Institut für Biochemie und Biologie
Arbeitsgruppe Prof. Dr. Bernd Müller-Röber

Evolutionary and functional analysis of transcription factors controlling leaf development

Dissertation
zur Erlangung des akademischen Grades
"doctor rerum naturalium"
(Dr. rer. nat.)
in der Wissenschaftdisziplin "Molekularbiologie"

eingereicht an der
Mathematisch-Naturwissenschaftlichen Falkultät
der Universität Potsdam

von
Luiz Gustavo Guedes Corrêa
aus Ipatinga, Brazil

Potsdam-Golm Winter 2008

Gutachter: Prof. Dr. Bernd Müller-Röber (Universität Potsdam)

Gutachter: Prof. Dr. Jesús Vicente Carbajosa (Universidad Politécnica de Madrid)

Gutachter: Prof. Dr. Karin Koch (University of Florida)

Tag der mündlichen Prüfung: Mai 20. 2009

"Gedruckt mit Unterstützung des Deutschen Akademischen Austauschdienstes"

Εὕρηκα! Εὕρηκα!
—Aristotle

To Vânia and José Luiz,
my safe port!

Online published at the
Institutional Repository of the Potsdam University:
http://opus.kobv.de/ubp/volltexte/2009/4003/
urn:nbn:de:kobv:517-opus-40038

[http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-40038]

Erklärung

I hereby declare that this Ph.D. thesis is the result of my own work carried out between the sommer semester of 2005 and winter semester 2008 in the group of Prof. Dr. Bernd Mueller-Roeber at the University of Potsdam in Golm, Germany. It has not been submitted for any degree or Ph.D. at any other university.

Potsdam, 04.12.2008

Luiz Gustavo Guedes Corrêa

Acknowledgements

It has been quite a long time since I first tought about coming to Germany for my PhD. I still remember the letter from the DAAD, where it stated I was awarded my PhD scholarship. A new country, a new work and new people, with whom I shared the past four years. And here is the opportunity to thank them.

To Prof. Bernd Mueller-Roeber for his supervision and giving me the chance to pursue my scientific interests and the opportunity to try innovations.

To Prof. Martin Steup, Prof. Mark Stitt and Dr. Joachim Fisahn for the excellent discussions and scientific input.

To Prof. Michel Vincentz (Unicamp) for his extra dedication and time spent with me, giving me extra moral to go on.

To Diego Mauricio Riaño Pachón, due to our good team–work and discussions during all my PhD. Much of this work has some of your touch.

To Dr. Karin Kohl and the Green Team in MPI for all the assistance giving during phetotyping and growing plants. This work would not be possible without you.

To all AG MüRö, in special to Ingo, Judy, Fernando, Diego, Salma and Barbara for the nice atmosphere and discussions.

To Dr. Dagmar Krupper, for her assistance on my experiments and friendship.

To Oliver Reinmann, Stephan Spindler, Madlen Wisky and Anja Schüler for their excellent technical assistance.

To Dr. Eugenia Maximova for enthusiastic help with microscopy.

To Dr. Joachin Kopka and Alexander Erban for the metabolite profiling.

To Dr. Joerg Fettke and Marie Caroline Durand-Steinhauser for helping and teaching me with all starch, sugar and enzymatic activity measurements.

To my friends from Caleidoscopio Latino and Turbine Potsdam, for proving me that it is still possible to have life outside the laboratory.

To Anni, Elle and Nils, for their friendship.

To Dr. Babete Regierer, for her enthusiastic support and discussions. It is not possible to thank you enough for everything.

To Natália, Sylvia and Thais, for our discussions and mutual support. We should have

a new ultimate!

To the IPP–IPS programme, for allowing me increasing my skills and taking part into conferences.

To Dr. Heicke Kuchmeister, the coordinator of IPP-IPS, for her excellent work and personal friendship.

To the Deutsche Akademische Austaschedienst (DAAD), for the fellowship and the support during my time in Germany.

To the Kunt, my Turkish family, for their support and admiration. *Sizi seviyorum*.

To the Ronzis, my French family, for their close support and the wonderfull time together. *Je vous adore!*

To all my family, for their love and support, in special to my grandparents, Helenita and Eduardo (*in memorian*), for being for me there when I needed.

To José Luiz and Vânia, my parents, for their unconditional love, support and courage. You are the best thing in my life. *Amo vocês*.

$\Sigma\tau\eta\ M\alpha\rho\iota\alpha\nu\nu\alpha$, $\gamma\iota\alpha\ \tau\eta\ \varphi\iota\lambda\iota\alpha\ \kappa\alpha\iota\ \tau\eta\nu\ \alpha\gamma\alpha\pi\eta\ \tau\eta\varsigma$.

To everyone, that in a way or another, contribute for this thesis.

Contents

List of Figures

List of Tables

List of Abbreviations

λ	Rate of gene gain and loss per million years
μL	Microlitres
ω	Ratio of non-synonymous mutations to synonymous mutations
ABA	Abscisic acid
bp	Base pairs
CRE	*cis*-regulatory element
DAS	Days after sowing
GFP	Green fluorescent protein
h	Hours
kb	Kilo base
min	Minutes
ML	Maximum likelihood
mL	Millilitres
mm	Milimetres
MRCA	Most recent common ancestor
mya	Million years ago
NJ	Neighbor-joining
ORF	Open reading frame
PoGO	Possible Group of Orthologues
qRT-PCR	Quantitative reverse transcription-polymerase chain reaction
TF	Transcription factor
TR	Transcription regulator
TSS	Transcription start site

Summary

Leaves are the main photosynthetic organs of vascular plants, and leaf development is dependent on a proper control of gene expression. Transcription factors (TFs) are global regulators of gene expression that play essential roles in almost all biological processes among eukaryotes. This PhD project focused on the characterization of the sink-to-source transition of Arabidopsis leaves and on the analysis of TFs that play a role in early leaf development.

The sink-to-source transition occurs when the young emerging leaves (net carbon importers) acquire a positive photosynthetic balance and start exporting photoassimilates. We have established molecular and physiological markers (i.e., *CAB1* and *CAB2* expression levels, *AtSUC2* and *AtCHoR* expression patterns, chlorophyll and starch levels, and photosynthetic electron transport rates) to identify the starting point of the transition, especially because the sink-to-source is not accompanied by a visual phenotype in contrast to other developmental transitions, such as the mature-to-senescent transition of leaves. The sink-to-source transition can be divided into two different processes: one light dependent, related to photosynthesis and light responses; and one light independent or impaired, related to the changes in the vascular tissue that occur when leaves change from an import to an export mode. Furthermore, starch, but not sucrose, has been identified as one of the potential signalling molecules for this transition.

The expression level of 1880 TFs during early leaf development was assessed by qRT-PCR, and 153 TFs were found to exhibit differential expression levels of at least 5-fold. GRF, MYB and SRS are TF families, which are overrepresented among the differentially expressed TFs. Additionally, processes like cell identity acquisition, formation of the epidermis and leaf development are overrepresented among the differentially expressed TFs, which helps to validate the results obtained. Two of these TFs were further characterized.

bZIP21 is a gene up-regulated during the sink-to-source and mature-to-senescent transitions. Its expression pattern in leaves overlaps with the one observed for *AtCHoR*, therefore it constitutes a good marker for the sink-to-source transition. Homozygous null mu-

tants of *bZIP21* could not be obtained, indicating that the total absence of *bZIP21* function may be lethal to the plant. Phylogenetic analyses indicate that *bZIP21* is an orthologue of *Liguleless2* from maize. In these analyses, we identified that the whole set of bZIPs in plants originated from four founder genes, and that all bZIPs from angiosperms can be classified into 13 groups of homologues and 34 Possible Groups of Orthologues (PoGOs).

bHLH64 is a gene highly expressed in early sink leaves, its expression is down-regulated during the mature-to-senescent transition. Null mutants of *bHLH64* are characterized by delayed bolting when compared to the wild-type; this indicates a possible delay in the sink-to-source transition or the retention of a juvenile identity.

A third TF, *Dof4*, was also characterized. *Dof4* is neither differentially expressed during the sink-to-source nor during the senescent-to-mature transition, but a null mutant of *Dof4* develops bigger leaves than the wild-type and forms a greater number of siliques. The Dof4 null mutant has proven to be a good background for biomass accumulation analysis.

Though not overrepresented during the sink-to-source transition, NAC transcription factors seem to contribute significantly to the mature-to-senescent transition. Twenty two NACs from Arabidopsis and 44 from rice are differentially expressed during late stages of leaf development. Phylogenetic analyses revealed that most of these NACs cluster into three big groups of homologues, indicating functional conservation between eudicots and monocots. To prove functional conservation of orthologues, the expression of ten *NAC* genes of barley was analysed. Eight of the ten *NAC* genes were found to be differentially expressed during senescence. The use of evolutionary approaches combined with functional studies is thus expected to support the transfer of current knowledge of gene control gained in model species to crops.

1

General introduction

1.1 Leaf development

Together with flowers, leaves are the most prominent plant organs. The leaf is the main photosynthetic organ of vascular plants and its modifications throughout evolution led to the formation of other structures such as spines, tendrils and flowers. There are five different phases during leaf lifespan: (1) acquisition of leaf identity, given by cell differentiation after leaf emergence from the shoot apical meristem (SAM); (2) proliferation, characterized by a prominent cell duplication and metabolic activity; (3) expansion, with the enlargement of the vacuole and the multiplication of chloroplasts; (4) maturation, when the leaf reaches its maximum photosynthetic activity; and (5) senescence, leading to cellular death (Figure 1.1).

Additionally, leaves undergo two main transitions: (1) the sink-to-source and (2) the mature-to-senescent, commonly addressed as senescence. There is a comprehensive literature about the latter transition, which is characterized by a systematic cell death programme and the redistribution of metabolites to sink parts of the plant (MASCLAUX *et al.* 2000). Due to the degradation of the chlorophyll content, leaves turn yellow during this transition, constituting a visual marker.

During early development, young leaves are dependent on carbohydrate imported from other parts of the plant, as they are not able to produce as much photoassymilate as they consume. As the developmental program proceeds, the rate of photosynthesis increases and dark respiration decreases, reducing the dependence of leaves from other plant parts. Meanwhile, the vein system undergoes a major transformation, changing the direction of the major metabolite flux, from importing to exporting. Cell-to-cell traffcking is also more difficult, as plasmosdesmata progressively close, thus requiring transport to occur through veins. The particular phase when the immature sink leaf (a net car-

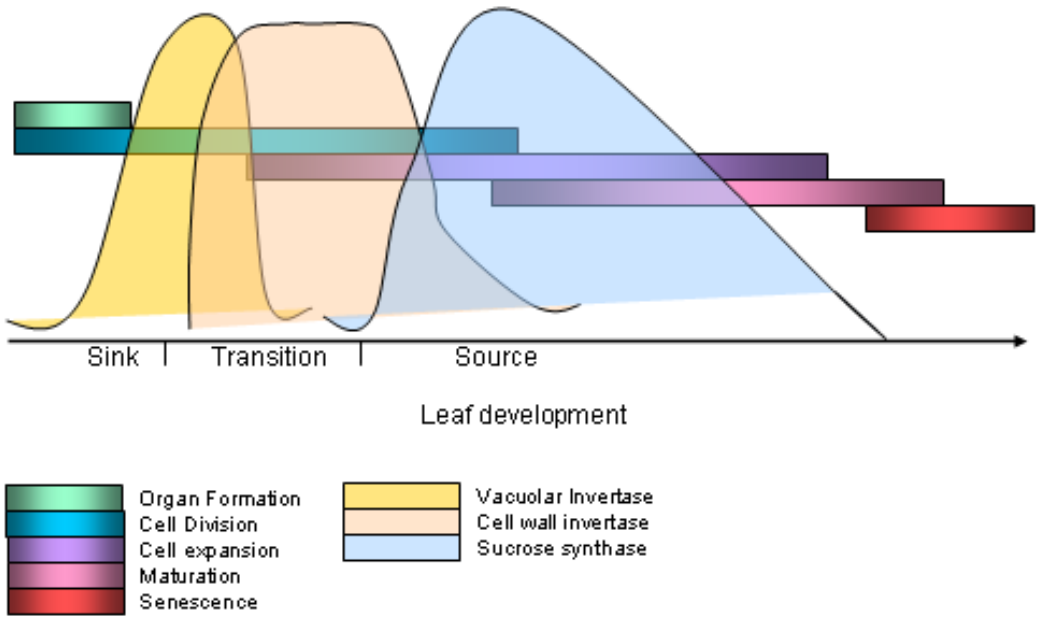


Figure 1.1: ***Stages of leaf development****. During the early phase of leaf development, cells are in an intense process of cell division and photosynthetic maturity is not yet achieved. Invertase activity is important during this phase to ensure sugar supply to cells. The curves indicate qualitatively enzyme activities. During the transition, the venation pattern changes, and leaves turn from net importers to exporters of sugar. At the same time, photosynthesis reaches a positive balance and cell expansion overrides cell division. As leaves enter the maturation phase, the activity of sucrose synthase increases following the increasing exporting role from leaves. At the end of maturation, leaves enter a phase of systematic degradation of molecules, leading to senescence and finally cell death.*

bon importer) starts exporting carbohydrates to other parts of the plant, thus becoming auto-sufficient on carbon production, is called sink-to-source transition. This transition, during which a depletion of both organic and inorganic nitrogen is observed, occurs in a basipetal (tip to base) manner, following a maturation gradient (ADE-ADEMILUA and BOTHA 2007, MASCLAUX *et al.* 2000, MENG *et al.* 2001, OPARKA *et al.* 1999, ROBERTS *et al.* 1997, TURGEON 1989, TURGEON and WEBB 1973).

In the last 15 years, relatively little attention has been given to the sink-to-source transition, which is surprising when one considers the importance of this developmental phase for carbon partioning, plant metabolism and the gain of overall plant biomass. Very little information has been added to the knowledge established in the 80's to understand better the transition from a sink catabolic to a source anabolic status and the activity of enzymes that play a role in these events. Moreover, most of the data available rely on pulse-and-chase experiments, where labelled ^{14}C is given as CO_2 to one source leaf, where it is converted to sugar. The labelled sugar is exported and the uptake of ^{14}C occurs only in

sink leaves, thus allowing identifying the transition (TURGEON 1989). Additionally, almost no information is available for *Arabidopsis thaliana*. Surprisingly, no substantial efforts have been made to use this plant as a model to study the leaf sink-to-source transition after the completion of its genome sequencing (ARABIDOPSIS GENOME INITIATIVE 2000)). Due to the availability of a large number of mutants and data deposited in public databases, it would be also interesting to use Arabidopsis as a model for the sink-to-source transition.

1.2 The sink-to-source transition

Early studies have shown that, in eudicot plants, the sink-to-source transition occurs in a basipetal manner right after the unfolding of the leaf (FELLOWS and GEIGER 1974, JONES and EAGLES 1962, PITCHER and DAIE 1991). It is completed when the leaf reaches 30-60% of its final length (HO 1988, HO and SHAW 1977, HOPKINSON 1966, TURGEON 1989). During this transition, the tip starts exporting carbon to other parts of the plant, whereas the base still imports carbon from other source leaves (VOGELMANN *et al.* 1982). Autoradiography from pulse-and-chase experiments of leaves from Populus deltoides (LARSON and DICKSON 1973), maize (EVERT *et al.* 1996), aspen (KLEINER KW 1999), sugarbeet (PITCHER and DAIE 1991), *Moricandia arvensis* (GAGNON and BEEBE 1996), *Curcubita pepo* and tobacco (TURGEON *et al.* 1975), beans and *Coleus blumei* (TURGEON and WIMMERS 1988) do support the basipetal direction of the transition. In transition leaves, a bidirectional transport of carbon is observed, representing the carbon import in the immature base, and export from the apical region (JONES and EAGLES 1962, LARSON *et al.* 1972). During leaf ontogeny, the photosynthetic capacity of immature leaves increases and the growth rate decreases, resulting in a positive carbon balance in the leaf and a diminution of carbon import (PITCHER and DAIE 1991, TURGEON and WEBB 1973). Meanwhile, unloading from major veins reduces or even ceases (IMLAU *et al.* 1999, OPARKA *et al.* 1999, ROBERTS *et al.* 1997, WRIGHT *et al.* 2003) and the number of plasmodesmata is drastically reduced (ROBERTS *et al.* 2001), compromising the import of photoassymilates and the cell-to-cell trafficking.

It is accepted that the sink-to-source transition coincides with a net positive accumulation of carbon in the leaf. It has been speculated that the overall carbon balance in the plant would be the triggering signal for the transition. However, analyses of tobacco albino plants have shown that the venation pattern changes at the same developmental point as it occurs in wild-type plants, indicating that this change in leaf vein patterning is not controlled by a carbon balance (TURGEON 1984, 1989). Moreover, the veins formed in the albino mutant are able to export if provided external sugar source (TURGEON 1984).

Similar results were obtained with sugar beet under shading treatment, though leaves were able to export sugar just when exposed to light (PITCHER and DAIE 1991). Interestingly, it was also observed that leaves without access to light had normal cell division, but were not able to enter the cell expansion phase, hence, the transition of the venation pattern is not light dependent; proper leaf development may rely on an accurate timing of plasmodesmata closure/positive net carbon balance. Sugar beet leaves that had access to light after dark treatment were able to grow only up to 70% of the size of their control counterparts (PITCHER and DAIE 1991).

1.2.1 Photosynthesis during the sink-to-source transition

Photosynthesis was probably one of be the first plant phenomena that has been systematically characterized. BOUSSINGAULT (1868) was the first to identify the regulatory process of photosynthesis, which is controlled by its end products. Since then, much advance has been made towards a better understanding of its regulation (ARAYA *et al.* 2006, BAKER *et al.* 2007, BOLLIVAR 2006, JIAO *et al.* 2007, KRAPP *et al.* 1993, NIINEMETS 2007, PAUL and PELLNY 2003, PAUL *et al.* 2001, SMITH and STITT 2007). The regulation of photosynthesis is a two-way process, where light induces genes and their activity (feedforward control), and the end-product of the process are used to determine the rate in which the process (photosynthesis) occurs (feedback control).

It is well-known that high sugar concentration leads to a suppression of the expression of genes related to photosynthesis and to a lower photosynthesis rate. On the contrary, lack of sugar induces gene expression and a higher photosynthesis rate. This regulation is commonly addressed as feast and famine (ARAYA *et al.* 2006, KOCH 2004, 1996, PAUL and PELLNY 2003). Although much information is available, the mechanism that leads to a proper maturation of the photosystem and a positive photosynthesis balance in the leaf is not fully resolved. Sugar has been proposed as one of the signalling molecules for the sink-to-source transition. ARAYA *et al.* (2006) suggested that if excess of sugar is applied to the plant, there should be a delay in the transition, a decrease of photosynthesis and a repression of photosynthesis related genes should occur as part of a feast effect. Interestingly, the feast effect is just seen in source leaves, but sink leaves do not seem to be affected by an extra source of sugar. Alternatively, it has been proposed that starch could play a role in the transition or even the C/N balance (ARAYA *et al.* 2006, PAUL *et al.* 2001).

One other important player to be considered are invertases. These enzymes cleave sucrose, thus they are an active part in the sugar importing process. Invertase activity is high in sink tissues and this allows establishing a sucrose gradient from a plant source

to the sink. Normally, the activity of invertases is directly related to the "sink strength", which is a measure of the ability of the sink to import sugar. There is an intrinsic relation between the offer and demand of C in a plant, and this is caused by cross-talk between sink and source parts, leading to the "sink regulation of photosynthesis" (KOCH 2004, KRAPP *et al.* 1993, PAUL *et al.* 2001, ROLLAND *et al.* 2006). An increase on sugar uptake in Arabidopsis roots due to the local expression of a carrot cell wall invertase led to higher production of sugars in leaves (VON SCHWEINICHEN and BÜTTNER 2005), whereas when an increase on photosynthesis is not accompanied by an increase of the sink import, this leads to the repression of sucrose transporter gene expression (PAUL *et al.* 2001).

Studies using albino mutants would help indentifying to which extension the sink-to-source transition is modulated by photosynthesis and possible players that may be involved (TIAN *et al.* 2007).

1.2.2 Markers for the sink-to-source transition

As the sink-to-source transition has no easily visible phenotype, in contrast to senescence, one has to rely on markers that help identifying when the transition takes place. To date, not many genes have been assigned as markers for the transition, or playing a role in it. These markers present a basipetal expression pattern and can be divided into two groups: (i) genes that are expressed in sink tissues and become repressed during the transition, e.g., *AtSUC3* (MEYER *et al.* 2000) *AtNSI* (CARVALHO *et al.* 2006); and (ii) genes not expressed during early leaf growth, but that are activated upon transition such as *AtSUC2* (IMLAU *et al.* 1999) and *AtISI1* (ROOK *et al.* 2006). Currently, of these genes, the best investigated is *AtSUC2*. AtSUC2 is a sucrose symporter expressed in the companion cells of mature leaves being necessary for sucrose transport from source to sink tissues in Arabidopsis (IMLAU *et al.* 1999, SRIVASTAVA *et al.* 2008). Arabidopsis and tobacco plants transformed with a chimeric gene consisting of the *AtSUC2* promoter fused to the *GFP* ORF indicate that *AtSUC2* expression follows a basipetal gradient. Additionally, pulse-and-chase experiments to identify sink leaves and the changes on the venation pattern (from importing to exporting) in tobacco leaves have indicated a strong correlation with the expression of *AtSUC2* (WRIGHT *et al.* 2003), indicating that *AtSUC2* is a good marker for the transition. Recent work on the *AtSUC2* promoter led to the identification of cis-elements required to drive a basipetal expression pattern following the leaf sink-to-source transition (SCHNEIDEREIT *et al.* 2008), indicating that Dof and HD-ZIP could be transcription factors responsible for controlling this expression pattern.

1.3 Control of gene expression

Growth and development of all organisms depend on a proper control of gene expression. The regulation of gene expression occurs at different levels as: transcription, mRNA processing, mRNA transport from the nucleus to the cytoplasm, translation, post-translational modifications, protein multimerization and protein degradation. The control of the transcription initiation rate by transcription regulators constitutes a major mechanism for modulating gene expression (WARREN 2002, WRAY *et al.* 2003). Transcription regulators are known as trans elements and can be defined as proteins able to modulate the transcription initiation rate either by binding a specific DNA sequence (*cis* element) in the upstream region of the gene (i.e., transcription factors) or by altering the activity of other proteins (i.e., transcription regulators) (BECKETT 2001, RIAÑO PACHÓN *et al.* 2008).

Transcription factors are classified into families according to similarities in the primary or tri-dimensional structures of their DNA binding and multimerization domains (RIAÑO PACHÓN *et al.* 2008, WARREN 2002). Some transcription factors families, such as bHLH (basic Helix Loop Helix), bZIP (basic leucine zipper) and MYB, are common to all main eukaryotic lineages (plants, fungi and animals), which diverged around 1.5 billion years ago (WINGENDER *et al.* 2001). Nevertheless, other lineages have developed their own specific sets of transcription factors involved in transcriptional regulation as AP2 (APETALA2) in plants, NHR in animals and Swi4/Swi6 in fungi (RIAÑO PACHÓN *et al.* 2008, RIECHMANN and RATCLIFFE 2000). The emergence of transcription factors that are present only in plants is of particular importance for the evolution of eukaryotic photosynthetic organisms after the original symbiosis that led to the emergence of the plant kingdom (LOPEZ-JUEZ *et al.* 2007, LOPEZ-JUEZ and PYKE 2005). These transcription factors might have important roles in controlling light-dependent processes or new biochemical pathways related to their autothrophic status, including sugar production and accumulation of starch (RIAÑO PACHÓN *et al.* 2008). For example, the G2-like family is present across the whole plant kingdom and some members of this family have been identified as playing a role in chloroplast development, in a process that requires involvement of plastidial and nuclear genomes (YASUMURA *et al.* 2005). Some other transcription factor families, though plant specific, are not present in all plant lineages. NACs are transcription factors that can be found from bryophytes onwards, being absent in algae. It has been shown that NACs play an important role in cell differentiation (OLSEN *et al.* 2005) and more recently in cell senescence/aging (BALAZADEH *et al.* 2008, GREGERSEN and HOLM 2007, UAUY *et al.* 2006), suggesting that multicellularity in embryophytes has a different origin than that in algae.

1.4 Groups of orthologues

Gene duplication is a key process in the diversification and adaptation of an organism during the course of evolution. Duplicated genes are a source of new genetic material for generating new function through mutation and natural selection (ALVAREZ-BUYLLA *et al.* 2000, KELLIS *et al.* 2004, LAWTON-RAUH 2003). Generation of new gene copies through non-homologous recombination and/or partial or whole-genome duplication are mechanisms that lead to the formation of multigenic families (BENNETZEN 2000, KELLIS *et al.* 2004). Multigene families have a big impact on the formation of eukaryotes, considering that they make up 65% of the whole Arabidopsis genome (WENDEL 2000).

The genome sequencing of the red alga *Cyanidioschyzon merolae*, the green alga *Chlamydomonas reinhardtii*, the moss *Physcomitrella patens*, the fern ally *Selaginella moellendorffii*, the monocots *Oryza sativa* (rice) and *Sorghum bicolour* (sorghum) and the eudicots *Arabidopsis thaliana*, *Carica papaya* (papaya), *Medicago truncatula* (barrel medic), *Populus trichocarpa* (black cottonwood) and *Vitis vinifera* (grape) have opened the opportunity for exploring the genetic bases of the main divisions of the plant kingdom (Figure 1.2; ARABIDOPSIS GENOME INITIATIVE 2000, CANNON *et al.* 2006, GOFF *et al.* 2002, JAILLON *et al.* 2007, MERCHANT *et al.* 2007, MING *et al.* 2008, NOZAKI *et al.* 2007, RENSING *et al.* 2008, TUSKAN *et al.* 2006, YU *et al.* 2002; `http://genome.jgi-psf.org/Selaginella`; `http://genome.jgi-psf.org/sorghum`) A possible comparative approach involves the establishment of relationships among genes from different genomes in a system of homologous genes, including orthologues and paralogues (PENNACCHIO 2003, VINCENTZ *et al.* 2004). The term orthologue is related to homologous genes that have diverged after a speciation event. Paralogues are homologues resulting from duplication events in the same genome, they tend to undergo sub- or neofunctionalization (MEYEROWITZ 2002, TATUSOV *et al.* 1997, THORNTON and DESALLE 2000).

To improve the resolution of the evolutionary relationship among genes belonging to the same multigenic family it is necessary to define possible groups of orthologous genes (PoGOs) via phylogenetic approches. Each resulting group is assumed to be the result of the evolution of an ancestral gene through speciation and duplication events. Considering that genes inside the same group of orthologues share the same common ancestor, the delimitation of the group should allow the transfer of biochemical, structural and functional information from one protein to another inside the same group (TATUSOV *et al.* 1997). This approach can optimize the analyses of proteins whose functions are still unknown. Moreover, the investigation of the evolution of gene families allows a better understanding of gene function, which generates a new paradigm in functional biology:

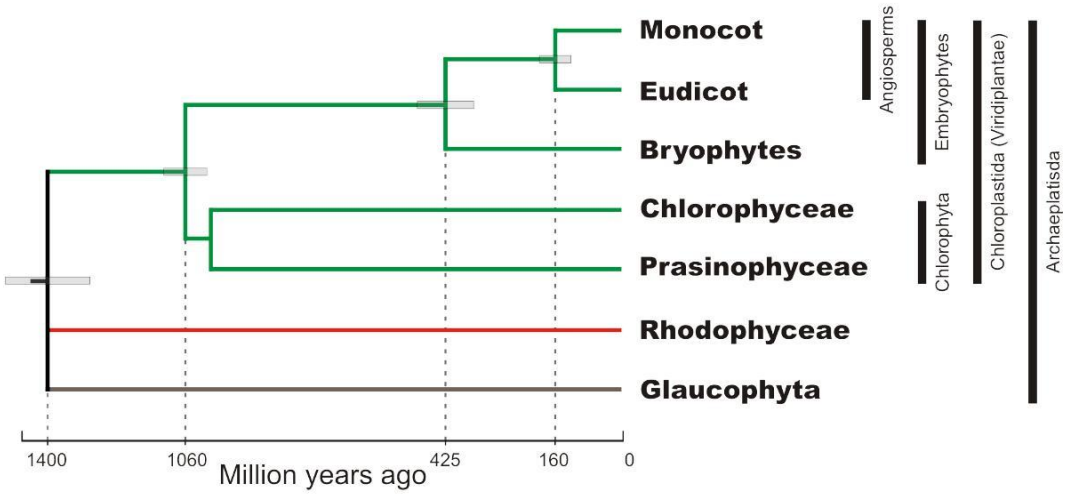


Figure 1.2: ***Emergence of plants.*** *The original symbiosis between a eukaryotic cell and a cyanobacterium leading to the emergence of the plant kingdom occurred around 1,400 million years ago (MYA), corresponding approximately to the split of the three eukaryotic kingdoms. The plant kingdom is subdivided into three main divisions: glaucophytes (freshwater algae), rhodophytes (red algae) and viridiplantae (green plants). Most of the plants currently sequenced belong to the viridiplantae clade. Boxes represent the split from the most recent common ancestor (MRCA) to algae and embryophytes (1,060 MYA), bryophytes and tracheophytes (425 MYA), and monocots and dicots (160 MYA), respectively. The split from pteridophytes and angiosperms is not depicted in this figure.*

instead of having a phenotype and then searching the gene responsible for it, it is possible to select candidate genes first, and through reverse genetics then confirm the function associated to this gene (HENIKOFF *et al.* 1997).

Many databases, which deal with genome comparison and gene families, have become publicly available, facilitating genome-wide comparisons and functional studies (CONTE *et al.* 2008, DUVICK *et al.* 2008, HARTMANN *et al.* 2006, RIAÑO PACHÓN *et al.* 2007, WALL *et al.* 2008).

1.5 Evolution of development

The advances made on genome sequencing allowed exploring not only the relationship among genes and their evolutionary history, but also to associate their evolution to a phenotype or a developmental trait. This new area of research is called evolution of development or "evo-devo", and it is a trial to compare developmental processes between different organisms in an attempt to determine their ancestral relationships and how their developmental processes evolved (CANESTRO *et al.* 2007, GOODMAN and COUGHLIN 2000, RAFF 2000). Evo-devo is supported by, at least, three concepts: (i) phenotypic

plasticity, which is the ability of a given genotype to change its phenotype in response to external conditions; (ii) canalisation, i.e. the rate at which a given population is able to reproduce the same phenotype regardless of a change in the environment; and (iii) modularity, which reflects the concept that anatomical and biochemical parts can be subdivided into modules, which can interact to form a new network (GILBERT 2006, MULLER *et al.* 2008, ROBERTIS 2008, RUTHERFORD and LINDQUIST 1998, SULTAN 2000, VON DASSOW and MUNRO 1999, WAGNER *et al.* 2007).

Modularity is one of the main aspects of evo-devo that has been increasingly accessed. The identification of ancestral networks, conserved gene interactions and their consequences for the phenotype should allow identifying key elements on the course of evolution and, most likely, the pressure exerted towards a certain phenotype and its adaptability to a certain condition or stress. Homeotic genes, like homeobox in animals (NEGRE and RUIZ 2007) and MADS-box in plants (IRISH 2003), represent the main toolkit used in the bauplan formation of these organisms and their function can be traced back to the raise of these lineages, differing just in the complexity of the network (CANESTRO *et al.* 2007, FROHLICH 2003).

Transcription factors constitute one of the main players in the toolkit of development, as they are able to control different processes. The involvement of a transcription factor in a given network can greatly increase its complexity, due the presence of paralogues and the multimerization of transcription factors. Recently, VÁZQUEZ-LOBO *et al.* (2007) have investigated the role of LEAFY/FLORICAULA and NEEDLY leading to reproductive characters in conifers. They have uncovered that the expression of these genes in conifers does not follow the pattern observed in angiosperms, where it seems to be restricted to the male parts, thus weakening the "mostly male theory". SINGER and ASHTON (2007) and SINGER *et al.* (2007) have done an interesting work on the role of MADS-box and homeobox genes with respect to the development of the moss Physcomitrella. They observed that MIKCc-type II MADS-box genes from bryophytes to spermatophytes are involved in the differentiation of reproductive organs from non-reproductive tissues. Additionally, it is possible that the KNOX subfamily is closely involved in the evolutionary transition from a predominant gametophyte in algae and bryophytes to a predominant sporophyte in spermatophytes.

Using an evo-devo approach for transcription factors involved in photosynthesis or leaf formation may uncover new elements of the networks they control.

1.6 References

ADE-ADEMILUA, O., and C. BOTHA, 2007 Sink-to-source transition of *Pisum sativum* leaves in relation

to leaf plastochron index. American Journal of Plant Physiology **2**: 27–35.

ALVAREZ-BUYLLA, E. R., S. PELAZ, S. J. LILJEGREN, S. E. GOLD, C. BURGEFF, *et al.*, 2000 An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. Proc Natl Acad Sci U S A **97**: 5328–5333.

ARABIDOPSIS GENOME INITIATIVE, 2000 Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408**: 796–815.

ARAYA, T., K. NOGUCHI, and I. TERASHIMA, 2006 Effects of carbohydrate accumulation on photosynthesis differ between sink and source leaves of *Phaseolus vulgaris* L. Plant Cell Physiol **47**: 644–652.

BAKER, N. R., J. HARBINSON, and D. M. KRAMER, 2007 Determining the limitations and regulation of photosynthetic energy transduction in leaves. Plant Cell Environ **30**: 1107–1125.

BALAZADEH, S., D. M. R. NO PACHÓN, and B. MUELLER-ROEBER, 2008 Transcription factors regulating leaf senescence in *Arabidopsis thaliana*. Plant Biol (Stuttg) **10 Suppl 1**: 63–75.

BECKETT, D., 2001 Regulated assembly of transcription factors and control of transcription initiation. J Mol Biol **314**: 335–352.

BENNETZEN, J. L., 2000 Comparative sequence analysis of plant nuclear genomes:m microcolinearity and its many exceptions. Plant Cell **12**: 1021–1029.

BOLLIVAR, D. W., 2006 Recent advances in chlorophyll biosynthesis. Photosynth Res **90**: 173–194.

BOUSSINGAULT, J., editor, 1868 *Agronomie, chimie agricole et physiologie*. Mallet Bachelier.

CANESTRO, C., H. YOKOI, and J. H. POSTLETHWAIT, 2007 Evolutionary developmental biology and genomics. Nat Rev Genet **8**: 932–942.

CANNON, S. B., L. STERCK, S. ROMBAUTS, S. SATO, F. CHEUNG, *et al.*, 2006 Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. Proc Natl Acad Sci U S A **103**: 14959–14964.

CARVALHO, M. F., R. TURGEON, and S. G. LAZAROWITZ, 2006 The geminivirus nuclear shuttle protein NSP inhibits the activity of AtNSI, a vascular-expressed arabidopsis acetyltransferase regulated with the sink-to-source transition. Plant Physiol **140**: 1317–1330.

CONTE, M. G., S. GAILLARD, N. LANAU, M. ROUARD, and C. PÉRIN, 2008 GreenPhylDB: a database for plant comparative genomics. Nucleic Acids Res **36**: D991–D998.

DUVICK, J., A. FU, U. MUPPIRALA, M. SABHARWAL, M. D. WILKERSON, *et al.*, 2008 PlantGDB: a resource for comparative plant genomics. Nucleic Acids Res **36**: D959–D965.

EVERT, R. F., W. A. RUSSIN, and A. M. BOSABALIDIS, 1996 Anatomical and Ultrastructural Changes Associated with Sink-to-Source Transition in Developing Maize Leaves. International Journal of Plant Sciences **3**: 247.

FELLOWS, R., and D. GEIGER, 1974 Structural and physiological changes in sugar beet leaves during sink to source conversion. Plant Physiology **54**: 877–885.

FROHLICH, M. W., 2003 An evolutionary scenario for the origin of flowers. Nat Rev Genet **4**: 559–566.

GAGNON, M.-J., and D. BEEBE, 1996 Establishment of a plastochron index and analysis of the sink-to-source transition in leaves of *Moricandia arvensis* (L.) DC. (Brassicaceae). International Journal of Plant Sciences **157**: 262–268.

GILBERT, S. F., 2006 *Developmental Biology*. Sinauer Associates, Inc.

GOFF, S. A., D. RICKE, T.-H. LAN, G. PRESTING, R. WANG, *et al.*, 2002 A draft sequence of the rice

genome (*Oryza sativa* L. ssp. *japonica*). Science **296**: 92–100.

GOODMAN, C. S., and B. C. COUGHLIN, 2000 Introduction. The evolution of evo-devo biology. Proc Natl Acad Sci U S A **97**: 4424–4425.

GREGERSEN, P. L., and P. B. HOLM, 2007 Transcriptome analysis of senescence in the flag leaf of wheat (*Triticum aestivum* L.). Plant Biotechnol J **5**: 192–206.

HARTMANN, S., D. LU, J. PHILLIPS, and T. J. VISION, 2006 Phytome: a platform for plant comparative genomics. Nucleic Acids Res **34**: D724–D730.

HENIKOFF, S., E. A. GREENE, S. PIETROKOVSKI, P. BORK, T. K. ATTWOOD, *et al.*, 1997 Gene families: the taxonomy of protein paralogs and chimeras. Science **278**: 609–614.

HO, L. C., 1988 Metabolism and compartmentation of imported sugars in sink organs in relation to sink strength. Annual Review of Plant Physiology and Plant Molecular Biology **39**: 355–378.

HO, L. C., and A. F. SHAW, 1977 Carbon economy and translocation of ^{14}C in leaflets of the seventh leaf of tomato during leaf expansion. Annals of Botany **41**: 833–848.

HOPKINSON, J. M., 1966 Studies on the expansion of the leaf surface. IV. The carbon and phosphorus economy of a leaf. J. Exp. Bot. **15**: 125–137.

IMLAU, A., E. TRUERNIT, and N. SAUER, 1999 Cell-to-cell and long-distance trafficking of the green fluorescent protein in the phloem and symplastic unloading of the protein into sink tissues. Plant Cell **11**: 309–322.

IRISH, V. F., 2003 The evolution of floral homeotic gene function. Bioessays **25**: 637–646.

JAILLON, O., J.-M. AURY, B. NOEL, A. POLICRITI, C. CLEPET, *et al.*, 2007 The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature **449**: 463–467.

JIAO, Y., O. S. LAU, and X. W. DENG, 2007 Light-regulated transcriptional networks in higher plants. Nat Rev Genet **8**: 217–230.

JONES, H., and J. E. EAGLES, 1962 Translocation of ^{14}Carbon within and between leaves. Annals of Botany **26**: 505–510.

KELLIS, M., B. W. BIRREN, and E. S. LANDER, 2004 Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. Nature **428**: 617–624.

KLEINER KW, RAFFA KF, D. R., 1999 Partitioning of ^{14}C-labeled photosynthate to allelochemicals and primary metabolites in source and sink leaves of aspen: evidence for secondary metabolite turnover. Oecologia **119**: 408–418.

KOCH, K., 2004 Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar sensing and plant development. Curr Opin Plant Biol **7**: 235–246.

KOCH, K. E., 1996 Carbohydrate-modulated gene expression in plants. Annu Rev Plant Physiol Plant Mol Biol **47**: 509–540.

KRAPP, A., B. HOFMANN, C. SCHÄFER, and M. STITT, 1993 Regulation of the expression of *rbcS* and other photosynthetic genes by carbohydrates: a mechanism for the "sink regulation" of photosynthesis? The Plant Journal **3**: 817–828.

LARSON, P., and R. DICKSON, 1973 Distribution of imported ^{14}C in developing leaves of eastern cottonwood according to phyllotaxy. Planta **115**: 95–112.

LARSON, P. R., J. G. ISEBRANDS, and R. E. DICKSON, 1972 Fixation patterns of ^{14}C within developing leaves of eastern cottonwood. Planta **107**: 301–314.

LAWTON-RAUH, A., 2003 Evolutionary dynamics of duplicated genes in plants. Mol Phylogenet Evol **29**: 396–409.

LOPEZ-JUEZ, E., J. R. BOWYER, and T. SAKAI, 2007 Distinct leaf developmental and gene expression responses to light quantity depend on blue-photoreceptor or plastid-derived signals, and can occur in the absence of phototropins. Planta **227**: 113–123.

LOPEZ-JUEZ, E., and K. A. PYKE, 2005 Plastids unleashed: their development and their integration in plant development. Int J Dev Biol **49**: 557–577.

MASCLAUX, C., M. H. VALADIER, N. BRUGIÈRE, J. F. MOROT-GAUDRY, and B. HIREL, 2000 Characterization of the sink/source transition in tobacco (*Nicotiana tabacum* L.) shoots in relation to nitrogen management and leaf senescence. Planta **211**: 510–518.

MENG, Q., K. SIEBKE, P. LIPPERT, B. BAUR, U. MUKHERJEE, *et al.*, 2001 Sink-source transition in tobacco leaves visualized using chlorophyll fluorescence imaging. New Phytologist **151**: 585–595.

MERCHANT, S. S., S. E. PROCHNIK, O. VALLON, E. H. HARRIS, S. J. KARPOWICZ, *et al.*, 2007 The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science **318**: 245–250.

MEYER, S., M. MELZER, E. TRUERNIT, C. HÜMMER, R. BESENBECK, *et al.*, 2000 AtSUC3, a gene encoding a new Arabidopsis sucrose transporter, is expressed in cells adjacent to the vascular tissue and in a carpel cell layer. Plant J **24**: 869–882.

MEYEROWITZ, E. M., 2002 Plants compared to animals: the broadest comparative study of development. Science **295**: 1482–1485.

MING, R., S. HOU, Y. FENG, Q. YU, A. DIONNE-LAPORTE, *et al.*, 2008 The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). Nature **452**: 991–996.

MULLER, M., M. OBEYESEKERE, G. B. MILLS, and P. T. RAM, 2008 Network topology determines dynamics of the mammalian MAPK1,2 signaling network: bifan motif regulation of C-Raf and B-Raf isoforms by FGFR and MC1R. FASEB J **22**: 1393–1403.

NEGRE, B., and A. RUIZ, 2007 HOM-C evolution in Drosophila: is there a need for Hox gene clustering? Trends Genet **23**: 55–59.

NIINEMETS, U., 2007 Photosynthesis and resource distribution through plant canopies. Plant Cell Environ **30**: 1052–1071.

NOZAKI, H., H. TAKANO, O. MISUMI, K. TERASAWA, M. MATSUZAKI, *et al.*, 2007 A 100%-complete sequence reveals unusually simple genomic features in the hot-spring red alga *Cyanidioschyzon merolae*. BMC Biol **5**: 28.

OLSEN, A. N., H. A. ERNST, L. L. LEGGIO, and K. SKRIVER, 2005 NAC transcription factors: structurally distinct, functionally diverse. Trends Plant Sci **10**: 79–87.

OPARKA, K. J., A. G. ROBERTS, P. BOEVINK, S. S. CRUZ, I. ROBERTS, *et al.*, 1999 Simple, but not branched, plasmodesmata allow the nonspecific trafficking of proteins in developing tobacco leaves. Cell **97**: 743–754.

PAUL, M., and T. PELLNY, 2003 Carbon metabolite feedback regulation of leaf photosynthesis and development. Journal of Experimental Botany **54**: 539–547.

PAUL, M., T. PELLNY, and G. O., 2001 Enhancing photosynthesis with sugar signals. Trends in Plant Science **6**: 197–200.

PENNACCHIO, L. A., 2003 Insights from human/mouse genome comparisons. Mamm Genome **14**: 429–

436.

PITCHER, L., and J. DAIE, 1991 Growth and sink-to-source transition in developing leaves of sugarbeet. Plant and Cell Physiology **32**: 335–342.

RAFF, R. A., 2000 Evo-devo: the evolution of a new discipline. Nat Rev Genet **1**: 74–79.

RENSING, S. A., D. LANG, A. D. ZIMMER, A. TERRY, A. SALAMOV, *et al.*, 2008 The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science **319**: 64–69.

RIAÑO PACHÓN, D. M., L. G. G. CORRÊA, R. TREJOS-ESPINOSA, and B. MUELLER-ROEBER, 2008 Green transcription factors: a chlamydomonas overview. Genetics **179**: 31–39.

RIAÑO PACHÓN, D. M., S. RUZICIC, I. DREYER, and B. MUELLER-ROEBER, 2007 PlnTFDB: An integrative plant transcription factor database. BMC Bioinformatics **8**: 42.

RIECHMANN, J. L., and O. J. RATCLIFFE, 2000 A genomic perspective on plant transcription factors. Curr Opin Plant Biol **3**: 423–434.

ROBERTIS, E. M. D., 2008 The molecular ancestry of segmentation mechanisms. Proc Natl Acad Sci U S A **105**: 16411–16412.

ROBERTS, A. G., S. S. CRUZ, I. M. ROBERTS, D. A. M. PRIOR, R. TURGEON, *et al.*, 1997 Phloem unloading in sink leaves of *Nicotiana benthamiana*: Comparison of a fluorescent solute with a fluorescent virus. Plant Cell **9**: 1381–1396.

ROBERTS, I. M., P. BOEVINK, A. G. ROBERTS, N. SAUER, C. REICHEL, *et al.*, 2001 Dynamic changes in the frequency and architecture of plasmodesmata during the sink-source transition in tobacco leaves. Protoplasma **218**: 31–44.

ROLLAND, F., E. BAENA-GONZALEZ, and J. SHEEN, 2006 Sugar sensing and signaling in plants: conserved and novel mechanisms. Annu Rev Plant Biol **57**: 675–709.

ROOK, F., F. CORKE, M. BAIER, R. HOLMAN, A. G. MAY, *et al.*, 2006 *Impaired sucrose induction1* encodes a conserved plant-specific protein that couples carbohydrate availability to gene expression and plant growth. Plant J **46**: 1045–1058.

RUTHERFORD, S. L., and S. LINDQUIST, 1998 Hsp90 as a capacitor for morphological evolution. Nature **396**: 336–342.

SCHNEIDEREIT, A., A. IMLAU, and N. SAUER, 2008 Conserved cis-regulatory elements for DNA-binding-with-one-finger and homeo-domain-leucine-zipper transcription factors regulate companion cell-specific expression of the *Arabidopsis thaliana SUCROSE TRANSPORTER 2* gene. Planta **228**: 651–662.

SINGER, S. D., and N. W. ASHTON, 2007 Revelation of ancestral roles of KNOX genes by a functional analysis of Physcomitrella homologues. Plant Cell Rep **26**: 2039–2054.

SINGER, S. D., N. T. KROGAN, and N. W. ASHTON, 2007 Clues about the ancestral roles of plant MADS-box genes from a functional analysis of moss homologues. Plant Cell Rep **26**: 1155–1169.

SMITH, A. M., and M. STITT, 2007 Coordination of carbon supply and plant growth. Plant Cell Environ **30**: 1126–1149.

SRIVASTAVA, A. C., S. GANESAN, I. O. ISMAIL, and B. G. AYRE, 2008 Functional characterization of the arabidopsis AtSUC2 Sucrose/H+ symporter by tissue-specific complementation reveals an essential role in phloem loading but not in long-distance transport. Plant Physiol **148**: 200–211.

SULTAN, S. E., 2000 Phenotypic plasticity for plant development, function and life history. Trends Plant Sci **5**: 537–542.

TATUSOV, R. L., E. V. KOONIN, and D. J. LIPMAN, 1997 A genomic perspective on protein families. Science **278**: 631–637.

THORNTON, J., and R. DESALLE, 2000 Gene family evolution and homology: genomics meets phylogenetics. Annu Rev Genomics Hum Genet **1**: 41–73.

TIAN, C., E. CHIKAYAMA, Y. TSUBOI, T. KUROMORI, K. SHINOZAKI, *et al.*, 2007 Top-down phenomics of *Arabidopsis thaliana*: metabolic profiling by one- and two-dimensional nuclear magnetic resonance spectroscopy and transcriptome analysis of albino mutants. J Biol Chem **282**: 18532–18541.

TURGEON, R., 1984 Termination of nutrient import and development of vein loading capacity in albino tobacco leaves. Plant Physiol **76**: 45–48.

TURGEON, R., 1989 The sink-source transition in leaves. Annual Review of Plant Physiology and Plant Molecular Biology **40**: 118–138.

TURGEON, R., and J. A. WEBB, 1973 Leaf development and phloem transport in *Cucurbita pepo*: Transition from import to export. Planta **113**: 197–191.

TURGEON, R., J. A. WEBB, and R. F. EVEN, 1975 Ultrastructure of minor veins in *Cucurbita pepo* leaves. Protoplasma **83**: 217–232.

TURGEON, R., and L. E. WIMMERS, 1988 Different patterns of vein loading of exogenous [^{14}C]sucrose in leaves of *Pisum sativum* and *Coleus blumei*. Plant Physiol **87**: 179–182.

TUSKAN, G. A., S. DIFAZIO, S. JANSSON, J. BOHLMANN, I. GRIGORIEV, *et al.*, 2006 The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science **313**: 1596–1604.

UAUY, C., A. DISTELFELD, T. FAHIMA, A. BLECHL, and J. DUBCOVSKY, 2006 A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. Science **314**: 1298–1301.

VÁZQUEZ-LOBO, A., A. CARLSBECKER, F. VERGARA-SILVA, E. R. ALVAREZ-BUYLLA, D. P. NERO, *et al.*, 2007 Characterization of the expression patterns of LEAFY/FLORICAULA and NEEDLY orthologs in female and male cones of the conifer genera Picea, Podocarpus, and Taxus: implications for current evo-devo hypotheses for gymnosperms. Evol Dev **9**: 446–459.

VINCENTZ, M., F. A. A. CARA, V. K. OKURA, F. R. DA SILVA, G. L. PEDROSA, *et al.*, 2004 Evaluation of monocot and eudicot divergence using the sugarcane transcriptome. Plant Physiol **134**: 951–959.

VOGELMANN, T. C., P. R. LARSON, and R. E. DICKSON, 1982 Translocation pathways in the petioles and stem between source and sink leaves of *Populus deltoides* Bartr. ex Marsh. Planta **156**: 345–358.

VON DASSOW, G., and E. MUNRO, 1999 Modularity in animal development and evolution: Elements of a conceptual framework for EvoDevo. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution **285**: 307–325.

VON SCHWEINICHEN, C., and M. BÜTTNER, 2005 Expression of a plant cell wall invertase in roots of Arabidopsis leads to early flowering and an increase in whole plant biomass. Plant Biol (Stuttg) **7**: 469–475.

WAGNER, G. P., M. PAVLICEV, and J. M. CHEVERUD, 2007 The road to modularity. Nat Rev Genet **8**: 921–931.

WALL, P. K., J. LEEBENS-MACK, K. F. MÜLLER, D. FIELD, N. S. ALTMAN, *et al.*, 2008 PlantTribes: a gene and gene family resource for comparative genomics in plants. Nucleic Acids Res **36**: D970–D976.

WARREN, A. J., 2002 Eukaryotic transcription factors. Curr Opin Struct Biol **12**: 107–114.

WENDEL, J. F., 2000 Genome evolution in polyploids. Plant Mol Biol **42**: 225–249.

WINGENDER, E., X. CHEN, E. FRICKE, R. GEFFERS, R. HEHL, *et al.*, 2001 The TRANSFAC system on gene expression regulation. Nucleic Acids Res **29**: 281–283.

WRAY, G. A., M. W. HAHN, E. ABOUHEIF, J. P. BALHOFF, M. PIZER, *et al.*, 2003 The evolution of transcriptional regulation in eukaryotes. Mol Biol Evol **20**: 1377–1419.

WRIGHT, K. M., A. G. ROBERTS, H. J. MARTENS, N. SAUER, and K. J. OPARKA, 2003 Structural and functional vein maturation in developing tobacco leaves in relation to AtSUC2 promoter activity. Plant Physiol **131**: 1555–1565.

YASUMURA, Y., E. C. MOYLAN, and J. A. LANGDALE, 2005 A conserved transcription factor mediates nuclear control of organelle biogenesis in anciently diverged land plants. Plant Cell **17**: 1894–1907.

YU, J.-Y., S. L. DERUITER, and D. L. TURNER, 2002 RNA interference by expression of short-interfering RNAs and hairpin RNAs in mammalian cells. Proc Natl Acad Sci U S A **99**: 6047–52.

2

Transcription factors in *Chlamydomonas reinhardtii*

Green transcription factors: a Chlamydomonas overview

Diego Mauricio Riaño-Pachón[1,2,†], Luiz Gustavo Guedes Corrêa[1,3,†], Raúl Trejos-Espinosa[1,3], Bernd Mueller-Roeber[1,3]

[†] These authors contributed equally to this work and should thus both be considered as first authors.

[1]Department of Molecular Biology, University of Potsdam, Potsdam-Golm, Germany, [2]Bioinformatics Research Group, GabiPD Team, Max-Planck Institute of Molecular Plant Physiology, Potsdam–Golm, Germany,[3]Cooperative Research Group, Max-Planck Institute of Molecular Plant Physiology, Potsdam–Golm, Germany

Published in *Genetics* (2008) **179**(1):31-39. doi:10.1534/genetics.107.086090

Author contributions

BMR, LGGC and DMRP conceived and designed the study. BMR coordinated the project. LGGC and RTJ identified groups of orthologues by phylogenetic analyses using NJ. DMRP identified groups of orthologues by symmetrical BLAST hits, compared the results of phylogenetics and symmetric similarity matches, identified TFs and TRs in both photosynthetic and non-photosynthetic species. All authors discussed and analysed the data.

DOI: 10.1534/genetics.107.086090

Green Transcription Factors: A Chlamydomonas Overview

Diego Mauricio Riaño-Pachón,*,†,[1] Luiz Gustavo Guedes Corrêa,*,‡,[1] Raúl Trejos-Espinosa*,‡ and Bernd Mueller-Roeber*,‡,[2]

**Department of Molecular Biology, University of Potsdam, 14476 Potsdam-Golm, Germany, †Bioinformatics Research Group, Max-Planck Institute of Molecular Plant Physiology, GabiPD Team, 14476 Potsdam-Golm, Germany and ‡Cooperative Research Group, Max-Planck Institute of Molecular Plant Physiology, 14476 Potsdam-Golm, Germany*

Manuscript received December 15, 2007
Accepted for publication January 29, 2008

ABSTRACT

Transcription factors (TFs) control gene expression by interacting with *cis*-elements in target gene promoters. Transcription regulators (TRs) assist in controlling gene expression through interaction with TFs, chromatin remodeling, or other mechanisms. Both types of proteins thus constitute master controllers of dynamic transcriptional networks. To uncover such control elements in the photosynthetic green alga *Chlamydomonas reinhardtii*, we performed a comprehensive analysis of its genome sequence. In total, we identified 234 genes encoding 147 TFs and 87 TRs of ~40 families. The set of putative TFs and TRs, including their transcript and protein sequences, domain architectures, and supporting information about putative orthologs, is available at http://plntfdb.bio.uni-potsdam.de/v2.0/. Twelve of 34 plant-specific TF families were found in at least one algal species, indicating their early evolutionary origin. Twenty-two plant-specific TF families and one plant-specific TR family were not observed in algae, suggesting their specific association with developmental or physiological processes characteristic to multicellular plants. We also analyzed the occurrence of proteins that constitute the light-regulated transcriptional network in angiosperms and found putative algal orthologs for most of them. Our analysis provides a solid ground for future experimental studies aiming at deciphering the transcriptional regulatory networks in green algae.

THE regulation of growth and development and the coordination of these processes in response to hormonal or environmental stimuli, including adverse conditions, requires a dynamic control of the expression of hundreds to thousands of genes in each organism (Lemon and Tjian 2000; Chen *et al.* 2002; Li *et al.* 2007). Transcription factors (TFs) are master control proteins that regulate gene expression levels by binding to specific DNA sequences, so-called *cis*-acting elements, in the promoters of target genes, thereby enhancing or repressing their transcriptional rates. The genomewide identification of TF genes through computational methods, and genomewide comparative studies, are important tasks that not only provide an insight into existing TF families within individual species or organism lineages but also help to understand how evolution shaped developmental and physiological diversification. TFs, as well as other transcriptional regulators (TRs) that generally do not directly bind DNA but assist in gene expression regulation through interaction with *cis*-element-binding proteins, can be grouped into different protein families according to their primary and/or three-dimensional structure similarities in the DNA-binding and multimerization domains. TF genes represent a considerable fraction of the genomes of all eukaryotic organisms, including angiosperms (Riechmann *et al.* 2000; Goff *et al.* 2002). In *Oryza sativa* (rice), for example, ~2.6% of the identified genes encode TFs (Goff *et al.* 2002). Currently, the genome sequences of four angiosperms (*Arabidopsis thaliana*, *O. sativa*, *Populus trichocarpa*, and *Vitis vinifera*) are in the public domain (Arabidopsis Genome Initiative 2000; Goff *et al.* 2002; Yu *et al.* 2002; Tuskan *et al.* 2006; Jaillon *et al.* 2007). Additionally, the genomes of various algae, including the red alga *Cyanidioschyzon merolae* (Nozaki *et al.* 2007), the green algae *Ostreococcus tauri* (Derelle *et al.* 2006), *Chlamydomonas reinhardtii* (Merchant *et al.* 2007), and the moss *Physcomitrella patens* (Rensing *et al.* 2008) have become available.

To facilitate the analysis of plant TFs and TRs, we have recently established the Plant Transcription Factor Database (PlnTFDB) (Riano-Pachon *et al.* 2007) and updated it by including additional plant species (available at http://plntfdb.bio.uni-potsdam.de/v2.0). Here we report about the occurrence of putative transcriptional regulators in Chlamydomonas. We identified 147 putative TFs that belong to 29 different protein families and 87 putative TRs that are members of 10 families. Of 34 plant-specific families, 3 (G_2-like, PLATZ, RWP-RK) predate the split between green and red algae. Nine

[1]These authors contributed equally to this work.

[2]*Corresponding author:* Department of Molecular Biology, University of Potsdam, Karl-Liebknecht-Strasse 24-25, Haus 20, 14476 Potsdam-Golm, Germany. E-mail: bmr@uni-potsdam.de

additional families, *i.e.*, ABI3/VP1, AP2-EREBP, ARR-B, C2C2-CO-like, C2C2-Dof, PBF-2-like/Whirly, Pseudo ARR-B, SBP, and WRKY, predate the split between chlorophyta (green algae) and streptophyta (land plants and charophycean algae). In total, 12 families were identified from algal groups onward. Interestingly, 22 plant-specific TF families and one TR family are not present in algae, indicating their particular importance for plant multicellularity and tissue organization.

MATERIALS AND METHODS

Identification of transcription factors: Putative complete sets of transcription factors of the following species were retrieved from the Plant Transcription Factor Database v2.0, PlnTFDB (http://plntfdb.bio.uni-potsdam.de/v2.0; Riano-Pachon *et al.* 2007): the red alga *C. merolae* (Nozaki *et al.* 2007), the green algae *O. tauri* (Derelle *et al.* 2006) and *C. reinhardtii* (Merchant *et al.* 2007), the moss *P. patens* (Rensing *et al.* 2008), and the angiosperms *A. thaliana* (Arabidopsis Genome Initiative 2000), *P. trichocarpa* (black cottonwood) (Tuskan *et al.* 2006), and *O. sativa* (rice) (Goff *et al.* 2002). PlnTFDB has two divisions: one providing information about transcription factors, defined as proteins that directly bind to DNA and affect the level of transcription (called TFs here), and the other providing information about transcriptional regulators that, for example, exert regulatory control through interaction with TFs or through chromatin remodeling (called TRs). Additionally, we identified TF and TR families common to all eukaryotes using the following model organisms: the protozoan *Giardia lamblia* (Best *et al.* 2004), the yeast *Saccharomyces cerevisiae* (Goffeau *et al.* 1996), the nematode *Caenorhabditis elegans* (*C. elegans* Sequencing Consortium 1998), the insect *Drosophila melanogaster* (Adams *et al.* 2000), and *Homo sapiens* (International Human Genome Sequencing Consortium 2004). For the identification of nonplant TFs, we used the procedure described by Riano-Pachon *et al.* (2007), using PFAM (Finn *et al.* 2006) release 20.0 for domain identification. Sequences were downloaded from Integr8 (http://www.ebi.ac.uk/integr8/; Kersey *et al.* 2005), except for *G. lamblia* sequences, which were downloaded from GiardiaDB (http://www.giardiadb.org).

Phylogenetic analysis: Protein sequences corresponding to the defining conserved domain of each TF and TR family were extracted from whole-protein sequences of the photosynthetic eukaryotes using the domain coordinates identified by the PFAM search described above. Alignment of protein sequences was performed employing ClustalX (Thompson *et al.* 1997), using default parameters. Phylogenetic analyses based on amino acid sequences were conducted using MEGA v3.1 (Kumar *et al.* 2004). Unrooted phylogenetic tree topologies were reconstructed by neighbor-joining (NJ), the distances were obtained using p-distances (Nei and Kumar 2000), and the resampling of the original protein set was a 1000-bootstrap repetition. These NJ analyses provide an overview of the general patterns of TF and TR evolution. All sequences and alignments used in this study are available upon request.

Identification of orthologs among green plants: We identified orthologs through pairwise comparisons of protein sequences in whole-protein sets of the green plants Chlamydomonas, Ostreococcus, Physcomitrella, rice, Arabidopsis, and black cottonwood, using a variation of the best BLAST bidirectional hit approach implemented in the program InParanoid (Remm *et al.* 2001). Orthologs identified in this way are presented in PlnTFDB (http://plntfdb.bio.uni-potsdam.de/v2.0). Pairs of orthologs (and direct paralogs) allowed us to identify possible clusters of orthologs (and paralogs) comprising genes from more than two species. This was achieved using a graph-theoretic approach as follows: (i) only InParanoid clusters containing at least one protein annotated as a TF in PlnTFDB were kept; (ii) all proteins identified in this way were represented as nodes in a network of orthologous relationships; edges were drawn between nodes when the InParanoid confidence score for the orthologous relationship was ≥ 0.9; (iii) connected components were extracted from the network; by definition, a connected component is a subgraph in which every node can be reached from every other node. The connected components (subgraphs) represent putative clusters of orthologs. Network visualization and analysis were carried out using the software package Pajek (De Nooy *et al.* 2005). The identification of orthologs through BLAST searches can lead to false positives; consequently, we made use of a phylogenetic approach to largely compensate for this fact. In addition to that, and affecting both approaches for ortholog detection (phylogenetics and InParanoid), false negatives can arise due to incomplete genome sequence information (gaps in the sequence) or misannotated genes.

As mentioned above, in addition to the BLAST approach, we performed phylogenetic analyses of each family, which allowed the identification of possible groups of orthologs (PoGOs). A PoGO is defined by the following criteria: (i) members of a PoGO have a monophyletic origin, indicated by a bootstrap support of $>50\%$; (ii) a PoGO conserved in all green plants possesses at least one representative gene of each of the main lineages analyzed here, including algae, bryophytes, and angiosperms, assuming that the putative complete sets of TF genes of these organisms were identified and no selective gene loss had occurred; (iii) the inferred phylogeny is consistent with the known phylogeny of plant species (Vincentz *et al.* 2003).

We evaluated the overlap between the clusters of orthologs identified by InParanoid and by phylogenetic analysis using the Adjusted Rand Index (Rand 1971; Hubert and Arabie 1985), implemented in the statistical package R (R Development Core Team 2007).

RESULTS AND DISCUSSION

Transcription factors in eukaryotes: We identified the putatively complete nonredundant sets of TFs and TRs in the algae Chlamydomonas and Ostreococcus, the moss Physcomitrella, and the angiosperms Arabidopsis, black cottonwood, and rice (Table 1). The genes were grouped into 66 gene families according to their characteristic conserved domains, as described by Riano-Pachon *et al.* (2007). We identified the putatively complete sets of genes for the same families in *G. lamblia* (protozoa), *S. cerevisiae* (yeast), *C. elegans* (nematodes), *D. melanogaster* (fruit flies), and *H. sapiens* (humans). Twenty TF and 11 TR families were also present in nonphotosynthetic eukaryotes. In contrast to the previous report by Riechmann *et al.* (2000), we observed that the Trihelix family is not restricted to the plant kingdom (Table 1). G_2-like and WRKY TFs are generally regarded as plant specific; our analysis largely confirms this view. However, we also identified genes encoding putative members of these families in the nonplant species *G. lamblia*, a

TABLE 1

Transcription factors and regulators present in different eukaryotic species

	Photosynthetic species							Nonphotosynthetic species				
Family	CME	OTA	CRE	PPA	OSAJ	ATH	PTR	HSA	DME	CEL	SCE	GLA
ABI3VP1			*1*	*30*	*59*	*59*	*81*					
Alfin-like				*7*	*12*	*8*	*9*					
AP2-EREBP		*8*	*11*	*150*	*174*	*160*	*206*					
ARF				*13*	*42*	*32*	*36*					
ARR-B		*1*	*1*	*5*	*9*	*15*	*15*					
BBR/BPC					*3*	*10*	*15*					
BES1				*6*	*6*	*10*	*12*					
bHLH	1	1	4	100	175	160	159	154	65	50	7	
bZIP	3	7	7	37	113	93	84	59	32	39	12	1
C2C2-CO-like		*3*	*1*	*11*	*21*	*19*	*14*					
C2C2-Dof		*2*	*1*	*20*	*33*	*42*	*41*					
C2C2-GATA	*6*	*4*	*6*	*12*	*37*	*30*	*36*	*15*	*7*	*14*	*10*	
C2C2-YABBY					*13*	*7*	*13*					
C2H2	7	4	5	56	103	104	113	644	312	150	39	5
CAMTA				1	7	6	7	2	1	2		
CCAAT	6	8	8	27	58	53	59	25	11	12	10	3
CPP	2	2	1	6	16	9	12	4	2	2		
CSD		4	1	3	3	4	7	16	4	5		
DBP					*7*	*5*	*10*					
E2F-DP	5	3	6	10	12	11	10	18	3	7		1
EIL				*2*	*7*	*6*	*6*					
FHA	2	7	12	15	19	17	18	44	22	12	14	2
G_2-like[a]	*1*	*2*	*4*	*41*	*52*	*48*	*66*					*1*
GeBP					*6*	*20*	*7*					
GRAS				*39*	*56*	*35*	*97*					
GRF				*2*	*17*	*9*	*9*					
HB	5	6	1	42	124	97	129	299	114	107	7	
HRT				*7*	*1*	*2*	*1*					
HSF	3	1	2	8	36	23	31	6	1	1	5	
LFY				2	1	1	1					
LIM	1	1		3	7	6	13	ND	ND	ND	ND	ND
MADS	1	1	2	22	82	122	108	9	3	2	4	
MYB	11	10	11	61	129	161	210	19	5	7	3	2
MYB-related	21	17	14	44	99	90	100	36	16	12	12	2
NAC				*32*	*140*	*115*	*163*					
NOZZLE						*1*						
PBF-2 like/Whirly		*1*	*1*		*3*	*4*	*3*					
PLATZ	*1*	*1*	*3*	*13*	*18*	*13*	*20*					
Pseudo ARR-B		*1*	*2*	*2*	*7*	*5*	*7*					
RWP-RK[b]	*1*	*4*	*14*	*8*	*13*	*14*	*18*					
S1Fa-like				*1*	*2*	*3*	*2*					
SAP						*1*	*1*					
SBP			*21*	*13*	*21*	*17*	*29*					
Sigma70-like[c]	4	1	1	5	9	6	9					
SRS				*2*	*5*	*11*	*10*					
TAZ			2	5	9	9	7	4	1	6		
TCP				6	22	26	33					
Trihelix				25	24	27	43	8	1			
TUB		1	3	6	17	12	11	6	3	2		
ULT					*2*	*2*	*2*					
VOZ				*2*	*2*	*2*	*4*					
WRKY[a]		*2*	*1*	*37*	*114*	*84*	*101*					*1*
zf-HD				*7*	*15*	*17*	*22*					
ZIM				*12*	*22*	*22*	*16*					
ARID[d]	*4*	*1*	*2*	*7*	*6*	*10*	*13*	*21*	*6*	*5*	*2*	
AUX/IAA[d]				*2*	*43*	*34*	*32*					

(*continued*)

TABLE 1

(Continued)

Family	Photosynthetic species							Nonphotosynthetic species				
	CME	OTA	CRE	PPA	OSAJ	ATH	PTR	HSA	DME	CEL	SCE	GLA
C3H[d]	7	18	15	44	97	75	96	85	33	40	7	5
DDT[d]	1		1	2	7	5	5	5	3	2	2	
HMG[d]		5	7	8	17	19	12	90	26	22	7	3
Jumonji[d]	3	5	7	10	17	19	20	38	11	15	3	
LUG[d]				1	12	3	5	6	1	1	1	
MBF1[d]			1	3	4	3	3	1	1	1	1	
PHD[d]	7	11	12	50	55	53	70	118	44	23	14	2
RB[d]	1	1	1	2	4	1	1	4	2	1		
SET[d]	6	10	22	26	32	38	44	49	17	29	7	3
SNF2[d]	13	20	19	35	44	43	48	48	22	23	17	6

Plant-specific TF and TR families are in italics; all other families are in roman. We also highlight TF families in italics that, in addition to plants, have members in early branching eukaryotes. Numbers represent distinct protein sequences. CME, *C. merolae*; OTA, *O. tauri*; CRE, *C. reinhardtii*; PPA, *P. patens*; OSAJ, *O. sativa* ssp. *japonica*; ATH, *A. thaliana*; PTR, *P. trichocarpa*; HAS, *H. sapiens*; DME, *D. melanogaster*; CEL, *C. elegans*; SCE, *S. cerevisiae*; GLA, *G. lamblia*. ND, not determined.

[a] Present in *G. lamblia*.

[b] Present in *D. discoideum* and *E. histolytica* (according to PFAM website).

[c] Present in bacteria.

[d] Transcription regulators (TRs).

protozoan that arose early in eukaryote evolution. In general, the number of TFs and TRs increases with the number of genes in the genome, following a power law as observed before (Van Nimwegen 2003). TFs and TRs were found to be similarly abundant in algae and yeast; however, in these lineages they are considerably less frequent than in animals. Numbers of TFs and TRs in many cases were similar in mosses and animals, whereas gene numbers were often greater in angiosperms. In Chlamydomonas, we identified 147 putative TF and 87 putative TR coding sequences from 29 and 10 protein families, respectively, totaling 234 distinct proteins involved in the regulation of transcription (Table 1; protein sequences are available at http://plntfdb.bio.uni-potsdam.de/v2.0/index.php?sp_id=CRE). A schematic of the transcriptional regulatory proteins identified in Chlamydomonas, including their defining domains, is given in supplemental Figure 1. To date, however, the biological functions of only a small number of these proteins have been analyzed (supplemental Table 1).

Chlamydomonas transcription factors: In animals, TFs of the C2H2 and HB families play important roles in growth-related and development processes (Wu 2002) and body-plan formation (Deutsch and Mouchel-Vielh 2003). These two families are the largest in animals, with >100 members each in humans, Drosophila, and Caenorhabditis. In animals, HB TFs function as homeotic genes that control the formation and differentiation of different body parts (Garcia-Fernandez 2005; Negre and Ruiz 2007). In contrast, in plants homeotic functions are carried out by TFs of the MADS-box family (Irish 2003). Typically, angiosperms have ~80–120 MADS-box proteins, whereas such TFs are largely absent from animals (<10). Similarly, MADS-box TFs are present in only small numbers in Chlamydomonas (two genes) and in all other unicellular organisms. In contrast, in these organisms, members of the C2H2 family are slightly more abundant than members of the MADS-box family, with five, four, and seven genes, respectively, in the algae Chlamydomonas, Ostreococcus, and Cyanidioschyzon, and 39 members in Saccharomyces (Table 1). C2H2 TFs contain a zinc-finger domain. The recruitment of this domain for transcriptional regulation occurred in prokaryotes, and members of the Ros family may have been the origin of C2H2 in eukaryotes (Bouhouche *et al.* 2000). In general, TFs bearing a zinc-finger domain have significantly contributed to the evolution of eukaryotic organisms (Riechmann *et al.* 2000) either through gene duplication leading to an increased gene number or through the modulation of other domains present in these proteins, resulting in the formation of new families of TFs.

The acquisition of chloroplasts represents an important step in the evolutionary path that separated plants from animals and fungi. Evidently, new regulatory networks had to be established through evolution to achieve an optimal integration of photosynthetic functions with other cellular processes. TFs and TRs constitute important elements of such networks. Three families of putative TFs predate the split between rhodophyta (red algae) and chlorophyta, *i.e.*, G_2-like, PLATZ, and RWP-RK. These families appear to be of particular importance for the evolution of eukaryotic photosynthetic organisms, as they are the only plant-specific TFs (with perhaps the exception of G_2-like, which might also be present in Giardia; see above) that are present in both red and green algae. Both algal groups derived from the original

primary endosymbiotic event that led to the establishment of plastids (Reyes-Prieto *et al.* 2007). Nine additional families, *i.e.*, ABI3/VP1, AP2-EREBP, ARR-B, C2C2-CO-like, C2C2-Dof, PBF-2-like/Whirly, Pseudo ARR-B, SBP, and WRKY (Table 1), predate the split between green algae and streptophytes. Plant-specific TF families might have important roles in the control of light-dependent processes and related biochemical pathways such as those involved in sugar production or starch accumulation.

TFs of the G_2-like family (a distinct group within the GARP superfamily of TFs; Rossini *et al.* 2001) are present in all plants, including red and green algae, and in *G. lamblia*, suggesting a deep evolutionary origin, but they are not found in animals or fungi. G_2-like TFs regulate chloroplast development in diverse plant species (*e.g.*, Physcomitrella, Arabidopsis, and *Zea mays*) through a process that requires a close coordination between plastidial and nuclear genomes. More specifically, GOLDEN 2-like (GLK) TFs are required for correct stacking of thylakoids within chloroplasts, although it is not known in detail how they exert their function in this process. One possible model is that GLKs regulate the transcription of genes encoding thylakoid-stabilizing factor(s) (Yasumura *et al.* 2005). We did not detect the ortholog of *GLK* in the sequenced Chlamydomonas genome, which is consistent with the fact that chloroplast thylakoid stacking is less advanced in this alga as compared to bryophytes and angiosperms, as previously discussed (Yasumura *et al.* 2005). *PHOSPHORUS STARVATION RESPONSE1* (*PSR1*) from Chlamydomonas and its ortholog *PHOSPHATE STARVATION RESPONSE1* (*PHR1*) from Arabidopsis encode TFs that control cellular responses to phosphate deprivation (Wykoff *et al.* 1999; Rubio *et al.* 2001). Both proteins were originally thought to be members of the MYB TF family, but subsequently were placed within the GARP superfamily (G_2-like) (Fitter *et al.* 2002). PSR1 targets include genes encoding chloroplast-localized proteins involved in photosynthesis, regulation of gene expression, and other processes (Moseley *et al.* 2006). Recently, PSR1 has also been shown to control the accumulation of chloroplast RNA under phosphorous limitation through control of the expression of ribonuclease polynucleotide phosphorylase (Yehudai-Resheff *et al.* 2007). Whether the angiosperm ortholog exerts a similar function is currently unknown.

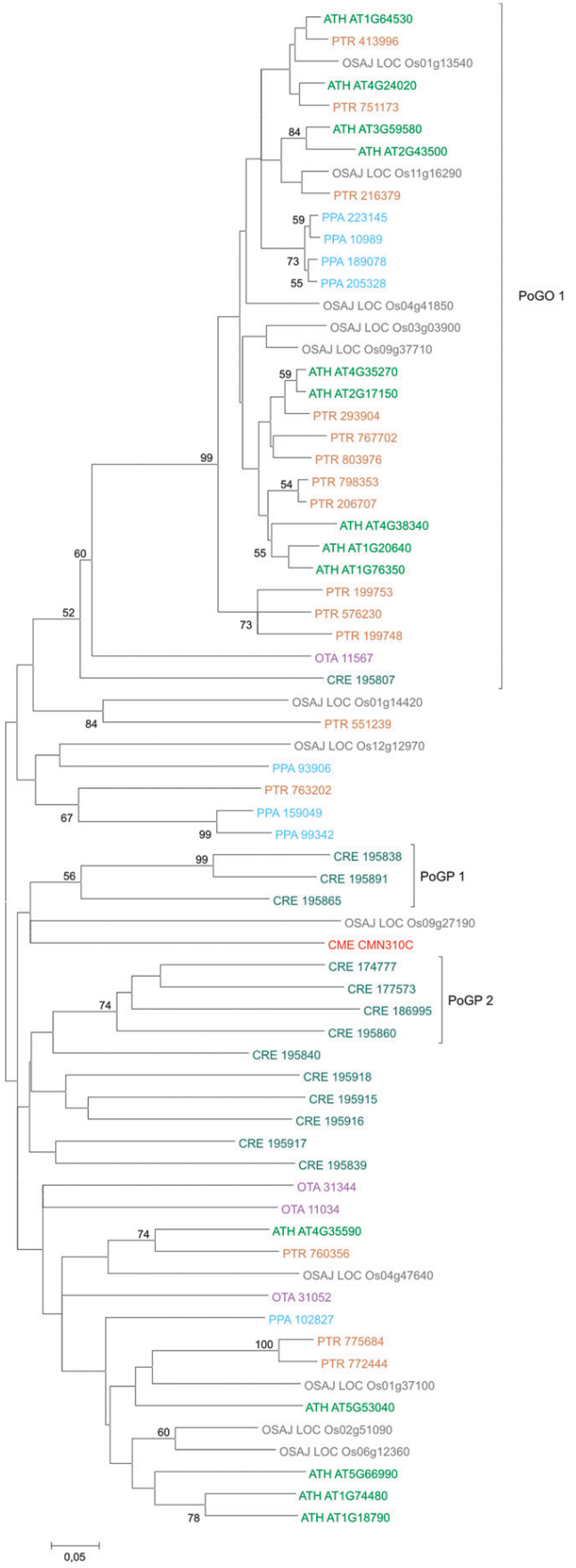
RWP-RK (Figure 1) is a TF family present in all green plants, as well as in red algae. It is also present in the early diverging amoebozoa *Dictyostelium discoideum* and *Entamoeba histolytica*, but not in animals or fungi, suggesting a deep evolutionary origin. In vascular plants, this family is involved in the regulation of genes in response to nitrogen status and nodule development in legumes (Schauser *et al.* 1995; Borisov *et al.* 2003). In Chlamydomonas, the gene *minus dominance* (*MID*; GenBank accession no. U92071; specific to mt^- strains and consequently not present in the sequenced strain, *i.e.*, CC-503 cw92 mt^+; Goodenough *et al.* 2007) is required for expression of *minus*-specific gamete-specific genes in response to nitrogen deprivation (Ferris and Goodenough 1997; Lin and Goodenough 2007). Another TF in this family, *NIT2*, is a positively acting regulatory gene of the nitrate assimilation pathway (Camargo *et al.* 2007) (GenBank accession no. DQ311647; this gene is mutated in the sequenced Chlamydomonas strain; Fernandez and Matagne 1984); the most similar entry in PlnTFDB is protein ID 195807).

Information regarding SBP TFs, Jumonji, and SET TRs, as well as microRNAs that often control TF genes in angiosperms but appear to be of minor importance in Chlamydomonas, is given in the supplemental text.

Transcription factors involved in hormone signaling: Phytohormones coordinate a vast spectrum of developmental and physiological processes in angiosperms. In contrast, knowledge about the occurrence of hormones in algae and their possible functions in cellular signaling is extremely limited. Some evidence indicates that auxins and cytokinins are present in algae (Tarakhovskaya *et al.* 2007), indicating their functional importance early in plant evolution. TF families known to participate in hormone signaling in angiosperms are also found in Chlamydomonas (Table 1). Recent work on ABSCISIC ACID INSENSITIVE 3 (ABI3) from Arabidopsis has indicated a possible role in cross talk of abscisic acid and auxin response pathways (Brady *et al.* 2003; Rock and Sun 2005). A similar observation was made for VP1 from maize, the ortholog of ABI3 (Suzuki *et al.* 2001). We observed a single *ABI3/VP1* gene in Chlamydomonas, whereas Physcomitrella has 30 *ABI3/VP1* genes, and angiosperms have ~60–80 (Table 1). To our knowledge the role of the *ABI3/VP1* gene in Chlamydomonas has not been characterized yet. TFs of the ARR-B and AP2-EREBP families are involved in cytokinin response pathways in angiosperms (Rashotte *et al.* 2006; Ishida *et al.* 2008). We detected one *ARR-B* gene and 11 *AP2-EREBP* genes in Chlamydomonas (Table 1). The role of these TFs has not been analyzed.

TF families absent from algae: Interestingly, 22 plant-specific TF families and 1 TR family are not present in algae (Table 1). These families may be related to the acquisition of multicellularity and tissue organization, invasion of the terrestrial environment, and long-distance trafficking. NAC TFs could be identified only from bryophytes onward. Functional studies have shown that several *NAC* genes play an important role in cell differentiation (Olsen *et al.* 2005). As we did not find any *NAC* gene in the *Volvox carteri* genome (not shown), we assume that TFs of this family were not important for establishing multicellularity in this organism.

Orthologs across green plants: The green plant lineage is a monophyletic group, its members having split from the red algal lineage ~1142 ± 167 million years ago (Zimmer *et al.* 2007). Tracing gene orthology relations across lineages provides a way to assess, to some



extent, the forces driving the functional diversification of multigene families. As reported previously, TF genes in plants have a higher retention rate after duplication than other genes (Seoighe and Gehring 2004; De Bodt *et al.* 2005). Additionally, genes functionally related to stress responses tend to undergo a more intense duplication process (Shiu *et al.* 2004). Therefore, TF gene families are well suited to trace back important events in evolution.

In our NJ analyses (for examples, see supplemental Figures 2–7), we have identified 120 clusters of orthologs with 1183 genes in total. Seventy-one of them are conserved in all green plants, and 26 are also common to red algae (see supplemental Table 2). Clusters to which functions could be assigned are involved mainly in light perception/response, control of plastidial gene expression, regulation of circadian rhythm, and the transition from the vegetative to the reproductive phase of growing plants (data not shown). Moreover, 38 of these clusters were found to have a one-to-one relationship (they do not possess any paralog inside the same group of orthologs). Such genes tend to exert key biological functions (Shiu *et al.* 2004). The greatest number of clusters, *i.e.*, 20, among all green plants is represented by the *SWI2/SNF2* gene family that encodes proteins involved in chromatin remodeling and thus the regulation of transcription, replication, and DNA recombination and repair. In plants, some Swi2/Snf2 proteins have been studied (Shaked *et al.* 2006), but a detailed functional analysis is missing for most of them. In the RWP-RK family, we found only one PoGO (Figure 1) in representatives from all green plants. The position of the *C. merolae* sequence is not evident from this analysis. In addition, groups of paralogs of Chlamydomonas are shown.

We also made a comparison between the clusters of orthologs obtained by phylogenetic analysis and by the InParanoid-Graph theoretic approach (see supplemental Table 3). In total, 446 genes from both classifications overlap, representing 99 of the 120 clusters obtained by the NJ analysis, and 98 of 168 clusters identified by the InParanoid approach (see supplemental Table 4). Thus, a large number of clusters was identified irrespective of the detection method used. We computed the Adjusted

Figure 1.—Phylogenetic tree of RWP-RK TFs in plants. We identified one PoGO (PoGO 1) conserved in all green plants, which includes the NIT2 TF (CRE 195807), a regulatory factor of genes involved in the nitrate assimilation pathway. Additionally, there are two possible groups of paralogs (PoGP 1 and PoGP 2) of Chlamydomonas genes. Red, *C. merolae* (CME); violet, *O. tauri* (OTA); light green, *C. reinhardtii* (CRE); light blue, *P. patens* (PPA); green, *A. thaliana* (ATH); brown, *P. trichocarpa* (PTR); gray, *O. sativa* ssp. *japonica* (OSAJ). The first three letters of the sequence name indicate the species (the first four letters in the case of OSAJ), and the remaining letters or numbers represent the accession code through which the respective sequence can be retrieved from the PlnTFDB (http://plntfdb.bio.uni-potsdam.de/v2.0).

Rand Index (*ar*) on the subset of common genes. The obtained value ($ar = 0.912$) indicates that the composition of the common clusters (gene membership) obtained by both methods is similar.

Evolution of photosynthetic networks: A recent review by Jiao *et al.* (2007) provides a good backbone for comparison of the light-regulated transcriptional networks of angiosperms and Chlamydomonas. The perception of light signals in dicots occurs through three cryptochromes and two phototropins, for which we found orthologs in Chlamydomonas (see supplemental Table 5). In contrast, phytochromes involved in the absorption of red and far-red light do not have homologs in green algae, consistent with previous findings (Mittag *et al.* 2005). One putative ortholog of the angiosperm bZIP protein COMMON PLANT REGULATORY FACTORS 1 (CPRF1), CMJ034C, was found in the red alga Cyanidioschyzon, although with a low InParanoid confidence score. The same Cyanidioschyzon protein is also orthologous to G-BOX BINDING FACTOR 1 (GBF1), suggesting subfunctionalization of the original multifunctional algal gene during angiosperm evolution; however, more detailed analyses are required to substantiate this hypothesis. GBF1 is phosphorylated by CASEIN KINASE II (CKII), which allows it to bind to target promoters containing the G-box, a well-defined light-response element. We found a putative CKII ortholog in Chlamydomonas, suggesting that light-dependent post-translational protein modification of the GBF1 ortholog was established early in plant evolution. Another important regulatory mechanism is the ubiquitin-mediated degradation of the bZIP TF ELONGATED HYPOCOTYL 5 (HY5) that is triggered by CONSTITUTIVE PHOTOMORPHOGENIC 1 (COP1) and its associated protein PHYTOCHROME A SUPPRESSOR 1 (SPA1). Both *HY5* and *SPA1* orthologs were found in green algae (see supplemental Table 5), whereas *COP1* has so far been found only in red algae. As the bZIP degradation mechanism triggered by COP1 is conserved throughout the plant kingdom (Yi and Deng 2005), one might speculate that COP1 is also present in Chlamydomonas. In summary, most of the components of the light-regulated transcriptional networks are shared between Chlamydomonas and seed plants, although phytochromes are missing in green algae. PHYTOCHROME INTERACTING FACTOR (PIF) TFs represent a subgroup of the bHLH family in angiosperms. No *PIF* ortholog could be found in Chlamydomonas. In addition, most other factors that are directly involved in phytochrome activity in land plants, such as FHY3, FAR1, HFR1, ATHB4, and PAR1, are absent from Chlamydomonas (see supplemental Table 5).

Conclusions: We have identified 147 putative TFs and 87 putative TRs in Chlamydomonas. Three TF families predate the rhodophyta–viridiplantae divide, while nine more of the TF families predate the chlorophyta–streptophyta divide and diversified further in bryophytes and angiosperms. However, we also observed that 22 plant-specific TF and 1 plant-specific TR family were not present in algae, highlighting their importance for the evolution of multicellular plants. Many of the elements of light-regulated transcriptional networks known from bryophytes and angiosperms are also present in Chlamydomonas, indicating an early evolutionary origin. Exceptions are elements of the phytochrome-mediated signaling pathways that are missing in algae. Our analysis provides a basis for further experimental studies on Chlamydomonas transcriptional regulators.

We thank the three anonymous reviewers for comments that helped to improve our manuscript. We are grateful to the Department of Energy Joint Genome Institute and the Chlamydomonas research community for sequencing and annotating the Chlamydomonas genome. L.G.G.C. and B.M.-R. thank the Interdisciplinary Research Centre, Advanced Protein Technologies, of the University of Potsdam and the International Ph.D. Programme, Integrative Plant Science [supported by the Deutscher Akademischer Austauschdienst and the Deutsche Forschungsgemeinschaft (DAAD), no. DAAD D/04/01336] for financial support. L.G.G.C. thanks the DAAD for providing a scholarship (no. A/04/34814). B.M.-R. thanks the Fonds der Chemischen Industrie for financial support (no. 0164389). D.M.R.-P. acknowledges financial support by the Bundesministerium fuer Bildung und Forschung (BMBF) (GABI-future grant 0315046). B.M.-R. and R.T.-E. thank the BMBF for funding of the systems biology research unit GoFORSYS–Potsdam-Golm BMBF Forschungseinrichtung zur Systembiologie [(Photosynthesis and Growth: A Systems Biology Based Approach (FKZ 0313924)].

LITERATURE CITED

Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne *et al.*, 2000 The genome sequence of *Drosophila melanogaster*. Science **287:** 2185–2195.

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

Best, A. A., H. G. Morrison, A. G. McArthur, M. L. Sogin and G. J. Olsen, 2004 Evolution of eukaryotic transcription: insights from the genome of *Giardia lamblia*. Genome Res. **14:** 1537–1547.

Borisov, A. Y., L. H. Madsen, V. E. Tsyganov, Y. Umehara, V. A. Voroshilova *et al.*, 2003 The Sym35 gene required for root nodule development in pea is an ortholog of Nin from *Lotus japonicus*. Plant Physiol. **131:** 1009–1017.

Bouhouche, N., M. Syvanen and C. I. Kado, 2000 The origin of prokaryotic C2H2 zinc finger regulators. Trends Microbiol. **8:** 77–81.

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.

Camargo, A., A. Llamas, R. A. Schnell, J. J. Higuera, D. Gonzalez-Ballester *et al.*, 2007 Nitrate signaling by the regulatory gene NIT2 in Chlamydomonas. Plant Cell **19:** 3491–3503.

C. elegans Sequencing Consortium, 1998 Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science **282:** 2012–2018.

Chen, W., N. J. Provart, J. Glazebrook, F. Katagiri, H. S. Chang *et al.*, 2002 Expression profile matrix of Arabidopsis transcription factor genes suggests their putative functions in response to environmental stresses. Plant Cell **14:** 559–574.

De Bodt, S., S. Maere and Y. Van de Peer, 2005 Genome duplication and the origin of angiosperms. Trends Ecol. Evol. **20:** 591–597.

de Nooy, W., A. Mrvar and A. Batagelj, 2005 *Exploratory Social Network Analysis With Pajek (Structural Analysis in the Social Sciences)*. Cambridge University Press, New York.

Derelle, E., C. Ferraz, S. Rombauts, P. Rouze, A. Z. Worden *et al.*, 2006 Genome analysis of the smallest free-living eukaryote

Ostreococcus tauri unveils many unique features. Proc. Natl. Acad. Sci. USA **103:** 11647–11652.

Deutsch, J. S., and E. Mouchel-Vielh, 2003 Hox genes and the crustacean body plan. BioEssays **25:** 878–887.

Fernandez, E., and R. F. Matagne, 1984 Genetic analysis of nitrate reductase-deficient mutants in *Chlamydomonas reinhardtii*. Curr. Genet. **8:** 635–640.

Ferris, P. J., and U. W. Goodenough, 1997 Mating type in Chlamydomonas is specified by mid, the minus-dominance gene. Genetics **146:** 859–869.

Finn, R. D., J. Mistry, B. Schuster-Bockler, S. Griffiths-Jones, V. Hollich *et al.*, 2006 Pfam: clans, web tools and services. Nucleic Acids Res. **34:** D247–D251.

Fitter, D. W., D. J. Martin, M. J. Copley, R. W. Scotland and J. A. Langdale, 2002 GLK gene pairs regulate chloroplast development in diverse plant species. Plant J. **31:** 713–727.

Garcia-Fernandez, J., 2005 The genesis and evolution of homeobox gene clusters. Nat. Rev. Genet. **6:** 881–892.

Goff, S. A., D. Ricke, T. H. Lan, G. Presting, R. Wang *et al.*, 2002 A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science **296:** 92–100.

Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon *et al.* 1996 Life with 6000 genes. Science **274:** 546, 563–567.

Goodenough, U., H. Lin and J. H. Lee, 2007 Sex determination in Chlamydomonas. Semin. Cell Dev. Biol. **18:** 350–361.

Hubert, L., and P. Arabie, 1985 Comparing partitions. J. Classification **2:** 193–218.

International Human Genome Sequencing Consortium, 2004 Finishing the euchromatic sequence of the human genome. Nature **431:** 931–945.

Irish, V. F., 2003 The evolution of floral homeotic gene function. BioEssays **25:** 637–646.

Ishida, K., T. Yamashino, A. Yokoyama and T. Mizuno, 2008 Three type-B response regulators, ARR1, ARR10 and ARR12, play essential but redundant roles in cytokinin signal transduction throughout the life cycle of *Arabidopsis thaliana*. Plant Cell Physiol. **49:** 47–57.

Jaillon, O., J. M. Aury, B. Noel, A. Policriti, C. Clepet *et al.*, 2007 The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature **449:** 463–467.

Jiao, Y., O. S. Lau and X. W. Deng, 2007 Light-regulated transcriptional networks in higher plants. Nat. Rev. Genet. **8:** 217–230.

Kersey, P., L. Bower, L. Morris, A. Horne, R. Petryszak *et al.*, 2005 Integr8 and Genome Reviews: integrated views of complete genomes and proteomes. Nucleic Acids Res. **33:** D297–D302.

Kumar, S., K. Tamura and M. Nei, 2004 MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. **5:** 150–163.

Lemon, B., and R. Tjian, 2000 Orchestrated response: a symphony of transcription factors for gene control. Genes Dev. **14:** 2551–2569.

Li, L., X. Wang, R. Sasidharan, V. Stolc, W. Deng *et al.*, 2007 Global identification and characterization of transcriptionally active regions in the rice genome. PLoS ONE **2:** e294.

Lin, H., and U. W. Goodenough, 2007 Gametogenesis in the *Chlamydomonas reinhardtii* minus mating type is controlled by two genes, MID and MTD1. Genetics **176:** 913–925.

Merchant, S. S., S. E. Prochnik, O. Vallon, E. H. Harris, S. J. Karpowicz *et al.*, 2007 The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science **318:** 245–250.

Mittag, M., S. Kiaulehn and C. H. Johnson, 2005 The circadian clock in *Chlamydomonas reinhardtii*. What is it for? What is it similar to? Plant Physiol. **137:** 399–409.

Moseley, J. L., C. W. Chang and A. R. Grossman, 2006 Genome-based approaches to understanding phosphorus deprivation responses and PSR1 control in *Chlamydomonas reinhardtii*. Eukaryot. Cell **5:** 26–44.

Negre, B., and A. Ruiz, 2007 HOM-C evolution in Drosophila: Is there a need for Hox gene clustering? Trends Genet. **23:** 55–59.

Nei, M., and S. Kumar, 2000 *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.

Nozaki, H., H. Takano, O. Misumi, K. Terasawa, M. Matsuzaki *et al.*, 2007 A 100%-complete sequence reveals unusually simple genomic features in the hot-spring red alga *Cyanidioschyzon merolae*. BMC Biol. **5:** 28.

Olsen, A. N., H. A. Ernst, L. L. Leggio and K. Skriver, 2005 NAC transcription factors: structurally distinct, functionally diverse. Trends Plant Sci. **10:** 79–87.

Rand, M. W., 1971 Objective criteria for the evaluation of clustering methods. J. Am. Stat. Assoc. **66:** 846–850.

Rashotte, A. M., M. G. Mason, C. E. Hutchison, F. J. Ferreira, G. E. Schaller *et al.*, 2006 A subset of Arabidopsis AP2 transcription factors mediates cytokinin responses in concert with a two-component pathway. Proc. Natl. Acad. Sci. USA **103:** 11081–11085.

R Development Core Team, 2007 *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Remm, M., C. E. Storm and E. L. Sonnhammer, 2001 Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. J. Mol. Biol. **314:** 1041–1052.

Rensing, S. A., D. Lang, A. D. Zimmer, A. Terry, A. Salamov *et al.*, 2008 The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science **319:** 64–69.

Reyes-Prieto, A., A. P. Weber and D. Bhattacharya, 2007 The origin and establishment of the plastid in algae and plants. Annu. Rev. Genet. **41:** 147–168.

Riano-Pachon, D. M., S. Ruzicic, I. Dreyer and B. Mueller-Roeber, 2007 PlnTFDB: an integrative plant transcription factor database. BMC Bioinformatics **8:** 42.

Riechmann, J. L., J. Heard, G. Martin, L. Reuber, C. Jiang *et al.*, 2000 Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science **290:** 2105–2110.

Rock, C. D., and X. Sun, 2005 Crosstalk between ABA and auxin signaling pathways in roots of *Arabidopsis thaliana* (L.) Heynh. Planta **222:** 98–106.

Rossini, L., L. Cribb, D. J. Martin and J. A. Langdale, 2001 The maize golden2 gene defines a novel class of transcriptional regulators in plants. Plant Cell **13:** 1231–1244.

Rubio, V., F. Linhares, R. Solano, A. C. Martin, J. Iglesias *et al.*, 2001 A conserved MYB transcription factor involved in phosphate starvation signaling both in vascular plants and in unicellular algae. Genes Dev. **15:** 2122–2133.

Schauser, L., L. Christensen, S. Borg and C. Poulsen, 1995 PZF, a cDNA isolated from *Lotus japonicus* and soybean root nodule libraries, encodes a new plant member of the RING-finger family of zinc-binding proteins. Plant Physiol. **107:** 1457–1458.

Seoighe, C., and C. Gehring, 2004 Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* proteome. Trends Genet. **20:** 461–464.

Shaked, H., N. Avivi-Ragolsky and A. A. Levy, 2006 Involvement of the Arabidopsis SWI2/SNF2 chromatin remodeling gene family in DNA damage response and recombination. Genetics **173:** 985–994.

Shiu, S. H., W. M. Karlowski, R. Pan, Y. H. Tzeng, K. F. Mayer *et al.*, 2004 Comparative analysis of the receptor-like kinase family in Arabidopsis and rice. Plant Cell **16:** 1220–1234.

Suzuki, M., C. Y. Kao, S. Cocciolone and D. R. McCarty, 2001 Maize VP1 complements Arabidopsis abi3 and confers a novel ABA/auxin interaction in roots. Plant J. **28:** 409–418.

Tarakhovskaya, E. R., Y. I. Maslov and M. F. Shishova, 2007 Phytohormones in algae. Russ. J. Plant Physiol. **54:** 163–170.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin and D. G. Higgins, 1997 The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:** 4876–4882.

Tuskan, G. A., S. Difazio, S. Jansson, J. Bohlmann, I. Grigoriev *et al.*, 2006 The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science **313:** 1596–1604.

van Nimwegen, E., 2003 Scaling laws in the functional content of genomes. Trends Genet. **19:** 479–484.

Vincentz, M., C. Bandeira-Kobarg, L. Gauer, P. Schlogl and A. Leite, 2003 Evolutionary pattern of angiosperm bZIP factors homologous to the maize Opaque2 regulatory protein. J. Mol. Evol. **56:** 105–116.

Wu, L. C., 2002 ZAS: C2H2 zinc finger proteins involved in growth and development. Gene Expr. **10:** 137–152.

Wykoff, D. D., A. R. Grossman, D. P. Weeks, H. Usuda and K. Shimogawara, 1999 Psr1, a nuclear localized protein that regulates phosphorus metabolism in Chlamydomonas. Proc. Natl. Acad. Sci. USA **96:** 15336–15341.

Yasumura, Y., E. C. Moylan and J. A. Langdale, 2005 A conserved transcription factor mediates nuclear control of organelle biogenesis in anciently diverged land plants. Plant Cell **17:** 1894–1907.

Yehudai-Resheff, S., S. L. Zimmer, Y. Komine and D. B. Stern, 2007 Integration of chloroplast nucleic acid metabolism into the phosphate deprivation response in Chlamydomonas reinhardtii. Plant Cell **19:** 1023–1038.

Yi, C., and X. W. Deng, 2005 COP1: from plant photomorphogenesis to mammalian tumorigenesis. Trends Cell Biol. **15:** 618–625.

Yu, J., S. Hu, J. Wang, G. K. Wong, S. Li *et al.*, 2002 A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science **296:** 79–92.

Zimmer, A., D. Lang, S. Richardt, W. Frank, R. Reski *et al.*, 2007 Dating the early evolution of plants: detection and molecular clock analyses of orthologs. Mol. Genet. Genomics **278:** 393–402.

Communicating editor: S. Dutcher

3

bZIP transcription factors in plants

The Role of bZIP Transcription Factors in Green Plant Evolution: Adaptive Features Emerging From Four Founder Genes

Luiz Gustavo Guedes Corrêa[1,2,3,†], Diego Mauricio Riaño-Pachón[2,4,†], Carlos Guerra Schrago[5], Renato Vicentini dos Santos[1], Bernd Mueller-Roeber[2,3], Michel Vincentz[1]

† These authors contributed equally to this work and should thus both be considered as first authors.

[1] Centro de Biologia Molecular e Engenharia Genética, Departamento de Genética e Evolução, Instituto de Biologia, Universidade Estadual de Campinas, Campinas SP, Brazil, [2] Department of Molecular Biology, University of Potsdam, Potsdam–Golm, Germany, [3] Cooperative Research Group, Max Planck Institute of Molecular Plant Physiology, Potsdam–Golm, Germany, [4]Bioinformatics Research Group, GabiPD Team, Max-Planck Institute of Molecular Plant Physiology, Potsdam–Golm, Germany, [5] Laboratório de Biodiversidade Molecular, Departamento de Genética, Universidade Federal do Rio de Janeiro, Cidade Universitária, Rio de Janeiro RJ, Brazil

Published in *PLoS ONE* (2008) **3**(8):e2944. doi:10.1371/journal.pone.0002944

Author contributions

MV, LGGC and DMRP conceived and designed the study. BMR and MV coordinated the project. LGGC carried out phylogenetic analyses by NJ, made the comparison based on expression profiles, identified conserved motifs using MEME. DMRP carried out phylogenetic analyses by ML, made the comparison based on MPSS data, carried out bZIP searches on EST data collections, identified putative pseudogenes of bZIP TFs. DMRP and CGS perform the analysis about gene family expansions. All authors discussed and analysed the data.

OPEN ACCESS Freely available online

PLoS one

The Role of bZIP Transcription Factors in Green Plant Evolution: Adaptive Features Emerging from Four Founder Genes

Luiz Gustavo Guedes Corrêa[1,2,3☯], Diego Mauricio Riaño-Pachón[2,4☯], Carlos Guerra Schrago[5], Renato Vicentini dos Santos[1], Bernd Mueller-Roeber[2,3], Michel Vincentz[1]*

1 Centro de Biologia Molecular e Engenharia Genética, Departamento de Genética e Evolução, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, Brazil, **2** Department of Molecular Biology, University of Potsdam, Potsdam-Golm, Germany, **3** Cooperative Research Group, Max-Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany, **4** GabiPD Team, Bioinformatics Group, Max-Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany, **5** Laboratório de Biodiversidade Molecular, Departamento de Genética, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

Abstract

Background: Transcription factors of the basic leucine zipper (bZIP) family control important processes in all eukaryotes. In plants, bZIPs are regulators of many central developmental and physiological processes including photomorphogenesis, leaf and seed formation, energy homeostasis, and abiotic and biotic stress responses. Here we performed a comprehensive phylogenetic analysis of bZIP genes from algae, mosses, ferns, gymnosperms and angiosperms.

Methodology/Principal Findings: We identified 13 groups of bZIP homologues in angiosperms, three more than known before, that represent 34 Possible Groups of Orthologues (PoGOs). The 34 PoGOs may correspond to the complete set of ancestral angiosperm bZIP genes that participated in the diversification of flowering plants. Homologous genes dedicated to seed-related processes and ABA-mediated stress responses originated in the common ancestor of seed plants, and three groups of homologues emerged in the angiosperm lineage, of which one group plays a role in optimizing the use of energy.

Conclusions/Significance: Our data suggest that the ancestor of green plants possessed four bZIP genes functionally involved in oxidative stress and unfolded protein responses that are bZIP-mediated processes in all eukaryotes, but also in light-dependent regulations. The four founder genes amplified and diverged significantly, generating traits that benefited the colonization of new environments.

Citation: Guedes Corrêa LG, Riaño-Pachón DM, Guerra Schrago C, Vicentini dos Santos R, Mueller-Roeber B, et al. (2008) The Role of bZIP Transcription Factors in Green Plant Evolution: Adaptive Features Emerging from Four Founder Genes. PLoS ONE 3(8): e2944. doi:10.1371/journal.pone.0002944

Editor: Shin-Han Shiu, Michigan State University, United States of America

Received February 18, 2008; **Accepted** July 22, 2008; **Published** August 13, 2008

Funding: L.G.G.C. thanks the DAAD for providing a scholarship (A/04/34814). D.M.R.P. acknowledges financial support from the BMBF (FKZ 0315046). This work was supported in part by grants from the Fundação de Amparo a Ciência do Estado de São Paulo (FAPESP), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) (to L.G.G.C and M.V.), the University of Potsdam Interdisciplinary Research Centre 'Advanced Protein Technologies' (to B.M.-R.), the DAAD/DFG International PhD Programme 'Integrative Plant Science' (DAAD D/04/01336; to B.M-R.), and the Fonds der Chemischen Industrie (N° 0164389; to B.M.-R.).

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: mgavince@unicamp.br

☯ These authors contributed equally to this work.

Introduction

Growth and development of all organisms depend on proper regulation of gene expression. The control of transcription initiation rates by transcription factors (TF) represents one of the most important means of modulating gene expression [1–4]. TFs can be grouped into different protein families according to their primary and/or three-dimensional structure similarities in the DNA-binding and multimerization domains [4–6]. The interplay between the amplification of the ancestral repertoire of TFs, the emergence of new TFs, the combination of protein domains and sequence divergence constitutes an important driving force towards the evolution of organismic complexity [7–10]. Understanding the detailed evolutionary history of these TFs and their corresponding functions is therefore crucial to reveal the changes and/or innovations in transcriptional regulatory circuits that underlie the biological diversity found among eukaryotes.

Large scale genomic comparisons revealed that angiosperm TF families undergo more intense gene expansion when compared to animals and fungi, possibly reflecting the ability of flowering plants to efficiently adapt to different and unstable environmental conditions. Moreover, gene expansion rates vary among plant TF families, indicating lineage-differential specializations [11,12]. For instance, MADS-box and homeodomain families, which exert similar functions in developmental control, expanded preferentially in the angiosperm and human lineages, respectively [13,14]. Contrariwise, the basic leucine zipper (bZIP) TF family apparently expanded to a similar extent in angiosperms and humans [15]. Currently we do not well understand why individual TF families underwent differential evolutionary expansions in the different

eukaryotic lineages. Therefore, a deep evolutionary analysis of TF families including the identification of the founding (ancestral) gene sets in combination with functional assignments will greatly assist in addressing this issue [16,17].

To our knowledge, however, only four families that are present in all green plants have until today been studied in a deep evolutionary scale, Dof [18], homeodomain [19], MADS-box [20,21] and WRKY [22]. As a matter of fact, groups of orthologues, for which functional equivalence is often assumed, are rarely identified in a systematic and direct manner, with the exception of the HD-Zip class III subfamily [23,24]. It is thus often difficult to infer ancestral functions at different time points of the evolutionary process. Here we performed a comprehensive analysis of the evolutionary relationships of TFs of the green plant bZIP family; homologous and orthologous relationships among bZIP TFs were established and ancestral functions were inferred.

The bZIP TFs are characterized by a 40- to 80-amino-acid-long conserved domain (bZIP domain) that is composed of two motifs: a basic region responsible for specific binding of the TF to its target DNA, and a leucine zipper required for TF dimerization [5,25]. Genetic, molecular and biochemical analyses indicate that bZIPs are regulators of important plant processes such as organ and tissue differentiation [26–30], cell elongation [31,32], nitrogen/carbon balance control [33,34], pathogen defence [35–40], energy metabolism [41], unfolded protein response [42,43], hormone and sugar signalling [44–47], light response [48–50], osmotic control [34,51], and seed storage protein gene regulation [52]. Initially, 50 plant bZIP proteins were classified into five families, taking into account similarities of their bZIP domain [53]. An original investigation of the complete *Arabidopsis thaliana* genome sequence indicated the presence of 81 putative *bZIP* genes [54,55]. However, further detailed studies revealed 75 to 77 bZIP proteins to be encoded by the Arabidopsis nuclear genome, representing members of ten groups of homologues [55,56].

The availability of the rice (*Oryza sativa*) [57,58], black cottonwood (*Populus trichocarpa*) [59] and Arabidopsis genomic sequences [54] provides an exciting opportunity for the large-scale investigation of the genetic bases that underlies the extensive physiological and morphological diversity amongst the two main angiosperm divisions: monocots and eudicots. A possible comparative approach involves the establishment of relationships between different genomes in a homologous gene system [60–62], in which each group of orthologues is derived from an ancestral gene that underwent numerous modifications throughout evolution, including duplication and subsequent functional diversification. Considering that all genes of a given group of orthologues have the same ancestral origin, the establishment of this classification should allow the transfer of biochemical, structural and functional information from one protein to another, inside the same group [63]. Moreover, the relationships within a group of orthologues constitute the basis for a better understanding of the evolution of ancestral functions (conservation versus neo- or sub-functionalization through duplication) [64–66].

In this study, we identified the possible non-redundant complete sets of bZIPs in rice, comprising 92 proteins, and in black cottonwood, comprising 89 proteins. These collections of bZIPs together with the 77 bZIPs from Arabidopsis [56] could be divided, based on bZIP domain and other conserved motifs similarities, into 13 groups of bZIP homologues in angiosperms, three more than previously reported [55]. The identified groups constituted a backbone for a more detailed analysis of each group, to which additional bZIP sequences reported from other plants, including those deduced from expressed sequence tags (ESTs), were added. In total, we defined 34 Possible Groups of Orthologues (PoGOs), which may represent 34 ancestral functions in angiosperms. Interestingly, one PoGO was found exclusively in monocots, whereas a Possible Group of Paralogues (PoGP) appears to be restricted to Arabidopsis.

To extend our bZIP analysis to all major lineages of green plants we additionally identified and incorporated bZIP sequences not only from two algal (*Chlamydomonas reinhardtii* [67] and *Ostreococcus tauri* [68]) and moss (*Physcomitrella patens* [69]) genomes, but also from ESTs of the ferns *Selaginella moellendorffii* and *Adiantum capillus-veneris* and the gymnosperms *Pinus taeda* and *Picea glauca*. Based on this investigation, a model for the evolution of *bZIP* genes in green plants, based on four founder genes representing an ancestral tool kit, was established. Its main points are discussed here. We also propose an updated classification of plant *bZIP* genes which should facilitate functional studies.

Results and Discussion

Groups of Homologues of Angiosperm *bZIP* Genes

The Arabidopsis genome encodes for a possible complete set of 77 unique bZIP proteins, representing an update of previous results [55,56,70]. *AtbZIP73* contains a premature stop codon and was thus not considered further in our analyses. As it appears to be a pseudogene it should be referred to as *ΨAtbZIP73*. Through iterated searches with tblastn and blastx algorithms, and PFAM bZIP Hidden Markov Models (HMM), we identified 92 *bZIP* genes in rice (Text S1a). Recently, Nijhawan *et al.* [71] reported the presence of 89 *bZIP* genes in rice and their phylogenetic relationship to the Arabidopsis *bZIPs*. Of the 89 bZIPs, 86 are also present in this study. Careful sequence analyses of both gene sets revealed complete sequence identity of the Os06g50480 and Os06g50830 TFs, and complete identity with TF Os06g50600 (OsbZIP14) along amino acids 1–143, indicating that these sequences were redundant in the Nijhawan *et al.* data set. *Os03g59460* has also been identified in our studies, however, the protein it encodes contains a proline residue at the beginning of its leucine zipper, precluding dimerization [25]; thus it may not function like other known bZIPs. Despite *OsbZIP24* and *OsbZIP75* being classified as retrotransposons in TIGR, we included them in our analysis as they possess a standard bZIP sequence in their open reading frame. Table S1 gives a summary of this information.

We identified 89 bZIP sequences in *P. trichocarpa*, some of which were incomplete. We therefore performed a more refined analysis of genomic data sets taking into account gene structures and conserved motifs. This allowed us to resolve the entire *bZIP* gene sequences in nine cases (Datasets S1 and S2).

Through Neighbour-Joining (NJ) analysis of the minimum bZIP domain (44 amino acids; Text S1a) of 257 unique bZIPs from Arabidopsis, rice and black cottonwood (bZARP data set) we identified seven clusters of proteins with bootstrap support greater than or equal to 50%, defining the groups of homologous genes B, D, F, G, H, J and K. The topology of the phylogenetic tree and a bootstrap support of 50% indicate that Groups D and F are sister groups that share a common ancestor (Figures 1A and S1). Although Group A has a weaker bootstrap support in NJ analyses (34% using PAM matrix data, and 58% using p-distance values), its members were kept together for two main reasons: (i) all its member genes share a common motif in accordance with previous results from Jakoby *et al.* [55]; (ii) all genes but *Gbf4* (*AtbZIP40*) and *AtbZIP13* from Arabidopsis share common intron positions, suggesting a single evolutionary origin (Text S1b, and Figure S2). In Group F a clear tendency for loss of introns was observed. None of the rice *bZIP* genes contains introns, nor do the black cottonwood genes *PtrbZIP39* and *PtrbZIP40*. Although *PtrbZIP38* and *PtrbZIP41* have introns, they lost it from the conserved basic

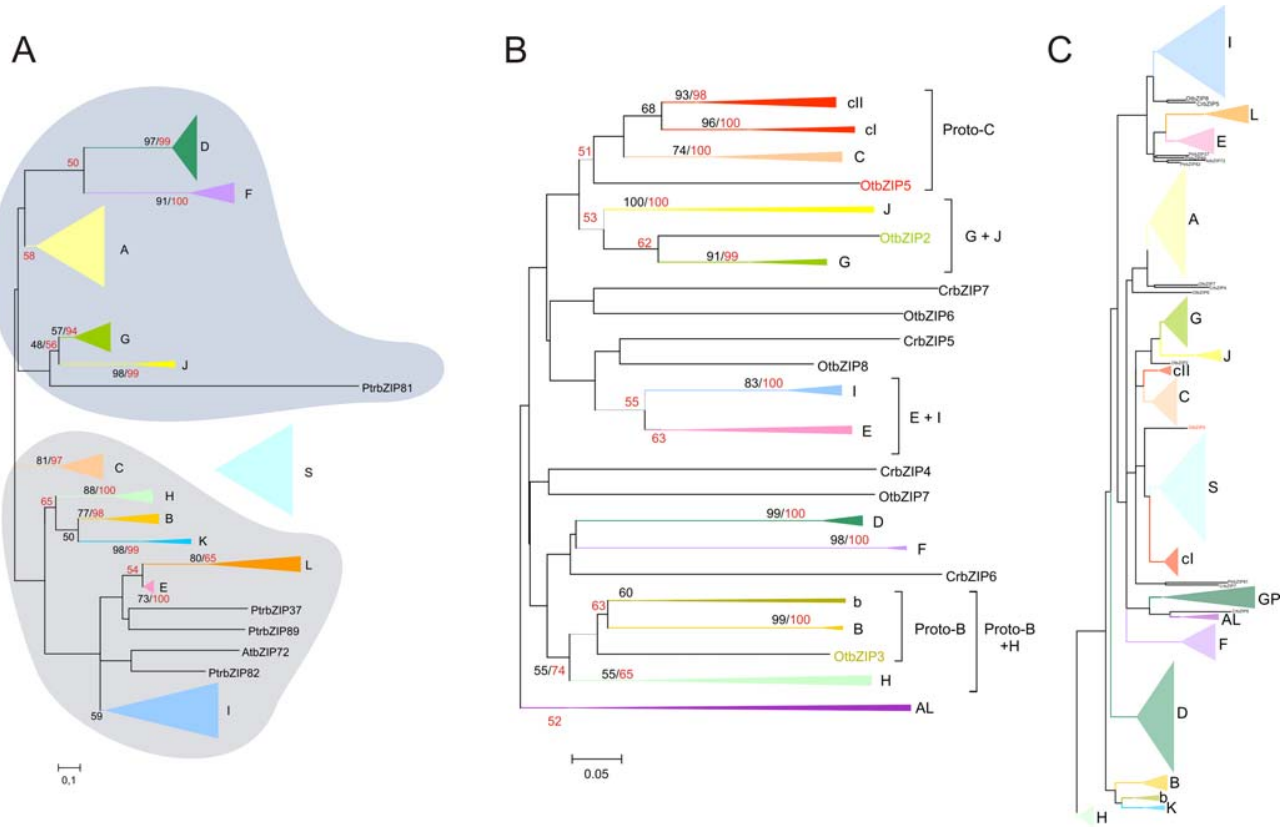


Figure 1. Phylogeny of bZIP transcription factors in green plants. (A) Model of angiosperm bZIP evolution with two large clades, one including groups A, D, F, G and J, and the other including groups B, C, E, H, I and L. Sister groups B and K, E and L, D and F, and G and J, respectively, were defined based on bootstrap support of >50%. The position of Group S could not be clearly defined. (B) Consensus tree inferred from NJ analyses of bryophyte and algal bZIP sequences. This tree reveals new evolutionary relationships among green plant bZIPs, which were not observed when the complete ViridiZIP set was analyzed. Group C appears to be related to two other groups (cI and cII) and members of these three groups are orthologues of *OtbZIP5*, constituting the Group Proto-C. Group b was identified as a sister group of Group B and genes of both groups are orthologous to the algal *OtbZIP3* gene, forming the Group Proto-B. Groups Proto-B and H have a common ancestral origin. Similarly, Groups G and J diverged from the same ancestor and are both orthologous to the algal gene *OtbZIP2*. Finally, Groups E and I show a sisterhood relation but no ancestral link to a bZIP from algae could be established. (C) Tree inferred from NJ analyses of the ViridiZIP data set (bZIPs from algae to angiosperms). This tree indicates that Group S probably originated from Proto-C, and Group K from Proto-B. Tree topology and functional data support these hypotheses. Bootstrap values were calculated from NJ analyses. Red, values obtained with p-distances and, black, with PAM matrix.
doi:10.1371/journal.pone.0002944.g001

motif. The only gene that possesses an intron in this motif is *AtbZIP24* from Arabidopsis.

Members of Groups A and D have a bZIP domain of only 44 amino acids. To refine our analysis we created a subset-of-bZARP (sbZARP) dataset that excluded groups A and D members but included all remaining 172 proteins with a bZIP domain of 60 amino acids (53, 60 and 59 bZIPs from Arabidopsis, rice and black cottonwood, respectively). NJ analyses revealed four new groups of homologues, Groups C, E, I and L, all supported by bootstrap values of >50% (Figure S3; note that Group L members harbor an atypical basic motif; see Figure S2, and Text S1c). The overall organization into twelve groups is further supported by the presence of at least one shared intron position among the members of each group, confirming a common ancestral origin of all its members (Figures 1A, 2 and S2). The twelve groups encompass 199 of the 257 bZIPs of the bZARP data set. Fifty-three of the remaining bZIPs (17, 17 and 19 from Arabidopsis, rice, and black cottonwood, respectively) tended to form a separate group, defined as Group S in agreement with previous data [55]. However, this group did not have significant bootstrap support. Members of Group S bZIPs share two characteristics: they harbor a long leucine zipper (eight to nine heptads) and are encoded by intron-less genes. Finally, *AtbZIP72* (Arabidopsis) and *PtrbZIP37*, *81*, *82* and *89* (black cottonwood) could not be classified into any of the above groups (Figure 1A).

In summary, our data suggest 13 groups of homologous angiosperm *bZIP* genes (A, B, C, D, E, F, G, H, I, J, K, L, and S), representing a unified classification of angiosperm bZIPs (Figure 3) [55,56,71]. This result is in agreement with previous analyses, but additionally revealed three new groups (J, K and L) (Figure S3). The name of each group of homologues follows the classification established by Jakoby *et al.* [55]. Similar conclusions were reached using Maximum Likelihood analyses.

Possible Groups of Orthologues (PoGOs) in Angiosperms

We next aimed at identifying Possible Groups of Orthologues (PoGOs) among the 13 groups of homologues. By definition, each PoGO represents a group of genes that diverged from an ancestral gene through speciation and duplication. Members of a given PoGO typically have closely related biological functions, and this allows making predictions for poorly characterized genes and rationalizes functional studies of the proteins they encode [72]. PoGOs also establish a basis for the definition of functional

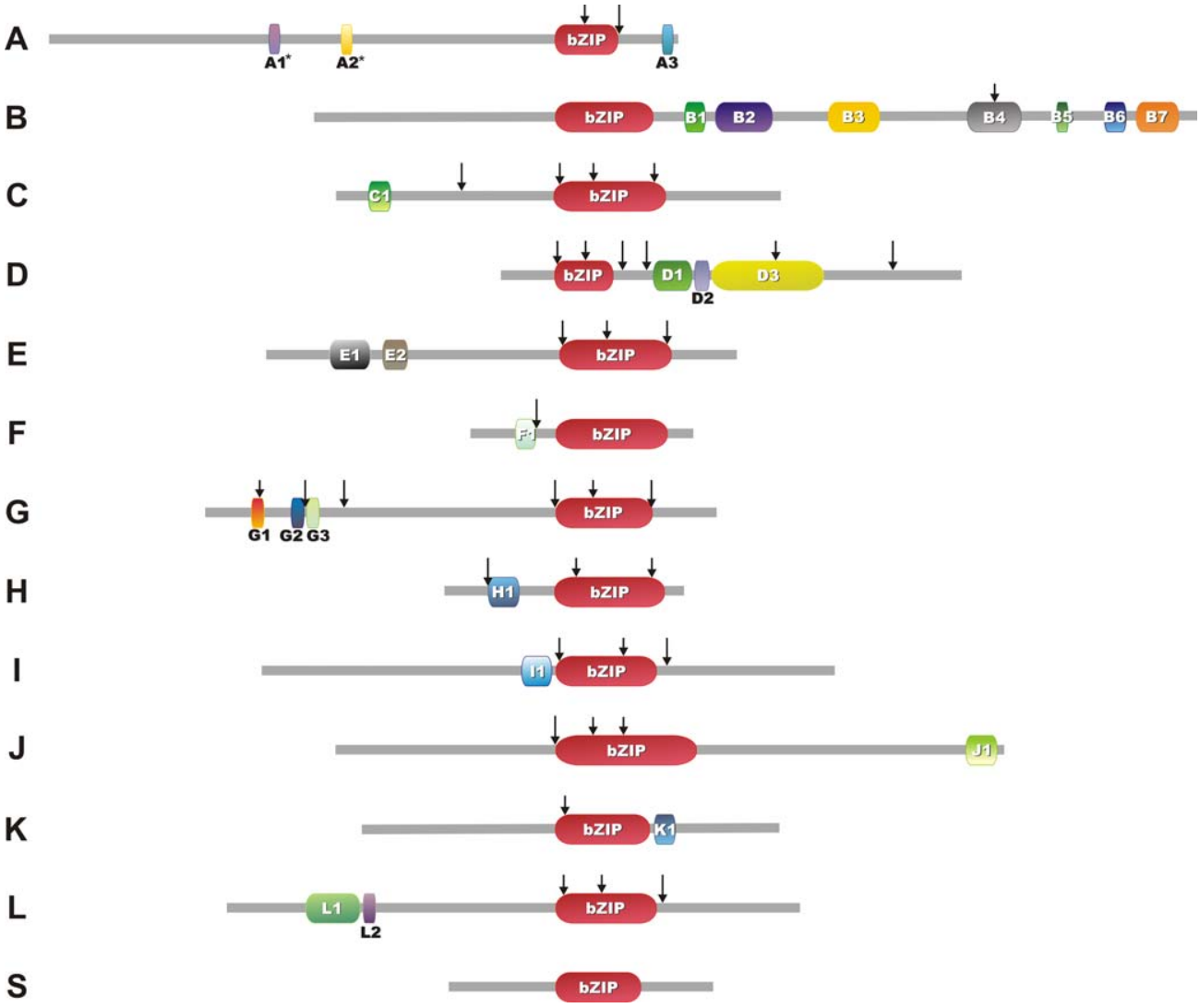


Figure 2. Motifs conserved in angiosperm bZIPs. A summary of the motif sequences is given in Table S2. Arrows indicate intron positions conserved among most members of each group. Representative bZIP sizes and positions of conserved motifs are shown. (*) Group A has two motifs (A1 and A2), that are important putative kinase phosphorylation sites involved in ABA responses. Both motifs appear to be conserved in most members of this group of homologues, except for OsbZIP8, 13, 14 and 15, and PtrbZIP5 and 10, which lack motif A1. The same sequences and also PtrbZIP9 lack motif A2. Due to the lack of complete sequences, no structures are shown for Groups AL, GP, b, cI and cII.
doi:10.1371/journal.pone.0002944.g002

diversification among genes. Here, we identified PoGOs by NJ analysis of each group of homologues separately, using the criteria defined in Material and Methods. To optimize the resolution of the evolutionary relationships, alignment lengths were extended by including conserved motifs specific to each group of homologues (Figure 2, and Table S2). Additionally, 636 further bZIP sequences, 260 from eudicots and 376 from monocots (Table S3), were extracted from EST databases. These new bZIPs were included in the respective groups of homologous genes according to their tblastn best matches against members of an upgraded Angiotot dataset that contained the rice and black cottonwood bZIPs.

Our analysis revealed 31 PoGOs distributed among Groups A to L (Figures 3 and S4, S5, S6, S7, S8, S9, S10, S11, S12, S13, S14 and S15). In all PoGOs except D2, at least one black cottonwood bZIP sequence could be included (Figure 3) further supporting the organization into PoGOs. The lack of a black cottonwood *bZIP* gene in PoGO D2 could be due to an absence of such a gene in its genome or to incomplete genome sequence availability. OsbZIP24, PtrbZIP86, 87 and 88 lack some of the motifs conserved in Group D members and were therefore assigned to the PoGO to which they showed the highest overall sequence similarity (as identified through blastp analysis).

We identified only one eudicot-monocot PoGO, S1, in Group S (Figure S16). The remaining sequences could be clustered into three PoGOs each restricted to either eudicots (SE1, SE2 and SE3) or monocots (SM1, SM2 and SM3) (Figure S16). Arabidopsis bZIP TFs of groups SE2 and SE3 are involved in energy metabolism and hypoosmolarity signaling (Table S4) further supporting the evolutionary relationship deduced from the phylogenetic analysis. Similarly, SM2 members play a role in cold signaling (Table S4), thus providing function-based support also for this group. Although further efforts to more precisely uncover the relationship between the three monocot (SM1, SM2 and SM3) and eudicot (SE1, SE2 and SE3) groups of orthologues

bZIP no.	Gene code	Synonym	GenBank	
OsbZIP1	Os03g20650			
OsbZIP2	Os07g48660		AK103188	
OsbZIP3	Os01g59760	DPBF3		
OsbZIP4	Os05g41070	AREB3os	AK063398	
AtbZIP12	At2g41070	DPBF4	AF334209	
AtbZIP66	At3g56850	AREB3	AB017162	
PtrbZIP8	754658			PoGO A1
PtrbZIP13	549022			
PtrbZIP14	558204			
PtrbZIP15	560286			
PtrbZIP16	808328			
PtrbZIP17	770717			
PtrbZIP18	803082			
PtrbZIP19	594286			
OsbZIP8	Os09g36910			
OsbZIP13	Os02g58670		AK061086	
OsbZIP14	Os06g50600		AK108991	
OsbZIP15	Os08g43600			
AtbZIP14	At4g35900	FD	BN000021	PoGO A2
AtbZIP15	At5g42910		AJ419599	
AtbZIP27	At2g17770	FDP	BN000022	
PtrbZIP5	550249			
PtrbZIP9	818828			
PtrbZIP10	642918			
OsbZIP5	Os01g64730	OSE2	AK067919	
OsbZIP6	Os05g36160	OSE2-like	AK120656	
AtbZIP13	At5g44080		BN000023	PoGO A3
AtbZIP40	At1g03970	GBF4	U01823	
PtrbZIP11	651568			
PtrbZIP12	754448			
OsbZIP7	Os01g64000	ABI5-2	AK070998	
AtbZIP39	At2g36270	ABI5	AF334206	
AtbZIP67	At3g44460	DPBF2	AJ419600	PoGO A4
PtrbZIP6	767006			
PtrbZIP7	801922			
OsbZIP9	Os09g28310	ABI5os	AK065873	
OsbZIP10	Os08g36790	TRAB1	NM001068553	
OsbZIP11	Os02g52780		AK072062	
OsbZIP12	Os06g10880		AK103188	
AtbZIP35	At1g49720	ABF1	AF093544	
AtbZIP36	At1g45249	ABF2	AF093545	PoGO A5
AtbZIP37	At4g34000	ABF3	AF093546	
AtbZIP38	At3g19290	ABF4	AF093547	
PtrbZIP1	551849			
PtrbZIP2	677861			
PtrbZIP3	267872			
PtrbZIP4	767577			
OsbZIP16	Os07g44950		AK121898	
OsbZIP17	Os05g34050		AK073142	
AtbZIP17	At2g40950		AV441374	PoGO B1
AtbZIP28	At3g10800		AJ419850	
AtbZIP49	At3g56660		AJ419851	
PtrbZIP20	255215			
OsbZIP22	Os03g58250	REB	AB021736	
OsbZIP23	Os07g08420	RISBZ1	AB053472	
AtbZIP63	At5g28770	BZO2H3		PoGO C1
PtrbZIP24	294737			
PtrbZIP25	729825			
OsbZIP18	Os12g40920	RBZO2H		
AtbZIP10	At4g02640	BZO2H1		
AtbZIP25	At3g54620	BZO2H4		PoGO C2
PtrbZIP22	551106			
PtrbZIP23	559630			
OsbZIP19	Os02g07840	RISBZ4	AB053473	
OsbZIP20	Os02g16680	RITA1	L34551	
OsbZIP21	Os06g45140	RISBZ5	AB053474	PoGO C3
AtbZIP9	At5g24800	BZO2H2	AF310223	
PtrbZIP21	271607			
OsbZIP24*	Os02g22280		AK103347	
OsbZIP25	Os09g10840			
OsbZIP26	Os09g31390		AK103174	
OsbZIP27	Os06g41100			PoGO D1
OsbZIP28	Os02g10140			
AtbZIP65	At5g06839		AJ314787	
PtrbZIP32	272608			
PtrbZIP86*	255651			
OsbZIP29	Os01g64020		AK101903	
OsbZIP30	Os05g37170		AK109520	
OsbZIP35	Os11g05480		AK102690	PoGO D2
OsbZIP36	Os12g05680	TGA-2.1	AK101620	
AtbZIP21	At1g08320		AJ314757	
OsbZIP32	Os08g07970	STGA	AK107028	
OsbZIP37	Os04g54474		AK066906	
AtbZIP22	At1g22070	TGA3	L10209	
AtbZIP47	At5g65210	TGA1	X68053	
AtbZIP50	At1g77920	TGA7	AJ315736	PoGO D3
AtbZIP57	At5g10030	OBF4	X69899	
PtrbZIP26	207609			
PtrbZIP27	217692			
PtrbZIP28	716556			
PtrbZIP29	830210			

bZIP no.	Gene code	Synonym	GenBank	
OsbZIP31	Os06g15480		AK109719	
AtbZIP46	At1g68640	PAN	AF111711	PoGO D4
PtrbZIP30	564507			
PtrbZIP88*	769678			
OsbZIP33	Os03g20310	NIF1	AB051294	
OsbZIP34	Os07g48820	NIF2	AB051295	
OsbZIP38	Os01g59350	NIF4	AB051297	
OsbZIP39	Os05g41280			
OsbZIP40	Os01g17260	NIF3	AB051296	
AtbZIP20	At5g06950	TGA2	D10042	PoGO D5
AtbZIP26	At5g06960	HBP-1b	X69900	
AtbZIP45	At3g12250	TGA6	AJ320540	
PtrbZIP31	712010			
PtrbZIP33	825048			
PtrbZIP87*	652586			
OsbZIP41	Os01g55150		AK108553	
OsbZIP42	Os01g11350	RF-2b like	AK100944	
OsbZIP43	Os02g14910			
AtbZIP34	At2g42380		AY074657	PoGO E1
AtbZIP61	At3g58120		AF401300	
PtrbZIP34	582775			
PtrbZIP35	836130			
PtrbZIP36	176347			
OsbZIP49	Os01g58760			
OsbZIP50	Os05g41540		AK104986	
OsbZIP51	OsIFCC032062		BX000502	
OsbZIP52	Os11g04390		AK103113	PoGO F1
AtbZIP24	At3g51960		AI994442	
PtrbZIP40	554977			
PtrbZIP41	812021			
OsbZIP48	Os06g50310		AK071639	
AtbZIP19	At4g35040		N65677	
AtbZIP23	At2g16770		AV544638	PoGO F2
PtrbZIP38	648793			
PtrbZIP39	649217			
OsbZIP53	Os01g46970	OSBZ8	U42208	
OsbZIP54	Os05g49420	Gbf	AK065440	
AtbZIP54	At4g01120	GBF2	AF053228	PoGO G1
AtbZIP55	At2g46270	GBF3	U51850	
PtrbZIP42	244814			
PtrbZIP43	411188			
OsbZIP56	Os03g13614	HBP-1a	AK066563	
AtbZIP41	At4g36730	GBF1	X63894	PoGO G2
PtrbZIP44	424322			
PtrbZIP45	719452			
OsbZIP57	Os02g03580		AK112009	PoGO G3
OsbZIP58	Os12g13170	osZIP-1a	U04295	
AtbZIP16	At2g35530		NM_179917	
AtbZIP68	At1g32150			
PtrbZIP46	757220			
PtrbZIP47	826637			
OsbZIP55	Os07g10890			PoGO G4
OsbZIP60	Os01g07880	THY5	BAB62558	
OsbZIP61	Os06g39960			PoGO H1
AtbZIP64	At3g17609	HY5-like	AF453477	
PtrbZIP50	657788			
OsbZIP59	Os02g10860			
AtbZIP56	At5g11260	HY5	AB005295	PoGO H2
PtrbZIP48	717128			
PtrbZIP49	809109			
OsbZIP67	Os11g06170		AY224425	
OsbZIP68	Os12g06520	RSG	AK065995	
AtbZIP51	At1g43700	VIP1	AF225983	PoGO I1
PtrbZIP53	204863			
PtrbZIP54	411874			
OsbZIP69	Os04g41820		AK064429	
OsbZIP70	Os09g34060	RF2a	AF005492	
AtbZIP59	At2g31370	PosF21	X61031	PoGO I2
AtbZIP69	At1g06070		AJ419854	
PtrbZIP55	718317			
PtrbZIP56	292756			
OsbZIP71	Os03g21800	RF2b	AY466471	
OsbZIP72	Os07g48180		AK102562	
AtbZIP18	At2g40620		AY0744269	
AtbZIP52	At1g06850		AAF63137	PoGO I3
PtrbZIP57	242954			
PtrbZIP58	739018			
PtrbZIP59	239991			
OsbZIP62	Os09g34880			
OsbZIP63	Os04g10260			
OsbZIP64	Os08g43090	vsf-1	AF467732	
OsbZIP65	Os03g03550			
OsbZIP66	Os10g38820		AK108607	PoGO I4
AtbZIP29	At4g38900		AF401297	
AtbZIP30	At2g21230		AF401298	
PtrbZIP51	556549			
PtrbZIP52	721835			
AtbZIP31	At2g13150		AF401301	
AtbZIP32	At2g12940	UNE4	AV566578	
AtbZIP33	At2g12900			PoGP I1
AtbZIP71	At2g24340			
AtbZIP74	At2g21235			

bZIP no.	Gene code	Synonym	GenBank	
OsbZIP73	Os11g05640	OsZIP-2a	U04296	
OsbZIP75	Os12g06010	OsZIP-2b	U04297	PoGO J1
AtbZIP62	At1g19490			
PtrbZIP60	826496			
OsbZIP74	Os06g41770		AK107021	
AtbZIP60	At1g42990		AY045964	PoGO K1
PtrbZIP61	818888			
OsbZIP44	OsIFCC014214			
OsbZIP45	Os02g33560		AK063644	
AtbZIP77	At1g35490		AY087319	PoGO L1
PtrbZIP83	820200			
PtrbZIP84	822688			
OsbZIP46	Os12g09270			
OsbZIP47	Os11g11100		AK072267	
AtbZIP76	At1g58110		BT015864	PoGO L2
AtbZIP78	at?		BT002467	
PtrbZIP85	584476			
OsbZIP88	Os08g38020		AK107150	
OsbZIP89	Os09g29820		AK108319	
OsbZIP90	Os02g49560			
OsbZIP91	Os06g42690			
OsbZIP92	Os02g09830			
AtbZIP3	At5g15830		AV549429	
AtbZIP8	At1g68880		AF400621	
AtbZIP42	AT3g30530		BAB01020	
AtbZIP43	At5g38800			PoGO S1
AtbZIP48	At2g04038		AC007178	
AtbZIP58	At1g13600		AF332430	
AtbZIP70	At5g60830			
AtbZIP75	At5g08141			
PtrbZIP72	251247			
PtrbZIP77	266015			
PtrbZIP78	590335			
PtrbZIP79	564461			
PtrbZIP80	566729			
AtbZIP4	At1g59530		AF400619	
AtbZIP5	At3g49760			
AtbZIP6	At2g22850			
AtbZIP7	At4g37730			SE1
PtrbZIP73	572012			
PtrbZIP74	754888			
PtrbZIP75	764916			
PtrbZIP76	774123			
AtbZIP2	At2g18160	GBF5	AF53939	
AtbZIP11	At4g34590	ATB2		
AtbZIP44	At1g75390		AV566155	
PtrbZIP62	710131			
PtrbZIP63	715285			SE2
PtrbZIP64	424048			
PtrbZIP65	719591			
PtrbZIP66	649375			
PtrbZIP67	818112			
AtbZIP1	At5g49450		AF400618	
AtbZIP53	At3g62420		AF400620	
PtrbZIP68	564400			SE3
PtrbZIP69	659068			
PtrbZIP70	245573			
PtrbZIP71	816720			
OsbZIP80	Os07g03220			
OsbZIP81	Os03g56010			SM1
OsbZIP82	Os12g43790			
OsbZIP83	Os03g47200			
OsbZIP84	Os01g36220		AK110526	
OsbZIP85	Os03g19370		AK109929	SM2
OsbZIP86	Os05g03860	LIP19	X57325	
OsbZIP87	Os12g37410	OBF1	AB185280	
OsbZIP76	Os08g26880		AK100580	
OsbZIP77	Os09g13570		AK064903	SM3
OsbZIP78	Os02g03960		AK070887	
OsbZIP79	OsIFCC038657			
AtbZIP72	At5g07160			
PtrbZIP37	767814			
PtrbZIP81	751080			
PtrbZIP82	767813			
PtrbZIP89	777882			

Figure 3. Classification of bZIPs from Arabidopsis, black cottonwood and rice. Thirteen groups of homologues (A to L, and S) were defined through NJ phylogenetic analyses with the bZARP set (Figures S1 and S3). The organization into Possible Groups of Orthologues (PoGOs) was done by more refined NJ phylogenetic analyses inside each group of homologues, including also sequences from other eudicots and monocots. The

alignment used for these analyses corresponds to a concatenated sequence of the group-specific conserved motifs identified employing MEME (http://meme.sdsc.edu/meme/website/intro.html; Figure 2). (*) Represents genes that lack group-wise conserved motifs, thus they were included inside a PoGO according to their best hit to another bZIP. Because the relation of AtbZIP72, PtrbZIP37, 81, 82 and 89 could not be clarified, they were not included in any of the groups of homologous or orthologous genes. One Possible Group of Paralogues (PoGP I1) was found in Arabidopsis. Column 'Gene code' provides the gene identifiers for Arabidopsis, black cottonwood and rice bZIP sequences taken from TAIR (http://www.arabidopsis.org/), JGI (http://www.jgi.doe.gov/) or TIGR (http://www.tigr.org/), respectively. 'Synonym' indicates published and often cites names of bZIP genes. The GenBank accession numbers of nucleotide sequences are given.
doi:10.1371/journal.pone.0002944.g003

proved unsuccessful, we propose that up to three additional eudicot-monocot PoGOs, besides S1, exist in Group S (as a minimal representation of the three possible monocot and eudicot PoGOs). The difficulty of organizing Group S bZIPs into PoGOs that comprise both eudicots and monocots sequences may reflect an increased evolutionary rate after their emergence. Rapid evolution can mainly be explained by relaxation of purifying selection or by positive selection. We used the Yang algorithm [73] to verify whether lineage-specific dN/dS ratios in Arabidopsis, black cottonwood and rice (the ω parameter, [74,75]) of Group S were different from that of all other groups. The ω value for Group S (0.12) was found to be significantly different from the average ω calculated for all other groups (0.03, likelihood ratio test $\chi^2_{df=1}$, $p<0.01$). Despite being under purifying selection ($\omega<1$), the value of ω for Group S is four times higher than the average. Thus it can be concluded that purifying selection is relaxed in this group, explaining the higher rate of sequence divergence among its members. Low selective constraint (i.e., low purifying selection) is a hallmark of more recently duplicated genes and can be correlated with functional diversification [76]. The extensive amplification of Group S members in angiosperms (see below) further supports the notion that functional diversification partly related to the control of energy metabolism is operating among Group S genes.

In Group G, we observed one PoGO that is restricted to monocots (PoGO G4; Figure S10). This may be explained by gene gain at an early phase of monocot radiation, or alternatively by gene loss in the ancestor of the eudicot lineage. Our analysis also revealed the existence of a Possible Group of Paralogues (PoGP) restricted to Group I in Arabidopsis (PoGP I1, Figure S12). This PoGP most probably reflects a recent duplication event followed by rapid divergence in the Arabidopsis lineage. As PoGO G4 and PoGP I1 are restricted to distinct evolutionary lineages, they probably do not play essential (common) roles in angiosperms as a whole. This conclusion is supported by the fact that EmBP from maize and wheat, both assigned to PoGO G4, control reserve protein (prolamin) production [77] which can be considered a monocot-specific function.

Gene duplication is an important means of evolutionary diversification. Therefore, PoGOs that preferentially expanded during angiosperm evolution are expected to include genes that were particularly important for establishing angiosperm-specific physiological or functional characteristics. Of the 13 groups of homologous genes, Groups A, D, E, I and S contain more genes per PoGO than the average (approximately six genes per PoGO, Figure S17), indicating their preferential contribution to the evolution of adaptive characteristics in angiosperms. Interestingly, Groups A, D and S include genes for responses and adaptation to environmental factors (abiotic and biotic stresses in Groups A/S and D, respectively; Table S4) and the control of energy use (Group S; Table S4). These observations raise the possibility that genes of these groups were particularly important for the colonization of new habitats and consequently for the radiation and expansion of angiosperms (Text S1d). Additionally, some PoGOs have a conserved one-to-one gene relationship, indicating that their genes may play a pivotal role during development (Text S1e)

In summary, we propose the existence of 31 monocot-eudicot PoGOs in Groups A to L, one monocot-specific PoGO (G5), one PoGP (I1) in Arabidopsis, and possibly three PoGOs in Group S. The 34 PoGOs are likely to be related to 34 possible ancestral functions of bZIPs in angiosperms (Figure 3, and Text S1d).

Tracing the Origin and Diversification of *bZIP* Genes in Green Plants

Based on the phylogenetic analyses and the *bZIP* gene structures from Arabidopsis, black cottonwood and rice, we propose a model for the evolution of angiosperm bZIPs (Figure 1A). This model proposes two large clades encompassing Groups A, D, F, G and J, and Groups B, C, E, H, I, K and L, respectively. Groups B, H and K, Groups E and L, and Groups D and F are sister groups, as evidenced by their bootstrap support. Furthermore, the conserved intron position in the bZIP domain shared by Groups A, D, G and J, as well as the one shared by Groups C, E, H, I, K and L (Figure S3) supports the hypothesis that these groups diverged from a common ancestor. We were not able to establish a clear relationship of Group S to any of the two larger groups. It may have an independent ancestral origin, constituting a third group, or may have evolved from one of the two large groups (Figure 1A).

To identify groups of homologues among the major eukaryotic lineages, i.e. animals, fungi, and plants, we performed a large-scale phylogenetic analysis using the conserved bZIP region of all bZIPs from *Homo sapiens* [78], *Caenorhabditis elegans* (http://www.wormbase.org/), *Drosophila melanogaster* [79], *Saccharomyces cerevisiae* (http://mips.gsf.de/genre/proj/yeast/), *A. thaliana* and *O. sativa*. This analysis revealed that *bZIPs* of each of these lineages share only one common ancestor (data not shown) which is in accordance with the fact that only a single *bZIP* sequence is present in the primitive eukaryote *Giardia lamblia* [80,81], perhaps representing the *bZIP* gene content prior to the plant/animal/fungal separation [80]. The function of this unique ancestral gene may be related to unfolded protein (UPR) and oxidative stress responses (see below). Deep evolutionary analyses have also been performed for the homeodomain and MADS-box families and it appears that their member TFs derived from at least two genes present in the last common ancestor of the three eukaryotic kingdoms [19,82]. It has been proposed that one of the ancestral functions of the MIKC^c class of MADS-box genes is an involvement in reproductive organ development [83,84]. Although this function appears to be conserved, it is still not clear whether it has a monophyletic origin.

We identified 7, 8, and 40 *bZIP* genes, respectively, in the genomes of the algae *Chlamydomonas reinhardtii* and *Ostreococcus tauri* and the moss *Physcomitrella patens* (however, a complete bZIP domain is missing in three of the moss proteins). Additionally, we identified bZIP sequences from assembled ESTs of species representing the most relevant divisions of the green plants from which sequences are available: four *bZIP* genes in the bryophyte *Marchantia polymorpha*, one each in the ferns *Selaginella moellendorffii* and *Adiantum capillus-veneris*, and 40 and nine, respectively, in the gymnosperms *Pinus taeda* and *Picea glauca* (Table S5). Although no complete genomic sequences were available for ferns or gymno-

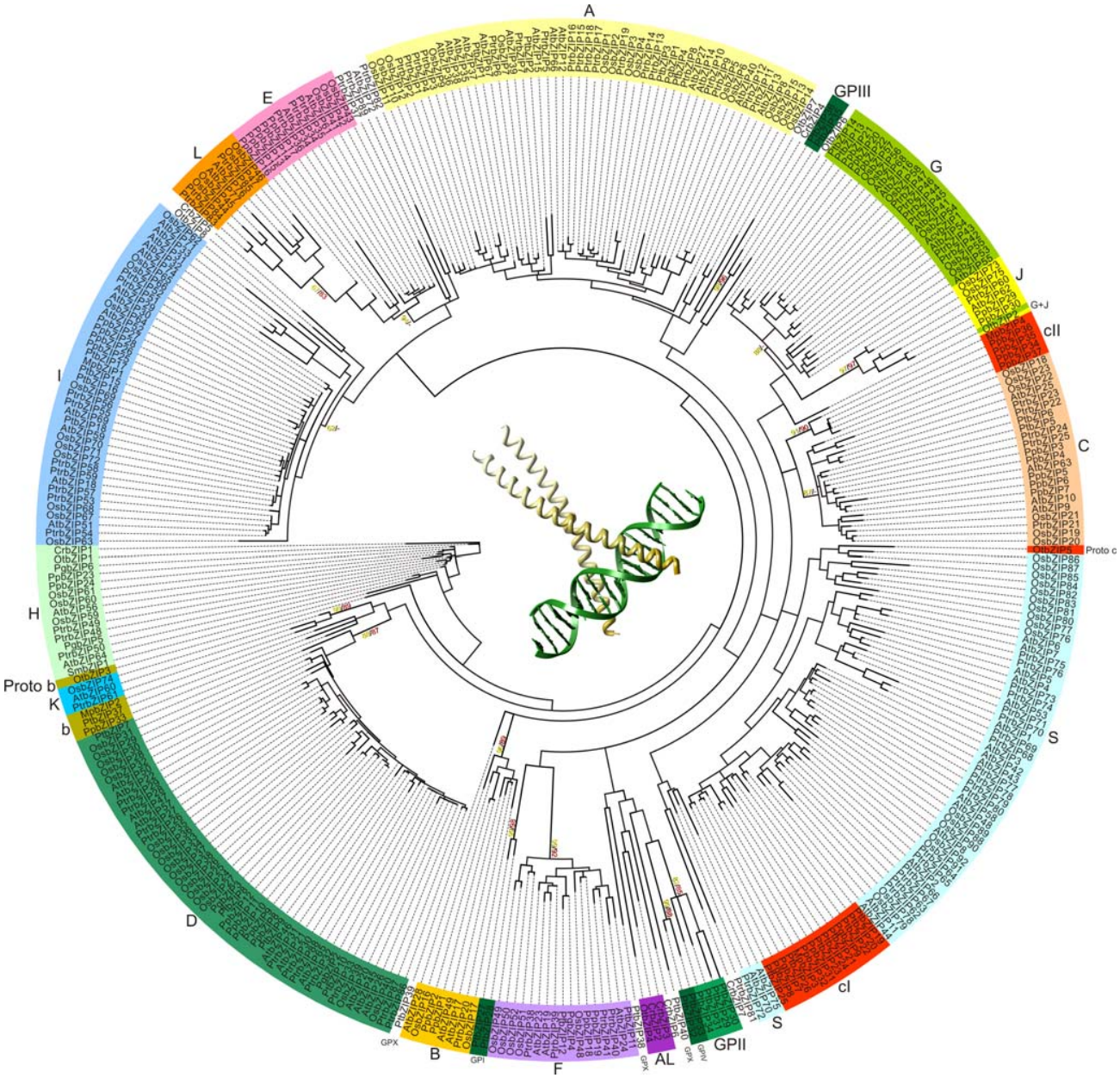


Figure 4. Global Phylogeny of bZIPs in green plants. This tree is a consensus of NJ analyses with p-distance performed with the ViridiZIP set. Bootstrap values in yellow were calculated from NJ analysis (PAM matrices, and with 44 and 60 amino acid alignments; only the highest bootstrap values are shown). Bootstrap values in red were calculated from ML analyses using the JTT+Γ evolutionary model (either with 44 or 60 amino acid alignments; only the highest bootstrap values are shown). GPX, GPI, GPII, GPIII, and GPIV indicate putative gymnosperm specific groups. Each group of homologues is colored following the same colour scheme used in Tables I and SV. The center of the tree depicts a typical bZIP dimer bound to DNA, representing the conserved bZIP domain (GCN4 from *Saccharomyces cerevisiae*; Protein Data Bank entry 2DGC).
doi:10.1371/journal.pone.0002944.g004

sperms, a considerable number of ESTs is available for the latter. We assembled a set of 345 bZIPs from algae to angiosperms (ViridiZIP set) for phylogenetic analyses (Figures 1B, 1C and 4).

Our study revealed that Group H is the most conserved group of *bZIP* homologues; members of this group are present in all green plant lineages. This observation is particularly interesting because Group H includes *HY5* and *HYH* that are important regulators of light responses and anthocyanin biosynthesis (Table S4). We therefore propose that Hy5-like bZIPs control light-dependent processes in all green plants. Similar to bZIPs in Group H, DOF transcription factors involved in light responses (subfamily A) also appear to be well conserved, suggesting that genes involved in light-related functions are under strong selective constraints [85]. In Arabidopsis Hy5-mediated photomorphogenesis is negatively regulated by the E3 ubiquitin ligase Cop1, which ubiquitylates Hy5 protein leading to its degradation [86]. We detected Cop1-related proteins in Physcomitrella, in agreement with previous results, as well as the Cop1-interaction motif in Physcomitrella Hy5-like bZIPs, suggesting that the genetic toolkit for photomorphogenesis described in angiosperms is also present

in mosses [87]. We also detected a single gene similar to *COP1* in Ostreococcus (ID 30007), but while in higher plants Cop1 protein contains a RING domain at the N-terminus, followed by multiple WD40 repetitions [88], this order is reversed in the Ostreococcus protein. Moreover, a Cop1 interaction site (Table S2) was not detected in the algal *HY5*-orthologues OtbZIP1 or CrbZIP1, or in any other green algae bZIP. Nevertheless, we found one Cop1-related protein in the red alga *Cyanidioschyzon merolae* (ID CMK039C; http://merolae.biol.s.u-tokyo.ac.jp/). Cop1-like proteins are also known in animals where they promote the degradation of the bZIP transcription factor c-Jun [88], suggesting Cop1-dependent protein degradation to be a regulatory scheme conserved in most eukaryotes.

Groups B, C, D, E, F, G, I and J were present in the most recent common ancestor (MRCA) of bryophytes and tracheophytes, indicating a functional connection to the colonization of the terrestrial environment (Figure 5). Some of these genes play a role in light responses (Group G), nitrogen/carbon balance control (Groups C and G), and ion responses (Group D), which are some of the important features that developed further in embryophytes (Table S4). Moreover, it appears that during the evolution from early land plants to angiosperms, Group D and I genes amplified

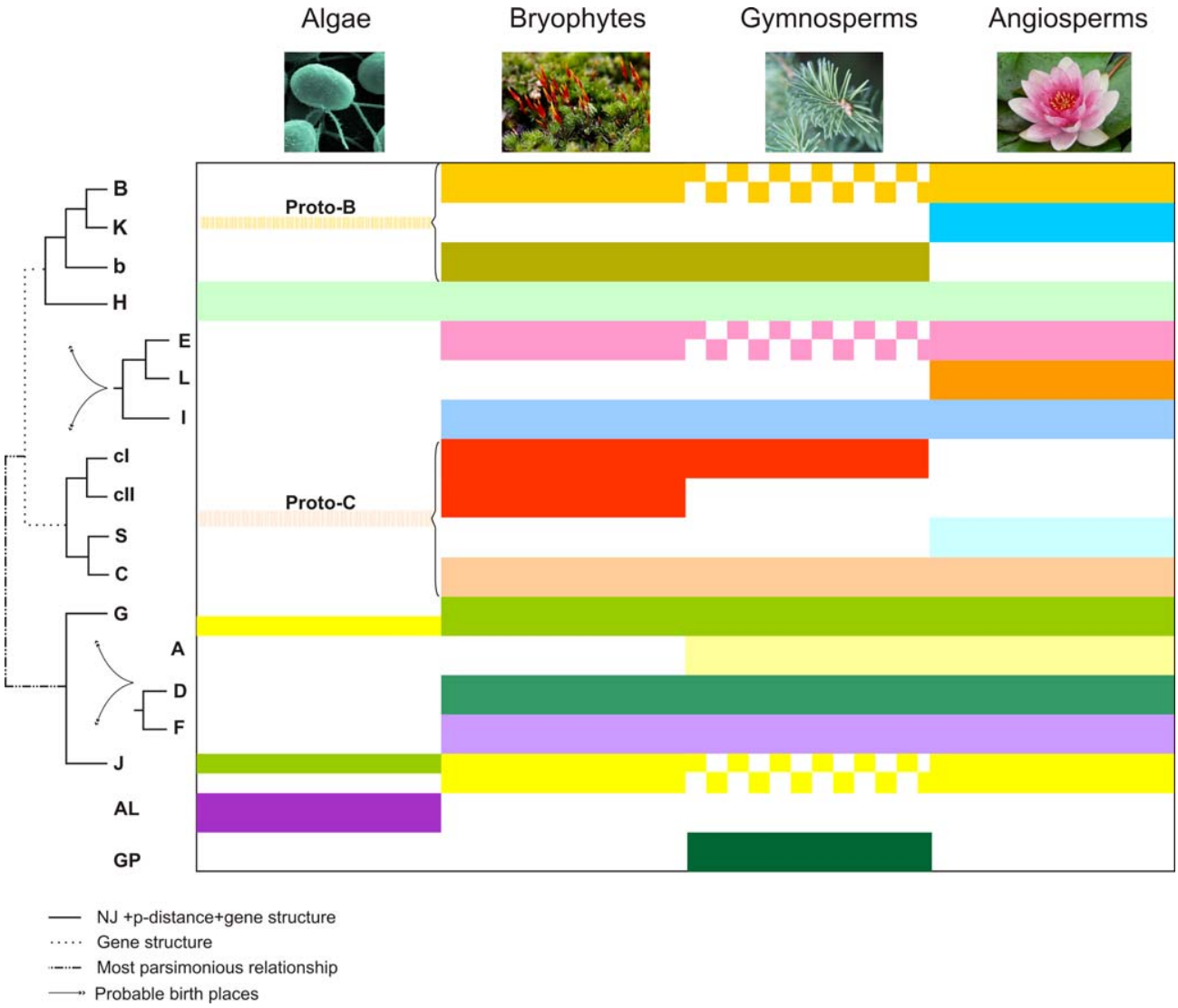


Figure 5. Phylogenetic profile and structure of bZIPs in green plants. Groups E, L and I belong to the same branch as Groups Proto-B, Proto-C and H but their exact position is not clear (Figure 1A). Similarly, Groups A, D and F do not have a clear position, though they belong to the same branch as Groups G and J (Figure 1A). The relation of Groups AL and GP to the other groups could not be established. bZIPs of the species studied here were grouped at the level of higher taxa, i.e., algae (represented by *C. reinhardtii* and *O. tauri*); bryophytes (*P. patens*); gymnosperms (*P. glauca* and *P. taeda*), and angiosperms (*O. sativa*, *A. thaliana* and *P. trichocarpa*). Solid boxes indicate that at least one bZIP was found for a given group of homologues in the respective taxon. Squared boxes indicate that homologous bZIP sequences were not yet observed in gymnosperms, possibly due to sampling limitations. Notably, however, sequences of the respective groups are conserved in bryophytes and angiosperms. Dashed lines with brackets shown in Groups Proto-B and Proto-C indicate that there is an orthologous bZIP in at least one of the algal species, although it does not strictly belong to any of the homologous groups. The half lines present in G and J indicate the presence of common orthologues in algae. Groups AL, GP, K, L and S appear to be lineage specific.
doi:10.1371/journal.pone.0002944.g005

more than genes of the other groups of homologues (5 to 10, and 4 to 11 genes in groups D and I, respectively), strongly suggesting that both groups were particularly important for this transition. Several Group D genes are involved in biotic stress responses (Table S4) indicating that improved pathogen defense was important for land plant evolution. Some *bZIP* genes of Group I control the expression of vascular genes (Table S4), which are central to vascular tissue development in tracheophytes.

Group A probably first appeared in the MRCA of spermatophytes and may thus be related to seed formation (Figure 5). As a matter of fact, Group A bZIPs often have functions in seed development, ABA responsiveness and fruit maturation (Table S4). Moreover, they are elements of ABA-dependent signaling pathways that coordinate responses to desiccation/dehydration and salt stress. ABA-mediated signaling is known in Physcomitrella [89,90], however, Group A bZIPs are not present in this organism (Figure 5), indicating a less developed ABA regulatory network (Text S1f).

According to our data Groups K, L and S are angiosperm-specific (Figure 5). However, due to sampling limitations we can not formally exclude the possibility that these groups are also present in gymnosperms. Additionally, this analysis eliminates the hypothesis that Group S has an independent ancestral origin (Figures 1A and 1C).

We also detected Group NA, a possible group of homologues exclusively present in non-angiosperm plants (Figure S18, and Text S1g). This finding is intriguing as genes conserved in mosses and gymnosperms are expected to represent general plant functions. Group NA bZIPs may thus have lineage-specific roles unimportant for angiosperms; the reduction of a dominant gametophyte during angiosperm evolution combined with a concomitant gene loss is an example for this. Alternatively, gene loss could have played a key role in the acquisition of important features in angiosperms, as seen for *KNOX* genes [91]; or, the roles played by bZIPs of Group NA could have been taken over by non-related but functionally analogous genes (non-orthologous gene displacement).

Ancestral Relationships in Groups B and C

The above analysis in combination with detailed sequential NJ analyses restricted to algal, moss and/or Arabidopsis sequences revealed two new groups, i.e. Groups Proto-B and Proto-C (Figure 1B). Group Proto-C encompasses Group C (Figure 1A) and two new Groups, cI and cII that correspond to the sequences previously identified in Group NA (Figure S18). While cI appears to be restricted to bryophytes, cII is found up to gymnosperms, and C is present up to angiosperms (Figures 1C and 5). Notably, in all phylogenetic analyses Group S appeared to be more attracted by Groups C, cI and cII (Figures 1C, 4 and 5), suggesting it originated from Group Proto-C, probably by gene duplication followed by rapid evolution. This finding is supported by the observation that bZIPs tend to dimerize with more similar partners, e.g. AtbZIP10 (Group C) with AtbZIP53 (Group S) [34,92]. Additionally, members of Group C (*AtbZIP63*) and S (*ATB2*, *GBF5*, *AtbZIP1* and *AtbZIP53*) participate in the control of energy metabolism and thus share similar functions (Table S4). Moreover, Group Proto-C possesses one *bZIP* gene, *OtbZIP5* from Ostreococcus, supporting the model that the biological functions played by bZIPs of Group C/S, such as oxidative stress responses associated with *AtbZIP10* [40] and energy metabolism control mediated for example by *GBF5* [41], are at least partially present in all green plants. Importantly, oxidative stress signaling involving bZIPs has been reported in yeast and men and thus appears to be conserved in all eukaryotes [93–97].

Group Proto-B consists of Group B, which includes members from bryophytes and angiosperms, a new group of homologues (Group b) that is apparently restricted to bryophytes and gymnosperms, and the Ostreococcus gene *OtbZIP3* (Figures 1B, 4 and 5). Based on our initial phylogenetic analysis of angiosperm sequences (Figure 1A) and tree topology (Figures 1C and 4) we concluded that angiosperm-specific Group K is not only a sister group of B, but very likely also emerged from Proto-B. Members of Group K are likely to have a role in the unfolded protein response (UPR), a cellular process involving the endoplasmic reticulum (ER) that counteracts cellular stress when incorrectly folded proteins accumulate [43]. bZIPs involved in this response are known in mammals and yeast and thus appear to be conserved in many lineages [98,99]. Recently, Liu et al. [42] demonstrated a role of Arabidopsis AtbZIP17 (Group B) in the UPR pathway, supporting the hypothesis that Group K emerged from Group B, and that *OtbZIP3* plays a similar role. Members of Groups B and K (like animal bZIP proteins involved in UPR) posses a trans-membrane domain for ER attachment (Table S2), but members of Group K lack the cleavage site recognized by the so-called site-1 protease (S1P). Most likely, the two groups function in different branches of the UPR pathway. Additionally, we looked for the presence of both trans-membrane and S1P interaction domains in other plant proteins. The trans-membrane domain is present in all Group B and K bZIPs from green plant lineages, whereas the S1P interaction domain was not found in some of them, perhaps due to missing sequence data.

Another important result of our analysis is that Ostreococcus sequences could be included, with significant bootstrap support, into Groups Proto-C (*OtbZIP5*) and Proto-B (*OtbZIP3*; Figure 1B). Moreover, Ostreococcus *OtbZIP2* was found to significantly cluster with Groups G and J, forming a new group named G+J (Figure 1B).

In conclusion, our results indicate that four Ostreococcus *bZIP* genes can be assigned to Groups Proto-C (*OtbZIP5*), Proto-B (*OtbZIP3*), G+J (*OtbZIP2*), and H (*OtbZIP1*), defining four orthologous relationships between algal and five groups of homologues from terrestrial plants (Figure 6). This data suggests the presence of at least four founder genes in the MRCA of green plants. Our analysis also indicates that Groups H (including *OtbZIP1* and *CrbZIP1*) and Proto-B (including *OtbZIP3*) originated from a common ancestral gene (Figure 1B). However, their relationship with Proto-C (*OtbZIP5*) and G+J (*OtbZIP2*), and the relationship of the four founder genes to the possible monophyletic origin of bZIPs in green plants could not be determined. The most parsimonious model that can explain the origin of the four ancestral bZIPs is shown in Figure 6. The assumption that Group Proto-C and Groups H/Proto-B share a common ancestral gene was inferred from the observation that angiosperm Groups C, B and H also cluster together (Figure 1A). Similarly, all DOF TFs appear to have originated from a single founder gene from subfamily A, which was present in the MRCA of green plants and might have played a role in light-regulated mechanisms [18]. In addition, MADS-box TYPE II (MIKC^{C}) and HD-Zip class III TF families each emerged from a single founder gene present in the MRCA of streptophytes that was possibly involved in haploid reproductive cell differentiation [84] or control of apical growth [23,24], respectively.

bZIP Evolution in Plants

Our data show that Group C and B members are elements of the oxidative stress signaling and UPR pathways, respectively, which appear to be crucial in all eukaryotes. This observation and the likely monophyletic origin of bZIPs of the main eukaryotic lineages (plants, animals, and fungi) suggest that the common bZIP ancestor was a multifunctional regulatory factor. An important

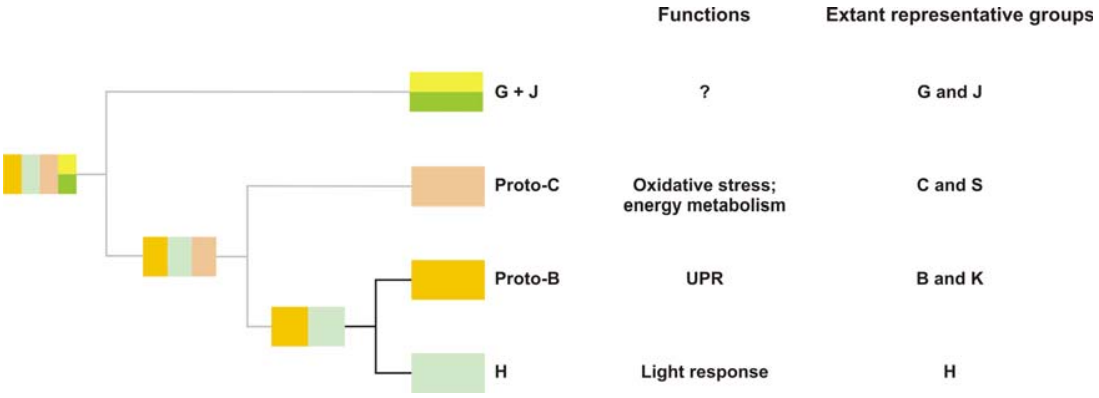


Figure 6. Most parsimonious model explaining the emergence of the four green plant founder *bZIP* genes. The four founder genes (in Groups G+J, Proto-C, Proto-B and H) are derived from a unique ancestral gene common to all eukaryotes. Groups Proto-B and Proto-C most likely derived from a multifunctional UPR/oxidative stress gene. Groups Proto-B and H are sister groups and their relationship to Group Proto-C was found by analyzing angiosperm bZIPs (Figure 1A). Group G+J is the ancestral group of a large set of *bZIP* genes included in Groups A, D and F, but the ancestral function played by this group is still largely unknown.
doi:10.1371/journal.pone.0002944.g006

consequence of this model is that Group H, which has a central role in light-mediated control, emerged from bZIPs of the oxidative stress and UPR regulatory modules. The integration of the branch leading to Group G+J, however, remains unclear which is partially due to the fact that functional information is limited and restricted to Group G that plays a role in light and ABA signaling.

From the extant algal sequences that do not cluster into any of the homologous groups of streptophytes, only a single group of homologues restricted to algae could be detected (Group AL; Figures 1C and 5). In most cases bZIP sequences from Chlamydomonas and Ostreococcus do not cluster together at all. This observation indicates that bZIPs evolved differently in the algal lineages, probably reflecting adaptations to different ecological niches; Chlamydomonas lives in fresh water, while Ostreococcus lives in sea water.

We estimated the number of bZIPs in the MRCA of all land plants (embryophytes), using the method of Hahn *et al.* [100]; the MRCA most likely had 64 bZIPs that expanded to 83 in the branch leading to seed plants. The rate of gene gain-loss, λ, in the seed plant lineage was found to be 2.01×10^{-3} per million years, which is similar to estimates for yeast (0.002) [100] and mammals (0.0016) [101]. We calculated expansions and contractions of the bZIP phylogenetic branches in the land plant lineage, using the estimated value for λ; this revealed a significant expansion ($p<0.05$) of the branch leading to the seed plant lineage. Finally, the evolution of the *bZIP* gene family is well explained by the random birth-and-death model in seed plants, i.e., no significant expansions/contractions occurred preferentially in any specific PoGO or group of homologues (Figure S19, and Text S1h).

Conclusions

In our analysis presented here we systematically classified bZIP TFs into PoGOs and considered existing knowledge about their biological functions to establish a robust methodology to reveal evolutionary relationships of this group of regulatory proteins. The moss Physcomitrella possesses almost five times more *bZIP* genes (37 genes, Table S5) than the alga Ostreococcus (8 genes), and half the number found in angiosperms (around 80 genes). Group A genes first appeared in the MRCA of spermatophytes and were recruited for seed development or germination but also to fine tune the responses to desiccation/dehydration and salt stress. Groups K, L and S are seemingly exclusive to angiosperms. Unexpectedly, Groups K and S control processes conserved in all eukaryotes, i.e. UPR and energy homeostasis. This apparent paradox can be explained by the fact that both, Groups K and S derived from the functionally related Groups Proto-B and Proto-C, respectively, that emerged early on during green plant evolution. Group S amplification likely contributed to refining the regulatory circuit controlling the organism's energy status. The most strongly conserved group of homologues in algae and angiosperms is Group H which includes light control factors *HY5* and *HYH*. Group H is representative of one of the four green plant founder *bZIP* genes. Our data thus establish the hypothesis that bZIP-controlled light responses of Group H emerged (through neofunctionalization) from a multifunctional ancestral gene of the UPR and oxidative stress response pathways (UPR/oxidative stress). The UPR/oxidative stress gene is also the ancestor of two other of the four founder genes, i.e. Groups Proto-B (UPR) and Proto-C (oxidative stress), which most likely diverged through subfunctionalization processes. The fourth founder gene, represented by Groups G and J, is the sister gene of the multifunctional UPR/oxidative stress gene. More functional data for Group G- and J-related bZIPs are required to further elaborate the model of green plant bZIP evolution.

Materials and Methods

Datasets of *bZIP* Genes

We generated a bZIP dataset (Angiotot) representing an updated version of the ABZ data set [56]. Plant bZIP sequences were identified as described by Riaño-Pachón et al. [102]. The whole proteomes deduced from the completely sequenced genomes of the algae *Ostreococcus tauri* [68] and *Chlamydomonas reinhardtii* [67], the bryophyte *Physcomitrella patens* [69], and the angiosperm *Populus trichocarpa* [59] were downloaded from the Joint Genome Institute/Department of Energy (JGI/DOE; http://www.jgi.doe.gov/). Protein sequences for the angiosperm *Arabidopsis thaliana* [54] were downloaded from The Arabidopsis Information Resource (TAIR; http://www.arabidopsis.org/), and from The J. Craig Venter Institute (http://www.tigr.org/) for the monocot *Oryza sativa* ssp. *japonica* [58].

Assembled ESTs from *Marchantia polymorpha*, *Physcomitrella patens*, *Adiantum capillus-veneris*, *Selaginella moellendorffii*, *Picea glauca*, *Pinus*

taeda, *Brassica napus*, *Glycine max*, *Heliathus annus*, *Medicago truncatula*, *Solanum lycopersicum*, *Solanum tuberosum*, *Hordeum vulgare*, *Saccharum ssp.*, *Sorghum bicolor*, *Triticum aestivum* and *Zea mays* were downloaded from the TIGR Plant Transcript Assemblies Database [103]. ESTs from *Oryza sativa* ssp. *indica* were downloaded from the Beijing Genomics Institute website (07.11.2006), and assembled into clusters using TGICL [104]. Additional rice bZIP sequences were obtained from the Full Length Rice cDNA Consortium [105]. Some sequences from completely sequenced genomes were re-annotated (Datasets S1 and S2), based on conserved protein motifs and gene structures of each family. The list of abbreviations of the organisms used is given in Table S6.

The tblastn program [106] was used to search for bZIP sequences in rice nucleotide databases (*Oryza sativa* ssp. *indica* [57]; Beijing Genomics Institute, http://btn.genomics.org.cn/rice, and *Oryza sativa* ssp. *japonica*; Syngenta, http://www.syngenta.com/; IRGSP, http://www.gramene.org/) using Angiotot as query. Sequences with an e-value $<10^{-4}$ were selected to form a subset (SeqZIP), from which false positive hits, corresponding mainly to low complexity regions, and hits that we initially identified using the above procedure were excluded. To identify the open reading frame and gene structure of each SeqZIP sequence, pairwise blastx analyses against their respective Angiotot best hits were performed. Gene structures were defined based on the alignments obtained, the conserved positions of introns in homologous bZIP genes, and the presence of canonical splicing sites (GT-AG). The protocol used for bZIP identification is described in Figure S20.

The procedure used to identify bZIPs in EST datasets was identical to that used for genomic sequences, except that the estwisedb program of the Wise2 package [107] was included to identify the most likely reading frames and its bZIP domains in a given cluster.

Phylogenetic Analyses

Alignment of bZIP protein sequences was performed by ClustalX [108], using default parameters, and subsequently adjusted manually. The alignments used for the analyses within each group of homologues represent a concatenated sequence of the different conserved motifs found within each group (Figure 2). The phylogenetic analyses based on amino acid sequences were conducted using MEGA v3.1 [109] and PHYLIP v3.6 [110]. Unrooted phylogenetic tree topologies were reconstructed by Neighbor-Joining (NJ), the distances were obtained using a PAM-like distance matrix [111], or alternatively, using p-distances [112], and the re-sampling of the original bZIP set was a 1,000 bootstrap repetition. Maximum Likelihood (ML) analyses of the bZIP domain (44 and 60 amino acids) were carried out using RAxML [113] with the distances computed using the JTT+Γ evolutionary model [114], and a re-sampling of the original bZIP set of 500 bootstrap repetitions. Bayesian approaches were not employed as they often lead to very liberal estimates of branch confidence that can result in wrong topologies [115]. Additionally, phylogenetic trees for nucleotide sequences, corresponding to the conserved motifs used for proteins, were inferred by means of the maximum likelihood method available in PAUP 4b10 [116]. The TrN+Γ [117] model of sequence evolution was used. Model choice was performed in MODELTEST 3.6 [118] by the likelihood ratio test with significance level set at 1%. ML trees are available upon request. Branch lengths of the tree comprising all species analyzed were estimated by Maximum Likelihood in TREE-PUZZLE v5.2 [119], using the consensus topology inferred by NJ analysis with PAM-like distances. All sequences and alignments used in this study are available upon request.

Identification of Conserved Motifs

The putative complete sets of unique bZIPs from Chlamydomonas, Ostreococcus, Physcomitrella, black cottonwood, Arabidopsis and rice served as input for a conserved motif analysis performed with MEME (http://meme.sdsc.edu/meme/meme.html) [120]. Whole protein sequences were employed for this search. A given motif was allowed to appear at any number of repetitions, the maximum width of a motif was set to 80, and the maximum number of motifs was set to 20. The other parameters were used as default. In a complementary approach, each group of homologues was analyzed individually with the parameters described above.

Phylogenetic Analyses and Identification of Possible Groups of Orthologoues (PoGOs)

The detailed evolutionary analysis of angiosperm bZIP sequence relationships within each group allowed the identification of PoGOs. A PoGO is defined by the following criteria: (i) members of a PoGO have a monophyletic origin, indicated by a bootstrap support greater than 50%; (ii) a PoGO possesses at least one representative gene each from *A. thaliana* and *O. sativa*, assuming that the putative complete sets of *bZIP* genes of these organisms were identified and no selective gene loss had occurred. In case a PoGO is found to be restricted to either monocots or eudicots, the presence of sequences from at least one other species of the same lineage in this PoGO is required; and (iii) the inferred phylogeny should be consistent with the known phylogeny of plant species [56].

Identification of Pseudogenes and Genomic Duplications

Search for pseudogenes in Chlamydomonas, Ostreococcus, black cottonwood, Arabidopsis and rice was performed by masking the genomic region for each identified bZIP. Blastx searches were performed against the masked sequences using the Angiotot bZIP database as query. A hit was considered as a pseudogene only if it possessed all or part of the bZIP domain; therefore all hits were compared against bZIP PFAM models [121] and manually cured, eliminating false positives. Genomic duplications in Arabidopsis were identified via "Paralogons in Arabidopsis thaliana" (http://wolfe.gen.tcd.ie/athal/dup) and 'MATDB: Segmental Duplications' from MIPS (Munich Information Center for Protein Sequences; http://www.mips.gsf.de/projects/plants) (Table S7).

Analysis of Gene Family Expansion and Contraction

The evolution of rates of *bZIP* gene gain and loss along the history of green plants was analyzed by the method of Hahn *et al.* [100], implemented in CAFÉ [122]. The method models gene family evolution as a stochastic birth-and-death process implemented as a probabilistic graphical model that allows for the inference of the most likely family sizes in the common ancestors of every branching point. In this way one can test the null hypothesis of random change in the family size. To avoid incomplete sampling, only plants with fully sequenced genomes were analyzed. The algorithm developed by Hahn *et al.* uses a birth-and-death parameter, λ, which was also estimated within CAFE. In addition to the parameter λ, CAFE needs divergence times to be entered along with the phylogeny of the organisms used. Since the inference of the size of gene families at deep evolutionary times is not reliable with any of the current methods available (Hahn, personal communication; [100]), we focused on land plants only. Tree topology and divergence times are shown in Figure S19. Significance of the contractions and expansions along branches was accessed by means of the three methodologies available in

CAFE: branch cutting, likelihood ratio test, and Viterbi assignments [122].

Gene Expression Analysis

Absolute signal intensity values from Arabidopsis ATH1_22K array (Affymetrix) was obtained through Meta-Analyzer from GENEVESTIGATOR (http://www.genevestigator.ethz.ch/) [123]. The developmental stages were as described by Boyes *et al.* [124]. Massively Parallel Signature Sequencing, MPSS, [125] was also verified for Arabidopsis and rice genes (Datasets S3 and S4).

Supporting Information

Figure S1 Definition of homologous gene groups A, D and F. This figure is a partial representation of the tree inferred from NJ analysis from the 258 non-redundant set of bZIPs from Arabidopsis, rice and black cottonwood using *p*-distance and 1000 bootstrap repetitions (indicated as percentages at the branch points). The alignment used corresponds to the minimum bZIP domain of 44 amino acids. Groups D and F are sister groups supported by a 50% bootstrap. Rice, black cottonwood and Arabidopsis sequences are represented in orange, dark blue and light blue, respectively.
Found at: doi:10.1371/journal.pone.0002944.s001 (1.01 MB TIF)

Figure S2 Conserved intron position in the basic motif region of angiosperm bZIP transcription factors. The first leucine of the leucine zipper is highlighted in green, and the conserved asparagine of the basic motif is shown in red. According to the position of the introns, indicated by arrows, four different groups can be observed (1 to 4). bZIPs from Group L have a basic motif five amino acids shorter than that of the other bZIPs, and the conserved asparagine, shown in red, is substituted either by lysine (K) or arginine (R). In bold, the first amino acid after the intron. The *bZIP* genes used in this figure are: *AtbZIP24* (Group F), *AtbZIP45* (Group D), *AtbZIP39* (Group A), *AtbZIP54* (Group G), *AtbZIP62* (Group J), *AtbZIP63* (Group C), *AtbZIP56* (Group H), *AtbZIP61* (Group E), *AtbZIP31* (Group I), *AtbZIP60* (Group K), *AtbZIP76* (Group L), *AtbZIP70* (Group S), and *AtbZIP49* (Group B).
Found at: doi:10.1371/journal.pone.0002944.s002 (1.85 MB TIF)

Figure S3 Unrooted phylogenetic tree inferred from a NJ analysis from a subset of 173 bZIPs of Arabidopsis, rice and black cottonwood using *p*-distance and 1000 bootstrap repetitions (indicated as percentages at the branches). The alignment used corresponds to the minimal bZIP domain extended by two leucine repetitions, totaling 60 amino acids. Groups B, K and H, as well as Groups E and L are sister groups supported by bootstrap analysis. Rice, black cottonwood and Arabidopsis sequences are represented in orange, dark blue and light blue, respectively.
Found at: doi:10.1371/journal.pone.0002944.s003 (1.11 MB TIF)

Figure S4 Phylogenetic tree of monocot and eudicot bZIPs of Group A. The unrooted tree was inferred by a NJ analysis from distances calculated with the PAM distance matrix. The bootstrap values correspond to 1000 repetitions and are indicated as percentage in every branch. The amino acid alignment used to generate this tree corresponds to the bZIP domain plus the conserved motif A1 (Figure 2 and Table S2). Rice, black cottonwood and Arabidopsis sequences are represented in orange, dark blue and light blue, respectively. Other eudicot sequences are shown in green. The organism from which the remaining monocot and eudicot bZIPs originated is indicated by the last two letters in each sequence. Abbreviations are explained in Table S6.
Found at: doi:10.1371/journal.pone.0002944.s004 (1.28 MB TIF)

Figure S5 Phylogenetic tree of Group B bZIPs from monocots and eudicots. An unrooted tree was inferred by a NJ analysis from distances obtained from the PAM distance matrix. The bootstrap values correspond to 1000 repetitions and are indicated as percentage in every branch. The amino acid alignment used to generate this tree corresponds to the bZIP domain plus the conserved motifs within this group (Figure 2 and Table S2). Rice, black cottonwood and Arabidopsis sequences are represented in orange, dark blue and light blue, respectively. Other monocot sequences are shown in red. The organism from which the remaining monocot and eudicot bZIPs originated is indicated by the last two letters in each sequence. Abbreviations are explained in Table S6.
Found at: doi:10.1371/journal.pone.0002944.s005 (0.31 MB TIF)

Figure S6 Phylogenetic tree of Group C bZIPs from monocots and eudicots. An unrooted tree was inferred by a NJ analysis from distances calculated with the PAM distance matrix. The bootstrap values correspond to 1000 repetitions and are indicated as percentage in every branch. The amino acid alignment used to generate this tree corresponds to the bZIP domain plus the conserved motif within this group (Figure 2 and Table S2). Rice, black cottonwood and Arabidopsis sequences are represented in orange, dark blue and light blue, respectively. Other eudicot and monocot sequences are show in green and red, respectively. The organism from which the remaining monocot and eudicot bZIPs originated is indicated by the last two letters in each sequence. Abbreviations are explained in Table S6.
Found at: doi:10.1371/journal.pone.0002944.s006 (2.03 MB TIF)

Figure S7 Phylogenetic tree of Group D bZIPs from monocots and eudicots. An unrooted tree was inferred by a NJ analysis from distances calculated with the PAM distance matrix. The bootstrap values correspond to 1000 repetitions and are indicated as percentage in every branch. The amino acid alignment used to generate this tree corresponds to the bZIP domain plus the conserved motifs within this group (Figure 2 and Table S2). Rice, black cottonwood and Arabidopsis sequences are represented in orange, dark blue and light blue, respectively. Other eudicot and monocot sequences are show in green and red, respectively. The organism from which the remaining monocot and eudicot bZIPs originated is indicated by the last two letters in each sequence. Abbreviations are explained in Table S6.
Found at: doi:10.1371/journal.pone.0002944.s007 (1.31 MB TIF)

Figure S8 Phylogenetic tree of Group E bZIPs from monocots and eudicots. An unrooted tree was inferred by a NJ analysis from distances calculated with the PAM distance matrix. The bootstrap values correspond to 1000 repetitions and are indicated as percentage in every branch. The amino acid alignment used to generate this tree corresponds to the bZIP domain plus the conserved motifs within this group (Figure 2 and Table S2). Rice, black cottonwood and Arabidopsis sequences are represented in orange, dark blue and light blue, respectively. The organism from which the remaining monocot and eudicot bZIPs originated is indicated by the last two letters in each sequence. Abbreviations are explained in Table S6.
Found at: doi:10.1371/journal.pone.0002944.s008 (0.31 MB TIF)

Figure S9 Phylogenetic tree of Group F bZIPs from monocots and eudicots. An unrooted tree was inferred by a NJ analysis from distances calculated with the PAM distance matrix. The bootstrap values correspond to 1000 repetitions and are indicated as percentage in every branch. The amino acid alignment used to generate this tree corresponds to the bZIP domain plus the conserved motif within this group (Figure 2 and Table S2). Rice,

black cottonwood and Arabidopsis sequences are represented in orange, dark blue and light blue, respectively. Other eudicot and monocot sequences are show in green and red, respectively. The organism from which the remaining monocot and eudicot bZIPs originated is indicated by the last two letters in each sequence. Abbreviations are explained in Table S6.
Found at: doi:10.1371/journal.pone.0002944.s009 (0.83 MB TIF)

Figure S10 Phylogenetic tree of Group G bZIPs from monocots and eudicots. An unrooted tree was inferred by a NJ analysis from distances calculated with the PAM distance matrix. The bootstrap values correspond to 1000 repetitions and are indicated as percentage in every branch. The amino acid alignment used to generate this tree corresponds to the bZIP domain plus the conserved motifs within this group (Figure 2 and Table S2). Rice, black cottonwood and Arabidopsis sequences are represented in orange, dark blue and light blue, respectively. Other eudicot and monocot sequences are show in green and red, respectively. The organism from which the remaining monocot and eudicot bZIPs originated is indicated by the last two letters in each sequence. Abbreviations are explained in Table S6.
Found at: doi:10.1371/journal.pone.0002944.s010 (1.03 MB TIF)

Figure S11 Phylogenetic tree of Group H bZIPs from monocots and eudicots. An unrooted tree was inferred by a NJ analysis from distances obtained from a PAM distance matrix. The bootstrap values correspond to 1000 repetitions and are indicated as percentage in every branch. The amino acid alignment used to generate this tree corresponds to the bZIP domain plus the conserved motif within this group (Figure 2 and Table S2). Rice, black cottonwood and Arabidopsis sequences are represented in orange, dark blue and light blue, respectively. Other eudicot and monocot sequences are show in green and red, respectively. The organism from which the remaining monocot and eudicot bZIPs originated is indicated by the last two letters in each sequence. Abbreviations are explained in Table S6.
Found at: doi:10.1371/journal.pone.0002944.s011 (0.85 MB TIF)

Figure S12 Phylogenetic tree of Group I bZIPs from monocots and eudicots. An unrooted tree was inferred by a NJ analysis from distances obtained from a PAM distance matrix. The bootstrap values correspond to 1000 repetitions and are indicated as percentage in every branch. The amino acid alignment used to generate this tree corresponds to the bZIP domain plus the conserved motifs within this group (Figure 2 and Table S2). Rice, black cottonwood and Arabidopsis sequences are represented in orange, dark blue and light blue, respectively. Other eudicot sequences are show in green. The organism from which the remaining monocot and eudicot bZIPs originated is indicated by the last two letters in each sequence. Abbreviations are explained in Table S6.
Found at: doi:10.1371/journal.pone.0002944.s012 (1.12 MB TIF)

Figure S13 Phylogenetic tree of Group J bZIPs from monocots and eudicots. An unrooted tree was inferred by a NJ analysis from distances obtained from a PAM distance matrix. The bootstrap values correspond to 1000 repetitions and are indicated as percentage in every branch. The amino acid alignment used to generate this tree corresponds to the bZIP domain plus the conserved motifs within this group (Figure 2 and Table S2). Rice, black cottonwood and Arabidopsis sequences are represented in orange, dark blue and light blue, respectively. The organism from which the remaining monocot and eudicot bZIPs originated is indicated by the last two letters in each sequence. Abbreviations are explained in Table S6.
Found at: doi:10.1371/journal.pone.0002944.s013 (0.14 MB TIF)

Figure S14 Phylogenetic tree of Group K bZIPs from monocots and eudicots. An unrooted tree was inferred by a NJ analysis from distances obtained from a PAM distance matrix. The bootstrap values correspond to 1000 repetitions and are indicated as percentage in every branch. The amino acid alignment used to generate this tree corresponds to the bZIP domain plus the conserved motif within this group (Figure 2 and Table S2). Rice, black cottonwood and Arabidopsis sequences are represented in orange, dark blue and light blue, respectively. Other eudicots and monocot sequences are show in green and red, respectively. The organism from which the remaining monocot and eudicot bZIPs originated is indicated by the last two letters in each sequence. Abbreviations are explained in Table S6.
Found at: doi:10.1371/journal.pone.0002944.s014 (0.82 MB TIF)

Figure S15 Phylogenetic tree of Group L bZIPs from monocots and eudicots. An unrooted tree was inferred by a NJ analysis from distances obtained from a PAM distance matrix. The bootstrap values correspond to 1000 repetitions and are indicated as percentage in every branch. The amino acid alignment used to generate this tree corresponds to the bZIP domain plus the conserved motifs within this group (Figure 2 and Table S2). Rice, black cottonwood and Arabidopsis sequences are represented in orange, dark blue and light blue, respectively. The organism from which the remaining monocot and eudicot bZIPs originated is indicated by the last two letters in each sequence. Abbreviations are explained in Table S6.
Found at: doi:10.1371/journal.pone.0002944.s015 (0.47 MB TIF)

Figure S16 Phylogenetic tree of Group S bZIPs from monocots and eudicots. An unrooted tree was inferred by a NJ analysis from distances obtained from a PAM distance matrix. The bootstrap values correspond to 1000 repetitions and are indicated as percentage in every branch. The amino acid alignment used to generate this tree corresponds to the bZIP domain. Rice, black cottonwood and Arabidopsis sequences are represented in orange, dark blue and light blue, respectively. Other eudicot and monocot sequences are show in green and red, respectively. The organism from which the remaining monocot and eudicot bZIPs originated is indicated by the last two letters in each sequence. Abbreviations are explained in Table S6.
Found at: doi:10.1371/journal.pone.0002944.s016 (2.04 MB TIF)

Figure S17 Gene amplification pattern in each angiosperm group of bZIP homologues.
Found at: doi:10.1371/journal.pone.0002944.s017 (0.77 MB TIF)

Figure S18 Identification of Groups cI and cII. Both trees are a partial representation of the whole tree obtained by NJ analyses. (A) In the initial phylogenetic analysis with the complete ViridiZIP set, we were able to identify two clusters of genes that did not posses any member from angiosperms; therefore, we called them NA (non-angiosperm). (B) Restricted analyses including bZIPs from algae and mosses uncovered the relationship of Groups NA and C; both groups share the same homologue in Ostreococcus (*OtbZIP5*), indicating it to be a common ancestor. Group NA was re-classified into Groups cI and cII. Their relation to members of Group NA shown in (A) is indicated by stars (* for Group cII, or ** for Group cI). Groups cI, cII, C and OtbZIP5 form the Group Proto-C. The bootstrap support of each group is shown in the figure.
Found at: doi:10.1371/journal.pone.0002944.s018 (2.44 MB TIF)

Figure S19 Evolution of the bZIP family of transcription factors in land plants. We estimated the birth-and-death parameter (λ) using CAFE, as described in Materials and Methods. (A) The examined values of λ ranged from 1.0×10^{-4} to 6.8×10^{-3}. The log probabilities obtained for each assayed value are shown. The

shadowed region is displayed at a higher scale in the inset, where a peak at $\lambda = 0.002011$ is observed. (B) Evolutionary relationships of land plants with divergence time points (Arabidopsis - black cottonwood, 100–120 million years ago (mya) (47); monocot - eudicot, 140–150 mya (57); Physcomitrella - angiosperms, 450 mya (58)). Numbers at the branch end points indicate the numbers of bZIPs observed in the extant species. Numbers at the nodes represent the expected number of bZIPs in the ancestral species. Using the three methods available in CAFE, i.e., Viterbi assignments, branch cutting and likelihood ratio test, we identified branches deviating from the background model. According to all three methods, the branch leading to angiosperms significantly deviates from the null model ($p<0.05$), which implies that there was a significant increase in the number of bZIPs in the lineage leading to that group. Similarly, the Viterbi and branch cutting methods identify the branch leading to bryophytes (Physcomitrella) exhibiting a significant reduction in the number of bZIPs ($p<0.05$). Finally, we did not observe any significant deviation of the model for the extant group of angiosperms which can be interpreted as an even diffusion of the number of bZIPs in each branch. However, one cannot exclude the effect of natural selection in accounting for the differences that are nevertheless occurring. The increased number of bZIPs in the branch leading to angiosperms might be, at least partly, related to the several genome-wide duplication events that took place in the history of that lineage.
Found at: doi:10.1371/journal.pone.0002944.s019 (1.62 MB TIF)

Figure S20 Scheme of the pipeline for bZIP identification in genomic sequences and ESTs. (I) Input genomic and EST sequences are compared by tblastn with the Angiotot protein dataset, generating a group of sequences that putatively code for bZIPs (SeqZIP). (II) Manual curation allowed subtracting sequences already present in Angiotot (redundancies) and false positives, which mainly correspond to low-complexity sequences. (III) The remaining sequences (true positives) are compared by tblastx against the best hit from Angiotot obtained in step I, allowing to identify the most probable ORF, and in the case of genomic sequences, to identify their gene structure, taking into account conserved intron positions and the presence of canonic splicing sites (GT-AG).
Found at: doi:10.1371/journal.pone.0002944.s020 (0.75 MB TIF)

Table S1 Comparison between bZIPs reported in this manuscript and in Nijhawan et al. (2008)
Found at: doi:10.1371/journal.pone.0002944.s021 (0.04 MB XLS)

Table S2 Conserved motifs in bZIP PoGOs.
Found at: doi:10.1371/journal.pone.0002944.s022 (0.01 MB PDF)

Table S3 Accession numbers and classification into groups of homologues of non-sequenced angiosperms.
Found at: doi:10.1371/journal.pone.0002944.s023 (0.03 MB PDF)

Table S4 Biological functions of genes in PoGOs.
Found at: doi:10.1371/journal.pone.0002944.s024 (0.02 MB PDF)

Table S5 Classification of non-angiosperm bZIPs.
Found at: doi:10.1371/journal.pone.0002944.s025 (0.02 MB XLS)

Table S6 Organism abbreviations.
Found at: doi:10.1371/journal.pone.0002944.s026 (0.03 MB XLS)

Table S7 Gene pairs resulting from segmental duplications of the Arabidopsis genome.
Found at: doi:10.1371/journal.pone.0002944.s027 (0.03 MB DOC)

Dataset S1 Re-annotated nucleotide sequences from rice and black cottonwood.
Found at: doi:10.1371/journal.pone.0002944.s028 (0.02 MB TXT)

Dataset S2 Re-annotated amino acid sequences from rice and black cottonwood.
Found at: doi:10.1371/journal.pone.0002944.s029 (0.01 MB TXT)

Dataset S3 MPSS Expression data for bZIP genes from rice.
Found at: doi:10.1371/journal.pone.0002944.s030 (0.02 MB PDF)

Dataset S4 MPSS Expression data for bZIP genes from Arabidopsis.
Found at: doi:10.1371/journal.pone.0002944.s031 (0.01 MB PDF)

Text S1 Supporting texts including further results and discussion.
Found at: doi:10.1371/journal.pone.0002944.s032 (0.06 MB DOC)

Acknowledgments

We thank Amanda Bortolini Silveira (Universidade Estadual de Campinas, Brazil) for nuclear localisation experiments on Group L bZIPs, and Stefanie Hartmann (University of Potsdam) for critical comments on our manuscript, Liam Childs (MPI of Molecular Plant Physiology, Potsdam) for improving our English and the two reviewers for their helpful comments on the manuscript.

Author Contributions

Conceived and designed the experiments: LGGC CGS RVRVdS MV. Performed the experiments: LGGC DMRP RVRVdS. Analyzed the data: LGGC DMRP CGS MV. Contributed reagents/materials/analysis tools: BMR. Wrote the paper: LGGC DMRP BMR MV.

References

1. Meshi T, Iwabuchi M (1995) Plant transcription factors. Plant Cell Physiol 36: 1405–1420.
2. Beckett D (2001) Regulated assembly of transcription factors and control of transcription initiation. J Mol Biol 314: 335–352.
3. Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, et al. (2003) The evolution of transcriptional regulation in eukaryotes. Mol Biol Evol 20: 1377–1419.
4. Warren AJ (2002) Eukaryotic transcription factors. Curr Opin Struct Biol 12: 107–114.
5. Wingender E, Chen X, Fricke E, Geffers R, Hehl R, et al. (2001) The TRANSFAC system on gene expression regulation. Nucleic Acids Res 29: 281–283.
6. Riechmann JL, Ratcliffe OJ (2000) A genomic perspective on plant transcription factors. Curr Opin Plant Biol 3: 423–434.
7. Hsia CC, McGinnis W (2003) Evolution of transcription factor function. Curr Opin Genet Dev 13: 199–206.
8. Chen K, Rajewsky N (2007) The evolution of gene regulation by transcription factors and microRNAs. Nat Rev Genet 8: 93–103.
9. Lawton-Rauh A (2003) Evolutionary dynamics of duplicated genes in plants. Mol Phylogenet Evol 29: 396–409.
10. Levine M, Tjian R (2003) Transcription regulation and animal diversity. Nature 424: 147–151.
11. Shiu SH, Shih MC, Li WH (2005) Transcription factor families have much higher expansion rates in plants than in animals. Plant Physiol 139: 18–26.
12. Riaño-Pachón DM, Corrêa LGG, Trejos-Espinosa R, Mueller-Roeber B (2008) Green transcription factors: a chlamydomonas overview. Genetics 179: 31–39.
13. Irish VF (2003) The evolution of floral homeotic gene function. Bioessays 25: 637–646.
14. Garcia-Fernandez J (2005) The genesis and evolution of homeobox gene clusters. Nat Rev Genet 6: 881–892.
15. Deppmann CD, Acharya A, Rishi V, Wobbes B, Smeekens S, et al. (2004) Dimerization specificity of all 67 B-ZIP motifs in *Arabidopsis thaliana*: a comparison to Homo sapiens B-ZIP motifs. Nucleic Acids Res 32: 3435–3445.
16. Floyd SK, Bowman JL (2007) The ancestral developmental tool kit of land plants. Int J Plant Sci 168: 1–35.

17. Bowman JL, Floyd SK, Sakakibara K (2007) Green genes-comparative genomics of the green branch of life. Cell 129: 229–234.
18. Moreno-Risueno MA, Martinez M, Vicente-Carbajosa J, Carbonero P (2007) The family of DOF transcription factors: from green unicellular algae to vascular plants. Mol Genet Genomics 277: 379–390.
19. Derelle R, Lopez P, Le GH, Manuel M (2007) Homeodomain proteins belong to the ancestral molecular toolkit of eukaryotes. Evol Dev 9: 212–219.
20. Martinez-Castilla LP, Alvarez-Buylla ER (2003) Adaptive evolution in the Arabidopsis MADS-box gene family inferred from its complete resolved phylogeny. Proc Natl Acad Sci U S A 100: 13407–13412.
21. Parenicova L, de Folter S, Kieffer M, Horner DS, Favalli C, et al. (2003) Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. Plant Cell 15: 1538–1551.
22. Zhang Y, Wang L (2005) The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants. BMC Evol Biol 5: 1.
23. Prigge MJ, Clark SE (2006) Evolution of the class III HD-Zip gene family in land plants. Evol Dev 8: 350–361.
24. Floyd SK, Zalewski CS, Bowman JL (2006) Evolution of class III homeodomain-leucine zipper genes in streptophytes. Genetics 173: 373–388.
25. Hurst HC (1995) Transcription factors 1: bZIP proteins. Protein Profile 2: 101–168.
26. Walsh J, Waters CA, Freeling M (1998) The maize gene liguleless2 encodes a basic leucine zipper protein involved in the establishment of the leaf blade-sheath boundary. Genes Dev 12: 208–218.
27. Chuang CF, Running MP, Williams RW, Meyerowitz EM (1999) The *PERIANTHIA* gene encodes a bZIP protein involved in the determination of floral organ number in *Arabidopsis thaliana*. Genes Dev 13: 334–344.
28. Abe M, Kobayashi Y, Yamamoto S, Daimon Y, Yamaguchi A, et al. (2005) FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. Science 309: 1052–1056.
29. Silveira AB, Gauer L, Tomaz JP, Cardoso PR, Carmello-Guerreiro S, et al. (2007) The Arabidopsis AtbZIP9 protein fused to the VP16 transcriptional activation domain alters leaf and vascular development. Plant Sci 172: 1148–1156.
30. Shen H, Cao K, Wang X (2007) A conserved proline residue in the leucine zipper region of AtbZIP34 and AtbZIP61 in *Arabidopsis thaliana* interferes with the formation of homodimer. Biochem Biophys Res Commun 362: 425–430.
31. Yin Y, Zhu Q, Dai S, Lamb C, Beachy RN (1997) RF2a, a bZIP transcriptional activator of the phloem-specific rice tungro bacilliform virus promoter, functions in vascular development. EMBO J 16: 5247–5259.
32. Fukazawa J, Sakai T, Ishida S, Yamaguchi I, Kamiya Y, et al. (2000) Repression of shoot growth, a bZIP transcriptional activator, regulates cell elongation by controlling the level of gibberellins. Plant Cell 12: 901–915.
33. Ciceri P, Locatelli F, Genga A, Viotti A, Schmidt RJ (1999) The activity of the maize Opaque2 transcriptional activator is regulated diurnally. Plant Physiol 121: 1321–1328.
34. Weltmeier F, Ehlert A, Mayer CS, Dietrich K, Wang X, et al. (2006) Combinatorial control of Arabidopsis proline dehydrogenase transcription by specific heterodimerisation of bZIP transcription factors. EMBO J 25: 3133–3143.
35. Zhang B, Foley RC, Singh KB (1993) Isolation and characterization of two related Arabidopsis ocs-element bZIP binding proteins. Plant J 4: 711–716.
36. Despres C, DeLong C, Glaze S, Liu E, Fobert PR (2000) The Arabidopsis NPR1/NIM1 protein enhances the DNA binding activity of a subgroup of the TGA family of bZIP transcription factors. Plant Cell 12: 279–290.
37. Pontier D, Miao ZH, Lam E (2001) Trans-dominant suppression of plant TGA factors reveals their negative and positive roles in plant defense responses. Plant J 27: 529–538.
38. Niggeweg R, Thurow C, Kegler C, Gatz C (2000) Tobacco transcription factor TGA2.2 is the main component of as-1-binding factor ASF-1 and is involved in salicylic acid- and auxin-inducible expression of as-1-containing target promoters. J Biol Chem 275: 19897–19905.
39. Thurow C, Schiermeyer A, Krawczyk S, Butterbrodt T, Nickolov K, et al. (2005) Tobacco bZIP transcription factor TGA2.2 and related factor TGA2.1 have distinct roles in plant defense responses and plant development. Plant J 44: 100–113.
40. Kaminaka H, Nake C, Epple P, Dittgen J, Schutze K, et al. (2006) bZIP10-LSD1 antagonism modulates basal defense and cell death in Arabidopsis following infection. EMBO J 25: 4400–4411.
41. Baena-Gonzalez E, Rolland F, Thevelein JM, Sheen J (2007) A central integrator of transcription networks in plant stress and energy signalling. Nature 448: 938–943.
42. Liu JX, Srivastava R, Che P, Howell SH (2007) Salt stress responses in Arabidopsis utilize a signal transduction pathway related to endoplasmic reticulum stress signaling. Plant J 51: 897–909.
43. Iwata Y, Koizumi N (2005) An Arabidopsis transcription factor, AtbZIP60, regulates the endoplasmic reticulum stress response in a manner unique to plants. Proc Natl Acad Sci U S A 102: 5280–5285.
44. Finkelstein RR, Lynch TJ (2000) Abscisic acid inhibition of radicle emergence but not seedling growth is suppressed by sugars. Plant Physiol 122: 1179–1186.
45. Uno Y, Furihata T, Abe H, Yoshida R, Shinozaki K, et al. (2000) Arabidopsis basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. Proc Natl Acad Sci U S A 97: 11632–11637.
46. Niggeweg R, Thurow C, Weigel R, Pfitzner U, Gatz C (2000) Tobacco TGA factors differ with respect to interaction with NPR1, activation potential and DNA-binding properties. Plant Mol Biol 42: 775–788.
47. Nieva C, Busk PK, Dominguez-Puigjaner E, Lumbreras V, Testillano PS, et al. (2005) Isolation and functional characterisation of two new bZIP maize regulators of the ABA responsive gene rab28. Plant Mol Biol 58: 899–914.
48. Wellmer F, Kircher S, Rugner A, Frohnmeyer H, Schafer E, et al. (1999) Phosphorylation of the parsley bZIP transcription factor CPRF2 is regulated by light. J Biol Chem 274: 29476–29482.
49. Osterlund MT, Hardtke CS, Wei N, Deng XW (2000) Targeted destabilization of HY5 during light-regulated development of Arabidopsis. Nature 405: 462–466.
50. Ulm R, Baumann A, Oravecz A, Mate Z, Adam E, et al. (2004) Genome-wide analysis of gene expression reveals function of the bZIP transcription factor HY5 in the UV-B response of Arabidopsis. Proc Natl Acad Sci U S A 101: 1397–1402.
51. Satoh R, Fujita Y, Nakashima K, Shinozaki K, Yamaguchi-Shinozaki K (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: 309–317.
52. Lara P, Onate-Sanchez L, Abraham Z, Ferrandiz C, Diaz I, et al. (2003) Synergistic activation of seed storage protein gene expression in Arabidopsis by ABI3 and two bZIPs related to OPAQUE2. J Biol Chem 278: 21003–21011.
53. Vettore AL, Yunes JA, Cord NG, da Silva MJ, Arruda P, et al. (1998) The molecular and functional characterization of an Opaque2 homologue gene from Coix and a new classification of plant bZIP proteins. Plant Mol Biol 36: 249–263.
54. Arabidopsis Genome Iniciative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796–815.
55. Jakoby M, Weisshaar B, Droge-Laser W, Vicente-Carbajosa J, Tiedemann J, et al. (2002) bZIP transcription factors in Arabidopsis. Trends Plant Sci 7: 106–111.
56. Vincentz M, Bandeira-Kobarg C, Gauer L, Schlogl P, Leite A (2003) Evolutionary pattern of angiosperm bZIP factors homologous to the maize Opaque2 regulatory protein. J Mol Evol 56: 105–116.
57. Yu J, Hu S, Wang J, Wong GK, Li S, et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296: 79–92.
58. Goff SA, Ricke D, Lan TH, Presting G, Wang R, et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296: 92–100.
59. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, et al. (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313: 1596–1604.
60. Bennetzen J (2002) The rice genome. Opening the door to comparative plant biology. Science 296: 60–63.
61. Pennacchio LA (2003) Insights from human/mouse genome comparisons. Mamm Genome 14: 429–436.
62. Vincentz M, Cara FA, Okura VK, da Silva FR, Pedrosa GL, et al. (2004) Evaluation of monocot and eudicot divergence using the sugarcane transcriptome. Plant Physiol 134: 951–959.
63. Tatusov RL, Koonin EV, Lipman DJ (1997) A genomic perspective on protein families. Science 278: 631–637.
64. Adams KL (2007) Evolution of duplicate gene expression in polyploid and hybrid plants. J Hered 98: 136–141.
65. Rijpkema AS, Gerats T, Vandenbussche M (2007) Evolutionary complexity of MADS complexes. Curr Opin Plant Biol 10: 32–38.
66. Woolfe A, Elgar G (2007) Comparative genomics using Fugu reveals insights into regulatory subfunctionalization. Genome Biol 8: R53.
67. Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, et al. (2007) The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science 318: 245–250.
68. Derelle E, Ferraz C, Rombauts S, Rouze P, Worden AZ, et al. (2006) Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. Proc Natl Acad Sci U S A 103: 11647–11652.
69. Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, et al. (2008) The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science 319: 64–69.
70. Corrêa LGG (2004) Análise Filogenética de Fatores de Transcrição bZIP em Angiospermas. (Phylogenetic analyses of bZIP transcription factors in angiosperms) [dissertation]. Universidade Estadual de Campinas, Campinas, Brazil.
71. Nijhawan A, Jain M, Tyagi AK, Khurana JP (2008) Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice. Plant Physiol 146: 333–350.
72. Thornton JW, DeSalle R (2000) Gene family evolution and homology: genomics meets phylogenetics. Annu Rev Genomics Hum Genet 1: 41–73.
73. Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol 15: 568–573.
74. Kellogg EA (2004) Evolution of developmental traits. Curr Opin Plant Biol 7: 92–98.
75. Nam J, Kim J, Lee S, An G, Ma H, et al. (2004) Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. Proc Natl Acad Sci U S A 101: 1910–1915.

76. Lynch M, Katju V (2004) The altered evolutionary trajectories of gene duplicates. Trends Genet 20: 544–549.
77. Carlini LE, Ketudat M, Parsons RL, Prabhakar S, Schmidt RJ, et al. (1999) The maize EmBP-1 orthologue differentially regulates opaque2-dependent gene expression in yeast and cultured maize endosperm cells. Plant Mol Biol 41: 339–349.
78. Vinson C, Myakishev M, Acharya A, Mir AA, Moll JR, et al. (2002) Classification of human B-ZIP proteins based on dimerization properties. Mol Cell Biol 22: 6321–6335.
79. Fassler J, Landsman D, Acharya A, Moll JR, Bonovich M, et al. (2002) B-ZIP proteins encoded by the Drosophila genome: evaluation of potential dimerization partners. Genome Res 12: 1190–1200.
80. Deppmann CD, Alvania RS, Taparowsky EJ (2006) Cross-species annotation of basic leucine zipper factor interactions: Insight into the evolution of closed interaction networks. Mol Biol Evol 23: 1480–1492.
81. Best AA, Morrison HG, McArthur AG, Sogin ML, Olsen GJ (2004) Evolution of eukaryotic transcription: insights from the genome of *Giardia lamblia*. Genome Res 14: 1537–1547.
82. Alvarez-Buylla ER, Pelaz S, Liljegren SJ, Gold SE, Burgeff C, et al. (2000) An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. Proc Natl Acad Sci U S A 97: 5328–5333.
83. Singer SD, Krogan NT, Ashton NW (2007) Clues about the ancestral roles of plant MADS-box genes from a functional analysis of moss homologues. Plant Cell Rep 26: 1155–1169.
84. Tanabe Y, Hasebe M, Sekimoto H, Nishiyama T, Kitani M, et al. (2005) Characterization of MADS-box genes in charophycean green algae and its implication for the evolution of MADS-box genes. Proc Natl Acad Sci U S A 102: 2436–2441.
85. Shigyo M, Tabei N, Yoneyama T, Yanagisawa S (2007) Evolutionary processes during the formation of the plant-specific Dof transcription factor family. Plant Cell Physiol 48: 179–185.
86. Holm M, Ma LG, Qu LJ, Deng XW (2002) Two interacting bZIP proteins are direct targets of COP1-mediated control of light-dependent gene expression in Arabidopsis. Genes Dev 16: 1247–1259.
87. Richardt S, Lang D, Reski R, Frank W, Rensing SA (2007) PlanTAPDB, a phylogeny-based resource of plant transcription-associated proteins. Plant Physiol 143: 1452–1466.
88. Yi C, Deng XW (2005) COP1 - from plant photomorphogenesis to mammalian tumorigenesis. Trends Cell Biol 15: 618–625.
89. Kamisugi Y, Cuming AC (2005) The evolution of the abscisic acid-response in land plants: comparative analysis of group 1 *LEA* gene expression in moss and cereals. Plant Mol Biol 59: 723–737.
90. Marella HH, Sakata Y, Quatrano RS (2006) Characterization and functional analysis of ABSCISIC ACID INSENSITIVE3-like genes from *Physcomitrella patens*. Plant J 46: 1032–1044.
91. Singer SD, Ashton NW (2007) Revelation of ancestral roles of KNOX genes by a functional analysis of Physcomitrella homologues. Plant Cell Rep 26: 2039–2054.
92. Vinson C, Acharya A, Taparowsky EJ (2006) Deciphering B-ZIP transcription factor interactions *in vitro* and *in vivo*. Biochim Biophys Acta 1759: 4–12.
93. Lawrence CL, Maekawa H, Worthington JL, Reiter W, Wilkinson CR, et al. (2007) Regulation of *Schizosaccharomyces pombe* Atf1 protein levels by Sty1-mediated phosphorylation and heterodimerization with Pcr1. J Biol Chem 282: 5160–5170.
94. Rodrigues-Pousada CA, Nevitt T, Menezes R, Azevedo D, Pereira J, et al. (2004) Yeast activator proteins and stress response: an overview. FEBS Lett 567: 80–85.
95. Jaiswal AK (2004) Nrf2 signaling in coordinated activation of antioxidant gene expression. Free Radic Biol Med 36: 1199–1207.
96. Warabi E, Takabe W, Minami T, Inoue K, Itoh K, et al. (2007) Shear stress stabilizes NF-E2-related factor 2 and induces antioxidant genes in endothelial cells: role of reactive oxygen/nitrogen species. Free Radic Biol Med 42: 260–269.
97. Makino C, Sano Y, Shinagawa T, Millar JB, Ishii S (2006) Sin1 binds to both ATF-2 and p38 and enhances ATF-2-dependent transcription in an SAPK signaling pathway. Genes Cells 11: 1239–1251.
98. Yoshida H, Haze K, Yanagi H, Yura T, Mori K (1998) Identification of the *cis*-acting endoplasmic reticulum stress response element responsible for transcriptional induction of mammalian glucose-regulated proteins. Involvement of basic leucine zipper transcription factors. J Biol Chem 273: 33741–33749.
99. Cox JS, Walter P (1996) A novel mechanism for regulating activity of a transcription factor that controls the unfolded protein response. Cell 87: 391–404.
100. Hahn MW, De Bie T, Stajich JE, Nguyen C, Cristianini N (2005) Estimating the tempo and mode of gene family evolution from comparative genomic data. Genome Res 15: 1153–1160.
101. Demuth JP, Wade MJ (2007) Maternal expression increases the rate of bicoid evolution by relaxing selective constraint. Genetica 129: 37–43.
102. Riaño-Pachón DM, Ruzicic S, Dreyer I, Mueller-Roeber B (2007) PlnTFDB: an integrative plant transcription factor database. BMC Bioinformatics 8: 42.
103. Childs KL, Hamilton JP, Zhu W, Ly E, Cheung F, et al. (2007) The TIGR Plant Transcript Assemblies database. Nucleic Acids Res 35: D846–D851.
104. Pertea G, Huang X, Liang F, Antonescu V, Sultana R, et al. (2003) TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 19: 651–652.
105. Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, et al. (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. Science 301: 376–379.
106. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.
107. Birney E, Clamp M, Durbin R (2004) GeneWise and Genomewise. Genome Res 14: 988–995.
108. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl Acids Res 25: 4876–4882.
109. Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform 5: 150–163.
110. Felsenstein J (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics 5: 164–166.
111. Dayhoff MO, Schwartz RC, Orcutt BC (1978) A model of evolutionary change in proteins. In: Dayhoff MO, ed. Atlas of protein sequence and structure. Silver Spring, MD: National Biomedical Research Fundation Silver. pp 301–310.
112. Nei M, Kumar S (2000) Molecular Evolution and Phylogenetics. New York: Oxford University Press.
113. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688–2690.
114. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8: 275–282.
115. Suzuki Y, Glazko GV, Nei M (2002) Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. Proc Natl Acad Sci U S A 99: 16138–16143.
116. Swofford DL (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates).
117. Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10: 512–526.
118. Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics 14: 817–818.
119. Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18: 502–504.
120. Bailey TL, Elkan C (1995) The value of prior knowledge in discovering motifs with MEME. Proc Int Conf Intell Syst Mol Biol 3: 21–29.
121. Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, et al. (2006) Pfam: clans, web tools and services. Nucleic Acids Res 34: D247–D251.
122. De Bie T, Cristianini N, Demuth JP, Hahn MW (2006) CAFE: a computational tool for the study of gene family evolution. Bioinformatics 22: 1269–1271.
123. Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W (2004) GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. Plant Physiol 136: 2621–2632.
124. Boyes DC, Zayed AM, Ascenzi R, McCaskill AJ, Hoffman NE, et al. (2001) Growth stage-based phenotypic analysis of Arabidopsis: a model for high throughput functional genomics in plants. Plant Cell 13: 1499–1510.
125. Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, et al. (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol 18: 630–634.

4

Phylogenetic analysis of NAC transcription factors and their role in plant senescence

Phylogenetic analysis of NAC transcription factors and their role in plant senescence

Luiz Gustavo Guedes Corrêa[1,2], Diego Mauricio Riaño-Pachón[1,2], Stefan Rensing [3], Salma Balazadeh[1,2], Bernd Mueller-Roeber[1,2,†]

[1]Department of Molecular Biology, University of Potsdam, Potsdam–Golm, Germany, [2]Max-Planck Institute of Molecular Plant Physiology, Potsdam–Golm, Germany,[3]University of Freiburg, Faculty of Biology, Hauptstr. 1, D-79104 Freiburg, Germany.

[†] Author for correspondence. Email: bmr@uni-potsdam.de

Manuscript to be submitted for publication.

Author contributions

LGGC conceived the idea for the study. The research was designed and planned by all authors. The experimental work was done by LGGC (expression analyses in rice and barley) and SB (expression analyses in Arabidopsis, rice and barley). DMRP and LGGC performed the phylogenetic studies. SR provided the data from Physcomitrella. BMR supervised the project. All authors contributed to manuscript writing.

Abstract

NACs constitute one of the biggest transcription factor families in plants. They are involved in most diverse processes in plants including tissue differentiation, senescence and hormone cross-talk. NAC is a plant specific transcription factor and can only be identified in embryophytes. Here we performed a comprehensive phylogenetic analysis of NACs from mosses, fern allies and angiosperms. We identified 45 groups of homologues, from which 13 are present in all angiosperms. The massive gene amplification and the high number of lineage-specific genes suggest that NACs played an important role in the acquisition of new features and conquering new environments. Expression analysis indicates that NACs are overrepresented among transcription factors that are differentially expressed during senescence. Similar expression analysis performed in rice allowed identifying 44 NACs differentially expressed. Most of NACs from Arabidopsis and rice cluster into four groups of homologues. This backbone was used to direct target ten orthologues in barley that could be differentially expressed in senescence; eight of these NACs showed high induction upon senescence.

4.1 Introduction

Over the course of evolution, the rise of new gene regulatory networks allowed conquering new environments and responding to different stresses. One of the main players of these networks are transcription factors (TFs), proteins that control the initiation rate of gene transcription by binding to specific sequences (*cis*-elements) upstream of the transcription start site (TSS) of a gene. TFs can be grouped into evolutionarily related gene families. Many TF families are present across all eukaryotic kingdoms, and, in some cases, it is possible to identify their common origin or even ancestral functions (ALVAREZ-BUYLLA *et al.* 2000, CORRÊA *et al.* 2008). Other families are kingdom-specific innovations, constituting a rich source of novelties. NACs (NAM, ATAF and CUC) are plant-specific TFs that play an important role in regulating processes like tissue differentiation (HE *et al.* 2005, KO *et al.* 2007, YAMAGUCHI *et al.* 2008, ZHONG *et al.* 2007), senescence (BALAZADEH *et al.* 2008, GUO and GAN 2006, UAUY *et al.* 2006, YOON *et al.* 2008), programmed cell death (ZHAO *et al.* 2008), defense response (BU *et al.* 2008), hormone cross-talk (BU *et al.* 2008, HE *et al.* 2005, KIM *et al.* 2008, NAKANO *et al.* 2006), seed germination (KIM *et al.* 2008, UAUY *et al.* 2006), cell wall synthesis (MITSUDA *et al.* 2007, ZHAO *et al.* 2008, ZHONG *et al.* 2007, ZHONG and YE 2007), and cell division (KIM *et al.* 2008). The NAC family of TFs is restricted to embryophytes (land plants), whereas other plant-specific TF families, such as DOF, are

found in all green plants (MORENO-RISUENO *et al.* 2007, RIAÑO PACHÓN *et al.* 2008). Previous studies have shown that NACs play a distinctively important role in cell differentiation (OLSEN *et al.* 2005); therefore, it has been hypothesized that NACs are important players on the origin of multicellularity in embryophytes, which independently appeared multiple times in evolution (KIRK 2005, LARROUX *et al.* 2008, MICHOD *et al.* 2006, SACHS 2008). Understanding the evolutionary history of TF families and their functions is crucial to reveal how innovations and/or changes in the gene regulatory networks led to the gains and losses of biological traits.

In this study, we identified the possible non-redundant complete sets of NACs in the angiosperms, black cottonwood (150 proteins) and grapevine (54), the fern-ally, *Selaginella moellendoerffii* (19) and in the moss, *Physcomitrella patens* (32), and revised the previous sets available for Arabidopsis and rice (OOKA *et al.* 2003). We identified additionally three genes in Arabidopsis (103 in total) and 41 in rice. Based on the similarity of their NAC domains, these 479 proteins were divided into 45 groups of homologues, representing an update of the groups presented before (OOKA *et al.* 2003). Curiously, the NAC family has a high number of lineage-specific groups of orthologues when compared with other transcription factor families (CORRÊA *et al.* 2008, MORENO-RISUENO *et al.* 2007), which may also reflect their late acquisition during evolution and an active role in the diversification and adaptation of plants into varied environments. Interestingly, NACs in Arabidopsis were identified to be preferentially expressed during the onset of senescence (BALAZADEH *et al.* 2008). Similar expression analyses were made for rice, suggesting that three groups of homologues play a conserved role in senescence in all angiosperms.

4.2 Materials and methods

4.2.1 Datasets of NAC genes

We generated a NAC dataset (AngioNAC) representing an updated version of the set previously used (OOKA *et al.* 2003). NAC TFs were identified as described by (RIAÑO PACHÓN *et al.* 2007). The whole proteomes deduced from the completely sequenced genomes of the bryophyte *Physcomitrella patens* (RENSING *et al.* 2008), the fern-ally *Selaginella moellendorffii* (JGI/DOE) and the angiosperm *Populus trichocarpa* (TUSKAN *et al.* 2006) were downloaded from the Joint Genome Institute/Department of Energy (JGI/DOE; `http://www.jgi.doe.gov/`). Sequences of *Vitis vinifera* (JAILLON *et al.* 2007, VELASCO *et al.* 2007), were downloaded from Istituto Agrario San Michele all'Adige (IASMA, `http://genomics.research.iasma.it/`) and Genoscope

(http://www.genoscope.cns.fr/). Protein sequences of the angiosperm *Arabidopsis thaliana* were downloaded from The Arabidopsis Information Resource (TAIR; http://www.arabidopsis.org/). Sequences of the monocot *Oryza sativa* ssp. *japonica* (GOFF *et al.* 2002) were downloaded from The J. Craig Venter Institute (http://www.tigr.org/).

4.2.2 Phylogenetic analyses

Alignment of NAC protein sequences was performed by MAFFT (KATOH *et al.* 2002), using default parameters. Alignments were subsequently adjusted manually using Jalview (CLAMP *et al.* 2004). The phylogenetic analyses based on amino acid sequences were conducted using MEGA v4.1 (TAMURA *et al.* 2007). Unrooted phylogenetic tree topologies were reconstructed by Neighbor-Joining (NJ), distances were obtained using a PAM-like distance matrix (DAYHOFF *et al.* 1978) and JTT (JONES *et al.* 1992), and the re-sampling of the original NAC set was a 1,000 bootstrap repetition. Maximum Likelihood (ML) analyses of the NAC domain were carried out using RAxML (STAMATAKIS 2006) with the distances computed using the JTT+Γ evolutionary model (JONES *et al.* 1992), and a re-sampling of the original set of 100 or 500 bootstrap repetitions, depending on the size of the set.

4.2.3 Identification of conserved motifs

The putative complete sets of unique NACs from *Physcomitrella, Selaginella*, black cottonwood, Arabidopsis, grape and rice were used as input for a conserved motif analysis performed with MEME (http://meme.sdsc.edu/meme/meme.html) (BAILEY and ELKAN 1995). Whole protein sequences were employed for this search. A given motif was allowed to appear at any number of repetitions, the maximum width of a motif was set to 80, and the maximum number of motifs was set to 20. The other parameters were used as default. In a complementary approach, each group of homologues was analyzed individually with the parameters described above.

4.2.4 Phylogenetic analyses and identification of Possible Groups of Orthologues (PoGOs)

The detailed evolutionary analysis of the evolutionary relationships among homologous groups of NAC TFs allowed the identification of PoGOs. A PoGO is defined by the following criteria: (i) members of a PoGO should have a monophyletic origin, indicated by

a bootstrap support greater than 50%; (ii) a PoGO should possesses at least one representative gene from *A. thaliana* and *O. sativa*, assuming that the putative complete sets of NAC genes of these organisms were identified and no selective gene loss had occurred. In case a PoGO is found to be restricted to either monocots or eudicots, the presence of sequences from at least one other species of the same lineage in this PoGO is required; and (iii) the inferred phylogeny should be consistent with the known phylogeny of plant species (CORRÊA *et al.* 2008, VINCENTZ *et al.* 2003).

4.2.5 Gene expression experiments

Expression of NAC genes during late stages of leaf development including senescence was assessed via qRT-PCR as in CZECHOWSKI *et al.* (2004) and BALAZADEH *et al.* (2008). Primers used for Arabidopsis are described in CZECHOWSKI *et al.* (2004) and for rice in CALDANA *et al.* (2007). An updated version of NAC primers for rice and the primers for barley are provided in Table SI. New primers were designed with QuantPrime (ARVIDSSON *et al.* 2008).

4.2.6 Conditions for growing plants

Rice plants (*Oryza sativa* L. ssp. *japonica*) were grown in hydroponic culture under a day-length of 12 h at 26/22C (day/night), 70% humidity and 700 mol m-2 s-1 light intensity. Senescing leaves were harvested and divided into three parts of 12 cm each: the top, corresponding to the senescing part, the middle, corresponding to the transition zone, and the base, corresponding to the mature zone. A scheme of the harvesting can be found in Figure S1. Barley plants (*Hordeum vulgare* L. cv Sebastien) were grown in soil under a day-length of 12 h at 26/22C (day/night), 70% humidity and 200 mol m-2 s-1 light intensity. Leaf number 5 was harvested at four developmental points: at 54 days after sowing (DAS), when leaves reached 50% of the full expanded size (15 cm); at 61 DAS, when leaves reached full expansion (30 cm); at 87 DAS, when leaves were ca. 5% senescent and 94 DAS, when leaves were 20% senescent. Arabidopsis plants were grown and harvested according to the conditions described in BALAZADEH *et al.* (2008).

4.2.7 Supplementary material

The supplementary material here cited can be found at `https://molbio00.bio.uni-potsdam.de/correa2008/`.

4.3 Results and Discussion

4.3.1 Group of Homologues of NAC genes

The NAM PFAM domain model (PF02365) was used to identify NAC proteins in the different species. This allowed us to identify three new NACs in Arabidopsis (ANAC106 to 108) compared to the previous report (OOKA *et al.* 2003), whereas five were excluded (ANAC21, 34, 39, 51 and 80) as they do not represent individual loci but alternative splicing forms (Table SII). In rice, 41 new NAC genes were identified, totaling 121 genes. Additionally we have identified 150 NAC genes in black cottonwood, 54 in grapevine, 19 in *Selaginella moellendoerffii*, and 32 in *Physcomitrella patens*. As previously observed, NAC genes were neither found in chlorophytes nor in rhodophytes (RIAÑO PACHÓN *et al.* 2008).

Neighbor-Joining (NJ) and Maximum Likelihood (ML) phylogenetic analyses were conducted for each species separately and in different combinations of organisms, including up to the whole set of 479 NACs. This approach (Figure S2) was followed to cope with the difficulties of dealing with many sequences at a time, which could lead to reduced bootstrap values for the groups of homologues, thus leaving much information uncovered. We identified 13 clusters of proteins with representatives of all angiosperms, with bootstrap support greater than or equal to 50%, defining the groups of homologues A to M (Figure 4.1 and Table SII). Groups of homologues C and D, E and F, and G and H are sister groups, respectively, as evidenced by bootstrap values linking them, thus each pair shares a common origin (Figure 4.1). Groups A and B, as well as the joint Groups CD, EF and GH include orthologues from Physcomitrella and Selaginella, thus, they might constitute the extant representatives of at least five ancestral NAC genes present in the most recent common ancestor (MRCA) of embryophytes.

Additionally, we have identified nine groups of homologues restricted to eudicots (EUDI1 to 9), three to Arabidopsis (ARA1 to 3), six to black cottonwood (POP1 to 6), three to grapevine (VIT1 to 3), ten to rice (OSA1 to 10) and one to Physcomitrella and Selaginella (SELAPHY). The fact that rice has the highest number of species-specific groups may be due to the fact that it is the only monocot representative used in these analyses. Rice is the only monocot, for which a fully sequenced genome is currently publicly available. The availability of new genomes in the near future may allow inferring functions related to the lineage-specific groups and help to better understand the gene gain/loss processes in the NAC family. The distribution of genes among the different groups of homologues can be seen in Figure 4.2.

Except for Group G, there is lack of systematic functional characterization of NACs from other groups of homologues. Group G is involved mainly in drought and salt stress,

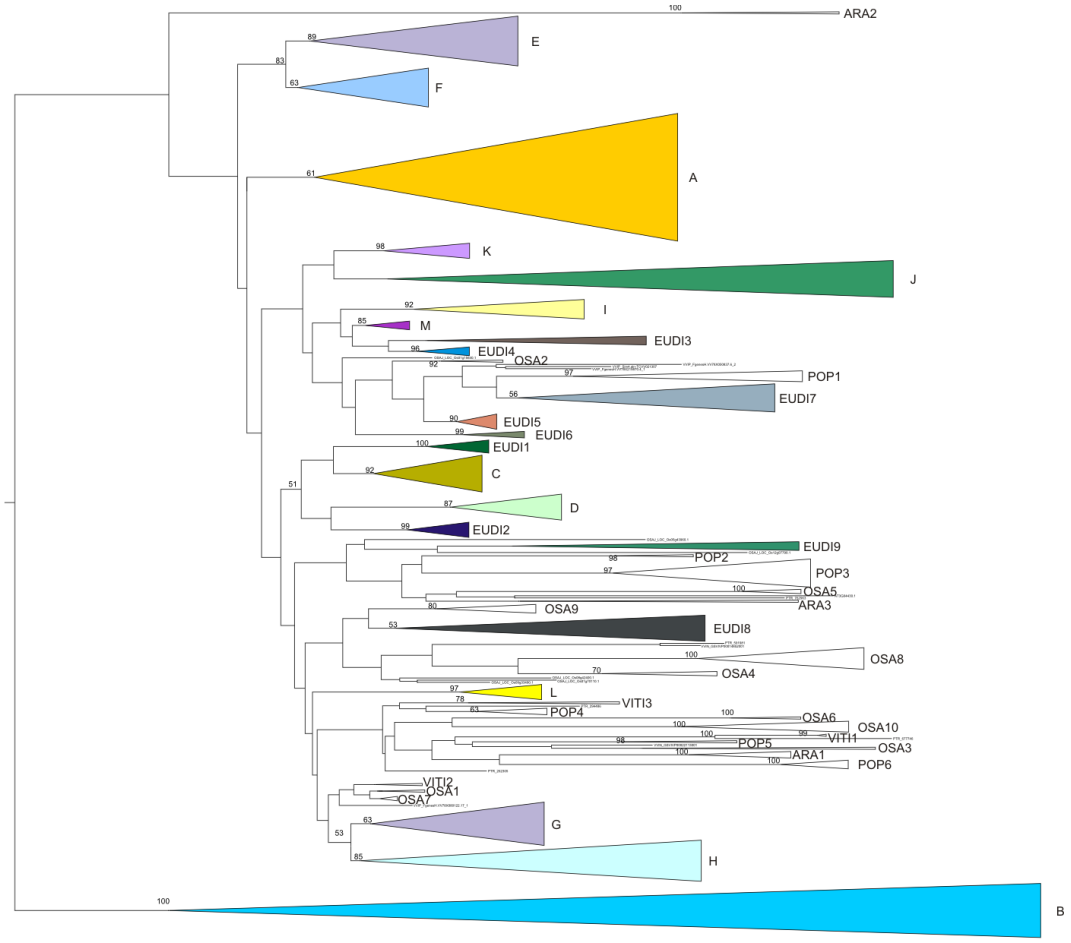


Figure 4.1: ***Phylogenetic tree of NACs in angiosperms.*** *This unrooted tree was inferred by Maximum-Likelihood methods using the corrected alignment of all NAC transcription factors from Arabidopsis, black cottonwood, grapevine and rice. All groups are supported by a bootstrap value ≥ 50. Groups E and F, G and H, and C, D, EUDI1 and EUDI2 are sister groups evidenced by a bootstrap ≥ 50. The size of each triangle corresponds to the number of genes in the group.*

and possibly, in the cross-talk between ABA and jasmonic acid signaling in these processes (BU *et al.* 2008, FUJITA *et al.* 2004, LU *et al.* 2007, MA *et al.* 2004, TRAN *et al.* 2007, 2004; Table SIII). The fact that functional information is mostly restricted to one group of homologues highlights the need for such systematic studies, which will improve our understanding of the evolution of regulatory networks in plants.

The classification presented here represents an up-dated version of previous work (OOKA *et al.* 2003). The name of each group of orthologues was changed to a more neutral classification, to avoid bias in gene function transfer or characterization.

A summary of the groups of homologues found in embryophytes can be seen in Figure 4.3.

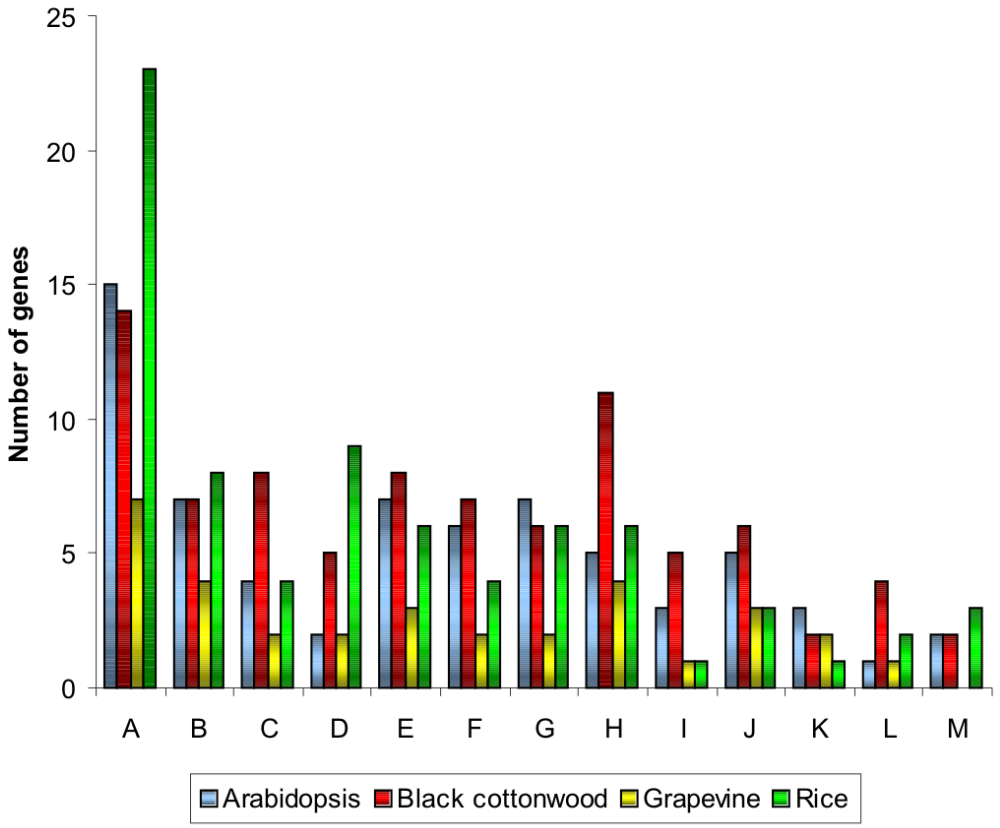


Figure 4.2: ***Distribution of NAC genes among the Groups of Homologues.***

4.3.2 Possible Groups of Orthologues (PoGOs)

To gain more information about the evolutionary relationship of NAC TFs, each group of homologues was analyzed separately, using an alignment corresponding to a conserved region within this family (Figure S3). Groups that are restricted to single species were not used for this approach, as, by definition, it is impossible to uncover Possible Groups of Orthologues (PoGOs), and only Possible Groups of Paralogues (PoGPs) can be identified. Forty five PoGOs were identified in Groups A to M; additionally one PoGP from Arabidopsis (A12), seven from rice (A14 to 17, D2, H3, M2) and two from Physcomitrella (A0 and B3) were uncovered (Figures S4 to S16). The high number of PoGPs in rice compared to other organisms may result from the fact that rice is the only representative from monocots. This observation is similar to the one made for the groups of homologues.

In these analyses we observed that Group A is sub-divided into sub-groups A0, A1, A2 and A3. Contrary to sub-groups A2 and A3, it was not possible to identify any PoGO in sub-group A1 containing sequences from monocots and eudicots. The sub-groups A0, A1 and A2 are the closest sister groups, and sub-group A3 seems to be an angiosperm-

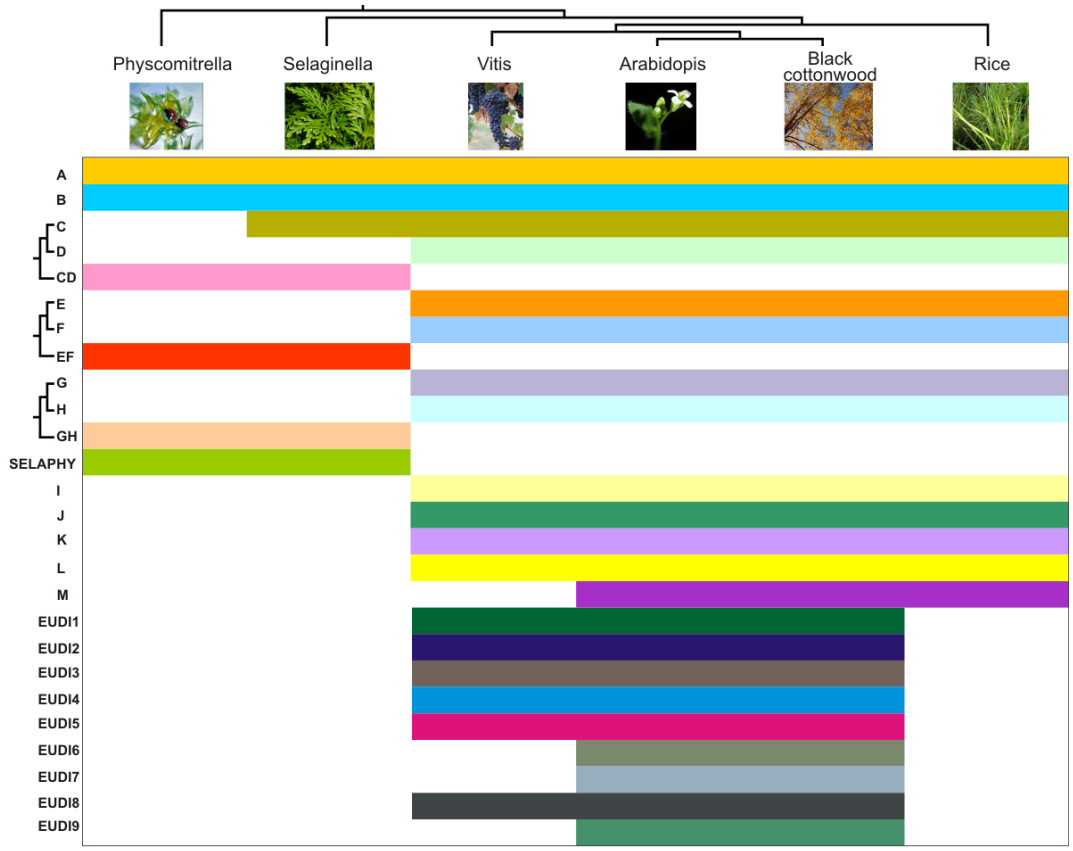


Figure 4.3: ***Phylogenetic profile of NACs in plants.*** *Groups C and D, E and F, and G and H are sister groups and share common extant representatives in Physcomitrella. Groups A and B are the only two groups present in all embryophytes. SELAPHY is constituted only by NACs from Physcometrella and Selaginella, most probably this group has been lost in angiosperms. The most recent common ancestor (MRCA) of all embryophytes possessed at least five NAC genes, corresponding to Groups A, B, CD, EF and GH. NACs of the species studied here were grouped at the level of higher taxa, i.e., bryophytes (P. patens); pteridophytes (S. moellendorffii), and angiosperms (A. thaliana, P. trichocarpa, V. vinifera and O. sativa). Solid boxes indicate that at least one NAC was found for a given group of homologues in the respective taxon. Groups EUDI1 to 9 appear to be restricted to eudicots.*

specific acquisition (Figure S4). The PoGO A2.3 seems to be the most conserved in Group A, as members of this group can be found in all tracheophytes. Similarly, Group B is also subdivided into three sub-groups (B1, B2 and B3), where sub-group B1 has representatives from all embryophytes (Figure S5).

Although the emergence of NAC factors occurred late in evolution when compared to other TFs presents in all eukaryotes, such as bZIP and bHLH (RIAÑO PACHÓN *et al.* 2008), they have experienced a strong expansion, being amongst the most numerous families in angiosperms. Such massive expansion and the high number of lineage-specific groups suggest that NACs played an important role in establishing plant diversity and/or

conquering new environments.

4.3.3 The evolution of NAC transcription factors

NACs constitute a plant-specific TF family, and efforts to identify this TF in animals and fungi have proven negative (RIAÑO PACHÓN *et al.* 2008). Interestingly, no NAC could be found in any algae, making this TF specific to land plants. Unexpectedly, in Physcomitrella we observed a high number of NACs when compared to other transcription factors which are common to all eukaryotes, such as bZIP and MADS. Recently, YAMASAKI *et al.* (2008) proposed that NACs have emerged from WRKY, a transcription factors family that was originally believed to be restricted to plants (RIECHMANN *et al.* 2000). A protein containing a WRKY domain was found in the genome of the protozoan *Giardia lamblia*, indicating that WRKY appeared earlier in evolution and that this TF evolved to the present GCM in animals and WRKY and NAC in plants (RIAÑO PACHÓN *et al.* 2008, YAMASAKI *et al.* 2008) .

Senescence is a very important process in plant development as it accounts for the distribution of micro- and macronutrient from the leaves to the reproductive organs, and thus has a large adaptive value. Recently, BALAZADEH *et al.* (2008) performed a TF expression analysis where they identified 185 TFs that are differentially expressed at late stages of leaf development in Arabidopsis. NACs are overrepresented in this set of genes, and, in most cases, their expression increases during the onset of senescence. One of the NACs identified is *AtNAP*, which has previously been shown to control leaf senescence (GUO *et al.* 2005). Null mutant *atnap* exhibits a strongly delayed senescence and reduced expression of the senescence marker gene *SAG12*. Full phenotypic complementation by a NAC orthologue from bean (*PvNAP*) indicates that eudicots have similar control mechanisms for senescence. Complementation with an orthologue from rice (*OsNAP*, LOC_Os03g21060) indicates a similar conservation at the angiosperm level GUO and GAN (2006).

As a rice orthologue was able to complement the phenotype of an Arabidopsis mutant, we were interested to uncover which NACs could play a similar role in rice. To this end, a qRT-PCR approach, similar to the one used in BALAZADEH *et al.* (2008), was taken to assess the expression level of the 121 NAC genes in rice leaves. Forty four genes were differentially expressed during the late stages of rice leaf development (Table SIV). It is important to note that most of the NACs from Arabidopsis and rice uncovered by these analyses tend to cluster into Groups A, E, G and H, indicating a possible set of common senescence regulatory players conserved in all angiosperms (Figure S3). Interestingly, these four groups belong to three bigger groups of homologues (A, EF and GH)

that posses extant representative in bryophytes (Figure 4.3), indicating that the network regulating senescence is, at least partially, present in Physcomitrella. Additionally, all genes tested in PoGO J1.2 are differentially expressed in senescence, but this observation might reflect an angiosperm acquisition, as Group J is formed only by angiosperm NACs. Though rice shares similarities with Arabidopsis, NACs from rice also have a tendency to cluster into rice-specific groups. This finding might reflect a monocot specialization or even an adaptation of crop species due to the nutrient requirement for grain filling.

The fact that Group G is one of the groups, whose members are differentially expressed during senescence, corroborates the observation that there is cross-talk between salt stress and senescence (GHANEM *et al.* 2008, MUNNS and TESTER 2008).

Besides NACs, WRKY TFs have also been identified as important players in regulating senescence in Arabidopsis (EULGEM and SOMSSICH 2007). WRKY53 is one of the most prominent WRKY TFs involved in senescence, controlling the expression of *SAG12* (HINDERHOFER and ZENTGRAF 2001, MIAO *et al.* 2004). Derived from the fact that NACs originated from WRKY and that both TF families are involved in the control of senescence, it is tempting to speculate that the control of senescence is an ancestral feature that raised before the split of both TF families. This opens the possibility to identify genes with similar functions in bryophytes. Physcomitrella does not show the same kind of senescence that is known from angiosperms, as its leaflets are always green. Nevertheless, loss of chlorophyll can be artificially induced by exposure to long dark periods. Additionally, in the beginning of development, the sporophytic tissue performs photosynthesis, but later on it loses chlorophyll and lives as a "parasite" of the gametophyte. At a later stage, the sporophyte undergoes programmed cell death (PCD). The identification of NACs related to these processes in mosses and the groups of homologues to which they belong would help to uncover a possible gene toolkit for senescence conserved in embryophytes. This will pave the way to better understand the evolutionary conserved core mechanism of leaf senescence in plants.

4.3.4 The identification of NACs involved in senescence in barley

One of the main goals of phylogenetic analyses is to provide a backbone for functional information that can be transferred from one organism to another CORRÊA *et al.* (2008). The ability to better identify candidate genes not only by virtue of their expression pattern, but also by taking into account information available from other organisms, may allow decreasing the time used for gene characterization thus resulting in a more efficient data analysis. To prove this concept, we designed primers for ten NACs from barley (Table SI), which are orthologues of NACs differentially expressed during senescence in rice,

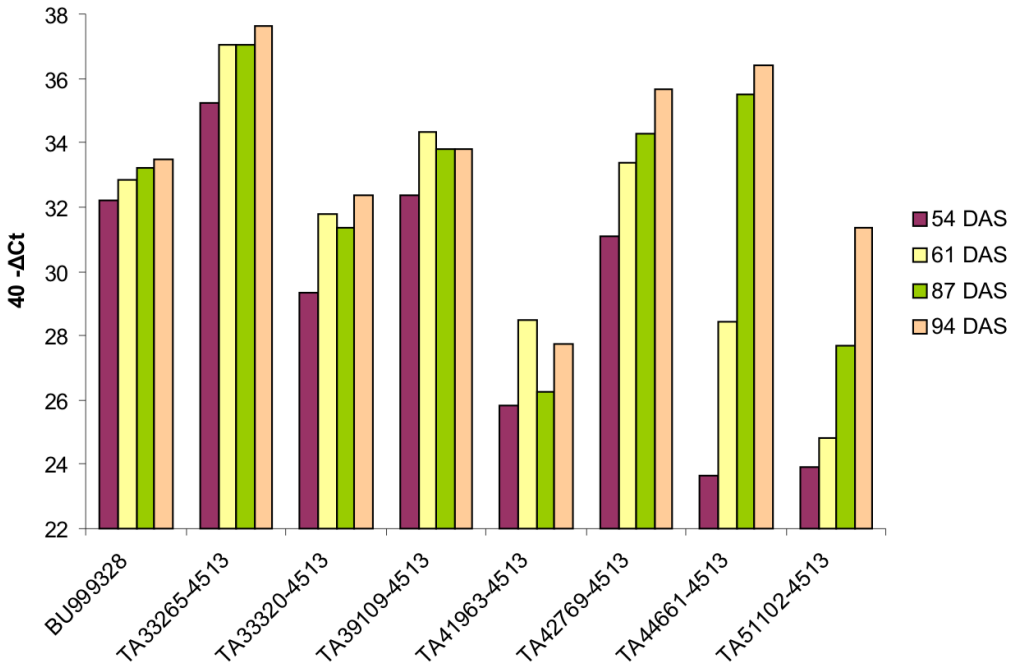


Figure 4.4: ***Expression level of NAC genes in late leaf development of barley.*** *The expression level of ten NACs was accessed via qRT-PCR and eight (indicated with their code numbers) of them show differential expression in the late stages of leaf development. The stages indicated by days after sowing (DAS) correspond to leaves, which are 50% of the full expanded leaf size (15 cm, 54 DAS), fully expanded leaf (30 cm, 61 DAS), 5% (87 DAS) and 20% (94 DAS) senescent. Expression levels were normalized to the expression of Elongation factor 1. Data represent the mean of two biological replicates.*

and assessed their expression via qRT-PCR. Of the ten barley genes tested, eight showed differential expression in late stages of leaf development; in one case a 173-fold increase in expression was observed (Figure 4.4). This result indicates that the prediction of a putative candidate gene via transfer of functional information in a phylogenetic background is a powerful and robust tool.

4.4 Conclusion

The identification of groups of homologues of NACs and the further characterization of its members into PoGOs may facilitate the transfer of functional information among different species. NACs are overrepresented among the TFs differentially expressed during senescence, which illustrates their importance in this process. So far, the control of senescence by NACs seems to be a conserved program in mono- and eudicots, which can be

observed not only by the expression pattern of genes, but also by functional complementation by orthologues. As most NAC genes from Arabidopsis and rice cluster into Groups A, EF and GH, it is possible that the core network leading to senescence is present already in bryophytes, but further investigations are needed to uncover the main elements of this network. Function prediction by phylogenetic analysis has proven to be a powerful approach to better target putative candidate genes. Thus the systematic usage of this approach may decrease the time spend on gene characterization and increase the chances of better choosing a candidate gene for detailed analyses.

4.5 References

ALVAREZ-BUYLLA, E. R., S. J. LILJEGREN, S. PELAZ, S. E. GOLD, C. BURGEFF, *et al.*, 2000 MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells, roots and trichomes. Plant J **24**: 457–466.

ARVIDSSON, S., M. KWASNIEWSKI, D. R. NO PACHÓN, and B. MUELLER-ROEBER, 2008 QuantPrime - a flexible tool for reliable high-throughput primer design for quantitative PCR. BMC Bioinformatics **9**: 465.

BAILEY, T. L., and C. ELKAN, 1995 The value of prior knowledge in discovering motifs with MEME. Proc Int Conf Intell Syst Mol Biol **3**: 21–29.

BALAZADEH, S., D. M. R. NO PACHÓN, and B. MUELLER-ROEBER, 2008 Transcription factors regulating leaf senescence in *Arabidopsis thaliana*. Plant Biol (Stuttg) **10 Suppl 1**: 63–75.

BU, Q., H. JIANG, C.-B. LI, Q. ZHAI, J. ZHANG, *et al.*, 2008 Role of the *Arabidopsis thaliana* NAC transcription factors ANAC019 and ANAC055 in regulating jasmonic acid-signaled defense responses. Cell Res **18**: 756–767.

CALDANA, C., W.-R. SCHEIBLE, B. MUELLER-ROEBER, and S. RUZICIC, 2007 A quantitative RT-PCR platform for high-throughput expression profiling of 2500 rice transcription factors. Plant Methods **3**: 7.

CLAMP, M., J. CUFF, S. M. SEARLE, and G. J. BARTON, 2004 The Jalview Java alignment editor. Bioinformatics **20**: 426–427.

CORRÊA, L. G. G., D. M. R. NO PACHÓN, C. G. SCHRAGO, R. V. DOS SANTOS, B. MUELLER-ROEBER, *et al.*, 2008 The role of bZIP transcription factors in green plant evolution: adaptive features emerging from four founder genes. PLoS ONE **3**: e2944.

CZECHOWSKI, T., R. P. BARI, M. STITT, W.-R. SCHEIBLE, and M. K. UDVARDI, 2004 Real-time RT-PCR profiling of over 1400 Arabidopsis transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes. Plant J **38**: 366–379.

DAYHOFF, M., R. SCHWARTZ, and B. ORCUTT, 1978 *Atlas of protein sequence and structure, v. 5, Suppl. 3*. National Biochemical Research Foundation, Silver Spring, MD.

EULGEM, T., and I. E. SOMSSICH, 2007 Networks of wrky transcription factors in defense signaling. Curr Opin Plant Biol **10**: 366–371.

FUJITA, M., Y. FUJITA, K. MARUYAMA, M. SEKI, K. HIRATSU, *et al.*, 2004 A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. Plant J **39**: 863–876.

GHANEM, M. E., A. ALBACETE, C. MARTINEZ-ANDUJAR, M. ACOSTA, R. ROMERO-ARANDA, *et al.*, 2008 Hormonal changes during salinity–induced leaf senescence in tomato (Solanum lycopersicum L.). J Exp Bot **59**: 3039–3050.

GOFF, S. A., D. RICKE, T.-H. LAN, G. PRESTING, R. WANG, *et al.*, 2002 A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science **296**: 92–100.

GUO, H.-S., Q. XIE, J.-F. FEI, and N.-H. CHUA, 2005 MicroRNA directs mRNA cleavage of the transcription factor NAC1 to downregulate auxin signals for arabidopsis lateral root development. Plant Cell **17**: 1376–1386.

GUO, Y., and S. GAN, 2006 AtNAP, a NAC family transcription factor, has an important role in leaf senescence. Plant J **46**: 601–612.

HE, X.-J., R.-L. MU, W.-H. CAO, Z.-G. ZHANG, J.-S. ZHANG, *et al.*, 2005 AtNAC2, a transcription factor downstream of ethylene and auxin signaling pathways, is involved in salt stress response and lateral root development. Plant J **44**: 903–916.

HINDERHOFER, K., and U. ZENTGRAF, 2001 Identification of a transcription factor specifically expressed at the onset of leaf senescence. Planta **213**: 469–473.

JAILLON, O., J.-M. AURY, B. NOEL, A. POLICRITI, C. CLEPET, *et al.*, 2007 The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature **449**: 463–467.

JONES, D. T., W. R. TAYLOR, and J. M. THORNTON, 1992 The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci **8**: 275–282.

KATOH, K., K. MISAWA, K. ICHI KUMA, and T. MIYATA, 2002 MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res **30**: 3059–3066.

KIM, S.-G., A.-K. LEE, H.-K. YOON, and C.-M. PARK, 2008 A membrane-bound NAC transcription factor NTL8 regulates gibberellic acid-mediated salt signaling in Arabidopsis seed germination. Plant J **55**: 77–88.

KIRK, D. L., 2005 A twelve-step program for evolving multicellularity and a division of labor. Bioessays **27**: 299–310.

KO, J.-H., S. H. YANG, A. H. PARK, O. LEROUXEL, and K.-H. HAN, 2007 ANAC012, a member of the plant-specific NAC transcription factor family, negatively regulates xylary fiber development in *Arabidopsis thaliana*. Plant J **50**: 1035–1048.

LARROUX, C., G. N. LUKE, P. KOOPMAN, D. S. ROKHSAR, S. M. SHIMELD, *et al.*, 2008 Genesis and expansion of metazoan transcription factor gene classes. Mol Biol Evol **25**: 980–996.

LU, P.-L., N.-Z. CHEN, R. AN, Z. SU, B.-S. QI, *et al.*, 2007 A novel drought-inducible gene, ATAF1, encodes a NAC family protein that negatively regulates the expression of stress-responsive genes in Arabidopsis. Plant Mol Biol **63**: 289–305.

MA, S., T. M. QUIST, A. ULANOV, R. JOLY, and H. J. BOHNERT, 2004 Loss of TIP1;1 aquaporin in Arabidopsis leads to cell and plant death. Plant J **40**: 845–859.

MIAO, Y., T. LAUN, P. ZIMMERMANN, and U. ZENTGRAF, 2004 Targets of the WRKY53 transcription factor and its role during leaf senescence in Arabidopsis. Plant Mol Biol **55**: 853–867.

MICHOD, R. E., Y. VIOSSAT, C. A. SOLARI, M. HURAND, and A. M. NEDELCU, 2006 Life-history evolution and the origin of multicellularity. J Theor Biol **239**: 257–272.

MITSUDA, N., A. IWASE, H. YAMAMOTO, M. YOSHIDA, M. SEKI, *et al.*, 2007 NAc transcription factors,

NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of Arabidopsis. Plant Cell **19**: 270–280.

MORENO-RISUENO, M. A., M. MARTÍNEZ, J. VICENTE-CARBAJOSA, and P. CARBONERO, 2007 The family of DOF transcription factors: from green unicellular algae to vascular plants. Mol Genet Genomics **277**: 379–390.

MUNNS, R., and M. TESTER, 2008 Mechanisms of salinity tolerance. Annu Rev Plant Biol **59**: 651–681.

NAKANO, T., K. SUZUKI, N. OHTSUKI, Y. TSUJIMOTO, T. FUJIMURA, *et al.*, 2006 Identification of genes of the plant-specific transcription-factor families cooperatively regulated by ethylene and jasmonate in *Arabidopsis thaliana*. J Plant Res **119**: 407–413.

OLSEN, A. N., H. A. ERNST, L. L. LEGGIO, and K. SKRIVER, 2005 NAC transcription factors: structurally distinct, functionally diverse. Trends Plant Sci **10**: 79–87.

OOKA, H., K. SATOH, K. DOI, T. NAGATA, Y. OTOMO, *et al.*, 2003 Comprehensive Analysis of NAC Family Genes in Oryza sativa and Arabidopsis thaliana. DNA Res **10**: 239–247.

RENSING, S. A., D. LANG, A. D. ZIMMER, A. TERRY, A. SALAMOV, *et al.*, 2008 The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science **319**: 64–69.

RIAÑO PACHÓN, D. M., L. G. G. CORRÊA, R. TREJOS-ESPINOSA, and B. MUELLER-ROEBER, 2008 Green transcription factors: a chlamydomonas overview. Genetics **179**: 31–39.

RIAÑO PACHÓN, D. M., S. RUZICIC, I. DREYER, and B. MUELLER-ROEBER, 2007 PlnTFDB: An integrative plant transcription factor database. BMC Bioinformatics **8**: 42.

RIECHMANN, J. L., J. HEARD, G. MARTIN, L. REUBER, C. JIANG, *et al.*, 2000 Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science **290**: 2105–2110.

SACHS, J. L., 2008 Resolving the first steps to multicellularity. Trends Ecol Evol **23**: 245–248.

STAMATAKIS, A., 2006 RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22**: 2688–2690.

TAMURA, K., J. DUDLEY, M. NEI, and S. KUMAR, 2007 MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol **24**: 1596–1599.

TRAN, L.-S. P., K. NAKASHIMA, Y. SAKUMA, Y. OSAKABE, F. QIN, *et al.*, 2007 Co-expression of the stress-inducible zinc finger homeodomain ZFHD1 and NAC transcription factors enhances expression of the ERD1 gene in Arabidopsis. Plant J **49**: 46–63.

TRAN, L.-S. P., K. NAKASHIMA, Y. SAKUMA, S. D. SIMPSON, Y. FUJITA, *et al.*, 2004 Isolation and functional analysis of Arabidopsis stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the *early responsive to dehydration stress 1* promoter. Plant Cell **16**: 2481–2498.

TUSKAN, G. A., S. DIFAZIO, S. JANSSON, J. BOHLMANN, I. GRIGORIEV, *et al.*, 2006 The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science **313**: 1596–1604.

UAUY, C., A. DISTELFELD, T. FAHIMA, A. BLECHL, and J. DUBCOVSKY, 2006 A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. Science **314**: 1298–1301.

VELASCO, R., A. ZHARKIKH, M. TROGGIO, D. A. CARTWRIGHT, A. CESTARO, *et al.*, 2007 A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS ONE **2**: e1326.

VINCENTZ, M., C. BANDEIRA-KOBARG, L. GAUER, P. SCHLÖGL, and A. LEITE, 2003 Evolutionary

pattern of angiosperm bZIP factors homologous to the maize Opaque2 regulatory protein. J Mol Evol **56**: 105–116.

YAMAGUCHI, M., M. KUBO, H. FUKUDA, and T. DEMURA, 2008 Vascular-related NAC-DOMAIN7 is involved in the differentiation of all types of xylem vessels in Arabidopsis roots and shoots. Plant J **55**: 652–664.

YAMASAKI, K., T. KIGAWA, M. INOUE, S. WATANABE, M. TATENO, *et al.*, 2008 Structures and evolutionary origins of plant-specific transcription factor DNA-binding domains. Plant Physiol Biochem **46**: 394–401.

YOON, H.-K., S.-G. KIM, S.-Y. KIM, and C.-M. PARK, 2008 Regulation of leaf senescence by NTL9-mediated osmotic stress signaling in Arabidopsis. Mol Cells **25**: 438–445.

ZHAO, C., U. AVCI, E. H. GRANT, C. H. HAIGLER, and E. P. BEERS, 2008 XND1, a member of the NAC domain family in *Arabidopsis thaliana*, negatively regulates lignocellulose synthesis and programmed cell death in xylem. Plant J **53**: 425–436.

ZHONG, R., E. A. RICHARDSON, and Z.-H. YE, 2007 Two NAC domain transcription factors, SND1 and NST1, function redundantly in regulation of secondary wall synthesis in fibers of Arabidopsis. Planta **225**: 1603–1611.

ZHONG, R., and Z.-H. YE, 2007 Regulation of cell wall biosynthesis. Curr Opin Plant Biol **10**: 564–572.

5

Molecular and physiological analysis of the sink-to-source transition in Arabidopsis leaves

Molecular and physiological analysis of the sink-to-source transition in Arabidopsis leaves

Luiz Gustavo Guedes Corrêa[1,2], Joerg Fettke[1], Joachim Fisahn[1], Diego Mauricio Riaño-Pachón[1,2], Judith Gomes-Porras[3], Eugenia Maximova[2], Marc Lohse[2], Bernd Mueller-Roeber[1,2,†]

[1]Department of Molecular Biology, University of Potsdam, Potsdam–Golm, Germany, [2]Max-Planck Institute of Molecular Plant Physiology, Potsdam–Golm, Germany,i [3]University of Bielefeld, Universitaetstr. 25, 33501, Bielefeld, Germany

[†] Author for correspondence. Email: bmr@uni-potsdam.de

Manuscript to be submitted for publication.

Author contributions

BMR and LGGC conceived the idea for the study and the research plan. The experimental work was done by LGGC. JFettke contributed to the analysis of sugar mobilization and ^{14}C labelling experiments. JFisahn contributed to the ETR analysis and gas exchange. DMRP performed the clustering of genes and statistical analyses. JLGP analysed the overrepresentation of cis–elements. EM contributed for the microscopical analysis. ML established the promAtCHoR::GUS plant lines. BMR supervised the project. All authors contributed to manuscript writing.

5.1 Abstract

Leaf development is one of the most studied processes in plants, although not very much is known about the control mechanisms involved in the sink-to-source transition. In this work, we have identified molecular and physiological markers that indicate this transition in leaves of *Arabidopsis thaliana*. Moreover, we identified 153 transcription factors differentially expressed during early leaf development that are putative candidates for playing an important role in the sink-to-source transition. Together, the data provided give a better overview about the sink-to-source transition and also open the possibility to use Arabidopsis as a model for this important biological process.

5.2 Introduction

The development of a plant is intrinsically related to the development of its leaves, which are the main photosynthetic organs of angiosperms. During their lifespan, leaves undergo two develpmental transitions: the sink-to-source and the mature-to-senescent transitions. The sink-to-source transition is characterized by an accumulation of carbohydrates and the depletion of nitrogen, when the immature sink leaves (net carbon importers) start exporting carbohydrates to other parts of the plant, becoming auto-sufficient on carbon production (ADE-ADEMILUA and BOTHA 2007, ROBERTS *et al.* 1997, TURGEON 1989, TURGEON and WEBB 1973). Senescence is characterized by the breakdown of chloroplasts, the progressive degradation of cell constituents and remobilization of metabolites to other parts of the plant (MASCLAUX *et al.* 2000). Moreover, due to the loss of chloroplasts, leaves acquire a yellowish colour, which is a visible marker of senescence. On the contrary, there is no visible marker for the sink-to-source.

Early studies have shown that, in eudicot plants, the sink-to-source transition occurs right after the unfolding of the leaf in a basipetal manner (JONES and EAGLES 1962), and that it is completed when the leaf reaches 30-60% of its full size (HO 1988, HO and SHAW 1977, HOPKINSON 1966, TURGEON 1989). During this transition, the leaf tip starts exporting carbon to other parts of the plant, whereas the base still imports carbon from other source leaves (VOGELMANN *et al.* 1982). Autoradiography of leaves from *Populus deltoides* (LARSON and DICKSON 1973), maize (EVERT *et al.* 1996), aspen (KLEINER KW 1999), sugarbeet (PITCHER and DAIE 1991), *Moricandia arvensis* (GAGNON and BEEBE 1996), *Curcubita pepo* and tobacco (TURGEON and WEBB 1973), beans and *Coleus blumei* (TURGEON and WIMMERS 1988) after the application of ^{14}C labelled CO2 support the basipetal direction of the transition. In transition leaves, a bidirectional transport of carbon is observed, as a result of the import to the immature base and

the export from the apical region (JONES and EAGLES 1962, LARSON *et al.* 1972); meanwhile, unloading from major veins reduces or even ceases (IMLAU *et al.* 1999, OPARKA *et al.* 1999, ROBERTS *et al.* 1997, WRIGHT *et al.* 2003) and the number of plasmodesmata is drastically reduced (ROBERTS *et al.* 2001).

In this study, we identified new markers for the sink-to-source transition in Arabidopsis leaves and established the relation between them, setting non-visual phenotypical parameters for further investigation of this process. Additionally, we identified 153 transcription factors differentially expressed during early leaf development, making them interesting candidates for the coordination of the sink-to-source transition.

5.3 Results

5.3.1 Identification of new markers for leaf sink-to-source transition

Unlike the mature-to-senescent transition, there is no apparent phenotypic change during the sink-to-source transition. Some genes were reported to exhibit an expression pattern that follows the sink-to-source transition: *AtSUC3* (MEYER *et al.* 2000) and *AtNSI* (CARVALHO *et al.* 2006) are expressed in sink leaves, but their expression decline in source leaves; conversely, expression of *AtSUC2* (IMLAU *et al.* 1999) and *ISI1* (ROOK *et al.* 2006) is low in sink leaves, but increases in source leaves. We observed that the ortholog of *StCHoR* in Arabidopsis (*AtCHoR*, AGI At3g55250) shows an expression pattern very similar to the one of *AtSUC2* (Figure 5.2). Plants carrying the promoter-reporter gene constructs promAtCHoR::GUS or promAtSUC2::GUS show GUS staining when leaves reached a size of 2 to 3 mm. Interestingly, a long term analysis of plants carrying the construct promAtSUC2::tmGFP revealed that not all leaves behave in the same manner; early leaves tend to undergo the transition at around 2 mm, whereas later emerging leaves undergo this transition a slightly later (Figure 5.1). This observation may indicate that the presence of already formed source leaves is able to provide enough photoassimilates to the plant, decreasing the need of the plant for new sources, thus leading to a later sink-to-source transition in later emerging leaves. Moreover, this indicates that different leaves formed at different time points throughout plant development are in different physiological stages even though they have the same size.

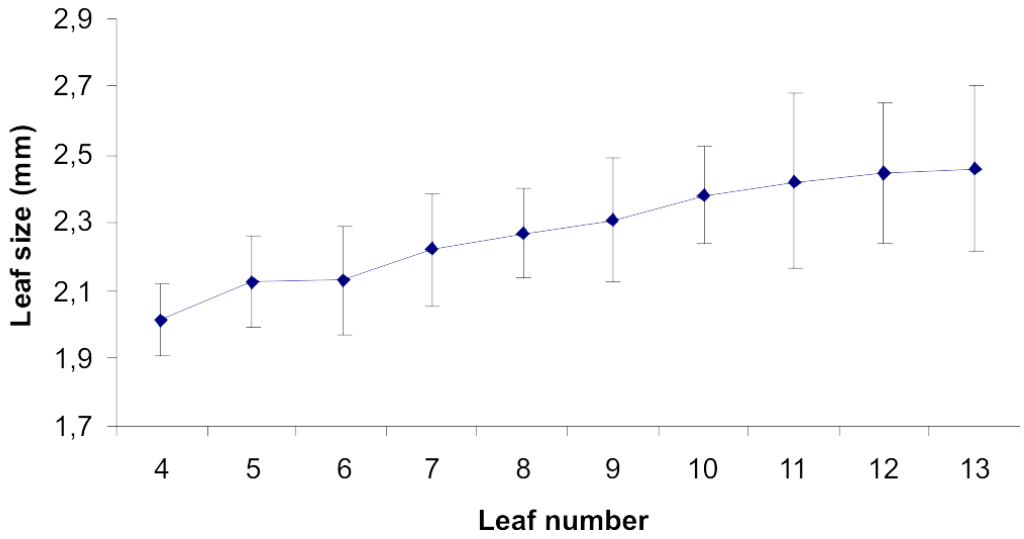


Figure 5.1: ***The sink-to-source transition in different leaves***. *The expression of GFP fused with AtSUC2 promoter was used as a marker for the sink-to-source transition. Thirty-three different plants were observed daily and the day when the GFP signal was first observed is taken as the beginning of the sink-to-source transition. As it can be seen early emerging leaves undergo the sink-to-source transition earlier than leaves developing later. Bars indicate the standard deviation.*

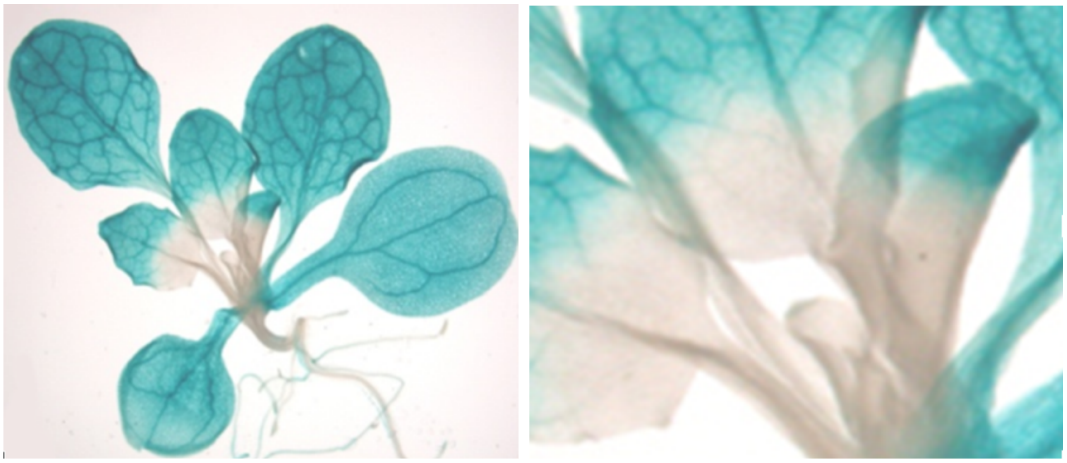


Figure 5.2: ***AtCHoR expression pattern in Arabidopsis plants***. *Plants carrying the construct promAtCHoR::GUS show staining at the tip when leaves are* $\sim 2-3$ *mm long. Staining increases in a basipetal manner as leaf develops. No staining was observed in leaves smaller than 2 mm. Similar result was obtained for plants carrying the AtSUC2::GUS construct (not shown).*

5.3.2 The sink-to-source transition depends on the overall sink/source relation

The observation that later developing leaves tend to undergo the sink-to-source transition when they reach a bigger size as compared to earlier emerging leaves led us to hypothesise

that the transition is depended on the overall the sink/source ratio of the plant. As the number of leaves increase, there is a higher source of carbon, but the sugar requirement from the sink parts of the plant does not increase in the same proportion. In order to test this hypothesis, we grew plants carrying the construct promAtSUC2::tmGFP and detached the fifth or the fifth and sixth leaves to decrease the carbon fixation capacity of the plant. We observed that leaves 9 to 12 undergwent the transition at a size of 1.83 $\pm$ 0.15 mm, whereas, in the control, the transition occurs at leaf size of 2.56 $\pm$ 0.2 mm. Another way of depriving plants from their carbon source is to change to a different light regime. Plants grown in long day were moved to a very short day condition (5 hours of light) for 24 hours. Leaves 9 to 12 underwent the transition when they were 1.60 $\pm$ 0.50 mm long, whereas in the control occurred at a leaf size 2.36 $\pm$ 0.26 mm. Curiously, a later sink-to-source transition was observed in plants shaded up to 50% of the standard light intensity, at a leaf size of 2.83 $\pm$ 0.12 mm.

One limiting point for the transition and photosynthesis would be the access of CO_2 to the leaves. If we consider later emerging leaves as having less access to CO_2 as early emerging leaves, we would then expect a later transition due to the availability of carbon source for photosynthesis. When C24 carrying the promAtSUC2::tmGFP construct are transferred to short day condition (8 h light) for one week, after three weeks in long day (16 h light), leaves 9 to 12 seems to undergo the sink-to-source earlier than in the normal condition, when leaves are 0.53 $\pm$ 0.24 mm. Thus, one we can conclude that the access of CO_2 is not the limiting factor for the transition.

One may also speculate that a delayed transition of later emerging leaves is influenced by the fact that later leaves reach bigger sizes at the end of their development. However, we could not find a correlation between the final size of a leaf and the time point of the sink-to-source transition (do the relation with diego, data already collected).

5.3.3 Chlorophyll content increases during the sink-to-source transition

The amount of chlorophyll in a given leaf is related to its photosynthetic capacity. Thus, one might expect to see an increase of the chlorophyll content upon the sink-to-source transition. An increased amount of chlorophyll was observed between 2 and 3 mm (Figure 5.3). This finding has a good correlation with the expression patterns observed for *AtSUC2* and *AtCHOR*. It also supports the idea that chlorophyll content is a good biomarker for estimating leaf age, as it follows the same tendency observed in tobacco and sugar beet (HARN *et al.* 1993, MASCLAUX *et al.* 2000). Unexpectedly, 1mm-leaves posses a higher amount of chlorophyll than 2mm leaves. There could be two main possible explanations

for this fact: (a) cell size of 1mm-leaves is smaller than in 2mm-leaves, implying a higher number of cells in the measured samples; or (b) chloroplasts do not divide at the same rate as cells, leading to a lower number of chloroplasts per cell at this early stage of leaf development (RAPP and MULLET 1991).

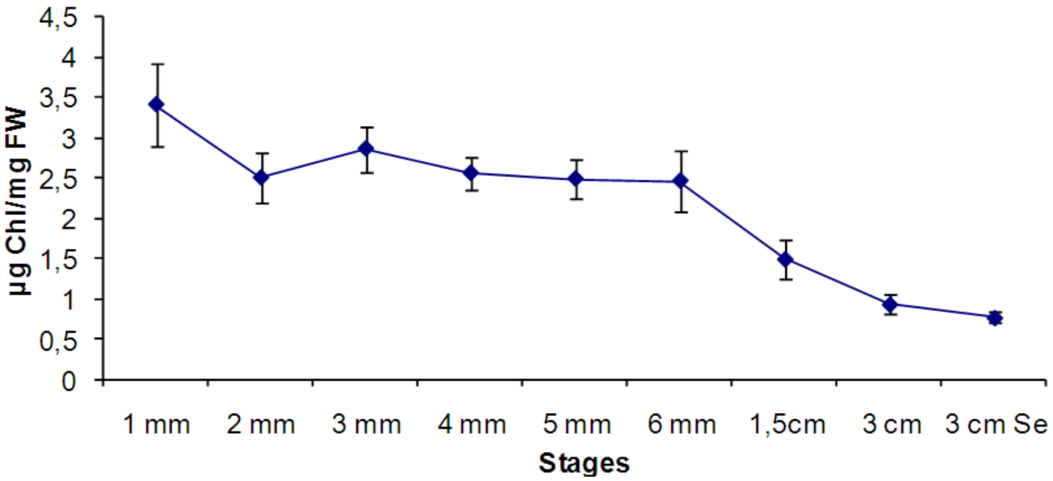


Figure 5.3: ***Amount of chlorophyll during leaf development.*** *It is possible to observe an increment of chlorophyll content between 2 and 3 mm that may correspond to the sink-to-source transition.*

Additionally to the amount of chlorophyll, we measured the expression level of Chlorophyll A/B binding proteins 1 and 2 (CAB1 and 2; Figure 5.4), which are subunits of the light harvesting complex, thus providing an indirect evidence of the photosynthetic activity in leaves. One would expect that during the sink-to-source transition the expression CAB1 and CAB2 increases in order to supply the needs of a new source leaf. In fact, it is possible to observe that the expression level of both genes increases when leaves grow from 2 to 3 mm (Figure 5.4).

Together, the expression patterns of *AtCHoR* and *AtSUC2*, the leaf chlorophyll content and the expression level of CAB1 and CAB2 indicate that the probable point of the initiation of the sink-to-source transition is between two and three millimetres. In the following, we classify leaves as being sink, when they have a size of 1 to 2 mm, transition leaves, at the size of 3 to 4 mm, and source, from 5 mm onward.

5.3.4 Leaf photosynthesis

The electron transport rate (ETR) is an indirect measure of the efficiency of photosystem II (PSII). Following the ETR of different leaves or parts of a leaf in different light intensities (light curve) allow identifying sink and source leaves, as sink leaves have a lower ETR

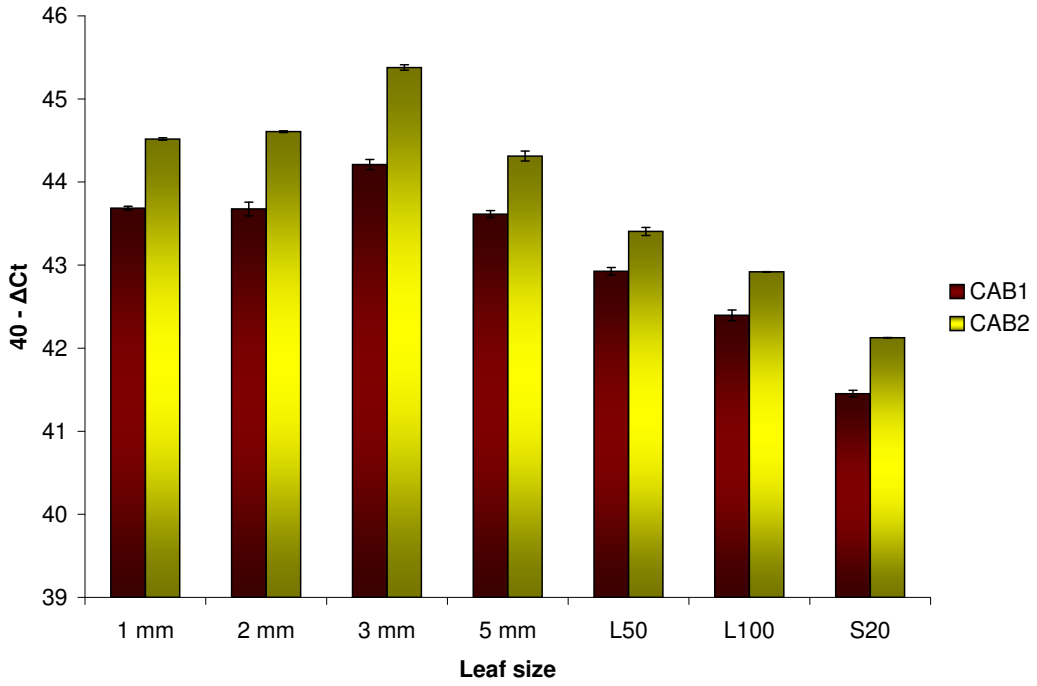


Figure 5.4: ***Expression of CAB1 and CAB2 during leaf development.*** *Expression levels were assesed via qRT-PCR using leaves of 1, 2, 3 and 5 mm, comprising the sink-to-source transition and leaves comprising 50% of leaf full expansion (L50, 1.5 cm), full expanded leaf (L100, 3 cm) and 20% senescent (S20). Expression of CAB1 and CAB2 increases between 2 and 3 mm, as expected for the sink-to-source transition. CAB genes are indirect markers of photosynthesis and the highest expression is observed when leaves reach 3 mm. Expression levels were normalized to the expression of UBQ10. Expression corresponds to the mean value of three biological replicates with two technical replicates and the error bars display the standard deviation.*

at saturating light intensities. As expected, leaves smaller than 2 mm have a low ETR when compared to bigger leaves (Figure 5.5; Table S1). Leaves smaller than 2 mm were measured just in one leaf region, as obtaining the light curve for more than one region in such small leaf led to inaccurate measurements. Leaves from 2.3 to 2.8 mm showed an increase in their ETR, but only at the leaf tip, whereas at the leaf base, the ETR was as low as in sink leaves (Figure 5.5). In leaves ranging of 4 to 6 mm the ETR at the leaf tip was similar to that in source leaves, whereas the base of the leaf is at transition (Figure 5.5). No significant differences were observed between different leaf regions, when leaves had a size of 6 to 10 mm. Due to technology constrains, the rate of CO_2 fixation could not be measured in small leaves.

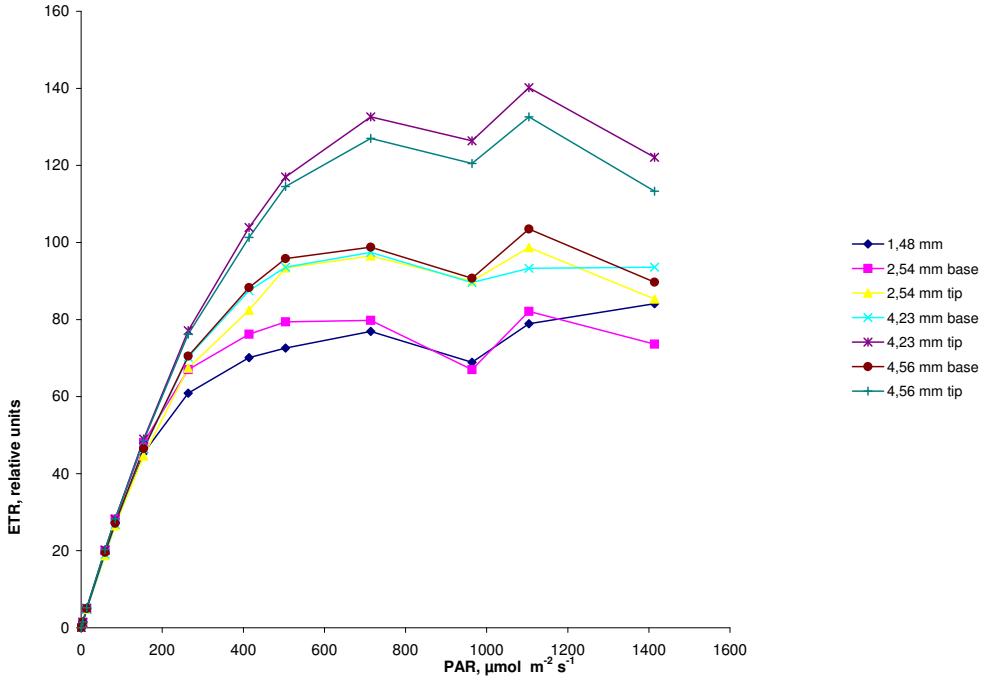


Figure 5.5: ***ETR measurement for the sink-to-source transition in Arabidopsis leaves.*** *Arabidopsis plants grown in a phytotron at long day condition (16 h light) were exposed to darkness for 1 h and an increasing amount of light incited the leaves leading increments on the photosynthetic electron transport rate (ETR). In sink leaves (smaller than ca. 2.3 mm), the ETR reaches a plateau of 70, whereas in source leaves (bigger than 5 mm), this plateau is almost the double of the value obtained for sink leaves. As leaves start undergoing the sink-to-source transition, ETR increases in the leaf tip, while it remains low at the base, similar to an sink leaf (as seen in the 2.54 mm leaf). As leaves grow, the tip reaches the source status while the base undergoes the sink-to-source transition. No significant difference was observed between tip and base regions of leaves bigger than 5.3 mm.*

5.3.5 Starch content

Starch is the main form in which fixed carbon is stored in Arabidopsis, but starch is expected to accumulate in Arabidopsis leaves when the carbon assimilation rate is higher then its consumption. Leaves in different stages of development were stained with Lugol, which indicates the presence of starch. We observed low starch content in leaves smaller than 2 mm. The intensity of Lugol staining increases progressively in leaves up to 6 mm (Figure 5.6). Although starch staining pattern was similar to the one expected for the sink-to-source transition, we also observed a gradient from the margin to the centre. This pattern is very similar to the expression pattern observed for the *Impaired Sucrose Induction 1* (*ISI1*, ROOK *et al.* 2006), which plays a role in carbohydrate partition.

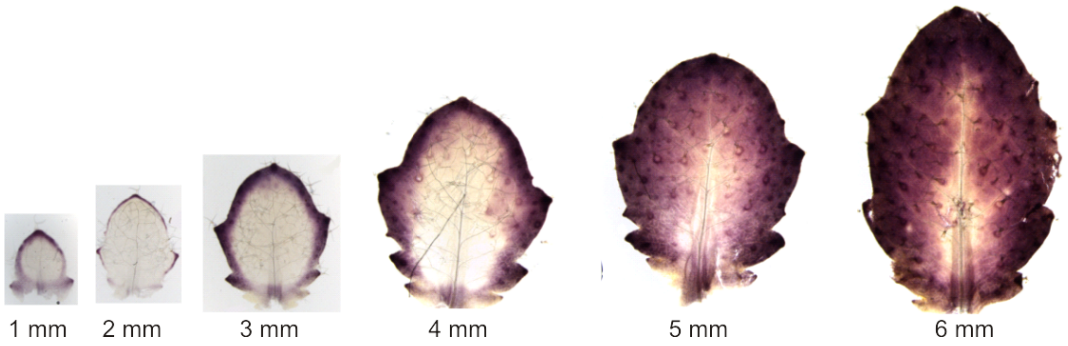


Figure 5.6: ***Lugol staining of Arabidopsis leaves***. *Leaves were stained with Lugol, as an indicator of the amount of starch in the leaf. A gradient of starch accumulation is seen between sink and source leaves, where sink leaves have low starch content, while source leaves have high starch content. Leaves illustrated here were harvested 6 hours in the light period. This result is similar top the one obtained for leaves harvested at 9 am (3 hours in the light period).*

The differences seen in the Lugol staining are also reflected partially in the starch content in leaves (Figure 5.7). Interestingly, the content of starch increases in leaves between 1 and 2 mm, slightly before the point expected for the sink-to-source transition. Starch may constitute a key player in the sink-to-source transition, as its accumulation may serve as a source for building new molecules in a phase of higher metabolism. A similar mechanism is observed in the sink-source transition in rice leaf sheaths, where they accumulate starch (sink) during development, and at the heading period, starch is remobilized (source) to the growing panicle (CHEN and WANG 2008). The amount of starch does not change significantly during the transition stages till a significant increase is seen when leaves underwent full transition (5mm; Figure 5.7). In agreement with this observation, the expression of genes related to the starch metabolism (Figure 5.8) are up regulated when leaves grow from 1 to 2 and 3 mm; and then were down regulated in bigger leaves (5 mm). This indicates that the starch production/degradation is more intense in transition leaves. Only two genes (i.e., *AMY2* and *XET*) do not follow this pattern; their expression increases during the transition and further during subsequent leaf development (Figure 5.8 and Table S2).

5.3.6 ^{14}C labeling

Photosynthesis-dependent carbon fluxes towards starch and soluble sugars were analyzed by in vivo labeling using intact Arabidopsis plants supplied with $^{14}CO_2$. In each experiment, 28 plants were placed into a sealed glass box and were continuously illuminated. Following 10, 20, 40 min of exposition to $^{14}CO_2$, the plants were withdrawn and the leaf material was used for the isolation of starch, soluble sugars and insoluble cell wall mate-

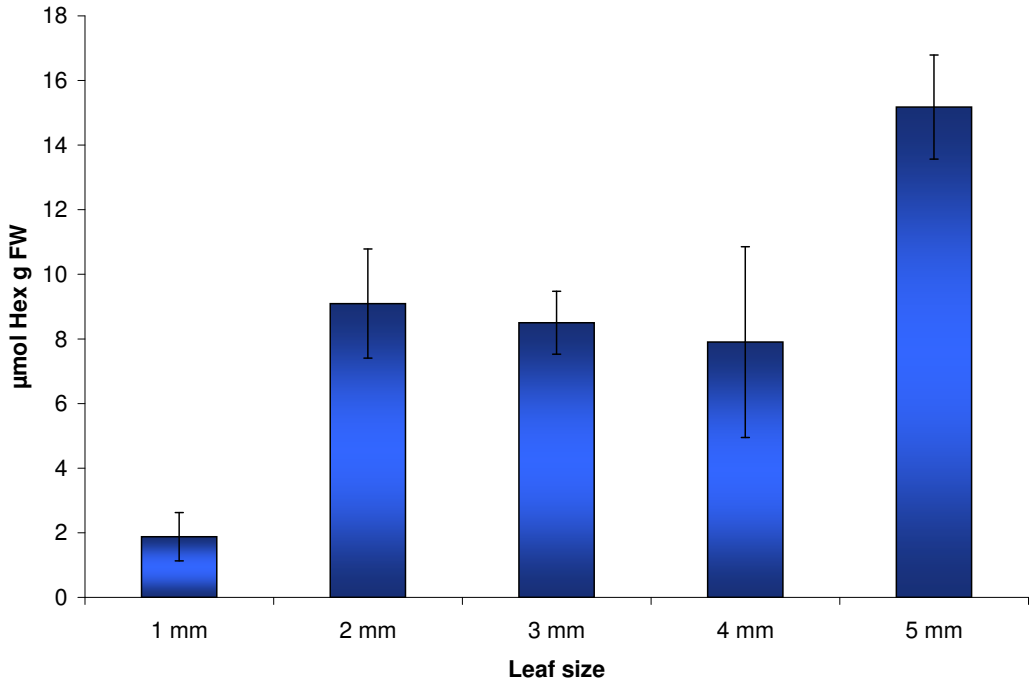


Figure 5.7: ***Starch concentration in Arabidopsis leaves***. *The total amount of starch in leaves from 1 to 5 mm was accessed according to* FETTKE et al. *(2005). Low starch content was observed in 1 mm leaves. The amount of starch increased between 1 and 2 mm, just before the sink-to-source transition. During the transition, the amount of starch remained almost stable and increased again at 5 mm, when leaves are mainly source leaves. The differences observed between the sink, transition and source leaves are statistically significant (Tukey HSD with p-values $\ll$0.01). The values correspond to the average of 5 biological replicates with 2 technical replicates and the bars display the standard deviation.*

rial. For the subfractions the ^{14}C labelling was monitored (Figure 5.9). It is possible to see that the carbon incorporation in starch in sink leaves is slower than in source leaves. Although the level of radioactivity observed in the soluble fraction is similar in all leaves after 40 minutes, short exposition to $^{14}CO_2$ leads to a low signal, which is in accordance with the fact that sink leaves are sugar importers; thus sugars are synthesised in source leaves and transported to sink leaves, causing the delay observed. On the other hand, accumulation of $^{14}CO_2$ in the insoluble fraction, representing the cell wall material, is more prominent in sink than source leaves.

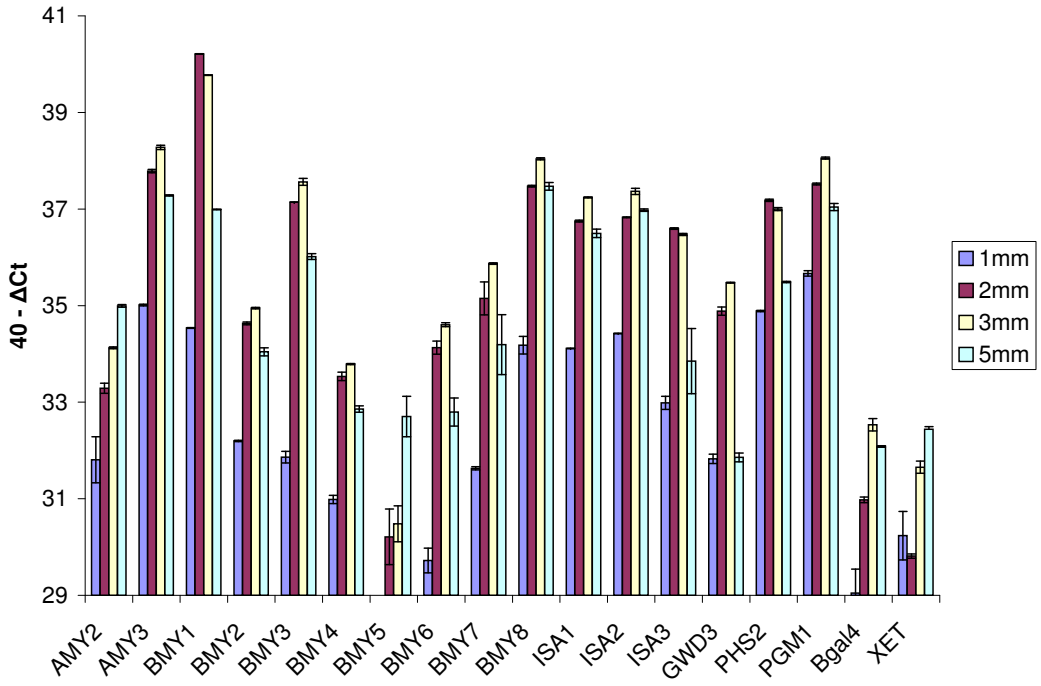


Figure 5.8: ***Expression level of genes related to the starch turnover****. Almost all genes represented here have a similar expression pattern, with an increase during the transition and a decrease afterwards. The complete set of genes tested can be found at the Table S2.*

5.3.7 Transcription factors differentially expressed during the sink-to-source transition

Transcription factors have a pivotal role as they control gene transcription initiation. Interestingly, altough many transcription factors have been assigned to leaf development (reviewed by BARKOULAS *et al.* 2007) and senescence (reviewed by BALAZADEH *et al.* 2008), no TF has been directly assigned to the sink-to-source transition. To uncover the TFs that could play a role during the sink-to-source transition we assessed their transcription level by qRT-PCR using an updated version of the platform originally described by CZECHOWSKI *et al.* (2004). Of the 1880 TFs analysed, ca. 1700 were found to be expressed in leaves (data not shown). Although expression of a large number of TFs could be detected in leaves, expression of only 153 TFs changed during early stages of leaf development (one, two, three, four and five millimetres, Table 5.1). Members of three TF families were found to be over-represented among the differentially expressed genes: GRF, MYB and SRS (Table S3). The Growth Regulatory Factor (GRF) family has been identified as playing an important role in cell proliferation and expansion. Mem-

bers of this family are expressed in actively growing regions, such as the leaf primordial (HORIGUCHI *et al.* 2005, KIM *et al.* 2003, KIM and KENDE 2004). Additionally, some GRF are involved in the initiation of de-etiolation, which, to some extension, can be interpreted as a cross-talk between light stimuli and growth (TEPPERMAN *et al.* 2006). The SHI-Related Sequence (SRS) is a small TF family (10 members) that seems to have an extensive functional conservation (KUUSK *et al.* 2006). Genes of this family have been linked to the control leaf morphogenesis, as well as to gibberellic acid and auxin responses, affecting cell growth and division (FRIDBORG *et al.* 2001, KUUSK *et al.* 2002, SOHLBERG *et al.* 2006, ULLAH *et al.* 2003). Both GRF and SRS are plant specific transcription factors, identified in tracheophytes only (RIAÑO PACHÓN *et al.* 2008) and MYB represents the most numerous TF family in angiosperms (RIAÑO PACHÓN *et al.* 2008). Its members are involved in a vast spectra of processes (YANHUI *et al.* 2006, and references therein). MYBs were also identified in animals and fungi, thus this family was present in the MRCA of eukaryotes, and may play conserved functions across the eukaryotic kingdom (CORRÊA *et al.* 2008). Of the 24 differentially expressed MYB, 12 have been functionally characterized; they are related mainly to light signalling/defence, leaf morphogenesis and trichome formation (COMINELLI *et al.* 2008, FU *et al.* 2007, GUO *et al.* 2008b, HAGA *et al.* 2007, HEMM *et al.* 2001, HONG *et al.* 2008, LEA *et al.* 2007, PETRONI *et al.* 2008, STRACKE *et al.* 2007, WANG *et al.* 2007, ZHAO *et al.* 2008).

To identify to overrepresentation of any GO category among the 153 TFs differentially expressed, we compared them to the ca. 1.700 TFs expressed during early leaf development. It is possible to observe that classes like system development, organ development, leaf development and epidermal cell differentiation are some of the processes overrepresented among the differentially expressed TFs (Figure 5.10 and Table S4). This overrepresentation not only indicates major processes during the sink-to-source transition, but also common points with other processes, such as flower development (Figure 5.10).

From these 153 TFs, 139 were grouped into seven different clusters according to their expression patterns; 14 genes could not be clustered (Figure 5.11 and Table 5.1). No family over-representation was observed in any cluster.

Clusters I and V contain genes that are up-regulated before the sink-to-source transition (UBeS), though genes from cluster I increase their expression throughout the early stages of leaf development, whereas genes from cluster V seem to be expressed at the same level after the transition (Figure 5.10). Clusters II, III and IV comprise genes that are down-regulated during the early stages of leaf development (Figure 5.10). Cluster II contains genes that are continuously decreasing their expression during the transition. Some genes from this cluster play a role in leaf morphogenesis (FU *et al.* 2007, GUO *et al.* 2008a, KIM *et al.* 2003, KUUSK *et al.* 2006, QIN *et al.* 2005, SADDIC *et al.* 2006, TEPER-

Cluster	AGI
1	At4g06746, At3g50260, At4g32280, At3g22100, At1g34180, At1g52830, At1g29600, At1g08320, At4g05170, At5g39610, At2g18300, At1g18400, At1g68800, At4g18960
2	At5g58280, At1g31320, At4g37750, At5g11190, At3g02550, At5g53290, At2g35310, At3g06220, At1g68480, At5g66940, At1g73360, At1g79840, At5g28640, At3g07260, At1g63100, At3g49950, At2g45480, At3g13960, At3g05690, At4g23800, At2g36890, At2g37630, At3g61250, At4g34990, At5g58610, At1g63650, At1g13400, At5g03790, At3g27920, At1g64000, At1g16060, At1g68640, At2g17770, At5g06510, At4g16750, At1g67100, At4g09820, At5g41315, At5g67110, At1g74430, At1g74650, At5g11510, At4g19630, At5g17800, At3g01530, At1g66380, At4g36260, At1g31310, At2g28610, At4g00480, At5g04400, At4g00250, At1g06180, At1g02065, At3g60670, At5g12330, At5g65510, At5g49330, At1g15360, At5g28650, At5g60440, At1g16070, At2g35700, At5g06650, At5g26660, At3g11260, At5g24330, At5g52600, At2g25820, At5g60910, At3g01140, At2g46410, At2g46990, At2g46990
3	At1g19850, At4g34400, At2g22770, At4g14560, At2g28160, At2g24700, At5g53660, At1g19790, At3g51060, At5g66700, At3g10040, At1g28360, At4g10240, At5g10280, At4g14770, At3g54340, At1g75520
4	At1g31640, At1g46264, At1g62975, At1g76420, At2g06200, At2g18120, At2g47460, At3g09600, At3g62610, At4g24540, At5g06250, At5g22570, At5g23260, At5g47670, At5g53950, At5g56840, At5g60830, At5g64810
5	At2g20880, At1g01380, At4g14550, At4g25560, At5g52020, At3g46130, At2g38300, At3g29035
6	At3g27810, At4g26030, At5g61620, At2g42410
7	At4g35610, At2g23660, At3g03260, At1g13300
Unassigned	At4g24150, At1g66390, At3g57920, At3g04850, At5g59780, At1g31040, At1g69180, At2g42150, At1g02030, At3g50410, At4g11660, At1g69120, At4g31640, At5g17810

Table 5.1: *Genes differentially expressed at least 5 fold between the different leaf stages measured.*

BAMNOLKER and SAMACH 2005), cell expansion and proliferation (HAGA *et al.* 2007, HORIGUCHI *et al.* 2005, KANG *et al.* 2007), and cell fate (GAN *et al.* 2007, MOROHASHI *et al.* 2007, TOMINAGA *et al.* 2007). This result confirms the expectation that genes controlling leaf identity and cell proliferation and fate should be highly expressed in the early stages of leaf development and decrease their expression as leaves grow. Cluster III contain genes that are sharply down-regulated before the sink-to-source transition (DoBeS). Genes from this cluster play a role in leaf morphogenesis (KIM and KENDE 2004, KUUSK *et al.* 2006), vascular tissue formation (CASSON *et al.* 2002, WENZEL *et al.* 2007), and meristem identity determination (LAMB *et al.* 2002). Cluster IV contains genes that decrease their expression by the end of the sink-to-source transition (DoES), and they play a role in leaf morphogenesis (KIM and KENDE 2004, VROEMEN *et al.* 2003), cell proliferation (PETRONI *et al.* 2008) and flavonoid accumulation (LEA *et al.* 2007). Though play similar roles, there is need of further functional analyses into understand why genes from clusters II, III and IV have a marked expression pattern. Clusters VI and VII have a cyclic expression pattern, and they are a mirror of each other (Figure 5.10); little is known about the genes in these clusters. From the physiological point of view, the sink-to-source transition is characterized by two main events: the increase of photosynthesis, allowing a higher sugar production rate; and the cessation of sugar import caused by a morphological change in the leaf vasculature from import to export and the closing of plasmodesmata (TURGEON 1989). In tobacco (TURGEON 1984), it has been observed that also in albino leaves import terminates in a basipetal manner, independent of a positive carbon balance, which is similar to plants where photosynthesis was prevented by darkening. It may be assumed that alteration of the venation pattern is developmentally regulated in a manner that is uncoupled with the achievement of a positive carbon balance. Thus the TFs identified in our analysis can be further classified into two groups: the one related to the carbon balance (CB), which is directly or indirectly involved in light response, and the one related to the venation pattern. In albino plants, we would expect that TFs related to the venation pattern (VP) would not have alteration in their expression whereas genes related to the carbon balance would have their expression deregulated. Recently TIAN *et al.* (2007) published metabolic profiling and transcriptome data for Arabidopsis albino mutants. Comparison between genes differentially expressed in albino plants and the transcription factors in early leaf development allowed identifying cluster I as CB (Table 5.1).

5.3.8 Promoter analysis of the 153 TFs differentially expressed during early leaf development

To control gene transcription, transcription factors must be able to recognize distinct *cis*-elements, which in the end will be largely responsible for the expression pattern of a downstream target gene. We analysed the promoter regions of the 153 TFs that were differentially expressed during early leaf development in an attempt to identify putative cis-elements that might play a role in the sink-to-source transition of Arabidopsis leaves. Taken the whole set of TF promoters, we were able to identify 50 overrepresented oligonucleotide sequences representing potential cis-elements. The most strongly and significantly overrepresented cis-element identified is SORLIP5AT, which stands for Sequences Over-Represented in Light-Induced Promoters (SORLIPs) in Arabidopsis (JIAO *et al.* 2005). This finding illustrates the importance of light for early leaf development. Additionally, 16 previously unknown, potential *cis*-elements were observed in this analysis. A resume of the putative cis-elements detected can be found in Table S5. When analysed individually, clusters I, VI and VII did not show any overrepresented cis-element. Analysis of cluster II revealed data similar to those obtained for the whole set of 153 TFs. Clusters III and IV also posses SORLIP5AT sequences, but they share with Cluster V another putative cis-element, whose function is still not described (Table S6).

5.3.9 Common players in sink-to-source and mature-to-senescent transition

Sink-to-source and mature-to-senescent transitions and the two main transitions during leaf development. Though regarded as complete independent processes, it is known that mature leaves can turn again into a sink during pathogen infection (BONFIG *et al.* 2006). In order to identify possible regulators of both processes, we compared our results with the one obtained for the TFs in the onset of senescence (BALAZADEH *et al.* 2008). Twenty seven TFs are differentially expressed in both sink-to-source and mature-to-senescent transitions (Table 5.2), thus these genes potentially play a role in both transitions. Moreover, we can assume that these genes are mainly involved in developmental aspects, rather then in biotic or abiotic stresses responses. Interestingly, 50% of these genes are bHLHs or MYBs. It was observed that bHLH are significantly represented in onset of senescence (BALAZADEH *et al.* 2008). bHLHs are related to light signalling via interaction with cytochromes (JIAO *et al.* 2007) what evidences the importance of light to both processes.

Cluster	AGI
1	At1g52830, At1g08320, At5g39610, At5g39610
2	At4g37750, At5g11190, At2g37630, At1g63650, At5g41315, At5g67110, At5g11510, At1g66380, At4g00480, At1g06180, At2g35700, At3g11260, At3g01140
3	At2g22770
4	At2g47460, At5g06250
5	At1g01380, At4g14550, At3g29035
6	At3g27810
7	At3g03260, At1g13300
Unassigned	At1g66390, At3g57920

Table 5.2: *Genes differentially expressed in both sink-to-source and mature-to-senescent transitions.*

5.4 Discussion

The sink-to-source transition is one of the most important transitions occurring during leaf development. However, progress in understanding the regulatory networks controlling this process has been modest. As the sink-to-source transition is not accompanied by a easily visible phenotype, its analysis has remained difficult. The identification of molecular or biochemical markers indicating the sink-to-source transition is thus a key element for functional studies of the transition process. *AtSUC2* expression has a positive correlation with vein maturation WRIGHT *et al.* (2003) and has been suggested as a marker for the sink-to-source transition IMLAU *et al.* (1999); therefore, the overlapping expression patterns of *AtSUC2* and *AtCHoR*, the expression levels of *CAB1* and *CAB2* and the leaf chlorophyll content lead to define sink-to-source transition in Arabidopsis leaves as occurring between 2 and 3 mm for plants growing under long day conditions. These observations are also in accordance with the data obtained for the electron transport rate (ETR) in early developing leaves, as there is a clear difference between sink, transition and source leaves. An increase of the ETR at saturating light intensity reflects a higher capacity for CO_2 assimilation and thus sugar production. As expected, sink leaves have a lower capacity for assimilation, therefore they are dependent on photoassimilates imported from source leaves. Curiously, sink leaves accumulate starch, which is in contrast to the view that starch accumulated in leaves comes from the surplus of photosynthesis. Therefore, starch accumulation observed in Arabidopsis sink leaves bears some analogy to starch accumulation in rice leaf sheaths CHEN and WANG (2008). Starch may be stored in the sink leaves to be used as a source of energy during the sink-to-source transition. Additionally, accumulation of starch would not disturb the gradient between sink and

source leaves necessary for the import of photoassimilates by the latter. This hypothesis is further supported by the fact that some genes regulating the cycling of starch are up-regulated during the transition, decreasing their expression levels when they are bigger than 5 mm (Figure 5.8). Considering that a direct supply of sugar to the plant affects only source, but not sink leaves ARAYA *et al.* (2006), it is tempting to conclude that starch, but not sucrose, would be one of the most important signalling molecules for the sink-to-source transition. The identification of 153 TFs differentially expressed during early leaf development opens a new possibility to uncover regulators of the sink-to-source transition. Further questions with respect to leaf development remain open, e.g. with respect to common regulatory pathways of both, sink-to-source and mature-to-senescent transitions. Additionally, it is possible to draw small networks to better understand early leaf development. One example is the gene *PRESSED FLOWER* (*PRS*), a member of cluster II. *PRS* is involved in cell proliferation and identity acquisition, but *PRS* represses *YABBY* required for proper leaf development (DAI *et al.* 2007). Thus the decreasing expression observed during early leaf development is directly associated to an important function played by this gene. One may thus hypothesize that TF genes with descreasing expression level identified here also have a role in this early phase of leaf development. Uncovering the cis-elements involved in early leaf development may also help identifying other important elements of the sink-to-source transition, such as hormones.

5.5 Materials and methods

5.5.1 Plant material

Seeds of *Arabidopsis thaliana* (L.) Heynh., accession Columbia-0 (Col-0) were sowed directly on soil (Einheitserde GS90; Gebrüder Patzer, Sinntal-Jossa, Germany) and put into a cold-night chamber (day 20°C, night 6°C, 16 hours light, 75% humidity) for seven days and another week into a short-day phytotron (day 20°C, night 16°C, 8 hours light, 75% humidity). Plants were transferred to trays (each containing 60 plants) into a greenhouse chamber (16 hours HPIT light). These conditions were always the same, if not stated elsewhere. All sample harvesting was performed with leaf number 11 at around 9 am (i.e., three hours into the light period) and frozen in liquid nitrogen, except when stated otherwise.

5.5.2 Constructs

A fragment of 1.5 kb from the *AtCHoR* promoter was amplified via PCR using the primers promATC-fwd 5' - GAATTCGACAAGTGGAGATTCTATTGACC - 3' and promATC-rev 5' - GACAACCACCATGGCTGTGATATATGG - 3' and cloned into the pCAMBIA 1305.1 (CAMBIA, Canberra, Australia) using the *Eco*RI site introduced by the forward primer and the *Nco*I constitutive present in the start of the *AtCHoR* gene. The promAtSUC2::tmGFP was obtained from Dr. Ruth Stadler (STADLER *et al.* 2005).

5.5.3 RNA extraction and cDNA synthesis

Total RNA was extracted from leaves with RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the protocol of the producer, and treated with TURBO DNA*free* DNase (Ambion, Foster City, USA). RNA integrity and DNA contamination were checked according to CZECHOWSKI *et al.* (2004). RT reactions were performed with SuperScript III reverse transcriptase (Invitrogen, Carlsbad, USA), according to the manufacturer's instructions. The efficiency of cDNA synthesis was tested by real-time PCR amplification of control genes encoding *UBQ10* (At4g05320, CZECHOWSKI *et al.* 2004). Only cDNA preparations with C_T values = 19 ± 1 were used for the comparison of transcription factor expression levels.

5.5.4 Real-time PCR primers and conditions

Transcription factor profiling was performed using an updated version of the platform described by CZECHOWSKI *et al.* (2004), which now covers 1880 TFs. Primers for other genes were designed using QuantPrime (ARVIDSSON *et al.* 2008) and synthesized by MWG Biotech AG (Ebersberg, Germany). All primers designed here are described in Suplementary Table II. PCRs were performed in an optical 384-well plate with an ABI PRISM 7900 HT Sequence Detection System (Applied Biosystems) using the same parameters as in CZECHOWSKI *et al.* (2004).

5.5.5 Lugol staining

Leaves were harvested 6 hours after the beginning of the light period. Single leaves were distained with 80% ethanol for 24 hours and subsequently stained with 10% Lugol solution (Merck) for two minutes and rinsed with water. Leaves were analyzed using a stereo-fluorescence microscope MZ 16FA (Leica, Wetzlar, Germany) and LAS suite software (Leica).

5.5.6 Fluorescence signal and determination of ETR

Two weeks after sowing (see above), plants were transferred to single pots and grown in a phytotron under long day conditions (16 hours light, 8 hours night). Plants used for the determination of ETR were adapted in darkness for at least 20 min and the light curve was obtained using an Imaging-PAM chlorophyll fluorimeter (Walz, Effeltrich, Germany). Fluorescence signals and leaf sizes were analyzed using the stereo-fluorescence microscope MZ 16FA (Leica) and LAS suite software (Leica).

5.5.7 Chlorophyll measurement

Leaves were grinded in liquid nitrogen and aliquot of ca. 15 mg fresh weight were made. Material was re-suspended in 1 mL of 96% (v/v) ethanol and homogenized. Probes were centrifuged for 3 min a 13.000 rpm and the supernatant was used to determine photometrically the chlorophyll content at 650 nm. Values were normalized to the initial fresh weight of each sample.

5.5.8 Starch measurements

Quantification of starch content was done with whole leaf grinded in liquid nitrogen, using 3 to 5 mg of material in each measurement. Extraction was performed according to FETTKE *et al.* (2005) and the release of glucose was determined according to STITT (1989).

5.5.9 In vivo $^{14}CO_2$ pulse labeling experiments

Single plants were placed in a pot and were grown under controlled conditions for four weeks. For in vivo labeling 28 intact plants were transferred into a sealed exsiccator (6.8 L) and were exposed to $^{14}CO_2$ (specific radioactivity 1.25 MBq mmol-1). The total CO_2 concentration in the gas phase upon the period of the pulse labeling was 400 l L-1. During pulse labeling plants were illuminated with white light (approximately 110 mol photons m-2s-1). Labeling was terminated by removing the $^{14}CO_2$ by exhaust pump into a flask with saturated barium hydroxide for 5 min and than the leaf material was transfered into liquid nitrogen. The material was stored frozen until use. For sample processing the leaf material was homogenized using a mortar and 1 mL 80% [v/v] ethanol was added to each sample. Following incubation at 80°C for 15 min the samples were centrifuged (14.000 x g; 10 min). The supernatants were transferred in to new tube and to the pellet again 1 ml 80% [v/v] ethanol was added, incubated and centrifuged as above. The supernatants (soluble sugars) were combined and the ^{14}C label were monitored by

using liquid scintillation counter (Rotiszint Mini, Roth, Karlsruhe, Germany; Beckman Coulter, Fullterton, USA). The pellets were incubated with 200 μl 200 mM KOH for 1h at 95°C. Following neutralization with 176 μl 1 M acetic acid and centrifugation (as above) the pellets were washed with 1mL water and centrifuged (as above). The ^{14}C label in the combined supernatants, representing starch, and in the pellets, representing mainly insoluble cell wall material, were monitored.

5.5.10 GO processes

To search for overrepresented processes among the differentially expressed TFs during early leaf development, we used the 1710 TFs expressed in leaves as a comparison set. We used the program Cytoscape (CLINE *et al.* 2007) and the plugin BiNGO (v. 2.3; MAERE *et al.* 2005) for identifying the overrepresented GOs and drawing the network.

5.5.11 Search for conserved *cis*-elements

To search for sequences over-represented in the set of transcription factors, the sequences 5' upstream of the Transcriptional Start Site (TSS) where analysed (up to 1 kb). We assayed the occurrence of all octamers in the subset of sequences and compared it with the expected number of occurrences (based on single nucleotide frequencies of all upstream sequences in the genome, 31921 sequences in total). Oligonucleotide occurrences were calculated using the program compseq from EMBOSS (RICE *et al.* 2000). To select the most significantly over-represented oligonucleotides a significance index was calculated. The index calculated the significance of a pattern irrespective of oligonucleotide size, number and size of the input sequences. To minimize the risk of selecting patterns that are highly significant due to the imbalance of A+T over C+G content in Arabidopsis upstream sequences, the significance of the over-represented patterns was calculated also in a set of 100 random sequences that had the same nucleotide composition as the transcription factor subset of sequences. Details of the protocol are described in GOMEZ-PORRAS *et al.* (2007).

5.5.12 Supplementary material

The supplementary material here cited can be found at `https://molbio00.bio.uni-potsdam.de/correa2008/`.

5.6 Acknowledgements

LGGC thanks the Deutscher Akademischer Austauschdienst (DAAD) fort a scholarship (n° A/04/34814). LGGC, JFettke and DMRP thank the Interdisciplinary Centre 'Advanced Protein Technologies' of the University of Potsdam, coordinated by Dr. Babette Regierer, and the International PhD Programme 'Integrative Plant Science' (supported by the DAAD and the Deutsche Forschungsgemeinschaft, n° DAAD D/04/01336) for funding. Thanks for funding through the BMBF for financial support of DMRP (GABI-Future grant 0315046). BMR thanks the Fonds der Chemischen Industrie for financial support (N° 0164389). The authors thank Dr. Ruth Stadler for seeds of transgenic Arabidopsis C24 promAtSUC2::tmGFP lines. The authors also thank Dr. Kathryn M Wright and Prof. Dr. Mark Stitt for fruitful discussions.

5.7 References

ADE-ADEMILUA, O., and C. BOTHA, 2007 Sink-to-source transition of *Pisum sativum* leaves in relation to leaf plastochron index. American Journal of Plant Physiology **2**: 27–35.

ARAYA, T., K. NOGUCHI, and I. TERASHIMA, 2006 Effects of carbohydrate accumulation on photosynthesis differ between sink and source leaves of *Phaseolus vulgaris* L. Plant Cell Physiol **47**: 644–652.

ARVIDSSON, S., M. KWASNIEWSKI, D. R. NO PACHÓN, and B. MUELLER-ROEBER, 2008 QuantPrime - a flexible tool for reliable high-throughput primer design for quantitative PCR. BMC Bioinformatics **9**: 465.

BALAZADEH, S., D. M. R. NO PACHÓN, and B. MUELLER-ROEBER, 2008 Transcription factors regulating leaf senescence in *Arabidopsis thaliana*. Plant Biol (Stuttg) **10 Suppl 1**: 63–75.

BARKOULAS, M., C. GALINHA, S. P. GRIGG, and M. TSIANTIS, 2007 From genes to shape: regulatory interactions in leaf development. Curr Opin Plant Biol **10**: 660–666.

BONFIG, K. B., U. SCHREIBER, A. GABLER, T. ROITSCH, and S. BERGER, 2006 Infection with virulent and avirulent *P. syringae* strains differentially affects photosynthesis and sink metabolism in Arabidopsis leaves. Planta **225**: 1–12.

CARVALHO, M. F., R. TURGEON, and S. G. LAZAROWITZ, 2006 The geminivirus nuclear shuttle protein NSP inhibits the activity of AtNSI, a vascular-expressed arabidopsis acetyltransferase regulated with the sink-to-source transition. Plant Physiol **140**: 1317–1330.

CASSON, S. A., P. M. CHILLEY, J. F. TOPPING, I. M. EVANS, M. A. SOUTER, *et al.*, 2002 The POLARIS gene of Arabidopsis encodes a predicted peptide required for correct root growth and leaf vascular patterning. Plant Cell **14**: 1705–1721.

CHEN, H.-J., and S.-J. WANG, 2008 Molecular regulation of sink-source transition in rice leaf sheaths during the heading period. Acta Physiologiae Plantarum **30**: 639–649.

CLINE, M. S., M. SMOOT, E. CERAMI, A. KUCHINSKY, N. LANDYS, *et al.*, 2007 Integration of biological networks and gene expression data using Cytoscape. Nat Protoc **2**: 2366–2382.

COMINELLI, E., T. SALA, D. CALVI, G. GUSMAROLI, and C. TONELLI, 2008 Over-expression of the arabidopsis AtMYB41 gene alters cell expansion and leaf surface permeability. Plant J **53**: 53–64.

CORRÊA, L. G. G., D. M. R. NO PACHÓN, C. G. SCHRAGO, R. V. DOS SANTOS, B. MUELLER-ROEBER, *et al.*, 2008 The role of bZIP transcription factors in green plant evolution: adaptive features emerging from four founder genes. PLoS ONE **3**: e2944.

CZECHOWSKI, T., R. P. BARI, M. STITT, W.-R. SCHEIBLE, and M. K. UDVARDI, 2004 Real-time RT-PCR profiling of over 1400 Arabidopsis transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes. Plant J **38**: 366–379.

DAI, M., Y. HU, Y. ZHAO, H. LIU, and D.-X. ZHOU, 2007 A *WUSCHEL-LIKE HOMEOBOX* gene represses a *YABBY* gene expression required for rice leaf development. Plant Physiol **144**: 380–390.

EVERT, R. F., W. A. RUSSIN, and A. M. BOSABALIDIS, 1996 Anatomical and Ultrastructural Changes Associated with Sink-to-Source Transition in Developing Maize Leaves. International Journal of Plant Sciences **3**: 247.

FETTKE, J., S. POESTE, N. ECKERMANN, A. TIESSEN, M. PAULY, *et al.*, 2005 Analysis of cytosolic heteroglycans from leaves of transgenic potato (*Solanum tuberosum* L.) plants that under- or overexpress the Pho 2 phosphorylase isozyme. Plant Cell Physiol **46**: 1987–2004.

FRIDBORG, I., S. KUUSK, M. ROBERTSON, and E. SUNDBERG, 2001 The Arabidopsis protein SHI represses gibberellin responses in Arabidopsis and barley. Plant Physiol **127**: 937–948.

FU, Y., L. XU, B. XU, L. YANG, Q. LING, *et al.*, 2007 Genetic interactions between leaf polarity-controlling genes and ASYMMETRIC LEAVES1 and 2 in Arabidopsis leaf patterning. Plant Cell Physiol **48**: 724–735.

GAGNON, M.-J., and D. BEEBE, 1996 Establishment of a plastochron index and analysis of the sink-to-source transition in leaves of *Moricandia arvensis* (L.) DC. (Brassicaceae). International Journal of Plant Sciences **157**: 262–268.

GAN, Y., C. LIU, H. YU, and P. BROUN, 2007 Integration of cytokinin and gibberellin signalling by Arabidopsis transcription factors GIS, ZFP8 and GIS2 in the regulation of epidermal cell fate. Development **134**: 2073–2081.

GOMEZ-PORRAS, J. L., D. M. RIAÑO-PACHÓN, I. DREYER, J. E. MAYER, and B. MUELLER-ROEBER, 2007 Genome-wide analysis of ABA-responsive elements ABRE and CE3 reveals divergent patterns in Arabidopsis and rice. BMC Genomics **8**: 260.

GUO, A.-Y., X. CHEN, G. GAO, H. ZHANG, Q.-H. ZHU, *et al.*, 2008a PlantTFDB: a comprehensive plant transcription factor database. Nucleic Acids Res **36**: D966–D969.

GUO, M., J. THOMAS, G. COLLINS, and M. C. P. TIMMERMANS, 2008b Direct repression of KNOX loci by the ASYMMETRIC LEAVES1 complex of Arabidopsis. Plant Cell **20**: 48–58.

HAGA, N., K. KATO, M. MURASE, S. ARAKI, M. KUBO, *et al.*, 2007 R1R2R3-Myb proteins positively regulate cytokinesis through activation of KNOLLE transcription in *Arabidopsis thaliana*. Development **134**: 1101–1110.

HARN, C., E. KHAYAT, and J. DAIE, 1993 Expression dynamics of genes encoding key carbon metabolism enzymes during sink-to-source transition of developing leaves. Plant and Cell Physiology **34**: 1045–1053.

HEMM, M. R., K. M. HERRMANN, and C. CHAPPLE, 2001 AtMYB4: a transcription factor general in the battle against UV. Trends Plant Sci **6**: 135–136.

HO, L. C., 1988 Metabolism and compartmentation of imported sugars in sink organs in relation to sink strength. Annual Review of Plant Physiology and Plant Molecular Biology **39**: 355–378.

HO, L. C., and A. F. SHAW, 1977 Carbon economy and translocation of ^{14}C in leaflets of the seventh leaf of tomato during leaf expansion. Annals of Botany **41**: 833–848.

HONG, S. H., H. J. KIM, J. S. RYU, H. CHOI, S. JEONG, *et al.*, 2008 CRY1 inhibits COP1-mediated degradation of BIT1, a MYB transcription factor, to activate blue light-dependent gene expression in Arabidopsis. Plant J **55**: 361–371.

HOPKINSON, J. M., 1966 Studies on the expansion of the leaf surface. IV. The carbon and phosphorus economy of a leaf. J. Exp. Bot. **15**: 125–137.

HORIGUCHI, G., G.-T. KIM, and H. TSUKAYA, 2005 The transcription factor atGRF5 and the transcription coactivator AN3 regulate cell proliferation in leaf primordia of *Arabidopsis thaliana*. Plant J **43**: 68–78.

IMLAU, A., E. TRUERNIT, and N. SAUER, 1999 Cell-to-cell and long-distance trafficking of the green fluorescent protein in the phloem and symplastic unloading of the protein into sink tissues. Plant Cell **11**: 309–322.

JIAO, Y., O. S. LAU, and X. W. DENG, 2007 Light-regulated transcriptional networks in higher plants. Nat Rev Genet **8**: 217–230.

JIAO, Y., L. MA, E. STRICKLAND, and X. W. DENG, 2005 Conservation and divergence of light-regulated genome expression patterns during seedling development in rice and Arabidopsis. Plant Cell **17**: 3239–3256.

JONES, H., and J. E. EAGLES, 1962 Translocation of ^{14}Carbon within and between leaves. Annals of Botany **26**: 505–510.

KANG, J., Y. MIZUKAMI, H. WANG, L. FOWKE, and N. G. DENGLER, 2007 Modification of cell proliferation patterns alters leaf vein architecture in *Arabidopsis thaliana*. Planta **226**: 1207–1218.

KIM, J. H., D. CHOI, and H. KENDE, 2003 The AtGRF family of putative transcription factors is involved in leaf and cotyledon growth in Arabidopsis. Plant J **36**: 94–104.

KIM, J. H., and H. KENDE, 2004 A transcriptional coactivator, AtGIF1, is involved in regulating leaf growth and morphology in Arabidopsis. Proc Natl Acad Sci U S A **101**: 13374–13379.

KLEINER KW, RAFFA KF, D. R., 1999 Partitioning of ^{14}C-labeled photosynthate to allelochemicals and primary metabolites in source and sink leaves of aspen: evidence for secondary metabolite turnover. Oecologia **119**: 408–418.

KUUSK, S., J. J. SOHLBERG, D. M. EKLUND, and E. SUNDBERG, 2006 Functionally redundant SHI family genes regulate Arabidopsis gynoecium development in a dose-dependent manner. Plant J **47**: 99–111.

KUUSK, S., J. J. SOHLBERG, J. A. LONG, I. FRIDBORG, and E. SUNDBERG, 2002 STY1 and STY2 promote the formation of apical tissues during Arabidopsis gynoecium development. Development **129**: 4707–4717.

LAMB, R. S., T. A. HILL, Q. K.-G. TAN, and V. F. IRISH, 2002 Regulation of APETALA3 floral homeotic gene expression by meristem identity genes. Development **129**: 2079–2086.

LARSON, P., and R. DICKSON, 1973 Distribution of imported ^{14}C in developing leaves of eastern cottonwood according to phyllotaxy. Planta **115**: 95–112.

LARSON, P. R., J. G. ISEBRANDS, and R. E. DICKSON, 1972 Fixation patterns of ^{14}C within developing

leaves of eastern cottonwood. Planta **107**: 301–314.

LEA, U. S., R. SLIMESTAD, P. SMEDVIG, and C. LILLO, 2007 Nitrogen deficiency enhances expression of specific MYB and bHLH transcription factors and accumulation of end products in the flavonoid pathway. Planta **225**: 1245–1253.

MAERE, S., K. HEYMANS, and M. KUIPER, 2005 BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics **21**: 3448–3449.

MASCLAUX, C., M. H. VALADIER, N. BRUGIÈRE, J. F. MOROT-GAUDRY, and B. HIREL, 2000 Characterization of the sink/source transition in tobacco (*Nicotiana tabacum* L.) shoots in relation to nitrogen management and leaf senescence. Planta **211**: 510–518.

MEYER, S., M. MELZER, E. TRUERNIT, C. HÜMMER, R. BESENBECK, *et al.*, 2000 AtSUC3, a gene encoding a new Arabidopsis sucrose transporter, is expressed in cells adjacent to the vascular tissue and in a carpel cell layer. Plant J **24**: 869–882.

MOROHASHI, K., M. ZHAO, M. YANG, B. READ, A. LLOYD, *et al.*, 2007 Participation of the Arabidopsis bHLH factor GL3 in trichome initiation regulatory events. Plant Physiol **145**: 736–746.

OPARKA, K. J., A. G. ROBERTS, P. BOEVINK, S. S. CRUZ, I. ROBERTS, *et al.*, 1999 Simple, but not branched, plasmodesmata allow the nonspecific trafficking of proteins in developing tobacco leaves. Cell **97**: 743–754.

PETRONI, K., G. FALASCA, V. CALVENZANI, D. ALLEGRA, C. STOLFI, *et al.*, 2008 The atMYB11 gene from Arabidopsis is expressed in meristematic cells and modulates growth in planta and organogenesis in vitro. J Exp Bot **59**: 1201–1213.

PITCHER, L., and J. DAIE, 1991 Growth and sink-to-source transition in developing leaves of sugarbeet. Plant and Cell Physiology **32**: 335–342.

QIN, G., H. GU, Y. ZHAO, Z. MA, G. SHI, *et al.*, 2005 An indole-3-acetic acid carboxyl methyltransferase regulates Arabidopsis leaf development. Plant Cell **17**: 2693–2704.

RAPP, J. C., and J. E. MULLET, 1991 Chloroplast transcription is required to express the nuclear genes *rbcS* and *cab*. Plastid Dna copy number is regulated independently. Plant Mol Biol **17**: 813–823.

RIAÑO PACHÓN, D. M., L. G. G. CORRÊA, R. TREJOS-ESPINOSA, and B. MUELLER-ROEBER, 2008 Green transcription factors: a chlamydomonas overview. Genetics **179**: 31–39.

RICE, P., I. LONGDEN, and A. BLEASBY, 2000 EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet **16**: 276–7.

ROBERTS, A. G., S. S. CRUZ, I. M. ROBERTS, D. A. M. PRIOR, R. TURGEON, *et al.*, 1997 Phloem unloading in sink leaves of *Nicotiana benthamiana*: Comparison of a fluorescent solute with a fluorescent virus. Plant Cell **9**: 1381–1396.

ROBERTS, I. M., P. BOEVINK, A. G. ROBERTS, N. SAUER, C. REICHEL, *et al.*, 2001 Dynamic changes in the frequency and architecture of plasmodesmata during the sink-source transition in tobacco leaves. Protoplasma **218**: 31–44.

ROOK, F., F. CORKE, M. BAIER, R. HOLMAN, A. G. MAY, *et al.*, 2006 *Impaired sucrose induction1* encodes a conserved plant-specific protein that couples carbohydrate availability to gene expression and plant growth. Plant J **46**: 1045–1058.

SADDIC, L. A., B. HUVERMANN, S. BEZHANI, Y. SU, C. M. WINTER, *et al.*, 2006 The LEAFY target LMI1 is a meristem identity regulator and acts together with LEAFY to regulate expression of

CAULIFLOWER. Development **133**: 1673–1682.

SAEED, A. I., V. SHAROV, J. WHITE, J. LI, W. LIANG, *et al.*, 2003 TM4: a free, open-source system for microarray data management and analysis. Biotechniques **34**: 374–378.

SOHLBERG, J. J., M. MYRENAS, S. KUUSK, U. LAGERCRANTZ, M. KOWALCZYK, *et al.*, 2006 STY1 regulates auxin homeostasis and affects apical-basal patterning of the Arabidopsis gynoecium. Plant J **47**: 112–123.

STADLER, R., K. M. WRIGHT, C. LAUTERBACH, G. AMON, M. GAHRTZ, *et al.*, 2005 Expression of GFP-fusions in Arabidopsis companion cells reveals non-specific protein trafficking into sieve elements and identifies a novel post-phloem domain in roots. Plant J **41**: 319–331.

STITT, M., 1989 Product inhibition of potato tuber pyrophosphate:fructose-6-phosphate phosphotransferase by phosphate and pyrophosphate. Plant Physiol **89**: 628–633.

STRACKE, R., H. ISHIHARA, G. HUEP, A. BARSCH, F. MEHRTENS, *et al.*, 2007 Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. Plant J **50**: 660–677.

TEPER-BAMNOLKER, P., and A. SAMACH, 2005 The flowering integrator FT regulates SEPALLATA3 and FRUITFULL accumulation in Arabidopsis leaves. Plant Cell **17**: 2661–2675.

TEPPERMAN, J. M., Y.-S. HWANG, and P. H. QUAIL, 2006 phyA dominates in transduction of red-light signals to rapidly responding genes at the initiation of Arabidopsis seedling de-etiolation. Plant J **48**: 728–742.

TIAN, C., E. CHIKAYAMA, Y. TSUBOI, T. KUROMORI, K. SHINOZAKI, *et al.*, 2007 Top-down phenomics of *Arabidopsis thaliana*: metabolic profiling by one- and two-dimensional nuclear magnetic resonance spectroscopy and transcriptome analysis of albino mutants. J Biol Chem **282**: 18532–18541.

TOMINAGA, R., M. IWATA, K. OKADA, and T. WADA, 2007 Functional analysis of the epidermal-specific MYB genes CAPRICE and WEREWOLF in Arabidopsis. Plant Cell **19**: 2264–2277.

TURGEON, R., 1984 Termination of nutrient import and development of vein loading capacity in albino tobacco leaves. Plant Physiol **76**: 45–48.

TURGEON, R., 1989 The sink-source transition in leaves. Annual Review of Plant Physiology and Plant Molecular Biology **40**: 118–138.

TURGEON, R., and J. A. WEBB, 1973 Leaf development and phloem transport in *Cucurbita pepo*: Transition from import to export. Planta **113**: 197–191.

TURGEON, R., and L. E. WIMMERS, 1988 Different patterns of vein loading of exogenous [^{14}C]sucrose in leaves of *Pisum sativum* and *Coleus blumei*. Plant Physiol **87**: 179–182.

ULLAH, H., J.-G. CHEN, B. TEMPLE, D. C. BOYES, J. M. ALONSO, *et al.*, 2003 The beta-subunit of the Arabidopsis G protein negatively regulates auxin-induced cell division and affects multiple developmental processes. Plant Cell **15**: 393–409.

VOGELMANN, T. C., P. R. LARSON, and R. E. DICKSON, 1982 Translocation pathways in the petioles and stem between source and sink leaves of *Populus deltoides* Bartr. ex Marsh. Planta **156**: 345–358.

VROEMEN, C. W., A. P. MORDHORST, C. ALBRECHT, M. A. C. J. KWAAITAAL, and S. C. DE VRIES, 2003 The CUP-SHAPED COTYLEDON3 gene is required for boundary and shoot meristem formation in Arabidopsis. Plant Cell **15**: 1563–1577.

WANG, S., S.-H. KWAK, Q. ZENG, B. E. ELLIS, X.-Y. CHEN, *et al.*, 2007 TRICHOMELESS1 regulates

trichome patterning by suppressing GLABRA1 in arabidopsis. Development **134**: 3873–3882.

WENZEL, C. L., M. SCHUETZ, Q. YU, and J. MATTSSON, 2007 Dynamics of MONOPTEROS and PIN-FORMED1 expression during leaf vein pattern formation in *Arabidopsis thaliana*. Plant J **49**: 387–398.

WRIGHT, K. M., A. G. ROBERTS, H. J. MARTENS, N. SAUER, and K. J. OPARKA, 2003 Structural and functional vein maturation in developing tobacco leaves in relation to AtSUC2 promoter activity. Plant Physiol **131**: 1555–1565.

YANHUI, C., Y. XIAOYUAN, H. KUN, L. MEIHUA, L. JIGANG, *et al.*, 2006 The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. Plant Mol Biol **60**: 107–124.

ZHAO, M., K. MOROHASHI, G. HATLESTAD, E. GROTEWOLD, and A. LLOYD, 2008 The TTG1-bHLH-MYB complex controls trichome cell fate and patterning through direct targeting of regulatory loci. Development **135**: 1991–1999.

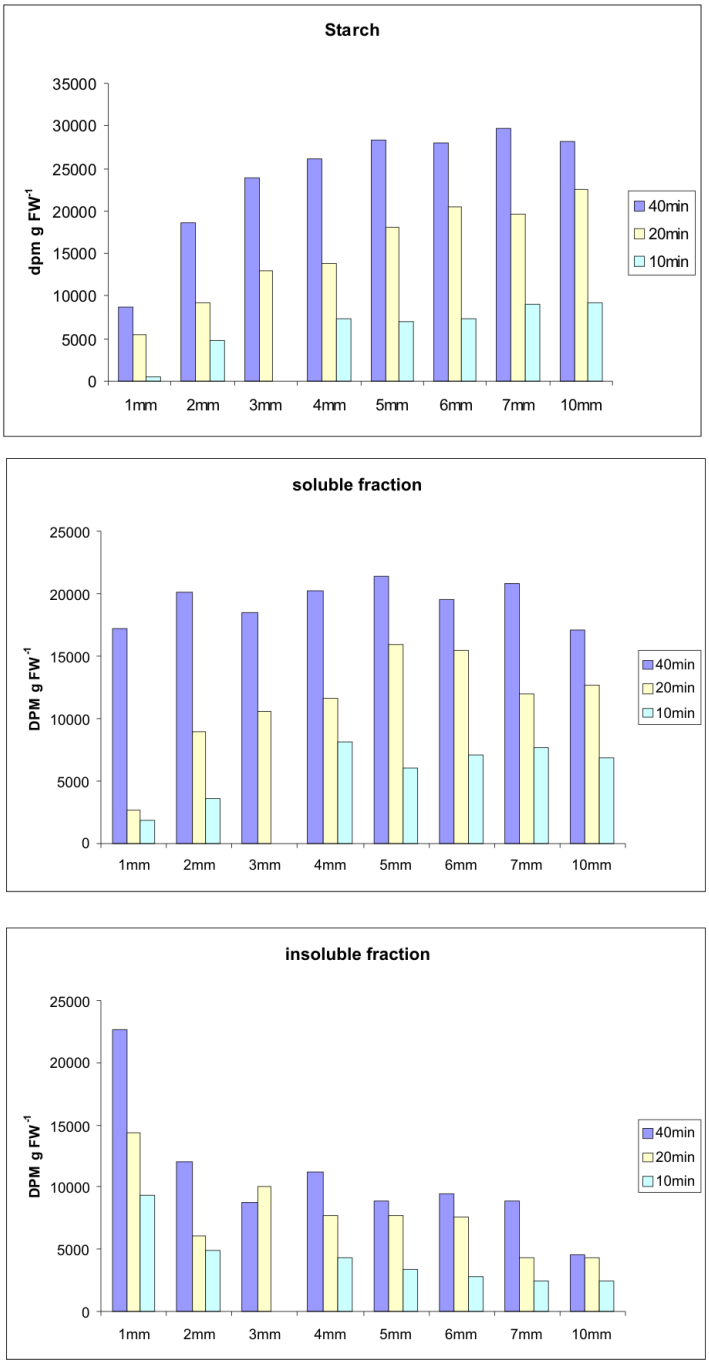


Figure 5.9: ***Carbon fluxes in Arabidopsis leaves***. *Arabidopsis plants were exposed to* ^{14}C *labeled* $CO-2$ *for 10, 20 and 40 min and levels of incorporation of radioactivity were measured in three different fractions: starch, soluble and insoluble. There is a clear increasing gradient in the labeled starch fraction towards bigger leaves, whereas this relation is inverted in the insoluble fraction, which represents cell wall material. In the soluble fraction, which corresponds mainly to sugars, the levels of incorporation are approximately the same at 40 min, but the amounts of sugar were less in sink leaves than in source leaves during the first 20 min, which may reflect the need to import sugar before achieving a stable value at 40 min.*

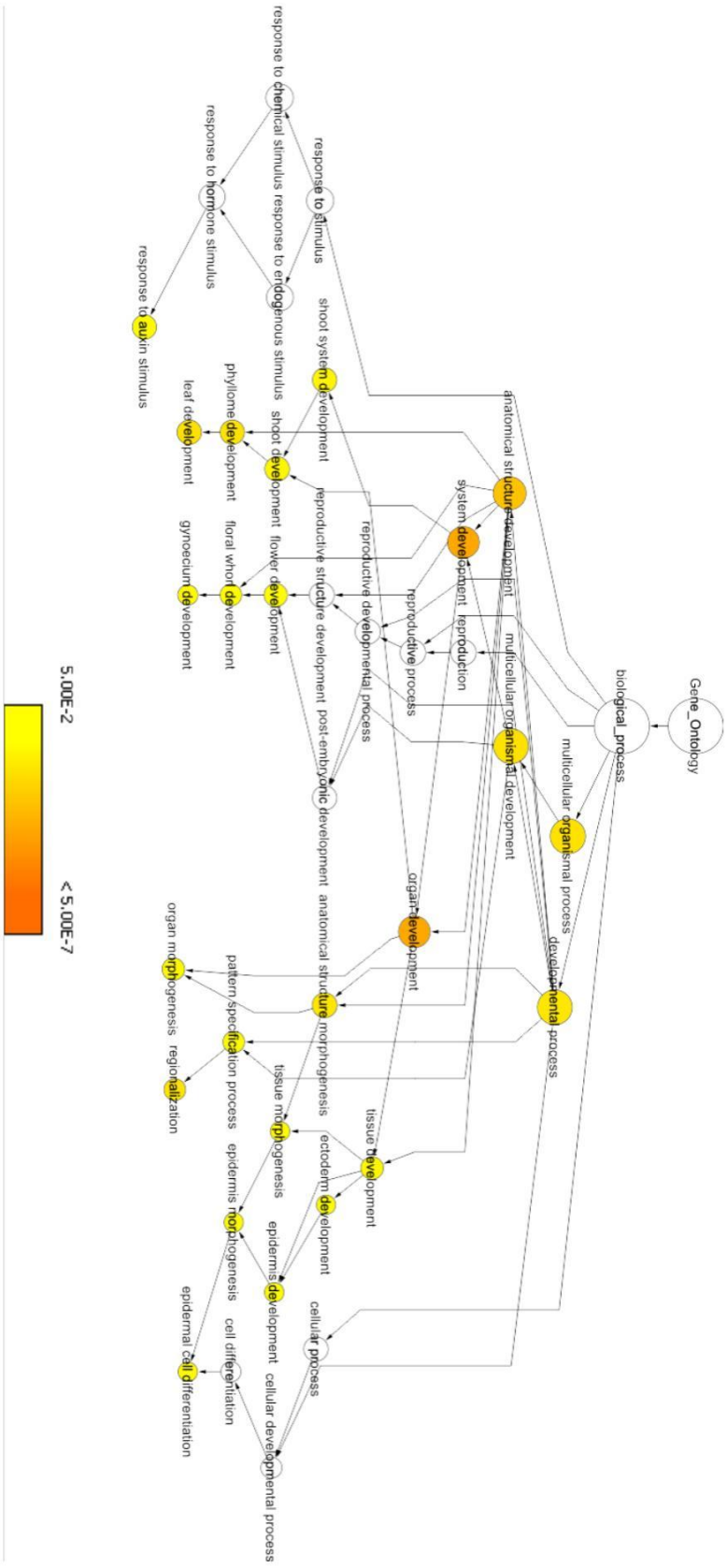


Figure 5.10: *GO overrepresentation among the 153 differentially expressed TFs.* *This is an hierarchical representation of the GO processes overrepresented among the differentially expressed TFs. GO processes that are statistically significant are colored according the heat map that indicates the p-value.*

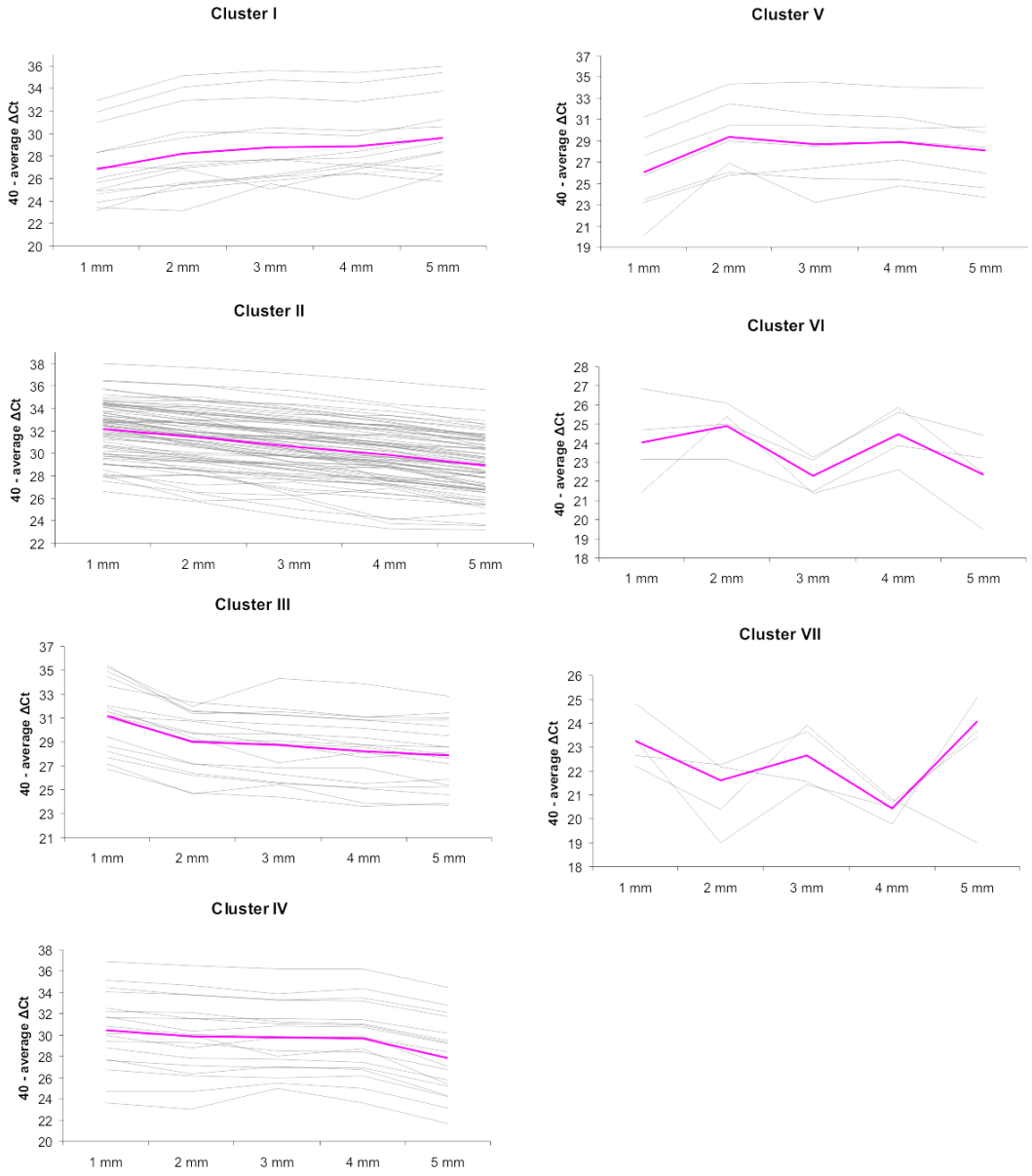


Figure 5.11: ***Clusters of the 153 transcription factors according to their expression pattern.*** *Clustering was performed using the K-means algorithm on Pearson correlations between genes. The best number of clusters was determined by the Figure of Merit (FOM). The average pattern for each cluster is shown in bold. All analyses were run in the MultiExperiment Viewer (MeV) part of the TM4 software from TIGR (*SAEED et al. *2003).*

6

Characterization of transcription factors that play a role in leaf growth

Leaf growth is a highly coordinated process. It has been clearly demonstrated that it is not only related to the genetic content of a given plant (i.e., different accessions; SCHMID *et al.* 2006). It is also regulated by light and nutrient availability, and affected by biotic and abiotic stresses. Additionally, leaves are responsible for detecting the day length and producing photoassimilates, and it is where much of the cross-talk between different phytohormones and nutrients take place. Coordinating the various signalling pathways involved requires a refined control of gene expression mainly by transcription factors, which modulate the initiation rate of gene transcription. In this chapter, three genes that play a role in leaf development are described: *bHLH64* (At2g18300), *bZIP21* (At1g08320) and *Dof4* (At4g38000).

6.1 Materials and Methods

6.1.1 Plant material and growth conditions

Arabidopsis thaliana (L.) Heynh. accession Col-0 was used in all experiments, except for plants carrying *35S::Dof4* and *Dof4RNAi* constructs, which are derived from the C24 cultivar. Tobacco (*Nicotiana tabacum*) cultivar SNN was used for transformation with the construct containing gene promoters fused to the Staphylococcus *GUS* plus open reading frame (ORF).

Mutant plants containing T-DNA insertions for *bHLH64* (Salk_015771), *bZIP21* (Salk_057570, Salk_057609 and Salk_141618) and *Dof4* (GK-384G11), as well as an RNAi mutant of *bHLH64* (CATMA2a16970) were ordered from NASC (`http://arabidopsis.info/`).

Seeds were sterilized in 70% ethanol followed by 50% sodium hypochlorite solution both for 20 min under continuous agitation. After washing four times with sterile water, seeds were dried. Sterile seeds were plated on AM medium (MURASHIGE and SKOOG 1962) containing 1% sucrose and 0.7% agar. For selection 50g/mL carbenicillin (T1 plants) and 20g/mL hygromycin, 50g/mL kanamycin or 1g/mL BASTA was added to the plates, depending on the resistance provided by the construct. Plates were left for 48h under vernalization conditions (16h light, 18C and 8h night, 4C) and then placed for two weeks under continuous light (100 mol m-2 s-1, 18C). After this period, plants were transplanted to soil (Einheitserde GS90; Gebrüder Patzer, Sinntal-Jossa, Germany) and cultivated at long (16h light, 8h darkness, relative humidity of 60/70%) or short (10h light, 14h darkness, relative humidity of 60/70%) day conditions. Alternatively, plants were cultivated on sterile AM media for hormone or sugar experiments.

6.1.2 Chemicals and enzymes

Most of the chemicals and reagents used were obtained from Sigma-Aldrich (Taufkirchen, Germany), Merck (Darmstadt, Germany), Carl-Roth (Karlsruhe, Germany) and Duchefa (Haarlen, The Netherlands). Restriction enzymes and T4 ligase were from New England Biolabs (Frankfurt am Main, Germany). *Taq* and *Pfu* polymerases were from Stratagene (Waldbronn, Germany). RNA extraction was done with RNeasy Mini Kit and cDNA synthesis with reverse transcriptase from Qiagen (Dusseldorf, Germany). SYBR Green for RT-PCR was obtained from Applied Biosystems (Darmstadt, Germany).

6.1.3 Sequencing

DNA sequencing was performed by MWG (Martinsried, Germany).

6.1.4 Bacteria

The following bacterial strains were used in this work: *Escherichia coli* DH5α (HANAHAN 1985)

Agobacterium tumefaciens pGV3101 (KONCZ and SCHELL 1986)

E. coli was cultivated in YT medium (SAMBROOK and RUSSELL 2001) and *A. tumefaciens* in YEB medium (VERVLIET *et al.* 1975).

6.1.5 Histochemical detection of reporter enzyme activities

Gene expression patterns were analysed via β-glucuronidase (GUS) assays. The promoter of the gene of interest was fused to the Staphylococcus GUS (β-glucuronidase)-reporter

gene in the pCAMBIA 1305 vector (CAMBIA, Canberra, Australia), and transformed into plants. GUS activity was detected by incubating plant tissue in X-Gluc buffer (2 mM X-Gluc, 100 mM sodium phosphate buffer pH 8.0, 10 mM EDTA, 1 mM potassium ferricyanide, 1 mM potassium ferrocyanide and 0.1% Triton) for 2 to 48h, depending on the intensity of gene expression and the tissue analysed. Tissue was cleared by successive washes with 70% ethanol.

6.1.6 Microscopy

Histological analyses of cell size and gene expression patterns were performed with the stereo-fluorescence microscope MZ 16FA from Leica (Wetzlar, Germany) using the LAS suite software from the same manufacturer; and the microscope AX70 from Olympus (Hamburg, Germany), using the CellP̂ software from the same manufacturer.

6.1.7 Constructs and cloning strategy

All fragments amplified in this work were cloned in pCR2.1 (Invitrogen, Karlsruhe, Germany) after PCR amplification using a mix of *Pfu* and *Taq* (9:1 ratio). The primers used for amplification are listed in Annex 1.

6.1.7.1 Promoter constructs

Fragments corresponding to the promoter of *bHLH64* (1.5 kb), *bZIP21* (1.5 kb) and *Dof4* (1.3 kb) were amplified from genomic DNA of Arabidopsis and cloned into pCR2.1. Vectors containing the promoters were digested using the enzymes corresponding to the restriction sites added (Annex 1). Fragments corresponding to the promoter were fused to the GusPlus reporter gene (BROOTHAERTS *et al.* 2005; PbZIP21::Gus+) in the pCAMBIA 1305 vector. Final constructs were verified by PCR and sequencing with a reverse primer annealing to the 300 bp region of the *GusPlus* coding region (Annex 1).

6.1.7.2 Overexpression constructs

Fragments corresponding to the ORFs of *bHLH64* (1,014 bp) and *bZIP21* (1,446 bp) were amplified from cDNA from early developing leaves and cloned into pCR2.1. Vectors containing the ORFs were digested using the enzymes corresponding to the restriction sites added (Annex 1). Fragments corresponding to the ORFs were fused to the 35S CaMV promoter in the pGREEN 029 vector (JIC, Norwich, England;HELLENS *et al.* 2000). Final constructs were verified by PCR and sequencing with a foward primer annealing to the 200 bp distal region of the 35S CaMV promoter (Annex 1).

6.2 Characterization of *bHLH64*

6.2.1 Introduction

The basic/helix-loop-helix (bHLH) family is one of the most numerous TF families in plants. bHLH transcription factors were present in the MRCA of all eukaryotes, as they can be also found in animals and fungi. In plants, these TFs regulate different processes including gynoeceum development (RAJANI and SUNDARESAN 2001), flavonoid biosynthesis (NESI *et al.* 2000), trichome differentiation (PAYNE *et al.* 2000), microspore development (SORENSEN *et al.* 2003), ABA response (ABE *et al.* 2003), tryptophan biosynthesis (SMOLEN *et al.* 2002), freezing tolerance (CHINNUSAMY *et al.* 2003), brassinosteroid signalling (FRIEDRICHSEN *et al.* 2002) and phytochrome signalling (FAIRCHILD *et al.* 2000, HUQ and QUAIL 2002, NI *et al.* 1998). This last feature is important, as it shows that bHLHs play a key role in light signal transduction (JIAO *et al.* 2007). However, bHLHs of the PHYTOCHROME INTERACTING FACTOR (PIF) subfamily are not present in algae (see Chapter 3, indicating that their emergence was associated with tissue/organ specialization for photosynthesis, as phylloids or leaves.

bHLH64 is phylogenetically close to *BEE1*, *BEE2* and *BEE3*, which are bHLHs involved in brassinosteroid regulated plant growth. Although the *BEE* genes seem to be directly involved in the response to brassinosteroids, *bHLH64* appears to have a different role (FRIEDRICHSEN *et al.* 2002).

Here, we describe the identification of bHLH64 as a possible player in the leaf sink-to-source transition and its initial characterization.

6.2.2 Results

6.2.2.1 *bHLH64* expression changes during early leaf development

Transcription factor expression profiling of leaves from 1 to 5 mm was performed using qRT-PCR (see Chapter 5). One of the genes that was found to be differentially expressed during the sink-to-source transition is *bHLH64*, which showed high expression at the onset of the transition (Figure 6.1 A). Additionally, expression level of *bHLH64* at late phase of leaf development (50% expanded leaf, 100% expanded leaf and 20% senescent leaf; BALAZADEH *et al.* 2008) have shown that *bHLH64* is also differentially expressed during the onset of senescence, but it is repressed during this process (Figure 6.1 B). Considering its expression pattern, *bHLH64* is an interesting gene that could play a role in both sink-to-source and senescence, thus in leaf lifespan.

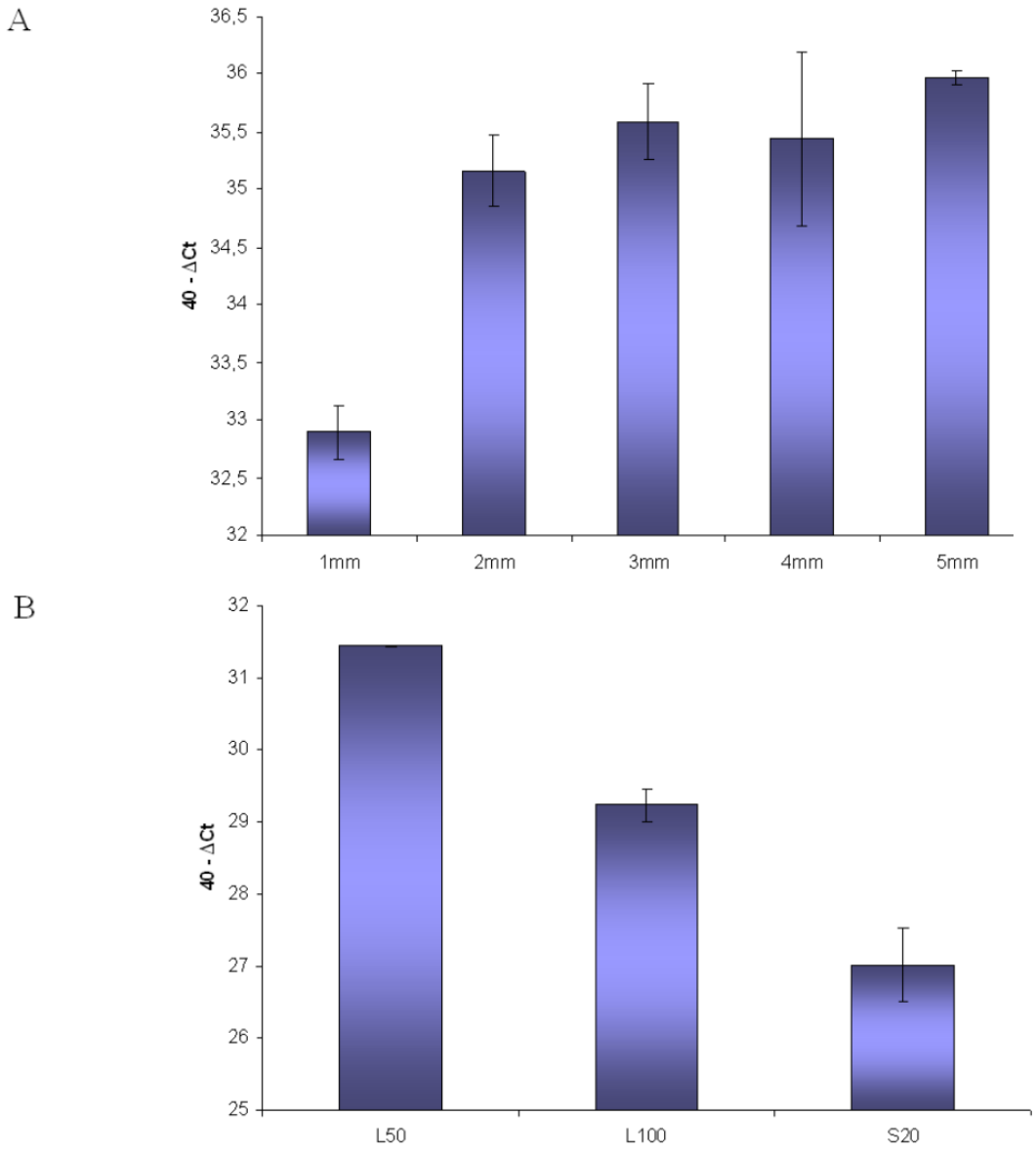


Figure 6.1: ***Expression of bHLH64 during the sink-to-source and mature-to-senescent transitions.*** *Expression level of bHLH64 was accessed via qRT-PCR. (A) Expression of bHLH64 increases at the onset of the sink-to-source transition, whereas (B) it decreases during the onset of senescence The stages L50, L100 and S20 correspond to leaves, which are 50% of the full expanded leaf size (1.5 cm), fully expanded leaf (3 cm) and 20% senescent leaf. These stages are described in* BALAZADEH et al. *(2008). Expression levels were normalized to the expression of UBQ10. Expression correspond to the mean value of three biologic replicates with two technical replicates and the error bars correspond to the standard deviation.*

6.2.2.2 Expression pattern of *bHLH64*

Plants carrying a construct of the *bHLH64* promoter fused to the reporter gene *GusPlus* (PbHLH64::Gus+) were selected on media containing hygromycin, and 20 lines were

tested in Gus assays, and 17 from these lines showed a strong overlap in the staining-pattern observed. Three lines, representing the most common pattern observed, were used for further characterization of *bHLH64* expression pattern. *bHLH64* was found to be expressed in the central vasculature of main and secondary roots; no expression was detected in root hairs or root tips (Figure 6.2). As expression could be detected in roots but not in the hypocotyl (Figure 6.2), *bHLH64* can be used as a marker for the hypocotyl/root transition. Additionally, this gene is expressed in the shoot apical meristem, where expression is prominent in the region of leaf initiation. Leaves smaller than 0.5 mm showed no expression, whereas GUS activity became visible at the tip of leaves whose length exceeded 0.5 mm. GUS activity increased in a basipetal manner as leaves grew further (Figure 6.2). Bigger leaves were completely stained, most notably in the veins, although staining intensity started to decline when leaves reached length of more than 7 mm. During leaf senescence, no expression could be detected. The expression observed by GUS staining overlaps with the one observed in Genevestigator and eFP browser, except for the expression in hypocotyls.

A similar expression pattern was found in leaves of tobacco plants transformed with the PbHLH64::Gus+ construct(Figure 6.3). During early development, no GUS staining was observed in leaves smaller than 3 cm; leaves bigger than 3 cm showed expression at the tip, and as leaves expanded, stain was observed in the whole leaf. No staining was observed in senescent regions of the leaf. Additionally, roots were stained, whereas no staining was observed in hypocotyl. The similarity of the expression patterns in both plants, Arabidopsis and tobacco, indicates that the cellular network controlling the expression of *bHLH64* is, at least partially, conserved in eudicots.

To identify if the expression of *bHLH64* is conserved in monocots, rice plants were transformed with the PbHLH64::Gus+ construct. Nevertheless, attempts to transform rice with this construct have not successful.

6.2.2.3 The expression of *bHLH64* follows the diurnal cycle

Leaf samples were collected at three different time points during the day (3 am, 9 am and 7 pm), corresponding to three hours before light, three hours into the light and three hours before darkness, respectively. Expression of *bHLH64* was found to be higher in the light than in the dark (Figure 6.4). This expression change during the day is in agreement with the data obtained by (BLÄSING *et al.* 2005) measuring the expression of genes in whole rosette every 4 h within 24h (Figure 6.4). Thus, *bHLH64* is not only regulated by developmental cues, i.e. the sink-to-source transition, but also according to light rhythm. This is an interesting feature, as we could cluster the TFs undergoing expression changes during the sink-to-source transition into two groups: (i) genes related to light-dependent

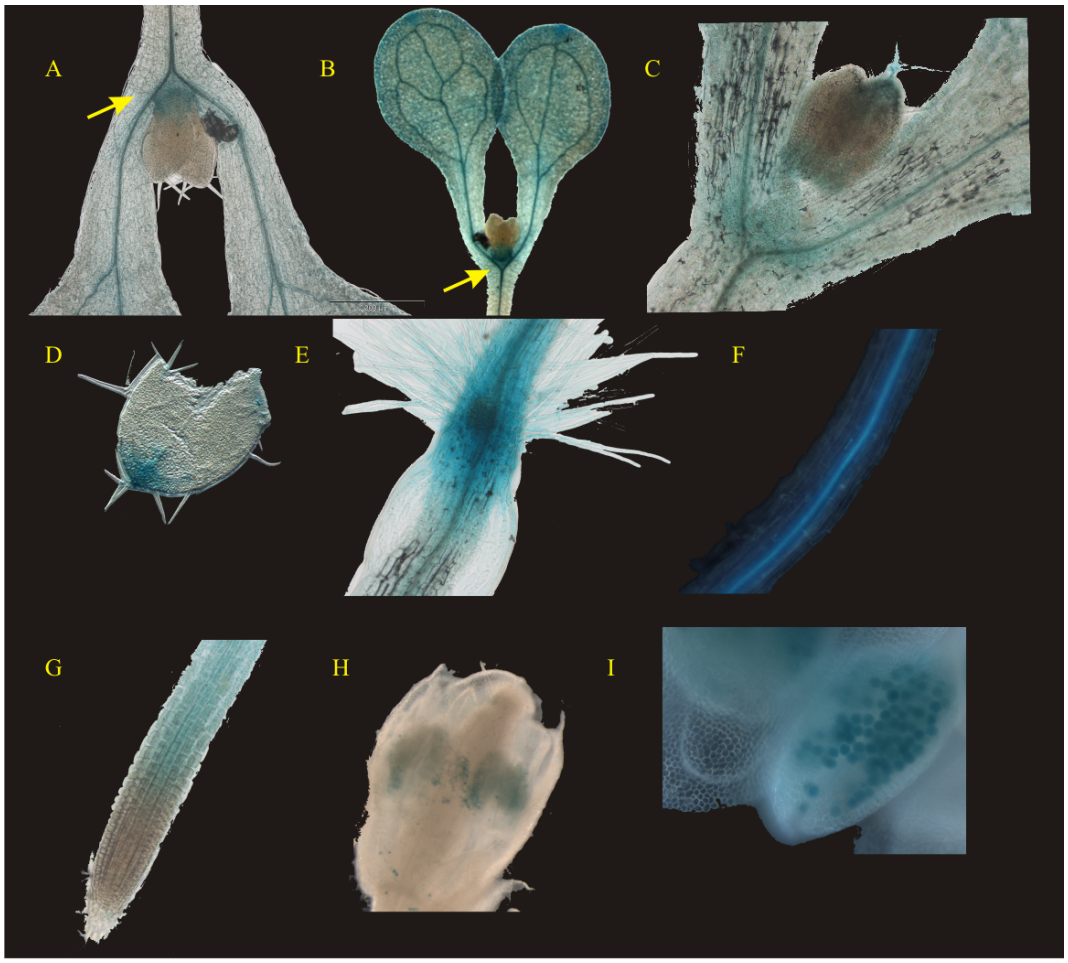


Figure 6.2: ***Expression pattern of the bHLH64 gene.*** *(A, B) bHLH64 is expressed in cotyledons, especially in veins. Expression is also observed in the SAM region (yellow arrows). (B) Leaves smaller than 0.4 mm do not show any GUS staining. (C, D) Expression starts at 0.5 mm and it develops in a basipetal manner, indicating a sink-to-source transition dependent expression. (E) Normally no or very pale staining is observed in the hypocotyl, and the transition between shoot and root is very prominent. (F) In roots, expression is mainly restricted to vascular tissue. (G) No expression is observed in root tips. (H, I) In flowers, expression is restricted to pollen grains.*

processes; and (ii) genes independent of light or leaf growth/expansion (see Chapter 5. Therefore, as bHLH64 is differentially expressed during the day, it could be involved in light dependent processes. TIAN *et al.* (2007) have analysed the transcriptome and metabolome of albino mutants of Arabidopsis, and found that *bHLH64* was one of the most strongly repressed genes when compared to wild-type plants, indicating that this gene is possibly regulated by light. Additionally, OSUNA *et al.* (2007) have shown that this gene is repressed after addition of sucrose to C-starved plants. Genes induced by light may also be part of the feast/famine control (KOCH 1996), therefore, they are generally

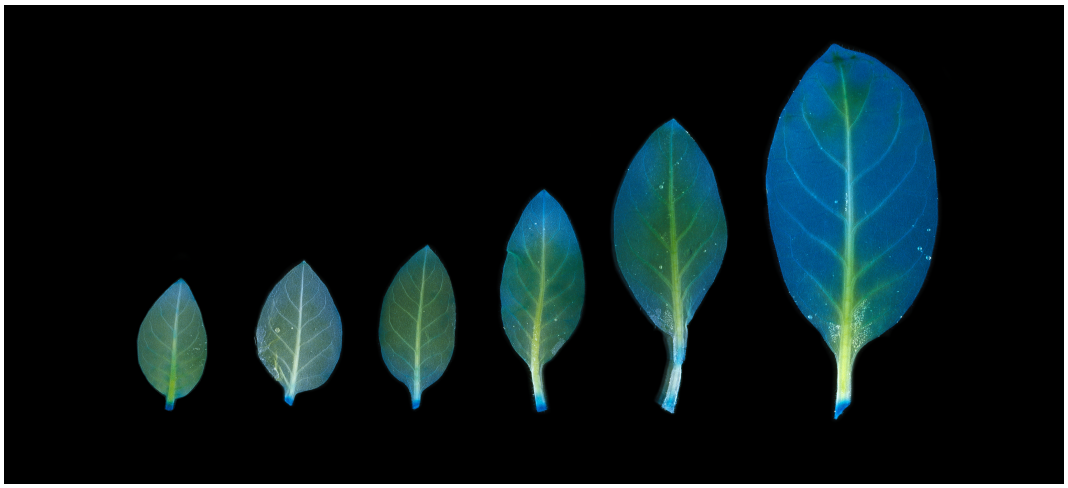


Figure 6.3: ***Expression pattern of the bHLH64 gene in tobacco leaves.*** *Tobacco plants transformed with the PbHLH64::Gus+ construct were analysed for the conservation of the expression pattern. No staining was visible in leaves smaller than 3 cm. In leaves bigger than 3 cm staining progressed not only in a basipetal manner (as expected for the sink-to-source transition) but also from the border to the centre.*

also sensitive to sugar.

Co-expression analysis performed using ATTED-II (`http://www.atted.bio.titech.ac.jp/`) revealed that *bHLH64* expression is strongly correlated with the expression of *PHOT1*, which encodes a blue light receptor, and *PIF4*, encoding a phytochrome interacting factor (Figure 6.5). These observations are in agreement with the results obtained for the albino mutants (TIAN *et al.* 2007). As PIF4 is also a TF of the bHLH family, one might hypothesise that light signal transduction is passed from PIF4 to bHLH64 via heterodimerization. However, interaction of PIF4 and bHLH64 has not yet been demonstrated. Also, *bHLH64* expression has not been analysed in *pif4* or *phot1* mutants.

We tested the expression level of *bHLH64* via qRT-PCR in mutants that have an altered sugar content or perception. Interestingly, *bHLH64* expression was found to be increased by at least three fold in *sex1* and *abi5* mutants, and 1.5-fold in *abi4* and *pwd* mutants (Figure 6.6 A). The ABA insensitive mutants allowed identifying a link between ABA and sugar signalling (DEKKERS *et al.* 2008). The *abi4* mutant was described as sucrose insensitive or uncoupled (HUIJSER *et al.* 2000). Thus lack of sugar sensing ability may block the repression of *bHLH64* in this mutant, at least to some extent. A similar conclusion could be drawn for abi5, though no sugar insensitivity was reported so far in this mutant. Starch excess mutants like *pwd* and *sex1* (FETTKE *et al.* 2006, YU *et al.* 2001) are less efficient in the degradation of starch to sugar. As more starch is accumulated, less sugar is left for signalling gene repressing through sugar level (Jörg Fettke,

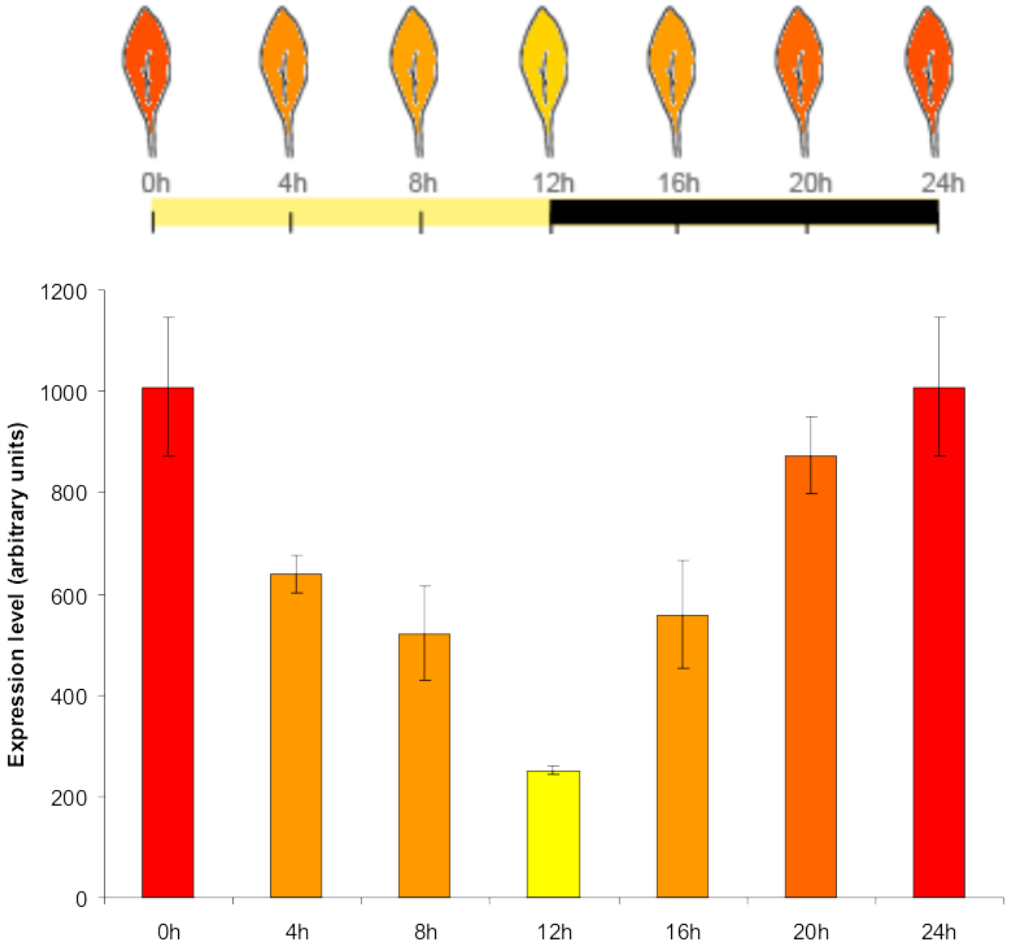


Figure 6.4: ***Diurnal expression of bHLH64.*** *Plants were grown in a 12/12 light regime and whole rosettes were harvested. The expression level of bHLH64 is reduced during the light period, reaching its lowest level before light is turned off. Data obtained from* BLÄSING et al. *(2005).*

personal communication); this might explain why the expression of *bHLH64* is higher in these mutants.

During leaf senescence, expression of *bHLH64* decreases (Figure 6.1 B). When senescence occurs, macromolecules are degraded and nutrients are mobilized to other parts of the plant, finally leading to a programmed cell death. To a certain extent, wounding mimics the effect of senescence, as there is cell death and remobilization of nutrients, especially sugar, whose concentration increases in the wounded area (put reference here). We tested the expression level of *bHLH64* in wounded leaves and observed that it is highly responsive to wounding. The expression level increased up to 11 fold within 60 minutes after wounding (Figure 6.6 B).

Figure 6.5: *Co-expression network of bHLH64.* *This co-expression network was obtained from ATTED-II* `http://www.atted.bio.titech.ac.jp/`. *Expression of PHOT1 and PIF4 are closely correlated with bHLH64 expression. Phylogenetic analyses have shown that bHLH64 belongs to a cluster of genes related to brassinosteroid response (BEE), however, it does not seem to respond directly to brassinosteroids* (FRIEDRICHSEN et al. 2002). *The position of bHLH64 in the network might indicate that it is involved in a cross-talk between auxin, brassinosteroids, ethylene, and light responses. In red, bHLH64. In yellow, the bHLH TFs involved in light response. In orange, the bHLH TFs involved in brassinosteroid response.*

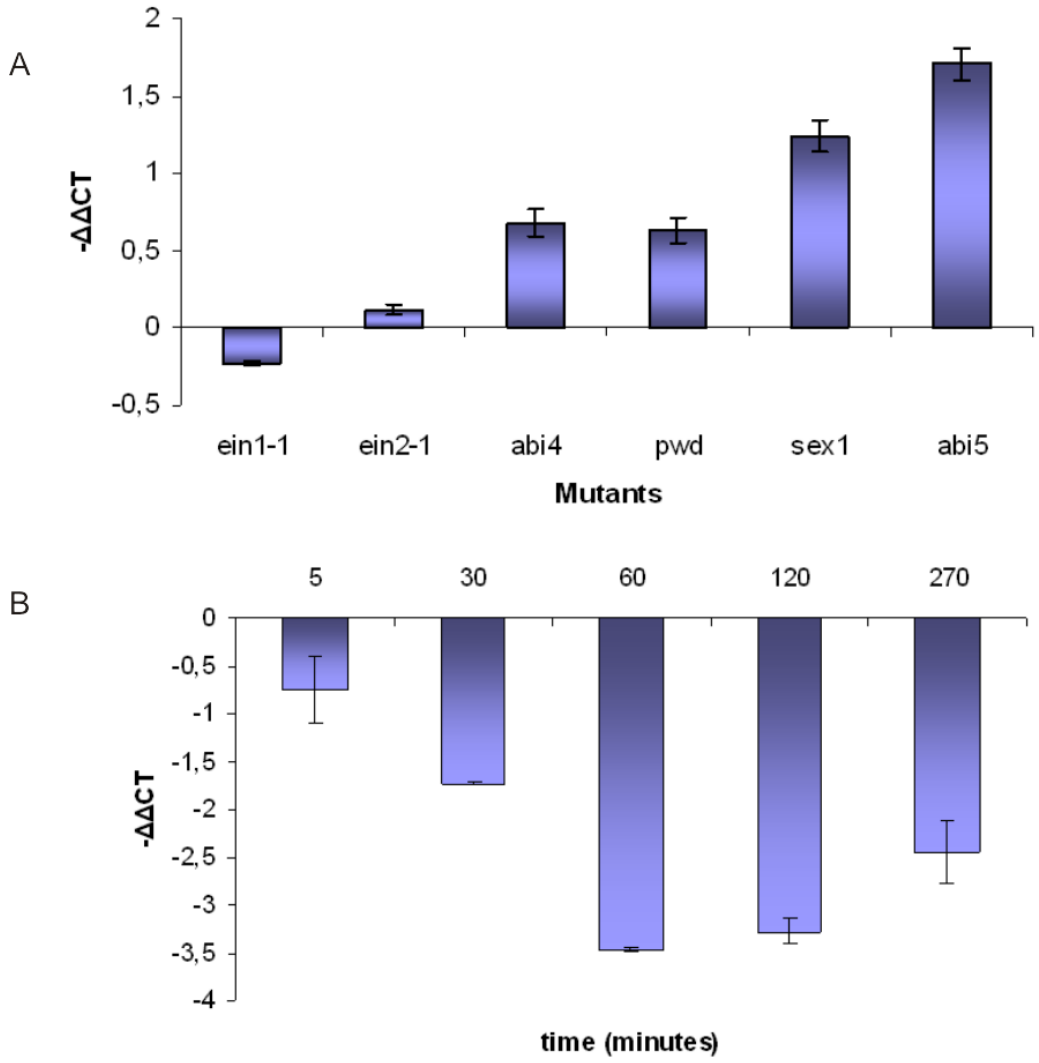


Figure 6.6: ***Expression of bHLH64 in different mutants and in wounding experiments.*** *(A) Expression of bHLH64 was measured in mutants insensitive to hormones (ein1-1, ein2-1, abi4, abi5) and with a defect in starch degradation (pwd and sex1). (B) Arabidopsis leaves of 1 cm were injured and harvested 5, 30, 60, 120 and 270 min later. Expression correspond to the mean value of two biologic replicates with three technical replicates and the error bars correspond to the standard deviation. Values in the graphics correspond to the average of bHLH64 expression in a treatment minus the average of bHLH64 expression in wild-type plants (A) or in time 0 of wounding (B).*

6.2.2.4 *bHLH64* is a regulator of bolting age

To gain information about the role of *bHLH64*, mutant plants with reduced or increased expression of *bHLH64* were analysed. Phenotype analyses of both knock-out and RNAi lines showed that these plants have a retarded development when compared to the wild-type. Bolting age is delayed by about 20 to 30 days in both RNAi and KO plants (Figure 6.7), thus we can assume that bHLH64 is an important regulator of bolting. As ex-

pression of *bHLH64* increases during early stages of leaf development (Figure 6.1 A), lack of *bHLH64* may lead to a later acquisition of leaf maturity. Furthermore, no apparent difference of leaf size and silique number was observed between the mutants and the wild-type, when plants reach 20% senescence of siliques. The overexpression lines do not show any significant difference in the bolting age when compared to the wild-type. Therefore, although *bHLH64* may play a role in bolting, its overexpression is not sufficient to induce early bolting, thus another limiting element must exist, preventing that *bHLH64* overexpression to alter the plant phenotype.

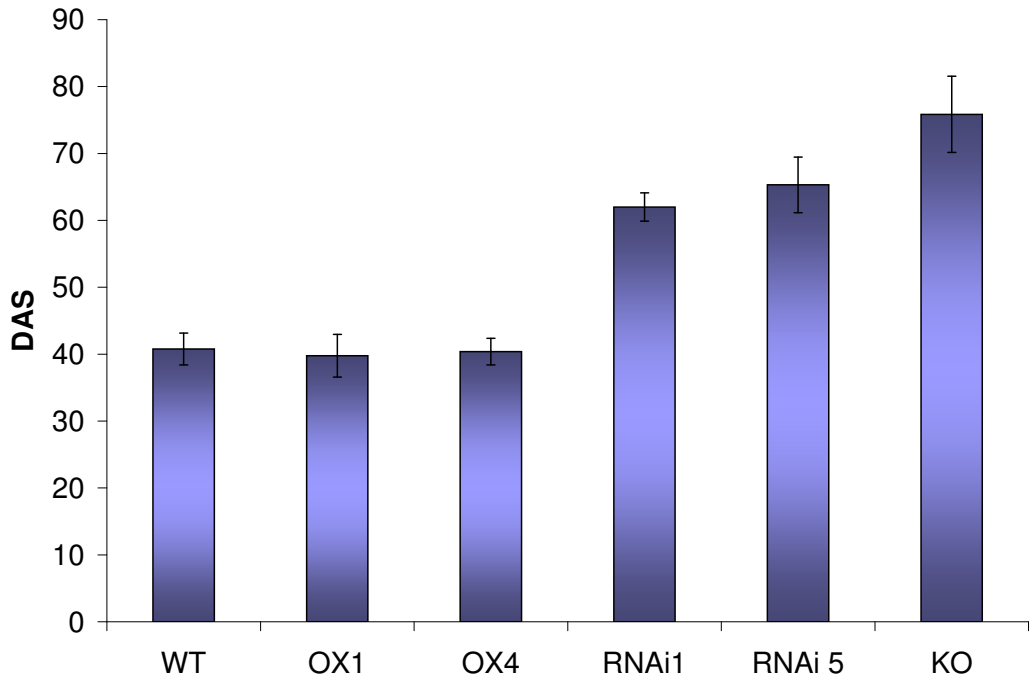


Figure 6.7: ***Comparision of bolting age between wild-type and OX, RNAi and KO bHLH64 mutants.*** *Plants were grown in the greenhouse under long-day condition. Plants were considered to have reached bolting age when the main flowering stalk was = 1cm. From each line, 14 plants were measured. Wild-type and overexpression plants bolted around 40 days after sowing, and RNAi and KO plants bolt 20 to 30 days after the wild-type. The error bars correspond the standard deviation. DAS, days after sowing.*

6.3 Characterization of *bZIP21*

6.3.1 Introduction

The basic leucine zipper (bZIP) family is one of the largest TF families in plants, and it is shared by all eukaryotes. Members of this TF family play a role in different processes during plant lifespan (see Chapter 3 for detailed description). Despite being one of the first TF families to be described in plants, with classical work on maize Opaque2 dating 20 years ago (MADDALONI *et al.* 1989), there is a great number of genes and PoGOs to which no function can be assigned yet. Efforts towards uncovering the evolutionary relationships among bZIPs has helped to achieve a more systematic characterization of this gene family in recent years. Especially the regulatory relations between bZIPs of Groups C and S have been analyzed in much detail (CORRÊA *et al.* 2008, EHLERT *et al.* 2006, JAKOBY *et al.* 2002, VETTORE *et al.* 1998, WELTMEIER *et al.* 2006). In our expression profiling study that we performed to uncover genes differentially expressed during the sink-to-source transition in Arabidopsis leaves, we have identified four bZIPs (*AtbZIP 21*, *27*, *46* and *70*) undergoing expression changes. AtbZIP27 is also known as FDP, a paralog of FD, which is a mediator of flower signalling at the shoot apex (ABE *et al.* 2005). AtbZIP46 is PAN, involved in the determination of floral organ number in Arabidopsis (CHUANG *et al.* 1999). AtbZIP70, though not characterized, may play a role in energy homeostasis due to its interaction with bZIPs from Group C (CORRÊA *et al.* 2008, EHLERT *et al.* 2006). Here, we describe a preliminary characterization of bZIP21.

6.3.2 Results

6.3.2.1 Expression of *bZIP21* during sink-to-source transition

Expression profiling of TF genes using qRT-PCR during early leaf development (see Chapter 5) indicates increased expression of *bZIP21* during the sink-to-source transition (Figure 6.8). Additionally, bZIP21 is apparently induced by senescence; its expression increases 23-fold during natural senescence and 87-fold during induced senescence (BALAZADEH *et al.* 2008; Genevestigator, respectively; Figure 6.9). The fact that *bZIP21* is differentially expressed during both the sink-to-source transition and at the onset of senescence is an important clue that this TF is regulating both processes.

6.3.2.2 The expression pattern of *bZIP21*

Plants carrying a construct with the promoter of *bZIP21* fused to the *GusPlus* reporter gene were selected on media containing hygromycin, and 25 lines were tested. The ex-

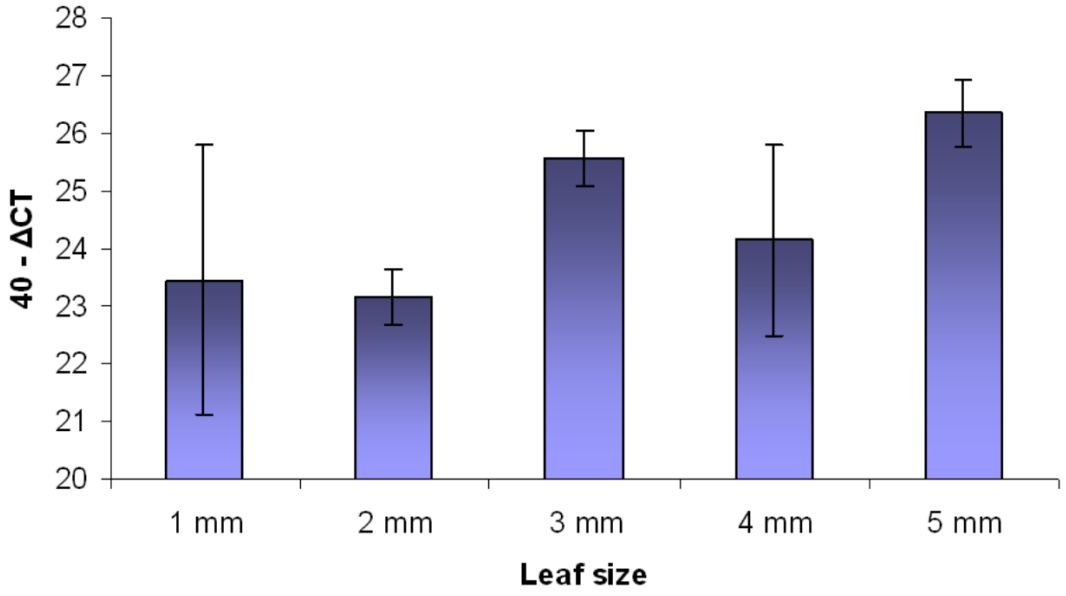


Figure 6.8: ***Expression of bZIP21 during the sink-to-source transition.*** *Expression levels were assesed via qRT-PCR using leaves from 1 to 5 mm, comprising the sink-to-source transition. Expression of bZIP21 increases alongside the sink-to-source transition, notably between 2 and 3 mm. Expression levels were normalized to the expression of UBQ10. Expression correspond to the mean value of three biological replicates with two technical replicates and the error bars correspond to the standard deviation.*

pression pattern was similar in 20 lines, differing basically in the presence (4 lines) or absence (16 lines) of staining in the tip of the root. The remaining 5 lines showed staining just in roots, but not in root tips. Three lines representing the most common pattern were further characterized. Expression of *bZIP21* is visible early in embryogenesis at the stage of single-terminal cell (Figure 6.10 A) and pre-globular (Figure 6.10 B). Expression is observed throughout embryogenesis in globular, heart and torpedo stages, and in mature seeds, in the whole embryo (Figures 6.10 C to F). During development, expression was detected in roots, but not in the hypocotyl (Figures 6.10 G and H). The expression pattern in leaves follows the sink-to-source basipetal gradient (Figure 6.10 I), and there is a significant overlap of the expression observed for *AtCHoR* and *AtSUC2* (see Chapter 5) in leaves of ca. 2.5 mm. In leaves larger than 7 mm the whole leaf blade was stained, but normally no stain was observed in the mid vein (Figure 6.10 J). In this part of the leaf there is the highest concentration of sugar, due to sugar export, but sugars do not seem to have a strong effect on the expression of *bZIP21*. In flowers, the expression is observed in pollen grains and in the stigma, but the latter is just observed when the pistil is longer than the stamen (Figures 6.10 K and M).

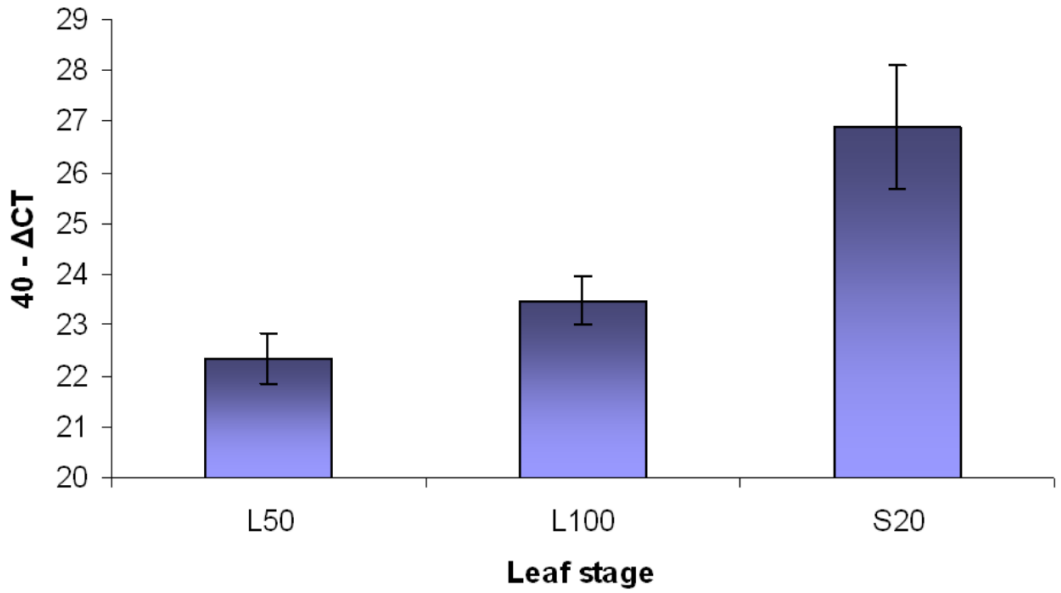


Figure 6.9: ***Expression of bZIP21 during late leaf development.*** *Expression analyses using qRT-PCR indicates that expression of bZIP21 is induced during leaf senescence. L50, 50% of fully developed leaf (1.5cm leaf length); L100, full-developed leaf (3cm); S20, 20% senescent leaf. The expression level of UBQ10 was used for normalizing the values. Data obtained from (*BALAZADEH et al. *2008).*

6.3.2.3 The evolution of *bZIP21*

bZIP21 belongs to the bZIP multigenic family of TFs, which is shared by all eukaryotes. Phylogenetic analyses have classified *bZIP21* into Group D, to which also TGAs and PAN belong, which are involved in pathogen responses and in floral organ number determination, respectively (HEPWORTH *et al.* 2005, KESARWANI *et al.* 2007). Interestingly, *bZIP21* is the only gene from Arabidopsis that belongs to PoGO D2 (Figure 6.11). The function of its four rice counterparts has not yet been identified, but it is an orthologue of the maize *Liguleless2* (*LG2*) gene that has a role in the formation of the ligule (HARPER and FREELING 1996). The ligule is a membrane-like tissue attached to the leaf of a grass at the point where the blade meets the leaf sheath. Ligules are observed mainly in monocots, although they are also present in some lycopods and some members of the Asteraceae family; the physiological role of the ligule is currently unknown. As most eudicots do not posses ligules, the function of bZIPs conserved in PoGO D2 remains open.

6.3.2.4 Analysis of bZIP21 function

Three possible transcripts are generated from the *bZIP21* gene (Figure 6.12). When comparing the three transcripts, two of them (1 and 3) have a longer 5' region with an intron in

Figure 6.10: ***Expression pattern of bZIP21.*** *bZIP21 is expressed in early developing embryos in the following stages: (A) single-terminal cell stage, (B) pre-globular, (C) globular, (D) heart, (E) torpedo (F) fully mature embryo. (G) Staining is detected in roots but never in hypocotyls, what makes bZIP21 a good marker for the transition between these two tissues (arrow). (H) Staining is not observed in leaves smaller than 2.5 mm, from this size onwards, expression is observed in the tip of the leaf (I) and from where it spreads towards the base in a basipetal manner. (J) Leaves bigger than 7 mm are fully stained, except for the main vein region. (L) In flowers, expression is observed in the pollen grains as well in the conductive tissue leading to the anther (M). (N) expression is also observed in the stigma, but expression is only visible when the pistili is bigger than the stamen.*

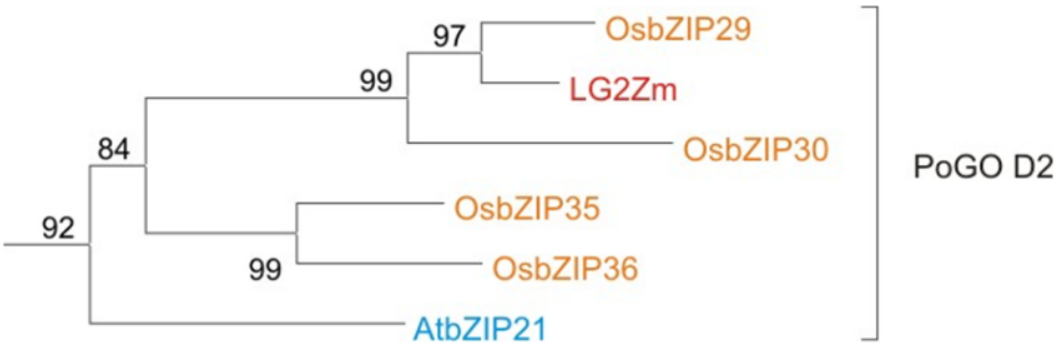


Figure 6.11: ***Evolutionary relationship of bZIP21.*** *Phylogenetic analyses of bZIPs from different plants (see Chapter 3) allowed clustering bZIP21 into PoGO D2. The only gene in this PoGO so far characterized is LG2 from maize, which plays a role in the formation of the ligule. This tissue is absent in most eudicots. The role of PoGO D2 has not yet been analysed in detail in eudicots so far. In orange, bZIPs from rice (OsbZIP) that belong to PoGO D2.*

the 5' UTRs. This intron possesses two distinct open reading frames (one with 185 bp and the other with 89 bp;`http://arabidopsis.org`), which might perhaps be involved in regulating the translation of the bZIP21 protein (LEE *et al.* 2006). Additionally, the intron region can also undergo differential splicing (Figure 6.13). PCR experiments with primers annealing to the first exon and to the junction between the second and third exon confirmed the existence of both transcripts (Figure 6.13). Thus, differential splicing of the primary transcript might also contribute to the regulation of the expression / translation of the *bZIP21* gene. A similar observation has been made for other bZIP genes, such as *AtbZIP11* (WIESE *et al.* 2004).

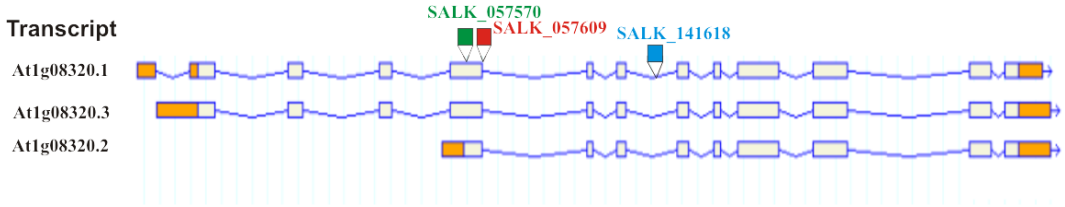


Figure 6.12: ***Three possible transcripts for bZIP21 and position of T-DNA insertions.*** *Information about these transcripts was obtained via databank analyses. Continuous lines, introns; yellow boxes, exons; orange boxes, untranslated regions; green, red and blue, T-DNA insertion regions. This scheme was partially obtained from http://atidb.org/cgi-perl/gbrowse/atibrowse/.*

It is worth mentioning that the expression pattern analyses done in this thesis were based only on the promoter region corresponding to transcripts 1 and 3. Expression pattern of the promoter region corresponding to the second transcript was not analysed. Al-

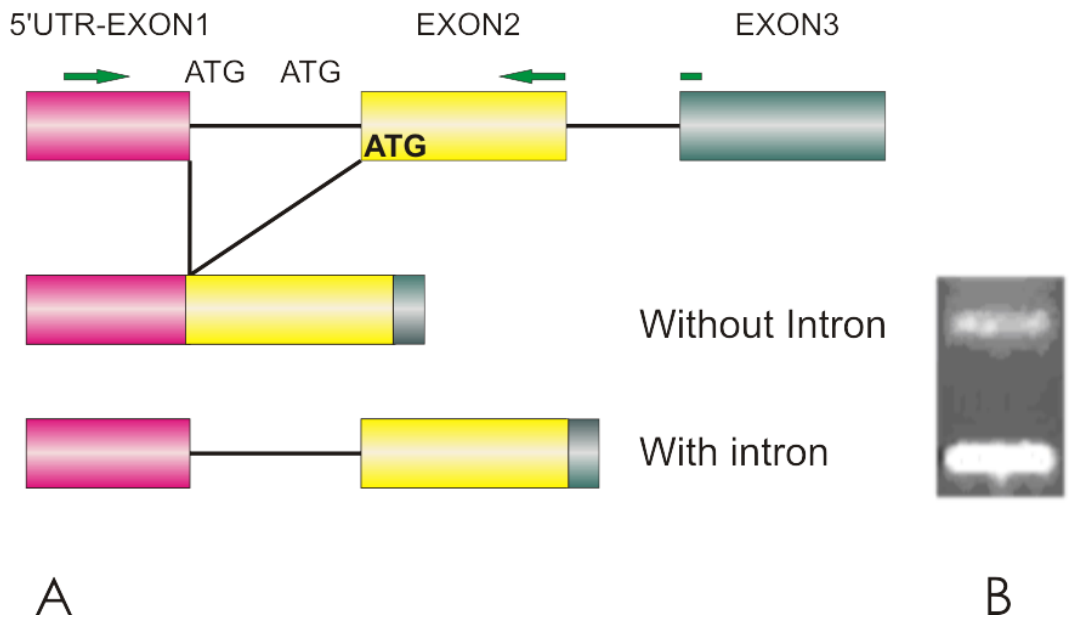


Figure 6.13: ***Scheme of the bZIP21 5' UTR and exons 1, 2 and 3.*** *(A) Possible transcripts resulting through differential splicing (the first without and the second with intron) are indicated. (B) Amplification products obtained by PCR using the primers indicated in (A) (green arrows), showing that two transcripts are generated through alternative splicing.*

though no subcelular localization experiment was conducted, bZIP21 is annotated to have both nuclear and chloroplastidic localization. Elucidation of the subcellular localization would provide means to link the function of bZIP21 to sugar/light control and to nucleus-chloroplast signaling.

Two heterozygous T-DNA insertion lines (SALK_057609.55.00.x and SALK_057570-.40.50.x, Figure 6.12) were analysed. Plants were grown in long-day condition in the greenhouse. The only visible phenotype identified in both mutant plants was a thicker flower stalk when compared to the wild-type. To identify homozygote lines, three specific primers (two annealing to the *bZIP21* gene and one to the T-DNA sequence) were used for a PCR with genomic DNA from the T-DNA lines. No homozygote line could be found among 60 plants tested. Lack of homozygote mutants for T-DNA insertion in an exon of *bZIP21* was also observed at Dr. Michel Vincentz' laboratory (UNICAMP, Brazil; personal communication). Whether lack of functional *bZIP21* is lethal is currently unknown. Nevertheless, no phenotype associated with lethality (embryoless seeds, aberrant seed form or lack of seeds) was observed in the siliques of heterozygous mutant plants.

Another T-DNA insertion line for *bZIP21* (SALK_141618.33.90.x) was analysed. This is a homozygote line for an insertion in the 6^{th} intron (Figure 6.12). In this line,

no phenotypic difference with respect to the leaf size and number, flowers, siliques and overall development was observed in comparison to the wild-type (data not shown).

Arabidopsis was transformed with a construct containing the CaMV 35S promoter fused to the *bZIP21* cDNA (35S::bZIP21). Ten independent transgenic lines were phenotypically analysed. To verify the effect of the constitutive overexpression of *bZIP21*, plants were grown at long-day condition in the greenhouse; *bZIP21* overexpressing plants (BOX21) reached 20% plant senescence (marked by the fact that 20% of the siliques in the plant are yellow) 4 to 7 days after the wild-type. Leaves had the same size as the ones from wild-type and produced the same number of siliques (Figure 6.14).

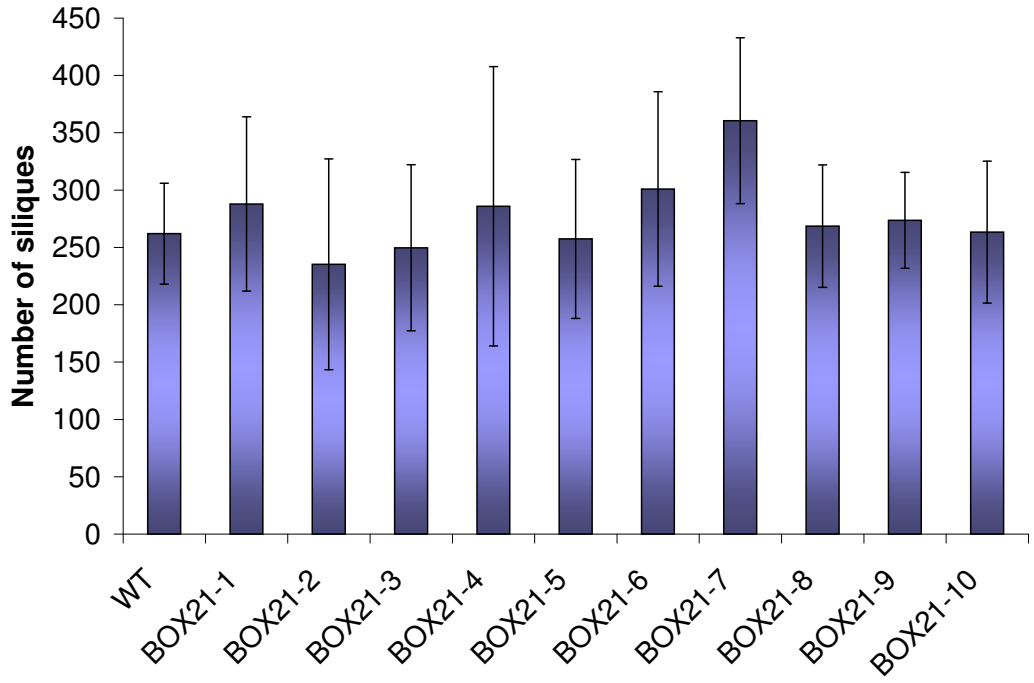


Figure 6.14: ***Number of siliques in BOX21 lines.*** *Total sillique number was determined at the stage when plants had 20% of their siliques yellow, around three months after sowing. No significant difference was observed between the mutants and the wild-type. Mutants from the T2 generation (10 plants per line) were used for this experiment. Error bars correspond to standard deviation.*

The expression level of bZIP21 was accessed in different mutants (i.e. *abi4*, *abi5*, *ein1-1*, *ein2-1*, *pwd* and *sex1*) via qRT-PCR. Almost no difference was observed in the expression of *bZIP21* in the mutants, except for *ein2-1*, where expression of *bZIP21* was 11-fold higher when compared to the wild-type (Figure 6.15). The ETHYLENE INSENSITIVE 2 protein is involved in ethylene signalling, and loss of function leads to insensitivity to ethylene, larger rosettes and delayed bolting (VANDENBUSSCHE *et al.* 2007).

Interestingly, BOX21 plants also show a tendency for delayed development, as observed in *ein2-1*. Moreover, *ein2-1* confers an enhanced response to ABA. Although the difference in expression observed is below 5-fold, there is a tendency that *bZIP21* is less expressed in the mutants insensitive to ABA (Figure 6.15).

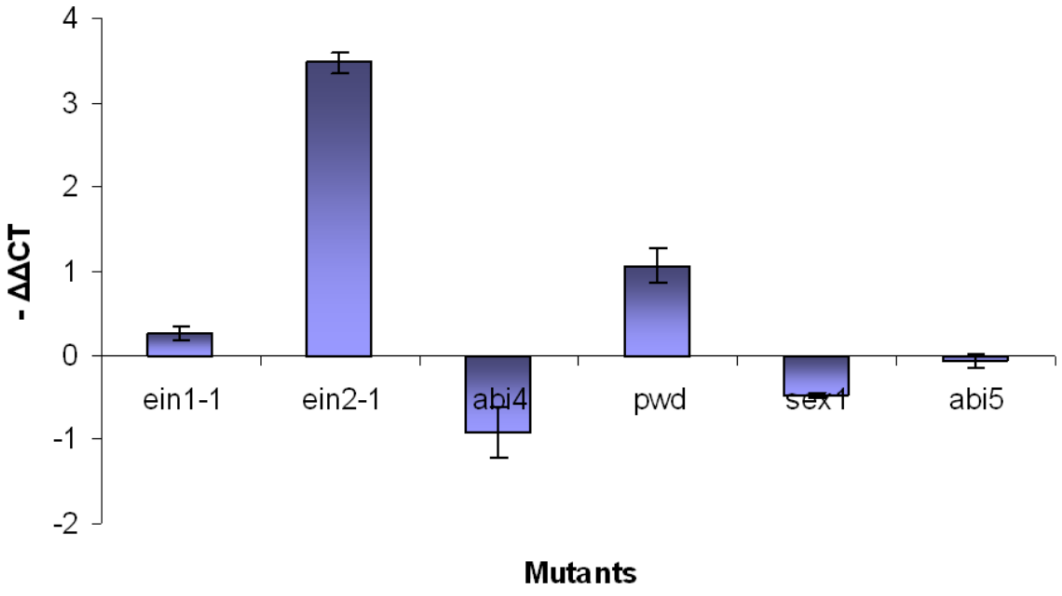


Figure 6.15: ***Expression levels of bZIP21 in different Arabidopsis mutants.*** *Hormone signalling (ein1-1, ein2-1, abi4 and abi5) and starch degradation mutants (pwd, sex1) and their respective wild-types were grown in long-day conditions in the greenhouse. Leaves (number 11 at 5 mm size) were collected at 9 am, and the RNA extracted from them was used to access the expression level of bZIP21. Values represent the difference in expression of bZIP21 in the wild-type and in the respective mutant. Expression correspond to the mean value of two biological replicates with three technical replicates and the error bars correspond to the standard deviation.*

Wounding strongly affects *bZIP21* expression. Its expression was found to increase 8-fold within five minutes after wounding (Figure 6.16), therefore it may be one of the first genes activated towards a wounding response. Therefore, *bZIP21* potentially plays a role similar to that of other members of Group D (see Chapter 3). After two hours *bZIP21* expression decreased. Reduction of the expression level of *bZIP21* may be required for induction of subsequent steps in the wounding response. Interestingly, the induction of bZIP21 by wounding is not consistent with data in Genevestigator. Apparently, this is an effect of the time points chosen for expression profiles publicly available, where the first time point usually is 30 minutes after the wounding (CHEONG *et al.* 2002, DELESSERT *et al.* 2004), when expression of *bZIP21* is almost the same as in the control plant (Figure 6.16).

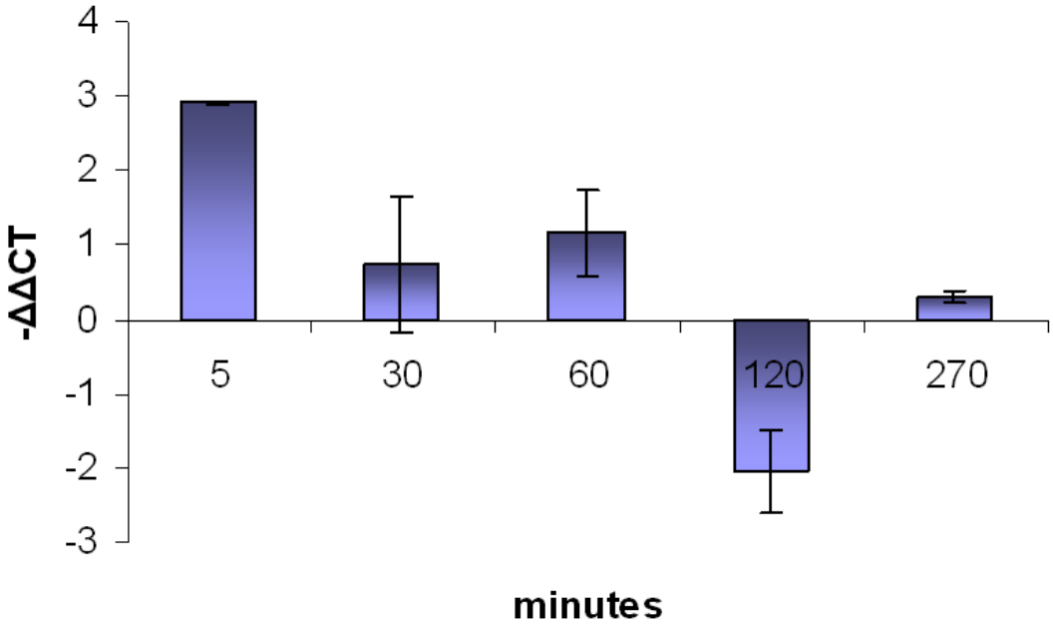


Figure 6.16: ***Expression levels of bZIP21 after wounding.*** *Arabidopsis leaves of 1 cm were injured and harvested 5, 30, 60, 120 and 270 min later. Expression correspond to the mean value of two biological replicates with three technical replicates and the error bars correspond to the standard deviation. Values in the graphics correspond to the average of bZIP21 expression in the treatment minus the average of bZIP21 expression in time 0 of wounding.*

6.4 Characterization of *Dof4*

6.4.1 Introduction

The origin of the green plant lineage, occurring after the original endosymbiosis of a cynobacterium into an eukaryotic cell (REYES-PRIETO *et al.* 2007), led the raise of a control network, that had not only to coordinate two different genomes, but also to deal with the acquisition of an autotrophic status. The DNA binding with one finger (Dof) transcription factor represents a particular class of zinc finger domain TFs that emerged in the early days of green plants (MORENO-RISUENO *et al.* 2007, RIAÑO PACHÓN *et al.* 2008). Functional characterization indicates that Dofs play a role in various processes, including stress responses (CHEN *et al.* 1996, KANG *et al.* 2003, ZHANG *et al.* 1995), light responses (PAPI *et al.* 2002, YANAGISAWA 2000, YANAGISAWA and SHEEN 1998), phytochrome signaling (PARK *et al.* 2003), responses to plant hormones (KISU *et al.* 1998, MENA *et al.* 2002, PAOLIS *et al.* 1996, WASHIO 2001), seed germination (GUALBERTI *et al.* 2002, PAPI *et al.* 2000) and cell cycle regulation (SKIRYCZ *et al.* 2008). Evolutionary analyses revealed that all Dofs identified in angiosperms evolved from a single gene,

represented by the only extant in Chlamydomonas, CrDof1 (MORENO-RISUENO *et al.* 2007). Dof TFs are classified into nine sub-families (A to G), and Dof4 belongs to sub-family D. From these analyses, it can be concluded that Dof4 possesses neither a close paralogue nor a direct orthologue in rice (LIJAVETZKY *et al.* 2003, MORENO-RISUENO *et al.* 2007). Here we describe a preliminary characterization of *Dof4*.

6.4.2 Results and Discussion

6.4.2.1 Identification of *Dof4*

The Arabidopsis genome codes for 36 unique Dof transcription factors. Evolutionary analyses including Dofs from the main green plant lineages allowed clustering these genes into nine sub-families (A to G; MORENO-RISUENO *et al.* 2007). Sub-family D is apparently the second cluster that emerged in the course of evolution, as it can be traced back to pteridophytes. The split between the most recent common ancestor (MRCA) from bryophytes and tracheophytes is marked by important features observed in all tracheophytes, i.e., the dominance of the sporophyte over the gametophyte, presence of real leaves and roots and a proper vascular tissue (thus the name tracheophytes or vascular plants). Therefore, from the evolutionary point of view, the characterization of Dofs from sub-family D could elucidate which genes or networks contributed in a significant way to the emergence of tracheophytes. Among all Dofs in sub-family D from Arabidopsis, Dof4 is the only one that does not posses any close paralogue (LIJAVETZKY *et al.* 2003), thus constituting a good candidate for reverse genetic approaches, as less redundancy with other genes is expected.

6.4.2.2 The expression pattern of *Dof4*

Plants carrying a construct with the promoter of *Dof4* (1.3 kb) fused to the reporter gene *GusPlus* (PDof4::Gus+) were selected on media containing hygromycin, and 10 lines, out of the 20 lines originally obtained, were selected for characterization of Dof4 expression pattern. We decided to analyse all 10 lines as the expression of *Dof4* seems to be very low, except for floral parts, where the expression is stronger. *Dof4* is expressed in very young leaves (Figure 6.17 A) and in the shoot apical meristem (Figure 6.17 B). As leaves grow, expression was restricted to vascular tissue (Figure 6.17 C) and trichomes (Figures 6.17 D to E). Expression in the hypocotyl is also strong, but it is restricted to the vascular tissue (Figures 6.17 F and G). In roots, expression is detected only in the zones where lateral roots emerge (Figures 6.17 H to J). In flowers, *Dof4* is expressed in the stigma and in the pollen before the flowers open (Figures 6.17 K to M). During embryogenesis, *Dof4* is expressed in very early stages of development untill the pre-globular stage (Figures 6.17

N and O), but no expression could be observed in the globular and the torpedo stages. Expression is observed again in fully developed embryos, in the tip of the cotyledon and in the radicle (Figure 6.17 P). A strong expression is also observed in the endosperm (Figure 6.17 O), which makes it difficult to visualize the embryo in the seed. This expression in endosperm is supported by the fact that recently *Dof4* has been found to be among 800 genes preferentially expressed in endosperm (DAY *et al.* 2008).

To uncover if the expression of *Dof4* is conserved in eudicots, tobacco plants were transformed with the PDof4::Gus+ construct. Twenty plants were recovered and used for characterization of the expression pattern. Expression is conserved in the flower, where stigma and anthers are stained (Figure 6.18 A). Expression in petals is only seen when flowers start to open and it is restricted to the apical inner part of petals (Figure 6.18 B). In leaves, a distinct expression pattern is observed, and staining is visible along the whole leaf blade and also in trichomes (Figure 6.18 C). In full-developed embryos, expression is restricted to the cotyledons (Figure 6.18 D), whereas in Arabidopsis embryos expression is also seen in the radicle (Figure 6.17 P).

6.4.2.3 *Dof4* plays a pleiotropic role in plant development

A reverse genetic approach was used to uncover the function of Dof4. One transgenic line carrying a T-DNA insertion in the *Dof4* ORF (GK-384G11) was identified and PCR analyses showed that it was a homozygous insertion line. Under long-day condition in the greenhouse, T-DNA plants show a faster development when compared to the wild-type (Figure 6.19 A), and senescence developed 10 days earlier than in the wild-type controls. This difference in development was observed early during plant growth and reached the 10-days difference when plants bolted. Bolting T-DNA plants had one or two leaves less than the wild-type, but leaves were approximately 10% bigger and had a higher chlorophyll content. Additionally, T-DNA plants were 23% bigger and posses more lateral branches than the wild-type, and the number of siliques was increased by approximately 30% (Figure 6.19 B and 6.20).

Plants overexpressing (35S::Dof4) or repressing (*Dof4*RNAi) *Dof4* were generated by (RUZICIC 2003) and three and two lines were further analysed, respectively. The phenotype observed in RNAi plants was similar to the one observed in the T-DNA insertion lines (Figures 6.19 B and D), although the difference in the number of leaves at the bolting age was not significant. On the other hand, plants overexpressing *Dof4* showed a delayed development (Figure 6.19 D), reaching 20% senescence 10 to 13 days after the wild-type. Plants produce smaller leaves and the number of leaves at bolting age increases tremendously, from 14 ± 2 in the wild-type to 65 ± 10 in the overexpression lines. Moreover, leaves contained less chlorophyll than the wild-type. No difference was observed neither

in the number of siliques nor in the height of the plants (Figures 6.21).

One important aspect observed in overexpression plants was the almost complete lack of apical dominance. In many plants, the main flower stalk did not develop or developed only after lateral stalks were formed. This effect was also visible in roots; they were small and the main root was not easily identifiable and acquired a fasciculated structure (data not shown).

6.4.2.4 *Dof4* and the heterosis effect

One well-known fact that influences biomass accumulation is heterosis. It is an old knowledge in agriculture that crossing two distinct phenotypes may result in a stronger progeny, a phenomenon called hybrid vigour. Nevertheless, crossings can result in negative effects also. Crosses made between the Arabidopsis ecotypes Col-0 and C24 resulted in plants that showed a strong positive heterosis effect (64% more biomass in comparison with the respective parental lines, Col-0 and C24), whereas crosses between the ecotypes Col-0 and Cvi led to negative heterosis (23% less biomass) (MEYER *et al.* 2004, SIMON *et al.* 2008; Rhonda Meyer, personal communication). As lack of Dof4 led to a higher number of siliques and bigger leaves (Figures 6.19 and 6.20), reflecting an increase of biomass accumulation, it is possible that *Dof4* plays a role in the heterosis effect.

In the Col-0 x Cvi cross, the transcript level of *Dof4* was found to be twice and four times as high as in the respective Col-0 and Cvi parental lines (Figures 6.22 A). The negative heterosis effect observed in this crossing is comparable to the phenotype observed in plants overexpressing *Dof4*. On the other hand, in plants obtained from a Col-0 x C24 cross *Dof4* was expressed at the same level as in the parental lines (Figure 6.22 B). To test the hypothesis that *Dof4* contributes to the heterosis effect, the *Dof4* T-DNA insertion line (KO) in the Col-0 background was crossed with RNAi lines inhibited for *Dof4* expression (C24 background), as well as with C24 and Cvi wild-type plants. Additionally, *Dof4* RNAi plants were crossed with KO and Col-0 plants. Crosses between Col-0 and Col-0, as well as between C24 and C24, were also generated as control for the crossing effect. A summary of the crosses performed can be seen in Table refTBL:Crosses. Crossed plants (KO x C24, RNAi x Col-0, KO x RNAi, and KO x Cvi) and respective control crosses and wild-type plants were analysed with respect to leaf size at bolding age and at the beginning of senescence. Plants originating from crosses between C24 and Col-0 have bigger leaves than the control parent plants, as previously observed (MEYER *et al.* 2004). Interestingly, leaves from plants originating from crosses involving KO or RNAi lines were even larger than in the wild-type crosses; therefore we can conclude that low expression level of *Dof4* has a positive effect on heterosis. Plants originating from the KO x Cvi cross had a less severe phenotype than of the Col-0 x Cvi crossing; thus expression

	Col-0	C24	Cvi	KO	RNAi
Col-0	X	X	X		X
C24	X	X		X	
Cvi	X	X	X	X	
KO	X	X	X		X
RNAi	X			X	

Table 6.1: *Crosses analysed in the heterosis experiment. Rows represent male plants and columns female plants.*

of both *Dof4* alleles may be required to full functionality of this gene. This effect was also observed in heterozygote *Dof4* T-DNA insertion lines; they showed a phenotype that was intermediate between the wild-type and the homozygous mutant.

6.4.2.5 The role of *Dof4* for the accumulation of biomass

Increasing in biomass accumulation is one of the biggest challenges in current plant sciences, as demand for food is increasing and the use of biomass has been seen as an alternative source of energy to the non-renewable sources used nowadays. Alterations of leaf size normally lead to alterations in plant biomass accumulation. Therefore, we are interested in understanding the process of biomass accumulation. As *Dof4* T-DNA insertion (KO) plants are bigger than wild-type plants, showing 30% higher number of siliques than the wild-type (Figure 6.20), they were used as a background for transformation with other genes known to have a positive effect on biomass accumulation, in order to uncover the possibility of an additive effect on the phenotype of the obtained supermutant.

We have screened the literature for genes that have already been reported to act on leaf size. A short-list of genes was chosen to test whether a positive interaction is possible. The following four genes were chosen for the first experiments:

Exordium (*EXO*, At4g08950), a phosphate-induced gene that regulates brassinosteroid–responsive genes, whose overexpression increased the vegetative growth of plants (COLL-GARCIA *et al.* 2004);

gibberellin (GA) oxidases *GA2-ox* (At1g78440) and *GA20-ox* (At4g25420), that when overexpressed, lead to a decrease and an increase of plant biomass, respectively (BIEMELT *et al.* 2004);

CrCIN, a cell wall invertase from carrot (Daucus carota) that, when expressed specifically in roots, leads to an increased leaf size (VON SCHWEINICHEN and BÜTTNER 2005).

The four genes were used to super-transform the *dof4* KO plants. Around 25 lines were recovered from each transformation and biomass formation was analysed. Phenotyping showed that KO:GA20-ox plants have an earlier development when compared to both controls (KO and Col-0, (Figure 6.23 A), *KO:CrCIN*, *KO:GA20-ox* and *KO:EXO* plants are higher than the controls, clearly indicating that these genes have synergetic effects with respect to biomass formation (Figures 6.23 B and 6.24). Interestingly, though height was increased, *KO:EXO* plants showed leaves that are smaller than the KO parental plant, but of the same size as the wild-type (Figure 6.23 B). Most probably, Dof4, CrCIN and GA20-ox act in different pathways that all lead to bigger plants, whereas EXO may overlap in some part of the network with Dof4, resulting in smaller leaves. The next step would be to obtain the triple mutant *KO4:CrCIN:GA20-ox* to test whether plant biomass accumulation can be further enhanced. On the other hand, KO4:GA2-ox lines were smaller than the controls. However, the decrease in size appeared to be less dramatic than reported in the literature (BIEMELT *et al.* 2004).

6.5 References

ABE, H., T. URAO, T. ITO, M. SEKI, K. SHINOZAKI, *et al.*, 2003 Arabidopsis atmyc2 (bhlh) and atmyb2 (myb) function as transcriptional activators in abscisic acid signaling. Plant Cell **15**: 63–78.

ABE, M., Y. KOBAYASHI, S. YAMAMOTO, Y. DAIMON, A. YAMAGUCHI, *et al.*, 2005 FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. Science **309**: 1052–1056.

BALAZADEH, S., D. M. R. NO PACHÓN, and B. MUELLER-ROEBER, 2008 Transcription factors regulating leaf senescence in *Arabidopsis thaliana*. Plant Biol (Stuttg) **10 Suppl 1**: 63–75.

BIEMELT, S., H. TSCHIERSCH, and U. SONNEWALD, 2004 Impact of altered gibberellin metabolism on biomass accumulation, lignin biosynthesis, and photosynthesis in transgenic tobacco plants. Plant Physiol **135**: 254–265.

BLÄSING, O. E., Y. GIBON, M. GÜNTHER, M. HÖHNE, R. MORCUENDE, *et al.*, 2005 Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in Arabidopsis. Plant Cell **17**: 3257–3281.

BROOTHAERTS, W., H. J. MITCHELL, B. WEIR, S. KAINES, L. M. A. SMITH, *et al.*, 2005 Gene transfer to plants by diverse species of bacteria. Nature **433**: 629–633.

CHEN, W., G. CHAO, and K. B. SINGH, 1996 The promoter of a H_2O_2-inducible, Arabidopsis glutathione S-transferase gene contains closely linked OBF- and OBP1-binding sites. Plant J **10**: 955–966.

CHEONG, Y. H., H.-S. CHANG, R. GUPTA, X. WANG, T. ZHU, *et al.*, 2002 Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in Arabidopsis. Plant Physiol **129**: 661–677.

CHINNUSAMY, V., M. OHTA, S. KANRAR, B.-H. LEE, X. HONG, *et al.*, 2003 ICE1: a regulator of cold-induced transcriptome and freezing tolerance in Arabidopsis. Genes Dev **17**: 1043–1054.

CHUANG, C. F., M. P. RUNNING, R. W. WILLIAMS, and E. M. MEYEROWITZ, 1999 The PERIANTHIA

gene encodes a bZIP protein involved in the determination of floral organ number in Arabidopsis thaliana. Genes Dev **13**: 334–344.

COLL-GARCIA, D., J. MAZUCH, T. ALTMANN, and C. MÜSSIG, 2004 EXORDIUM regulates brassinosteroid-responsive genes. FEBS Lett **563**: 82–86.

CORRÊA, L. G. G., D. M. R. NO PACHÓN, C. G. SCHRAGO, R. V. DOS SANTOS, B. MUELLER-ROEBER, *et al.*, 2008 The role of bZIP transcription factors in green plant evolution: adaptive features emerging from four founder genes. PLoS ONE **3**: e2944.

DAY, R. C., R. P. HERRIDGE, B. A. AMBROSE, and R. C. MACKNIGHT, 2008 Transcriptome analysis of proliferating arabidopsis endosperm reveals biological implications for the control of syncytial division, cytokinin signalling, and gene expression regulation. Plant Physiol .

DEKKERS, B. J. W., J. A. M. J. SCHUURMANS, and S. C. M. SMEEKENS, 2008 Interaction between sugar and abscisic acid signalling during early seedling development in Arabidopsis. Plant Mol Biol **67**: 151–167.

DELESSERT, C., I. W. WILSON, D. V. D. STRAETEN, E. S. DENNIS, and R. DOLFERUS, 2004 Spatial and temporal analysis of the local response to wounding in Arabidopsis leaves. Plant Mol Biol **55**: 165–181.

EHLERT, A., F. WELTMEIER, X. WANG, C. S. MAYER, S. SMEEKENS, *et al.*, 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.

FAIRCHILD, C. D., M. A. SCHUMAKER, and P. H. QUAIL, 2000 HFR1 encodes an atypical bHLH protein that acts in phytochrome A signal transduction. Genes Dev **14**: 2377–2391.

FETTKE, J., N. ECKERMANN, O. KÖTTING, G. RITTE, and M. STEUP, 2006 Novel starch-related enzymes and carbohydrates. Cell Mol Biol (Noisy-le-grand) **52 Suppl**: OL883–OL904.

FRIEDRICHSEN, D. M., J. NEMHAUSER, T. MURAMITSU, J. N. MALOOF, J. ALONSO, *et al.*, 2002 Three redundant brassinosteroid early response genes encode putative bHLH transcription factors required for normal growth. Genetics **162**: 1445–1456.

GUALBERTI, G., M. PAPI, L. BELLUCCI, I. RICCI, D. BOUCHEZ, *et al.*, 2002 Mutations in the Dof zinc finger genes DAG2 and DAG1 influence with opposite effects the germination of Arabidopsis seeds. Plant Cell **14**: 1253–1263.

HANAHAN, D., 1985 Techniques for transformation of *Escherichia coli*. In D. Glover, editor, *DNA Cloning, Volume I, a practical approach*. IRL Press, Oxford, 109–135.

HARPER, L., and M. FREELING, 1996 Interactions of liguleless1 and liguleless2 function during ligule induction in maize. Genetics **144**: 1871–1882.

HELLENS, R. P., E. A. EDWARDS, N. R. LEYLAND, S. BEAN, and P. MULLINEAUX, 2000 pGreen: a versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant transformation. Plant Mol. Bio. **42**: 819–832.

HEPWORTH, S. R., Y. ZHANG, S. MCKIM, X. LI, and G. W. HAUGHN, 2005 BLADE-ON-PETIOLE-dependent signaling controls leaf and floral patterning in Arabidopsis. Plant Cell **17**: 1434–1448.

HUIJSER, C., A. KORTSTEE, J. PEGO, P. WEISBEEK, E. WISMAN, *et al.*, 2000 The Arabidopsis *SUCROSE UNCOUPLED*-6 gene is identical to *ABSCISIC ACID INSENSITIVE-4*: involvement of abscisic acid in sugar responses. Plant J **23**: 577–585.

HUQ, E., and P. H. QUAIL, 2002 PIF4, a phytochrome-interacting bHLH factor, functions as a negative regulator of phytochrome B signaling in Arabidopsis. EMBO J **21**: 2441–2450.

JAKOBY, M., B. WEISSHAAR, W. DROEGE-LASER, J. VICENTE-CARBAJOSA, J. TIEDEMANN, *et al.*, 2002 bZIP transcription factors in Arabidopsis. Trends in Plant Science **7**: 106–111.

JIAO, Y., O. S. LAU, and X. W. DENG, 2007 Light-regulated transcriptional networks in higher plants. Nat Rev Genet **8**: 217–230.

KANG, H.-G., R. C. FOLEY, L. O. NATE SÁNCHEZ, C. LIN, and K. B. SINGH, 2003 Target genes for OBP3, a Dof transcription factor, include novel basic helix-loop-helix domain proteins inducible by salicylic acid. Plant J **35**: 362–372.

KESARWANI, M., J. YOO, and X. DONG, 2007 Genetic interactions of TGA transcription factors in the regulation of pathogenesis-related genes and disease resistance in Arabidopsis. Plant Physiol **144**: 336–346.

KISU, Y., T. ONO, N. SHIMOFURUTANI, M. SUZUKI, and M. ESAKA, 1998 Characterization and expression of a new class of zinc finger protein that binds to silencer region of ascorbate oxidase gene. Plant Cell Physiol **39**: 1054–1064.

KOCH, K. E., 1996 Carbohydrate-modulated gene expression in plants. Annu Rev Plant Physiol Plant Mol Biol **47**: 509–540.

KONCZ, C., and J. SCHELL, 1986 The promoter of TL-DNA gene 5 controls the tissue specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. Mol. Gen. Genet. **204**: 383–396.

LEE, J.-Y., J. COLINAS, J. Y. WANG, D. MACE, U. OHLER, *et al.*, 2006 Transcriptional and posttranscriptional regulation of transcription factor expression in Arabidopsis roots. Proc Natl Acad Sci U S A **103**: 6055–6060.

LIJAVETZKY, D., P. CARBONERO, and J. VICENTE-CARBAJOSA, 2003 Genome-wide comparative phylogenetic analysis of the rice and Arabidopsis Dof gene families. BMC Evol Biol **3**: 17.

MADDALONI, M., N. D. FONZO, H. HARTINGS, N. LAZZARONI, F. SALAMINI, *et al.*, 1989 The sequence of the zein regulatory gene *opaque-2* (*O2*) of *Zea mays*. Nucleic Acids Res **17**: 7532.

MENA, M., F. J. CEJUDO, I. ISABEL-LAMONEDA, and P. CARBONERO, 2002 A role for the DOF transcription factor BPBF in the regulation of gibberellin-responsive genes in barley aleurone. Plant Physiol **130**: 111–119.

MEYER, R. C., O. TÖRJÉK, M. BECHER, and T. ALTMANN, 2004 Heterosis of biomass production in Arabidopsis. Establishment during early development. Plant Physiol **134**: 1813–1823.

MORENO-RISUENO, M. A., M. MARTÍNEZ, J. VICENTE-CARBAJOSA, and P. CARBONERO, 2007 The family of DOF transcription factors: from green unicellular algae to vascular plants. Mol Genet Genomics **277**: 379–390.

MURASHIGE, T., and F. SKOOG, 1962 A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol Plant **3**: 473–497.

NESI, N., I. DEBEAUJON, C. JOND, G. PELLETIER, M. CABOCHE, *et al.*, 2000 The TT8 gene encodes a basic helix-loop-helix domain protein required for expression of DFR and BAN genes in Arabidopsis siliques. Plant Cell **12**: 1863–1878.

NI, M., J. M. TEPPERMAN, and P. H. QUAIL, 1998 PIF3, a phytochrome-interacting factor necessary for

normal photoinduced signal transduction, is a novel basic helix-loop-helix protein. Cell **95**: 657–667.

OSUNA, D., B. USADEL, R. MORCUENDE, Y. GIBON, O. E. BLÄSING, *et al.*, 2007 Temporal responses of transcripts, enzyme activities and metabolites after adding sucrose to carbon-deprived arabidopsis seedlings. Plant J **49**: 463–491.

PAOLIS, A. D., S. SABATINI, L. D. PASCALIS, P. COSTANTINO, and I. CAPONE, 1996 A rolB regulatory factor belongs to a new class of single zinc finger plant proteins. Plant J **10**: 215–223.

PAPI, M., S. SABATINI, M. M. ALTAMURA, L. HENNIG, E. SCHÄFER, *et al.*, 2002 Inactivation of the phloem-specific Dof zinc finger gene DAG1 affects response to light and integrity of the testa of Arabidopsis seeds. Plant Physiol **128**: 411–417.

PAPI, M., S. SABATINI, D. BOUCHEZ, C. CAMILLERI, P. COSTANTINO, *et al.*, 2000 Identification and disruption of an Arabidopsis zinc finger gene controlling seed germination. Genes Dev **14**: 28–33.

PARK, D. H., P. O. LIM, J. S. KIM, D. S. CHO, S. H. HONG, *et al.*, 2003 The Arabidopsis COG1 gene encodes a Dof domain transcription factor and negatively regulates phytochrome signaling. Plant J **34**: 161–171.

PAYNE, C. T., F. ZHANG, and A. M. LLOYD, 2000 *GL3* encodes a bHLH protein that regulates trichome development in arabidopsis through interaction with GL1 and TTG1. Genetics **156**: 1349–1362.

RAJANI, S., and V. SUNDARESAN, 2001 The Arabidopsis myc/bHLH gene ALCATRAZ enables cell separation in fruit dehiscence. Curr Biol **11**: 1914–1922.

REYES-PRIETO, A., A. P. M. WEBER, and D. BHATTACHARYA, 2007 The origin and establishment of the plastid in algae and plants. Annu Rev Genet **41**: 147–168.

RIAÑO PACHÓN, D. M., L. G. G. CORRÊA, R. TREJOS-ESPINOSA, and B. MUELLER-ROEBER, 2008 Green transcription factors: a chlamydomonas overview. Genetics **179**: 31–39.

RUZICIC, S., 2003 *Isolation and characterization of novel DOF genes from Arabidopsis thaliana (L.) Heynh.*. Ph.D. thesis, University of Potsdam.

SAMBROOK, J., and D. RUSSELL, 2001 *Molecular cloning: A Laboratory Manual*. Cold Sprng Harbor Laboratory, Cold Spring Harbor, New York.

SCHMID, K. J., O. TÖRJÉK, R. MEYER, H. SCHMUTHS, M. H. HOFFMANN, *et al.*, 2006 Evidence for a large-scale population structure of *Arabidopsis thaliana* from genome–wide single nucleotide polymorphism markers. Theor Appl Genet **112**: 1104–1114.

SIMON, M., O. LOUDET, S. DURAND, A. BÉRARD, D. BRUNEL, *et al.*, 2008 Quantitative trait loci mapping in five new large recombinant inbred line populations of *Arabidopsis thaliana* genotyped with consensus single-nucleotide polymorphism markers. Genetics **178**: 2253–2264.

SKIRYCZ, A., A. RADZIEJWOSKI, W. BUSCH, M. A. HANNAH, J. CZESZEJKO, *et al.*, 2008 The DOF transcription factor OBP1 is involved in cell cycle regulation in *Arabidopsis thaliana*. Plant J **56**: 779–792.

SMOLEN, G. A., L. PAWLOWSKI, S. E. WILENSKY, and J. BENDER, 2002 Dominant alleles of the basic helix-loop-helix transcription factor ATR2 activate stress-responsive genes in Arabidopsis. Genetics **161**: 1235–1246.

SORENSEN, A.-M., S. KRÖBER, U. S. UNTE, P. HUIJSER, K. DEKKER, *et al.*, 2003 The arabidopsis ABORTED MICROSPORES (AMS) gene encodes a MYC class transcription factor. Plant J **33**: 413–423.

TIAN, C., E. CHIKAYAMA, Y. TSUBOI, T. KUROMORI, K. SHINOZAKI, *et al.*, 2007 Top-down phenomics of *Arabidopsis thaliana*: metabolic profiling by one- and two-dimensional nuclear magnetic resonance spectroscopy and transcriptome analysis of albino mutants. J Biol Chem **282**: 18532–18541.

VANDENBUSSCHE, F., Y. HABRICOT, A. S. CONDIFF, R. MALDINEY, D. V. DER STRAETEN, *et al.*, 2007 HY5 is a point of convergence between cryptochrome and cytokinin signalling pathways in Arabidopsis thaliana. Plant J **49**: 428–441.

VERVLIET, G., M. HOLSTERS, H. TEUCHY, M. V. MONTAGU, and J. SCHELL, 1975 Characterization of different plaque-forming and defective temperate phages in Agrobacterium. J Gen Virol **26**: 33–48.

VETTORE, A. L., J. A. YUNES, G. C. NETO, M. J. DA SILVA, P. ARRUDA, *et al.*, 1998 The molecular and functional characterization of an Opaque2 homologue gene from Coix and a new classification of plant bZIP proteins. Plant Mol Biol **36**: 249–263.

VON SCHWEINICHEN, C., and M. BÜTTNER, 2005 Expression of a plant cell wall invertase in roots of Arabidopsis leads to early flowering and an increase in whole plant biomass. Plant Biol (Stuttg) **7**: 469–475.

WASHIO, K., 2001 Identification of Dof proteins with implication in the gibberellin-regulated expression of a peptidase gene following the germination of rice grains. Biochim Biophys Acta **1520**: 54–62.

WELTMEIER, F., A. EHLERT, C. S. MAYER, K. DIETRICH, X. WANG, *et al.*, 2006 Combinatorial control of Arabidopsis proline dehydrogenase transcription by specific heterodimerisation of bZIP transcription factors. EMBO J **25**: 3133–3143.

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**: 1717–1729.

YANAGISAWA, S., 2000 Dof1 and Dof2 transcription factors are associated with expression of multiple genes involved in carbon metabolism in maize. Plant J **21**: 281–288.

YANAGISAWA, S., and J. SHEEN, 1998 Involvement of maize Dof zinc finger proteins in tissue-specific and light-regulated gene expression. Plant Cell **10**: 75–89.

YU, T. S., H. KOFLER, R. E. HÄUSLER, D. HILLE, U. I. FLÜGGE, *et al.*, 2001 The Arabidopsis *sex1* mutant is defective in the R1 protein, a general regulator of starch degradation in plants, and not in the chloroplast hexose transporter. Plant Cell **13**: 1907–1918.

ZHANG, B., W. CHEN, R. C. FOLEY, M. BÜTTNER, and K. B. SINGH, 1995 Interactions between distinct types of dna binding proteins enhance binding to ocs element promoter sequences. Plant Cell **7**: 2241–2252.

Figure 6.17: ***Expression pattern of Dof4.*** *(A) Dof4 is expressed in early developing leaves and (B) in the shoot apical meristem (indicated by arrow). (C) As leaves grow, expression seems to be restricted to vascular tissue and (D) trichomes. (E) Fully developed leaves also conserve this expression pattern associated with vascular tissue and trichomes. (F) Strong expression is observed in the hypocotyl, mainly in the vascular tissue, but (G) no expression is seen in roots at the root-hypocotyl transition region. (H) Dof4 is not expressed in the root, except (I) for regions where lateral roots form or (J) at the emerging lateral root. (K) In closed flower buds, Dof4 is expressed in pollen and stigma, this expression persists until late flower stages (L), though, in some cases, very low expression was also observed in petals. (M) The expression of Dof4 is very prominent in the stigma. (N) Staining can be observed in embryos in early phases of development until (O) the pre-globular stage. Expression is also observed in the endosperm, which is accordance to the data from* DAY et al. *(2008). (P) Staining was also observed in fully developed embryos, especially in the tip of cotyledons and in the radicle.*

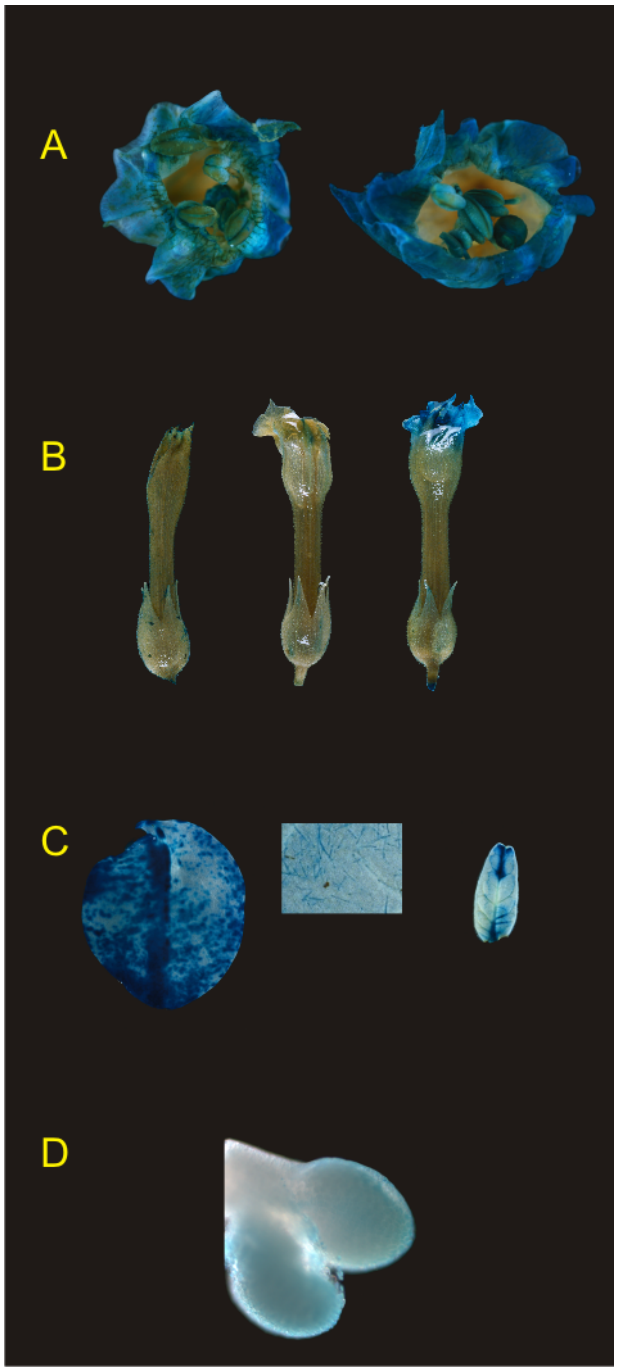


Figure 6.18: ***Expression pattern of Dof4 in tobacco.*** *(A) Dof4 is expressed in the stigma and anthers. (B) GUS staining is visible in the petals, but restricted to the apical part, and to later phases of flower development. (C) In leaves, Dof4 does not seems to have a clear pattern of expression, though it is possible to identify staining in the vascular tissue and in trichomes. (D) In fully developed embryos, expression is restricted to the cotyledons, not being observed in the radicle.*

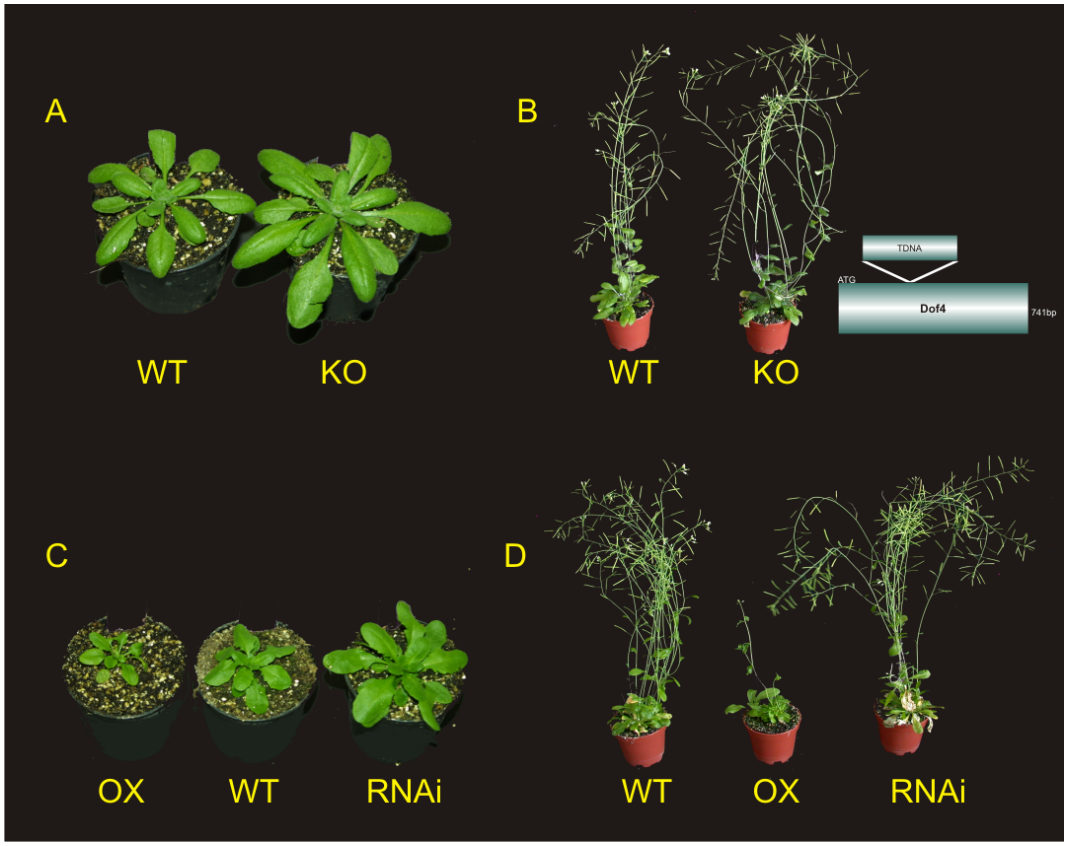


Figure 6.19: ***Phenotyping of Dof4 mutants.*** *(A) Plants carrying a T-DNA insertion for Dof4 (KO) have an early development and produce bigger leaves when compared to the wild-type (WT). (B) This difference persists until the end of development, when KO plants are also bigger than the WT and produce more siliques. (C) RNAi plants show a phenotype similar to that observed for the KO plants. On the other hand, overexpression lines (OX) show a slower development, with reduced leaf size. (D) These differences are observed until the end of development. Plants correspond to 6 (A and C) and 10 weeks (B and D) after sowing.*

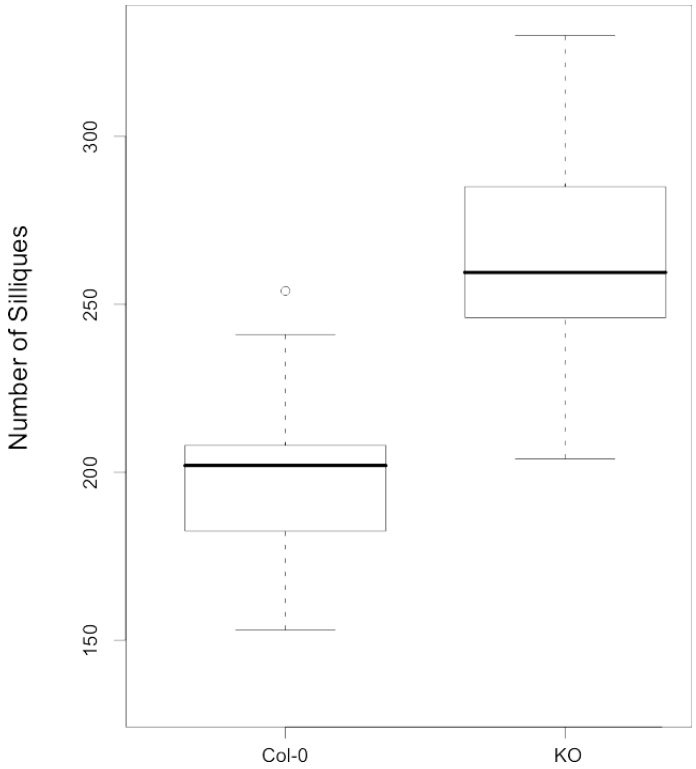


Figure 6.20: ***Number of siliques in the Dof4 T-DNA mutant.*** *In the T-DNA insertion plants (KO), the total number of siliques was 30% higher than in the wild-type control.*

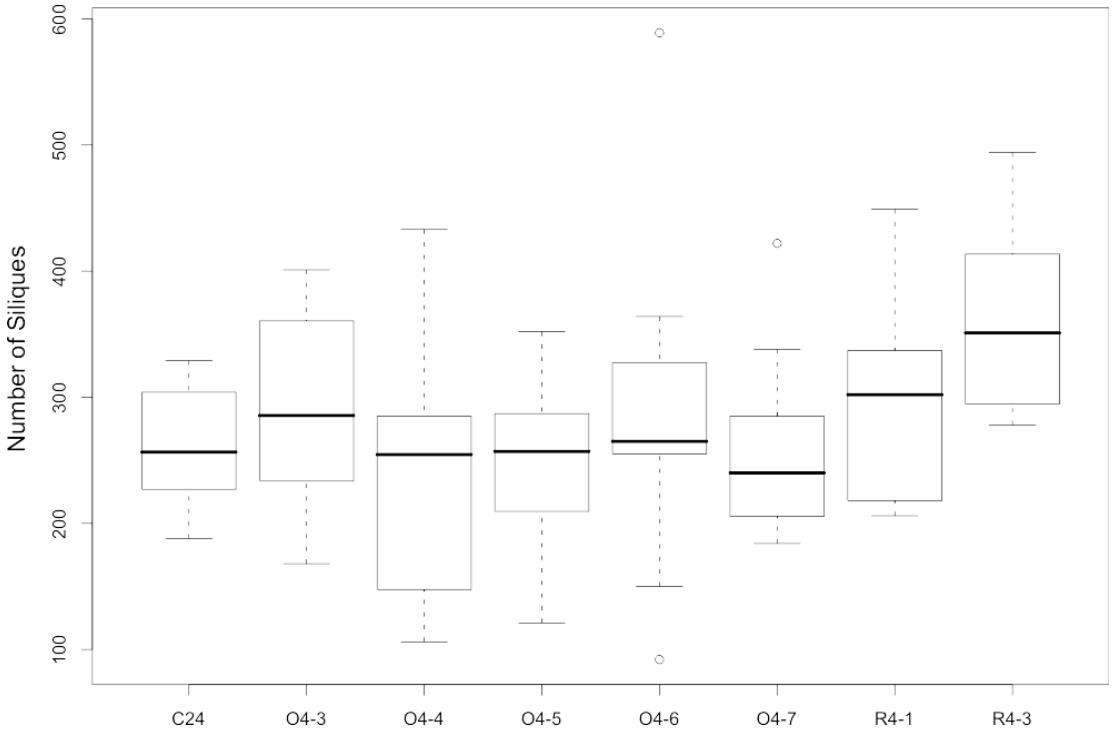


Figure 6.21: ***Number of siliques in Dof4 RNAi and OX mutants.*** *In the RNAi plants (R4-1 and R4-3), the total number of siliques tended to be higher than in the wild-type control (C24), whereas no difference was observed in the overexpression lines (O4-3 to O4-7). Three lines (O4-3, 4 and 6) were further used for the phenotype analysis. Levels of Dof4 expression were analyzed by* RUZICIC *(2003).*

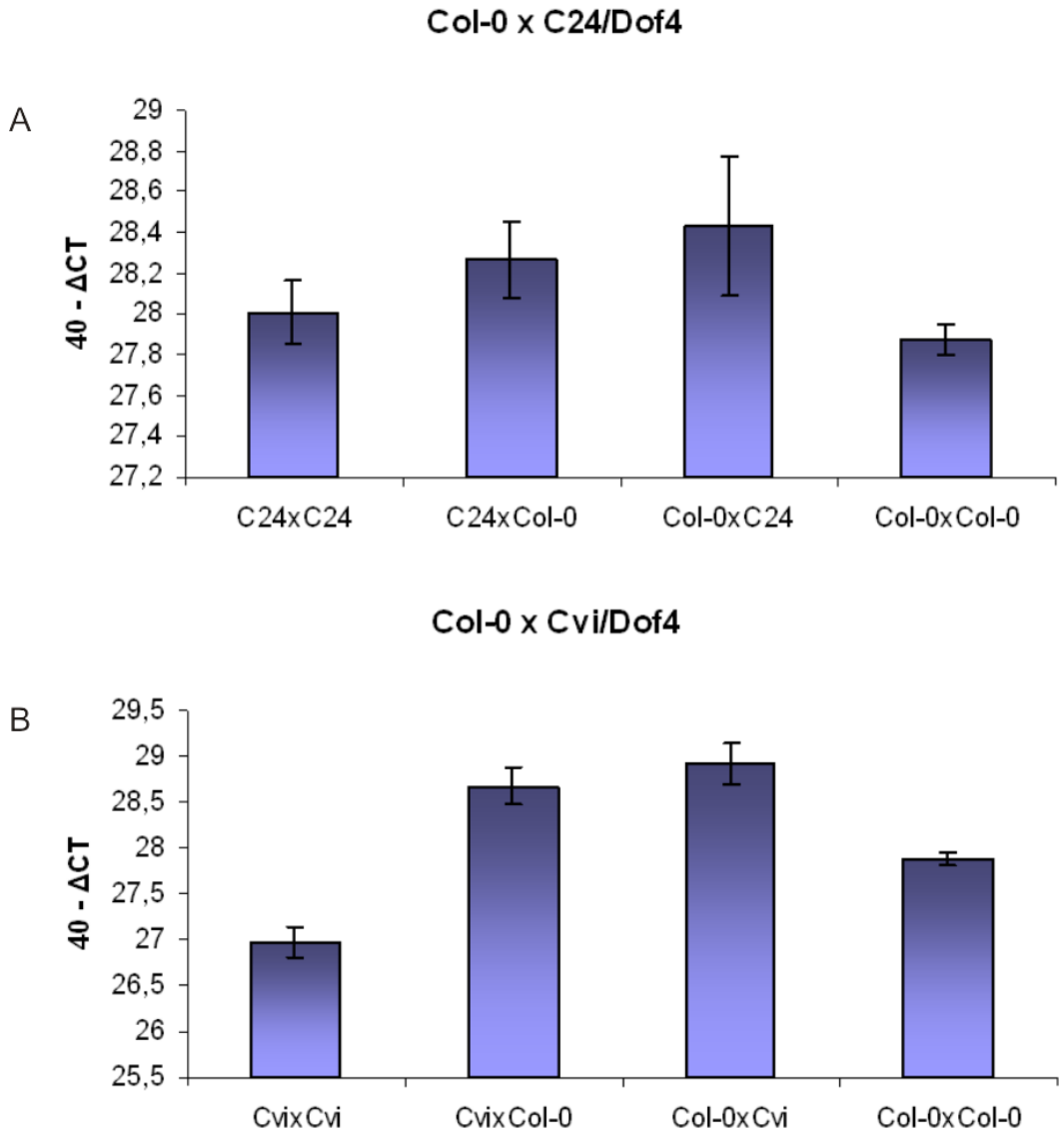


Figure 6.22: ***Expression of Dof4 in Col-0, C24 and Cvi crossings.*** *(A) Col-0 x C24 cross leads to a positive heterosis effect. Expression of Dof4 remained largely unaltered between the heterotic plants and the control crossings. (B) The crossing of Col-0 with Cvi leads to a negative heterosis effect, which is similar to the phenotype observed for the Dof4 overexpression lines. Heterotic plants have a higher expression of Dof4 when compared to control crossings (approximately 2-fold higher).*

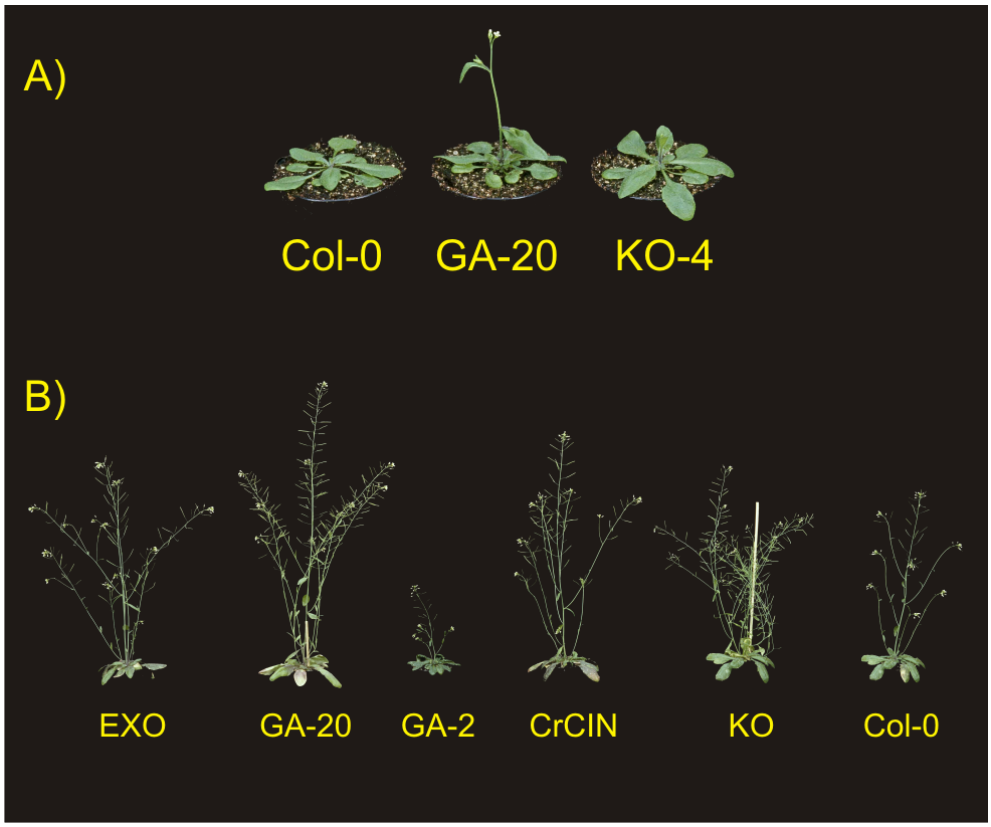


Figure 6.23: ***Phenotype of the supermutants generated from the T-DNA (KO) insertion lines.*** *(A) Differences of development are observed from early stages onward, specially in the KO:GA20-ox lines, where bolting takes place even before the KO lines. (B) EXO, GA20 and CrCIN had a positive effect on plant growth, though the increase in height observed in EXO is followed by a reduction of leaf size when compared to the dof4 KO line, but leaves had the same size as in the wild-type. GA2 plants are smaller than both controls, as expected, as this gene has a negative effect on plant growth.*

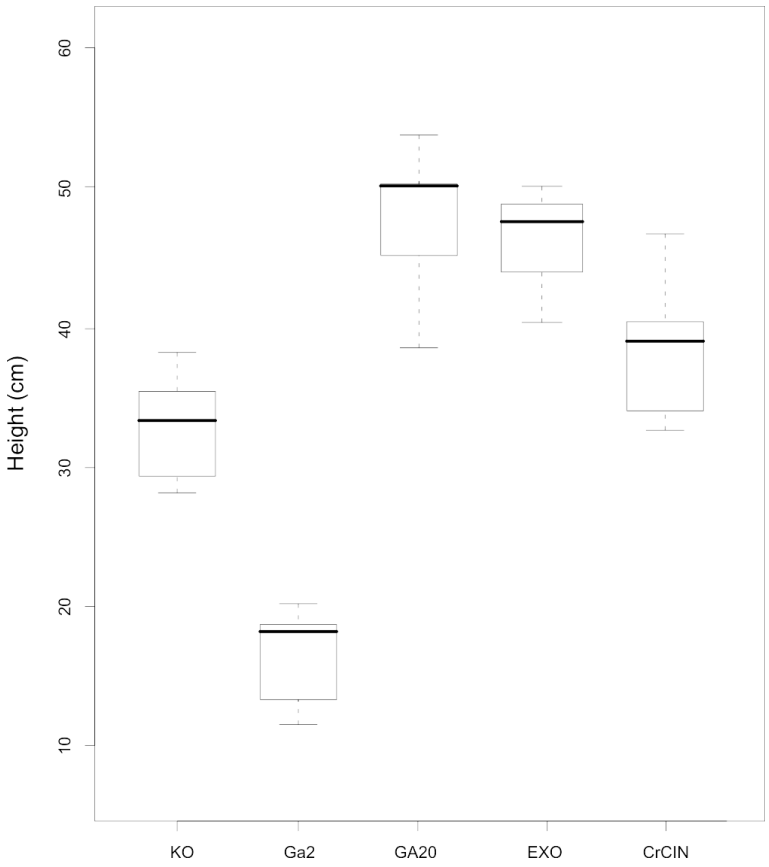


Figure 6.24: *Height distribution of the Dof4 T-DNA supermutants. Dof4 KO plants overexpressing GA20-ox, EXO and CrCIN showed an increase in shoot height when compared to the KO control plants, whereas plants overexpressing GA2 were smaller than the control.*

7

General Conclusion

The availability of completely sequenced and annotated genomes in public resources has greatly contributed to the comprehension of gene regulatory mechanisms, especially because the acquisition of new developmental traits can be interpreted as the result of modifications of pre-existing networks. Moreover, a more detailed understanding of the evolution of a given regulatory network may provide a basis for a more direct - and thus more efficient - approach in reverse genetic studies. Regulatory elements or networks conserved among all green plants or among eukaryotes may allow identifying the original set of genes/functions that generated the organism diversity observed nowadays. The identification of four founder genes for all bZIPs in green plants allowed tracing the putative function played by these genes in the most recent common ancestor of all green plants. Additionally, it has also been observed that regulatory pathways that were originally believed to be restricted to plants, such as for example the Cop1-mediated degradation of the bZIP transcription factor HY5, is also present in other eukaryotes, where Cop1-like proteins mediate the degradation of bZIPs. Therefore, one may speculate that the degradation of bZIP mediated by Cop-like proteins is a phenomenon shared by all eukaryotes, whereas the use of this regulatory system in the context of a light response network potentially evolved specifically in plants. The original function of this regulatory function is still unknown. The use of phylogenetic approaches for comparative studies at the gene and genome level is proving to be a good means for the optimization of gene characterization work and in the identification of genes controlling a given process, at least when a member of a Possible Groups of Orthologues (PoGO) representing an ancestral biological function in known. Therefore, the functional characterization of an individual member of a PoGO might be transferred to extant members of the same PoGO in other species. One good example where knowledge from an evolutionary background helped to assign functions to genes is provided by the NAC transcription factor genes. We have observed that NAC genes are overrepresented among the genes that are differentially ex-

pressed at the onset of leaf senescence in Arabidopsis. Phylogenetic analyses lead to the conclusion that most of the NAC genes differentially expressed during leaf senescence in Arabidopsis and rice mainly belong to three big groups of homologues. In a screen for senescence regulated TF genes in barley we found that eight, out of ten orthologous genes tested, showed a differential expression during late leaf development, indicating that the use of evolutionary tools is a powerful means to predict gene function. Leaf is the main photosynthetic organ in vascular plants; therefore a proper leaf development assures the plant with a source of photoassimilates required for the life of a plant. Transcription factors are important players in leaf development, as they are master controllers of gene expression. Identifying the transcription factors involved in leaf development may allow altering specific aspects of the plant such as for example leaf size or the overall time of development. The sink-to-source transition is one of the main transitions occurring during leaf lifespan. The identification of molecular and physiological markers that indicate the transition from a sink to a source leaf constitutes a key element needed for a better characterization of this transition. Sink-to-source transition in leaves lacks a visual phenotype, in contrast to the mature-to-senescent transition. Additionally, the characterization of the sink-to-source transition in Arabidopsis opens a new horizon in the field because for Arabidopsis, in contrast to tobacco (the model plant mostly used to analyze leaf sink-to-source transition), a large number of mutants are available; additionally, its genome is completely sequenced. Therefore working with Arabidopsis may provide an easier way of gaining knowledge about the sink-to-source transition; it may also allow transferring the knowledge obtained more easily to crop species, for example. To sum up, the functional characterization of transcription factors involved in leaf development will permit identifying elements controlling different aspects of leaf physiology, i.e., metabolite reallocation, sugar import and light perception, thus, allowing to gain a better knowledge about the networks that control biomass accumulation, the timing of plant senescence and allocation of nutrients.

Primers

bHLH64

BH64OX_PmeF GTTTAAACATGTTGGAAGGTCTTGTCTCTC

BH64OX_PacR TTAATTAATTAGTAATGAAAACCGAGGCTAGA

PBH64+_BamF GGATCCATTAATTTCTAAAATATTCAATGGTTGG

PBH64+_NcoR CCATGGATTGTTCTCTGCAGAAATCTTTCAA

bZIP21

ZIP21OX_XbaF TCTAGAATGGCGAACCATAGAATGAG

ZIP21OX_EcoR GAATTCTCAGAAATTGGAGAAGTGGTTC

PZP21+_EcoF GAATTCATGCATGATCCATGCTGCTA

PZP21+_NcoF CCATGGCTCTACTCAAGAAAACAGCTTCAA

Dof4

PDOF4+_BamF GGATCCTGTCGTACTAATTAAGTTTTGTGTTGTTG

PDOF4+_NcoR CCATGGGAGTTATAAACTCTTTTTATTTTTTTCTGAAGGAG

Sequencing

35S GAAGTTCATTTCATTTGGAGAGG

Gus+up AAGTTCCAGACGCCATTGAG

Allgemeinverständliche Zusammenfassung

Das Blatt ist das wichtigste photosynthetische Organ von Gefäßpflanzen und die Blattentwicklung ist von einer exakten Genexpression abhängig. Transkriptionsfaktoren (TFs) sind globale Regulatoren der Genexpression. Diese sind, in fast allen biologischen Vorgängen der Eukaryoten, von grundlegender Bedeutung. Die Promotionsarbeit legte den Schwerpunkt auf den sogenannten Sink-source-Übergang in Blättern der Modellpflanze *Arabidopsis thaliana*, zu deutsch Ackerschmalwand. Ein besonderer Fokus lag dabei auf der Analyse von TFs, welche eine wichtige Rolle in der frühen Blattentwicklung spielen.

Sehr junge Blätter befinden sich im sogenannten Sink-Status, sie müssen Photoassimilate aus älteren, sogenannten Source-Blättern importieren, da sie selbst noch nicht in der Lage sind, hinreichend viel Kohlendioxid über die Photosynthese zu binden. Der Übergang vom Sink- in den Source-Zustand eines Blattes ist ein hoch komplizierter biologischer Prozess, der bisher nur in Ansätzen verstanden ist. Im Rahmen der Doktorarbeit wurden molekulare und physiologische Marker identifiziert, die es erlauben, den für das bloße Auge nicht ohne weiteres sichtbaren Sink-Source-Übergang zu erkennen. Dazu wurde beispielsweise die Aktivität bestimmter Gene, unter anderem der Gene *AtSUC2* und *AtCHoR*, mittels molekularer Techniken verfolgt. Um den Übergang zwischen den beiden Entwicklungszuständen eingehend zu charakterisieren wurde die Aktivität von etwa 1900 Regulatorgenen mittels eines multiparallelen Verfahrens - der sogenannten quantitativen RT-PCR - untersucht. Bei den Regulatoren handelt es sich um Transkriptionsfaktoren, die die Aktivität anderer Gene der Pflanzen steuern. Von allen untersuchten Genen zeigten 153 ein vom Blattstadium abhängiges Aktivitätsmuster. Dabei waren Mitglieder der GRF, MYB und SRS Familien überrepräsentiert. Für die gefundenen Transkriptionsfaktoren zeigte sich besonders häufig eine Assoziation zu Prozessen wie Spezialisierung von Zellen, Entwicklung der Epidermis sowie der Blattentwicklung. Zwei ausgewählte Regulatorproteine - bZIP21 und bHLH64 - wurden detaillierter charak-

terisiert.

Das *bZIP21*-Gen zeigte eine starke Aktivität während des Sink-Source-Übergangs. Sein Expressionsmuster in Blättern deckt sich mit dem für *AtCHoR* beobachteten Expressionsmuster, so dass *bZIP21* als ein neuer Marker für die Sink-Source- Transition dienen kann. Es konnten keine homozygoten Null-Mutanten des Gens erhalten werden, was die Vermutung nahelegt, dass gänzliche Abwesenheit von *bZIP21* letal für die Pflanze sein kann. Phylogenetische Analysen ergaben, dass *bZIP21* ortholog zum Gen *Liguleless2* aus Mais ist. In diesen Analysen konnte gezeigt werden, dass alle pflanzlichen bZIP Transkriptionsfaktoren von vier Gründergenen abstammen und alle bZIPs der Angiospermen in 13 homologe Klassen und 34 mögliche orthologe Klassen (Possible Groups of Orthologues, PoGOs) eingeordnet werden können.

Das *bHLH64* Gen ist im unreifen Blatt stark aktiv und während des Alterungsprozesses herunterreguliert. Null-Mutationen von *bHLH64* zeigen eine verzögerte Blütenbildung im Vergleich zum Wildtyp; dies weist auf eine mögliche Verzögerung in des Sink-Source-Übergangs oder Aufrechterhaltung der jugendlichen Identität hin.

Ein dritter Transkriptionsfaktor, *Dof4*, wurde ebenfalls charakterisiert. *Dof4* wird weder während des Sink-Source-Übergangs noch während des Alterungsprozesses unterschiedlich exprimiert. Eine Null-Mutante von *Dof4* besaß größere Blätter und eine höhere Anzahl an Schoten im Vergleich zum Wildtyp. Diese Mutanten erwiesen sich als gut geeignet für die Analyse der Akkumulation pflanzlicher Biomasse.

Obwohl während der Sink-Source Transition nicht überrepräsentiert, scheinen NAC Transkriptionsfaktoren eine große Rolle während des Alterungsprozesses zu spielen. Zweiundzwanzig NAC-Gene von Arabidopsis und 44 von Reis sind in der späten Phase der Blattentwicklung verändert exprimiert. Phylogenetische Analysen erlaubten die Einordnung der meisten dieser NACs in vier homologe Gruppen, was auf einen funktionellen Erhalt zwischen einkeimblättrigen und zweikeimblättrigen Pflanzen hinweist. Um den funktionellen Erhalt von Orthologen zu untersuchen, wurde die Expression von zehn *NAC*-Genen aus Gerste analysiert. Acht dieser Gene zeigten eine von der Blattalterung abhängige Expression. Die Kombination von evolutionären Analysen und funktionellen Studien könnte den Wissenstransfer von Modellpflanzen auf Getreidepflanzen in Zukunft vereinfachen.

Publications

Peer reviewed papers

- **Corrêa L.G.G.**, Riaño-Pachón D.M., Schrago C.G., dos Santos R.V., Mueller-Roeber B., Vincentz M. (2008) The role of bZIP transcription factors in green plant evolution: adaptive features emerging from four founder genes. *PLoS ONE*, 3:e2944.
- Riaño-Pachón D.M., **Corrêa L.G.G.**, Trejos-Espinosa R., Mueller-Roeber B. (2008) Green transcription factors: a Chlamydomonas overview. *Genetics*, 179:31–39.
- Vasconcelos AT, Guimarães AC, Castelletti CH, Caruso CS, Ribeiro C, Yokaichiya F, Armôa GR, Pereira Gda S, da Silva IT, Schrago CG, Fernandes AL, da Silveira AR, Carneiro AG, Carvalho BM, Viana CJ, Gramkow D, Lima FJ, **Corrêa L.G.G.**, Mudado Mde A, Nehab-Hess P, Souza R, Corrêa RL, Russo CA. (2005) MamMiBase: a mitochondrial genome database for mammalian phylogenetic studies. *Bioinformatics*, 21:2566-7.
- Vincentz M., Schlögel P.S., **Corrêa L.G.G.**, Kühne F., Leite A. (2001) Phylogenetic relationships between *Arabidopsis* and sugarcane bZIP transcriptional regulatory factors. *Genetics and Molecular Biology*, 24 (1-4), 55-60

In preparation

- **Corrêa L.G.G.**, Riaño-Pachón D.M., Rensing S., Balazadeh S., Mueller-Roeber B.(in preparation) Phylogenetic analysis of NAC transcription factors and their role in plant senescence.
- **Corrêa L.G.G.**, Fettke J., Fisahn J., Riaño-Pachón D.M., Gomes-Porras J., Maximova E., Lohse M., Mueller-Roeber B. (in preparation) Molecular and physiological analysis of the sink-to-source transition in Arabidopsis leaves.
- **Corrêa L.G.G.**, Arvidsson S., Fettke J., Masclaux-Daubresse C., Mueller-Roeber B. (in preparation) A quantitative RT-PCR platform for identification of leaf developmental stage in Arabidopsis.
- Tomaz J.P., Gauer L., **Corrêa L.G.G.**, Meyer C., Santos D.B., Drumond R.D., Silveira A.B., Vincentz M. (in preparation) Glucose and Mannose regulate the expression of *AtbZIP25* and *AtbZIP63* bZIP genes through two distinct HXK1-independent pathways that involve *ABA2*, *ABI4* and *ABI5*.

Luiz Gustavo Guedes Corrêa

Contact Information

E-mail: correa@uni-potsdam.de — correa@mpimp-golm.mpg.de

Personal Information

Nationality	Brazilian.
Place of birth	Ipatinga – M.G.
Date of bith	October 27th, 1979.

Research Interests

Phylogenetic analysis, transcriptional regulation, evo-devo, transcriptomics, metabolimics.

Education

University of Potsdam, Potsdam, Germany
Department of Molecular Biology

PhD Student (2005–2008).

- Dissertation Topic: "Evolutionary and functional analysis of transcription factors controlling leaf development".
- Supervisor: Prof. Dr. Bernd Mueller-Roeber.

Universidade Estadual de Campinas, Campinas S.P., Brazil
Institute of Biology

Master in Genetics and Molecular Biology, July 2004

- Dissertation Topic: "Phylogenetic analyses of bZIP transcription factors in angiosperms".
- Supervisor: Prof. Dr. Michel Vincentz.

Universidade Estadual de Campinas, Campinas S.P., Brazil
Institute of Biology

Bachelor and License in Biology, February 2002

- Research Topic: "Characterization of ATBZS4 a bZIP transcription factor of Arabidopsis".
- Supervisor: Prof. Dr. Michel Vincentz.

Honors and Awards

Best work in Plant Genetics and Evolution from the Brazilian Genetics Society (SBG).2001.

Best work in Plant Genetics and Evolution from the Brazilian Genetics Society (SBG). Co-author. 2004.

Paul Harris Fellow from the Rotary Foundation for outsdanding humanitarian services.

Fellowships

2000–2001 **FAPESP** Scientific Initiation Fellowship. Undergraduate.

2001 Intership Fellowship from the Sao Paulo Association of French Teachers for outsdanding level of French.

2002–2004 **CAPES** Master Fellowship.

2003 **CNPq** Specialization Fellowship on Bioinformatics.

2004–2008 **DAAD** PhD Fellowship.

2005–2008 Participation in the International PhD Program Integrative Plant Science **(IPP-IPS)**.

Congresses

7–10/09/2004 50th National Congress of Genetics - Florianópolis - Brazil.

18–23/08/2006 8th International Congress of Plant Molecular Biology - Adelaide - Australia.

20–23/06/2007 18th Internation Conference on Arabidopsis Reasearch - Beijing - China.

26–28/09/2007 German Conference on Bioinformatics 2007 - Potsdam - Germany.

17–22/08/2007 XVI Congress of the Federation of European Societies of Plant Biology - Tampere - Finland.

Event Organization

4–6/09/2006 MolConnect Workshop - Bogotá - Colombia.

Computational Skills

- Programming languages: Perl, C++.
- Applications: Microsoft Windows software, LaTeX.
- Phylogenetic Applications: MEGA, Phylip, RAxML, JalView, MAFFT, EpOs.
- Genomic Analysis: TGICL, MeV, Cytoscape.

Molecular Biology Skills

- Molecular techniques: Cloning, RNA-DNA isolation, Northern, Southern and Western blot, PCR, RT-PCR, qRT-PCR, protein expression and purification in *E.coli* heterologous system.
- Microscopy tecnniques: Light and stereo microscopy, tissue sectioning .
- Biochemical techniques: Enzyme activity measurment, chlorophyll and starch measurment, ^{14}C labelling.