860.1	lipoic acid synthase 1
GRMZM5G891656	upregulated	2.79	protein_coding						AT5G05320.1	FAD/NAD(P)-binding oxidoreductase family protein
GRMZM5G899851	upregulated	2.00	protein_coding						AT1G04580.1	aldehyde oxidase 4
XLOC_000025	upregulated	inf		RLC						
XLOC_000920	upregulated	inf								
XLOC_001668	upregulated	inf		RLC						
XLOC_001779	upregulated	inf			lncRNA					
XLOC_003674	upregulated	inf								
XLOC_003693	upregulated	inf								
XLOC_004325	upregulated	inf		RLG;RLX						
XLOC_004497	upregulated	inf								
XLOC_004721	upregulated	inf			lncRNA					
XLOC_004722	upregulated	inf								
XLOC_004723	upregulated	inf								
XLOC_004898	upregulated	inf								
XLOC_005460	upregulated	inf			lncRNA					
XLOC_005501	upregulated	inf		RLG						
XLOC_005785	upregulated	inf		RLG;RLX						
XLOC_005786	upregulated	inf		RLX						

DE gene ID*	*rmr6-1*/wt	log2 fold change**	AGPv3.20 biotype	TE	lncRNA	Classical genes	MaizeGDB curated genes	ChromDB name	*Arabidopsis* homolog	*Arabidopsis* annotation
XLOC_005787	upregulated	inf		RLG;RLX						
XLOC_006109	upregulated	inf								
XLOC_006210	upregulated	inf		RLG;RLX						
XLOC_008157	upregulated	inf		RLC;RLG						
XLOC_009939	upregulated	inf								
XLOC_010082	upregulated	inf		RLX						
XLOC_010083	upregulated	inf		RLG;RLX						
XLOC_010221	upregulated	inf								
XLOC_011133	upregulated	inf		RLG;RLX						
XLOC_011418	upregulated	inf		RLX						
XLOC_012235	upregulated	inf								
XLOC_012299	upregulated	3.08		RLX						
XLOC_012853	upregulated	inf								
XLOC_013636	upregulated	inf								
XLOC_014589	upregulated	inf		RLC;RLX						
XLOC_014590	upregulated	inf		RLC;RLX						
XLOC_014591	upregulated	inf		RLC	lncRNA					
XLOC_015573	upregulated	inf		RLC						
XLOC_016040	upregulated	inf		RLG						
XLOC_017397	upregulated	inf		RLG						
XLOC_017809	upregulated	inf								
XLOC_019586	upregulated	inf								
XLOC_020370	upregulated	inf								
XLOC_021467	upregulated	inf								
XLOC_021468	upregulated	inf			lncRNA					
XLOC_021719	upregulated	inf								
XLOC_023323	upregulated	inf								
XLOC_023976	upregulated	inf								
XLOC_025177	upregulated	inf								
XLOC_025760	upregulated	inf		RLG						
XLOC_026041	upregulated	inf		RLG						
XLOC_026889	upregulated	inf		RLX	lncRNA					
XLOC_027333	upregulated	inf		RLG;RLX						
XLOC_027525	upregulated	inf		RLG						
XLOC_027529	upregulated	inf		RLC	lncRNA					
XLOC_027534	upregulated	inf		RLC						
XLOC_027840	upregulated	inf			lncRNA					
XLOC_029954	upregulated	inf			lncRNA					
XLOC_031937	upregulated	inf			lncRNA					
XLOC_032378	upregulated	inf								
XLOC_032509	upregulated	inf			lncRNA					
XLOC_033757	upregulated	inf								
XLOC_034499	upregulated	inf		RLG;RLX						
XLOC_035168	upregulated	inf			lncRNA					
XLOC_035497	upregulated	inf		RLG						
XLOC_036069	upregulated	inf								
XLOC_036246	upregulated	inf								
XLOC_036686	upregulated	inf								
XLOC_036939	upregulated	inf			lncRNA					

DE gene ID*	*rmr6-1*/wt	log2 fold change**	AGPv3.20 biotype	TE	lncRNA	Classical genes	MaizeGDB curated genes	ChromDB name	*Arabidopsis* homolog	*Arabidopsis* annotation
XLOC_037631	upregulated	inf								
XLOC_038003	upregulated	inf								
XLOC_038180	upregulated	inf								
XLOC_038181	upregulated	inf								
XLOC_038652	upregulated	inf		RLG						
XLOC_039046	upregulated	3.66								
XLOC_039655	upregulated	inf		RLG;RLX						
XLOC_039771	upregulated	inf			lncRNA					
XLOC_040274	upregulated	7.31								
XLOC_041407	upregulated	inf								
XLOC_041515	upregulated	inf		RLX						
XLOC_042817	upregulated	inf								
XLOC_042866	upregulated	inf								
XLOC_043499	upregulated	inf		RLG						
XLOC_043975	upregulated	inf		RLC						
XLOC_045293	upregulated	inf			lncRNA					
XLOC_048120	upregulated	inf			lncRNA					
XLOC_048272	upregulated	inf								
XLOC_048837	upregulated	inf			lncRNA					
XLOC_050965	upregulated	inf			lncRNA					
XLOC_051408	upregulated	inf			lncRNA					
XLOC_051992	upregulated	inf								
XLOC_052748	upregulated	inf		RLX						
XLOC_052864	upregulated	inf		RLG						
XLOC_053485	upregulated	inf			lncRNA					
XLOC_054167	upregulated	inf		RLC						
XLOC_054857	upregulated	inf			lncRNA					
XLOC_056458	upregulated	inf								
XLOC_056914	upregulated	inf								
XLOC_058903	upregulated	inf		RLG	lncRNA					
XLOC_059015	upregulated	1.36								
XLOC_059934	upregulated	inf			lncRNA					
XLOC_059935	upregulated	inf			lncRNA					
XLOC_059937	upregulated	inf								
XLOC_060458	upregulated	inf								
XLOC_060500	upregulated	4.88								
XLOC_063112	upregulated	inf		RLC						
XLOC_063550	upregulated	inf								
XLOC_064798	upregulated	inf		RLG;RLX						
XLOC_065059	upregulated	inf			lncRNA					
XLOC_065289	upregulated	inf			lncRNA					
XLOC_065292	upregulated	inf			lncRNA					
XLOC_065293	upregulated	inf			lncRNA					
XLOC_065675	upregulated	inf		RLC						
XLOC_066274	upregulated	inf			lncRNA					
XLOC_066469	upregulated	inf								
XLOC_066470	upregulated	inf			lncRNA					
XLOC_066729	upregulated	inf			lncRNA					
XLOC_066999	upregulated	8.50		RLG						

DE gene ID*	*rmr6-1*/wt	log2 fold change**	AGPv3.20 biotype	TE	lncRNA	Classical genes	MaizeGDB curated genes	ChromDB name	*Arabidopsis* homolog	*Arabidopsis* annotation
XLOC_067056	upregulated	inf		RLG						
XLOC_068310	upregulated	inf								
XLOC_068861	upregulated	inf		RLG;RLX	lncRNA					
XLOC_069791	upregulated	inf			lncRNA					
XLOC_070361	upregulated	inf		RLG;RLX						
XLOC_070875	upregulated	inf		RLX						
XLOC_070961	upregulated	inf		RLG						
XLOC_070963	upregulated	inf		RLG						
XLOC_070964	upregulated	inf								
XLOC_070965	upregulated	inf								
XLOC_070966	upregulated	inf								
XLOC_072688	upregulated	inf								
XLOC_073220	upregulated	inf								
XLOC_073307	upregulated	inf								
XLOC_073811	upregulated	inf								
XLOC_074067	upregulated	inf		RLC						
XLOC_075719	upregulated	inf		RLG;RLX						
XLOC_077137	upregulated	inf								
XLOC_077409	upregulated	inf		DXX;RLX						
XLOC_078658	upregulated	inf								
XLOC_078667	upregulated	inf		RLC						
XLOC_079928	upregulated	1.99		RLC						
XLOC_080091	upregulated	inf								
XLOC_081700	upregulated	2.16								
XLOC_085523	upregulated	inf			lncRNA					
XLOC_087097	upregulated	2.88								
XLOC_087098	upregulated	inf								
XLOC_087455	upregulated	inf		RLX;RLG						
XLOC_087626	upregulated	inf								
XLOC_088399	upregulated	inf								
XLOC_089783	upregulated	inf		RLX						
XLOC_090008	upregulated	inf								
XLOC_090057	upregulated	inf		RLX						
XLOC_090166	upregulated	inf								
XLOC_090551	upregulated	inf		RLX						
XLOC_091182	upregulated	inf		RLG;RLX						
XLOC_091183	upregulated	inf								
XLOC_091249	upregulated	inf		RLC						
XLOC_091703	upregulated	2.51		RLX						
XLOC_091833	upregulated	inf		RLC						
XLOC_093768	upregulated	inf		RLC						
XLOC_093958	upregulated	inf			lncRNA					
XLOC_094212	upregulated	inf								
XLOC_094884	upregulated	inf								
XLOC_095227	upregulated	inf		RLG						
XLOC_095427	upregulated	inf		RLX						
XLOC_096197	upregulated	inf								
XLOC_096333	upregulated	inf			lncRNA					
XLOC_096734	upregulated	inf		RLX						

DE gene ID*	*rmr6-1*/wt	log2 fold change**	AGPv3.20 biotype	TE	lncRNA	Classical genes	MaizeGDB curated genes	ChromDB name	*Arabidopsis* homolog	*Arabidopsis* annotation
XLOC_097236	upregulated	inf		RLC						
XLOC_097730	upregulated	inf		RLC						
XLOC_097741	upregulated	inf		RLC						
XLOC_098196	upregulated	inf								
XLOC_099714	upregulated	inf			lncRNA					
XLOC_101230	upregulated	inf		RLC						
XLOC_101231	upregulated	inf		RLC						
XLOC_101642	upregulated	inf		RLC						
XLOC_102492	upregulated	inf		RLC						
XLOC_102493	upregulated	inf			lncRNA					
XLOC_104774	upregulated	inf		RLC;RLG						
XLOC_105860	upregulated	inf		RLX	lncRNA					
XLOC_106398	upregulated	inf								
XLOC_106404	upregulated	inf		DTC						
XLOC_106540	upregulated	inf		RLC						
XLOC_107181	upregulated	inf								
XLOC_107843	upregulated	inf								
XLOC_108376	upregulated	inf								
XLOC_109841	upregulated	inf		RLG;RLX						
XLOC_110289	upregulated	inf		RLG;RLX						
XLOC_110780	upregulated	inf								
XLOC_111737	upregulated	3.24								
siRNA_Z27kG1_19135	upregulated	2.27								
AC187262.4_FG007	downregulated	-2.32	protein_coding							
AC194158.3_FG005	downregulated	-3.14	protein_coding						AT2G39220.1	PATATIN-like protein 6
AC194172.3_FG002	downregulated	−inf	low_confidence							
AC194264.3_FG013	downregulated	−inf	low_confidence		lncRNA					
AC200063.3_FG001	downregulated	−inf	low_confidence		lncRNA					
AC210731.3_FG002	downregulated	−inf	protein_coding		lncRNA					
AC234528.1_FG001	downregulated	-3.25	protein_coding						AT3G21215.1	RNA-binding (RRM/RBD/RNP motifs) family protein
ClusterV2_24	downregulated	-1.80								
Cluster_121	downregulated	-1.79		RLG						
Cluster_36	downregulated	-3.10								
GRMZM2G002368	downregulated	−inf	transposable_element	TXX						
GRMZM2G006765	downregulated	-1.68	protein_coding		lncRNA				AT5G09810.1	actin 7
GRMZM2G007936	downregulated	-2.15	low_confidence						AT5G05980.1	DHFS-FPGS homolog B
GRMZM2G012221	downregulated	−inf	low_confidence		lncRNA				AT1G11870.2	Seryl-tRNA synthetase
GRMZM2G012413	downregulated	-1.49	protein_coding						AT4G35730.1	Regulator of Vps4 activity in the MVB pathway protein
GRMZM2G012860_O_1	downregulated	-2.03								
GRMZM2G012970	downregulated	-1.80	protein_coding						AT2G38370.1	Plant protein of unknown function (DUF827)
GRMZM2G013555	downregulated	-1.72	protein_coding							
GRMZM2G013764	downregulated	−inf	transposable_element	TXX						
GRMZM2G015503	downregulated	-2.04	low_confidence						AT3G46730.1	NB-ARC domain-containing disease resistance protein
GRMZM2G017536	downregulated	-1.62	protein_coding						AT5G26910.1	
GRMZM2G020054	downregulated	-2.14	protein_coding				ereb54		AT3G20310.1	ethylene response factor 7
GRMZM2G020150	downregulated	-1.45	protein_coding				ereb196		AT3G15210.1	ethylene responsive element binding factor 4
GRMZM2G021069	downregulated	-1.62	protein_coding						AT5G44635.1	minichromosome maintenance (MCM2/3/5) family protein
GRMZM2G021885	downregulated	-1.46	protein_coding						AT1G50660.1	
GRMZM2G024653	downregulated	-1.53	low_confidence						AT4G02480.1	AAA-type ATPase family protein

DE gene ID*	*rmr6-1*/wt	log2 fold change**	AGPv3.20 biotype	TE	lncRNA	Classical genes	MaizeGDB curated genes	ChromDB name	*Arabidopsis* homolog	*Arabidopsis* annotation
GRMZM2G027728	downregulated	-1.36	protein_coding						AT1G77940.1	Ribosomal protein L7Ae/L30e/S12e/Gadd45 family protein
GRMZM2G027976	downregulated	-1.44	protein_coding				bzip53		AT2G42380.2	Basic-leucine zipper (bZIP) transcription factor family protein
GRMZM2G028346	downregulated	-1.52	protein_coding						AT2G27020.1	20S proteasome alpha subunit G1
GRMZM2G030009	downregulated	-1.51	protein_coding							
GRMZM2G030839	downregulated	-1.89	protein_coding						AT1G31910.1	GHMP kinase family protein
GRMZM2G031310	downregulated	−inf	low_confidence		lncRNA					
GRMZM2G031441	downregulated	−inf	protein_coding						AT1G75250.2	RAD-like 6
GRMZM2G031724	downregulated	-3.33	protein_coding						AT1G78440.1	Arabidopsis thaliana gibberellin 2-oxidase 1
GRMZM2G032878	downregulated	-1.61	protein_coding						AT1G54630.1	acyl carrier protein 3
GRMZM2G033828	downregulated	-1.64	protein_coding						AT3G12280.2	retinoblastoma-related 1
GRMZM2G034828	downregulated	-1.53	protein_coding						AT3G23670.1	phragmoplast-associated kinesin-related protein, putative
GRMZM2G035421	downregulated	-1.70	protein_coding						AT5G04040.1	Patatin-like phospholipase family protein
GRMZM2G036092	downregulated	-1.70	protein_coding				bhlh30		AT1G26945.1	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
GRMZM2G041732	downregulated	-4.21	protein_coding						AT5G08610.1	P-loop containing nucleoside triphosphate hydrolases superfamily protein
GRMZM2G042502	downregulated	-1.29	protein_coding						AT1G76550.1	Phosphofructokinase family protein
GRMZM2G044128	downregulated	-1.52	protein_coding				pza01623		AT3G57150.1	homologue of NAP57
GRMZM2G049021	downregulated	-1.99	protein_coding						AT3G16190.1	Isochorismatase family protein
GRMZM2G049781	downregulated	-2.51	protein_coding		lncRNA				AT1G23000.1	Heavy metal transport/detoxification superfamily protein
GRMZM2G049913	downregulated	-1.75	protein_coding		lncRNA					
GRMZM2G050307	downregulated	-2.22	protein_coding							
GRMZM2G055193	downregulated	-3.54	low_confidence						AT5G62000.1	auxin response factor 2
GRMZM2G055999	downregulated	-1.41	protein_coding						AT1G23000.1	Heavy metal transport/detoxification superfamily protein
GRMZM2G056750	downregulated	-1.60	protein_coding						AT2G38440.1	SCAR homolog 2
GRMZM2G057111	downregulated	-2.19	transposable_element	TXX						
GRMZM2G060451	downregulated	-1.51	protein_coding						AT5G42480.1	Chaperone DnaJ-domain superfamily protein
GRMZM2G061280	downregulated	-1.39	protein_coding						AT5G65270.1	RAB GTPase homolog A4A
GRMZM2G061791	downregulated	-2.40	protein_coding							
GRMZM2G065355	downregulated	-1.84	protein_coding		lncRNA		hsbp2		AT4G15802.1	heat shock factor binding protein
GRMZM2G067713	downregulated	-1.55	protein_coding						AT4G26090.1	NB-ARC domain-containing disease resistance protein
GRMZM2G068193	downregulated	-1.50	protein_coding						AT1G76540.1	cyclin-dependent kinase B2;1
GRMZM2G071089	downregulated	-1.74	protein_coding						AT4G20260.1	plasma-membrane associated cation-binding protein 1
GRMZM2G071277	downregulated	-1.69	protein_coding						AT3G61460.1	brassinosteroid-responsive RING-H2
GRMZM2G071387	downregulated	−inf	low_confidence		lncRNA					
GRMZM2G073401	downregulated	-1.95	protein_coding						AT1G14980.1	chaperonin 10
GRMZM2G075456	downregulated	-1.83	protein_coding						AT1G70670.1	Caleosin-related family protein
GRMZM2G076323	downregulated	−inf	low_confidence							
GRMZM2G076985	downregulated	-1.86	protein_coding						AT1G76160.1	SKU5 similar 5
GRMZM2G077227	downregulated	-1.71	protein_coding						AT2G26180.1	IQ-domain 6
GRMZM2G077755	downregulated	-1.82	protein_coding				tcptf26		AT3G47620.1	TEOSINTE BRANCHED, cycloidea and PCF (TCP) 14
GRMZM2G077858	downregulated	-1.76	protein_coding							
GRMZM2G078170	downregulated	-2.70	low_confidence						AT1G27170.1	transmembrane receptors;ATP binding
GRMZM2G078314	downregulated	-1.44	protein_coding					HTR103	AT1G09200.1	Histone superfamily protein
GRMZM2G079468	downregulated	-2.01	protein_coding						AT5G19840.1	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
GRMZM2G079638	downregulated	-3.15	protein_coding							
GRMZM2G081322	downregulated	-1.82	protein_coding						AT2G41740.1	villin 2
GRMZM2G083475	downregulated	-1.69	protein_coding						AT3G45850.1	P-loop containing nucleoside triphosphate hydrolases superfamily protein

DE gene ID*	*rmr6-1*/wt	log2 fold change**	AGPv3.20 biotype	TE	lncRNA	Classical genes	MaizeGDB curated genes	ChromDB name	*Arabidopsis* homolog	*Arabidopsis* annotation
GRMZM2G083590	downregulated	-1.70	low_confidence							
GRMZM2G083749	downregulated	-1.56	protein_coding						AT4G29100.1	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
GRMZM2G086614	downregulated	-1.58	protein_coding				c3h6		AT3G51950.1	Zinc finger (CCCH-type) family protein / RNA recognition motif (RRM)-containing protein
GRMZM2G086934	downregulated	-1.92	protein_coding				rfa1		AT5G08020.1	RPA70-kDa subunit B
GRMZM2G090149	downregulated	-2.00	protein_coding						AT1G79610.1	Na+/H+ antiporter 6
GRMZM2G095655	downregulated	-1.72	protein_coding						AT3G17020.1	Adenine nucleotide alpha hydrolases-like superfamily protein
GRMZM2G096247	downregulated	-1.45	protein_coding				gst9		AT2G47730.1	glutathione S-transferase phi 8
GRMZM2G098714	downregulated	-1.76	protein_coding						AT5G08020.1	RPA70-kDa subunit B
GRMZM2G099295	downregulated	-1.63	protein_coding						AT5G06860.1	polygalacturonase inhibiting protein 1
GRMZM2G099820	downregulated	-4.21	protein_coding						AT5G45480.1	Protein of unknown function (DUF594)
GRMZM2G104146	downregulated	-2.54	low_confidence						AT2G22730.1	Major facilitator superfamily protein
GRMZM2G105587	downregulated	-1.61	protein_coding							
GRMZM2G106792_O_1	downregulated	-1.43								
GRMZM2G107003	downregulated	-1.57	protein_coding						AT5G14920.1	Gibberellin-regulated family protein
GRMZM2G108149	downregulated	-1.40	protein_coding						AT3G57880.1	Calcium-dependent lipid-binding (CaLB domain) plant phosphoribosyltransferase family protein
GRMZM2G109814	downregulated	-1.49	protein_coding							
GRMZM2G110726	downregulated	-1.68	protein_coding						AT5G53400.1	HSP20-like chaperones superfamily protein
GRMZM2G112210	downregulated	-2.84	protein_coding						AT5G42510.1	Disease resistance-responsive (dirigent-like protein) family protein
GRMZM2G112686	downregulated	-1.56	protein_coding						AT5G45910.1	GDSL-like Lipase/Acylhydrolase superfamily protein
GRMZM2G113191	downregulated	-2.17	protein_coding						AT1G08080.1	alpha carbonic anhydrase 7
GRMZM2G113840	downregulated	-1.65	protein_coding						AT4G39170.1	Sec14p-like phosphatidylinositol transfer family protein
GRMZM2G114552	downregulated	-2.08	protein_coding							
GRMZM2G116632	downregulated	-1.73	protein_coding						AT2G32300.1	uclacyanin 1
GRMZM2G117028	downregulated	-1.43	protein_coding						AT2G02480.1	AAA-type ATPase family protein
GRMZM2G117281	downregulated	-2.17	protein_coding							
GRMZM2G121088	downregulated	−inf	low_confidence							
GRMZM2G121929	downregulated	−inf	low_confidence		lncRNA					
GRMZM2G124718	downregulated	-2.24	protein_coding				parp1		AT4G02390.1	poly(ADP-ribose) polymerase
GRMZM2G125020	downregulated	-1.59	protein_coding						AT3G63210.1	Protein of unknown function (DUF581)
GRMZM2G126285	downregulated	-2.21	protein_coding						AT4G13010.1	Oxidoreductase, zinc-binding dehydrogenase family protein
GRMZM2G127695	downregulated	-1.70	protein_coding							
GRMZM2G131756	downregulated	-3.68	protein_coding						AT2G36490.1	demeter-like 1
GRMZM2G132371	downregulated	-1.68	protein_coding						AT2G36200.1	P-loop containing nucleoside triphosphate hydrolases superfamily protein
GRMZM2G139399	downregulated	-2.44	low_confidence						AT5G59030.1	copper transporter 1
GRMZM2G139662	downregulated	−inf	low_confidence						AT5G13520.1	peptidase M1 family protein
GRMZM2G139963	downregulated	-1.84	protein_coding				hb102		AT1G69780.1	Homeobox-leucine zipper protein family
GRMZM2G141386	downregulated	-1.74	protein_coding						AT3G54770.1	RNA-binding (RRM/RBD/RNP motifs) family protein
GRMZM2G145444	downregulated	-1.80	protein_coding				myb50		AT5G67300.1	myb domain protein r1
GRMZM2G145690	downregulated	-1.74	protein_coding				hb71		AT1G73360.1	homeodomain GLABROUS 11
GRMZM2G145756	downregulated	-1.60	protein_coding						AT5G18700.1	Protein kinase family protein with ARM repeat domain
GRMZM2G149024	downregulated	-2.55	protein_coding						AT3G50760.1	galacturonosyltransferase-like 2
GRMZM2G149624	downregulated	-1.43	protein_coding						AT2G21140.1	proline-rich protein 2
GRMZM2G151390	downregulated	-3.77	transposable_element	DTC						
GRMZM2G151430	downregulated	-3.57	transposable_element	DTC						
GRMZM2G152447	downregulated	-2.10	protein_coding						AT1G14700.1	purple acid phosphatase 3

DE gene ID*	*rmr6-1*/wt	log2 fold change**	AGPv3.20 biotype	TE	lncRNA	Classical genes	MaizeGDB curated genes	ChromDB name	*Arabidopsis* homolog	*Arabidopsis* annotation
GRMZM2G152789	downregulated	−inf	low_confidence		lncRNA					
GRMZM2G153286	downregulated	-2.29	transposable_element	RLX						
GRMZM2G153823	downregulated	-1.98	protein_coding						AT3G02520.1	general regulatory factor 7
GRMZM2G153945	downregulated	-2.04	protein_coding						AT3G59350.1	Protein kinase superfamily protein
GRMZM2G154124	downregulated	-1.65	protein_coding						AT1G09890.1	Rhamnogalacturonate lyase family protein
GRMZM2G154414	downregulated	-1.82	protein_coding						AT1G49620.1	Cyclin-dependent kinase inhibitor family protein
GRMZM2G154442	downregulated	-2.47	protein_coding							
GRMZM2G158316	downregulated	-1.38	protein_coding						AT3G50950.1	HOPZ-ACTIVATED RESISTANCE 1
GRMZM2G159105	downregulated	-1.64	protein_coding						AT1G51940.1	protein kinase family protein / peptidoglycan-binding LysM domain-containing protein
GRMZM2G159369	downregulated	-1.33	protein_coding						AT2G47470.1	thioredoxin family protein
GRMZM2G163771	downregulated	-5.78	low_confidence							
GRMZM2G169859	downregulated	-3.08	transposable_element	TXX						
GRMZM2G169914	downregulated	-1.99	protein_coding						AT1G34065.1	S-adenosylmethionine carrier 2
GRMZM2G171408	downregulated	-1.43	protein_coding						AT4G31890.1	ARM repeat superfamily protein
GRMZM2G172734	downregulated	-2.49	protein_coding						AT5G45470.1	Protein of unknown function (DUF594)
GRMZM2G176307	downregulated	-1.31	protein_coding				gpc4		AT1G13440.1	glyceraldehyde-3-phosphate dehydrogenase C2
GRMZM2G178686	downregulated	-2.57	protein_coding							
GRMZM2G179638	downregulated	-1.49	protein_coding						AT2G22795.1	
GRMZM2G180246	downregulated	-2.95	protein_coding						AT5G28640.1	SSXT family protein
GRMZM2G301663	downregulated	-1.90	protein_coding						AT1G02900.1	rapid alkalinization factor 1
GRMZM2G304362	downregulated	-1.36	protein_coding						AT2G21790.1	ribonucleotide reductase 1
GRMZM2G304575	downregulated	-1.46	protein_coding						AT3G45980.1	Histone superfamily protein
GRMZM2G324248	downregulated	-1.39	protein_coding						AT5G35090.1	
GRMZM2G324973	downregulated	-2.39	protein_coding							
GRMZM2G326328	downregulated	-1.63	protein_coding					CPD101	AT3G57060.1	binding
GRMZM2G326825	downregulated	−inf	protein_coding							
GRMZM2G330455	downregulated	-1.78	protein_coding							
GRMZM2G333923	downregulated	-1.91	protein_coding						AT3G02430.1	Protein of unknown function (DUF679)
GRMZM2G333980	downregulated	-1.43	protein_coding						AT5G06860.1	polygalacturonase inhibiting protein 1
GRMZM2G337190	downregulated	-2.92	protein_coding							
GRMZM2G342515	downregulated	-1.61	protein_coding				his2b5	HTB104	AT3G45980.1	Histone superfamily protein
GRMZM2G346865	downregulated	-1.71	protein_coding						AT3G51980.1	ARM repeat superfamily protein
GRMZM2G355572	downregulated	-2.09	protein_coding						AT5G12010.1	
GRMZM2G357660	downregulated	-2.27	protein_coding							
GRMZM2G372074	downregulated	-1.46	protein_coding						AT3G22142.1	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
GRMZM2G374475	downregulated	-1.66	protein_coding						AT1G75900.1	GDSL-like Lipase/Acylhydrolase superfamily protein
GRMZM2G378653	downregulated	-1.85	protein_coding				bhlh69		AT2G42300.1	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
GRMZM2G398807	downregulated	-1.42	protein_coding						AT4G12520.1	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
GRMZM2G400126	downregulated	-2.33	low_confidence						AT5G66150.1	Glycosyl hydrolase family 38 protein
GRMZM2G401147	downregulated	-2.07	protein_coding					HTB105	AT3G45980.1	Histone superfamily protein
GRMZM2G402987	downregulated	-2.52	protein_coding						AT1G12380.1	
GRMZM2G406024	downregulated	−inf	protein_coding							
GRMZM2G413113	downregulated	-1.66	protein_coding						AT1G52770.1	Phototropic-responsive NPH3 family protein
GRMZM2G423555	downregulated	-1.31	protein_coding						AT2G30870.1	glutathione S-transferase PHI 10
GRMZM2G427722	downregulated	-1.62	low_confidence							
GRMZM2G429213	downregulated	-1.71	protein_coding						AT3G06290.1	SAC3/GANP/Nin1/mts3/eIF-3 p25 family

DE gene ID*	*rmr6-1*/wt	log2 fold change**	AGPv3.20 biotype	TE	lncRNA	Classical genes	MaizeGDB curated genes	ChromDB name	*Arabidopsis* homolog	*Arabidopsis* annotation
GRMZM2G429857	downregulated	−inf	protein_coding		lncRNA					
GRMZM2G433731	downregulated	-2.22	protein_coding						AT4G26090.1	NB-ARC domain-containing disease resistance protein
GRMZM2G440643	downregulated	−inf	transposable_element	TXX						
GRMZM2G440902	downregulated	-2.28	protein_coding						AT2G30140.1	UDP-Glycosyltransferase superfamily protein
GRMZM2G447984	downregulated	-1.66	protein_coding					HTR106	AT1G09200.1	Histone superfamily protein
GRMZM2G449123	downregulated	-2.44	protein_coding						AT3G20250.1	pumilio 5
GRMZM2G457411	downregulated	-2.08	protein_coding						AT3G18930.1	RING/U-box superfamily protein
GRMZM2G460861	downregulated	-1.88	protein_coding						AT5G20820.1	SAUR-like auxin-responsive protein family
GRMZM2G468016	downregulated	−inf	low_confidence							
GRMZM2G470307	downregulated	-1.84	protein_coding		lncRNA		myb117		AT4G32730.2	Homeodomain-like protein
GRMZM2G470499	downregulated	-2.67	protein_coding						AT3G24550.1	proline extensin-like receptor kinase 1
GRMZM2G477325	downregulated	-1.43	protein_coding				pmpm5		AT3G05880.1	Low temperature and salt responsive protein family
GRMZM2G478140	downregulated	-1.80	transposable_element	TXX						
GRMZM2G482621	downregulated	-2.15	transposable_element	RLG;RLX;XXX						
GRMZM2G490278	downregulated	−inf	protein_coding							
GRMZM2G491632	downregulated	-1.60	protein_coding						AT2G36930.1	zinc finger (C2H2 type) family protein
GRMZM2G492387	downregulated	−inf	low_confidence							
GRMZM2G519229	downregulated	-1.44	low_confidence							
GRMZM2G520154	downregulated	-1.70	transposable_element	TXX						
GRMZM2G533098	downregulated	−inf	transposable_element	TXX						
GRMZM2G548306	downregulated	−inf	low_confidence		lncRNA					
GRMZM2G550100	downregulated	−inf	transposable_element	TXX						
GRMZM2G553422	downregulated	-1.70	low_confidence							
GRMZM2G560870	downregulated	−inf	low_confidence							
GRMZM2G563405	downregulated	−inf	low_confidence							
GRMZM2G567718	downregulated	−inf	low_confidence							
GRMZM2G569010	downregulated	−inf	transposable_element	DTC;RLX						
GRMZM2G575619	downregulated	-1.50	transposable_element	RLG;RLX						
GRMZM2G583274	downregulated	−inf	low_confidence							
GRMZM2G583828	downregulated	−inf	transposable_element	TXX						
GRMZM2G585315	downregulated	-1.71	low_confidence							
GRMZM5G811972	downregulated	−inf	low_confidence		lncRNA					
GRMZM5G813011	downregulated	-1.87	low_confidence							
GRMZM5G822388	downregulated	-1.67	low_confidence							
GRMZM5G826838	downregulated	-1.49	protein_coding						AT5G17330.1	glutamate decarboxylase
GRMZM5G826899	downregulated	-1.90	low_confidence							
GRMZM5G828987	downregulated	-2.97	protein_coding						AT1G26560.1	beta glucosidase 40
GRMZM5G833523	downregulated	−inf	low_confidence							
GRMZM5G835629	downregulated	-1.55	protein_coding						AT3G26700.1	Protein kinase superfamily protein
GRMZM5G838666	downregulated	-1.93	low_confidence							
GRMZM5G848696	downregulated	-1.71	protein_coding						AT1G17650.1	glyoxylate reductase 2
GRMZM5G851266	downregulated	-1.95	protein_coding							
GRMZM5G852504	downregulated	-1.35	low_confidence							
GRMZM5G858454	downregulated	-1.58	protein_coding						AT4G26970.1	aconitase 2
GRMZM5G860761	downregulated	-1.94	protein_coding		lncRNA				AT2G43670.1	Carbohydrate-binding X8 domain superfamily protein
GRMZM5G861100	downregulated	-4.43	low_confidence	RLX						
GRMZM5G862602	downregulated	-1.29	protein_coding							
GRMZM5G869453	downregulated	-2.69	low_confidence		lncRNA				AT3G52990.2	Pyruvate kinase family protein

DE gene ID*	*rmr6-1*/wt	log2 fold change**	AGPv3.20 biotype	TE	lncRNA	Classical genes	MaizeGDB curated genes	ChromDB name	*Arabidopsis* homolog	*Arabidopsis* annotation
GRMZM5G872443	downregulated	-1.81	protein_coding						AT3G51520.1	diacylglycerol acyltransferase family
GRMZM5G873259	downregulated	−inf	transposable_element	DTC;RLC						
GRMZM5G873586	downregulated	-1.86	protein_coding				lbd44		AT4G23150.1	cysteine-rich RLK (RECEPTOR-like protein kinase) 7
GRMZM5G875516_O_1	downregulated	-1.88								
GRMZM5G882225	downregulated	-1.85	low_confidence							
GRMZM5G890025	downregulated	-1.48	transposable_element	TXX						
GRMZM5G895702	downregulated	-2.02	transposable_element	TXX						
XLOC_003494	downregulated	−inf								
XLOC_008366	downregulated	−inf								
XLOC_009976	downregulated	−inf			lncRNA					
XLOC_012839	downregulated	-3.03								
XLOC_013136	downregulated	-1.91								
XLOC_013725	downregulated	−inf		RLG						
XLOC_016699	downregulated	−inf			lncRNA					
XLOC_022016	downregulated	−inf								
XLOC_025961	downregulated	−inf			lncRNA					
XLOC_039294	downregulated	−inf								
XLOC_056886	downregulated	−inf		RLG						
XLOC_068648	downregulated	−inf								
XLOC_074129	downregulated	−inf		RLG;RLX						
XLOC_082596	downregulated	-1.93								
XLOC_088986	downregulated	-1.70								
XLOC_088992	downregulated	-2.89			lncRNA					
XLOC_095055	downregulated	-1.56								
XLOC_098110	downregulated	-1.82								