UNIVERSITY OF MODENA AND REGGIO EMILIA
PhD School in *AgriFood Sciences, Technologies and Biotechnologies*

Identification of microRNAs involved in plant development and abiotic stress response

Ph.D. student: Lorenzo Maria Giusti

XXVI cycle

Supervisor: Prof. Enrico Francia

Co-supervisor: Dott. Luigi Cattivelli

Dean of Ph.D School: prof. Andrea Pulvirenti

"We find after years of struggle that we do not take a trip; a trip takes us"

(Le persone non fanno i viaggi, sono i viaggi che fanno le persone)

J. Steinbeck, *Travels with Charley: In Search of America*

To my beautiful family

and my beautiful future wife, Sara

Abstract

MicroRNAs are a class of small RNAs with fundamental roles in key plant biological processes, such as development and environmental stress response. They act on gene regulation mainly at post-transcriptional level through sequence-based interaction with target mRNAs.
Until now several studies on plant microRNAs have been carried out, but only few of them focused on a deep characterization of their functions in species of agricultural interest like wheat and barley.
This thesis set out to unravel the roles of microRNAs in different environmental conditions and different plant species. To do so, four ad-hoc experiments were set up and a deep sequencing of small RNA transcriptomes was carried out.
The first experiment concerned the study of the peach tree development, taking into consideration five different organs and/or phenological stages, in which the small RNA population was characterized. The bioinformatic analysis focused on isomiRs generated from each locus, and outlined the surprising complexity of the peach miRNome.
In the second experiment the barley response to acclimation was explored. Two cultivars, Nure and Tremois, a winter and a spring type, respectively, were exposed to low temperatures (4° C) at the first-leaf stage for a time ranging from 2 hours to 6 days. Several cold-related microRNAs were found. Moreover, an interesting result was found for miR397. It showed an expression trend likely associated with the growth habit. This correlation was further analyzed in a new panel of 24 spring, winter and alternative barley cultivars.
The third experiment consisted in a comparison between bread wheat euplasmic and alloplasmic lines, in which the substitution of the wheat cytoplasm with alien cytoplasm occurred. It aimed at the investigation of a possible regulation of microRNAs involved the crosstalk between the nuclear and the organellar genomes.
The fourth experiment explored the durum wheat response to high temperatures and drought stress at microRNA level. Two cultivars, Ofanto and Cappelli, were chosen for their contrasting behavior in these conditions, and exposed to a stress treatment, *i.e.* heat stress (3 hours at 36 °C) or drought stress (two levels with final Soil Water Content of 18% and 12.5%, respectively).
Sixty-nine small RNA libraries were prepared from samples coming from these four experiments and sequenced on an Illumina Genome Analyzer IIx.
Concerning the last two experiments, all bioinformatic analyses were carried out with a free, web-available tool that allowed the identification of microRNAs whose mature sequence was already annotated in miRBase.
Subsequently a standardized and reliable qRT-PCR method based on the use of stem-loop primers was developed to validate Illumina results. MicroRNAs expression data were normalized with different classic housekeeping genes. Reference genes belonging to the small RNA class were considered as well, and good results came from some barley and wheat small nucleolar RNAs (snoRNAs).
An in silico target search for differentially expressed barley and wheat microRNAs was carried out with a free web-based tool, and a 5' RACE procedure for experimental validation of these targets was optimized.
The expression trend of identified putative target genes in control and stressed samples was checked by qPCR, using primers straddling the miRNA cleavage site.
Finally, special degradome libraries from the wheat experiment were prepared and sequenced on an Illumina HiSeq 2500.
The results obtained shed new light on the complex networks underlying plants behavior during processes such as development and response to stresses, and represent a step forward in the understanding of the mechanism of action of these tiny regulators. They also provide a useful starting point for future studies on genes interactions regulating these key biological processes.

Riassunto

I microRNA costituiscono una classe di small RNA con un ruolo di fondamentale importanza in processi biologici chiave nelle piante, come lo sviluppo e la risposta a stress ambientali. Regolano l'espressione genica agendo principalmente a livello post-trascrizionale, attraverso il riconoscimento di una sequenza a loro complementare nell'RNA messaggero target.
Fino ad oggi sono stati prodotti molti lavori aventi come argomento i microRNA in pianta, ma solo pochi di questi hanno riguardato specie di interesse agrario come orzo e frumento.
Questa tesi aveva l'obiettivo di esplorare il ruolo dei microRNA in diverse condizioni ambientali e in diverse specie vegetali. Per fare ciò, sono stati predisposti quattro esperimenti.
Il primo esperimento ha riguardato lo studio dello sviluppo del pesco, considerando campioni provenienti da 5 fasi fenologiche tra rottura delle gemme e fioritura. L'analisi bioinformatica si è focalizzata sulle isoforme generate dai diversi loci genomici, ed ha delineato la sorprendente complessità del miRNoma di pesco.
Il secondo esperimento si proponeva di esplorare la risposta dell'orzo all'acclimatamento. Nure e Tremois, due cultivar di orzo rispettivamente invernale e primaverile, sono state esposte a basse temperature (4° C) per un tempo variabile tra 2 ore e 6 giorni. Un interessante risultato ha riguardato il miR397, il cui trend di espressione è probabilmente associato all'habitus di crescita. Questa correlazione è stata ulteriormente studiata in un panel di 24 varietà di orzo comprendente tipi primaverili, invernali e facoltativi.
Il terzo esperimento ha visto il confronto di linee euplasmiche di frumento tenero con corrispondenti linee alloplasmiche, nelle quali era avvenuta una sostituzione di citoplasma con quello di altre specie evolutivamente collegate. L'obiettivo era lo studio di eventuali microRNA coinvolti nel cross-talk tra il genoma nucleare e quello degli organelli citoplasmatici.
Il quarto esperimento aveva come obiettivo la caratterizzazione della risposta del frumento duro allo stress da siccità e da calore a livello di microRNA. Sono state scelte due cultivar, Ofanto e Cappelli, per il loro comportamento contrastante in tali condizioni. Le piante sono state esposte a stress in terza foglia (calore – 3 ore a 36° – oppure siccità, con due livelli finali di Soil Water Content del 18% e del 12.5%).
Da questi quattro esperimenti sono state costruite 69 librerie, poi sequenziate con un Illumina Genome Analyzer IIx.
Per quanto riguarda gli ultimi due esperimenti, i dati provenienti dalla corsa sono stati analizzati con un programma freeware che ha permesso l'identificazione dei microRNA la cui sequenza matura fosse già annotata e presente nel database miRBase.
E' stata messa a punto una procedura standardizzata e ripetibile per la validazione dei dati Illumina, basata su una strategia di qPCR facente uso di primer stem-loop. I dati di espressione dei microRNA sono stati normalizzati con geni housekeeping classici e con normalizzatori appartenenti alla classe degli small RNA.
Successivamente è stata effettuata una ricerca bioinformatica dei target dei miRNA differenzialmente espressi negli esperimenti di orzo e frumento ed è stata ottimizzata una procedura di RACE per la validazione sperimentale di queste predizioni. Il trend di espressione dei putativi geni target identificati è stato analizzato mediante qPCR.
Infine sono state preparate delle particolari librerie di prodotti di degradazione, che sono state sequenziate con un Illumina HiSeq 2500.
I risultati ottenuti costituiscono un passo in avanti nella comprensione del meccanismo d'azione di questi piccoli regolatori endogeni e dei network complessi che governano il comportamento delle piante in processi come sviluppo e risposta a stress ambientali, e pongono le basi per studi futuri sulle interazioni geniche che regolano questi processi.

Acknowledgements

First and foremost, I would like to thank Dr. Luigi Cattivelli, director of the Genomics Research Centre of Fiorenzuola d'Arda, for providing me with these three years of research experience. They have been fantastic. He gave me the opportunity to work in a very good environment and in an excellent project, to improve my scientific background and to know much about myself and my future. I am also sincerely grateful to him, and always will be, for the help with the choice of the foreign laboratory in which I spent six months last year, and for finding the financial support for the internship, which have been indeed a priceless experience.

I would like to acknowledge Prof. Andrea Pulvirenti, dean of the Doctorate School, for his activity of coordinator, and Prof. Enrico Francia for his tutoring activity and for his aid in the writing.

I want to thank Dr. Cristina Crosatti for her fundamental help in the thesis project. Without her invaluable lab assistance and supervision, from the RNA extraction to the Illumina libraries construction, this work would have not been possible. She always urged me to broaden my point of view instead of focusing on small things, and has been a true reference point, teaching me how to work in a laboratory. I thank her also for the help in the revision of the present manuscript. In addition, I wish to acknowledge Dr. Primetta Faccioli and Dr. Moreno Colaiacovo for their essential bioinformatic support, a thing I could not do without. Special thanks also to Dr. Luigi Orrù for the technical support with the Genome Analyzer.

My heartfelt thanks must go to Prof. Antonio Michele Stanca, a daily motivator and an invaluable counselor, he always encouraged me from the first day and gave me a lot of energy and good suggestions. Thank you to all the lab members in Fiorenzuola, from the undergraduates to the PI's, for the aid given in support to my thesis research, their support and their friendship – from the soccer matches to the RACE protocol adjustments.

I want to thank Prof. Peter Langridge for welcoming me into his lab (the ACPFG), at the Plant Research Centre in Adelaide, and Dr. Bujun Shi and Dr. Jannatul Ferdous for tutoring me during the internship. I learnt a lot of new things about experiments planning and lab practices, and I got involved into many really instructive discussions. The ACPFG is an ideal work environment, it holds an unparalleled knowledge on plant genetics and genomics together with a fantastic and relaxed atmosphere. I had a great, unforgettable time down in South Australia, and for this I have to thank also my "australian" mates – Juan Carlos, Mathieu, Amelie, Agathe, John, Mark and 30 Cudmore Avenue, Iman, Vahid, Domenico, Fabio, Judith, Lucile, Shefat, Andy and the soccer guys, Debbie, Tony and the waterpolo team, and so on…

No words can truly express my infinite gratefulness to my family. Their constant support, their trust in me and in what I have been doing for years – maybe even without comprehending why I was doing that – made everything possible. They encouraged me to accept challenges and solve problems, they cheered up with me at every great moment and supported me whenever I needed. Thank you!

Last, but not least, I want to say thank you to the most special person on Earth, my future wife Sara. For her unconditional love and understanding during the past years, for lighting up the way, "for bringing me here, for showing me home. Finally I've found that I belong here" (thank you Dave). A beautiful trip ends, but a new, exciting, sensational one is just about to start, and I am looking forward to it.

Table of contents

Abstract ..... *III*

Riassunto ..... *IV*

Acknowledgements ..... *V*

1. Introduction 4

1.1 A brief history of RNA interference 4

1.2 RNAi effectors: core machinery 6

1.2.1 Argonaute superfamily 7

1.2.2 Dicer family 8

1.2.3 RDR family 9

1.3 RNAi effectors: small silencing RNA classes 9

1.3.1 miRNAs 10

1.3.2 siRNAs 11

1.3.3 piRNAs 13

1.3.4 Other small silencing RNAs 13

1.4 Plant miRNAs 14

1.4.1 Biogenesis 14

1.4.2 Evolution 15

1.4.3 Modes of action 16

1.4.4 Role in development 18

1.4.5 Role in abiotic stress response 21

2. Objectives and thesis outline 24

3. Materials and Methods 25

3.1 Experiments setup 25

3.2 RNA Isolation 27

3.3 Small RNA libraries preparation 28

3.4 Illumina data analysis 30

3.5 Stem-loop qRT-PCR on microRNAs 31

3.6 5'-RLM-RACE 33

3.7 Degradome libraries preparation 34

3.8 qPCR on target genes 36

4. Results – *part I: In silico miRNA discovery* 38

4.1 Introduction 38

4.2 Paper: *On the complexity of miRNA-mediated regulation in plants: novel insights into the genomic organization of plant miRNAs* 39

5. Results – *part II: NGS-based miRNA discovery* 46

5.1 General introduction 46

5.2 Experiment 1 - *Peach and development* 48

Introduction 48

*PAPER: **A survey of microRNA length variants contributing to miRNome complexity in peach (Prunus persica L.)** 49*

5.3 Experiment 2 – *Barley response to low temperatures* 66

Introduction 66

Deep sequencing – miRNA families 66

Deep Sequencing – Differential constitutive expression of microRNAs 68

Further tests 69

Deep sequencing – Nure and Tremois response to low temperatures 70

Discussion 72

5.4 Experiment 3 – *MicroRNAs involved in nucleus-organelles crosstalk* 76

Introduction 76

Results and discussion 76

5.5 Experiment 4 – *Durum wheat response to drought and heat stress* 79

Introduction 79

Deep sequencing – miRNA families 79

Deep sequencing – Differential constitutive expression of microRNAs 80

Deep sequencing – Ofanto and Cappelli response to drought stress 81

Deep sequencing – Ofanto and Cappelli response to heat stress 82

Results confirmation – Radioso and Trinakria experiment 84

Target search 87

Discussion 87

6. General Discussion 94

7. References 98

8. Annexes 106

PAPER: *Post-transcriptional and post-translational modifications controlling cold response* 107

PAPER: ***Cytoplasmic genome substitution in wheat affects the nuclear-cytoplasmic cross-talk leading to transcript and metabolite alterations*** **..... 118**

Introduction

1.1 A brief history of RNA interference

[reviewed in Matzke and Matzke, 2004; Montgomery, 2004; Eamens *et al.*, 2008; Almeida *et al.*, 2011]

RNA interference (RNAi) is a mechanism that inhibits gene expression at the stage of translation or transcription of specific genes. It has become a major focus of molecular biology and biomedical research around the world, and it represents nowadays one of the most interesting and innovative research issues in the field of functional genomics. This phenomenon was first noted as a surprise observation in the late 1980s by plant and animal scientists working on plant transformation experiments and nematode developmental mutants.

In an early study by Matzke *et al.* (1989) two different T-DNA complexes were introduced into the tobacco (*Nicotiana tabacum*) genome. Each complex encoded different proteins, but contained identical promoter regions. The authors reported that, unexpectedly, the first transgene complex, stably active on its own, often became inactive in the presence of the second. It was shown that silencing occurred via DNA methylation at promoter level, and that it was reversed when the transgene complexes segregated from each other in the progeny. This suggested that initiation of gene inactivation was dependent on the genomic integration of the second inserted complex, and that it was probably due to interactions between the common promoter regions.

One year after, Jorgensen and colleagues noted another strange phenomenon in petunia (*Petunia hybrida*) plants. They inserted additional copies of a key flower pigmentation gene, chalcone synthase (CHS). The overexpression of this gene intensified the purple coloration of petals, which unexpectedly led to partial or complete loss of color (Figure 1). Further analyses demonstrated that in some plant lines both introduced and endogenous forms of the CHS gene were “turned off”, in a phenomenon the authors called “co-suppression” (Napoli *et al.*, 1990; Van der Krol *et al.*, 1990).

Fig. 1 *Transformation of wild-type petunia (left) with a transgene encoding a pigment protein can lead to loss of pigment (white areas), owing to co-suppression of the transgene and homologous endogenous plant gene (from Matzke and Matzke, 2004).*

In the following years co-suppression was found to be present in other plants and in other kingdoms, like in the case of the so called “quelling” in the filamentous fungus *Neurospora crassa* (Romano and Macino, 1992). The underlying mechanism responsible for these curious observations was unknown at the time, and so remained for many years.

In 1993, a study by Lindbo and colleagues allowed a step forward in the comprehension of this phenomenon. They showed that plant RNA viruses could be both initiators and targets of silencing. Plants expressing a transgene encoding a truncated viral coat protein became resistant to the corresponding virus, owing to the mutual degradation of viral RNA and transgene mRNA. These observations led the authors to the conclusion that this cosuppression was localized to the cytoplasm and occurred at the post-transcriptional level.

At the same time, in the early 1990s, other laboratories were working on the same issue, RNAi, after the discovery of a null mutation in a gene of the nematode *Caenorhabditis elegans*. The joint efforts of Victor Ambros's and Gary Ruvkun's laboratories led to the discovery of a peculiar mechanism underlying gene regulation, and set a research milestone in the field of RNAi. The *C. elegans lin-4* gene negatively regulates the level of the LIN-14 protein, which has to decrease to allow the starting of the first larval stage. *lin-4* loss of function causes therefore a failure in the larval development of the nematode. The authors interestingly noted that *lin-4* does not encode a protein, but a couple of short transcripts of 22 and 61 nucleotides that show a sequence complementarity to the 3' untranslated region (UTR) of the lin-14 mRNA. They suggested a model in which *lin-4* regulates lin-14 translation via an antisense RNA-RNA interaction: it is able to create a temporal gradient of the LIN-14 protein by modulating its translation (Lee *et al.*, 1993; Wightman *et al.*, 1993).

The authors themselves were surprised about this discovery (Lee *et al.*, 2004). At that time they ignored that such a mechanism is one of the most fundamental and conserved gene control mechanisms in eukaryotic organisms. *lin-4* is considered the founding member of a class of tiny gene regulators subsequently dubbed microRNAs (miRNAs).

To this point, the three possible levels of action of the so-called RNAi were already elucidated: transcriptional, that is, DNA methylation (Matzke *et al.*, 1989), post-transcriptional, *i.e.* cosuppression (Napoli *et al.* 1990; Van der Krol *et al.*, 1990; Lindbo *et al.*, 1993) or quelling (Romano and Macino, 1992), and translational (Lee *et al.*, 1993; Wightman *et al.*, 1993).

The term RNAi however was coined some years after, in 1998, when Andrew Fire and Craig Mello published their discovery of a mechanism that can degrade mRNA from a specific gene. This mechanism is activated when RNA molecules occur as double-stranded pairs in the cell, directing the destruction of mRNAs containing similar sequences.

Fire and Mello were investigating how gene expression is regulated in the nematode *C. elegans*. The injection of mRNA molecules encoding a muscle protein led to no changes in the behaviour of the worms, and so did 'antisense' RNA molecules, which can pair with the endogenous 'sense' mRNA. But when they injected 'sense' and 'antisense' RNA together, the worms displayed peculiar, twitching movements, similar to those that were seen in worms that completely lacked a functioning gene for the muscle protein (Figure 2). This happened also for several other worm proteins: the injection of double-stranded RNA (dsRNA) directed the

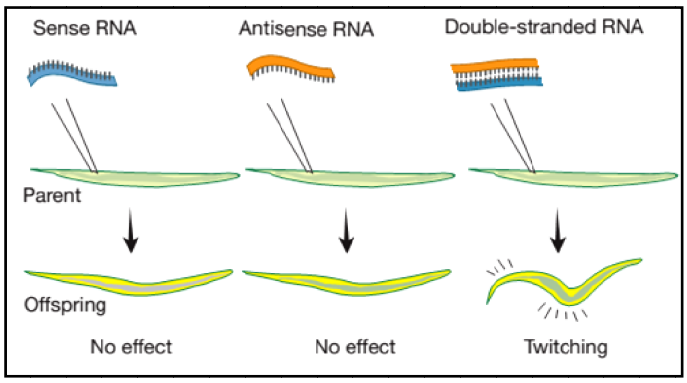


Fig. 2 *mRNA of a muscle protein is injected into the nematodes. Single-stranded RNA has no effect, but when dsRNA is injected, the worms start twitching like worms carrying a defective gene for that muscle protein.*

destruction of mRNAs containing similar sequences, therefore causing gene-silencing. The proteins encoded by those genes were no longer formed (Fire *et al.*, 1998).

This mechanism, dsRNA-mediated silencing, was then employed as a technique for gene functional analyses in a broad variety of organisms. For this discovery Fire and Mello were awarded a Nobel Prize in Physiology and Medicine in 2006.

It is very likely that RNA silencing evolved as a defense mechanism towards viral infection and transposon mobilization. Many viruses have dsRNA intermediates. When such a virus infects a cell these dsRNA intermediates trigger the host RNAi machinery, which thereby degrades the invading viral RNA allowing the cell to survive the infection. The evolution of a natural protection was shown for the first time in 1997 in plants, with two research papers published in Science (Ratcliff *et al.*, 1997) and Nature (Covey *et al.*, 1997).

Transposons often act as parasites of a genome. When mobilized, they can disrupt protein-coding genes, alter transcriptional regulatory networks, and cause chromosomal breakage and large-scale genomic rearrangements (McClintock, 1951). Cells therefore evolved mechanisms to protect their genomic integrity. RNAi is particularly effective towards those transposons that operate by copying their DNA to RNA, which is then reverse-transcribed back to DNA and inserted at another site in the genome. Part of this RNA molecule is often double-stranded, and can be easily targeted by the silencing machinery (Montgomery, 2004; Malone and Hannon, 2009).

To recap, RNAi is a complex process that regulates gene expression, exerting a silencing action; it can act at three different levels (transcriptional, post-transcriptional or translational); it is widely conserved among all eukaryotes, and it probably arose as a genome defense mechanism. Another fundamental aspect of RNAi is the central role played by dsRNA, which is chopped into tiny RNAs by the RNAi machinery.

Other steps forward in unraveling the silencing pathways were made in the early 2000s, when the endonuclease class responsible for processing these tiny RNAs (termed small RNAs, or sRNAs) from longer dsRNA transcripts was found during *in vitro* experiments on the fruit fly *Drosophila melanogaster* (Bernstein *et al.*, 2001). This class of ribonucleases was called Dicer (DCR), and soon thereafter homologous genes were also found in plants, where they were named Dicer-like (DCL) (Finnegan *et al.*, 2003). DCL-1 is considered the founding member of the plant-specific RNase III-like endonuclease family, which cleaves endogenous dsRNAs to produce sRNAs (Eamens *et al.*, 2008).

The last piece of the puzzle came shortly afterwards. A protein family named Argonaute (AGO), known for being involved in developmental processes in *Arabidopsis thaliana* (Bohmert *et al.*, 1998), was found to be the homolog of the *C. elegans* RNAi-deficient (RDE), identified in a mutant screening for proteins required for the RNAi pathway (Tabara *et al.*, 1999). The final evidence that AGO/RDE is crucial for sRNAs to exert their silencing action, selectively recruiting one strand of the sRNAs duplexes chopped by DCR/DCL to target complementary RNAs (Baumberger and Baulcombe, 2005), allowed the characterization of a primitive simple form of the RNA silencing puzzle.

1.2 RNAi effectors: core machinery

In the last 15 years our understanding of the RNAi endogenous mechanisms grew exponentially, and today an overall scheme of the events occurring in a cell during silencing is almost complete.

Invariably, dsRNA precursors are sliced by a member of the Dicer (or Dicer-like, in plants) gene family and unwound. Then, one selected strand is selected and enters the core effector complex named RISC (RNA-induced Silencing Complex), made up essentially by a member of the Argonaute protein superfamily, which can guide the silencing action. The RNA-dependent RNA polymerase (RDR) gene family is instead involved in a subset of silencing actions regarding the short-interfering RNA (siRNA) class.

A point of potential confusion is that an exact molecular composition for RISC has never been defined, and furthermore, the term RISC has been used to describe several biochemically distinct gene-silencing complexes. The minimal RISC, sufficient for target RNA recognition and cleavage, was demonstrated to be simply an Argonaute protein bound to a small RNA. However, Argonaute proteins can have dozens of associated binding partners, which do not assemble as one distinct complex *in vivo* (Pratt and MacRae, 2009).

Several complexes can be classified under the name RISC; the underlying feature common to all of them is that they contain a small regulatory RNA and they use it to guide an Argonaute protein.

1.2.1 Argonaute superfamily

Argonaute (AGO) proteins can be classified into three paralogous groups based on the phylogenetic analysis (Carthew and Sontheimer, 2009; Wei *et al.*, 2012). A first clade, named Piwi, binds the so called Piwi-associated small RNAs (piRNAs), which can be found only in animal germline cells. The Ago clade associates with the two main classes of small RNAs, that is, microRNAs (miRNAs) and short-interfering RNAs (siRNAs). Finally, there is a third subfamily that is nematode-specific, called WAGO.

Members of this family provide the small RNA binding component of silencing effector complexes (RISC). The small RNA duplex originated by the Dicer cleavage unwinds and one of the two stands gets caught in a stable association with the AGO effector protein. The selected strand (guide strand) directs target recognition, while the other one (passenger strand) is discarded and, usually, rapidly degraded.

AGO proteins are defined by the presence of four domains: N, PAZ, Mid and PIWI (Figure 3, left). The PAZ domain has RNA 3'-terminus binding activity, while the 5'-terminus gets caught by the Mid domain. The PIWI domain, instead, catalyzes the endonucleolytic cleavage of a base-paired target, at the post-transcriptional level.

Not all AGO proteins have endonucleolytic activity, and those that lack it usually also lack certain critical active sites.

There are 27 AGO proteins in *C. elegans*, 10 in *A. thaliana*, 8 in humans, 5 in *D. melanogaster* and 2 in *N. crassa*. The proliferation of these gene families in multicellular eukaryotes indicates a subfunctionalization of the encoded proteins into diverse processes, and provides the possibility of separate nuclear or cytoplasmic distributions in the cell (Matzke and Birchler, 2005; Carthew and Sontheimer, 2009).

Regarding *A. thaliana*, the ten different AGO proteins show preferences for distinct classes of small RNAs. AGO1 receives 18-21 nt long sRNAs excised by DCL1. AGO4 and AGO6 contain a specific class of small RNAs termed heterochromatic siRNAs, which are involved in small RNA-directed DNA methylation (RdDM) processes. AGO2 is mainly loaded with trans-acting siRNAs (ta-siRNAs) and repeat associated siRNAs (ra-siRNAs). AGO7 was suggested to function in ta-siRNA generation. AGO5 contains various siRNAs that originate from intergenic regions.

The small RNA content of the remaining *A. thaliana* AGO proteins has not yet been analyzed in detail, mainly owing to the highly specific expression patterns in germ cells. Interestingly, AGO9 is expressed in somatic companion cells of female gametes and preferentially interacts with siRNAs derived from transposable elements. It is required for transposable element silencing in the gametes, most probably owing to transport processes between the two cell types. AGO10 seems to sequester miR-166 and miR-165 specifically, regulating shoot apical meristem development (Meister, 2013).

1.2.2 Dicer family

Proteins of the Dicer family are ribonuclease III enzymes that process dsRNA into small RNAs. They are usually characterized by several domains in a specific order from the N- to C- terminus (even though some member of this family differs slightly from this arrangement): a DEXD/H ATPase domain, a DUF283 domain, a PAZ domain, two tandem RNase III domains, and a dsRNA-binding domain (dsRBD) (Figure 3, right).

The PAZ domain, which is shared with Argonaute proteins, is specialized to recognize duplexes with a short 3' overhang. It binds RNA ends, and, along with the RNase III domains, it plays an essential role in excising sRNAs from longer precursors. Once inserted, the duplex extends approximately two helical turns along the surface of the protein and lies in the space between the two RNase III domains, each of which cleaves one of the two strands. A new smaller duplex is therefore created, each strand being exactly the same length – that is, with both a 3' and a 5' 2-nucleotide-overhang.

As a general rule, organisms with multiple Dicers exhibit a functional specialization of them. Different Dicers can give rise to small RNAs of slightly different lengths. Not surprisingly, this appears to be dictated by the distance between the PAZ domain and the two tandem RNase III domains (Carthew and Sontheimer, 2009).

There are four such proteins in *A. thaliana*, which are called Dicer-like (DCL) and are specialized in the production of different small RNAs. In general, DCL1 and DCL4 produce 21-nt small RNAs, whereas DCL2 produces 22-nt small RNAs, and DCL3 produces 24-nt small RNAs (Vaucheret, 2008).

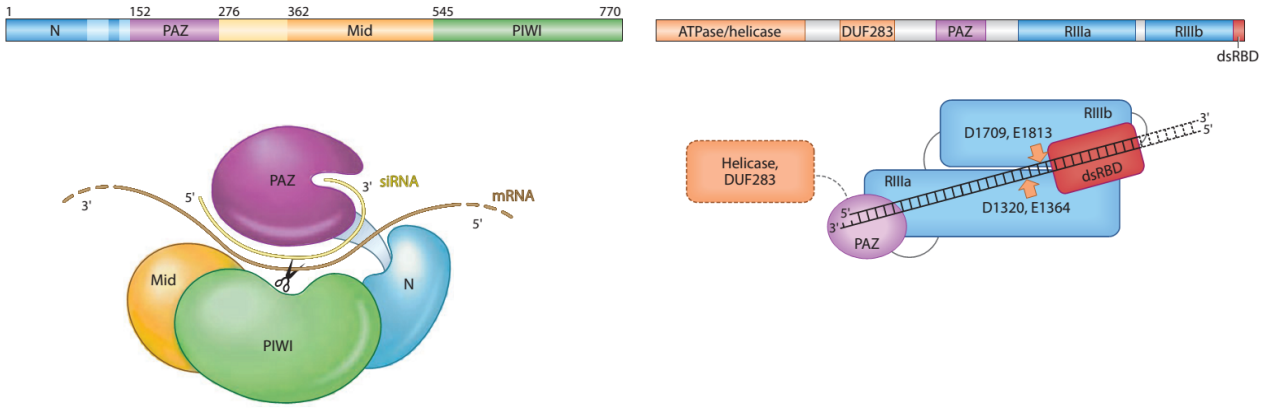


Fig. 3 ***(left)*** *Domain structure of an AGO protein with a schematic representation of mRNA cleavage. The 3' end of the sRNA is positioned in the cleft of the PAZ domain. The mRNA situates between the upper PAZ domain and the lower crescent-shaped base formed by the N-terminal, PIWI, and middle (Mid) domains. The catalytic site (scissors) slices mRNA at a position that corresponds to the ninth and tenth nucleotides of the guide sRNA.*
(right) *Domain structure of a Dicer protein with a single processing center model of Dicer catalysis. The PAZ domain recognizes the terminus of the dsRNA substrate. The two RNase III domains approximate to form one catalytic center and cleave the opposing strands of dsRNA in an offset manner, resulting in a characteristic 2-nt 3' overhang.*
From Liu and Paroo, 2010.

1.2.3 RDR family

RNA-dependent RNA polymerase (RDR) is the core class of enzymes that mediates silencing amplification. In *A. thaliana* there are six known RDR genes. Three of them (RDR1, RDR2 and RDR6) have been biologically characterized.

RDR1 is associated primarily with antiviral defense mechanisms, while RDR2 plays a role in DNA methylation, and RDR6 functions in both endogenous *trans*-acting siRNA biogenesis as well as antiviral silencing (Lowery *et al.*, 2011).

Once detected, the unwanted RNA is copied into double stranded forms and becomes a substrate for the ribonuclease Dicer, which converts it into small interfering RNA (siRNA) duplexes. The end-product is a multitude of highly reactive small RNA species, which, upon their incorporation into a RISC, execute effective silencing of the unwanted RNA species, or of any potential siblings (Voinnet, 2008).

In plants and nematodes this secondary pool of siRNAs created by these enzymes can lead to systemic silencing that spreads throughout the organism. It has to be noted that recognizable RDR genes are present in the genomes of many RNAi-competent eukaryotes, with the notable exceptions of insect and vertebrate species (Carthew and Sontheimer, 2009).

1.3 RNAi effectors: small silencing RNA classes

The cleavage products of the Dicer enzymes were first called “small-temporal RNAs” by their discoverers (Lee *et al.*, 1993), who were not fully aware of the importance of such a finding at that time, a real breakthrough in gene regulation research.

Some years after, a stable association of small RNAs (sRNAs) with all kinds of gene silencing was demonstrated by Hamilton and Baulcombe (1999), who screened for sRNA species in four different silencing backgrounds. They identified sRNAs approximately 25 nucleotides in length and specific to the nucleic acid sequence undergoing silencing in all four experiments. Interestingly, the authors also showed a correlation between the level of sRNAs accumulated and the silencing efficiency conferred by each of these systems (Eamens *et al.*, 2008).

This second major breakthrough marked the division between two eras, with the rising of a new concern around these small, non-protein-coding, widespread regulatory elements of eukaryotic genomes.

sRNAs can be classified into many categories but a universal, definitive classification still lacks nowadays. New small RNA classes and new examples of existing classes continue to be discovered. Different small RNA classes guide diverse and complex schemes of gene regulation, differ in the proteins required for their biogenesis, the constitution of the Argonaute-containing complexes that execute their regulatory functions, their modes of gene regulation and the biological functions in which they participate. Moreover, some small silencing RNAs derive from dsRNA, whereas others, such as Piwi-interacting RNAs (piRNAs), do not (Ghildiyal and Zamore, 2009).

There is a general agreement in considering three main categories of small silencing RNAs: microRNAs (miRNAs), short-interfering RNAs (siRNAs) and Piwi-interacting RNAs (piRNAs) (Carthew and Sontheimer, 2009; Chen, 2009).

Plant microRNAs are the focus of this thesis, and they will be described more in-depth in the sub-chapter 1.4.

1.3.1 miRNAs

[reviewed in Ghildiyal and Zamore, 2009; Almeida *et al.*, 2011; Axtell *et al.*, 2011]

lin-4 was the first miRNA to be discovered, in 1993, in a screen for genes that are required for postembryonic development in *C. elegans* (Lee *et al.*, 1993). Seven years after, the second one, *let-7*, was discovered. Likewise *lin-4*, it is a heterochronic gene controlling larval development in nematodes (Reinhart *et al.*, 2000).

In 2001, tens of miRNAs were identified in humans, flies and worms by small RNA cloning and sequencing, thereby establishing miRNAs as a new class of small silencing RNAs (Ghildiyal and Zamore, 2009). Today miRBase (release 20 – June 2013), the registry that coordinates miRNA naming, lists 24521 distinct miRNA genes in algae, plants, animals, protists and viruses.

miRNAs arise from primary transcripts with self complementarity called primary miRNAs (pri-miRNAs) that are typically transcribed by RNA polymerase II (RNA Pol II) and processed in the nucleus into a 60–70-nucleotide precursor miRNA (pre-miRNA) by the enzyme Drosha (an RNase III), acting with its dsRBD partner — DGCR8 in mammals and Pasha in invertebrates.

The resulting pre-miRNA has a hairpin structure: a loop flanked by base-paired arms that form a stem. Pre-miRNAs have a two-nucleotide overhang at their 3’ ends and a 5’ phosphate group, which are indicative of their production by an RNase III.

The nuclear export protein Exportin 5 carries the pre-miRNA to the cytoplasm where Dicer and its dsRBD partner protein (TRBP in mammals and Loqs in flies) cleave it, thus generating a duplex containing two strands, termed miRNA (guide strand) and miRNA* (or passenger strands), corresponding to the two sides of the base of the stem. The duplex is then unwound by a helicase and the mature miRNA, most of the time 21 to 24 nucleotides in length, is incorporated into the RISC. Usually only one strand of the duplex is selected for RISC loading, based on thermodynamic criteria, such as 5’ end stability. In some cases they can both be eligible for selection.

Many aspects of biogenesis and mode of action mark the distinction between plant and animal microRNAs (Figure 4). Firstly, plants lack the Drosha-like enzyme of metazoans, and DCL1 is the only enzyme responsible for both the pri- to pre-miRNA conversion and mature miRNA processing. Secondly, the whole process is completed in the nucleus. Thirdly, the 3’-most nucleotides of the initial miRNA/miRNA* duplex are 2’-O-methylated by the nuclear HEN1 protein, which is lacking in animals.

Besides, the genomic patterns of animal miRNA genes are significantly different from those of plants, where most miRNA-encoding loci comprise independent, non-protein-coding transcription units, that is, very often multiple hairpins are expressed from a single pri-miRNA. In animals approximately 30% of miRNA genes are located on the sense strands of introns although many derive from stand-alone non-protein-coding loci. Clustering of miRNA genes is more common in animals than in plants. Unlike in plants, there are also occasional examples of miRNA biogenesis from exons of animal protein-coding genes, including UTRs and coding sequences.

It has long been known that plant miRNAs often have targets with perfect or near-perfect complementarity. Canonical plant miRNA target sites are found in 5’ UTRs, ORFs and 3’ UTRs, as well as

within non-protein-coding transcripts, suggesting that all RNA contexts are equally amenable to miRNA-directed regulation in plants. By contrast, it is rare for animal miRNAs to identify targets with 'plant-like' complementarity, being translational inhibition the most frequent mode of action (Axtell *et al.*, 2011).

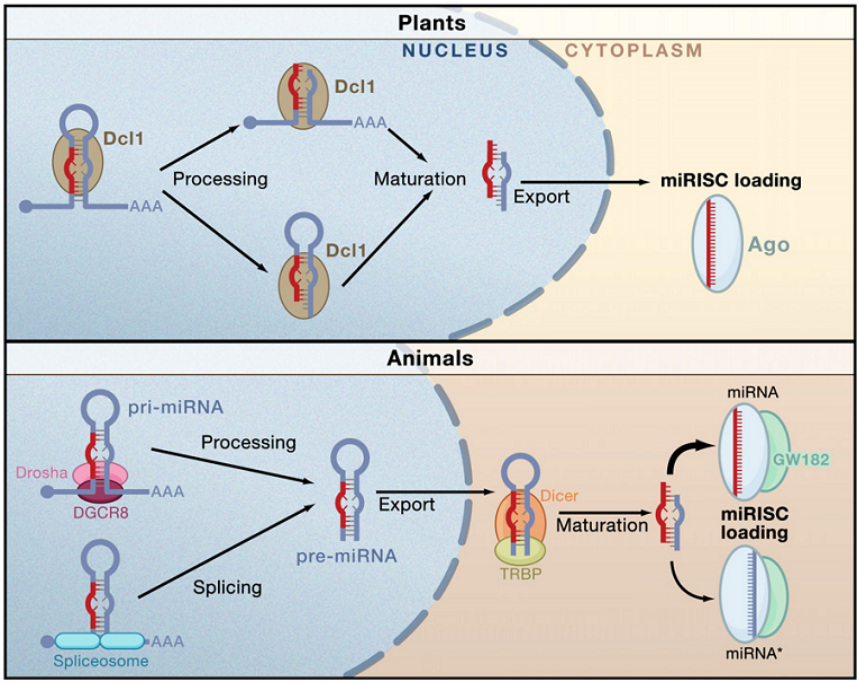


Fig. 4 ***(Top)*** *In plants, DCL1 processes the pri-miRNA in succession. The terminal loop and the flanking segments are excised to give birth to a pre-miRNA. The second processing step by DCL1 yields a mature miRNA/miRNA* duplex that becomes methylated and exported in the cytoplasm, where it is loaded on a RISC.* ***(Bottom)*** *In animals, the pri-miRNA is processed by Drosha with the aid of DGCR8 to generate a pre-miRNA species. This is exported from the nucleus and processed by Dicer to form the mature miRNA/miRNA* duplex. After processing, miRNAs are assembled into a RISC.*

Usually only one strand is selected, and the other is degraded, although exceptions to this general rule are known.

From Carthew and Sontheimer, 2009.

Regarding the differences with siRNAs, they can be distinguished in two primary ways. First, miRNAs are always endogenous and purposefully expressed products of an organism's own genome, whereas siRNAs can be exogenous in origin, derived directly from the virus, transposon, or transgene trigger. Second, miRNAs are processed from stem-loop precursors with incomplete double-stranded character, whereas siRNAs are excised from long, fully complementary double-stranded RNAs (dsRNAs) (Carthew and Sontheimer, 2009).

1.3.2 siRNAs

[reviewed in Ghildyial and Zamore, 2009; Axtell, 2013]

siRNAs are the most abundant class of plant small RNAs (Voinnet, 2009) and can be classified according to their origin, the proteins involved in their biogenesis, their mode of regulation or their size.

Their origin can be considered as a first layer of classification.

Exogenous siRNAs (exo-siRNAs) are usually the products of long dsRNA derived from viruses, transposons, or transgenes that are cleaved into double stranded siRNAs, approximately 21 nucleotides in length, by a Dicer enzyme. Similar to miRNAs, one strand (guide strand) exerts the silencing action, whereas the other (passenger strand) is ultimately degraded.

In plants, inverted-repeat transgenes or coexpressed sense and antisense transcripts produce three sizes of siRNAs: 21, 22 and 24 nucleotides. The 21-nucleotide siRNAs are produced by DCL4, but in the absence of DCL4, DCL2 can substitute, producing 22-nucleotide siRNAs. These typically associate with AGO1 and guide mRNA cleavage. The 24-mers are produced by DCL-3 and associate with AGO4 or AGO6 and promote DNA methylation, to silence the transgene that produce them.

In plants, exogenous sources of siRNAs are not confined to dsRNAs. Single-stranded sense transcripts from tandemly repeated or highly expressed single-copy transgenes are converted to dsRNA by RDR6, a member of the RNA-dependent RNA polymerase (RdRP) family that transcribe ssRNAs from an RNA template. RDR6 and RDR1 also convert viral ssRNA into dsRNA, initiating an antiviral RNAi response.

In a recent review paper by Axtell (2013), currently known **endogenous siRNAs** (endo-siRNAs, derived from endogenous loci) are classified into heterochromatic siRNAs, secondary siRNAs, and natural antisense transcript siRNAs (nat-siRNAs). A further subdivision regards secondary siRNAs, which can be "phased" or "trans-acting" (ta-siRNAs), and nat-siRNAs, which can act in *cis* (*cis*-nat-siRNAs) or in trans (*trans*-nat-siRNAs).

- **Heterochromatic siRNAs** are derived from intergenic and/or repetitive genomic regions and are associated with chromatin modifications leading to silencing of target DNA loci. They are usually 23-24 nt in length, and require RDR2 and DCL3 for their biogenesis and AGO4 (or AGO6 and AGO9 in *Arabidopsis*) for their functions. Their appearance in the genome likely depends on the response to the rapid changes in transposon position and copy number that occur during plant evolution.
- **Secondary siRNAs** derive from a dsRNA precursor whose production is due to the intervention of a small RNA trigger and an RDR activity (Figure 5). They are 21 nucleotides long and their formation is mediated by DCL4 and RDR6. A further distinction is possible (even if blurry, since many of the known trans-acting siRNAs are also phased):

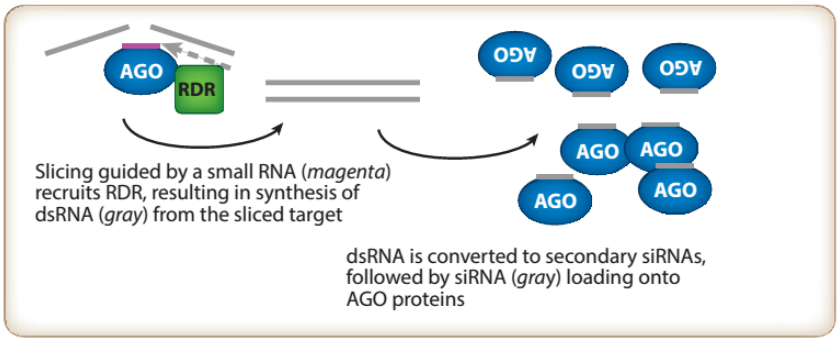


Fig. 5 *Detailed schematic of secondary siRNA production.* **From Axtell, 2013.**

  - Many secondary siRNAs are **phased**, meaning that they derive from successive DCL-catalyzed processing from a consistent dsRNA terminus. The consistency of the dsRNA terminus is defined by a discrete initial small-RNA-directed cleavage event on the primary transcript.
  - Some secondary siRNAs are also capable of acting in *trans* (***trans*-acting siRNAs**) to direct repression of distinct mRNA targets. The most extensively studied case is the miR390-triggered TAS3 family of secondary siRNA loci, which produce two nearly identical ta-siRNAs that target Auxin Response Factor 3 (ARF3) and ARF4 (Allen *et al.*, 2005).
- **Nat-siRNAs** originate when two complementary and independently transcribed RNAs hybridize, thus forming a dsRNA precursor. They are produced in response to stresses in plants: typically, one transcript is expressed constitutively, whereas the complementary RNA is transcribed only when the plant is subject to environmental stress. Their production requires DCL2 and/or DCL1, RDR6 or RDR2. The nat-siRNAs then direct cleavage of one of the mRNAs of the pair.
  - In the case of ***cis*-nat-siRNAs** the separate RNAs complementarity is due to the fact that they are transcribed from opposite strands of the same locus. Three cases of *cis*-nat-siRNAs have been reported.

- Alternatively, the hybridizing RNAs can arise from genes that possess no overlap; these are the ***trans*-nat-siRNAs**. Only *cis*-nat-siRNAs have been described in plants; *trans*-nat-siRNAs remain only a hypothetical possibility.

1.3.3 piRNAs

piRNAs are the most recently discovered class of small RNAs and, as their name suggests, they bind to the Piwi clade of AGO proteins, therefore they are specific to the animal kingdom. They are produced through successive cleavage of long single stranded RNA precursors through the action of PIWI members (a mechanism referred to as 'ping-pong') (Brenneke *et al.*, 2007, reviewed in Ghildiyal and Zamore, 2009), and, unlike miRNAs and siRNAs, they do not require a Dicer enzyme for their production.

They protect germline cells from transposons in organisms as diverse as flies, fish or mammals and are 25 to 33 nt in length, depending on which PIWI protein they bind to.

Some other subclasses of piRNA-like small RNAs were found in nematodes and in the fruit fly. 21-nt-long piRNAs were found in *C. elegans*, while somatic piRNA-like siRNAs (24-30 nt long) were identified in *D. melanogaster* (Ghildiyal and Zamore, 2009).

1.3.4 Other small silencing RNAs

Besides these three main categories, there is a plethora of other small silencing RNAs with less ubiquitous expression but not less important.

'Long' siRNAs (**lsiRNAs**) in *Arabidopsis* species originate from natural antisense transcript pairs and are stress induced. Unlike nat-siRNAs, lsiRNAs are 30–40 nucleotides and require DCL1, DCL4, AGO7 and RDR6 for their production (Ghildiyal and Zamore, 2009).

scanRNAs are found in ciliated protozoans in the form of 25-30 nt species and are processed from dsRNA produced during meiosis. They are responsible for the extensive, programmed excision of excess DNA in the genome reorganization that occurs during the sexual process of conjugation.

In a recent study, Plante *et al.* (2012) propose a model whereby RNA species half the size of microRNAs (approximately 12 nucleotides), tentatively termed semi-microRNAs (or **smiRNAs**), may be generated along the microRNA pathway and participate to the control of mRNA translation by regulating microRNA activity *in vivo*. In other words, according to the authors, smiRNAs can cleave miRNAs, thus silencing the silencers. They have only been found in human.

Another important discovery regarding small-RNA mediated gene regulation was made in the mid-2000s, by two research groups working on synthetic antigene RNAs (agRNAs). They noted that dsRNA can also activate gene expression at transcriptional level, a mechanism termed RNA activation, or RNAa, as a contrast to RNAi. Small-activating RNAs (**saRNAs**) can target promoter regions in chromosomal DNA resulting in transcriptional activation of genes. It has been shown that RNAa exists in several mammalian species other than human, suggesting that it is a mechanism conserved in mammals (Angaji and Darvishani, 2012).

1.4 Plant miRNAs

1.4.1 Biogenesis

miRNAs are a well-studied subset of hairpin-derived RNAs defined by the highly precise excision of one or sometimes a few functional products, termed mature miRNAs. Most plant miRNAs are 21 nt long, require a DCL1-clade DCL for their biogenesis, and an AGO1-clade AGO for function, although exceptions have been described for each of these general trends (Axtell, 2013, and references therein).

The majority of miRNA genes are intergenic and transcribed individually from their own region, but a few genes are organized into polycistronic transcription units and cotranscribed from a single promoter at the end of a miRNA gene cluster.

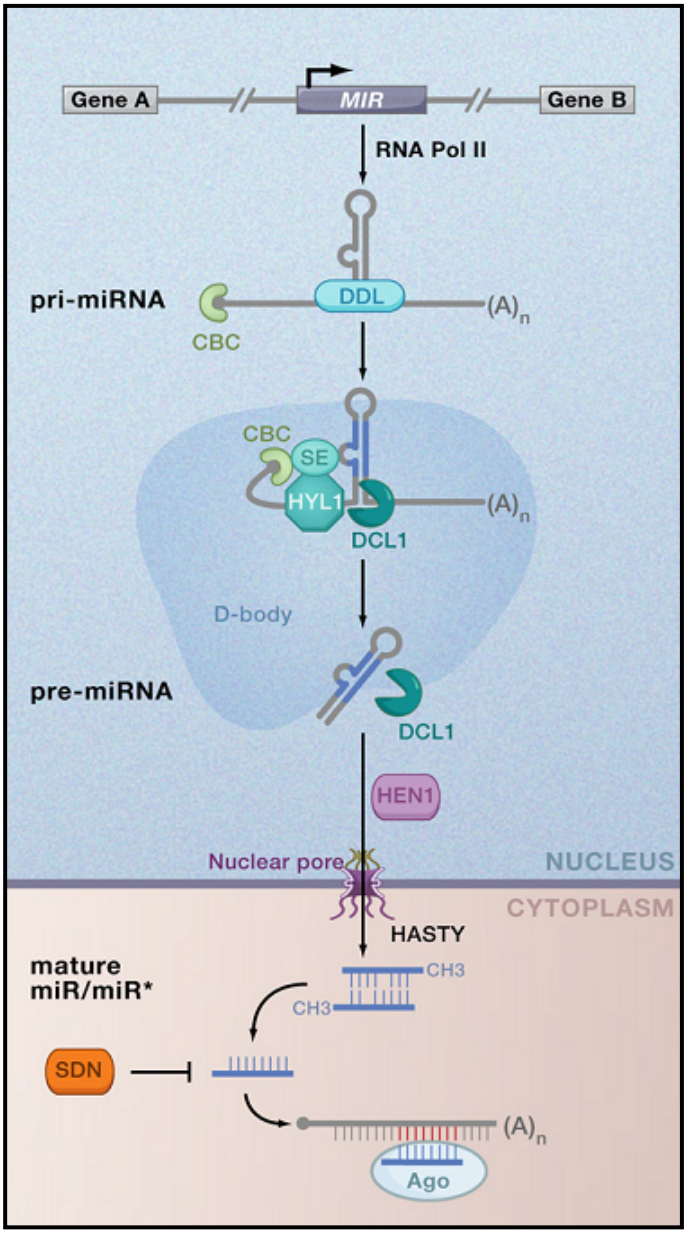


Fig. 6 *Molecular pathway for the processing and stability of conserved plant microRNAs.*
From Voinnet, 2009

Plant miRNA genes are transcribed into primary miRNA transcripts (**pri-miRNA**) by RNA polymerase II. The length of pri-miRNA molecules can range from 70 to a few hundreds of bases. They typically contain a region of imperfect self-complementarity and before being processed into stem-loop secondary structures, pri-miRNAs are stabilized by the nuclear protein Dawdle (**DDL**). **DCL1** then cleaves these long transcripts to hairpin-like miRNA precursors (**pre-miRNAs**) with the aid of double-stranded RNA binding protein Hyponastic leaves1 (**HYL1**) and a C2H2 zinc finger protein Serrate (**SE**). The SE and HYL1 proteins interact with DCL1 in special nuclear regions termed "dicing bodies" (**D-bodies**) which serve as nuclear processing centers.

DCL1 also mediates the excision of miRNA:miRNA* duplexes, containing a 2-nt overhang at the 3´-end, which are methylated by Hua Enhancer1 (**HEN1**) that adds methyl groups to 2´-OH of the 3´-terminal nucleotide of each strand. Methylation is crucial since it prevents miRNA uridylation and degradation from exonuclease activity of the small RNA degrading nuclease (SDN) family.

The methylated miRNA/miRNA* duplex is then exported from the nucleus into the cytoplasm by the action of the **HASTY** protein, an *Arabidopsis* homolog of the animal Exportin-5, which seems to cooperate with RanGTP. Once in the cytoplasm, one strand of the duplex is selectively

incorporated in the RNA-induced Silencing Complex (**RISC**), while the other strand is usually degraded (Eldem *et al.*, 2013).

The strand that is loaded into the RISC complex is selected based on several factors. The first is the preferential association of AGO1 with miRNA having 5´ terminal uridine residues. The thermodynamic stability of the 5'-end is another preference factor: as a general rule, the lowest, the higher the likelihood to be selected.

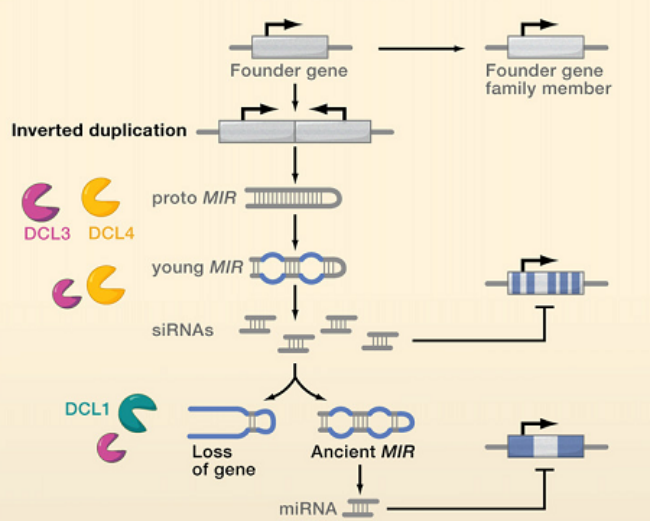
Eventually, functional mature miRNA guides the RISC to the exertion of the silencing action. Due to its rapid degradation, miRNA* sequences abundance is generally low compared with that of their corresponding miRNA sequences. Recent studies showed that plant miRNAs* can accumulate at high levels under certain conditions, and can also intervene in the same pathway as their counterpart (Sunkar *et al.*, 2012).

1.4.2 Evolution

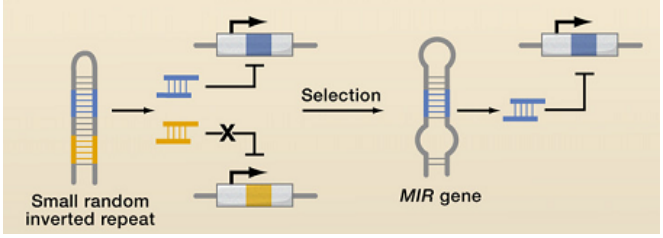
[reviewed in Voinnet, 2009; Cuperus *et al.*, 2011; Axtell, 2013]

Several miRNA families are conserved between multiple plant species, sometimes even across the huge evolutionary distances between mosses and flowering plants. At the same time, the majority of miRNAs present in any given plant species appear to be unique to that species, and many more are present only within a few closely related species, implying that there is a great deal of birth and death among miRNA genes (Axtell, 2013).

There are several possible ways in which a new miRNA gene can arise.

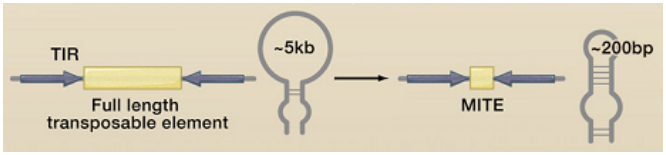


Inverted duplication events can form self-complementary regions with the potential to spawn miRNA genes. Initial duplication events would result in loci, termed proto-MIR, with perfect or near-perfect self-complementarity and produce siRNAs with the aid of DCL3 and DCL4. Over time, the accumulation of mutations in the stem-loop region could result in an affinity for the miRNA biogenesis machinery and decreased similarity with the locus of origin (young MIR, prevalently diced by DCL4). The concerted evolution of one of this new sRNAs and one or more target transcripts can cause the birth of a productive miRNA target node and, in rare instances, it can be incorporated into new or existing regulatory networks (ancient MIR, now preferentially diced by DCL1). Young MIRs that lack a co-evolution with a target, therefore a function, encounter genetic drift, and are eventually lost.



Another possibility is the **spontaneous evolution** of a MIR gene from small, random inverted repeats spread along plant genomes. siRNAs generated from such repeats might match complementary regions found in potential target genes. Selection of the targeting event under appropriate circumstances would then contribute to the isolation of a single miRNA molecule within the

stem-loop structure. Selection could conceivably act on mutations that modify miRNA processing efficiency and alter the affinity of the *de novo*–generated miRNA for a target if there were effects on existing regulatory networks.



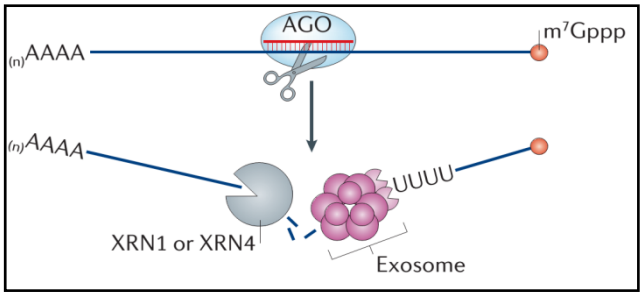
A third way of evolution of a young MIR gene is through **generation of miniature inverted-repeat transposable elements** (MITEs). A full-length DNA-type transposable element (TE) with terminal inverted repeats (TIRs) flanking a long open reading frame can form a secondary structure owing to the annealing of the TIRs. This stem-loop structure is too big to be processed by the sRNA machinery, but TE may give rise to MITEs, non autonomous derivatives of full-length DNA-type elements, and the small imperfect hairpin resulting from them could resemble a pre-miRNA precursor, entering the young miRNA biogenesis pathway.

1.4.3 Modes of action

[reviewed in Voinnet, 2009; Axtell, 2013; Ameres and Zamore, 2013]

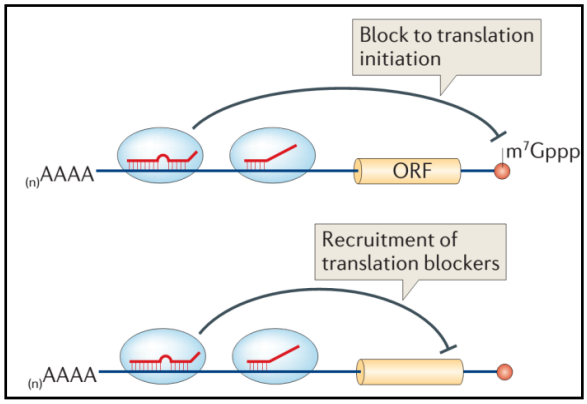
MiRNAs can mediate gene silencing at either transcriptional, post-transcriptional or translational level.

Post-Transcriptional Gene Silencing (PTGS)



Most plant miRNAs possess perfect or near-perfect complementarity to their target mRNAs, and **mRNA cleavage** is believed to be a predominant mechanism by which miRNAs regulate gene expression in plants. This mechanism provides the microRNA-loaded RISC binding to the target mRNA through Watson-Crick base-pairing. The AGO proteins involved cleave target RNAs in a process catalyzed by their RNase H-like PIWI domain. Cleavage occurs at the phosphodiester bonds mainly between nucleotides 10 and 11 of the miRNA, relatively to its 5'-end. The 3'-end of the 5'-cleavage product is frequently uridylated, and hence degraded. The 5'-to-3' exoribonuclease XRN4 subsequently carries out the degradation process.

Translational inhibition



One potential alternative paradigm comes from animal miRNA/target pairs, where miRNAs recognize their targets solely through their seed sequence (between nucleotides 2–7, or 2–9, of the miRNA). This partial complementarity does not allow target cleavage, but the stable association of the RISC with the target mRNA can prevent its translation or cause its decay. This mechanism is termed '**translational inhibition**', and it requires the recruitment of other factors in the RISC besides an AGO protein. These include the microtubule-severing enzyme KATANIN (KTN) and the P-body component VARICOSE (VCS) in *Arabidopsis* (Figure 7, left).

Transcriptional Gene Silencing (TGS)

Some 24-nt long microRNAs can enter into the heterochromatic siRNA effector pathway and direct chromatin modifications at their target gene loci. This mechanism relies on the action of a RISC containing AGO4 or AGO6. Interestingly, most of the long miRNAs described to date in *Arabidopsis* and rice are also lineage-specific miRNAs, *i.e.* they are not conserved. This suggests the possibility that newly evolving "proto-MIRs" are passing through a phase where they produce 24-nt species.

Excluding particular cases, like long miRNAs exerting transcriptional gene silencing, the degree of miRNA-mRNA complementarity is considered a key determinant of the regulatory mechanism. Perfect complementarity allows AGO-catalyzed cleavage of the mRNA strand, whereas central mismatches exclude cleavage and promote repression of mRNA translation. All experimentally verified plant miRNA targets described to date have extensive complementarity to their cognate miRNAs (perfect complementarity is rare as well as examples with more than five mismatches between miRNA and target). Therefore, it can be predictable that plant microRNAs regulate target genes mainly via the PTGS mechanism.

It seems likely, however, that most plant miRNAs repress their targets via some combination of RNA destabilization (either slicing dependent or not) and translational repression at the same time, being slicing the key mechanism anyway (Voinnet, 2009; Axtell, 2013) (Figure 7, center).

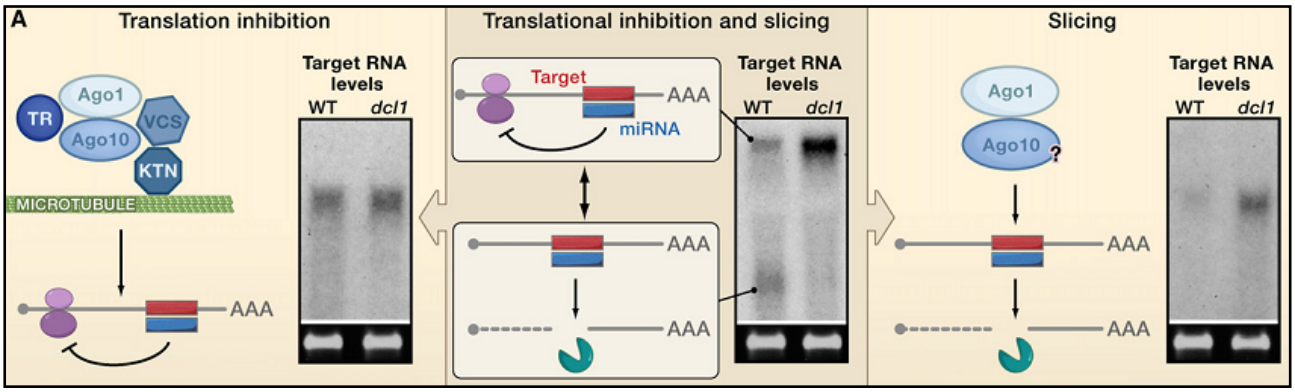


Fig. 7 ***(Center)*** *Dual modes of miRNA action entail a combination of slicing and translational repression. Shown is the typical molecular output of regulations affecting most plant miRNAs targets, as detected by a northern blot. The right lane on the blot contains RNA extracted from the Arabidopsis miRNA-deficient mutant dcl1, used as a negative control. The left lane contains RNA from wild-type (WT) plants. The accumulation of a fast-migrating 3' cleavage mRNA fragment diagnoses the occurrence of slicing, but high levels of full-length mRNAs remain and may be subject to translational repression.*
(Right) *One possible extreme case where the miRNA entirely regulates its target at the mRNA level.*
(Left) *One possible extreme case where the miRNA entirely regulates its target at the protein level, pending the presence of hypothetical translational repressors (TR). Translation inhibition also specifically requires the microtubule (MT) severing activity of KATANIN (KTN) and the action of VARICOSE (VCS).*
From Voinnet, 2009.

Besides these kinds of regulation, in which the induction of a microRNA implies a negative modulation of the related targets, there are at least two other patterns of miRNA/target complementarity that are functional in plants. A microRNA can lack a target, or at least identify its target based on currently unknown criteria. This can be the case of the so called 'lineage-specific' microRNAs (Axtell, 2013).

Secondly, there is a subset of functional sites that have an alternative arrangement, with very high complementarity at both the 5' and 3' regions, with central mismatches and/or bulged nucleotides. This conformation prevents AGO slicing, and is therefore able to sequester active microRNAs in the cytoplasm. The mechanism at issue is called 'target mimicry', it has been described for the IPS1/At4 family of non

protein-coding RNAs, targeted by miR399 (Franco-Zorrilla *et al.*, 2007), and, so far, it is the only documented example of natural miRNA target mimics in plants. Whether this is a more widespread mechanism in plants awaits further research (Axtell, 2013).

1.4.4 Role in development

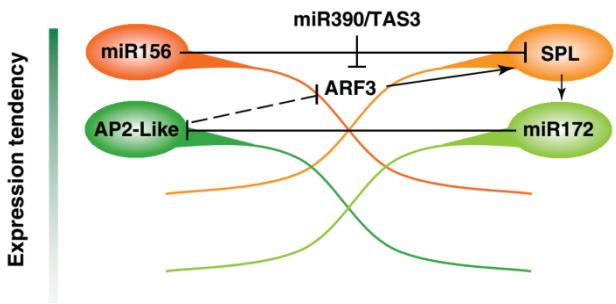
[reviewed in Chen, 2009; Kidner, 2010; Rubio-Somoza and Weigel, 2011; Eldem *et al.*, 2013]

The first evidence that small RNAs play roles in plant development came from mutants impaired in small RNA biogenesis or function. Several genes central to miRNA function, including DCL1, AGO1, HEN1, SE, DDL and HYL1, have been identified in plants based on the developmental consequences of their mutations even before they were known to be important for small RNA biogenesis or function (Jones-Rhoades *et al.*, 2006; Chen, 2009; reviewed in Ramachandran and Chen, 2008).

Several developmental programs involving miRNAs were detected in plants. They comprise root initiation and development, vascular development, leaf morphogenesis and polarity, floral differentiation and phase transition from vegetative growth to reproductive growth.

1.4.4.1 Phase transition

In *Arabidopsis thaliana* and *Zea mays* developmental transitions are coordinated by the antagonistic activities of **miR156** and **miR172**. miR156 expression levels decrease with leaf age, while miR172 increases.



Their targets, Squamosa Promoter Binding Protein-Like (SPL) and Apetala 2 (AP2) transcription factors, respectively, are expressed in complementary patterns as well.

Increased levels of miR156, along with reduced miR172 activity, restrain developmental transitions, prolong juvenile features and delay flowering. In turn, decreased levels of miR156 lead to premature acquisition of adult leaf features and early flowering, resembling effects observed in plants with reduced activity of AP2-like miR172 targets. An increase in miR172 reduces activity of potent floral repressors (SMZ, SNZ, TOE1 and TOE2), therefore inducing flowering. This sequence of genetic events places miR172 action downstream of miR156. **miR390** is another actor of this network, triggering the production of ta-siRNAs from *TAS3* transcripts. These sRNAs cause the destruction of the AUXIN RESPONSE FACTOR 3 and 4 (ARF3/4). Rubio-Somoza and Weigel (2011) speculate that ARF3 might promote expression of SPLs, and are repressed by AP2-like proteins. They put forward a model in which miR172 targets AP2 and diminish SPL levels through down-regulation of its activator ARF3.

miR156 and SPLs regulates an additional aspect of developmental timing, plastochron length, which is the time interval that elapses between the initiation of two successive leaves.

1.4.4.2 Leaf development

The **miR390**-TAS3-ARF3/4 node is also involved in the specification of abaxial/adaxial polarity during leaf development (Kidner, 2010). In this process, this node is helped by another class of transcription factors, HD-ZIP III, namely PHABULOSA (PHB), PHAVOLUTA (PHV) and REVOLUTA (REV), which are regulated by **miR165/miR166**. The miR165/miR166 action allows the containment of these proteins to the adaxial leaf side; disruption of this regulation would lead to loss of abaxial identity.

PHB and REV also play an important role in leaf vascular development, and defects are seen when they escape miR165/miR166 regulation. Other defects are found when **miR159** is not functional. Its negative regulation on MYB TFs (specifically MYB33 and MYB65) is needed for a proper development of leaf vasculature.

CUP-SHAPED COTYLEDON1 (CUC1) and CUC2 are fundamental in cell differentiation during leaf development. Plants having their action compromised show developmental abnormalities like fused cotyledons, unseparated stamens (Figure 8 – left panel) or anomalous number of petals and sepals (Figure 8 – right panel). CUC1 and CUC2 are specifically regulated by **miR164**, with another layer of complexity provided by the **miR319**-TCP transcription factors node, in the fact that TCPs can bind to miR164 regulatory sequences, modulating its expression levels. These two microRNAs act in concert in a regulatory loop (below depicted), in which the lack of activity of one actor eventually results in the appearance of aberrant phenotypes (Figure 8).

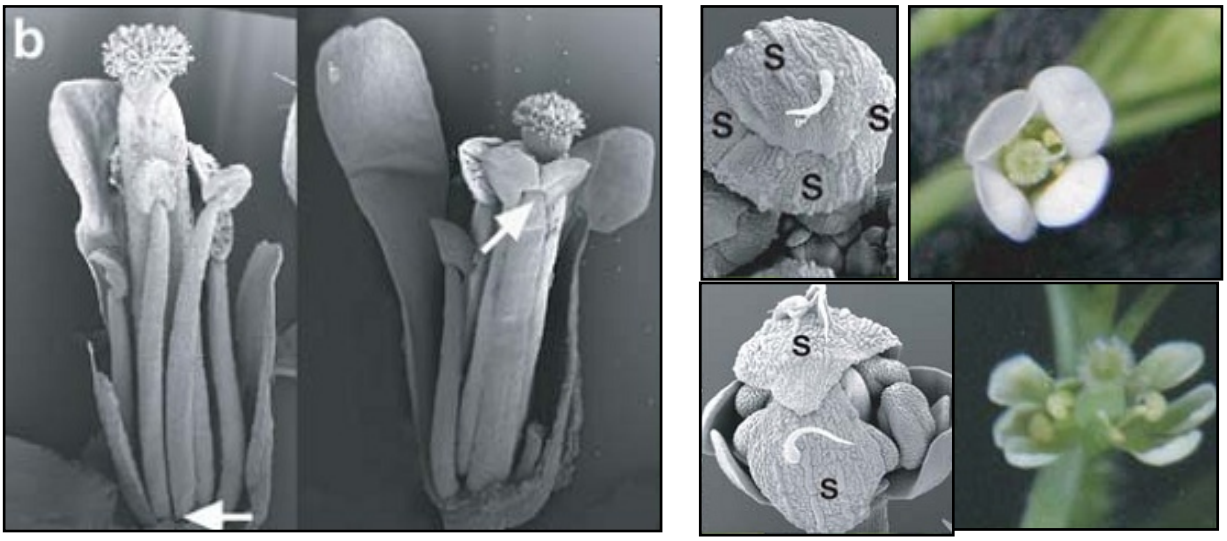
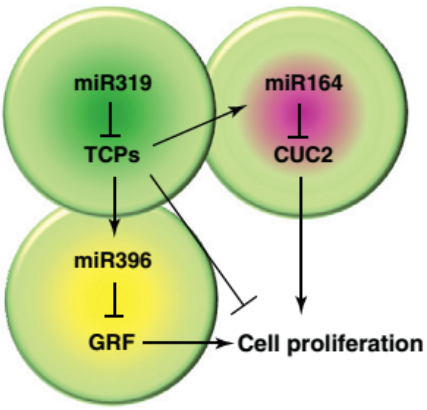


Fig. 8 (left panel) Impaired CUCs action leads to the stamens fused together. Left, wild type; right, miR164 overexpressing plants. Arrows denote point of stamen separation. (right panel) Wild-type flowers have the expected four sepals (S, top left) and four petals (top right), whereas flowers with impaired miR164 action often display two sepals and six petals (bottom). **From Jones-Rhoades *et al.*, 2006.**

In addition, CUC TFs function in the growth of the leaf front, a process in which TCPs have an important role as well. Deregulation of this network leads to crinkled and more serrated leaves. This ability of the miR319-TCP nodes to regulate cell proliferation is partially mediated in *Arabidopsis* by the conserved **miR396**-GRF (Growth-Regulating Factor) node. Compromised miR319-TCP4 regulation leads to reduced GRF expression and a subsequent reduction in mitotic activity (Rubio-Somoza and Weigel, 2011).



miR396 antagonizes the expression of GRF2 in the distal part of the leaf, thus attenuating cell proliferation (Eldem *et al.*, 2013). However, how these three nodes interact is still not clear. It would be interesting to determine at what level the function of the miR164, miR319 and miR396 nodes converges on the regulation of the cell cycle (Rubio-Somoza and Weigel, 2011).

miR824 is a newly evolved miRNA found only in *Brassicaceae*, where it regulates pattern formation during stomatal development by targeting the MADS-box gene *AGL16* (Chen, 2009).

1.4.4.3 Reproductive development

Several miRNA genes control various aspects of reproductive development, such as inflorescence architecture, flower organogenesis, sex determination and organ maturation. The prominent role is played by the **miR156**-**miR172** network. In maize inflorescences, spikelet meristem specification and organogenesis are regulated by these two nodes. Furthermore, in *Arabidopsis* inflorescences, miR172 target, AP2, binds not only to miR156e *cis*-regulatory sequences, as in leaf cells, but also to miR172a locus, thus creating a loop in the regulatory network (Rubio-Somoza and Weigel, 2011). In this way AP2 specifies the identities of the perianth organs, sepals, and petals in the *Arabidopsis* flower (Chen, 2009).

miR159-MYB deregulation can lead to male sterility and delayed flowering under short day in *Arabidopsis* (Chen, 2009).

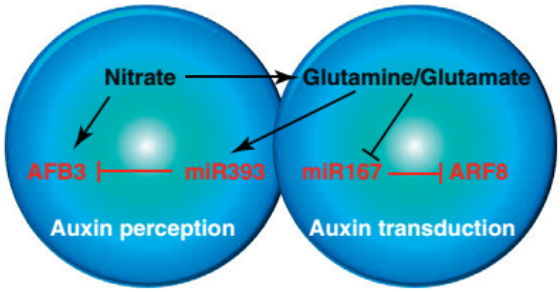
miR167 targets ARF6 and ARF8, which act redundantly to regulate ovule and anther development in *Arabidopsis thaliana* (Chen, 2009).

The identity of the reproductive organs in the flowers of *Petunia hybrida* and *Antirrhinum majus* is controlled by the homeotic class C genes. **miR169** targets the NF-YA class of transcription factors, which are likely activators of these genes, thereby regulating the spatial restriction of the proteins they encode. In the *miR169* mutants, ectopic expression of the class C genes in the outer whorls transforms petals into stamens (Chen, 2009).

In *Brachypodium distachyon*, flowering time is controlled by **miR5200**, a *Pooideae*-specific miRNA. It targets the highly conserved florigen gene FLOWERING LOCUS T (FT) for mRNA cleavage. miR5200 accumulates abundantly to repress FT mRNAs in short day conditions, while it must be repressed to allow FT protein accumulation during long day (Wu *et al.*, 2013).

1.4.4.4 Root development

Root system plasticity mainly relies on secondary growth of lateral roots and root hairs. Auxins play a key role in lateral root development and several miRNAs are linked to auxin signaling. Initiation of root growth is mediated by **miR164**-NAC1 along with **miR160**, which targets different members of the ARF family (ARF10, ARF16, ARF17). In a following phase **miR390** triggers the production of ta-siRNA from *TAS3*, which finally target ARFs. Together with it, two nitrate-regulated miRNA nodes, **miR167**-ARF8 and **miR393**-AFB3 (Auxin-signaling F-Box), act subsequently during emergence and elongation of the lateral root.



Auxin-based plasticity of the root system is well exemplified by the interactions of these two miRNA nodes. At low nitrate concentration in the soil, the transmembrane protein NTR1.1 pumps auxins out of the root cells, preventing their accumulation. When the soil nitrate concentration rises, these transporters show a higher affinity for nitrate instead of auxins, which eventually can accumulate and mediate the growth of lateral roots. Nitrate absorption turns on the auxin receptor AFB3. Glutamine and glutamate derived from Nitrate activate the expression of miR393, which targets AFB3 thus dampening the initial

signal, but at the same time they cause the increase of ARF8 activity through reduction of miR167 levels (Rubio-Somoza and Weigel, 2011).

1.4.4.5 Senescence

Senescence is a process mainly controlled by two miRNAs, **miR164** and **miR319**. In *Arabidopsis*, the former decreases with age, resulting in growing accumulation of its targets: NAC1, ORE1 (AtNAC2) and *At5g61430*. Accordingly, overexpression of miR164, or loss of ORE1 activity, promotes leaf longevity. In contrast, miR319 has several activities, one of which is to repress the onset of senescence. It seems likely that miR319 levels gradually decline with age. Its targets, TCPs, affect senescence by activating LIPOXYGENASE2 (LOX2), one of the main jasmonic acid (JA) biosynthetic enzymes.

Like ethylene and JA, auxins also play a big role in the regulation of plant longevity, delaying degenerative processes like leaf senescence or floral organ abscission. The **miR390**-triggered TAS3 targets ARF2, which is a negative regulator of auxin responses, and is involved in controlling the onset of leaf senescence and floral organ abscission.

It is striking that despite these three miRNA nodes, miR319-TCP, miR164-NAC and miR390-TAS3-ARF2, regulate similar processes, it has not been explored whether they form a miRNA regulatory network (Rubio-Somoza and Weigel, 2011).

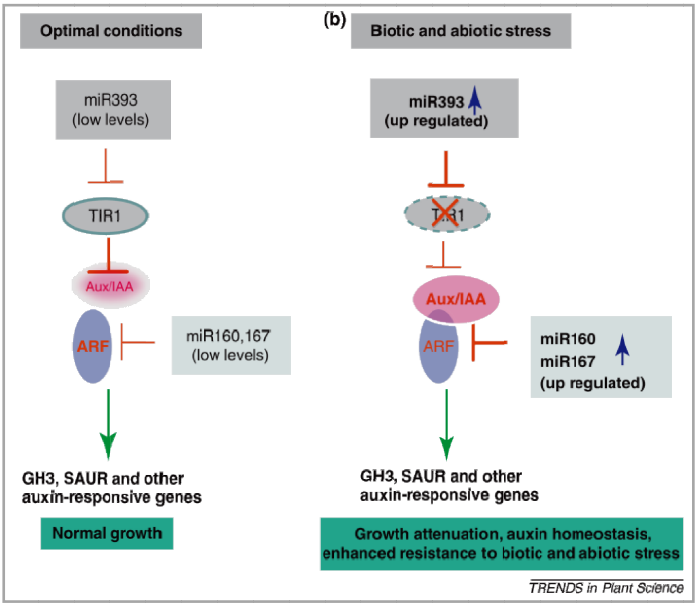
1.4.5 Role in abiotic stress response

Being sessile organisms, plants are exposed to a vast array of environmental stresses, both biotic and abiotic. Responses to these stresses are mediated via multiple mechanisms, and miRNAs are often considered as the restorer of cellular homeostasis upon such environmental changes (Leung and Sharp, 2010).

Under stress, growth and development are stalled and the resources are mobilized towards adaptive responses. Most conserved miRNAs target mRNAs encoding diverse families of transcription factors. Notably, the levels of those conserved miRNAs appear to be regulated during stress, and their target genes appear to be stress regulated as well. This suggests that plant growth and development are modulated during stress (Khraiwesh *et al.*, 2011).

Lots of studies have been carried out on abiotic stresses such as drought, heat and cold (these three stresses are a major focus of this thesis). Results varied between different studies: for evolutionarily close species this fact may be due to the employment of different experimental conditions (e.g. age of plants, way of stress induction, stress intensity and duration, growth conditions, tissue examined, methods employed to monitor the miRNA expression, and so on), but at the same time it is also possible that specific miRNAs play a specific role in different species.

Up-regulation of miR393, miR160 and miR167 has been commonly observed during drought stress and other abiotic stresses like salt, ABA, UV-B, cold and heat. Their altered expression in response to stress seems to be important for the attenuation of plant growth and development (Sunkar et al., 2012).



As previously mentioned, these three miRNAs are implicated in auxin signaling, therefore in plant growth, since they target TIR1/AFB, and ARF transcription factors. During stress, up-regulated miR393 contributes to the repression of auxin signaling by lowering TIR1 levels, in order to increase AUX/IAA-ARF hetero-dimerization and ARF release. In addition, miR160 and miR167 down-regulate ARF levels.

In this way, the overall ARF-mediated gene expression is suppressed by these three miRNAs. The consequence is the attenuation of plant growth and development under stress, and possibly the promotion of plant stress tolerance as well (from Sunkar *et al.*, 2012).

miR169 is well known for playing an important role in abiotic stress (drought and salinity in particular). In *Arabidopsis*, miR169 down-regulation allows expression of the mRNA of *NFYA5*, which is important for drought responses. It has been hypothesized that its expression in guard cells controls stomatal aperture, while in other cells it may contribute to the expression of stress-associated genes (Khraiwesh *et al.*, 2011). One curious finding regarding this microRNA expression during drought stress is that, unlike in *Arabidopsis* and *Medicago truncatula*, it is upregulated in rice (*Oryza sativa*) (Sunkar *et al.*, 2012).

miR168 has been frequently found to be regulated under different stresses, such as drought, heat, salinity and UV-B radiation. It seems to be transcriptionally co-regulated with its target, AGO1, therefore allowing maintenance of the homeostasis in every cell where the miRNA pathway is functioning (Khraiwesh *et al.*, 2011).

miR171 is modulated in response to drought and other abiotic stresses (Zhang *et al.*, 2010; Hwang *et al.*, 2011). It targets transcription factors of the Scarecrow-like (SCL) family, which antagonize master growth repressors DELLA in *Arabidopsis* (Zhang *et al.*, 2011), thus promoting gibberellin signaling. It is also known that these TFs are induced by drought stress (Ma *et al.*, 2010), putting forward the involvement of this miRNA in plant growth and development networks under stress conditions.

miR396 is a salt responsive plant miRNA, but also responds to other stresses like drought and cold. Overexpression of miR396 in rice and *Arabidopsis* results in reduced salt and alkali stress tolerance compared to wild-type plants.

In a recent study, expression levels of **miR398**, which targets two Cu/Zn superoxide dismutases (CSD1 and CSD2), were responsive to salt stress. Salt stress decreased the expression of miR398 and this correlated with the increased expression of its target genes. Under stress conditions, reactive oxygen species (ROS) tend to accumulate and their detoxification requires increased CSD1 and CSD2 expression, which is mediated by suppression of miR398 levels.

Interestingly this microRNA is also involved in a tightly regulated process that aims at the conservation of copper homeostasis. Copper is an essential micronutrient that serves as a cofactor for proteins involved in various plant functions, including photosynthesis, scavenging of reactive oxygen species and sensing of

ethylene. Several miRNAs were described as copper-sensitive, besides **miR398**: **miR397**, **miR408** and **miR857**. They are all known to target transcripts encoding copper containing proteins such as superoxide dismutases, plantacyanin and several laccases (laccase-2, 3, 4, 7, 12, 13 and 17). Copper deprivation causes the induction of these miRNAs, with the consequent down-regulation of target genes, suggesting that plants evolved this mechanism to save copper for more essential proteins, such as plastocyanin and cytochrome C oxidase (Sunkar *et al.*, 2012).

In summary, miR398 levels are up- or downregulated depending on the physiological status of a plant; down-regulation occurs in response to lower oxidative stress, whereas upregulation occurs to save copper.

miR397 is involved in the response to many abiotic stresses, like heat, cold, salinity and drought.

The expression of **miR395** was increased upon sulphate starvation. It targets the genes that encode ATP sulfurylases APS1, APS3, and APS4. These enzymes catalyze the first step of inorganic sulfate assimilation.

Interestingly, some miRNAs respond differently in different tissues during stress, and even in cultivars contrasting for stress sensitivity.

For instance, four miRNAs displayed tissue-specific regulation during dehydration in barley: **miR166** was upregulated in leaves, but downregulated in roots; **miR156a**, **miR171** and **miR408** were induced in leaves, but unaltered in roots. Similarly, in *M. truncatula*, **miR398** upregulation was more pronounced in the leaves than in the roots of water-stressed plants (Sunkar *et al.*, 2012).

miRNA profiles in drought-tolerant and drought-sensitive soybean genotypes yielded striking differences. The same set of microRNAs was upregulated in the sensitive genotype and downregulated in the tolerant one (Sunkar *et al.*, 2012).

Many other microRNAs have been found to be modulated under drought stress in different species (for a comprehensive list see Khraiwesh et al. (2011), Sunkar et al. (2012), Eldem et al. (2013) and references therein).

Objectives and thesis outline

The overall aim of this thesis was to identify microRNAs involved in the complex networks underlying plant development and response to abiotic stresses such as drought, heat and low temperatures. Two approaches were employed to reach this aim: a) bioinformatic searches in online sequence datasets of different species, and b) microRNA discovery based on Next Generation Sequencing in proper experimental systems. Details of these experimental procedures and results are described below.

Chapter 3 describes materials and methods adopted in the present study. Some of the methods were newly created (e.g., the special protocol for degradome libraries construction), while some other were only optimized from existing protocols (RNA isolation, stem-loop qRT-PCR).

Chapter 4 reports on the results of the first strategy. A computational identification of putative exonic miRNAs in 30 plant species and an analysis of possible mechanisms involved in their regulation were carried out.

In **Chapter 5** we describe the results coming from a deep-sequencing run. Sixty-nine small RNA libraries were prepared and sequenced on an Illumina® Genome Analyzer II_x. **Chapter 5.2** focuses on the wide assortment of miRNA sequence variants that arise from the same precursor (isomiRs). Peach samples coming from five developmental stages were collected. **Chapters 5.3** to **5.5** deal with the miRNome adaptations to abiotic stresses studied in well characterized experimental systems. Nure and Tremois are two barley cultivars that differ for cold tolerance and vernalization requirements, and their response to low temperatures at miRNA level is characterized in **Chapter 5.3**. In the following Chapter (**5.4**) miRNAs involved in the nucleus-organelles cross-talk were searched in a collection of bread wheat euplasmic and alloplasmic lines. **Chapter 5.5** reports on the miRNA genes modulated in response to drought and heat stresses in two durum wheat cultivars, Ofanto and Cappelli, previously studied for their contrasting efficiency in the use of water during the exposure to such conditions.

The ultimate goal is to reach a progress in our understanding of these regulatory mechanisms. We took into consideration species of agricultural interest (peach, barley and wheat), with an eye towards a possible exploitation of these genetic relations for breeding.

Materials and Methods

3.1 Experiments setup

Experiment 1: *Peach development*

A 12-year-old tree grafted on wild seedling of the yellow-fleshed cultivar Maycrest (*Prunus persica* (L.) Batsch), grown in Palazzolo di Sona, Verona, Italy (45.457˚N, 10.822˚E), was used as plant source material. Each sample was collected pooling together material from three different branches of the same plant. Four phenological stages (Chapman and Catlin, 1976) were considered: swollen bud, half-inch green, pink, bloom. Leaf and flower swollen buds were collected 41 days before flowering (DBF), half-inch leaves were collected 21 DBF, pink flower buds were collected six DBF. Tissues were frozen in liquid nitrogen immediately after drawing.

Experiment 2*: Barley response to low temperatures*

Two *Hordeum vulgare* L. cultivars, Nure and Tremois, were used as plant materials. The winter cv. Nure [(Fior40 x Alpha) x Baraka] is a modern, high yielding two-rowed feed-barley cultivar released by the Genomic Research Centre (formerly Istituto Sperimentale per la Cerealicoltura) of the Agricultural Research Council (CRA) of Fiorenzuola d'Arda, Italy. The spring cv. Tremois [(Dram x Aramir) x Berar] is a modern, high yielding French two-rowed malting variety. Three biological replicates per group were considered, and 25 seeds per vase (11 cm diameter) were sown. Seedlings were grown under controlled conditions as follows: 10 hours of light at 20° C and 14 hours of dark at 15° C. After 7 days plants were at the first leaf stage and low temperature treatment started. Leaves from control plants were harvested immediately, while all the other groups were moved to a growth chamber with the following conditions: 10 hours of light at 3.5° C and 14 hours of dark at 2.5° C. Six time-points were considered: 0 (control), 2, 8 and 24 hours, 3 and 6 days. The last two time points were not considered for the deep sequencing run.

Experiment 3*: MicroRNAs involved in nucleus-organelles crosstalk*

The *Hordeum chilense* (Roem. et Schultz) - *Triticum aestivum* L. alloplasmic line TH237 was used. It was developed at IAS-CSIC as described by Atienza *et al.* (2007), using common wheat T20 as nucleus donor and *H. chilense* H7 as cytoplasm donor. The euplasmic background was used as the control for the experiment. Fifty seeds were sown in three biological replicates for each sample in pots of 11 cm diameter and grown in controlled conditions as follows: 12 hours of light at 22° C with a light intensity of 600 µE m^{-2} s^{-1}, 12 hours of dark at 15°C. Fully expanded second leaves were collected two weeks after sowing in the middle of the light period and frozen in liquid nitrogen.

Experiment 4*: Durum wheat response to drought and heat stress*

The experiment was performed using two durum wheat (*Triticum turgidum* subsp. *durum*) cultivars, Ofanto and Cappelli, that were grown in a growth chamber under controlled conditions as follows: 16 hours of light at 20° C and 8 hours of dark at 18° C, with a relative humidity of 50% and a Soil Water Content (SWC) kept at 28%. Standard peat soil was used. Stress treatments were applied when the plants reached the developmental stage of fully expanded mature 3rd leaf.

Heat stress consisted in 3 hours at 36° C in presence of light, and leaves were harvested immediately after. For the drought stress, water was withheld until the RSWC dropped to 18% (mild drought stress) and 12.5% (severe water stress). Leaves were harvested 2 days and 3 days, respectively, for mild and severe stress, after RSWC reached the desired threshold. Leaves from control plants were harvested the same day and hour of leaves plants subjected to severe stress (that moment they were exposed to light).

The experiment was repeated with slight modifications during an internship at the Australian Centre for Plant Functional Genomics (ACPFG) of Adelaide, Australia. Two new cultivars (Radioso and Trinakria) were chosen based on their pedigree, since Ofanto and Cappelli were not available at the ACPFG. Radioso is a modern italian cultivar, which has got Creso in its pedigree (Creso x Isa1). Trinakria is an old, tall and low yielding cultivar and derives from a cross between a landrace (B14) and Capeiti (Cappelli x Eiti8). The system formed by Creso and Trinakria has been studied in several papers dealing with drought tolerance, where the former was considered drought-susceptible and the latter more drought-tolerant (De Leonardis *et al.*, 2007; Burling *et al.*, 2013). Radioso was the best choice since having no availability of Creso seeds. Moreover, another cultivar, Yallaroi, was added to the present study. Its function was to provide evidence that a given microRNA is indeed related to drought, and not to a breeding group, since Yallaroi does not belong to the same breeding group of the italian cultivars.

Coco-peat soil was used and field capacity (FC) was calculated. Twelve vases were filled with 400 g of soil. Six of them were placed in an oven at 50° for 4 days to calculate the dry weight, while 6 of them were watered to excess to calculate the wet weight. The difference between the averages of wet and dry weights gave the FC of the soil, a percentage of which should have been used as threshold for watered controls and drought stress treatment. Given some difficulties in determining the wet weight though, another formula (De Leonardis, personal communication) for Soil Water Content (SWC) calculation was considered. It takes into consideration the dry weight only, and the following thresholds:

- Watered controls: 28% of SWC
- Mild drought: 18% of SWC
- Severe drought: 12.5% of SWC

The formula is as follows:

$$\mathrm{x}= dw + dw * \frac{t}{1-t}$$

where

- x = final weight to reach
- t = desired threshold (0.28, 0.18 or 0.125)
- dw= dry weight (212 g on average)

Plants were therefore organized into 4 groups (two levels of stress and two controls) made of 6 biological replicates. Two developmental stages (3rd-leaf and booting) were taken into consideration. In total 48 plants per cultivar were grown. Plants were grown in a bigfoot in long day conditions: 16 hours of light at 20° C and 8 hours of dark at 18° C. Dawn was set at 10.30 am to allow for measurements of physiological parameters, like chlorophyll fluorescence, that have to be taken during the dark period.

Vases were weighed every day, at 5 pm, and water was added to compensate for evaporation. When the first lot of plants reached the 3rd leaf stage, stress was applied. Control plants were kept watered to the 28% threshold of SWC, while water was withheld for the other two groups. When mild-drought group

reached 18%, plants were watered to maintain that condition for two days, and then they were harvested. Concerning severe drought stress, water was withheld until the end. Harvesting occurred when plants showed a clearly visible phenotype (Figure 9). Control plants were harvested along with their respective drought treatment. All aboveground tissues were cut and put into falcon tubes. Three biological replicates out of six were harvested.

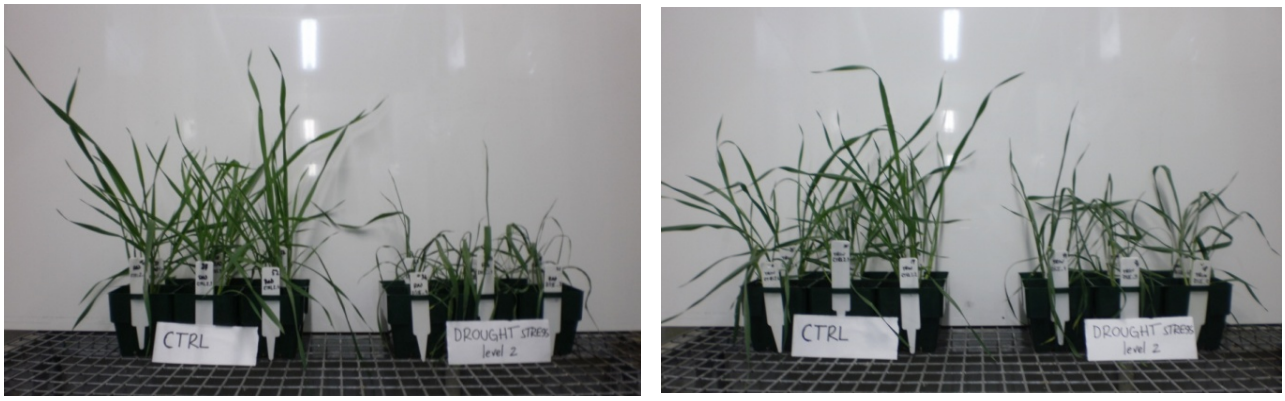


Fig. 9 *Radioso (left) and Trinakria (right) plants on the harvest day*

Concerning the second lot of plants, drought treatment was carried out as described above (with the exception of the threshold for mild stress, which was 12.5% of SWC instead of 18%), and a correction was made to take into consideration the increasing weight of the growing plants. To do so, some spare plants from the Radioso pool were cut and weighed every week. Different tissues were harvested separately: flag leaves, older leaves and culms, spikes were harvested from the 3 most similar biological replicates out of 6, and put into falcon tubes.

To check for varietal differential behaviour under drought stress, also physiological measurements were taken on plants subjected to severe drought. Measurements of chlorophyll fluorescence, chlorophyll content and stomatal conductance were taken with a Mini-Pam (Walz), a SPADmeter (Konica Minolta), and a Leaf Porometer (Decagon), respectively.

3.2 RNA Isolation

Trizol protocol (modified)

This protocol was adopted for experiments 2, 3 and 4.

- **Preparation.** Heat Trizol (1 ml per 100 mg of tissue) at 55-60° C for 1 hour in the fume hood. Pre-chill the centrifuge at 4 °C.
- **Homogenization**. Grind 100 mg of tissue in liquid nitrogen with mortar and pestle. Transfer to a 2 ml eppendorf tube and leave at -80° C until further processing takes place. Remove samples from the -80° C freezer in liquid nitrogen and immediately pour 1 ml of preheated Trizol into each tube. Mix well by vortexing and incubate for 5 min at 60° C in the fume hood.
- **Phase Separation.** Centrifuge samples at 12000 x g at 4°C for 10 min (increase speed if centrifuge and tubes allow it). Transfer supernatant to a new eppendorf.
- **Chloroform**. Add 200 µl of chloroform and vortex for 15 sec. Leave tubes at room temp for 2-3 min and centrifuge at 12000 x g at 4° C for 5 min.
- **Precipitation.** Transfer the aqueous phase to a new tube and add 2 volumes of isopropanol. Mix and incubate at -80 °C for 20 min. Centrifuge tubes at max rpm at 4° C for 15 min.

- **Wash.** Wash pellet with 1 ml of 75% ethanol. Centrifuge at max rpm at 4° C for 5 min, briefly dry pellet (5-10 min; not longer - no centrifugation under vacuum). Resuspend the pellet in 35 μl of DEPC-MQ water and add RNAase inhibitor (8 u per 100 mg).
- **Resuspension.** Incubate at 55 – 60° C for 10 min to resuspend pellet. Centrifuge samples at maximum speed at 4° C for 10 min. Pipette supernatant into new Eppendorf tubes.
- **Concentration and quality.** Determine RNA concentration with a spectrophotometer by putting 2 μl of sample in 198 μl of 10 mM Tris-HCl (pH 7.5). Determine quality by running a 1 μl aliquot of every sample on an Agilent 2100 Bioanalyzer RNA 6000 Nano chip.

Norgen's Plant Total RNA Purification Kit

This protocol was adopted for experiment 1.

- **Lysate Preparation.** Transfer 50-100 mg of plant tissue into a mortar with 600 μL of Lysis Solution. Grind the sample using a pestle until the tissue is completely macerated. Using a pipette, transfer the lysate into an RNAase-free microcentrifuge tube.
  Spin the lysate for 2 minutes to pellet any cell debris. Transfer the supernatant to another tube.
  Add an equal volume of 70% ethanol to the lysate collected above. Vortex to mix.
- **Binding to Column.** Assemble a column with a collection tube. Apply up to 600 μL of the clarified lysate with ethanol onto the column and centrifuge for 1 minute at 14000 × g. Discard the flowthrough and reassemble the spin column with the collection tube. Depending on your lysate volume, repeat previous step if necessary.
- **Column Wash.** Apply 400 μL of Wash Solution to the column and centrifuge for 1 minute. Discard the flowthrough and reassemble the spin column with its collection tube. Repeat the wash step three times, then spin the column for 2 minutes in order to thoroughly dry the resin. Discard the collection tube.
- **RNA Elution.** Place the column into a fresh 1.5 ml eppendorf tube. Add 35 μL of Elution Buffer to the column. Centrifuge for 2 minutes at 200 x g, followed by a 1 minute spin at 14000 x g.
- **Concentration and quality.** Determine RNA concentration with a spectrophotometer by putting 2 μl of sample in 198 μl of 10 mM Tris-HCl (pH 7.5). Determine quality by running a 1 μl aliquot of the sample on an Agilent 2100 Bioanalyzer RNA 6000 Nano chip.

3.3 Small RNA libraries preparation

The TruSeq Small RNA Sample Prep Kit (Illumina®) workflow was followed as per the manufacturer's instructions.

Adapter ligation. 1 μg of total RNA was ligated with two adapters at 3' and 5' ends, without any size fractionation.

Reverse Transcription and amplification. Adapter-ligated RNA was reverse transcribed with SuperScript II Reverse Transcriptase (Invitrogen), then PCR-amplified. Samples were barcoded using different variants of the reverse primer provided with the kit. An aliquot of every PCR reaction was checked with a High Sensitivity chip on the Agilent 2100 Bioanalyzer, to look for the desired peaks.

Libraries purification. Libraries from the same experiment were pooled together to a maximum plexity of 12, and then purified on a 6% TBE PAGE gel after electrophoresis. Multiplexing levels were as follows: 9+9 (experiment 1, peach development), 12+12 (experiment 2, barley response to low temperatures), 6 (experiment 3, eu-alloplasmic lines), 12+12 (experiment 4, durum wheat response to drought and heat stresses). Libraries quality and concentration were evaluated with the Agilent 2100 Bioanalyzer DNA1000 assay.

Loading. Libraries were diluted to 10 nM and finally loaded at 12 pM on an Illumina Genome Analyzer II_x. They subsequently underwent 30 cycles of sequencing.

Several runs were prepared during the PhD project. For the sake of simplicity they are here reported as one.

#	Sample name	Species	Index #	Index Seq	Lane	Recipe	Sample Project
1	Gemme fiore #1	Peach	RPI1	ATCACG	**1**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Peach-development
2	Gemme fiore #2	Peach	RPI2	CGATGT	**1**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Peach-development
3	Gemme fiore #3	Peach	RPI3	TTAGGC	**1**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Peach-development
4	Gemme legno #1	Peach	RPI4	TGACCA	**1**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Peach-development
5	Gemme legno #2	Peach	RPI5	ACAGTG	**1**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Peach-development
6	Gemme legno #3	Peach	RPI6	GCCAAT	**1**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Peach-development
7	Bottoni fiorali #1	Peach	RPI7	CAGATC	**1**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Peach-development
8	Bottoni fiorali #2	Peach	RPI8	ACTTGA	**1**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Peach-development
9	Bottoni fiorali #3	Peach	RPI9	GATCAG	**1**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Peach-development
1	Gemme fiore #1	Peach	RPI1	ATCACG	**2**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Peach-development
2	Gemme fiore #2	Peach	RPI2	CGATGT	**2**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Peach-development
3	Gemme fiore #3	Peach	RPI3	TTAGGC	**2**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Peach-development
10	Orecchiette #1	Peach	RPI13	AGTCAA	**2**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Peach-development
11	Orecchiette #2	Peach	RPI14	AGTTCC	**2**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Peach-development
12	Orecchiette #3	Peach	RPI15	ATGTCA	**2**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Peach-development
13	Fiore #1	Peach	RPI16	CCGTCC	**2**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Peach-development
14	Fiore #2	Peach	RPI17	GTAGAG	**2**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Peach-development
15	Fiore #3	Peach	RPI18	GTCCGC	**2**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Peach-development
16	Nure_ctrl_1	Barley	RPI19	GTGAAA	**3**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Barley_Cold
17	Nure_ctrl_2	Barley	RPI20	GTGGCC	**3**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Barley_Cold
18	Nure_ctrl_3	Barley	RPI21	GTTTCG	**3**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Barley_Cold
19	Nure_2h_1	Barley	RPI22	CGTACG	**3**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Barley_Cold
20	Nure_2h_2	Barley	RPI23	GAGTGG	**3**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Barley_Cold
21	Nure_2h_3	Barley	RPI24	GGTAGC	**3**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Barley_Cold
22	Nure_8h_1	Barley	RPI25	ACTGAT	**3**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Barley_Cold
23	Nure_8h_2	Barley	RPI26	ATGAGC	**3**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Barley_Cold
24	Nure_8h_3	Barley	RPI27	ATTCCT	**3**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Barley_Cold
25	Nure_24h_1	Barley	RPI28	CAAAAG	**3**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Barley_Cold
26	Nure_24h_2	Barley	RPI29	CAACTA	**3**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Barley_Cold
27	Nure_24h_3	Barley	RPI30	CACCGG	**3**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Barley_Cold
	phiX	-	-	-	**4**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	phi_control
28	Tremois_ctrl_1	Barley	RPI31	CACGAT	**5**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Barley_Cold
29	Tremois_ctrl_2	Barley	RPI32	CACTCA	**5**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Barley_Cold
30	Tremois_ctrl_3	Barley	RPI33	CAGGCG	**5**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Barley_Cold
31	Tremois_2h_1	Barley	RPI34	CATGGC	**5**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Barley_Cold
32	Tremois_2h_2	Barley	RPI35	CATTTT	**5**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Barley_Cold
33	Tremois_2h_3	Barley	RPI36	CCAACA	**5**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Barley_Cold
34	Tremois_8h_1	Barley	RPI37	CGGAAT	**5**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Barley_Cold
35	Tremois_8h_2	Barley	RPI38	CTAGCT	**5**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Barley_Cold
36	Tremois_8h_3	Barley	RPI39	CTATAC	**5**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Barley_Cold
37	Tremois_24h_1	Barley	RPI40	CTCAGA	**5**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Barley_Cold
38	Tremois_24h_2	Barley	RPI41	GACGAC	**5**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Barley_Cold
39	Tremois_24h_3	Barley	RPI42	TAATCG	**5**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Barley_Cold
40	Eu 20.2	Wheat	RPI1	ATCACG	**6**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Eu-Alloplasmic lines
41	Eu 20.3	Wheat	RPI2	CGATGT	**6**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Eu-Alloplasmic lines
42	Eu 20.4	Wheat	RPI3	TTAGGC	**6**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Eu-Alloplasmic lines

43	Allo 237.2	Allopl.	RPI10	TAGCTT	**6**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Eu-Alloplasmic lines
44	Allo 237.3	Allopl.	RPI11	GGCTAC	**6**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Eu-Alloplasmic lines
45	Allo 237.4	Allopl.	RPI12	CTTGTA	**6**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Eu-Alloplasmic lines
46	Of ctrl #1	Wheat	RPI25	ACTGAT	**7**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Wheat_drought-heat
47	Of ctrl #2	Wheat	RPI26	ATGAGC	**7**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Wheat_drought+heat
48	Of ctrl #3	Wheat	RPI27	ATTCCT	**7**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Wheat_drought+heat
49	Of drought_I #1	Wheat	RPI28	CAAAAG	**7**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Wheat_drought+heat
50	Of drought_I #2	Wheat	RPI30	CACCGG	**7**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Wheat_drought+heat
51	Of drought_I #3	Wheat	RPI31	CACGAT	**7**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Wheat_drought+heat
52	Of drought_II #1	Wheat	RPI1	ATCACG	**7**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Wheat_drought+heat
53	Of drought_II #2	Wheat	RPI32	CACTCA	**7**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Wheat_drought+heat
54	Of drought_II #3	Wheat	RPI33	CAGGCG	**7**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Wheat_drought+heat
55	Of heat #1	Wheat	RPI34	CATGGC	**7**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Wheat_drought+heat
56	Of heat #2	Wheat	RPI35	CATTTT	**7**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Wheat_drought+heat
57	Of heat #3	Wheat	RPI17	GTAGAG	**7**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Wheat_drought+heat
58	Capp ctrl #1	Wheat	RPI13	AGTCAA	**8**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Wheat_drought+heat
59	Capp ctrl #2	Wheat	RPI14	AGTTCC	**8**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Wheat_drought+heat
60	Capp ctrl #3	Wheat	RPI15	ATGTCA	**8**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Wheat_drought+heat
61	Capp drought_I #1	Wheat	RPI16	CCGTCC	**8**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Wheat_drought+heat
62	Capp drought_I #2	Wheat	RPI17	GTAGAG	**8**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Wheat_drought+heat
63	Capp drought_I #3	Wheat	RPI18	GTCCGC	**8**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Wheat_drought+heat
64	Capp drought_II #1	Wheat	RPI19	GTGAAA	**8**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Wheat_drought+heat
65	Capp drought_II #2	Wheat	RPI20	GTGGCC	**8**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Wheat_drought+heat
66	Capp drought_II #3	Wheat	RPI21	GTTTCG	**8**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Wheat_drought+heat
67	Capp heat #1	Wheat	RPI22	CGTACG	**8**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Wheat_drought+heat
68	Capp heat #2	Wheat	RPI23	GAGTGG	**8**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Wheat_drought+heat
69	Capp heat #3	Wheat	RPI24	GGTAGC	**8**	GA2_PEM2X_MP_30+7Cycle_SR_v8.3px.xml	Wheat_drought+heat

Table 1 *Hypothetical sample sheet collecting all the information from several Illumina runs performed during the PhD project.*

3.4 Illumina data analysis

Data analyses were carried out with different pieces of software.

Experiment 1

Specifications about Illumina data analyses are reported in the enclosed paper (Chapter 5.2.2).

Experiment 2

Data analysis, from read filtering to miRNA annotation, was carried out with the software CLC Genomics Workbench 5. Briefly, reads were first filtered to remove adaptor sequences. Those shorter than 18 nt or longer than 24, low-complexity reads, reads matching rRNAs or tRNAs were eliminated. Remaining reads were aligned with miRBase version 20 (June 2013), allowing 2 mismatches at most. The program performed subsequent alignments considering the following hierarchical evolutionary order: *Hordeum vulgare*, *Triticum aestivum*, *Brachypodium distachyon*, *Oryza sativa*, *Zea mays*, *Arabidopsis thaliana*. The evolutionarily closer alignment with the smaller number of mismatches was kept. Fold changes were calculated from normalized read counts, and Baggerley's test was applied to determine statistical significance.

Experiment 3 and 4

Free, web-available tools were used for data analysis: Sequence pre-processing, Filter and miRProf from the UEA sRNA toolkit, plant version[1]. Annotation of known miRNAs was carried out referring to version 18 of miRBase (November 2011). Read counts in each library were normalized dividing by the total number of

[1] http://srna-tools.cmp.uea.ac.uk/

qualified miRNA reads. Fold changes were calculated from normalized read counts, and t test was applied to determine statistical significance.

Target search

MiRNA target identification was carried out with the psRNATarget tool[2] with default parameters and the most recent Gene Index releases (Dai and Zhao, 2011). To score the complementarity between miRNAs and their target transcript, psRNATarget applies the scoring schema of miRU by Zhang (2005). The maximum expectation is the threshold of the score. A small RNA/target site pair will be discarded if its score is greater than the threshold. The default cut-off threshold is 3.0.

3.5 Stem-loop qRT-PCR on microRNAs

Protocol adapted from Varkonyi-Gasic *et al.* (2007) – *modified for ABI SYBR Green PCR Master Mix.*

Primer design

Primers were designed according to the above-mentioned protocol. An excel macro was developed to speed up the procedure. The universal reverse primer was always used except in case of a high level of complementarity with the forward primer. Modified reverse primers were used in these situations, and the backbone of the stem-loop RT primers was adjusted accordingly.

Forward primers were checked with Vector NTI Advance 10 (Invitrogen). They were designed in order to have a melting temperature ranging between 56 and 58° C, and to avoid self-complementarity and reverse complementarity (dG had to be more than -2).

Housekeeping genes

Different housekeeping genes were considered for qPCR data normalization. They were both classic and smallRNA-like. Polyubiquitin, Actin, GAPDH were used in experiments 2 and 3. In experiment 4 OEP16-3 was employed, since it showed a better stability than other normalizers. It was taken from a previous work on the same experiment (Aprile et al., 2013). A microRNA (miR827) was employed for experiment 3. Small-nucleolar RNAs (snoRNAs) were considered as well. Their sequences were taken from an online database[3] and primers were designed on the loop region (carefully avoiding regions partly on the loop and partly on the stem). *Hv-U59* and *Hv-U61* were used for barley while *Ta-snoR10* was used in durum wheat experiment. Reverse primers were added in the retrotranscription reaction (see following table) instead of the "reference stem-loop primer" (1 μl of a 10 μM dilution).

All primers used in these experiments for microRNA trend validations are reported in Table 3 at the end of this chapter.

Retrotranscription

- Prepare the first Master Mix in a 0.2 ml eppendorf tube. Multiply the volumes by the number of samples. Include 10% overage to cover pipetting errors (s.l. = stem loop).

[2] http://plantgrn.noble.org/psRNATarget
[3] http://bioinf.scri.sari.ac.uk/cgi-bin/plant_snorna/home

Reagent	[ ]	ul
dNTP	10 mM	0.50
H_2O	-	10.15
s.l. primer 1	1 uM	1.00
s.l. primer ref.	1 uM	1.00

- For a 'no RT primer' master mix, add 1ul of RNA instead of s.l. primer
- Heat mixture to 65°C for 5 minutes
- Incubate on ice for 2 minutes
- Centrifuge briefly to bring solution to the bottom and transfer to a 1.5 ml tube
- Then add (+10% for pipetting errors)

Reagent	[ ]	ul
1st Strand Buffer	5x	4.00
DTT	0,1 M	2.00
RNaseOUT	40 u/ul	0.10
SuperScript III	200 u/ul	0.25

- Mix gently and centrifuge to bring solution to the bottom of the tube.
- Aliquot 19 µl of the Master mix in PCR tubes

- Add 1 µl RNA template to 'no RNA' Master mix (usually 100-200 ng) **OR** Add 1 µl of appropriate stem-loop RT primer (1µM) to 'no RT primer' Master mix

- Mix gently and centrifuge to bring solution to the bottom of the tube.
- Perform pulsed RT as per the following table.

Temperature	Time	Cycles
16°C	30 min	1
30°C	30 sec	
42°C	30 sec	60
50°C	1 sec	
85°C	5 min	1
4°C	∞	

End-point PCR

- Prepare a PCR master mix by scaling the volumes listed below to the desired number of amplification reactions. Include 10% overage to cover pipetting errors. Also prepare water controls by adding nuclease-free water in place of the RT product.
- Add the following components to a nuclease-free microcentrifuge tube:

Reagent	[ ]	ul
H_2O	-	10.8
Buffer	5x	4.0
MgCl2	25 mM	1.6
Pr fwd	10 uM	0.4
Pr rev	10 uM	0.4
dNTP	10 mM	0.4
Taq pol	3 u/ul	0.4
cDNA		2.0

- Place reactions in a preheated (94°C) thermal cycler heat block and incubate as follows

Temperature	Time	Cycles
94°C	2 min	1
94°C 60°C 72°C	30 sec 30 sec 30 sec	33
72°C	5 min	1
4°C	∞	

- Analyze reaction products by electrophoresis on a 3% agarose gel in 1× TAE.

qPCR

- Prepare the following mix for every sample. Three technical replicates are considered, and the procedure has to be repeated for every primer couple (this case, two). Considering a little excess, volumes can therefore be multiplied by 6.5. Two replicates only for NTC (multiply by 4.5).

Reagent	ul	x 6,5
cDNA	2.0	13.00
ABI Master Mix	12.5	81.25
H_2O	5.5	35.75

Reagent	ul	x 4,5
ABI Master Mix	12.5	56.25
H_2O	7.5	33.75

- Aliquot 20 μl of the mix in the qPCR plate wells.
- Centrifuge briefly.
- Add 2.5 μl of each 10 μM primer. Forward on left border, reverse on right border.
- Cover with the optical film.
- Centrifuge briefly.
- Incubate on the qPCR machine.
- Use the following running conditions

Temperature	Time	Cycles
95°C	10 min	1
95°C 60°C	15 sec 1 min	40
Dissociation stage		1

3.6 5'-RLM-RACE

A 5' RNA Ligase Mediated - Rapid Amplification of cDNA Ends (5'-RLM-RACE) procedure was developed with the FirstChoice® RLM-RACE Kit by Invitrogen. Slight modifications were made from the original protocol.

Briefly, CIP/TAP treatment was avoided in order to favor 5' adaptor ligation to already decapped products only. 800 μg of total RNA were used for the 5'-Adapter ligation. The obtained products were reverse-transcribed following protocol instructions. cDNA was then amplified using PCR reagents from the kit. Go-Taq DNA Polymerase (Promega) as thermostable DNA polymerase (not included in the kit).

PCR conditions were the same reported in the instructions. PCR products were visualized on a 1.5% Agarose gel. Selected bands were excised for sequencing.

Primers were designed according to the suggestions contained in the protocol. Outer primers were approximately 300 bp far from the putative cleavage site, while inner primers were 150 bp far. They were ~22-25 nt long, no terminal Gs at 3' end, and had a melting temperature of at least 60° C. They were checked for primer dimmers formation with Vector NTI Advance 10 (Invitrogen). Selected targets accessions are reported in the following table.

miRNA name	Exp.	miRNA sequence	Target Accession	Expectation	Target Description
bdi-miR159-3p	4	CTTGGATTGAAGGGAGCTCT	TC368630	2	Transcription factor Myb3
			TC421314	3	Transcription factor Myb3
osa-miR171h	4	TGAGCCGAACCAATATCACTC	TC460317	2.5	sensor histidine kinase
hvu-miR397	2	TTGAGTGCAGCGTTGATGAAC	TC262144	0.5	Laccase-23 precursor
			TC274250	2.5	Laccase-10 precursor
			CK568938	2.5	
osa-miR444e	4	TGCAGTTGCTGCCTCAAGCTT	TC396893	0	Zinc finger (C3HC4-type RING finger) protein-like
			TC368997	**0**	**MIKC-type MADS-box transcription factor WM32B**
			TC403194	0.5	Zinc finger (C3HC4-type RING finger) protein-like
osa-miR1432	4	ATCAGGAGAGATGACACCGAC	CK195521	0	
			TC388166	0.5	

Table 2 *Selected targets tested with the described RACE procedure. Experiment number is reported (2 = barley response to low temperatures; 4 = durum wheat response to drought and heat stress). Expectation is a parameter describing the number of mismatches occurring between a microRNA and its respective target (G-U mismatches weight 0.5). The only validated target is reported in bold.*

3.7 Degradome libraries preparation *(intellectual property belongs to ACPFG)*

A special protocol for degradome libraries preparation was developed by Dr. Bujun Shi at the ACPFG of Adelaide. It was employed to prepare degradome libraries from the Radioso-Trinakria experiment. It enables the analysis of longer RNA fragments compared to the standard fragment length obtained with commercially available kits.

1 DNAase-treated total RNA coming from three different biological replicates of leaf samples of the two cultivars was pooled to reach a final quantity of 20 µg. Control and severe drought samples were used.

2 Ribosomal RNA (rRNA: 5S, 5.8S, 18S and 28S) was removed using the RiboMinus™ Plant Kit for RNA-Seq (Invitrogen), following the manufacturer's instructions.

3 Purified RNAs were ligated to a 5' Adaptor using a T4 RNA Ligase 1 (New England Biolabs). The RNAs from step 2 were incubated at 65° C for 15 minutes and then cooled down to room temperature. Then they were added to a solution containing 10X T4 RNA Ligase Reaction Buffer, 1mM ATP and 15% polyethylene glycol 8000. The reaction was incubated at 37°C for 60 minutes, and then transferred at 16° C overnight (about 14 hours).
The adaptor sequence is (5'-GAGUUCUACAGUCCGACGAUCCAAC-3')

4 Unincorporated RNA adapters and residual tRNAs, 5S rRNA, and 5.8S rRNA were removed by S-400 MicroSpin Columns (GE Healthcare), following the manufacturer's instructions.

5 After the centrifugation, the treated RNAs were extracted with an equal volume of phenol/chloroform. The supernatant was precipitated by adding 3 volumes of ethanol, 20 µg glycogen, 0.1 volume of 0.3 M Sodium Acetate pH 5.2 and dissolved in 11 µl of RNase-free water.

6 The obtained adapter-ligated RNA was reverse-transcribed using Superscript III (Invitrogen), following manufacturer's instructions. Two special primers (P1 and P2) were added in the reaction of 20µl. Their sequences are reported below.

P1 5' GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTTTTTTTTTTTTTTTT 3'

P2 5' GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCNNNNNN 3'

"ℕ" represents random nucleotides.

P1 is able to bind polyA tails, while P2 can anneal everywhere thanks to its random sequence. The reaction was incubated at 37°C for 15 minutes and then 50°C for 60 minutes.

7 A short PCR was performed using special primer pairs.

Radioso - control	**Primer Sequences**
F-long	5' AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTatgtcaGAGTTCTACAGTCCGACGATCCAAC 3'
R-long	5' CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC 3'
F-short	5' AATGATACGGCGACCACCGAGATCTA 3'
R-short	5' CAAGCAGAAGACGGCATACGAGATCGT 3'
Radioso - drought stress	**Primer Sequences**
F-long	5' AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTccgtccGAGTTCTACAGTCCGACGATCCAAC 3'
R-long	5' CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC 3'
F-short	5' AATGATACGGCGACCACCGAGATCTA 3'
R-short	5' CAAGCAGAAGACGGCATACGAGATCGT 3'
Trinakria - control	**Primer Sequences**
F-long	5' AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTgtccgcGAGTTCTACAGTCCGACGATCCAAC 3'
R-long	5' CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC 3'
F-short	5' AATGATACGGCGACCACCGAGATCTA 3'
R-short	5' CAAGCAGAAGACGGCATACGAGATCGT 3'
Trinakria – drought stress	**Primer Sequences**
F-long	5' AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTgtgaaaGAGTTCTACAGTCCGACGATCCAAC 3'
R-long	5' CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC 3'
F-short	5' AATGATACGGCGACCACCGAGATCTA 3'
R-short	5' CAAGCAGAAGACGGCATACGAGATCGT 3'

Sequences reported in lower case represent the index that allows read sorting after multiplexing.

The reactions were composed of 20 µl cDNA, 24 µl 2X Phusion Master Mix (New England Biolabs) and 1 µl each of the four primers. PCR conditions were as follows.

- 98°C for 30 seconds
- 6 cycles of
  - 98°C 10 seconds
  - 55°C 30 seconds
  - 72°C 30 seconds
- 72°C for 10 minutes.

8 Unincorporated primers were removed by S-400 MicroSpin Columns (GE Healthcare), following the manufacturer's instructions.
9 The elution was precipitated as in step 5 and finally dissolved in 20 µl of 10mM Tris-HCl, pH8.5 with 0.1% Tween-20.

The obtained sequences underwent 150 cycles of sequencing.

3.8 qPCR on target genes

Quantitative PCRs were carried out for *in silico* identified targets (Table 2) as follows. 400 ng of total RNA of each sample were reverse transcribed using Superscript II (Invitrogen) according to the manufacturer's standard protocol. Gene specific reverse primers were designed and checked with Vector NTI Advance 10 (Invitrogen) for optimal melting temperature (~58° C) and absence of primer dimmers (dG > -2). The reaction was incubated at 42 °C for 50 min, then heat-inactivated at 70°C for 15 min. The obtained cDNAs were used for qPCR amplification without dilution.

PCR reactions were performed in a final volume of 25 µl containing 2x SYBR Green PCR Master Mix (Life Technologies), Forward/Reverse primers (10mM; 2.5 µl each), 2 µl of template cDNA and DNase-RNase free water.

PCR amplification protocol was carried out on a 7300 Real-Time PCR System (Applied Biosystems) using standard cycling conditions: 95°C (10 min) and 40 cycles of amplification (95°C (30 s), 60°C (1 min)) followed by a dissociation stage. All reactions were run in triplicate and included no template and no reverse transcription controls. Quantification results were expressed in terms of the cycle threshold (C_t) value. Barley actin was used as normalization control for experiment 2, while *OEP-16* was used for experiment 4.

Exp.	Gene name	RT primer	Forward primer	Reverse primer
2	hvu-miR168	GTCGTATCCAGAGCTGGGTCCGAGGTATTCGCTCTGGATACGACGTCCCG	TCGTTCGCTTGGTGCAGAT	GAGCTGGGTCCGAGGT
	hvu-miR166	GTCGTATCCAGTGCAGGGAGGGAGGTATTCGCACTGGATACGACGGAATG	TGTCGTCGGACCAGGCTT	GTGCAGGGAGGGAGGT
	bdi-miR167h	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACCCAGAT	TTTGATGAAGCTGCCAGCATG	**GTGCAGGGTCCGAGGT**
	hvu-miR159a*	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACTCATTG	GCGTGCGAGCTACTATCATTC	**GTGCAGGGTCCGAGGT**
	hvu-miR397	GTCGTATCCAGTCGAGGGTCCGAGGTATTCCGACTGGATACGACGTTCAT	TAGCTTTGAGTGCAGCGTTG	GTCGAGGGTCCGAGGT
	hvu-miR408	GTCGTATCCAGAGCTGGGTCCGAGGTATTCGCTCTGGATACGACGCCAGG	TCGCTTGCACTGCCTCTTC	GAGCTGGGTCCGAGGT
	hvu-miR399	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACCAGGGC	GTGCGTGCCAAAGGAGAGTT	**GTGCAGGGTCCGAGGT**
	bdi-miR396e	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACCCGTTC	CAACGGTCCACAGGCTTTCTT	**GTGCAGGGTCCGAGGT**
	osa-miR5079	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACCCCAAC	CGGGCGGTCCTAAATACTTGTT	**GTGCAGGGTCCGAGGT**
	Hv-Actin		GGCAGTGCTTTCCCTATATGCT	CATCACCAGAATCCAACACGAT
	Hv-Polyubiquitin		AGACCATCACGCTGGAGGTT	ATCTTCGCCTTCACATTGTCAA
	Hv-U61		GCAATGAGGAAACGAAACCT	AGAGGGTTGTGTTACCGACAC
	Hv-U59		TCTTCTGATGTGCTATCTCATGG	TCCTCAGTTCAGGTCTCCAAA
3	miR827	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACTGTTTG	GCGCGTTAGATGACCATCAG	**GTGCAGGGTCCGAGGT**
	miR166f	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACGGAATG	TGCGTTTCGGACCAGGCTT	**GTGCAGGGTCCGAGGT**
	tae-miR395	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACGAGTTC	CAAGGCTGAAGTGTTTGGGG	**GTGCAGGGTCCGAGGT**
4	osa-miR1432	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACGTCGGT	GGGCGATCAGGAGAGATGAC	**GTGCAGGGTCCGAGGT**
	tae-miR156	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACGTGCTC	CCGTCTCGTGACAGAAGAGAGT	**GTGCAGGGTCCGAGGT**
	tae-miR159a/b	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACCAGAGC	TGTGGTTTGGATTGAAGGGA	**GTGCAGGGTCCGAGGT**
	bdi-miR159-3p	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACAGAGCT	ACGAGCTTGGATTGAAGGG	**GTGCAGGGTCCGAGGT**
	ttu-miR160	GTCGTATCCAGAGCTGGGTCCGAGGTATTCGCTCTGGATACGACTGGCAT	TCTTGCCTGGCTCCCTGT	GAGCTGGGTCCGAGGT
	bdi-miR167c	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACTCAGAT	TTTGATGAAGCTGCCAGCATG	**GTGCAGGGTCCGAGGT**
	hvu-miR168	GTCGTATCCAGAGCTGGGTCCGAGGTATTCGCTCTGGATACGACGTCCCG	TCGTTCGCTTGGTGCAGAT	GAGCTGGGTCCGAGGT
	osa-miR169c-1	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACGCCGGCA	AACGCAGCCAAGGATGACTT	**GTGCAGGGTCCGAGGT**
	osa-miR169n	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACCAGGCA	TCGTGCGTAGCCAAGGATGACT	**GTGCAGGGTCCGAGGT**
	osa-miR171h	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACGAGTGA	TGTCGTGAGCCGAACCAATA	**GTGCAGGGTCCGAGGT**
	gma-miR393h	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACATCAAT	GACACTTCCAAAGGGATCGC	**GTGCAGGGTCCGAGGT**
	bdi-miR399b	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACCAGGGC	GTGCGTGCCAAAGGAGAGTT	**GTGCAGGGTCCGAGGT**
	osa-miR399d	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACCAGGGC	AAGCGTGCCAAAGGAGAATT	**GTGCAGGGTCCGAGGT**
	osa-miR444e	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACAAGCTT	TATCGTGCAGTTGCTGCCTC	**GTGCAGGGTCCGAGGT**
	bdi-miR5200	GTCGTATCCAGAGCTGGGTCCGAGGTATTCGCTCTGGATACGACAAGCCT	CGGCGGTGTAGATACTCCCTA	GAGCTGGGTCCGAGGT
	bdi-miR827-3p	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACTGTTTG	TTGTCCGTTAGATGACCATCAG	**GTGCAGGGTCCGAGGT**
	Ta-snoR10		GGATGAAACCTTCAAACAATCTTA	TCATTCACAGGAAAATAAGCACTG
	Ta-OEP16-3		GATGCACCTAGTTCCTTTGCAA	GCAAGATATTTATTCCAGGTCTGTTG

Table 3 *List of primers used for microRNA expression trend validations in the three experiments. Experiment number is reported (2 = barley, 3 = eu/alloplasmic lines, 4 = durum wheat). Reverse primers reported in bold share the same sequence. Normal type is for specific primers.*

miRNA name	Target Accession	RACE outer primer	RACE inner primer	qPCR forward primer	qPCR reverse primer
osa-miR444e	TC368997	CCTGGTGGATTAGACTTCCCTTCCT	GCTCCTCTCCCATCAGTTGCTTGT	TGGCGTGAATGCAACTTCAGAGA	TAGGCTGGAAAGCTCCTCTCCC

Table 4 *Primers used in the only successful 5'-RACE and qPCR validation of a microRNA target.*

Results – part I

In silico miRNA discovery

4.1 Introduction

MicroRNA discovery in plants has always been carried out following basically two different approaches: direct cloning and bioinformatics. They start with different data sets and use different rules, but work synergistically in the miRNA discovery process (Barrera-Figueroa *et al.*, 2013).

The common bioinformatic strategy consists in a homology search. Mature miRNAs sequences already identified in one species are used as a query in genomic sequences, expressed sequence tags (ESTs), or other databases from other species to find potential new miRNAs that match the known ones (Barrera-Figueroa *et al.*, 2013, and references therein).

Search based on evolutionary conservation has allowed the identification of miRNA families in many plant species, and has proved to be a powerful tool especially for those species where the complete genome sequence is not available. Another evidence, though indirect, of the existence of a newly identified microRNA consists in the identification of its target genes, therefore a complete bioinformatic search should not leave target analysis out of consideration.

For instance, our bioinformatic group published a study in which barley ESTs were screened for microRNAs and related to Unigene clusters, *i.e.* clusters containing sequences that appear to come from the same transcription locus, linking miRNA identification with target search (Colaiacovo *et al.*, 2010).

Besides homology search, another bioinformatic approach – *de novo* identification – has been often employed in miRNA discovery efforts in the last decade, with two different procedures. First, it is possible to identify all potential hairpin structures that exist in a genome, successively filtering the results on the basis of miRNA genes structural features (loop length, number of mismatches in miRNA:miRNA* duplex, GC content). Other groups instead looked for matches between intergenic regions and annotated protein coding gene transcripts, then evaluating these regions for characteristics of miRNA genes (Bonnet *et al.*, 2004; Wang *et al.*, 2004). This approach takes advantage of the fact that plant miRNAs and their target regions usually have nearly perfect match (Barrera-Figueroa *et al.*, 2013).

Finally, many cases are known of microRNAs coded by intragenic loci. As previously mentioned (see chapter 1.3.1) this feature is quite common in animal genomes but quite uncommon in plants, thus far.

We recently published a paper in which we provide the evidence that such a mechanism is spread also in plants. Exonic miRNAs were identified searching in EST databases belonging to 30 different plant species.

BIOLOGY DIRECT

REVIEWERS Open Access

On the complexity of miRNA-mediated regulation in plants: novel insights into the genomic organization of plant miRNAs

Moreno Colaiacovo†, Antonella Lamontanara†, Letizia Bernardo, Renzo Alberici, Cristina Crosatti, Lorenzo Giusti, Luigi Cattivelli and Primetta Faccioli*

Abstract

MicroRNAs (miRNAs) are endogenous small non-coding RNAs of about 20–24 nt, known to play key roles in post-transcriptional gene regulation, that can be coded either by intergenic or intragenic loci. Intragenic (exonic and intronic) miRNAs can exert a role in the transcriptional regulation and RNA processing of their host gene. Moreover, the possibility that the biogenesis of exonic miRNAs could destabilize the corresponding protein-coding transcript and reduce protein synthesis makes their characterization very intriguing and suggests a possible novel mechanism of post-transcriptional regulation of gene expression.

This work was designed to carry out the computational identification of putative exonic miRNAs in 30 plant species and the analysis of possible mechanisms involved in their regulation.

The results obtained represent a useful starting point for future studies on the complex networks involved in microRNA-mediated gene regulation in plants.

Keywords: Gene regulation, Exonic miRNA, miRNA self regulation, Plants

Findings

MicroRNAs (miRNAs) are endogenous small non-coding RNAs of about 20–24 nt that can be coded either by intergenic or intragenic loci.

In plants a few examples of intronic miRNAs have been found in Arabidopsis and rice. In Arabidopsis, miR402 has been identified within the intron of gene At1g77230 in the same orientation as the hosting pre-mRNA [1]. A putative mirtron, miR1429.2, has also been identified in rice by Zhu et al. [2] as part of a homeobox gene. In Arabidopsis the intronic miR838 is particularly interesting being located in the intron of Dicer-like 1 (*Dcl1*) gene (which codes the key enzyme involved in plant miRNA maturation,[3]), and thus enabling a self-regulatory mechanism that helps maintain DCL1 homeostasis [4].

The rice miR3981 was detected in the exon of a putative glyoxalase gene and its biogenesis pathway might be involved in the regulation of glyoxalase expression. Excessive levels of glyoxalase mRNA should lead to an increase in pre-miR3981 processing by the enzyme DCL1 and with glyoxalase also being a putative target for miR3981, an increase in the amount of miRNA3981 could lead to a decrease in glyoxalase mRNA levels, thus preventing durable high expression of the gene [5]. This suggests the possibility of a balanced competition, mediated by DCL1, between glyoxalase protein production and miR3981 biogenesis leading to a fine-tuning of glyoxalase expression in rice. The possibility that the biogenesis of exonic miRNAs destabilizes the corresponding protein-coding transcript and reduces protein synthesis makes their characterization very intriguing and suggests a new mechanism of post-transcriptional regulation of gene expression.

This work was designed to carry out the computational identification of putative exonic miRNAs in plants by taking advantage of the huge amount of ESTs available for many species. The analysis of putative miRNA target sites in the host mRNA was also performed in order to gain some insight into possible novel mechanisms of gene regulation.

TCs (Tentative Consensus), from 30 plant species, publicly available from the DFCI Gene Index Project

* Correspondence: primetta.faccioli@entecra.it
†Equal contributors
CRA-Genomics Research Centre, Agricultural Research Council, via S.Protaso 302, Fiorenzuola d'Arda, PC, Italy

(http://compbio.dfci.harvard.edu/tgi/), were computationally screened for the presence of miRNA precursors and miRNA target sites (Additional file 1 detailed methods can be found as additional file). TCs hosting miRNA precursors are reported in Additional file 2. Some of the annotated TCs refer both to a miRNA and a protein and thus represent the starting point for the identification of candidate exonic miRNAs.

The relationships between miRNAs and transcription factors (TFs) is particularly relevant, being the basis of complex regulatory networks [6,7] and several examples of TF genes hosting microRNA precursor sequence were found and are reported in aAdditional file 2: examples include AP-2, NAC and SPL coding genes. Interestingly, in some of the miRNA-containing TCs identified in this study, the miRNA host gene is also its conserved target, for example miR444, a monocot-specific miRNA identified in several monocots (barley, maize, purple false brome, rice, wheat), is part of TCs coding for MADS-box transcription factors (Additional file 2) and this family of transcription factors is also a well-known conserved target for miR444. Additional file 3 contains the example of the barley EST called GH228935 [GenBank: GH228935], which is a MADS-box-related sequence containing both the target site for miR444b (+/−) and the precursor sequence for miR444a (+/+).

All TCs or singletons that include both a miRNA precursor sequence and one or more miRNA target sites are presented in Table 1. TCs that include both a miRNA target site with +/− orientation and the miRNA precursor sequence with +/+ orientation on the coding strand can be referred to those cases where the miRNA is oriented on the same strand of the target gene. The presence of the target sequence upstream or downstream of the precursor sequence could affect both host gene post-transcriptional regulation as well as miRNA regulation or self-regulation in the case that the target site belongs to the same miRNA family as the miRNA precursor. The latter mechanism could be a further buffering system for modulation of miRNA expression and could be in accordance with the model of miRNA self-regulation proposed by Meng et al. [6,7] that suggests a feedback model in which the miRNA* binds to the complementary sites on their precursor thus exerting a cleavage-based modulating role. Examples of putative self-regulation based on the presence of a target site upstream or downstream of the precursor have been found in our data for miR444 as reported above, miR395 (rice), miR159/319 (soybean), miR1118 (wheat) and miR1219 (moss).

On the other hand, those TCs where the orientation of miRNA precursor sequence with respect to coding strand is +/− should refer to those miRNAs coded in the antisense orientation with respect to the host gene. Antisense transcription associated with miRNA target mRNAs have been previously discovered in Arabidopsis [8] and may serve as a link between miRNA and RNA silencing pathways, suggesting that miRNAs may have additional roles in post-transcriptional regulation that are independent of cleavage.

TCs or singletons coding for more than one miRNA precursor (miRNA clusters) have also been identified both in the +/+ and +/− orientations (sense/antisense miRNAs) and are summarized in Additional file 4. The occurrence of pairs of antisense miRNAs in plants has been previously reported [9]. Most of the miRNA clusters highlighted in plants contain several members of a specific miRNA family rather than encoding miRNA with unrelated sequences, as is frequently the case for animal miRNA clusters [10]. Examples of plant miRNA clusters with members of different families are reported in Additional file 4.

Among the exonic miRNAs identified, we chose to explore, as an interesting example, the relationship between miR1118 and calmodulin-coding genes in wheat (Table 1). Calmodulin genes are of particular interest due to the fundamental role of calmodulin in the calcium-dependent regulation during plant response to endogenous and exogenous stimuli [11]. Calmodulin-binding proteins have already been shown to be targeted by miRNAs in plants [12], while no report is available on miRNA targeting to calmodulin itself.

Many plant species possess calmodulin (CaM) multigene families composed of several genes: some of which encode identical proteins, while others encode different isoforms of calmodulin-related proteins. In wheat, at least 13 CaM-related genes have been identified and classified into four subfamilies (SF-1:4 on the basis of their nucleotide sequence) coding for three distinct isoforms named TaCaM-I, TaCaM-II and TaCaM-III [13]. Genes from SF-1, SF-3 and SF-4 encode for the same isoform (TaCaM-I, the most similar to calmodulin from other plant species and organisms), while SF-2, the most divergent calmodulin subfamily with most recent origin in wheat, encodes at least two isoforms (TaCaM-II and TaCaM-III).

To deeply characterize the calmodulin gene family for the presence of miRNA-coding sequences, ten cDNA clones corresponding to 10 different genes encoding calmodulin-like proteins were screened against the wheat miRNA mature sequences present in miRBase. Each cDNA clone contained at least part of the 5′UTR, the full coding region, the 3′-UTR and the polyA tail. Blast output is reported in Additional file 5. The most significant match was found for tae-miR1118 mature sequence (score 43, e-value 2e-006), furthermore, an additional miRNA (tae-miR1125 mature sequence) showed a very significant match (score 39; e-value 2e-005). Both miRNAs were found in the two genes belonging to SF-2 (*TaCaM2-1* and *TaCaM2-2*), but not in the third sequence belonging to the same subfamily (*TaCaM2-3*). Mature sequences for both miRNAs are included in the 3′UTR regions of the two calmodulin

Table 1 Reports those TCs or singletons that include both a miRNA precursor sequence and one or more miRNA target sites

Species name	Scientific name	Sequence identifier	Annotation from DFCI Gene Index	miRNA family	Target site (+/-)	Precursor site
Arabidopsis	*Arabidopsis thaliana*	TC360779	pentatricopeptide (PPR) repeat-containing protein	5652	1102–1122	261-708 (+/-)
				161	1093–1113	
				400	1314–1334	
		TC383702		5636		89–166 (+/–)
				173	444–465	
Barley	*Hordeum vulgare*	GH228935	MIKC-type MADS-box transcription factor WM32A	444	246–266	280–393 (+/+)
Cotton	*Gossypum hirsutum*	TC244560		2949		366–547 (+/–)
						399–511 (+/–)
				3476	469–488	
		TC273521	Uncharacterized protein YJR115W	2949		1544–1723 (+/–)
						1576–1689 (+/–)
				3476	1647–1666	
Maize	*Zea mays*	TC498468	Transcriptional regulator, TetR family	168		42–145 (+/+)
				396	814–833	
Moss	*Physcomitrella patens*	NP13127877		1219	535–555	263–414 (+/–)
Medicago truncatula	*Medicago truncatula*	EX529600		2629		171–231 (+/–)
				2627	173–194	
		NP7251801		5241		1186–1260 (+/+)
				1510	674–694	
				5242	1341–1361	
				2678	86–106	
		NP7252398		5241		1163–1280 (+/+)
						1186–1260 (+/+)
						1189–1255 (+/+)
				1510	674–694	
				5242	1341–1361	
				2678	86–106	
		NP7252404		5241		1189–1255 (+/+)
				1510	674–694	
				2678	86–106	
		NP7254391		2674		467–760 (+/+)
				2089	200–221	
Moss	*Physcomitrella patens*	TC19973	RNA polymerase II largest subunit	899		323–544 (+/+)
				2083	416–436	
					217–236	
		TC24438	Predicted protein	1062		378–561 (+/+)
				529	581–600	

Table 1 Reports those TCs or singletons that include both a miRNA precursor sequence and one or more miRNA target sites *(Continued)*

		TC27593	Ribulose bisphosphate carboxylase small chain	899		48–269 (+/–)
				2083	354–374	
					154–174	
Rice	*Oryza sativa*	CB677501	Os05g0439200 protein	156/157		219–338 (+/+)
						252–352 (+/+)
				168	217–236	
		CX099912		156		187–351 (+/+)
						220–320 (+/+)
				168	205–224	
		EG709456	Conserved protein	2118		109–285 (+/+)
				1869	239–259	
		CA760441		395	539–559	373–441 (+/–)
						230–397 (+/–)
						87–154 (+/–)
Soybean	*Glycine max*	BW657280	Transcriptional regulator, MarR family	4414		164–291 (+/+)
				159/319	123–144	
		TC425684	Chromosome chr14 scaffold_190, whole genome shotgun sequence	4409		479–539 (+/+)
				5372		346–410 (+/+)
				5674	373–393	
		TC428562	Transcriptional regulator, MarR family	4414		563–690 (+/+)
				319	522–543	
		TC442024	MATE efflux family protein	159/319	324–343	358–460 (+/+)
		TC443643		5668		717–829 (+/–)
				1510	697–717	
		TC451566		5038		424–574 (+/+)
						482–556 (+/+)
				5670	567–586	
Spruce	*Picea abies*	EX343185		3700		281–358 (+/+)
				947	215–236	
		EX369017		3700		217–294 (+/+)
				947	151–172	
Wheat	*Triticum aestivum*	TC369361	Calmodulin TaCaM2-1	1118	727–749	918–1057 (+/+)
		TC370181	Calmodulin TaCaM2-3	1118	644–666	835–974 (+/+)

genes. Although the coding sequences of *TaCaM2-1* and *TaCaM2-2* are substantially different and produce two distinguishable calmodulin isoforms (TaCaM-III and TaCaM-II, respectively), the corresponding 3′UTR regions differ only by one nucleotide. The nucleotide sequence alignment between the 3′ UTR regions of the three calmodulin genes belonging to SF-2 revealed a striking difference between *TaCaM2-3* and the other two genes: *TaCaM2-1* and *TaCaM2-2* share a long insertion of 239 bp that is absent in *TaCaM2-3* (Figure 1 and Additional file 6). A Blast search against the complete TREP database, containing a collection of repetitive DNA sequences from different *Triticeae* species, highlighted a strong match (score 315, e-value 1e-86) of the inserted sequence with a DNA transposon of the Mariner class called DTT_Polyphemus_464G14-3. This finding is in agreement with a possible model for the evolution of MIR genes that proposes their origin from transposable elements, on the basis of the observation that the DNA-type nonautonomous elements called MITE (miniature inverted-repeat transposable element) can fold into imperfect stem loops that are typical of miRNA precursors [3]. Several pieces of evidence pointed to plant *mariner*-like elements (MLEs) as the autonomous partners of the nonautonomous *Stowaway* MITEs [14].

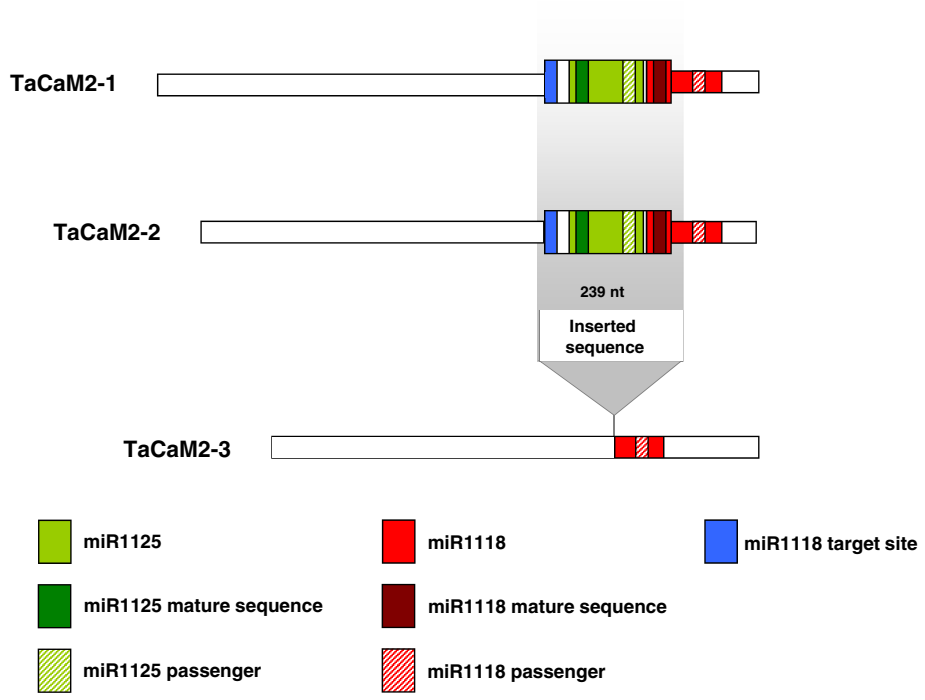


Figure 1 Reports the positions of miR1118 and miR1125 coding sequences inside wheat SF-2 calmodulin genes (*TaCaM2-1*, *TaCaM2-2*, *TaCaM2-3*).

The presence of the transposon-related insertion in calmodulin genes from other plant species has was investigated, and a blast search using the sequence of *TaCaM2-2* gene as query was performed. The results are reported in Additional file 7: a significant match with the 3′UTR region of the wheat gene was detected only with a barley gene, where, however, the transposon-related sequence was absent.

Moreover, while the miR1125 precursor sequence deposited in the miRBase found a significant match (81% identity) entirely included in the inserted sequence, the miR1118 precursors sequence matched (100% identity) only partially into the insertion while it spanned also the very last part of the 3′UTR of the *TaCaM* genes (Figure 1). The reverse complement sequence for miR1118 found inside the 239 nt insertion could thus represent a target site for the same miRNA. This is very possible since only 3 mismatches were found when compared with the mature sequence, and the target site is also predicted by psRNAtarget. A second reverse complement site is also shared by all three SF-2 genes (*TaCaM2-1*, *TaCaM2-2* and *TaCaM2-3*), being located in the last part of the 3′ UTR region. This site corresponds to the miRNA* of the precursor deposited in the miRBase and it can be considered a less likely target site due to the high number of mismatches with the mature sequence. However, the prediction of the RNA secondary structures by mFold software (Additional file 8) highlighted the worse energetic stability of the miR1118 precursor stored in miRBase compared to that of the inserted sequence (−32.5 kcal/mol vs.−117.6 kcal/mol). Although the transposable element is longer than the deposited precursor (239 nt vs. 140 nt) it has a Minimal Folding Energy Index (MFEI) higher than the miRNA precursor (1.16 vs. 0.59) and it is in better agreement with the values known for plant miRNAs. The existence of two alternative miR1118 precursors could thus be possible. The inserted sequence has acquired some mutations since the insertion event and these variations have led to a better stability of the candidate miRNA precursor, thus confirming this hypothesis. In fact, compared with the inserted sequence, the predicted secondary structure of the transposable element stored in the TREP database shows a worse stability (−84.4 kcal/mol), a lower MFEI (0.88) and a higher number of mismatches (4) between miRNA and miRNA*.

Conclusions

Recently Meng et al. [6,7] reported on the high dynamicity of the regulatory activities mediated by miRNAs which are themselves strictly regulated, thus making the miRNA-involved networks more robust.

The results presented in this work can be considered as a valuable support for future studies on the complex

networks involved in miRNA-mediated gene regulation in plants. In particular the relationships between miRNAs and transcription factors (TFs) represent a key node of these complex networks. Most of the currently available studies have been involved in the identification of TF recognition sites in the promoter of miRNA genes and in the identification of miRNA target sites in TF coding genes. The present work adds a novel level of complexity pointing out the possibility that miRNA coding sequences can also be hosted by TF genes.

Additional files

Additional file 1: Reports the results of blasting each species-specific Gene Index against the corresponding species-specific miRNA precursor sequences.

Additional file 2: Reports barley EST named GH228935 which includes both the target site for miR444b (+/−) and the precursor sequence for miR444a (+/+).

Additional file 3: Reports TCs or singletons coding for more than one miRNA precursor (miRNA clusters) both in the +/+ and +/− orientations (sense/antisense miRNAs).

Additional file 4: Blast results for ten wheat cDNA clones (corresponding to 10 different genes encoding calmodulin-like proteins) against the wheat miRNA mature sequences present in miRBase.

Additional file 5: Sequences of wheat SF-2 calmodulin genes (*TaCaM2-1*, *TaCaM2-2*, *TaCaM2-3*). Precursors, mature and reverse complement sequences are reported for miR1118 and miR1125.

Additional file 6: Blast results related to the presence of the transposon-related insertion in calmodulin genes from plant species other than wheat using the sequence of *TaCaM2-2* gene as query.

Additional file 7: Reports the predicted secondary structures of the three miRNA precursors putatively coded by the calmodulin mRNA: the miR1118 precursor identical to the one present in miRBase, the miR1118 and miR1225 precursors putatively coded by the transposon insertion.

Additional file 8: Reports detailed on the adopted computational methods [15-17].

Competing interests
The authors declare that they have no competing interests.

Acknowledgements
We thank Keith Anthony Grimaldi for helping with the preparation of the manuscript. This work has been supported by "Mappa 5A" and "Drupomics" projects (Italian Ministry for Agriculture).

Authors' contribution
PF conceived the idea and wrote the paper. MC and AL did most of the computational work and were supported by LB. RA gave the informatics support. LC, CC, LG contributed to the interpretation of the results and to the preparation of manuscript. All authors read and approved the final manuscript.

Reviewers' comments

Reviewer's report 1
Dr Alexander Max Burroughs (nominated by Dr L Aravind)

The authors probe TCs and identify several miRNA precursors. What is missing from this analysis is a sense of the novelty of the findings. How many of these precursors have previously been predicted as being present at in the genome at the identified locations? For reasonably well-annotated genomes like Arabidopsis or rice it should be relatively easy to look at this to provide an overview of how many times novel locations are identified for known miRNAs. If this count is large, it may be interesting to discuss why these have not been identified in the past.

Author response: *In order to clarify this aspect, we checked all the miRNA genomic locations as they are reported in miRBase, and we compared them with the location of the identified TCs. We performed the analysis only for Arabidopsis sequences, which are supposed to be better annotated. Almost all pairs miRNA-TC show overlapping locations, meaning that the two pieces of information were previously known to the scientific community. However, the novelty of our work is that we merged these existing data and we were able to extract novel findings which are relevant for the miRNA analysis, e.g. the characterization of these miRNA precursors as exonic miRNAs.*

As an additional form of evidence it might be useful, when available, to map deep-sequencing data to the TCs in order to confirm the presence and active processing of the identified precursors, particularly any identified in novel locations. In addition, one limitation in the analysis is that it precludes detection of previously unidentified miRNA precursors, an angle which could be addressed by feeding the deep-sequencing data mappings into a miRNA prediction program (based on parameters like mature/star sequence count distribution, folding properties, etc.) like miRanalyzer or miRdeep. This could increase the number of transcripts potentially harboring exonic miRNA precursors.

Author response: *Small RNA sequencing is an interesting further development for this work and we plan to have new data in the near future which will be part of another publication. At the moment we are in the process of sequencing small RNAs from several plant species, and this will give us the opportunity to study also novel microRNAs.*

I was struck by the general lack of conservation in the locations of the identified precursors: few seemed to be located in homologous transcripts across different plant species. While the authors discuss an example of conservation of the miR444 precursor, it might be informative to give an overview of the general conservation/lack of conservation observed in the localization of precursors in host transcripts. Are transcripts harboring both the precursor and the miRNA target sequence more likely to be conserved? This could provide some general insight into the mobility of the precursors and their role in "cis-regulation" of transcripts.

Author response: *According to our analysis, there are few miRNA precursors which are present in conserved transcripts across different species. Of all the miRNA families which were identified in multiple species, approximately 20% show host conservation in at least two species. The situation is similar if we consider transcripts harboring both the precursor and the target site. A possible explanation for this finding could be the different level of completeness of species-specific gene indices (e.g. accuracy of TC annotations).*

The authors note "On the other hand, those TCs where the orientation of miRNA precursor sequence with respect to coding strand is +/− should refer to those miRNAs coded in the antisense orientation with respect to the host gene. Antisense transcription associated with miRNA target mRNAs have been previously discovered in Arabidopsis and may serve as a link between miRNA and RNA silencing pathways. . ." This is a very interesting observation and quite consistent with a growing number of studies linking antisense transcription at gene loci to RNAi pathways. It might be worthwhile to scan existing gene annotations for evidence of downstream, noncoding antisense transcription which could overlap the gene body in +/− cases; this could provide some global evidence in support of the statement.

Author response: *We manually inspected all the +/− cases which we identified in Arabidopsis (chosen for the same reasons reported above, i.e. the accuracy of genome annotation), then exploring the TAIR genome browser and looking for evidences of antisense transcription at those locations. We found evidence for antisense transcription in 38% of the cases.*

Outside of the analysis of the calmodulin transcripts, there was little discussion of the predicted folding energies for the identified precursors. It

could be instructive to see how secondary structure stabilities in the identified precursor compare to other precursors in individual genomes, foremost as a gauge of the reliability of the identified precursors, particularly those precursors identified in novel locations (see first comment).

Author response: *The criteria which we used for the prediction of miRNA precursors are very stringent, since we selected as putative miRNAs only those sequences which have a nucleotide identity greater than 95% with a known precursor. Because of this, we don't expect big changes in the energetic stability of the identified sequences compared to the ones stored in miRBase. The variations are supposed to be minimal and should not affect dramatically the miRNA stability. However, in order to confirm our expectation, we selected some miRNAs at random and we compared their energy values to the corresponding sequence stored in miRBase, by using the mFold software. The differences are small: ath-MIR416 has a free energy equal to −27.9 kcal/mol if we consider the miRBase sequence, and −31 kcal/mol if we consider our sequence; the same is true also for ath-MIR398b (−51.5 vs −51.4), ath-MIR319b (−81.3 vs −80.4) and osa-MIR439d (−57.8 vs −60.8).*

Reviewer's report 2
Dr Raya Khanin (nominated by Arcady Mushegian)

Results are given as flat files. The paper would benefit from organizing the results in a searchable database.

Author response: *We are planning to set up a searchable database with all the relevant information.*

Received: 2 February 2012 Accepted: 8 May 2012
Published: 8 May 2012

References

1. Sunkar R, Zhu JK: **Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis.** *Plant Cell* 2004, **16:**2001–2019.
2. Zhu Q-H, Spriggs A, Matthew L, Fan L, Kennedy G, Gubler F, Helliwell C: **A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains.** *Genome Res* 2008, **18**(9):1456–1465.
3. Voinnet O: **Origin, biogenesis and activity of plant microRNAs.** *Cell* 2009, **136**(4):669–687.
4. Rajagopalan R, Vaucheret H, Treio J, Bartel DP: **A diverse and evolutionary fluid set of microRNAs in *Arabidopsis thaliana*.** *Gene Dev* 2006, **20** (24):3407–3425.
5. Li T, Li H, Zhang YX, Liu JY: **Identification and analysis of seven H2O2-responsive miRNAs and 32 new miRNAs in the seedlings of rice (Oryza sativa L. ssp indica).** *Nucleic Acids Res* 2011, **39**(7):2821–2833.
6. Meng Y, Shao C, Wang H, Chen M: **The regulatory activities of plant microRNAs: a more dynamic perspective.** *Plant Physiol* 2011, **157**(4):1583–1595.
7. Meng Y, Shao C, Chen M: **Toward microRNA-mediated gene regulatory networks in plants.** *Brief Bioinform* 2011, **12**(6):645–659.
8. Luo QJ, Samanta MP, Köksal F, Janda J, Galbraith DW, Richardson CR, Ou-Yang F, Rock CD: **Evidence for antisense transcription associated with microRNA target mRNAs in Arabidopsis.** *PLOS Genet* 2009, **5**(4):e1000457.
9. Zhang B, Pan X, Stellwag EJ: **Identification of soybean microRNAs and their targets.** *Planta* 2008, **229**(1):161–182.
10. Axtell MJ, Snyder JA, Bartel DP: **Common functions for diverse smallRNAs of land plants.** *Plant Cell* 2007, **19**(6):1750–1769.
11. Kim MC, Chung WS, Yun DJ, Cho MJ: **Calcium and calmodulin-mediated regulation of gene expression in plants.** *Mol Plant* 2009, **2**(1):13–21.
12. Manila TM, Riju A, Lakshmi Priya Darshini K, Chandrasekar A, Eapen SJ: **In silico microRNA identification from paprika (*Capsicum annuum*) ESTs.** *Nature Precedings* 2009, http://hdl.handle.net/10101/npre.2009.3737.1.
13. Yang T, Segal G, Abbo S, Feldman M, Fromm H: **Characterization of the calmodulin gene family in wheat: structure, chromosomal location and evolutionary aspects.** *Mol Gen Genet* 1996, **252**(6):684–694.
14. Feschotte C, Swamy L, Wessler SR: **Genome-wide analysis of *mariner*-like transposable elements in rice reveals complex relationships with *Stowaway* Miniature Inverted Repeat Transposable Elements (MITEs).** *Genetics* 2003, **163**(2):747–758.
15. Dai X, Zhao PX: **psRNATarget: a Plant Small RNA Target Analysis Server.** *Nucleic Acids Res* 2011, Web Server Issue.
16. Kozomara A, Griffiths-Jones S: **miRBase: integrating microRNA annotation and deep-sequencing data.** *Nucleic Acids Res* 2011, **39:**D152–D157.
17. Zuker M: **Mfold web server for nucleic acid folding and hybridization prediction.** *Nucleic Acids Res* 2003, **31:**3406–3415.

doi:10.1186/1745-6150-7-15
Cite this article as: Colaiacovo *et al.*: On the complexity of miRNA-mediated regulation in plants: novel insights into the genomic organization of plant miRNAs. *Biology Direct* 2012 **7**:15.

Results – part II

NGS-based miRNA discovery

5.1 General introduction

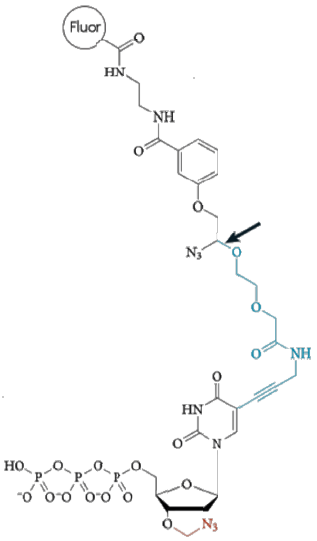
Before the beginning of the Next Generation Sequencing (NGS) era, the most common approach to the discovery of new miRNAs was to clone and sequence individual small RNAs using traditional molecular methods. The basic procedure included mapping small RNAs to reference genomic sequences, examining the secondary structure of candidate precursors, and checking for precise cleavage of precursors (Meyers *et al.*, 2006; Barrera-Figueroa *et al.*, 2013)

Such a procedure was quite laborious and expensive considered the amount of information gained (throughput). A great step forward was made possible by next-generation sequencing techniques, which led to the identification of many microRNAs in a great number of species, and allowed a precise profiling of their expression trend as well.

Small RNA sequencing begins with the preparation of a small RNA cDNA library from RNA samples of interest followed by the 'massively parallel' sequencing of millions of individual cDNA molecules from the library. Bioinformatic analysis of the sequence reads identifies both known and novel miRNAs in the data sets and provides relative quantification using a digital approach (that is, the number of sequence reads for a given miRNA relative to the total reads in the sample is an estimate of relative abundance of the miRNA). (Pritchard *et al.*, 2012).

From mid 2000s lots of sequencing platforms were developed, in order to improve the sequencing quality (*i.e.* reducing sequencing errors), to minimize costs and increase the final throughput. Different platforms rely on different sequencing techniques, each of which presents its own advantages and drawbacks (reviewed in Mardis, 2008; Metzker, 2010; Mardis, 2013). Illumina technology has provided one of the most successful and widely-adopted next-generation sequencing platforms worldwide.

It utilizes a "sequencing by synthesis" approach in which all four nucleotides are added simultaneously to the flow cell channels, along with a DNA polymerase, for incorporation into the oligo-primed cluster fragments. Specifically, the nucleotides carry a base-unique fluorescent label and the 3'-OH group is chemically blocked, such that each incorporation is a unique event (diagram on the left). The blocking group attached to the 3' end causes a bias against incorporation with DNA polymerase. A mutated DNA polymerase is required to facilitate the incorporation of 3'-blocked terminators. In the figure, arrows indicate the site of cleavage separating the fluorophore from the nucleotide, and the blue chemical structures denote residual linker structures that are attached to the base and accumulate with subsequent cycles.



An imaging step follows each base incorporation step (Figure 10), during which each flow cell lane is imaged by the instrument optics. After each imaging step, the 3' blocking group is chemically removed to prepare each strand for the next incorporation by DNA polymerase. This series of steps continues for a specific number of cycles, as determined by user-defined instrument settings, and allows discrete read

lengths of up to 200-300 bases. A base-calling algorithm assigns sequences and associated quality values to each read and a quality checking pipeline evaluates the Illumina data from each run, removing poor-quality sequences (Mardis, 2008).

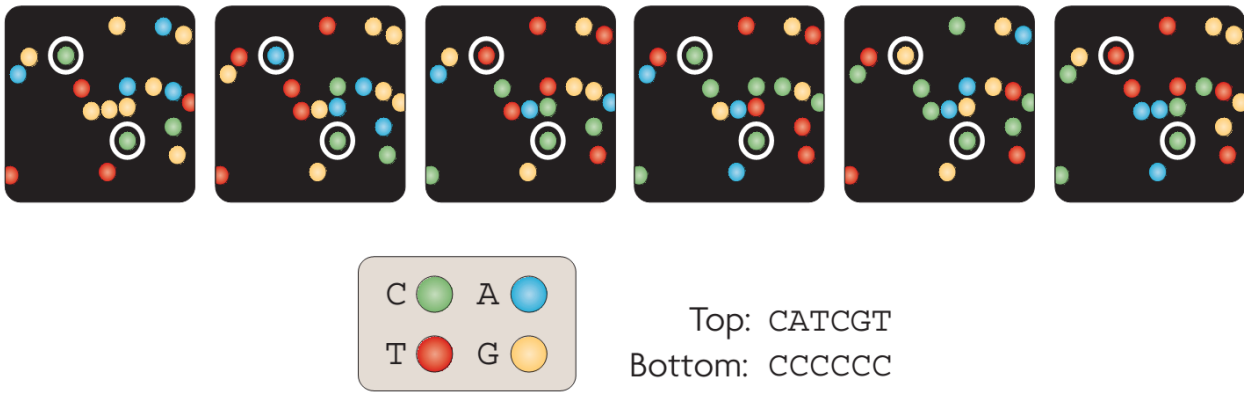


Fig. 10 *Exemplification of Illumina imaging technology. The four-colour images highlight the sequencing data from two clonally amplified templates. Each square represents one cycle of sequencing, in which only one base can be incorporated on the same cluster.*
From Metzker, 2010.

As for this thesis, miRNA discovery and expression profiling were carried out with an Illumina® Genome Analyzer II_x. Several runs were made, as summarized in Chapter 3, and four different experiments were considered.

5.2 Experiment 1 - *Peach and development*

5.2.1 Introduction

Peach (*Prunus persica*) has distinct advantages that make it suitable as a model species for genomics study in the *Rosaceae* family, which includes as well a number of economically important fruit trees such as apple, cherry and plum. Peach is diploid with n=8 and has a comparatively small genome (it is currently estimated to be ~230 Mb). It has a relatively short juvenility period of 2-3 years compared to most other fruit tree species that require 6-10 years. In addition, a great number of ESTs is available, and many genes for important traits have been genetically described, e.g. for the control of flower and fruit development, tree growth habit, dormancy, cold hardiness, and disease and pest resistance.

The first release of the peach genome was made available in 2010, as crowning achievement of the efforts of the International Peach Genome Initiative (IPGI) consortium. Early online access to the draft assembled and annotated peach genome was allowed prior to peer-reviewed publication of the data, which arrived only three years after (Verde *et al.*, 2013). These years, from 2010 to 2014, the concern about this species started to rise, and several papers on peach microRNAs appeared.

Peach microRNA discovery started with bioinformatic analyses based on ESTs only (Zhang *et al.*, 2012a), and throve with the genome availability. Computational *de novo* identification (exploiting miRNA-target complementarity) (Gao *et al.*, 2012b), and homology searches (Zhenlin, 2013) led to a significant advance in our knowledge of the peach microRNA population. Deep sequencing studies coupled with degradome analysis, aided by subsequent read alignment on the genome, were carried out (Zhu *et al.*, 2012; Luo *et al.*, 2013) as well. Other groups decided not to focus on discovery only, moving to the profiling of microRNAs involved in development (Bonghi *et al.*, 2011), drought stress (Eldem *et al.*, 2012), chilling requirement and cold stress (Bakarat *et al*., 2012).

In this scenario, starting from a deep-sequencing run, we highlighted the remarkable complexity of the peach miRNome, in terms of sequence and length variability of microRNAs among different tissues and developmental stages.

IsomiRs, the population of mature variants from what is generally regarded as the “reference” sequence, can result from alternative DCL cleavage, nucleotide additions or deletions to 3’ ends, trimming of 5’ ends, or post-transcriptional editing resulting in base substitutions (Morin *et al.*, 2008; Pritchard *et al.*, 2012; Jeong *et al.*, 2013). They can greatly contribute in broadening cell regulatory network, since these modifications can have substantial effects on miRNA functions. Additions or deletions can first of all cause a shift of the seed region, typically made up by nucleotides 2-8 of the miRNA, which is the predominant determinant of target selection. Secondly, alteration of 3’ ends may affect miRNA stability and mRNA targeting efficacy. Thirdly, it is known that different AGO proteins show preference for microRNAs with a certain length (e.g. 21-nt miRNAs are loaded on AGO1 RISCs, whereas 24-nt miRNAs are redirected to AGO4 RISCs) and 5’-terminal residues.

frontiers in
PLANT SCIENCE

ORIGINAL RESEARCH ARTICLE
published: 26 July 2012
doi: 10.3389/fpls.2012.00165

A survey of microRNA length variants contributing to miRNome complexity in peach (*Prunus persica L.*)

Moreno Colaiacovo†, Letizia Bernardo†, Isabella Centomani, Cristina Crosatti, Lorenzo Giusti, Luigi Orrù, Gianni Tacconi, Antonella Lamontanara, Luigi Cattivelli and Primetta Faccioli*

CRA Genomics Research Centre, Fiorenzuola d'Arda, Italy

Edited by:
Takuji Sasaki, National Institute of Agrobiological Sciences, Japan

Reviewed by:
Zhixi Tian, Chinese Academy of Sciences, China
Takeshi Itoh, National Institute of Agrobiological Sciences, Japan

***Correspondence:**
Primetta Faccioli, CRA Genomics Research Centre, via S.Protaso 302, I-29017 Fiorenzuola d'Arda (Pc), Italy.
e-mail: primetta.faccioli@entecra.it

†*Moreno Colaiacovo and Letizia Bernardo have contributed equally to this work.*

MicroRNAs (miRNAs) are short non-coding RNA molecules produced from hairpin structures and involved in gene expression regulation with major roles in plant development and stress response. Although each annotated miRNA in miRBase (www.mirbase.org) is a single defined sequence with no further details on possible variable sequence length, isomiRs – namely the population of variants of miRNAs coming from the same precursors – have been identified in several species and could represent a way of broadening the regulatory network of the cell. Next-gen-based sequencing makes it possible to comprehensively and accurately assess the entire miRNA repertoire including isomiRs. The aim of this work was to survey the complexity of the peach miRNome by carrying out Illumina high-throughput sequencing of miRNAs in three replicates of five biological samples arising from a set of different peach organs and/or phenological stages. Three hundred-ninety-two isomiRs (miRNA and miRNA*-related) corresponding to 26 putative miRNA coding loci, have been highlighted by mirDeep-P and analyzed. The presence of the same isomiRs in different biological replicates of a sample and in different tissues demonstrates that the generation of most of the detected isomiRs is not random. The degree of mature sequence heterogeneity is very different for each individual locus. Results obtained in the present work can thus contribute to a deeper view of the miRNome complexity and to better explore the mechanism of action of these tiny regulators.

Keywords: microRNA, isomiRs, next generation sequencing

INTRODUCTION

MicroRNAs (miRNAs) are short non-coding RNA molecules produced from hairpin structures and involved in gene expression regulation with major roles in plant development and stress response. MiRNAs are transcribed into a primary transcript which folds into a bulge with stem-loop conformation that is then cleaved by a Dicer-like (DCL) RNase III enzyme named DCL1. The cleavage results in a short duplex: one of the two strands forming the duplex and designated as miRNA* is then typically degraded while the other strand is incorporated into the RNA-induced silencing (RISC) complex where it mediates mRNA recognition and cleavage or translational repression (Jones-Rhoades et al., 2006; Voinnet, 2009; Xie et al., 2010).

Although each annotated miRNA in miRBase[1] is a single defined sequence with no further details on possible variable sequence length, isomiRs – namely the population of variants of miRNAs coming from the same precursors – have been identified in several species and could represent a way of broadening the cell regulatory network (Ebhardt et al., 2009; Guo and Lu, 2010).

Vaucheret (2009) demonstrated the biological significance of mature miRNA length heterogeneity in *Arabidopsis* where the athmiR168 can be processed as two different miRNA isoforms of 21 nt and 22 nt in length with different activities on AGO1 homeostasis (AGO1 is the Argonaute1 protein, a component of RISC complex that catalyzes broad miRNA- and siRNA-guided mRNA cleavage and translation repression Voinnet, 2009).

Alteration in miRNA end sequences can have strong effects on miRNA function due to the fact that the identity of the first 5′ nucleotide is the major determinant for AGO protein association (Takeda et al., 2008). As an example, Mi et al. (2008) found that AGO1 (which predominates in the miRNAs-mediated pathway) harbors miRNAs that favor a 5′ terminal uridine. A change at the 5′ terminal nucleotide of a miRNA predictably redirected it into a different AGO complex and altered its biological activity. Additionally, it was reported that the thermodynamic stability at the 5′ end of the strand is likely to affect the loading in the AGO complex (Eamens et al., 2009).

An accurate profile of the entire miRNA population of a biological sample provides useful information on miRNA activity and it can be used to compare the distribution of miRNA sequence variants in different samples. In fact, although the distribution of isomiRs across samples has been previously shown to be generally similar, examples in which the dominant isomiR is different from sample to sample have been found in animals (Lee et al., 2012) and could imply a functional role for specific isomiR sequences, besides affecting the accuracy and consistency of miRNA measurement.

This work aims to survey, by carrying out Illumina high-throughput sequencing, the complexity of peach miRNome

[1] www.mirbase.org

through the analysis of the miRNA population of a set of samples representative of different tissues and developmental stages.

MATERIALS AND METHODS

PLANT MATERIAL AND RNA EXTRACTION

A 12-year-old tree grafted on wild seedling of the yellow-fleshed cv. Maycrest (*Prunus persica* (L.) Batsch), grown in Palazzolo di Sona, Verona, Italy (45.457°N, 10.822°E), was used as plant source material. Each sample was collected pooling together material from three different branches of the same plant. Four phenological stages (Chapman and Catlin, 1976) were considered: swollen bud, half-inch green, pink, bloom. Leaf and flower swollen buds were collected 41 days before flowering (DBF), half-inch leaves were collected 21 DBF, pink flower buds were collected six DBF. Codes were assigned to each samples: BF, pink; F, bloom; GF, swollen flower bud; O, half-inch green; GL, swollen leaf bud. Tissues were frozen in liquid nitrogen immediately after drawing. Total RNA was extracted from three independent samples with the Plant Total RNA Purification Kit (NORGEN Biotek Corp., Thorold, ON, Canada) following manufacturer instructions. RNA quality and concentration were evaluated with the Agilent 2100 Bioanalyzer RNA 6000 Nano assay (Agilent Technologies, Santa Clara, CA, USA).

SMALL RNA LIBRARIES CONSTRUCTION AND SEQUENCING

Preparation of small RNA libraries was performed with the TruSeq Small RNA Sample Prep Kit (Illumina, San Diego, CA, USA) following manufacturer instructions. Briefly, 1 μg of total RNA was ligated with adapters at 3′ and 5′ ends, without any size fractionation. Adapter-ligated RNA was reverse-transcribed with SuperScript II Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA), then PCR-amplified (15 cycles). Samples were barcoded using 15 variants of the reverse primer provided with the kit. Libraries were pooled together and then purified on a 6% TBE PAGE gel after electrophoresis. Libraries quality and concentration were evaluated with the Agilent 2100 Bioanalyzer DNA 1000 assay. The obtained cDNAs were sequenced using the Illumina Genome Analyzer IIx.

DATA ANALYSIS

Reads were filtered with UEA sRNA plant toolkit[2] (Moxon et al., 2008) to remove adaptor sequences, reads shorter than 18 nt or longer than 24, low-complexity reads, reads matching rRNAs or tRNAs and reads that did not match the peach genomic sequence available at "The International Peach Genome Initiative – www.rosaceae.org/peach/genome" (only those sequences with a full-length perfect match to the selected genome were retained). Reads from one replicate (randomly chosen) of each biological sample were then analyzed with the software miRDeep-P (Yang and Li, 2011, default parameters) to identify miRNA loci expressed in all the five tested samples (GF, GL, B, F, O). Reads associated to these loci were then also screened in the remaining two replicates.

Read counts for each variant were divided by the total number of reads with a match in the peach genome in each sample and normalized to 1,000,000 reads. Reads that could be related to more than one locus were assigned by MiRDeep-P to all possible involved loci.

IsomiRs for each putative locus were blasted against miRBase (release 18) to search for the loci related to previously known miRNAs (Kozomara and Griffith-Jones, 2011). Blast vs. mature sequences was based on the following parameters: outfmt 6, task blastn, dust "no," *e*-value 10, word_size 7, reward 2, num_alignments 10. Blast vs. precursor sequences was based on the following parameters: outfmt 6, task blastn, dust "no," *e*-value 10e−3, num_alignments 10.

The correlation between biological replicates was evaluated by calculating the Pearson coefficient for all the possible pairs of replicates belonging to the same biological sample, as well as samples from different tissues for sake of comparison. We decided to remove from the set a sequence (HE860285) whose expression level was abnormally high, because its presence caused the Pearson correlation to be almost one in every comparison, irrespective of the tissue.

To identify miRNAs isomiRs that were differentially expressed among the biological samples, a *t*-test was performed for all the possible comparisons. An isomiR was considered as differentially expressed in a specific comparison if its *p*-value was less than 0.05. The whole set of reads associated with the miRNA loci was then used to perform a hierarchical clustering with R software, by applying the Canberra metric to calculate the distances between the expression vectors of the samples across the reads.

MiRNA target identification was carried out by psRNATarget[3] (Dai and Zhao, 2011), with default parameters. To score the complementarity between small RNA and their target transcript, psRNATarget applies the scoring schema of miRU by Zhang (2005). The *maximum expectation* is the threshold of the score. A small RNA/target site pair will be discarded if its score is greater than the threshold. The default cut-off threshold is 3.0.

The accessibility of the mRNA target site to small RNA has been identified as one of the important factors involved in target recognition because the secondary structure (stem, etc.) around the target site will prevent the small RNA and mRNA target from contacting. The psRNATarget server employes RNAup to calculate target accessibility, which is represented by the energy required to open (unpair) secondary structure around the target site (usually the complementary region with small RNA and up/downstream) on target mRNA. The lower the energy the higher the possibility that small RNA is able to contact (and cleave) target mRNA. PsRNATarget uses a software, namely RNAup, described by Mückstein et al. (2006) to calculate this value, denoted as UPE.

All the miRNA-related sequences were submitted to the EMBL database, whereas the sequencing raw reads were submitted to the NCBI SRA (BioProject accession: PRJNA167962).

RESULTS

SEQUENCING PEACH SMALL RNA LIBRARIES

Illumina deep sequencing was used to profile the whole miRNA set of five different samples corresponding to different organs and/or phenological stages. Three replicates were analyzed for

[2]http://srna-tools.cmp.uea.ac.uk/plant/cgi-bin/srna-tools.cgi

[3]http://plantgrn.noble.org/psRNATarget

each sample. A total number of 40,764,330 sequence reads were obtained and filtered as reported in Section "Materials and Methods." Details on the results of each filtering step are reported in **Table 1**. On average, 2,717,622 raw reads and 664,777 clean reads perfectly matching the genome were obtained in each of the 15 samples.

One technical replicate (randomly chosen and numbered as "1") of each sample was subsequently analyzed with miRDeeP-P which highlighted the putative miRNA coding loci of the peach genome expressed in the five tested samples (reported in Files S1–S5 in Supplementary Material and summarized in File S6 in Supplementary Material). Twenty-six putative miRNA coding loci were expressed in all samples according to miRDeep-P results. The length of the putative precursors was between 41 nt and 227 nt (average length of 104 nt), while average mature miRNAs size was 22 nt.

These 26 miRNAs were selected and, for each of them, the corresponding associated reads were searched in all the replicates of each sample. The results (miRNAs and miRNAs* associated reads) are reported in **Table 2** for each locus, the link between locus name and locus position can be found in File S7 in Supplementary Material and retrieved at www.rosaceae.org/peach/genome.

ISOMIRs IDENTIFICATION AND ANALYSIS

IsomiRs at each locus were blasted against miRBase. In some cases no mismatches were reported with the conserved sequences present in miRBase (e.g., miR403, miR394, miR166, miR156) while in some others mismatches were present and related to differences in the sequence and/or in its length. Detailed blast results are reported in File S8 in Supplementary Material which reports blast results based both on mature sequences (sheet "mature") and precursor sequences (sheet "precursors") deposited in miRBase. The file reports the matching sequence with the lowest *e*-value. When more than one matching sequence, belonging to different miRNA families, were found to have the same *e*-value all of them were reported.

Some miRNA families have more than one putative locus, therefore miRDeeP assigned common reads to all the possible loci. Both miRNA and miRNA*-related reads were identified at each locus. In some cases putative miRNAs* were identified on the basis of the alignment orientation (± with miRNA mature sequence deposited in miRBase) in some others the miRNA* sequences were already deposited in miRBase. The results of **Table 2** highlight that some loci are characterized by a larger set of variants than others.

In the majority of the loci the most frequent read for a specific locus was the same in all the tested samples and across all the replicates of a sample (**Table 2**). Only in a few cases were some differences detected among samples or among replicates belonging to the same sample. Locus named 3_16 is particularly interesting because all the replicates of sample O have as the most frequent read the one corresponding to miRNA* (**Table 2**).

In some loci also the second most frequent read referred to the mature miRNA was the same in all the replicates of a sample and in all the samples. The second most frequent read was often obtained by a different cutting site at 5′ or 3′ ends. As reported above, miRNA*-related reads have also been identified by miRDeep-P for most of the 26 loci and length variability was detected for both 5′ and 3′ends.

Target analysis was carried out by psRNATarget. The whole set of targets identified is reported in File S9 in Supplementary Material.

INTRA- AND INTER-SAMPLES ANALYSIS

The average Pearson correlation between all the possible pairs of replicates belonging to the same biological sample was calculated, in order to evaluate whether it was in agreement with the "Standards, guidelines, and best practices for RNA-seq" adopted

Table 1 | Reports the number of filtered reads perfectly matching the peach genome in each of the tested samples.

Sample	Raw reads	Matching adaptors	Matching adaptors (18–24 nt)	Low-complexity filtered (non-redundant)	rRNA/tRNA removed (non-redundant)	Matching peach genome (non-redundant)
BF1	2842653	2332661	1592671	1582689 (356161)	1335567 (333478)	797297 (207233)
BF2	2553116	2124140	1570583	1560824 (370103)	1401892 (351620)	817438 (220223)
BF3	2641037	2184480	1646360	1635958 (387915)	1466005 (368373)	853043 (235873)
F1	2523898	1819981	1180424	1173273 (215571)	858130 (194546)	368958 (78332)
F2	2898014	2361457	1389585	1381108 (233414)	1040918 (214222)	630870 (115644)
F3	3613383	2923242	1768783	1757830 (320484)	1362239 (297258)	700729 (143030)
GF1	2696289	2170904	1354915	1346476 (307349)	1200810 (289976)	774146 (190043)
GF2	3722325	2848155	1962881	1950819 (429497)	1570664 (405303)	928251 (248399)
GF3	2952035	2254276	1327544	1319295 (308967)	1145330 (291401)	609728 (168939)
GL1	2357377	1707677	1072569	1065997 (241873)	870779 (220909)	481523 (122824)
GL2	2304406	1729282	1080322	1073580 (253835)	929602 (234995)	460632 (125267)
GL3	1822754	1345521	745352	740785 (207198)	616318 (190449)	338168 (111111)
O1	3704334	2907197	1811057	1799817 (327514)	1384994 (302970)	890186 (182703)
O2	1896816	1509747	1068609	1061991 (226499)	911190 (208620)	624084 (129672)
O3	2235893	1753142	1286833	1278775 (282053)	1093002 (263673)	696601 (162637)

BF, pink; F, bloom; GF, swollen flower bud; O, half-inch green; GL, swollen leaf bud.

Table 2 | Reports the read count (divided by the total number of reads with a perfect match to the peach genome and normalized to 1,000,000 reads) of 26 putative miRNA coding loci that were expressed in all the 15 samples according to miRDeep-P results.

miRNA	Reads	EMBL accession number	BF1	BF2	BF3	F1	F2	F3	GF1	GF2	GF3	GL1	GL2	GL3	O1	O2	O3
1_10	AGTTTGTGCGTGAATCGAACC	HE862997	2.5	1.2	1.2	35.2	11.1	45.7	2.6	4.3	4.9	0	10.9	5.9	1.1	0	0
	CAGTTTGTGCGTGAATCGAAC	HE860429	6.3	9.8	19.9	24.4	7.9	20	3.9	10.8	14.8	8.3	17.4	5.9	3.4	0	8.6
	TTAGATTCACGCACAAAC	HE862999	0	0	0	0	0	0	1.3	2.2	1.6	0	0	0	0	0	0
	TTAGATTCACGCACAAACT	HE860429	3.8	6.1	4.7	5.4	0	4.3	3.9	4.3	3.3	2.1	6.5	0	1.1	1.6	0
	TTAGATTCACGCACAAACTC	HE863001	2.5	6.1	4.7	8.1	0	5.7	1.3	4.3	1.6	0	4.3	0	4.5	3.2	4.3
	TTAGATTCACGCACAAACTCG	HE860293	66.5	93	83.2	208.7	71.3	125.6	67.2	106.7	132.8	105.9	125.9	230.7	27	40.1	58.9
1_15	AACCACAAATCTCTTGGACTCCTG	HE860430	1.3	0	1.2	0	0	0	1.3	1.1	0	0	0	0	0	1.6	0
	AAGAGATTTGTGGTTACTCAC	HE863003	0	0	1.2	0	0	0	1.3	2.2	1.6	0	4.3	3	0	1.6	1.4
	AAGAGATTTGTGGTTACTCACC	HE860430	2.5	2.4	0	0	0	1.4	0	1.1	0	0	0	0	0	0	0
	AAGAGATTTGTGGTTACTCACCG	HE863005	0	7.3	1.2	0	0	0	2.6	1.1	1.6	0	0	0	1.1	0	2.9
	AAGAGATTTGTGGTTACTCACCGT	HE860431	12.5	18.4	12.9	5.4	3.2	2.9	14.2	9.7	6.6	4.2	2.2	0	6.7	6.4	7.2
	AGAGATTTGTGGTTACTCAC	HE863007	6.3	2.4	3.5	2.7	0	0	0	1.1	0	2.1	0	0	1.1	1.6	1.4
	AGAGATTTGTGGTTACTCACCG	HE860431	1.3	4.9	4.7	0	1.6	0	2.6	0	0	2.1	0	3	1.1	4.8	4.3
	AGAGATTTGTGGTTACTCACCGT	HE863009	0	1.2	2.3	0	1.6	0	0	1.1	1.6	0	0	0	2.2	0	2.9
	AGAGATTTGTGGTTACTCACCGTT	HE860297	117.9	162.7	126.6	27.1	36.5	25.7	71	65.7	49.2	45.7	43.4	20.7	80.9	52.9	34.5
	ATTTACATCCAACGGTGAGTAACC	HE860432	0	0	0	0	0	0	0	0	0	2.1	2.2	0	0	0	0
	CAAGAGATTTGTGGTTACTCA	HE863011	0	0	0	0	1.6	0	0	0	0	0	0	3	1.1	0	0
	CAAGAGATTTGTGGTTACTCACC	HE860432	1.3	0	1.2	0	0	0	0	2.2	0	0	0	0	1.1	0	0
	CAAGAGATTTGTGGTTACTCACCG	HE863013	0	2.4	2.3	2.7	3.2	4.3	2.6	2.2	0	0	0	0	0	0	0
	CCAAGAGATTTGTGGTTACTCA	HE860433	0	0	0	0	0	0	0	0	0	0	0	0	1.1	0	0
	TCCAAGAGATTTGTGGTTACTCAC	HE863015	1.3	0	0	0	0	1.4	0	0	0	2.1	0	0	0	0	1.4
1_25	CGAAACCTCCCATTCCAA	HE860433	1.3	0	0	0	0	0	1.3	2.2	0	2.1	2.2	3	0	0	0
	GAGAGGTTGCCGGAAAGA	HE863017	0	0	0	0	0	0	0	0	0	2.1	0	0	0	0	0
	GGGTGAGAGGTTGCCGGAAA	HE860434	0	0	2.3	0	0	0	0	0	0	2.1	0	0	0	1.6	4.3
	GGGTGAGAGGTTGCCGGAAAG	HE863019	2.5	9.8	16.4	0	0	0	10.3	14	8.2	12.5	15.2	11.8	15.7	30.4	38.8
	GGGTGAGAGGTTGCCGGAAAGA	HE860434	32.6	24.5	52.8	0	3.2	0	96.9	206.8	157.4	105.9	121.6	106.5	75.3	91.3	208.2
	GGGTGAGAGGTTGCCGGAAAGAA	HE863021	0	0	0	0	0	0	1.3	0	0	0	0	3	0	0	0
	GGTGAGAGGTTGCCGGAAAGAAT	HE860435	0	0	0	0	0	0	0	0	0	4.2	0	0	0	0	0
	TCCGAAACCTCCCATTCCAA	HE863023	1.3	0	1.2	0	0	1.4	3.9	0	0	0	0	0	2.2	0	1.4
	TCCGAAACCTCCCATTCCAAT	HE860435	3.8	1.2	1.2	0	0	0	1.3	1.1	0	0	0	0	0	0	1.4
	TCCGAAACCTCCCATTCCAATG	HE863025	0	0	0	0	0	0	3.9	0	0	0	0	0	0	0	0
	TTCCGAAACCTCCCATTCCAA	HE860436	17.6	30.6	17.6	5.4	6.3	15.7	29.7	49.6	27.9	33.2	39.1	26.6	13.5	11.2	8.6
	TTCCGAAACCTCCCATTCCAAT	HE863027	1.3	2.4	1.2	2.7	0	1.4	6.5	1.1	3.3	0	2.2	3	0	3.2	0
	TTGGGTGAGAGGTTGCCGGAA	HE860436	0	0	0	0	0	0	0	0	0	0	0	0	1.1	0	0
	TTGGGTGAGAGGTTGCCGGAAA	HE863029	2.5	0	0	0	0	2.9	1.3	0	0	0	0	0	0	0	0
	TTTCCGAAACCTCCCATT	HE860437	2.5	2.4	1.2	0	0	1.4	0	2.2	1.6	0	4.3	3	1.1	0	0
	TTTCCGAAACCTCCCATTC	HE863031	3.8	1.2	4.7	0	0	8.6	7.8	5.4	3.3	2.1	4.3	3	1.1	0	0
	TTTCCGAAACCTCCCATTCC	HE860437	8.8	8.6	23.4	5.4	9.5	17.1	34.9	23.7	23	10.4	21.7	3	2.2	4.8	4.3
	TTTCCGAAACCTCCCATTCCA	HE863033	15.1	11	12.9	0	4.8	5.7	29.7	22.6	13.1	16.6	21.7	3	1.1	0	5.7
	TTTCCGAAACCTCCCATTCCAA	HE860304	1135.1	1196.4	1134.8	401.1	391.5	687.9	1802	2047.9	1835.2	1277.2	1417.6	777.7	433.6	387.8	446.5
	TTTCCGAAACCTCCCATTCCAAT	HE860438	18.8	15.9	16.4	8.1	6.3	12.8	14.2	21.5	14.8	20.8	28.2	3	3.4	4.8	5.7

1_26	AAAAAGACTCAACAACCCATGTTT	HE863035	0	0	0	0	0	0	0	0	0	2.1	0	0	0	0	0
	AAAAGACTCAACAACCCATGT	HE860438	0	1.2	1.2	2.7	0	1.4	0	0	0	0	2.2	0	0	1.6	0
	AAAAGACTCAACAACCCATGTTT	HE863037	0	0	0	0	0	0	0	0	0	2.1	0	0	0	0	0
	AAAGACTCAACAACCCATGT	HE860439	1.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	AAAGGCATAGTAGGGTTTAGGA	HE863039	0	0	0	0	0	0	1.3	0	0	0	0	0	0	0	1.4
	AAAGGCATAGTAGGGTTTAGGAAG	HE860439	3.8	0	1.2	0	0	0	6.5	8.6	3.3	2.1	4.3	0	0	0	1.4
	AAGGCATAGTAGGGTTTAGGAAGT	HE863041	1.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	ACCCCGCCCATTCCAAATATT	HE860440	0	0	0	2.7	0	1.4	1.3	0	0	0	0	0	0	1.6	0
	ACCCCGCCCATTCCAAATATTT	HE863043	0	0	0	0	0	1.4	1.3	0	0	0	0	0	0	0	0
	ATATTTTCTAAGCCTACTGTC	HE860440	7.5	3.7	5.9	8.1	3.2	8.6	22	20.5	26.2	16.6	13	11.8	13.5	4.8	7.2
	CAAATATTTTCTAAGCCTACTGTC	HE863045	0	0	0	2.7	0	0	0	0	0	0	0	0	0	0	0
	CATAGTAGGGTTTAGGAA	HE860441	0	0	0	0	0	0	1.3	0	0	2.1	0	0	0	0	0
	CATAGTAGGGTTTAGGAAGTT	HE863047	0	0	0	0	0	0	1.3	0	0	0	0	0	0	0	0
	CATAGTAGGGTTTAGGAAGTTT	HE860441	0	0	2.3	0	0	1.4	0	0	0	2.1	0	0	0	0	0
	CATAGTAGGGTTTAGGAAGTTTT	HE863049	1.3	1.2	2.3	0	1.6	2.9	5.2	3.2	1.6	2.1	0	0	1.1	0	2.9
	CATAGTAGGGTTTAGGAAGTTTTT	HE860442	7.5	7.3	4.7	5.4	3.2	4.3	10.3	11.9	3.3	10.4	0	3	2.2	1.6	4.3
	CTTTGCCAACCCCGCCCATTCC	HE863051	2.5	0	0	0	1.6	2.9	6.5	5.4	3.3	0	0	0	2.2	0	1.4
	CTTTGCCAACCCCGCCCATTCCA	HE860442	1.3	0	0	0	0	0	0	0	1.6	0	0	0	0	0	0
	CTTTGCCAACCCCGCCCATTCCAA	HE863053	0	1.2	1.2	0	0	0	5.2	1.1	4.9	16.6	2.2	8.9	1.1	0	1.4
	GAAAGGCATAGTAGGGTTTAGGA	HE860443	1.3	0	1.2	0	0	0	0	1.1	0	2.1	0	0	1.1	1.6	0
	GAAAGGCATAGTAGGGTTTAGGAA	HE863055	1.3	0	5.9	2.7	0	5.7	3.9	6.5	0	0	0	0	0	1.6	4.3
	GCCAACCCCGCCCATTCCAA	HE860443	0	0	0	0	0	0	0	0	0	0	0	0	1.1	0	0
	GGAATGAGCGTGTTGGAAA	HE863057	1.3	0	1.2	0	0	0	0	0	0	0	0	0	0	0	1.4
	GGAATGAGCGTGTTGGAAAA	HE860444	1.3	1.2	0	0	0	0	1.3	0	0	0	2.2	0	1.1	1.6	1.4
	GGAATGAGCGTGTTGGAAAAG	HE863059	7.5	4.9	3.5	2.7	3.2	4.3	5.2	8.6	4.9	10.4	6.5	3	5.6	3.2	12.9
	GGAATGAGCGTGTTGGAAAAGA	HE860444	2.5	4.9	5.9	0	0	0	9	2.2	8.2	29.1	15.2	11.8	0	1.6	0
	GGAATGAGCGTGTTGGAAAAGAA	HE863061	0	0	0	0	0	0	1.3	0	0	2.1	0	0	1.1	0	0
	TATTTTCTAAGCCTACTGTC	HE860445	0	0	1.2	0	0	0	0	3.2	0	4.2	0	0	0	0	0
	TCTAAGCCTACTGTCTTTCCC	HE863063	0	0	0	0	0	2.9	2.6	1.1	0	2.1	0	0	2.2	1.6	0
	TCTAAGCCTACTGTCTTTCCCT	HE860445	0	0	0	0	0	1.4	1.3	0	0	0	0	0	1.1	1.6	0
	TGCCAACCCCGCCCATTCCA	HE863065	1.3	0	0	0	0	1.4	0	0	0	0	0	0	0	0	0
	TGCCAACCCCGCCCATTCCAA	HE860446	6.3	1.2	2.3	0	0	4.3	2.6	2.2	1.6	6.2	2.2	5.9	2.2	3.2	1.4
	TGCCAACCCCGCCCATTCCAAA	HE863067	6.3	4.9	2.3	5.4	0	4.3	10.3	11.9	16.4	14.5	28.2	8.9	6.7	8	7.2
	TGGAATGAGCGTGTTGGAAAA	HE860446	1.3	0	0	0	0	0	0	1.1	0	0	0	0	0	0	0
	TTCTTTGCCAACCCCGCCCATT	HE863069	1.3	0	0	0	0	1.4	2.6	0	1.6	4.2	4.3	3	1.1	0	0
	TTGCCAACCCCGCCCATT	HE860447	1.3	1.2	0	0	0	1.4	0	2.2	4.9	2.1	0	0	0	0	0
	TTGCCAACCCCGCCCATTC	HE863071	3.8	2.4	2.3	2.7	0	0	7.8	10.8	6.6	4.2	2.2	0	1.1	0	0
	TTGCCAACCCCGCCCATTCC	HE860447	26.3	24.5	15.2	10.8	4.8	17.1	71	40.9	78.7	35.3	26.1	53.2	10.1	24	12.9
	TTGCCAACCCCGCCCATTCCA	HE863073	5	3.7	3.5	5.4	0	1.4	10.3	5.4	9.8	2.1	8.7	5.9	3.4	6.4	5.7
	TTGCCAACCCCGCCCATTCCAA	HE860305	180.6	163.9	138.3	140.9	85.6	119.9	260.9	266.1	477.3	344.7	382.1	275	126.9	203.5	189.5
	TTGCCAACCCCGCCCATTCCAAA	HE860448	0	0	2.3	0	1.6	1.4	2.6	3.2	3.3	0	2.2	0	1.1	1.6	2.9
	TTGCCAACCCCGCCCATTCCAAAT	HE863075	1.3	0	0	0	0	0	0	0	0	0	0	0	0	1.6	0

(Continued)

Table 2 | Continued

miRNA	Reads	EMBL accession number	BF1	BF2	BF3	F1	F2	F3	GF1	GF2	GF3	GL1	GL2	GL3	O1	O2	O3
	TTTGAAGCAGATGATGGAAC	HE860448	0	0	0	0	0	0	1.3	0	0	0	0	0	0	0	0
	TTTGCCAACCCCGCCCAT	HE863077	0	2.4	0	0	0	0	1.3	0	3.3	0	2.2	0	1.1	0	1.4
	TTTGCCAACCCCGCCCATT	HE860449	2.5	0	0	2.7	0	0	6.5	0	8.2	2.1	2.2	0	0	6.4	0
	TTTGCCAACCCCGCCCATTC	HE863079	5	1.2	1.2	2.7	0	1.4	3.9	2.2	8.2	4.2	0	3	2.2	4.8	1.4
	TTTGCCAACCCCGCCCATTCC	HE860449	20.1	19.6	14.1	10.8	7.9	12.8	67.2	26.9	73.8	33.2	41.2	50.3	15.7	19.2	7.2
	TTTGCCAACCCCGCCCATTCCA	HE863081	5	8.6	3.5	2.7	1.6	1.4	22	9.7	13.1	10.4	13	5.9	5.6	3.2	5.7
	TTTGCCAACCCCGCCCATTCCAA	HE860450	115.4	71	70.3	59.6	36.5	75.6	148.6	136.8	239.5	240.9	223.6	171.5	57.3	134.6	83.3
	TTTGCCAACCCCGCCCATTCCAAA	HE863083	0	1.2	1.2	0	0	0	3.9	2.2	11.5	0	4.3	11.8	0	0	0
1_29	AGGTGGGCATACTGCCAACTG	HE860450	3.8	2.4	2.3	13.6	4.8	10	1.3	0	0	0	0	0	3.4	1.6	1.4
	ATTGGCATTCTGTCCACCTCC	HE863085	0	1.2	0	0	1.6	0	0	0	0	0	0	0	1.1	0	0
	TGGCATTCTGTCCACCTCC	HE860451	1.3	0	0	0	0	1.4	0	0	0	0	2.2	0	0	0	0
	TTGGCATTCTGTCCACCT	HE863087	6.3	6.1	7	13.6	11.1	7.1	2.6	1.1	1.6	2.1	0	5.9	0	0	0
	TTGGCATTCTGTCCACCTC	HE860451	18.8	24.5	15.2	13.6	28.5	27.1	1.3	5.4	8.2	2.1	4.3	0	0	1.6	8.6
	TTGGCATTCTGTCCACCTCC	HE860307	89.1	121.1	89.1	127.4	141.1	225.5	80.1	53.9	59	27	26.1	20.7	20.2	35.3	25.8
	TTGGCATTCTGTCCACCTCCT	HE863089	36.4	29.4	36.3	10.8	26.9	62.8	23.3	15.1	13.1	2.1	4.3	3	4.5	4.8	1.4
	TTGGCATTCTGTCCACCTCCTC	HE860452	1.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1_3	AACATGATCATCCGAATGAT	HE863091	0	0	0	0	0	0	0	0	0	0	0	0	1.1	0	0
	AATGCTGTCTGGTTCGAGA	HE860452	1.3	2.4	1.2	0	1.6	2.9	0	1.1	3.3	2.1	2.2	0	0	1.6	1.4
	ACCAGGCTTCATTCCCCC	HE863093	1.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	ATCCGAATGATCTCGGACCAGGCT	HE860453	0	0	0	0	0	0	0	0	0	0	0	0	1.1	0	0
	ATCTCGGACCAGGCTTCATTCCCC	HE863095	6.3	14.7	17.6	0	9.5	11.4	2.6	1.1	0	6.2	6.5	0	1.1	4.8	5.7
	ATGCTGTCTGGTTCGAGA	HE860453	0	0	0	2.7	0	0	0	0	0	0	0	0	0	0	0
	CGGACCAGGCTTCATTCC	HE863097	0	0	0	0	0	0	0	1.1	0	0	2.2	0	2.2	0	0
	CGGACCAGGCTTCATTCCC	HE860454	1.3	2.4	0	2.7	1.6	0	1.3	5.4	1.6	0	0	3	1.1	0	0
	CGGACCAGGCTTCATTCCCC	HE863099	282.2	208	289.6	336.1	187	182.7	235.1	213.3	203.4	265.8	251.8	174.5	449.3	299.6	328.7
	CGGACCAGGCTTCATTCCCCC	HE860454	1.3	1.2	0	2.7	3.2	1.4	1.3	1.1	0	2.1	4.3	0	0	1.6	0
	CTCGGACCAGGCTTCATTCC	HE863101	0	2.4	1.2	2.7	1.6	1.4	1.3	0	1.6	0	0	0	1.1	0	0
	CTCGGACCAGGCTTCATTCCC	HE860455	66.5	83.2	65.6	132.8	42.8	129.9	9	9.7	9.8	10.4	17.4	17.7	16.9	46.5	43.1
	CTCGGACCAGGCTTCATTCCCC	HE863103	151.8	165.2	175.8	192.4	130	189.8	165.3	159.4	146	126.7	147.6	136	75.3	187.5	183.7
	CTCGGACCAGGCTTCATTCCCCC	HE860455	0	0	0	0	0	1.4	1.3	0	0	2.1	0	0	0	0	0
	GAATGCTGTCTGGTTCGAGAC	HE863105	3.8	1.2	0	5.4	0	1.4	0	0	0	0	0	0	1.1	1.6	0
	GACCAGGCTTCATTCCCC	HE860456	1.3	0	0	0	0	0	0	0	1.6	0	0	0	0	0	0
	GGAATGCTGTCTGGTTCGA	HE863107	0	0	0	0	0	0	1.3	0	0	0	0	0	0	0	0
	GGAATGCTGTCTGGTTCGAGA	HE860456	12.5	17.1	12.9	24.4	4.8	30	3.9	4.3	4.9	8.3	17.4	29.6	4.5	0	10
	GGAATGCTGTCTGGTTCGAGAC	HE863109	5	6.1	8.2	2.7	4.8	17.1	0	1.1	3.3	4.2	4.3	0	0	0	2.9
	GGACCAGGCTTCATTCCC	HE860457	1.3	0	1.2	0	0	0	0	0	0	0	0	0	0	0	0
	GGACCAGGCTTCATTCCCC	HE863111	146.7	154.1	143	184.3	136.3	148.4	155	161.6	134.5	189	191	147.9	171.9	174.7	193.8
	TCGGACCAGGCTTCATTC	HE860457	58.9	64.8	65.6	43.4	28.5	31.4	42.6	54.9	39.4	24.9	28.2	29.6	47.2	54.5	40.2
	TCGGACCAGGCTTCATTCC	HE863113	440.2	408.6	385.7	417.4	261.5	412.4	384.9	339.3	332.9	180.7	256.2	283.9	159.5	280.4	254.1
	TCGGACCAGGCTTCATTCCC	HE860458	706.1	675.3	720.9	441.8	321.8	449.5	586.5	627	546.1	388.4	525.4	387.4	410	512.8	541.2
	TCGGACCAGGCTTCATTCCCC	HE860285	282744.1	279717.9	296598.2	273500	243934.6	239360.7	306036.6	275122.8	248117.2	282794.4	317557.2	199637.5	217491.6	348033.3	350352.6
	TCGGACCAGGCTTCATTCCCCC	HE863115	209.5	212.9	233.3	384.9	160.1	276.9	170.5	137.9	165.6	272.1	191	168.6	174.1	282	236.9

	TCTCGGACCAGGCTTCATTCC	HE860458	110.4	179.8	155.9	273.7	190.2	332.5	11.6	8.6	14.8	24.9	99.9	11.8	28.1	57.7	45.9
	TCTCGGACCAGGCTTCATTCCC	HE863117	0	1.2	1.2	5.4	0	5.7	0	0	0	0	2.2	0	0	1.6	1.4
	TCTCGGACCAGGCTTCATTCCCC	HE860459	2.5	11	9.4	8.1	4.8	12.8	9	4.3	8.2	6.2	4.3	5.9	4.5	9.6	10
	TCTCGGACCAGGCTTCATTCCCCC	HE863119	1.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1_32	ATTGACAGAAGAGAGTGAGCAC	HE860459	1.3	0	0	0	0	0	1.3	0	0	0	0	0	0	0	0
	GACAGAAGAGAGTGAGCAC	HE863121	1.3	1.2	0	0	0	0	0	0	0	0	0	0	0	0	0
	GCTCATGTCTCTTTCTGTCAGC	HE860460	5	2.4	7	2.7	4.8	4.3	0	1.1	0	2.1	0	0	1.1	1.6	1.4
	GCTCATGTCTCTTTCTGTCAGCT	HE863123	2.5	0	0	0	0	1.4	0	0	0	0	0	0	0	0	0
	TGACAGAAGAGAGTGAGCA	HE860460	1.3	0	1.2	0	0	0	0	0	0	0	0	0	0	0	0
	TGACAGAAGAGAGTGAGCAC	HE860311	71.5	80.7	97.3	181.6	38	44.2	29.7	53.9	31.2	49.8	17.4	11.8	11.2	11.2	20.1
	TGACAGAAGAGAGTGAGCACA	HE863125	2.5	3.7	0	0	1.6	0	0	0	1.6	2.1	0	0	0	0	0
	TGCTCATGTCTCTTTCTGTCAGC	HE860461	0	2.4	2.3	8.1	1.6	0	1.3	0	0	2.1	0	0	0	1.6	0
	TTGACAGAAGAGAGTGAGCAC	HE863127	8.8	14.7	15.2	21.7	3.2	7.1	2.6	0	1.6	10.4	0	0	0	0	0
1_44	AGGTGGTCAGCATGTCAAACT	HE860461	3.8	2.4	3.5	2.7	0	0	5.2	9.7	0	0	0	3	2.2	3.2	5.7
	TGGCATTCTGTCCACCTCC	HE860451	1.3	0	0	0	0	1.4	0	0	0	0	2.2	0	0	0	0
	TTGGCATTCTGTCCACCT	HE863087	6.3	6.1	7	13.6	11.1	7.1	2.6	1.1	1.6	2.1	0	5.9	0	0	0
	TTGGCATTCTGTCCACCTC	HE860451	18.8	24.5	15.2	13.6	28.5	27.1	1.3	5.4	8.2	2.1	4.3	0	0	1.6	8.6
	TTGGCATTCTGTCCACCTCC	HE860307	89.1	121.1	89.1	127.4	141.1	225.5	80.1	53.9	59	27	26.1	20.7	20.2	35.3	25.8
	TTTGGCATTCTGTCCACCTCC	HE863129	1.3	0	1.2	0	0	0	1.3	1.1	1.6	2.1	0	0	1.1	0	0
1_5	TACAATGAAATCACGGCC	HE860462	0	0	0	0	0	0	0	0	0	2.1	0	0	0	0	0
	TATAAAGAGATGTACTGGACC	HE863131	3.8	2.4	2.3	2.7	0	0	2.6	2.2	0	2.1	4.3	0	2.2	1.6	1.4
	TTATACAATGAAATCACGG	HE860462	0	0	0	0	0	0	0	0	0	0	0	0	1.1	0	0
	TTATACAATGAAATCACGGC	HE860288	8.8	23.2	27	5.4	7.9	7.1	14.2	15.1	3.3	16.6	23.9	20.7	20.2	17.6	18.7
	TTATACAATGAAATCACGGCC	HE860287	125.4	174.9	219.2	70.5	130	81.3	107.2	131.4	147.6	280.4	230.1	378.5	179.7	126.6	277.1
	TTATACAATGAAATCACGGCCG	HE860286	129.2	116.2	148.9	29.8	71.3	37.1	63.3	75.4	62.3	170.3	141.1	195.2	85.4	83.3	208.2
10_1	ACAGGGAACAGGTAGAGCA	HE863133	2.5	1.2	2.3	0	0	0	0	1.1	0	0	0	0	1.1	1.6	0
	ACAGGGAACAGGTAGAGCATG	HE860463	2.5	3.7	8.2	21.7	0	2.9	1.3	2.2	0	0	0	0	0	4.8	0
	ATGCACTGCCTCTTCCCTGGC	HE863135	2.5	3.7	4.7	2.7	1.6	5.7	1.3	0	0	0	0	0	2.2	8	1.4
	TGCACTGCCTCTTCCCTG	HE860463	11.3	3.7	7	5.4	0	8.6	3.9	3.2	0	0	0	3	9	4.8	5.7
	TGCACTGCCTCTTCCCTGG	HE863137	26.3	22	36.3	16.3	1.6	8.6	7.8	10.8	3.3	2.1	0	3	6.7	24	11.5
	TGCACTGCCTCTTCCCTGGC	HE860464	47.7	35.5	32.8	146.4	14.3	65.6	11.6	12.9	8.2	2.1	8.7	8.9	15.7	16	15.8
	TGCACTGCCTCTTCCCTGGCT	HE860428	154.3	137	137.2	168	53.9	119.9	95.6	72.2	90.2	16.6	19.5	14.8	83.1	81.7	64.6
	TGCACTGCCTCTTCCCTGGCTG	HE863139	2.5	6.1	3.5	10.8	1.6	5.7	1.3	2.2	1.6	0	2.2	0	2.2	1.6	4.3
2_31	CCAAAGGGATCGCATTGATCT	HE860464	0	0	0	2.7	1.6	0	0	0	0	0	0	0	0	0	0
	TCCAAAGGGATCGCATTGA	HE863141	0	1.2	1.2	0	0	0	0	0	0	2.1	0	0	0	0	0
	TCCAAAGGGATCGCATTGAT	HE860465	0	1.2	0	0	0	1.4	1.3	1.1	0	0	0	0	0	0	0
	TCCAAAGGGATCGCATTGATC	HE860334	12.5	22	18.8	46.1	44.4	44.2	37.5	43.1	16.4	22.8	23.9	41.4	1.1	3.2	7.2
	TCCAAAGGGATCGCATTGATCT	HE863143	3.8	12.2	15.2	19	20.6	11.4	11.6	19.4	8.2	8.3	4.3	3	1.1	0	2.9
	TCCAAAGGGATCGCATTGATCTA	HE860465	0	0	0	5.4	1.6	4.3	0	0	0	0	0	0	0	0	0
	TCGATGCGATCCCTTGGGA	HE863145	1.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	TCGATGCGATCCCTTGGGAAG	HE860337	1.3	4.9	3.5	67.8	30.1	55.7	1.3	4.3	1.6	12.5	6.5	3	2.2	1.6	5.7
	TCGATGCGATCCCTTGGGAAGT	HE860466	0	0	0	2.7	1.6	0	0	0	0	0	0	0	0	0	0

(Continued)

Table 2 | Continued

miRNA	Reads	EMBL accession number	BF1	BF2	BF3	F1	F2	F3	GF1	GF2	GF3	GL1	GL2	GL3	O1	O2	O3
	TGATATTGGATCGATGCGATC	HE863147	0	0	0	0	0	0	1.3	0	0	0	0	0	0	0	0
3_16	ATTGTAGGAATGGGCTGTTTG	HE860466	2.5	0	1.2	0	0	0	1.3	0	0	0	0	0	0	0	0
	CCCAAGCCCGCCCATTCC	HE863149	0	0	0	0	0	0	2.6	0	0	0	0	0	0	0	0
	CCCAAGCCCGCCCATTCCA	HE860467	0	0	0	0	0	0	1.3	1.1	0	0	0	0	0	0	0
	CTTCCCAAGCCCGCCCATTCCA	HE863151	0	0	0	0	0	0	1.3	0	0	0	0	0	0	0	0
	GGAATGGGCTGTTTGGGA	HE860467	13.8	7.3	17.6	16.3	28.5	12.8	3.9	7.5	11.5	10.4	17.4	20.7	11.2	14.4	10
	GGAATGGGCTGTTTGGGAT	HE863153	5	1.2	4.7	0	1.6	2.9	1.3	2.2	0	6.2	4.3	0	2.2	1.6	5.7
	GGAATGGGCTGTTTGGGATG	HE860468	70.2	56.3	92.6	168	68.2	92.8	10.3	24.8	23	22.8	17.4	20.7	69.6	56.1	61.7
	GGAATGGGCTGTTTGGGATGA	HE860347	100.3	84.4	150.1	311.7	111	177	22	49.6	50.8	47.8	34.7	53.2	197.7	86.5	104.8
	GGAATGGGCTGTTTGGGATGAA	HE863155	0	0	0	2.7	0	1.4	0	0	0	0	0	0	1.1	0	0
	GGAATGGGCTGTTTGGGATGAAAG	HE860468	0	0	0	2.7	0	0	0	0	0	0	0	0	2.2	0	1.4
	TAGGAATGGGCTGTTTGGGA	HE863157	2.5	1.2	3.5	0	1.6	0	0	0	0	0	0	0	0	0	1.4
	TTCCCAAGCCCGCCCATT	HE860469	0	0	0	0	0	0	0	0	0	2.1	0	0	0	0	0
	TTCCCAAGCCCGCCCATTC	HE863159	1.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	TTCCCAAGCCCGCCCATTCC	HE860469	0	3.7	0	10.8	0	7.1	6.5	4.3	4.9	6.2	8.7	3	0	0	1.4
	TTCCCAAGCCCGCCCATTCCA	HE863161	0	0	2.3	0	1.6	1.4	0	0	1.6	2.1	0	0	0	0	0
	TTCCCAAGCCCGCCCATTCCAA	HE860348	112.9	130.9	99.6	216.8	123.6	182.7	217	225.2	288.7	278.3	212.8	115.3	47.2	38.5	63.2
	TTGTAGGAATGGGCTGTTTGGGA	HE860470	0	0	0	0	0	0	0	0	0	2.1	0	0	0	0	0
	TTTCTTTCATCCCAAACAGCC	HE863163	0	0	0	0	0	0	1.3	0	0	0	0	0	0	0	0
3_28	ATGGTGTCATCCCTCCTGTGACC	HE860470	0	0	0	0	0	0	0	0	0	2.1	0	0	0	0	0
	CCAAATTGAGAGAGAGAGAGAGAG	HE863165	1.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	CCATCTTCCTGTGACATGAAC	HE860471	0	0	0	2.7	0	1.4	0	0	0	2.1	0	0	0	0	0
	CGCAGGAGAGATGGCACTG	HE863167	0	0	0	0	0	0	0	0	0	2.1	2.2	3	0	0	0
	GGTGTCATCCCTCCTGTGACC	HE860471	0	0	0	0	1.6	0	0	0	0	2.1	2.2	0	0	0	0
	TCCATCTTCCTGTGACATGA	HE863169	0	0	0	2.7	0	0	0	0	1.6	0	2.2	3	0	0	0
	TCGCAGGAGAGATGGCAC	HE860472	2.5	0	0	0	0	0	0	1.1	0	0	4.3	3	0	0	0
	TCGCAGGAGAGATGGCACTG	HE863171	7.5	2.4	2.3	5.4	0	0	5.2	14	9.8	35.3	39.1	47.3	0	0	5.7
	TCGCAGGAGAGATGGCACTGT	HE860472	1.3	0	2.3	8.1	0	10	1.3	5.4	3.3	8.3	15.2	17.7	1.1	0	1.4
	TCGCAGGAGAGATGGCACTGTC	HE860356	15.1	9.8	15.2	51.5	19	38.5	40	74.3	44.3	110.1	165	139	3.4	4.8	7.2
	TCGCAGGAGAGATGGCACTGTCT	HE863173	0	0	0	2.7	0	0	0	0	1.6	0	0	0	0	0	0
	TGGTGTCATCCCTCCTGTGACC	HE860473	0	0	0	8.1	1.6	4.3	0	2.2	4.9	10.4	8.7	29.6	1.1	0	1.4
	TTCCATCTTCCTGTGACATGA	HE863175	0	0	0	2.7	3.2	1.4	2.6	1.1	1.6	2.1	4.3	23.7	0	0	0
	TTCGCAGGAGAGATGGCAC	HE860473	0	0	0	0	0	0	0	0	0	2.1	0	0	0	0	0
	TTCGCAGGAGAGATGGCACTGTC	HE863177	0	0	1.2	0	0	1.4	2.6	1.1	1.6	2.1	2.2	3	0	0	1.4
4_21	CCCTGCAGTACCTTCCTTTACCC	HE860474	0	0	1.2	2.7	0	0	0	0	0	0	0	0	0	0	0
	GGAGCGACCTGGGATCACATG	HE863179	0	1.2	1.2	21.7	1.6	15.7	0	0	0	0	0	0	1.1	0	0
	GTGTTCTCAGGTCGCCCCTG	HE860474	0	0	2.3	0	0	1.4	0	1.1	3.3	0	0	0	2.2	0	0
	TGTGTTCTCAGGTCGCCCC	HE863181	3.8	2.4	3.5	8.1	4.8	4.3	0	0	0	2.1	0	3	0	14.4	7.2
	TGTGTTCTCAGGTCGCCCCT	HE860475	0	1.2	0	2.7	0	0	0	2.2	0	0	0	3	3.4	1.6	8.6
	TGTGTTCTCAGGTCGCCCCTG	HE860366	356.2	210.4	280.2	311.7	313.9	216.9	31	71.1	54.1	60.2	36.9	35.5	540.3	700.2	723.5
5_14	AATGTTGTCTGGCTCGAG	HE863183	0	1.2	3.5	0	0	1.4	0	0	0	0	0	0	1.1	1.6	1.4
	AATGTTGTCTGGCTCGAGG	HE860475	6.3	13.5	7	8.1	0	1.4	6.5	8.6	0	6.2	10.9	0	6.7	28.8	11.5

AATGTTGTCTGGCTCGAGGCC	HE863185	0	0	0	2.7	0	0	1.3	1.1	1.6	0	4.3	0	1.1	1.6	0
AATGTTGTCTGGCTCGAGGCCC	HE860476	0	0	0	0	0	0	1.3	1.1	0	0	0	0	0	0	0
AATGTTGTCTGGCTCGAGGCCCCT	HE863187	0	0	0	0	0	0	0	2.2	3.3	2.1	0	3	0	0	0
ACCAGGCTTCATTCCCCC	HE863093	1.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ACGTCGGACCAGGCTTCATTC	HE860476	0	0	0	0	0	0	0	0	0	2.1	0	0	0	0	0
ACGTCGGACCAGGCTTCATTCCCC	HE863189	1.3	0	1.2	0	0	0	1.3	0	1.6	0	0	0	0	1.6	1.4
ATGTTGTCTGGCTCGAGG	HE860477	1.3	2.4	3.5	0	0	0	0	0	0	0	2.2	0	1.1	1.6	2.9
ATTTGGTTCTACATTTAGTGAC	HE863191	1.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CGGACCAGGCTTCATTCC	HE863097	0	0	0	0	0	0	0	1.1	0	0	2.2	0	2.2	0	0
CGGACCAGGCTTCATTCCC	HE860454	1.3	2.4	0	2.7	1.6	0	1.3	5.4	1.6	0	0	3	1.1	0	0
CGGACCAGGCTTCATTCCCC	HE863099	282.2	208	289.6	336.1	187	182.7	235.1	213.3	203.4	265.8	251.8	174.5	449.3	299.6	328.7
CGGACCAGGCTTCATTCCCCC	HE860454	1.3	1.2	0	2.7	3.2	1.4	1.3	1.1	0	2.1	4.3	0	0	1.6	0
CGTCGGACCAGGCTTCATTCC	HE860477	1.3	0	1.2	0	1.6	0	0	2.2	0	0	0	0	0	0	0
CGTCGGACCAGGCTTCATTCCC	HE863193	0	0	0	0	0	1.4	0	0	0	0	0	0	3.4	0	0
CGTCGGACCAGGCTTCATTCCCC	HE860478	3.8	2.4	1.2	0	0	0	1.3	0	3.3	0	2.2	0	2.2	1.6	4.3
GAATGTTGTCTGGCTCGA	HE863195	1.3	1.2	0	0	0	0	1.3	0	1.6	0	0	0	1.1	0	0
GAATGTTGTCTGGCTCGAGG	HE860478	5	13.5	7	10.8	1.6	2.9	1.3	2.2	0	0	2.2	3	3.4	14.4	5.7
GAATGTTGTCTGGCTCGAGGC	HE863197	0	1.2	2.3	2.7	0	0	0	2.2	0	0	0	3	0	0	1.4
GAATGTTGTCTGGCTCGAGGCC	HE860479	0	2.4	0	0	0	1.4	0	0	0	0	0	0	1.1	1.6	0
GAATGTTGTCTGGCTCGAGGCCCC	HE863199	0	0	0	0	0	0	0	0	3.3	2.1	0	0	0	0	0
GACCAGGCTTCATTCCCC	HE860456	1.3	0	0	0	0	0	0	0	1.6	0	0	0	0	0	0
GGAATGTTGTCTGGCTCG	HE860479	38.9	24.5	44.5	16.3	1.6	5.7	9	26.9	18	14.5	21.7	20.7	22.5	49.7	63.2
GGAATGTTGTCTGGCTCGA	HE863201	10	11	24.6	5.4	4.8	1.4	11.6	16.2	14.8	22.8	8.7	32.5	5.6	8	24.4
GGAATGTTGTCTGGCTCGAG	HE860480	6.3	3.7	10.6	16.3	4.8	0	6.5	8.6	3.3	8.3	4.3	11.8	4.5	6.4	14.4
GGAATGTTGTCTGGCTCGAGG	HE863203	165.6	156.6	174.7	73.2	26.9	49.9	74.9	106.7	80.4	58.1	117.2	263.2	175.2	211.5	353.1
GGAATGTTGTCTGGCTCGAGGC	HE860480	1.3	1.2	1.2	0	0	0	2.6	1.1	1.6	2.1	0	3	0	4.8	0
GGACCAGGCTTCATTCCC	HE860457	1.3	0	1.2	0	0	0	0	0	0	0	0	0	0	0	0
GGACCAGGCTTCATTCCCC	HE863111	146.7	154.1	143	184.3	136.3	148.4	155	161.6	134.5	189	191	147.9	171.9	174.7	193.8
GTCGGACCAGGCTTCATTC	HE863205	0	0	0	0	0	0	1.3	0	0	0	0	0	0	0	0
GTCGGACCAGGCTTCATTCC	HE860481	0	0	0	0	0	0	0	1.1	0	0	0	0	1.1	0	0
GTCGGACCAGGCTTCATTCCC	HE863207	64	117.4	86.7	86.7	49.1	109.9	49.1	56	42.6	29.1	28.2	3	97.7	94.5	68.9
GTCGGACCAGGCTTCATTCCCC	HE860481	21.3	26.9	29.3	29.8	9.5	14.3	31	29.1	23	33.2	30.4	17.7	51.7	62.5	37.3
GTCGGACCAGGCTTCATTCCCCC	HE863209	0	2.4	7	0	0	1.4	7.8	8.6	0	2.1	4.3	0	4.5	4.8	7.2
GTTGTCTGGCTCGAGGCC	HE860482	0	0	0	0	0	0	1.3	0	0	0	0	0	0	0	0
TAAATGTAGAACCAAATGATCT	HE863211	1.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TCACTAAATGTAGAACCAAATG	HE860482	0	0	0	0	0	0	0	0	0	0	0	0	1.1	0	0
TCGGACCAGGCTTCATTC	HE860457	58.9	64.8	65.6	43.4	28.5	31.4	42.6	54.9	39.4	24.9	28.2	29.6	47.2	54.5	40.2
TCGGACCAGGCTTCATTCC	HE863113	440.2	408.6	385.7	417.4	261.5	412.4	384.9	339.3	332.9	180.7	256.2	283.9	159.5	280.4	254.1
TCGGACCAGGCTTCATTCCC	HE860458	706.1	675.3	720.9	441.8	321.8	449.5	586.5	627	546.1	388.4	525.4	387.4	410	512.8	541.2
TCGGACCAGGCTTCATTCCCC	HE860285	282744.1	279717.9	296598.2	273500	243934.6	239360.7	306036.6	275122.8	248117.2	282794.4	317557.2	199637.5	217491.6	348033.3	350352.6
TCGGACCAGGCTTCATTCCCCC	HE863115	209.5	212.9	233.3	384.9	160.1	276.9	170.5	137.9	165.6	272.1	191	168.6	174.1	282	236.9
TGTCTGGCTCGAGGCCCCTA	HE863213	0	0	0	2.7	0	0	0	0	0	0	0	0	0	0	0

(Continued)

Table 2 | Continued

miRNA	Reads	EMBL accession number	BF1	BF2	BF3	F1	F2	F3	GF1	GF2	GF3	GL1	GL2	GL3	O1	O2	O3
5_3	CCCGCCTTGCATCAACTG	HE860483	0	0	2.3	0	0	0	0	1.1	0	6.2	0	3	0	0	0
	CCCGCCTTGCATCAACTGAA	HE863215	0	0	1.2	0	0	2.9	1.3	1.1	3.3	6.2	8.7	5.9	0	1.6	2.9
	CCCGCCTTGCATCAACTGAAT	HE860483	35.1	44	38.7	75.9	30.1	95.6	62	263.9	242.7	255.4	455.9	257.3	20.2	139.4	208.2
	CCGCCTTGCATCAACTGAAT	HE863217	2.5	0	1.2	0	0	0	2.6	0	0	2.1	4.3	0	0	0	0
	CGCTTGGTGCAGGTCGGGA	HE860484	0	0	0	0	0	0	0	1.1	1.6	2.1	0	3	0	0	0
	CGCTTGGTGCAGGTCGGGAA	HE863219	0	0	0	0	0	0	0	1.1	3.3	6.2	2.2	3	1.1	0	1.4
	CGCTTGGTGCAGGTCGGGAAC	HE860484	0	0	0	2.7	0	0	0	0	0	0	2.2	0	0	0	0
	GCTTGGTGCAGGTCGGGAA	HE863221	0	0	0	0	0	0	1.3	0	0	0	0	0	0	0	0
	GGGTCCCGCCTTGCATCAAC	HE860485	0	0	0	0	0	0	0	0	0	4.2	2.2	0	0	0	0
	GGTCCCGCCTTGCATCAACTGAAT	HE863223	0	0	2.3	0	0	0	0	0	0	2.1	2.2	0	0	0	0
	TCGCTTGGTGCAGGTCGGGA	HE860485	11.3	20.8	15.2	48.8	6.3	14.3	27.1	26.9	68.9	78.9	56.4	65.1	5.6	19.2	18.7
	TCGCTTGGTGCAGGTCGGGAA	HE860370	259.6	254.5	184	219.5	136.3	186.9	193.8	339.3	501.9	494.3	579.6	520.5	104.5	208.3	249.8
	TCGCTTGGTGCAGGTCGGGAACT	HE863225	0	0	0	0	1.6	0	3.9	1.1	8.2	10.4	4.3	3	0	1.6	4.3
	TGGGTCCCGCCTTGCATCAAC	HE860486	6.3	9.8	12.9	24.4	14.3	28.5	28.4	24.8	39.4	45.7	52.1	82.8	4.5	12.8	10
	TGGGTCCCGCCTTGCATCAACT	HE863227	0	0	0	2.7	0	0	0	0	1.6	0	0	0	0	0	0
	TGGTGCAGGTCGGGAACTGCT	HE860486	1.3	1.2	1.2	0	0	0	0	0	0	4.2	0	0	0	0	0
	TTGGTCGGTGGGTGCGAAATGGGT	HE863229	0	0	0	0	0	0	0	0	0	2.1	0	0	0	0	0
6_29	AAGCTCAGGAGGGATAGC	HE860487	1.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	AAGCTCAGGAGGGATAGCGC	HE863231	0	0	0	2.7	0	1.4	0	1.1	1.6	0	0	0	0	0	0
	AAGCTCAGGAGGGATAGCGCC	HE860388	70.2	64.8	86.7	178.9	218.7	118.4	153.7	149.7	173.8	211.8	254	224.7	141.5	158.6	113.4
	AGCTCAGGAGGGATAGCGCC	HE860487	0	1.2	0	5.4	1.6	0	0	0	0	2.1	0	0	1.1	0	0
	CGCTATCCATCCTGAGTTTC	HE863233	0	1.2	1.2	8.1	19	2.9	0	0	1.6	0	0	0	0	0	0
	CGCTATCCATCCTGAGTTTCA	HE860488	1.3	1.2	4.7	13.6	42.8	27.1	2.6	5.4	1.6	0	2.2	0	1.1	0	0
	TATTGCGCTATCCATCCTGAGTT	HE863235	0	0	0	0	0	0	1.3	0	0	0	0	0	0	0	0
	TCCATCCTGAGTTTCATGGCT	HE860488	1.3	0	1.2	0	0	0	0	0	0	0	0	0	0	0	0
	TTGCGCTATCCATCCTGAG	HE863237	0	0	0	0	0	0	1.3	0	0	0	0	0	0	0	0
6_30	AAGCTGCCAGCATGATCTGAGC	HE860489	0	0	0	0	0	0	0	0	0	0	0	0	1.1	4.8	0
	AGATCATGTGGTAGCTTCATC	HE863239	5	0	0	0	0	0	0	2.2	1.6	0	0	0	1.1	6.4	7.2
	CTAGATCATGTGGTAGCTTCATC	HE860489	1.3	0	0	0	0	0	1.3	0	1.6	0	0	0	1.1	1.6	0
	GAAGCTGCCAGCATGATCTG	HE863241	0	0	0	0	0	0	0	0	0	0	0	0	1.1	0	2.9
	GAAGCTGCCAGCATGATCTGA	HE860490	1.3	0	1.2	0	0	0	0	0	0	0	0	0	1.1	0	2.9
	GATCATGTGGTAGCTTCATC	HE863243	15.1	11	14.1	0	0	0	10.3	8.6	8.2	0	0	0	37.1	51.3	41.6
	GCTAGATCATGTGGTAGCTTCATC	HE860490	0	0	0	0	1.6	0	0	0	0	0	0	0	1.1	0	4.3
	TAGATCATGTGGTAGCTTCATC	HE863245	0	0	0	0	0	0	0	0	0	0	0	0	1.1	0	0
	TGAAGCTGCCAGCATGAT	HE860491	0	0	0	0	0	0	0	0	0	0	0	0	1.1	0	0
	TGAAGCTGCCAGCATGATC	HE863247	76.5	48.9	62.1	2.7	4.8	2.9	25.8	45.2	41	2.1	8.7	0	21.3	25.6	31.6
	TGAAGCTGCCAGCATGATCT	HE860284	170.6	172.5	161.8	19	17.4	30	153.7	153	88.6	10.4	15.2	14.8	35.9	36.9	48.8
	TGAAGCTGCCAGCATGATCTG	HE860399	1434.8	1105.9	1518.1	43.4	174.4	99.9	496	627	321.5	10.4	13	11.8	775.1	796.4	785.2
	TGAAGCTGCCAGCATGATCTGA	HE860400	1056.1	677.7	907.3	59.6	57.1	34.3	511.5	667.9	408.4	16.6	6.5	3	302.2	427.8	447.9
	TGAAGCTGCCAGCATGATCTGAGC	HE860491	2.5	0	0	0	0	0	1.3	2.2	1.6	0	0	0	2.2	1.6	2.9
	TGTTGAAGCTGCCAGCATGATC	HE863249	1.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6_4	AAGCTCAGGAGGGATAGC	HE860487	1.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0

	AAGCTCAGGAGGGATAGCGC	HE863231	0	0	0	2.7	0	1.4	0	1.1	1.6	0	0	0	0	0	0
	AAGCTCAGGAGGGATAGCGCC	HE860388	70.2	64.8	86.7	178.9	218.7	118.4	153.7	149.7	173.8	211.8	254	224.7	141.5	158.6	113.4
	AGCTCAGGAGGGATAGCGCC	HE860487	0	1.2	0	5.4	1.6	0	0	0	0	2.1	0	0	1.1	0	0
	CGCTATCTATCCTGAGTTTCA	HE860492	0	0	0	0	0	0	1.3	0	1.6	8.3	4.3	0	0	0	0
6_7	AATTACTACTTTTGAGTGGTTA	HE863251	1.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	ATCTTTCCCAATCCACCCA	HE860492	0	0	0	0	0	0	0	0	0	0	0	0	1.1	0	0
	ATCTTTCCCAATCCACCCATGCC	HE863253	10	12.2	3.5	2.7	1.6	1.4	11.6	16.2	11.5	10.4	6.5	3	14.6	33.6	20.1
	CATGGGTAAGTGGGGAAGA	HE860493	0	0	2.3	0	0	1.4	0	0	0	2.1	0	0	0	0	0
	CATGGGTAAGTGGGGAAGATG	HE863255	18.8	15.9	19.9	37.9	23.8	35.7	1.3	4.3	6.6	4.2	0	3	5.6	8	12.9
	CATGGGTAAGTGGGGAAGATGA	HE860493	5	6.1	3.5	2.7	6.3	2.9	2.6	4.3	0	2.1	6.5	0	2.2	1.6	1.4
	CTTTCCCAATCCACCCATGC	HE863257	0	0	0	0	0	0	0	0	0	2.1	0	0	0	0	0
	CTTTCCCAATCCACCCATGCC	HE860494	0	1.2	1.2	2.7	0	0	1.3	1.1	1.6	4.2	0	5.9	1.1	0	0
	TCCCAATCCACCCATGCC	HE863259	0	0	0	0	0	0	0	0	1.6	2.1	0	0	0	0	0
	TCTTTCCCAATCCACCCA	HE860494	3.8	4.9	3.5	2.7	3.2	1.4	0	1.1	9.8	8.3	0	8.9	0	11.2	7.2
	TCTTTCCCAATCCACCCAT	HE863261	1.3	0	0	0	0	0	0	1.1	0	0	0	0	1.1	0	0
	TCTTTCCCAATCCACCCATG	HE860495	1.3	0	0	0	0	0	2.6	0	0	2.1	0	0	1.1	0	0
	TCTTTCCCAATCCACCCATGC	HE863263	2.5	1.2	3.5	0	1.6	0	1.3	2.2	1.6	6.2	2.2	0	1.1	4.8	1.4
	TCTTTCCCAATCCACCCATGCC	HE860389	652.2	652	726.8	379.4	271.1	298.3	746.6	554.8	657.7	830.7	640.4	505.7	497.6	645.7	551.2
	TCTTTCCCAATCCACCCATGCCT	HE860495	5	0	0	0	0	0	2.6	0	1.6	0	0	0	1.1	0	1.4
	TGGCATGGGTAAGTGGGGAAGA	HE863265	0	0	0	0	0	0	0	0	0	2.1	0	0	1.1	0	0
	TTAGGTTTCCTCTTATTCATCC	HE860496	16.3	39.1	23.4	2.7	4.8	2.9	0	1.1	1.6	8.3	4.3	17.7	10.1	16	10
	TTCCCAATCCACCCATGCCT	HE863267	2.5	0	0	0	0	1.4	1.3	1.1	0	0	0	0	0	0	0
	TTCCCAATCCACCCATGCCTT	HE860496	0	1.2	1.2	0	1.6	0	0	1.1	4.9	0	2.2	0	1.1	1.6	2.9
	TTTCCCAATCCACCCATGCCT	HE863269	0	0	0	2.7	0	0	1.3	1.1	1.6	2.1	4.3	0	0	1.6	0
	TTTCCCAATCCACCCATGCCTT	HE860497	3.8	3.7	3.5	2.7	4.8	1.4	0	3.2	3.3	10.4	4.3	0	4.5	3.2	4.3
	TTTCCCAATCCACCCATGCCTTA	HE863271	0	1.2	1.2	0	4.8	1.4	1.3	2.2	1.6	0	2.2	3	0	4.8	1.4
	TTTCCTCTTATTCATCCCTCT	HE860497	0	0	0	0	0	0	1.3	0	0	0	0	0	0	0	0
7_23	AAGAAAGCTGTGGGAGAACAT	HE863273	0	0	0	2.7	0	0	0	0	0	0	0	0	0	0	0
	AAGAAAGCTGTGGGAGAACATGGC	HE860498	0	1.2	0	0	0	0	1.3	0	0	2.1	0	0	1.1	0	0
	CACAGCTTTCTTGAACTT	HE863275	1.3	3.7	1.2	2.7	3.2	1.4	0	0	0	0	0	0	0	0	1.4
	CCACAGCTTTCTTGAACT	HE860498	0	0	0	2.7	0	1.4	0	0	0	0	0	0	0	0	0
	CTCAAGAAAGCTGTGGGAGA	HE863277	2.5	1.2	4.7	2.7	0	4.3	0	0	0	2.1	2.2	0	1.1	4.8	4.3
	GCTCAAGAAAGCTGTGGGAGA	HE860499	6.3	18.4	10.6	8.1	6.3	14.3	1.3	1.1	1.6	6.2	2.2	8.9	5.6	8	5.7
	TATAAACAAGTCCTGGTCATGCTT	HE863279	0	0	0	2.7	0	0	0	0	0	0	0	0	0	0	0
	TCCACAGCTTTCTTGAACT	HE860499	1.3	0	0	2.7	0	0	0	1.1	0	0	2.2	0	0	0	1.4
	TCCACAGCTTTCTTGAACTT	HE863281	3.8	3.7	4.7	5.4	4.8	4.3	1.3	0	3.3	0	0	3	3.4	0	4.3
	TTCCACAGCTTTCTTGAA	HE860500	0	0	0	2.7	0	0	0	1.1	0	0	0	0	0	0	0
	TTCCACAGCTTTCTTGAAC	HE863283	10	15.9	12.9	27.1	46	37.1	2.6	4.3	4.9	2.1	0	5.9	4.5	4.8	5.7
	TTCCACAGCTTTCTTGAACT	HE860500	76.5	132.1	87.9	146.4	271.1	289.7	16.8	20.5	11.5	8.3	4.3	23.7	40.4	35.3	51.7
	TTCCACAGCTTTCTTGAACTT	HE860413	1138.8	1645.4	1659.9	1658.7	2919.8	2253.4	198.9	377.1	216.5	211.8	149.8	165.6	949.2	741.9	808.2
	TTCCACAGCTTTCTTGAACTTC	HE863285	2.5	0	2.3	5.4	0	4.3	0	1.1	0	0	0	0	3.4	0	1.4
7_24	ACACTGTGGCTCGTTGTGTTGTCA	HE860501	0	0	0	0	0	0	0	0	0	0	0	0	1.1	0	0

(Continued)

Table 2 | Continued

miRNA	Reads	EMBL accession number	BF1	BF2	BF3	F1	F2	F3	GF1	GF2	GF3	GL1	GL2	GL3	O1	O2	O3
	ACGTTATGTTGTCAAATTGTC	HE863287	0	0	0	0	0	0	1.3	1.1	0	0	0	0	0	0	0
	ATGTTGTCAAATTGTCAATC	HE860501	0	0	0	0	1.6	1.4	1.3	0	1.6	0	0	5.9	0	0	0
	CAACGTGACAACACAACGAGC	HE863289	1.3	0	1.2	0	0	2.9	0	0	1.6	0	0	0	0	0	0
	CAACGTGACAACACAACGAGCC	HE860502	10	15.9	12.9	8.1	7.9	12.8	2.6	5.4	6.6	8.3	0	0	0	1.6	2.9
	CACGTTATGTTGTCAAATTGTC	HE863291	0	0	0	2.7	0	1.4	0	0	0	0	0	0	0	0	0
	TATGTTGTCAAATTGTCAAT	HE860502	0	0	0	0	0	0	1.3	0	0	0	0	0	0	0	0
	TATGTTGTCAAATTGTCAATC	HE860414	16.3	9.8	24.6	24.4	12.7	20	24.5	21.5	29.5	10.4	17.4	14.8	12.4	4.8	5.7
	TGAACACAAAGATACATGCCCG	HE863293	0	0	0	0	0	0	1.3	0	0	0	0	0	0	0	0
	TTGACAACGTGACAACACAAC	HE860503	0	1.2	1.2	0	0	0	5.2	2.2	1.6	0	4.3	3	0	0	0
7_25	GACAGAAGAGAGTGAGCAC	HE863121	1.3	1.2	0	0	0	0	0	0	0	0	0	0	0	0	0
	GCTCACTTCTCTCTCTGTCAGC	HE863295	0	0	0	0	0	0	0	0	0	2.1	0	0	0	0	0
	TGACAGAAGAGAGTGAGCA	HE860460	1.3	0	1.2	0	0	0	0	0	0	0	0	0	0	0	0
	TGACAGAAGAGAGTGAGCAC	HE860311	71.5	80.7	97.3	181.6	38	44.2	29.7	53.9	31.2	49.8	17.4	11.8	11.2	11.2	20.1
	TGACAGAAGAGAGTGAGCACA	HE863125	2.5	3.7	0	0	1.6	0	0	0	1.6	2.1	0	0	0	0	0
8_16	CCGACAAGCGTGCTCTCTCTCGTT	HE860503	1.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	GTGCTCTCTCTTGTTGTCATG	HE863297	1.3	3.7	4.7	13.6	4.8	0	1.3	0	3.3	2.1	6.5	0	2.2	6.4	5.7
	TGACAACGAGAGAGAGCAC	HE860504	2.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	TGACAACGAGAGAGAGCACG	HE863299	0	0	0	2.7	0	1.4	0	0	3.3	0	0	0	0	0	0
	TGACAACGAGAGAGAGCACGC	HE860423	75.3	35.5	21.1	29.8	23.8	41.4	20.7	45.2	32.8	39.5	69.5	62.1	5.6	14.4	8.6
	TTGACAACGAGAGAGAGCAC	HE860504	2.5	1.2	1.2	0	0	1.4	0	0	1.6	0	2.2	0	0	0	1.4
	TTGACAACGAGAGAGAGCACG	HE863301	6.3	0	0	8.1	4.8	2.9	3.9	6.5	3.3	2.1	10.9	0	1.1	1.6	0
	TTGACAACGAGAGAGAGCACGC	HE860505	0	0	0	0	0	0	0	0	0	0	0	0	1.1	0	0
	TTGTCGGCACCCATGAAAGGGCCA	HE863303	0	0	0	2.7	0	0	0	0	0	0	0	0	0	0	0
	TTTGACAACGAGAGAGAGCAC	HE860505	2.5	0	1.2	2.7	0	4.3	1.3	3.2	1.6	4.2	6.5	3	1.1	1.6	0
8_19	AATGTCGTCTGGTTCGAGA	HE863305	0	0	1.2	2.7	3.2	1.4	0	0	0	0	0	0	0	1.6	0
	AATGTCGTCTGGTTCGAGATC	HE860506	1.3	0	0	0	0	1.4	0	0	0	0	0	0	0	0	0
	ATTTCGGACCAGGCTTCATTC	HE863307	3.8	3.7	3.5	2.7	1.6	4.3	1.3	0	0	0	0	3	0	0	0
	ATTTCGGACCAGGCTTCATTCCCC	HE860506	0	1.2	0	2.7	0	0	0	0	0	0	2.2	0	0	0	0
	CGGACCAGGCTTCATTCC	HE863097	0	0	0	0	0	0	0	1.1	0	0	2.2	0	2.2	0	0
	CGGACCAGGCTTCATTCCC	HE860454	1.3	2.4	0	2.7	1.6	0	1.3	5.4	1.6	0	0	3	1.1	0	0
	CGGACCAGGCTTCATTCCCC	HE863099	282.2	208	289.6	336.1	187	182.7	235.1	213.3	203.4	265.8	251.8	174.5	449.3	299.6	328.7
	CGGACCAGGCTTCATTCCCCT	HE863309	2.5	6.1	1.2	8.1	15.9	4.3	0	1.1	4.9	2.1	2.2	0	1.1	3.2	0
	GAATGTCGTCTGGTTCGAGA	HE860507	1.3	7.3	2.3	2.7	0	8.6	0	0	0	0	0	0	0	0	2.9
	GACCAGGCTTCATTCCCC	HE860456	1.3	0	0	0	0	0	0	0	1.6	0	0	0	0	0	0
	GACCAGGCTTCATTCCCCTCA	HE863311	0	0	0	2.7	0	0	0	0	0	0	0	0	0	0	0
	GATTTCGGACCAGGCTTCATTCCC	HE860507	1.3	0	0	2.7	0	0	0	0	0	0	0	0	0	0	0
	GGAATGTCGTCTGGTTCGA	HE863313	1.3	1.2	0	0	1.6	2.9	1.3	1.1	1.6	2.1	0	3	1.1	0	1.4
	GGAATGTCGTCTGGTTCGAGA	HE860508	20.1	15.9	19.9	21.7	28.5	35.7	2.6	5.4	8.2	4.2	0	3	3.4	4.8	5.7
	GGAATGTCGTCTGGTTCGAGAT	HE863315	0	0	0	2.7	0	0	0	0	0	0	0	0	0	0	0
	GGACCAGGCTTCATTCCC	HE860457	1.3	0	1.2	0	0	0	0	0	0	0	0	0	0	0	0
	GGACCAGGCTTCATTCCCC	HE863111	146.7	154.1	143	184.3	136.3	148.4	155	161.6	134.5	189	191	147.9	171.9	174.7	193.8
	GGGAATGTCGTCTGGTTCGAG	HE860508	1.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0

	TCGGACCAGGCTTCATTC	HE860457	58.9	64.8	65.6	43.4	28.5	31.4	42.6	54.9	39.4	24.9	28.2	29.6	47.2	54.5	40.2	
	TCGGACCAGGCTTCATTCC	HE863113	440.2	408.6	385.7	417.4	261.5	412.4	384.9	339.3	332.9	180.7	256.2	283.9	159.5	280.4	254.1	
	TCGGACCAGGCTTCATTCCC	HE860458	706.1	675.3	720.9	441.8	321.8	449.5	586.5	627	546.1	388.4	525.4	387.4	410	512.8	541.2	
	TCGGACCAGGCTTCATTCCCC	HE860285	282744.1	2797179	296598.2	273500	243934.6	239360.7	306036.6	275122.8	248117.2	282794.4	3175572	1996375	217491.6	348033.3	350352.6	
	TCGGACCAGGCTTCATTCCCCT	HE863317	37.6	41.6	39.9	27.1	12.7	21.4	59.4	47.4	41	31.2	28.2	35.5	33.7	32	40.2	
	TCGGACCAGGCTTCATTCCCCTC	HE860509	0	0	1.2	0	0	0	0	0	0	2.1	0	0	0	0	0	
	TTCGGACCAGGCTTCATTCC	HE863319	1.3	0	2.3	0	4.8	1.4	0	1.1	0	0	0	0	0	0	0	
	TTCGGACCAGGCTTCATTCCC	HE860509	184.4	223.9	195.8	273.7	416.9	299.7	49.1	62.5	47.6	24.9	17.4	23.7	35.9	49.7	43.1	
	TTCGGACCAGGCTTCATTCCCC	HE863321	153	168.8	158.3	219.5	280.6	191.2	117.5	101.3	95.1	27	32.6	20.7	35.9	57.7	44.5	
	TTCGGACCAGGCTTCATTCCCCT	HE860510	1.3	0	0	0	1.6	0	0	0	0	0	0	0	0	0	0	
	TTGAGGGGAATGTCGTCTGG	HE863323	1.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	TTTCGGACCAGGCTTCATTCC	HE860510	67.7	104	85.6	103	187	114.2	41.3	25.9	24.6	4.2	6.5	20.7	14.6	11.2	15.8	
8_21	CACGTGCTCCCCTTCTCC	HE863325	0	0	0	0	0	0	0	0	0	2.1	0	0	0	0	0	
	CACGTGCTCCCCTTCTCCAAC	HE860511	2.5	1.2	2.3	0	4.8	10	10.3	4.3	11.5	12.5	17.4	20.7	3.4	1.6	4.3	
	TGGAGAAGCAGGGCACGTGCA	HE860424	55.2	30.6	24.6	62.3	7.9	45.7	14.2	28	21.3	20.8	17.4	29.6	6.7	8	4.3	

BF, pink; F, bloom; GF, swollen flower bud; O, half-inch green; GL, swollen leaf bud.

by ENCODE Consortium.[4] Average correlation coefficients were equal to 0.98 for BF, 0.95 for F, 0.98 for GF, 0.95 for GL, and 0.97 for O. For the sake of completeness and in order to allow a comparison between related and unrelated samples, we also calculated the average Pearson correlation between samples of different tissues, which was equal to 0.66 on the basis of the reads reported in **Table 2**. All the Pearson coefficients are reported in File S10 in Supplementary Material. **Figure 1A** reports the results obtained from clustering the five samples on the basis of all the reads frequencies (average frequencies of three replicates, reads included miRNA*-related reads; reads assigned by miRDeep-P to more than one locus were counted once) at the 26 loci analyzed. Additionally, a clustering analysis was performed by considering only the count of the most frequent read in each locus. The analysis included those loci where the most frequent read was the same in all the samples (16 different reads, **Figure 1B**). **Figure A1** in Appendix reports clustering results obtained without averaging the three replicates of each sample. As it can be seen, replicates are always grouped correctly.

A *t*-test was also performed for all the possible comparisons of biological samples (File S11 in Supplementary Material). The most frequent isomiR (highlighted in yellow in File S11 in Supplementary Material) is frequently the one able to distinguish the higher number of samples (e.g., locus 4_21, locus 6_4). Some miRNA-related reads are able to differentiate most of the analyzed samples: e.g., miR398 and miR167 got 8 significant comparisons out of 10.

DISCUSSION

To assess the putative biological significance of isomiRs in peach, in the present study we carried out miRNAs profiling by sequencing three replicates of five biological samples arising from a set of different organs and/or phenological stages. Actually, variants of miRNAs are commonly found in deep sequencing experiments but their functional meaning and stability is still under investigation in plants.

Twenty-six miRNA putative loci expressed in all samples analyzed have been identified by miRdeep-P and analyzed for miRNA population heterogeneity. The average length of miRNA associated reads was included between 18 nt and 24 nt. Several previous works reported a miRNA length in plants included between 22 nt and 24 nt. The identification of miRNA* associated reads provides more evidence about reliability of the loci identified by miRDeep-P.

All the analyzed loci show miRNA length variants but tend to maintain the uridine at the 5′ end, in those cases where uridine is the first base of the most abundant isomiR. As reported above, uridine is the most frequent nucleotide in AGO1 association, perhaps explaining the drive to maintain it at the 5′ end. Ebhardt et al. (2009) reported examples of miRNA with 5′ deletions and 3′ uridine additions that create a different distribution in AGO complexes. As an example, ath-miR822 was determined to reside almost exclusively in the AGO1 complex while its modified variant with a U deletion at 5′ end and a UU addition at 3′end was found equally in AGO1 and AGO4 complexes.

The difference in read count between the first most frequent read and the second most frequent read varies among loci being

[4]http://encodeproject.org/ENCODE/protocols/dataStandards/RNA_standards_v1_2011_May.pdf

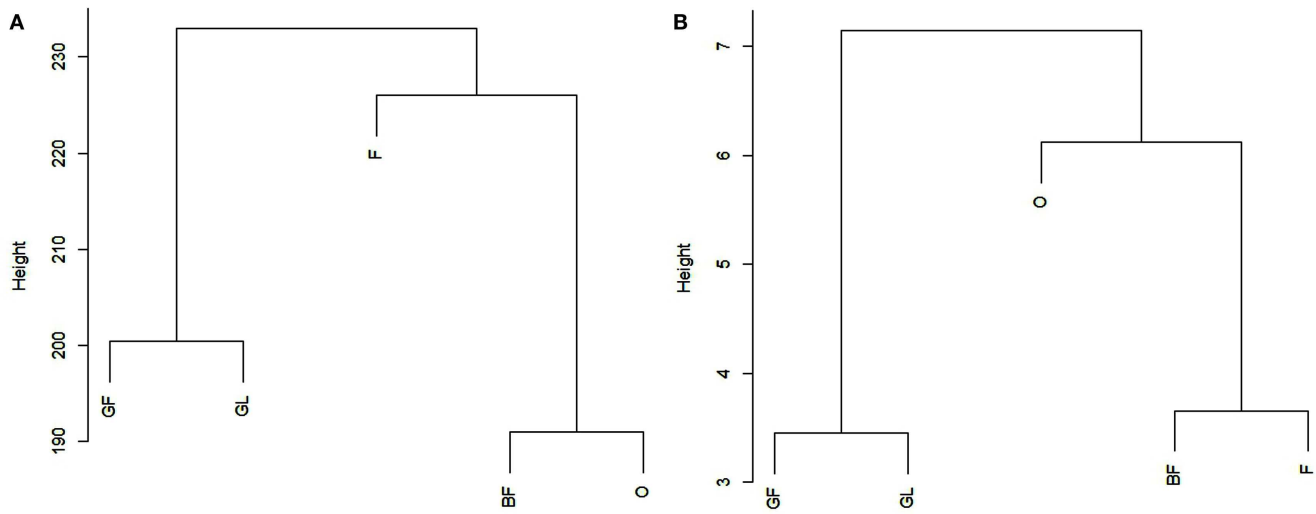


FIGURE 1 | Reports the results obtained by cluster analysis of the five tissues on the basis of the frequencies of all the reads (A) or on the basis of the frequency of the most frequent read in each locus (B). For each tissue, the average frequency across the three replicates was considered (miRNA* -related reads are included). In **(B)**, the analysis included only those loci where the most frequent read was the same in all the samples. BF, pink; F, bloom; GF, swollen flower bud; O, half-inch green; GL, swollen leaf bud.

in some cases minimal (e.g., locus 1_5) while in some others it is quite consistent (e.g., loci 4_21, 6_4). In some loci the second most frequent read was the same in all the replicates of a sample and in all the different samples. The presence of the same isomiRs in different biological replicates of a sample and in different tissues demonstrate that the generation of most of the detected isomiRs is not random. The importance of evaluating the correlation between biological replicates from RNA-seq experiments has been discussed previously in several papers (Oshlack et al., 2010; Hansen et al., 2011). As above reported, the correlation among biological replicates has been calculated to check the reliability of the experiment on the basis of the "Standards, guidelines, and best practices for RNA-seq" adopted by ENCODE Consortium which requires that the Pearson correlation of gene expression between two biological replicates for RNAs that are detected in both samples using RPKM or read counts should be between 0.92 and 0.98. Regarding the present work, the average Pearson correlation between all the possible pairs of replicates belonging to the same biological sample was greater than or equal to 0.95 for all the tested samples, in agreement with the required standards. Clustering results and *t*-test reported in **Figure 1** and File S11 in Supplementary Material, respectively, show that it is possible to clearly distinguish among samples and to group them in a functional way. However, when considering **Figure A1** in Appendix obtained without averaging replicates of each sample, it should be noted that clustering results seem to be more confident when only the most frequent read is taken into account: BF (pink) and F (bloom) are more strictly related being two subsequent phenological stages so it is expected to find a closer relationship between them.

The co-existence of different variants with a similar level of expression could imply a biological role for all of them. Locus 1_26 shows such an example: in this case there are two prevalent isomiRs (HE860305 and HE860450) that differ for one T at the 5′ end. For both the isomiRs there are then variants at the 3′ end with different lengths.

Target analysis carried out by psRNATarget (File S9 in Supplementary Material) revealed that in many cases isomiRs share the same target. However, because AGO invariably catalyzes the cleavage of targets opposite the bond between nucleotides 10 and 11 from the 5′ end of the miRNA, the cleavage products are different when there is a shift toward the 5′ end or nucleotide addition at the 5′ end of the miRNA mature sequence. Differences in cleavage sites among members of the same miRNA family have been recently studied in rice by Jeong et al. (2011) highlighting a different abundance of specific cleavage sites among plant organs.

A very interesting finding is related to the biological role of miRNA*. Despite the general consensus that miRNAs* have no regulatory activity, several recent publications have provided evidence about their biological function (Mah et al., 2010). In our results, isomiRs have been found also for miRNAs*. As an example, at locus 3_16 the conserved miRNA* has a high number of length variants, most due to a variable 3′end. Locus 3_16 codes for miR482: the miRNA* sequence deposited in miRBase was actually the most frequent read (HE860347) in all the three replicates of sample O (half-inch green) with an average ratio miRNA/miRNA* equal to 0.4. GF and GL showed an average ratio of miRNA/miRNA* equal to 5.7, while in BF and F the ratio was close to one in two out of three replicates. Similar results have been previously found in mammals by Kuchenbauer et al. (2012) that classified miRNA/miRNA* ratios into groups showing that about 50% of all miRNA duplexes revealed high ratios (>100) consistent with a strong preferential processing of one dominant miRNA strand. About 24% had intermediate ratios (between 100 and 10), about 13% showed low ratios (between 10 and 1), while another 13% showed inverted ratios (<1). The finding that miRNAs can

display tissue-dependent miRNA arm selection opposes the general consensus that only one strand is highly dominant for any given miRNA duplex and opens insights into the possible biological function of selective accumulation of miRNA*. A recent review of Sunkar et al. (2012), discusses several studies showing that miRNA* tend to accumulate at a high level under particular conditions. As an example, miR393* accumulates at a high level during infection of *P. syringae* in *Arabidopsis* leaves and promotes plant resistance to bacterial infection. Mir399* is accumulated at high levels during phosphate deprivation in *Arabidopsis* and miR395* accumulates at high levels in *Sorghum* grown in optimal nutrient conditions.

PsRNATarget has been used to investigate possible target genes for miR482 and miR482* at locus 3_16. MiRNA482 target a peach sequence coding for a "probable receptor-like protein kinase" (expectation = 2, target accessibility = 17.288), while miRNA482* targets a NADH dehydrogenase gene (expectation = 3, target accessibility = 8.463). Examples of different targets for a pair of miRNA/miRNA* are reported in previous studies (Sunkar et al., 2012). Mir393 and miR393* target two entirely different gene families (TIR1 and SNARE) both involved in pathogen resistance of host plant. The possibility that a target-dependent strand selection based on the presence in the cell of miRNA or miRNA* targets might influence the selection of the active miRNA arm has been discussed by other authors. For instance Chatterjee and Grosshans (2009) reported that mRNAs can stabilize their cognate miRNAs thus suggesting coordinated RISC assembly which depends on a miRNA and its target levels.

Results obtained in the present work contribute to a deeper view of the miRNome complexity and to a better exploitation of the mechanism of action of these tiny regulators. The exact definition of the entire repertoire of peach miRNAs is in fact a prerequisite for a correct description of miRNAs whose expression is altered in response to specific developmental conditions or environmental stimuli. Future experiments based on small RNA-seq coupled with RNA-seq on the same samples will be carried out to highlight more clearly the possible biological role of miRNA isomiRs in plants.

ACKNOWLEDGMENTS

We thank Keith Anthony Grimaldi for helping with the preparation of the manuscript. The present work has been supported by Drupomics Project (Italian Ministry for Agriculture). We acknowledge the International Peach Genome Initiative for pre-publication access to the peach genome sequence.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/Plant_Genetics_and_Genomics/10.3389/fpls.2012.00165/abstract

File S1 | Reports the miRNA coding loci identified by miRDeep-P in pink sample.

File S2 | Reports the miRNA coding loci identified by miRDeep-P in bloom sample.

File S3 | Reports the miRNA coding loci identified by miRDeep-P in swollen flower bud sample.

File S4 | Reports the miRNA coding loci identified by miRDeep-P in half-green sample.

File S5 | Reports the miRNA coding loci identified by miRDeep-P in swollen leaf bud sample.

File S6 | Reports a summary of the miRNA coding loci identified by miRDeep-P.

File S7 | Reports the link between locus name and locus position.

File S8 | Reports the results of the blast analysis against known plant miRNAs.

File S9 | Reports target analysis for all the identified isomiRs.

File S10 | Reports Pearson correlation coefficients between all the possible pairs of replicates belonging to the same biological sample, as well as samples from different tissues.

File S11 | Reports the results of the *t*-test which was performed for all the possible comparisons of biological samples. The most frequent isomiR in each locus is highlighted in yellow.

REFERENCES

Chapman, P. J., and Catlin, G. A. (1976). Growth stages in fruit trees from dormant to fruit set. *N. Y. Food Life Sci. Bull.* 58.

Chatterjee, S., and Grosshans, H. (2009). Active turnover modulates mature miRNA activity in *Caenorhabditis elegans*. *Nature* 461, 546–549.

Dai, X., and Zhao, P. X. (2011). PsRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res.* 39, W155–W159.

Eamens, A. L., Smith, N. A., Curtin, S. J., Wang, M. B., and Waterhouse, P. M. (2009). The *Arabidopsis thaliana* double-stranded RNA binding protein DRB1 directs guide strand selection from microRNA duplexes. *RNA* 15, 2219–2235.

Ebhardt, H. A., Tsang, H. H., Dai, D. C., Liu, Y., Bostan, B., and Fahlman, P. (2009). Meta-analysis of small RNA-sequencing errors reveals ubiquitous post-transcriptional RNA modifications. *Nucleic Acids Res.* 37, 2461–2470.

Guo, L., and Lu, Z. (2010). Global expression analysis of miRNA gene cluster and family based on isomiRs from deep sequencing data. *Comput. Biol. Chem.* 34, 165–171.

Hansen, K. D., Wu, Z., Irizarry, R. A., and Leek, J. T. (2011). Sequencing technology does not eliminate biological variability. *Nature Biotechnol.* 29, 572–573.

Jeong, D.-H., Park, S., Zhai, J., Gurazada, S. G. R., de Paoli, E., Meyers, B. C., and Green, P. J. (2011). Massive analysis of rice small RNAs: mechanistic implications of regulated microRNAs and variants for differential target RNA cleavage. *Plant Cell* 23, 4185–4207.

Jones-Rhoades, M. W., Bartel, D. P., and Bartel, B. (2006). MicroRNAs and their regulatory roles in plants. *Annu. Rev. Plant Biol.* 57, 19–53.

Kozomara, A., and Griffith-Jones, S. (2011). miRBase: Integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* 39, D152–D157.

Kuchenbauer, F., Mah, S. M., Heuser, M., McPherson, A., Rüschmann, J., Rouhi, A., Berg, T., Bullinger, L., Argiropoulos, B., Morin, R. D., Lai, D., Starczynowski, D. T., Karsan, A., Eaves, C. J., Watahiki, A., Wang, Y., Aparicio, S. A., Ganser, A., Krauter, J., Döhner, H., Döhner, K., Marra, M. A., Camargo, F. D., Palmqvist, L., Buske, C., and Humphries, R. K. (2012). Comprehensive analysis of mammalian miRNA* species and their role in myeloid cells. *Blood* 118, 3350–3358.

Lee, L. W., Zhang, S., Etheridge, A., Ma, L., Martin, D., and Galas, D. (2012). Complexity of the microRNA repertoire revealed by next-generation sequencing. *RNA* 16, 2170–2180.

Mah, S. M., Buske, C., Humphries, R. K., and Kuchenbauer, F. (2010). miRNA*: a passenger stranded in RNA-indiced silencing complex? *Crit. Rev. Eukaryot. Gene Expr.* 20, 141–148.

Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C., Chen, S., Hannon, G. J., and Qi, Y. (2008). Sorting of small RNAs into *Arabidopsis* argonaute complexes is directed by the 5′ terminal nucleotide. *Cell* 133, 116–127.

Moxon, S., Schwach, F., Dalmay, T., MacLean, D., Studholme, D. J., and Moulton, V. (2008). A toolkit for analyzing large-scale plant small RNA datasets. *Bioinformatics* 24, 2252–2253.

Mückstein, U., Tafer, H., Hackermüller, J., Bernhart, S. H., Stadler, P. F., and Hofacker, I. L. (2006). Thermodynamics of RNA–RNA binding. *Bioinformatics* 22, 1177–1182.

Oshlack, A., Robinson, M. D., and Young, M. D. (2010). From RNA-seq reads to differential expression results. *Genome Biol.* 11, 220.

Sunkar, R., Li, Y.-F., and Jagadeeswaran, G. (2012). Functions of microRNAS in plant stress responses. *Trends Plant Sci.* 17, 196–203.

Takeda, A., Iwasaki, S., Watanabe, T., Utsumi, M., and Watanabe, Y. (2008). The mechanism selecting the guide strand from small RNA duplexes is different among argonaute proteins. *Plant Cell Physiol.* 49, 493–500.

Vaucheret, H. (2009). AGO1 homeostasis involves differential production of 21-nt and 22-nt miR168 species by miR168a and miR168b. *PLoS ONE* 4, e6442. doi:10.1371/journal.pone.0006442

Voinnet, O. (2009). Origin, biogenesis, and activity of plant microRNAs. *Cell* 136, 669–687.

Xie, Z., Khanna, K., and Ruan, S. (2010). Expression of microRNAs and its regulation in plants. *Semin. Cell Dev. Biol.* 21, 790–797.

Yang, X., and Li, L. (2011). miRDeep-P: A computational tool for analyzing the microRNA transcriptome in plants. *Bioinformatics* 27, 2614–2615.

Zhang, Y. (2005). miRU: An automated plant miRNA target prediction server. *Nucleic Acids Res.* 33(Suppl. 2), W701–W704.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 03 April 2012; accepted: 04 July 2012; published online: 26 July 2012.
*Citation: Colaiacovo M, Bernardo L, Centomani I, Crosatti C, Giusti L, Orrù L, Tacconi G, Lamontanara A, Cattivelli L and Faccioli P (2012) A survey of microRNA length variants contributing to miRNome complexity in peach (Prunus persica L.). Front. Plant Sci. **3**:165. doi: 10.3389/fpls.2012.00165*
This article was submitted to Frontiers in Plant Genetics and Genomics, a specialty of Frontiers in Plant Science.

APPENDIX

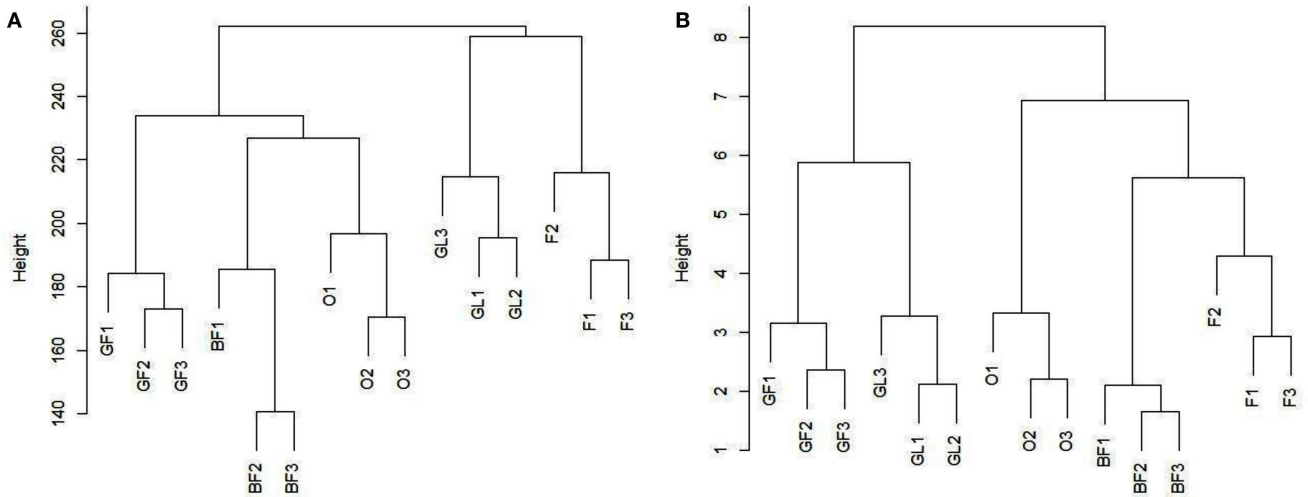


FIGURE A1 | Reports the results obtained by cluster analysis as in Figure 1 but in this case without averaging the replicates of each sample. In details, **(A)** reports the results obtained by cluster analysis of the five tissues on the basis of the frequencies of all the reads, while **(B)** reports results obtained on the basis of the frequency of the most frequent read in each locus. Abbreviations: pink = BF, bloom = F, swollen flower bud = GF, half-inch green = O, swollen leaf bud = GL.

5.3 Experiment 2 – *Barley response to low temperatures*

Introduction

Barley (*Hordeum vulgare*) is an economically important diploid model for the *Triticeae* and a better understanding of low temperature tolerance mechanisms could significantly improve the yield of cereals. Temperature fluctuations over the course of each day and during changes in the seasons are common and plants have to reprogram their gene expression profile to adjust to such shifts in temperature. Transcriptional control of the expression of cold-responsive genes is well known (Chinnusamy *et al.*, 2010) and miRNAs have recently been added to the suite of cold responsive gene regulatory networks. In this regard, in Chapter 8 (Annexes) we report our recently published review dealing with post-transcriptional and post-translational modifications controlling cold response. It contains a sub-chapter about degradation of stress-related transcripts by small RNAs.

Two barley cultivars, Nure and Tremois, were chosen to identify microRNAs involved in the response to low temperature stress. They are parents of a mapping population that has been used in many previous studies on low-temperature tolerance in barley (Francia *et al.*, 2004; Marè *et al.*, 2004; Tondelli *et al.*, 2006; Stockinger *et al.*, 2007; Francia *et al.*, 2007; Limin *et al.*, 2007). These two cultivars differ for vernalization requirement and cold tolerance, being Nure cold-tolerant with a winter growth habit, and Tremois cold-susceptible with a spring growth habit.

It has been proved that the difference between Nure and Tremois is at least partially due to many duplication events that occurred in the genomic region harboring CBF (C-repeat Binding Factor) loci in frost-tolerant genotypes (Knox *et al.*, 2010; Tondelli *et al.*, 2011). Another possible explanation refers to the fact that winter-type barley is able to express CBFs at a higher rate compared to spring-type barley (Skinner *et al.*, 2005; Stockinger *et al.*, 2007; Tondelli *et al.*, 2011).

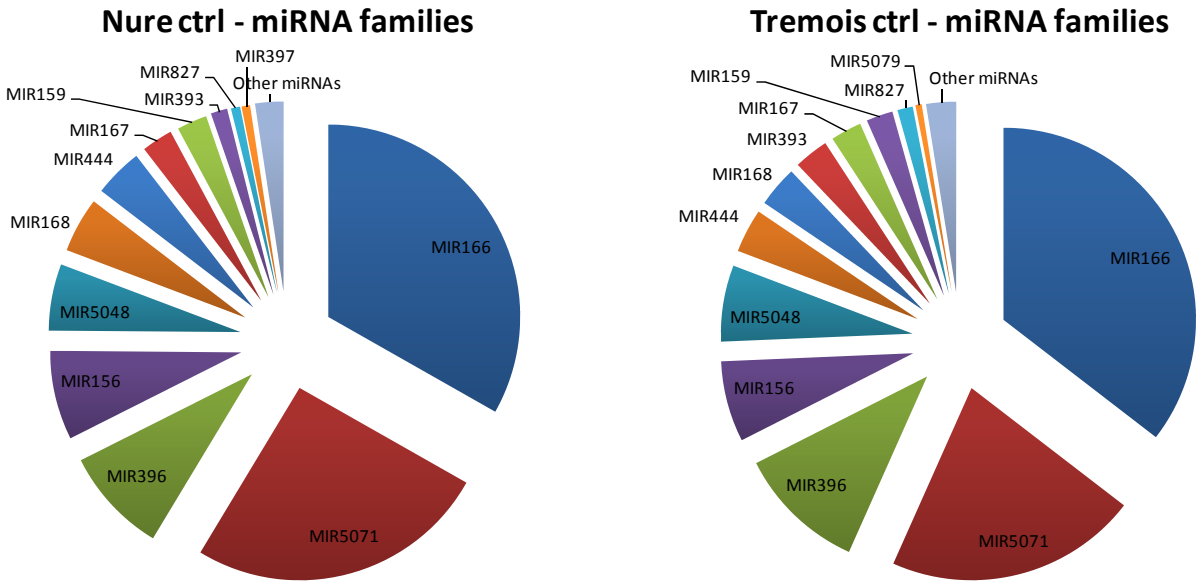
The CBF transcription factors and their target genes, the so-called “CBF regulon”, have been described as a fundamental part of the signal cascade leading to acclimation and acquisition of frost tolerance in many different plant species. Genes specific of this pathway encode, for example, other transcription factors, late embryogenesis abundant (LEA) proteins, cold-regulated (COR) and cold-inducible (KIN) proteins, osmoprotectant biosynthesis proteins, carbohydrate metabolism-related proteins, phospholipase C enzymes, sugar transport proteins. In addition to the CBF pathway components, many other key regulatory genes involved in the switch from the vegetative to the reproductive phase of fall-sown crops have been described, thus showing the interconnection between frost resistance and vernalization requirement in temperate cereals (Tondelli *et al.*, 2011; reviewed in Galiba *et al*, 2009).

Deep sequencing – miRNA families

Deep sequencing allowed the identification of many microRNAs modulated in response to low temperatures, and, interestingly, significant differences between the two cultivars were highlighted. They regarded both constitutive expression in control plants and regulation during treatment.

As for miRNA families, their number was different (42 in Nure, 36 in Tremois), but little diversity could be seen in their overall abundance between the two cultivars. Sequences belonging to *miR166*, *miR5071*, *miR396*, *miR156* and *miR5048* families account for more than 80% of the total miRNA reads (Graphs 1 and

2). As it can be noticed, two of the most expressed families out of five are specific of the *Triticeae* tribe (*miR5071* and *miR5048*), whereas the others are widely conserved among higher plants.



Graphs 1 and 2 *miRNA families abundance in the two cultivars. Abundance is expressed as a percentage of the total microRNA read number.*

If little diversity can be seen among the most expressed miRNA families, some constitutive differences between Nure and Tremois are evident in two cases in particular (Figure 11, right side).

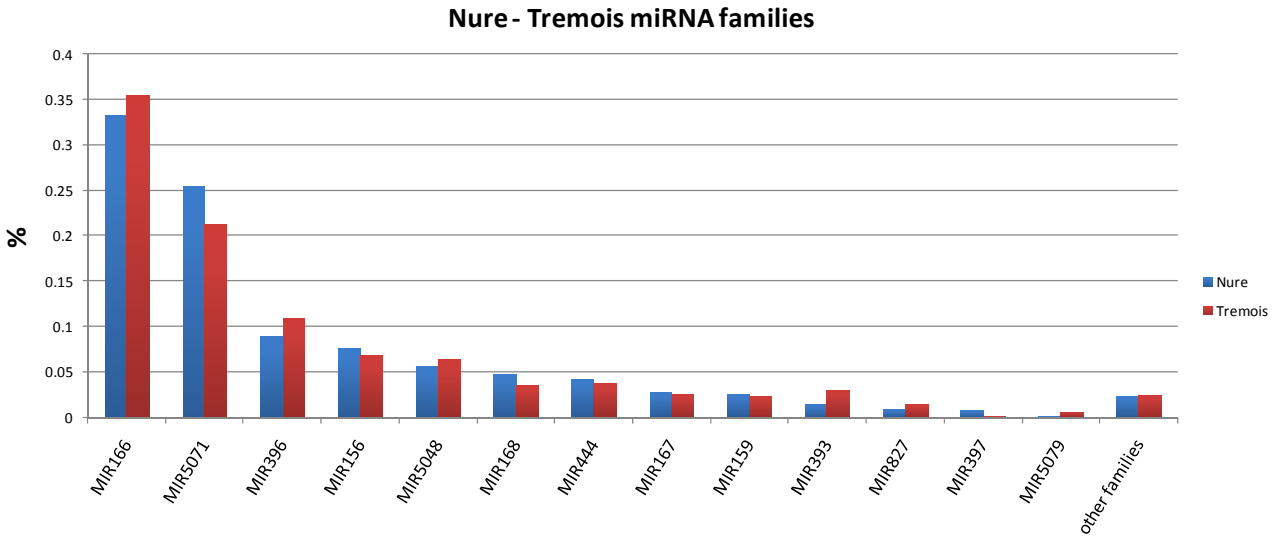


Fig 11 *Direct comparison of microRNA families' abundance between the two cultivars. miR397 and miR5079 families (right side) showed a significant differential expression that was judged worthy of further investigations.*

miR397 and *miR5079* have middle-low expression level (less than 1% of the total miRNA population) but still worthy of further investigations. The former has two isoforms almost equally expressed, 21 and 24 nucleotides long, probably arising from the same precursor. Their total expression is about 450 RPM (Read Per Million) in Nure, but it falls down in Tremois (13 RPM on average in control samples). The latter has many isoforms – 6 in Tremois, 3 in Nure; the most abundant ones are 22- and 23- nt long. All the reads belonging to the family are about 176 RPM in total, on average, in Tremois, versus 38 RPM in Nure. These

findings were subsequently tested in a wider panel of genotypes, to verify a possible correlation with the growth habit (see "Further tests").

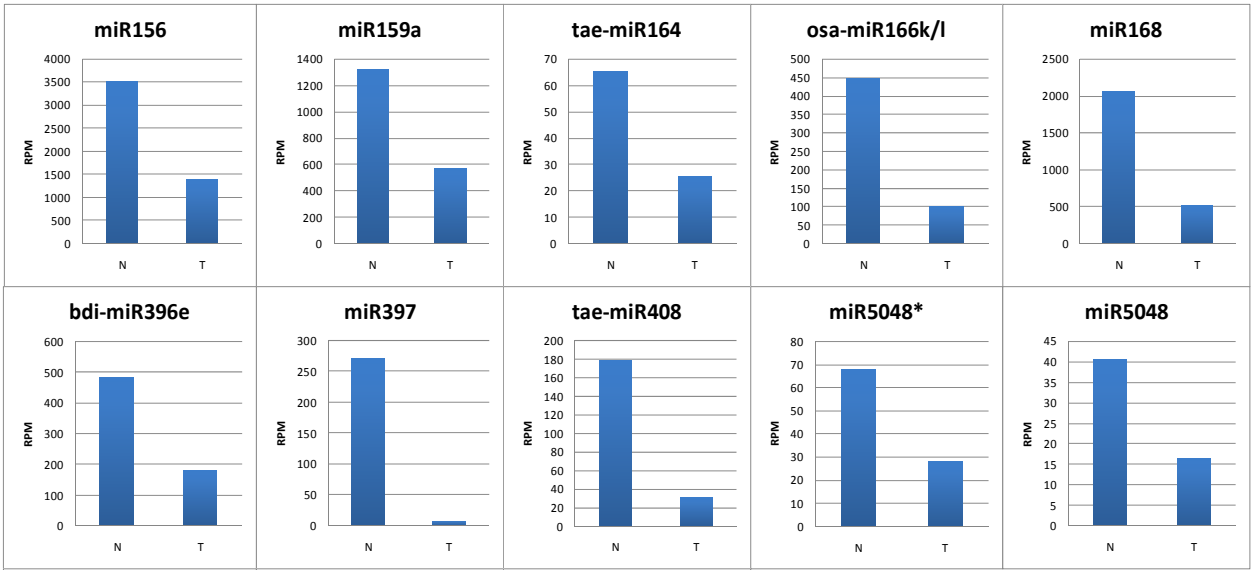
Finally, 8 miRNA families were found in Nure only (*miR395*, *miR437*, *miR528*, *miR1117*, *miR5050*, *miR5054*, *miR5067*, *miR5072*) and 2 in Tremois only (*miR1318*, *miR6201*). They globally make up the 0.1% of the total miRNA reads of the database.

Deep Sequencing – Differential constitutive expression of microRNAs

Many other microRNAs showed a differential expression between the control samples of the two varieties, though no significant overall change could be seen in their respective families. Interestingly, results showed that Nure constitutively overexpress many microRNAs compared to Tremois, which in turn keeps only one miRNA, *i.e.* the above-mentioned *miR5079*, at a higher level than Nure. Results are reported in the following Table and in the following Figure.

Name	Sequence	Lenght	Source	Mismatches
miR156	TGACAGAAGAGAGTGAGCAC	20	***Hordeum vulgare***	0
miR159a	TTTGGATTGAAGGGAGCTCTG	21	***Hordeum vulgare***	0
miR164	TGGAGAAGCAGGGCACGTGCA	21	*Triticum aestivum*	0
miR166k/miR166l	TCGGACCAGGCTTCAATCCC	20	*Oryza sativa*	0
miR168	TCGCTTGGTGCAGATCGGGAC	21	***Hordeum vulgare***	0
miR396	TTCCACAGCTTTCTTGAACTT	21	*Brachypodium distachyon*	0
miR397	TTGAGTGCAGCGTTGATGAAC	21	***Hordeum vulgare***	0
miR408	TGCACTGCCTCTTCCCTGGC	20	*Triticum aestivum*	0
miR5048*	TTTAGACCTAGACATGCAAGT	21	***Hordeum vulgare***	0
miR5048	TATATTTGCAGGTTTTAGGTC	21	***Hordeum vulgare***	0
miR5071	TCAAGCATCATGTCATGGACA	21	*Oryza sativa*	2
miR5079-1	TTTGAGAACCCCTTGAGCGCCT	22	*Oryza sativa*	1
miR5079-2	ATTTGAGAACCCCTTGAGCGCCT	23	*Oryza sativa*	1

Table 5 *Differentially expressed microRNAs in the control samples of the two barley cultivars. Source organisms used to annotate microRNAs during bioinformatic analyses are reported along with respective mismatches with the deposited sequence in miRBase. Sequences already deposited in miRBase for barley are highlighted in bold. Sequences reported in normal type are not yet reported for barley.*



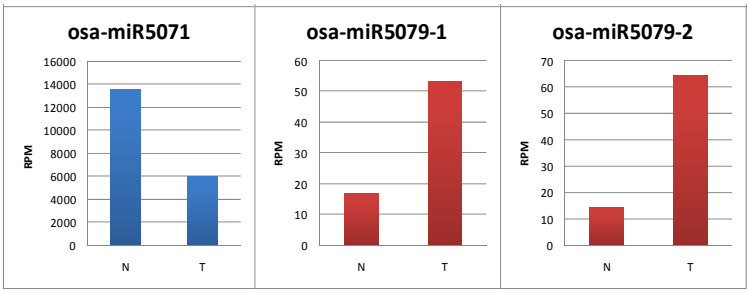


Fig. 12 *Differentially expressed microRNAs in the control samples of Nure (N) and Tremois (T). Normalized expression values (RPM) are reported. The expression level of some of these microRNAs was tested in a wider panel of genotypes, to test a possible correlation with traits explaining the difference between the two cultivars.*

Further tests

Some of the microRNAs which showed a constitutive different expression profile between Nure and Tremois were further tested in a wider panel of genotypes comprising winter, spring and facultative type cultivars (Table 6). Barley varieties of the latter kind flower quickly in long day conditions, but their floral induction is more or less inhibited by short day conditions.

#	Cultivar	Growth habit	Row type	Released	Country of origin	Pedigree	Breeding Institute
1	**Nure**	**Winter**	**2-row**	**1999**	**Italy**	**(Fior 40 x Alpha) x Baraka**	**CRA -Fiorenzuola**
2	**Tremois**	**Spring**	**2-row**	**1989**	**France**	**(Dram x Aramir) x Berac**	**Verneuil**
3	Arda	Winter	2-row	1985	Italy	Igri x HJ 51-15-3	CRA -Fiorenzuola
4	Fridericus	Winter	6-row	2006	Germany	Carola x LP 6-564	KWS
5	Frost	Winter	6-row	1988	Sweden	Pella x Astrix	Weibull
6	Heligan	Winter	2-row	1998	United Kingdom	Tosca x Intro	CPB Twyford
7	Ketos	Winter	6-row	2002	France	(Gotic x Orblonde) x (12813 x 91H595)	Limagrain
8	Labea	Winter	2-row	1992	Germany	Br.269 c x LBP 5906	Breun
9	Lonni	Winter	6-row	2005	Denmark	Aviron x Carola	Pajbjergfonden
10	Manolia	Winter	2-row	2001	France	Labea x Astrid	Secobra
11	Pearl	Winter	2-row	1997	United Kingdom	Puffin x Angora	Nickerson
12	Trebbia	Winter	6-row	1990	Italy	Selection from Fior Synt 3	CRA
13	Vogelsanger Gold	Winter	6-row	1965	Germany	(Isaria x H.spontaneum nigrum) x Wintergerste-5	Hauptsaaten
14	Balder	Spring	2-row	1942	Sweden	(Gull x Scanian Barley) x Maja	Weibull
15	Gitane	Spring	2-row	1976	Netherlands	Zephyr x (C I 1237 x Cambrinus)	Wiersum
16	Beatrix	Spring	2-row	2004	Germany	Viskosa x Pasadena	Nordsaat
17	Publican	Spring	2-row	2006	United Kingdom	Drum x Sebastian	Syngenta
18	Kilta	Spring	6-row	1981	Finland	(Seger x Vega) x Bonus	Boreal Plant Breeding Ltd
19	Alliot	Spring	2-row	1999	Denmark	Chariot x Alexis	Pajbjergfonden
20	Triumph	Spring	2-row	1973	Germany	Diamant x St. 14024-64/6	Hadmersleben
21	Atribut	Spring	2-row	1996	Czech Republic	KM V 3/83 x BR 2174	Selgen Stupice
22	Proctor	Spring	2-row	1953	United Kingdom	Plumage Archer x Kenia	PBI
23	Ceylon	Spring	2-row	2003	Netherlands	Portia x Amber	Cebeco
24	Rupal	Spring	2-row	1972	Sweden	Pallas*3 x Rupee	Svalof
25	Pamina	Facultative	6-row	1969	Germany	(Hohenthurmer x Mahnd.Viktoria) x Dominator	
26	Tarm92	Facultative	2-row	1993	Turkey	Tokak x Landrace 4875	FCCRI-A
27	Tidone	Facultative	2-row	1994	Italy	(Okos x 273 cat.) x Igri	CRA -Fiorenzuola
28	U259	Facultative	6-row	-	Hungary	-	-

Table 6 *Expanded panel of winter-type, spring-type and facultative cultivars for testing of microRNAs showing a constitutive different expression profile.*

miR166, *miR168*, *miR397*, *miR408* and *miR5079* expression level was assessed in 7-day-old leaves of these barley varieties via a stem-loop qRT-PCR (see Chapter 3). Their profiles did not match any possible hypothesis of correlation with plant vernalization requirement, or any other kind of grouping, except, partially, in the case of *miR397*.

Results are reported in figure 13. The spring-type cultivars (green bars) are generally characterized by a common trend, *i.e.* a lower *miR397* expression, compared to the winter-type group (blue bars). Facultative cultivars (yellow bars) tend to behave in a different way, sometimes showing an intermediate expression

level (Tidone, U259, Tarm92), or a relatively high level (Pamina is close to Nure, the genotype with the highest *miR397* expression observed). As it can be noticed, though, some winter genotypes (Manolia, Pearl, Labea and Heligan) do not respect this correlation, being not significantly different from spring-type ones.

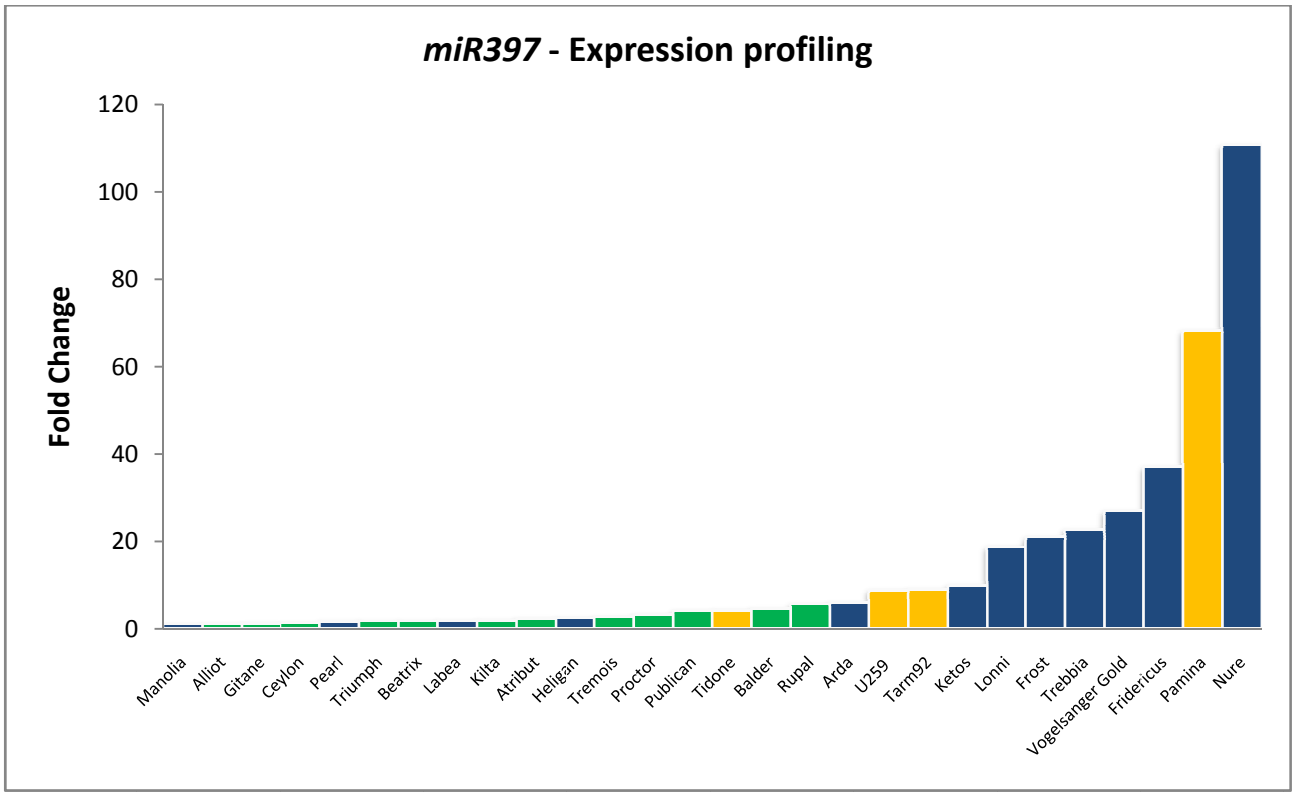


Fig. 13 *Expression levels of miR397 in a panel of 28 barley cultivars. Values are expressed as Fold Change referred to the miRNA level in the cv. Manolia (therefore Manolia miR397 level equals 1, whereas Nure level is approximately 100 times higher). Green bars indicate spring-type varieties, blue bars are for winter-type varieties, and yellow bars for facultative genotypes.*

Deep sequencing – Nure and Tremois response to low temperatures

Many microRNAs were modulated as a consequence of low temperatures exposure. Sometimes the two varieties behaved in the same way, but more often than not their response was quite different.

miR166, *bdi-miR396e*, *osa-miR444c* and *osa-miR5079* were downregulated, while *miR5051* and *bdi-miR5200* were upregulated in both cultivars (Figure 14). Activation or repression started at different timing: expression peaks occurred at the mid-time point (8 hours) more often, even if early regulation cases (2 hours) were present. This applied for the cultivars as well: sometimes Nure activates/repress miRNAs earlier than Tremois, and vice versa.

MicroRNAs belonging to the *miR156* and *miR159* families were modulated in both cultivars but showed different trends, being induced in one and repressed in the other. *Bdi-miR156c* is always static, but its star sequence is upregulated in Nure and downregulated in Tremois (peak at 24 hours). Many other *miR156*-like sequences, probably arising from the same precursor and having lower expression levels compared to the "reference" sequence, behaved differently in the two cultivars.

miR159a was not modulated in Nure, but slightly upregulated (after 24 hours) in Tremois. Interestingly, the star sequence (*miR159a**) was found greatly induced by the treatment. Again, a little difference was observed between the two cultivars: *miR159a** was 21-nt-long in Nure and 20-nt-long in Tremois. Another interesting finding concerns all those sequences that are not reported as microRNAs in miRBase, but align on precursor sequences there deposited. They are presumably byproducts generated by further DCL cuttings of pre-miRNAs after the "main" miRNA duplex excision. It is worth noting that, despite an essential stasis of the *miR159a*, all the sequences mapping on its precursor, including *miR159a**, showed a strong upregulation in response to low temperatures (Figure 15).

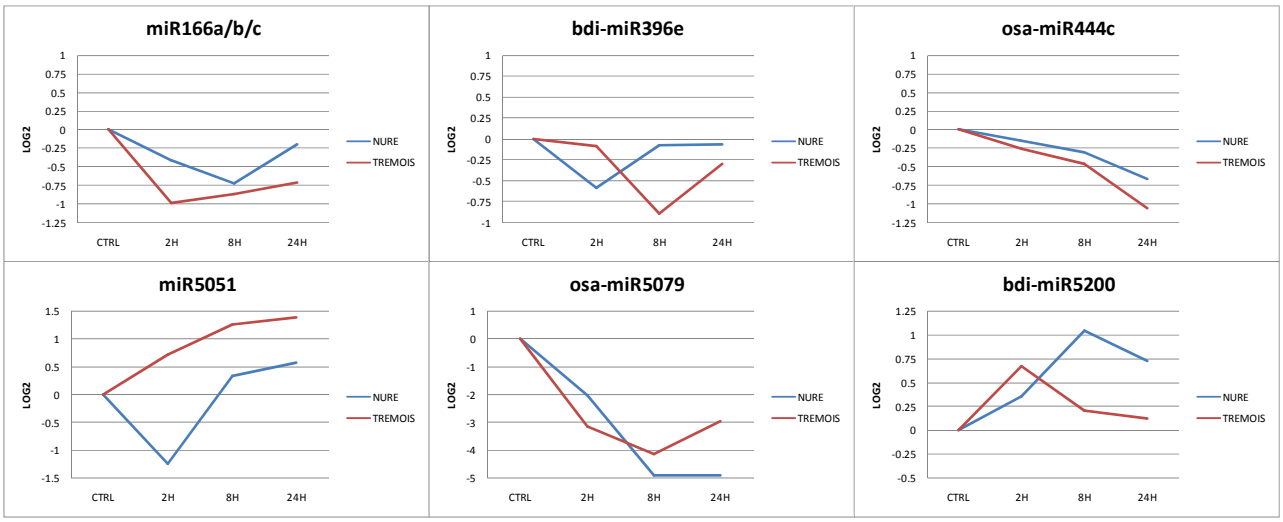


Fig. 14 *Expression profiles of microRNAs involved in low temperatures stress that showed the same trend in the two cultivars. Reference species are reported for miRNAs not deposited in miRBase for barley.*

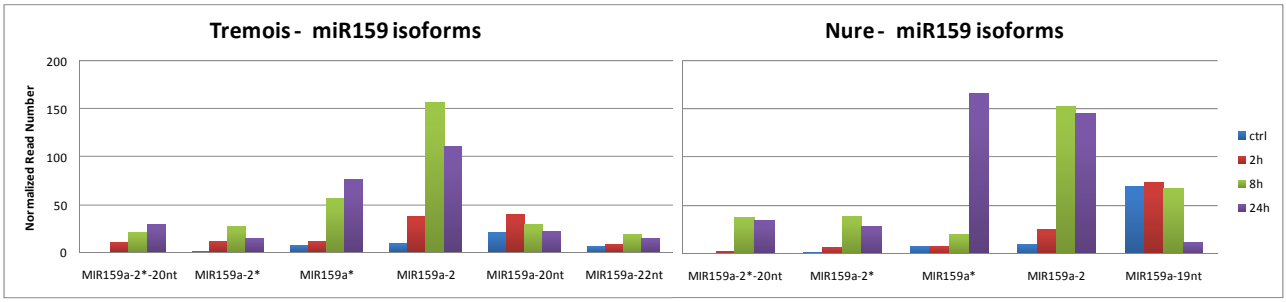


Fig. 15 *Tremois and Nure miR159 isoforms are upregulated by the treatment, despite the fact that miR159 (about 1000 RPM, not shown) is essentially static*

Many microRNAs were found to be regulated in one cultivar only. *Bdi-miR164*, *osa-miR167h*, *osa-miR319b*, and *miR5048* were found to be repressed in Tremois, being 2 hours sufficient to observe a change in their expression profile (with the exception of *miR5048*). Accumulation of *bdi-miR169c* and *bdi-miR827** sequences occurred, with a peak at 24 hours. As for Nure, *bdi-miR171a* and *osa-miR399d* were both upregulated during treatment.

Expression trends of selected microRNAs were validated with a stem-loop qPCR approach. Classic housekeeping genes (*Hv-Actin*, *Hv-polyubiquitin*) were taken into consideration as reference genes, along with small RNAs belonging to the small-nucleolar RNAs (snoRNAs) class (*Hv-U59*, *Hv-U61*).

Discussion

In the present study, barley miRNome was analyzed and a profiling of its movements in response to low temperature stress was carried out. The software used for data analysis allowed the identification of already annotated microRNAs only (referring to miRBase version 20). In recent years some papers describing barley miRNome appeared, and data concerning miRNA families composition and relative abundance were different from this experiment. As said, in Nure and Tremois the 80% of total miRNA reads was made up by *miR166*, *miR5071*, *miR396*, *miR156* and *miR5048* families. Schreiber and colleagues (2011) and Lv and colleagues (2012) found that *miR167* and *miR168* families made up the most of the total miRNA reads. They described also a lower number of families. These differences are essentially normal since different cultivars, growth conditions, age of plants, tissue sampled (e.g. Lv included roots and other plant tissues besides leaves, the only tissue considered in this study) and the kind of the bioinformatic analysis employed can greatly affect miRNome composition. A supporting evidence to the fact that even the cultivar choice can have a substantial impact can be seen in this experiment as well: *miR397* is expressed in Nure, and in general in winter genotypes, while it remains at very low levels in spring varieties.

Like in other recent studies different cultivars, contrasting for their tolerance to a given stress, behaved in a different way (Barrera-Figueroa *et al.*, 2011; Zhao *et al.*, 2012; Yin *et al.*, 2012). Few microRNAs showed the same expression profile in Nure and Tremois and, if so, they were often activated/repressed with different timing (e.g. Nure showed an earlier down-regulation of *bdi-miR396e* compared to Tremois).

Another layer of complexity is given by the great variability of microRNA sequences in the barley miRNome. This is consistent with our findings (see chapter 5.2.2) and with what has been reported in literature. Often the most abundant sequence for a given microRNA is different from the one deposited in miRBase (e.g. *miR156* has 20 nt instead of 21), and the same family member can appear in an incredibly wide assortment of different rearrangements. This plasticity is underlined also by the above described *miR159a** case: it was strongly upregulated by the treatment in both cultivars, but its sequence had a different length in Nure and in Tremois.

Targets were searched for all these microRNAs and results are shown in Table 7. Sometimes discrepancies between putative targets found and known targets reported in literature can be seen. Theoretically this could be due to the fact that *Hordeum vulgare* microRNAs target different mRNAs compared to other, more studied plants. Practically, more often, it is due to the limitation of the barley EST databases commonly available in the web. An example comes from *miR168*. *miR168*-mediated AGO1 regulation is a well known and conserved feedback mechanism that controls the microRNA machinery (Vaucheret 2009 and references therein), but the available online databases at the time of writing (DFCI Gene Index (HVGI), version 12)[4] gave "flame chlorosis virus-like agent" as most likely result. By the way, a recent paper including a degradome analysis finally ascertained that AGO1 is no doubt *miR168* target also in barley (Ozhuner *et al.*, 2013).

Without a direct prove of the target cleavage by a microRNA (that is, a degradome sequencing or a 5'-RACE) speculation is the only possibility. It may be interesting to note that *miR397* and *miR169*, known to be cold-regulated in different plant species (*Arabidopsis*, *Brachypodium* and *Populus trichocarpa*), are static or not expressed at all here. In many other cases, expression trends differing from this study can be seen in literature: *miR166, miR319* and *miR396* upregulation has been reported for *Arabidopsis*, while *miR171* down-regulation has been described in rice and poplar (Sunkar *et al.*, 2012).

[4] ftp://occams.dfci.harvard.edu/pub/bio/tgi/data/Hordeum_vulgare/HVGI.release_12.zip

At the same time, *miR171* was found to be upregulated in *Arabidopsis*, like in this study. This could be consistent with a scenario in which its target (SCL) is repressed, thus causing the disruption of the blocking action on the master growth repressor DELLA (Zhang *et al.*, 2011). In other words, upon cold sensing plants stop their growth through *miR171* activation.

However, opposing patterns of miRNA regulation during cold stress were often reported in literature. Some of these conflicting reports could be genuine responses that may be attributed to the differential responses of unrelated plant species. However, variations in experimental conditions and the methods employed to measure changes in miRNA expression cannot be excluded as possible causes of the observed discrepancies (Sunkar *et al.*, 2012).

Table 7. *Putative targets for studied microRNAs*

miRNA	Sequence	Target Accession	Expectation	Inhibition	Target Description
miR156	TGACAGAAGAGAGTGAGCAC	TC245963	1	Cleavage	Squamosa promoter-binding-like protein 16
		TC265421	1	Cleavage	Squamosa promoter-binding-like protein 18
		CA032492	1	Cleavage	Squamosa promoter-binding-like protein 17
		BF258419	1	Cleavage	Promoter binding protein
		EX598089	1	Cleavage	Squamosa promoter-binding-like protein 18
		TC252995	2	Cleavage	Squamosa promoter-binding-like protein 13
		TC246441	2	Cleavage	Squamosa promoter-binding-like protein 3
		BF630636	3	Cleavage	SBP transcription factor
		TC241815	3	Cleavage	unknown protein
miR159a*	GAGCTCCTATCATTCCAATGA	AV835332	3	Cleavage	OSA15 protein
		TC280007	3	Cleavage	OSA15 protein
		TC239304	3	Cleavage	OSA15 protein
		BM443914	2.5	Cleavage	TRNA (Uracil-5-)-methyltransferase-related protein
miR159a	TTTGGATTGAAGGGAGCTCTG	TC259549	2.5	Cleavage	GAMyb protein
		TC238438	2.5	Cleavage	Transcription factor GAMyb
		CV062223	2	Cleavage	60S ribosomal protein L30
tae-miR164	TGGAGAAGCAGGGCACGTGCA	TC240039	1	Cleavage	NAC transcription factor
		TC248477	1	Cleavage	NAC domain protein NAC1
		TC251797	1	Cleavage	NAC domain protein NAC1
		TC256476	1.5	Cleavage	Mitogen-activated protein kinase 4
		TC254954	2	Cleavage	Os02g0579000 protein
		TC269623	3	Cleavage	Phytosulfokine-alpha 1 precursor
		BI956904	3	Cleavage	Pherophorin-C1 protein precursor
		BY866386	3	Cleavage	Os07g0582400 protein
		TC248533	2.5	Cleavage	unknown protein
		BU993554	3	Cleavage	XS domain containing protein-like
		BG367052	2.5	Cleavage	Non-specific lipid-transfer protein 1 precursor
miR166	TCGGACCAGGCTTCATTCCCC	CV063912	3	Cleavage	unknown protein
osa-miR166k/l	TCGGACCAGGCTTCAATCCC	TC247931	3	Cleavage	S-receptor kinase-like
osa-miR167h	AGGTCATGCTGGAGTTTCATC	BQ762315	3	Cleavage	Os01g0775500 protein
		CD663814	3	Cleavage	Os02g0818000 protein
		TC266641	3	Cleavage	AgrC
miR168	TCGCTTGGTGCAGATCGGGAC	NP315934	1.5	Translation	flame chlorosis virus-like agent [H. vulgare]
		TC258904	3	Translation	Os07g0571100 protein

bdi-miR169c	TTGGCTACACCTAGTTCTCTT	TC267130	3	Translation	CCT motif family protein, expressed
osa-miR319b	CTTGGACTGAAGGGTGCTCCCT	BM370389	3	Cleavage	Predicted protein
		TC264933	2.5	Cleavage	Os07g0152000 protein
		TC259549	2.5	Cleavage	GAMyb protein
		TC238438	2.5	Cleavage	Transcription factor GAMyb
		TC240929	3	Cleavage	Rab geranylgeranyl transferase like protein
bdi-396e	TTCCACAGCTTTCTTGAACTT	CA026441	2	Cleavage	Transcription activator
		GH213244	2	Cleavage	Growth-regulating factor 6
		CB860915	2	Cleavage	Growth-regulating factor 1
		TC279628	2	Cleavage	Growth-regulating factor 3
		TC272166	2	Cleavage	Growth-regulating factor 1
		TC270705	2	Cleavage	Growth-regulating factor 3
		TC260953	2	Cleavage	Growth-regulating factor 8
		AV913958	2.5	Cleavage	Uncharacterized protein At1g22700.1
		TC250287	2.5	Cleavage	Os10g0460900 protein
		BG342891	2.5	Cleavage	Molecular chaperone Hsp90-2
		TC269943	3	Cleavage	unknown protein
		TC260189	3	Cleavage	Os05g0373700 protein
		BF627785	3	Cleavage	Os05g0373700 protein
		TC257291	3	Cleavage	Alpha chain of nascent polypeptide associated complex
miR397	TTGAGTGCAGCGTTGATGAAC	TC262144	0.5	Cleavage	Laccase-23 precursor
		BI960433	2.5	Cleavage	Laccase-10 precursor
		BY839519	2.5	Translation	Laccase-12/13 precursor
		BI960140	2.5	Translation	Laccase-12/13 precursor
		TC274250	2.5	Cleavage	Laccase-10 precursor
		BY835984	2.5	Translation	Laccase LAC5-6
		BI957301	2.5	Translation	Laccase LAC5-6
		TC263463	2.5	Translation	Laccase LAC5-6
		CK568938	2.5	Cleavage	
		CV057686	3	Cleavage	Cystatin Hv-CPI8
		CV060414	3	Cleavage	Cystatin Hv-CPI8
		CV058406	3	Cleavage	Cystatin Hv-CPI8
		CV064261	3	Cleavage	Cystatin Hv-CPI8
osa-miR399d	TGCCAAAGGAGAGTTGCCCTG	AV835204	1.5	Cleavage	DNA primase
		TC244059	3	Cleavage	Os02g0537400 protein
tae-miR408	TGCACTGCCTCTTCCCTGGC	TC272169	3	Cleavage	INDETERMINATE-related protein 9
osa-miR444c	CGGCAAGCTAGAGACAGCAAC	TC266173	3	Cleavage	Transducin /WD-40 repeat protein family-like protein
		GH228935	2.5	Translation	MIKC-type MADS-box transcription factor WM32A
		TC274417	2.5	Translation	MIKC-type MADS-box transcription factor WM32A
		TC240629	2.5	Translation	MIKC-type MADS-box transcription factor WM32A
		TC254311	3	Cleavage	Serine/threonine kinase-like protein ABC1040
		TC274849	3	Cleavage	Beclin 1 protein
		TC241864	3	Cleavage	Predicted protein
miR5048	TATATTTGCAGGTTTTAGGTCT	TC250387	0	Cleavage	Os11g0274100 protein
		TC238421	2	Cleavage	Resistance protein
		TC238427	2	Translation	Serine/threonine kinase-like protein ABC1036
		TC238429	2	Translation	Serine/threonine kinase-like protein ABC1040
		TC243354	2.5	Cleavage	Os06g0283300 protein
		TC255513	2.5	Cleavage	Serine/threonine protein kinase

		TC243305	2.5	Cleavage	D-mannose binding lectin family protein, expressed
		TC269313	2.5	Cleavage	SRPK4
		TC277472	2.5	Cleavage	Aquaporin
		TC240164	3	Cleavage	Os03g0422800 protein
		CD053914	3	Cleavage	Acyl-CoA dehydrogenase
		BM816872	3	Cleavage	Os09g0460200 protein
		TC252527	3	Cleavage	Acyl-CoA dehydrogenase
miR5048*	TTTAGACCTAGACATGCAAGT	TC250387	0	Cleavage	Os11g0274100 protein
		EY275010	3	Cleavage	CG1
osa-miR5071	TCAAGCATCATGTCATGGACA	TC243080	1	Cleavage	NB-ARC domain containing protein, expressed
		TC254067	1.5	Cleavage	Stripe rust resistance protein Yr10
		TC254272	1.5	Cleavage	Mla-like protein
		DN181339	2	Cleavage	Os10g0131000 protein
		TC240513	2	Cleavage	MLA10
		AV835525	2	Cleavage	Os10g0131000 protein
		EX595033	2.5	Cleavage	
		TC241599	2.5	Cleavage	NB-ARC domain containing protein
		TC241734	3	Cleavage	Os10g0124300 protein
		TC269181	3	Cleavage	Disease resistance protein RPM1 homolog
		BU970049	3	Cleavage	Os10g0124300 protein
		BJ483403	3	Cleavage	
bdi-miR5200	TAGATACTCTCTAAGGCTTGG	TC253913	3	Cleavage	FT Flowering Locus T
bdi-miR827*	TTTTGTTGGTTGTCATCTAACC	BQ464351	2.5	Cleavage	Ubiquitin carrier protein
		TC239709	2.5	Cleavage	Ubiquitin carrier protein
		CA023518	2.5	Cleavage	Hydroxyproline-rich glycoprotein presumed cell wall component
		CV064193	3	Cleavage	G-patch domain-containing protein C890.05
		DN156209	3	Cleavage	Pentatricopeptide repeat-containing protein-like protein
		BF263382	3	Cleavage	Os01g0929100 protein
		DN156788	3	Cleavage	

Table 7 *Putative targets for studied microRNAs are reported. Results were obtained from http://plantgrn.noble.org/psRNATarget/. Target inhibition method is indicated along with Expectation, i.e. the likelihood that a given mRNA is targeted by the microRNA considered (0 = perfect match, 3 = 3 mismatches. G-U mismatches weight 0.5). miRNAs for which no target was found are not reported.*

5.4 Experiment 3 – *MicroRNAs involved in nucleus-organelles crosstalk*

Introduction

The genetic information of a plant species is not limited to the nuclear DNA only, but it is implemented by mitochondrial and chloroplastic genomes. These three genomes coevolved and built a complex network of interactions that guide the cell functioning (reviewed in Greiner and Bock, 2013). It has been showed that, over evolutionary times, gene transfer from an organelle to the nucleus or to another organelle has been a common feature in plants (Zhang *et al.*, 2012b), as exemplified by the chimeric multiprotein complexes in plastids and mitochondria (one part of the subunits being nuclear-encoded and another one being encoded in the organellar genome). Furthermore, nucleus controls organellar functions, in a process referred to as 'anterograde signaling', and at the same time organelles can influence the expression of the nuclear genome, in a process called 'retrograde signaling'. This enormous complexity requires an extraordinarily tight coordination between the actors (Greiner and Bock, 2013).

Plant organellar genomes are quite small if compared to the nuclear one. Over the past years, lots of them have been fully sequenced, and a characterization of the genes they harbor has been obtained (Zhang *et al.*, 2012b). With this came the identification of genes involved in many processes linked with fitness and adaptation to the environment. Some of them were deemed useful for plant breeding, especially in cereals. Concern about these traits started to rise in the '50s, with Kihara's experiments on wheat (1951). In the following years several research groups aimed at exploiting the information contained in the plasmon of wild relatives of cultivated species for crop breeding.

The cytoplasm is maternally inherited in *Triticeae* species (Kihara, 1951). Firstly, new hybrids were created by crossing maize, wheat and barley with wild relatives like teosinte (*Zea spp.*) (Allen, 2005), *Hordeum chilense* and *Aegilops spp.* (Martin and Chapman, 1977; Martin and Sanchez-Monge Laguna, 1982; Atienza *et al.*, 2007), using wild species as maternal parent. At a later stage, serial backcrossing with the pollen parent (the cultivated species) brought to the creation of "alloplasmic lines", that is, new plants with the nucleus of the pollen donor and the cytoplasm of the maternal parent. These plants represent a very powerful tool to investigate nuclear-cytoplasm interactions, besides bringing new solutions to breeders. It has been demonstrated that alien cytoplasm introgression in cultivated species can bring great advantages ranging from enhancement of biotic and abiotic stress tolerance to a better grain quality (e.g. for carotenoid content) (Atienza *et al.*, 2007; Atienza *et al.*, 2008).

We recently published a paper in which we report the results of a microarray-based transcriptome analysis of three alloplasmic lines compared to their euplasmic control (see Annexes). This study showed that the expression of many genes was altered in response to the replacement of wheat cytoplasm with the cytoplasm of a related species. Remarkably, the more evolutionary distant the species was, the higher the number of these modifications.

Within this project, a NGS-based microRNA profiling was carried out to identify a new potential level of gene regulation involved in the nucleus-organelles crosstalk.

Results and discussion

The alloplasmic line TH237 was produced by introgressing the *Hordeum chilense* accession H7 cytoplasm into the nuclear background of *Triticum aestivum* accession T20. Illumina sequencing regarded this alloplasmic-euplasmic couple only (in three biological replicates).

This couple was chosen based on preliminary results coming from the microarray sequencing. *Hordeum chilense* is evolutionarily more distant from wheat than *Aegilops uniaristata* and *Aegilops tauschii*, the other cytoplasm donors used in the experiment. As above mentioned, this wider distance resulted in a higher number of modulated genes, therefore the possibility to find altered microRNA expression trends was expected to be higher.

Data analysis highlighted only four microRNAs that were moving, all up-regulated, with a statistical significance. They are reported in Table 8.

miRNA name	Sequence	Length	Fold Change (Alloplasmic)
hvu-miR166	TTCGGACCAGGCTTCATTCCC	21	1.6584
bdi-miR166f	TTTCGGACCAGGCTTCATTCC	21	2.8332
ath-miR165a/b	TTCGGACCAGGCTTCATCCCC	21	2.0242
tae-miR395a/b	TGAAGTGTTTGGGGGAACTC	20	6.3018

Table 8 *Differentially expressed microRNAs between the wheat alloplasmic line T20 (harboring H. chilense cytoplasm) and its euplasmic background. Reported Fold Changes were calculated subtracting the expression value in the euplasmic background from the expression value in the alloplasmic one. Being all more than 1, they all describe an up-regulation occurring in the alloplasmic background.*

As it can be noticed, *hvu-miR166* and *bdi-miR166f* are almost identical and probably arise from the same precursor – they might be generated by alternative DCL cleavage. At the moment of writing though, no *miR166* sequence for bread wheat is deposited in miRBase (version 20, June 2013), so this cannot be more than a sheer hypothesis.

ath-miR165a/b showed lower expression level and a very similar sequence compared to *miR166* family members – actually these families are often reported as one, dubbed *miR166/miR165*. It is up-regulated, and this could be consistent with the modulation of the other two miRNAs, considering that, owing to their sequence similarity, they probably share the same target mRNAs.

Finally, *tae-miR395a/b* was expressed at low levels but showed a significant up-regulation in an alloplasmic background.

Unfortunately, qPCR validations of these results failed to show any difference between samples. We were therefore unable to draw any kind of conclusion different from the fact that, in this experiment, no microRNA expression trend was significantly affected by cytoplasm substitution. The divergence between sequencing data and qPCR results could be due to many possible factors. They rely on different mechanisms: theoretically deep-sequencing should be more precise, but many steps in libraries preparation, such as construction methods, final quantification and multiplexing level on a flowcell lane, can greatly affect the final results. The company itself (Illumina) released two completely different protocols for libraries preparation in the last three years, solving troubles (e.g. small RNAs ligation efficiency), tricks and biases (Hansen *et al.*, 2010) due to the use of a technology or another. Moreover, different software used for the profiling can give different results, based also on available miRNA databases.

Another possible explanation for this difference might be implied in the method chosen for the qPCR validation procedure. Stem-loop primers ensure specificity in targeting mature miRNAs thanks to their spatial conformation – it does not allow the annealing to pri- and pre-miRNA sequences – and their six bases overhang, which are the reverse complement of the last six bases (3' end) of a specific microRNA.

Stem-loop primers guide the small RNA retrotranscription: cDNA sequences will result from whatever has been bound by the RT-primer overhang. Also unwanted sequences that share the same last six nucleotides at the 3' end will become part of the cDNA pool that will undergo amplification. At this stage, fifteen bases, considering a 21-nt-long miRNA, are specified by the forward primer used. It is not hard to imagine that two miRNAs differing for one or two nucleotides at the 5' end will not be discriminated by the system. At the same time, also sequences with a 3' deletion will anneal on the RT-primer overhang and enter the cDNA pool. This is why reaching the same precision of the deep-sequencing could be hampered in a stem-loop qPCR strategy.

Finally, analyses of other experiments have often highlighted an overall discrepancy between Illumina fold changes and qPCR fold changes. Real time PCR generally seems to underestimate the values calculated by the Illumina.

5.5 Experiment 4 – *Durum wheat response to drought and heat stress*

Introduction

Durum wheat (*Triticum turgidum subsp. durum*) is an important cereal crop grown mainly in semi-arid environments, like mediterranean regions, characterized by water scarcity and high temperatures.

Drought and heat stress are among the most important yield-limiting factors of arable lands worldwide. They induce a range of physiological and biochemical responses in plants (reviewed in Rowley and Mockler, 2011). Briefly, from a physiological point of view, mild drought induces regulation of water loss and uptake, allowing maintenance of leaf relative water content within the limits where the photosynthetic capacity shows no or little changes, but severe drought induces unfavorable changes in plants, leading to inhibition of photosynthesis and growth. Heat stress disturbs cellular homeostasis and can lead to severe retardation in growth and development, and even death: direct injuries include protein denaturation and aggregation, and increased fluidity of membrane lipids. Indirect or slower heat injuries can lead to starvation, inhibition of growth, reduced ion flux, production of toxic compounds and reactive oxygen species (ROS) (Yordanov *et al.*, 2003).

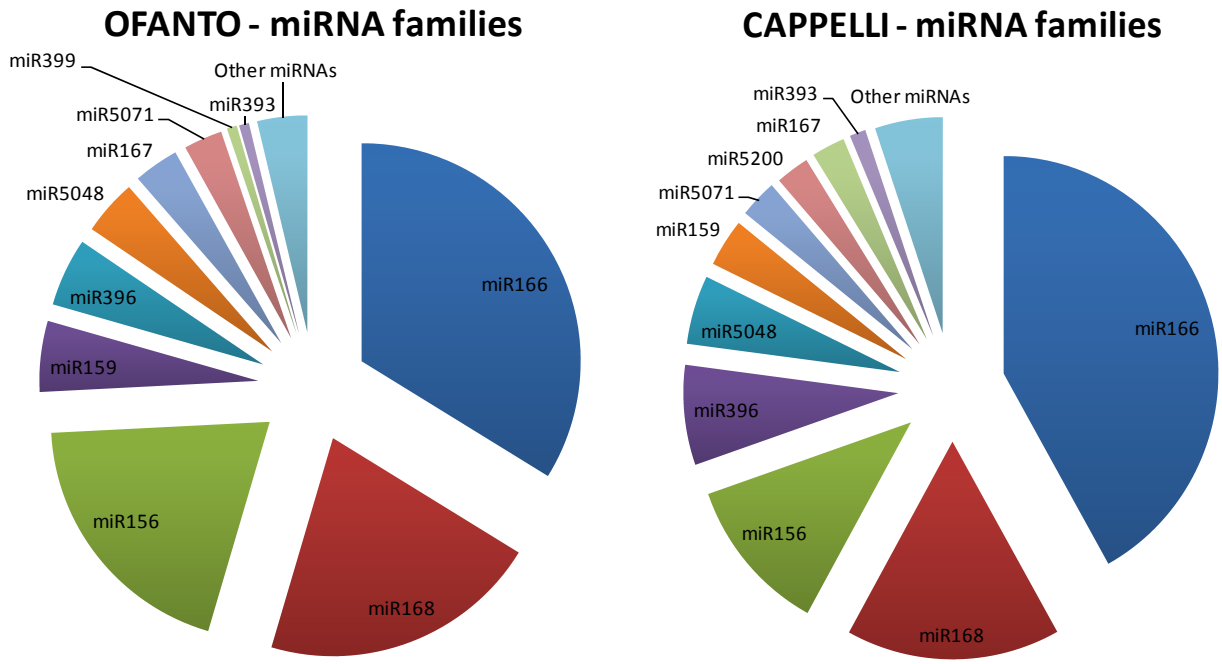
Two durum wheat cultivars, Ofanto and Cappelli, were extensively characterized from an agronomical and physiological point of view (De Vita and Nigro, 2007; Rizza *et al.*, 2012; Panio *et al.*, 2013; Aprile *et al.*, 2013), since they form an interesting experimental system to study the response to such stresses in cereals. Cappelli consistently showed a higher water use efficiency (WUE, *i.e.* the ability of a crop to produce biomass per unit of water evapotranspired) compared to Ofanto and this finding correlated with the fact that Cappelli showed a lower stomatal conductance compared to Ofanto over a range of relative soil water contents (Rizza *et al.*, 2012). According to Levitt (1980) plants can avoid drought by two mechanisms: “water savers” try to conserve water, whereas “water spender” try to absorb water rapidly enough to keep up with their extremely quick water loss. According to this classification, Cappelli can be considered a “water saver”, since on incipient drought it is able to close stomata earlier than Ofanto (Rizza *et al.*, 2012).

Aprile and colleagues (2013) recently discovered that Ofanto activates a large set of well-known drought-related genes after drought treatment, while Cappelli shows the constitutive expression of several genes that in Ofanto are induced by drought and a minimal modulation of gene expression in response to stress. This can basically explain the diversity at molecular level between these two cultivars, which cope with stress using different strategies.

Durum wheat miRNome under mild and severe drought stress and under heat stress was characterized to identify microRNAs involved in stress response, and, possibly, contributing to the explanation of the difference between these two cultivars.

Deep sequencing – miRNA families

Data analysis identified 54 families of microRNAs (expressed with at least 5 RPM). Ofanto and Cappelli showed some differences in miRNA families composition (Graphs 3 and 4 and Figure 16). While *miR166*, *miR168* and *miR156*, widely conserved miRNA families among all plant species, made up more than 70% of the total miRNome of both cultivars, things changed in less abundant families. Ofanto showed a constitutively higher expression of *miR159* and *miR167* compared to Cappelli, which in turn showed a much greater number of sequences belonging to *miR5200* family (2000 RPM versus 300).



Graph 3 and 4 *miRNA families abundance in the two cultivars. Abundance is expressed as a percentage of the total microRNA read number.*

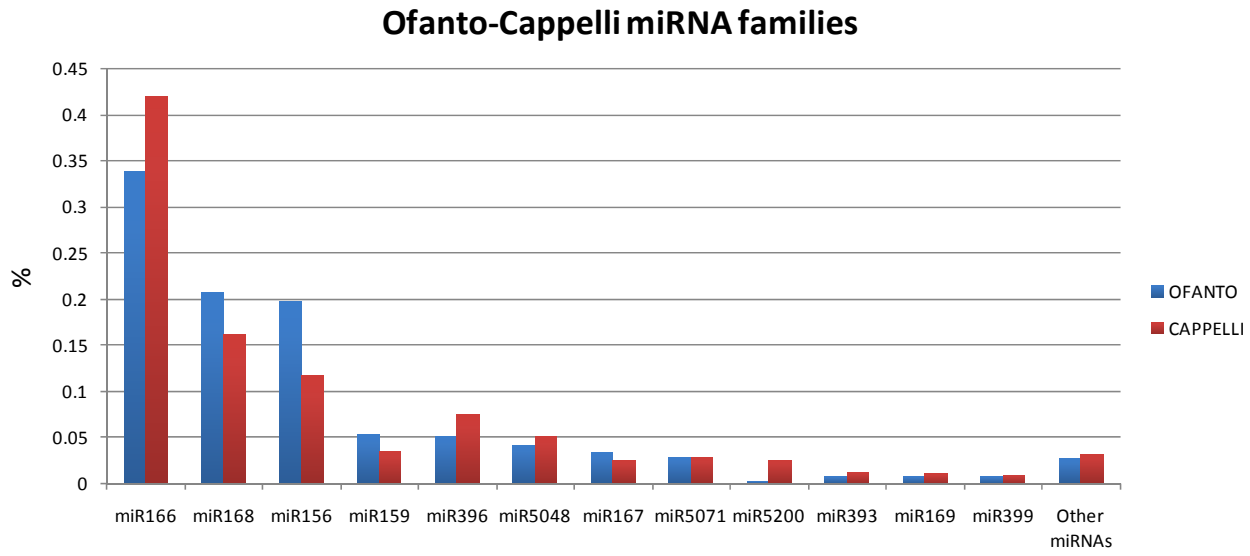


Fig. 16 *Direct comparison of miRNA families abundance between the two cultivars.*

Deep sequencing – Differential constitutive expression of microRNAs

Seven microRNAs showed a different behaviour in the control samples of the two varieties, being constitutively expressed at different levels. Some specifications and their expression trends are reported in table 9 and Figure 17, respectively.

It is worth noting that, so far, miRBase reports only *miR160* mature and precursor sequence for *Triticum turgidum*. All the microRNAs identified in this work have been dubbed based on alignments with known mature sequences deposited in miRBase (version 20, June 2013). The evolutionarily closest species which showed the sequence with the smallest number of mismatches was considered as “source”.

Name	Sequence	Length	Source	Mismatches
miR156	TGACAGAAGAGAGTGAGCAC	20	*Triticum aestivum*	0
miR159a/miR159b	TTTGGATTGAAGGGAGCTCTG	21	*Triticum aestivum*	0
miR167c	TGAAGCTGCCAGCATGATCTGA	22	*Brachypodium distachyon*	0
miR171a	TGATTGAGCCGCGCCAATATC	21	*Oryza sativa*	0
miR5062a	TGAACCTTGGGGAAAAGCCGCAT	23	*Brachypodium distachyon*	1
miR5083	CAGACTACAATTATCTGATCA	21	*Oryza sativa*	1
miR5200	TGTAGATACTCCCTAAGGCTT	21	*Brachypodium distachyon*	1

Table 9 *Differentially expressed microRNAs between Ofanto and Cappelli in control samples. Source organisms used to annotate microRNAs during bioinformatic analyses are reported along with respective mismatches with the deposited sequence in miRBase. So far (miRBase v. 20) only one sequence for durum wheat is deposited in miRBase.*

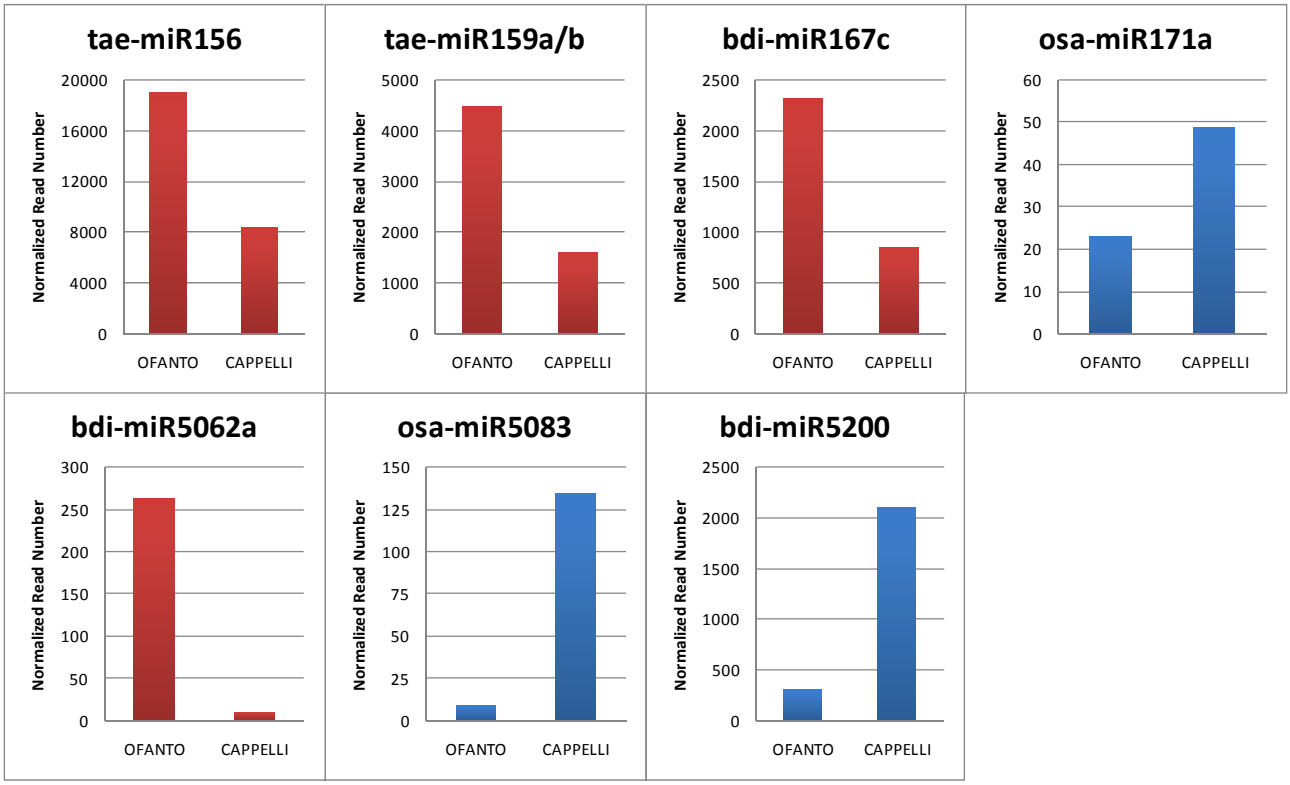


Fig. 17 *Differentially expressed microRNAs in the control samples of Ofanto and Cappelli. Normalized expression values (RPM) are reported. Red bars mark microRNAs with a constitutively higher expression in Ofanto, blue bars are for the opposite situation.*

Deep sequencing – Ofanto and Cappelli response to drought stress

Two stress levels were considered, a mild and a severe drought stresses (see Chapter 3), characterized by two different soil water contents. The most dramatic effects in terms of microRNA activation or repression could generally be seen when water deficit reached the maximum, but cases of early regulation are still visible.

Modulated miRNAs are here classified into two main groups: the first group contains miRNAs regulated in both cultivars, while in the second miRNAs showing cultivar-specific regulation are listed.

miR169 family members were generally all down-regulated by the treatment both in Ofanto and in Cappelli. Particularly, *osa-miR169c-1*, *osa-miR169h* and *osa-miR169n* (Figure 18, top) showed a medium-high expression level with an interesting repression trend, usually visible in the first stress level already. *Osa-miR1432* and *osa-miR171h* were down-regulated as well, while a 20-nt long microRNA belonging to the

miR159 family (here dubbed *bdi-miR159-3p*) was the only one to show a substantial up-regulation (Figure 18, center). *Ttu-miR160*, *bdi-miR827-3p* and *bdi-miR5200* were drought-modulated in both cultivars, but showed opposite trends, being induced in one and repressed in the other (Figure 18, bottom).

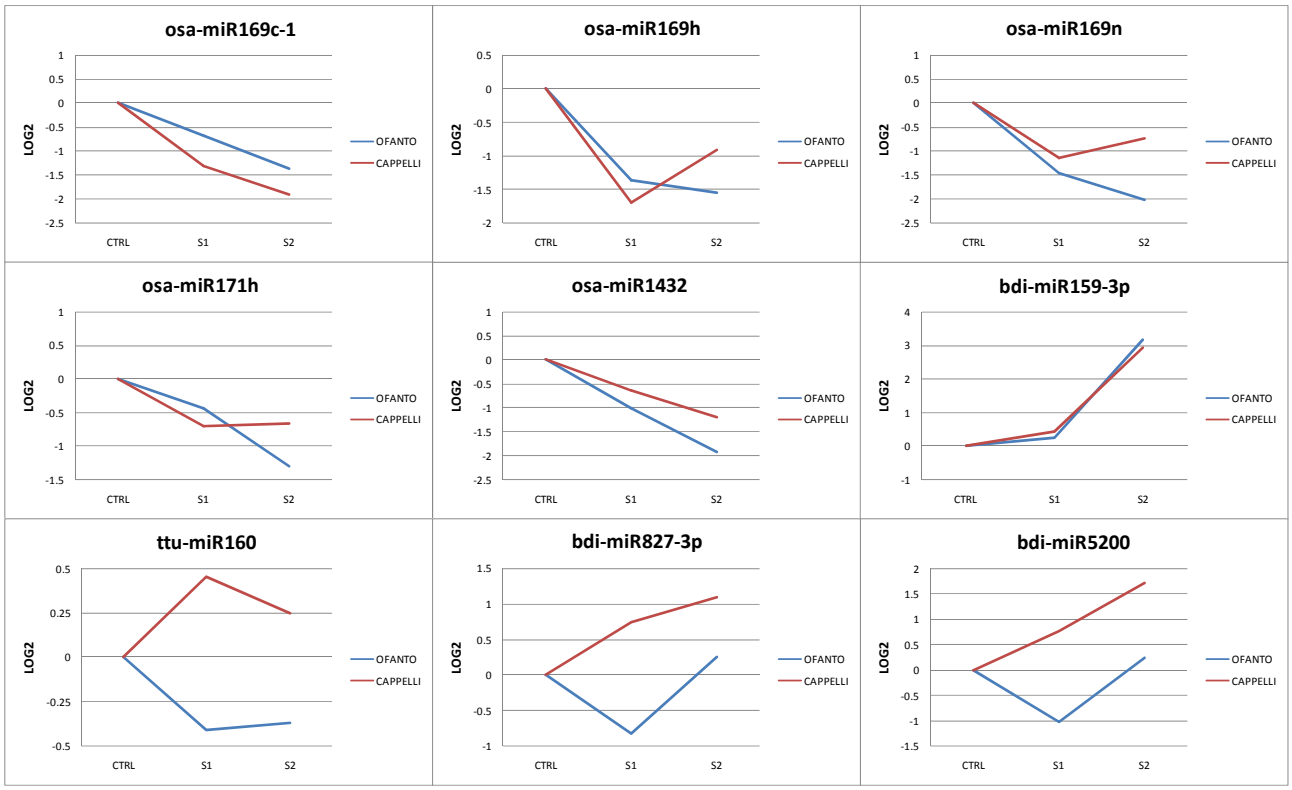


Fig. 18 *Expression profiles of microRNAs involved in drought stress that showed the same trend (**top** and **center**) or an opposite trend (**bottom**) in the two cultivars. S1 = mild drought; S2 = severe drought.*

The second group is made up by microRNAs showing a cultivar-specific modulation. *Tae-miR156* is a 21 nt-long isoform of the more abundant *tae-miR156-2*, which has only 20 nt. While the latter remained static throughout the treatments, the former was induced under severe water stress in Cappelli. *Tae-miR159a/b*, three members of the *miR167* family and *gma-miR393h* were all induced in Cappelli only as well, while *osa-miR5072* was repressed. Ofanto in turn down-regulated *hvu-miR168*, *osa-miR399d* and *bdi-miR399b*, and up-regulated *ath-miR165a*.

All drought-regulated microRNAs found in this study are reported in Table 10.

Deep sequencing – Ofanto and Cappelli response to heat stress

Plants underwent a heat stress treatment for which they were kept at 36° C for 3 hours. The expression trend of many microRNAs was affected. Three miRNAs had a common trend in the two cultivars: *bdi-miR159-3p* and *osa-miR1432* were repressed, while *osa-miR444e* showed a significant up-regulation (Figure 19).

In the cultivar Cappelli up-regulation of a member of *miR166* family (*bdi-miR166h-2*) and of *miR169* family (*osa-miR169n*) were seen, along with *bdi-miR399b* induction.

Ofanto in turn modulates a great number of other microRNAs. The only up-regulation was seen in the case of *bdi-miR166h*, which seems to be an isoform of the *miR166* member induced in Cappelli (they differ for a nucleotide at the 3' end). Down-regulation was seen in miR156 (*tae-miR156-2*, which is the most abundant

member, and two of its isoforms, namely *tae-miR156* and *ath-miR156g*), miR159 (*tae-miR159a/b*, besides the above-mentioned *bdi-miR159-3p*, which is likely a shorter, 5'-modified isoform), miR160 (*ttu-miR160*, *bdi-miR160e-5p*), miR166 (*osa-miR166g*), miR167 (*ath-miR167d-2*, *ath-miR167d* and *bdi-miR167c*, which only differ for a nucleotide at the 3' end), miR168 (*hvu-miR168* and its star sequence) families, and for *osa-miR5071 and bdi-miR5200*.

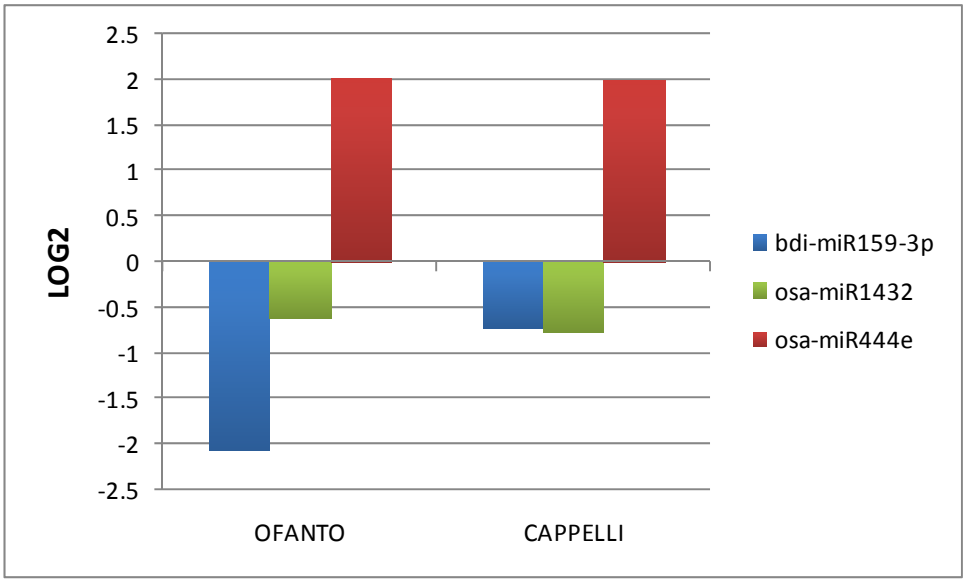


Fig. 19 *Expression trend of heat-regulated microRNAs that showed the same trend in both cvs.*

All drought and heat-modulated microRNAs found in this study are reported in the following table.

Name	Sequence	Length	Source	Mism.	Stress
tae-miR156-2	TGACAGAAGAGAGTGAGCACA	21	*Triticum aestivum*	0	**H**
tae-miR156	TGACAGAAGAGAGTGAGCAC	20	*Sorghum bicolor*	0	**D, H**
ath-miR156g	CGACAGAAGAGAGTGAGCAC	20	*Arabidopsis thaliana*	0	**H**
tae-miR159a/b	TTTGGATTGAAGGGAGCTCTG	21	*Triticum aestivum*	0	**D, H**
bdi-miR159-3p	CTTGGATTGAAGGGAGCTCT	20	*Brachypodium distachyon*	0	**D, H**
ttu-miR160	TGCCTGGCTCCCTGTATGCCA	21	*Triticum turgidum*	0	**D, H**
bdi-miR160e-5p	TGCCTGGCTCCCTGAATGCCA	21	*Brachypodium distachyon*	0	**H**
ath-miR165a	TCGGACCAGGCTTCATCCCCC	21	*Arabidopsis thaliana*	0	**D**
osa-miR166g	TCGGACCAGGCTTCATTCCTT	21	*Oryza sativa*	1	**H**
bdi-miR166h	TCGGACCAGGCTTCAATCCCT	21	*Brachypodium distachyon*	0	**H**
bdi-miR166h-2	TCGGACCAGGCTTCAATCCCG	21	*Brachypodium distachyon*	1	**H**
tae-miR167a	TGAAGCTGCCAGCATGATCTA	21	*Triticum aestivum*	0	**D**
bdi-miR167c	TGAAGCTGCCAGCATGATCTGA	22	*Brachypodium distachyon*	0	**D, H**
ath-miR167d	TGAAGCTGCCAGCATGATCTGG	22	*Arabidopsis thaliana*	0	**D, H**
ath-miR167d-2	TGAAGCTGCCAGCATGATCTGC	22	*Arabidopsis thaliana*	1	**H**
hvu-miR168	TCGCTTGGTGCAGATCGGGAC	21	*Hordeum vulgare*	0	**D, H**
osa-miR169c-1	CAGCCAAGGATGACTTGCCGGC	22	*Oryza sativa*	0	**D**
osa-miR169n	TAGCCAAGAATGACTTGCCTG	21	*Oryza sativa*	1	**D, H**
osa-miR169h	TAGCCAAGGATGACTTGCCTG	21	*Oryza sativa*	0	**D**
osa-miR171h	TGAGCCGAACCAATATCACTC	21	*Oryza sativa*	0	**D**
gma-miR393h	TTCCAAAGGGATCGCATTGAT	21	*Glycine max*	0	**D**
bdi-miR399b	TGCCAAAGGAGAATTGCCCTG	21	*Brachypodium distachyon*	0	**D, H**
osa-miR399d	TGCCAAAGGAGAGTTGCCCTG	21	*Oryza sativa*	0	**D**
bdi-miR399b	TGCCAAAGGAGAATTGCCCTG	21	*Brachypodium distachyon*	0	**D, H**

osa-miR444e	TGCAGTTGCTGCCTCAAGCTT	21	*Oryza sativa*	0	**H**
bdi-miR827-3p	TTAGATGACCATCAGCAAACA	21	*Brachypodium distachyon*	0	**D**
osa-miR1432	ATCAGGAGAGATGACACCGAC	21	*Oryza sativa*	0	**D, H**
osa-miR5071	TCAAGCATCATGTCATGGACA	21	*Oryza sativa*	2	**D, H**
osa-miR5072	CGTTCCCCAGCGGAGTCGCCA	21	*Oryza sativa*	2	**D**
bdi-miR5200	TGTAGATACTCCCTAAGGCTT	21	*Brachypodium distachyon*	1	**D, H**

Table 10 *Stress-responsive microRNAs found in this study. Stress causing modulation is reported (D = drought, H = heat). Many microRNAs were found to be activated or repressed by both stresses.*

Results confirmation – Radioso and Trinakria experiment

Preparation

The experiment was repeated during an internship at the Australian Centre for Plant Functional Genomics (ACPFG) of Adelaide, Australia. Two new cultivars were chosen based on their pedigree, since Ofanto and Cappelli were not available.

Radioso is a modern italian cultivar, which has got Creso in its pedigree. Trinakria is an italian old, tall and low yielding cultivar that derives from Capeiti (Cappelli x Eiti8). The system formed by Creso and Trinakria has been already studied in experiments dealing with drought tolerance, where the former was considered drought-susceptible and the latter more drought-tolerant (De Leonardis *et al.*, 2007; Burling *et al.*, 2013). Radioso was the best choice since having no availability of Creso seeds. Moreover, an australian modern cultivar, Yallaroi, was added to the present study. Its function was to provide evidence that a given microRNA is indeed related to drought, and not to a breeding group, since Yallaroi does not belong to the same breeding group of the italian cultivars.

To check for varietal differential behaviour under drought stress, also physiological measurements were taken on plants subjected to severe drought stress. Trinakria plants took one day more to reach the first drought level. They were losing less water per day, as can be seen in figures 20 and 21, but when the Soil Water Content (SWC) decreased below 18% the significance of this difference dropped down.

Daily increments													
	DAY 1	day 2	day 3	day 4	day 5		day 6	day 7	day 8	day 9	day 10	day 11	day 12
	stress started					S1 harvest							S2 harvest
RADIOSO		-13.7	-13.9	-12.5	-13		-11	-9	-10	-8	-6	-2	
		1.272	1.595	1.304	4.179		1.211	0.408	2.608	0.983	0.894	0.837	
TRINAKRIA		-13.0	-11.8	-9	-9		-10	-9	-12	-10	-9	-3	
		1.514	1.834	1.368	2.121		2.000	3.317	1.517	2.345	2.074	1.304	
YALLAROI		-13.3	-10.6	-7.0	-6.3		-7.3	-6.0	-10.3	-5.3	-5.5		
		2.464	2.251	0.408	3.055		1.155	1.732	3.512	3.000	1.528		

Fig. 20 *Average daily water loss of drought-stressed plants of the three cultivars. Water withholding started on day 1. First harvest (S1, mild stress) occurred 4 days after for Radioso and Yallaroi, while Trinakria was harvested one day later. Second harvest (S2, severe stress) was carried out 11 days after for all cultivars. Data represent average values of 6 biological replicates (green cells), and standard deviations are reported as well (white cells).*

Physiological parameters were checked as well on plants subjected to severe drought stress. Measurements of chlorophyll fluorescence (figure 22, left), chlorophyll content and stomatal conductance were taken (figure 22, right). The aim was to observe plant behaviour and detect any sufferings before wilting symptoms appearance. Although errors were not negligible, data were in agreement with the kind of treatment applied.

As it can be seen, stressed plants showed a lower chlorophyll fluorescence, that highlighted a damage occurred in the photosystem II, and a lower stomatal conductance, in an ultimate attempt to save water before wilting. Yallaroi data seem not to be very reliable, being small plant size the most likely explanation for such a low daily water loss. The day after these measurements were taken, plants were harvested.

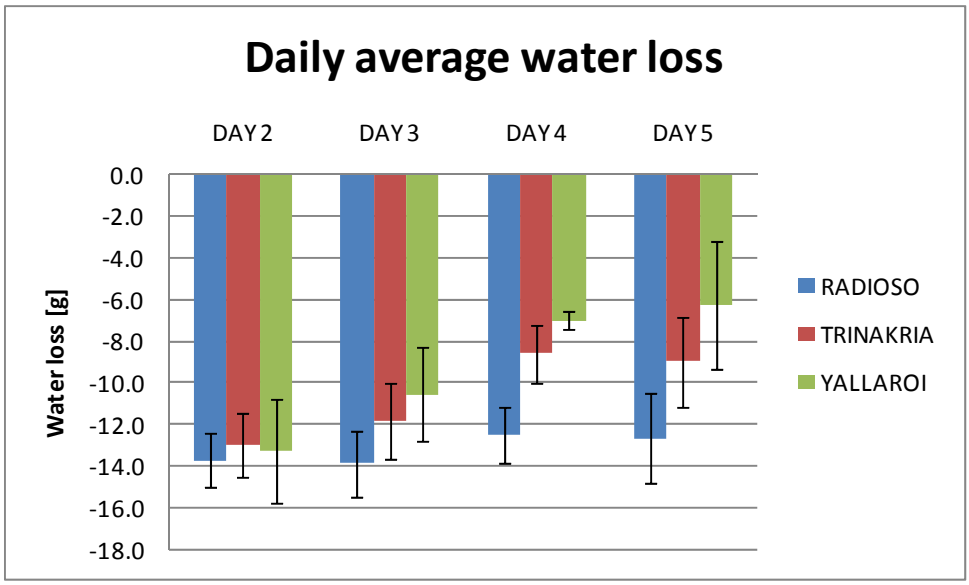


Fig. 21 *Average daily water loss of drought-stressed plants of the three cultivars from the imposition of the stress to the harvest of stress 1. Data represent average values of 6 biological replicates.*

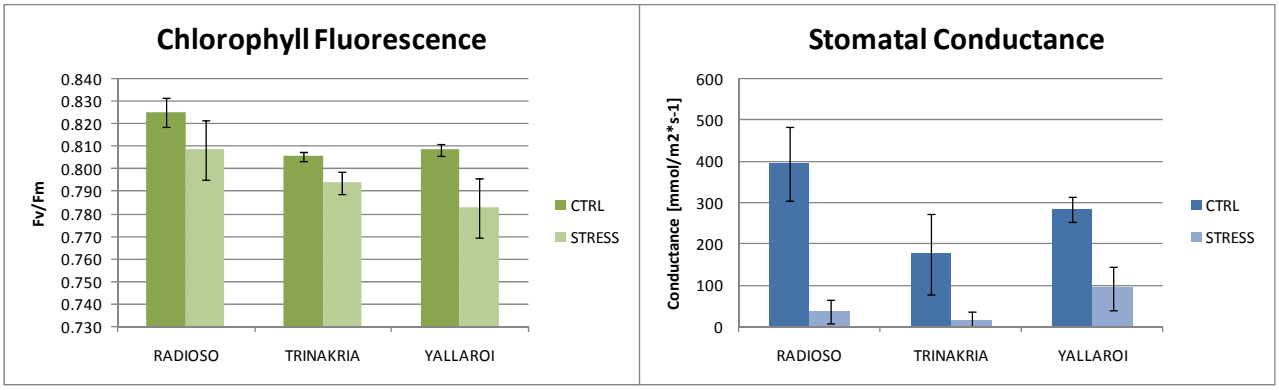


Fig. 22 (left) *Chlorophyll fluorescence measured with a Mini-Pam on dark-adapted leaves (3 biological replicates per measure) of plants subjected to severe drought stress. Values reported refer to the day before harvest (day 12).*
(right) *Stomatal conductance measured 5 hours after dawn with a leaf porometer (6 biological replicates per measure) on plants subjected to severe drought stress. High error bars show a general unreliability of these measures. Values reported refer to the day before harvest (day 12).*

Analysis of drought-related microRNAs expression trend

The expression trend of ten microRNAs that were found to be differentially expressed in Ofanto-Cappelli experiment (*osa-miR1432*, *tae-miR156*, *bdi-miR159-3p*, *ttu-miR160*, *tae-miR167a*, *ath-miR167d*, *osa-miR169c-1*, *osa-miR171h*, *osa-miR444e*, *bdi-miR5200*) was assessed with a qPCR approach.

Results outlined a high degree of similarity with previous data coming from the Ofanto-Cappelli experiment, especially in the cases of *osa-miR1432*, *bdi-miR159-3p* and *osa-miR171h*, which are reported in figure 23. It is worth noting that in Yallaroi these microRNAs exhibited the same expression trend as in the other four italian cultivars

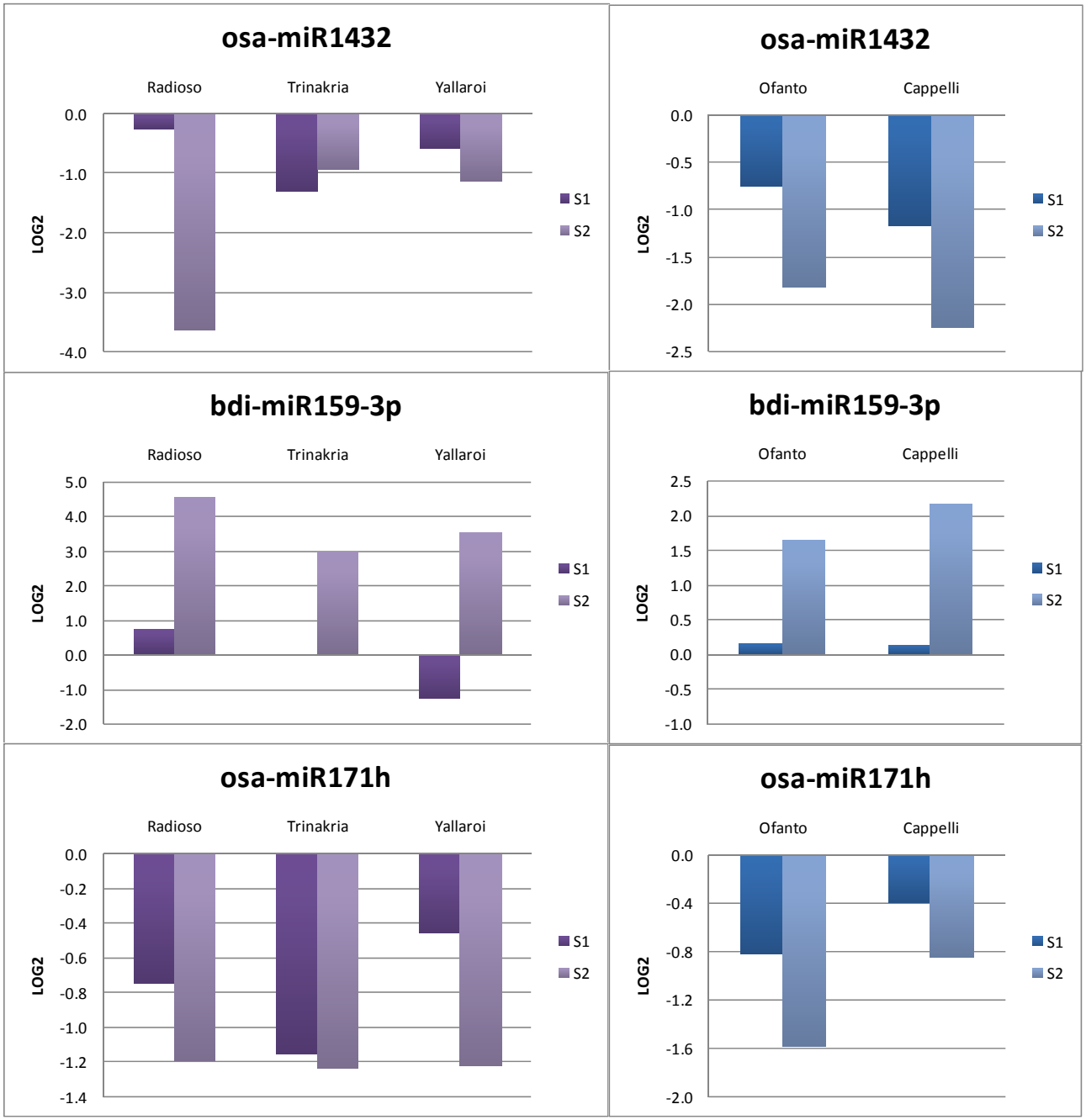


Fig. 23 *The expression trend of three selected microRNAs is reported. Left panel (purple bars) shows the behaviour of these microRNAs in the Radioso-Trinakria experiment, whereas right panel (blue bars) shows the expression trend of the same microRNAs in the Ofanto-Cappelli experiment. In both cases, data come from stem-loop qPCR validations.*
DS1 = mild drought stress, DS2 = severe drought stress.

The only exceptions concerned *bdi-miR5200* and *osa-miR444e*. The former was static instead of being down-regulated in the drought-susceptible genotype (Radioso) as it was in Ofanto. Moreover it is strangely upregulated in Yallaroi. The latter was induced (with a calculated Fold Change of 2.55) in the drought-tolerant cultivar instead of being static. These two microRNAs have never been reported for being modulated under drought stress, and they may be not really associated with the drought response.

Target search

Different approaches were considered in the search for target mRNAs of selected stress-modulated microRNAs. Initially, *in silico* analyses allowed the identification of many putative targets, which are reported in Table 11. Most of them are transcription factors and in many cases they are consistent with what is reported in literature, though some new unexplored possibility is present.

Target validation is a crucial step. A 5'-RACE procedure was developed, along with a qPCR confirmation. Moreover, degradome libraries were prepared during the internship in the foreign laboratory.

Since mRNA slicing is the preferential way of action of plant microRNAs, different targets identified *in silico* were checked via a 5-RLM-RACE (RNA-Ligase-Mediated Rapid Amplification of cDNA Ends) to verify that a microRNA-mediated cleavage occurred. Unfortunately, no definitive and reliable results were obtained: PCR products were uncertain and sequencing of the gel-eluted PCR-bands failed.

At the same time, a qPCR approach was tried. It was initially thought to be only a confirmation of RACE results, but due to time scarcity it was utilized concurrently with the other validation method. *Ad-hoc* primer couples were designed in order to have a forward primer upstream the putative miRNA cleavage site and a reverse one downstream the cleavage site. In this way only full-length mRNAs could be amplified, thus avoiding the "background noise" due to target parts that had not been caught by the degradation machinery. The expression trend of the real target is expected to be inversely correlated with the one of the microRNA that slices it. With this technique one likely target was identified: it was a MIKC-type MADS-box transcription factor (WM32B), which is targeted by *osa-miR444e* during heat stress. Its RACE-mediated validation failed to give a unique band to sequence.

Finally, degradome libraries were prepared at the ACPFG laboratory with Radioso and Trinakria leaf samples subjected to severe drought, with a new protocol prepared by Dr. Bujun Shi. The peculiarity of this experimental protocol consists in the length of the generated reads. Standard degradome sequencing output is made up of (approximately) 22-nt-long reads, since it is based on an enzymatic cleavage of adapter-ligated degradation products. The consequent difficulty lies in the bioinformatic analysis of the results, a time-consuming operation that becomes a hard task when the genome sequence of the studied species is unavailable. This protocol allows instead the full retrotranscription of adapter-ligated products, and therefore the amplification of longer transcripts (see Chapter 3). The prepared libraries underwent 150 cycles of sequencing on an Illumina Hi-Seq platform. Results analysis is still in progress.

Discussion

Wheat is the most widely grown crop world-wide; much work has gone into understanding its genetics and genomics, yet its small RNAs are just starting to be characterized (Yao and Sun, 2012). To date, few reports on wheat microRNAs were published, and none regarded durum wheat. This is reflected also in the number of annotated microRNAs deposited in miRBase for *Triticum turgidum* (only one), a fact that made miRNA name assignation in this work a task not easy to solve.

Drought and heat stresses are two major threats of durum wheat, given the fact that it is mainly cultivated in Mediterranean or Mediterranean-like areas (like the southern part of Australia). These regions are often characterized by water scarcity and high temperatures, frequently occurring at the same time (Aprile *et al.*, 2013). A first consideration regards the likely interconnection between the responses to these stresses. As it can be noticed in table 10, almost 60% of drought-related microRNAs found in this study is modulated also during heat stress. This set of stress-responsive microRNAs specify a pool of target genes likely

belonging to the same adaptation pathways that plants activate to cope with such abiotic stresses. A list of *in silico* identified targets is reported in table 11 at the end of this chapter. In this experiment a combination of both drought and heat stresses was not considered, since they were applied independently. The study of the plant miRNome in such a situation could result in interesting outcomes; previous findings highlighted that this combination is a unique stress sharing a marginal portion of the molecular responses activated by singular stresses alone (Aprile *et al.*, 2013 and references therein).

Owing to difficulties encountered during validation procedures, we have almost no validated targets, therefore a description of the situation coming out from miRNome analysis can only be theoretical. Accumulation of the microRNA and of its target mRNA is usually inferred to be inversely correlated, but their interaction is indeed a complex process. miRNA and target expression domains could coincide in the sampled tissues, but this might not be true in every mode of regulation, or in every time point considered. The study of one putative target at a time can easily not come to anything, especially if available online EST databases contain limited information, likely out of date. Moreover, in this specific case, another complication resided in the fact that the database utilized by the target prediction tool[5] contained hexaploid wheat ESTs. A hypothetic target encoded by the D genome cannot be found in *Triticum turgidum*. For these reasons, despite its drawbacks, like the need of time-consuming and often complicated bioinformatic analyses, degradome sequencing still represents the most advisable procedure to identify miRNA cleavage products.

Against this background, the scenario outlined by the expression trend of the drought-responsive microRNAs found in this experiment is in general agreement with findings reported in literature. *miR169*, *miR159*, *miR167* and miR1432 can be grouped as microRNAs involved in regulation of stomatal movements.

The repression of *miR169* (three family members showed strong down-regulation, figure 18) has been reported for several species and this miRNA is known to be a key regulator in plants adaptation to drought. It has been hypothesized that its expression in guard cells controls stomatal aperture, while in other cells it may contribute to the expression of stress-associated genes (Khraiwesh *et al.*, 2011; Sunkar *et al.*, 2012; Ding *et al.*, 2013). It is not a case, therefore that a stronger down-regulation can be seen in Cappelli compared to Ofanto, given Cappelli's attitude of "water saver".

The strong upregulation of a *miR159* family member (*bdi-miR159-3p*, described in figure 18) has been already described in drought response and linked with abscissic acid (ABA) signaling. ABA is produced *de novo* under water-deficit conditions and plays a central role in mediating the expression of stress-related genes and in the initiation of stomatal closure (Ding *et al.*, 2013).

A possible role in stomatal movement can be held also by *miR167*. It has been reported that, besides Auxin Response Factors (ARFs), this miRNA can cleave Phospholipase D (PLD) mRNAs, which are involved in stomatal closure under drought stress (Ding *et al.*, 2013). Recalling Cappelli's and Ofanto's different ability to quickly close stomata under incipient drought, their differential behaviour regarding *miR167* expression trend was not so surprising. Three *miR167* family members (*tae-miR167a*, *bdi-miR167c* and *ath-miR167d*) were induced in Cappelli only.

Finally, *osa-miR1432* showed a substantial decrease in drought- and heat-stressed samples. Its down-regulation under drought was described also in *Triticum dicoccoides*, the wild ancestor of durum wheat,

[5] (DFCI Gene Index (TAGI), version 12, released on 18-04-2010).
ftp://occams.dfci.harvard.edu/pub/bio/tgi/data/Triticum_aestivum/TAGI.release_12.zip

(Kantar *et al.*, 2011), while it was not reported in a recent study on bread wheat response to heat stress (Xin *et al.*, 2010). It is predicted to target a calmodulin-like binding protein (Table 11), a fact that suggests its involvement in calcium signaling. This network is implicated in the response to abiotic stresses like drought and heat: Ca^{2+} signals are core regulators of stomatal aperture (Dodd *et al.*, 2010; Batistic and Kudla, 2012).

miR171 has been already characterized for being regulated under drought stress. A general trend could not be described since up- and down-regulations were seen in different studies even for the same species, e.g. rice and poplar (Sunkar *et al.*, 2012). Down-regulation has been reported in *Triticum dicoccoides* subjected to shock drought stress (Kantar *et al.*, 2011). Accordingly, here *miR171* was found to be modulated in the same way by both cultivars. *In silico* searches failed to find a convincing target, but a recent paper by Li and colleagues (2013), reporting the results of a degradome sequencing of bread wheat (*Triticum aestivum*), stated that *miR171* targets belong to the Scarecrow-like (SCL) family of transcription factors. As reported in Chapter 1, SCLs are involved in gibberellin signaling and in the control of plant development under stress conditions.

Few reports on plant miRNAs modulated by heat stress have been published so far. The closest one to the present work is a paper by Xin *et al*. (2010), who investigated the response to high temperatures (40° C for one hour) of a heat-tolerant *Triticum aestivum* cultivar. Their findings show only partial agreement to ours; it is restricted to the up-regulation of *miR166* (both cultivars) and *miR169* (Cappelli only) family members. This can be due to many factors, ranging from the different species considered, the duration and the intensity of the stress, the age of plants etc.

Induction of *osa-miR444e* was observed in both cultivars. One of its targets, WM32B, was identified via 5'-RACE, but sequencing of the PCR band failed, and its expression, verified by a qPCR, turned out to be very low; WM32B belongs to the MADS-box family of transcription factors, which has often been linked with various aspects of development and reproductive processes, and with abiotic stress responses (Tardif *et al.*, 2007).

A final consideration concerns the Radioso-Trinakria experiment that was carried out during the abroad internship. It confirmed that the identified microRNAs are indeed drought-responsive, since their expression profiles remained unchanged when the Ofanto-Cappelli system was reproduced with different cultivars. Moreover, these results were strengthened by the behaviour of the australian durum wheat cultivar, Yallaroi. Its microRNAs expression profile was in agreement with that of the other four italian cultivars, thus demonstrating the lack of relatedness of the results with the breeding group.

As for physiological parameters, it is worth noting that Trinakria showed the same water-saving attitude of Cappelli. On incipient drought, Trinakria plants were losing less water per day (figures 20 and 21) and started to close their stomata earlier than Radioso, but these differences dropped down when the SWC (Soil Water Content) decreased below 18%, *i.e.* the selected threshold for mild drought stress. Stomatal conductance measurements were carried out with a porometer. The high error bars show a general unreliability of this measure, but the depicted scenario is consistent with the meaning of this parameter: plants close their stomata upon drought sensing, and the extent of this closure is greater in the water saver cultivar. MiRNome analysis showed that this mechanism is mediated by rapid activation/repression of a core set of microRNAs.

***Table 11**. Putative targets for drought and heat modulated microRNAs found in this study*

miRNA	Sequence	Target Accession	Exp.	Inhibition	Target Description
tae-miR156-2	TGACAGAAGAGAGTGAGCACA	TC453361	1	Cleavage	Teosinte glume architecture 1
		CK196549	1	Cleavage	Squamosa promoter-binding-like protein 2
		AL810223	1	Cleavage	Squamosa promoter-binding-like protein 16
		TC409846	1	Cleavage	Isoform 2 of Q6H509
		TC384445	2	Cleavage	Squamosa promoter-binding-like protein 3
		TC420438	2	Cleavage	unknown protein
		TC412204	2	Cleavage	Squamosa promoter-binding-like protein 11
		TC441570	2	Cleavage	Alr2394 protein
		DR739383	2.5	Cleavage	Glycosyl/glycerophosphate transferase
		TC373290	3	Cleavage	unknown protein
		TC398965	3	Cleavage	Predicted permease
		TC460639	3	Cleavage	
		TC372857	3	Cleavage	
		TC390294	3	Cleavage	Cob(I)alamin adenolsyltransferas
		CA741955	3	Cleavage	Cytochrome P450
tae-miR156	TGACAGAAGAGAGTGAGCAC	TC409846	1	Cleavage	Isoform 2 of Q6H509
		TC453361	1	Cleavage	Teosinte glume architecture 1
		CK196549	1	Cleavage	Squamosa promoter-binding-like protein 2
		AL810223	1	Cleavage	Squamosa promoter-binding-like protein 16
		TC384445	2	Cleavage	Squamosa promoter-binding-like protein 3
		TC420438	2	Cleavage	unknown protein
		TC412204	2	Cleavage	Squamosa promoter-binding-like protein 11
		TC441570	2	Cleavage	Alr2394 protein
		DR739383	2.5	Cleavage	Glycosyl/glycerophosphate transferase
		TC373290	3	Cleavage	unknown protein
		TC390294	3	Cleavage	Cob(I)alamin adenolsyltransferas
		CA741955	3	Cleavage	Cytochrome P450
		TC398965	3	Cleavage	Predicted permease
		TC370322	3	Cleavage	Telomere binding protein
		TC460639	3	Cleavage	
ath-miR156g	CGACAGAAGAGAGTGAGCAC	TC409846	2	Cleavage	Isoform 2 of Q6H509
		TC453361	2	Cleavage	Teosinte glume architecture 1
		CK196549	2	Cleavage	Squamosa promoter-binding-like protein 2
		AL810223	2	Cleavage	Squamosa promoter-binding-like protein 16
		TC395061	3	Cleavage	F-box family protein
		TC409511	3	Cleavage	Predicted protein
		TC390294	2.5	Cleavage	Cob(I)alamin adenolsyltransferas
		TC398965	2.5	Cleavage	Predicted permease
		TC384445	3	Cleavage	Squamosa promoter-binding-like protein 3
		TC420438	3	Cleavage	unknown protein
		TC412204	3	Cleavage	Squamosa promoter-binding-like protein 11
		TC441570	3	Cleavage	Alr2394 protein
		TC370322	2.5	Cleavage	Telomere binding protein
		TC413555	2.5	Cleavage	Telomere binding protein
		CA612886	2.5	Cleavage	Initiator binding protein
tae-miR159a/b	TTTGGATTGAAGGGAGCTCTG	CA679118	2	Cleavage	Alkaline phosphatase family protein
		CD898447	2.5	Cleavage	Os10g0392700 protein
		TC388353	2.5	Cleavage	NB-ARC domain containing protein
		TC368630	2.5	Cleavage	Transcription factor Myb3
		CA484819	2.5	Cleavage	Acyl-CoA synthetase
		TC377621	3	Cleavage	Ubiquitin carboxyl-terminal hydrolase
		TC421314	3	Cleavage	Transcription factor Myb3
		TC393503	3	Cleavage	Ubiquitin carboxyl-terminal hydrolase
		CA593588	3	Cleavage	Os01g0212500 protein
		CA671152	2.5	Cleavage	TLD family protein, expressed
		CD915742	3	Cleavage	Mxi1 protein
		CD916046	3	Cleavage	Mob1-like protein
		TC393182	3	Cleavage	
		TC402088	3	Cleavage	
		TC400267	3	Cleavage	
bdi-miR159-3p	CTTGGATTGAAGGGAGCTCT	TC368630	2	Cleavage	Transcription factor Myb3

		TC388353	2	Cleavage	NB-ARC domain containing protein
		TC377621	3	Cleavage	Ubiquitin carboxyl-terminal hydrolase
		TC421314	3	Cleavage	Transcription factor Myb3
		TC393503	3	Cleavage	Ubiquitin carboxyl-terminal hydrolase
		CA679118	3	Cleavage	Alkaline phosphatase family protein
		CA683024	3	Cleavage	
		TC452060	3	Cleavage	Topoisomerase 6 subunit A
ttu-miR160	TGCCTGGCTCCCTGTATGCCA	TC412104	0	Cleavage	Auxin response factor 8
		TC416747	0	Cleavage	Auxin response factor 8
		TC425315	0	Cleavage	Auxin response factor 22
bdi-miR160e-5p	TGCCTGGCTCCCTGAATGCCA	TC412104	1	Cleavage	Auxin response factor 8
		TC416747	1	Cleavage	Auxin response factor 8
		TC425315	1	Cleavage	Auxin response factor 22
		TC411588	2	Translation	Peroxidase
ath-miR165a	TCGGACCAGGCTTCATCCCCC	TC426980	1.5	Cleavage	Homeodomain leucine-zipper protein Hox9
		TC404537	1.5	Cleavage	Class III HD-Zip protein 4
		CK204430	1.5	Cleavage	Class III HD-Zip protein 4
		TC443436	1.5	Cleavage	Class III HD-Zip protein 4
		TC394512	3	Cleavage	Os10g0118900 protein
		TC407496	3	Cleavage	
		TC445130	2.5	Cleavage	unknown protein
		GH732388	3	Cleavage	2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase
		TC389346	3	Cleavage	unknown protein
		CK153712	3	Cleavage	
osa-miR166g	TCGGACCAGGCTTCATTCCTT	CK204430	1	Cleavage	Class III HD-Zip protein 4
		TC443436	1	Cleavage	Class III HD-Zip protein 4
		TC426980	2	Cleavage	Homeodomain leucine-zipper protein Hox9
		TC404537	2	Cleavage	Class III HD-Zip protein 4
		GH732388	2.5	Cleavage	2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase
		TC397559	3	Translation	Ubiquitin carboxyl-terminal hydrolase
		TC369648	3	Translation	Ubiquitin carboxyl-terminal hydrolase
		TC371066	3	Translation	Ubiquitin carboxyl-terminal hydrolase
		CD903435	3	Translation	unknown protein
		TC427244	3	Cleavage	
		TC457108	2.5	Cleavage	
bdi-miR166h	TCGGACCAGGCTTCAATCCCT	CK204430	1.5	Cleavage	Class III HD-Zip protein 4
		TC443436	1.5	Cleavage	Class III HD-Zip protein 4
		TC429172	3	Cleavage	Hydroxyproline-rich glycoprotein DZ-HRGP precursor
		TC389346	2.5	Cleavage	unknown protein
		TC426980	3	Cleavage	Homeodomain leucine-zipper protein Hox9
		TC404537	3	Cleavage	Class III HD-Zip protein 4
bdi-miR166h-2	TCGGACCAGGCTTCAATCCCG	CK204430	1.5	Cleavage	Class III HD-Zip protein 4
		TC443436	1.5	Cleavage	Class III HD-Zip protein 4
		TC429172	3	Cleavage	Hydroxyproline-rich glycoprotein DZ-HRGP precursor
		TC447927	3	Cleavage	
		TC389346	2.5	Cleavage	unknown protein
		TC426980	3	Cleavage	Homeodomain leucine-zipper protein Hox9
		TC404537	3	Cleavage	Class III HD-Zip protein 4
tae-miR167a	TGAAGCTGCCAGCATGATCTA	TC422705	3	Cleavage	
	TGAAGCTGCCAGCATGATCTGA	TC402576	3	Cleavage	
		CV773862	3	Cleavage	
		TC422705	3	Cleavage	
ath-miR167d	TGAAGCTGCCAGCATGATCTGG	TC422705	3	Cleavage	
		TC402576	3	Cleavage	
		CV773862	3	Cleavage	
ath-miR167d-2	TGAAGCTGCCAGCATGATCTGC	TC402576	3	Cleavage	
		CV773862	3	Cleavage	
		TC422705	3	Cleavage	
osa-miR169c-1	CAGCCAAGGATGACTTGCCGGC	TC426343	2.5	Cleavage	CCAAT-box transcription factor complex WHAP12
		TC370257	2.5	Cleavage	CCAAT-box transcription factor complex WHAP12
		TC430923	2.5	Cleavage	CCAAT-box transcription factor complex WHAP6

		NP958308	2.5	Cleavage	CCAAT-box transcription factor complex WHAP9
		TC377727	2.5	Cleavage	RAPB protein
		TC387436	2.5	Cleavage	CCAAT-box transcription factor complex WHAP12
		TC399072	2.5	Cleavage	
		TC380679	2.5	Cleavage	
		BJ268074	2	Cleavage	Os03g0430500 protein
		TC391576	3	Cleavage	
		CA664884	3	Translation	Photosystem II 10 kDa polypeptide, chl. precursor
osa-miR169n	TAGCCAAGAATGACTTGCCTG	TC391576	3	Translation	
osa-miR169h	TAGCCAAGGATGACTTGCCTG	TC391576	2	Cleavage	
		TC452786	2.5	Cleavage	
		TC448322	2.5	Cleavage	
		TC426343	3	Cleavage	CCAAT-box transcription factor complex WHAP12
		TC370257	3	Cleavage	CCAAT-box transcription factor complex WHAP12
		TC430923	3	Cleavage	CCAAT-box transcription factor complex WHAP6
		NP958308	3	Cleavage	CCAAT-box transcription factor complex WHAP9
		CA664884	3	Translation	Photosystem II 10 kDa polypeptide, chl. precursor
		TC377727	3	Cleavage	RAPB protein
		TC387436	3	Cleavage	CCAAT-box transcription factor complex WHAP12
		TC399072	3	Cleavage	
		TC380679	3	Cleavage	
		TC456486	3	Cleavage	Transport energizing protein, ExbD/TolR family
osa-miR171h	TGAGCCGAACCAATATCACTC	TC460317	2.5	Cleavage	sensor histidine kinase
		CA624788	3	Translation	
		CJ529083	3	Translation	Oxysterol-binding protein
gma-miR393h	TTCCAAAGGGATCGCATTGAT	CA484228	2	Cleavage	TIR1-like auxin receptor protein
		TC422531	2	Cleavage	TIR1-like auxin receptor
		TC371524	2	Cleavage	Os05g0150500 protein
		TC388306	2	Cleavage	Os05g0150500 protein
		CK158797	2.5	Cleavage	Ubiquitin fusion degradation protein
		TC432387	3	Cleavage	unknown protein
		TC381513	3	Cleavage	unknown protein
		TC387372	3	Cleavage	unknown protein
		TC426992	3	Cleavage	unknown protein
bdi-miR399b	TGCCAAAGGAGAATTGCCCTG	CK210056	1.5	Cleavage	Uncharacterized protein At1g11850.3
		TC434302	2.5	Cleavage	Peptidoglycan-associated lipoprotein
		TC429934	3	Cleavage	unknown protein
		CK164434	3	Cleavage	SJCHGC01964 protein
		TC398128	3	Cleavage	Expressed protein
osa-miR399d	TGCCAAAGGAGAGTTGCCCTG	CK210056	1.5	Cleavage	Uncharacterized protein At1g11850.3
		CK156995	3	Cleavage	4Fe-4S ferredoxin, iron-sulfur binding
		CA682916	3	Translation	Predicted protein
		TC434302	3	Cleavage	Peptidoglycan-associated lipoprotein
		BE403144	3	Cleavage	Non-structural protein 5A
osa-miR444e	TGCAGTTGCTGCCTCAAGCTT	TC368997	0	Cleavage	MIKC-type MADS-box transcription factor WM32B
		TC396893	0	Cleavage	Zinc finger (C3HC4-type RING finger) protein-like
		TC403194	0.5	Cleavage	Zinc finger (C3HC4-type RING finger) protein-like
		CA484682	2	Cleavage	
		TC424265	2	Translation	MIKC-type MADS-box transcription factor WM30
		CN011704	2	Cleavage	
		NP9351348	2.5	Cleavage	MADS-box transcription factor TaAGL6
		TC370370	2	Translation	MIKC-type MADS-box transcription factor WM30
		TC368730	2.5	Cleavage	MIKC-type MADS-box transcription factor WM31B
		TC446683	2.5	Cleavage	Ribulose bisphosphate carboxylase small chain
		CA649090	2.5	Cleavage	unknown protein
		DR733291	2	Cleavage	Os05g0499500 protein
		CK171762	3	Cleavage	Predicted protein
		TC438971	2	Translation	
		TC399203	2	Translation	MIKC-type MADS-box transcription factor WM32A
		NP13081097	3	Cleavage	MIKC-type MADS-box transcription factor WM31C
		TC409271	2.5	Cleavage	CDK5 activator-binding protein-like
		BF202063	3	Cleavage	
		TC400959	3	Cleavage	Os07g0525900 protein
		BE404243	3	Cleavage	Myosin heavy chain-like protein
bdi-miR827-3p	TTAGATGACCATCAGCAAACA	CK163248	3	Cleavage	Predicted protein

		BU099421	3	Translation	
		TC389532	3	Cleavage	ATP-dependent Clp protease ATP-binding subunit
osa-miR1432	ATCAGGAGAGATGACACCGAC	CK195521	0	Cleavage	
		TC459334	0.5	Cleavage	Pilus-associated protein, PilW-related precursor
		TC388166	0.5	Cleavage	
		TC402640	0.5	Cleavage	
		CN011453	1.5	Cleavage	
		TC397132	2.5	Cleavage	unknown protein
		TC405736	2.5	Cleavage	unknown protein
		BQ838967	2	Cleavage	unknown protein
		TC401422	2	Cleavage	unknown protein
		CA659674	3	Translation	Mitochondrial phosphate transporter
osa-miR5071	TCAAGCATCATGTCATGGACA	TC436573	0	Cleavage	Os06g0683600 protein
		CJ627170	0	Cleavage	Os10g0132700 protein
		TC448757	1	Cleavage	Os12g0273300 protein
		CJ689070	1	Cleavage	Os10g0131000 protein
		CK163066	1	Cleavage	Os10g0131000 protein
		TC424148	1.5	Cleavage	
		TC376415	1.5	Cleavage	Os10g0131000 protein
		BQ752781	1.5	Cleavage	
		FL488865	2	Translation	NB-ARC domain containing protein
		CA501245	2	Cleavage	NB-ARC domain containing protein, expressed
		DR736474	2	Cleavage	NB-ARC domain containing protein, expressed
		TC413791	2	Cleavage	
		TC444623	2	Cleavage	
		CV765869	2	Translation	unknown protein
		TC436025	2	Translation	unknown protein
		TC393731	2	Translation	unknown protein
		GH727096	2	Cleavage	Os01g0603100 protein
		CD910548	2.5	Cleavage	NB-ARC domain containing protein
		TC444204	2.5	Cleavage	OmpA/MotB
		TC426546	2.5	Cleavage	Os08g0202400 protein
		BE607074	2.5	Translation	60S acidic ribosomal protein P0
		CD933947	2.5	Cleavage	Predicted protein
		BF484266	3	Cleavage	Os11g0130900 protein
		CA600148	3	Cleavage	Os11g0130900 protein
		TC416230	3	Cleavage	Predicted protein
bdi-miR5200	TGTAGATACTCCCTAAGGCTT	TC407412	2.5	Cleavage	Flowering locus T
		TC442593	2.5	Cleavage	Flowering locus T
		TC414319	2.5	Cleavage	Flowering locus T
		TC383902	3	Translation	Serine/threonine-protein kinase HT1

Table 11 *Putative targets for studied microRNAs are reported. Results were obtained from http://plantgrn.noble.org/psRNATarget/. Target inhibition method is indicated along with Expectation, i.e. the likelihood that a given mRNA is targeted by the microRNA considered (0 = perfect match, 3 = 3 mismatches. G-U mismatches weight 0.5). miRNAs for which no target was found are not reported.*

General Discussion

Results reported allowed the characterization of microRNAs involved in plant development and response to abiotic stresses like drought, heat and low temperatures and, in most cases, the outline of possible scenarios in which they exert their functions.

The common thread of all the experiments analyzed, however, was no doubt the striking complexity of plant miRNomes, an aspect that can be dissected into many points.

The chapter 4.2 can provide the first example. Positioning microRNA loci in the genome has been one hot topic of the past years in the microRNA field. At the beginning, intergenic regions only were taken into consideration, but soon clear evidence appeared that they were not the sole source of miRNAs. Many cases of miRNA-coding intragenic loci were then described. This feature was initially thought to be a prerogative of animal miRNA machinery only, but recently it has been reported also for plants. Meng and Shao (2012) identified miRtrons, microRNAs produced in a Dicer-independent way by splicing from intronic regions, in *Arabidopsis* and rice. Different studies suggested different roles for intronic miRNAs: they can play an antagonistic or a synergetic role as an enemy or a partner of their host genes (Gao *et al.*, 2012a). In this background, we carried out a computational analysis that identified plant microRNAs putatively arising from exons of protein coding transcripts. The excision of a hairpin from an exon is, necessarily, a destructive event and the destabilization of the corresponding protein-coding mRNA, which leads to the depletion of protein synthesis, can suggest a new mechanism of post-transcriptional regulation of gene expression. This mechanism adds a new layer of complexity: it can give rise to a loop controlling the "source" mRNA, originating at the same time a mature sequence that can be incorporated in a RISC and target other genes of the pathway. The monocot-specific *miR444* family, identified in barley, wheat and other cereals, provides an interesting example. Its members are part of TCs coding for the MADS-box transcription factors family, which, at the same time, represent one of its conserved target. A barley EST, for instance, was found to contain both the target site for *miR444b* and the precursor sequence of *miR444a*.

Complexity is given also by another aspect that emerged from the analyses carried out in peach, barley and durum wheat experiments (Chapters 5.2, 5.3 and 5.5). Often, the same microRNA appeared with a great variety of different rearrangements, called isomiRs. The population of variants of miRNAs coming from the same precursors can greatly contribute in expanding the set of possible targets in the cell regulatory network. Additions or deletions can cause a shift of the seed region, or the recruitment of the miRNA in a different AGO-loaded RISC, finally affecting its efficacy and even its way of action. Most of the times, isomiRs generation did not seem to be random, being the most frequent read for a specific locus the same across different tissues and different replicates. Cases have been found, however, in which this was not true, meaning that a certain level of randomness is still present. The most likely explanation is that, in these situations, sequence modifications do not affect the microRNA function and therefore were not discarded by the evolution. Broadening the concept, these differences in isomiRs abundance were frequently seen among different cultivars (e.g. the cold-induced barley *miR159a** was 21-nt-long in Nure and 20-nt-long in Tremois), but this is only a face of the diversity at molecular level occurring between them. A consideration that can certainly be made concerns the difficulty to state which isoform is the "real" microRNA, making the most of the silencing action: frequently they all concur to the final result. Their annotation is a tricky task too. Often the reported sequence for a given miRNA differed from the most abundant sequence in our

databases. IsomiRs are commonly found in deep-sequencing studies but their functional meaning and stability is still under investigation in plants.

The third evidence of the complexity is given by the constitutive expression trends of plant microRNAs, which can change in different cultivars. Lots of examples were seen in barley and in durum wheat experiments. The continuous gradient characterizing *miR397* expression in a wide panel of barley genotypes did not come as a big surprise in the light of these considerations. Different family abundance among cultivars was often a fact, though no cases of missing loci were seen – at least for miRNAs whose sequence was deposited in miRBase. Furthermore, comparisons with published papers reporting on barley deep sequencing runs (Schreiber *et al.*, 2011; Lv *et al.*, 2012) remarked this feature. Sometimes highly frequent miRNA families in their analyses were barely visible in our databases, (e.g. *miR172* sequences did not pass the cutoff threshold (5 RPM) in our database, whereas they ranked third in Lv *et al.* (2012) analysis) and, in general, different relative families abundance was seen. No comparison could be made for durum wheat since no reports on its miRNome were found, except a study by Kantar *et al.* (2011), dealing with its wild ancestor (*Triticum dicoccoides*). This diversity at miRNA level between different studies is however only partly due to miRNome complexity; the most of it is likely conferred by the variety of experimental conditions employed and, last but not least, by the bioinformatic component (e.g. release of new miRBase versions, newly developed pieces of software relying on different grouping criteria etc.).

The absolute necessity of a bioinformatic support is a distinctive trait of this part of miRNA research field, as witnessed by its history. Bioinformatic analyses provided the starting point, with the identification of miRNA homologues in a number of species, then allowed the following step, *i.e.* the development of the next-generation sequencing technologies, thus bringing the attention also to those species with no genome availability, or anyway less considered. Lastly, degradome analysis provides a very powerful tool in understanding miRNA functions in different plants, but it has to be aided by computational tools. Its fundamental advantage over other methods like 5'-RACE is that it does not require an *a priori* determination of target sequences, it allows to skip the prediction step, and it even provides information on the relative abundance of cleaved targets (Addo-Quaye *et al.*, 2008).

In this thesis, a degradome sequencing was prepared for the durum wheat experiment, but results were not ready at the moment of writing. Quite a lot of trials have been made with methods based on computational predictions of targets, but most of the times they came up to nothing. As already mentioned, several explanations can be put forward, ranging from the lack of up to date (or species-specific) target databases, to the lack of spatio-temporal co-expression of a microRNA and its targets.

A core set of abiotic stress related microRNAs was identified, however. Most of it was made up of widely conserved microRNAs that are well known for their role in general response to such stresses, *i.e. miR156*, *miR159*, *miR167*, *miR169*, *miR171* and *miR399*, which were found to be modulated in every experiment. *miR166* regulation was reported only for extreme temperatures (Khraiwesh *et al.*, 2011), and so it was in our databases. Finally, we found other less conserved and less known microRNAs showing a common responsiveness to abiotic stresses, that is, *miR444* (only in high or low temperature stress), *miR827* and *miR5200*. These three microRNAs are no doubt less characterized than the others above reported, and they can maybe represent a new field to explore in unraveling cereals adaptation mechanisms to abiotic stress.

The differential behaviour of contrasting cultivars found under stress was not surprising, since it has been reported already for many species, like *Vigna unguiculata* (Barrera-Figueroa *et al.*, 2011) *Brassica napus* (Zhao *et al.*, 2012) and cotton (Yin *et al.*, 2012). The differences at molecular level in Nure-Tremois and

Ofanto-Cappelli systems were known, but this is the first inquiry into their miRNomes adaptation to the stress for which they show a different tolerance. Results here reported lacked of final evidence regarding targeted genes, therefore only hypotheses partially contributing to an explanation of these differential behaviours could be put forth.

qPCR validation of the results was carried out with a specific protocol to target microRNAs (Varkonyi-Gasic *et al.*, 2007; see materials and methods). Though the protocol was quite easy to adapt to our experimental systems, a key node was represented by the normalization system. It has been highlighted the crucial importance of choosing endogenous controls that share similar properties to miRNAs, such as RNA stability and size, and are amenable to the miRNA assay design (Ceccardi *et al.*, 2008). Several papers reported the abundant and stable expression of small non-coding RNAs (ncRNAs) classes, like small-nucleolar RNAs (snoRNAs), small-nuclear RNAs (snRNAs) and even microRNAs, that make them good candidates for use as endogenous controls (Davoren *et al.*, 2008; Galiveti *et al.*, 2010; Kulcheski *et al.*, 2010; Feng *et al.*, 2012; Kou *et al.*, 2012). Following this line, we identified a snoRNAs database and tested many of those for suitability as normalization controls in our qPCR-based validation procedures. Some of them, namely Hv-U59 and Hv-U61 for barley, and Ta-snoR10 for wheat, showed an expression level similar to mid-expressed microRNAs and a good stability in all the stress conditions of the present experiment, and were therefore chosen as endogenous controls along with classical housekeeping genes (*i.e.*, *Hv-Actin* and *Hv-Polyubiquitin*, *Ta-polyubiquitin*) or other genes taken from previous works (*OEP16-3*).

Future perspectives

This work represents a step forward towards the comprehension of the complex gene regulatory mechanisms acting in plant development and adaptation to stresses, but still it is only a starting point for future studies, a tool for focused searches on traits potentially useful for the improvement of a species.

The most immediate direction could be the validation of housekeeping genes belonging to the small RNA class for quantitative PCRs. We tested some small nucleolar RNAs (snoRNAs) taken from an online snoRNA database[6]. Their expression profiles were compared to those of classic housekeeping genes (like actin, polyubiquitin and GAPDH) and a very good correlation was found. The next step could be the test of their suitability in different stress conditions.

A target analysis of stress-responsive microRNAs would be fundamental for a progress of our knowledge on stress response and differential behaviour of stress tolerant and stress susceptible genotypes. Ofanto and Cappelli have already been characterized at transcriptome level to find drought and heat responsive genes (Aprile *et al.*, 2013), and so were Nure and Tremois in respect of cold tolerance (Tondelli *et al.*, 2011). In the present work, hypotheses on microRNAs contributing to this difference have been made, but they require a confirmation through a target analysis. A degradome sequencing would be the most advisable way. The final goal, however, is the exploitation of these newly discovered regulatory mechanisms for the improvement of plants of agricultural interest.

[6] http://bioinf.scri.sari.ac.uk/cgi-bin/plant_snorna/home

References

Addo-Quaye C., Eshoo T.W., Bartel D.P., Axtell M.J., (2008), Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. *Current biology* 18(10): 758-62.

Allen E., Xie Z., Gustafson A.M., Carrington J.C., (2005), MicroRNA-directed phasing during *trans*-acting siRNA biogenesis in plants. *Cell* 121(2): 207-221.

Allen J.O., (2005), Effect of teosinte cytoplasmic genomes on maize phenotype. *Genetics* 169(2): 863-880.

Almeida M.I., Reis R.M., Calin G.A., (2011), MicroRNA history: discovery, recent applications, and next frontiers. *Mutation Research* 717(1-2): 1-8.

Ameres S.L., Zamore P.D., (2013), Diversifying microRNA sequence and function. *Nature Reviews Molecular Cell Biology* 14(8): 475-88.

Angaji, S.A., and Darvishani S., (2012), RNAa: A Few Key Points. *Annals of Biological Research* 3 (9): 4367-4373

Aprile A., Havlickova L., Panna R., Marè C., Borrelli G.M., Marone D., Perrotta C., Rampino P., De Bellis L., Curn V., Mastrangelo A.M., Rizza F., Cattivelli L., (2013), Different stress responsive strategies to drought and heat in two durum wheat cultivars with contrasting water use efficiency. *BMC Genomics* 14: 821.

Atienza S.G., Martín A.C., Ramírez M.C., Martín A., Ballesteros J., (2007), Effects of *Hordeum chilense* cytoplasm on agronomic traits in common wheat. *Plant Breeding* 126: 5-8.

Atienza S.G., Martín A., Pecchioni N., Platani C., Cattivelli L., (2008), The nuclear-cytoplasm interaction controls carotenoid content in wheat. *Euphytica* 159: 325-331.

Axtell M.J., Westholm J.O., Lai E.C., (2011), Vive la différence: biogenesis and evolution of microRNAs in plants and animals. *Genome Biology* 12(4): 221.

Axtell M.J., (2013), Classification and Comparison of Small RNAs from Plants. *Annual Review of Plant Biology,* 64: 137-159.

Barakat A., Sriram A., Park J., Zhebentyayeva T., Main D., Abbott A., (2012), Genome wide identification of chilling responsive microRNAs in *Prunus persica* . *BMC Genomics* 13: 481.

Barrera-Figueroa B.E., Gao L., Diop N.N., Wu Z., Ehlers J.D., Roberts P.A., Close T.J., Zhu J., and Liu R., (2011), Identification and comparative analysis of drought-associated microRNAs in two cowpea genotypes. *BMC Plant Biology* 11: 127.

Barrera-Figueroa B.E., Wu Z., Liu R., (2013), Abiotic stress-associated microRNAs in plants: discovery, expression analysis, and evolution. *Frontiers in Biology* 8(2): 189-197.

Batistič O., Kudla J., (2012), Analysis of calcium signaling pathways in plants. *Biochimica et Biophysica Acta* 1820(8): 1283-93.

Baumberger N., Baulcombe D.C., (2005), Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *PNAS* 102(33): 11928-33.

Bernstein E., Caudy A.A., Hammond S.M., Hannon G.J., (2001), Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409(6818): 363-366.

Bohmert K., Camus I., Bellini C., Bouchez D., Caboche M., Benning C., (1998), AGO1 defines a novel locus of Arabidopsis controlling leaf development. *EMBO Journal,* 17(1): 170-80.

Bonghi C., Trainotti L., Botton A., Tadiello A., Rasori A., Ziliotto F., Zaffalon V., Casadoro G., Ramina A., (2011), A microarray approach to identify genes involved in seed-pericarp cross-talk and development in peach. *BMC Plant Biology* 11: 107.

Bonnet E., Wuyts J., Rouzé P., Van de Peer Y., (2004), Detection of 91 potential conserved plant microRNAs in *Arabidopsis thaliana* and *Oryza sativa* identifies important target genes. *PNAS*, 101(31): 11511–11516

Brennecke J., Aravin A.A., Stark A., Dus M., Kellis M., Sachidanandam R., Hannon G.J., (2007), Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* 128(6): 1089-103.

Bürling K., Cerovic Z.G., Cornic G., Ducruet J. M., Noga G., and Hunsche M., (2013), Fluorescence-based sensing of drought-induced stress in the vegetative phase of four contrasting wheat genotypes. *Environmental and Experimental Botany* 89(3): 51-59.

Carthew R.W., Sontheimer E.J., (2009), Origins and Mechanisms of miRNAs and siRNAs. *Cell* 136(4): 642-55.

Ceccardi T.L., Goldrick M.M., Ren P., Conrad R.C., and Chen C., (2008), Real-Time Quantitation of MicroRNAs by TaqMan MicroRNA Assays. In "The Handbook of Plant Functional Genomics: Concepts and Protocols". Edited by G. Kahl and K. Meksem, Copyright © 2008 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim ISBN: 978-3-527-31885-8.

Chapman, P. J., and Catlin, G. A. (1976). Growth stages in fruit trees from dormant to fruit set. *N. Y. Food Life. Sci. Bull. 58*

Chen X., (2009), Small RNAs and their roles in plant development. *Annual review of cell and developmental biology* 25: 21-44.

Chinnusamy V, Zhu JK, Sunkar R.(2010) Gene regulation during cold stress acclimation in plants. *Methods Mol Biol*., 639:39-55.

Colaiacovo M., Subacchi A., Bagnaresi P., Lamontanara A., Cattivelli L., Faccioli P., (2010), A computational-based update on microRNAs and their targets in barley (*Hordeum vulgare* L.). *BMC Genomics* 11: 595.

Covey S.N., Al-Kaff N.S., Langara, A., and Turner D.S. (1997), Plants combat infection by gene silencing. *Nature* 385: 781-782.

Cuperus J.T., Fahlgren N., Carrington J.C., (2011), Evolution and functional diversification of MIRNA genes. *Plant Cell* 23(2): 431-42.

Davoren P.A., McNeill R.E., Lowery A.J., Kerin M.J., Miller N., (2008), Identification of suitable endogenous control genes for microRNA gene expression analysis in human breast cancer. *BMC Molecular Biology* 9: 76.

De Leonardis A.M., Marone D., Mazzucotelli E., Neffar F., Rizza F., Di Fonzo N., Cattivelli L., Mastrangelo A.M., (2007), Durum wheat genes up-regulated in the early phase of cold stress are modulated by drought in a developmental and genotype dependent manner. *Plant Science* 172: 1005-1016.

De Vita P., Nigro F.P., (2007), Breeding progress in morpho-physiological, agronomical and qualitative traits of durum wheat cultivars released in Italy during the 20th century. *European Journal of Agronomy* 26: 39-53.

Ding Y, Tao Y, Zhu C., (2013), Emerging roles of microRNAs in the mediation of drought stress response in plants. *J Exp Bot; 64(11):3077-86*.

Dodd A.N., Kudla J., Sanders D., (2010), The language of calcium signaling. *Annual Review of Plant Biology* 61: 593-620.

Eamens A., Wang M.B., Smith N.A., Waterhouse P.M., (2008), RNA silencing in plants: yesterday, today, and tomorrow. *Plant Physiology* 147(2): 456-468.

Eldem V., Çelikkol Akçay U., Ozhuner E., Bakır Y., Uranbey S., Unver T., (2012), Genome-Wide Identification of miRNAs Responsive to Drought in Peach (*Prunus persica*) by High-Throughput Deep Sequencing. *PLoS ONE* 7(12): e50298.

Eldem V., Okay S., Űnver T., (2013), Plant microRNAs: new players in functional genomics. *Turkish Journal of Agriculture and Forestry* 37: 1-21.

Feng H, Huang X, Zhang Q, Wei G, Wang X, Kang Z., (2012), Selection of suitable inner reference genes for relative quantification expression of microRNA in wheat. *Plant Physiol Biochem*; 51:116-22.

Finnegan E. J., Margis R. and Waterhouse P., (2003), Posttranscriptional gene silencing is not compromised in the Arabidopsis CARPEL FACTORY (DICER-LIKE1) mutant, a homolog of Dicer-1 from *Drosophila*. *Current Biology* 13(3): 236-40.

Fire A., Xu S., Montgomery M.K., Kostas S.A., Driver S.E., Mello C.C., (1998), Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391(6669): 806-11.

Francia E., Rizza F., Cattivelli L., Stanca A.M., Galiba G., Tòth B., Hayes P.M., Skinner J.S., Pecchioni N., (2004), Two loci on chromosome 5H determine low-temperature tolerance in a 'Nure' (winter) x 'Tremois' (spring) barley map. *Theoretical and Applied Genetics*, 108(4): 670-680.

Francia E., Barabaschi D., Tondelli A., Laido' G., Rizza F., Stanca A.M., Busconi M., Fogher C., Stockinger E.J., Pecchioni, N., (2007), Fine mapping of a HvCBF gene cluster at the frost resistance locus Fr-H2 in barley. *Theoretical and applied genetics* 115(8): 1083-1091.

Franco-Zorrilla J.M., Valli A., Todesco M., Mateos I., Puga M.I., Rubio-Somoza I., Leyva A., Weigel D., García J.A., Paz-Ares J., (2007), Target mimicry provides a new mechanism for regulation of microRNA activity. *Nature Genetics* 39(8): 1033-7.

Galiba G., Vágújfalvi A., Li C., Soltész A., and Dubcovsky J., (2009), Regulatory genes involved in the determination of frost tolerance in temperate cereals. *Plant Science,* 176(1): 12–19.

Galiveti C.R., Rozhdestvensky T.S., Brosius J., Lehrach H., Konthur Z., (2010), Application of housekeeping npcRNAs for quantitative expression analysis of human transcriptome by real-time PCR. *RNA* 16(2): 450-461.

Gao X, Qiao Y, Han D, Zhang Y, Ma N. (2012**a**) Enemy or partner: relationship between intronic microRNAs and their host genes. *IUBMB Life*; 64(10):835-40

Gao Z., Luo X., Shi T., Cai B., Zhang Z., Cheng Z., Zhuang W., (2012**b**), Identification and validation of potential conserved microRNAs and their targets in peach (*Prunus persica*). *Molecules and cell* 34(3): 239-2349.

Ghildiyal M., Zamore P.D., (2009), Small silencing RNAs: an expanding universe. *Nature Reviews Genetics,* 10(2): 94-108.

Greiner S. and Bock R., (2013), Tuning a ménage à trois: co-evolution and co-adaptation of nuclear and organellar genomes in plants. *BioEssays* 35(4): 354-365.

Hansen K.D., Brenner S.E., and Dudoit S., (2010). Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Research* 38(12): e131.

Hwang E.W., Shin S.J., Yu B.K., Byun M.O., Kwon H.B., (2011), miR171 Family Members are Involved in Drought Response in *Solanum tuberosum*. *Journal of Plant Biology* 54: 43-48.

Jeong DH, Thatcher SR, Brown RS, Zhai J, Park S, Rymarquis LA, Meyers BC, Green PJ, (2013), Comprehensive investigation of microRNAs enhanced by analysis of sequence variants, expression patterns, ARGONAUTE loading, and target cleavage. *Plant Physiol.* 162(3):1225-45.

Jones-Rhoades M.W., Bartel D.P., Bartel B., (2006), MicroRNAs and their regulatory roles in plants. *Annual review of plant biology* 57: 19-53.

Kantar M., Lucas S.J., Budak H., (2011), miRNA expression patterns of *Triticum dicoccoides* in response to shock drought stress. *Planta* 233(3): 471–484.

Khraiwesh B., Zhu J.K., Zhu J., (2011), Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochimica and Biophysica Acta* 1819(2): 137-48.

Kidner C.A., (2010), The many roles of small RNAs in leaf development. *Journal of genetics and genomics,* 37(1): 13-21.

Kihara H., (1951), Substitution of nucleus and its effects on genome manifestations. *Cytologia* 16: 177-193.

Knox A.K., Dhillon T., Cheng H., Tondelli A., Pecchioni N., Stockinger E.J., (2010), CBF gene copy number variation at Frost Resistance-2 is associated with levels of freezing tolerance in temperate-climate cereals. *Theoretical and Applied Genetics* 121(1): 1 21–35.

Kou S.J., Wu X.M., Liu Z., Liu Y.L., Xu Q., Guo W.W., (2012), Selection and validation of suitable reference genes for miRNA expression normalization by quantitative RT-PCR in citrus somatic embryogenic and adult tissues. *Plant Cell Reports* 31(12): 2151-63.

Kulcheski F.R., Marcelino-Guimaraes F.C., Nepomuceno A.L., Abdelnoor R.V., Margis R., (2010), The use of microRNAs as reference genes for quantitative polymerase chain reaction in soybean. *Analytical Biochemistry* 406(2): 185-92.

Lee R.C., Feinbaum R.L., and Ambros V., (1993), The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75(5): 843-854.

Lee R., Feinbaum R., Ambros V., (2004), A short history of a short RNA. *Cell* 116: S89-92.

Leung A.K., and Sharp P.A., (2010). MicroRNA functions in stress responses. *Molecular cell* 40(2): 205-215.

Levitt J., (1980), Response of plants to environmental stresses. Water, Salt and other Stresses. *Academic Press*, NY, 1: 129-186.

Li Y.F., Zheng Y., Jagadeeswaran G., Sunkar R., (2013), Characterization of small RNAs and their target genes in wheat seedlings using sequencing-based approaches. *Plant Science* 203-204: 17-24.

Limin A., Corey A., Hayes P., Fowler D., (2007), Low-temperature acclimation of barley cultivars used as parents in mapping populations: response to photoperiod, vernalization and phenological development. *Planta* 226(1): 139-146.

Lindbo J.A., Silva-Rosales L., Proebsting W.M., Dougherty W.G., (1993), Induction of a Highly Specific Antiviral State in Transgenic Plants: Implications for Regulation of Gene Expression and Virus Resistance. *Plant Cell* 5(12): 1749-1759.

Liu Q., Paroo Z., (2010), Biochemical principles of small RNA pathways. *Annual Review of Biochemistry,* 79: 295-319.

Lowery N.C., Garcia-Ruiz H., Carrington J.C., (2011), The Role of Arabidopsis RNA-Dependent RNA Polymerase Genes 3, 4 and 5 in Antiviral Defense. Oregon State University, 12-Jul-2011.

Luo X., Gao Z., Shi T., Cheng Z., Zhang Z., Ni Z., (2013), Identification of miRNAs and Their Target Genes in Peach (*Prunus persica* L.) Using High-Throughput Sequencing and Degradome Analysis. *PLoS ONE* 8(11): e79090.

Lv S., Nie X., Wang L., Du X., Biradar S.S., Jia X., Weining S., (2012), Identification and Characterization of MicroRNAs from Barley (*Hordeum vulgare* L.) by High-Throughput Sequencing. *International Journal of Molecular* 13(3): 2973-84.

Malone C.D., Hannon G.J., (2009), Small RNAs as guardians of the genome. *Cell* 136(4): 656-668.

Mardis E.R., (2008), Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics* 9: 387-402.

Mardis E.R., (2013), Next-Generation Sequencing Platforms. *Annual Review of Analytical Chemistry* 6: 287-303.

Maré C., Mazzucotelli E., Crosatti, C., Francia E., Stanca A.M. And Cattivelli L., (2004), Hv-WRKY38: a new transcription factor involved in cold- and drought-response in barley. *Plant Molecular Biology* 55(3): 399-416.

Martin A. and Chapman V., (1977), A hybrid between *Hordeum chilense* and *Triticum aestivum*. *Cereal Research Communications* 5: 365-368.

Martin A. and Sanchez-Monge Laguna E., (1982), Cytology and morphology of the amphiploid *Hordeum chilense* x *Triticum turgidum* conv. *durum*. *Euphytica* 31:261- 267.

Matzke M.A. and Matzke A.J.M., (2004), Planting the seeds of a new paradigm. *PLoS Biology* 2(5): 582-586.

Matzke M.A., Birchler J.A., (2005), RNAi-mediated pathways in the nucleus. *Nature Reviews Genetics* 6(1): 24-35.

Matzke M.A., Primig M., Trnovsky J. and Matzke A.J.M., (1989), Reversible methylation and inactivation of marker genes in sequentially transformed tobacco plants. *EMBO Journal* 8(3): 643-649.

McClintock B., (1951), Mutable loci in maize. *Carnegie Institution of Washington,* Yearbook 50: 174-181.

Meister G., (2013), Argonaute proteins: functional insights and emerging roles. *Nature Reviews Genetics* 14, 447-459.

Meng Y. and Shao C., (2012), Large-scale identification of mirtrons in *Arabidopsis* and rice. *PLoS One* 7, e31163.

Metzker M.L., (2010), Sequencing technologies - the next generation. *Nature Reviews Genetics* 11(1): 31-46.

Meyers B.C., Souret F..F, Lu C., Green P.J., (2006), Sweating the small stuff: microRNA discovery in plants. *Current Opinion in Biotechnology* 17(2): 139-146.

Montgomery M.K., (2004), RNA interference: historical overview and significance. *Methods in Molecular Biology* 265: 3-21.

Morin R.D., O'Connor M.D., Griffith M., Kuchenbauer F., Delaney A., Prabhu A.L., Zhao Y., McDonald H., Zeng T., Hirst M., Eaves C.J., Marra M.A., (2008), Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Research* 18(4): 610-621.

Napoli C., Lemieux C., and Jorgensen R.A., (1990), Introduction of a chimeric chalcone synthase gene into *Petunia* results in reversible co-suppression of homologous genes in trans. *Plant Cell* 2(4): 279-289.

Ozhuner E., Eldem V., Ipek A., Okay S., Sakcali S., Zhang B., Boke H., Unver T., (2013), Boron Stress Responsive MicroRNAs and Their Targets in Barley. *PLoS ONE* 8(3): e59543.

Panio G., Motzo R., Mastrangelo A.M., Marone D., Cattivelli L., Giunta F., De Vita P., (2012), Molecular mapping of stomatal-conductance-related traits in durum wheat (*Triticum turgidum ssp. durum*). *Annals of Applied Biology* 162: 258-270.

Plante I., Plé H., Landry P., Gunaratne P.H., Provost P., (2012), Modulation of microRNA Activity by Semi-microRNAs. *Frontiers in Genetics* 3: 99.

Pratt A.J., MacRae I.J., (2009), The RNA-induced silencing complex: a versatile gene-silencing machine. *The Journal of Biological Chemistry* 284(27): 17897-901

Pritchard C.C., Cheng H.H., Tewari M., (2012), MicroRNA profiling: approaches and considerations. *Nature reviews. Genetics* 13(5): 358-69.

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

Ratcliff F., Harrison B.D., Baulcombe D.C., (1997), A Similarity Between Viral Defense and Gene Silencing in Plants. *Science,* 276(5318): 1558-60.

Reinhart B.J., Slack F.J., Basson M., Pasquinelli A.E., Bettinger J.C., Rougvie A.E., Horvitz H.R., Ruvkun G., (2000), The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 403(6772): 901-6.

Rizza F., Ghashghaie J., Meyer S., Matteu L., Mastrangelo A.M. & Badeck F.-W., (2012), Constitutive differences in water use efficiency between two durum wheat cultivars. *Field Crops Research* 49-60.

Romano N., Macino G., (1992), Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. *Molecular Microbiology* 6(22): 3343-3353.

Rowley E.R. and Mockler T.D., (2011), Plant Abiotic Stress: Insights from the Genomics Era, Abiotic Stress Response in Plants - Physiological, Biochemical and Genetic Perspectives, Prof. Arun Shanker (Ed.), ISBN: 978-953-307-672-0, InTech, DOI: 10.5772/23215.

Rubio-Somoza I., Weigel D., (2011), MicroRNA networks and developmental plasticity in plants. *Trends in Plant Science* 16(5): 258-64.

Schreiber A.W., Shi B.J., Huang C.Y., Langridge P., Baumann U., (2011), Discovery of barley miRNAs through deep sequencing of short reads. *BMC Genomics*, 12: 129.

Skinner J.S., von Zitzewitz J., Szűcs P., Marquez-Cedillo L., Filichkin T., Amundsen K., Stockinger E.J., Thomashow M.F., Chen T.H.H., Hayes P.M., (2005), Structural, functional, and phylogenetic characterization of a large CBF gene family in barley. *Plant Molecular Biology* 59(4): 533–551.

Stockinger E.J., Skinner J.S., Gardner, K.G., Francia E. and Pecchioni N., (2007), Expression levels of barley Cbf genes at the Frost resistance-H2 locus are dependent upon alleles at Fr-H1 and Fr-H2. *The Plant Journal* 51(2): 308–321.

Sunkar R., Li Y.F., Jagadeeswaran G., (2012), Functions of microRNAs in plant stress responses. *Trends in plant science* 17(4): 196-203.

Tabara H., Sarkissian M., Kelly W.G., Fleenor J., Grishok A., Timmons L., Fire A., Mello C.C., (1999), The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* 99(2): 123-32.

Tardif G., Kane N.A., Adam H., Labriel L., Major G., Gulick P., Sarhan F., Laliberté J.F., (2007), Interaction network of proteins associated with abiotic stress response and development in wheat. *Plant Molecular Biology* 63(5): 703-708.

Tondelli A., Francia E., Barabaschi D., Pasquariello M., Pecchioni N., (2011), Inside the CBF locus in *Poaceae*. *Plant Science* 180(1): 39-45.

Tondelli A., Francia, E., Barabaschi D., Aprile A., Skinner J.S., Stockinger E.J., Stanca, A.M.; Pecchioni, N., (2006), Mapping regulatory genes as candidates for cold and drought stress tolerance in barley. *Theoretical and Applied Genetics* 112(3): 445-454.

Van der Krol A. R., Mur L. A., Beld, M., Mol J. N. M. and Stuitje A. R., (1990), Flavonoid genes in petunia: addition of a limited number of genes copies may lead to a suppression of gene expression. *Plant Cell* 2(4): 291-299.

Varkonyi-Gasic E., Wu R., Wood M., Walton E.F., Hellens R.P., (2007), Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. *Plant Methods* 3: 12.

Vaucheret H., (2008), Plant ARGONAUTES. *Trends in Plant science,* 13(7): 350-8.

Vaucheret H., (2009), AGO1 Homeostasis Involves Differential Production of 21-nt and 22-nt *miR168* Species by *MIR168a* and *MIR168b*. *PLoS ONE* 4(7): e6442.

Verde I., Abbott A.G., Scalabrin S., Jung S., Shu S., et al., (2013), The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature Genetics* 45(5): 487-494.

Voinnet O., (2008), Post-transcriptional RNA silencing in plant-microbe interactions: a touch of robustness and versatility. *Current Opinion in Plant Biology,* 11(4): 464-70.

Voinnet O., (2009), Origin, Biogenesis, and Activity of Plant MicroRNAs. *Cell* 136(4), 669-687.

Wang X.J., Reyes J.L., Chua N.H., Gaasterland T., (2004), Prediction and identification of Arabidopsis thaliana microRNAs and their mRNA targets. *Genome Biology* 5(9): R65.

Wei K.F., Wu L.J., Chen J., Chen Y.F., Xie D.X., (2012), Structural evolution and functional diversification analyses of argonaute protein. *Journal of Cellular Biochemistry* 113(8): 2576-85.

Wightman B., Ha I. and Ruvkun G.,(1993), Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* 75(5): 855-862.

Wu L., Liu D., Wu J., Zhang R., Qin Z., Liu D., Li A., Fu D., Zhai W., Mao L., (2013), Regulation of FLOWERING LOCUS T by a MicroRNA in *Brachypodium distachyon*. *Plant cell* 25(11): 4363-77.

Xin M.M., Wang Y., Yao Y.Y., Xie C.J., Peng H.R., Ni Z.F., Sun Q.X., (2010), Diverse set of microRNAs are responsive to powdery mildew infection and heat stress in wheat (*Triticum aestivum L.*). *BMC Plant Biology* 10: 123.

Yao Y., Sun Q., (2012), Exploration of small non coding RNAs in wheat (*Triticum aestivum* L.). *Plant molecular biology* 80(1): 67-73.

Yin Z., Li Y., Yu J., Liu Y., Li C., Han X., Shen F., (2012), Difference in miRNA expression profiles between two cotton cultivars with distinct salt sensitivity. *Molecular biology reports* 39(4): 4961-70.

Yordanov I., Velikova V. and Tsonev T., (2003), Plant responses to drought and stress tolerance. *Bulgarian Journal of Plant Physiology* Special issue 187-206.

Zhang, Y. (2005). miRU: An automated plant miRNA target prediction server. *Nucleic Acids Res*. 33 (Suppl. 2), W701–W704.

Zhang S., Zhou J., Han S., Yang W., Li W., Wei H., Li X., Qi L., (2010), Four abiotic stress-induced miRNA families differentially regulated in the embryogenic and non-embryogenic callus tissues of *Larix leptolepis*. *Biochemical and Biophysical Research Communications* 398(3): 355-60.

Zhang Z.L., Ogawa M., Fleet C.M., Zentella R., Hu J., Heo J.O., Lim J., Kamiya Y., Yamaguchi S., Sun T.P., (2011), Scarecrow-like 3 promotes gibberellin signaling by antagonizing master growth repressor DELLA in Arabidopsis. *PNAS* 108(5): 2160-5.

Zhang Y., Yu M., Yu H., Han J., Song C., Ma R., Fang J., (2012**a**), Computational identification of microRNAs in peach expressed sequence tags and validation of their precise sequences by miR-RACE. *Molecular biology reports* 39(2): 1975-87.

Zhang T., Fang Y., Wang X., Deng X., Zhang X., et al., (2012**b**), The Complete Chloroplast and Mitochondrial Genome Sequences of *Boea hygrometrica*: Insights into the Evolution of Plant Organellar Genomes. *PLoS ONE* 7(1): e30531.

Zhao Y.T., Wang M., Fu S.X., Yang W.C., Qi C.K., Wang X.J., (2012), Small RNA profiling in two *Brassica napus* cultivars identifies microRNAs with oil production- and development-correlated expression and new small RNA classes. *Plant physiology* 158(2): 813-23.

ZhenLin W., (2013), Identification and characterization of microRNAs and their targets in peach (*Prunus persica*). *International Journal of Agriculture and Biology* 15(5): 1017-1020.

Zhu H., Xia R., Zhao B., An Y.Q., Dardick C.D., Callahan A.M., Liu Z., (2012), Unique expression, processing regulation, and regulatory network of peach (*Prunus persica*) miRNAs. *BMC Plant Biology* 12: 149.

Annexes

List of other publications:

- **Mazzucotelli E, Crosatti C, Giusti L, Guerra D, Cattivelli L** (2013). Post-transcriptional and post-translational modifications controlling cold response. In: "Plant and Microbe Adaptations to Cold in a Changing World: Proceedings of the Plant and Microbe Adaptation to Cold Conference, 2012". Edited by Ryozo Imai, Midori Yoshida, and Naoyuki Matsumoto. Published by Springer, New York, USA
- **Crosatti C., Quansah L., Marè C., Giusti L., Roncaglia E., Atienza S. G., Cattivelli L. and Fait A.** (2013) Cytoplasmic genome substitution in wheat affects the nuclear-cytoplasmic cross-talk leading to transcript and metabolite alterations. *BMC Genomics* 14:868
- **Colaiacovo, M., Crosatti, C., Giusti, L., Alberici, R., Cattivelli, L., & Faccioli, P**. (2013). Characterization of the Barley (*Hordeum vulgare* L.) miRNome: A Computational-Based Update on MicroRNAs and Their Targets. In "Advance in Barley Sciences" (pp. 427-431). Springer Netherlands.

Post-transcriptional and Post-translational Modifications Controlling Cold Response

Elisabetta Mazzucotelli, Cristina Crosatti, Lorenzo Giusti, Davide Guerra and Luigi Cattivelli

Introduction

Plants become tolerant to environmental stresses by reprogramming metabolism and gene expression, to establish a new equilibrium between growth, development, and survival. In the past two decades, important advances have been made in the understanding of transcriptional changes induced by exposition to low temperature and in the identification of signaling proteins and transcription factors which regulate the cold-induced gene expression. Outcomes depicted a complex process where several pathways connecting stress perception and specific transcriptional changes play a main role. Typical and famous components of the cold response are the C-repeat (CRT) Binding Factors (CBFs), universally recognized as key regulators of response to low temperature, ICE1 (Inducer of CBF Expression 1) and the downstream *COR* (COld-Regulated) genes (Huang et al. 2012).

The characterization of new mutants as well as the integration of gene expression data and protein activities indicates that further levels of regulation based on post-transcriptional and post-translational mechanisms are responsible for the transcriptional changes related to the plant response to cold and other stresses. This chapter focuses on the most recent findings about these mechanisms—alternative splicing, RNA-mediated silencing, ubiquitination, and sumoylation—activated by plants after the perception of a cold stress. The network of such mechanisms targets transcription factors and other regulatory components of the stress signaling, resulting in either activation or repression of their activities (Fig. 1, Mazzucotelli et al. 2008). This ensures temporally and spatially appropriate patterns of downstream gene expression, thus the shaping of proteome and metabolome complexity, to ultimately switch on adaptive response.

E. Mazzucotelli (✉) · C. Crosatti · L. Giusti · D. Guerra · L. Cattivelli
Genomics Research Centre, Consiglio per la ricerca e la sperimentazione in agricoltura (CRA), Fiorenzuola d'Arda (PC), Italy
e-mail: elisabetta.mazzucotelli@entecra.it

R. Imai et al. (eds.), *Plant and Microbe Adaptations to Cold in a Changing World,*
DOI 10.1007/978-1-4614-8253-6_10,

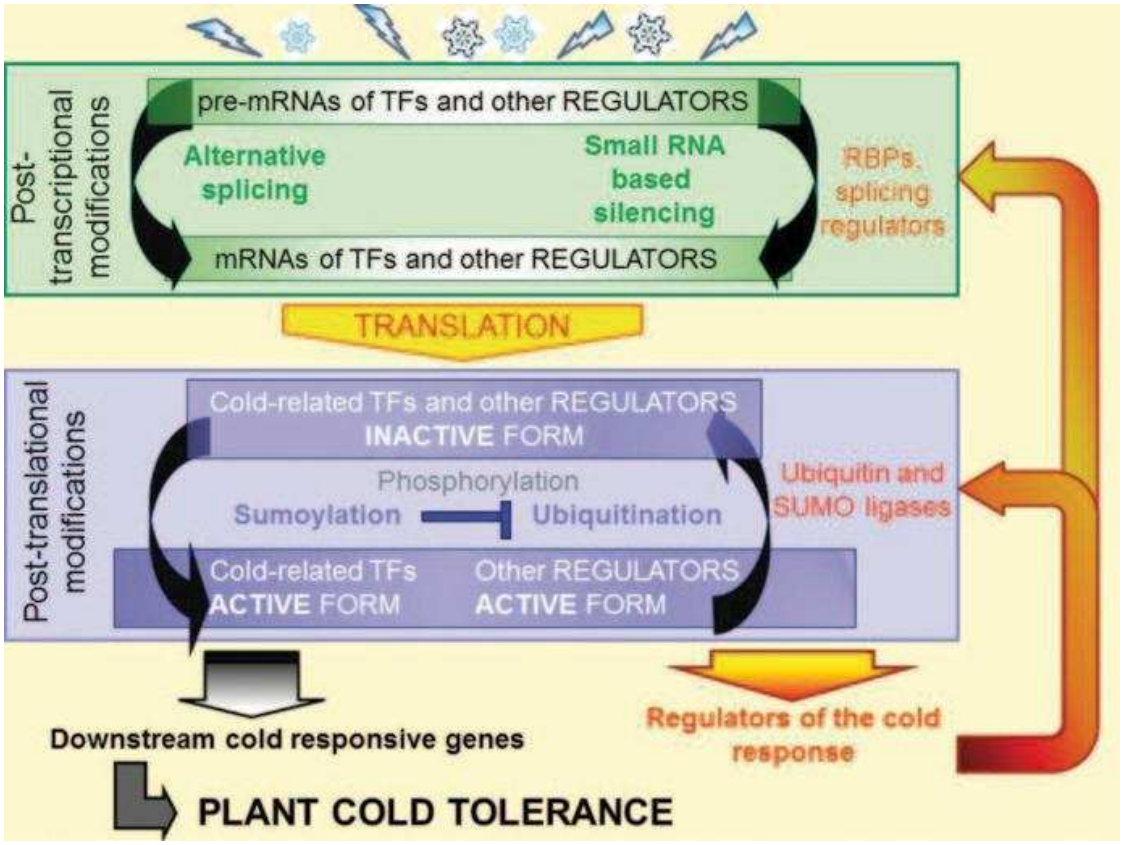


Fig. 1 Model describing the cross-talking among post-transcritpional and post-translational regulations involved in the control of the plant response to cold stress. See the text for details. (Modified from Mazzucotelli et al. 2008)

Post-transcriptional Processes Affecting mRNA Availability

Different mechanisms regulate synthesis, maturation, and stability of transcript, ranging from alternative splicing to association with polysomes, from the formation of RNA granules and mechanisms of RNA degradation, guided by noncoding RNAs and nonsense-mediated decay (NMD). All these steps are emerging as important point in the regulation of the plant stress responses.

RNA-Binding Proteins (RBP) are at the heart of all these post-transcriptional processes. They bind RNA molecules immediately after the transcription till the translation, in order to protect, organize, and prepare mRNAs for posttranscriptional processes. More than 200 RBP genes have been predicted in *Arabidopsis* and rice genomes (Ambrosone et al. 2012). RBPs contain different kinds of RNA-binding and auxiliary domains, which enable them to bind RNA in a versatile way, such as RNA recognition motif, glycine-rich motif, zinc finger motif, cold shock domain (CSD), and double-stranded RNA-binding motif. RBPs have a role in normal cellular functions particularly in regulating several developmental processes. Further, a large body of evidence is supporting their key function in plant adaptation to various environmental conditions when they regulate the stress-dependent export of mRNAs from the nucleus, selective translation of stress-associated genes, and

increased stability of related transcripts (Ambrosone et al. 2012). RBPs execute this activity functioning as RNA chaperones (Kang et al. 2012). Indeed, when cells are exposed to low temperatures, misfolded RNA molecules become overstabilized and cannot assume a native conformation without the help of RNA chaperones. In *Arabidopsis*, a positive effect on cold and freezing tolerance associated with RNA chaperone activity has been shown for the Glycine-Rich Proteins GRP7 (Kim et al. 2008), GRP2 (Kim et al. 2007), and AtRZ-1a, b, c (Kim and Kang 2006), and for the CSD proteins CSP2 (Sasaki et al. 2007) and CSP3 (Kim et al. 2009). GRP7 acts in the export of mRNAs from the nucleus to the cytoplasm under cold stress conditions in the guard cells. GRP2 modulates the expression and activity of mitochondrial-encoded genes by exhibiting transcription antitermination activity. AtRZ-1a affects the translation of several genes involved in stress resistance and in RNA and protein metabolism.

Also RNA-helicases belong to the class of RBPs. The *Arabidopsis* gene *Los4* (Low expression of osmotically responsive genes 4) encodes an RNA-helicase acting on nuclear mRNA export, with particular effects in the response to temperature stress (Gong et al. 2005). The *los4-1* mutant has a reduced expression of *CBF3* and a delayed expression of *CBF1* and *CBF2* during cold acclimation resulting in chilling sensitivity, while cryophyte/los4-2 mutant (allelic to los4-1) promotes a superinduction of *CBF2* under cold stress and an enhanced freezing tolerance. Both mutants are heat sensitive. This divergent response is mediated through a differential effect on nuclear mRNA export: inhibited by los4-1 and enhanced by los4-2 at low temperatures, while both mutations disrupt the mRNA export at high temperatures.

Stress-Related Transcripts from Alternative Splicing Events

Alternative splicing (AS), a mechanism by which different RNA transcripts from a single gene are produced, occurs in up to 42 % of *Arabidopsis* genes containing introns. Alternative mRNA forms have different fates according to whether the protein still maintain an activity or not, which usually depends on the kind of splicing which takes place. Nonfunctional transcripts incur in degradation by NMD, which is both a surveillance mechanism avoiding accumulation of incomplete, potentially harmful proteins, and a fine modulator of gene expression (Mastrangelo et al. 2011).

A large number of gene networks related to abiotic stresses, involving, among others, protein kinases, transcription factors, splicing regulators have demonstrated the centrality of AS in the fine-tuning of plant responses to the environment (Mastrangelo et al. 2011). For example, the *Arabidopsis STA1* gene encodes a splicing regulator upregulated by cold. STA1 protein can regulate the stability and splicing pattern of a number of endogenous gene transcripts related to abiotic stress response (Lee et al. 2006).

AS can act to control the abundance of the active protein. For example, AS of the wheat transcription factor gene *Wdreb2* generates three transcripts in response

to cold, only one of which still maintains its transcriptional activity (Egawa et al. 2006). Alternatively, AS can represent an interconnecting mechanism between the response to stress and other cellular processes. A very interesting example is provided by the interaction between circadian clock and cold tolerance, in which alternative splicing bridges the two pathways. CCA1 (CIRCADIAN CLOCK ASSOCIATED 1) and LHY (LATE ELONGATED HYPOCOTYL) are MYB transcription factors belonging to *Arabidopsis* circadian clock. Among other functions, they positively affect the expression of *CBF* pathway genes (Dong et al. 2011). *CCA1* transcript can be alternatively spliced into two variants: *CCA1α*, the fully spliced one, and *CCA1β*, which retains the fourth intron. The latter encodes a protein lacking a DNA-binding domain, but still able to form nonfunctional homodimers and heterodimers with CCA1α and LHY, suggesting a CCA1 self-regulation model in which CCA1β variant competitively inhibits CCA1α and LHY activity by sequestering them. Intriguingly, *CCA1* splicing is regulated by temperature, and the fully spliced *CCA1α* variant becomes dominant during cold stress, therefore promoting the expression of the *CBF* pathway genes (Seo et al. 2012). Interestingly, cold stress is responsible also for the alternative splicing of LHY, in this case leading to the downregulation of the functional transcript (James et al. 2012).

The alternative splicing of *IDD14* (INDETERMINATE DOMAIN14) generates a self-controlled regulatory loop that modulates starch accumulation in response to cold (Seo et al. 2011). *IDD14* binds to Qua-Quine Starch (QQS) promoter, which regulates starch accumulation. Under cold conditions, *IDD14* is alternatively spliced producing a nonfunctional form (IDD14β) lacking a DNA-binding domain, but able to form heterodimers with the fully spliced variant (IDD14α), thus preventing its binding to QQS.

Besides transcription factors, all levels of regulation of gene expression can be affected by AS in response to abiotic stresses. Splicing events are being described for genes coding for E3 ubiquitin ligases. For instance, the transcripts of the E3 ligase 6g2, later called TdRf1, undergoes cold-dependent alternative splicing (Mastrangelo et al. 2005). Similarly, the gene *HOS1* (High expression of Osmotically responsive gene1), encoding an E3 ligase with a negative regulative function on the cold response, produces two differential transcripts differentially regulated in both temporal, spatial manners, and at different temperatures (Lee et al. 2012).

Many splicing regulators, like *SR* (ser/arg) genes (Palusa et al. 2007) promote alternative splicing of their own transcript as well as of other gene products. An example is provided by the *Arabidopsis* genes encoding the RNA-binding proteins AtGRP8 and AtGRP7 (Schöning et al. 2008). During cold stress, AtGRP8 promotes the use of a 5' splice site of its own transcript, producing an inactive form (*as_AtGRP8*) that is immediately targeted by NMD machinery. At the same time, AtGRP8 promotes the alternative splicing of *AtGRP7* transcript, which leads to NMD of *as_AtGRP7*, and then to a decrease of AtGRP7 abundance. Furthermore, also AtGRP7 does the same with its own and *AtGRP8* mRNAs.

Degradation of Stress-Related Transcripts by miRNAs and tasiRNAs

The discovery of microRNAs (miRNAs) and tasiRNA (trans acting small interfering RNA) revealed their fundamental role in the regulatory network of the plant response to cold (Sunkar et al. 2012). The small noncoding RNAs post-transcriptionally silence target genes either by guiding degradation or repressing translation of target mRNAs.

Many plant cold-stress-regulated miRNAs have been identified by cloning, bioinformatic and high-throughput sequencing approaches in *Arabidopsis* (Zhou et al. 2008), *Populus* (Chen et al. 2012), *Brachypodium* (Zhang et al. 2009), rice (Lv et al. 2010). miRNAs are conserved among dicotyledonous and monocotyledonous plants, nevertheless, differences in their regulation suggest a significant level of species specificity. Comparative studies of miRNAs in *Brachypodium* and rice highlighted that the composition and location of miRNA families are different even in closely related plant species (Zhang et al. 2009). Even in sugarcane, the miR319 was shown to be differentially expressed among sugarcane genotypes contrasting for cold tolerance (Thiebaut et al. 2012).

Despite identification of many cold-related miRNAs, little is known about their target genes and their role in the response to cold. In general, miRNA targets are mainly represented by transcription factors, regulators of auxin signaling, RNA silencing, but also enzymes (Khraiwesh et al. 2012).

A cold-responsive miRNAs can act in different signaling pathways and simultaneously regulate genes with diverse functions linking together several pathways. For example, several cold-responsive miRNAs are also involved in the response of plant to biotic stress suggesting a cross-talk between pathways for biotic and abiotic stress responses (Lu et al. 2008).

Auxin signaling is crucial for plant growth and development, and several studies suggested that auxin is also involved in plant stress tolerance (Achard et al. 2006). The *Arabidopsis* miR393, upregulated in response to most stress conditions (including cold stress), is putatively directed against TIR1 and other closely related F-box proteins (Sunkar et al. 2004). Because TIR1 is the auxin receptor that targets repressors of the Auxin-Responsive Factor (ARF) for ubiquitin-mediated degradation in response to auxin, the miRNA inhibition of *TIR1* would downregulate auxin signaling and seedling growth.

Tang et al. (2012) sequenced a set of both miRNA and tasiRNAs from spike tissues of a wheat thermosensitive genic male-sterile line exposed to cold. tasiRNAs are noncoding RNA generated by miRNA processing of a *Tas* gene transcript resulting in the production of 21-nt RNAs phased with respect to the miRNA cleavage site. They found that miR167 and tasiRNA-ARF play roles in regulating the auxin-signaling pathway and possibly in the developmental response to cold stress.

Protein Degradation in Response to Stress: The Role of Ubiquitination

Ubiquitination is a unique, widely conserved molecular mechanism to degrade proteins. The covalent attachment of ubiquitin to a target substrate involves the sequential action of three enzymes: ubiquitin-activating enzyme (E1), ubiquitin-conjugating enzyme (E2), and ubiquitin ligase (E3). Usually once a target protein has been labeled with a polyubiquitin chain (at least four ubiquitin), it is recognized and then degraded by the 26 S proteasome. The E3 ubiquitin ligases are the enzymes which specifically recruit the target proteins ensuring the specificity of the entire process. More than 1,400 E3 enzymes are coded by *Arabidopsis* genome (Mazzucotelli et al. 2006). E3-mediated protein degradation regulates almost all cellular processes, including abiotic stress response (Lyzenga and Stone 2012).

One of the first E3 ligases identified in stress response is HOS1 which exerts a negative control on cold response being responsible for ICE1 degradation (Dong et al. 2006). Indeed, *hos1* mutation enhances the induction of *CBFs* and of the downstream cold-regulated genes by low temperatures.

Further studies identified other E3 ligases involved in low temperature response. In *Arabidopsis*, the overexpression of the U-Box E3 ubiquitin ligase AtCHIP (Carboxyl terminus of HSC70-Interacting Protein) causes higher sensitivity to low temperatures than wild type and also a growth-impaired phenotype suggesting its role as negative regulator of low temperature response (Yan et al. 2003).

Often E3 ligases represent a connecting point between different signaling pathways. The RING-finger E3 ligase TdRF1 (*Triticum durum* RING-finger protein 1) represents an interesting example (Guerra et al. 2012). *TdRF1* is induced upon exposure to low temperatures and dehydration. TdRF1 is phosphorylated by the kinase TdWNK5 (With No Lysine [K]5) a member of the *Arabidopsis* WNK family of MAP kinases involved in flowering time and circadian clock regulation (Wang et al. 2008). TdRF1 interacts with another E3 ligase, WVIP2 (Wheat Viviparus1 Interacting Protein2) strongly upregulated upon cold treatment and sharing high amino acid similarity with the wild oat VIP2 (Jones et al. 2000). Finally, it degrades the transcription factor WBLH1 (Wheat Bel1-Type Homeodomain1), a previously described protein belonging to KNOX (Knotted1-like homeobox) gene family (Mizumoto et al. 2011). The overexpression of *TdRF1* increases tolerance of barley cells to dehydration, suggesting that it could protect plant under drought and freezing conditions.

Control of Stress Response by Sumoylation

Sumoylation is a post-translational transient modification of protein substrates based on the covalent conjugation of the SUMO (Small Ubiquitin-like Modifier) peptide. The biochemical steps are similar to those operating in the ubiquitination pathway, but sumoylation is specifically regulated at the target level, with phosphorylation

accomplishing a critical role, besides the target consensus motif CKxE/D (C: hydrophobic amino acid; K: SUMO target lysine; x: any amino acid; D/E acidic amino acids), while the dynamic aspects are regulated by SUMO peptidase activities (Miura and Hasegawa 2010).

A wide variety of biological consequences of sumoylation have been observed, including changes in enzymatic activity, subcellular relocalization, and protection from ubiquitin-mediated degradation of regulatory proteins. Both loss and gain of function analyses, as well as the pattern of SUMO-conjugates revealed a key role of sumoylation in plants in response to environmental signals. *Arabidopsis siz1* (for SAP and Miz domain) mutants are hypersensitive to chilling and freezing stresses (Miura et al. 2007). Sumoylation is essential for freezing tolerance through the stabilization of ICE1, the inducer of *CBF* and repressor of *MYB15* expression. This modification blocks the ubiquitin-mediated degradation of ICE1 allowing it to activate *CBF* transcription. The sumoylated isoform of ICE1 also has a negative effect on the transcription of *MYB15*, which functions as repressor of *CBF* genes. The final effect of the AtSIZ1-mediated sumoylation is therefore the attenuation of a repressor system that in normal growing conditions blocks part of the transcriptional response to cold.

A general accumulation of SUMO conjugates is an early effect of the exposure to extreme temperatures, and other stress, thus many proteins are target of sumoylation upon exposure to stress (Miura et al. 2007). Studies on SUMO proteome list hundreds of potentially sumoylated proteins (Budhiraja et al. 2009; Elrouby and Coupland 2010; Miller et al. 2010; Park et al. 2011). The list is enriched in transcription factors and RNA-related factors regulating processing, splicing, export, and translation of mRNA. Thus many regulators of the plant cold response could represent the SUMO conjugates which accumulate under stress conditions. One example is STABILIZED1 (STA1), a splicing factor of the cold-induced gene *COR15 A* also involved in mRNA turnover. Its expression is induced by cold stress and the protein is sumoylated (Lee et al. 2006; Meier et al. 2012).

A Complex Network of Regulations

The overall regulations interacting at several levels represent a complex and dynamic network (Fig. 1). Factors responsible of post-transcriptional and post-translatrional mechanisms are often controlled at transcriptional level (i.e., many E3 ubiquitin ligases are stress induced), subjected to gene silencing by action of miRNA (Sunkar et al. 2012) or to alternative splicing (Mastrangelo et al. 2011). They might be modified by SUMO (many RNPs, RNA helicases, and components of the RNA export are substrates of sumoylation; Meier 2012), ubiquitin (ubiquitin conjugation controls transcription to mRNP assembly down to export through the nuclear pore complex; Babour 2012). Multiple signaling pathways may converge on the same target protein by multisite modifications, resulting in complex combinatorial regulatory patterns that dynamically and reversibly affect the activity of the target protein. Different post-translational mechanisms may act synergically or have antagonistic

effects. The final output is a new shape of transcriptome, proteome, and metabolome which constitute the plant response to the stress.

The regulation of the activity of the transcription factor ICE1 offers a well-characterized example of the complexity of these regulatory systems. ICE1 is constitutively expressed, nevertheless, it activates the expression of CBF genes only upon cold treatment (Chinnusamy et al. 2012). Three different modifications are known, so far, to control the activity of ICE1 protein, and thus the expression of *CBFs*. At low temperature ICE1 is sumoylated through the action of AtSIZ1 (Miura et al. 2007), resulting in a fully active transcription factor. Alternatively, HOS1 ubiquitinates ICE1 leading to its proteasomal degradation (Dong et al. 2006). ICE1 may be more or less available for ubiquitination and sumoylation depending on its phosphorylation status, which is most likely temperature dependent (Zhu et al. 2007). The balance between activation and degradation allows a perfect tuning of ICE1 activity which in turn leads to the activation of the cold-induced molecular response. The function of HOS1, which is also subjected to alternative splicing, is activated by nuclear localization mediated by an unknown post-translational modification upon cold signal, in order to attenuate the stress response (Lee et al. 2012).

Epigenetic changes have been implicated in the acclimation process (Santos et al. 2011). In barley histone genes constitute an overrepresented functional category in the cold transcriptome of crown, the key organ for cereal overwintering (Janska et al. 2011), among them specific histone variants acting as temperature sensors in *Arabidopsis* (Kumar and Wigge 2010). Nontranscriptional regulations are implicated in epigenetic processes. Indeed, the final effect of RNA-mediated gene silencing is often the methylation of the genomic region producing the target RNA. Ubiquitination and sumoylation, beyond acetylation and methylation, act on nucleosome core histones and sumoylation regulates the activity of the chromatin-remodeling complexes. All together these modifications constitute a histone code which activates or silences gene expression by modifying chromatin structure and constituting a memory of the stress.

New Targets for Engineering Stress-Tolerant Plants?

A new generation of transgenic plants with improved performance under challenging environments could be developed using the increased knowledge on the network of post-transcriptional and post-translational regulations. Plants evolved this network to strictly regulate and perfectly fine-tune the molecular responses to abiotic stresses to achieve stress tolerance avoiding strongly physiological alteration and futile metabolic costs. Current transgenic strategies based on a rough manipulation of regulatory factors often produced plants with some increase of stress tolerance at the expense of growth and production. Indeed, many signaling pathways cross-talk with other developmental and metabolic pathways, thus increasing the risk of inadvertently affecting other routes producing yield penalties or developmental constraints. Future aims will be the development of plants with a finer and more specific regulation of upstream regulators of the stress response.

Acknowledgments This research was supported by Progetto Strategico MIPAF "OLEA – Genomica e Miglioramento genetico dell'olivo," D.M. 27011/7643/10.

References

Achard P, Cheng H, De Grauwe L, Decat J, Schoutteten H, Moritz T, Van Der Straeten D, Peng J, Harberd NP (2006) Integration of plant responses to environmentally activated phytohormonal signals. Science 311:91–94

Ambrosone A, Costa A, Leone A, Grillo S (2012) Beyond transcription: RNA-binding proteins as emerging regulators of plant response to environmental constraints. Plant Sci 182:12–8

Babour A, Dargemont C, Stutz F (2012) Ubiquitin and assembly of export competent mRNP. Biochim et Biophys Acta 1819:521–530

Budhiraja R, Hermkes R, Muller S, Schmidt J, Colby T, Panigrahi K, Coupland G, Bachmair A (2009) Substrates related to chromatin and to RNA-dependent processes are modified by *Arabidopsis* SUMO isoforms that differ in a conserved residue with influence on desumoylation. Plant Physiol 149:1529–1540

Chen L, Zhang Y, Ren Y, Xu J, Zhang Z, Wang Y (2012) Genome-wide identification of cold-responsive and new microRNAs in *Populus tomentosa* by high-throughput sequencing. Biochem Biophys Res Commun 417:892–896

Chinnusamy V, Ohta M, Kanrar S, Lee BH, Hong X, Agarwal M, Zhu JK (2003) ICE1: a regulator of cold-induced transcriptome and freezing tolerance in *Arabidopsis*. Genes Dev 17:1043–1054

Dong CH, Agarwal M, Zhang Y, Xie Q, Zhu JK (2006) The negative regulator of plant cold responses, HOS1, is a RING E3 ligase that mediates the ubiquitination and degradation of ICE1. Proc Natl Acad Sci U S A 103:8281–8286

Dong MA, Farre´ E, Thomashow MF (2011) Circadian clock-associated 1 and late elongated hypocotyl regulate expression of the C-repeat binding factor (CBF) pathway in *Arabidopsis*. Proc Natl Acad Sci U S A 108:7241–7246

Egawa C, Kobayashi F, Ishibashi M, Nakamura T, Nakamura C, Takumi S (2006) Differential regulation of transcript accumulation and alternative splicing of a DREB2 homolog under abiotic stress conditions in common wheat. Genes Genet Syst 81:77–91

Elrouby N, Coupland G (2010) Proteome-wide screens for small ubiquitin-like modifier (SUMO) substrates identify *Arabidopsis* proteins implicated in diverse biological processes. Proc Natl Acad Sci U S A 107:17415–17420

Gong Z, Dong CH, Lee H, Zhu J, Xiong L, Gong D, Stevenson B, Zhu JK (2005) A DEAD box RNA helicase is essential for mRNA export and important for development and stress responses in *Arabidopsis*. Plant Cell 17:256–267

Guerra D, Mastrangelo AM, Lopez-Torrejon G, Marzin S, Schweizer P, Stanca AM, del Pozo JC, Cattivelli L, Mazzucotelli E (2012) Identification of a protein network interacting with TdRF1, a wheat RING ubiquitin ligase with a protective role against cellular dehydration. Plant Physiol 158:777–89

Huang GT, Ma SL, Bai LP, Zhang L, Ma H, Jia P, Liu J, Zhong M, Guo ZF (2012) Signal transduction during cold, salt, and drought stresses in plants. Mol Biol Rep 39:969–87

James AB, Syed NH, Bordage S, Marshall J, Nimmo GA, Jenkins GI, Herzyk P, Brown JWS, Nimmo HG (2012) Alternative splicing mediates responses of the *Arabidopsis* circadian clock to temperature changes. Plant Cell 24:961–981

Janská A, Aprile A, Zámečnàk J, Cattivelli L, Ovesná J (2011) Transcriptional responses of winter barley to cold indicate nucleosome remodelling as a specific feature of crown tissues. Funct Integr Genomics 11:307–325

Jones HD, Kurup S, Peters NC, Holdsworth MJ (2000) Identification and analysis of proteins that interact with the Avena fatua homologue of the maize transcription factor VIVIPAROUS 1. Plant J 21:133–142

Kang H, Park SJ, Kwak KJ (2012) Plant RNA chaperones in stress response. Trends Plant Sci 4:130 (http://dx.doi.org/10.1016/j.tplants. 2012. 08.004)
Khraiwesh B, Zhu JH, Zhu J (2012) Role of miRNAs and siRNA in biotic and biotic and abiotic stress responses of plants. Biochim Biophys Acta 1819:137–148
Kim YO, Kang H (2006) The role of a zinc finger-containing glycine-rich RNA-binding protein during the cold adaptation process in *Arabidopsis thaliana*. Plant Cell Physiol 47:793–798
Kim JY, Park SJ, Jang B, Jung CH, Ahn SJ, Goh CH, Cho K, Han O, Kang H (2007) Functional characterization of a glycine-rich RNA-binding protein 2 in *Arabidopsis* thaliana under abiotic stress conditions. Plant J 50:439–451
Kim JS, Jung HJ, Lee HJ, Kim KA, Goh C-H, Woo Y, SH Oh, Han YS, Kang H (2008) Glycine-rich RNA-binding protein 7 affects abiotic stress responses by regulating stomata opening and closing in *Arabidopsis* thaliana. Plant J 55:455–466
Kim MH, Sasaki K, Imai R (2009) Cold shock domain protein 3 regulates freezing tolerance in Arabidopsis thaliana. J Biol Chem 284:23454–23460
Kumar SV, Wigge PA (2010) H2 A.Z-containing nucleosomes mediate the thermosensory response in *Arabidopsis*. Cell 140:136–147
Lee BH, Kapoor A, Zhu J, Zhu JK (2006) STABILIZED1, a stress-upregulated nuclear protein, is required for pre-mRNA splicing, mRNA turnover, and stress tolerance in *Arabidopsis*. Plant Cell 18:1736–1749
Lee JH, Kim SH, Kim JJ, Ahn JH (2012) Alternative splicing and expression analysis of high expression of osmotically responsive genes1 (HOS1) in Arabidopsis. BMB Rep 45:515–520
Lu S, Sun YH, Chiang VL (2008) Stress-responsive microRNAs in *Populus*. The Plant J 55:131–151
Lv DK, Bai X, Li Y, Ding XD, Ge Y, Cai H, Ji W, Wu N, Zhu YM (2010) Profiling of cold-stress-responsive miRNAs in rice by microarrays. Gene 459:39–47
Lyzenga WJ, Stone SL (2012) Abiotic stress tolerance mediated by protein ubiquitination. J Exp Bot 63:599–616
Mastrangelo AM, Belloni S, Barilli S, Ruperti B, Di Fonzo N, Stanca AM, Cattivelli L (2005) Low temperature promotes intron retention in two e-cor genes of durum wheat. Planta 221:705–715
Mastrangelo AM, Marone D, Laidò G, De Leonardis AM, De Vita P (2011) Alternative splicing: enhancing ability to cope with stress via transcriptome plasticity. Plant Sci 185–186:40–49
Mazzucotelli E, Belloni S, Marone D, De Leonardis AM, Guerra D, Di Fonzo N, Cattivelli L, Mastrangelo AM (2006) The E3 ubiquitin ligase gene family in plants: regulation by degradation. Curr Genomics 7:509–522
Mazzucotelli E, Mastrangelo AM, Crosatti C, Guerra D, Stanca AM, Cattivelli L (2008) Abiotic stress response in plants: when post-transcriptional and post-translational regulations control transcription. Plant Sci 174:420–431
Meier I (2012) mRNA export and sumoylation-Lessons from plants. Biochim Biophys Acta 1819:531–537
Miller MJ, Barrett-Wilt GA, Hua Z, Vierstra RD (2010) Proteomic analyses identify a diverse array of nuclear processes affected by small ubiquitin-like modifier conjugation in *Arabidopsis*. Proc Natl Acad Sci U S A 107:16512–16517
Miura K, Hasegawa PM (2010) Sumoylation and other ubiquitin-like post-translational modifications in plants. Trends Cell Biol 20:223–232
Miura K, Jin JB, Lee J, Yoo CY, Stirm V, Miura T, Ashworth EN, Bressan RA, Yun DJ, Hasegawa PM (2007) SIZ1-mediated sumoylation of ICE1 controls CBF3/DREB1 A expression and freezing tolerance in *Arabidopsis*. Plant Cell 19:1403–1414
Mizumoto K, Hatano H, Hirabayashi C, Murai K, Takumi S (2011) Characterization of wheat Bell1-type homeobox genes in floral organs of alloplasmic lines with Aegilops crassa cytoplasm. BMC Plant Biol 11:2
Palusa SG, Ali GS, Reddy ASN (2007) Alternative splicing of pre-mRNAs of *Arabidopsis* serine/arginine-rich proteins: regulation by hormones and stresses. Plant J 49:1091–1107
Park HC, Choi W, Park HJ et al (2011) Identification and molecular properties of SUMO-binding proteins in *Arabidopsis*. Mol Cells 32:143–151

Santos AP, Serra T, Figueiredo DD, Barros P, Lourenço T, Chander S, Oliveira MM, Saibo NJ (2011) Transcription regulation of abiotic stress responses in rice: a combined action of transcription factors and epigenetic mechanisms. OMICS 15:839–57

Sasaki K, Kim MH, Imai R (2007) Arabidopsis COLD SHOCK DOMAIN PROTEIN2 is a RNA chaperone that is regulated by cold and developmental signals. Biochem Biophys Res Commun 364:633–638

Schöning JC, Streitner C, Meyer IM, Gao Y, Staiger D (2008) Reciprocal regulation of glycine-rich RNA-binding proteins via an interlocked feedback loop coupling alternative splicing to nonsense-mediated decay in *Arabidopsis*. Nucleic Acids Res 36:6977–6987

Seo PJ, Kim MJ, Ryu JY, Jeong EY, Park CM (2011) Two splice variants of the IDD14 transcription factor competitively form nonfunctional heterodimers which may regulate starch me-tabolism. Nat Commun 2:303

Seo PJ, Park MJ, Lim MH, Kim SG, Lee M, Baldwin IT, Park CM (2012) A self-regulatory circuit of CIRCADIAN CLOCK-ASSOCIATED1 underlies the circadian clock regulation of temperature responses in *Arabidopsis*. Plant Cell 24:2427–2442

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

Sunkar R, Li YF, Jagadeeswaran G (2012) Functions of microRNA in plant stress responses. Trends in. Plant Sci 17:196–203

Tang Z, Zhang L, Xu C, Yuan S, Zhang F, Zheng Y, Zhao C (2012) Uncovering small RNA-mediated responses to cold stress in a wheat thermosensitive genic male-sterile line by deep sequencing. Plant Physiol 159:721–738

Thiebaut F, Rojas CA, Almeida KL, Grativol C, Domiciano GC, Regina C, Lamb C, De Almeida Engler J, Hemerly AS, Ferreira PCG (2012) Regulation of miR319 during cold stress in sugarcane. Plant Cell Envir 35:502–512

Wang Y, Liu K, Liao H, Zhuang C, Ma H, Yan X (2008) The plant WNK gene family and regulation of flowering time in *Arabidopsis*. Plant Biol 10:548–562

Yan J, Wang J, Li Q, Hwang JR, Patterson C, Zhang H (2003) AtCHIP, a U-Box-Containing E3 ubiquitin ligase, plays a critical role in temperature stress tolerance in *Arabidopsis*. Plant Physiol 132:861–869

Zhang J, Xu Y, Huan Q, Chong K (2009) Deep sequencing of *Brachypodium* small RNAs at the global genome level identifies microRNAs involved in cold stress response. BMC Genomics 10:449

Zhou X, Wang G, Sutoh K, Zhu JK, Zhang W (2008) Identification of cold-inducible microRNAs in plants by transcriptome analysis. Biochim Biophys Acta 1779:780–788

Zhu J, Dong CH, Zhu JK (2007) Interplay between cold-responsive gene regulation, metabolism and RNA processing during plant cold acclimation,Curr. Opin. Plant Biol. 10: 290–295

Crosatti et al. BMC Genomics 2013, 14:868
http://www.biomedcentral.com/1471-2164/14/868

BMC Genomics

RESEARCH ARTICLE Open Access

Cytoplasmic genome substitution in wheat affects the nuclear-cytoplasmic cross-talk leading to transcript and metabolite alterations

Cristina Crosatti[2†], Lydia Quansah[1†], Caterina Maré[2], Lorenzo Giusti[2], Enrica Roncaglia[4], Sergio G Atienza[3], Luigi Cattivelli[2*] and Aaron Fait[1*]

Abstract

Background: Alloplasmic lines provide a unique tool to study nuclear-cytoplasmic interactions. Three alloplasmic lines, with nuclear genomes from *Triticum aestivum* and harboring cytoplasm from *Aegilops uniaristata, Aegilops tauschii* and *Hordeum chilense,* were investigated by transcript and metabolite profiling to identify the effects of cytoplasmic substitution on nuclear-cytoplasmic signaling mechanisms.

Results: In combining the wheat nuclear genome with a cytoplasm of *H. chilense,* 540 genes were significantly altered, whereas 11 and 28 genes were significantly changed in the alloplasmic lines carrying the cytoplasm of *Ae. uniaristata* or *Ae. tauschii*, respectively. We identified the RNA maturation-related process as one of the most sensitive to a perturbation of the nuclear-cytoplasmic interaction. Several key components of the ROS chloroplast retrograde signaling, together with the up-regulation of the ROS scavenging system, showed that changes in the chloroplast genome have a direct impact on nuclear-cytoplasmic cross-talk. Remarkably, the *H. chilense* alloplasmic line down-regulated some genes involved in the determination of cytoplasmic male sterility without expressing the male sterility phenotype. Metabolic profiling showed a comparable response of the central metabolism of the alloplasmic and euplasmic lines to light, while exposing larger metabolite alterations in the *H. chilense* alloplasmic line as compared with the *Aegilops* lines, in agreement with the transcriptomic data. Several stress-related metabolites, remarkably raffinose, were altered in content in the *H. chilense* alloplasmic line when exposed to high light, while amino acids, as well as organic acids were significantly decreased. Alterations in the levels of transcript, related to raffinose, and the photorespiration-related metabolisms were associated with changes in the level of related metabolites.

Conclusion: The replacement of a wheat cytoplasm with the cytoplasm of a related species affects the nuclear-cytoplasmic cross-talk leading to transcript and metabolite alterations. The extent of these modifications was limited in the alloplasmic lines with *Aegilops* cytoplasm, and more evident in the alloplasmic line with *H. chilense* cytoplasm. We consider that, this finding might be linked to the phylogenetic distance of the genomes.

Keywords: Wheat, Alloplasmic line, Nuclear-cytoplasmic interaction, ROS signaling, Cytoplasmic male sterility

* Correspondence: luigi.cattivelli@entecra.it; fait@bgu.ac.il
†Equal contributors
[2]Consiglio per la Ricerca e la Sperimentazione in Agricoltura -Genomics Research Centre, via S. Protaso 302, 29017, Fiorenzuola d' Arda, (PC), Italy
[1]Jacob Blaustein Institutes for Desert Research, French Associates Institute for Agriculture and Biotechnology of Drylands, Ben-Gurion University of the Negev, Midreshet Ben-Gurion, 84990, Sde Boqer, Israel
Full list of author information is available at the end of the article

Background

The genetic information of the eukaryotic organisms is divided into a nuclear genome and cytoplasmic genomes (mitochondrion and chloroplast, sometimes referred to as plasmon). Nevertheless, mitochondrion and chloroplast contain only a few hundred genes, and their functionality is largely dependent on nuclear genes. This condition imposes a complex coordination between nuclear and cytoplasmic gene expression activities to assure that all mitochondrion and chloroplast proteins are timely and correctly formed [1,2]. Alloplasmic lines, alien cytoplasm substitution [3], offer a unique opportunity to study the nuclear-cytoplasmic interactions and the traits whose expression is dependent on the coordinate activity of nuclear and cytoplasmic genes [4-6]. The most relevant cytoplasm-inherited trait is cytoplasmic male sterility (CMS) [1], although other traits are also known (e.g., insect resistance [7]).

Alloplasmic lines created using a nuclear genome from *Triticum spp*. and a cytoplasm from *Aegilops* were the first ones created in grasses [8]. Nevertheless, the instability of some of the traits and a detrimental phenotype reported in the *Aegilops-Triticum* system has fostered the investigation of new cytoplasmic-nuclear systems. Due to its high crossability with other members of the *Triticeae* tribe [9], *Hordeum chilense* Roem et Schultz has been of interest as a source of new traits to be transferred to wheat, and it has also been suggested as a basis for new cytoplasmic-nuclear systems [10-12]. Aside from the possible beneficial effect that can be derived from the introgression of cytoplasm genomes from wild species into crops, a detrimental effect has also been consistently reported up to this point. It was observed that the alloplasmic lines carrying the *Aegilops spp*. or *H. chilense* cytoplasm performed worse than the euplasmic line in regard to many agronomic traits [6,13].

A non-natural combination of nuclear and cytoplasm genomes results in an essential alteration of the gene balance that infringes on the adaptability of the new genotype and brings about changes in quantitative traits and biological functions under the influence of mitochondrial and chloroplast genomes [14], which, in turn, can have an adverse effect on the responses to biotic and abiotic stresses [15]. The chloroplast and mitochondrial genomes work in coordination with the nuclear genome to support their functions and their adaptation to the changeable environmental conditions [16]. Conversely, alteration of the cytoplasm genome has a relevant impact on nuclear gene expression. For instance, many nuclear genes were found altered in a *Brassica napus* CMS line, particularly those encoding proteins involved in the organelles protein import machinery and the genes expressed in stamens and pollens, as well as the genes implicated in cell-wall remodeling or architecture [17]. Although substantial work has been carried out to investigate gene expression in alloplasmic lines [18-20], a majority of these studies were devoted to genes associated with floral organ development in CMS *Brassicaceae* [21]. As such, the absence of published comparative studies investigating the alteration in transcript and metabolite profiles, due to a foreign cytoplasm, limits the understanding of nuclear-cytoplasmic interaction and its output on the cellular processes. The central role of the chloroplast and mitochondria on plant cell metabolism and the known potential of CMS in crop breeding call for a comprehensive understanding of the global changes occurring as a result of foreign cytoplasm integration.

Very often the analysis of nuclear-cytoplasmic interaction is carried out by means of nuclear mutants [22], on the contrary, differences detectable between an alloplasmic line *vs*. the corresponding euplasmic can reveal the impact of small variations of the cytoplasm genome on the nuclear-cytoplasmic cross-talk. This study was carried out to assess the impact of alloplasmic lines on cell functionality. Parallel transcript and metabolite profiles were performed on three wheat alloplasmic lines obtained by cytoplasmic genome introgression from wild species (*Ae. uniaristata, Ae. tauschii* and *H. chilense*) into cultivated wheat and in their respective euplasmic genotypes. The effect that phylogenetic relatedness has on gene expression and metabolic regulation of common wheat is discussed. We hypothesize that, in order to maintain its functionality, the cell must integrate the perturbation caused by a non-natural nuclear-cytoplasmic interaction and reorganize its regulatory system and metabolism accordingly.

Results

The substitution of a wheat cytoplasm with an alien cytoplasms has an impact on the nuclear transcriptome

Three wheat alloplasmic lines and the corresponding euplasmic controls were subjected to transcriptome analysis to assess the effect of cytoplasm substitution on gene expression. The alloplasmic lines T183 and T195 were developed through the introgression of the cytoplasm from *Ae. uniaristata* (T183) and *Ae. tauschii* (T195) in the nuclear background of *T. aestivum* cv. Chris by Prof. S.S. Maan [23]. The alloplasmic line TH237 was produced by introgressing the *H. chilense* accession H7 cytoplasm into the nuclear background of *T. aestivum* accession T20 [13]. The seeds of the alloplasmic lines were multiplied at IAS-CSIC and verified for cytoplasm identity with specific cytoplasm markers. T183, T195 and TH237 were selected, based on published data [13,24], in order to represent the largest cytoplasmic diversity compatible with an overall normal phenotype (no evidence of development or flower abnormalities). A global gene expression analysis was performed on RNA samples isolated from the fully expanded

leaves of two-week old plants grown at 600 μE m^{-2} s^{-1}. Array hybridizations were carried out using the Affymetrix Wheat Genome Array. A Principal Component Analysis (PCA, a mathematical representation of the dataset diversity [25]), performed on the transcriptome dataset, identified two main components that explain 29.5% and 9.6% of the variance (Figure 1). The first component separated the alloplasmic lines developed in the genetic background of the accession T20 from those developed in the genetic background of the cultivar Chris. The second component allowed for the separation of the alloplasmic line TH237 (*H. chilense* cytoplasm) from the corresponding euplasmic, but not the separation of the other two alloplasmic lines (*Aegilops* spp. cytoplasm) from the cv. Chris. The PCA results suggest the existence of important differences between the two genetic backgrounds T20 and Chris; therefore, all the findings of this work are based only on the comparisons between each alloplasmic line and its corresponding euplasmic.

A Welch t-test, using a two-fold change cut-off and a false discovery rate correction for multiple testing, identified 562 probe sets, each of them representing a putative gene, differentially modulated in at least one experimental comparison (Additional file 1: Table S1). When the wheat nuclear genome was combined with the cytoplasm of *H. chilense*, 540 genes significantly altered their expression (295 up- and 245 down-regulated), whereas 11 (2 up- and 9 down-regulated) and 28 (5 up- and 23 down-regulated) genes were significantly modulated in the alloplasmic lines carrying the cytoplasm of *Ae. uniaristata* and *Ae.tauschii*, respectively. Two up-regulated (Ta.24245.1.A1_at and TaAffx.57297.1.S1_at annotated as maturase with chloroplast and mitochondrial localization) and five down-regulated genes (TaAffx.6196.2.S1_s_at and TaAffx.6196.1.S1_atannotated as cyt b_{559}, TaAffx.128795.10.S1_x_at annotated as ribosomal protein rps12, TaAffx.128617.1.S1_x_at annotated as NADPH-quinone oxidoreductase and TaAffx.128795.12.S1_x_atannotated as ABC transporter permease) were common in all comparisons (Figure 2), underlining some common responses across all alloplasmic lines. Significantly, the five down-regulated sequences were all encoded by the chloroplast genome.

The up- and down-regulated genes were divided into classes according to their expression behavior and classified using wheat gene identifier categories of the MapMan software [26]. Figure 3 summarizes the metabolism overview of up- and down-regulated genes in the *H. chilense* alloplasmic line *vs.* the corresponding euplasmic control according to the MapMan classification. The results

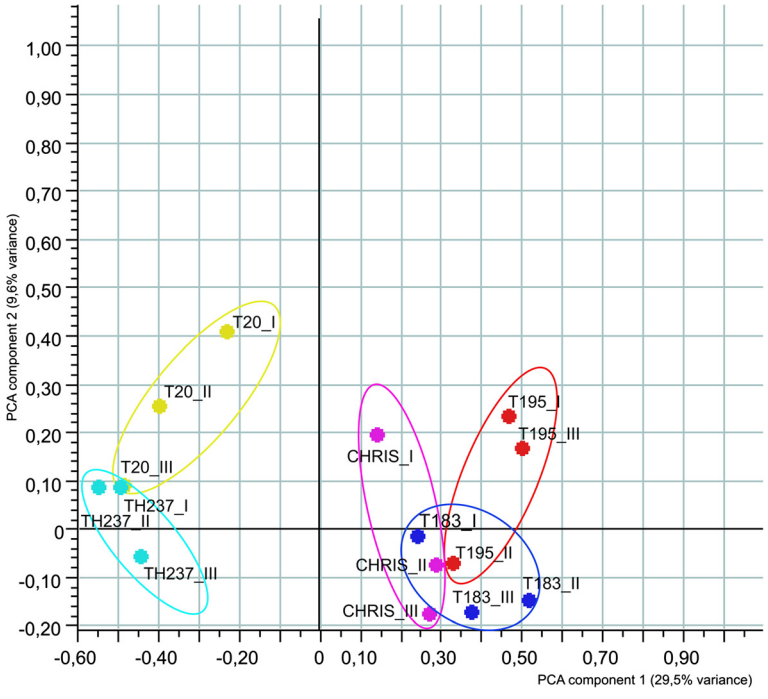


Figure 1 PCA plots of wheat alloplasmic array hybridization data. The samples have been represented according to the variance explained by the two principal component (29.5% and 9.6% on x and y axis, respectively) variations detected through transcriptome comparison. Biological replicates of each sample are surrounded by lines of different colors.TH237: alloplasmic line with *H. chilense* cytoplasm; T20: wheat accession representing the euplasmic control for TH237; T183: alloplasmic line with *Ae. uniaristata* cytoplasm; T195: alloplasmic line with *Ae. tauschii* cytoplasm; Chris: wheat cultivar representing the euplasmic control for T183 and T195.

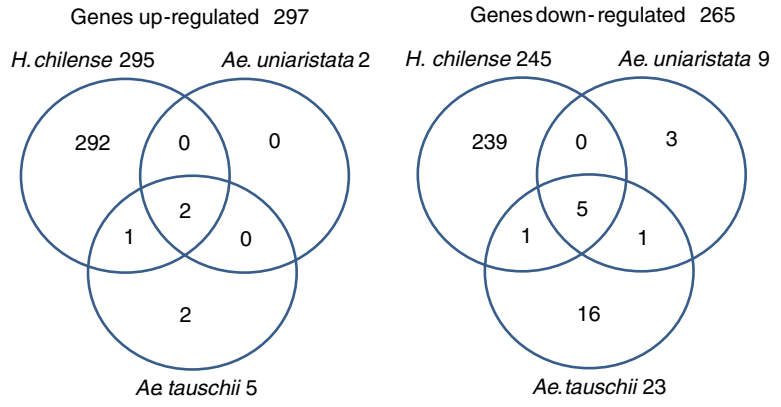


Figure 2 Representation of up- and down-regulated genes. The Venn diagrams show the genes up- and down-regulated in the alloplasmic lines developed with the cytoplasm of *Ae. uniaristata* (T183), *Ae. tauschii* (T195) and *H. chilense* (TH237) in comparison with the corresponding euplasmic controls. A single gene, up-regulated in some comparisons and down-regulated in others, has been excluded in this analysis.

showed that the genes associated with mitochondrial electron transport, photosynthesis, redox and hormones were generally up-regulated, while the genes associated with lipid metabolism (above all, lipid metabolism in pollen development), secondary metabolism, and signaling were mainly down-regulated.

The analysis of the sub-cellular localization of the proteins, encoded by the genes that modulated in response to cytoplasm substitution, identified 65 genes coding for plastid-targeted proteins and 22 genes coding for mitochondria-targeted proteins (15.2% of the 562 genes modulated in the experiment, Additional file 1: Table S1).

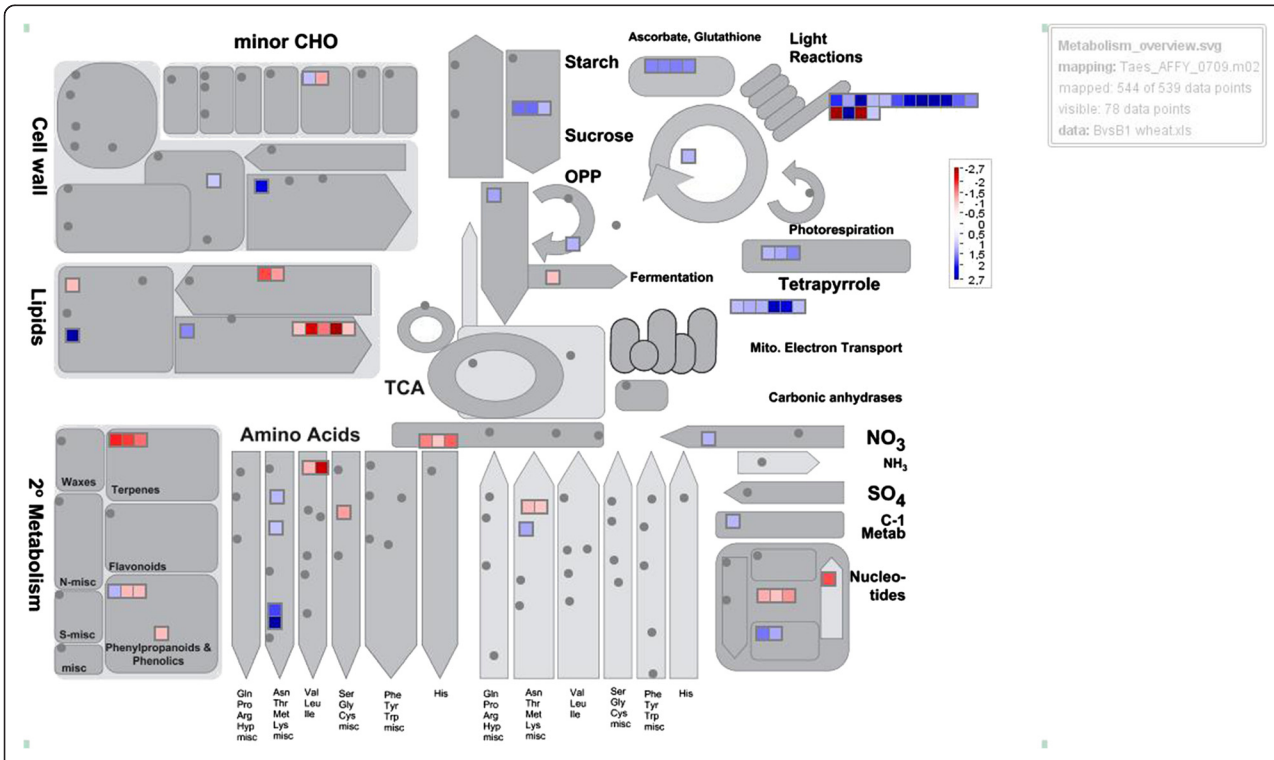


Figure 3 MapMan classification of differentially expressed genes. MapMan Metabolism overview maps showing the differences in transcript levels between the alloplasmic line TH237 (*H. chilense* cytoplasm) and the T20 euplasmic control [26]. The $\log_2$ expression ratios of genes altered in the comparison were loaded into the MapMan Image Annotator Module to map the wheat transcriptome data to define functional categories and to display them onto pathway diagrams. In the color scale, red and blue represent decreased and increased gene expression, respectively, in the *H. chilense* alloplasmic line as compared with the corresponding T20 euplasmic line.

Among them, 18 plastid and 11 mitochondrial-localized proteins were encoded by cytoplasmic genomes.

A validation experiment was carried out with six differentially expressed genes (*AtnMat3*, *Executer2*, *Flu*, *psbM*, *FAR*, *LSD1*, Additional file 2: Table S2) selected for their role in the chloroplast-retrograde signaling pathway (see below for their descriptions). The tested genes showed a value of 0.78 of the Pearson correlation coefficient between array and qRT-PCR expression values with an expression trend conserved across the samples between array and qRT-PCR results (detailed results are reported in Additional file 2: Table S2).

A significant proportion of the genes that were altered in response to the substitution of a wheat cytoplasm with an alien cytoplasm can be traced back to specific cellular processes or metabolic pathways. The most relevant ones are described in the next sections.

RNA maturation-related processes in the organelles are sensitive to perturbations of the nuclear-cytoplasmic cross-talk

The nucleus controls organelle gene expression through a number of nuclear-encoded factors that act post-transcriptionally on specific organelle target transcripts (anterograde signaling pathway [27]) and that regulate the maturation, stability and/or translation of organelle transcripts. RNA maturation in mitochondrion and chloroplast, including RNA splicing, is highly dependent on nuclear-encoded splicing factors. A general consequence of the perturbation induced in the alloplasmic lines was a strong modulation of the nuclear genes coding for organelle-targeted maturases. The probe set TaAffx.57297.1.S1_at homologous to *AtnMat3*, a maturase with demonstrated mitochondrion localization [28], was up-regulated with a $\log_2$ fold change of 5.3, 3.7, and 1.6 in alloplasmic lines carrying *Ae. uniaristata*, *Ae. tauschii* and *H. chilense* cytoplasm, respectively. Similarly, Ta.24245.1.A1_at homologous to *AtnMat4*, localized in both mitochondrion and chloroplast [28], was also up-regulated with a $\log_2$ fold change of 6.5, 5.1and 3.0 in the same comparisons indicated above.

Although the targets of AtnMat3 and AtnMat4 are still unknown, it should be noted that the RNA coding for five chloroplast encoded genes, cyt b_{559} (two probe sets), ribosomal protein rps12, NADPH-quinone oxidoreductase and ABC transporter permease, were down-regulated in all alloplasmic lines, an expression trend that suggests a possibly functional link between the up-regulation of nuclear-encoded maturase genes and chloroplast expression.

Furthermore, the wheat chip array carried a probe set (TaAffx.109850.1.A1_s_at) annotated as chloroplast maturase K (MatK) in Plexdb (http://www.plexdb.org/), an essential component for the splicing of plastid RNAs. This probe set was down-regulated in the alloplasmic line carrying the *H. chilense* cytoplasm. The plastidial MatK is associated with several chloroplast genes with class II introns [29]; one of them, the mRNA precursor coding for ribosomal protein rpl2 (TaAffx.128896.8.A1_at), was also down-regulated with a $\log_2$ fold change of −3.3 in the same alloplasmic line.

The substitution of a wheat cytoplasm with an alien cytoplasms alters the chloroplast retrograde signaling

To coordinate nuclear gene expression with the functional or metabolic state of plastids, the plant cells have acquired retrograde signaling pathways from plastid to nucleus, including redox signaling, Mg-protoporphyrin IX (an intermediate of the tetrapyrrole biosynthetic pathway), signals that are generated by inhibiting plastid gene expression or by accumulating various reactive oxygen species (ROS), such as hydrogen peroxide (H_2O_2), singlet oxygen ($^1O_2$), hydroxyl radical (OH) and superoxide anion (O_2^-) [30]. Excess light can potentially regulate nuclear gene expression by affecting tetrapyrrole biosynthesis. Plants exposed to high light, as was the condition for the transcriptomic experiment here reported, require specific adaptations that are mediated by chloroplast retrograde signaling [30]. Therefore, we reasoned that alterations in this signaling pathway should be evident in the experimental dataset, and a specific mining work was dedicated to finding genes that represent the key components of this signaling with an altered expression in response to a cytoplasm substitution.

HEMA1 encodes one of the three isoforms of the glutamyl-tRNA reductase, which catalyzes the first key regulatory step of the tetrapyrrole biosynthesis, and three probe sets (Ta.3243.1.S1_at, Ta.3243.1.S1_x_at and TaAffx.8262.1.S1_x_at) corresponding to *HEMA1* were up-regulated in the alloplasmic line carrying the cytoplasm of *H. chilense*. It is known that the accumulation of chlorophyll precursors, such as Mg-Protoporphyrin IX, may stimulate the expression of the nuclear gene *HSP70* [31]. Importantly, in the *H. chilense* alloplasmic line, the up-regulation of *HEMA1* was associated with a parallel up-regulation of *HSP70-1* (Ta.24154.1.S1_s_at).

Low concentrations of ROS act as signaling molecules for a number of regulated processes during plant growth and development [32], as well as in response to a variety of environmental stimuli [33,34]. The existence of a specific $^1O_2$ signaling pathway has been shown in *Arabidopsis* through a mutant analysis of *flu* (*fluorescent in blue light*) [35] and its regulators named *executer* (*ex1*and *ex2*) [36,37]. Three probe sets (Ta.24301.1.A1_x_at, TaAffx.128618.1.S1_at andTa.24301.1.A1_s_at), corresponding to *Flu*, and one probe set (TaAffx.110498.1.S1_at), corresponding to*Ex2*, were all up-regulated in the alloplasmic line carrying the *H. chilense* cytoplasm, in comparison with the corresponding euplasmic. Furthermore, the same alloplasmic line was characterized by an

up-regulation of three probe sets (Ta.28069.1.A1_at, Ta.6594.1.S1_at and Ta.6594.1.S1_s_at), corresponding to *LSD1* (*LESION SIMULATING DISEASE 1*), a putative regulator of the hypersensitive response and light acclimation driven by ROS accumulation [38]. Notably, the up-regulation of *LSD1* was associated with an altered expression of 26 probe sets involved in the hypersensitive response to biotic stress belonging to the pathway of methyl jasmonate and salicylic acid (Table 1). Among these sequences, there were enzymes involved in the synthesis of salicylic acid (isochorismate synthases, one probe set), the WRKY70 transcription factor (two probe sets), lipoxigenase 2 (three probe sets), DIR1 (Defective Induced Resistance 1, a gene involved in the systemic acquired resistance -SAR, two probe sets), FMO1 (Flavin-dependent Monoxigenase 1, an essential component of SAR, three probe sets) and several other pathogen responsive proteins. Furthermore, in the *H. chilense* alloplasmic line, we found several down-regulated genes involved in programmed cell death: ACD1 (Accelerated Cell Dead 1, TaAffx.67544.1.S1_at), STP13 (Sugar Transport Protein 13, Ta.5766.1.S1_at), VEP1 (Vein Patterning 1, Ta.13682.1.A1_at) and two probe sets annotated as putative metacaspases (Ta.3154.1.S1_at and Ta.10581.1.A1_at) [39-42].

Since *Flu, Ex* and *LSD1* are key components of the ROS chloroplast retrograde signaling [37,43,44], the identification of these genes, among those altered in the *H. chilense* alloplasmic line, as well as of other downstream genes, suggests that a change in the chloroplast genome has a direct impact on the chloroplast retrograde signaling mediated by $^1O_2$.

The substitution of a wheat cytoplasm with an alien cytoplasm alters the expression of photosynthetic components

Retrograde signaling is known to control the expression of many genes coding for components of the PSII, and indeed, a number of these genes, mostly encoded by the chloroplast genome, were changed in the alloplasmic lines (Table 2 and Additional file 1: Table S1). Two genes (probe sets TaAffx.6196.1.S1_at and TaAffx.6196.2.S1_s_at) coding for the cytochrome b_{559} alpha chain (psbE) were down-regulated in all alloplasmic lines, while a gene coding for a component of the psbM protein (TaAffx.48564.2.S1_s_at) was strongly down–regulated in the *H. chilense* alloplasmic line (log_2 fold change of −7.9). Furthermore, eleven genes coding for chlorophyll a/b-binding protein (LHCII), five genes related to LT29/APX4 (a protein associated with photosystem II [45] and a gene coding for a D1 precursor (TaAffx.128757.1.S1_at) were up-regulated in the alloplasmic line carrying the *H. chilense* cytoplasm. Lastly, the alloplasmic line carrying the *Ae. tauschii* cytoplasm showed a down-regulation of a gene encoding a psbX-related protein (Ta.20878_at).

NADH-plastoquinone oxidoreductase chain 5 (TaAffx.128617.1.S1_x_at), a component of the chloroplast NADH-ubiquinone oxidoreductase, was down-regulated in all alloplasmic lines. Furthermore, a probe set, corresponding to ATP synthase (TaAffx.128565.9.S1), was up-regulated in the alloplasmic line with the *H. chilense* cytoplasm.

To conclude, the findings described above highlight that the substitution of the wheat cytoplasm with the cytoplasm of *H. chilense*, impacts the expression of several nuclear genes encoding photosynthetic components, suggesting a modification of the nuclear-chloroplast cross-talk, at least under high light conditions.

The substitution of a wheat cytoplasm with an alien cytoplasm modifies the expression of genes coding for components of the mitochondrial electron transport chains

The mitochondrial electron transport chain is comprised of four large protein complexes (I, II, III and IV) that interact with each other via the small lipid ubiquinone and the small protein cytochrome c [46]. Three probe sets (Ta.24230.1.S1_at, Ta.28812.1.S1_at and TaAffx.87200.1.S1_at) related to NADH dehydrogenase (EC 1.6.5.3, complex I), the first step of the oxidative phosphorylation, had an altered expression profile in the alloplasmic lines. The first two genes were up-regulated in the alloplasmic line with the *H. chilense* cytoplasm, while the last one was up-regulated in the alloplasmic lines with the *Ae. uniaristata* and the *H. chilense* cytoplasms. Furthermore, four probe sets (TaAffx.69955.1.S1_at, Ta.28701.1.A1_at, TaAffx.114431.1.S1_s_at and Ta.28645.1.S1_at), related to different subunits of the cytochrome c oxidase (COX; EC 1.9.3.1, complex IV) [45], were up-regulated in the alloplasmic line with the *H. chilense* cytoplasm. TaAffx.4544.2.S1_s_at, involved in cytochrome c oxidase biogenesis [47], was, instead, down-regulated in the *Ae. tauschii* and *H. chilense* alloplasmic lines. Notably, most of the genes described in this paragraph are encoded by the mitochondrial genome (Additional file 1: Table S1).

The expression of genes involved in cytoplasm male sterility is dependent on the interaction between nuclear and cytoplasm genomes

The transcriptome comparison of the alloplasmic line, carrying the cytoplasm of *H. chilense* with its euplasmic control, highlighted the alteration of several genes known to be involved in the determination of cytoplasm male sterility (CMS). Three up-regulated probe sets (Ta.1967.1.S1_x_at, TaAffx.104812.1.S1_s_at and Ta.1967.2.A1_x_at) were related to LOX2 (chloroplast 13-lipoxygenase), while two down-regulated probe sets (Ta.8097.1.A1_at and Ta.13650.1.A1_at) were annotated as LOX3 (chloroplast lipoxygenase). LOX2 and LOX3 play a role in the biosynthesis of jasmonate, and

Table 1 Probe sets encoding proteins involved in biotic stress responses significantly up- or down-regulated in the alloplasmic line TH237 carrying the *H. chilense* cytoplasm compared to the corresponding euplasmic line T20

Probe set ID	Log2 fold change	Probe set description
Ta.24195.3.S1_at	+1.19	Vrga1-disease resistance gene homologues
Ta.24195.1.A1_at	+1.99	Vrga1-disease resistance gene homologues
TaAffx.55225.1.S1_at	+1.50	Leucine-rich repeat family protein similar to disease resistance family protein
TaAffx.94000.1.S1_at	−1.40	Barley stem rust resistance protein
Ta.21533.1.A1_at	+2.18	Powdery mildew resistance protein
Ta.21314.1.S1_at	−1.62	Disease resistance response protein (PR protein)
Ta.21314.1.S1_x_at	−1.74	Disease resistance response protein (PR protein)
Ta.22687.1.A1_at	−1.75	Disease resistance response protein (PR protein)
TaAffx.28302.2.S1_at	−1.00	Disease resistance-responsive family protein (PR protein)
TaAffx.78606.1.S1_at	−1.81	Disease resistance protein (NBS-LRR class -PR protein)
TaAffx.107480.1.S1_at	−1.61	Cold induced ice recrystallisation inhibition protein. Phytohormones involved in pathogen defense pathways (jasmonic acid and ethylene)
Ta.19859.1.S1_at	+1.36	FMO1 (FLAVIN-DEPENDENT MONOOXYGENASE 1); mono-oxygenase FMO1 is required for full expression of TIR-NB-LRR conditioned resistance to avirulent pathogens and for basal resistance to invasive virulent pathogens. Functions in an EDS1-regulated but SA-independent
TaAffx.104885.1.S1_at	+1.61	FMO1 (FLAVIN-DEPENDENT MONOOXYGENASE 1); mono-oxygenase FMO1 is required for full expression of TIR-NB-LRR conditioned resistance to avirulent pathogens and for basal resistance to invasive virulent pathogens. Functions in an EDS1-regulated but SA-independent
TaAffx.104885.2.S1_at	+1.69	FMO1 (FLAVIN-DEPENDENT MONOOXYGENASE 1); mono-oxygenase FMO1 is required for full expression of TIR-NB-LRR conditioned resistance to avirulent pathogens and for basal resistance to invasive virulent pathogens. Functions in an EDS1-regulated but SA-independent
TaAffx.131611.1.S1_at	+4.29	DIR1 (DEFECTIVE IN INDUCED RESISTANCE 1); lipid binding encodes a putative apoplastic lipid transfer protein that is involved in systemic acquired resistance (SAR). Mutants in this gene exhibit wild-type local resistance to avirulent and virulent Pseudomonas
TaAffx.131611.4.S1_at	+2.76	DIR1 (DEFECTIVE IN INDUCED RESISTANCE 1); lipid binding encodes a putative apoplastic lipid transfer protein that is involved in systemic acquired resistance (SAR). Mutants in this gene exhibit wild-type local resistance to avirulent and virulent Pseudomonas
Ta.1967.1.S1_x_at	+1.59	LOX2 (LIPOXYGENASE 2) Chloroplast lipoxygenase required for wound-induced jasmonic acid accumulation in Arabidopsis. Mutants are resistant to Staphylococcus aureus and accumulate salicylic acid upon infection. Identical to Lipoxygenase, chloroplast precursor
TaAffx.104812.1.S1_s_at	+1.63	LOX2 (LIPOXYGENASE 2) Chloroplast lipoxygenase required for wound-induced jasmonic acid accumulation in Arabidopsis. Mutants are resistant to Staphylococcus aureus and accumulate salicylic acid upon infection. Identical to Lipoxygenase, chloroplast precursor
Ta.1967.2.A1_x_at	+1.64	LOX2 (LIPOXYGENASE 2) Chloroplast lipoxygenase required for wound-induced jasmonic acid accumulation in Arabidopsis. Mutants are resistant to *Staphylococcus aureus* and accumulate salicylic acid upon infection. Identical to lipoxygenase, chloroplast precursor
Ta.8097.1.A1_at	+1.02	LOX3 (Lipoxygenase 3); iron ion binding/lipoxygenase
Ta.13650.1.A1_at	+1.93	Lipoxygenase, putative similar to LOX3 (Lipoxygenase 3), iron ion binding/lipoxygenase
Ta.21353.2.S1_at	−1.01	Putativesalicylic acid-binding protein PP
Ta.8614.1.S1_at	−1.30	WRKY70 (WRKY DNA-binding protein 70); transcription factor member of WRKY Transcription Factor; Group III. Function as activator of SA-dependent defense genes and a repressor of JA-regulated genes. WRKY70-controlled suppression of JA-signaling is partly executed by NPR1.
Ta.8614.2.S1_x_at	−1.66	WRKY70 (WRKY DNA-binding protein 70); transcription factor member of WRKY Transcription Factor; Group III. Function as activator of SA-dependent defense genes and a repressor of JA-regulated genes. WRKY70-controlled suppression of JA-signaling is partly executed by NPR1.
TaAffx.37536.1.S1_at	+1.03	Isochorismate synthases involved in salicylic acid synthesis
Ta.22221.1.S1_a_at	−1.51	ADC2 (ARGININE DECARBOXYLASE 2) encodes an arginine decarboxylase (ADC), a rate-limiting enzyme that catalyzes the first step of polyamine (PA) biosynthesis via ADC pathway in Arabidopsis thaliana.

Table 2 Probe sets encoding components of the photosynthetic apparatus significantly up- or down-regulated in the three alloplasmic lines compared with the corresponding euplasmic line

Probe set ID	Log2 fold change alloplasmic T183 vs. euplasmic Chris	Log2 fold change alloplasmic T195 vs. euplasmic Chris	Log2 fold change alloplasmic TH237 vs. euplasmic T20	Probe set description
TaAffx.6196.1.S1_at	−5.81	−5.65	−5.26	Cytochromeb559 alpha chain; psbE
TaAffx.6196.2.S1_s_at	−5.92	−5.63	−4.96	Cytochromeb559 alpha chain; psbE
TaAffx.48564.2.S1_s_at			−7.93	Photosystem II M protein; psbM
Ta.20878.1.S1_at				Protein Photosystem II reaction centre X protein (PsbX), putative
Ta.22984.1.S1_x_at		−1.07		Chlorophyll a/b-binding protein (LHCII
Ta.28496.1.A1_at			+2.61	Chlorophyll a/b-binding protein (LHCII
Ta.28496.1.A1_x_a			+2.54	Chlorophyll a/b-binding protein (LHCII
Ta.29587.2.S1_x_at			+1.33	Chlorophyll a/b-binding protein (LHCII
Ta.29587.3.A1_at			+3.23	Chlorophyll a/b-binding protein (LHCII
Ta.30702.1.S1_x_at			+1.99	Chlorophyll a/b-binding protein (LHCII
Ta.3249.1.S1_at			+3.41	Chlorophyll a/b-binding protein (LHCII
Ta.3249.2.S1_x_at			+1.72	Chlorophyll a/b-binding protein (LHCII
Ta.3249.3.A1_at			+3.47	Chlorophyll a/b-binding protein (LHCII
Ta.3795.1.S1_x_at			+1.13	Chlorophyll a/b-binding protein (LHCII
TaAffx.114127.1.S1_x_at			+1.83	Chlorophyll a/b-binding protein (LHCII
Ta.488.1.S1_at			+1.50	APX4 (ASCORBATE PEROXIDASE 4)
Ta.488.1.S1_x_at			+1.51	APX4 (ASCORBATE PEROXIDASE 4)
Ta.488.2.S1_at			+1.67	APX4 (ASCORBATE PEROXIDASE 4)
Ta.488.3.S1_a_at			+1.54	APX4 (ASCORBATE PEROXIDASE 4)
Ta.488.3.S1_at			+1.48	APX4 (ASCORBATE PEROXIDASE 4)
TaAffx.128757.1.S1_at			+1.48	Photosystem II protein D1 precursor
TaAffx.128617.1.S1_x_at	−5,14	−4,65	−4.62	Chloroplast (NAD(P)H dehydrogenase, chain 5
TaAffx.128565.9.S1_at			+4.98	Protein ATPase I subunit
Ta.25398.1.S1_at			−1.10	ELIP1 (early light-inducible protein); chlorophyll binding
TaAffx.128418.14.A1_at			−1.26	ELIP1 (early light-inducible protein); chlorophyll binding
Ta.26997.1.S1_at			−1.12	Putative yearly light-inducible protein

both are involved in the determination of male sterility [48]. LOXes catalyze the addition of oxygen to polyunsaturated fatty acids to produce the unsaturated fatty acid hydroperoxide. Linoleic and linolenic acids are the most common substrates for LOX [49], which favors free fatty acids to sterified ones [50]. LOXes are also involved in oxylipin biosynthesis during stress, a process mainly associated with wounding and pathogen attack [51].

Lipid metabolism is essential for normal pollen development [52,53]. A starting point in fatty-acid synthesis is the conversion of acetyl-CoA into malonyl-CoA. Six down-regulated probe sets (Ta.12415.3.S1_at, Ta.23988.1.A1_at, Ta.26144.1.A1_at, Ta.26144.1.A1_s_at; TaAffx.114108.1.S1_at and TaAffx.38271.1.A1_at) are all related to fatty acyl-CoA reductase (FAR), a putative male sterility protein involved in wax biosynthesis [54]. Furthermore, a probe set (Ta.16047.1.S1_at), corresponding to *male sterility 2* (*ms2*, chloroplastic fatty acid reductase), and two probe sets (Ta.1456.1.A1_at and TaAffx.38271.1.A1_at), related to the *CER4*, both involved in the synthesis of the lipids of the outer pollen wall [55], were down-regulated in the alloplasmic line with the *H. chilense* cytoplasm; these two genes belong to a family of genes encoding FAR enzymes. In the same sample, three additional probe sets (Ta.635.2.A1_at, Ta.635.2.A1_x_at and Ta.635.3.S1_a_at), related to *POP2* (*pollen-pistil incompatibility*) [56], an enzyme of the GABA-shunt involved in the conversion of the non-protein amino acid GABA to succinyl-semi-aldehyde [57], were also down-regulated, while a probe set (Ta.20591.2.

S1_a_at), related to plantacyanin involved in anther development and pollination, was up-regulated.

GC-MS analysis reveals the impact nuclear-cytoplasmic interaction has on leaf metabolism

Considering that the nucleus-chloroplast cross-talk is essential for the adaptation of the leaf metabolism to environmental variations, GC-MS-based metabolite profiling was done on the alloplasmic lines and corresponding euplasmic controls grown at two light intensities (low light 150 μE m^{-2} s^{-1}, and high light 600 μE m^{-2} s^{-1}) to test if a substitution of the cytoplasm impairs light response.

To acquire an overview of the global changes in leaf metabolism across genotypes, and conditions, the GC-MS dataset was subjected to PCA [25]. The analysis revealed a clear separation of the samples due to light effect on PC1 (Figure 4A and B) in both euplasmic backgrounds, accounting for more than 60% of the variance. These results suggest a comparable response to light intensity of the alloplasmic lines and their respective euplasmic controls, implying relatively functional mechanisms of electron dissipation and chloroplast functioning.

Nevertheless, substituting the wheat cytoplasm with the *H. chilense* cytoplasm led to a perturbation of the central metabolism in response to high light (HL), resulting in a separation of the euplasmic control from the *H. chilense* alloplasmic line with the PCA not observed under low light (LL) (Figure 4A). On the genetic background of Chris, the alloplasmic line carrying the *Ae. uniaristata* cytoplasm showed a slight separation from the euplasmic control under LL, while separating from the *Ae. tauschii* alloplasmic line. Under HL, the alloplasmic line carrying the *Ae. tauschii* cytoplasm grouped together with the euplasmic control and separated from the other alloplasmic line (Figure 4B), implying the existence of processes that affect the central metabolism in a species-specific manner.

The alteration in the metabolite profiles in response to light intensity was validated by a two-way ANOVA, separately testing the factors of cytoplasmic substitution, light treatment and their interaction (Additional file 3: Table S3 and Additional file 4: Table S4). As expected, light was confirmed as having a major impact on the metabolite profiles of all genotypes (Additional file 3: Table S3 and Additional file 4: Table S4). Substitution in T20 with the *H. chilense* cytoplasm led to the alteration in the abundance of 33% (17) of the annotated metabolites, while the interaction arising between genotype and light accounted for the change in 31% (16) of metabolites (Additional file 3: Table S3). In the Chris series, cytoplasm substitution caused a significant change in the abundance of 22% (11) and 24% (12) metabolites in alloplasmic lines with *Ae. uniaristata* and *Ae. tauschii* cytoplasm, respectively. The interaction between genotype and light treatment caused 16% (8) and 33% (17) of metabolites to be significantly altered in alloplasmic lines with *Ae. uniaristata* and *Ae. tauschii* cytoplasm, respectively (Additional file 4: Table S4). The average extent of the change in metabolite abundance was also dependent on the nuclear-cytoplasmic interaction. For example, the average fold change in the *H. chilense* cytoplasm-substituted plants was 2.6-fold, while in the

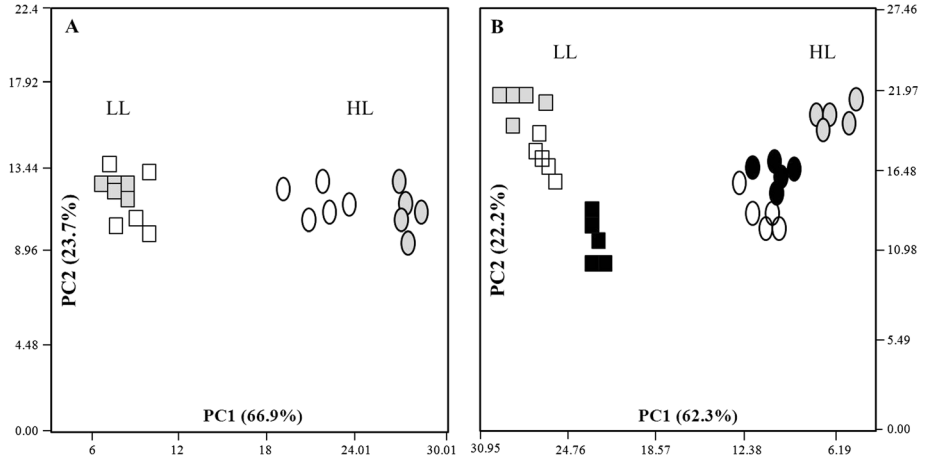


Figure 4 PCA plots of the wheat alloplasmic metabolic dataset. The metabolic dataset of leaves from plants grown at 150 (LL) and 600 (HL) μE m^{-2} s^{-1} is shown. Each spot on the plot is a representative of a sample (that is, five biological samples for each category) **A**: PCA of alloplasmic TH237 (*H. chilense* cytoplasm) *vs*. T20 euplasmic control. The rectangles and circles represent low light and high light, respectively. The white and grey colors represent T20 and TH237, respectively. **B**: PCA of alloplasmic lines T183 (*Ae. uniaristata* cytoplasm) and T195 (*Ae. tauschii* cytoplasm) in comparison with the wheat cv. Chris used as the euplasmic control. Rectangles and circles represent low light and high light, respectively. White, grey and black denote Chris, T183 and T195, respectively.

Chris series, the average fold change in metabolite abundance was 0.5 and 1.9-fold in *Ae. uniaristata* and *Ae. tauschii,* respectively.

Substitution of a wheat cytoplasm with a distant relative H. chilense cytoplasm alters leaf metabolic response to light

Under LL, the comparison between the alloplasmic line carrying the cytoplasm from *H. chilense* and the corresponding euplasmic revealed significant variations in the abundance of amino acids (aspartate, glycine, glutamate and homoserine), sugars (arabinose, glucose and glucose 1, 6 anhydro-β), sugar phosphates, mannitol, myo-inositol, organic acids (mainly citrate, malate, succinate and itaconate) and putrescine (a stress-related polyamine) [58,59] (Figure 5, Additional file 5: Table S5). Markedly, in the alloplasmic line, homoserine, an intermediate in the synthesis of the aspartate-derived amino acids [60], was reduced to 10% of its content in the euplasmic control. A three-step pathway converts L-aspartate into homoserine; hence, the accumulation of aspartate and the low homoserine level might suggest a compromised step in the biosynthesis. The coordinated up-regulation of myo-inositol pathway-related genes was shown under environmental stresses [61], and often, the role of these metabolites is reported in the literature to be associated with stress, as they serve as osmoprotectants [62].

Ae. uniaristata and *Ae. tauschii* cytoplasm substitutions showed milder effects on the metabolism of their euplasmic lines, compared to the effect of *H. chilense*. In several instances, the two *Aegilops* cytoplasms affected the metabolism in an opposite manner, while not showing significant alterations in stress-associated processes (Figures 6 and 7, Additional file 6: Table S6). Yet, a decreased level of intermediates associated with the TCA cycle (citrate, fumarate and malate) characterized both lines.

In the *Aegilops* series under HL, only a few metabolites were significantly altered, and as observed under LL, *Ae. uniaristata* and *Ae. tauschii* had opposite trends on some of the most significantly altered metabolites, e.g., homoserine and 5-caffeoylquinate (Figures 6 and 7, Additional file 6: Table S6). More significant metabolic alterations were found in the comparison between the *H. chilense* alloplasmic line and the corresponding euplasmic control under HL (Figure 5). In the *H. chilense* alloplasmic line, HL-induced alterations included a drop in the abundance of amino acid alanine, aspartic acid, homoserine and serine, compared to the euplasmic line. Significant increases in content were observed in sugars (except for arabinose and turanose), sugar phosphates and mannitol. The increase in glycine, glucose and most of the sugar phosphates in the *H. chilense* alloplasmic line could also reflect an enhancement in the photorespiratory metabolism, a finding supported by the up-regulation in the same alloplasmic line of several key genes of the photorespiration, e.g., glycolate oxidase (GOX), serine hydroxyl-methyl transferase (SHMT) and glutamine synthetase (GS). The remarkable enhancement in raffinose abundance (33-fold) and its downstream metabolite galactose is suggestive of a stress response. Under HL conditions, organic acids (butanoate-2, 4-dihydroxy, succinate, glycerate, quinate and shikimate) were at lower levels in the *H. chilense* alloplasmic line. Succinate can be produced from within the TCA cycle or by the conversion of GABA to sucinyl-semialdehyde and succinate, a reaction governed by GABA-transaminase and succinic semialdehyde-dehydrogenase; the former gene was shown to be down-regulated at the transcript level (Ta.635.2.A1_at, Ta.635.2.A1_x_at and Ta.635.3.S1_a_at). The analytical details are reported in Figure 5 and Additional file 4: Tables S4 and Additional file 5: Table S5. The decrease in shikimate intermediates, quinate and shikimate, may have supported the production of secondary metabolites.

The direct comparison between the transcriptomic and metabolomic data in the *H. chilense* alloplasmic line exposed to HL allows the identification of some correspondences between changes in the expression of genes coding for enzymes and the corresponding metabolites (Figure 8). The most clear was the up-regulation of phospho-glucomutase (EC 5.4.2.2), sucrose synthase (EC 2.4.1.13) and β-fructofuranosidase (3.2.1.26) mRNAs that were associated with an increased accumulation of sugars and sugar phosphates.

Substitution of a wheat cytoplasm with an H. chilense cytoplasm had no effect on fatty acid accumulation

Following the alteration of enzymes involved in lipid biosynthesis in the gene expression profiling of the alloplasmic line carrying the *H. chilense* cytoplasm, a fatty acid analysis was carried out for this alloplasmic line and its euplasmic control. Eleven detectable fatty acids were annotated with linolenic acid (18:3ω3) having the highest level in the alloplasmic line and its euplasmic control, compare to the other annotated fatty acids. Nevertheless, no significant differences were detected between the alloplasmic line and the euplasmic control grown at HL (Additional file 7: Table S7). The possible explanation for the differences in the fatty acid regulation at the metabolite and gene level could be that there are yet unknown regulatory genes that may participate in a post-translational process and hence do not affect the final metabolite level.

Discussion

The development of alloplasmic lines via alien cytoplasm substitution provides a unique tool with which to study

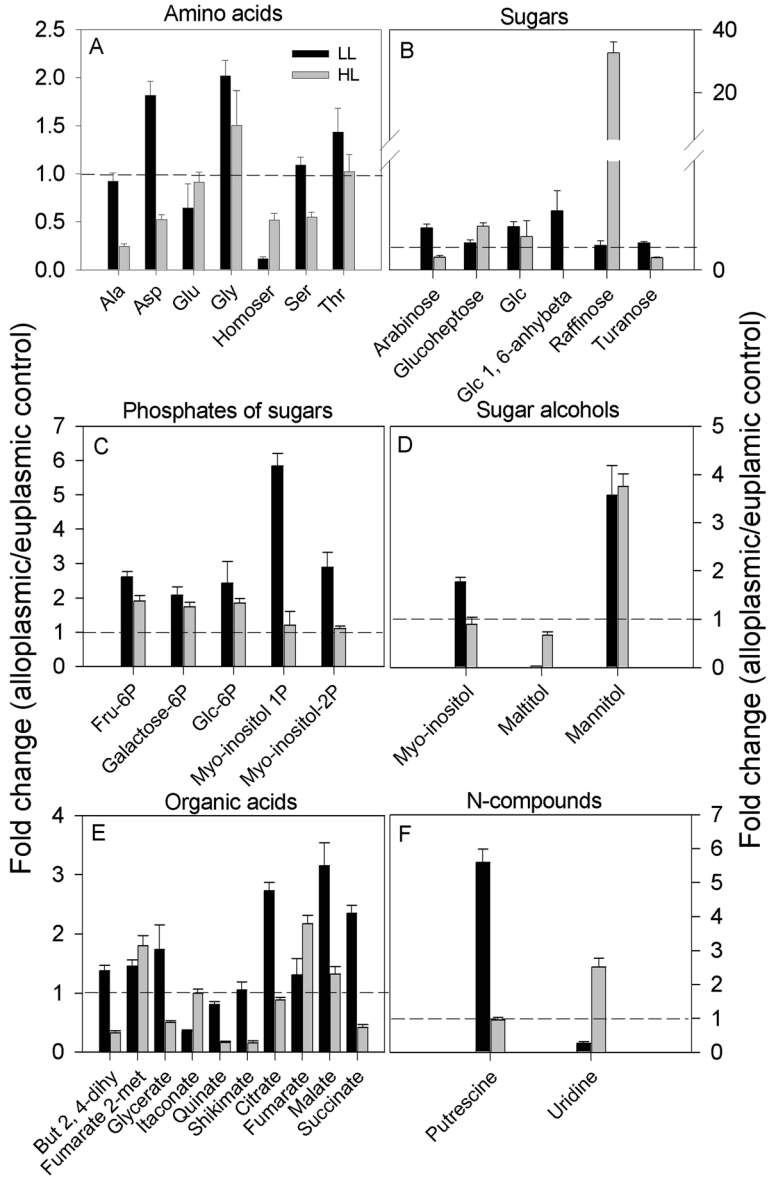


Figure 5 Metabolites altered in the TH237 (*H. chilense* cytoplasm) alloplasmic line. A-**F** denotes amino acids, sugars, sugar phosphates, sugar alcohols, organic acids, and N-compounds, respectively. Metabolites in each category that were significantly ($p \leq 0.05$) altered in either low light (LL, 150 μE m^{-2} s^{-1}) or high light (HL, 600 μE m^{-2} s^{-1}) in the mature leaf of the alloplasmic line TH237 (*H. chilense* cytoplasm) in comparison with the wheat accession T20 used as the euplasmic control. Bars are fold changes of the alloplasmic line on the euplasmic control. The euplasmic control is given the threshold 1. Bars above or below the dashed line (threshold value 1) indicate an increase or a decrease of each metabolite in the alloplasmic genotype compared with the euplasmic control. Each bar is presented as means ± SEM of five biological repetitions. Glc, glucose; glc1, 6-anhybeta, glucose-1, 6-anhydro β.

the nuclear-cytoplasmic interactions in plants. Hexaploid wheat resulted from a combination of three nuclear genomes (A genome from *T. urartu*, B genome from *Ae. speltoides* and D genome from *Ae. tauschii*) with a cytoplasm genome from *Ae. speltoides* [63,64]. The crossability of hexaploid wheat with most *Triticeae* species has permitted the development of a panel of alloplasmic lines where the wheat nuclear genome has been coupled with cytoplasm genomes from species with different degrees of phylogenetic relatedness. The current work investigated the transcript and the metabolite profiling of three wheat alloplasmic lines carrying the cytoplasm of two very closely related species (*Ae. uniaristata* and *Ae. tauschii)* and one of a more distant species (*H. chilense)*. The results showed that the *H. chilense* cytoplasm had a greater effect on the transcriptomic and metabolite

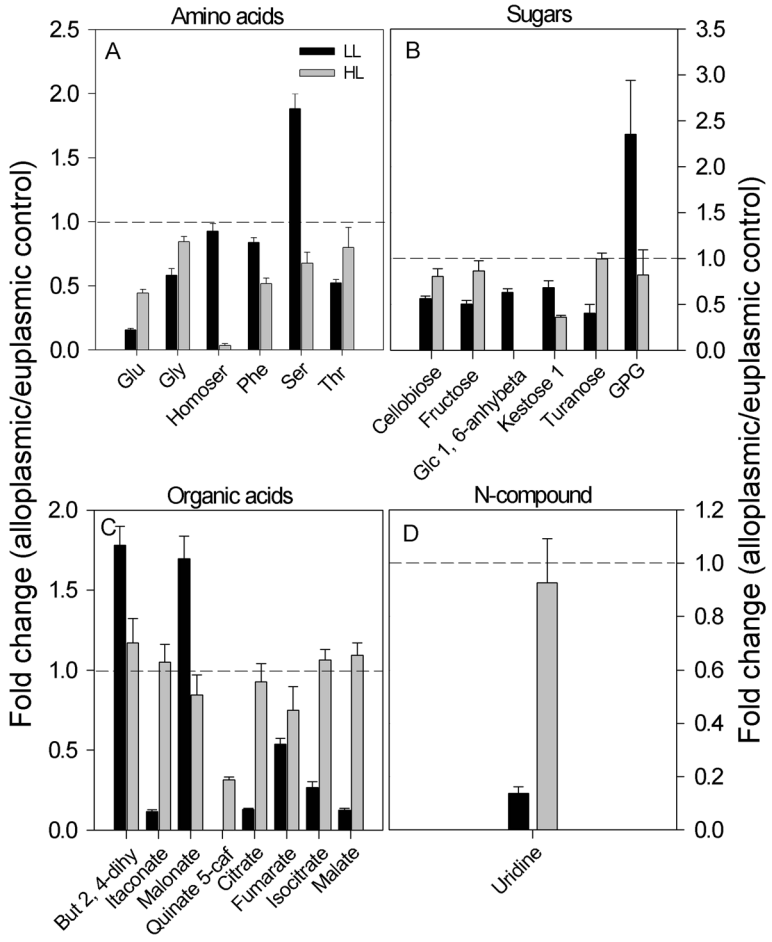


Figure 6 Metabolites altered in the T183 (*Ae. uniaristata* cytoplasm) alloplasmic line. A-D denotes amino acids, sugars, organic acids, and N-compounds, respectively Metabolites in each category that were significantly ($p \leq 0.05$) altered in either low light (LL, 150 μE m^{-2} s^{-1}) or high light (HL, 600 μE m^{-2} s^{-1}) in the mature leaf of the alloplasmic line T183 with the *Ae. uniaristata* cytoplasm in comparison with the wheat cv. Chris used as the euplasmic control. Bars are fold changes of the alloplasmic line on the euplasmic control. The euplasmic control is given the threshold 1. Bars above or below the dashed line (threshold value 1) indicate an increase or a decrease of each metabolite of the alloplasmic genotype compared with the euplasmic control. Each bar is presented as means ± SEM of five biological repetitions. GPG, glycerophosphoglycerol.

profiles of its euplasmic control than did the *Aegilops* cytoplasms on theirs. It has been suggested that the negative effects produced by substituting a native cytoplasm with a cytoplasm from a more or less related species are dependent on the evolutionary distance between the two species [4,13]. These features were generally considered as the consequence of an incompatibility between nuclear and organelle-encoded components of mitochondria and/or chloroplast after their evolution into separate species. During evolutionary processes, nuclear genomes may have accumulated mutations, while organelle genomes accumulated compensatory mutations and vice versa. Therefore, the juxtaposition of cytoplasm and nucleus from distantly related genera exposes those non-compensated mutations that lead to detrimental effects on those traits determined by nuclear-cytoplasmic interactions [4]. Hence, the replacement of the wheat cytoplasm with the cytoplasm of the closely related species *Aegilops* had a limited impact on gene expression and central metabolism. In contrast, the cytoplasm of the phylogenetically more distant species *H. chilense* altered the expression of functionally relevant genes and led to impaired cellular functioning, which in turn, activated stress-related processes at the metabolite level. Nevertheless, the changes at the transcript level of the lipid biosynthetic genes were not reflected in the abundance of fatty acid. This discordance in omics data is not exclusive to this study. In fact, an extensive review by Fernie and Stitt [65] postulated that quite often, there is little linear relation between transcriptomic and metabolomic data, implying a higher level of the regulation and structure of metabolic networks. We suggest that phylogenetic relatedness infers the degree of compatibility between nucleus and cytoplasm genomes, and in turn, the

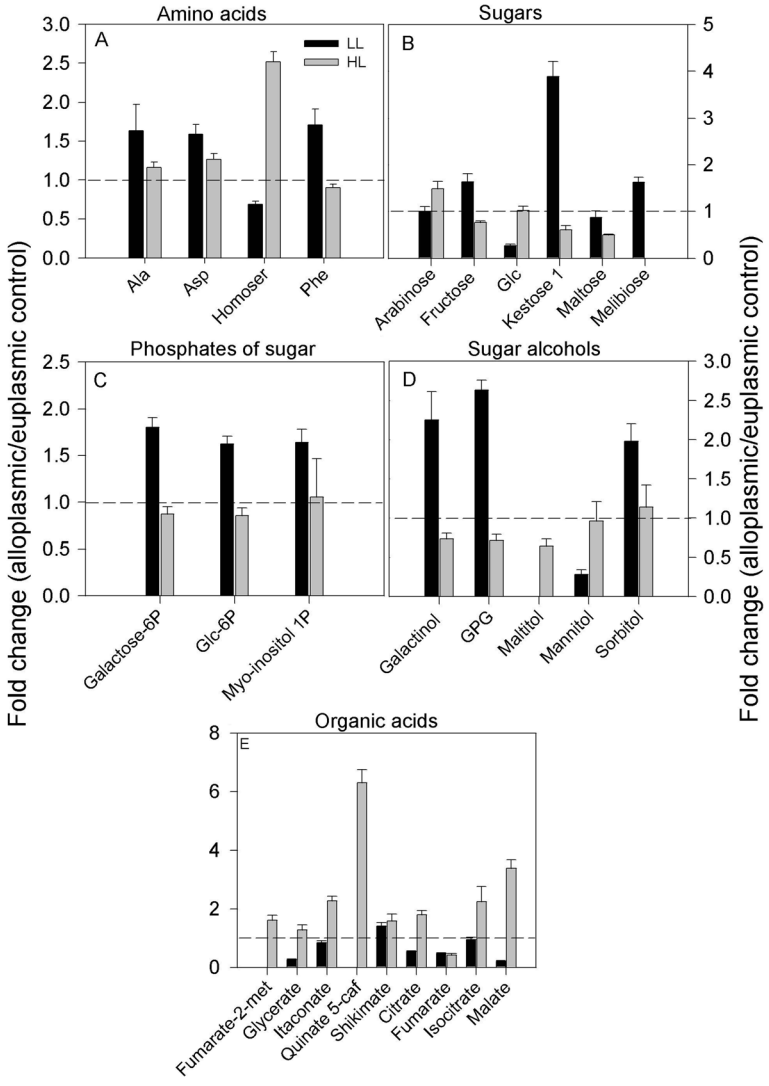


Figure 7 Metabolites altered in the T195 (*Ae. tauschii* cytoplasm) alloplasmic line. A-E denotes amino acids, sugars, sugar phosphates, sugar alcohols, and organic acids, respectively. Metabolites in each category that were significantly ($p \leq 0.05$) altered in either low light (LL, 150 μE m^{-2} s^{-1}) or high light (HL, 600 μE m^{-2} s^{-1}) in the mature leaf of the alloplasmic line T195 (*Ae. tauschii* cytoplasm) in comparison with the wheat cv. Chris used as the euplasmic control. Bars are fold changes of the alloplasmic line on the euplasmic control. The euplasmic control is given the threshold 1. Bars above or below the dashed line (threshold value 1) indicate an increase or a decrease of each metabolite of the alloplasmic genotype compared with the euplasmic control. Each bar is presented as means ± SEM of five biological repetitions.

co-evolution of nucleus and cytoplasm has shaped the regulatory mechanisms that control the cross-talk between the cellular genomes.

Specific signaling mechanisms are affected by cytoplasm substitution

The present study was carried out in the fully expanded leaves of plants grown at 600 μE m^{-2} s^{-1}, a condition that enables the highlighting of differences due to the activation of light-stress-associated processes known to be relevant for chloroplast and mitochondrion functionality. Since the cross-talk between nucleus and cytoplasm genomes is controlled by defined anterograde and retrograde signaling mechanisms [66], the alterations detected with transcriptomic and metabolomic analysis can be used to highlight specific signaling pathways sensitive to small variations in the cytoplasm genome.

The plastid transcripts undergo an extensive and complex maturation process that includes the splicing of introns and RNA editing which require specific proteins encoded exclusively in the nucleus. Most introns in organelles are group II introns, whose catalytic mechanism closely resembles that of the nuclear spliceosome. The plant nuclear genome contains four organelle maturase-

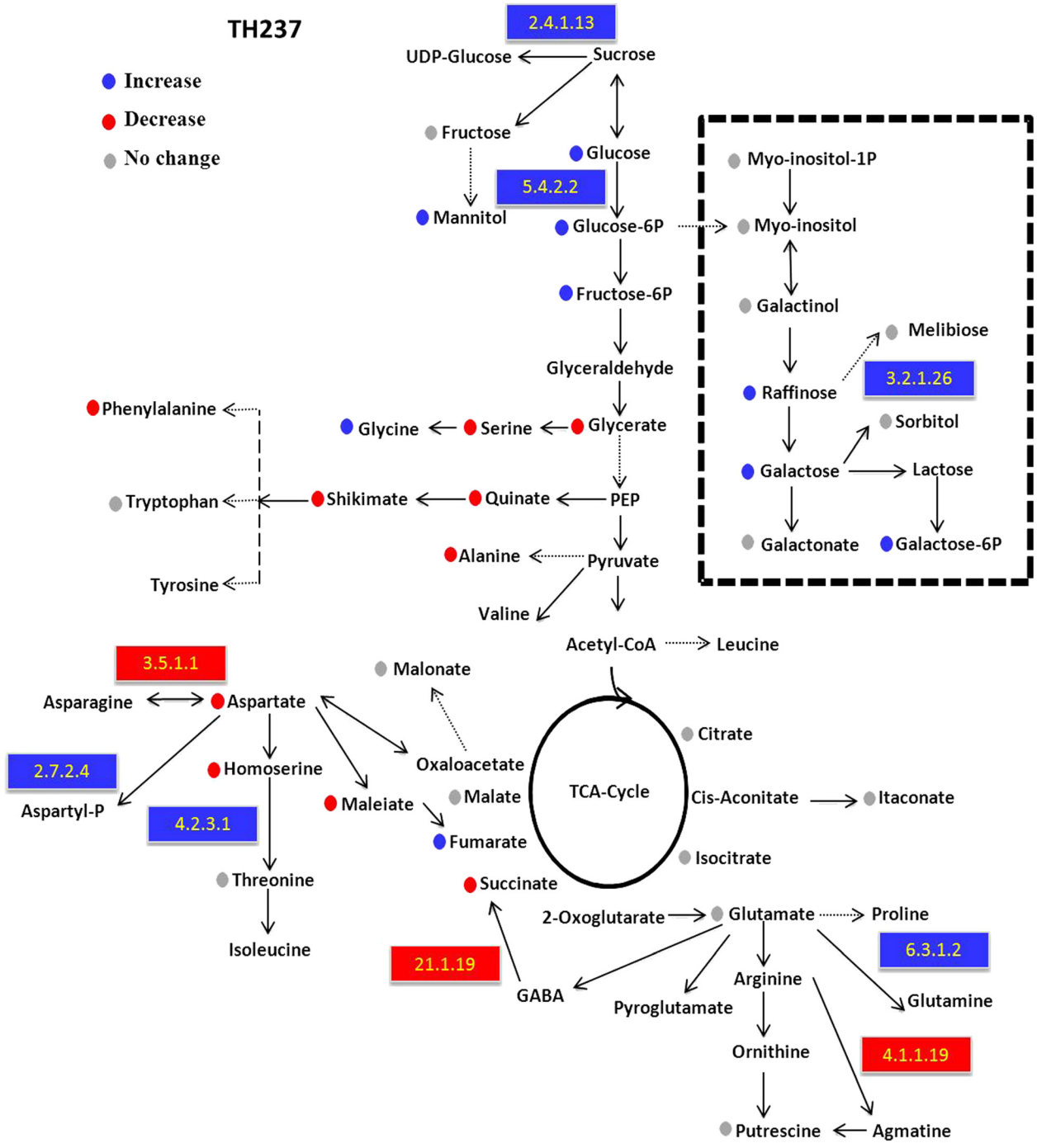


Figure 8 Metabolic pathway map for the alloplasmic genotype TH237 (*H. chilense* cytoplasm) subjected to high light treatment. Each data point is the fold change of the alloplasmic genotype on the euplasmic control. Circles and rectangles denote measured metabolites and metabolic genes, respectively. Blue, red and grey points indicate significant ($p \leq 0.05$) increases, decreases and no change of metabolite and enzyme levels, respectively. Metabolic genes measured are mapped and indicated with enzyme EC numbers: arginine decarboxylase (4.1.1.19), asparaginase (3.5.1.1), glutamine synthetase (6.3.1.3), aspartyl-phosphate (2.7.2.4), threonine synthase (4.2.3.1), sucrose synthase (2.4.1.13), phospho-glucomutase (5.4.2.2), and β-fructofuranosidase (3.2.1.26).

related genes, out of the context of their evolutionary-related group-II introns dedicated to the nuclear genes [67]. All the encoded proteins are targeted to mitochondria/chloroplast [28] and act *in-trans* in the splicing of organellar-encoded introns. The finding that two (*AtnMat3* and *AtnMat4*) out of four nuclear

maturase-related genes were up-regulated in all alloplasmic lines identifies the anterograde signaling that controls the RNA-maturation-related processes as one of the most sensitive to a perturbation of the nuclear-cytoplasmic interaction. Although no specific targets have been demonstrated so far for *AtnMat3* and *AtnMat4*, correlative evidence supporting a defective maturation process for some cytoplasmic-encoded genes can be found in the down-regulation of five chloroplast-encoded genes in all alloplasmic lines, a trend opposite to the maturase up-regulated expression.

Although no chloroplast retrograde pathway is well-understood mechanistically, several signals of various natures have been reported to trigger retrograde signaling from chloroplasts to nucleus, among them the signals generated by ROS accumulation. The different forms of ROS cause similar cellular damages, but each of them is involved in the activation of a specific signaling pathway [30]. High light conditions, as those employed in this work, are known to generate ROS, and consequently, the results from our study identifying a number of modifications can be linked to ROS-mediated retrograde signaling pathways. Light acclimation processes in plants act to dissipate the excess excitation energy and optimize photosynthesis under variable light intensity. Under high light conditions, the excitation energy might not be sufficiently quenched by the PSII reaction center or by carotenoids, promoting the formation of the triplet-excited state of the chlorophyll and leading to the generation of $^1O_2$ [68,69]. We found that the retrograde signaling pathway triggered by $^1O_2$ was specifically altered in the alloplasmic line carrying the *H. chilense* cytoplasm. Transcriptomic analysis revealed an up-regulation of *flu* and *ex2*, two of the key genes identified as essential and specific components of the $^1O_2$ signaling.

Flu is a negative regulator of chlorophyll biosynthesis that prevents the accumulation of intermediate that are potentially extremely destructive when illuminated, causing a massive release of $^1O_2$. The accumulation of $^1O_2$ leads to rapid and selective transcriptional reprogramming and finally induces programmed cell death in *flu* plants. The primary effect of $^1O_2$ generation in the Arabidopsis *flu* mutant is the activation of a broad range of signaling pathways known to be involved in biotic and abiotic stress responses [36,37]. *Flu* and *ex* are considered specific markers of the $^1O_2$ pathway; therefore, their up-regulation demonstrates that a change in the chloroplast genome has a direct impact on the chloroplast retrograde signaling controlled by $^1O_2$. The up-regulation of D1 can also be interpreted as a consequence of the production of $^1O_2$ which, in turn, promotes the irreversible oxidation of the D1 protein [68]. In cyanobacterial cells, singlet oxygen has been shown to cause direct photo-damage to PSII and the D1 protein [70,71] and to prevent PSII repair by suppressing the elongation of the D1 protein [72].

A number of *flu*-downstream genes were also found up-regulated in the *H. chilense* alloplasmic line. *LSD1* is a key gene controlling the cross-talk between ROS and SA-dependent pathways, leading to light acclimation and pathogen defense responses [44]. The up-regulation of the *LSD1* gene, a negative regulator of cell death, in the *H. chilense* alloplasmic line was associated with the altered expression of 26 probe sets involved in the hypersensitive response to biotic stress belonging to methyl jasmonate and salicylic acid pathways (Table 1) and with the down-regulation of several genes involved in programmed cell death. These results suggest that a change in the chloroplast genome has a direct impact on the chloroplast retrograde signaling that integrates light acclimation, cell death and immune defense responses, including genes engaged in ethylene, ROS, salicylic acid, glutathione, ABA, sugar and auxin signaling.

In line with the observation reported above, several significant metabolic alterations found in the *H. chilense* alloplasmic line can be associated with a typical response to high light mediated by the ROS signal [73]. It is known that soluble sugars, especially sucrose, glucose and fructose, are involved in the responses to a number of stresses, and they act as nutrient and metabolite signaling molecules that activate specific hormone-cross-talk transduction pathways, thus resulting in important modifications of gene expression and proteomic patterns. Various metabolic reactions and regulation directly link soluble sugars with the production rates of ROS, such as mitochondrial respiration and photosynthesis regulation and, conversely, with anti-oxidative processes. A metabolic analysis of the *H. chilense* alloplasmic line showed significant increases in sugars, sugar phosphates and mannitol, with a remarkable enhancement in raffinose abundance (33-fold) and of its downstream metabolite galactose. The accumulation of raffinose is in agreement with the up-regulation of two key enzymes (sucrose synthase and β-fructofuranosidase) of the raffinose metabolism (Figure 8). These oligosaccharides are associated with a number of stresses in plants [74] e.g., drought, salt, temperature [75] and light [76]. They function mainly as osmoprotectants, and raffinose is known to play a key role in ROS scavenging and the protection of the metabolism in chloroplasts from the oxidative damage caused by HL [77]. The increased abundance of a number of polyols could further support a limited response to light of the alloplasmic line [78,79]. For example, polyols are reported to play an important role in the protection of sensitive enzymes and membranes from ROS in the cytoplasm [80]. Taken together, these results suggest that, the *H. chilense* alloplasmic line is likely impaired in electron dissipation, and thus, exposure to light leads to enhanced ROS generation and the induction of

ROS-scavenging processes and the accumulation of membrane protective compounds.

Cytoplasm male sterility-associated genes are modulated in response to cytoplasm substitution

Although the alloplasmic lines here tested were all fully male fertile, the finding that some genes known to be involved in the determination of CMS have an altered expression due to the substitution of a wheat cytoplasm with the cytoplasm of *H. chilense* further demonstrates the involvement of the nuclear-cytoplasmic cross-talk in CMS, as already reported for CMS that originated after inter-species crosses [10,81,82]. It can be suggested that in the specific *H. chilense* alloplasmic line here employed (TH237), the mechanisms leading to CMS have been only partially modified without a phenotypic effect, while a more strong de-regulation of the genes here described (or of some of them) might lead to male sterility. Indeed, the male fertility of the *H. chilense* alloplasmic lines varies depending on the *H. chilense* accession used. Some *H. chilense* accessions give full fertility, while others lead to male sterility as was the case of the H1 accession [10]. The fertility restoration of CMS in the msH1 system is caused by genes located on chromosome $6H^{ch}$ [10,11]. Notably, ten probe sets, corresponding to genes modulated in the *H. chilense* alloplasmic line and related to CMS (*LOX2*, acyl CoA reductase, *CER4* and *POP2*), matched to *Brachypodium* genes from chromosome 3 in the region syntenic with barley chromosome 6H [83]. One of them (*LOX2*) also found a match in the barley genome zipper on chromosome 6H (http://mips.helmholtz-muenchen.de/plant/barley/gz/searchjsp/index.jsp). These genes may be good candidates by which to investigate the restoration of CMS in the msH1 system.

Conclusions

The replacement of the wheat cytoplasm with the cytoplasm of a related species affects the nuclear-cytoplasmic cross-talk leading to transcript and metabolite alterations. The alteration of the nuclear-cytoplasmic cross-talks might reflect an impaired light response capacity that exposes the alloplasmic lines to some degree of stress, as suggested by the altered expression of ROS-responsive genes and the accumulation of stress-related metabolites. The extent of these modifications was limited in the alloplasmic lines with the *Aegilops* cytoplasm, and more relevant in the alloplasmic line with the *H. chilense* cytoplasm. We consider that, this finding might be linked to the phylogenetic distance of the genomes.

Methods

Plant material and growth conditions

Three alloplasmic lines (T183, T195 and TH237) were used in this work. The alloplasmic lines T183 and T195 were developed through the introgression of the cytoplasm from *Aegilops uniaristata* (T183) and *Aegilops tauschii* (T195) in the nuclear background of *Triticum aestivum* cv. Chris by Prof. S.S. Maan (North Dakota State University, USA) after 10 backcrosses as described by Busch and Maan [23]. The alloplasmic line TH237 was produced by introgressing the *Hordeum chilense* accession H7 cytoplasm into the nuclear background of *T. aestivum* accession T20 [13]. The seeds of the alloplasmic lines were multiplied at IAS-CSIC and verified for cytoplasm identity with specific cytoplasm markers. T183 and T195 were checked using the cleaved amplified polymorphic sequence (CAPS) psbE & psbF-HpaII [64] that distinguishes between the *T. aestivum* and *Aegilops* cytoplasm, while TH237 was confirmed using a ccSSR marker [84] as described by Atienza et al. [13]. The alloplasmic lines have overall normal phenotypes that differ from their euplasmic control only for a few phenotypic traits (e.g., plant height) and yielding capacity [13].

Alloplasmic line TH237 was produced by transferring the *Hordeum chilense* accession H7 cytoplasm into the nuclear background of the *Triticum aestivum* accession T20 [13]. Since *H. chilense* and wheat chromosomes do not pair during meiosis, the alloplasmic line was obtained after two backcrosses as previously reported [13]. The euplasmic backgrounds were used as the control. Fifty seeds were sown in five biological replicates for each sample in pots of 11 cm diameter and grown in controlled conditions under two light intensities, 150 and 600 $\mu E\ m^{-2}\ s^{-1}$, denoted low and high light intensity, respectively, in a daily regime of 12 h light at 22°C and 12 h darkness at 15°C. Plants were bulked from each pot, and three biological replicates were used for the transcriptomics, while five biological replicates were used for the metabolic profiling. Fully expanded second leaves were collected two weeks from germination in the middle of the light period and frozen in liquid nitrogen until further analysis.

RNA isolation and array hybridization

RNA samples from the leaves of alloplasmic and the corresponding euplasmic plants grown under 600 $\mu E\ m^{-2}\ s^{-1}$ were prepared using a TRIZOL reagent according to the method published at the *Arabidopsis* Functional Genomics Consortium (http://www.arabidopsis.org/portals/masc/AFGC/RevisedAFGC/site2RnaL.htm#isolation) and further cleaned using RNeasy columns according to the Qiagen RNeasy Mini Handbook. RNA quantity and integrity were confirmed on an Agilent Bioanalyzer 2100 using the RNA 6000 nano Kit. RNA samples were processed following the AffymetrixGeneChip Expression Analysis Technical Manual (Affymetrix, Inc., Santa Clara, CA). Single-stranded, then double-stranded, cDNAs were synthesized from the

poly (A) mRNA isolated from 5 μg of total RNA for each sample using the Affymetrix One-Cycle Labelling Kit and Control reagents. The resulting *ds*-cDNA was column purified and then used as a template to generate biotin-tagged cRNA from an *in vitro* transcription reaction (IVT), using the Affymetrix Gene-Chip IVT Labelling Kit. The resulting biotin-tagged cRNA was fragmented and then hybridized at 45°C for 16 h (Affymetrix Gene-Chip Hybridization Oven 640) to the probe sets present on an Affymetrix Gene-Chip® Wheat Genome Array. The arrays were washed and then stained (SAPE, Streptavidin-phycoerythrin) on an Affymetrix Fluidics Station 450, followed by scanning with a Gene-Chip Scanner 3000.

Hybridization quality was verified using the standard Affymetrix controls, and all hybridizations showed the expected checkerboard pictures. The collected data were normalized using an RMA algorithm. The average background was 40.07, well within recommended levels. The percentage of "present" calls ranged between 42.9% and 46.9% of the 60 K probe sets of the array. To value the quality of biological replicates, R-squared was calculated between the replicates of the same sample, and the values ranged between 0.97 and 0.99 with an average value of 0.976.

Wheat microarray design and expression profiling data are available in Plexdb (http://www.plexdb.org) as experiment TA49: "alteration in transcript level in wheat alloplasmic lines".

Transcript data processing and analysis

Raw intensity values were normalized by RMA [85], using the R package Affymetrix library [86]. The same library was used to run the MAS 5.0 algorithm on raw data to produce a detection call for each probe set. These detection calls (P = "present", M = "marginal" or A = "absent") were used to apply an initial filtering step, since genes not expressed ("absent") represent experimental noise and can generate false positives. All probe sets that didn't show "present" calls in all three reps of at least one sample were removed. R-squared linear correlation coefficients were computed on the RMA expression values (log_2-transformed) for each set of biological triplicates. RMA-filtered data were imported to the Genespring GX 7.3 (Agilent Technologies) software for all subsequent analyses. Three comparisons were done to evaluate the impact of different cytoplasms on nuclear gene expression at 600 μE m^{-2} s^{-1}: *Ae. tauschii* (alloplasmic line) *vs.* Chris (euplasmic control), *Ae. uniaristata* (alloplasmic line) *vs.* Chris (euplasmic control), and *H. chilense* (alloplasmic line) *vs.* T20 (euplasmic control). Differentially expressed probe sets were identified through a Welch t-test with a Benjamini and Hochberg false discovery rate correction for multiple tests [87]. Differences in gene expression were considered to be significant when the p-value <0.05 and the induction or repression ratio was equal to or higher than two-fold. A Principal Component Analysis (PCA), a mathematical procedure that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components, is often used to analyse array and metabolomic data [25]. A PCA was carried out with Genespring GX 7.3 software and employed to assess the contribution of the genetic backgrounds and of the cytoplasms on the observed variations in gene expression. The array data were analysed using a two-fold change as the cut-off, followed by an analysis for statistically significant changes using a Welch t-test and a false discovery rate correction for multiple testing [87].

Blast searches were done using a HarvEST Affymetrix Wheat1 Chip 1.50 (http://www.harvest.ucr.edu), and only the annotations of wheat probe sets with a homology level cut-off equal to or lower than E-value = e^{-10} were considered. The functional classification was based on the wheat gene identifier categories (Affymetrix Gene-Chip Wheat probe Taes_0709) reported by the MapMan 3.1.0 software (http://gabi.rzpd.de/projects/MapMan/), a user-driven tool for displaying genomics datasets on diagrams of metabolic pathways and other biological processes [26]. The log_2 expression ratio of genes modulated in the comparison between alloplasmic line TH237 (*H. chilense* cytoplasm) and T20 euplasmic control were loaded into the MapMan Image Annotator Module to map the wheat transcriptome data into define functional categories and to display them onto pathway diagrams. In the color scale, red represents a lower gene expression and blue represents a higher gene expression in the *H. chilense* alloplasmic line, as compared with the corresponding T20 euplasmic. The complete set of genes submitted to the MapMan analysis is given in Additional file 1: Table S1.

Probe set descriptions, functional classification and mitochondrial/chloroplast localization were then verified and implemented manually and by other bioinformatics resources, such as MIPS *Arabidopsis thaliana* database (MAtDB) Functional Catalogue (FunCat) (http://mips.helmholtz-muenchen.de/proj/funcatDB) and Plexdb (plant expression database http://www.plexdb.org).

Validation of array data with real-time reverse transcription PCR (qRT-PCR)

An array validation was carried out with six genes (*AtnMat3*, *Executer2*, *Flu*, *psbM*, *FAR*, *LSD1*, Additional file 2: Table S2), selected because they are representative of key pathways altered in the alloplasmic lines. Total RNA (400 ng) was reverse transcribed using gene specific primers with SuperScript II (Invitrogen) according to the manufacturer's protocol. Subsequently, the cDNAs

were used for q-PCR amplifications with SYBR Green fluorescence detection. The reactions (final volume of 25 μl) were set with a QuantiFast SYBR Green PCR Master Mix (Qiagen) together with forward and reverse primers (10 mM; 2.5 μl each, see Additional file 2: Table S2) and amplified using standard cycling conditions: 95°C (10 min) and 40 cycles of amplification (95°C for 30 s and 60°C for 1 min). Expression was determined for three analytical replicates. The wheat polyubiquitin gene (Ta.24299.1.S1_at) was used as a reference.

Metabolic profiling using GC-MS-based method

Metabolite analyses from the leaves of alloplasmic and the corresponding euplasmic plants grown at 150 and 600 μE m^{-2} s^{-1} were carried out using a GC-MS-based protocol described by Lisec et al. [88] and Roesnner et al. [89] with a few modifications. Leaf tissues (100 mg) were homogenized using a ball mill pre-cooled with liquid nitrogen and extracted in 1400 μl of methanol, 60 μl of 0.2 mg ribitol ml^{-1} in methanol as an internal quantitative standard for the polar phase. The mixture was extracted for 15 min at 70°C, mixed vigorously, and centrifuged for 10 min at 14000 rpm. The supernatant was transferred to clean 2 ml tubes. In order to separate polar and non-polar metabolites, 750 μl chloroform and 1500 μl of water was added to the supernatant. After centrifugation at 4000 rpm for 15 min, 200 μl of the upper methanol/water phase was transferred to new tubes and reduced to dryness *in vacuo*. Dried samples were re-dissolved in 40 μl of 20 mg ml^{-1}methoxyaminehydrochloride in pyridine and derivatized for 120 min at 37°C followed by a 30 min treatment with 70 μl MSTFA (*N*-methyl-*N*-[trimethylsilyl] trifluoroacetamide) at 37°C and 8 μl of a retention time standard mixture (3.7% [w/v] hepatonic acid, 3.7% [w/v] nonanoic acid, 3.7% [w/v] undecanoic acid, 3.7% [w/v] tridecanoic acid, 3.7% [w/v] pentadecanic acid, 7.4% [w/v] nonadeanoic acid, 7.4% [w/v] tricosanoic acid, 22.2% [w/v] heptacosanoic acid, and 55.5% [w/v] hentriacontanoic acid dissolved in 10 mg ml^{-1}tetrahydrofuran) was added before trimethylsilylation. Sample volumes of 1 μl were then injected with a split ratio of 25:1, using a hot needle technique. The gas chromatography–mass spectrometry (GC-MS) system was composed of an AS 2000 auto-sampler, a GC 8000 gas chromatograph, and a Voyager quadrole mass spectrometer (Thermo-Quest, Manchester, UK). The mass spectrometer was tuned according to the manufacturer's recommendations, using tris-(perfluorobutyl)-amine (CF43). GC was performed on a 30-m SPB-50 column with 0.25-μm film thickness (Superlco, Bellfonte, CA). The injection temperature was set at 230°C, the interface at 250°C, and the ion source adjusted to 200°C. Helium was used as the carrier gas at a flow rate of 1 ml min^{-1}. The analysis was performed under the following temperature program: 5 min of isothermal heating at 70°C, followed by a 5°C min^{-1} oven temperature ramp to 310°C, and a final 1 min of heating at 310°C. The system was then temperature equilibrated for 6 min at 70°C before the injection of the next sample. Mass spectra were recorded at 2 scan sec^{-1} with a scanning range of 50 to 600 m/z. Data was analysed using Tagfinder software [90] and statistical analysis including an ANOVA, a student t-test and a principle component analysis that were performed using dedicated software, Microsoft Excel, R software environment (http://www.R-project.org; R Development Core Team, 2004) and TMEv (http://www.tm4.org/mev/) [91].

Data analysis using principal component analysis

Principal component analysis (PCA) visualizes the dispersion of metabolites/genes and the formation of groups of co-dispersed metabolites/genes. A PCA was performed with the software package TMEv [91] for metabolite data using the default weighted covariance estimation function. PCA is an unsupervised multivariate method that allows patterns, trends, groups and outliers in a large datasets to be easily identified. The dimensionality of complex data is reduced to what are called Principal Components (PCs) that retain the maximal amount of variation within a samples' population. For PC and average based statistical analysis, the data was $\log_{10}$ transformed and normalized to the median of the entire sample set for each metabolite before analysis. This transformation reduces the influence of outliers and increases symmetry in a dataset [25].

Availability of supporting data

The datasets supporting the results of this article are included within the article and its additional files.

Additional files

Additional file 1: Table S1. Probe sets detecting transcripts more than two-fold up-or down-regulated in the comparisons between each alloplasmic line and the corresponding euplasmic control.

Additional file 2: Table S2. Genes selected for array validation and the corresponding primers employed in the qRT-PCR analysis. Expression data detected with microarray and qRT-PCR analysis are reported for comparison.

Additional file 3: Table S3. Data of two-way ANOVA metabolites that significantly changed in the alloplasmic line TH237 (*H. chilense* cytoplasm) and its euplasmic control T20 due to genotype, light treatment and the interaction effect between them.

Additional file 4: Table S4. Data of two-way ANOVA metabolites that significantly changed in the alloplasmic lines T183 (*Ae. uniaristata* cytoplasm) and T195 (*Ae. tauschii* cytoplasm) and their euplasmic control (Chris) due to genotype, light treatment and the interaction effect between them.

Additional file 5: Table S5. Metabolite variations in the comparison between the alloplasmic line with the *H. chilense* cytoplasm (TH237) and the corresponding wheat euplasmic control (T20). Values expressed as fold change for plants grown at low light and high light.

Additional file 6: Table S6. Metabolite variations in the comparison between the alloplasmic lines with *Ae. uniaristata* (T183) or *Ae. tauschii* (T195) and the corresponding wheat euplasmic control (Chris). Values expressed as fold change for plants grown at low light and high light.

Additional file 7: Table S7. Fatty acids measured in the alloplasmic line TH237 (*H. chilense* cytoplasm) and its euplasmic control T20. Data was subjected to a student t-test to test for fatty acids that were significant ($p \leq 0.05$) between the alloplasmic line and its euplasmic control.

Abbreviations

CMS: Cytoplasmic male sterility; HL: High light; LL: Low light,$^1O_2$, singlet oxygen; PCA: Principal component analysis; ROS: Reactive oxygen species; SAR: Systemic acquired resistance.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CC and LG performed the experiment and carried out all gene expression analyses; CC and CM carried out the bioinformatics analysis and interpreted the array results. LQ carried out and interpreted the metabolic analysis. ER carried out the array hybridization. SA, LC and AF participated in the design of the study. CC, LQ, SA, LC and AF wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors thank Prof. S.S. Maan (North Dakota State University, USA) for providing the alloplasmic lines T183 and T195 and Dr. Alessio Aprile (University of Salento, Lecce, Italy) for help in data analysis. LQ was partly supported by grant # 2770367, the Chief Scientist Fund, the Ministry of Agriculture, Israel.

Author details

[1]Jacob Blaustein Institutes for Desert Research, French Associates Institute for Agriculture and Biotechnology of Drylands, Ben-Gurion University of the Negev, Midreshet Ben-Gurion, 84990, Sde Boqer, Israel. [2]Consiglio per la Ricerca e la Sperimentazione in Agricoltura -Genomics Research Centre, via S. Protaso 302, 29017, Fiorenzuola d' Arda, (PC), Italy. [3]Departamento de Mejora Genética, IAS-CSIC, Apdo. 4084, Cordoba 14980, Spain. [4]Center for Genome Research, Biomedical Sciences Department, Biological Chemistry Section, University of Modena and Reggio Emilia, Via G. Campi 287, 41125, Modena, Italy.

Received: 25 June 2013 Accepted: 6 December 2013
Published: 10 December 2013

References

1. Hanson MR, Bentolila S: **Interactions of mitochondrial and nuclear genes that affect male gametophyte development.** *Plant Cell* 2004, **16**(Suppl): S154–S169.
2. Rodríguez-Suárez C, Giménez MJ, Ramírez MC, Martín AC, Gutierrez N, Ávila CM, Martín A, Atienza SG: **Exploitation of nuclear and cytoplasm variability in *Hordeum chilense* for wheat breeding.** *Plant Genetic Res* 2011, **9:**313–316.
3. Tsunewaki K: **Plasmon analysis in the *Triticum-Aegilops* complex.** *Breed Sci* 2009, **59:**455–470.
4. Allen JO: **Effect of teosinte cytoplasmic genomes on maize phenotype.** *Genetics* 2005, **169:**863–880.
5. Tsunewaki K, Wang GZ, Matsuoka Y: **Plasmon analysis of *Triticum* (wheat) and *Aegilops*. 1. Production of alloplasmic common wheats and their fertilities.** *Genes Genet Syst* 1996, **71:**293–311.
6. Tsunewaki K, Wang G-Z, Matsuoka Y: **Plasmon analysis of *Triticum* (wheat) and *Aegilops*. 2. Characterization and classification of 47 plasmons based on their effects on common wheat phenotype.** *Genes Genet Syst* 2002, **77:**409–427.
7. McBride KE, Svab Z, Schaaf DJ, Hogan PS, Stalker DM, Maliga P: **Amplification of a chimeric *Bacillus* gene in chloroplasts leads to an extraordinary level of an insecticidal protein in tobacco.** *Biotechnology (NY)* 1995, **13:**362–365.
8. Kihara H: **Substitution of nucleus and its effects on genome manifestations.** *Cytologia* 1951, **16:**177–193.
9. Von Bothmer R, Giles BE, Jacobsen N: **Crosses and genome relationship in the *Hordeum patagonicum* group.** *Genetica* 1986, **71:**75–80.
10. Martín AC, Atienza SG, Ramírez MC, Barro F, Martín A: **Male fertility restoration of wheat in *Hordeum chilense* cytoplasm is associated with 6HchS chromosome addition.** *Aust J Agric Res* 2008, **59:**206–213.
11. MartinAC ASG, Ramirez MC, Barro F, Martin A: **Chromosome engineering in wheat to restore male fertility in the msH1 CMS system.** *Mol Breed* 2009, **24:**397–408.
12. Martín AC, Atienza SG, Ramírez MC, Barro F, Martín A: **Molecular and cytological characterization of an extra acrocentric chromosome that restores male fertility of wheat in the msH1 CMS system.** *Theor Appl Genet* 2010, **121:**1093–1101.
13. Atienza SG, Martín AC, Ramírez MC, Martín A, Ballesteros J: **Effects of *Hordeum chilense* cytoplasm on agronomic traits in common wheat.** *Plant Breed* 2007, **126:**5–8.
14. Fujii S, Toriyama K: **Genome barriers between nuclei and mitochondria exemplified by cytoplasmic male sterility.** *Plant Cell Physiol* 2008, **49:**1484–1494.
15. Li S, Wan C, Hu C, Gao F, Huang Q, Wang K, Wang T, Zhu Y: **Mitochondrial mutation impairs cytoplasmic male sterility rice in response to H_2O_2 stress.** *Plant Sci* 2012, **195:**143–150.
16. Bassene J-B, Froelicher Y, Navarro L, Ollitrault P, Ancillo G: **Influence of mitochondria on gene expression in a citrus cybrid.** *Plant Cell Rep* 2011, **30:**1077–1085.
17. Carlsson J, Lagercrantz U, Sundström J, Teixeira R, Wellmer F, Meyerowitz EM, Glimelius K: **Microarray analysis reveals altered expression of a large number of nuclear genes in developing cytoplasmic male sterile *Brassica napus* flowers.** *Plant J* 2007, **49:**452–462.
18. Leino M, Landgren M, Glimelius K: **Alloplasmic effects on mitochondrial transcriptional activity and RNA turnover result in accumulated transcripts of Arabidopsis orfs in cytoplasmic male-sterile *Brassica napus*.** *Plant J* 2005, **42:**469–480.
19. Fujii S, Yamada M, Fujita M, Itabashi E, Hamada K, Yano K, Kurata N, Toriyama K: **Cytoplasmic-nuclear genomic barriers in rice pollen development revealed by comparison of global gene expression profiles among five independent cytoplasmic male sterile lines.** *Plant Cell Physiol* 2010, **51:**610–620.
20. Yang J, Liu X, Xu B, Zhao N, Yang X, Zhang M: **Identification of miRNAs and their targets using high-throughput sequencing and degradome analysis in cytoplasmic male-sterile and its maintainer fertile lines of *Brassica juncea*.** *BMC Genomics* 2013, **14:**9.
21. Kumar P, Vasupalli N, Srinivasan R, Bhat SR: **An evolutionarily conserved mitochondrial orf108 is associated with cytoplasmic male sterility in different alloplasmic lines of *Brassica juncea* and induces male sterility in transgenic *Arabidopsis thaliana*.** *J Exp Bot* 2012, **63:**2921–2932.
22. Nott A, Jung H-S, Koussevitzky S, Chory J: **Plastid- to- nucleus retrograde signaling.** *Annu Rev Plant Biol* 2006, **57:**739–759.
23. Busch RH, Maan SS: **Effects of alien cytoplasms on agronomic and bread-making traits of two spring wheat cultivars.** *Crop Sci* 1978, **18:**864–866.
24. Atienza SG, Martín A, Pecchioni N, Platani C, Cattivelli L: **The nuclear–cytoplasmic interaction controls carotenoid content in wheat.** *Euphytica* 2008, **159:**325–331.
25. Yeung KY, Ruzzo WL: **Principal component analysis for clustering gene expression data.** *Bioinformatics* 2001, **17:**763–774.
26. Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M: **MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes.** *Plant J* 2004, **37:**914–939.
27. Bräutigam K, Dietzel L, Pfannschmidt T: **Plastid-nucleus communication: anterograde and retrograde signalling in the development and function of plastids.** *Cell Mol Biol Plastids* 2007, **19:**409–455.
28. Keren I, Bezawork-Geleta A, Kolton M, Maayan I, Belausov E, Levy M, Mett A, Gidoni D, Shaya F, Ostersetzer-Biran O: ***AtnMat2*, a nuclear-encoded maturase required for splicing of group-II introns in Arabidopsis mitochondria.** *RNA* 2009, **15:**2299–2311.

29. Zoschke R, Nakamura M, Liere K, Sugiura M, Börner T, Schmitz-Linneweber C: **An organellar maturase associates with multiple group II introns.** *Proc Natl Acad Sci U S A* 2010, **107:**3245–3250.
30. Fernández AP, Strand A: **Retrograde signaling and plant stress: plastid signals initiate cellular stress responses.** *Curr Opin Plant Biol* 2008, **11:**509–513.
31. Kropat J, Oster U, Rüdiger W, Beck CF: **Chloroplast signalling in the light induction of nuclear HSP70 genes requires the accumulation of chlorophyll precursors and their accessibility to cytoplasm/nucleus.** *Plant J* 2000, **24:**523–531.
32. Foreman J, Demidchik V, Bothwell JHF, Mylona P, Miedema H, Torres MA, Linstead P, Costa S, Brownlee C, Jones JDG, Davies JM, Dolan L: **Reactive oxygen species produced by NADPH oxidase regulate plant cell growth.** *Nature* 2003, **422:**442–446.
33. Dat J, Vandenabeele S, Vranová E, Van Montagu M, Inzé D, Van Breusegem F: **Dual action of the active oxygen species during plant stress responses.** *Cell Mol Life Sci* 2000, **57:**779–795.
34. Gapper C, Dolan L: **Control of plant development by reactive oxygen species.** *Plant Physiol* 2006, **141:**341–345.
35. Op den Camp RGL, Przybyla D, Ochsenbein C, Laloi C, Kim C, Danon A, Wagner D, Hideg E, Göbel C, Feussner I, Nater M, Apel K: **Rapid induction of distinct stress responses after the release of singlet oxygen in Arabidopsis.** *Plant Cell* 2003, **15:**2320–2332.
36. Wagner D, Przybyla D, Op den Camp R, Kim C, Landgraf F, Lee KP, Würsch M, Laloi C, Nater M, Hideg E, Apel K: **The genetic basis of singlet oxygen-induced stress responses of *Arabidopsis thaliana*.** *Science* 2004, **306:**1183–1185.
37. Lee KP, Kim C, Landgraf F, Apel K: **EXECUTER1- and EXECUTER2-dependent transfer of stress-related signals from the plastid to the nucleus of Arabidopsis thaliana.** *Proc Natl Acad Sci U S A* 2007, **104:**10270–10275.
38. Mateo A, Mühlenbock P, Rustérucci C, Chang CC-C, Miszalski Z, Karpinska B, Parker JE, Mullineaux PM, Karpinski S: **LESION SIMULATING DISEASE 1 is required for acclimation to conditions that promote excess excitation energy.** *Plant Physiol* 2004, **136:**2818–2830.
39. Tanaka R, Hirashima M, Satoh S, Tanaka A: **The *Arabidopsis-accelerated cell death* gene *ACD1* is involved in oxygenation of pheophorbide *a*: inhibition of the pheophorbide *a* oxygenase activity does not lead to the "Stay-Green" phenotype in *Arabidopsis*.** *Plant Cell Physiol* 2003, **44:**1266–1274.
40. Norholm MHH, Nour-Eldin HH, Brodersen P, Mundy J, Halkier BA: **Expression of the *Arabidopsis* high-affinity hexose transporter STP13 correlates with programmed cell death.** *FEBS Lett* 2006, **580:**2381–2387.
41. Hyung JH, Ha CM, Nam HG: **Involvement of the *VEP1* gene in vascular strand development in *Arabidopsis thaliana*.** *Plant Cell Physiol* 2002, **43:**323–330. 41.
42. Coll NS, Vercammen D, Smidler A, Clover C, Van Breusegem F, Dangl JL, Epple P: ***Arabidopsis* type I metacaspases control cell death.** *Science* 2010, **330:**1393–1396.
43. Meskauskiene R, Nater M, Goslings D, Kessler F, op den Camp R, Apel K: **FLU: a negative regulator of chlorophyll biosynthesis in *Arabidopsis thaliana*.** *Proc Natl Acad Sci U S A* 2001, **98:**12826–12831.
44. Karpiński S, Szechyńska-Hebda M, Wituszyńska W, Burdiak P: **Light acclimation, retrograde signalling, cell death and immune defences in plants.** *Plant Cell Environ* 2013, **36:**736–744.
45. Granlund I, Storm P, Schubert M, García-Cerdán JG, Funk C, Schröder WP: **The TL29 protein is lumen located, associated with PSII and not an ascorbate peroxidase.** *Plant Cell Physiol* 2009, **50:**1898–1910.
46. Millar AH, Whelan J, Soole KL, Day DA: **Organization and regulation of mitochondrial respiration in plants.** *Ann Rev Plant Biol* 2011, **62:**79–104.
47. Attallah CV, Welchen E, Pujol C, Bonnard G, Gonzalez DH: **Characterization of *Arabidopsis thaliana* genes encoding functional homologues of the yeast metal chaperone Cox19p, involved in cytochrome c oxidase biogenesis.** *Plant Mol Biol* 2007, **65:**343–355.
48. Caldelari D, Wang G, Farmer EE, Dong X: **Arabidopsis *lox3 lox4* double mutants are male sterile and defective in global proliferative arrest.** *Plant Mol Biol* 2011, **75:**25–33.
49. Siedow JN: **Plant lipoxygenase: structure and function.** *Annu Rev Plant Physiol Plant Mol Biol* 1991, **42:**145–188.
50. Feussner I, Kühn H, Wasternack C: **Lipoxygenase-dependent degradation of storage lipids.** *Trends Plant Sci* 2001, **6:**268–273.
51. Porta H, Rocha-Sosa M: **Plant lipoxygenases. Physiological and molecular features.** *Plant Physiol* 2002, **130:**15–21.
52. Ariizumi T, Hatakeyama K, Hinata K, Inatsugi R, Nishida I, Sato S, Kato T, Tabata S, Toriyama K: **Disruption of the novel plant protein NEF1 affects lipid accumulation in the plastids of the tapetum and exine formation of pollen, resulting in male sterility in *Arabidopsis thaliana*.** *Plant J* 2004, **39:**170–181.
53. Yui R, Iketani S, Mikami T, Kubo T: **Antisense inhibition of mitochondrial pyruvate dehydrogenase E1alpha subunit in anther tapetum causes male sterility.** *Plant J* 2003, **34:**57–66.
54. Doan TTP, Carlsson AS, Hamberg M, Bülow L, Stymne S, Olsson P: **Functional expression of five Arabidopsis fatty acyl-CoA reductase genes in *Escherichia coli*.** *J Plant Physiol* 2009, **166:**787–796.
55. Aarts MG, Hodge R, Kalantidis K, Florack D, Wilson ZA, Mulligan BJ, Stiekema WJ, Scott R, Pereira A: **The Arabidopsis MALE STERILITY 2 protein shares similarity with reductases in elongation/condensation complexes.** *Plant J* 1997, **12:**615–623.
56. Palanivelu R, Brass L, Edlund AF, Preuss D: **Pollen tube growth and guidance is regulated by POP2, an Arabidopsis gene that controls GABA levels.** *Cell* 2003, **114:**47–59.
57. Fait A, Fromm H, Walter D, Galili G, Fernie AR: **Highway or byway: the metabolic role of the GABA shunt in plants.** *Trends Plant Sci* 2008, **13:**14–19.
58. Kusano T, Berberich T, Tateda C, Takahashi Y: **Polyamines: essential factors for growth and survival.** *Planta* 2008, **228:**367–381.
59. Capell T, Bassie L, Christou P: **Modulation of the polyamine biosynthetic pathway in transgenic rice confers tolerance to drought stress.** *Proc Natl Acad Sci U S A* 2004, **101:**9909–9914.
60. Aubert S, Curien G, Bligny R, Gout E, Douce R: **Transport, compartmentation, and metabolism of homoserine in higher plant cells. Carbon-13- and phosphorus-31-nuclear magnetic resonance studies.** *Plant Physiol* 1998, **116:**547–557.
61. Ishitani M, Majumder AL, Bornhouser A, Michalowski CB, Jensen RG, Bohnert HJ: **Coordinate transcriptional induction of myo-inositol metabolism during environmental stress.** *Plant J* 1996, **9:**537–548.
62. Fujii S, Hellebust JA: **Growth and osmoregulation of *Boekelovia hooglandii* in relation to salinity.** *Can J Bot* 1994, **72:**823–828.
63. Golovina KA, Glushkov SA, Blinov AG, Mayorov VI, Adkison LR, Goncharov NP: **Molecular phylogeny of the genus *Triticum* L.** *Plant Syst Evol* 2007, **264:**195–216. 63.
64. Haider N: **Evidence for the origin of the B genome of bread wheat based on chloroplast DNA.** *Turk J Agric Forestry* 2012, **36:**13–25.
65. Fernie AR, Stitt M: **On the discordance of metabolomics with proteomics and transcriptomics: coping with increasing complexity in logic, chemistry, and network interactions scientific correspondence.** *Plant Physiol* 2012, **158:**1139–1145.
66. Woodson JD, Chory J: **Coordination of gene expression between organellar and nuclear genomes.** *Nat Rev Genet* 2008, **9:**383–395.
67. De Longevialle AF, Small ID, Lurin C: **Nuclearly encoded splicing factors implicated in RNA splicing in higher plant organelles.** *Mol Plant* 2010, **3:**691–705.
68. Krieger-Liszkay A: **Singlet oxygen production in photosynthesis.** *J Exp Bot* 2005, **56:**337–346.
69. Mullineaux PM, Baker NR: **Oxidative stress: antagonistic signaling for acclimation or cell death?** *Plant Physiol* 2010, **154:**521–525.
70. Nogushi T: **Dual role of triplet localization on the accessory chlorophyll in the photosystem II reaction center: photoprotection and photodamage of the D1 protein.** *Plant Cell Physiol* 2002, **43:**1112–1116.
71. Lupinkova L, Komenda J: **Oxidative modifications of the photosystem II D1 protein by reactive oxygen species: from isolated protein to cyanobacterial cells.** *Photochem Photobiol* 2004, **79:**152–162.
72. Nishiyama Y, Allakhverdiev SI, Yamamoto H, Hayashi H, Murata N, Nishiyama Y: **Singlet oxygen inhibits the repair of photosystem II by suppressing the translation elongation of the D1 protein in Synechocystis sp. PCC 6803.** *Biochemistry* 2004, **43:**11321–11330.
73. Gordon MJ, Carmody M, Albrecht V, Pogson B: **Systemic and local responses to repeated HL stress-induced retrograde signaling in Arabidopsis.** *Front Plant Sci* 2012, **3:**303.
74. Bartels D, Sunkar R: **Drought and salt tolerance in plants.** *Crit Rev Plant Sci* 2005, **24:**23–58.

75. Krasenski J, Jonak C: **Drought, salt and temperature stress-induced metabolic rearrangements and regulatory networks.** *J Exp Bot* 2012, **63:**1593–1608.
76. Jänkänpää HJ, Mishra Y, Schröder WP, Jansson S: **Metabolic profiling reveals metabolic shifts in Arabidopsis plants grown under different light conditions.** *Plant Cell Environ* 2012, **35:**1824–1836.
77. Nishizawa A, Yabuta Y, Shigeoka S: **Galactinol and raffinose constitute a novel function to protect plants from oxidative damage.** *Plant Physiol* 2008, **147:**1251–1263.
78. Vinocur B, Altman A: **Recent advances in engineering plant tolerance to abiotic stress: achievements and limitations.** *Curr Op Biotech* 2005, **16:**123–132.
79. Abebe T, Guenzi AC, Martin B, Cushman JC: **Tolerance of mannitol-accumulating transgenic wheat to water stress and salinity.** *Plant Physiol* 2003, **131:**1748–1755.
80. Sheveleva E, Chmara W, Bohnert HJ, Jensen RG: **Increased salt and drought tolerance by D-Ononitol production in transgenic *Nicotiana tabacum* L.** *Plant Physiol* 1997, **115:**1211–1219.
81. Zubko MK, Zubko EI, Ruban AV, Adler K, Mock HP, Misera S, Gleba YY, Grimm B: **Extensive developmental and metabolic alterations in cybrids *Nicotiana tabacum* (+ *Hyoscyamus niger*) are caused by complex nucleo-cytoplasmic incompatibility.** *Plant J* 2001, **25:**627–639.
82. Teixeira RT, Farbos I, Glimelius K: **Expression levels of meristem identity and homeotic genes are modified by nuclear-mitochondrial interactions in alloplasmic male-sterile lines of *Brassica napus*.** *Plant J* 2005, **42:**731–742.
83. Mayer KFX, Martis M, Hedley PE, Simková H, Liu H, Morris JA, Steuernagel B, Taudien S, Roessner S, Gundlach H, Kubaláková M, Suchánková P, Murat F, Felder M, Nussbaumer T, Graner A, Salse J, Endo T, Sakai H, Tanaka T, Itoh T, Sato K, Platzer M, Matsumoto T, Scholz U, Dolezel J, Waugh R, Stein N: **Unlocking the barley genome by chromosomal and comparative genomics.** *Plant Cell* 2011, **23:**1249–1263.
84. Chung S-M, Staub JE: **The development and evaluation of consensus chloroplast primer pairs that possess highly variable sequence regions in a diverse array of plant taxa.** *Theor Appl Genet* 2003, **107:**757–767.
85. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP: **Exploitation, normalization and summaries of high density oligonucleotide array probe level data.** *Biostatistics* 2003, **4:**249–264.
86. Gautier L, Cope L, Bolstad BM, Irizarry RA: **Affy-analysis of Affymetrix Gene-Chip data at the probe level.** *Bioinformatics* 2004, **20:**307–315.
87. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J Royal Stat Soc Series B (Methodological)* 1995, **57:**289–300.
88. Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR: **Gas chromatography mass spectrometry-based metabolite profiling in plants.** *Nat Protoc* 2006, **1:**387–396.
89. Roessner U, Luedemann A, Brust D, Feihn O, Linke T, Willmitzer L, Fernie R: **Metabolic profiling and phenotyping of genetically and environmentally modified systems.** *Plant Cell* 2001, **13:**11–29.
90. Luedemann A, Strassburg K, Erban A, Kopka J: **TagFinder for the quantitative analysis of gas chromatography–mass spectrometry (GC-MS)-based metabolite profiling experiments.** *Bioinformatics* 2008, **24:**732–737.
91. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J: **TM4: a free, open-source system for microarray data management and analysis.** *Biotechniques* 2003, **34:**374–378.

doi:10.1186/1471-2164-14-868
Cite this article as: Crosatti *et al.*: Cytoplasmic genome substitution in wheat affects the nuclear-cytoplasmic cross-talk leading to transcript and metabolite alterations. *BMC Genomics* 2013 **14**:868.