Unraveling transcriptional and translational control in plant energy homeostasis:

A BIOINFORMATIC APPROACH

Alessia Peviani

Alessia Peviani (2016)
Unraveling transcriptional and translational control in plant energy homeostasis: a bioinformatic approach.
PhD thesis, Utrecht University
Cover and layout by Alessia Peviani
Printed by: PrintSupport4U (www.printsupport4u.nl)
ISBN: 978-90-393-6584-7

Unraveling transcriptional and translational control in plant energy homeostasis: a bioinformatic approach

Onderzoek naar transcriptionele en translationele controle van energie homeostase in planten: een bioinformatische studie

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. G. J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op vrijdag 8 juli 2016 des middags te 4.15 uur

door

Alessia Peviani

geboren op 23 april 1987 te Reggio Emilia, Italië

Promotoren:	Prof. dr. B. Snel Prof. dr. J.C.M. Smeekens
Copromotor:	Dr. S.J. Hanson

The work presented in this thesis was financially supported by the European Union via the MERIT (MEtabolic Reprogramming by Induction of Transcription) Marie Curie Initial Training Network, grant agreement number 264474 under call FP7-PEOPLE-2010-ITN.

Table of Contents

1

General Introduction

partially adapted from:

The low energy signaling network

Filipa Tomé[1], Thomas Nägele[2], Mattia Adamo[3], Abhroop Garg[4], Carles Marco-llorca[4], Ella Nukarinen[2], Lorenzo Pedrotti[5], **Alessia Peviani**[6], Andrea Simeunovic[2], Anna Tatkiewicz[7], Monika Tomar[8], and Magdalena Gamm[8]

[1]*Bayer CropScience NV, Innovation Center, Ghent, Belgium*

[2]*Department of Ecogenomics and Systems Biology, University of Vienna, Vienna, Austria*

[3]*Instituto Gulbenkian de Ciência, Oeiras, Portugal*

[4]*Zentrum für Molekularbiologie der Pflanzen, Eberhard Karls Universität Tübingen, Tübingen, Germany*

[5]*Julius-von-Sachs-Institut, Julius-Maximilians-Universität Würzburg, Würzburg, Germany*

[6]*Theoretical Biology and Bioinformatics, Utrecht University, Utrecht, The Netherlands*

[7]*Universidad Politécnica de Madrid - Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, Centro de Biotecnología y Genómica de Plantas, Madrid, Spain*

[8]*Molecular Plant Physiology, Institute of Environmental Biology, Utrecht University, Utrecht, The Netherlands*

Frontiers in Plant Science 5, 353 (2014)

Controlling energy homeostasis is a challenge for all organisms, as they must constantly sense and integrate internal and external signals to optimize growth and development, often under suboptimal conditions (Polge and Thomas 2007). This is particularly critical for plants which, being sessile organisms, are exposed to a changing environment without possibility of escaping or sheltering themselves. Unraveling the molecular regulatory systems responsible for plant energy homeostasis is an important first step to achieve significant advances in agricultural production, necessary to secure food resources for a growing world population subjected to the whims of climate change.

Plants evolved complex signaling networks to finely coordinate energy consumption and preservation in all tissues (Baena-González et al. 2007; Baena-González 2010), which reflects into a distinctive regulatory genes repertoire when compared to other eukaryotes (Riechmann et al. 2000). Particularly during the evolution of land plants, several new families of transcription factors emerged, while others experienced a dramatic expansion, allowing for a more complex morphogenesis and adaptation to new environments (Shiu et al. 2005; Lang et al. 2010). In addition, post-transcriptional, translational, and post-translational mechanisms provide further levels in the fine-tuning of energy homeostasis. However, even for long-known and intensively studied processes much of the molecular effectors and mechanistic details are far from fully characterized, making necessary continuing efforts in fundamental research.

Especially in recent years, more accessible high-throughput techniques have led to the production of vast amounts of molecular plant research data, such as full genomic sequences, transcript and metabolite levels in different experimental settings, transcription factors DNA-binding sites, protein-protein interaction affinities, and more. The concurrent growth of online repositories created the opportunity for the integrated computational analysis of a variety of biological data, making bioinformatic methods essential to extract higher-level information from independent experiments. This is reflected in the continuing development of computational tools for data visualization, analysis, and integration, including several specific to the plant biology domain (Mochida and Shinozaki 2011). One of the most significant advances is represented by the accessibility of newly sequenced plant genomes, which makes it possible to explore the evolution and conservation of gene families in much greater depth than previously possible.

The work presented in this thesis investigates different energy-responsive systems in plants using a variety of bioinformatic approaches. The next two chapters focus on transcriptional control, investigating the evolutionary history of sugar-regulated transcription factors involved in seed development and stress response. In the last research chapter we move to translational control, analyzing the sequence features affecting the translational efficiency of certain transcripts during seed germination and sucrose treatment. The following paragraphs introduce the regulatory systems and bioinformatic approaches described in later chapters, to make the reader familiar with the content of this thesis.

TRANSCRIPTIONAL CONTROL OF ENERGY HOMEOSTASIS

A tight regulation of energy homeostasis appears crucial at almost every stage of the plant life cycle: for instance during seed germination, when seedling establishment relies on the limited nutrients embedded within the seed itself; or when severe stress conditions alter the balance between source (energy-producing) and sink (storage or energy-consuming) tissues. More specifically, in plants exposed to stress, sink organs such as seeds or tubers often suffer from reduced sugar import and are impaired in biomass production (Pinheiro et al. 2001; Cuellar-Ortiz et al. 2008). Accordingly, stress factors like nutrient limitation, hypoxia, excess of salt, and low or high temperatures have been shown to impair fruit and seed development by interfering with the source-sink balance (Gibon et al. 2002; Geigenberger 2003; Bailey-Serres et al. 2012; Lemoine et al. 2013; Liu, Chen, et al. 2013). Furthermore, stress conditions often affect photosynthesis and respiration in source leaves, which can further accentuate the source-sink imbalance. The resulting energy deprivation was suggested to be common to most types of stress and to trigger specific responses (Baena-González and Sheen 2008), leading to the massive alteration of plant cellular processes. The effects include growth arrest and metabolic reprogramming, comprising the repression of biosynthetic activities and sugar storage, as well as the induction of photosynthesis, sugar remobilization, and catabolic processes (Paul and Pellny 2003; Contento et al. 2004; Thimm et al. 2004; Gibon et al. 2006).

At the transcriptional level, alterations in the cellular energy status correlate with dramatic changes in gene expression patterns. Sugars specifically act not only as energy substrates, but also as signaling molecules with an active role in stress adaptation and other physiological processes (Eveland and Jackson 2012; Lastdrager et al. 2014). Accordingly, experiments have shown that artificial changes in the concentration of various sugars can induce or repress the activity of several regulatory proteins, triggering specific transcriptional programs (Price et al. 2004; Blasing et al. 2005; Osuna et al. 2007; Usadel et al. 2008). Sugars also cross-talk with phytohormone signaling to modulate critical developmental processes, such as embryo establishment, seed germination, and seedling growth (Eveland and Jackson 2012); accordingly, mutant plants defective in phyotohormone signaling components tend to display altered sugar response phenotypes, and *vice versa*, indicating an extensive overlap between these signaling pathways.

In the following paragraphs we discuss two families of transcription factors regulated by sugars that are involved in the regulation of energy homeostasis during seed germination and stress response.

The transcription factor ABI4

ABI4 role in ABA and sugar signaling

In Chapter 2 we introduce ABSCIC ACID-INSENTITIVE 4 (ABI4), a transcription factor known for its role in seed germination and signaling in response to the plant hormone abscisic acid (ABA). ABI4 was first discovered as an ABA-insensitive mutant in the model plant *Arabidopsis thaliana* (Finkelstein 1994). Later studies also showed involvement in various sugar-insensitive phenotypes, indicating a role for ABI4 in the cross-talk between sugars and ABA signaling. More specifically, ABI4 has been characterized as a mediator of ABA in the repression of seed germination (Finkelstein et al. 2011); this control system is thought to exist to prevent seed activation in conditions that are unsuitable for plant growth. Typically, ABI4 expression reaches its maximum at seed maturity and gradually decreases during germination, unless stress signals interfere with its degradation (Finkelstein et al. 2011). Interestingly, the repressive effect of ABI4 is observed also when seeds are exposed to high sugar concentrations, a response termed sugar-directed growth arrest (Dekkers et al. 2004). It has been proposed that this is due to the stabilizing effect of glucose on the ABI4 protein, which remains expressed at high concentrations (Finkelstein et al. 2011). Consistently, the artificial application of sugars is less effective at later germination stages, when most ABI4 is already degraded (Gibson et al. 2001).

Much less is known about the role of ABI4 in other plant developmental stages, in part due to the low expression levels in organs other than seed. Studies showed that the transcription factor is involved among other things in photosynthesis and lipid biosynthesis, and that it regulates the expression of a sodium transporter in roots, which can impact the response to salt stress (Shkolnik-Inbar et al. 2013). ABI4 appears also implicated in the protection of the plant from oxidative stress, through the regulation of enzymes involved in the degradation of reactive oxygen species (Cheng et al. 2011). Additional functions of ABI4 in seeds and other tissues are described in more detail in Chapter 2.

Open questions on ABI4 evolution and conservation

ABA signaling components have been found in green algae (Bajguz 2009) and lower plants, such as mosses and ferns (Hanada et al. 2011). In fact the ABA pathway is highly conserved in the latter group, where *Physcomitrella patens* and *Selaginella moellendorfii* have been shown to host a wide array of genes central to ABA signaling (Hanada et al. 2011). The role of these genes however is not fully equivalent to that of higher plants, since lower species present important physiological differences, such as the lack of a seed development stage. ABA appears involved in heat stress response in algae (Bajguz 2009), and previous studies showed a similar function in thermotolerance experiments in cereals (Maestri et al. 2002). It has been hypothesized that this specific role of ABA signaling, if conserved in lower plants, might have contributed to the desiccation resistance necessary for the colonization of land

(Hanada et al. 2011).

The survey of ABA signaling genes in lower plants (Hanada et al. 2011) did not include ABI4, however another study showed the presence of putative ABI4 genes in *P. patens* and *S. moellendorffii* (Mizoi et al. 2012). Another experiment showed that the wild-type ABI4 gene from maize complements a mutated version in *A. thaliana* (Niu et al. 2002), indicating that its functions are conserved at least between the two major groups of flowering plants, monocots and dicots. However not much has been done in terms of sequence comparison between species, as the only study of this kind involves just *A. thaliana* and rice (*Oryza sativa*) (Nakano et al. 2006). Given the crucial role of ABI4 in seed development and germination, a survey of ABI4 proteins across higher plant species is timely and desirable to help identifying conserved sequence features involved in its mechanism of action. Also, the comparison of these ABI4 sequences with the ones identified in lower plants could provide insights into the still unexplored role of this transcription factor in ABA signaling during land plants evolution.

The C/S1 bZIP transcription factor network

Role of the C/S1 bZIP transcription factor network in plant stress

In Chapter 3 we present an in-depth phylogenetic analysis of the C and S1 groups of plant basic leucine zipper (bZIP) transcription factors. The proteins encoded by these two groups of genes have a strong tendency to form heterodimers (Ehlert et al. 2006), creating a "C/S1 bZIP transcription factor network" of interacting proteins. Unfortunately, in spite of the evidence for dimerization as a necessary condition for bZIP transcriptional activity (Deppmann et al. 2004; Deppmann et al. 2006), most experimental studies remain focused on the functions of individual C and S1 bZIPs rather than of their dimers. We believe that more efforts should be directed towards the characterization of the C/S1 transcription factor network as a single regulatory entity, since its components are bound to have interrelated, or even redundant, functions (Dietrich et al. 2011; Hartings et al. 2011; Iglesias-Fernandez et al. 2013). This approach is not an easy one though: compared to the single copy gene ABI4, the integrated analysis of C and S1 bZIPs presents a more difficult challenge for functional and phylogenetic studies, as each bZIP group counts multiple members. For instance, the *A. thaliana* genome hosts four C (bZIP9, bZIP10, bZIP25, and bZIP63) and five S1 (bZIP1, bZIP2, bZIP11, bZIP44, and bZIP53) genes, respectively (Jakoby et al. 2002).

The few studies that investigated the functions of C/S1 bZIP dimers indicated a strong synergistic effect of dimerization on the up- and down-regulation of selected target genes, as shown in a series of double overexpression experiments (Weltmeier et al. 2006; Alonso et al. 2009; Kang et al. 2010). Individually, members of both C and S1 groups have been characterized as regulators in seed development (Lara et al. 2003; Alonso et al. 2009; Iglesias-Fernandez et al. 2013; Veerabagu et al. 2014), sugar signaling (Kang et al. 2010;

Ma et al. 2011; Matiolli et al. 2011), and stress responses, in particular in relation to amino acid metabolism (Satoh et al. 2004; Kaminaka et al. 2006; Weltmeier et al. 2006; Baena-González et al. 2007; Hanson et al. 2008; Dietrich et al. 2011; Sun et al. 2012). Target amino acid metabolism genes include the glutamate dehydrogenases GDH1 and GDH2, induced by bZIP11 (Hanson et al. 2008), DIN6/ASN1 (Baena-González et al. 2007), and proline dehydrogenase (ProDH) (Satoh et al. 2004). While ProDH is thought to be involved in stress recovery (Weltmeier et al. 2006), the accumulation of asparagine and other amino acids during dark-induced starvation was proposed to result from protein degradation in order to provide an alternative to carbon as energy source (Caldana et al. 2011; Dietrich et al. 2011). S1 bZIPs in particular were proposed to mediate some of the transcriptional changes induced by the sucrose-non-fermenting-1-related protein kinase-1 (SnRK1), a master regulator in stress response (Baena-González et al. 2007). SnRK1 acts as a metabolic sensor that can decode energy deficiency signals and induce an extensive metabolic reprogramming during stress responses. This is mediated by a number of transcription factors and downstream targets starting an energy-saving program at different levels, including transcriptional, translational, and post-translational (Baena-González et al. 2007). Simultaneous overexpression of SnRK1 and individual S1 bZIPs was shown to have a strong synergistic effect on SnRK1 target genes activation, with the exception of bZIP44, which was not included in the study (Baena-González et al. 2007). C group bZIP9, bZIP25, and bZIP63 were also tested, and bZIP63 in particular showed a mild synergistic activity with SnRK1. The observation was confirmed in a more recent study, which showed the direct phosphorylation of bZIP63 by SnRK1 (Mair et al. 2015).

Regulation of the C/S1 bZIP transcription factor network members by sugars and SIRT

C and S1 bZIPs activity is also influenced by sugars. The expression of C group bZIP63 was shown to be repressed by glucose in an ABA-dependent manner (Matiolli et al. 2011). In the presence of sucrose and glucose, the transcript levels of S1 group bZIP1 and bZIP53 decrease, and energy availability also seems to affect the phenotypes of their mutants (Usadel et al. 2008; Alonso et al. 2009; Kang et al. 2010; Dietrich et al. 2011). Under normal growth conditions, bZIP1 gene expression is limited to sink tissues (Weltmeier et al. 2009), but transcript levels were found to increase in source leaves after sugar starvation (Usadel et al. 2008; Dietrich et al. 2011). bZIP11 transcript levels show an opposite behavior, increasing in the presence of glucose and sucrose, and decreasing when the plant suffers from starvation stress (Rook et al. 1998; Usadel et al. 2008). At the same time, sucrose represses the translation of S1 bZIP transcripts, including bZIP11, by means of an inhibitory mechanism termed Sucrose-Induced Repression of Translation (SIRT) (Rook et al. 1998; Wiese et al. 2004; Rahmani et al. 2009). It has been shown that SIRT requires a specific upstream open reading frame (uORF), which in the presence of sucrose provides an alternative translation start site for the ribosome and prevents protein production from the main bZIP transcript (Rahmani et al. 2009; Schepetilnikov et al. 2013). While the mechanistic details are still to be completely clarified, both SIRT and the uORF sequences on which it relies are considered

unique and characterizing features of S1 bZIPs (von Arnim et al. 2014).

Open questions on the evolution and conservation of the C/S1 bZIP transcription factor network

Given its important role in seed development and stress signaling, the C/S1 bZIP transcription factor network appears as a promising target for crop research. However, most of its regulatory features remain poorly investigated in non-model plants. SIRT for instance has been experimentally confirmed in only one other species apart from *A. thaliana* (Thalor et al. 2012). As previously argued for ABI4, the comparison of C and S1 sequences from multiple species can be used to identify conserved regulatory elements, which could be tested experimentally to gain further insights into the function and regulation of these transcription factors. Unfortunately, collecting sequences from different plants to perform such a comparison is not as straightforward as for ABI4. The identification of S1 bZIPs is particularly challenging, due to the high similarity between these sequences and related members from the S bZIP subfamily (Côrrea et al. 2008); in fact to date it is not even established whether S1 bZIPs are present in all flowering plants (Côrrea et al. 2008). Moreover, while some sequence features could be conserved across all C or S1 bZIPs, others are likely specific to individual genes, which can explain the differences in function and expression patterns between related bZIPs. Comparisons between species should therefore take into account the fine structure of C and S1 bZIP groups, which however is not fully resolved at this time (Côrrea et al. 2008). The problem affects not only sequence comparisons, but also the integration of experimental information between different species, hindering the transfer of functional information from model plants to crops.

The regulatory features of the C/S1 bZIP transcription factor network genes can also be investigated from an evolutionary perspective. It is still unclear at what point during plant evolution C and S1 bZIPs started to form dimers with each other, and what functions these early dimers performed in ancestral plant species. The origin of SIRT is also completely unexplored. Previous phylogenetic analyses could detect C bZIP genes in mosses and ferns, but S1 bZIP and related S sequences were found only in flowering plants (Côrrea et al. 2008), which excludes both the formation of C/S1 bZIP dimers and the presence of SIRT in more basal species. However, the results did not appear conclusive due to the limited number and quality of analyzed genomes (Côrrea et al. 2008).

In summary, establishing when C and S1 bZIP transcription factors and their regulatory features first appeared is a necessary step to understand the emergence and role of the C/S1 bZIP transcription factor network during plant evolution. Moreover, a comprehensive survey of C and S1 bZIP genes in plants is desirable both to identify conserved regulatory elements, and to facilitate the integration of functional information between model and crop species.

Phylogenetics and comparative genomics approaches to the study of plant transcription factors

Questions on the evolution and conservation of transcription factors, such as those encountered in discussing ABI4 and C and S1 bZIPs, pertain to the domain of phylogenetics and comparative genomics. Plant genome projects generated in recent years a wealth of sequence information, leading to the growth of specialized online repositories, such a Phytozome (Goodstein et al. 2012). Together with the increase in computational power and the development of specialized analysis tools, this made it easier to perform automated comparisons of gene sequences across a wide variety of species, and to reconstruct their evolutionary history. Below we introduce some general concepts to understand how these analyses are performed.

Phylogenetics is the study of the evolutionary history of species and their genes, while comparative genomics is the comparison of genome features across multiple species, again including gene sequences. These fields are linked by the concept of gene homology: from a phylogenetic perspective, two genes are defined homologous when they share a common ancestry; homology typically implies a high sequence similarity, allowing homologous genes to be compared across different organisms using comparative genomics methods. Two homologous genes are called orthologous if they originated from a speciation event, or paralogous if they originated from a duplication event, either involving the gene alone (gene duplication) or the entire genome (whole genome duplication). The evolution of plants is characterized by several independent rounds of whole genome duplications (Van de Peer et al. 2009; Jiao et al. 2011; Van de Peer 2011; Jiao et al. 2012), which had an important role is shaping their gene repertoires (Hufton and Panopoulou 2009). While orthologous genes tend to retain the same functions in different species, paralogous genes within the same species are less constrained by selective pressure. This can lead to the acquisition of new functions, or to the partition of ancestral functions among the paralogous copies, through the processes of neo- or sub-functionalization, respectively (Force et al. 1999; He and Zhang 2005). These apparently trivial definitions can therefore provide important insights into the functional diversification of homologous genes.

The first step in performing a phylogenetic or comparative study is to collect the homologous gene sequences to analyze. The starting point is usually a single gene or collection of homologous genes, defined as “query” or “seed”, for which the researcher wants to find other homologous sequences. Different algorithms exist to compare the queries against databases containing millions of DNA and protein sequences in a fast and efficient way. These programs can compute pairwise alignment scores either between database sequences and individual queries, or between database sequences and a single profile built integrating the information from all the query sequences; BLAST (Altschul et al. 1990; Camacho et al. 2009) and HMMER (Eddy 1995; Wheeler and Eddy 2013) are the most popular software to perform either analysis types, respectively. Importantly, both methods score database

sequences based on similarity with the query. However, while homology implies similarity, similarity *per se* does not imply homology; for instance, two genes without shared ancestry might experience convergent evolution due to similar structural constraints on the folding of their respective proteins. After collecting all putative homologous sequences, these are aligned together with the queries to generate a multiple sequence alignment (MSA). Also in this case several algorithms are available (Katoh et al. 2002; Edgar 2004; Sievers et al. 2011), whose goal is to match conserved portions of the sequences across the homologous genes, down to the level of single amino acid or DNA residues. Usually protein sequences are preferred to build MSAs, since the 20-letter amino acid code is more informative than the 4 DNA bases. MSAs can be visually inspected to exclude obvious non-homologous genes that were wrongly included in the initial search results; the remaining sequences can then be used as queries to perform new searches and find matches iteratively, until no new homologous genes can be identified. It is important to note that in certain cases, especially when the analysis is focused on a group of orthologous genes within a family of homologs, the visual inspection of MSAs might not be sufficient to exclude unwanted sequences from the alignment. In this case, phylogenetic analyses are needed to tell apart orthologs from more distantly related homologs.

MSAs are a source of valuable information for comparative studies. They offer a direct way to visualize highly conserved residues and motifs between homologs, which are likely crucial for gene function or regulation. Conversely, if the function of a certain residue or motif has already been tested in a given species, MSAs are useful to assess the conservation of that same residue or motif in other homologs, allowing the integration of functional information between species. The highly conserved regions of MSAs are also used in the phylogenetic analysis of gene families. Mutations in these regions are not likely, since they will likely have an impact on gene function or regulation; however they still happen at low frequency, which is convenient to reconstruct sequence evolution over long periods of time. Different algorithms are capable of reconstructing the phylogenetic tree of a gene family based on MSAs, by applying suitable models of sequence evolution (Guindon and Gascuel 2003; Ronquist and Huelsenbeck 2003; Stamatakis et al. 2005). Typically these programs perform multiple independent tree reconstructions to identify the most likely tree configuration, or topology. The final tree will report bootstrap support values for each branch, indicating how frequently that branch is observed in that configuration in the independent reconstructions. Molecular phylogenetic trees are usually compared against the known species tree, a process termed tree reconciliation, to reconstruct the most likely order of speciation and duplication events that shaped a gene family. Reconciled trees provide a convenient representation of complex orthologous and paralogous relationships that cannot be easily inferred from the original MSA.

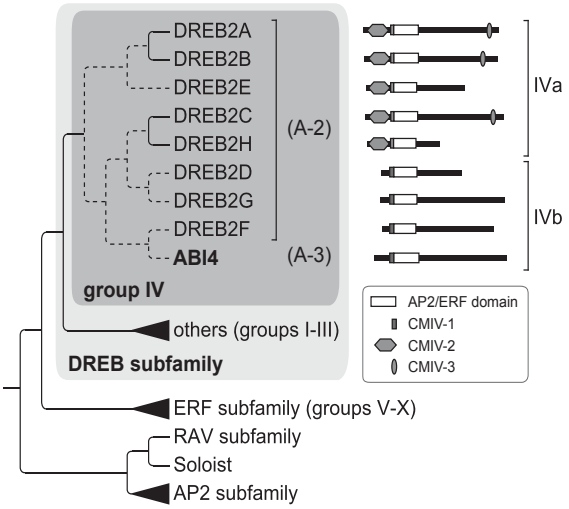


Figure 1.1. Phylogenetic relationship of ABI4 to other AP2/ERF superfamily genes. *The phylogenetic tree shows the classification of ABI4 within the AP2/ERF superfamily of transcription factors, and its relationship to other group IV DREB subfamily genes, according to Nakano et al. (2006); dotted lines indicate branches with a low bootstrap support value (< 90%). Group IV genes are further classified into group IVa and IVb based on the gene structure, also shown (Nakano et al. 2006). The previous classification from Sakuma et al. (2002) is reported for comparison.*

Resolution of the ABI4 phylogeny

ABI4, discussed previously, is present as a single copy gene in the *A. thaliana* genome, which makes it a simple case for phylogenetic studies. The transcription factor belongs to the APETALA 2/ethylene-responsive element binding factor (AP2/ERF) superfamily, whose members all share the characteristic AP2/ERF DNA-binding domain (**Figure 1.1**). In Chapter 2 we performed a phylogenetic reconstruction to study ABI4 evolution and identify conserved sequence features of its orthologs in different species. The analyses followed the steps described above, including the removal of non-orthologous sequences from the initial search results; this required prior knowledge of the relationship between ABI4 and other AP2/ERF superfamily genes, which we explain in more detail here.

In *A. thaliana*, AP2/ERF members are encoded by 145 loci divided in the families ERF, AP2, RAV, and Soloist (Nakano et al. 2006; Licausi et al. 2013). Most proteins containing a single AP2/ERF DNA-binding domain are assigned to the ERF family, which can be further subdivided into actual ERFs, Dehydration-Responsive Element-Binding proteins (DREBs), and a few related subfamilies, including ABI4. Currently two independent nomenclatures are being used in the classification of ERF family members: one developed by Sakuma and colleagues

in 2002, which assigns ABI4 to its own group (A-3) based on phylogenetic reconstructions alone; and the other from Nakano et al. (2006), incorporating gene structure information. According to the latter, ABI4 falls into group IVb ERFs together with other DREB2 genes, and it might be more closely related to them than it appears from the previous classification system (**Figure 1.1**). In Chapter 2 we deliberately collected sequences encompassing the entire group IV (both IVa and IVb) ERFs, and used them to build a phylogenetic tree that allowed us to distinguish true ABI4 orthologs from closely related DREB2 genes (Chapter 2, see Figure S3 in the Supplementary Material). This was particularly important to assess the identity of the less conserved ABI4 sequences found in the lower species *P. patens* and *S. moellendorffii*. Confirmed ABI4 orthologs were then re-aligned to perform more focused comparative and phylogenetic analyses.

In brief, our phylogenetic reconstruction showed that most plant species possess a single ABI4 gene, which suggests a highly conserved function. We also identified a new conserved "ABI4" motif in orthologs from higher plants, which is likely important for ABI4 activity. Confirmed ABI4 orthologs from the lower plants *P. patens* and *S. moellendorffii* did not have this motif, and also possessed an additional copy of the gene, likely indicating that its role might be different than in higher plants. The results and possible follow-ups are discussed in more detail in Chapters 2 and 5 (summarizing discussion).

Resolution of the C and S1 bZIPs phylogeny

Compared to ABI4, the study of C and S1 bZIPs evolution and conservation is more challenging, due to the higher number of orthologous genes within each group. S1 bZIPs in particular have a complex history of gene duplications, which hinders the assessment of their phylogenetic relationships between and within species, as described below. In Chapter 3 we performed a high-resolution phylogenetic analysis of C and S1 bZIP orthologs, which allowed us to resolve previous classification hurdles, and to gain significant insights into the evolution and conservation of these transcription factors. Here we introduce the more specifically the challenges we needed to overcome.

Members of the bZIP family of transcription factors are recognized as such based on the presence of a bZIP domain, a conserved stretch of 60–80 amino acids composed of a N-terminal basic region, necessary for DNA binding, and a leucine zipper, which mediates homo- and heterodimerization with other bZIPs (Deppmann et al. 2004; Deppmann et al. 2006). bZIPs are thought to have evolved from a single gene in the common ancestor of eukaryotes, which expanded independently in different lineages (Jindrich and Degnan 2016). In flowering plants, bZIP genes are classified into 13 groups of orthologs, or subfamilies, covering a wide range of regulatory functions (Jakoby et al. 2002; Côrrea et al. 2008). While C bZIPs are considered a subfamily of their own, S1 genes are actually a subset of closely related orthologous sequences within the larger S bZIP subfamily. From a functional perspective, S1 bZIPs can be distinguished from other S sequences based on

their unique functional properties, such as SIRT and the heterodimerization with C bZIPs; other S bZIPs tested so far do not seem capable of heterodimerization with C bZIPs (Ehlert et al. 2006). However, at the phylogenetic level the distinction between S1 and other S bZIPs does not appear as clear (**Figure 1.2 A**). One reason for this is that genes in the S bZIP subfamily are characterized by a shorter bZIP domain (Jakoby et al. 2002; Côrrea et al. 2008), which reduces the information content of sequence alignments, causing a drop in the resolution of reconstructed phylogenetic trees. This characteristic, together with the high rate of duplications and sequence divergence within the S subfamily, makes the reconstruction of its evolutionary history extremely challenging. The most comprehensive phylogenetic reconstruction of the plant bZIP transcription factor family to date, by Corrêa and colleagues (2008), could not provide definitive results; in particular, the study left unanswered whether S1 bZIPs are present in both eudicots and monocots, the two main subdivisions of flowering plants (**Figure 1.2 A**). The authors also encountered difficulties in establishing the relationship of the S subfamily to other bZIP groups, and proposed that these genes might have originated from the related C bZIP subfamily or proto-C genes (Côrrea et al. 2008) (**Figure 1.2 B**).

An important limitation in the Corrêa study (2008) was the small set of plant sequences available at the time of the analyses. The situation changed dramatically since then, as several plant genomes have been sequenced and published on online databases in the past years; these include not only commercially relevant species, but also poorly characterized lineages which are however valuable to evolutionary studies (Der et al. 2011; Banks et al. 2011; Birol et al. 2013; Bushart et al. 2013; Chamala et al. 2013; Nystedt et al. 2013; Zimmer et al. 2013; Brouwer et al. 2014; Hori et al. 2014; Neale et al. 2014; Wolf et al. 2015). In Chapter 3, we include the wealth of new data to investigate the presence of C and S1 bZIP orthologs in species where this was simply not possible before. For instance, the new genome from the charophyte *Klebsormidium flaccidum* (Hori et al. 2014) allowed us to demonstrate the existence of both C and S bZIP orthologs already in non-terrestrial species (**Figure 1.2 B**). Another important achievement was the discovery of S1 bZIP orthologs in gymnosperm species, for which we performed experiments and confirmed regulation by SIRT. Therefore, the increased availability of species allowed us to completely reexamine the evolution of the C/S1 bZIP transcription factor network and its functional properties.

In spite of the significant species sampling advantage, the high sequence divergence of C and S1 bZIPs still affected our phylogenetic reconstructions, especially in flowering plants. To overcome this problem, we used an unconventional topology-based approach: next to a standard phylogenetic reconstruction including all flowering plant species, we created multiple independent phylogenetic trees of C and S1 bZIP orthologs for major lineages, such as eudicots, monocots, and asterids, and also for much smaller groups of closely related species, such as brassicales; we then compared the topology from different trees to infer the relationship between orthologs within both C and S1 bZIP groups, and repeated the phylogenetic reconstructions for each of the previous plant lineages using a more restricted

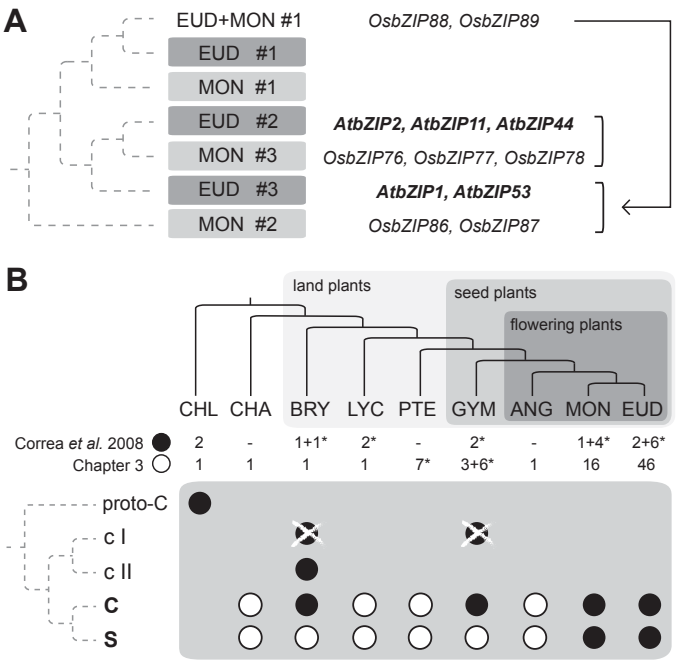


Figure 1.2. Previous (Corrêa et al. 2008) and proposed (Chapter 3) classification of C and S bZIPs. A. *Classification of S bZIPs from flowering plants into possible groups of orthologs (PoGOs) according to Corrêa et al. (2008); the tree is traced using dotted lines to indicate the low bootstrap support for this topology. A single PoGO including both eudicots and monocots (EUD+MON), 3 including only eudicots (EUD), and 3 including only monocots (MON) are shown. Known S1 bZIPs from A. thaliana (in bold) and rice S1 bZIP orthologs identified in Chapter 3 are indicated next to their corresponding PoGO. The relationship between S1 bZIP sequences according to our phylogenetic reconstructions (Chapter 3) is shown for comparison to the PoGOs.* ***B.*** *Conservation pattern of C and S bZIP genes in green plants. The gene tree shows the relationship between C and S bZIPs (in bold), ancestral algal proto-C sequences, and the C-related groups c I and c II proposed by Corrêa et al. (2008); the tree is traced used dotted lines to indicate that the represented topology is fully hypothetical. Circles indicate the presence of a bZIP group in the corresponding plant lineage, shown at the top; black: presence according to Corrêa et al. (2008); white: additional presence according to our results (Chapter 3); white crosses indicate previously identified groups that were not recovered in our analyses. The number of plant genomes used in the phylogenetic reconstructions is indicated below the species tree for both Corrêa et al. (2008) and our analyses (Chapter 3); numbers marked with an asterisk indicate species without a complete genome (CHL = chlorophyta, CHA = charophyta, BRY = bryophyta, LYC = lycopodiophyta, PTE = pteridophyta, GYM = gymnosperms, ANG = basal angiosperms, MON = monocots, EUD = eudicots).*

set of orthologs, which gave us even more resolution. While individual phylogenetic trees still failed to yield high bootstrap support values, we observed surprisingly consistent results

between independent phylogenetic reconstructions, which allowed us to extract a sound consensus topology for both bZIP groups. Overall, results allowed us to provide detailed phylogenetic trees that can be used to correctly compare and integrate experimental information from different flowering plant species.

We discuss the significance and possible applications of our phylogenetic reconstruction in the study of the C/S1 bZIP transcription factor network, as well as of its individual members, in Chapters 3 and 5 (summarizing discussion).

TRANSLATIONAL CONTROL OF ENERGY HOMEOSTASIS

So far we treated the molecular regulation of energy homeostasis from the perspective of transcriptional control, introducing specific transcription factor families. Another level at which energy homeostasis is regulated is translational control, discussed in Chapter 4. The translation of messenger RNAs (mRNAs) into proteins is one of the most energy-demanding cellular processes in the plant cell (Pal et al. 2013), and in situations of limited energy availability resources need to be directed towards the production of proteins that are immediately necessary for plant survival. Accordingly, stress conditions affect translation by causing a decrease in protein synthesis rates for most mRNA species (Kawaguchi et al. 2003; Kawaguchi et al. 2004; Branco-Price et al. 2005; Nicolaï et al. 2006; Branco-Price et al. 2008; Mustroph et al. 2009; Matsuura et al. 2010; Ueda et al. 2012; Yanguez et al. 2013; Juntawong et al. 2014). While these changes can be generally correlated to mRNA levels, the translation of some mRNA species appears to decrease or even increase independently from the amount of transcript, indicating that regulatory mechanisms are operating directly at the translational stage (Bailey-Serres et al. 2009; Roy and von Arnim 2013; Juntawong et al. 2014). This uncoupled regulation of transcription and translation is an efficient mechanism for the plant to rapidly adjust its protein production in specific conditions; regulating translational efficiency while leaving mRNA levels unchanged is also convenient to quickly resume normal cell functions once the stress source is removed (Piques et al. 2009; Juntawong and Bailey-Serres 2012). More recently, studies have shown that subsets of transcripts are translationally regulated also in a variety of plant cell types and physiological processes (Jiao and Meyerowitz 2010; Liu et al. 2012; Ribeiro et al. 2012; Liu, Wu, et al. 2013; Galland et al. 2014; Layat et al. 2014; Lin et al. 2014; Rajasundaram et al. 2014; Basbouss-Serhal et al. 2015; Yamasaki et al. 2015), indicating that translation represents a commonly regulated step in protein synthesis.

Translation of an mRNA into protein can be divided into four stages (Kapp and Lorsch 2004), simplified as follows: initiation, consisting in the scanning of the 5'UTR and assembly of the ribosome at the translation start site of the mRNA; elongation, in which the ribosome progresses along the mRNA to synthesize the corresponding protein sequence; termination, when the ribosome encounters a stop codon on the mRNA causing the release of the

protein; and recycling, consisting in the dissociation of the ribosome from the mRNA, so that another round of initiation can start. Actively translated mRNAs are normally associated with multiple ribosomes at the same time, forming so-called polysomes. The translational efficiency of an mRNA species can be measured in terms of its polysomal occupancy, also defined as ribosome loading, and ribosome density; the first refers to the fraction of mRNA associated to polysomes, the second to the average number of ribosomes per length unit of mRNA. Initiation is considered the most important stage for translational regulation (Kawaguchi and Bailey-Serres 2002; Esposito et al. 2010), having the strongest effect on polysomal occupancy and ribosome density; elongation and termination rates can also affect these indicators, but their impact in studied performed so far appears secondary (Ma and Blenis 2009; Jackson et al. 2010).

Several mechanisms underlying condition-dependent translational regulation have been proposed to operate directly at the initiation step. For instance, translation can be regulated at the transcript level through the activation of alternative translation start sites, such as internal ribosome entry sites (IRESs) and uORFs (Xue and Barna 2012). IRESs are structured elements favoring the recruitment of the ribosome directly at a given position along the transcript, reducing the requirement for canonical initiation factors. Depending on their localization, IRESs can favor translational initiation either at the main translation start site, maintaining high translation rates during stress for instance, or at alternative start sites, causing variations in translational efficiency but also the production of different protein isoforms from the same transcript. Differently from IRESs, uORFs are sequence- rather than structure-based features, which can recruit the ribosome at an alternative upstream initiation site to prevent the translation of the main ORF. An example of this is provided by the previously discussed SIRT in S1 bZIPs, in which a conserved uORF drives translational repression of the main ORF in the presence of sucrose (Rahmani et al. 2009). However, the presence of uORFs with a non-canonical start codon or an unfavorable sequence context can actually enhance the translation of the main ORF (Roy and von Arnim 2013).

Increasing evidence suggests that translational regulation can also take place directly at the ribosome level, in contrast with the classic view of the ribosome as a static “molecular machine” (Alberts 1998). Recently, it has been shown that ribosomes can change their composition in different experimental conditions (Hummel et al. 2012; Rodriguez-Celma et al. 2013; Wang et al. 2013), leading to ribosome heterogeneity (Xue and Barna 2012). Condition-dependent changes in the translational machinery offer a plausible explanation for the differential interaction of ribosomes with IRESs, uORFs, and other mRNA sequence signals, justifying the observed changes in translational efficiency in response to environmental signals. Ribosome heterogeneity can arise from the differential expression of ribosomal proteins (RPs), encoded by more than 240 genes in *A. thaliana* (Barakat et al. 2001; Carroll et al. 2008; Hummel et al. 2015); their post-translational modifications; and diversity in ribosomal RNAs (rRNAs) and ribosome-associated factors (Turkina et al. 2011; Xue and Barna 2012; Boex-Fontvieille et al. 2013). The variety of RPs paralogs seems

a driving element of ribosome heterogeneity in plants (Velez-Bermudez and Schmidt 2014): while animal RPs are encoded mostly by a single gene, each plant RP family consists of 2–7 paralogs with seemingly distinct functions (Degenhardt and Bonham-Smith 2008; Schippers 2010) and expression patterns (Williams and Sussex 1995; Weijers et al. 2001; Hulm 2005; McIntosh and Bonham-Smith 2005; Degenhardt and Bonham-Smith 2008; Mustroph et al. 2009; Wang et al. 2013), offering a wider scope for regulation.

RNA-binding proteins (RBPs) are another class of molecules with emerging potential in translational regulation, in virtue of their differential expression in various growth conditions (Lorkovic 2009; Gerstberger et al. 2014). These proteins seem involved in nearly all aspects of post-transcriptional regulation, including pre-mRNA splicing, polyadenylation, and mRNA stability and export (Lorkovic 2009; Gerstberger et al. 2014). For instance, UBP1 in *A. thaliana* has been shown to bind a specific subpopulation of mRNA molecules during hypoxia, causing their sequestration in separate cellular compartments termed stress granules (Sorenson and Bailey-Serres 2014); other RBPs are involved in a similar mechanism, where target mRNAs are sequestered in processing bodies (P-bodies) (Maldonado-Bonilla 2014). This type of regulation results in the removal of selected mRNAs from the pool of actively translated transcripts, causing translational inhibition. More than 200 RBPs have been identified in *A. thaliana*, most of which are not present in other eukaryotic organisms, and are therefore likely to be involved in plant-specific regulatory mechanisms (Lorkovic 2009).

Open questions on sequence signals responsible for differential translational regulation

Translational control mechanisms based on the interaction of mRNAs with RPs and RBPs likely rely on specific sequence signals in the target transcripts, which remain largely elusive in plants. So far, the majority of plant translatome studies in both stress and developmental conditions focused mainly on the functional annotation of translationally regulated transcripts, rather than on the sequence features responsible for their regulation. However, an increasing number of studies is going further by investigating mRNA motifs and secondary structure (Kawaguchi and Bailey-Serres 2005; Jiao and Meyerowitz 2010; Matsuura et al. 2010; Liu et al. 2012; Liu, Wu, et al. 2013; Basbouss-Serhal et al. 2015), which are more likely to provide insights into the detailed mechanisms of plant translational control. Nonetheless, specific sequence motifs and structural elements related to translational regulation remain difficult to identify. The *in silico* prediction of structural features, such as IRESs, from the mRNA sequence is not entirely reliable, as mRNA folding is likely dependent on specific cellular conditions. uORFs can also be elusive, namely because plants can use non-canonical translation start sites which are not fully documented; however these and other regulatory sequences, such as mRNA localization signals, could be identified with dedicated motif searches if sufficiently conserved in a set of translationally regulated mRNAs. While the lack of specific experimental data is certainly a limitation, more advanced bioinformatic analyses than those performed in previous studies could still help to shed light on the sequence signatures involved in plant translational regulation.

Bioinformatic analysis of translationally regulated transcripts

In Chapter 4 we performed both basic and advanced bioinformatic analyses of translationally regulated transcripts during sucrose treatment and seed development. Here we introduce the concepts and methods behind them, with particular attention to motif and structure analyses.

Early studies on translational regulation in bacteria, yeast and other eukaryotic species showed the correlation of several sequence features with the basal translational efficiency of an mRNA species. For instance, it is established that long transcripts tend to have lower translation rates, and the same is generally thought for transcripts with a high content of guanine and cytosine (GC) nucleotides, associated with more structured regions. Codon usage bias, which indicates the preference for certain synonymous codons over others in a nucleotide sequence, has also been extensively studied. While encoding identical amino acid, synonymous codons recruit distinct tRNAs with distinct abundances in the cell, ultimately affecting translation rates (Ikemura 1985; Kanaya et al. 2001). Together with uORF enrichment and translation start site context, these features are regularly assessed in plant translational regulation studies (Branco-Price et al. 2005; Kawaguchi and Bailey-Serres 2005; Juntawong and Bailey-Serres 2012; Liu et al. 2012; Yanguez et al. 2013; Basbouss-Serhal et al. 2015). However, to investigate condition-dependent translational regulation operated by RPs and RBPs, these general indicators may not be sufficient, namely because transcript length and GC content are unlikely to provide specific signals for protein interactions.

More informative signals can be discovered through motif analyses techniques, which apply statistical methods to detect overrepresented nucleotide or amino acid patterns within a group of sequences. A good example of motif discovery software is MEME (Bailey et al. 2009), probably the most widely used tool for this kind of analyses. The only input requirement is a list of genes or proteins, as motifs are predicted *de novo* from the set of sequences. Unfortunately, for structural analyses the available *de novo* prediction tools are not as informative. While the linear sequence of a motif remains the same in different experimental conditions, this is unlikely the case for structured elements, as mRNA folding can be extremely sensitive to cellular conditions. This means that the most reliable way to investigate the role of mRNA structure in translationally regulated transcripts would be to obtain *in vivo* mRNA structure data in the same experimental conditions, in order to constrain structural predictions.

Analysis of translationally regulated mRNAs during sucrose treatment and seed development

In Chapter 4 we performed a bioinformatic analysis of translationally regulated transcripts in two very different experimental settings: one dealt with the sucrose treatment of young seedlings, representing an artificial state of high energy availability (Gamm dataset); the other with the physiological process of seed development (Bai dataset). The datasets differ

also in that the Bai one is structured as a time series rather than a single measurement. While most literature focuses on translational regulation during stress, the experimental data provided by our collaborators allowed us to perform bioinformatics analyses in these less characterized conditions. It is worth mentioning that the two datasets were analyzed at different times, and not all tests performed in the more recent Bai study were covered as thoroughly in the Gamm dataset.

Given the differences in the two datasets, the number of translationally regulated transcripts identified in each of them was also quite different: in the Gamm dataset the vast majority of regulated transcripts showed an increased polysomal occupancy, while in the Bai datasets transcripts with decreased polysomal occupancy outnumbered the positively regulated ones; the number of regulated genes was also much greater in the latter study, indicating that negative translational regulation is a major force in the control of seed development. We performed the analysis of basic sequence features, such as length and GC content, as well as codon bias, enrichment for uORFs, and start site sequence context.

As mentioned before, we additionally performed advanced motif and structural analyses to individuate specific sequence signals responsible for translational regulation in these sets of transcripts. The most important difference between our motif analyses and previous work of this kind is that we did not stop at the MEME output, but double-checked enriched motifs against the entire gene background using another tool, called FIMO (Bailey et al. 2009). The rationale for this is that MEME captures enriched motifs in a set of sequences, but if these motifs are already overrepresented among *A. thaliana* transcripts, they might be wrongly confused with translational regulatory features while in fact they are not. It is surprising that so many studies ignore this second control step, which is important to filter out false signals. In addition to this we also determined the position of enriched motifs along the transcripts, to identify trends in their localization. This is also an uncommon analysis strategy, which however can provide important insights on the dependence of regulatory mechanism on the localization of sequence features.

For the analysis of structural features, we used experimentally determined per-nucleotide structure scores (Li et al. 2012) to calculate the average score for each translationally regulated transcript. It is to note that the use of this *in vitro* dataset represents an important improvement on previous translational regulation studies, which relied on less reliable *de novo* structural predictions instead (Kawaguchi and Bailey-Serres 2005). As for motif analyses, we also checked the localization of mRNA structure along the transcripts, and indeed verified that some transcripts are prone to fold in specific regions of their sequence.

Overall the analyses we performed on the Gamm and Bai datasets revealed a number of motifs and structural features related to translational regulation; however, these could not explain changes in translation for the entire set of regulated mRNAs. Basic transcript characteristics, such as length and GC content, also seemed to play a role. Our results

and considerations for future analyses are discussed in more detail in Chapters 4 and 5 (summarizing discussion).

CHAPTERS OUTLINE

In the next two chapters we discuss specific examples of transcriptional regulators involved in the control of energy homeostasis: the transcription factor ABI4 in **Chapter 2**, well known for its role in seed development and germination, and the stress-responsive C/S1 bZIP transcription factor network in **Chapter 3**. Both transcription factor families are analyzed from an evolutionary and comparative genomics perspective, although in the ABI4 chapter they are provided as contextual to functional information. In the C/S1 bZIP chapter, we do not limit ourselves to investigate the evolution of each group of orthologs, but also provide experimental evidence for the conservation of SIRT in non-flowering plant species, and speculate on the emergence of C/S1 bZIP dimers. In **Chapter 4** we move to translational control, and present an integrated bioinformatic analysis aimed at individuating sequence features responsible for the differential translational regulation of specific subsets of transcripts. The chapter combines data from two different studies, one on sucrose treatment, and the other on seed development. Finally, in **Chapter 5** we summarize our findings and discuss in detail the implications for both future experimental and bioinformatic studies.

2

ABI4: versatile activator and repressor

Julia J. Wind[1], **Alessia Peviani**[2], Berend Snel[2,4], Johannes Hanson[1,3,4], and Sjef C. Smeekens[1,4]

[1]*Molecular Plant Physiology, Institute of Environmental Biology, Utrecht University, Utrecht, The Netherlands*
[2]*Theoretical Biology and Bioinformatics, Utrecht University, Utrecht, The Netherlands*
[3]*UPSC Umea˚ Plant Science Center, Umeå University, Umeå, Sweden*
[4]*Centre for Biosystems Genomics, Wageningen, The Netherlands*

Trends in Plant Science 18, 125–132 (2013)

ABSTRACT

The ABSCISIC ACID INSENSITIVE4 (ABI4) gene was discovered to be an abscisic acid (ABA) signaling responsive transcription factor active during seed germination. The evolutionary history of the ABI4 gene supports its role as an ABA signaling intermediate in land plants. Investigating the ABI4 protein-cis element interaction supports the proposal that ABI4 binding to its known CE1 cis-element competes with transcription factor binding to the overlapping G-Box element. Recent publications report on ABI4 as a regulatory factor in diverse processes. In developing seedlings, ABI4 mediates sugar signaling, lipid breakdown, and plastid-to-nucleus signaling. Moreover, ABI4 is a regulator of rosette growth, redox signaling, cell wall metabolism and the effect of nitrate on lateral root development.

RESULTS AND DISCUSSION

The transcription factor ABI4 is evolutionarily conserved

The ABI4 gene was first discovered as the abi4-1 mutation in a genetic screen for ABA insensitive mutants (Finkelstein 1994). This discovery pinpointed important features of ABI4 as a non-redundant gene that is essential for the response of seedlings to ABA. Green algae contain and respond to ABA (Bajguz 2009) and in land plants most ABA signaling components are conserved (Hanada et al. 2011). ABI4 function appears to be conserved as well, as ABI4 from the monocot maize (*Zea mays*) can complement the dicot *Arabidopsis thaliana* abi4-1 mutation (Niu et al. 2002). Studying the phylogenetic history of ABI4 narrows down its origin during the evolution of land plants. The analysis of ABI4 homologs from published proteomes and genomes shows that most plants possess a single ABI4 (**Figure 2.1 A**); the few exceptions seem mainly a consequence of recent whole genome duplications (e.g., *Populus trichocarpa* and *Glycine max*). ABI4 is a member of the AP2/ERF family and members of this family bind specifically to the ABRE elements in the promoters of abiotic stress-responsive genes and regulate their expression (Mizoi et al. 2012). The sequence alignment between species shows the conservation of the characteristic AP2/ERF domain in the N-terminal region, in which the glutamic acid residue at the 69th position was shown to be essential for the function of ABI4 (Laby et al. 2000). Preceding the AP2/ERF domain is the flanking CMIV-1 motif that is shared between all group IV members of the dehydration responsive element binding (DREB) subfamily of AP2/ERF transcription factors (Nakano et al. 2006) (**Figure 2.1**). ABI4 is targeted for proteasomal degradation and most likely, the region for protein degradation resides in the N-terminal half as well (Finkelstein et al. 2011). Interestingly, the alignment allows the identification of another highly conserved motif only present within the ABI4 group. This 'ABI4 motif' consists of an eight- residue core LRPLLPRP, typically preceded by a low complexity region enriched in serine (**Figure 2.1 B**; Suppl. Figure S1). The remarkable ABI4 motif conservation within angiosperms suggests an important selective pressure on this short amino acidic stretch, but its contribution to the function of ABI4 needs to be established.

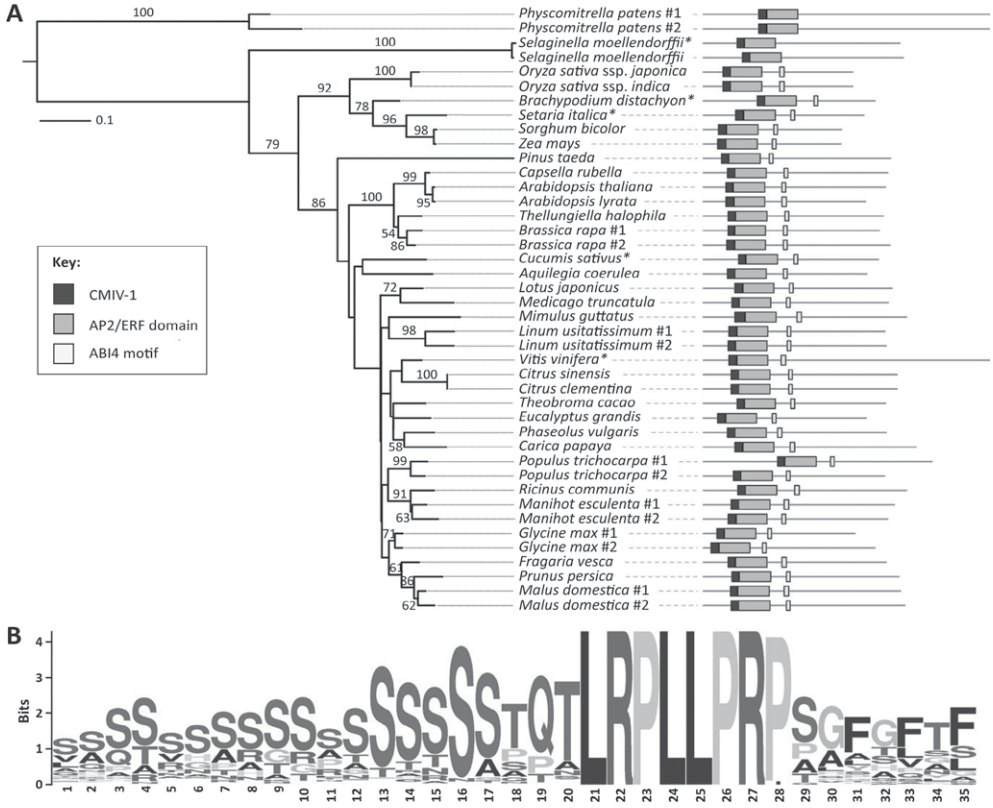


Figure 2.1. The ABI4 sequence is strongly conserved among land plants. *(**A**) ABI4 homologs were predicted from genomic sequences available at Phytozome (*www.phytozome.net*, release 8.0) or Plaza (*http://bioinformatics.psb.ugent.be/plaza*, release 2.5) databases for a total of 33 species. Pinus taeda ABI4-like protein sequence was additionally retrieved from GenBank (*www.ncbi.nlm.nih.gov/genbank*). Tree leaves are annotated with the species name and a numeric identifier in case of duplication events (see Suppl. Figure S4 for the original identifiers). Sequences indicated with * represent ABI4 members predicted from the genomic sequence. ABI4 is typically unique within species, with the exception of L. usitatissimum, P. trichocarpa, M. truncatula, G. max, M. domestica and B. rapa; two ABI4-related sequences were also found in S. moellendorffii and P. patens. A schematic representation of the ABI4 protein domains is reported on the right of the leaves. Boxes show the position of the AP2 domain and conserved motifs, as indicated. Alignments of the conserved regions are shown in Suppl. Figure S1. Sequence alignment of the full-length ABI4 proteins was performed using the Tcoffee webtool (Notredame et al. 2000). The tree reconstruction was done using MEGA5 software (Tamura et al. 2011) with the Neighbor-Joining method (1000 bootstrap replicates); support values above 50 are reported on the branches. (**B**) The figure shows the core of the motif (LRPLLPRP) found in the ABI4 sequences from Spermatophytes. The consensus for the flanking regions is shown as well, as analyzed with MEME (Bailey et al. 2009); the gapped alignment is shown in Suppl. Figure S1.*

The first seven residues of the ABI4 motif are conserved in the gymnosperm *Pinus taeda*. However, the motif is absent in the protein sequences of *Selaginella moellendorffii* (a lycophyte) and *Physcomitrella patens* (a moss). Although the ABI4 motif is absent in these ancient species, the presence of putative ABI4 proteins in *S. moellendorffii*, *P. patens*, and *Chlamydomonas reinhardtii* (a green alga) was suggested in a phylogenetic analysis based on whole protein sequences (Mizoi et al. 2012). Similarly, a phylogenetic reconstruction based on the AP2 domain in a wider selection of species suggested close ABI4 orthologs in *S. moellendorffii* and *P. patens* (Suppl. Figures S3 and S4). However, the clustering of *P. patens* and *S. moellendorffii* sequences with the ABI4 group should be interpreted cautiously since the ABI4-like proteins present in these ancient species might be ancestral to the entire group IV of DREB transcription factors and their sorting could represent an artifact of the phylogenetic reconstruction. The low bootstrap values in support of the tree topology leave space for this alternative interpretation on the proposed absence of ABI4 orthologs in lower plants (Suppl. Figures S3 and S4). Additional sequence information from primitive land plants is needed to resolve this issue of the order of duplications within the DREB subfamily. Thus, the question whether ABI4 originated before the land plants, or afterwards in the ancestor of the seed plants, remains unresolved at this time.

Role in ABA signaling

ABA signaling requires ABA sensing by the PYR/PYL family of START proteins leading to inhibition of type 2C protein phosphatases (PP2Cs) (Park et al. 2009). These inhibited PP2Cs no longer repress SnRK2 protein kinases leading to activation of ABA-responsive binding factors (ABFs) (Cutler et al. 2010). ABI4 initially was discovered as one of five ABA insensitive mutants (Finkelstein 1994; Finkelstein et al. 1998) and later as the salt- (NaCl and Na_2SO_4), mannitol-, and ABA-tolerant mutant sañ5 (Quesada et al. 2000). The five abi mutants identified all contain mutations in genes that encode components of ABA signaling: abi1-abi3 (Koornneef 1984), and abi4 and abi5 (Finkelstein 1994). ABI1 (Leung et al. 1994) and ABI2 (Rodriguez et al. 1998) encode PP2Cs and the ABI3 (Giraudat et al. 1992), ABI4 (Finkelstein et al. 1998), and ABI5 (Finkelstein and Lynch 2000) encode transcription factors. The SnRK2 kinases SRK2D/ SnRK2.1, SRK2E/SnRK2.2, and SRK2I/SnRK2.3 act redundantly and the triple SnRK2 mutant srk2d srk2e srk2i is ABA insensitive. Together, the SnRK2 kinases phosphorylate ABI5 and affect a set of transcripts that overlap with ABI5 and ABI3 affected sets of transcripts, suggesting that these SnRK2s affect ABA-responsive transcripts through regulation of ABI5 and ABI3 (Nakashima et al. 2009). During seed germination ABI3, 4, and 5 either repress or induce promoters that contain ABRE motifs (Nakashima et al. 2006; Reeves et al. 2011; Monke et al. 2012), suggesting that these transcription factors act as ABFs. ABI3 is a B3 domain containing transcription factor which is involved in ABA signaling during seed maturation (Ooms et al. 1993), seed germination (Koornneef et al. 1989), and in the control of chlorophyll accumulation (Clerkx et al. 2003), and its target promoters are enriched for G-box-like and RY-like elements (Monke et al. 2012). ABI5 is a bZIP transcription factor expressed during seed germination and is a confirmed target of ABI4 (Bossi et al. 2009).

ABI4 and ABI5 synergistically controls germination and sugar sensitivity, and a subset of ABI4 target genes are ABI5 targets as well (Reeves et al. 2011). The ABI3, 4, and 5 genes show extensive cross-regulation (Soderman et al. 2000; Arroyo et al. 2003) and are co-expressed while regulating ABA controlled processes (**Figure 2.3**). In yeast two-hybrid assays, only ABI3 and ABI5 were shown to interact with each other, but not with ABI4 (Nakamura et al. 2001). Phenotypic analysis showed that abi3, 4, and 5 mutant embryos all have green cotyledons in the seed coat while germinating (Finkelstein 1994). ABI4 was also proposed to play another role in ABA signaling through the ABA-induced production of phosphatidic acid. In conclusion, ABI4 plays an essential role as one of the transcription factors mediating ABA-dependent transcriptional changes. Most ABI4 requiring developmental processes are ABA dependent and in these processes abi4 mutants are discernible from wild type following ABA addition. However, some phenotypes of abi4 mutants are independent of ABA addition.

Table 2.1. Published transcriptome analyses of ABI4 overexpressing and abi4 mutant lines compared to wild type.

Genotype	Conditions	Upregulated [b]	Downregulated [b]	Processes affected	Refs
abi4-11	Dry seed	49*	11*	Enzyme inhibitor activity, hydrolase activity, cell wall, endomembrane system	[41]
abi4-11	Imbibed seeds	405*	291*	lipid localization, postembryonic development, photosynthesis, seed development, pigment metabolic process, oxidation reduction, catalytic activity	[41]
abi4-102	5 day old seedlings	43	71	Chloroplast, anthocyanin-containing compound biosynthetic process, catalytic activity	[37]
35S:ABI4 line 114A	11 day old seedlings treated with cycloheximide and ABA, or cycloheximide alone	95	n.d.	postembryonic development, anatomical structure development, metabolic process, response to salt stress	[19]
abi4-102	Rosette, 42 days old	277	269	lipid localization, postembryonic development, photosynthesis, seed development, response to abscisic acid stimulus, chlorophyll metabolic process, wax biosynthetic process, response to water	[61]

[a]Genes were considered differently expressed when the change exceeded twofold and was associated with an adjusted P-value of less than 0.05.

[b]Numbers with an asterisk were re-analyzed to fit these requirements (see also Suppl. File S1).

The AP2 domain of ABI4 interacts with a specific DNA sequence

ABI4 is a unique protein but its AP2 domain is a well-conserved domain present in all 145 proteins in the DREB/ERF-related family (Sakuma et al. 2002). The AP2 domain consists of one a-helix and three b-sheets and the first two b-sheets interact with DNA (Allen et al. 1998; Lindner et al. 2010). In Arabidopsis, ABI4 binds the CE1 element CACC(G) in the promoter of its target genes (Niu et al. 2002; Bossi et al. 2009; Giraud et al. 2009; Yang et al. 2011). These target genes have been identified using microarray experiments (**Table 2.1**). All but one of the 95 ABI4 target genes identified contain the CACC sequence in their promoters, and 80% contain the CACCG sequence (Reeves et al. 2011). In the ABI4 target gene promoters a longer, less conserved region of 11 nucleotides overlaps with the CE1 elements (**Figure 2.2 A**). In the 11-nucleotide conserved region a higher occurrence of CACA(G) than CACC(G)

was found in the motif analysis. This CACA(G) is named here CE1*. ABI4 has repeatedly been shown to bind the CE1 element CACC(G), but the binding to CE1* has not been experimentally demonstrated. A docking analysis (de Vries et al. 2007; Dominguez, Boelens, and Bonvin 2003) was performed to understand how ABI4 interacts with DNA. This analysis uses all known information on ABI4 DNA binding (summarized in Suppl. Figure S2). The resulting model (**Figure 2.2 B and C**) closely resembles the determined 3D structure of the AtERF1 AP2 domain-DNA complex (Allen et al. 1998). Both transcription factors belong to the same family of transcription factors, although they have different binding preferences (Mizoi et al. 2012). This model predicts that ABI4 binds to the nine-nucleotide conserved near-palindromic sequence GACACGTGT (**Figure 2.2 A and C**) and this suggests that the DNA binding domain of ABI4 binds within the G-Box (**Figure 2.2 C**). Due to the palindromic sequence of the binding site, the binding domain could theoretically bind in two orientations, one 180° rotated opposed to the other, although amino acids outside of the DNA binding domain in the ABI4 protein excluded from the modeling might restrict the binding to one of the orientations by interacting with bases outside of the palindromic sequence. A role of

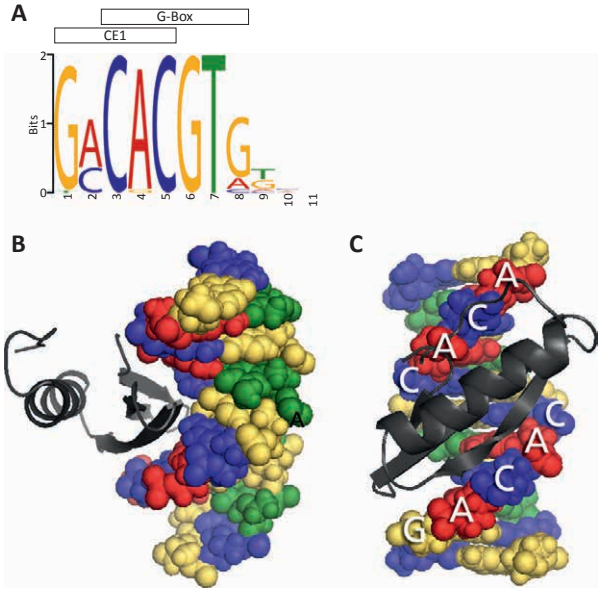


Figure 2.2. Modeling the ABI4 AP2 domain-DNA interaction*. (**A**) Sequence conservation of the motif found in the 500 nucleotides upstream region of 95 ABI4 target genes using the program MEME (Bailey et al. 2009). (**B**) Side view and (**C**) front view of the docking model that indicates how the AP2 domain interacts with a conserved motif containing both the CE* element and a G-Box. The docking input is described in Suppl. Figure S2. The clustered model with the lowest HADDOCK score, demonstrating most probable interactions, was visualized with PyMOL. Color code bases: red: A; Blue: C; Yellow: G; Green: T. Colors are the same as in Figure S2.*

ABI4 in repressing genes activated by G-box binding transcription factors has been proposed where ABI4 binding to the CE1 element blocks the binding of other transcription factors to the G-Box (Acevedo-Hernandez et al. 2005; Rook et al. 2006; Koussevitzky et al. 2007). This is exemplified by how ABI4 represses the CAB2 promoter by binding the CE1 element, which overlaps with the G-Box as observed in yeast one-hybrid assays (Koussevitzky et al. 2007). Other transcription factors were shown to bind and induce this same promoter, like CCA1 and HY5 (Andronis et al. 2008) and CCA1 and LHY (Xu et al. 2007). ABI4 might compete on the promoter with the G-Box binding transcription factors and represses the genes, due to the relatively weak inducing activity of ABI4 (Rook et al. 2006), which is in agreement with this model.

ABI4 acts as an activator and repressor of gene expression

ABI4 is an activator of gene expression (Soderman et al. 2000; Bossi et al. 2009; Reeves et al. 2011) as well as a repressor (Nakabayashi et al. 2005; Koussevitzky et al. 2007; Giraud et al. 2009; Reeves et al. 2011). ABI4-regulated transcripts in microarray experiments using genotypes with altered ABI4 activity are both up-regulated and down-regulated. The affected genes are involved in various processes including lipid metabolism, postembryonic development, photosynthesis, response to ABA stimulus, and chlorophyll metabolic processes (**Table 2.1**). The identified ABI4 targets (Reeves et al. 2011) are affected in the abi4 null mutants as well, showing highest overlap with water imbibed abi4 null mutant seeds (Suppl. File S1). ABI4 induces genes involved in seed maturation and ABA signaling (Reeves et al. 2011). ABI4 represses genes involved in photosynthesis, fatty acid biosynthesis, pigment metabolic processes, and wax metabolic processes. Although, differential expression in mutants is no proof of direct regulation by ABI4, compelling amount of evidence indicates ABI4 is able to both repress and induce transcription upon DNA binding.

ABI4 activates and represses metabolic and developmental processes

ABI4 transmits various signals to mediate repression and induction of transcription. These transcriptional changes can be linked to its repressing and activating role in metabolic and developmental processes during the life cycle of the plant (**Figure 2.3**).

Seed germination

ABI4 expression is detectable already in the globular stage during embryo development (Soderman et al. 2000). In germinating seeds ABI4 levels decrease with seedling establishment (Soderman et al. 2000; Shkolnik-Inbar and Bar-Zvi 2011). During germination and seedling establishment ABI4 protein levels go down as well, through proteasomal degradation of the protein. ABI4 overexpression does not affect the speed of germination. However, overexpression of ABI4 does repress germination when exogenous ABA is added to geminating seeds (Finkelstein et al. 2011).

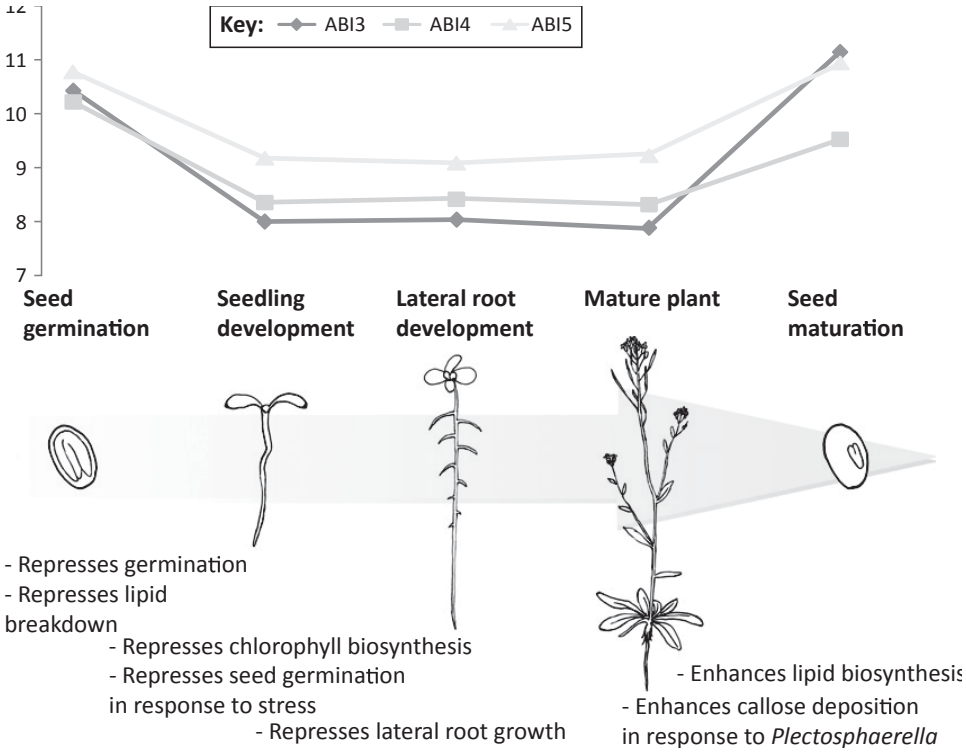


Figure 2.3. ABI4 is mainly active during seed germination and seed maturation. *The expression levels of ABI3, ABI4, and ABI5 during the plant life cycle show that the genes are tightly co-expressed. The role of ABI4 during the different developmental stages is indicated. Expression data taken from Genevestigator (Hruz et al. 2008). Scale indicates log_2 level of absolute signal in the database.*

Sugar-directed growth arrest

ABI4 was first identified as an abi germination mutant. Subsequently, abi4 mutants were identified in screens for reduced sucrose sensitivity as the sun6 mutant (Huijser et al. 2000), for sucrose insensitivity as the sis5 mutant (Laby et al. 2000), for impaired sucrose induction as the isi3 mutant (Rook et al. 2001), and for glucose insensitivity as the gin6 mutant (Arenas-Huertero et al. 2000). sis and gin screens were performed in conditions that cause sugar-directed growth arrest in wild type plants, which is a response of seeds to very high sugar levels present during germination. This response is characterized by delayed seed germination (Dekkers et al. 2004) followed by slow growth of the seedling, impaired chloroplast development and absence of cotyledon greening (To et al. 2003). abi4 mutants are insensitive to 6% glucose and lack all these inhibitory characteristics mentioned above (Arenas-Huertero et al. 2000; Huijser et al. 2000), except for sugar induced delay of germination (Price et al. 2003). abi4-1 mutants are fructose insensitive as well (Li et al.

2011) showing that ABI4 mediates the response to a variety of sugars. Accessions harboring particular natural variants of DE- LAY OF GERMINATION1 (DOG1) have elevated ABI4 mRNA levels, and these plants are oversensitive to 6% glucose (Teng et al. 2008). Similarly, in the SCARECROW mutant, ABI4 mRNA levels are elevated leading to glucose oversensitivity (Cui et al. 2012). Glucose stabilizes ABI4 protein levels in young seedlings and this is possibly the mechanism by which sugars promote sugar-directed growth arrest (Finkelstein et al. 2011). Sucrose might act through ABI4 as well. Diglyceride acyltransferase1 (DGAT1) is a direct target of ABI4 and a rate-limiting enzyme in triacylglycerol (TAG) biosynthesis. High levels of sucrose increase DGAT1 expression, possibly through increased activity of ABI4 in wild type seedlings. ABI4 is involved in trehalose signaling, as well. Root growth is inhibited in young Arabidopsis seedlings exposed to trehalose containing media due to the accumulation of T6P (Schluepmann et al. 2004). The abi4 mutant is less sensitive to the inhibitory effects of trehalose on root growth indicating the involvement of ABI4 in this process (Ramon et al. 2007).

Plastid-to-nucleus signaling and photosynthesis

Genes encoding photosynthesis components must be coordinately expressed for optimal photosynthesis. Most proteins involved in photosynthesis are encoded in the nucleus and are imported into the chloroplast. Plastid-to-nucleus signaling is needed to initiate and balance Photosynthesis-Associated Nuclear Gene (PhANG) expression. Several mutants have been identified that have a perturbed nuclear feedback and misexpress the PhANG gene CAB2 (Susek et al. 1993). The genes of several such genome uncoupled (gun) mutants were cloned and shown to be involved mainly in plastidic processes. The nature of the signal that is transported from the plastid-to-nucleus to balance PhANG expression is still not fully resolved (Kakizaki et al. 2009; Pfannschmidt 2010). ABI4 is a repressor of the PhANG gene CAB2 in young seedlings and a role for ABI4 was proposed in plastid-to-nucleus signaling as a negative regulator of PhANGs acting downstream of GUN1 (Koussevitzky et al. 2007). It was proposed that upon stress treatment a chloroplast-bound PHD type transcription factor with transmembrane domains is cleaved from the chloroplast outer membrane and is transported to the nucleus where it promotes ABI4 expression (Sun et al. 2011). Two other transcription factors regulating PhANG genes positively, Golden2-like 1 (AtGLK1) (Kakizaki et al. 2009) and HY5 (Kindgren et al. 2012), were proposed to respond to retrograde signals as well. The promoters of the PLASTOCYANIN1 (PC) and other PhANG genes are repressed by the sugar analog 2-deoxy-D-glucose (2DG). This repression is lost in the abi4 allele sucrose uncoupled6 (sun6) (Huijser et al. 2000). PC is no longer repressed by sucrose in the sun6 mutant (Oswald et al. 2001), showing that ABI4 is essential for the repressing effect of sugars on PC gene expression. 2DG represses overall photosynthetic activity in mature rosette plants and this repression is reduced in the sun6 mutant (Van Oosten et al. 1997). Possibly, the lack of sugar-repressed photosynthesis in abi4 explains the slightly increased growth rate, despite increased ABA levels in the abi4-1 mutant (Kerchev et al. 2011). These findings clearly show the importance of ABI4 in regulating PhANG expression at the physiological

level. Moreover, since most phenotypic differences of abi4 mutants compared to wild type are independent of added ABA, ABI4 most likely acts independent of ABA signaling in the plastid-to-nucleus signaling process.

Role in redox regulation

ABI4 function in plastid-to-nucleus signaling is linked to redox regulation (Pontier et al. 2007). Reactive oxygen species are produced in electron transfer pathways associated with photosynthesis. ABI4 binds to the CE1 motif in the promoter of ALTERNATIVE OXIDASE 1A (AOX1A) and represses this gene. AOX1A enhances reactive oxygen species (ROS) levels (Strodtkotter et al. 2009) and in the abi4-1 mutant, AOX1A is de-repressed, showing that ABI4 represses this gene (Giraud et al. 2009). However, ABI4 also plays a role in ROS metabolism through control of ascorbic acid levels. The vitamin c defective mutants vtc1 and vtc2 are mutated in genes encoding enzymes in the ascorbate biosynthesis pathway. In vtc1, vtc2, and abi4 mutant rosettes similar changes in transcriptional profiles were observed and in abi4 rosettes ascorbic acid levels are elevated, demonstrating a role for ABI4 in the regulation of the ascorbic acid pathway (Kerchev et al. 2011) and providing another mechanism by which ABI4 controls ROS levels in the plant.

Lipid biosynthesis and breakdown

In maturing Arabidopsis seeds, reserves in the embryo are mainly synthesized and stored as TAG. ABI4 both induces TAG biosynthesis and inhibits TAG breakdown. During seed maturation, DGAT1 is the rate-limiting enzyme in TAG biosynthesis. ABI4 binds to the promoter of DGAT1 and induces this gene, which is enhanced in the presence of ABA (Yang et al. 2011). ABI4 also promotes seed maturation by inducing oleosin- and dehydrin-encoding genes (Reeves et al. 2011), which are associated with lipid biosynthesis. In germinating seeds, TAG is mobilized and used in the gluconeogenesis pathway. This provides the embryo with energy before the initiation of photosynthesis. During germination, TAG is converted to eicosenoic acids and in abi4-1 mutants reduced eicosenoic acid levels were found when germinating on exogenously supplied ABA. Thus, ABI4 inhibits TAG breakdown in an ABA dependent way (Penfield et al. 2006).

Lateral root development

ABI4 is predominantly expressed in the seed and mRNA and protein levels rapidly decline during germination. However, a basic level of expression remains present during later developmental stages of the plant, including lateral root development. The number of lateral roots is higher in abi4 mutants and lower in ABI4 overexpressing plants (Shkolnik-Inbar and Bar-Zvi 2010). This effect of ABI4 on lateral root development is possibly mediated through repression of PIN1, which encodes the dominating auxin efflux carrier and thereby affects auxin distribution (Shkolnik-Inbar and Bar-Zvi 2010). Interestingly, the abi3-6 mutant

has reduced lateral root development in the presence of auxin, showing the involvement of ABI3 in the control of lateral root development. ABI3 is an ABI4 induced gene (Brady et al. 2003) and possibly ABI4 controls PIN1 expression through ABI3. Low levels of nitrate stimulate lateral root formation, whereas high levels suppress lateral root formation. ABI4 mediates this nitrate signaling effect on root branching because the inhibition of lateral root formation by high nitrate levels is reduced in abi4-1 mutants, while no difference in lateral root number can be detected in plants grown on low nitrate levels (Signora et al. 2001).

Cell wall modifications

Two reports link ABI4 to cell wall modifications. ABI4 can mediate biotic stress responses as well, in addition to its role in abiotic stress tolerance. Beta-amino-butyric acid (BABA) is a non-protein amino acid that induces callose deposition in Arabidopsis and protects the plant from pathogen attack. In abi4-1 plants callose does not accumulate, making abi4 more susceptible to infection by the pathogen *Plectosphaerella cucumerina* (Ton and Mauch-Mani 2004). ABI4 is also involved in pectin rearrangement. The LM6 1,5-Arabinan epitope (LM6) is a marker of pectin rearrangement and LM6 levels are reduced in the abi4-1 mutant (Talboys et al. 2011).

CONCLUDING REMARKS

Here, we discuss the function of ABI4 at the molecular level by proposing its evolutionary history, its DNA binding and its effect on transcription, and at the physiological and developmental level by describing its role in various processes. ABI4 appears to have an activating and repressing effect at both the molecular and physiological level. The phylogenetic analysis revealed three conserved regions, the AP2 DNA binding domain, the CMIV-1 domain and a remarkably conserved 'ABI4 motif' in the ABI4 orthologs in angiosperms. *P. patens* and *S. moellendorffii* encode proposed ABI4 orthologs lacking this motif, while in *P. taeda* the ABI4 motif is conserved as well. Domain swap and cross species complementation experiments should demonstrate functionality of lower plant ABI4 proteins, describe a function for the ABI4 motif and answer the question of whether ABI4 is needed for drought responses typically required in land plants.

ABI4 activates and represses genes by binding the CE1 element CACC(G) (Reeves et al. 2011). However, the 11-nucleotide conserved region present in target promoters has a higher occurrence of the CE1* element, CACA(G) and likely, ABI4 binds this element in vivo as well. The docking model predicts that ABI4 binds within the G-Box. This model needs experimental validation by structure determination of DNA bound ABI4. The ABI4 mode of binding is in contrasts with bZIP transcription factors, that bind as heterodimers to the G-box. Whether ABI4 can form functional heterodimers with bZIPs is not known to date; ABI4 does not interact with ABI5 in yeast two-hybrid experiments (Nakamura et al. 2001).

ABI4 is part of a highly complex network of transcription factors that together integrate signaling responses to light, plastid-to-nucleus communication, ABA, and sugars. These processes together are important in coordinating germination, seedling establishment, and onset of photosynthesis. Revealing the interaction of ABI4 with other transcriptional regulators might help in further understanding its role in this complex network and in these different processes. In several processes, ABI4 is an essential ABA signaling component required for the response to ABA, such as its role in the synthesis and breakdown of lipids. However, there are several processes where ABI4 seems to act ABA independently, including controlling photosynthesis, lateral root development, sugar-directed growth arrest, and redox regulation. The precise role of ABI4 in plastid-to-nucleus is not fully understood but its repressing role in photosynthesis is evident as abi4 seedlings green prematurely (Finkelstein 1994), abi4 rosettes have higher biomass (Kerchev et al. 2011), and abi4 plants show less repression on the rate of photosynthesis in response to 2DG. This suggests that it plays an important role as a repressor of photosynthesis during the life cycle of the plant. ABI4 also mediates the response to sugar-directed growth arrest leading to repression of visible greening of the cotyledons, pointing to its repressive effect on photosynthesis. Overall, the versatile transcription factor ABI4 is most important during seed maturation and germination but it is required in many other developmental and physiological processes throughout the life cycle of plants.

ACKNOWLEDGEMENTS

We would like to thank MSc. Arjen Jakobi and Dr Marc van Dijk (Bijvoet Center for Biomolecular Research, Utrecht University) for their advice in 3D modeling of the AP2 domain of ABI4 and for performing the docking experiment to predict ABI4 DNA binding. The authors also want to thank three anonymous reviewers, whose comments significantly improved the manuscript. Work in the authors' laboratories was supported by the Dutch Research Council, NWO, through TOP-GO program (854.10.011) and by Bio4Energy, a strategic research environment appointed by the Swedish government.

SUPPLEMENTARY MATERIAL

All Supplementary Figures are shown in the following pages. **Supplementary File S1**, associated to Table 2.1, can be found at the publisher's website:
http://dx.doi.org/10.1016/j.tplants.2012.10.004

Supplementary Figure S1. ABI4 sequence alignment.
Supplementary Figure S2. ABI4-DNA binding.
Supplementary Figure S3. Phylogenetic tree of the DREB group IV subfamily members.
Supplementary Figure S4. Original sequence identifiers for the DREB group IV sequences.
Supplementary File S1. Original data shown in Table 2.1.

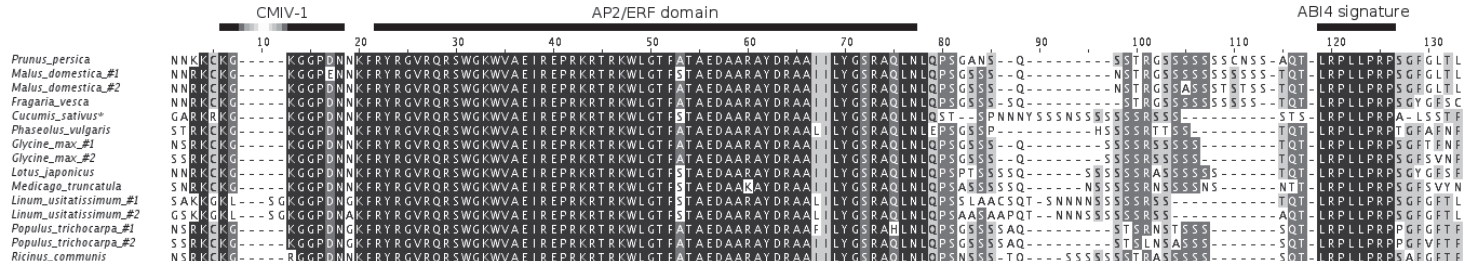


Figure S1. ABI4 sequence alignment. *The figure shows the aligned ABI4 sequences in the region including the AP2/ERF domain and the conserved motifs, as indicated. Protein identifiers are mapped to species names as reported in Suppl. Figure S4; an asterisk indicates proteins predicted from genome. The 56 residues belonging to the AP2/ERF domain were identified according to the domain model from the NCBI Conserved Domain Database (www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml; domain ID: cd00018). The position of the CMIV-1 motif (consensus [R/K]GRGGPxN) and the newly discovered ABI4 motif are also shown. Shaded columns represent residues with at least 50% identity (the higher the darker). Residues conserved in all ABI4 sequences are marked at the bottom of the figure with a black star; additional residues conserved in spermatophytes (including the ABI4 motif) are marked with a grey star. Original protein identifiers are reported in Suppl. Figure S4.*

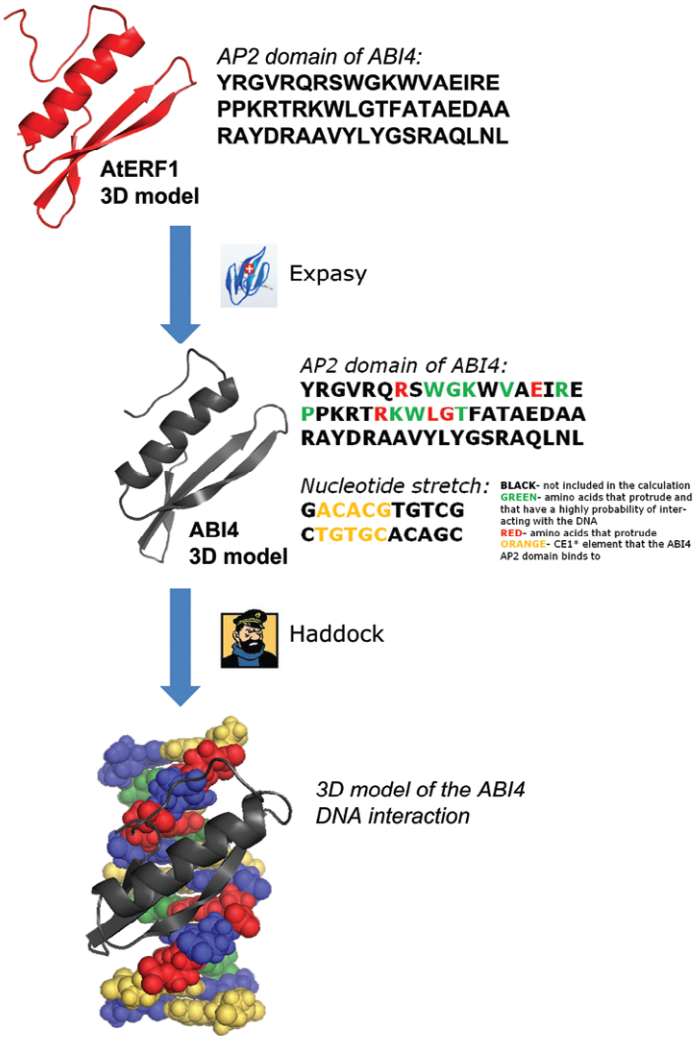


Figure S2. ABI4-DNA binding. *3D model of the AP2 domain of ABI4: The model of the AP2 domain of ABI4 was based on the NMR derived 3D structure of AtERF1. The structure of AtERF1 was determined by heteronuclear multidimensional NMR. This structure was used by the SwissModel program accessible at* http://swissmodel.expasy.org *to predict the 3D structure of the AP2 domain of ABI4. This model of the AP2 domain has a Z-score of -0.33. 3D model of the ABI4 DNA interaction: The ABI4 AP2 domain model was used for docking on the DNA helix using Haddock* http://haddock.science.uu.nl*. The complete set of parameters used for the input of the haddock program can be found in Suppl. File S1 (online). Most nucleotides included in the modeling were selected if they had a tail pointing in the direction of the DNA, except for one amino acid: the Glumatic acid in the second strand is a confirmed essential amino acid required for ABI4 function. The abi4-104 has a single nucleotide substitution leading to missense mutation of a glutamic acid to a lysine. The clustered model with the lowest Haddock score was visualized with PyMOL (*www.pymol.org*). Each nucleotide has the same color as assigned by the MEME program, described in Figure 2.2. The amino acids are depicted in dark grey.*

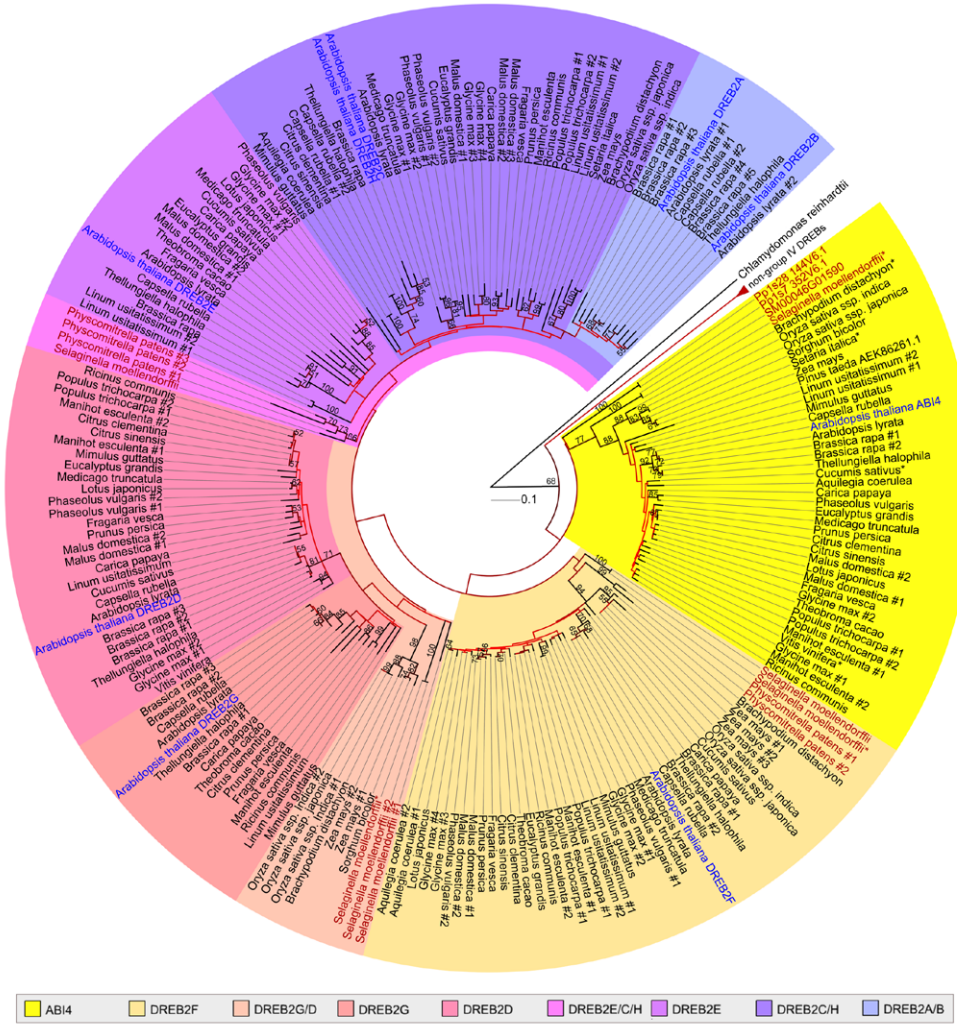


Figure S3. Phylogenetic tree of the DREB group IV subfamily members. *Group IV DREB sequences were collected from different source databases (see Suppl. Figure S4) based on the similarity with the AP2/ERF domain of A. thaliana ABI4. For S. moellendorffii and P. patens, additional protein sequences were predicted from the genome; non-group IV DREB representatives were also collected for these two species and A. thaliana (the group is shown as a collapsed branch in the phylogeny). The closest C. reinhardtii sequence was retrieved to provide a root for the tree. The core AP2/ERF domain was extended to the conserved flanking regions to allow a more sensitive alignment of the sequences, performed using MAFFT v6.864b algorithm (Katoh et al. 2002). The final phylogenetic tree was built using MEGA5 software (Tamura et al. 2011) with the Neighbor-Joining method (1000 bootstrap replications). Support values higher than 50 are shown above the branches; a color gradient from red to black gives an additional indication of branches reliability (from low to high, respectively). DREB group IV sequences from A. thaliana are colored in blue. Proteins from other species that clustered with known DREB members were highlighted using different background colors, to improve the visualization of the different subgroups in the transcription factor family (ABI4 proteins in yellow, see color code). Clusters not clearly belonging to a specific DREB group were also colored. All sequences from P. patens and S. moellendorffii are shown in red print to allow the comparison between the positions of DREB- and ABI4-like sequences from these two species in the tree. Leaves are annotated with species names, followed by a number in case of multiple occurrences within the same DREB group. The original protein identifiers from Phytozome, Plaza and Genbank are reported in Suppl. Figure S4. The phylogenetic tree graphics was generated with the iTOL webtool (*http://itol.embl.de/index.shtml*) (Letunic and Bork 2011).*

DREB GROUP	SPECIES ID	ORIGINAL ID	SOURCE DATABASE
root	Chlamydomonas reinhardtii	Cre01.g063350.t1.1	Phytozome 8.0
ABI4	Aquilegia coerulea	AcoGoldSmith_v1.017021m	Phytozome 8.0
	Arabidopsis lyrata	AL4G26970	Plaza 2.5
	Arabidopsis thaliana (ABI4)	AT2G40220	Plaza 2.5
	Brachypodium distachyon*	predicted protein	Phytozome 8.0
	Brassica rapa #1	Bra004978	Phytozome 8.0
	Brassica rapa #2	Bra000178	Phytozome 8.0
	Capsella rubella	Carubv10025253m	Phytozome 8.0
	Carica papaya	CP01288G00020	Plaza 2.5
	Citrus sinensis	orange1.1g038194m	Phytozome 8.0
	Citrus clementina	clementine0.9_031263m	Phytozome 8.0
	Cucumis sativus*	predicted protein	Phytozome 8.0
	Eucalyptus grandis	Eucgr.J00187.1	Phytozome 8.0
	Fragaria vesca	FV7G29790	Plaza 2.5
	Glycine max #1	GM02G43240	Plaza 2.5
	Glycine max #2	GM14G06290	Plaza 2.5
	Linum usitatissimum #1	Lus10009834	Phytozome 8.0
	Linum usitatissimum #2	Lus10040944	Phytozome 8.0
	Lotus japonicus	LjZG033340	Plaza 2.5
	Malus domestica #1	MD04G0020700	Plaza 2.5
	Malus domestica #2	MD00G426860	Plaza 2.5
	Manihot esculenta #1	ME00435G00110	Plaza 2.5
	Manihot esculenta #2	ME10689G00620	Plaza 2.5
	Medicago truncatula	MT5G082950	Plaza 2.5
	Mimulus guttatus	mgv1a024901m	Phytozome 8.0
	Oryza sativa ssp. japonica	OS05G28350	Plaza 2.5
	Oryza sativa ssp. indica	OSINDICA_05G25070	Plaza 2.5
	Phaseolus vulgaris	Phvulv091021931m	Phytozome 8.0
	Physcomitrella patens #1	Pp1s7_352V6.1	Phytozome 8.0
	Physcomitrella patens #2	Pp1s28_144V6.1	Phytozome 8.0
	Pinus taeda	AEK862261.1	GenBank
	Populus trichocarpa #1	PT10G18230	Plaza 2.5
	Populus trichocarpa #2	PT08G08930	Plaza 2.5
	Prunus persica	ppa015230m	Phytozome 8.0
	Ricinus communis	RC29814G00250	Plaza 2.5
	Selaginella moellendorffii*	predicted protein (scaffold 96)	Plaza 2.5
	Selaginella moellendorffii	SM00046G01590	Plaza 2.5
	Setaria italica*	predicted protein	Phytozome 8.0
	Sorghum bicolor	SB09G016550	Plaza 2.5
	Thellungiella halophila	Thhalv10017614m	Phytozome 8.0
	Theobroma cacao	TC10G001200	Plaza 2.5
	Vitis vinifera*	predicted protein	Plaza 2.5
	Zea mays	ZM06G22750	Plaza 2.5
DREB2F	Aquilegia coerulea #1	AcoGoldSmith_v1.015954m	Phytozome 8.0
	Aquilegia coerulea #2	Aquca_009_00282.1	Phytozome 8.0
	Arabidopsis lyrata	AL5G28090	Plaza 2.5
	Arabidopsis thaliana (DREB2F)	AT3G57600	Plaza 2.5
	Brachypodium distachyon	BD1G72990	Plaza 2.5
	Brassica rapa #1	Bra007341	Phytozome 8.0
	Brassica rapa #2	Bra014617	Phytozome 8.0
	Capsella rubella	Carubv10017800m	Phytozome 8.0
	Carica papaya	CP00092G00160	Plaza 2.5
	Citrus clementina	clementine0.9_017423m	Phytozome 8.0
	Citrus sinensis	orange1.1g042353m	Phytozome 8.0
	Cucumis sativus	Cucsa.336890.1	Phytozome 8.0
	Eucalyptus grandis	Eucgr.A02390.1	Phytozome 8.0
	Fragaria vesca	FV7G27980	Plaza 2.5
	Glycine max #1	GM10G21850	Plaza 2.5
	Glycine max #2	GM02G31350	Plaza 2.5
	Glycine max #3	GM19G32380	Plaza 2.5
	Glycine max #4	GM03G29530	Plaza 2.5
	Linum usitatissimum #1	Lus10011319	Phytozome 8.0
	Linum usitatissimum #2	Lus10012210	Phytozome 8.0
	Lotus japonicus	LjTG045820	Plaza 2.5
	Malus domestica #1	MD00G199220	Plaza 2.5
	Malus domestica #2	MD12G005970	Plaza 2.5
	Manihot esculenta #1	ME07046G00090	Plaza 2.5
	Manihot esculenta #2	ME04526G00120	Plaza 2.5
	Mimulus guttatus	mgv1a011288m	Phytozome 8.0
	Medicago truncatula	MT7G097190	Plaza 2.5
	Oryza sativa ssp. indica	OSINDICA_03G06780	Plaza 2.5
	Oryza sativa ssp. japonica	OS03G07830	Plaza 2.5
	Phaseolus vulgaris #1	Phvulv091021718m	Phytozome 8.0
	Phaseolus vulgaris #2	Phvulv091021082m	Phytozome 8.0
	Physcomitrella patens #1	Pp1s162_92V6.1	Phytozome 8.0
	Physcomitrella patens #2	Pp1s292_5V6.1	Phytozome 8.0
	Populus trichocarpa #1	PT06G05130	Plaza 2.5
	Populus trichocarpa #1	PT16G05170	Plaza 2.5
	Prunus persica	ppa009616m	Phytozome 8.0
	Ricinus communis	RC27810G00100	Plaza 2.5
	Selaginella moellendorffii	SM00021G00250	Plaza 2.5
	Selaginella moellendorffii*	predicted protein (scaffold31)	Phytozome 8.0
	Thellungiella halophila	Thhalv10006489m	Phytozome 8.0
	Theobroma cacao	TC00G061640	Plaza 2.5
	Zea mays #1	ZM09G27590	Plaza 2.5
	Zea mays #2	ZM09G27620	Plaza 2.5
	Zea mays #3	ZM01G05280	Plaza 2.5
DREB2G/D	Brachypodium distachyon	BD3G12680	Plaza 2.5
	Oryza sativa ssp. Indica #1	OSINDICA_08G44000	Plaza 2.5
	Oryza sativa ssp. Indica #2	OSINDICA_08G43950	Plaza 2.5
	Oryza sativa ssp. japonica	OS08G45110	Plaza 2.5
	Selaginella moellendorffii #1	SM00009G00130	Plaza 2.5
	Selaginella moellendorffii #2	SM00011G00650	Plaza 2.5
	Selaginella moellendorffii*	predicted protein (scaffold102	Phytozome 8.0
	Sorghum bicolor	SB07G023575	Plaza 2.5
	Zea mays #1	ZM01G35700	Plaza 2.5
	Zea mays #2	ZM04G10250	Plaza 2.5

DREB GROUP	SPECIES ID	ORIGINAL ID	SOURCE DATABASE
DREB2G	Arabidopsis lyrata	AL6G18490	Plaza 2.5
	Arabidopsis thaliana (DREB2G)	AT5G18450	Plaza 2.5
	Brassica rapa #1	Bra006460	Phytozome 8.0
	Brassica rapa #2	Bra002159	Phytozome 8.0
	Brassica rapa #3	Bra023674	Phytozome 8.0
	Capsella rubella	Carubv10002425m	Phytozome 8.0
	Carica papaya	CP00051G00230	Plaza 2.5
	Citrus clementina	clementine0.9_021778m	Phytozome 8.0
	Fragaria vesca	FV0G27060	Plaza 2.5
	Linum usitatissimum	Lus10003581	Phytozome 8.0
	Manihot esculenta	ME03614G02750	Plaza 2.5
	Mimulus guttatus	mgv1a018752m	Phytozome 8.0
	Prunus persica	ppa026178m	Phytozome 8.0
	Ricinus communis	RC29680G00760	Plaza 2.5
	Thellungiella halophila	Thhalv10015907m	Plaza 2.5
	Theobroma cacao	TC05G025670	Plaza 2.5
DREB2D	Arabidopsis lyrata	AL2G24420	Plaza 2.5
	Arabidopsis thaliana (DREB2D)	AT1G75490	Plaza 2.5
	Brassica rapa #1	Bra003752	Phytozome 8.0
	Brassica rapa #2	Bra008200	Phytozome 8.0
	Brassica rapa #3	Bra015840	Phytozome 8.0
	Capsella rubella	Carubv10020960m	Phytozome 8.0
	Carica papaya	CP00099G00190	Plaza 2.5
	Citrus clementina	clementine0.9_019752m	Phytozome 8.0
	Citrus sinensis	orange1.1g036655m	Phytozome 8.0
	Cucumis sativus	Cucsa.231820.1	Phytozome 8.0
	Eucalyptus grandis	Eucgr.F01607.1	Phytozome 8.0
	Fragaria vesca	FV2G38990	Plaza 2.5
	Glycine max #1	GM17G37350	Plaza 2.5
	Glycine max #2	GM14G07620	Plaza 2.5
	Linum usitatissimum	Lus10010632	Phytozome 8.0
	Lotus japonicus	LJ0G19890	Plaza 2.5
	Malus domestica #1	MD00G3424880	Plaza 2.5
	Malus domestica #2	MD12G015320	Plaza 2.5
	Manihot esculenta #1	ME09657G00130	Plaza 2.5
	Manihot esculenta #2	ME07012G00750	Plaza 2.5
	Medicago truncatula	MT1G023170	Plaza 2.5
	Mimulus guttatus	mgv1a018624m	Phytozome 8.0
	Phaseolus vulgaris #1	Phvulv091010724m	Phytozome 8.0
	Phaseolus vulgaris #2	Phvulv091004912m	Phytozome 8.0
	Populus trichocarpa #1	PT05G24330	Plaza 2.5
	Populus trichocarpa #2	PT02G02900	Plaza 2.5
	Prunus persica	ppa023248m	Phytozome 8.0
	Ricinus communis	RC30170G04700	Plaza 2.5
	Thellungiella halophila	Thhalv10019142m	Phytozome 8.0
	Vitis vinifera	VV18G09710	Plaza 2.5
DREB2E/C/H	Physcomitrella patens #1	Pp1s120_13V6.1	Phytozome 8.0
	Physcomitrella patens #2	Pp1s23_27V6.1	Phytozome 8.0
	Physcomitrella patens #3	Pp1s81_44V6.1	Phytozome 8.0
	Selaginella moellendorffii	SM00000G05140	Plaza 2.5
DREB2E	Arabidopsis lyrata	AL4G24680	Plaza 2.5
	Arabidopsis thaliana (DREB2E)	AT2G38340	Plaza 2.5
	Brassica rapa	Bra005113	Phytozome 8.0
	Capsella rubella	Carubv10024477m	Phytozome 8.0
	Carica papaya	CP00005G02250	Phytozome 8.0
	Cucumis sativus	Cucsa.130200.1	Phytozome 8.0
	Eucalyptus grandis	Eucgr.K02952.1	Phytozome 8.0
	Fragaria vesca	FV6G09650	Plaza 2.5
	Glycine max #1	GM18G43750	Plaza 2.5
	Glycine max #2	GM07G19220	Plaza 2.5
	Linum usitatissimum #1	Lus10012226	Phytozome 8.0
	Linum usitatissimum #2	Lus10002855	Phytozome 8.0
	Lotus japonicus	LJ1G025460	Plaza 2.5
	Malus domestica #1	MD00G426580	Plaza 2.5
	Malus domestica #2	MD04G013230	Plaza 2.5
	Medicago truncatula	MT7G070220	Plaza 2.5
	Phaseolus vulgaris	Phvulv091017860m	Phytozome 8.0
	Thellungiella halophila	Thhalv10017658m	Phytozome 8.0
	Theobroma cacao	TC05G001640	Plaza 2.5
DREB2C/H	Aquilegia coerulea	Aquca_004_00452.1	Phytozome 8.0
	Arabidopsis lyrata	AL4G27210	Plaza 2.5
	Arabidopsis thaliana (DREB2C)	AT2G40340	Plaza 2.5
	Arabidopsis thaliana (DREB2H)	AT2G40350	Plaza 2.5
	Brachypodium distachyon	BD2G29960	Plaza 2.5
	Brassica rapa	Bra017011	Phytozome 8.0
	Capsella rubella #1	Carubv10023556m	Phytozome 8.0
	Capsella rubella #2	Carubv10023590m	Phytozome 8.0
	Carica papaya	CP00019G01230	Plaza 2.5
	Citrus clementina	clementine0.9_015025m	Phytozome 8.0
	Citrus sinensis	orange1.1g019546m	Phytozome 8.0
	Cucumis sativus	Cucsa.136780.1	Phytozome 8.0
	Eucalyptus grandis	Eucgr.G03094.1	Phytozome 8.0
	Fragaria vesca	FV7G30250	Plaza 2.5
	Glycine max #1	GM02G42960	Plaza 2.5
	Glycine max #2	GM14G06080	Plaza 2.5
	Glycine max #3	GM06G45680	Plaza 2.5
	Glycine max #4	GM12G11150	Plaza 2.5
	Linum usitatissimum #1	Lus10023633	Phytozome 8.0
	Linum usitatissimum #2	Lus10034902	Phytozome 8.0
	Malus domestica #1	MD00G250980	Plaza 2.5
	Malus domestica #2	MD04G002330	Plaza 2.5
	Malus domestica #3	MD00G048410	Plaza 2.5
	Manihot esculenta	ME05009G00110	Plaza 2.5
	Medicago truncatula	MT5G083340	Plaza 2.5
	Mimulus guttatus	mgv1a010327m	Phytozome 8.0
	Oryza sativa ssp. indica	OSINDICA_05G24500	Plaza 2.5
	Oryza sativa ssp. japonica	OS05G27930	Plaza 2.5
	Phaseolus vulgaris #1	Phvulv091021762m	Phytozome 8.0
	Phaseolus vulgaris #2	Phvulv091021763m	Phytozome 8.0
	Populus trichocarpa #1	PT08G07170	Plaza 2.5
	Populus trichocarpa #2	PT10G17960	Plaza 2.5
	Prunus persica	ppa007606m	Phytozome 8.0
	Ricinus communis	RC29729G00250	Plaza 2.5
	Setaria italica	Si002067m	Phytozome 8.0
	Thellungiella halophila	Thhalv10016913m	Phytozome 8.0
	Zea mays	ZM08G14120	Plaza 2.5
DREB2A/B	Arabidopsis lyrata #1	AL6G05000	Plaza 2.5
	Arabidopsis lyrata #2	AL3G11450	Plaza 2.5
	Arabidopsis thaliana (DREB2A)	AT5G05410	Plaza 2.5
	Arabidopsis thaliana (DREB2B)	AT3G11020	Plaza 2.5
	Brassica rapa #1	Bra009112	Phytozome 8.0
	Brassica rapa #2	Bra005852	Phytozome 8.0
	Brassica rapa #3	Bra028759	Phytozome 8.0
	Brassica rapa #4	Bra029889	Phytozome 8.0
	Brassica rapa #5	Bra034159	Phytozome 8.0
	Capsella rubella #1	Carubv10001099m	Phytozome 8.0
	Capsella rubella #2	Carubv10013771m	Phytozome 8.0
	Thellungiella halophila	Thhalv10021161m	Phytozome 8.0

Figure S4. Original sequence identifiers for the DREB group IV sequences. *The table reports the mapping between the species-based name of the protein (as found in the figures) and the original identifier and source database. Protein sequences were retrieved or predicted from Phytozome (*www.phytozome.net*, version 8.0), Plaza (*http://bioinformatics.psb.ugent.be/plaza*, version 2.5) or Genbank (*www.ncbi.nlm.nih.gov/genbank*) (Pinus (Pinus taeda only).*

3

The phylogeny of C/S1 bZIP transcription factors reveals a shared algal ancestry and the pre-angiosperm translational regulation of S1 transcripts

Alessia Peviani[1], Jeroen Lastdrager[2], Johannes Hanson[2,3], and Berend Snel[1]

[1] *Theoretical Biology and Bioinformatics, Utrecht University, Utrecht, The Netherlands*

[2] *Molecular Plant Physiology, Institute of Environmental Biology, Utrecht University, Utrecht, The Netherlands*

[3] *Umeå Plant Science Centre, Department of Plant Physiology, Umeå University, Umeå, Sweden*

under review in Scientific Reports

ABSTRACT

Basic leucine zippers (bZIPs) form a large plant transcription factor family. C and S1 bZIP groups can heterodimerize, fulfilling crucial roles in seed development and stress response. S1 sequences also harbor a unique regulatory mechanism, termed Sucrose-Induced Repression of Translation (SIRT). The conservation of both C/S1 bZIP interactions and SIRT remains poorly characterized in non-model species, leaving their evolutionary origin uncertain and limiting crop research. In this work, we explored recently published plant sequencing data to establish a detailed phylogeny of C and S1 bZIPs, investigating for the first time their intertwined role in plant evolution, and the origin of SIRT.

Our analyses clarified C and S1 bZIP orthology relationships in angiosperms, and identified S1 sequences in gymnosperms. We experimentally showed that the gymnosperm orthologs are regulated by SIRT, tracing back the origin of this unique regulatory mechanism to the ancestor of seed plants. Additionally, we discovered an earlier S ortholog in the charophyte algae *Klebsormidium flaccidum*, together with a C ortholog. This suggests that C and S groups originated by duplication from a single algal proto-C/S ancestor. Based on our observations, we propose a model wherein the C/S1 bZIP dimer network evolved in seed plants from pre-existing C/S bZIP interactions.

INTRODUCTION

The basic leucine zipper (bZIP) family of transcription factors is one of the largest in plants, represented in angiosperms by 13 subfamilies involved in the regulation of fundamental physiological and developmental processes (Jakoby et al. 2002; Côrrea et al. 2008). The participation of several members in the response to environmental cues makes this transcription factor family a promising research subject for crop yield and stress resistance improvement, motivating the recent burst of genome-wide bZIP analyses in a variety of cultivated species (Liao et al. 2008; Nijhawan et al. 2008; Wang et al. 2011; Wei et al. 2012; Baloglu et al. 2014; Hwang et al. 2014; Jin, Xu, and Liu 2014; Liu et al. 2014; Liu and Chu 2015; Pourabed et al. 2015).

Within the S bZIP subfamily, the group of orthologs named S1 (Ehlert et al. 2006) is specifically involved in seed development and metabolic reprogramming in response to stress (Hanson et al. 2008; Alonso et al. 2009; Weltmeier et al. 2009; Dietrich et al. 2011; Ma et al. 2011; Iglesias-Fernandez et al. 2013), probably downstream of SnRK kinase signaling (Baena-González et al. 2007; Tomé et al. 2014). The specificity of S1 bZIP activity in low energy conditions is achieved via a unique regulatory mechanism, termed Sucrose-Induced Repression of Translation (SIRT) (Rook et al. 1998; Wiese et al. 2004; Rahmani et al. 2009), which relies on the presence of a characteristic uORF at the 5' leader of the bZIP transcript (5'uORF). According to the current SIRT model, high sugar availability increases the affinity of the ribosome for the 5'uORF, possibly triggering its translation into a Sucrose Control (SC) peptide, and preventing protein synthesis at the bZIP main ORF (mORF) (Rahmani et al. 2009; Schepetilnikov et al. 2013). While the SC-peptide remains to be isolated, key

amino acid residues and other 5'uORF features appear necessary for SIRT to take place, and accordingly are conserved across S1 bZIP orthologs (Rahmani et al. 2009). Importantly, SIRT is considered a characterizing feature of these transcription factors, as no similar uORF sequences or uORF-based repression mechanism by sucrose has been observed in other bZIP groups, or more generally in other plant gene families (von Arnim et al. 2014).

A further layer of regulation is achieved through the formation of heterodimers between S1 and C bZIP subfamily members, also characterized as regulators in seed development (Weltmeier et al. 2006; Alonso et al. 2009) and stress response (Kaminaka et al. 2006). The interaction between S1 and C bZIP sequences relies on the compatibility of their bZIP domains (Deppmann et al. 2004), likely as a consequence of the close phylogenetic relationship between C and S subfamilies; indeed it has been proposed that S subfamily emerged from the C subfamily in an angiosperm-specific duplication event (Côrrea et al. 2008). Remarkably, while other bZIP subfamilies are mainly involved in simple homo- or quasi-homodimerization events (i.e. dimerization between close paralogs) (Deppmann et al. 2006), C/S1 interactions give rise to a more complex dimerization network, as shown in the model species arabidopsis (*Arabidopsis thaliana*) (Ehlert et al. 2006; Hanson et al. 2008). Moreover, C and S1 bZIP members interact specifically while avoiding dimerization with other S sequences (Ehlert et al. 2006; Weltmeier et al. 2006), indicating the presence of selective pressure to prevent promiscuity.

The functional importance of C/S1 bZIP heterodimers for transcriptional activity was reported in a series of double overexpression experiments, which showed a strong synergistic effect on both up- and down-regulation of selected targets (Weltmeier et al. 2006; Alonso et al. 2009; Kang et al. 2010). At the same time, the lack of drastic phenotypes in loss-of-function mutants indicates that the system allows for a certain degree of redundancy (Dietrich et al. 2011; Hartings et al. 2011; Iglesias-Fernandez et al. 2013). These observations, together with the individual C and S1 bZIPs interaction preferences (Ehlert et al. 2006), tissue- and condition-dependent expression patterns (Weltmeier et al. 2009), translational regulation by SIRT (Rahmani et al. 2009), and phosphorylation of C group members (Schutze et al. 2008; Kirchler et al. 2010), seem to allow for a signal integration system endowed with tremendous flexibility in the fine-tuning of target genes. It has been proposed that such system could help optimizing the effectiveness of stress response depending on the specific environmental threat (Llorca et al. 2014). However, experimental studies of heterodimerization preferences and other bZIP-specific properties such as SIRT are still limited to a handful of dicots (Rugner et al. 2001; Strathmann et al. 2001; Ehlert et al. 2006; Rahmani et al. 2009; Thalor et al. 2012).

Phylogenetics represents a powerful resource to assess the C/S1 bZIP transcription factor network conservation across plant species, providing a framework for the transfer of functional information between model organisms and crops. Unfortunately, while the identification of C bZIP orthologs appears relatively straightforward, putative S1 sequences

proved troublesome in previous phylogenetic studies (Côrrea et al. 2008). Duplications within the S subfamily gave rise to the largest and most diverse group among bZIP transcription factors, hindering the assessment of orthology relationships between plant lineages. For example, BLAST searches of S1 bZIP proteins suggest the presence of orthologs in both dicots and monocots (Hayden and Jorgensen 2007; Rahmani et al. 2009), while current phylogenetic reconstructions are unable to validate this (Côrrea et al. 2008). The identification and study of S1 bZIP orthologs in novel species would therefore benefit from a solid classification reference.

A high-resolution phylogeny yielding reliable ortholog identification could also clarify the early evolution of the C/S1 bZIP transcription factor network, which remains unexplored. Currently, we can extrapolate that heterodimer formation was not possible before the emergence of flowering plants, as the first C and S bZIP representatives were reported in bryophytes and angiosperms, respectively (Côrrea et al. 2008). However, the limited quantity and quality of plant sequencing data available at that time (Côrrea et al. 2008) suggests that this might be an incomplete picture. The accurate characterization of both C and S bZIP early members, and especially of the younger S subfamily, is therefore crucial to generate sensible hypotheses on the dimerization network origin.

Recently, a plethora of plant genome projects made available valuable sequence information for comparative studies, not only among crops, but also in poorly characterized early branching lineages (Banks et al. 2011; Der et al. 2011; Birol et al. 2013; Bushart et al. 2013; Chamala et al. 2013; Nystedt et al. 2013; Zimmer et al. 2013; Brouwer et al. 2014; Hori et al. 2014; Neale et al. 2014; Wolf et al. 2015). By taking advantage of this resource, we investigated for the first time the intertwined role of the C/S1 bZIP transcription factor network members during plant evolution. The new data allowed us to resolve S subfamily classification in angiosperms, finally confirming the presence of S1 bZIP orthologs in both dicots and monocots. The analysis of C bZIP orthologs also clarified the pattern of gene duplications and losses in flowering plants, providing additional information for comparative studies. Surprisingly, our results showed S subfamily to be much older than previously thought, with basal S sequences discovered before the emergence of land plants, and possibly originating at the same time as the C subfamily. We additionally proved the presence of S1 bZIPs in the common ancestor of spermatophytes, as supported by *in vivo* experiments confirming regulation of putative S1 orthologs from gymnosperms by SIRT. Finally, we outlined new hypotheses on the origin and specialization of the C/S1 bZIP dimerization network during plant evolution, discussing in particular its putative emergence from pre-existing C/S interactions.

RESULTS

C and S1 bZIP subfamilies show lineage-specific patterns of gene duplications and losses in angiosperms

The C/S1 bZIP transcription factor network is still poorly characterized in non-model species. Here we analyze a large set of recently published plant genomes to thoroughly assess C and S1 bZIP transcription factors conservation across flowering plants, and in particular S1 bZIP ortholog relationships between dicots and monocots. First, we assembled a database of sequenced angiosperm species encompassing previously unexplored lineages, such as basal angiosperms, asterids, and non-poales monocots (Suppl. Table S1). C and S1 orthlogs were then collected through iterative BLAST and HMMER searches using reference bZIPs as queries, and aligned to generate phylogenetic trees of each subfamily (see Methods and Suppl. Figure S1 and S2). The combined results of independent reconstructions for different ortholog groups and plant lineages are shown in **Figure 3.1**.

Our phylogenetic trees showed a clear subdivision of S1 bZIPs into two further groups of orthologs, each conserved between dicots and monocots as showed by consistent topologies between several independent lineage-specific reconstructions (**Figure 3.1** and Suppl. Figure S2). The first identified group included AtbZIP2, AtbZIP11, and AtbZIP44 from the model dicot arabidopsis (*Arabidopsis thaliana),* and OsbZIP76, OsbZIP77, and OsbZIP78 from the representative monocot rice (*Oryza sativa*), while the second group included AtbZIP1 and AtbZIP53, and OsbZIP84, OsbZIP85, OsbZIP86, and OsbZIP87 from the same two species, respectively. Orthology between dicot and monocot sequences from each group can be described as a "many-to-many" relationship, as further within-group duplications appear independent between the two plant lineages (**Figure 3.1**; see also **Figure 3.2**). The duplication which created these two groups of orthologs likely occurred early in angiosperm evolution, as we found two S1 bZIP genes in the amborella (*Amborella trichopoda*) genome. These sequences clustered consistently with either S1 bZIP ortholog group in most lineage-specific trees, although with low bootstrap support values (see Suppl. Figure S2).

The C bZIP subfamily is divided in two main groups of orthologs: one represented by arabidopsis AtbZIP9, and the other by AtbZIP10, AtbZIP25, and AtbZIP63. The corresponding sequences for each group in rice were OsbZIP19, OsbZIP20, and OsbZIP21, and OsbZIP18, OsbZIP22, and OsbZIP23, respectively (**Figure 3.1**; see also **Figure 3.2** and Suppl. Figure S1). The basal angiosperm amborella also harbors two C bZIP sequences, with strongly supported membership in either C ortholog group (**Figure 3.1** and Suppl. Figure S1). In more recent angiosperm lineages, the second group of orthologs might be further split into two separate clusters: one including AtbZIP63 and OsbZIP19, and the other collecting AtbZIP10 and AtbZIP25, and OsbZIP20 and OsbZIP21, as previously proposed (Côrrea et al. 2008). Importantly, our analyses allowed us to detect unexpected patterns of gene duplications and losses in both C and S1 bZIP phylogeny (**Figure 3.1** and Suppl. Figure S1 and S2). For

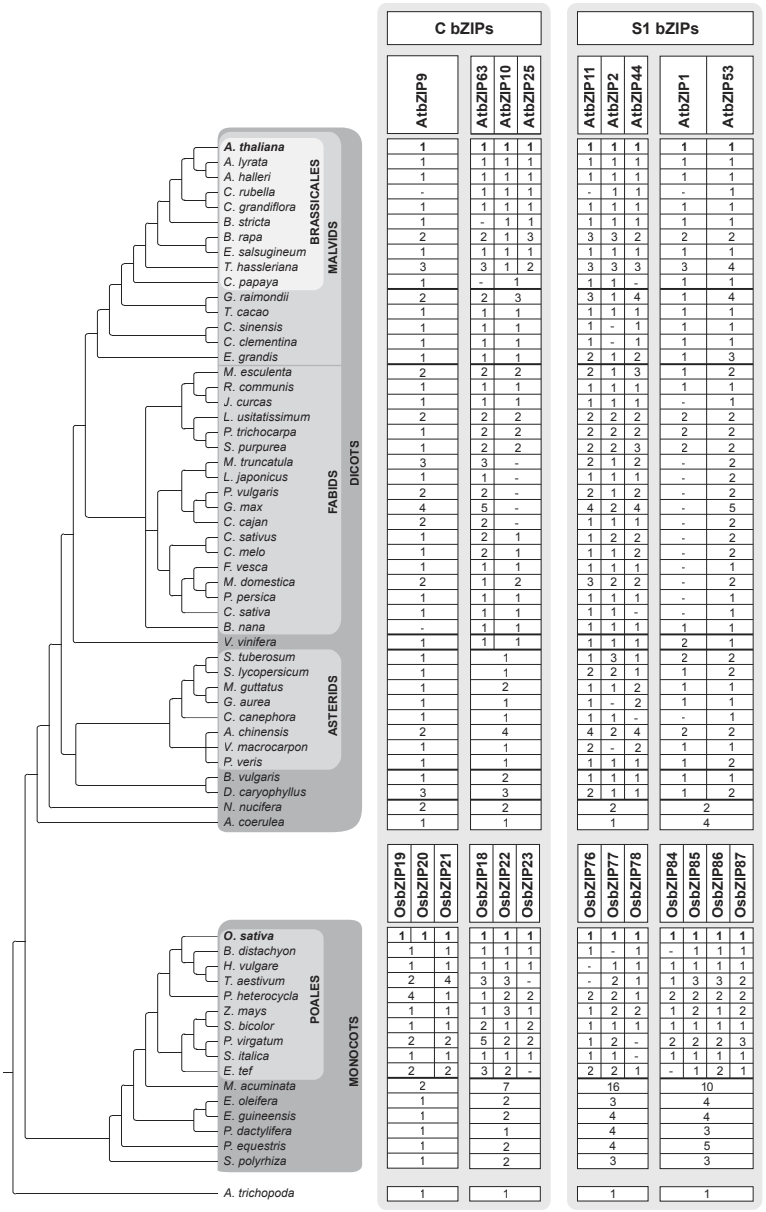


Figure 3.1. Conservation of C and S1 bZIP orthologs across angiosperm species. *The figure combines consistent topologies from independent lineage-specific Maximum Likelihood phylogenetic trees of angiosperm C and S1 bZIP subfamilies. For each species, values indicate the number of orthologs directly related to arabidopsis (A. thaliana) or rice (O. sativa) as reference bZIPs (in bold) for dicots or monocots, respectively. Notice that no direct correspondence exists between individual dicot and monocot orthologs (individual columns), but only between groups of orthologs (white blocks). Dashes indicate missing orthologs. See Suppl. Figure S1 and S2 for original C and S1 bZIP trees, respectively.*

instance, C subfamily AtbZIP10 and AtbZIP25 turned out to be the result of a gene duplication event specific to brassicales, constituting in fact a lineage-specific pair of paralogs (**Figure 3.1**). Other eudicots on the contrary seem to possess a depleted set of orthologs for these sequences; we observed this in the subgroup of fabales termed faboideae, which includes soybean (*Glycine max*), and in the early branching eudicots *Aquilegia coerulea, Vitis vinifera*, and asterids (**Figure 3.1** and Suppl. Figure S1). Absence in these latter species might be potentially explained by a poor assignment of their AtbZIP63 orthologs, which could be ancestral not only to AtbZIP63 but also to AtbZIP10 and AtbZIP25; however, our phylogenetic reconstruction points more convincingly to a secondary gene loss (see Suppl. Figure S1).

Among S1 bZIPs, we revealed the complete lack of AtbZIP1 orthologs in the eudicot groups fabales (soybean, *Cajanus cajan*, *Lotus japonicus*, *Medicago truncatula*, *Phaseolus vulgaris*), cucurbitales (*Cucumis melo* and *C. sativus*), and rosales (*Cannabis sativa*, *Fragaria vesca*, *Malus domestica*, *Prunus persica*), which include many species of agricultural importance. This is likely due to a secondary gene loss in the ancestor of these lineages, as AtbZIP1 orthologs are found in other fabids and in their sister lineage malvids. As a consequence, direct S1 orthologs comparisons with lineages branching before *V. vinifera*, such as solanales (*Solanum lycopersicum* and *S. tuberosum*) among asterids, should be considered carefully. A similar situation is present in monocots: while poales could be easily compared with rice reference sequences, species that diverged early on (e.g. *Musa acuminata*) show an independent pattern of duplications for each of the C and S1 ortholog subgroups (**Figure 3.1** and Suppl. Figure S1 and S2). The simple scheme presented here (**Figure 3.1**) condenses the information from the fully detailed phylogeny available in the Supplementary Material. Together, our results illustrate the important role of phylogenetics for reliable ortholog identification, especially in the case of gene families shaped by multiple duplication and loss events, such as S1 bZIPs.

C and S bZIPs originated as sister groups before the emergence of land plants

Similar to the recent phylogenetic history of C and S1 bZIPs in angiosperms, the early evolution of C and S groups was also compromised by species sampling (Côrrea et al. 2008). We included in our analyses a larger set of early branching species, encompassing the multicellular algae *Klebsormidium flaccidum* (charophyte), the bryophyte *Physcomitrella patens*, the spikemoss *Selaginella moellendorffii* (lycopodiophyte), and several pteridophyte and gymnosperm species (see Suppl. Table S1). For the identification of C and S sequences we adopted the same strategy used for angiosperms C and S1 bZIPs (see Methods), additionally including amborella, rice, and arabidopsis C and S bZIP sequences as flowering plants representatives in the phylogenetic reconstructions.

Our analyses resulted in the discovery of unambiguous C and S sequences from gymnosperms to charophytes (Figure 3.2). For the S subfamily in particular, our findings completely abolish the current view of these sequences as an angiosperm-specific innovation (Côrrea et al.

2008). Some of the S class orthologs we identified in early branching plant lineages were previously classified as a separate bZIP sequence subfamily (cI), closer to C rather than to S bZIP orthologs (Côrrea et al. 2008); hence our results invalid the definition of cI subfamily. S sequences from more ancestral species did not show any tendency to cluster specifically with S1 or any other S group of orthologs from seed plants (**Figure 3.2**), indicating that these groups emerged from spermatophyte-specific duplications. Importantly, the discovery of an S bZIP sequence in the charophyte *K. flaccidum* creates a tremendous gap between the first appearance of an S subfamily member and the later expansion of the group, indicating these transcription factors are likely endowed with more ancestral functions than previously thought. *K. flaccidum* was also the earliest diverging species to harbor a C bZIP sequence in our phylogenetic reconstruction, bringing back the origin of the subfamily from bryophytes (Côrrea et al. 2008) to charophytes, and therefore from after to before the land colonization event. Importantly, the discovery of both C and S earliest known members in a charophyte

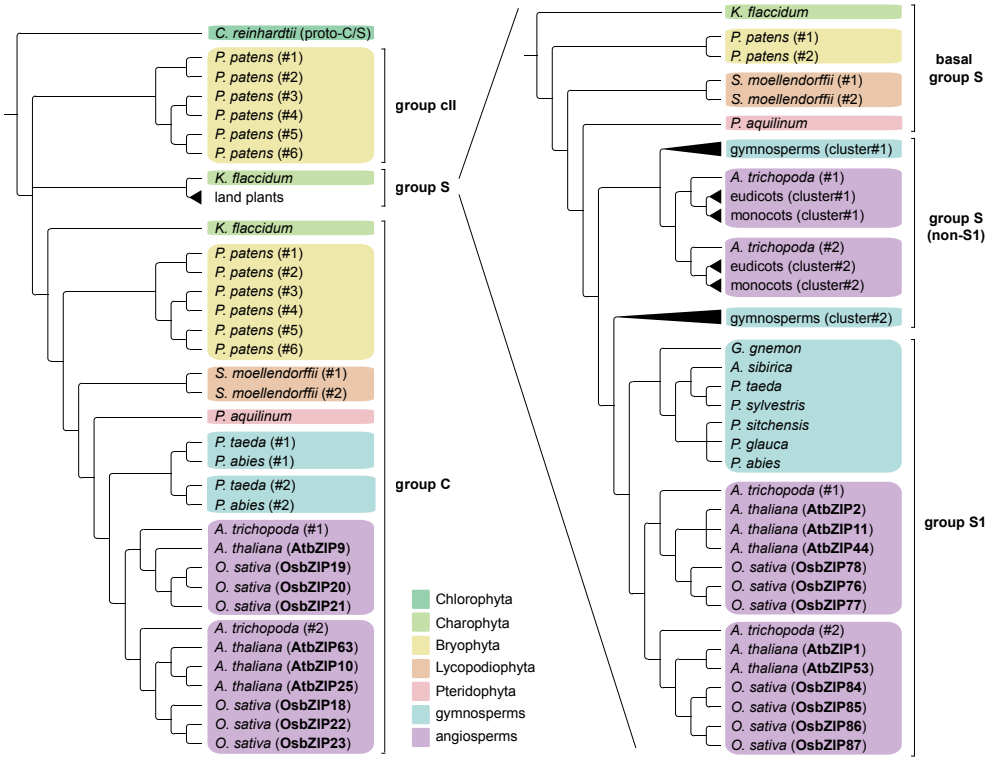


Figure 3.2. A phylogeny of C and S bZIP subfamilies in green plants. *Simplified phylogenetic trees of the C and S bZIP subfamilies showing the relationship between the two groups of orthologs across green plants. Known bZIP orthologs from arabidopsis and rice are shown in bold. Topology represents the consensus of independent phylogenetic reconstructions shown in Suppl. Figure S3.*

(**Figure 3.2**) is an unprecedented clue to the origin of these subfamilies by duplication from a shared ancestor, and suggests a possible role of both subfamilies in the later emergence of land plants.

In addition to C and S orthologs, we found a third group of sequences in bryophytes, corresponding to previously identified cII bZIPs (Côrrea et al. 2008). In our analyses, cII bZIPs were missing in charophytes and vascular plant species, confirming this group is bryophyte-specific (**Figure 3.2**). Finally, chlorophytes harbored a more ancestral type of sequences, termed "proto-C" (Côrrea et al. 2008), which we deemed appropriate to rename "proto-C/S" in order to reflect its parental relationship to both C and S subfamilies (**Figure 3.2** and Suppl. Figure S3). Surprisingly, among the S bZIPs identified in gymnosperms we observed a group of sequences clustering together with angiosperms S1 bZIPs, suggesting they might be S1 orthologs (**Figure 3.2** and Suppl. Figure S3). Previous phylogenetic analyses indicated S1 bZIP transcription factors as a recent innovation in plant evolution, likely restricted to angiosperms as the rest of the S subfamily (Côrrea et al. 2008). However, our discovery of putative S1 orthologs in both gymnosperms and angiosperms indicates the possible origin of S1 bZIPs in the ancestor of spermatophytes. We also found a putative S1 ortholog from *Pteridium aquilinum*, but the interpretation is less reliable as we did not observe members in other fern species. Rather than an angiosperm-specific innovation (Côrrea et al. 2008), S1 bZIPs might therefore be regarded as a common toolkit of seed plants, which potentially contributed to the radiation of the entire lineage.

In vivo reporter gene assays confirm 5'ORF-mediated SIRT in gymnosperms S1 bZIP orthologs

Our discovery of putative S1 bZIP members in gymnosperms prompted us to assess the conservation of SIRT-mediating 5'uORFs in these species. Given their ancestral relationship to the entire group of angiosperm S bZIPs, we also hypothesized that the S sequences from early branching species might harbor 5'uORFs with similar properties. Therefore we extracted the upstream region of each identified S (including putative gymnosperm S1) sequence from our genomic database, whenever available, and searched for the presence of 5'uORFs similar to those observed in arabidopsis and rice S1 bZIP 5'UTRs (see Methods). Our results showed matching 5'uORFs in candidate S1 bZIPs from most gymnosperms (**Figure 3.3 A**); however, no hit was found in other S orthologs from these species, nor in S subfamily representatives from more early branching plants. Thus, this feature was not inherited by S1 bZIPs from more ancestral S sequences. The newly discovered gymnosperm 5'uORFs shared with angiosperms only two of the four residues thought to be essential for S1 bZIP regulation (Rahmani et al. 2009), i.e. Leu-35 and Tyr-39 from arabidopsis bZIP11, while Ser-29 and Ser-31 were not conserved (**Figure 3.3 A**). The termination codon position, another invariable feature of angiosperm S1 5'ORFs (Rahmani et al. 2009), was also different in gymnosperm species (**Figure 3.3 A**). Still, the extent of sequence conservation within gymnosperms themselves appears striking, suggesting functionality (**Figure 3.3 A**).

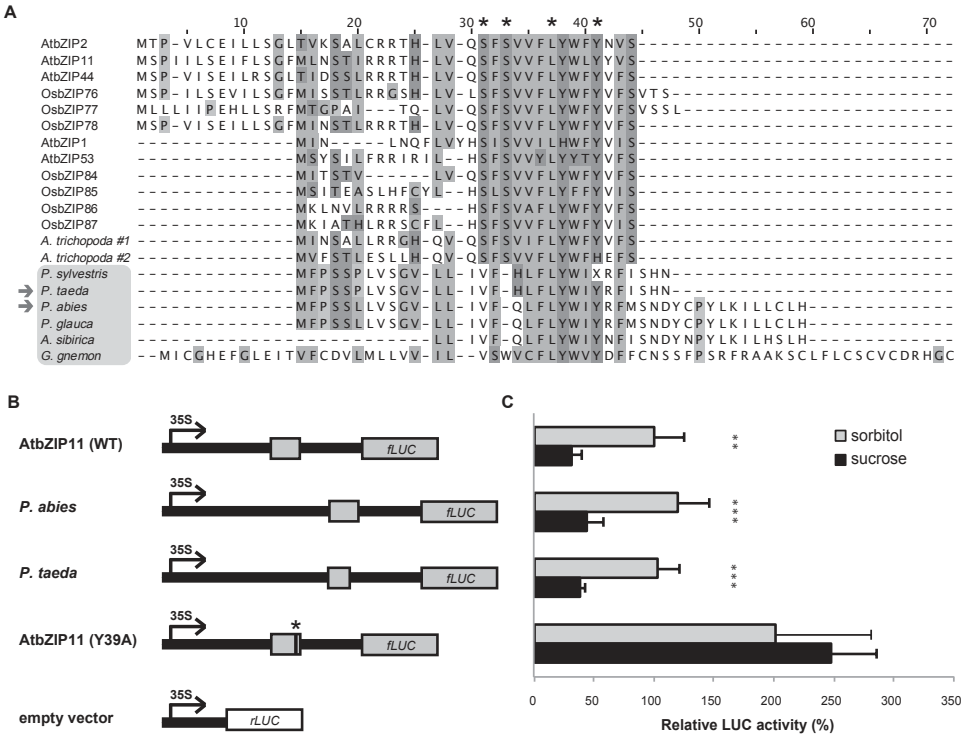


Figure 3.3. Sequence alignment of 5'ORFs and experimental confirmation of SIRT from gymnosperms putative S1 bZIPs. (A) *Alignment of 5'uORFs from angiosperms and gymnosperms S1 bZIPs showing conservation across each lineage; gymnosperm sequences are enclosed in a grey box. Residues known to be necessary for SIRT in arabidopsis (Rahmani et al. 2009) are marked with a black star. Gymnosperms sequences tested for SIRT are indicated with an arrow.* ***(B)*** *Schematic representation of the constructs used to test SIRT in the transient LUC expression assay. Rectangles represent 5'ORFs shown in panel A, with proportions reflecting length and distance from the main ORF. A star and a line in the mutant AtbZIP11 (Y39A) indicate the substitution point (negative control). An empty 35S:rLUC construct was used to normalize LUC activity data.* ***(C)*** *Results of relative LUC activity assays. Normalized LUC activity is presented relative to arabidopsis WT construct results in sorbitol. Values represent the average of at least four biological replicates. Error bars indicate SD from the mean. Stars indicate the significance of a two-tailed distribution t-test with unequal variance (** p < 0.01, *** p < 0.005). For the original results see Suppl. Table S2.*

To test the ability of gymnosperm S1 5'uORFs to mediate SIRT, we proceeded with specific experimental assays. For the test we selected two representative sequences, from *Picea abies* and *Pinus taeda*, based on their quality and completeness (see sequences in **Figure 3.3 A**). Wild type (WT) and SIRT loss of function mutant (Y39A) AtbZIP11 5'uORFs were additionally included as positive and negative controls, respectively. The effect of each sequence on

translational regulation was then tested in a transient luciferase (LUC) expression assay in the presence of either sucrose or sorbitol, the latter serving as an osmotic control (see Methods).

Our results showed that both *P. abies* and *P. taeda* S1 5'uORF sequences could efficiently mediate the translational repression of the LUC reporter gene in the presence of sucrose (**Figure 33. B,C** and Suppl. Table S2). The magnitude of the effect was similar to that observed for WT AtbZIP11, i.e. more than 2-fold decrease in expression at the given experimental conditions, with little variation between replicates (**Figure 3.3 C** and Suppl. Table S2). Thus, we confirmed that the newly found gymnosperm S1 5'uORFs are capable of mediating SIRT. The conservation of SIRT in gymnosperm S1 sequences might reflect their involvement in the metabolic adaptation to low energy conditions (i.e. sucrose depletion), as observed for angiosperm orthologs, and the need for downscaling their activity when energy levels are restored. Importantly, this is the first time SIRT is experimentally reported for not non-dicot sequences, extending the relevance of this regulatory mechanism to distant plant lineages. In particular, our discovery indicates that SIRT-regulated S1 orthologs were present in the common ancestor of spermatophytes, and that monocot 5'uORFs found through previous similarity searches (Rahmani et al. 2009; von Arnim et al. 2014) are likely to be functional. Moreover, we suggest that an in-depth comparison of the newly identified gymnosperm sequences with known angiosperms S1 5'uORFs might help clarifying the mechanistic details of SIRT. For instance, it is unclear whether 5'uORF-encoded SC peptides sense sugar molecules directly, or through association to a more complex regulatory machinery. Notably, our assays of gymnosperm sequences took place in arabidopsis cells (see Methods), exploiting the molecular machinery available in this species, and successfully reproduced SIRT in spite of significant differences with angiosperm 5'uORF-encoded peptides, i.e. in the conservation of two specific amino acid residues and the C-terminal position. Our results might therefore indicate that the SIRT mechanism relies on structural conformation rather than on the recognition of specific sequence motifs.

DISCUSSION

In this work, we presented an integrated phylogenetic reconstruction of the C/S1 bZIP transcription factor network members across plant evolution. In a previous publication (Côrrea et al. 2008), which at the time represented the most comprehensive study of plant bZIPs available, the phylogenetic details of the S subfamily were particularly elusive, limiting the investigation of this regulatory system. Here we finally confirmed the presence of S1 orthologs in both eudicots and monocots, providing phylogenetic evidence for previous observations based on simple sequence similarity searches (Rahmani et al. 2009). Moreover, we were able to identify SIRT-mediating S1 orthologs in gymnosperms, showing the conservation of this group in all seed plants. This is notably the first time a SIRT assay is performed for non-eudicot sequences, and our results might therefore provide clues

for further research on the mechanism by which S1 5'ORFs regulate translation in these sequences.

We also showed that, while S1 orthologs likely appeared in the common ancestor of spermatophytes, the S subfamily as a whole is even older, dating back to charophytes. This finding completely abolishes the previous notion of S bZIPs as an angiosperm-specific group, and by making possible the study of ancient family members, could facilitate our understanding of modern S ortholog functions, possibly inherited and exploited through subfunctionalization.

A model for the early evolution of the C/S1 bZIP transcription factor network

Due to the long coexistence of C and S bZIPs before the appearance of S1 orthologs (**Figure 3.4 A**), we propose that S1 bZIPs may have inherited their specific heterodimerization preferences from ancestral S members; this seems more likely than S1 bZIPs abruptly acquiring dimerization capabilities with C bZIP partners, and vice versa. Ancestral C/S bZIP interactions in turn might have been retained from a homodimerizing proto-C/S bZIP ancestor, after the duplication event at the origin of *K. flaccidum* C and S sequences. We combined such observations into a consistent scenario for the emergence of C/S1 bZIP interactions, which we propose here (**Figure 3.4 B**). More importantly, our discovery of the first C subfamily representative also in charophytes uncovered the possible origin of C and S bZIPs as sister groups, duplicated from a common ancestral sequence. This hypothesis is again a novel insight, as the S bZIPs are currently proposed to have evolved from duplications within the C subfamily (Côrrea et al. 2008).

To recapitulate, the steps leading to the C/S1 bZIP dimerization network emergence according to our model are proposed as follows: originally, an ancestral algal proto-C/S bZIP existed with homodimerizing properties. A later duplication of this sequence led to the generation of two paralogs, i.e. the ancestral C and S bZIP sequences, still able to interact. Such heterodimerization capability was maintained in the course of evolution, up to the S subfamily duplications in the common ancestor of spermatophytes, which created the conditions for the subfunctionalization of different S ortholog groups. Among them, S1 bZIPs specialized as C dimerization partners, while other S orthologs lost the heterodimerization capability. The latter step is the most hypothetical in our scenario, as the specificity of C/S1 interactions has been documented only in the model plant *A. thaliana* (Ehlert et al. 2006). While it is likely that other eudicots present the same dimerization specificity, more distant species, such as monocots or gymnosperms, might still allow promiscuous C/S bZIP dimers. Independently from the exact steps correctly describing the C/S1 bZIP dimerization network emergence, comparing its timing with major events in plant evolution (e.g. speciation and colonization of new environments) would help clarifying its original role. For instance, the presence of C and S sequences in charophytes might have provided ancestral plants with an additional toolset for land colonization; later on, specialized C/S1 bZIP dimers

in spermatophytes could have contributed to the complex adaptive features of both angiosperms and gymnosperms. While we showed the presence of both C and S bZIP sequences in basal species, the role of putative early C/S bZIP dimers remains a hypothesis.

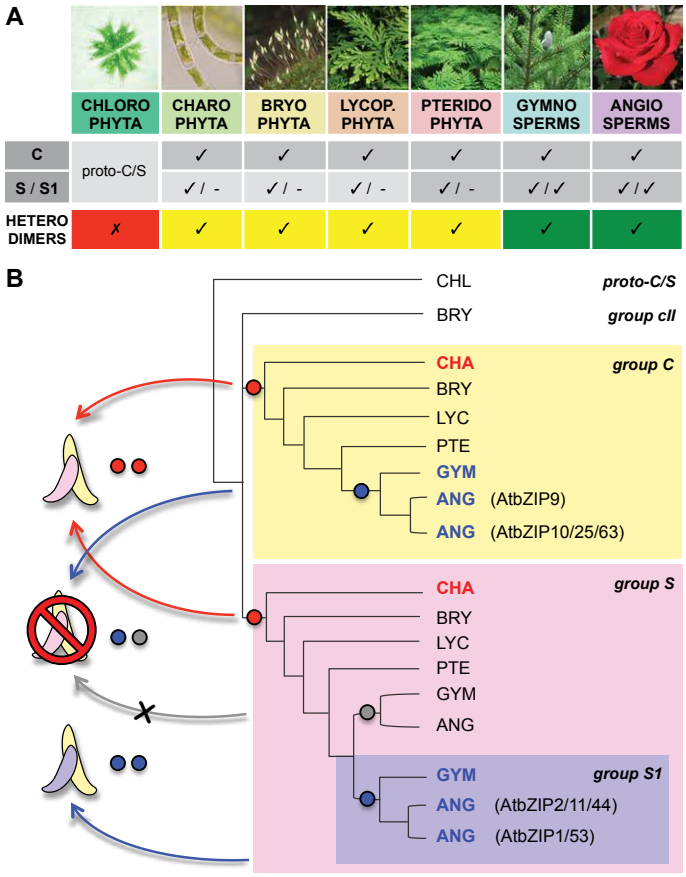


Figure 3.4. Presence pattern of C and S bZIP orthologs in different plant lineages and a model for the evolution of the C/S1 dimerization network. (A) *Schematic summary of C and S or S1 bZIP orthologs coexistence in different plant lineages according to our results. Ticks and dashes indicate presence and absence, respectively. Based on the presence pattern, the putative formation of C/S or C/S1 heterodimers can be postulated (lowest row, yellow and green background respectively). Plant pictures from Wikimedia Commons (https://commons.wikimedia.org).* ***(B)*** *The gene tree summarizes the findings presented in this paper, and illustrates a possible model for the emergence of the C/S1 bZIP dimerization network as described in the text. Abbreviations refer to the plant lineage names used in panel A. Dot pairs and dimer cartoons provide information on approximate time of appearance and ortholog group of the bZIP sequences involved in hypothetical interactions, respectively.*

A reference for comparative studies on C and S1 bZIPs in non-model species

Our analyses clarified the details of C and S1 orthologs conservation in angiosperms, revealing lineage-specific duplications and gene losses in both subfamilies. Knowledge of such gain and loss patterns is necessary for the accurate transfer of functional information between model and crop species, and our results provide a more reliable classification framework than individual genome-specific bZIP catalogs published in recent years. We believe that these findings will facilitate the study of the C/S1 bZIP transcription factor network in non-model species, suggesting new directions for experimental research on SIRT and heterodimers formation, and possibly leading to useful agricultural applications.

METHODS

Sequences collection

Genome, transcriptome, and annotated protein data for several green plant species were collected from Phytozome v10 (Goodstein et al. 2012) and other online public repositories; included species, resources, and abbreviations are listed in Suppl. Table S1.

Identification of C and S1 bZIP orthologs in angiosperms

Candidate angiosperm C and S1 bZIP transcription factors were retrieved from our protein or DNA sequence database using BLASTP v2.25+ (Camacho et al. 2009) and HMMER v3.0 (Finn et al. 2011) searches, or TBLASTN v2.25+ searches, respectively, with default settings. We used as queries annotated bZIP sequences (main ORF peptide sequence) from the arabidopsis TAIR10 release (Lamesch et al. 2010) as dicot representative, from the rice TIGR release 7.0 (Ouyang et al. 2007) as monocot representative, and from the basal angiosperm amborella (Chamala et al. 2013); notice that rice and amborella sequences were used only in later search iterations, after phylogenetic assessment of their identity as C or S1 orthologs, and re-annotation of the peptide sequence for one of the amborella C bZIP sequences with the web version of AUGUSTUS (http://bioinf.uni-greifswald.de/augustus) (Stanke and Morgenstern 2005). The previously annotated S subfamily OsbZIP79 (Côrrea et al. 2008) was excluded from the queries for two reasons: no match in our version of the rice genome, and divergent sequence features pointing to a pseudogene. Genomic hits were extended upstream and downstream by 1000 nucleotides and translated to amino acid sequence with the web version of AUGUSTUS (http://bioinf.uni-greifswald.de/augustus) (Stanke and Morgenstern 2005); when this failed to produce satisfactory *de novo* predictions, guided protein predictions based on reference C and S1 bZIPs were generated using Exonerate v2.2.0 (www.ebi.ac.uk/~guy/exonerate). The same tools were used to correct original protein annotations which appeared incomplete or mispredicted, after extracting their corresponding genomic region. Reverse BLASTP searches of the translated hits versus arabidopsis and rice proteomes (including labeled bZIP sequences) were used to remove obvious false positive hits, i.e. those without C or S bZIP sequences in the top 5 reverse matches. Given the high sequence similarity between S bZIP groups, we chose to include hits with best reverse matches to any S subfamily member, and not just to S1 bZIPs, to prevent the possible exclusion of relevant sequences. Reference arabidopsis and rice C

and S bZIP sequences are shown in Suppl. Data S1.

Identification of C and S bZIP orthologs in early branching species

The identification of C and S bZIP orthologs in chlorophytes, charophytes, bryophytes, lycopodiophytes, pterydophytes, and gymnosperms was performed as described above for angiosperms C and S1 bZIPs. Confirmed C and S orthologs from each species were iteratively used as queries to identify more distant sequences, potentially missed during the initial search with angiosperm queries.

Phylogenetic analyses

C and S hits were aligned to the entire set of annotated arabidopsis and rice orthologs from the 13 known bZIP subfamilies, and phylogenetic reconstructions were performed to assess their identity as C or S subfamily orthologs. Confirmed candidates and reference arabidopsis and rice C or S bZIP were therefore re-aligned without members from other bZIP subfamilies to generate final high-resolution phylogenetic trees (not shown). For angiosperms S1 bZIP hits, a further phylogenetic reconstruction against reference S sequences from arabidopsis and rice was performed to better distinguish actual S1 candidates from other S hits (not shown). In addition to general C and S1 bZIP trees including all angiosperms species, lineage-specific alignments were generated independently to achieve a higher resolution of both C and S1 bZIP orthologs in different groups of flowering plants (Suppl. Figure S1 and S2). The trees obtained from these phylogenetic reconstructions were compared to the general C and S1 bZIP subfamilies trees, providing additional evidence for nodes with low bootstrap support values based on independent consistent topologies. C and S1 angiosperm sequences used in the building of phylogenetic trees are shown in Suppl. Data S1. For early branching species (gymnosperms and earlier), alignment were initially generated using only annotated protein sequences, and new translated genome hits were aligned to these in a later step (both version are shown in Suppl. Figure S3); this strategy allowed us to build more reliable trees than by directly aligning a large number of newly predicted C and S bZIPs from genome, which would be strongly affected by an over-representation of gymnosperm sequences. Because of low bootstrap support values, multiple independent tree topologies for C, S, and C+S orthologs were again compared to infer consensus trees (Suppl. Figure S3). C and S sequences from early branching species used in the building of phylogenetic trees are shown in Suppl. Data S1. Sequence alignments were performed using MAFFT v7.040, einsi algorithm (Katoh and Standley 2013), and manually trimmed in Jalview v2.8.2 (Waterhouse et al. 2009). Maximum likelihood phylogenetic reconstructions were performed with RAxML v7.2.8 (Stamatakis et al. 2005) using 1000 bootstrap replicates, after selecting an appropriate amino acid substitution model with ProtTest v3.2 (Darriba et al. 2011). The JTT+I+G model was used for all the trees shown in the Supplemental Information. Tree graphics was generated in iTOL v2.1 (Letunic and Bork 2011) and TreeGraph v2.4.0-456 beta (Stöver and Müller 2010).

Identification of 5'ORFs in S1 bZIP candidates

For each identified gymnosperms and amborella candidate S1 bZIP ortholog, the corresponding genomic region, including 1500 upstream nucleotides, was scanned with Exonerate v2.2.0 (www.ebi.ac.uk/~guy/exonerate) for the presence of 5'ORFs similar to reference S1 sequences from arabidopsis and rice. The S1 bZIP ortholog from *Picea sitchensis* had to be excluded from the analysis, as the

required upstream sequence was missing. Confirmed S1 5'ORFs were used as queries in a second search round to obtain additional hits, which allowed the identification of the truncated 5'ORF sequence from *Pinus abies.* Other gymnosperm S bZIP hits and more basal sequences from charophytes, bryophytes, lycopodiophytes, and pterydophytes were also analyzed to assess the presence of S1-like 5'ORFs. False positives were removed through manual inspection of sequence alignments, which were performed using MAFFT v7.040, linsi algorithm (Katoh and Standley 2013). Identified 5'ORF sequences from gymnosperms are shown in Suppl. Data S1.

Plant material and growth conditions

Plant material was obtained from arabidopsis Columbia-0 (Col-0) ecotype. Seeds were stratified in the dark at 4°C for 2 days on soil, after which adult plants were grown at 22°C under a 16h light/8h dark regime (100 µmol m^{-2} s^{1}). Leaves of 4-week old plants were used in transient expression experiments.

Construction of *P. abies* and *P. taeda* 5'-leader vectors

GeneArt® gene synthesis service (Life Technologies, Carlsbad, CA, USA) was used to custom synthesize the 5'-leader sequences Arabidopsis S1 bZIP homologous genes from gymnosperm species *Picea abies* and *Pinus taeda*. Synthesized 5'-leader sequences contained the full 5'-leader, starting 500 nucleotides upstream of the arabidopsis *bZIP11* 5'ORF sequence. Gateway® cloning sites attL1 and attL2 flanked the sequences, and the resulting constructs were cloned in the pMK-RQ vector backbone. Construct sequences are shown in the Suppl. Methods. Sequences were cloned into the pUC19 based p35S-ccdB-fLUC destination vector (Rahmani et al. 2009), using Gateway® LR cloning according to manufacturer instructions (Invitrogen, ThermoFisher Scientific, Waltham, USA) to create transient *LUC* expression vectors.

Transient transformation of arabidopsis material

Transient transformation of arabidopsis seedlings was performed as previously described (Rahmani et al. 2009), with minimal adjustments. DNA coating of gold particles was performed accordingly, using 1.2 mg of fLUC vector and 0.4 mg of rLUC normalizing vector per transient expression experiment. Plant material was transformed using the Biolistic particle delivery system, model PDS-1000 He (Bio-Rad, Hercules, CA, USA). Leaves from 4 weeks old plants were transformed using 900psi rupture discs. Two leaves were simultaneously transformed, after which one was incubated in 10mL liquid one-half strength MS-medium supplemented with 6% sorbitol, and the other in medium containing 6% sucrose. Incubations were performed in 100mL flasks, which were placed on a rotary shaker (50rpm) in constant light for 24 hours. Material was harvested, washed with demi-water and frozen in liquid nitrogen. Samples were stored at -80°C.

LUC activity assays

Protein extracts from transgenic leaf material expressing *LUC* were made from approximately 25 mg ground tissue using 100µl of Cell Culture Lysis (CCL) reagent (Promega, Madison, WI, USA, #E1531). Plant powder was incubated in extraction buffer for 10 minutes at room temperature, followed by 5 minutes of centrifugation (16,000 xg). Twenty microliters of supernatant was transferred to a white 96 well luminometer plate (Promega, #Z3291). LUC activity was measured with a Glomax 96 microplate

luminometer (Promega), using the "LUC assay system with injector" protocol of the Glomax software. Relative LUC-levels of transiently transformed plant material was determined by the ratio of fLUC to rLUC activity, as previously described (Rahmani et al. 2009). 100 µL of the substrates supplied in the Dual Luciferase assay kit (#E1960, Promega), was applied to measure fLUC and rLUC activity. LUC activity was assayed with a 10 second integration time and a 2 second delay between injection and measurement.

ACKNOWLEDGEMENTS

This work was supported by the MERIT Marie Curie ITN, Grant Agreement Number 264474. The authors would like to thank Dr. Lidija Berke, Jolien van Hooff, Eelco Tromer, and Dr. Bas Dutilh for their constructive comments on the manuscript organization and readability.

AUTHOR CONTRIBUTIONS

A.P. wrote the paper and prepared the paper figures. B.S. edited the paper. J.L. and J.H. designed the SIRT assays. J.L. performed the SIRT assays.

SUPPLEMENTARY MATERIAL

All Supplementary materials for this chapter are made available online at the following link: http://bioinformatics.bio.uu.nl/alessia/THESIS.html

Supplementary Methods. Gymnosperms constructs used for the SIRT experiments.
Supplementary Figure S1. Phylogenetic trees of angiosperm C bZIPs.
Supplementary Figure S2. Phylogenetic trees of angiosperm S1 bZIPs.
Supplementary Figure S3. Phylogenetic trees of proto-C/S, cII, C, and S bZIPs from green plants.
Supplementary Table S1. Species used in the phylogenetic analyses and abbreviations.
Supplementary Table S2. Results of SIRT reporter gene experiments.
Supplementary Dataset S1. Sequences used in the analyses (fasta format), including reference arabidopsis and rice C and S sequences, identified C and S1 angiosperm sequences, identified C, S, proto-C/S and cII green plant sequences, and identified S1 5'ORF sequences from gymnosperms.

4

Multiple sequence signals affect translational efficiency during sucrose treatment and seed germination in *A. thaliana*

adapted from:

Increased sucrose levels mediate selective mRNA translation in Arabidopsis
Magdalena Gamm[1], **Alessia Peviani**[2], Anne Honsel[3], Berend Snel[2], Sjef Smeekens[1], and Johannes Hanson[1,3]

BMC Plant Biology 14, 306 (2014)

and:

Extensive translational regulation during seed germination revealed by translational profiling
Bing Bai[1,4], **Alessia Peviani**[2], Sjors van der Horst[2], Magdalena Gamm[1], Berend Snel[2], Johannes Hanson[1,3], and Leónie Bentsink[4]

(submitted)

[1]Molecular Plant Physiology, Institute of Environmental Biology, Utrecht University, Utrecht, The Netherlands
[2]Theoretical Biology and Bioinformatics, Utrecht University, Utrecht, The Netherlands
[3]Umeå Plant Science Centre, Department of Plant Physiology, Umeå University, Umeå, Sweden
[4]Wageningen Seed Laboratory, Laboratory of Plant Physiology, Wageningen University, Wageningen, The Netherlands

ABSTRACT

Translation is an extremely energy-consuming process in the cell. While differential translational regulation has been shown to occur in many conditions, the molecular mechanisms and sequence signals by which this is achieved are only beginning to be understood. In this study we performed a bioinformatic investigation of putative mRNA sequence features responsible for translational regulation during sucrose treatment and seed germination, in contrast to the majority of the literature focused on stress conditions. We performed both classic analyses such as transcript length, GC content, and codon bias, and novel tests for motif and structural elements, for which we additionally determined the localization along the transcript.

Our results show that general sequence properties, particularly transcript length and GC content, correlate with translational efficiency in some conditions. However, these can be overwritten by more specific signals, such as sequence motifs and mRNA structure. Although not covering the entire set of translationally regulated mRNAs, motifs appear as driving features for translational control in virtue of their specificity and localization. The distribution of mRNA secondary structure also shows significant differences between translationally regulated and non-regulated transcripts, indicating an important role of mRNA folding. Overall our results indicate that differential translational regulation is likely to arise from the interplay between multiple sequence signals. mRNA structure in particular might have a central role, either as direct target of regulation, or by controlling the accessibility of sequence motifs by RNA-binding proteins. High-resolution in vivo mRNA structural data are needed to bring further advances in the field.

INTRODUCTION

Plant growth and development depend on energy-consuming processes, which, given the variability of environmental conditions, must be tightly coordinated to avoid the premature consumption of energy supplies. One of the most demanding cellular processes in this regard is protein synthesis, and particularly mRNA translation: energy has to be provided for the synthesis of amino acids and tRNAs, the formation of peptide bonds, and the biogenesis of the translational machinery (Pal et al. 2013). However, compared to the regulation of transcription and mRNA stability, the translational control of mRNA molecules remains a much more elusive process.

The translational efficiency of a mRNA species can be generally measured in terms of its loading on multiple ribosomes (polysomes), which in turn relies mostly on translation initiation rates (Kawaguchi and Bailey-Serres 2002; Esposito et al. 2010); the impact of elongation and termination rates is more limited (Ma and Blenis 2009; Jackson et al. 2010). Several studies addressed the adaptation of translational efficiency to environmental conditions limiting energy availability, such as mild dehydration (Kawaguchi et al. 2003; Kawaguchi et al. 2004; Kawaguchi and Bailey-Serres 2005), hypoxia (Branco-Price et al. 2005; Branco-Price et al. 2008; Mustroph et al. 2009; Juntawong et al. 2014), salt and high

temperature (Matsuura et al. 2010; Ueda et al. 2012; Yanguez et al. 2013), phosphate and iron deficiency (Wang et al. 2013), and sucrose starvation (Nicolaï et al. 2006). Stress appears to affect translation with the decrease of both polysomal occupancy (fraction of a mRNA species associated to polysomes) and ribosome density (number of ribosomes per mRNA). Interestingly, a substantial fraction of mRNAs escapes this general stress-mediated reduction of translation and maintains or increases polysomal occupancy, ensuring the production of proteins necessary for stress adaptation (Bailey-Serres et al. 2009; Roy and von Arnim 2013). The translation of other mRNAs is strongly repressed under stress conditions, while their steady-state transcript levels do not change (Juntawong et al. 2014). Therefore it appears that in specific conditions a relevant subpopulation of mRNA species is subjected to positive or negative translational regulation, independent of transcript abundance. This indicates that specific regulatory mechanisms must be operating directly at the translational level to influence protein production.

Recently, some translationally regulated mRNAs were found to associate with the RNA-binding protein (RBP) UBP1 in *Arabidopsis thaliana* (Sorenson and Bailey-Serres 2014). UBP1 has been shown to participate in the reversible sequestration of specific mRNAs in cytoplasmic granules during hypoxia, thus providing a mechanism for condition-dependent translational regulation (Sorenson and Bailey-Serres 2014). Other RBPs have been found to be involved in various steps of post-transcriptional mRNA processing (Lorkovic 2009; Gerstberger et al. 2014), however many members still lack a complete functional characterization, making them potential candidates as regulators in differential translational control. The *A. thaliana* genome contains more than 200 RBPs, most of which are not found in other eukaryotes, and might therefore be involved in plant-specific responses (Lorkovic 2009).

Another fascinating hypothesis is that the translational selection of specific mRNA subsets might be operated by ribosomes themselves. More specifically, it has been proposed that the ribosome composition could change in response to different environmental cues, allowing for condition-dependent translational regulation (Byrne 2009; Hummel et al. 2012; Xue and Barna 2012). In *A. thaliana*, ribosomal proteins (RPs) are encoded by more than 240 genes, divided into 81 families comprising between 2 and 7 paralogs (Barakat et al. 2001; Carroll et al. 2008; Hummel et al. 2015). Generally, a single protein paralog from each family is present in the functional ribosome, allowing for a number of possible RP combinations; this variability is at the base of the concept of ribosomal heterogeneity (McIntosh and Bonham-Smith 2005; Xue and Barna 2012). Previous research showed that only specific RP paralogs are expressed in certain tissues or conditions (Williams and Sussex 1995; Weijers et al. 2001; Hulm 2005; Degenhardt and Bonham-Smith 2008; Mustroph et al. 2009; Wang et al. 2013), and that mutants lacking specific paralogs display severe and distinguishable phenotypes (Degenhardt and Bonham-Smith 2008; Schippers 2010). Post-translational modifications might also contribute to ribosomal heterogeneity, as recently shown for a number of plant RPs and translation initiation factors in response to photosynthetic activity (Turkina et al. 2011; Boex-Fontvieille et al. 2013). Overall, a large number of regulatory

proteins and pathways might be responsible for the observed effects, and research only started to identify the molecular players involved in the process.

Independently from the mechanism involved, it is reasonable to assume that the translational selection of specific mRNAs operates on signals present within the transcripts themselves. Early studies have explored the relationship between general sequence properties and the basal translational efficiency of mRNAs, uncovering for instance an inverse correlation between sequence length and translation rates (Valleriani et al. 2011); the same was observed for guanine and cytosine (GC) content (Qu et al. 2011), typically associated with more structured mRNA regions, which are less accessible by the translational machinery. More detailed studies on secondary mRNA structure have indeed shown that structured and unstructured regions could favor or disfavor mRNA accessibility to the ribosome, respectively, affecting basal translational efficiency (Li et al. 2012; Ding et al. 2014). Codon usage is also a widely explored factor in translational efficiency. Due to the redundancy of the genetic code, most amino acids are encoded by several synonymous codons, some of which are translated more efficiently than others due to the different abundance of their corresponding tRNAs (Wright et al. 2004). It has been proposed that plants might express tRNA species at different levels in different tissues, influencing the basal translation rates of mRNAs with specific codon preferences (Camiolo et al. 2012). The described sequence features, while useful to predict the basal translational efficiency of an mRNA molecule, are however unlikely to explain alone the differential translational regulation observed in specific stress and growth conditions; the hypotheses of differential regulation operated by RBPs and RPs, in particular, would imply the presence of more specific signals within regulated transcripts, to allow for specific protein-RNA interactions.

More recent studies on plant translational regulation have started to include analyses of mRNA motifs and structured elements as putative sources of specific signals for differential translational control (Kawaguchi and Bailey-Serres 2005; Jiao and Meyerowitz 2010; Liu et al. 2012; Liu, Wu, et al. 2013; Matsuura et al. 2013; Basbouss-Serhal et al. 2015). Researchers identified several sequence motifs in translationally regulated transcript, although not necessarily strongly enriched (Liu et al. 2012; Liu, Wu, et al. 2013; Matsuura et al. 2013; Basbouss-Serhal et al. 2015). The identification of structured sequence elements remains more challenging, mainly because of the sensitivity of mRNA folding on specific cellular conditions. While a changing conformation could provide a convenient mechanism for condition-dependent regulation, it makes difficult to investigate mRNA structure without experimental data to constrain *in silico* predictions. Overall, the features driving the differential translation of specific mRNAs are still marginally characterized and mostly interpreted in isolation, while regulation could be driven by multiple signals at once.

Another limit to the understanding of translational control comes from the strong research focus on energy-limiting stress conditions, while translational information about normal physiological processes has only recently started to emerge (Jiao and Meyerowitz 2010; Liu

et al. 2012; Ribeiro et al. 2012; Liu, Wu, et al. 2013; Galland et al. 2014; Layat et al. 2014; Lin et al. 2014; Rajasundaram et al. 2014; Basbouss-Serhal et al. 2015; Yamasaki et al. 2015). Previous studies showed that sucrose concentration correlates with general translational activity (Pal et al. 2013) and affects RPs composition (Hummel et al. 2012), suggesting a direct function for sucrose in translational control and, possibly, mRNA selection. Light was also shown to induce translational activity in etiolated seedlings (Liu et al. 2012; Liu, Wu, et al. 2013), during the diurnal cycle (Piques et al. 2009), or after dark treatment (Juntawong and Bailey-Serres 2012). However, the specific roles of light and sugars in translational control have not been investigated so far. Seed germination has also begun to be explored from a translational perspective only very recently (Galland et al. 2014; Layat et al. 2014; Basbouss-Serhal et al. 2015). The question remains of which regulatory mechanisms are responsible for differential translational regulation during physiological and stress conditions, and whether they are specific or present some degree of overlap between conditions.

Here we present an in depth study on the role of signals affecting translational control, focusing in particular on mRNA secondary structure and sequence motifs. Ribosome fractionation can be used to extract polysome-bound mRNAs, which represent actively translated transcripts. Microarray hybridization of this fraction and total mRNA can then be performed to derive polysomal occupancy for individual mRNA species, and monitor its changes between samples and time points (Mustroph et al. 2009; Jiao and Meyerowitz 2010; Liu et al. 2012; Liu, Wu, et al. 2013; Juntawong et al. 2014; Layat et al. 2014; Lin et al. 2014; Basbouss-Serhal et al. 2015). We performed our bioinformatic analyses on two similarly generated datasets, both focusing on translational regulation in non-stress conditions. The first explores and uncouples the effect of light and sugar in *A. thaliana* seedlings; the second investigates the translational changes taking place during seed germination.

RESULTS

Quantification of translationally regulated genes during sucrose treatment and seed germination

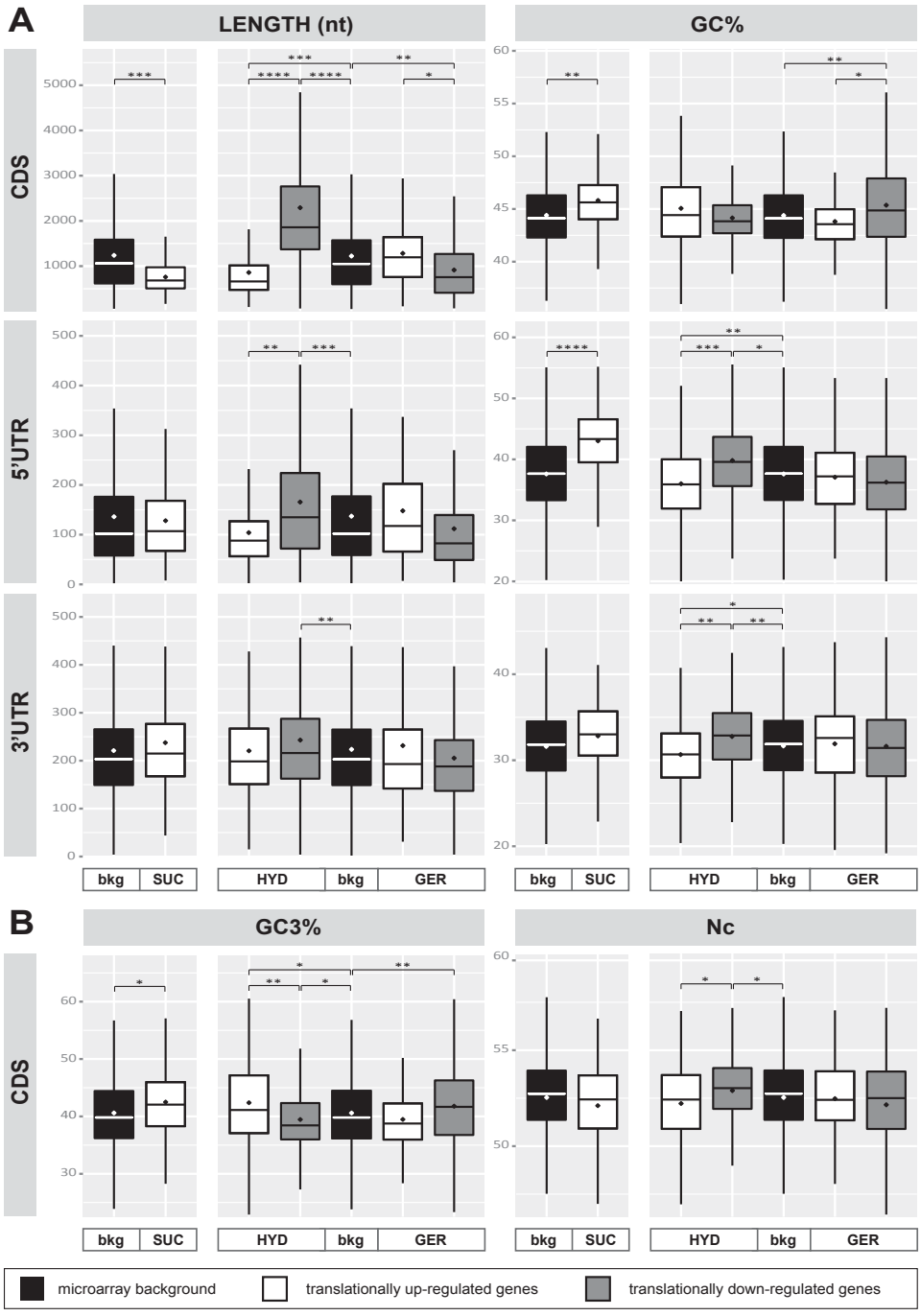
The bioinformatic analyses presented in this study were performed on two different datasets, from Gamm *et al.* (2014) and Bai *et al.* (submitted). In the rest of the chapter we will refer to them as the Gamm and Bai datasets or studies. The Gamm study focused on uncoupling the effect of light and sucrose on translation, as well as on the identification of RP targets to discuss the possible involvement of ribosome heterogeneity. Shortly, *A. thaliana* seedlings were grown for 10 days in constant light, and then exposed for 6 additional hours to either light or dark while being treated with 150 mM of sucrose or sorbitol as an osmotic control (see Methods). Microarray analyses were performed both on the polysome-bound and total mRNA to determine the polysomal occupancy for each mRNA species, which revealed 243 translationally sucrose-regulated genes in the light and 12 in the dark. The dark

sample was excluded from further bioinformatic analyses, as it would have been difficult to produce reliable statistics with such a small gene list. For the same reason, we also excluded the few down-regulated genes in the light sample, focusing on the characterization of the remaining 233 mRNAs translationally enhanced by sucrose.

The Bai study was conceived as a time series, following the germination of *A. thaliana* seeds after imbibition (see Methods). Polysomal-bound and total mRNA were quantified at the stage of dry seed (time 0), early imbibition (6 hours after imbibition, HAI), testa rupture (26 HAI), 80% radicle protrusion (48 HAI), and 80% green seedlings (72 HAI); polysomal occupancy was then compared between consecutive time points. Results showed extensive translational regulation between dry seed and 6 HAI, and between testa rupture and 48 HAI, which we renamed hydration and germination shifts respectively. Our bioinformatic analyses focused on four resulting gene lists: 1204 translationally down- and 435 up-regulated in the hydration shift, and 717 translationally down- and 195 up-regulated in the germination shift. Interestingly, basically no translational regulation was detected between other time points, in spite of widespread transcriptional changes; this indicates that translational regulation is highly specific for certain stages of the germination process.

The Gamm and Bai datasets presented striking differences in the magnitude and direction of translational regulation: seed germination seemed to affect a much larger subset of mRNAs, and mostly through translational repression. Also, no significant overlap could be detected between the regulated genes from each dataset (not shown). Of course the conditions under investigation, as well as the experimental settings, were quite different between the two studies; however, this also means that we can expect different signals to operate in the two different studies, to generate condition-specific translational regulation.

Figure 4.1. Comparison of sequence features between translationally regulated genes and microarray background. (A) *Length and GC content comparison between the coding sequences (CDS), 5'UTRs, and 3'UTRs of translationally regulated genes and microarray background transcripts.* ***(B)*** *GC3 content and codon bias (Nc) comparison between the coding sequences (CDS) of translationally regulated and microarray background transcripts. Black: microarray background; white: up-regulated transcripts; grey: down-regulated transcripts. The sucrose-regulated genes (SUC) from the Gamm dataset are shown in the left panel of each comparison; hydration (HYD) and germination (GER) shift genes from the Bai dataset are shown on the right. Note that background genes are shown separately for each study, as they are based on different microarray platforms (see Methods). P- value ranges: <= e-5 (*), e-10 (**), <= e-20 (***), <= e-50 (****). P-values refer to the difference in median, as calculated using a 1-tailed Wilcoxon test. Diamonds in the boxplots indicate the mean of the distribution; outliers (more than 1.5 IQR lower than the 1st or higher than the 3rd quartile) were excluded to improve the visualization of the results. Actual values are reported in Suppl. Table S1.*


A
LENGTH (nt)
GC%
CDS
5'UTR
3'UTR
bkg
SUC
HYD
bkg
GER
B
GC3%
Nc
microarray background
translationally up-regulated genes
translationally down-regulated genes

Translationally regulated transcripts show significant biases in length and GC content, but not in their codon usage

To investigate whether transcript features correlate with translational regulation, we have compared sequence length, GC content, and codon bias of the translationally regulated mRNAs versus the microarray background for both datasets (see Methods). It has been established that short transcripts and transcripts with low GC content are in general more efficiently translated than long ones (Valleriani et al. 2011; Qu et al. 2011). GC3 content (GC content in the third codon position of synonymous sites) is a related metric, giving information on codon degeneracy. The assessment of codon bias can also be informative, as translational efficiency might be affected by tRNA availability (Wright et al. 2004). Currently one of the most reliable statistics for codon bias is the effective number of codons (Nc); extremely biased genes that use only one codon per amino acid will be scored 20, while less biased genes will obtain higher scores (Sun et al. 2013).

In the Gamm study, transcripts with increased translational efficiency after sucrose treatment were significantly shorter in their total mRNA length, as well as in the length of the coding sequence (CDS), in agreement with the general effect of sequence length on basal translational efficiency (**Figure 4.1** and Suppl. Table S1). In contrast to expectations with regards to basal translational efficiency, the GC content of the CDS and especially 5'UTR of translationally up-regulated transcripts was significantly higher than the one of the background. No significant codon bias could be detected, and the difference in GC3 content observed between translationally regulated and background transcripts is could be likely explained by the difference in GC content.

In the Bai study, we found significantly longer transcripts in the set of translationally down-regulated mRNAs at the hydration shift compared to up-regulated and background transcripts. However, mRNAs in the germination shift showed an opposite trend, showing longer CDS sequences in the up-regulated set, and vice versa for the down-regulated ones (see **Figure 4.1** and Suppl. Table S1). This indicates that translation at the two shifts is likely regulated by distinct mechanisms. Significantly higher than background GC contents were observed in the 5' UTR and 3' UTR of the hydration shift and in the CDS of the germination shift, in both cases in translationally down-regulated mRNAs; the observation is consistent with the general association between GC content and basal translational efficiency. Following an opposite trend, GC3 content was significantly higher in the CDS of translationally up-regulated mRNAs in the hydration shift, for which a marginally significant codon bias, as measured by Nc, was also detected when comparing these transcripts to the down-regulated set (**Figure 4.1** and Suppl. Table S1); however, Nc scores did not indicate a significant difference between the up-regulated group of transcripts and the background (**Figure 4.1** and Suppl. Table S1). The effect is rather counterintuitive and indicates that codon bias is probably not, or not the only, source of regulation.

Neither study showed enrichment for uORFs, nor differences in the context of the start codon when comparing translationally regulated mRNAs against background genes (not shown).

Enriched motifs differ in their distribution and localization across datasets

To investigate more specific sequence signals, we performed a *de novo* motif enrichment search for each set of translationally regulated mRNAs (see Methods). Motif analysis was performed on four regions of the transcripts, namely 5'UTR, CDS, 3'UTR and the whole sequence. We compared the occurrence of the discovered motifs between our gene lists and the background, to filter out sequence signals that are highly represented in the *A. thaliana* genome. The most significant motifs were then mapped on the transcript sequences to observe possible patterns in their localization (this part of the analysis was performed only for the more recent Bai dataset).

Table 4.1. Motif enrichment in translationally regulated genes. *Enriched motifs identified in sucrose (SUC) and hydration shift (HYD) translationally regulated genes; the direction of the regulation is indicated (up or down), as well as the location along the transcript (5'UTR or entire mRNA)*

	reg.	motif #	consensus	e-value*	location	list	tot	BKG	TOT	fold enrich.	p-value**
SUC	up	1a	[ga][ga]AGA[ga]	5.1E-31	5'UTR	29	226	784	19250	3.15	4.62E-08
	up	1b	GAAGAAgAag	1.1E-55	5'UTR	72	226	3262	19250	1.88	2.51E-08
	up	1c	GaaGAAG[ac]	2.0E-49	5'UTR	59	226	2688	19250	1.87	9.24E-07
	up	2a	CG[ga]CG[ag]	1.5E-04	5'UTR	82	226	3413	19250	2.05	1.93E-11
	up	2b	[uca]CGnCGn[ca][gc]	7.5E-15	5'UTR	55	226	1476	19250	3.17	7.72E-15
	up	3a	C[gu]UC[gu][uc]C[gu]UC	5.7E-93	full transcript	160	233	15406	26405	1.34	3.33E-08
	up	3b	[cu]UCU[uc][uc][cu]UCU	1.9E-116	full transcript	166	233	15975	26405	1.35	7.78E-09
HYD	down	1a	CuUCuUCuUC	0.0E+00	full transcript	1041	1204	19835	27827	1.21	9.88E-54
	down	1b	UC[uac]UC[uac]UC[uac]u	8.7E-60	full transcript	1073	1204	20418	27827	1.21	2.63E-61
	down	2a	UCuC[uca]nU[cu]U[cu]	5.3E-152	full transcript	1089	1204	20735	27827	1.21	2.82E-65
	down	3a	GA[auc]GA[auc]GA[au]G	8.1E-216	full transcript	1064	1204	21490	27827	1.14	8.53E-39
	up	3b	[ag][ag]A[gc]AA[ga]AA[ag]	3.8E-113	5'UTR	197	343	8858	19159	1.24	2.99E-06

**original MEME e-value for the motif overrepresentation in the given set of genes, as computed by MEME*

***p-value for the fold enrichment of the motif versus the microarray background, as computed by a one-tailed Fisher's exact test.*

In the Gamm study, de novo motif discovery identified a number of motifs enriched in the sequences of translationally regulated mRNAs after sucrose treatment (**Table 4.1**). Two motifs, 1a and 2b, were found enriched more than threefold in the 5'UTR region. Although overrepresented among translationally regulated transcripts, such motifs do not fully explain translational regulation, as only between 13-25% of the regulated mRNAs showed occurrences.

In the Bai study, a total of five enriched motifs with occurrence significantly higher (p-value $< 10^{-5}$) than expected from the gene background were detected, all of which were identified in the hydration shift (**Table 4.1**). One of these motifs, 3b, was found in the 5'UTR of

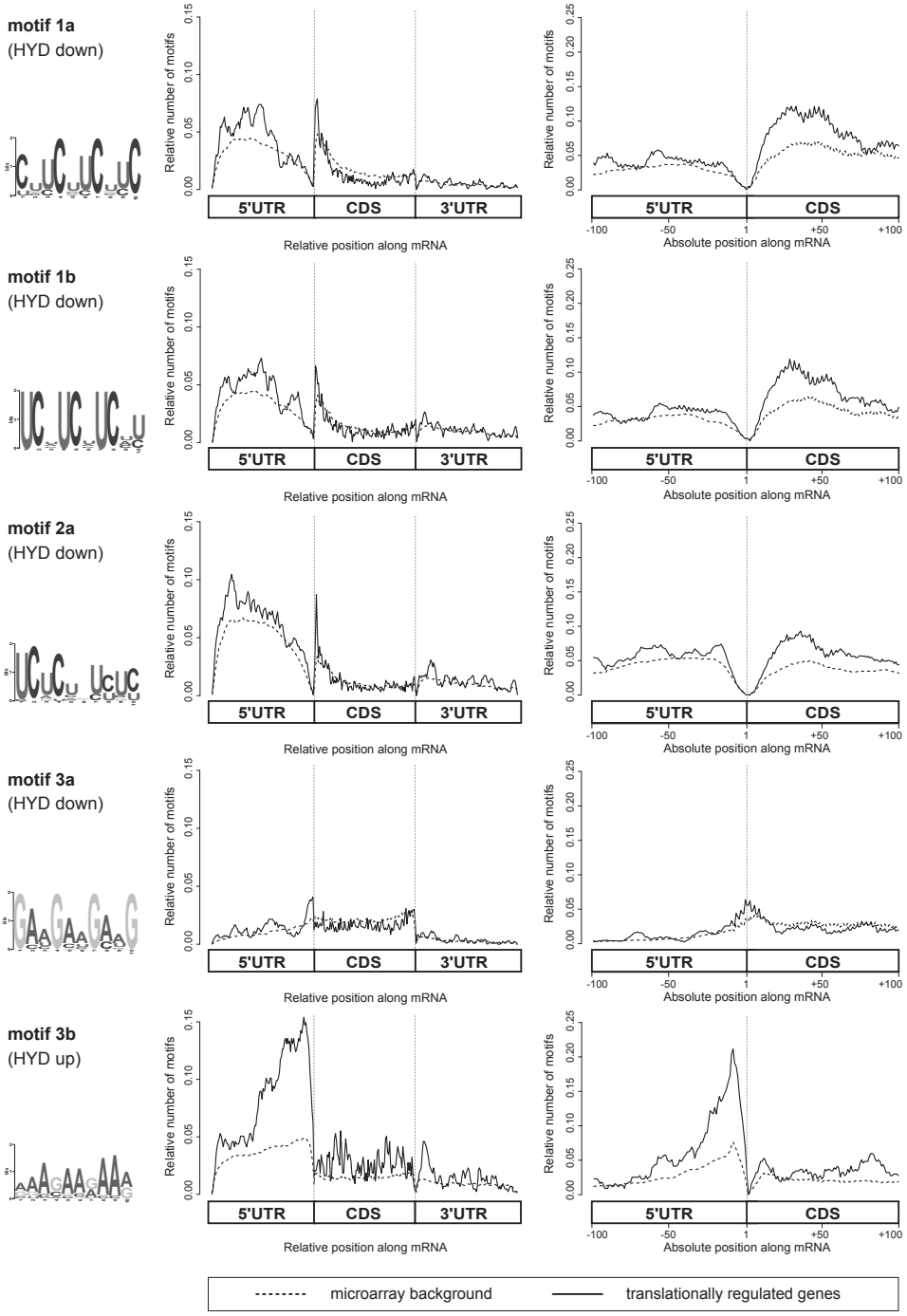

motif 1a
(HYD down)
motif 1b
(HYD down)
motif 2a
(HYD down)
motif 3a
(HYD down)
motif 3b
(HYD up)
Relative number of motifs
5'UTR
CDS
3'UTR
Relative position along mRNA
Absolute position along mRNA
-100
-50
1
+50
+100
microarray background
translationally regulated genes

Figure 4.2. Localization of motifs identified in the Bai dataset. *The plots show the frequency of each motif along the set of translationally regulated transcript in which the motif was found enriched (hydration shift (HYD), either up- or down-regulated genes). Motif codes correspond to those showed in Table 4.1, Bai dataset. A graphical representation of each motif's consensus is also shown. Plots on the left show the relative position of the motif along the transcript, with each region (5'UTR, CDS, 3'UTR) scaled to 100 bins. Plots on the right show the absolute position of the motif around the start codon (+1), expressed in nucleotides. Motif occurrences are shown with a dashed line for the background transcripts, and with a continuous line for translationally regulated transcripts.*

translationally up-regulated transcripts, and the remaining in the full transcript sequence of the down-regulated set. It is to note that in this case motifs were present in the majority of the transcripts, however with lower fold enrichment than detected in the Gamm study. The five hydration shift motifs fell in two groups: UC (motifs 1a, 1b and 2a) and GA (motif 3a and 3b) repeats (**Table 4.1**); the latter group is somehow related to the motifs 1a, 1b, and 1c enriched in the Gamm datasets, however Bai's motif 3a is found in the entire transcript rather than at the 5'UTR like the remaining GA motifs. Motif localization analyses showed that Bai's UC-rich motifs (1a, 1b and 2a) were overrepresented in the entire transcript sequence (**Figure 4.2**). Given their enrichment in translationally down-regulated transcripts, these motifs might have a role in limiting the progression of ribosomes, for instance by recruiting RNA binding proteins, and thus inhibiting translation elongation. Motif 3b is significantly enriched in the 5'UTR of translationally up-regulated transcripts. This adenosine enriched motif is mainly localized in the 50 nt region upstream of the start codon, where due to nucleotide composition it could contribute to a more open secondary structure, potentially favoring the recruitment of translation initiation factors (Xia et al. 2011).

Structure signals correlate with translational regulation

We investigated the role of mRNA secondary structure in translational control by mapping experimentally derived structural scores (Li et al. 2012) to the 5'UTR, CDS and 3'UTR of translationally regulated mRNAs (see Methods). These structure scores represent the ratio of double (ds) to single strandedness (ss) at each nucleotide position along the mRNA, as detected in specific *in vitro* conditions (Li et al. 2012). It has been shown that experimentally detected structure scores are a better predictor of polysome occupancy than *in silico* predicted free energies (Vandivier et al. 2013). Generally, a decrease in structure at the start and stop codon is a conserved feature among eukaryotes, facilitating the association and dissociation of the ribosome to the transcript and thereby correlating with an increase in basal translational efficiency (Kozak 2005; Kertesz et al. 2010; Li et al. 2012).

Translationally up-regulated transcripts in the Gamm study appeared significantly less structured than background transcripts, both at the 5'UTR and towards the end of the CDS region (**Figure 4.3**). While the trend is also visible in the background set, the overall lower

structural score of sucrose-regulated genes is remarkable. Given the enrichment for motifs in the 5'UTR region, this could favor the accessibility of such motifs by regulatory proteins; however the observation is somehow inconsistent with the higher GC content of this group of transcripts, which can be expected to correspond to a higher structure score on average.

Significantly lower average structure scores were also detected for the translationally up-regulated transcripts in the Bai study (**Figure 4.3**). Especially at the hydration shift, both 5'UTR and CDS scores appeared dramatically lower than the background, while the germination shift showed significantly less structure than the background only at the 5'UTR. Interestingly, at this time point the 3'UTR of translationally up-regulated mRNAs appeared slightly more structured than expected. Translationally down-regulated transcripts showed an opposite behavior, with higher than background structure scores at both shifts. At the hydration shift the structural effect seemed less marked in this case, with significantly higher structure than background only at the 5'UTR. However, at the germination shift structure scores in the CDS region of the transcript appear markedly higher. The high structure of the CDS may slow down the progression of the ribosome during the elongation phase, thereby inhibiting the translation of these mRNAs. The higher structure score of translationally up-regulated versus down-regulated transcripts at the 3'UTR however is more difficult to interpret, although the effect is not extremely significant, and might as well represent noise.

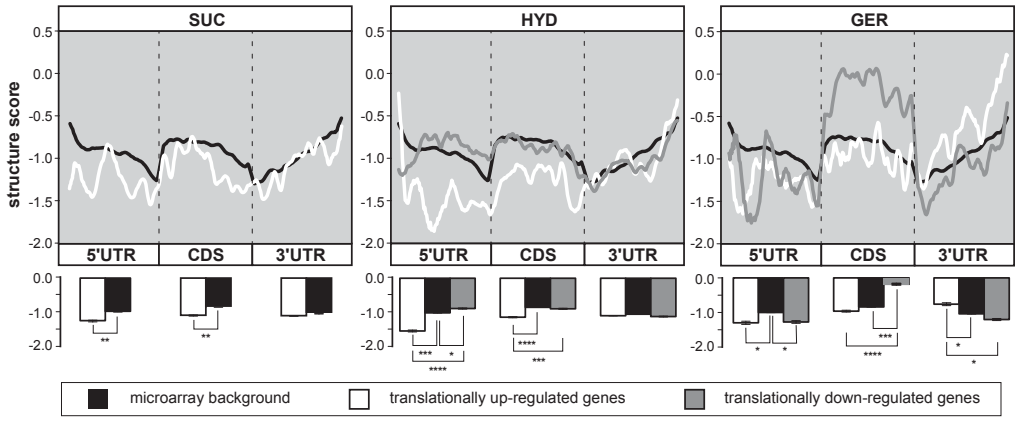


Figure 4.3. Structure score distributions along the sequence of translationally regulated transcripts. *The plots show the structure score distribution along the sequence of sucrose (SUC), hydration (HYD), and germination (GER) shift translationally regulated transcripts; black: background score distribution; white: translationally up-regulated transcripts; grey: translationally down-regulated transcripts. Each region of the transcript (5'UTR, CDS, and 3'UTR) was scaled to 100 bins. The bar plots represent the mean structure scores per transcript region; error bars ± SE are shown. P-value ranges: <= e-10 (*), <= e-20 (**), <= e-50 (***), <= e-100 (****), as obtained from a t-test.*

DISCUSSION

Plant translational regulation has been the subject of an increasing number of studies in recent years. However, these focused mostly on the functional characterization of regulated transcripts rather than on the underlying mechanisms responsible for the observed effects. Especially information about the role of motifs and structured sequence signals remains scarce; moreover, the majority of these studies were based on stress conditions, while translational regulation during physiological processes or in condition of energy abundance is less explored. In this study we focused on two such datasets, investigating the role of both general sequence properties and more specific signals in differential translational regulation.

Both general and specific signals contribute to translational regulation

It is known that long transcript length and high GC content correlate with low basal translational efficiency (Qu et al. 2011; Valleriani et al. 2011), while codon bias may indicate the presence of selective pressure to use more abundant tRNAs in order to increase elongation rates (Wright et al. 2004). In the two datasets analyzed here we observed both agreement and disagreement with expected trends: for instance, in the Gamm study shorter transcripts showed higher polysomal occupancy, as expected; however, in the same dataset we also observed a higher GC content in the 5'UTR, unexpectedly correlated with increased translational efficiency. Up-regulated transcripts in the germination shift of the Bai study were also longer than down-regulated ones, showing again an opposite behavior to the one associated with basal translational efficiency. Codon bias seemed to play a role in regulation, although marginal, only in the hydration shift up-regulated genes; GC3 content differences were more significant, but likely explained by the even stronger trends observed for the general GC content. Interestingly, the decreased CDS length and increased 5'UTR GC content of translationally up-regulated mRNAs in the Gamm study were previously observed for genes translationally repressed under dark or hypoxia treatments (Kawaguchi and Bailey-Serres 2005; Juntawong and Bailey-Serres 2012). It is possible that the features responsible for a decrease in translational efficiency during stress conditions are also limiting under control conditions.

The observation of significant trends both in agreement and opposite to the expected association between general sequence features and direction of translational regulation indicates that, at least in certain conditions, some degree of regulation is potentially achieved by amplifying basal translational efficiency; however in other conditions this is obviously overwritten by different signals. Interestingly, 238 of the 243 sucrose-regulated genes in the Gamm dataset were detectable in the control samples of a recently published dataset obtained by ribosomal profiling of *A. thaliana* seedlings (Juntawong et al. 2014), indicating that these genes are translated also in normal growth conditions. However, their translational efficiency in the control (1% sucrose) was markedly lower than the

dataset average, indicating that normally these genes might not be efficiently translated. Whether this is related to their higher GC content or other sequence features remains to be established. More generally, it is unlikely that basic sequence properties alone would account for the differential translation of an mRNA in different environments; particularly mechanisms based on the intervention of RPs and RBPs would be incompatible with a simplistic model of regulation, as protein-RNA interactions would require specific sequence- or structure-based binding sites.

Condition-dependent changes in translational efficiency likely rely on multiple regulatory mechanisms

The analysis of mRNA motifs and structured elements revealed several sequence signals in the two datasets. Enriched motifs appeared sufficiently distinct between sets of translationally regulated genes, however they could not be identified in all of these sets: for instance, in the Bai study only the hydration shift showed motif enrichment, whereas for the germination shift no significant sequence motif could be identified. This might indicate that different mechanisms prevail at different stages of seed germination, for instance motif-based at one shift, and structure-based at the other. Two motifs identified in the Gamm study were enriched more than threefold, however they were not widespread in the regulated gene set. Therefore, in spite of the strong signal, they cannot explain translational regulation for all of the sucrose-regulated mRNAs. A different situation was observed for the hydration shift in the Bai study, where nearly all translationally regulated transcripts showed motif occurrences, however with low enrichment when compared to the background. It is possible that, although such motifs are present also in non-regulated transcripts, in the hydration datasets they act synergistically with other signals to achieve specific regulation. For instance, mRNA folding might limit their accessibility by regulatory proteins in the general transcriptome background, preventing binding. Although widespread, the presence of these motifs might therefore be relevant only for a subset of transcripts, in which the motifs are accessible. Another difference between the two datasets is that, in the Gamm study, most enriched motifs were discovered in the 5'UTR, while in the germination shift the majority of motifs seemed to be spread over the whole transcript sequence. This might indicate that regulatory signals operate differently in the two experimental conditions: for instance, the localization of motifs over the entire sequence of translationally down-regulated transcripts in the Bai dataset might suggest a role in limiting the progression of the ribosomes, possibly through the recruitment of RBPs. On the contrary, the localization of Bai's motif 3b, as well as of those identified in the Gamm study, at the 5'UTR of translationally up-regulated transcripts might indicate the involvement of translation initiation factors, with the effect of enhancing translation initiation (Xia et al. 2011). Once again, the presence of structure might also have an effect on recognition; for instance, in spite of their presence on the entire transcript, not all motifs identified in Bai's down-regulated genes might be accessible by regulatory proteins.

RBPs have been shown to be able to interact with several RNA motifs, and might be responsible for various regulatory mechanisms, such as mRNA sequestration to prevent translation, but also translational enhancement by facilitating the recruitment of initiation factors. For instance, the pyrimidine (UC) and purine (GA) enriched motifs identified among some of the regulated transcripts may represent binding sites for polypyrimidine tract-binding protein (PTB) (Singh et al. 1995; Perez et al. 1997; Oberstrass et al. 2005). Bai's motif 3a is also similar to the target sequence GAAGAAGAAGCUC) of SERINE/ARGININE-RICH PROTEIN SPLICING FACTOR 40, which acts as exon enhancer mediated by a complex of nuclear proteins (Yeakley et al. 1996). The *A. thaliana* SR paralog SERINE/ARGININE-RICH SC35-LIKE SPLICING FACTOR 33 has been identified and plays a role in regulating alternative splicing (Thomas et al. 2012), which might also be involved in the regulation of translational efficiency in our datasets. The purine-rich 5'UTR motifs identified in the translationally up-regulated 5'UTR transcripts identified both in the Gamm and Bai datasets is in agreement with the observation of increased transcript stability and translation rates of transcripts bearing AG-rich motifs in their 5' leaders (Vaughn et al. 2012); the study also showed the conservation of such motifs across six different families of dicots. To establish a clear link between these many clues and actual motif-based mechanisms of regulation in our datasets, it would be extremely useful to experimentally determine the accessibility of motifs by regulatory proteins. Regulation by direct recognition of structured mRNA elements, rather then linear motifs, should also be considered.

Our structure analyses showed that translationally up-regulated transcripts in the Gamm and Bai studies were both significantly less structured than background ones at the 5'UTR and in the CDS region. Translationally down-regulated transcripts at each of the hydration and germination shifts in the Bai study showed an opposite behavior; however significant structural differences with the background were driven by 5'UTR and CDS, respectively, indicating once again that regulatory mechanisms are not the same at each stage of seed germination. For instance, a high structure score at the 5'UTR can prevent translational initiation, while a structured CDS is more likely to influence elongation by slowing down ribosome progression. The sensitivity of translational efficiency to mRNA structure at different stages of seed germination might be mediated by the differential activity of RNA helicases, through their ability to unravel mRNA structures. RNA helicases have mostly been described in abiotic stress studies, however the *A. thaliana* RNA helicase LOW EXPRESSION OF OSMOTICALLY RESPONSIVE GENES 4 (LOS4) has been linked to the regulation of seed germination and other developmental processes (Gong et al. 2002; Gong et al. 2005). Whether LOS4 or other helicases play a role in the translational regulation of physiological processes, such as the ones described, remains to be investigated. As previously mentioned, regulation of translation by RPs and RBPs might also rely on structured elements. The experimental data on mRNA secondary structure used in our analyses provided information about the ratio of double- to single-strandedness for each mRNA nucleotide position. While this information can be used to identify patterns of high and low structure scores along translationally regulated mRNAs, as we reported, interactions with RPs and RBPs

might rely on more specific structured elements, that are difficult to predict from simple scores. Accordingly, complex mRNA structures composed of hairpins and loops are generally unstable and likely subjected to cellular conditions, making *in silico* predictions not completely reliable. While these predictions are still informative, for instance providing clues on whether a given sequence stretch is likely to assume a stable structural conformation or not, it would be desirable to constrain such predictions with *in vivo* data, rather than relying on *de novo* reconstructions. Also, structural data should be ideally obtained for both control and treatment conditions. The structural dataset we used (Li et al. 2012) is an improvement in this sense, but still suboptimal given that the data were obtained *in vitro*, therefore not necessarily reflecting real cellular conditions, and for different experimental treatments than the one used for our transcript measures.

Overall, the changes in translational efficiency observed during sucrose treatment and seed germination are likely the result of the interplay between several different regulatory mechanisms, some of which we have just started to identify. Different combinations of motifs and structural features seem to indicate that regulatory signals are specific for each group of translationally regulated transcripts, however more experimental work is necessary to capture their exact mechanism of action.

CONCLUSION

In this study we performed an in depth analysis of transcripts showing differential translational regulation during sucrose treatment and seed germination. Both general and specific sequence signals seemed to play a role, with motif and structured elements likely driving the observed condition-specific changes in translational efficiency. Sequence features such as motifs and structure alone might not result in strong signals, however the observed effects on translational regulation could be achieved with the interplay of multiple mechanisms. Structured elements remain particularly challenging to identify, while likely to play an important role both as direct recognition elements, and in regulating the accessibility of sequence motifs by putative RP and RBP regulators. Given the complexity of signal interactions, future studies on translational regulation might benefit from the use of modeling and multivariate analysis methods, in order to better integrate and understand the interplay between different sources of regulation.

METHODS

Sample preparation

Gamm dataset

A. thaliana seeds (accession Col-0) were sterilized in 20% bleach for 20 min, washed 5 times with sterile

deionized water, and stratified in water at 4°C for 2 days. Seedlings were grown in 250 ml Erlenmeyer bottles containing 100 ml 0.5 Murashige-Skoog medium for 10 days under constant shaking and in constant light at 22°C. Flasks were covered with aluminum foil or left in the light, and treated with 150 mM sorbitol or sucrose for six hours. Material was harvested by washing seedlings with deionized water and snap-freezing in liquid nitrogen.

Ribosomes and ribosome-bound mRNA were extracted and separated over sucrose gradient as described before (Mustroph et al. 2009). Approximately 10 ml of powdered plant material were extracted using 20 ml polysome extraction buffer PEB (0.2 M Tris, pH 9.0, 0.2 M KCl, 0.025 M EGTA, 0.035 M $MgCl_2$, 1% Brij-35, 1% Triton X-100, 1% Igepal CA 630, 1% Tween 20, 1% PTE, 5 mM DTT, 50 mg/mL Cycloheximide, 50 mg/mL Chloramphenicol, 80 mM betaglycerophosphate, 1 mM Sodium Molybdate, protease inhibitor cocktail (Sigma Aldrich), phosphatase inhibitor cocktail 3 (Sigma Aldrich)) and the extract was cleared by filtering through MiraclothTM and centrifugation. An aliquot of 500 µl was taken for total mRNA extraction; the extracts were loaded on top of a sucrose cushion (1.75 M sucrose in PEB without detergents) and centrifuged (18 h, 90000 g) in a Beckman Ti70 rotor. The resulting pellet was taken up in wash buffer (0.2 M Tris, pH 9.0, 0.2 M KCl, 0.025 M EGTA, 0.035 M $MgCl_2$, 5 mM DTT, 50 mg/mL Cycloheximide, 50 mg/mL Chloramphenicol, 80 mM betaglycerophosphate, 1 mM Sodium Molybdate) and loaded on a 20-60% sucrose gradient. After ultracentrifugation (1.5 h, 190000 g) in a Beckman Sw55Ti rotor, the gradients were fractionated into 12 fractions using a Teledyne Isco Density Gradient Fractionation System with online spectrophotometric detection of A_{254}. These fractions were used separately or pooled into three samples containing non-polysomal material, small polysomes and large polysomes. Areas under the curves were calculated after subtracting the baseline obtained by measuring a blank gradient and normalizing to total area under the curve to account for possible uneven loading of the gradients. When no cushion enrichment step was performed, 1 ml of packed material was extracted with 1 ml of polysome extraction buffer as described above. Extract was filtered and centrifuged before loading of 750 µl directly on 20-60% sucrose gradients. Ultracentrifugation and fractionation were performed as described. RNA was extracted using Guanidine-HCl extraction, followed by a clean-up step using the Spectrum Plant Total RNA Kit (Sigma Aldrich). To each sample (single or pooled fractions), 50 pg of luciferase RNA was added prior to mRNA extraction to allow normalization independent of the original RNA content of the sample.

Bai dataset

A. thaliana fully after-ripened seeds (accession Col-0) were used for all assays. The timing of testa and endosperm rupture and seedling greening of fully after-ripened was performed as described previously (Joosen et al. 2010). In brief, two layers of Blue blotter paper (Anchorpaper company, www.seedpaper.com) were equilibrated with 48 ml demineralized water in plastic trays (15 x 21 cm). Six samples of approximately 50 to 150 seeds were spread on wetted papers using a mask to ensure accurate spacing. Piled up trays were wrapped in a closed transparent plastic bag. The experiment was carried out in a 22°C incubator under continuous light (143 µmol $m^{-2}s^{-1}$). Germination parameters were manually counted. For ribosome analyses, dry seeds were imbibed as mentioned above. Seeds and seedlings were harvested at each physiological state during seed to seedling transition, frozen in liquid nitrogen followed by freeze-drying. The dry material was stored at -80°C until further analyses.

For the isolation of polysomal RNA, about 400 mg of freeze-dried tissue was extracted with 8 ml

of polysome extraction buffer, PEB (400 mM Tris, pH 9.0, 200 mM KCl, 35 mM $MgCl_2$, 5 mM EGTA, 50 µg/mL Cycloheximide, 50 µg/mL Chloramphenicol) modified from (Subramanian 1978). The extracts were loaded on top of a sucrose cushion (1.75 M sucrose in PEB) and centrifuged (18h, 90,000 g) using a Beckman Ti70 rotor for 18 h (Beckman Coulter, Brea, USA). The resulting pellet was resuspended in wash buffer (200 mM Tris, pH 9.0, 200 mM KCl, 0.025 M EGTA, 35 mM $MgCl_2$, 5 mM DTT, 50 µg/mL Cycloheximide, 50 µg/mL Chloramphenicol) and loaded on a 20-60% linear sucrose gradient, centrifuged at 190,000 g for 1.5 h at 4°C using Beckman SW55 rotor (Beckman Coulter). After ultracentrifugation, the gradients were fractionated into 20 fractions using a Teledyne Isco Density Gradient Fractionation System (Teledyne Isco, Lincoln, USA) with online spectrophotometric detection of A_{254}. The fractions corresponding to the polysome region in the ribosome profile were pooled for future analysis. The ribosome abundance is reflected by the area under the curve and was calculated after subtracting the baseline obtained by measuring a blank gradient and normalizing to total area under the curve to account for possible uneven loading of the gradients.

Microarray data analysis

Gamm dataset

Microarray analysis was performed in 3 biological replicates using Affymetrix GeneChip Arabidopsis AGRONOMICS Genome Arrays by ServiceXS, Leiden, The Netherlands. Microarray analysis was performed for three fractions of the sucrose density gradient (non-polysomal, small and large polysome). However, as the non-polysomal fraction contained much less RNA than the others, data obtained for these samples was not used in background correction and normalization to avoid the introduction of artifacts. Furthermore, initial analyses revealed that the list of genes found affected in small and large polysomes were nearly identical. Therefore, data for the small polysomal fraction was used to represent the polysomal fraction. Labeled sense stranded cDNA was synthesized with the Ambion WT Expression kit using 100 ng RNA. Fragmentation and terminal labeling with 5.5 µg sscDNA was performed using the Affymetrix Terminal Labeling Kit. Nanodrop and BioAnalyzer were used to assess the concentration and the quality of the cRNA and fragmented sscDNA. 5.5 µg fragmented sscDNA were used for hybridization on the Affymetrix GeneChips Arabidopsis AGRONOMICS Genome Array. Hybridization, washing, and staining were performed with the GeneChip Hybridization, Wash and Stain Kit. Affymetrix GeneChip Command Console (v3.1) software was used to operate the Affymetrix fluidics stations and the scanned array images were analyzed using Affymetrix Command Console Viewer software. The GeneChip data were analyzed using the R statistical programming environment and the Bioconductor packages (www.r-project.org) (Irizarry et al. 2003; Smyth 2004). The aroma.affymetrix package was used to perform RMA background correction, normalization, probe summarization, and quality check of the data. The LIMMA package was used to obtain gene expression data (Smyth 2004). Steady-state mRNA alterations were calculated using the ratio between the values obtained for total RNA in the sucrose treatment with the corresponding control (light or dark). Polysome association alterations were calculated using the ratio between the values obtained for polysomal RNA in the sucrose treatment with the corresponding control (light or dark). Polysomal occupancy alterations were calculated using the ratio between the polysome association and the steady-state levels. P-values were adjusted for multiple testing with the Benjamini and Hochberg

method (Benjamini and Hochberg 1995). Raw data is deposited along with description of the experimental setup in the GEO repository under the accession nr: GSE59306.

Bai dataset

Two translational shifts were identified by comparing the transcripts changes in polysome-bound and total mRNA during consecutive imbibition stages. Affymetrix Arabidopsis Gene 1.1 ST Arrays (Affymetrix, Santa Clara, USA) were hybridized using the GeneChip® 3' IVT Express kit (cat. # 901229) according to instructions from the manufacturer. Hybridization data were analyzed and gene specific signal intensities were computed using the R statistical programming environment (www.r-project.org), the BioConductor package Affy (Gautier et al. 2004), and the Brainarray cdf file ver. 17.1.0 (http://brainarray.mbni.med.umich.edu). Following RMA normalization (Irizarry et al. 2003) constant low intensity signals (logExprs<4 in all samples) were removed. The LIMMA package was used to obtain gene expression data (Smyth 2005; Diboun et al. 2006). Lists of translationally regulated genes identified by comparing total and polysomal mRNA are provided in the Supplementary Material (Suppl. Dataset S1).

Bioinformatic sequence feature analysis

Translationally regulated genes were compared to the appropriate microarray background for each study for several sequence features using custom scripts. The distributions of sequences length and GC content were evaluated separately for CDS, 5'UTR, 3'UTR, and full transcript. CDSs were additionally analyzed for GC3 content, after removing sequences missing the start codon and/or containing premature stop codons. CDSs shorter than 100 codons were further removed for the codon bias analysis, measured using a recent and more informative implementation of the Effective Number of Codons (Nc) index (Sun et al. 2013). In the Bai study, all of the above analyses were performed on two additional CDS subsets, i.e. protein-coding genes having both annotated 5'UTR and 3'UTR, or no annotated UTR (UTRs were called present when having length > 1 nt). Enrichment for annotated uORFs (von Arnim et al. 2014) and analyses of the sequence context around uORFs and main ORF start codon position (-5 to +8) were also computed, using custom scripts comparing the gene lists against the background. A Wilcoxon signed-rank test was adopted (median as test statistic) to account for deviations from normality in the distribution of values.

Motif enrichment

DNA motif analyses were performed using the MEME suite (Bailey et al. 2009) for full transcripts, 5'UTR, CDS and 3'UTR sequences, extracted from the TAIR10 database (www.arabidopsis.org) whenever the corresponding gene annotation was available. For genes with multiple isoforms, only the TAIR10 representative gene model was used. Background dinucleotide frequencies were provided separately for each sequence type. Translationally regulated genes from the Gamm study were scanned for short and long DNA motifs (4-6 and 8-10 nucleotides, respectively) separately; given the observed redundancy of short motifs, gene lists from the Bai study were scanned for motifs with minimum and maximum width set to 6 and 10, respectively. To test specificity of the resulting motifs, FIMO (Bailey et al. 2009) was used to scan all genes represented on the microarray for motif hits in the corresponding sequence type. For the Gamm study, motifs with FIMO p-value ≤0.001 (short motifs)

or ≤0.0001 (long motifs) were considered significant, whereas in the Bai study all motifs with p-value ≤0.001 were retained. Obtained motif counts were used to compute the enrichment p-value for the gene lists versus the background using a one-tailed Fisher's exact test, performed with a custom script and the R software package (www.r-project.org).

Structure and motif position analyses

Experimentally determined per-nucleotide structure scores (Li et al. 2012) were used to calculate the average structure score of the genes with significantly increased and decreased ribosomal association for both the Gamm and Bai datasets. Relative scaling was achieved by averaging the structure scores for each region (5'UTR, CDS and 3'UTR) over 100 bins. Standard errors and t-tests were calculated using the Python SciPy module (www.scipy.org). For the Bai dataset, we extracted the localization of each significantly enriched motif from the FIMO output, and used it to calculate the relative number of motifs per position along the mRNA. Relative scaling of 5'UTR, CDS, and 3'UTR was performed as for the structure scores.

SUPPLEMENTARY MATERIAL

Supplementary materials for this chapter are made available online at the following link:
http://bioinformatics.bio.uu.nl/alessia/THESIS.html
These include additional data and information from the Bai study that we would like to make accessible prior to publication; it should be noted that not all of the material is strictly relevant to our bioinformatic analyses.

Gamm and Bing datasets
Supplementary Table S1. Input data for Figure 4.1.

Bai dataset
Supplementary Figure S1. Translatome profiling during seed germination.
Supplementary Figure S2. Dynamics of the two identified translational shifts.
Supplementary Figure S3. Transcriptional and translational expression shift following seed germination.
Supplementary Figure S4. Temporal differences between transcriptome and translatome using over- representation analysis (ORA).
Supplementary FIgure S5. Overlap of the shift genes with other datasets.
Supplementary Data S1. List of translationally regulated genes during the two shifts.

Additional material from the Gamm study can be found at the publisher's website:
http://bmcplantbiol.biomedcentral.com/articles/10.1186/s12870-014-0306-3

5

Summarizing Discussion

As introduced in Chapter 1, we explored various systems involved in the regulation of plant energy homeostasis, both from a transcriptional and translational perspective. The biological distinction between the two levels of regulation corresponded in our work to different computational analysis strategies, as we investigated the first by means of phylogenetic reconstruction methods, and the second with an integrative bioinformatics approach. Here we summarize the results described in previous chapters, discussing the biological relevance of our specific findings as well as the main questions that will need to be addressed in the future. We will also cover issues of general interest, such as the technical limitations linked to each type of analysis, and the improvements needed to overcome them.

Phylogenetic analyses of plant transcription factors

Evolution and conservation of ABI4 orthologs

In Chapter 2 we investigated the evolution and conservation of ABI4, a well-known regulator involved in seed development and germination. We noticed that in spite of extensive experimental work, a comprehensive survey of its orthologs in higher plants was still missing. Our analyses aimed to fill this gap, as well as to identify conserved sequence elements of possible functional relevance.

Our results showed that most plant species possess a single ABI4 gene, and only occasionally two as a consequence of recent whole genome duplication events. This suggests that the function of ABI4 is highly conserved, and that it probably relies on finely balanced transcript levels, since additional paralogous copies in most cases are not retained after duplication. Our main finding however was the discovery of a new conserved "ABI4" motif in all analyzed higher plants, which has not been described before, nor does appear in other gene families. The motif could not be found in the ABI4 orthologs from the lower plants *Physcomitrella patens* and *Selaginella moellendorffii*. These species present additional peculiarities, such as a second copy of ABI4, and lower overall sequence similarity, likely explained by evolutionary distance. These observations indicate that ABI4 might present important functional differences between lower and higher plants; the new "ABI4" motif in particular could have contributed to specific regulatory properties in the latter group.

In summary, our analyses showed the widespread presence of ABI4, as well as the conservation of a putative regulatory motif in the orthologs from higher plants. We believe the most natural follow-up to these findings should be the experimental characterization of the "ABI4" motif. A year after the publication of our results, another article was published where the authors experimentally showed that the motif, termed "LRP" in this study, has indeed an impact in ABI4 activity (Gregorio et al. 2014). The results ruled out that the motif functions by affecting subcellular localization or ABI4 mRNA stability, and hypothesized instead a role in the phosphorylation status of the ABI4 protein, or its interaction with

other proteins belonging to the transcriptional machinery. The functional characterization of the "ABI4" motif is therefore still an open subject for further experimental research. Another interesting follow-up question would be the characterization of the functions of the ABI4 orthologs from the lower plants *P. patens* and *S. moellendorffii*. Complementation experiments between the ABI4 orthologs from these species and higher plants, similar to what has been done between *Arabidopsis thaliana* and maize (Niu et al. 2002), could help shedding light on the ancestral function of ABI4, and understanding its role in evolutionary terms.

Evolution and conservation of the C/S1 bZIP transcription factor network

In Chapter 3 we investigated the evolution and conservation of members of the C and S1 groups of bZIP transcription factors. Our analyses were motivated by the lack of consideration for the interconnected functions of these transcriptional regulators in the C/S1 bZIP transcription factor network, which among other things has not been studied in evolutionary terms. The specific regulation of S1 bZIP through the translational control mechanism termed SIRT also appeared completely uncharacterized from a phylogenetic perspective, despite its high relevance for studies in stress signaling. Additionally, as for ABI4, we argued that a comprehensive survey of C and S1 bZIP genes in plants would be desirable, both to identify novel regulatory elements conserved between orthologs, and to enable the integration of functional information between model and crop species.

Our analyses were able to resolve or at least bring significant improvements on each of the points listed above; nevertheless, we also encountered some issues. Our phylogenetic reconstruction was still affected by low bootstrap support values, despite the vastly improved and larger species sampling. As previously described (Chapter 1 and 3), we overcame the problem by using an unconventional approach, consisting in the comparison of independent phylogenetic trees in order to obtain a consensus topology. In flowering plants, our results not only confirmed the conservation of S1 orthologs in both dicots and monocots, but also allowed us to establish detailed orthology relationships between lineages. For instance, we now know that *A. thaliana* bZIP11 has directly comparable orthologs in most eudicots; however, when considering rice, the comparison with its orthologs OsbZIP76, OsbZIP77, and OsbZIP78 makes sense only if including the *A. thaliana* sequences bZIP2 and bZIP44, which are closely related to bZIP11. This is because our phylogeny shows that all rice orthologs and all *A. thaliana* orthologs originated from independent duplications in each plant lineage. In other cases, we uncovered that certain comparisons between species are not even possible: for instance several fabids, in the dicot lineage, entirely lack orthologs for *A. thaliana* bZIP1. The list of observations could go on with further examples. Overall, our main accomplishment from a comparative genomic perspective was to provide detailed phylogenetic trees that can be used to correctly compare and integrate experimental information from different species, which for such complex groups of orthologs would not possible using standard similarity searches.

Our survey of C and S1 bZIPs in flowering plants is also a precious resource to identify novel regulatory elements, of which however we did not take advantage ourselves due to time and research scope limitations. During the manual inspection of the alignments, we noticed that C bZIP sequences in particular seemed to host many conserved motifs outside of the bZIP domain, which might hint to regulatory functions. S1 bZIPs on the contrary showed barely any conservation outside of the bZIP domain, except for the uORFs involved in SIRT. Perhaps this difference is related to different roles within the C/S1 bZIP transcription factor network, with C bZIPs possibly being targets or effectors in additional regulatory protein-protein interactions; this in turn might have consequences for the activity of C/S1 bZIP dimers, via mechanisms operating either before or after their formation. We expect other researchers to make use of this information with more specific follow-up studies.

From a functional as well as evolutionary perspective, our most significant result was the discovery of S1 bZIPs in gymnosperms, supported by experimental evidence confirming their regulation by SIRT. The grouping of predicted gymnosperm S1 bZIP with angiosperm S1 bZIPs received bootstrap support below recommended reference values; however, the matching results between tree topology, the conservation of uORFs with S1-like features, and the experimental validation of SIRT, gave us confidence to state that the ones identified in gymnosperms were genuine S1 orthologs. Although gymnosperms do not possess as many S1 bZIPs as angiosperm species, our findings set the basis for studies on the physiological roles of S1 bZIPs in different seed plant lineages, especially in relation to stress signaling. We also envisage that our experimental results could be especially relevant for the clarification of the molecular mechanisms responsible for SIRT itself. While research on SIRT noticeably progressed since its first observation (Rook et al. 1998), the details of how sucrose and S1 uORFs interact with the translational machinery to prevent the translation of S1 bZIP proteins are still unclear. It has been hypothesized that regulation relies on the partial translation of uORFs into an arrest peptide (termed SC-peptide, for sucrose control); however this has not been isolated as yet, preventing further characterization of the mechanism. The fact that in our experiments gymnosperm uORFs could mediate SIRT in an *A. thaliana* background, in spite of hosting only 2 of the 4 fundamental residues known in bZIP11 (Rahmani et al. 2009), might indicate that regulation relies on conserved structural characteristics of the SC peptide rather than on individual amino acid residues. As a follow-up, it would be useful to test mutated versions of the gymnosperm uORF sequences to individuate crucial residues for SIRT in these species, and compare them to *A. thaliana* S1 uORFs. The identified gymnosperm sequences might also serve as a useful tool to identify putative SC-peptide binding factors in *A. thaliana* with biochemical methods.

Another important discovery we made was the existence of both C and S bZIP orthologs in the multicellular algae *Klebsormidium flaccidum*, as well as in early branching land plants. This finding indicates that the first C and S bZIPs likely originated as reciprocal sister groups by duplication from a previous hybrid C/S bZIP algal sequence, contradicting the previously proposed hypothesis that they emerged at different times (Côrrea et al. 2008). The

presence of C and S bZIPs in an aquatic species such as *K. flaccidum* also indicates that these transcription factors were already present before the emergence of land plants, and could have represented an important component of their adaptation toolkit. We also observed that both groups remained limited in size until the appearance of seed plants, where they underwent a dramatic expansion, especially in the S bZIP subfamily; this includes the duplication events responsible for the origin of S1 bZIP orthologs. Hence S1 and other S bZIP groups have experienced a large expansion not only in flowering plants, likely contributing to their adaptation and diversification, but potentially also in other seed plants, such as gymnosperms.

This finding also represents an important step forward in the study of the C/S1 bZIP transcription factor network. The discovery that C and S bZIPs were born as sister groups provides a direct explanation for the origin of heterodimerization in the C/S1 bZIP transcription factor network, given the tendency for bZIPs to interact amongst close paralogs. In our opinion, the sister group hypothesis favors a scenario where the network emerged as a specialized system retaining pre-existing heterodimerization capabilities between ancestral C and S bZIPs, rather than reinventing them later on. While interactions between C and non-S1 bZIP orthologs have not been detected in *A. thaliana* (Ehlert et al. 2006), our finding predicts that other S bZIPs might have inherited the capability to form dimers with C sequences in different plant species. This in turn would require researchers to take into consideration additional potential interactions when performing functional experiments on members of either group. An interesting follow-up experiment would be the testing of interactions between reconstructed ancestral C and S bZIPs, to confirm that heterodimerization was present early on, as hypothesized; another useful experiment would be the assessment of interactions in basal seed plants, such as *Amborella trichopoda*. Since this organism possesses both S1 and non-S1 orthologs and emerged before the speciation of flowering plants, results could be useful to infer a likely scenario for the conservation of interactions between C and S bZIPs in several other species. Ultimately these experimental tests would allow researchers to formulate more detailed hypotheses on the role of the C/S1 bZIP transcription factor network in plant evolution.

Availability of data for plant phylogenetic studies

During the search for ABI4 and C and S bZIP orthologous sequences for our phylogenetic analyses, we were faced with large biases in the distribution of available sequenced species in different plant lineages. In contrast to the increasing availability of genome annotation data in flowering plants, at the time of our analyses only a handful of species spanning the enormous evolutionary gap between algae and seed plants had been sequenced, leaving groups such as mosses and ferns largely underrepresented. While it is understandable that sequencing efforts tend to be directed towards more economically relevant species, we are of the strong opinion that including other groups of plants is an absolute necessity to achieve significant progresses in evolutionary and comparative genomics studies. In a

specific example, the increased availability of non-flowering plant genomes, such as those from gymnosperms, allowed our phylogenetic reconstruction of the C and S bZIP groups to bring substantial improvements on previous knowledge of these transcription factors (Côrrea et al. 2008). The availability of sequences from more basal species is pivotal to answer biological questions and raise hypotheses that are of interest also to researchers working on more economically relevant species. For instance, our hypothesis on the emergence of C/S1 bZIP dimers from previous C/S interactions would have not been possible without sequencing data from the multicellular alga *K. flaccidum*. If this scenario is confirmed, the implication would be that C bZIPs might be able to interact with other S orthologs apart from S1, providing an important clue for experimental work on bZIP dimers in flowering plants. Recently, the OneKP project (www.onekp.org) was started with the specific aim to provide transcriptome data for a large number of underrepresented plant lineages. Although sequences are not made fully accessible at this time, this project is likely to become an important resource for future plant phylogenetic studies.

Plant translational regulation

In Chapter 4, we investigated the sequence features responsible for the translational selection of mRNAs in specific conditions, using data from sucrose treatment (Gamm) and seed germination (Bai) studies. Overall, our analyses identified several enriched motifs among translationally regulated transcripts, as well as significant structural differences with the total transcriptome. General transcript properties, particularly length and GC content, also appeared involved in the regulation, sometimes showing very strong biases in the set of translationally regulated transcripts. Different experimental conditions and time points were characterized by different combination of sequence features, and motifs in particular showed specific localization and coverage. We hypothesized that, rather than relying on individual motifs or features, translational regulation during seed development, as well as in high energy availability conditions, is likely achieved through the interplay between different levels of regulation, from basic sequence properties, to motifs, to mRNA folding.

The question remains whether it is possible to identify more specific transcript signals involved in translational regulation, in the analyzed conditions as well as in different ones. The accurate strategy we adopted for motif analyses did not reveal strong signals, suggesting that motif enrichment is not the main, or the only, factor involved in translational regulation, as hypothesized. A more complex and biologically plausible approach would be to consider motif accessibility by putative regulatory molecules. It is possible that some motifs, which do not appear enriched in translationally regulated transcripts when compared to the general transcriptome, are however more “visible” to RPs or RBPs due to their favorable localization in unstructured regions in the set of regulated transcripts. Conversely, motif occurrences in non-regulated transcripts might be less accessible due to their localization in relatively more structured regions. If we computed motif enrichment only taking into consideration

those located in accessible regions, perhaps we would be able to detect stronger motif signals. In our study we plotted the localization of motif occurrences along translationally regulated and background transcripts to detect positional effects; however, this approach did not directly take into account the presence of local structure. To investigate the interplay between motifs and structure, we could have matched structure scores to the nucleotide position of each motif occurrence, in order to assign an accessibility score to each of them, and include in the motif enrichment tests only the occurrences above a certain accessibility threshold.

This potential follow-up however raises another important issue, which was also of our concern during the analysis of mRNA structure scores distributions: the applicability of mRNA structural information to specific experimental conditions. While the nucleotides composing a mature mRNA sequence stay the same, the folding of an mRNA into secondary and tertiary structures is unstable and prone to changes between different conformations. Several factors, from small changes in cellular conditions to the specific activity of RNA chaperones (Kang et al. 2013), can affect the structure of an mRNA species. The folding operated by RNA chaperones, which belong to the RBP class, might in turn expose new structured elements or hide binding sites that are the target of different RBPs or RPs. This situation makes it difficult to establish cause and effect, and it is possible that multiple mRNA conformation steps may be involved in the translational regulation process. All considered, it is important to recognize that *de novo* computational predictions of mRNA structures cannot be taken as fully reliable, although still useful to get clues on the general stability of certain regions of the mRNA sequence. However, the reliability of *in silico* predictions can be substantially improved by constraining folding algorithms with *in vitro*, or preferably *in vivo*, experimental data on mRNA structure. Recently, one such method allowing the *in vivo* detection of structured mRNA, termed structure-seq, was developed (Ding et al. 2014); we tried to include this information in our analyses in Chapter 4, as presumably less noisy than *in vitro* data from Li *et al.* (2012). Unfortunately, an incomplete description of the data processing steps, which were made available only in raw output format, prevented us from using this resource. In addition, the short reads from the control dataset where far from randomly distributed across the transcripts, which casted doubts on the quality of the data. Hopefully, more condition- and tissue-specific *in vivo* mRNA structural datasets will become available in the future, allowing more informative large-scale bioinformatic analyses on the role of mRNA structure in differential translational regulation.

On a final note, it is worth mentioning another regulatory level that we did not investigate in our analyses, but which appears promising for plant translational control: epitranscriptomics. The term encompasses several different post-transcriptional modifications of RNA molecules, such as pseudouridylation and the methylation of adenosine at the N6 position (m^6A); the latter is currently the most characterized modification of this kind in plants (Fray and Simpson 2015). Preliminary experimental evidence in various species indicates that epitranscriptomics might have a role in the regulation of mRNA interactions with

RBPs and RPs, possibly by influencing the folding of transcripts. The reversible nature of epitranscriptomic modifications also makes them a suitable medium to regulate condition-specific translational efficiency: the activity of enzymes responsible for the modifications could be regulated by specific cellular conditions, causing the introduction or removal of modifications in a controlled manner. This in turn would have an impact on the interaction of the modified mRNA with regulatory proteins and the ribosome. Genome-wide epitranscriptomics data are still missing for plants, but the situation will likely change in the near future, allowing bioinformatic analyses on differential translational regulation to integrate information on this regulatory level.

CONCLUDING REMARKS

The variety of regulatory aspects covered in this thesis arose from the close collaboration with plant biologists working on different research topics, showing the versatility of bioinformatic methods in answering independent biological questions. An important message from our work is that the collaboration between computational and experimental scientists should be intended as a bi-directional process: bioinformatics should not only be seen as a tool to help researchers interpret and integrate laboratory results, but also as a source of new testable hypotheses, which in turn require the help of experimentalists to be validated. Given the necessity of high-quality datasets and technical advances for specific computational studies, as we saw for our evolutionary and translational regulation analyses, bioinformatics will hopefully assume an increasingly leading role in the planning of relevant large-scale experiments, and in stimulating the development of novel experimental techniques.

References

Acevedo-Hernandez, G. J., P. Leon, and L. R. Herrera-Estrella. 2005. 'Sugar and ABA responsiveness of a minimal RBCS light-responsive unit is mediated by direct binding of ABI4', *Plant J*, 43: 506-19.

Alberts, B. 1998. 'The cell as a collection of protein machines: preparing the next generation of molecular biologists', *Cell*, 92: 291-4.

Allen, M. D., K. Yamasaki, M. Ohme-Takagi, M. Tateno, and M. Suzuki. 1998. 'A novel mode of DNA recognition by a beta-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA', *EMBO J*, 17: 5484-96.

Alonso, R., L. Onate-Sanchez, F. Weltmeier, A. Ehlert, I. Diaz, K. Dietrich, J. Vicente-Carbajosa, and W. Droge-Laser. 2009. 'A pivotal role of the basic leucine zipper transcription factor bZIP53 in the regulation of Arabidopsis seed maturation gene expression based on heterodimerization and protein complex formation', *Plant Cell*, 21: 1747-61.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. 'Basic local alignment search tool', *J Mol Biol*, 215: 403-10.

Andronis, C., S. Barak, S. M. Knowles, S. Sugano, and E. M. Tobin. 2008. 'The clock protein CCA1 and the bZIP transcription factor HY5 physically interact to regulate gene expression in Arabidopsis', *Mol Plant*, 1: 58-67.

Arenas-Huertero, F., A. Arroyo, L. Zhou, J. Sheen, and P. Leon. 2000. 'Analysis of Arabidopsis glucose insensitive mutants, gin5 and gin6, reveals a central role of the plant hormone ABA in the regulation of plant vegetative development by sugar', *Genes Dev*, 14: 2085-96.

Arroyo, A., F. Bossi, R. R. Finkelstein, and P. Leon. 2003. 'Three genes that affect sugar sensing (abscisic acid insensitive 4, abscisic acid insensitive 5, and constitutive triple response 1) are differentially regulated by glucose in Arabidopsis', *Plant Physiol*, 133: 231-42.

Baena-González, E. 2010. 'Energy signaling in the regulation of gene expression during stress', *Mol Plant*, 3: 300-13.

Baena-González, E., F. Rolland, J. M. Thevelein, and J. Sheen. 2007. 'A central integrator of transcription networks in plant stress and energy signalling', *Nature*, 448: 938-42.

Baena-González, E., and J. Sheen. 2008. 'Convergent energy and stress signaling', *Trends Plant Sci*, 13: 474-82.

Bailey, T. L., M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li, and W. S. Noble. 2009. 'MEME SUITE: tools for motif discovery and searching', *Nucleic Acids Res*, 37: W202-8.

Bailey-Serres, J., S. C. Lee, and E. Brinton. 2012. 'Waterproofing crops: effective flooding survival strategies', *Plant Physiol*, 160: 1698-709.

Bailey-Serres, J., R. Sorenson, and P. Juntawong. 2009. 'Getting the message across: cytoplasmic ribonucleoprotein complexes', *Trends Plant Sci*, 14: 443-53.

Bajguz, A. 2009. 'Brassinosteroid enhanced the level of abscisic acid in Chlorella vulgaris subjected to short-term heat stress', *J Plant Physiol*, 166: 882-6.

Baloglu, M. C., V. Eldem, M. Hajyzadeh, and T. Unver. 2014. 'Genome-wide analysis of the bZIP transcription factors in cucumber', *PLoS One*, 9: e96014.

Banks, J. A., T. Nishiyama, M. Hasebe, J. L. Bowman, M. Gribskov, C. dePamphilis, V. A. Albert, N. Aono, T. Aoyama, B. A. Ambrose, et al. 2011. 'The Selaginella genome identifies genetic changes associated with the evolution of vascular plants', *Science*, 332: 960-3.

Barakat, A., K. Szick-Miranda, I. F. Chang, R. Guyot, G. Blanc, R. Cooke, M. Delseny, and J. Bailey-Serres. 2001. 'The organization of cytoplasmic ribosomal protein genes in the Arabidopsis genome', *Plant Physiol*, 127: 398-415.

Basbouss-Serhal, I., L. Soubigou-Taconnat, C. Bailly, and J. Leymarie. 2015. 'Germination Potential of Dormant and Nondormant Arabidopsis Seeds Is Driven by Distinct Recruitment of Messenger RNAs to Polysomes', *Plant Physiol*, 168: 1049-65.

Benjamini, Y.; Hochberg, Y. 1995. 'Controlling the false discovery rate: a practical and powerful approach to multiple testing', *Journal of the Royal Statistical Society*, Series B (Methodological): 289-300.

Birol, I., A. Raymond, S. D. Jackman, S. Pleasance, R. Coope, G. A. Taylor, M. M. Yuen, C. I. Keeling, D. Brand, B. P. Vandervalk, et al. 2013. 'Assembling the 20 Gb white spruce (Picea glauca) genome from whole-genome shotgun sequencing data', *Bioinformatics*, 29: 1492-7.

Blasing, O. E., Y. Gibon, M. Gunther, M. Hohne, R. Morcuende, D. Osuna, O. Thimm, B. Usadel, W. R. Scheible, and M. Stitt. 2005. 'Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in Arabidopsis', *Plant Cell*, 17: 3257-81.

Boex-Fontvieille, E., M. Daventure, M. Jossier, M. Zivy, M. Hodges, and G. Tcherkez. 2013. 'Photosynthetic control of Arabidopsis leaf cytoplasmic translation initiation by protein phosphorylation', *PLoS One*, 8: e70692.

Bossi, F., E. Cordoba, P. Dupre, M. S. Mendoza, C. S. Roman, and P. Leon. 2009. 'The Arabidopsis ABA-INSENSITIVE (ABI) 4 factor acts as a central transcription activator of the expression of its own gene, and for the induction of

ABI5 and SBE2.2 genes during sugar signaling', *Plant J*, 59: 359-74.
Brady, S. M., S. F. Sarkar, D. Bonetta, and P. McCourt. 2003. 'The ABSCISIC ACID INSENSITIVE 3 (ABI3) gene is modulated by farnesylation and is involved in auxin signaling and lateral root development in Arabidopsis', *Plant J*, 34: 67-75.
Branco-Price, C., K. A. Kaiser, C. J. Jang, C. K. Larive, and J. Bailey-Serres. 2008. 'Selective mRNA translation coordinates energetic and metabolic adjustments to cellular oxygen deprivation and reoxygenation in Arabidopsis thaliana', *Plant J*, 56: 743-55.
Branco-Price, C., R. Kawaguchi, R. B. Ferreira, and J. Bailey-Serres. 2005. 'Genome-wide analysis of transcript abundance and translation in Arabidopsis seedlings subjected to oxygen deprivation', *Ann Bot*, 96: 647-60.
Brouwer, P., A. Brautigam, C. Kulahoglu, A. O. Tazelaar, S. Kurz, K. G. Nierop, A. van der Werf, A. P. Weber, and H. Schluepmann. 2014. 'Azolla domestication towards a biobased economy?', *New Phytol*, 202: 1069-82.
Bushart, T. J., A. E. Cannon, A. Ul Haque, P. San Miguel, K. Mostajeran, G. B. Clark, D. M. Porterfield, and S. J. Roux. 2013. 'RNA-seq analysis identifies potential modulators of gravity response in spores of Ceratopteris (Parkeriaceae): evidence for modulation by calcium pumps and apyrase activity', *Am J Bot*, 100: 161-74.
Byrne, M. E. 2009. 'A role for the ribosome in development', *Trends Plant Sci*, 14: 512-9.
Caldana, C., T. Degenkolbe, A. Cuadros-Inostroza, S. Klie, R. Sulpice, A. Leisse, D. Steinhauser, A. R. Fernie, L. Willmitzer, and M. A. Hannah. 2011. 'High-density kinetic analysis of the metabolomic and transcriptomic response of Arabidopsis to eight environmental conditions', *Plant J*, 67: 869-84.
Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden. 2009. 'BLAST+: architecture and applications', *BMC Bioinformatics*, 10: 421.
Camiolo, S., L. Farina, and A. Porceddu. 2012. 'The relation of codon bias to tissue-specific gene expression in Arabidopsis thaliana', *Genetics*, 192: 641-9.
Carroll, A. J., J. L. Heazlewood, J. Ito, and A. H. Millar. 2008. 'Analysis of the Arabidopsis cytosolic ribosome proteome provides detailed insights into its components and their post-translational modification', *Mol Cell Proteomics*, 7: 347-69.
Chamala, S., A. S. Chanderbali, J. P. Der, T. Lan, B. Walts, V. A. Albert, C. W. dePamphilis, J. Leebens-Mack, S. Rounsley, S. C. Schuster, R. A. Wing, N. Xiao, R. Moore, P. S. Soltis, D. E. Soltis, and W. B. Barbazuk. 2013. 'Assembly and validation of the genome of the nonmodel basal angiosperm Amborella', *Science*, 342: 1516-7.
Cheng, J., C. X. He, Z. W. Zhang, F. Xu, D. W. Zhang, X. Wang, S. Yuan, and H. H. Lin. 2011. 'Plastid signals confer Arabidopsis tolerance to water stress', *Z Naturforsch C*, 66: 47-54.
Clerkx, E. J., H. B. Vries, G. J. Ruys, S. P. Groot, and M. Koornneef. 2003. 'Characterization of green seed, an enhancer of abi3-1 in Arabidopsis that affects seed longevity', *Plant Physiol*, 132: 1077-84.
Contento, A. L., S. J. Kim, and D. C. Bassham. 2004. 'Transcriptome profiling of the response of Arabidopsis suspension culture cells to Suc starvation', *Plant Physiol*, 135: 2330-47.
Côrrea, L. G., D. M. Riano-Pachon, C. G. Schrago, R. V. dos Santos, B. Mueller-Roeber, and M. Vincentz. 2008. 'The role of bZIP transcription factors in green plant evolution: adaptive features emerging from four founder genes', *PLoS One*, 3: e2944.
Cuellar-Ortiz, S. M., M. De La Paz Arrieta-Montiel, J. Acosta-Gallegos, and A. A. Covarrubias. 2008. 'Relationship between carbohydrate partitioning and drought resistance in common bean', *Plant Cell Environ*, 31: 1399-409.
Cui, H., Y. Hao, and D. Kong. 2012. 'SCARECROW has a SHORT-ROOT-independent role in modulating the sugar response', *Plant Physiol*, 158: 1769-78.
Cutler, S. R., P. L. Rodriguez, R. R. Finkelstein, and S. R. Abrams. 2010. 'Abscisic acid: emergence of a core signaling network', *Annu Rev Plant Biol*, 61: 651-79.
Darriba, D., G. L. Taboada, R. Doallo, and D. Posada. 2011. 'ProtTest 3: fast selection of best-fit models of protein evolution', *Bioinformatics*, 27: 1164-5.
de Vries, S. J., A. D. van Dijk, M. Krzeminski, M. van Dijk, A. Thureau, V. Hsu, T. Wassenaar, and A. M. Bonvin. 2007. 'HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets', *Proteins*, 69: 726-33.
Degenhardt, R. F., and P. C. Bonham-Smith. 2008. 'Arabidopsis ribosomal proteins RPL23aA and RPL23aB are differentially targeted to the nucleolus and are disparately required for normal development', *Plant Physiol*, 147: 128-42.
Dekkers, B. J., J. A. Schuurmans, and S. C. Smeekens. 2004. 'Glucose delays seed germination in Arabidopsis thaliana', *Planta*, 218: 579-88.
Deppmann, C. D., A. Acharya, V. Rishi, B. Wobbes, S. Smeekens, E. J. Taparowsky, and C. Vinson. 2004. 'Dimerization specificity of all 67 B-ZIP motifs in Arabidopsis thaliana: a comparison to Homo sapiens B-ZIP motifs', *Nucleic Acids Res*, 32: 3435-45.
Deppmann, C. D., R. S. Alvania, and E. J. Taparowsky. 2006. 'Cross-species annotation of basic leucine zipper factor interactions: Insight into the evolution of closed interaction networks', *Mol Biol Evol*, 23: 1480-92.
Der, J. P., M. S. Barker, N. J. Wickett, C. W. dePamphilis, and P. G. Wolf. 2011. 'De novo characterization of the gametophyte transcriptome in bracken fern, Pteridium aquilinum', *BMC Genomics*, 12: 99.
Diboun, I., L. Wernisch, C. A. Orengo, and M. Koltzenburg. 2006. 'Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma', *BMC Genomics*, 7: 252.
Dietrich, K., F. Weltmeier, A. Ehlert, C. Weiste, M. Stahl, K. Harter, and W. Droge-Laser. 2011. 'Heterodimers of the Arabidopsis transcription factors bZIP1 and bZIP53 reprogram amino acid metabolism during low energy stress', *Plant Cell*, 23: 381-95.
Ding, Y., Y. Tang, C. K. Kwok, Y. Zhang, P. C. Bevilacqua, and S. M. Assmann. 2014. 'In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features', *Nature*, 505: 696-700.
Dominguez, C., R. Boelens, and A. M. Bonvin. 2003. 'HADDOCK: a protein-protein docking approach based on biochemical or biophysical information', *J Am Chem Soc*, 125: 1731-7.

Eddy, S. R. 1995. 'Multiple alignment using hidden Markov models', *Proc Int Conf Intell Syst Mol Biol*, 3: 114-20.
Edgar, R. C. 2004. 'MUSCLE: a multiple sequence alignment method with reduced time and space complexity', *BMC Bioinformatics*, 5: 113.
Ehlert, A., F. Weltmeier, X. Wang, C. S. Mayer, S. Smeekens, J. Vicente-Carbajosa, and W. Droge-Laser. 2006. 'Two-hybrid protein-protein interaction analysis in Arabidopsis protoplasts: establishment of a heterodimerization map of group C and group S bZIP transcription factors', *Plant J*, 46: 890-900.
Esposito, A. M., M. Mateyak, D. He, M. Lewis, A. N. Sasikumar, J. Hutton, P. R. Copeland, and T. G. Kinzy. 2010. 'Eukaryotic polyribosome profile analysis', *J Vis Exp*.
Eveland, A. L., and D. P. Jackson. 2012. 'Sugars, signalling, and plant development', *J Exp Bot*, 63: 3367-77.
Finkelstein, R.R. 1994. 'Mutations at two new Arabidopsis ABA response loci are similar to the abi3 mutations', *Plant J*, 5: 765-71.
Finkelstein, R. R., M. L. Wang, T. J. Lynch, S. Rao, and H. M. Goodman. 1998. 'The Arabidopsis abscisic acid response locus ABI4 encodes an APETALA 2 domain protein', *Plant Cell*, 10: 1043-54.
Finkelstein, R. R., and T. J. Lynch. 2000. 'The Arabidopsis abscisic acid response gene ABI5 encodes a basic leucine zipper transcription factor', *Plant Cell*, 12: 599-609.
Finkelstein, R., T. Lynch, W. Reeves, M. Petitfils, and M. Mostachetti. 2011. 'Accumulation of the transcription factor ABA-insensitive (ABI)4 is tightly regulated post-transcriptionally', *J Exp Bot*, 62: 3971-9.
Finn, R. D., J. Clements, and S. R. Eddy. 2011. 'HMMER web server: interactive sequence similarity searching', *Nucleic Acids Res*, 39: W29-37.
Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan, and J. Postlethwait. 1999. 'Preservation of duplicate genes by complementary, degenerative mutations', *Genetics*, 151: 1531-45.
Fray, R. G., and G. G. Simpson. 2015. 'The Arabidopsis epitranscriptome', *Curr Opin Plant Biol*, 27: 17-21.
Galland, M., R. Huguet, E. Arc, G. Cueff, D. Job, and L. Rajjou. 2014. 'Dynamic proteomics emphasizes the importance of selective mRNA translation and protein turnover during Arabidopsis seed germination', *Mol Cell Proteomics*, 13: 252-68.
Gautier, L., L. Cope, B. M. Bolstad, and R. A. Irizarry. 2004. 'affy--analysis of Affymetrix GeneChip data at the probe level', *Bioinformatics*, 20: 307-15.
Geigenberger, P. 2003. 'Regulation of sucrose to starch conversion in growing potato tubers', *J Exp Bot*, 54: 457-65.
Gerstberger, S., M. Hafner, and T. Tuschl. 2014. 'A census of human RNA-binding proteins', *Nat Rev Genet*, 15: 829-45.
Gibson, S. I., R. J. Laby, and D. Kim. 2001. 'The sugar-insensitive1 (sis1) mutant of Arabidopsis is allelic to ctr1', *Biochem Biophys Res Commun*, 280: 196-203.
Gibon, Y., H. Vigeolas, A. Tiessen, P. Geigenberger, and M. Stitt. 2002. 'Sensitive and high throughput metabolite assays for inorganic pyrophosphate, ADPGlc, nucleotide phosphates, and glycolytic intermediates based on a novel enzymic cycling system', *Plant J*, 30: 221-35.
Gibon, Y., B. Usadel, O. E. Blaesing, B. Kamlage, M. Hoehne, R. Trethewey, and M. Stitt. 2006. 'Integration of metabolite with transcript and enzyme activity profiling during diurnal cycles in Arabidopsis rosettes', *Genome Biol*, 7: R76.
Giraud, E., O. Van Aken, L. H. Ho, and J. Whelan. 2009. 'The transcription factor ABI4 is a regulator of mitochondrial retrograde expression of ALTERNATIVE OXIDASE1a', *Plant Physiol*, 150: 1286-96.
Giraudat, J., B. M. Hauge, C. Valon, J. Smalle, F. Parcy, and H. M. Goodman. 1992. 'Isolation of the Arabidopsis ABI3 gene by positional cloning', *Plant Cell*, 4: 1251-61.
Gong, Z., H. Lee, L. Xiong, A. Jagendorf, B. Stevenson, and J. K. Zhu. 2002. 'RNA helicase-like protein as an early regulator of transcription factors for plant chilling and freezing tolerance', *Proc Natl Acad Sci U S A*, 99: 11507-12.
Gong, Z., C. H. Dong, H. Lee, J. Zhu, L. Xiong, D. Gong, B. Stevenson, and J. K. Zhu. 2005. 'A DEAD box RNA helicase is essential for mRNA export and important for development and stress responses in Arabidopsis', *Plant Cell*, 17: 256-67.
Goodstein, D. M., S. Shu, R. Howson, R. Neupane, R. D. Hayes, J. Fazo, T. Mitros, W. Dirks, U. Hellsten, N. Putnam, and D. S. Rokhsar. 2012. 'Phytozome: a comparative platform for green plant genomics', *Nucleic Acids Res*, 40: D1178-86.
Gregorio, J., A. F. Hernandez-Bernal, E. Cordoba, and P. Leon. 2014. 'Characterization of evolutionarily conserved motifs involved in activity and regulation of the ABA-INSENSITIVE (ABI) 4 transcription factor', *Mol Plant*, 7: 422-36.
Guindon, S., and O. Gascuel. 2003. 'A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood', *Syst Biol*, 52: 696-704.
Hanada, K., T. Hase, T. Toyoda, K. Shinozaki, and M. Okamoto. 2011. 'Origin and evolution of genes related to ABA metabolism and its signaling pathways', *J Plant Res*, 124: 455-65.
Hanson, J., M. Hanssen, A. Wiese, M. M. Hendriks, and S. Smeekens. 2008. 'The sucrose regulated transcription factor bZIP11 affects amino acid metabolism by regulating the expression of ASPARAGINE SYNTHETASE1 and PROLINE DEHYDROGENASE2', *Plant J*, 53: 935-49.
Hartings, H., M. Lauria, N. Lazzaroni, R. Pirona, and M. Motto. 2011. 'The Zea mays mutants opaque-2 and opaque-7 disclose extensive changes in endosperm metabolism as revealed by protein, amino acid, and transcriptome-wide analyses', *BMC Genomics*, 12: 41.
Hayden, C.A., and R.A. Jorgensen. 2007. 'Identification of novel conserved peptide uORF homology groups in Arabidopsis and rice reveals ancient eukaryotic origin of select groups and preferential association with transcription factor-encoding genes', *BMC Biol*, 5: 32.
He, X., and J. Zhang. 2005. 'Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution', *Genetics*, 169: 1157-64.
Hori, K., F. Maruyama, T. Fujisawa, T. Togashi, N. Yamamoto, M. Seo, S. Sato, T. Yamada, H. Mori, N. Tajima, et al.

2014. 'Klebsormidium flaccidum genome reveals primary factors for plant terrestrial adaptation', *Nat Commun*, 5: 3978.
Hruz, T., O. Laule, G. Szabo, F. Wessendorp, S. Bleuler, L. Oertle, P. Widmayer, W. Gruissem, and P. Zimmermann. 2008. 'Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes', *Adv Bioinformatics*, 2008: 420747.
Hufton, A. L., and G. Panopoulou. 2009. 'Polyploidy and genome restructuring: a variety of outcomes', *Curr Opin Genet Dev*, 19: 600-6.
Huijser, C., A. Kortstee, J. Pego, P. Weisbeek, E. Wisman, and S. Smeekens. 2000. 'The Arabidopsis SUCROSE UNCOUPLED-6 gene is identical to ABSCISIC ACID INSENSITIVE-4: involvement of abscisic acid in sugar responses', *Plant J*, 23: 577-85.
Hulm, J. L.; McIntosh, K. B.; Bonham-Smith, P. C. . 2005. 'Variation in transcript abundance among the four members of the Arabidopsis thaliana RIBOSOMAL PROTEIN S15a gene family', *Plant Sci*, 169: 267-78.
Hummel, M., J. H. Cordewener, J. C. de Groot, S. Smeekens, A. H. America, and J. Hanson. 2012. 'Dynamic protein composition of Arabidopsis thaliana cytosolic ribosomes in response to sucrose feeding as revealed by label free MSE proteomics', *Proteomics*, 12: 1024-38.
Hummel, M., T. Dobrenel, J. J. Cordewener, M. Davanture, C. Meyer, S. J. Smeekens, J. Bailey-Serres, T. A. America, and J. Hanson. 2015. 'Proteomic LC-MS analysis of Arabidopsis cytosolic ribosomes: Identification of ribosomal protein paralogs and re-annotation of the ribosomal protein genes', *J Proteomics*, 128: 436-49.
Hwang, I., H. J. Jung, J. I. Park, T. J. Yang, and I. S. Nou. 2014. 'Transcriptome analysis of newly classified bZIP transcription factors of Brassica rapa in cold stress response', *Genomics*, 104: 194-202.
Iglesias-Fernandez, R., C. Barrero-Sicilia, N. Carrillo-Barral, L. Onate-Sanchez, and P. Carbonero. 2013. 'Arabidopsis thaliana bZIP44: a transcription factor affecting seed germination and expression of the mannanase-encoding gene AtMAN7', *Plant J*, 74: 767-80.
Ikemura, T. 1985. 'Codon usage and tRNA content in unicellular and multicellular organisms', *Mol Biol Evol*, 2: 13-34.
Irizarry, R. A., B. Hobbs, F. Collin, Y. D. Beazer-Barclay, K. J. Antonellis, U. Scherf, and T. P. Speed. 2003. 'Exploration, normalization, and summaries of high density oligonucleotide array probe level data', *Biostatistics*, 4: 249-64.
Jackson, R. J., C. U. Hellen, and T. V. Pestova. 2010. 'The mechanism of eukaryotic translation initiation and principles of its regulation', *Nat Rev Mol Cell Biol*, 11: 113-27.
Jakoby, M., B. Weisshaar, W. Droge-Laser, J. Vicente-Carbajosa, J. Tiedemann, T. Kroj, F. Parcy, and Z. I. P. Research Group b. 2002. 'bZIP transcription factors in Arabidopsis', *Trends Plant Sci*, 7: 106-11.
Jiao, Y., and E. M. Meyerowitz. 2010. 'Cell-type specific analysis of translating RNAs in developing flowers reveals new levels of control', *Mol Syst Biol*, 6: 419.
Jiao, Y., N. J. Wickett, S. Ayyampalayam, A. S. Chanderbali, L. Landherr, P. E. Ralph, L. P. Tomsho, Y. Hu, H. Liang, P. S. Soltis, et al 2011. 'Ancestral polyploidy in seed plants and angiosperms', *Nature*, 473: 97-100.
Jiao, Y., J. Leebens-Mack, S. Ayyampalayam, J. E. Bowers, M. R. McKain, J. McNeal, M. Rolf, D. R. Ruzicka, E. Wafula, N. J. Wickett, et al. 2012. 'A genome triplication associated with early diversification of the core eudicots', *Genome Biol*, 13: R3.
Jin, Z., W. Xu, and A. Liu. 2014. 'Genomic surveys and expression analysis of bZIP gene family in castor bean (Ricinus communis L.)', *Planta*, 239: 299-312.
Jindrich, K., and B. M. Degnan. 2016. 'The diversification of the basic leucine zipper family in eukaryotes correlates with the evolution of multicellularity', *BMC Evol Biol*, 16: 28.
Juntawong, P., and J. Bailey-Serres. 2012. 'Dynamic Light Regulation of Translation Status in Arabidopsis thaliana', *Front Plant Sci*, 3: 66.
Juntawong, P., T. Girke, J. Bazin, and J. Bailey-Serres. 2014. 'Translational dynamics revealed by genome-wide profiling of ribosome footprints in Arabidopsis', *Proc Natl Acad Sci U S A*, 111: E203-12.
Kakizaki, T., H. Matsumura, K. Nakayama, F. S. Che, R. Terauchi, and T. Inaba. 2009. 'Coordination of plastid protein import and nuclear gene expression by plastid-to-nucleus retrograde signaling', *Plant Physiol*, 151: 1339-53.
Kaminaka, H., C. Nake, P. Epple, J. Dittgen, K. Schutze, C. Chaban, B. F. Holt, T. Merkle, E. Schafer, K. Harter, and J. L. Dangl. 2006. 'bZIP10-LSD1 antagonism modulates basal defense and cell death in Arabidopsis following infection', *EMBO J*, 25: 4400-11.
Kanaya, S., Y. Yamada, M. Kinouchi, Y. Kudo, and T. Ikemura. 2001. 'Codon usage and tRNA genes in eukaryotes: correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis', *J Mol Evol*, 53: 290-8.
Kang, S. G., J. Price, P. C. Lin, J. C. Hong, and J. C. Jang. 2010. 'The arabidopsis bZIP1 transcription factor is involved in sugar signaling, protein networking, and DNA binding', *Mol Plant*, 3: 361-73.
Kang, H., S. J. Park, and K. J. Kwak. 2013. 'Plant RNA chaperones in stress response', *Trends Plant Sci*, 18: 100-6.
Kapp, L. D., and J. R. Lorsch. 2004. 'The molecular mechanics of eukaryotic translation', *Annu Rev Biochem*, 73: 657-704.
Katoh, K., K. Misawa, K. Kuma, and T. Miyata. 2002. 'MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform', *Nucleic Acids Res*, 30: 3059-66.
Katoh, K., and D. M. Standley. 2013. 'MAFFT multiple sequence alignment software version 7: improvements in performance and usability', *Mol Biol Evol*, 30: 772-80.
Kawaguchi, R., and J. Bailey-Serres. 2002. 'Regulation of translational initiation in plants', *Curr Opin Plant Biol*, 5: 460-5.
Kawaguchi, R., Williams, A. J., Bray, E. A. & Bailey-Serres, J. 2003. 'Water-deficit-induced translational control in Nicotiana tabacum.', *Plant and Cell Physiology*, 26: 221-29.
Kawaguchi, R., T. Girke, E. A. Bray, and J. Bailey-Serres. 2004. 'Differential mRNA translation contributes to gene regulation under non-stress and dehydration stress conditions in Arabidopsis thaliana', *Plant J*, 38: 823-39.
Kawaguchi, R., and J. Bailey-Serres. 2005. 'mRNA sequence features that contribute to translational regulation in

Arabidopsis', *Nucleic Acids Res*, 33: 955-65.
Kerchev, P. I., T. K. Pellny, P. D. Vivancos, G. Kiddle, P. Hedden, S. Driscoll, H. Vanacker, P. Verrier, R. D. Hancock, and C. H. Foyer. 2011. 'The transcription factor ABI4 Is required for the ascorbic acid-dependent regulation of growth and regulation of jasmonate-dependent defense signaling pathways in Arabidopsis', *Plant Cell*, 23: 3319-34.
Kertesz, M., Y. Wan, E. Mazor, J. L. Rinn, R. C. Nutter, H. Y. Chang, and E. Segal. 2010. 'Genome-wide measurement of RNA secondary structure in yeast', *Nature*, 467: 103-7.
Kindgren, P., L. Noren, D. Lopez Jde, J. Shaikhali, and A. Strand. 2012. 'Interplay between Heat Shock Protein 90 and HY5 controls PhANG expression in response to the GUN5 plastid signal', *Mol Plant*, 5: 901-13.
Kirchler, T., S. Briesemeister, M. Singer, K. Schutze, M. Keinath, O. Kohlbacher, J. Vicente-Carbajosa, M. Teige, K. Harter, and C. Chaban. 2010. 'The role of phosphorylatable serine residues in the DNA-binding domain of Arabidopsis bZIP transcription factors', *Eur J Cell Biol*, 89: 175-83.
Koornneef, M.; Reuling, G.; Karssen, C. M. 1984. 'The isolation and characterization of abscisic acid-insensitive mutants of Arabidopsis thaliana', *Physiologia Plantarum*, 61: 377-83.
Koornneef, M., C. J. Hanhart, H. W. Hilhorst, and C. M. Karssen. 1989. 'In Vivo Inhibition of Seed Development and Reserve Protein Accumulation in Recombinants of Abscisic Acid Biosynthesis and Responsiveness Mutants in Arabidopsis thaliana', *Plant Physiol*, 90: 463-9.
Koussevitzky, S., A. Nott, T. C. Mockler, F. Hong, G. Sachetto-Martins, M. Surpin, J. Lim, R. Mittler, and J. Chory. 2007. 'Signals from chloroplasts converge to regulate nuclear gene expression', *Science*, 316: 715-9.
Kozak, M. 2005. 'Regulation of translation via mRNA structure in prokaryotes and eukaryotes', *Gene*, 361: 13-37.
Laby, R. J., M. S. Kincaid, D. Kim, and S. I. Gibson. 2000. 'The Arabidopsis sugar-insensitive mutants sis4 and sis5 are defective in abscisic acid synthesis and response', *Plant J*, 23: 587-96.
Lamesch, P., K. Dreher, D. Swarbreck, R. Sasidharan, L. Reiser, and E. Huala. 2010. 'Using the Arabidopsis information resource (TAIR) to find information about Arabidopsis genes', *Curr Protoc Bioinformatics*, Chapter 1: Unit1 11.
Lang, D., B. Weiche, G. Timmerhaus, S. Richardt, D. M. Riano-Pachon, L. G. Côrrea, R. Reski, B. Mueller-Roeber, and S. A. Rensing. 2010. 'Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a timeline of loss, gain, expansion, and correlation with complexity', *Genome Biol Evol*, 2: 488-503.
Lara, P., L. Onate-Sanchez, Z. Abraham, C. Ferrandiz, I. Diaz, P. Carbonero, and J. Vicente-Carbajosa. 2003. 'Synergistic activation of seed storage protein gene expression in Arabidopsis by ABI3 and two bZIPs related to OPAQUE2', *J Biol Chem*, 278: 21003-11.
Lastdrager, J., J. Hanson, and S. Smeekens. 2014. 'Sugar signals and the control of plant growth and development', *J Exp Bot*, 65: 799-807.
Layat, E., J. Leymarie, H. El-Maarouf-Bouteau, J. Caius, N. Langlade, and C. Bailly. 2014. 'Translatome profiling in dormant and nondormant sunflower (Helianthus annuus) seeds highlights post-transcriptional regulation of germination', *New Phytol*, 204: 864-72.
Lemoine, R., S. La Camera, R. Atanassova, F. Dedaldechamp, T. Allario, N. Pourtau, J. L. Bonnemain, M. Laloi, P. Coutos-Thevenot, L. Maurousset, et al. 2013. 'Source-to-sink transport of sugar and regulation by environmental factors', *Front Plant Sci*, 4: 272.
Letunic, I., and P. Bork. 2011. 'Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy', *Nucleic Acids Res*, 39: W475-8.
Leung, J., M. Bouvier-Durand, P. C. Morris, D. Guerrier, F. Chefdor, and J. Giraudat. 1994. 'Arabidopsis ABA response gene ABI1: features of a calcium-modulated protein phosphatase', *Science*, 264: 1448-52.
Li, F., Q. Zheng, L. E. Vandivier, M. R. Willmann, Y. Chen, and B. D. Gregory. 2012. 'Regulatory impact of RNA secondary structure across the Arabidopsis transcriptome', *Plant Cell*, 24: 4346-59.
Li, P., J. J. Wind, X. Shi, H. Zhang, J. Hanson, S. C. Smeekens, and S. Teng. 2011. 'Fructose sensitivity is suppressed in Arabidopsis by the transcription factor ANAC089 lacking the membrane-bound domain', *Proc Natl Acad Sci U S A*, 108: 3436-41.
Liao, Y., H. F. Zou, W. Wei, Y. J. Hao, A. G. Tian, J. Huang, Y. F. Liu, J. S. Zhang, and S. Y. Chen. 2008. 'Soybean GmbZIP44, GmbZIP62 and GmbZIP78 genes function as negative regulator of ABA signaling and confer salt and freezing tolerance in transgenic Arabidopsis', *Planta*, 228: 225-40.
Licausi, F., M. Ohme-Takagi, and P. Perata. 2013. 'APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs', *New Phytol*, 199: 639-49.
Lin, S. Y., P. W. Chen, M. H. Chuang, P. Juntawong, J. Bailey-Serres, and G. Y. Jauh. 2014. 'Profiling of translatomes of in vivo-grown pollen tubes reveals genes with roles in micropylar guidance during pollination in Arabidopsis', *Plant Cell*, 26: 602-18.
Lindner, S. E., E. K. De Silva, J. L. Keck, and M. Llinas. 2010. 'Structural determinants of DNA binding by a P. falciparum ApiAP2 transcriptional regulator', *J Mol Biol*, 395: 558-67.
Liu, J., N. Chen, F. Chen, B. Cai, S. Dal Santo, G. B. Tornielli, M. Pezzotti, and Z. M. Cheng. 2014. 'Genome-wide analysis and expression profile of the bZIP transcription factor gene family in grapevine (Vitis vinifera)', *BMC Genomics*, 15: 281.
Liu, M. J., S. H. Wu, H. M. Chen, and S. H. Wu. 2012. 'Widespread translational control contributes to the regulation of Arabidopsis photomorphogenesis', *Mol Syst Biol*, 8: 566.
Liu, M. J., S. H. Wu, J. F. Wu, W. D. Lin, Y. C. Wu, T. Y. Tsai, H. L. Tsai, and S. H. Wu. 2013. 'Translational landscape of photomorphogenic Arabidopsis', *Plant Cell*, 25: 3699-710.
Liu, X., C. Y. Chen, K. C. Wang, M. Luo, R. Tai, L. Yuan, M. Zhao, S. Yang, G. Tian, Y. Cui, et al. 2013. 'PHYTOCHROME INTERACTING FACTOR3 associates with the histone deacetylase HDA15 in repression of chlorophyll biosynthesis and photosynthesis in etiolated Arabidopsis seedlings', *Plant Cell*, 25: 1258-73.
Liu, X., and Z. Chu. 2015. 'Genome-wide evolutionary characterization and analysis of bZIP transcription factors and their expression profiles in response to multiple abiotic stresses in Brachypodium distachyon', *BMC Genomics*, 16: 227.

Llorca, C. M., M. Potschin, and U. Zentgraf. 2014. 'bZIPs and WRKYs: two large transcription factor families executing two different functional strategies', *Front Plant Sci*, 5: 169.
Lorkovic, Z. J. 2009. 'Role of plant RNA-binding proteins in development, stress response and genome organization', *Trends Plant Sci*, 14: 229-36.
Ma, J., M. Hanssen, K. Lundgren, L. Hernandez, T. Delatte, A. Ehlert, C. M. Liu, H. Schluepmann, W. Droge-Laser, T. Moritz, et al. 2011. 'The sucrose-regulated Arabidopsis transcription factor bZIP11 reprograms metabolism and regulates trehalose metabolism', *New Phytol*, 191: 733-45.
Ma, X. M., and J. Blenis. 2009. 'Molecular mechanisms of mTOR-mediated translational control', *Nat Rev Mol Cell Biol*, 10: 307-18.
Maestri, E., N. Klueva, C. Perrotta, M. Gulli, H. T. Nguyen, and N. Marmiroli. 2002. 'Molecular genetics of heat tolerance and heat shock proteins in cereals', *Plant Mol Biol*, 48: 667-81.
Mair, A., L. Pedrotti, B. Wurzinger, D. Anrather, A. Simeunovic, C. Weiste, C. Valerio, K. Dietrich, T. Kirchler, T. Nagele, et al. 2015. 'SnRK1-triggered switch of bZIP63 dimerization mediates the low-energy response in plants', *Elife*, 4.
Maldonado-Bonilla, L. D. 2014. 'Composition and function of P bodies in Arabidopsis thaliana', *Front Plant Sci*, 5: 201.
Matiolli, C. C., J. P. Tomaz, G. T. Duarte, F. M. Prado, L. E. Del Bem, A. B. Silveira, L. Gauer, L. G. Côrrea, R. D. Drumond, A. J. Viana, et al. 2011. 'The Arabidopsis bZIP gene AtbZIP63 is a sensitive integrator of transient abscisic acid and glucose signals', *Plant Physiol*, 157: 692-705.
Matsuura, H., Y. Ishibashi, A. Shinmyo, S. Kanaya, and K. Kato. 2010. 'Genome-wide analyses of early translational responses to elevated temperature and high salinity in Arabidopsis thaliana', *Plant Cell Physiol*, 51: 448-62.
Matsuura, H., S. Takenami, Y. Kubo, K. Ueda, A. Ueda, M. Yamaguchi, K. Hirata, T. Demura, S. Kanaya, and K. Kato. 2013. 'A computational and experimental approach reveals that the 5'-proximal region of the 5'-UTR has a Cis-regulatory signature responsible for heat stress-regulated mRNA translation in Arabidopsis', *Plant Cell Physiol*, 54: 474-83.
McIntosh, K. B., and P. C. Bonham-Smith. 2005. 'The two ribosomal protein L23A genes are differentially transcribed in Arabidopsis thaliana', *Genome*, 48: 443-54.
Mizoi, J., K. Shinozaki, and K. Yamaguchi-Shinozaki. 2012. 'AP2/ERF family transcription factors in plant abiotic stress responses', *Biochim Biophys Acta*, 1819: 86-96.
Mochida, K., and K. Shinozaki. 2011. 'Advances in omics and bioinformatics tools for systems analyses of plant functions', *Plant Cell Physiol*, 52: 2017-38.
Monke, G., M. Seifert, J. Keilwagen, M. Mohr, I. Grosse, U. Hahnel, A. Junker, B. Weisshaar, U. Conrad, H. Baumlein, and L. Altschmied. 2012. 'Toward the identification and regulation of the Arabidopsis thaliana ABI3 regulon', *Nucleic Acids Res*, 40: 8240-54.
Mustroph, A., M. E. Zanetti, C. J. Jang, H. E. Holtan, P. P. Repetti, D. W. Galbraith, T. Girke, and J. Bailey-Serres. 2009. 'Profiling translatomes of discrete cell populations resolves altered cellular priorities during hypoxia in Arabidopsis', *Proc Natl Acad Sci U S A*, 106: 18843-8.
Nakabayashi, K., M. Okamoto, T. Koshiba, Y. Kamiya, and E. Nambara. 2005. 'Genome-wide profiling of stored mRNA in Arabidopsis thaliana seed germination: epigenetic and genetic regulation of transcription in seed', *Plant J*, 41: 697-709.
Nakamura, S., T. J. Lynch, and R. R. Finkelstein. 2001. 'Physical interactions between ABA response loci of Arabidopsis', *Plant J*, 26: 627-35.
Nakano, T., K. Suzuki, T. Fujimura, and H. Shinshi. 2006. 'Genome-wide analysis of the ERF gene family in Arabidopsis and rice', *Plant Physiol*, 140: 411-32.
Nakashima, K., Y. Fujita, K. Katsura, K. Maruyama, Y. Narusaka, M. Seki, K. Shinozaki, and K. Yamaguchi-Shinozaki. 2006. 'Transcriptional regulation of ABI3- and ABA-responsive genes including RD29B and RD29A in seeds, germinating embryos, and seedlings of Arabidopsis', *Plant Mol Biol*, 60: 51-68.
Nakashima, K., Y. Fujita, N. Kanamori, T. Katagiri, T. Umezawa, S. Kidokoro, K. Maruyama, T. Yoshida, K. Ishiyama, M. Kobayashi, et al. 2009. 'Three Arabidopsis SnRK2 protein kinases, SRK2D/SnRK2.2, SRK2E/SnRK2.6/OST1 and SRK2I/SnRK2.3, involved in ABA signaling are essential for the control of seed development and dormancy', *Plant Cell Physiol*, 50: 1345-63.
Neale, D. B., J. L. Wegrzyn, K. A. Stevens, A. V. Zimin, D. Puiu, M. W. Crepeau, C. Cardeno, M. Koriabine, A. E. Holtz-Morris, J. D. Liechty, et al. 2014. 'Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies', *Genome Biol*, 15: R59.
Nicolaï, M., M. A. Roncato, A. S. Canoy, D. Rouquie, X. Sarda, G. Freyssinet, and C. Robaglia. 2006. 'Large-scale analysis of mRNA translation states during sucrose starvation in arabidopsis cells identifies cell proliferation and chromatin structure as targets of translational control', *Plant Physiol*, 141: 663-73.
Nijhawan, A., M. Jain, A. K. Tyagi, and J. P. Khurana. 2008. 'Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice', *Plant Physiol*, 146: 333-50.
Niu, X., T. Helentjaris, and N. J. Bate. 2002. 'Maize ABI4 binds coupling element1 in abscisic acid and sugar response genes', *Plant Cell*, 14: 2565-75.
Notredame, C., D. G. Higgins, and J. Heringa. 2000. 'T-Coffee: A novel method for fast and accurate multiple sequence alignment', *J Mol Biol*, 302: 205-17.
Nystedt, B., N. R. Street, A. Wetterbom, A. Zuccolo, Y. C. Lin, D. G. Scofield, F. Vezzi, N. Delhomme, S. Giacomello, A. Alexeyenko, et al. 2013. 'The Norway spruce genome sequence and conifer genome evolution', *Nature*, 497: 579-84.
Oberstrass, F. C., S. D. Auweter, M. Erat, Y. Hargous, A. Henning, P. Wenter, L. Reymond, B. Amir-Ahmady, S. Pitsch, D. L. Black, and F. H. Allain. 2005. 'Structure of PTB bound to RNA: specific binding and implications for splicing regulation', *Science*, 309: 2054-7.

Ooms, J., K. M. Leon-Kloosterziel, D. Bartels, M. Koornneef, and C. M. Karssen. 1993. 'Acquisition of Desiccation Tolerance and Longevity in Seeds of Arabidopsis thaliana (A Comparative Study Using Abscisic Acid-Insensitive abi3 Mutants)', *Plant Physiol*, 102: 1185-91.
Osuna, D., B. Usadel, R. Morcuende, Y. Gibon, O. E. Blasing, M. Hohne, M. Gunter, B. Kamlage, R. Trethewey, W. R. Scheible, and M. Stitt. 2007. 'Temporal responses of transcripts, enzyme activities and metabolites after adding sucrose to carbon-deprived Arabidopsis seedlings', *Plant J*, 49: 463-91.
Oswald, O., T. Martin, P. J. Dominy, and I. A. Graham. 2001. 'Plastid redox state and sugars: interactive regulators of nuclear-encoded photosynthetic gene expression', *Proc Natl Acad Sci U S A*, 98: 2047-52.
Ouyang, S., W. Zhu, J. Hamilton, H. Lin, M. Campbell, K. Childs, F. Thibaud-Nissen, R. L. Malek, Y. Lee, L. Zheng, J. Orvis, B. Haas, J. Wortman, and C. R. Buell. 2007. 'The TIGR Rice Genome Annotation Resource: improvements and new features', *Nucleic Acids Res*, 35: D883-7.
Pal, S. K., M. Liput, M. Piques, H. Ishihara, T. Obata, M. C. Martins, R. Sulpice, J. T. van Dongen, A. R. Fernie, U. P. Yadav, J. E. Lunn, B. Usadel, and M. Stitt. 2013. 'Diurnal changes of polysome loading track sucrose content in the rosette of wild-type arabidopsis and the starchless pgm mutant', *Plant Physiol*, 162: 1246-65.
Park, S. Y., P. Fung, N. Nishimura, D. R. Jensen, H. Fujii, Y. Zhao, S. Lumba, J. Santiago, A. Rodrigues, T. F. Chow, et al. 2009. 'Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins', *Science*, 324: 1068-71.
Paul, M. J., and T. K. Pellny. 2003. 'Carbon metabolite feedback regulation of leaf photosynthesis and development', *J Exp Bot*, 54: 539-47.
Penfield, S., Y. Li, A. D. Gilday, S. Graham, and I. A. Graham. 2006. 'Arabidopsis ABA INSENSITIVE4 regulates lipid mobilization in the embryo and reveals repression of seed germination by the endosperm', *Plant Cell*, 18: 1887-99.
Perez, I., C. H. Lin, J. G. McAfee, and J. G. Patton. 1997. 'Mutation of PTB binding sites causes misregulation of alternative 3' splice site selection in vivo', *RNA*, 3: 764-78.
Pfannschmidt, T. 2010. 'Plastidial retrograde signalling--a true "plastid factor" or just metabolite signatures?', *Trends Plant Sci*, 15: 427-35.
Pinheiro, C., M. M. Chaves, and C. P. Ricardo. 2001. 'Alterations in carbon and nitrogen metabolism induced by water deficit in the stems and leaves of Lupinus albus L', *J Exp Bot*, 52: 1063-70.
Piques, M., W. X. Schulze, M. Hohne, B. Usadel, Y. Gibon, J. Rohwer, and M. Stitt. 2009. 'Ribosome and transcript copy numbers, polysome occupancy and enzyme dynamics in Arabidopsis', *Mol Syst Biol*, 5: 314.
Polge, C., and M. Thomas. 2007. 'SNF1/AMPK/SnRK1 kinases, global regulators at the heart of energy control?', *Trends Plant Sci*, 12: 20-8.
Pontier, D., C. Albrieux, J. Joyard, T. Lagrange, and M. A. Block. 2007. 'Knock-out of the magnesium protoporphyrin IX methyltransferase gene in Arabidopsis. Effects on chloroplast development and on chloroplast-to-nucleus signaling', *J Biol Chem*, 282: 2297-304.
Pourabed, E., F. Ghane Golmohamadi, P. Soleymani Monfared, S. M. Razavi, and Z. S. Shobbar. 2015. 'Basic leucine zipper family in barley: genome-wide characterization of members and expression analysis', *Mol Biotechnol*, 57: 12-26.
Price, J., T. C. Li, S. G. Kang, J. K. Na, and J. C. Jang. 2003. 'Mechanisms of glucose signaling during germination of Arabidopsis', *Plant Physiol*, 132: 1424-38.
Price, J., A. Laxmi, S. K. St Martin, and J. C. Jang. 2004. 'Global transcription profiling reveals multiple sugar signal transduction mechanisms in Arabidopsis', *Plant Cell*, 16: 2128-50.
Qu, X., J. D. Wen, L. Lancaster, H. F. Noller, C. Bustamante, and I. Tinoco, Jr. 2011. 'The ribosome uses two active mechanisms to unwind messenger RNA during translation', *Nature*, 475: 118-21.
Quesada, V., M. R. Ponce, and J. L. Micol. 2000. 'Genetic analysis of salt-tolerant mutants in Arabidopsis thaliana', *Genetics*, 154: 421-36.
Rahmani, F., M. Hummel, J. Schuurmans, A. Wiese-Klinkenberg, S. Smeekens, and J. Hanson. 2009. 'Sucrose control of translation mediated by an upstream open reading frame-encoded peptide', *Plant Physiol*, 150: 1356-67.
Rajasundaram, D., J. Selbig, S. Persson, and S. Klie. 2014. 'Co-ordination and divergence of cell-specific transcription and translation of genes in arabidopsis root cells', *Ann Bot*, 114: 1109-23.
Ramon, M., F. Rolland, J. M. Thevelein, P. Van Dijck, and B. Leyman. 2007. 'ABI4 mediates the effects of exogenous trehalose on Arabidopsis growth and starch breakdown', *Plant Mol Biol*, 63: 195-206.
Reeves, W. M., T. J. Lynch, R. Mobin, and R. R. Finkelstein. 2011. 'Direct targets of the transcription factors ABA-Insensitive(ABI)4 and ABI5 reveal synergistic action by ABI4 and several bZIP ABA response factors', *Plant Mol Biol*, 75: 347-63.
Ribeiro, D. M., W. L. Araujo, A. R. Fernie, J. H. Schippers, and B. Mueller-Roeber. 2012. 'Translatome and metabolome effects triggered by gibberellins during rosette growth in Arabidopsis', *J Exp Bot*, 63: 2769-86.
Riechmann, J. L., J. Heard, G. Martin, L. Reuber, C. Jiang, J. Keddie, L. Adam, O. Pineda, O. J. Ratcliffe, R. R. Samaha, et al. 2000. 'Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes', *Science*, 290: 2105-10.
Rodriguez, P. L., G. Benning, and E. Grill. 1998. 'ABI2, a second protein phosphatase 2C involved in abscisic acid signal transduction in Arabidopsis', *FEBS Lett*, 421: 185-90.
Rodriguez-Celma, J., I. C. Pan, W. Li, P. Lan, T. J. Buckhout, and W. Schmidt. 2013. 'The transcriptional response of Arabidopsis leaves to Fe deficiency', *Front Plant Sci*, 4: 276.
Ronquist, F., and J. P. Huelsenbeck. 2003. 'MrBayes 3: Bayesian phylogenetic inference under mixed models', *Bioinformatics*, 19: 1572-4.
Rook, F., N. Gerrits, A. Kortstee, M. van Kampen, M. Borrias, P. Weisbeek, and S. Smeekens. 1998. 'Sucrose-specific signalling represses translation of the Arabidopsis ATB2 bZIP transcription factor gene', *Plant J*, 15: 253-63.
Rook, F., F. Corke, R. Card, G. Munz, C. Smith, and M. W. Bevan. 2001. 'Impaired sucrose-induction mutants reveal

the modulation of sugar-induced starch biosynthetic gene expression by abscisic acid signalling', *Plant J*, 26: 421-33.
Rook, F., S. A. Hadingham, Y. Li, and M. W. Bevan. 2006. 'Sugar and ABA response pathways and the control of gene expression', *Plant Cell Environ*, 29: 426-34.
Roy, B., and A. G. von Arnim. 2013. 'Translational Regulation of Cytoplasmic mRNAs', *Arabidopsis Book*, 11: e0165.
Rugner, A., H. Frohnmeyer, C. Nake, F. Wellmer, S. Kircher, E. Schafer, and K. Harter. 2001. 'Isolation and characterization of four novel parsley proteins that interact with the transcriptional regulators CPRF1 and CPRF2', *Mol Genet Genomics*, 265: 964-76.
Sakuma, Y., Q. Liu, J. G. Dubouzet, H. Abe, K. Shinozaki, and K. Yamaguchi-Shinozaki. 2002. 'DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene expression', *Biochem Biophys Res Commun*, 290: 998-1009.
Satoh, R., Y. Fujita, K. Nakashima, K. Shinozaki, and K. Yamaguchi-Shinozaki. 2004. 'A novel subgroup of bZIP proteins functions as transcriptional activators in hypoosmolarity-responsive expression of the ProDH gene in Arabidopsis', *Plant Cell Physiol*, 45: 309-17.
Schepetilnikov, M., M. Dimitrova, E. Mancera-Martinez, A. Geldreich, M. Keller, and L. A. Ryabova. 2013. 'TOR and S6K1 promote translation reinitiation of uORF-containing mRNAs via phosphorylation of eIF3h', *EMBO J*, 32: 1087-102.
Schippers, J. H., Mueller-Roeber, B. 2010. 'Ribosomal composition and control of leaf development', *Plant Sci*, 179: 307-15.
Schluepmann, H., A. van Dijken, M. Aghdasi, B. Wobbes, M. Paul, and S. Smeekens. 2004. 'Trehalose mediated growth inhibition of Arabidopsis seedlings is due to trehalose-6-phosphate accumulation', *Plant Physiol*, 135: 879-90.
Schutze, K., K. Harter, and C. Chaban. 2008. 'Post-translational regulation of plant bZIP factors', *Trends Plant Sci*, 13: 247-55.
Shiu, S. H., M. C. Shih, and W. H. Li. 2005. 'Transcription factor families have much higher expansion rates in plants than in animals', *Plant Physiol*, 139: 18-26.
Shkolnik-Inbar, D., and D. Bar-Zvi. 2010. 'ABI4 mediates abscisic acid and cytokinin inhibition of lateral root formation by reducing polar auxin transport in Arabidopsis', *Plant Cell*, 22: 3560-73.
Shkolnik-Inbar, D., and D. Bar-Zvi. 2011. 'Expression of ABSCISIC ACID INSENSITIVE 4 (ABI4) in developing Arabidopsis seedlings', *Plant Signal Behav*, 6: 694-6.
Shkolnik-Inbar, D., G. Adler, and D. Bar-Zvi. 2013. 'ABI4 downregulates expression of the sodium transporter HKT1;1 in Arabidopsis roots and affects salt tolerance', *Plant J*, 73: 993-1005.
Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson, and D. G. Higgins. 2011. 'Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega', *Mol Syst Biol*, 7: 539.
Signora, L., I. De Smet, C. H. Foyer, and H. Zhang. 2001. 'ABA plays a central role in mediating the regulatory effects of nitrate on root branching in Arabidopsis', *Plant J*, 28: 655-62.
Singh, R., J. Valcarcel, and M. R. Green. 1995. 'Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins', *Science*, 268: 1173-6.
Smyth, G. K. 2004. 'Linear models and empirical bayes methods for assessing differential expression in microarray experiments', *Stat Appl Genet Mol Biol*, 3: Article3.
Smyth, G.K. 2005. 'Limma: linear models for microarray data.' in, *Bioinformatics and computational biology solutions using R and Bioconductor* (Springer: New York).
Soderman, E. M., I. M. Brocard, T. J. Lynch, and R. R. Finkelstein. 2000. 'Regulation and function of the Arabidopsis ABA-insensitive4 gene in seed and abscisic acid response signaling networks', *Plant Physiol*, 124: 1752-65.
Sorenson, R., and J. Bailey-Serres. 2014. 'Selective mRNA sequestration by OLIGOURIDYLATE-BINDING PROTEIN 1 contributes to translational control during hypoxia in Arabidopsis', *Proc Natl Acad Sci U S A*, 111: 2373-8.
Stamatakis, A., T. Ludwig, and H. Meier. 2005. 'RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees', *Bioinformatics*, 21: 456-63.
Stanke, M., and B. Morgenstern. 2005. 'AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints', *Nucleic Acids Res*, 33: W465-7.
Stöver, B. C., and K. F. Müller. 2010. 'TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses', *BMC Bioinformatics*, 11: 7.
Strathmann, A., M. Kuhlmann, T. Heinekamp, and W. Droge-Laser. 2001. 'BZI-1 specifically heterodimerises with the tobacco bZIP transcription factors BZI-2, BZI-3/TBZF and BZI-4, and is functionally involved in flower development', *Plant J*, 28: 397-408.
Strodtkotter, I., K. Padmasree, C. Dinakar, B. Speth, P. S. Niazi, J. Wojtera, I. Voss, P. T. Do, A. Nunes-Nesi, A. R. Fernie, et al. 2009. 'Induction of the AOX1D isoform of alternative oxidase in A. thaliana T-DNA insertion lines lacking isoform AOX1A is insufficient to optimize photosynthesis when treated with antimycin A', *Mol Plant*, 2: 284-97.
Subramanian, V. 1978. 'Polyribosome formation during early germination of bean embryonic axes', *Indian J Biochem Biophys*, 15: 235-7.
Sun, X., P. Feng, X. Xu, H. Guo, J. Ma, W. Chi, R. Lin, C. Lu, and L. Zhang. 2011. 'A chloroplast envelope-bound PHD transcription factor mediates chloroplast signals to the nucleus', *Nat Commun*, 2: 477.
Sun, X., Y. Li, H. Cai, X. Bai, W. Ji, X. Ding, and Y. Zhu. 2012. 'The Arabidopsis AtbZIP1 transcription factor is a positive regulator of plant tolerance to salt, osmotic and drought stresses', *J Plant Res*, 125: 429-38.
Sun, X., Q. Yang, and X. Xia. 2013. 'An improved implementation of effective number of codons (nc)', *Mol Biol Evol*, 30: 191-6.
Susek, R. E., F. M. Ausubel, and J. Chory. 1993. 'Signal transduction mutants of Arabidopsis uncouple nuclear CAB and RBCS gene expression from chloroplast development', *Cell*, 74: 787-99.

Talboys, P. J., H. M. Zhang, and J. Paul Knox. 2011. 'ABA signalling modulates the detection of the LM6 arabinan cell wall epitope at the surface of Arabidopsis thaliana seedling root apices', *New Phytol*, 190: 618-26.
Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar. 2011. 'MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods', *Mol Biol Evol*, 28: 2731-9.
Teng, S., S. Rognoni, L. Bentsink, and S. Smeekens. 2008. 'The Arabidopsis GSQ5/DOG1 Cvi allele is induced by the ABA-mediated sugar signalling pathway, and enhances sugar sensitivity by stimulating ABI4 expression', *Plant J*, 55: 372-81.
Thalor, S. K., T. Berberich, S. S. Lee, S. H. Yang, X. Zhu, R. Imai, Y. Takahashi, and T. Kusano. 2012. 'Deregulation of sucrose-controlled translation of a bZIP-type transcription factor results in sucrose accumulation in leaves', *PLoS One*, 7: e33111.
Thimm, O., O. Blasing, Y. Gibon, A. Nagel, S. Meyer, P. Kruger, J. Selbig, L. A. Müller, S. Y. Rhee, and M. Stitt. 2004. 'MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes', *Plant J*, 37: 914-39.
Thomas, J., S. G. Palusa, K. V. Prasad, G. S. Ali, G. K. Surabhi, A. Ben-Hur, S. E. Abdel-Ghany, and A. S. Reddy. 2012. 'Identification of an intronic splicing regulatory element involved in auto-regulation of alternative splicing of SCL33 pre-mRNA', *Plant J*, 72: 935-46.
To, J. P.; Reiter, W. D.; & Gibson, S. I. 2003. 'Chloroplast biogenesis by Arabidopsis seedlings is impaired in the presence of exogenous glucose', *Physiologia Plantarum*, 118: 456-63.
Tomé, F., T. Nagele, M. Adamo, A. Garg, C. Marco-Llorca, E. Nukarinen, L. Pedrotti, A. Peviani, A. Simeunovic, A. Tatkiewicz, M. Tomar, and M. Gamm. 2014. 'The low energy signaling network', *Front Plant Sci*, 5: 353.
Ton, J., and B. Mauch-Mani. 2004. 'Beta-amino-butyric acid-induced resistance against necrotrophic pathogens is based on ABA-dependent priming for callose', *Plant J*, 38: 119-30.
Turkina, M. V., H. Klang Arstrand, and A. V. Vener. 2011. 'Differential phosphorylation of ribosomal proteins in Arabidopsis thaliana plants during day and night', *PLoS One*, 6: e29307.
Ueda, K., H. Matsuura, M. Yamaguchi, T. Demura, and K. Kato. 2012. 'Genome-wide analyses of changes in translation state caused by elevated temperature in Oryza sativa', *Plant Cell Physiol*, 53: 1481-91.
Usadel, B., O. E. Blasing, Y. Gibon, K. Retzlaff, M. Hohne, M. Gunther, and M. Stitt. 2008. 'Global transcript levels respond to small changes of the carbon status during progressive exhaustion of carbohydrates in Arabidopsis rosettes', *Plant Physiol*, 146: 1834-61.
Valleriani, A., G. Zhang, A. Nagar, Z. Ignatova, and R. Lipowsky. 2011. 'Length-dependent translation of messenger RNA by ribosomes', *Phys Rev E Stat Nonlin Soft Matter Phys*, 83: 042903.
Van de Peer, Y., J. A. Fawcett, S. Proost, L. Sterck, and K. Vandepoele. 2009. 'The flowering world: a tale of duplications', *Trends Plant Sci*, 14: 680-8.
Van de Peer, Y. 2011. 'A mystery unveiled', *Genome Biol*, 12: 113.
Van Oosten, J. J., A. Gerbaud, C. Huijser, P. P. Dijkwel, N. H. Chua, and S. C. Smeekens. 1997. 'An Arabidopsis mutant showing reduced feedback inhibition of photosynthesis', *Plant J*, 12: 1011-20.
Vandivier, L., F. Li, Q. Zheng, M. Willmann, Y. Chen, and B. Gregory. 2013. 'Arabidopsis mRNA secondary structure correlates with protein function and domains', *Plant Signal Behav*, 8: e24301.
Vaughn, J. N., S. R. Ellingson, F. Mignone, and Av Arnim. 2012. 'Known and novel post-transcriptional regulatory sequences are conserved across plant families', *RNA*, 18: 368-84.
Veerabagu, M., T. Kirchler, K. Elgass, B. Stadelhofer, M. Stahl, K. Harter, V. Mira-Rodado, and C. Chaban. 2014. 'The interaction of the Arabidopsis response regulator ARR18 with bZIP63 mediates the regulation of PROLINE DEHYDROGENASE expression', *Mol Plant*, 7: 1560-77.
Velez-Bermudez, I. C., and W. Schmidt. 2014. 'The conundrum of discordant protein and mRNA expression. Are plants special?', *Front Plant Sci*, 5: 619.
von Arnim, A. G., Q. Jia, and J. N. Vaughn. 2014. 'Regulation of plant translation by upstream open reading frames', *Plant Sci*, 214: 1-12.
Wang, J., J. Zhou, B. Zhang, J. Vanitha, S. Ramachandran, and S. Y. Jiang. 2011. 'Genome-wide expansion and expression divergence of the basic leucine zipper transcription factors in higher plants with an emphasis on sorghum', *J Integr Plant Biol*, 53: 212-31.
Wang, J., P. Lan, H. Gao, L. Zheng, W. Li, and W. Schmidt. 2013. 'Expression changes of ribosomal proteins in phosphate- and iron-deficient Arabidopsis roots predict stress-specific alterations in ribosome composition', *BMC Genomics*, 14: 783.
Waterhouse, A. M., J. B. Procter, D. M. Martin, M. Clamp, and G. J. Barton. 2009. 'Jalview Version 2--a multiple sequence alignment editor and analysis workbench', *Bioinformatics*, 25: 1189-91.
Wei, K., J. Chen, Y. Wang, Y. Chen, S. Chen, Y. Lin, S. Pan, X. Zhong, and D. Xie. 2012. 'Genome-wide analysis of bZIP-encoding genes in maize', *DNA Res*, 19: 463-76.
Weijers, D., M. Franke-van Dijk, R. J. Vencken, A. Quint, P. Hooykaas, and R. Offringa. 2001. 'An Arabidopsis Minute-like phenotype caused by a semi-dominant mutation in a RIBOSOMAL PROTEIN S5 gene', *Development*, 128: 4289-99.
Weltmeier, F., A. Ehlert, C. S. Mayer, K. Dietrich, X. Wang, K. Schutze, R. Alonso, K. Harter, J. Vicente-Carbajosa, and W. Droge-Laser. 2006. 'Combinatorial control of Arabidopsis proline dehydrogenase transcription by specific heterodimerisation of bZIP transcription factors', *EMBO J*, 25: 3133-43.
Weltmeier, F., F. Rahmani, A. Ehlert, K. Dietrich, K. Schutze, X. Wang, C. Chaban, J. Hanson, M. Teige, K. Harter,et al. 2009. 'Expression patterns within the Arabidopsis C/S1 bZIP transcription factor network: availability of heterodimerization partners controls gene expression during stress response and development', *Plant Mol Biol*, 69: 107-19.
Wheeler, T. J., and S. R. Eddy. 2013. 'nhmmer: DNA homology search with profile HMMs', *Bioinformatics*, 29: 2487-

9.
Wiese, A., N. Elzinga, B. Wobbes, and S. Smeekens. 2004. 'A conserved upstream open reading frame mediates sucrose-induced repression of translation', *Plant Cell*, 16: 1717-29.
Williams, M. E., and I. M. Sussex. 1995. 'Developmental regulation of ribosomal protein L16 genes in Arabidopsis thaliana', *Plant J*, 8: 65-76.
Wolf, P. G., E. B. Sessa, D. B. Marchant, F. W. Li, C. J. Rothfels, E. M. Sigel, M. A. Gitzendanner, C. J. Visger, J. A. Banks, D. E. Soltis, et al. 2015. 'An Exploration into Fern Genome Space', *Genome Biol Evol*, 7: 2533-44.
Wright, S. I., C. B. Yau, M. Looseley, and B. C. Meyers. 2004. 'Effects of gene expression on molecular evolution in Arabidopsis thaliana and Arabidopsis lyrata', *Mol Biol Evol*, 21: 1719-26.
Xia, X., V. MacKay, X. Yao, J. Wu, F. Miura, T. Ito, and D. R. Morris. 2011. 'Translation initiation: a regulatory role for poly(A) tracts in front of the AUG codon in Saccharomyces cerevisiae', *Genetics*, 189: 469-78.
Xu, X., C. T. Hotta, A. N. Dodd, J. Love, R. Sharrock, Y. W. Lee, Q. Xie, C. H. Johnson, and A. A. Webb. 2007. 'Distinct light and clock modulation of cytosolic free Ca2+ oscillations and rhythmic CHLOROPHYLL A/B BINDING PROTEIN2 promoter activity in Arabidopsis', *Plant Cell*, 19: 3474-90.
Xue, S., and M. Barna. 2012. 'Specialized ribosomes: a new frontier in gene regulation and organismal biology', *Nat Rev Mol Cell Biol*, 13: 355-69.
Yamasaki, S., H. Matsuura, T. Demura, and K. Kato. 2015. 'Changes in Polysome Association of mRNA Throughout Growth and Development in Arabidopsis thaliana', *Plant Cell Physiol*, 56: 2169-80.
Yang, Y., X. Yu, L. Song, and C. An. 2011. 'ABI4 activates DGAT1 expression in Arabidopsis seedlings during nitrogen deficiency', *Plant Physiol*, 156: 873-83.
Yanguez, E., A. B. Castro-Sanz, N. Fernandez-Bautista, J. C. Oliveros, and M. M. Castellano. 2013. 'Analysis of genome-wide changes in the translatome of Arabidopsis seedlings subjected to heat stress', *PLoS One*, 8: e71425.
Yeakley, J. M., J. P. Morfin, M. G. Rosenfeld, and X. D. Fu. 1996. 'A complex of nuclear proteins mediates SR protein binding to a purine-rich splicing enhancer', *Proc Natl Acad Sci U S A*, 93: 7582-7.
Zimmer, A. D., D. Lang, K. Buchta, S. Rombauts, T. Nishiyama, M. Hasebe, Y. Van de Peer, S. A. Rensing, and R. Reski. 2013. 'Reannotation and extended community resources for the genome of the non-seed plant Physcomitrella patens provide insights into the evolution of plant gene structures and functions', *BMC Genomics*, 14: 498.

Samenvatting

Omdat planten sessiel zijn, moeten zij zich continue aanpassen aan abiotische en biotische stressfactoren uit hun directe omgeving. De productie en opslag van energie in planten wordt sterk negatief beïnvloed door nadelige omstandigheden zoals droogte, hitte, kou, overstromingen, tekorten of overvloeden van bepaalde nutriënten, en plagen en pathogenen. Aangezien planten niet weg kunnen lopen van deze stressfactoren, hebben zij in plaats daarvan complexe regulatienetwerken geëvolueerd waarmee zij hun interne energiebalans nauwkeurig kunnen controleren, ook in omgevingen waarin energie schaars is. Onderzoek naar deze regulatienetwerken leidt tot nieuwe inzichten in de fundamentele plantenbiologie, en is daarnaast belangrijk voor de ontwikkeling en verbetering van landbouwgewassen, en daarmee voor het garanderen van voedselproductie voor de groeiende wereldbevolking.
Bioinformatica speelt een steeds grotere rol in plantbiologisch onderzoek. Deze ontwikkeling hangt direct samen met het beschikbaar komen van steeds meer publiek toegankelijke online databanken. De publicatie van de volledige genomen van steeds meer planten heeft er in het bijzonder voor gezorgd dat de evolutionaire geschiedenis van genfamilies steeds beter onderzocht kan worden, en verschaft onderzoekers steeds meer mogelijkheden om de vroege molecularie evolutie van planten te bestuderen. In dit proefschrift worden zowel transcriptionele als translationele regulatiesystemen voor de energie-homeostase in planten onderzocht. Hiervoor gebruiken wij verschillende bioinformatische methodes.

In hoofdstuk 2 bestuderen we de evolutie en conservatie van de transcriptiefactor ABI4 in verscheidene plantensoorten. ABI4 is goed gekarakteriseerd in de modelplant A. thaliana, en is betrokken bij zaadontwikkeling en de repressie van ontkieming. Door homologe sequenties van verschillende plantensoorten te verzamelen tonen wij aan dat ABI4 in de meeste genomen slechts in enkelvoud aanwezig is, wat een sterk geconserveerde functie suggereert. Onze alignments laten ook een sterk geconserveerd “ABI4” motief zien, waarvan de rol is onderzocht in een vervolgpublicatie. Kortom, we verschaffen hier een grondige omschrijving van de ABI4 familie en hebben daarbij een motief gevonden dat potentieel een belangrijke rol speelt voor regulatie. Beide resultaten bieden een belangrijke basis voor vervolgonderzoek naar ABI4 in zowel modelplanten als plantensoorten in het algemeen.

In hoofdstuk 3 passen we opnieuw fylogenetische en comparatieve genomics toe, dit keer om de C en S1 groepen van bZIP transcriptiefactoren te onderzoeken. Hoewel bekend is dat de eiwitten uit beide groepen hetereodimeriseren met elkaar in A. thaliana, waarbij

het zogeheten “C/S1 bZIP transcriptiefactor-netwerk” wordt gevormd, is over de evolutie en conservatie van dit regulatienetwerk in andere soorten vooralsnodig niets bekend. Onze resultaten laten zien dat vroeger C en S bZIPs, waar later S1 bZIP uit is ontstaan, waarschijnlijk op hetzelfde moment zijn ontstaan door duplicatie van een voorouderlijk bZIP gen uit de algenfamilie. Gebaseerd op deze ontdekking hypothetiseren wij vervolgens een plausibel evoultionair scenario voor de afkomst van latere C/S1 bZIP heterodimeren, wat richting kan geven aan vervolgexperimenten. Onze vondst laat ook zien dat zowel C als S bZIPs een rol hebben gehad in de adaptatie van planten aan terrestrische omgevingen; wat vooral voor S bZIPs uitermate verrassend is, omdat werd aangenomen dat deze een meer recente bedektzadigen-“ontdekking” waren. Tussen de nieuw ontdekte sequenties in niet-bedektzadigen identificeren wij vervolgens mogelijke S1 bZIP orthologen in naaktzadigen, door sequenties te selecteren met een geconserveerd 5’uORF. Van deze genen wordt experimenteel aangetoond dat zij een rol spelen in de repressie van translatie door sucrose (SIRT), wat een karakteristieke S1 bZIP respons is. Deze resultaten suggereren dat SIRT waarschijnlijk geconserveerd is in alle zaadplanten. Gegeven enkele verschillen in de sequentie van de 5’uORFs van bedektzadigen en naaktzadigen, zou onze vinding ook licht kunnen schijnen op enkele, nog slecht begrepen, mechanismen van SIRT.
Tot slot geven we een gedetailleerde beschrijving van de orthologische relatie van C en S1 bZIP groepen in bedektzadigen. Naast het combineren van eerdere functionele observaties met de fylogenetische geschiedenis van elke groep, levert ons werk een referentiekader voor toekomstig onderzoek naar bZIP in soorten die geen model-organisme zijn.

In hoofdstuk 4 presenteren we een geïntegreerde bioinformatische analyse van de sequenties die verantwoordelijk zijn voor de regulatie van translatie van specifieke mRNA subsets in verschillende groeiomstandigheden, namelijk tijdens behandeling met sucrose en ontkieming. Onze resultaten laten zien dat zowel algemene eigenschappen van het transcript, zoals lengte en GC content, als specifiekere eigenschappen, zoals motieven en gestructureerde elementen in het mRNA, bijdragen aan de geobserveerde translationele selectie. In plaats van gebruik te maken van een enkele of een klein aantal verrijkte signaalsequenties, lijkt regulatie afhankelijk te zijn van een combinatie van verschillende eigenschappen, welke op hun beurt weer erg specifiek zijn voor elke groep van door translatie gereguleerde transcripten. We stellen voor dat het ontbrekene van verrijkte sequenties het resultaat zou kunnen zijn van beperkingen in de data van secundaire mRNA structuren, en bediscussiëren alternatieve benaderingen om meer informatieve resultaten te krijgen.

Ons onderzoek heeft enerzijds de evolutie van bekende energie-afhankelijke transcriptie regulatoren verhelderd, en anderzijds belangrijke aanwijzingen gegeven voor het minder bekende gebied van regulatie van translatie, wat een weg opent voor meer onderzoek in dit vakgebied. Het gepresenteerde onderzoek is ook een goed voorbeeld van het toenemende belang van computationele analyses in het genereren van nieuwe hypotheses voor experimentele studies, die bioinformatica langzaam maar zeker een leidende rol laten spelen in plantenonderzoek.

Curriculum Vitae

Alessia Peviani was born on April 23rd, 1987 in Reggio Emilia, Italy. In 2006, she enrolled in the Biotechnology bachelor program of the University of Modena and Reggio Emilia. Here, she performed her project internship in Immunology under the supervision of Dr. Marcello Pinti and PhD candidate Linda Bertoncelli. In September 2009, she graduated *cum laude* with the thesis: "*Differentiation, activation, and gag-specific response of T lymphocytes in HIV+ patients during CD4-guided structured treatment interruption*".

Following a genuine interest for Bioinformatics, triggered by one of her bachelor courses, in October of the same year Alessia enrolled in the English-taught International Master in Bioinformatics offered by the University of Bologna, Italy. In the second year she was awarded an Erasmus fellowship, which allowed her to undertake a 6-month project at the Stockholm Bioinformatics Center, SciLifeLab, Solna, Sweden. Here, under the supervision of Prof. Erik Sonnhammer and PhD candidate Kristoffer Forslund, and with the external review of Dr. Pier Luigi Martelli from the Bologna Biocomputing Group, she completed her master thesis on the "*Integration of proteinprotein interaction networks and gene expression data in cancer outcome prediction*". Alessia received her M.Sc. in Bioinformatics *cum laude* in July 2011.

The stimulating international experience of the master project, together with the desire to develop her computational skills, motivated Alessia to pursue a PhD in Bioinformatics abroad. In December 2011, she joined the Theoretical Biology and Bioinformatics group at Utrecht University under the supervision of Prof. Berend Snel. Here, she worked as a fellow of the Marie Curie initial training network "MEtabolic Reprogramming by Induction of Transcription" (MERIT), an international project aimed at unraveling energy signaling in plants. The project gave her the opportunity to collaborate with young international plant scientists, and to acquire experience with different bioinformatic methods. In the Netherlands, Alessia worked in close collaboration with members of the Molecular Plant Physiology group at Utrecht University, and of the Seed Lab at Wageningen University. Prof. Sjef Smeekens and Dr. Johannes Hanson, from the Molecular Plant Physiology group, supervised her together with Prof. Berend Snel during her PhD.

The results of Alessia's PhD research are presented in this thesis.

List of publications

Bai B, **Peviani A**, van der Horst S, Gamm M, Snel B, Hanson J, Bentsink L. Extensive translational regulation during seed germination revealed by translational profiling. *Submitted*

Peviani A, Lastdrager J, Hanson J, Snel B. The phylogeny of C/S1 bZIP transcription factors reveals a shared algal ancestry and the pre-angiosperm translational regulation of S1 transcripts. *Under review in* Scientific Reports

Gamm M, **Peviani A**, Honsel A, Snel B, Smeekens SC, Hanson J. Increased sucrose levels mediate selective mRNA translation in Arabidopsis. BMC Plant Biology (2014) 14: 306

Tomé F, Nägele T, Adamo M, Garg A, Marco-llorca C, Nukarinen E, Pedrotti L, **Peviani A**, Simeunovic A, Tatkiewicz A, Tomar M, Gamm M. The low energy signaling network. Frontiers in Plant Science (2014) 5: 353

Wind JJ, **Peviani A**, Snel B, Hanson J, Smeekens SC. ABI4: versatile activator and repressor. Trends in Plant Science (2013) 18: 125–132

Acknowledgements

Everything comes to an end.. And so does a PhD! These four and a half years have been an intensive formative experience, which taught me about myself at least as much as about science. During this time I've met many smart, kind, talented, extraordinary people, who broadened my thinking and enriched my life experience. After spending so much energy on the previous 100+ pages, I would like to fill a few more with some well-deserved acknowledgements (and a bit of funny things).

First of all, I would like to thank my supervisor and promotor Berend Snel. Berend, you have been patient and encouraging at times where it wasn't easy to do either. Being a good supervisor is hard, but you gave me the guidance I needed to take the right decisions for my project. Some students can only catch their supervisor in the office late at night (in their dreams :P), but that wasn't the case with you. You were always available to give valuable feedback, so spot-on in fact that a few times I wished I had recorded it to keep track of the details! You are a truly inspired and inspiring bioinformatic scientist, and I admire your passion. If I could change something from the past 4 years, that would be to have started out more relaxed and outspoken at our meetings (I really don't know what I was worrying about).

Next, I would like to thank my second promotor Sjef Smeekens and my copromotor Johannes Hanson, who have been my guides in the plant biology side of my project. Sjef, your manners and diplomacy have the feel of another epoch. I appreciated the time you spent discussing with me, and your detailed feedback on presentations. Johannes, I've been sitting in your office many times to make plans and discuss results; your friendly and positive attitude made it easier for me to believe in my project. I also very much enjoyed our informal chats :)

This thesis wouldn't have come into existence without the help of several collaborators. Julia, I remember when you appeared at the door to ask me about phylogenetic trees; as a matter of fact, I've learned most of what I know about them while working with you on the ABI4 paper. You are an enthusiastic plant scientist, I wish you the best for your career! Magdalena, you got me involved in translational regulation, which ended up becoming a project on its own. Besides a smart and determined researcher, I will remember you as my first flatmate in this country; now indeed I have these oddly mixed memories of scientific talking and cake baking. Jeroen, you performed the tests for the bZIP chapter; you too have been an important collaborator, as well as a familiar presence at home. I wish you both a

bright and happy future together in Delft. Leonie and Bing, after a casual discussion over lunch at the annual EPS meeting, you enthusiastically took up my proposal for bioinformatic analyses of the seed translatome; thank you for letting me and my student Sjors contribute. Sjors, you did a great job starting from where I left, I hope you'll get a stimulating PhD position to suit your interests! Marcel, thank you for your help with advanced motif analyses, I just wish they had ended up in a paper! Filipa, Monika, Lorenzo, Mattia, Carles, Anna, Abhi, Ella, Andrea, Thomas, and Magdalena again, besides pulling out a review, we had a great time at MERIT meetings and courses! (together with "associates" Anne and Andrea Mair). Filipa, we have been keeping in touch since we shared that room at the meeting in Madrid, which is an amazingly long time! It was always a pleasure to come visit you in Gent, and I have some really nice memories with you. Be sure that I'll come again to Köln :) Monika, you didn't have it easy with your project, I wish you the best of luck with finishing your PhD!

Doing research in the Snel group gave me the chance to work side by side with many smart and talented people. Adrian and Gabino, you left soon after I started, but I've learned a lot from you. Michael, you kept on saying "You'll be fine", which got me seriously irritated (how would you know?!?), but as it turned out you were right :P I wish you a brilliant academic career! Lidija, I was struck many times by your insightful, clear thinking, and sometimes I wish I could be as sharp and confident. I hope the best for you too! Like, you have an incredible personality, we should have had some more chats! Leny, you are such a nice person, perhaps you should be louder about it :) You became a mother last year, but even with this big extra duty I'm sure you will make it. Eelco and Jolien, being busy between two labs didn't help, but I enjoyed having conversations with you whenever you had the time. Bas, you are the most friendly and outgoing group leader I've ever met! Bruno, Diego, Felipe, Juline, you brought life directly from Brazil into my quiet weekends (sometimes even too much), and I cannot thank you enough for that! Bruno, you are something. A bit hyperactive at times, but sure you know how to have fun with friends! I had a good time with you and Sai, despite most of it consisting in the two of you making fun of my beer pouring skills (or, more precisely, lack thereof). You know I consider you my honorary second Evil paranymph ;) It's sad you won't be here for my defense, but also sort of relieving; Lidija (my Good paranymph) would have had a hard time balancing out your and Sai's cumulative bad influence at the party! I'm looking forward to have you back on this side of the ocean, I'm sure you'll find the chance :) Felipe, you are a cool guy! I wish you best of luck with your PhD, and I'd be really happy to come visit in Brazil (since you are unlikely to come back to this rainy country.. My sympathies!). Diego, you thought you didn't have any hobby, until I helped you realize how much you are into languages compared to the average person :) You are smart, you'll get to the end of this thing! Juline, you gave me plenty of advice on some delicate matters :P Although things solved unexpectedly for me, I appreciated the support (also, I'm still trying to grow that sort of tumor that fell off your cactus). Ashmita and Bastiaan, I didn't spend as much time with you as with other Bas' students, but those times were fun! (I'd still like to kick someone's ass at lasergame though). Rianne, Esther, Ethel, Andrew, Sjors, Yee Ying, it has been a pleasure to meet such mature and promising young minds.

When I started at the TBB, I was impressed with the informal environment and the consideration for everyone's opinion, no matter the role and level of experience. Paulien, Can, Rutger, Rob, Kirsten, I certainly appreciated your humor and open-mindedness when chatting with you in the multipurpose room. Jan-Kees, I admire your patience with us computer illiterates, I was just grateful we have such a system administrator. Sandro, Renske, (tall) Bram, (pipeline) Bram, Ioana, Hilje, Aridaman, Chris, Thomas, Sai, Paola, Daniel, Hanneke, Nobuto: I've had the weirdest conversations with some of you, and more than a few beers with some others; overall I wish there had been more of both! Sandro, you seriously crept me out at the beginning, but I eventually realized you can be a nice person :P Bram (the tall one), I wouldn't be offended even if you said I looked like someone bald with mustache (well, maybe a little bit!). Daniel, it's nice that you rejoined to the group; I hope your research exodus will come to a happy end sooner or later! Rutger, you did a great job as our Movie Master, I appreciated your effort and I was happy to join more than a few times :)

Bram (again the tall one), Hilje, Bastiaan, I need to thank you for your help with the samenvatting. I didn't imagine I would end up with such a team of translators! Bram, also many thanks for accepting to be my photographer :)

Lidija and Sai, I didn't choose you as my paranymphs without reason. Lidija, you were the unexpected help in some pretty awful period; perhaps you thought you did nothing, but someone to talk to was all I could hope for. Thank you for bothering to check what was going on; it didn't solve everything at once of course, but it made me feel less hopeless and isolated. More recently, I was really impressed by how seriously you took your role as paranymph.. Almost a kg of Zotter chocolate!! :D (I'm proud to say it took me more than 2 weeks to finish it). Sai, it's so weird that we have been sitting in the same corridor for 3 years before getting to know each other. I wish it had happened earlier! You are a surprisingly witty and thoughtful friend, and I really appreciated your detailed help with the defense deadlines. You can also be terrible thought, especially when going out for beers.. Never saw someone this fast with a smartphone!! (I didn't call you my Evil paranymph for nothing). I hope you'll find friends with a good sense of humor in Düsseldorf, you really deserve some!

Being welcome at the Molecular Plant Physiology group talks and social activities gave me the chance to meet several other people here in Utrecht: Monika, Magdalena, Jeroen, Julia, Nicole, Wouter, Jebasingh, Jolanda, Evelien, Sylvia, and quite a few nice students. I enjoyed the many dinners, game nights, movies, and city trips with you in the first couple years in Utrecht! Jolanda and Evelien, you really put an effort in bringing the group together for social activities and presents; I think it's an amazing and absolutely worthy task! Sylvia, you did the same for the MERIT fellows; although I complained about that first course accommodation (ehm!), I really appreciated your work in organizing activities and keeping track of so many people during the 3 years of the project :)

Kristoffer, ever since I left Stockholm you've been kind of an academic guardian angel to me, or at least I like to think that way :P I was really happy to meet you in Utrecht! Next time I come to Heidelberg (or wherever you'll be at the time) I'll make sure to give you a call.

Carol, we met towards the end of your PhD, but even in such a short time you managed to become a good friend to me! You made an extraordinary effort to bring people together, and I was happy to be part of the group :) Silvia, it was a pleasure to meet you as well!

Aleksandra, I'm so glad no one showed up at your appointment on that day (sorry), otherwise I would have missed a great flatmate! You are a strange combination of thoughness and kindness; in spite of your inflexible attitude, and your complaining about my complaining, you always listened when I needed someone to talk to. I really enjoyed our trips to Roma and Wrocław, the Polish breakfasts in the weekend, and the adventurous night trips to/ from Leiden. Thank you for your guidance in fashion matters and your tolerance of food parasitism. I hope you can finish your PhD without too much struggle, and then go for whatever job and country you like! Just make sure to take a few breaks in between; you really look different (happier) when you allow yourself to be a bit crazy :) Krzysiek, you are crazy enough instead!! I had a good time when you stayed at our place, you are a seriously cool guy! Moving abroad was a challenge, but I believe you will make it together :)

Gaurav, you have been my flatmate only for a few months now, but it feels like so much longer! I really appreciate your sense of humor and positive attitude to life, as well as your cooking skills (*drools profusely*). If there ever was an online portal for flatmate reviews, I would give you 5/5 stars :) I'm looking forward to some trips together, and to add other lekker recipes to our list!

Marianna e Francesco, Teresa e Corrado (e Bernardo), Matylda e Raffaele, nonostante la distanza mi siete rimasti incredibilmente vicini; se possibile, ancora più vicini. Ogni volta che sono tornata mi avete fatto sentire accolta e coccolata, che non è poco (qualcuno mi ha addirittura voluta come testimone di nozze :P). Un grazie sincero per non avermi perso di vista. Spero di essere riuscita a trasmettervi e di farvi sentire in futuro la stessa accoglienza, qualsiasi cosa faccia ed ovunque finisca nel mondo. Vi auguro buona fortuna nella vita e nel lavoro; ne avete tutti bisogno in un modo o nell'altro, e vedervi felici e realizzati renderebbe sicuramente più felice anche me. Pippi, un giorno anche noi riusciremo ad incontrarci! (Ho l'inquietante sensazione che potrebbe trattarsi del giorno dell'Apocalisse)

Cari mamma e papà, quante volte mi avete sentito lamentare su Skype! Mi spiace di avervi fatto preocuppare per la mia salute (mentale soprattutto) mentre eravate lontani ed impotenti. Nonostante la distanza siete sempre riusciti a rendermi partecipe della vita domestica, tra foto della campagna e aggiornamenti dettagliati. Ne sono cambiate di cose, ma ogni volta che torno mi sento sempre a casa! Anche se dovessi rimanere all'estero, per me l'importante è continuare a sapere che state bene e che vi divertite un po' anche voi.

Moni, vorrei esserti stata più vicina in questi anni. Purtroppo non è facile trovare la propria strada; nemmeno io so ancora cosa voglio fare nella vita dopo 9 anni di studi universitari! Però ho due consigli da darti: primo, continua a provarci (come stai facendo ora); potresti essere uno di quei fortunati che la strada la trovano, anche a 50 anni. Secondo, non lasciare comunque che tutta la tua felicità dipenda dal trovare la cosa giusta (n.b.: vale lo stesso con le persone!!). Spero di vederti sempre più matura e serena!

Scott, I am so, so thankful that I've met you! (mostly to Ola, if it wasn't for her persistence we wouldn't be here today :P) I'd been warned that these things happen when you least expect them; indeed, we came to be in what was possibly the busiest and most disoriented year of my life. I wish I had been better at thesis deadlines; I felt really sorry for all the times we had to change plans due to the status of my writing flow. Thank you for having been so patient and responsible, and for all your creative ideas to boost my motivation and keep me on track. I'm looking forward to give you back all that time and dedication now!!!
:**********

Description of photographic material

Cover:
Flowers from two different *Drosera* species
The grey/green dendrogram is a topologically accurate representation
of the phylogenetic relationships between currently recognized carnivorous plant genera;
a version annotated with species names is shown in the following page

Chapter 1:
Dionaea muscipula, close-up of snap trap

Chapter 2
Sarracenia sp.

Chapter 3
Drosera sp., close-up of flower

Chapter 4
Sarracenia sp.

Chapter 5
Dionaea muscipula, close-up of flower

Bookmark
Drosera sp., close-up of sticky traps

All photographs by Alessia Peviani.

Plants come from the author's collection, with the exception of the Sarracenias of Chapter 4 (photographed at Keith Wilson's Hungry Plants, Rosmalen, The Netherlands).

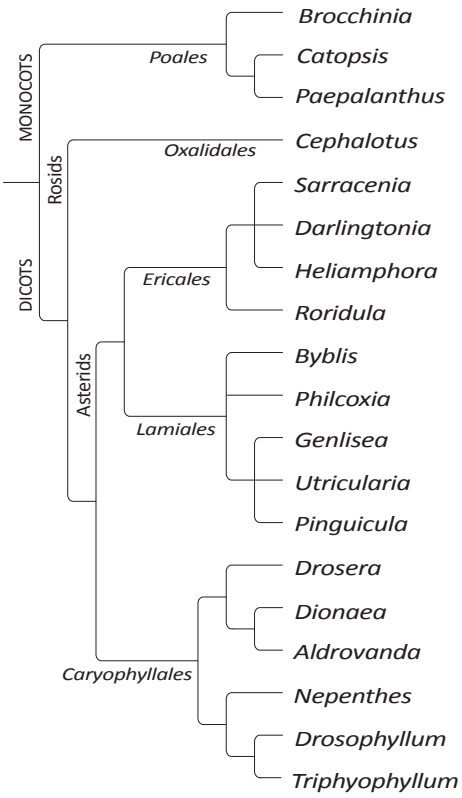


Figure: inferred phylogenetic relationships between currently recognized carnivorous plant genera[1,2]; the tree topology corresponds to the grey/green tree shown in the cover figure.

---

[1] *Heubl G, Bringmann G, Meimberg H (2006) Molecular phylogeny and character evolution of carnivorous plant families in Caryophyllales - revisited. Plant Biol 8: 821–830*

[2] *Givnish TJ (2015) New evidence on the origin of carnivorous plants. Proc Natl Acad Sci USA 112: 10–11*