Cloning and Functional Characterization of Temperature Responsive Genes of *Ricinus communis*?

Thesis by Eva Goudsmit
Date: 27th May 2015
MSc student Plant Biotechnology
Supervisors: Paulo Ribeiro de Jesus,
Second Supervisor: Wilco Ligterink
Student Number: 920214272120
Course code: PPH-80436

Cloning and Functional Characterization of Temperature Responsive Genes of *Ricinus communis?*

By Eva Goudsmit MSc Plant Biology

Ricinus communis **is an important oilseed crop grown in (sub)-tropical regions worldwide.** *R. communis* **is tolerant to a wide variety of stresses including heat, drought and soil pollution, which makes it very suitable for agricultural production in sub-optimal conditions. However,** *R. communis* **is thermosensitive at its germination stage. The objective of this research was to clone and functionally characterize temperature responsive genes in** *R. communis.* **A genome-wide transcriptomics approach was used to unravel the effect of the temperature on important biochemical and molecular mechanisms required for successful seedling establishment. The selected candidate genes are coding for a (putative) amino acid transporter (***RcAAT),* **glycerol kinase (***RcGK),* **malate synthase (***RcMLS),* **multiprotein bridging factor (***RcMBF),* **three heat shock proteins (***RcHSP1-3),* **two transcription factors (***RcTF1-2)* **and two zinc finger proteins (***RcZFP)***. They were selected based on results of a microarray experiment. These candidate were cloned into** *Agrobacterium tumefaciens* **using the Gateway cloning-system. Subsequently, the transformed bacteria were used to (stably) transform** *Arabidopsis thaliana* **by a floral dip procedure. A pilot study was conducted testing wild type** *A. thaliana* **(Col-0) for their heat tolerance at germination and seedling stage. Furthermore a phylogenetic (bioinformatics) research was performed for all the selected genes to get clues about their function. For all candidate genes,** *A. thaliana* **was successfully transformed. The heat stress pilot showed that the germination of wild type** *A. thaliana (***Col-0***)* **was decreased to 40% at 35°C. In the future, the thermotolerance of germinating seeds and seedlings of the transformed** *A. thaliana* **lines overexpressing the candidate genes should be phenotyped at this temperature. Promising candidates to improve the thermotolerance of the seedlings are** *RcMBF* **and genes encoding the heat shock proteins (***RcHSPs***), because of their involvement in thermotolerance responses as was indicated by the phylogenetic study. Furthermore, the phylogenetic study indicates** *RcTF2***,** *RcGK, RcMLS* **and** *RcZFP1* **as promising candidates to influence thermotolerance because of their reported involvement in major developmental processes including seed germination and seedling.**

Table of Contents

Introduction

Ricinus communis versatility

Ricinus communis (Castor bean) is an important oilseed crop grown in several (sub)-tropical regions worldwide. This species belongs to the Euphorbiacea family and, despites what the name suggests, its seeds are not real beans and they are not edible. In fact, they contain high levels of ricin, ricinine and certain allergens that are highly toxic to humans and animals (Ogunniyi, 2006, Severino et al., 2012). The exact origin of the species is uncertain, but it is generally assumed to be originated from eastern Africa (Ethiopia) where the plants still occur in the wild with high genetic diversity (Harlan, 1976, Foster et al., 2010, Anjani, 2012). This region is also considered to be the origin of domestication, which is proposed to have happened early in history, because *R. communis* seeds have been found in Egyptian tombs dating from 4000 BC (Purseglove, 1976). Additionally, ethno-botanical information dating back to times as early as 500 BC shows that *R. communis* was used in vast areas of the world (including India, China, tropical Africa and later, around 1500 AD, in Europe) where it had a wide array of uses in traditional medicine ranging from healing inflammatory to cardiovascular diseases (Scarpa and Guerci, 1982).

Nowadays, its oil is used for numerous industrial and cosmetic applications such as for biofuels, plastics and soaps (Foster et al., 2010, Ogunniyi, 2006). Major producers of *R. communis* are Brazil, China and India and to a lesser extend some African countries (including Mozambique, South Africa and Ethopia) and Russia (FAOstat, 2014). Its versatility and the interest in biofuels have caused the world-wide production of *R. communis* to almost triple in the past ten years. However, in the most recent years the growth in the production of the crop has stagnated and is now decreasing again (figure 1) (FAOstat, 2014).

Brazilian National Program for Production and Use of Biodiesel (PNPB) has stimulated the production of *R. communis* seeds by creating a huge demand for *R. communis* oil as resource for the biodiesel industry (da Silva César and Otávio Batalha, 2010). *R. communis* was selected by the PNPB program as THE biodiesel crop to promote social development (of small family farms) in semi-arid regions in the north east and the center-west of Brazil. However the mechanisms proposed by the program were insufficient to promote effective production by the family farms (da Silva César and Otávio Batalha, 2010, Rousseff, 2004). One of the main problems that arose was making contracts between the family growers and the companies, causing the castor oil, that was initially produced for the PNPB, to be often sold for higher prices to the chemical industry (da Silva César and Otávio Batalha, 2010). These problems could have led to reduced interest of the biodiesel industry, therewith reducing the demand for the *R. communis* seeds. This could explain the observation that, after peaking in 2011, the *R. communis* production is declining (figure 1).

R. communis: a multitolerant crop

R. communis is tolerant to a diverse range of environmental stresses, including drought stress, heat stress and soil pollution (Langevin, 2010, da Silva César and Otávio Batalha, 2010). For example, *R. communis* plants are still able to partially maintain their photosynthetic functions during severe drought stress and when drought is relieved, they completely recover to normal photosynthesis within one day (Sausen and Rosa, 2010). Furthermore, *R. communis* can deal with soil pollution by heavy metals such as lead, nickel and zinc (Romeiro et al., 2006, Liu et al., 2008). For some metals it is even possible to use the crop for phytoremediation, therewith improving the quality of the soil (de Souza Costa et al., 2012, Olivares et al., 2013). These (tolerance) properties make *R. communis* a very promising crop for countries dealing with arid regions with polluted soils (including a lot of developing countries).

However, during germination of seeds and seedling establishment, *R. communis* is vulnerable to stress (Severino and Auld, 2013). If the seeds suffer from heat stress (35°C) during imbibition, the seedlings will die before they can establish themselves (Ribeiro de Jesus et al., 2014a). If this thermosensitivity could be reduced, it would mean that *R. communis* could be produced in even warmer, more arid, regions than it currently is. This would be a very beneficial trait for an industrial crop like *R. communis,* because it would increase the now limited area which is suitable for its production. Especially since most areas with optimal conditions are reserved for the 'more important' and less tolerant food crops. Furthermore, improvement of *R. communis* could help sustain farmers and their families in areas with harsh environmental conditions. Until then the susceptibility of imbibing and germinating seeds to heat could have a large agricultural impact. When seeds suffer from heat stress in that particular stage *R. communis* will not establish at all

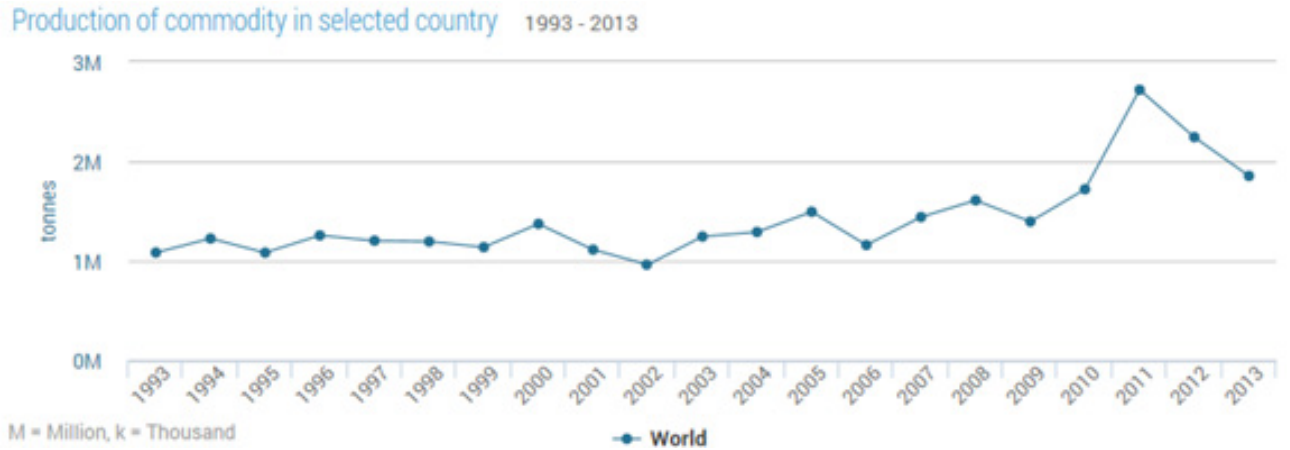

Figure 1: Statistical information of R. communis production in the world obtained by FAOSTAT (FAOstat, 2014).

which will result in complete loss of the crop. To cope with this problem, understanding the molecular mechanisms by which *R. communis* responds to heat stress is essential.

Heat stress

Heat stress can be defined as a rise in temperature beyond a threshold level for a period of time sufficient enough to cause irreversible damages to plant growth and development (Wahid et al., 2007). Heat stress damage consists of various sets of complex perturbations of the homeostasis of the plant (Larkindale et al., 2005c, Kotak et al., 2007). The effects that heat stress can have differs based on the developmental stages, plant tissues, and organs (Larkindale et al., 2005c). Damages caused by heat stress can be detected on different levels; morph-anatomical, physiological and biochemical and the observed damaging effects are all inter-related. The severity of the damage in certain organs or developmental stages is dependent on the dominant, active cellular processes in that organ. The effect of the heat on these cellular processes are dependent on the disturbance of the cellular homeostasis as a result of disruption of biochemical processes.

Disruption of cellular homeostasis by heat stress alters membrane stability and fluidity (Kotak et al., 2007, Wahid et al., 2007, Sangwan et al., 2002). Because of the increased fluidity, some membrane related processes, such as trans-membrane transport, may be disrupted (Larkindale et al., 2005c). Another important adverse effect arises when the temperature window, at which proteins can function well, is exceeded. This causes a disruption of the balance of enzyme activity leading to disruption of metabolic pathways, including the production of primary and secondary metabolites. When temperatures are too high it can eventually lead to complete protein inactivation due to denaturation (Kotak et al., 2007, Larkindale et al., 2005c, Wahid et al., 2007). Furthermore, Reactive Oxygen Species (ROS) are produced because of all the damage to the cell and perturbations in the metabolism (Agarwal et al., 2005). These reactive oxygen molecules cause oxidative damage to even more proteins, DNA and lipids in addition to the direct effects of heating (Suzuki and Mittler, 2006). All this accumulative damage affects major (physiological) processes such as photosynthesis, respiration, regulation of hormone levels and more. On plant level this can cause altered phenology including reproductive failure and early appearance of senescence (Hall, 2000)

Thermotolerance

Thermotolerance is the ability of a plant to actively prevent damage caused by heat and to be able to repair damage caused to heat-sensitive components (Larkindale et al., 2005a). To acquire thermotolerance, the plant has to react to temperature stress with activating heat stress response pathways. These heat stress response pathways lead to two types of thermotolerance for the plants. Firstly there is basal thermotolerance which is the inherent ability of the plant to survive in temperatures above the optimal temperature for growth, and to complete their lifecycle. The other type of thermotolerance is acquired thermo-tolerance which is induced by a short acclimation period of elevated, not yet lethal, temperatures, where after the plant can sustain otherwise lethal temperatures.

More specifically, acquired thermotolerance functions by producing heat shock factors (a form of heat stress transcription factors) that stimulate production of heat stress proteins. The most predominant function of the acquired thermotolerance is stabilizing proteins due to chaperone functions of the heat shock proteins (Kregel, 2002, Wahid et al., 2007). Whereas basal thermotolerance has shown to be involved in heat stress signal transduction, phytohormone regulation and production of ROS-scavenging enzymes (antioxidants) that disarm the ROS molecules by binding and converting them (Larkindale et al., 2005a, Miller et al., 2008, Suzuki and Mittler, 2006).

In previous research the optimal temperature and the heat stress threshold for *R. communis* germination and seedling establishment have been established. Optimal germination temperature for *R. communis* is 25°C and at 35°C proper seedling establishment is severely reduced (Cheema et al., 2010, Moshkin, 1986, Severino and Auld, 2013, Mutimawurugo, 2014, Ribeiro de Jesus et al., 2014a). Despite the fact that germination of *R. communis* is the fastest at 35°C with a ~100% germination percentage, seedlings were affected when developing at this temperature. After germination the seedlings displayed poor growth and malformation resulting in less than 10% of normal seedling establishment (Mutimawurugo, 2014, Ribeiro de Jesus et al., 2014a).

Metabolomics research and Reverse Transcript quantitative Polymerase Chain Reactions (RT-qPCR) were used to analyse the molecular and biochemical causes of this poor seedling establishment. An increase of heat shock protein expression in seeds imbibed at 35°C (Ribeiro de Jesus, 2014) indicates that the seedlings suffered from heat stress and tried to acclimate to their environment (Kotak et al., 2007, Wahid et al., 2007). Furthermore, it was found that a set of genes, involved in the metabolic pathway of seed stored oil, was down-regulated in the seeds imbibed at 35°C in comparison to optimal temperatures (Ribeiro de Jesus et al., 2014). This metabolic pathway includes the mobilization of triacylglyceride molecules by converting them to starch and sugars, as energy supply for the germinating seedling. Disturbance of this process poses a good hypothesis to why the seedlings imbibed under heat stress (35°C) developed so poorly. If the seeds cannot rely on their reserves, they cannot manage to grow into the stage in which photosynthesis can be performed and therefor it will never reach its autotrophic stage. In this research a genome-wide transcriptomics approach was used to unravel the effect of the temperature on even more biochemical and molecular mechanisms that are required for successful seedling establishment of *R. communis*. The objective of this research was to select genes with temperature responsive expression patterns and to functionally characterize them. The candidate genes are cloned to stable transform *A. thaliana* in order to overexpress the candidate genes. Pilot studies, testing *A. thaliana* thermotolerance, were conducted to support future experiments for phenotyping thermotolerance of the produced Arabidopsis transformants. Furthermore, to start the gene function analysis, phylogenetic research is conducted.

For future research, the next step is to perform the phenoty-

ping experiments for the *A. thaliana* transformants overexpressing the selected candidate genes. This data, combined with data of Arabidopsis homolog knockout lines, could serve as a strong basis for conclusions about the functions of the cloned *R. communis* genes. This information will be very valuable for breeding more heat tolerant *R. communis* cultivars to improve the *R. communis* cultivation and therewith the quality of life of the farmers growing it.

Material and methods

Plant material

R. communis seeds (genotypes MPB01 and MPA11) were supplied by Empresa Baiana de Desenvolvimento Agricola SA (EBDA), Salvador-Bahia, Brazil. *A. thaliana* (Col-0) was used for *Agrobacterium tumefaciens*-mediated stable transformation.

Selection of Genes

Nine genes were selected based on their possible involvement in heat-tolerance or heat-sensitivity of *R. communis* seeds during germination (table 1). For target gene selection, a microarray experiment has been conducted on *R. communis* seeds during germination at three different temperatures (20°C, 25°C and 35ºC) as part of an ongoing PhD project conducted by MSc. Paulo Ribeiro at the Plant Physiology department of Wageningen University. Four different stages were defined: dry seeds (Dry), 6-hour-imbibed seeds (6hIS), seeds at radicle protrusion (RP) and seeds with a radicle of 2 cm (R2). From the microarray data, nine genes were selected that showed a significant difference in expression between 35°C and 25°C treated samples in the RP or R2 stage. Additionally, genes encoding for malate synthase 1 (*RcMLS1* and glycerol kinase3 (*RcGK3*) were selected (Mutimawurugo, 2014).

Validation of microarray expression data via RT-qPCR

The cDNA samples used for the RT-qPCR analysis were provided by Paulo Ribeiro, who produced this samples as part of his PhD project at the Plant Physiology department. cDNA samples were prepared from seeds germinating at three different temperatures (20°C, 25°C and 35°C) collected at two different germination stages (radicle protrusion (RP) and 2 cm roots (R2)). Genes were selected that showed significantly elevated or reduced levels at 35°C in comparison to 20°C and 25°C.

Primer design

Primers for RT-qPCR

Primers for RT-qPCR (forward and reverse) were designed using the CLC Workbench 7.5 software (CLC bio, Aarhus, Denmark) with melting temperatures (Tm) of 59–61°C, primer lengths of 18–22 bp and amplicon lengths of 70-200 bp (table 2).

Table 1: Overview of information of genes used for transformation including gene identification number, systematic name, length of coding region of the gene and the putative gene description.

Genembol	Gene identifier	Systematic name	Coding region length (bp)	Putative gene description
RcTF1	255555946	29851.m002359	1068	Transcription factor
RcTF2	255579161	29966.m000228	594	Ccaat-binding transcription factor subunit A
RcZFP1	255585073	27585.m000151	1107	Zinc finger protein
RcZFP2	255586980	28147.m000025	468	Zinc finger protein
RcHSP1	255539774	30147.m014221	890	Heat-shock protein
RcHSP2	255560518	29869.m001176	573	Heat-shock protein
RcHSP3	255585825	29520.m000092	471	Heat-shock protein
RcAAT	255537951	30170.m013661	1749	Amino acid transporter
RcMBF	255546542	29912.m005549	441	Multiprotein-bridging factor
RcGK3*	255553276	29841.m002903	1566	Glycerol kinase
RcMLS1**	255540320	30147.m013773	1704	Malate synthase

** For the RcGK gene the cloning process was completed by Mutimawurugo (2014), transformed A. tumefaciens was available.*
*** For RcMLS1 the cloning and the transformation was completed by Mutimawurugo (2014) and Ribeiro de Jesus (2014). Fifteen transformed plants (T1) plants overexpressing RcMLS1 were available.*

Table 2: Overview of the efficient RT-qPCR primers that were designed for the target genes including the primer number, size of the amplified fragment and the primer sequences.

Gene	Primer #	Fragment size (bp)	5' end primer sequence (Fwd)	3' end primer sequence (Rev)
RcTF1	2667	128	CACTGCCAATAAACCCAAAACC	ATTTTGCATACCATCACCTCCT
RcTF2	2685	121	GAGTATGGTGGTGGTGGTG	GATGCTGATGCTGATGTTGTT
RcZFP1	2669	88	GCTTGGGTTCATTGGGTTGT	ACCGTAACTTCTTCCATGCATC
RcZFP2	2723	192	CTTCTTCTTCTTCATCGCCACA	ATGATCCGGACTTCCCTCC
RcHSP1	2673	177	GGCATTGGAGGTGGATTAGG	CCTGCCACTCTCTTCATCTT
RcHSP2	2675	199	TCCTTCCAACACTCGTCTTTCT	TGCGCTGATGGTGTCTCT
RcHSP3	2721	194	CTTCAACTCCGCAAATTTACCT	CCTGCTCTTTGCTTCTTTCT
RcAAT	2681	150	CCCTTCTTGTTTTTGCTCCTCT	GCATTCCAATACCTCCTTCCAT
RcMBF	2727	160	CAAGTCCAAAACCAAAGCCCA	CCCTTCGTCCAGCTTCTTC

Table 3: Primers for cloning of the target genes. Numbers of primer working solution (10x diluted mix of the forward and reverse primer) are given. Furthermore the (functional) annealing temperatures for which the primers were shown to amplify the target genes are given. Lastly the forward and reverse primer sequences and the number of the matching attB primer working solution are shown for all genes.

Gene	Primer #	Anneal. temp	Sequence F	Sequence R	attB #
RcTF1	2508	53.8-56	ATGGAGAGTACTGATTCATCCT	TTAAGAGTTCCAGTTCATGCTT	2637
RcTF2	2600	58,0	ATGGCGGACTCGGACAAC	CTAAGGTGGCCTACCAGAGGAT	2655
RcZFP1	2512	53.8-56	ATGGCAAGTTTAAGATCTCATT	TTATGCCTCACATAGTTCTTTAAG	2639
RcZFP2	2514	50,6	ATGGATTTCCAACCCAATAC	TCAAAGTCTCAAAGACAAGTC	2641
RcHSP1	2580	51,8	ATGGCATCTTCTCTAGCTTT	CTACTCAACCTTAATTTGAACCG	2643
RcHSP2	2584	55,0	ATGGCCAGCCCAAGATTC	TCACATCTCAGACTTCACAGTCTT	2645
RcHSP3	2586	51,8	ATGTCTCTAATTCCTAGCATCTT	TTAGCCGGAAATTTGAACAG	2647
RcAAT	2594	53,2	ATGACCGTTGTATCCTTATCTAGC	TCAGCTTCCAACACCACT	2651
RcMBF	2689	50,4	ATG CCA AGC AGA TCT ACA	TTA CTT GCC AAT CTT TCC C	2731

The PCR amplification efficiency was tested with dilutions of a pooled cDNA sample and/or plasmid DNA containing the target gene. If the test with the pooled cDNA did not show proper amplification, the test with the plasmid DNA was performed to establish if the concentration of the target sequence was too low in the cDNA or if the primers were inefficient. Initial concentration of the vector solution was 1.25 pg/µl. This concentration was established by making a dilution series of the plasmid DNA samples with dilution steps of 20x until the Cq was approximately 21.

Primer design for target gene amplification

For each target gene the *R. communis* DNA sequence of that gene was obtained from the NCBI (Pruitt et al., 2005). Primers were designed matching the first 18-24 and the last 18-24 base pairs of this reference sequence (table 3). Primers matching the last 18-24 basepairs were converted into reverse complement sequences (bioinformatics.org) to obtain a reverse primer. The exact number of base pairs for the forward and reverse primers was determined by prediction of the annealing temperature by the Oglioanalyzer (IDT Company) which differed with different lengths of primers. Primers sets for all target genes were selected with annealing temperatures between 50°C and 60°C with less than 1°C difference. After designing the primers the forward and reverse primers were mixed together in a 10x dilution with DEPC water.

Initially, to assess whether the designed primers could be used to amplify the full-length coding regions of the target genes from a cDNA library, a PCR was done using (per reaction) 6.65 µl of MQ, 1,5 µl of PCR buffer and $MgCL_2$, 0,8 µl of the primers at 10µM, 0,4 µl of dNTP's (10 mM), 0.15 µl of Firepol (5 units/µl) and 4 µl of cDNA. Each primer was tested for the cDNA produced from seeds and leaves as described in the section 'RNA isolation and cDNA synthesis' below. PCRs were run with a gradient program: 95°C for 5 minutes (for separation of the two DNA strands) followed by 40 cycles of [30 seconds of the primer annealing temperatures (ranging from 50-60°C) followed by 72°C for DNA synthesis with a period dependent on the length of the gene with as directive 1000 bp/min]. Eight different annealing temperatures were tested using gradient PCR programs (50,4°C, 50,6°C, 52,8°C, 53,2°C, 53,8°C, 55°C, 56°C and 58°C). Which specific annealing temperatures was tested for which primers was dependent on the predicted primer annealing temperature. For each gene, a specific temperature (range) was determined for which the primers (working solution) successfully annealed to the cDNA and amplified the target gene (table 3).

The amplicon products of the PCR were tested for length and abundance with gel-electrophoresis on 1% agarose (TEA) gels. After electrophoresis the gels were post stained in a 60x diluted solution of gelred in demi-water for approximately 30 minutes. Visualization of the bands was done using the Bio Rad Gel Doc™ XT. Aplicon length was determined by comparison to a 10000 bp SmartLadder (Eurogentech).

For cloning purposes these primers were also designed with additional attB-sites for the 5'end of the forward and the reverse primer. The addition attB sequence for the forward primer was: GGGG ACA AGT TTG TAC AAA AAA GCA GGC T, and for the reverse primer: GGGG AC CAC TTT GTA CAA GAA AGC TGG GT. The attB primers were tested for amplification of the target genes similar as described for the original primers.

All primers, for RT-qPCR as well as cloning, were ordered from the IDT DNA Company (Leuven, Belgium).

Production of *A. thaliana* transformants overexpressing the *R. communis* genes

In this research a range of different techniques was applied to complete the whole process from target gene amplification until selection of stable transformed *A. thaliana* plants overexpressing *R. communis* genes. In this process four different research phases can be distinguished; firstly the target gene amplification and isolation, second the cloning of the target genes using the Gateway system. Thirdly, the transformed *A. tumefaciens* (containing the plasmids with the target genes) was used for the floral dip procedure to transform *A. thaliana,* for which, subsequently single insert homozygous lines were selected in phase four (figure 2).

1. Target gene amplification and isolation
- RNA isolation
- cDNA library synthesis
- Primer design and PCR with Firepol
- Gateway primer design and PCR with Firepol
- PCR with Phusion
- Extraction attB-flanked fragment from gel
- Second PCR with Phusion
- Second extraction attB-flanked fragment from gel

2. Gateway cloning procedure
- BP reaction
- Transformation of E. coli
- Colony PCR
- Vector isolation
- Sequencing of vector
- LR reaction
- Transformation of E. coli
- Colony PCR
- Vector isolation
- Enzyme digestion
- Transformation of A. tumefaciens
- Preparing glycerol stock

3. Floral dip procedure
- Growing A. thaliana plants
- Grow 100ml of A. tumefaciens culture
- Perform floral dip
- Seal plant with plastic bag
- Remove bag
- Allow plants to produce seeds
- Harvest T1 seeds

4. Homozygot selection
- Test T1 seeds for kanamycin resistance
- Harvest and test T2 seeds for a 75/25 resistance ratio
- Harvest and test T3 seeds for 100% resistant lines
- Grow and harvest T4 seeds under experimental conditions, including growing them simultaneously with the control

Figure 2: An overview of the different phases of the research for stable transformation of Arabidopsis overexpressing the candidate genes is given.

Target gene amplification and isolation

RNA isolation and cDNA synthesis

RNA was extracted from *R. communis* leaf and seed samples. Seeds (of genotype MPB01) that germinated at 20ºC and that showed a radicle of 2 cm were used. Leaves (genotype MPA11) were collected from seedlings that grew at 20ºC for 14 days.

RNA extraction was performed using the hot-Borate protocol following Jawdat and Karajoli (2012). The Hot-Borate buffer was pre-made containing 0.2M Na borate decahydrate (Borax), 30mM EGTA, 1% SDS and 1% Na deoxycholate (Na-DOC) in DEPC water. Of this buffer 800 µl was combined with 48 mg PVP and 1.6 mg DDT to complete the extraction buffer (XT) after which it was added to the plant material. After addition of the XT buffer, the samples were ground. Four microliter of proteinase K solution (0.28 mg/µl) was added to the sample and the mixture was vortexed for 15 minutes, which was followed by an incubation step of 12 minutes at 42°C. Subsequently, 64 µl of 2M KCL was added to the samples and it was left for 1 hour on ice. Hereafter, the samples were centrifuged for 20 minutes at 12000g at 4ºC. The supernatant containing the DNA was then transferred to a new tube and 259 µl of ice-cold 8M LiCl was added. Samples were incubated on ice overnight. Following incubation, samples were centrifuged for 20 minutes at 12000g at 4ºC. The remaining pellet was washed with 750 µl of ice cold 2M LiCl. Each pellet was resuspended in 80 µl of DEPC-water. DNAse digestion was performed on 0.125 µg of RNA by adding 10 µl of DNAse buffer (premade) and 10 µl of DNAse to a final volume of 100 µl. Samples were incubated for 30 minutes at 37°C.

RNA quantification and quality control were performed spectrophotometrically using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies). RNA quality was further inspected by the integrity of ribosomal RNA bands on a 1% agarose gel. Sharp and intense 18S and 28S ribosomal RNA bands without visible degradation, confirmed the quality of the isolate RNA.

cDNA synthesis and target gene amplificaition

The cDNA synthesis from the isolated RNA was performed using the Bio Rad iScript™ cDNA synthesis kit according to the manufacturer's instructions. This kit (for performing a reverse transcriptase reaction) includes the iScript™ mix which consists of a RNase inhibitor, oligo(dT) and random hexamers as primers and RNase H^+ iScript reverse transcriptase. For each sample 7 µl of RNA template (total of 490 ng) was mixed with 8 µl of DEPC (nuclease free) water. Then, 4 µl of iScript reaction mix and 1 µl of the enzyme was added. The RT-PCR program consisted of five minutes at 25ºC, 40 minutes at 42ºC and 5 minutes 85ºC. cDNA samples were diluted 20 times prior to RT-qPCR analysis.

Amplification of the target genes from this cDNA was performed using the primers with attB sites for cloning (table 3). For the PCR, a high-Fidelity DNA Polymerase containing proofreading system was used. For each sample the reaction mix consisted of 10,4 µl of MQ water, 4 µl of HF

buffer, 4 µl of cDNA, 1 µl of primer (1 nM), 0,4 µl of dNTP's (10 mM) and 0,2 µl of Phusion enzyme (2 units/µl). The PCR program and electrophoresis experiments were performed as described in the 'Primer design for cloning purposes' section. The gels were post stained with the 60x diluted Gelred solution. The bands formed after electrophoresis, with the right fragment length, were excised from the gel and amplified fragments were isolated using the QIAquick Gel Extraction Kit according to the manufacturers' instructions (QIAGEN). The concentration of the amplified fragments was measured using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies) and ranged from 5 to 55 ng/µl.

Gateway cloning procedure

BP reaction and first *E.coli* transformation

Amplified (target gene) products were recombined into the donor plasmid pDONr207 by performing a BP reaction. For the BP reaction the following components were mixed; 15-150 ng of attB-flanked-PCR product, ~150 ng of the donor vector pDONR207 and the necessary volume of the Gateway® TE Buffer to complete the volume to 8 µl in total. The reaction was incubated at 37°C overnight. The next morning the reaction was stopped using 1 µl of Proteinase K. The pDONR207 contains a gene for gentamycin resistance as selection marker. After the BP reaction, 1 µl of this donor plasmid was used to transform ~100 µl of *E. coli* competent cells (strain DH5α) using electroporation (180 Volts, 2 sec pulse). After electroporation, bacteria suspensions were incubated for one hour at 37°C with 250 µl of unselective liquid LB medium to regain strength. Thereafter, they were plated on solid (1% agar) LB medium with 25 µg/ml gentamycin and grown overnight at 37°C for selection of transformed cells. Of the colonies that were formed, four were selected and transferred to 5 ml liquid LB medium with 25 µg/ml gentamycin.

Subsequently, each of the selected cultures was tested for presence of a target gene by performing a colony PCR. This PCR reaction is a modification of the PCR as described in the section 'Primer design for target gene amplification'. For the colony PCR the 4 µl of cDNA was replaced by a sample of living bacteria. To do so first 50 µl of the liquid bacteria cultures was spun down and the supernatant was removed. 10.65 µl MQ water was added to the bacteria-pellet, which was re-suspended and then the rest of the ingredients were added and used as PCR mixture. The only adaption made to the PCR program is the prolonged 95°C at the start, which was extended from 5 to 10 minutes. The presence of the amplified gene was checked by electrophoresis.

If the colony PCR showed amplicons of the expected length for the target genes, two of these 'positive' strains were selected for plasmid isolation. This plasmid isolation was performed using the QIAprep Spin Miniprep Kit according to the manufacturers quick start protocol (QIAGEN). In this protocol bacteria of the 5 ml liquid culture are pelleted by centrifugation. Thereafter, they were resuspended in 250 µl of P1, after which the P2 buffer for lysis of the cells. The lysis reaction was stopped using 350 µl of buffer N3. After lysis of the cells, the DNA dissolved in the extraction buffers. Centrifugation was used to separate the DNA-containing supernatant from the bacteria cell residues. Hereafter, the DNA was bound to the QIAprep 2.0 spin column (provided) by centrifugation of the supernatant through this column. After binding the DNA was washed using 750 µl of Buffer PE followed by the elution of the DNA from the membrane using 30 µl DEPC water. The concentration of the isolated plasmid DNA was measured using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies).

As an extra control, the isolated plasmids were sent for sequencing to Macrogen Europe. For each plasmid containing one target gene two samples were sent; one with a primer annealing to pDonr207 upstream of the target gene sequence (Forward) and one sample with a reverse complement primer annealing downstream of the target gene. Samples for sequencing were prepared by adding 5 µl of one of the two pDonr207 primers (forward or reverse) to 500 ng of plasmid (volume dependent on the concentration). MQ water was added to reach the required 10 µl. Sequencing results were analysed using CLC Workbench 7.5 software (CLC bio, Aarhus, Denmark), in which the sequence of the gene cloned into the plasmid was compared to a reference sequence of the *R. communis* target genes.

LR reaction and second *E.coli* transformation

After confirmation of the gene being present in the BP plasmid with the right sequence, the isolated plasmid was used to start the LR reaction as is described in the Gateway® protocol. For this reaction 50-150 ng of the pDONR207 vector containing a target gene was combined with 50-150 ng of the destination vector pGD625. The volume was completed to 8 µl using the Gateway® TE Buffer. The LR reaction was incubated overnight at 37°C and was stopped the next morning by adding 1 µl of proteinase K. Expression plasmid pGD625 harbours a Cauliflower Mosaic Virus (CaMV) 35S promoter, and as selection marker it harbours a gene for kanamycin resistance. Competent cells (strain: DH5α) of *E. coli* bacteria were transformed with the LR product using electroporation as described in the section 'BP reaction and first *E.coli* transformation'.
Transformed bacteria were selected by growing them overnight on 1% agar LB medium with 50 µg/ml Kanamycin.

Four colonies were selected for each *R. communis* target gene and these were grown overnight in liquid LB medium with Kanamycin (50 µg/ml). A colony PCR test and the subsequent plasmid isolation of plasmids was performed as described in the section 'BP reaction and first *E. coli* transformation'.

As a final check the expression plasmids were subjected to an enzyme digestion using the restriction endonuclease enzyme ECOR1-HF® (high fidelity of New England Biolabs). Therefore, a 3-hour digestion protocol was used with for each sample 500ng of plasmid, 1,5 µl of Cutsmart buffer, 1 µl of ECOR1 and MQ water to complete the sample to 15 µl. Samples were incubated at 37°C as this is the optimal functioning temperature of the restriction enzyme. After digestion of the plasmids, the sizes of restriction fragments were analyzes using gel electrophoresis.

***A. tumefaciens* mediated transformation**

The isolated pGD625 plasmids (each containing one of the target genes) were used for transformation of *A. tumefaciens*. For each gene 1 µl of the pGD625-gene solution was mixed with ~75 µl of competent *A. tumefaciens* cells (Agl0 strain) which subsequenlty underwent electroporation (1.8 V). Cells in the suspension were allowed to recover in 250 µl of unselective LB medium for 2 hours at 28°C. Thereafter, cell suspensions were placed on selective medium with kanamycin (50 µg/ml) and rifampicin (25 µg/ml) and were allowed to grow for two days at 28ºC. Four colonies were selected for each gene and were grown in liquid LB medium

with kanamycin (50 µg/ml) and rifampicin (25 µg/ml). All colonies were tested with a colony PCR as described in the section of 'BP reaction and first *E. coli* transformation'. Two of the four selected colonies for each gene that showed amplification fragments with the length of the target genes were selected to use for transformation of *A. thaliana.*

Glycerol stock

Within this experiment, bacterial strains of all phases for each cloned gene were stored as glycerol stock (*E coli* transformed with the BP and the LR reaction and two independent *A. tumefaciens* clones containing the LR plasmid). To do so, 50 µl aliquots were taken of the target gene containing bacterial strains before plasmid isolation. These aliquots were used to grow new 2 ml liquid bacteria cultures in LB medium with Gentamycin (25 µg/ml) overnight. The fresh bacteria cultures were used the next day to produce glycerol stock by adding 250 µl to 750 µl of a 60% glycerol solution after which the bacteria stocks were immediately frozen in liquid nitrogen and stored at -80°C.

Floral dip procedure

The Arabidopsis floral dip procedure was performed following Clough and Bent (1998) and Logemann et al. (2006) on *A. thaliana* Colombia-Zero (Col-0) plants with some modifications. For the floral dip procedure *A. tumefaciens* bacteria carrying the target genes were isolated from a 100ml liquid culture. The bacterial pellet was re-suspended in 100 ml 5% Sucrose solution with Silwet L-77 (0,05%). For transformation of *A. thaliana* newly formed flowers were dipped in this suspension. After the flower dip, plants were kept in humid conditions by sealing them for a day with a plastic bag, which was removed thereafter. Plants were allowed to develop further, form siliques and subsequently dry for 3-5 weeks after which the seeds (T1) were harvested. For each gene, independent plants were dipped using two independent Agrobacterium strains.

Homozygote selection

T1: selection of transformants

Seeds of the first generation (T1) were sterilized using a liquid sterilization protocol. Seeds were sterilized in an eppendorf tube by adding 1 ml of ethanol and vortexing the tubes for 5 minutes. Seeds were spun down and the ethanol was removed. Then, 1 ml of 50% commercial bleach with 0.1% Triton was added. Seeds were vortexed for 3 minutes and spun down again. Bleach was removed and autoclaved demi water was used for five subsequent washing steps. Sterilized seeds were sown on a 1/2 MS plate with kanamycin (50 µg/ml) for selection. These plates were put at 4°C for 3 days for stratification. Thereafter, they were transferred to a chamber with controlled temperature (24ºC; 16h/8h photoperiod). Plants that were able to grow on the media were considered to be transformed. These transformants were transferred to rockwool soaked in nutrient solution (Hyponex) and allowed to grow at 24ºC with a 16h/8h photoperiod in a climate chamber, in a plastic propagator box to maintain high humidity. Every day the humidity was reduced by slightly opening the containers until the plants were habituated and were ready to grow in open air. After 3 days of growing without a lid the plants were moved to the Unifarm Greenhouse where the plants received Hyponex three times a week until the seeds could be harvested.

T2 and T3: selection of homozygous single copy insert lines

For the following selection steps, seeds (T2 and T3) were sterilized using gas sterilization. This was performed by placing open eppendorf tubes with seeds in a closed incubator with 100 ml undiluted commercial bleach mixed with 3ml of HCl (37%) overnight. Thereafter, seeds were scattered on selective 1/2 MS plates with 50 µg/ml kanamycin and were placed at 4°C for 3 days.

For T2 selection approximately 100 seeds were sown on selective 1/2 MS medium containing 50 µg/ml Kanamycin. After two weeks T2 lines that showed a 75/25 ratio for resistant / sensitive seedlings were selected after counting the seedlings as single copy insert lines. The ratio was tested statistically for a 75/25 ratio using a chi-square test. Seedlings were transferred to rockwool with nutrient medium and were placed in growing containers to contain the high humidity and slowly habituate them to an environment with lower humidity. This habituation was performed as described in the section of the T1 selection. These containers were placed in a climate chamber. After this habituation the plants were transferred to less controlled conditions: a flooding table that floods 2 times a day with Hyponex, in a greenhouse. These plants are currently growing in the flooding table until their siliques are formed. After maturation of the first siliques the plants will not receive any water or Hyponex to allow the plants and their T3 seeds to dry.

Homozygote selection of Arabidopsis plants overexpressing the *RcMLS* gene

The homozygote selection for the Arabidopsis plants overexpressing the *RcMLS* gene was part of a pilot research. The homozygote selection procedure for this plant was quite different from the selection procedure of the other plants. Firstly the T1 transformants of the *RcMLS* overexpression line were already selected and grown by Paulo Ribeiro and Chantal Mutimawurugo (Mutimawurugo, 2014). After the selection these plants were placed directly in open air in the 24ºC and 16h/8h light regime conditions of the climate chamber where they completed their lifecycle.

T2 selection was performed similar to the described T2 selection for the other lines, but growth of the T2 generation of transformants overexpressing *RcMLS* was different than for the other lines. The T2 plants of the MLS lines were transferred directly from the selection Petridish to soil pots in the greenhouse after which the Aracon System was used to grow the plants and harvest the T3 seeds.

Harvested T3 seeds were sterilized, stratified and selected for 100% transformant lines to ensure the selection of homozygous lines. Of the selected lines, new (unsterilized) seeds together with WT Arabidopsis (Col-0) seeds for control were directly sown onto rockwool soaked in Hyponex, and were placed in the greenhouse. After ~7 weeks plants were deprived

of water to allow siliques to dry. After ~8 weeks the seeds were harvested and used for a pilot phenotypic experiment.

Arabidopsis Heat Stress Pilot

Heat Stress Treshold experiment

As a pilot for the phenotypic experiments testing the *A. thaliana* lines (that were transformed with the selected genes in this experiment), a threshold experiment for germination under heat stress with *A. thaliana* WT Col-0 was conducted. Germination under a range of temperatures was tested using the Germinator system. For each temperature, three samples with approximately 50-100 seeds were placed in a plastic tray with two blue filter papers which were soaked using 50 ml of demi water. The trays were closed with a lid and placed for three days at 4°C for stratification. Thereafter, the trays were stacked with two other trays above and below containing white filter paper and the same amount of water to avoid unequal evaporation, and the whole pile was closed with a lid. These stacks were placed in a plastic bag to keep the environment humid for the germinating seeds. These bags were then placed in incubators with temperatures ranging from 30°C to 37°C and in 22°C for control. During the period of germination several pictures were made using the Germinator set-up. These photos were later analysed using the Germinator package with help of ImageJ to establish the germination rate and percentage which were subsequently analysed with excel using a student t-test for the statistics.

Life Cyle experiment

For the lifecycle experiment, it was tested if the seeds germinated under threshold heat stress conditions were able to complete their lifecycle in that temperature. Ten seeds germinated in 34°C, 35°C and 22°C (at protrused radicle stage) were transferred to rockwool with Hyponex placed in a growing container for the humidity which in its case was placed in the same incubator (with the same temperature) as in which the seeds germinated. Of each line also 10 germinated seeds were placed in a climate chamber growing the plants at ~22°C after their heat stressed germination to see if they could complete their lifecycle after a germination with heat stress.

Phylogenetic trees

Phylogenetic trees were made to provide additional clues regarding the putative functions of the genes. For orientation purposes each gene was first blasted using tBlastx against the database of the NCBI (Pruitt et al., 2005). Thereafter the genes were also checked against the PFAM database using their translated protein sequence using the translation tool from nucleotide sequence to protein sequence by Expasy (http://expasy.org/tools/dna.html).

For the construction of the phylogenetic tree, Phytozome V10.1 (Goodstein et al., 2012) was used as source of homologous protein sequences of different organisms. Similar sequences were searched using the BLOSUM62 matrix to blast protein sequences (pBlast) of the cloned *R. communis* genes against the protein genome data of four species with high quality sequencing/annotation; *Manihot esculenta* (cassava), *Populus trichocarpum*, *A. thaliana* (coming from the previously used TAIR database) and *Oryza sativa* (rice).

M. esculenta and *P. trichocarpa* were selected because of their close evolutionary relation to *R. communis*, and because of the sequence quality. *O. sativa* and *A. thaliana* were chosen because of the quality of the sequencing and annotation of their genome. For each group of genes the Blast results with E values lower than E^{-10} were selected. For ZFP1 and TF1, however, because of the high similarity with many genes only the best 100 hits were selected.

These sequences were then put in a new database and were converted into Fasta formats using two letters to annotate the species of the gene and the unique PAC code from the Phytozome database (Goodstein et al., 2012), for example Rc12345678). The Fasta file was uploaded into the HMMscan for detection of conserved domains to annotate the types of proteins. This was conducted to confirm the protein types and the families of the related proteins (Finn et al., 2011). (Appendix 8)

The same Fasta file was uploaded into Cello (Yu et al., 2006) to predict the subcellular localization of the proteins of each phylogenetic tree. After collection of that data, sequences were aligned using ClustalX2 (Thompson et al., 2002). If the remaining gaps were assessed as being interfering with an appropriate alignment, the Trimal tool (Capella-Gutiérrez et al., 2009) provided by the Phylemon website (Sánchez et al., 2011) was used to trim all sequence so that at least 40% of the most conserved regions of the sequence were kept, whereas the most differing regions were cut out. This was done for the protein sequences used for the HSP, TF and the ZFP phylogenetic trees (Appendix 1). Trimmed sequences were thereafter used to build an alignment and a neighbour joining tree using MEGA 6 (Kumar et al., 2008), with 500x bootstrapping.

Clustal X2 (Thompson et al., 2002) was used to alignment *R. Communis* protein sequences of similarly annotated genes. If shared conserved domains between the proteins were observed (as for RcHSP and RcTF) one phylogenetic three was constructed. If not, as for RcZFP, two separate phylogenetic trees were constructed. When combining two *R. communis* target genes within one phylogenetic tree, conflicts appeared because some of the homologs of one gene were too different from the homologs of the other gene. These proteins could therefore not be aligned. If this arose, protein sequences of both genes were compared to the *R. communis* protein sequences. The conflicting genes with the lowest similarity (to the *R. communis* genes), or genes causing multiple conflicts were removed in order to have as little impact on the phylogenetic tree as possible.

After construction of the phylogenetic three, Phytozome V10.1 (Goodstein et al., 2012), Plant Transcription Factor Database (TFDB) (Jin et al., 2013) and the Database of Rice Transcription Factors (DRTF) (Gao et al., 2006) were used to find homology groups. Homology groups were based on described (named) *A. thaliana* proteins or their close undescribed *A. thaliana* homologs.

Results and Discussion

For selection of good candidate genes, two separate transcriptomics approaches (microarray and RT-qPCR) were used to ensure the effect of temperature on the expression of the selected genes. Comparable data of two very different techniques validates the temperature responsive expression patterns of the nine selected candidate genes.

Verification of the expression patterns of the candidate genes

For each candidate gene, the expression profile observed in the microarray experiment was compared with data from a RT-qPCR experiment. Relative expression of the genes was measured for two germination stages; Radicle Protrusion (RP) and 2 cm root (R2) each for three temperatures; 20°C, 25°C and 35°C. Candidate genes where selected when expression at 20°C and 25°C differed significantly from 35°C. In these two stages the effect of heat stress on gene expression was most prominent (data not shown) despite the fact that seedlings suffered most when they experienced heat stress during imbibition (Ribeiro de Jesus et al., 2014a). There seems to be a delay between the moment in which the heat stress responsive pathways are triggered and the moment where the actual difference in expression of the genes can be perceived. This delay could be explained by the time needed for the signal transduction pathways to activate the downstream targets. One of the delaying factors in the signal transduction pathways are switching delays. This is the time required to accumulate the necessary concentration of proteins in their effective form to activate or repress controlled promoters to alter the actual gene expression. These switching delays show high variability (McAdams and Arkin, 1997).

Genes with elevated expression by heat stress temperatures (35°C)

For the genes; *RcMBF, RcHSP1, RcHSP2* and *RcHSP3* the same gene expression pattern was observed for the microarray and RT-qPCR experiment (figure 3). For *RcMBF, RcHSP2*

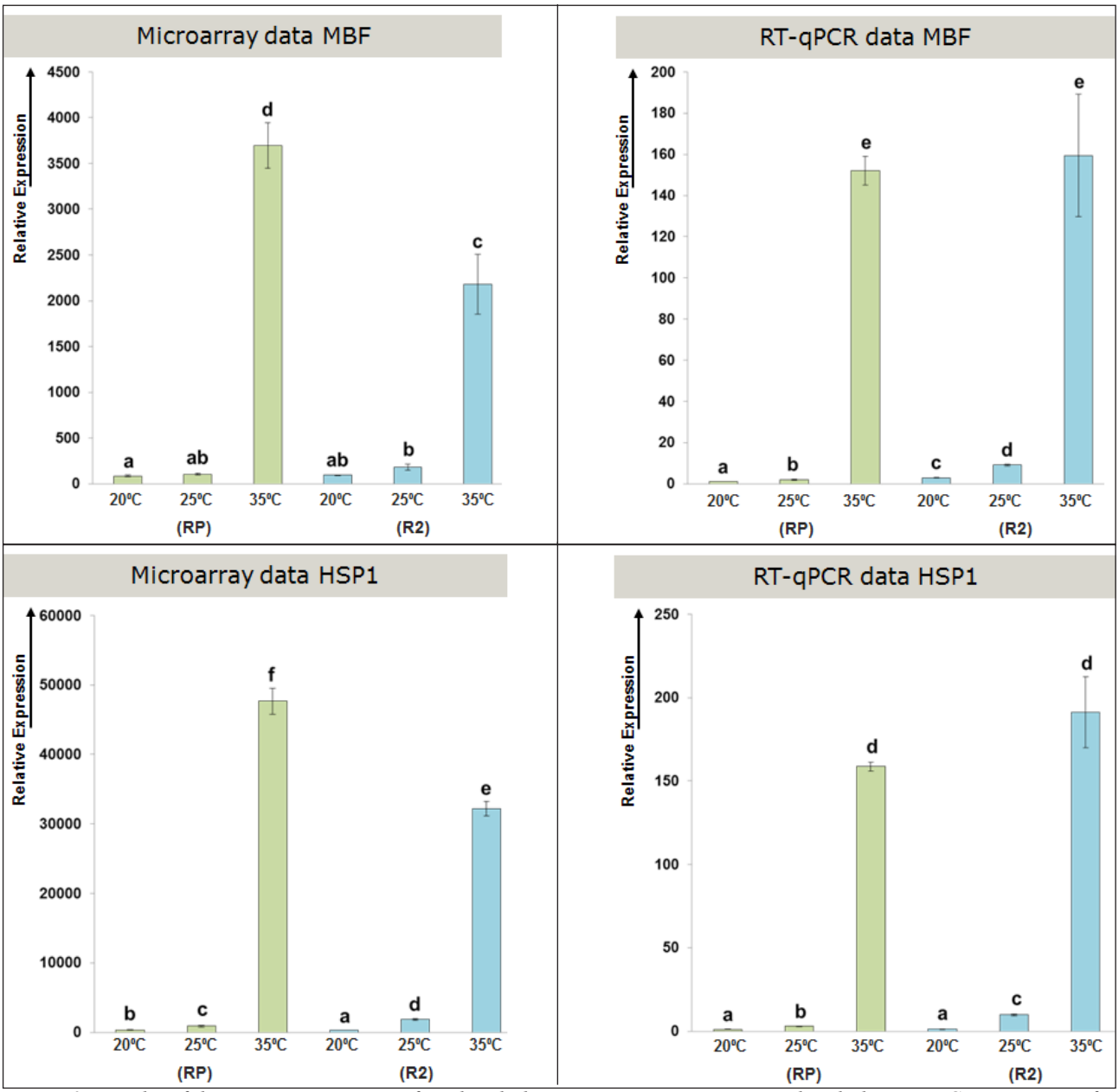

Figure 3: Results of the expression patterns found with the microarray experiment and with the RT-qPCR experiment for validation for genes up-regulated by heat stress.

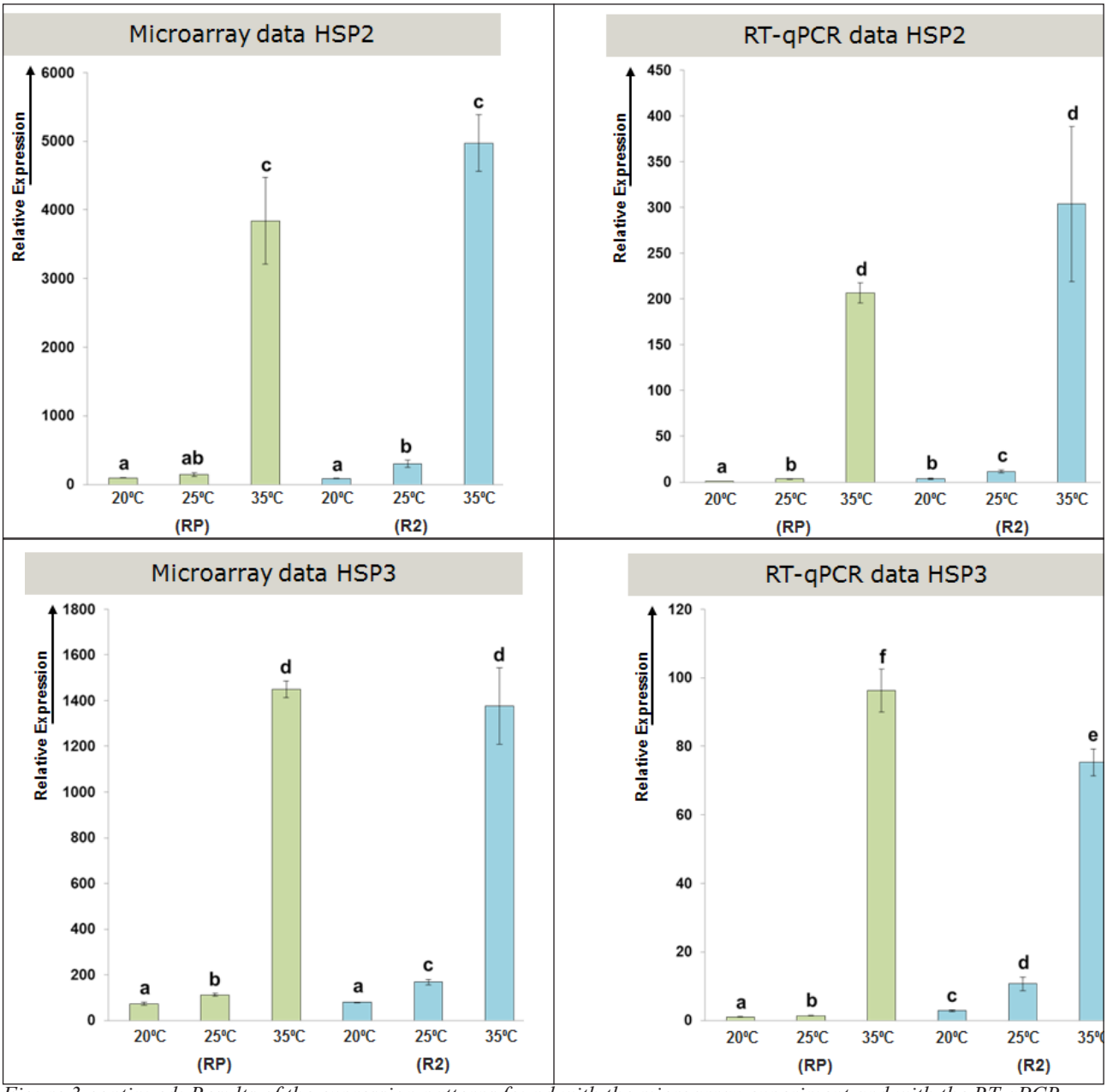

Figure 3 continued: Results of the expression patterns found with the microarray experiment and with the RT-qPCR experiment for validation for genes up-regulated by heat stress.

and *RcHSP3* expression levels in seeds germinating at 35°C were elevated in comparison to the expression in seeds at 20°C in RP and in R2 (up to 57-fold). For *RcHSP1* the effect of the high temperature on expression was most significant showing elevated expression levels up to a 135 fold for 35°C in comparison to 20°C. Furthermore, for all four genes the RT-qPCR results show significant elevated levels at 25°C in comparison to 20°C and for *RcHSP1* and *RcHSP3* this is also true for the microarray data (up to 3-fold).

Genes with reduced expression at heat stress temperatures (35°C)

For the genes *RcTF1, RcZFP1 and RcZFP2* the same trend in the expression patterns of these genes germinating under different temperatures was observed for the microarray and RT-qPCR experiments (figure 4). Levels of *RcTF1* were greater in seeds germinating at 20°C than at 25°C and 35°C (up to 2.1-fold) in RP and R2 samples. For *RcZFP1* levels were higher at 20°C and 25°C than at 35°C (up to 2.9-fold) in RP and R2. Furthermore *RcZFP1* levels were also significantly lower in R2 seeds germinating at 25°C than at 20°C (1.5-fold). Levels of *RcZFP2* were significantly higher in seeds germinating at 20°C and 25°C than at 35°C (up to 2.6-fold). For *RcZFP2* relative expression data of the RT-qPCR also shows reduction in expression in seeds germinating at 25°C in comparison to seeds germinating at 20°C but this is not true for the data of the microarray experiment.

For *RcAAT* the microarray shows higher relative expression in seeds germinating at 20°C and 25°C than at 35°C (up to 2.1 fold). For *RcTF2* the microarray shows higher relative expression in seeds germinating at 20°C and 25°C than at 35°C (up to 4.5 fold). Moreover, for *RcTF2* levels are already significantly reduced in seeds germinating at 25°C in comparison to 20°C (up to 2-fold) (figure 5).

However the results for these two genes are not confirmed

by the RT-qPCR experiment. For *RcAAT*, the discrepancy between the two experiments is expected to be a result of the efficiency of the RT-qPCR primers. This is hypothesized since the primer efficiency could not be determined with the

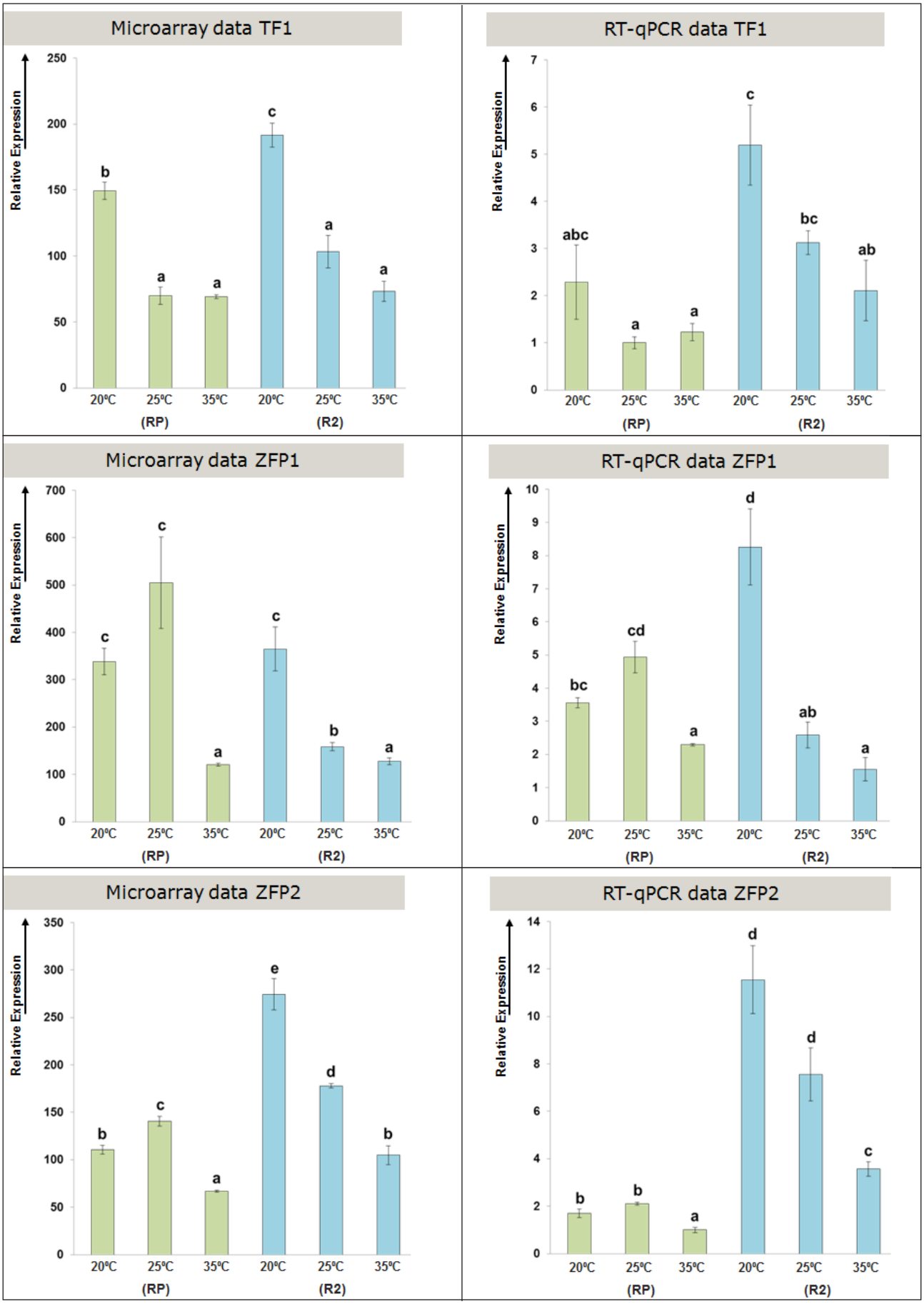

Figure 4: The results of the expression patterns found with the microarray experiment and with the RT-qPCR experiment for validation for genes down-regulated as by heat stress.

pooled cDNA sample, but only with the *RcAAT* containing plasmid dilutions and therefore for RT-qPCR, the cDNA concentration is expected to be too low for (efficient) amplification (Cq levels started at 26,6 compared to ~21-24.5 for the other genes) .

For *RcTF2* the primer was functional for the pooled cDNA samples. To confirm the different trend in expression, the RT-qPCR experiment for *RcTF2* was repeated. The results of both replicas of the RT-qPCRs for *RcTF2* were very similar. It is not expected that the different expression patterns are a result of low concentration of the target gene in the cDNA samples since the Cq values were only one cycle higher (Cq values started 25.3 compared to ~21-24.5 for the other genes Discrepancies between microarray and RT-qPCR data often arise as a result of inherent pitfalls and the large variability between technical procedures (e.g. with different qualities of RNA as a result) (Morey et al., 2006) (figure 5).

Conclusion

In conclusion the microarray data shows that *RcMBF, RcHSP1, RcHSP2* and *RcHSP3* are up-regulated in seeds germinating at 35°C in comparison to seeds germinating at 20°C for RP as well as R2. The genes *RcAAT, RcTF1, RcTF2, RcZFP1* and *RcZFP2* all show reduced expression in seeds germinating at 20°C in comparison to seeds germinating at 35°C. Although the exact values are impossible to compare between the two techniques, the trend of gene expression of the genes was similar for the majority of results of the microarray and the RT-qPCR experiments. In summary, for all genes the relative expression pattern of the data of the microarray was validated by the pattern in the data of the RT-qPCR experiment except for *RcAAT* and *RcTF2*. The high similarity between the microarray data and RT-qPCR data highlights the robustness of the responses that were observed. If the difference in gene expression is clear for different germination states (RP and R2) and perceivable with these two distinct methods it validates the quality of the results. It is therefore established that the selected genes are proper candidates to be involved in the response of the plant to heat stress.

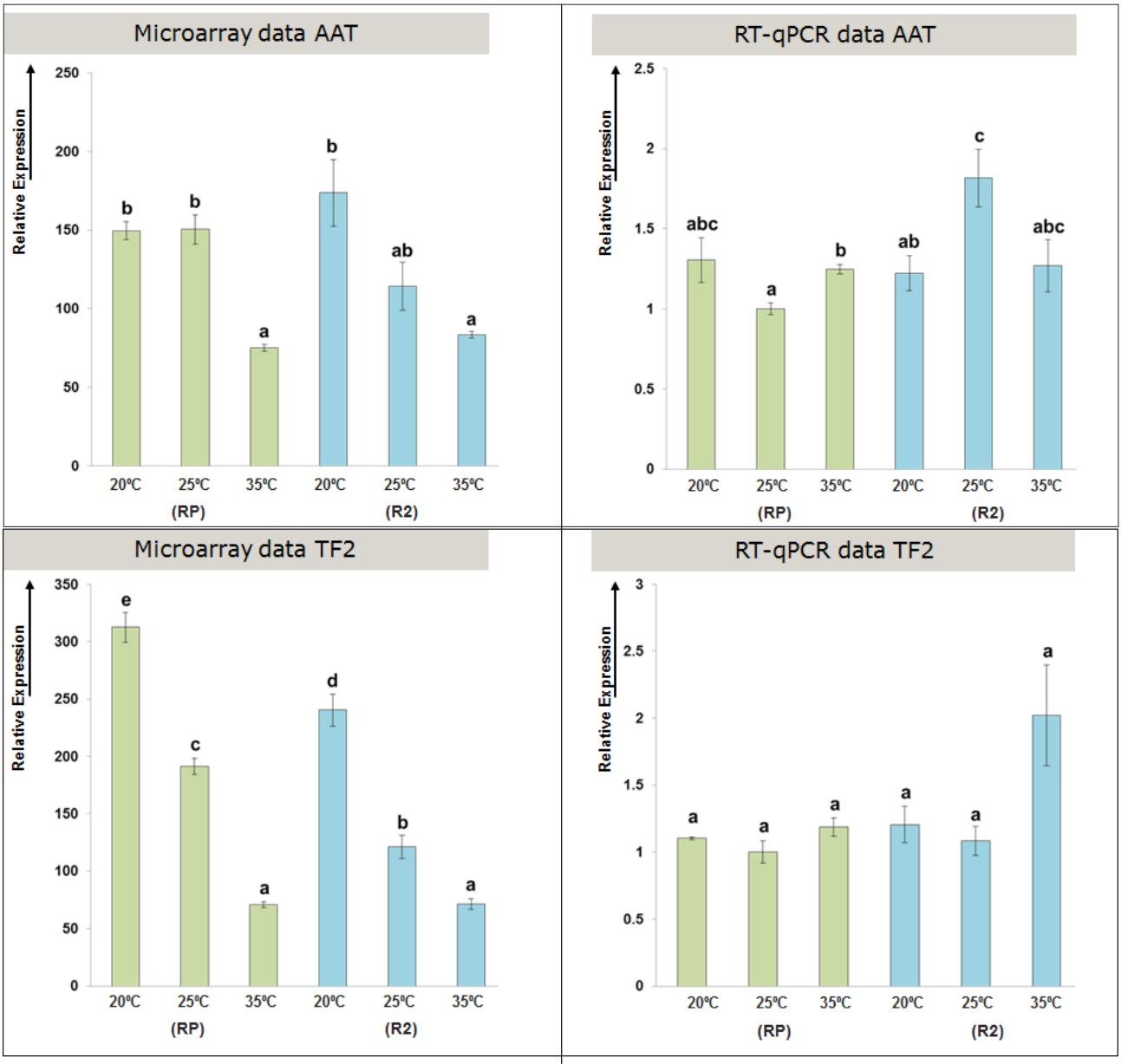

Figure 5: Results of the expression patterns found with the microarray experiment and with the RT-qPCR experiment for validation of genes down-regulated in the microarray experiment

Production of *A. thaliana* transformants overexpressing the *R. communis* genes

Cloning of the selected *R. communis genes*

The sequences for the *R. communis* amino acid transporter (*RcAAT*), heat shock proteins (*RcHSP*), multiprotein bridging factor (*RcMBF*), transcription factors (*RcTF*) and the zinc finger 2 (*RcZFP2)* were all cloned identical to the reference sequence described by the NCBI (Pruitt et al., 2005)? with two exceptions (Appendix 2). Firstly, the cloned *RcZFP*1 gene has one base pair substitution (bp 33, Adenine (A) for Thyrosine (T)), however, this substitution does not give rise to a different amino acid sequence for the RcZFP1 protein. Secondly, there is also one remaining conflict in the sequence cloned for *RcAAT (*bp 771, cytosine (C) for adenine (A)). This conflict does lead to a substitution in the amino acid translation (amino acid 257) of a histidine (H) into a glutamine (Q) in the cloned sequence (Appendix 3). Whereas histidine has positive electrically charged side chains, the side chains of glutamine are not polar. This is not expected to have any implication because this substitution has been shown to occur in this conserved domain of homologs of the *RcAAT* gene as well, according to the ClustalX2 amino acid sequence alignment for all the genes in the three (Appendix 3).

Nevertheless, it was decided to continue with this line since this sequence change was assumed to be an allelic difference. This was decided after it was observed that another cloned plasmid of *Rc*AAT showed the exact same conflict (Appendix 3).

Transformation of *A. thaliana* and selection of homozygote lines

Ten *A. thaliana* lines were successfully transformed overexpressing the nine cloned candidate genes and the *RcGK* (Mutimawurugo, 2014). For all ten lines the T1 selection procedure was repeated until approximately 20-50 transformants could be selected for each target gene (table 4). By designing and applying the transplantation method (with a phase of growth on rock wool in high humidity and lowering the humidity for hardening of the seedling) the survival rate of the transformants was very high (approximately 90%). All these lines were tested for being a single insert line in the T2 selection. The target number of 10 single insertion lines was reached for *RcAAT*, *RcHSP3, RcMBF* and *RcZFP1*.

For the other lines less single insert lines were selected ranging from one to nine selected single insertion lines with the 75/25 resistant/susceptible ratio. For the transcription factors many of the transformants were sterile and therefore could not be tested in the T2 selection. For the *RcHSP1* transformants (although many plants were transformed) only 2 lines were single insert lines. New T1 selection for all lines with less than 10 single insertion lines is advised. The T3 selection for all lines (except for *RcMLS*) has not yet been completed, the T2 plants are currently growing.

For the *A. thaliana* lines overexpressing *RcMLS*, the homozygote selection was completed. Of the 15 transformed plants that were selected in the T1 selection, two showed to be single insert lines in the T2 selection. After transplanting 40 of the transformants, 13 homozygous lines could be selected that showed 100% transformed progeny (table 4). It was possible to phenotype the T4 seeds harvested from these 13 lines within this research.

Table 4: Selection of homozygous single insertion transformants. The number of selected plants for each of the tested generations (T1 – T3) is given. For T1 all surviving transformants were selected, for T2 single insert lines (showing a 75/25 resistance/ sensitive ratio) were selected. For the T3 selection homozygous lines with a 100% survival rate were selected.

Gene	T1	T2 (75/25%)	T3 (100%)
RcTF1	36	5	-
RcZFP1	38	10	-
RcZFP2	26	1	-
RcHSP1	23	2	-
RcHSP2	48	9	-
RcHSP3	46	10	-
RcAAT	43	10	-
RcTF2	34	5	-
RcMBF	54	10	-
RcGK	20	3	-
RcMLS	15	2	13

Application of transgene ectopic overexpression in *A. thaliana* of a gene of interest is a technique which is very useful for its functional characterization (Ahrazem et al., 2015). Currently, this technique is widely used in functional genomics research and it has been applied for *R. communis* genes before, as well as for genes related to thermotolerance (Sakamoto et al., 2004, Rolletschek et al., 2005, Kim et al., 2014, Ahrazem et al., 2015, Prändl et al., 1998). What should be taken into account assessing the phenotypes is that functional redundancy might play a role for the *R. communis* genes that showed to be up-regulated during heat. Homologs of Arabidopsis are probably functional and up-regulated too. If overexpression of these genes does improve the heat tolerance in *A. thaliana*, the overexpression of these genes in *R. communis* might positively influence *R. communis* heat tolerance in similar ways.

Insertion mutagenesis

Phenotypes caused by disturbance of a gene as a result of insertion rather than overexpression of the target gene (insertional mutagenesis) was observed in one of the *A. thaliana* single insertion lines overexpressing *RcMLS.* This was recognized in the T4 generation, where this insertional mutagenesis phenotype was only observed for one of the two single insert lines (figure 6). This phenotype was expressed by only 42% of the plants within this line and was not at all present in the other *A. thaliana* single insertion lines selected for *RcMLS* (Appendix 4). Remarkably, this phenotype was not consistent over the whole insertion line, some plants showed the phenotype within this line whereas others didn't. This is expected to be a result of penetrance of the genotype which means that only a proportion of individuals that carrying this insertion mutagenesis actually expresses the mutated phenotype (Griffiths, 2005).

Figure 6: Insertion phenotype of A. thaliana overexpressing RcMLS. Within this single insertion line 42% of the plants showed this dwarfing phenotype with excessive branching and increased production of siliques.

Heat stress threshold experiment

Before testing the thermotolerance phenotypes of the *A. thaliana* lines overexpressing the *R. communis* target genes, thermotolerance in wild type *A. thaliana* (Col-0) had to be described, therefore this heat stress threshold experiment was conducted.

For seeds germinating in 22°C, 31°C and 33°C the total germination was statistically the same: 100%. Germination speed, measured by the time it took for 50% of the seeds to germinate (T_{50}) was significantly different for all temperatures. At 31°C germination was the quickest with a T_{50} value of 19.3 hours. Thereafter germination was quickest at 33°C with a T_{50} of 21.9 hours followed by the control temperature of 22°C with 25.2 hours (figure 7). At temperatures above 33ºC the germination speed was lower and the maximal germination percentage decreased. At 34°C total germination had lowered to 92% with a T_{50} of 31 hours, at 35°C the germination percentage decreased to 43% with a T_{50} of 34.6 hours. At temperatures of 36°C or above none of the seeds germinated, not even after relocation to 22°C for a week (figure 7).

Temperature is known to have very important influences on seed dormancy and germination. The relation between temperature and seed germination is described by the mathematical thermo time model (Alvarado and Bradford, 2002, Bewley and Black, 1994). Within this model two temperature thresholds are described; the base temperature (T_b) which is the lowest temperature at which germination can occur and the ceiling temperature (T_c) which is the highest temperature at which germination can occur. From the results of the heat stress threshold experiment it can be concluded that for *A. thaliana* the T_c threshold is 35°C since at this temperatures still some seeds germinate, whereas in 36°C they do not at all. This model also describes that germination speed is lower in sub-optimal temperature conditions (Bewley and Black, 1994). However, what should be taken into account is that T_b and T_c are dependent on dormancy levels (Alvarado and Bradford, 2002). It is therefore essential that all plants are in the same conditions and all seeds are harvested simultaneously for the thermotolerance phenotyping experiments for the *A. thaliana* lines overexpressing the *R. communis* candidate genes.

Additionally, it was shown that the water potential of the seeds is affected by supra-optimal temperatures in the hydrothermal time model (Alvarado and Bradford, 2002). The water potential of the seeds is high when they are dry and lowers during imbibition to facilitate germination and seedling growth. When seeds are germinating under high temperatures the water potential remains higher, which could be an explanation for reduced germination rates and percentages (Alvarado and Bradford, 2002). This could also contribute to the explanation for the finding that in *R. communis* seedlings suffered most damage from heat stress when the seeds are subjected to high temperatures during their imbibition (Mutimawurugo, 2014, Ribeiro de Jesus et al., 2014a)

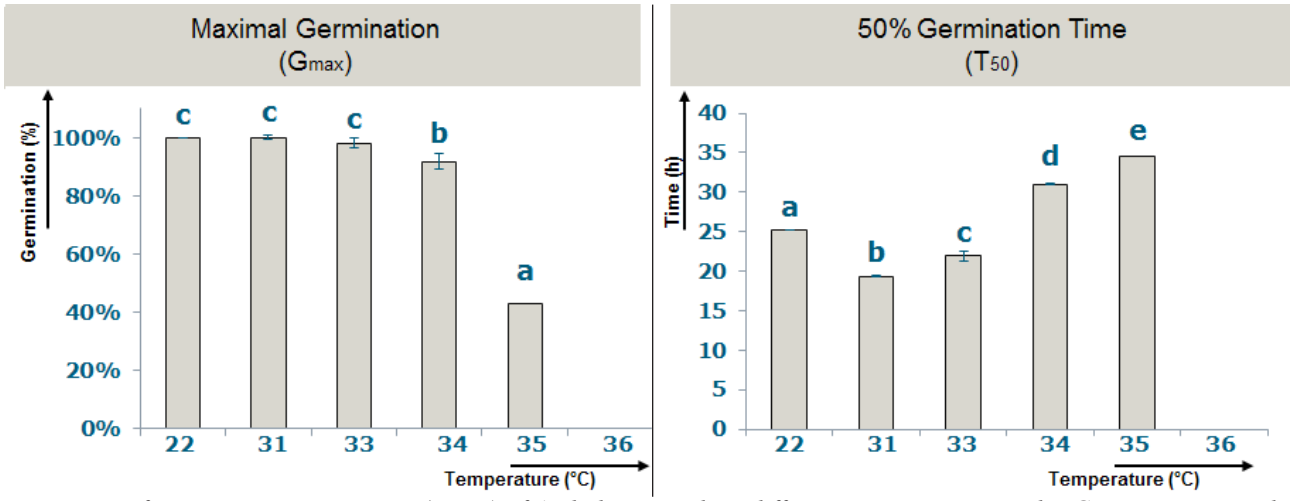

Figure 7: Left: Maximum germination (gMax) of A. thaliana seeds at different temperatures. Right: Germination speed of A, thaliana seeds at different temperatures measured by the time until 50% germination was reached (T50)

Life Cycle Pilot

When the threshold temperatures for proper germination were established it was tested if the seeds that did germinate were able to complete their lifecycle under those environmental conditions. Seedling establishment was tested for both continuous heat stress and translocation to optimal temperatures after germination. None of the seeds germinated at 34°C and 35°C could establish seedlings in continuous heat stress (table 5). When seeds were translocated to optimal temperatures after germination in heat stress conditions 34°C and 35°C, 90% respectively 80% of the seedlings were able to establish. Approximately 50% of the plants that were transplanted from 34°C and 35°C showed a period of slower growth after transplantation.

Table 5: Lifecycle experiment with the number of plants surviving under continuous heat and after transfer to more optimal temperatures.

Incubator temperature	# of Survivors - Complete lifecycle in incubator	# of Survivors - Transferred to B9 after germination
22°C	10/10	10/10
34°C	0/10	9/10 (one small)
35°C	0/10	8/10 (one small)

From this pilot it can be concluded that seeds germinated under the threshold temperatures are not damaged too much and can still form proper seedlings, but only when they are transfered to more optimal temperatures. In *R. communis* seedlings were also heavily affected at this temperatures and were not able to establish themselves under continuous heat stress.

A difference between *A. thaliana* and *R. communis* seedlings is that after germination under high temperatures, Arabidopsis seedlings are still able to establish themselves after transfer to colder temperatures and complete their lifecycle whereas *R. communis* seedlings are not (Mutimawurugo, 2014). This difference can maybe be explained by the cold imbibition of the *A. thaliana* seeds during the necessary cold treatment whereas the *R. communis* seeds also imbibe in high temperatures. Especially since high temperatures mostly affects germination and seedling establishment of *R. communis* when applied during imbibition (Ribeiro de Jesus et al., 2014a)

Phylogenetic analysis of gene function

An in-silica analysis of the proteins encoded by the *R. communis* candidate genes was performed by means of phylogenetic research in order to have a basic understanding of the putative functions of the proteins.

HMM-scan results and prediction of subcellular domains

For most proteins their putative annotation agreed with the domains that gave the highest hits in the HMM-scan. Additionally, the expected subcellular localization was predominantly in agreement with the Cello subcellular predictions. Cello is a bioinformatics tool to predict the localization of the proteins based on the biochemical properties (for example polarity and acidity) of its amino acids (Yu et al., 2006). If the expected localization of the protein (based on its function) matches the Cello predicted sub-cellular localization it means that the putative function of the protein is supported by the biochemical properties of the protein itself.

The four candidate proteins for which the genes were up-regulated in *R. communis* seeds germinating at 35°C in comparison to 20°C and 25°C were RcMBF with as highest HMM-scan hit the multiprotein bridging factor 1 domain with localization in the cytoplasm and mitochondria, and the *R. communis* heat shock proteins (RcHSPs) for which the predicted subcellular locations were the chloroplast (RcHSP1), the cytoplasm (RcHSP2 and RcHSP3) or extracellular (RcHSP2) (table 6).

For the second group of proteins the expression of the genes encoding these proteins were down regulated in *R. communis* suffering from heat stress. Within this group three transcription factor proteins were present; RcTF1, RcTF2 and RcZFP2 which all had a predicted nuclear localization. More specifically RcTF1 was found to be a No Apical Meristem (NAM) protein and RcTF2 had highest hits with domains of the histone-like transcription factor (CBF or Nuclear factor-Y (NF-Y)). The HMM-scan results for RcZFP2 classify the zinc finger as the C_2H_2 type (table 6).

Although RcZFP1 was annotated as putative zinc finger protein the results of the HMM-scan as well as the sub-cellular

Table 6: HMM-scan hits (identifier and description) and the Cello Subcellular prediction for the target *R. communis* genes.

Protein	HMM Identifier	HMM Description (top hit)	Cello Prediction
RcMBF	MBF1	- Multiprotein bridging factor 1	Cytoplasmic, Mitochondrial
RcHSP1	HSP20	- Hsp20/alpha crystallin family	Chloroplast
RcHSP2	HSP20	- Hsp20/alpha crystallin family	Cytoplasmic, Extracellular
RcHSP3	HSP20	- Hsp20/alpha crystallin family	Cytoplasmic
RcTF1	NAM	- No apical meristem (NAM) protein	Nuclear
RcTF2	CBFD_NFYB_HMF	- Histone-like transcription factor (CBF/NF-Y) and archaeal	Nuclear
RcZFP2	Zf-C2H2_6	- C2H2-type zinc finger	Nuclear
RcZFP1	Lipase_GDSL	- GDSL-like Lipase/Acylhydrolase	Vacuole
RcGK3	FGGY_N	- FGGY family of carbohydrate kinases, N-terminal domain	Cytoplasmic, Mitochondrial
RcMLS1	Malate_synthase	- Malate Synthase	Peroxisomal
RcAAT	Aa_trans	- Transmembrane amino acid transporter	Plasma membrane

prediction oppose this assumption. RcZFP1 was predicted to be a GDSL lipase with as predicted subcellular localization the vacuole. RcGK had highest similarity with the FGGY family of carbohydrate kinases domain and is predicted to be localized in the cytoplasm or mitochondria. RcMLS has been found to be a malate synthase located in peroxisomes. RcAAT1 was found to be a transmembrane amino acid transporter (table 6).

The results of the HMM-scans were confirmed by building the phylogenetic trees for these proteins and their homologs. For some genes these HMM-scan results of functional domains together with the subcellular location prediction explain the branching structure of the phylogenetic tree. For other genes more information about protein homology to *A. thaliana* (to form homology groups) was needed to understand the structure.

These 11 genes are all expected to be involved in processes of seedling establishment, heat response signalling pathways or both. Genes with high levels of expression under heat stress conditions are hypothesized to be involved in the heat stress response, whereas genes with low levels of expression under heat stress are hypothesized to be involved in (seedling) developmental pathways. Down-regulation of the expression of these genes could therefore contribute to malformation and the inability to establish of *R. communis* seedlings. Phenotyping experiments for thermotolerance of the *A. thaliana* lines overexpressing the candidate genes will be needed to see if overexpression of the genes encoding these proteins could improve the thermotolerance of seedlings or not. Until then, the functional analysis by phylogenetic research will provide insight in the possible influence of these genes on thermotolerance. For this, the two types of thermotolerance have to be taken into account. First, there is the basal thermotolerance, which deals with continuous heat and secondly there is acquired thermotolerance. Basal thermotolerance is most applicable to the situation of high temperature germination of *R. communis* tested by Ribeiro de Jesus et al. (2014a). One of the genes that could be important to improve the basal thermotolerance, is the multiple protein bridging factor (MBF or MBF1).

Basal thermotolerance: MBF

(MBF1) – gene identity and function

For RcMBF and all its protein homologs, the HMM-scan results were clear, all had highest hits with the multiprotein bridging factor 1 domain (MBF1) (PF08523.5). Additionally RcMBF (and 16 out of 18 homologs) had significant hits with a helix-turn-helix domain in their sequence, subsequent to the MBF1 domain. A helix-turn-helix domain is a structural motif recognized by two alpha helixes that are joined together by a short strand of amino acids. According to the HMM-scan, the helix-turn-helix motif found in these proteins is mostly similar to those of the superfamily of λ-repressor like DNA binding domains. This structural motif is generally associated with the binding of DNA in a process in which the second helix of the motif is proposed to be most involved in operator DNA recognition by binding within the grooves of the DNA (Brennan and Matthews, 1989). GO-terms associated with this super family of λ-repressor like DNA binding domains are 'transcription co-activator activity', sequence-specific 'DNA binding transcription factor activity' and 'RNA polymerase II transcription co-activator activity' in the field of molecular function (Sillitoe et al., 2012).

MBF1 proteins were found to have transcriptional (co)activator functions that are evolutionarily conserved for a broad range of species. A considerable degree of sequence homology was found between sequences of *MBF1* genes of plants species (including *A. thaliana*, *R. communis* and *Z. mays*), yeast (*S. cerevisiae*) and animals (including *H. sapiens)* (Takemaru et al. (1997). This sequence homology, together with the presence of the DNA binding domain found in the RcMBF protein sequence strongly support the transcription regulatory function for this *R. communis* gene.

In the phylogenetic tree constructed with RcMBF proteins three main branches can be distinguished (figure 8). One branch with nuclear, cytoplasmic and mitochondrial predicted locations and two branches of proteins with only nuclear localization (that both belong to the same homolog group). RcMBF falls within the branch with the three different sub-cellular locations, together with homologs of all other researched species. Two Arabidopsis proteins AT19661131 (*AT3G24500.1*) and AT19661132 (*AT3G24500.2*) are closely related to the *R. communis* MBF and are described as two splicing forms of the AtMBF1c protein. According to the homology search, all the other proteins in this branch are homologs of this AtMBF1c protein. However, the RcMBF protein seems to be missing the nuclear localization domain. Looking at the reliability values of the Cello prediction the cytoplasmic (1,890) and mitochondrial (1.439) localizations are significantly higher than for the nucleus (1.136), but the nuclear localization is still higher than the reliability values of the other possible sub-cellular locations for which the reliability values ranging from 0.010 to 0.256 (Appendix 5).

For the other two branches (with only nuclear localization), one is made up out of *A. thaliana* proteins and the other of *M. esculentum*, *P. trichocarpa* and *O. sativa* proteins. Within these two branches are (homologs of) *A. thaliana* proteins AtMBF1b (*At19659923*) and AtMBF1a (*At19642090*). There is no clear distinction between the 1a and 1b group of proteins within this phylogenetic tree. This shows that intraspecific difference between the MBF1a and MBF1b are smaller than the interspecific differences between the sets of proteins.

The nuclear localization for all RcMBF homologs was expected because of their proposed function as transcriptional (co) activator. Additionally, the predicted localization of some homologs is confirmed by literature since it was shown that AtMBF1 proteins localize in the nucleolus (Sugikawa et al., 2005). The main function of the nucleolus is determined to be the transcription and processing of ribosomal RNA genes. Additionally the nucleolus is also implicated to function in other processes such as cell cycle control, mRNA maturation and stress responses (Olson et al., 2000, Rubbi and Milner, 2003). So it can be assumed that the multiple bridging factor has a function within processes regulated by the nucleolus (Sugikawa et al., 2005).

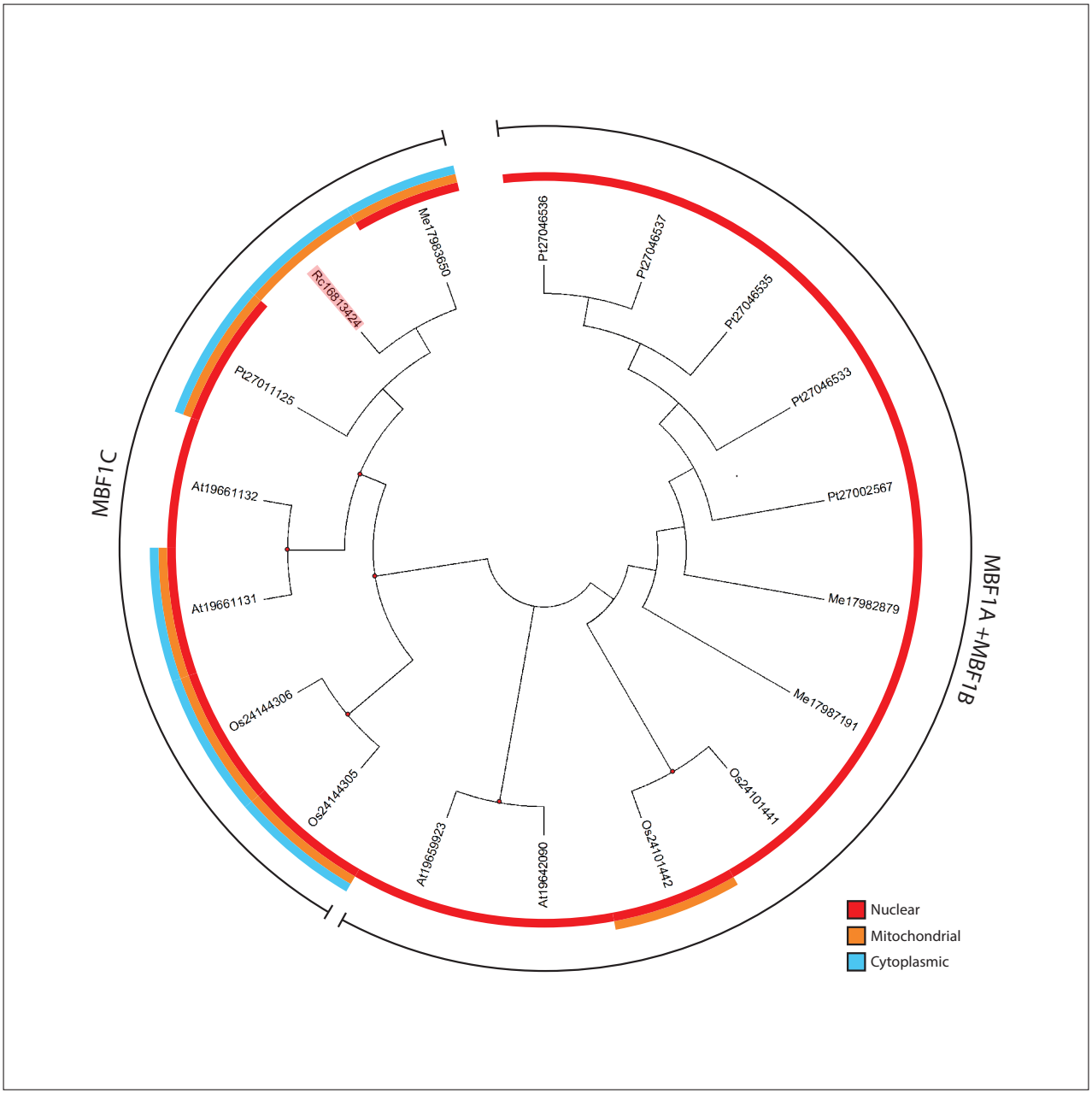

Figure 8: Phylogenetic tree constructed for the RcMBF protein. The phylogenetic tree depicts the homology and the high certainty branches (>95% after 500x bootstrapping) by the red dots. The subcellular prediction for each gene is represented by the coloured rings. Homology groups are depicted in the outer ring.

MBF – Relation to heat tolerance

Generally it is accepted that the MBF1c protein is very different from the MBF1a and MBF1b proteins. This was concluded from an alignment of the three different *MBF* genes (*MBF1a,-b and –c)* for many different species including *A. thaliana,* and *O. sativa* (Tsuda and Yamazaki, 2004). Additionally, the observed expression pattern for each *AtMBF* was distinct, although the expression patterns of *AtMBF1a* and *-1b* are more similar. For *AtMBF1c (*the closest homolog of RcMBF) expression is observed in all vegetative tissues and can be elevated by different stresses including salinity, drought, heat, hydrogen peroxide or application of salicylic acid or abscisic acid (Tsuda and Yamazaki, 2004, Tsuda et al., 2004, Suzuki et al., 2005). Expression patterns for *AtMBF1a* and *AtMBF1b* are specific for developmental stages and different tissues (Tsuda and Yamazaki, 2004, Suzuki et al., 2008, Sugikawa et al., 2005). The main function of AtMBF1a and AtMBF1b is described as regulators of development controlling cell cycle and leaf expansion, which also correlates to their presence in the nucleolus (Tojo et al., 2009, Olson et al., 2000).

The function proposed in literature for the MBF1c protein is regulator in stress tolerance responses, in particular heat (and drought) stress. For example, MBF1c is shown to be involved in the response to wounding stress or treatment with an ethylene precursor in potato (Godoy et al., 2001), in the response to heat shock of *Retama raetam* (Pnueli et al., 2002) and in the reaction to combined drought stress and heat shock in tobacco (Rizhsky et al., 2002). These findings are also in line with the results of the microarray experiment conducted for the selection of the candidate genes that showed significant up-regulation of *RcMBF* in seeds germinating under heat stress conditions.

MBF1c accumulates quickly during heat stress and is then localized from the cytosol to the nucleus (Suzuki et al., 2008). This re-localization is in accordance with the subcellular localization prediction of this protein placing these proteins in the nucleus as well as in the cytosol. After its movement to the nucleus, AtMBF1c is enabled to conduct its function as transcriptional co-activator. Moreover, AtMBF1c is actually shown to be a key regulator for basal thermotolerance in *A. thaliana* (Suzuki et al., 2008). Seedlings overexpressing *MBF1c* are proven to be more tolerant in experiments testing basal thermotolerance for 48 hours at 38°C and at 45°C (Suzuki et al., 2005, Suzuki et al., 2008). MBF1c is not required for the acquired thermotolerance response including Heat Shock Transcription Factors (HSF) and the expression of heat shock proteins (Suzuki et al., 2008). Additionally, germination and seedling development was improved in Arabidopsis seedlings overexpressing *AtMBF1c* under oxidative stress (being one of the effects of heat stress), even without functional MBF1a and -1b proteins (Arce et al., 2010).

The recognition of heat stress, upstream in the MBF1c signal transduction pathway, is proposed to work with ROS signalling (Miller et al., 2008). ROS molecules are known as toxic components reacting with cell components and cause considerable damage to the cell homeostasis. As a counter-reaction the plant is able to produce ROS scavenging molecules (antioxidants) that interact with the ROS molecules and inactivate them. However, besides the detrimental functions of ROS molecules, they act in abiotic stress signal transduction pathways too (Mittler et al., 2011, Miller et al., 2008). Genes associated with this network are called the ROS gene network which includes over 150 genes in *A. thaliana* (Mittler et al., 2004, Kotak et al., 2007). Acccumulation of ROS is proposed to be sensed by the heat shock factor HsfA4 which then (either directly or indirectly) stimulates *MBF1c* expression. This hypothesis is supported by the findings that ROS species increase expression of *AtMBF1c* (Tsuda and Yamazaki, 2004, Suzuki et al., 2008).

After up-regulation of the *MBF1c* gene and translocation of the MBF1c protein to the nucleus, thermotolerance pathways are stimulated. There is still some insecurity about the exact mode of action of MBF1c in these thermotolerance pathways. Proposed pathways involve the MBF1c protein as regulating factor with an upstream function in salicylic acid (SA), ethylene, abscisic acid (ABA) and trehalose signalling pathways (figure 8) (Suzuki et al., 2008, Arce et al., 2010). Literature supports the involvement of all of these factors within basal thermotolerance and underlies the complexity and the involvement of interaction between their signal transduction pathways. Transformants overexpressing *AtMBF1c* show increased levels of both SA and trehalose(Suzuki et al., 2008), whereas *AtMBF1c* mutants (*mbf1c)* are unable to accumulate these signalling molecules during heat stress and show increased basal thermosensitivity. This thermosensitivity can be partially rescued by applying exogenous SA or trehalose (Suzuki et al., 2008).

For the phytohormone salicylic acid it is known that it is an important signalling hormone in defence responses of plants to pathogens; it regulates the systemic acquired resistance (SAR). However, the function of SA in regulation of responses to abiotic stresses has been acknowledged as well. For heat stress it was shown that SA is involved in the basal thermotolerance of all developmental stages during heat shock and in the subsequent recovery (Clarke et al., 2004, Larkindale et al., 2005a). The exact signal cascade for this thermotolerance is unknown but is shown to involve PR-1 (Clarke et al., 2004, Miller et al., 2008, Suzuki et al., 2008). Therewith, there is crosstalk in the signalling cascade of SA to address both types of stress (biotic and abiotic) (Fujita et al., 2006, Kotak et al., 2007, Larkindale et al., 2005a, Clarke et al., 2004). The mode of action of salicylic acid is either to influence the activity of enzymes such as catalase involved in the thermotolerance or to induce genes involved in thermoprotective mechanisms (Horváth et al., 2007). These thermoprotective mechanisms include induction of changes in the antioxidant system for protection against oxidative damage and stabilization of membrane properties to minimize electrolyte leakage (Clarke et al., 2004).

Arabidopsis accumulates trehalose during heat stress. Trehalose is thought to play an important role as signalling molecule for abiotic stresses (Avonce et al., 2004, Paul et

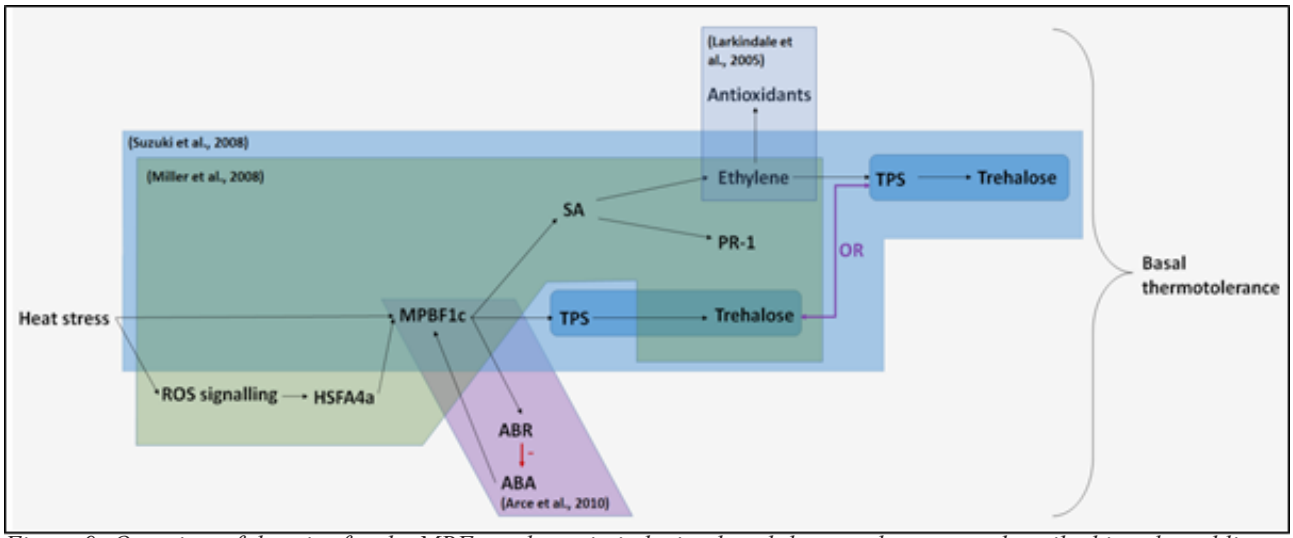

Figure 9: Overview of theories for the MBFc pathway in inducing basal thermotolerance as described in selected literature.

al., 2008). Trehalose is able to protect proteins and membranes from denaturation by replacing water with its hydrogen by forming bonds with polar residues of the proteins and membranes (Paul et al., 2008). A relation was found between AtTPS5 from the trehalose synthesis pathway and basal thermotolerance (Suzuki et al., 2008). TPS (Trehalose-6-Phosphatase Synthase) is an enzyme catalyzing the early steps of the transformation from glucose to trehalose (Paul et al., 2008). AtMBF1c is able to elevate the expression of *AtTPS5* which subsequently results in accumulation of trehalose. In the microarray experiments used to select genes for this research also a *RcTPS* (trehalose-6-phosphate synthase) was found with elevated transcription under heat stress.

Furthermore, relations are established between the ABA signalling pathways and AtMBF1c proteins. It was found that the AtMBF1c proteins elevate expression levels of the *AtABR* transcription factor, which is a member of the *AP2/ERF* transcription factor super family (encoding an APETALA2 (with AP2 domain) (Arce et al., 2010) (figure 9). The AP2 family is known for repression of ABA responses (Pandey et al., 2005). ABA is a phytohorme that is involved in early seedling development, vegetative growth and adaptation to a range of abiotic stresses. However, its function within the pathway for establishment of basal thermotolerance is still unclear. ABA mediates SA responses against biotic stresses and works almost antagonisticallly with ethylene in many different developmental processes (Fujita et al., 2006). Therefore, it could be argued that its down-regulation would be necessary for the basal thermotolerance establishment. Exogenous ABA, however, elevates expression of AtMBF1c in *A. thaliana* (Tsuda and Yamazaki, 2004). This up-regulation could explain the findings of Larkindale et al. (2005a) who find that ABA has a positive infuence on basal thermotolerance.

Crosstalk is expected between ABA and ethylene pathways. The ethylene-response signal transduction pathway is regulated or perturbed downstream of AtMBF1c as well. *AtMBF1c* overexpressing plants were shown to have elevated levels of ethylene-response-binding factors and the ethylene biosynthesis enzyme 1-aminocyclopropane-1-carboxylic acid (ACC) synthase (Suzuki et al., 2005).

All in all, it is obvious that there is yet much to be discovered within this complicated network of mutual regulation for establishment of thermotolerance. What is clear is that there is an important role for the MBF1c transcription factor, which might function as a regulatory component for the cross-talk between the different basal thermotolerance signalling pathways during germination and early developmental stages of seedlings in Arabidopsis and presumably in *R. communis* too (Arce et al., 2010).

Acquired thermotolerance: Heat shock proteins (HSP).

Besides basal thermotolerance, plants also fight damage by heat stress with acquired thermotolerance. This tolerance is acquired after a period of non-lethal heat which prepares plants for possible higher temperatures in the future. The acquired thermotolerance pathway involves heat shock factors (transcription factors) and heat shock proteins (Burke et al., 2000, Larkindale et al., 2005d, Larkindale and Vierling, 2008, Kotak et al., 2007). Three heat shock proteins were selected as *R. communis* candidate genes.

The heat shock response is one of the most conserved genetic systems known, existing in every living organism (Lindquist and Craig, 1988). These proteins are (as expected by their name) predominantly involved in thermotolerance response pathways but several heat shock proteins are also induced by a variety of other stresses and developmental processes (Lindquist and Craig, 1988, Vierling, 1991). Heat shock proteins range in their molecular mass from 15 to 110 kDa. Their molecular weight is used to devide the heat shock protein in several classes. Some of the major classes are HSP110, HSP90, HSP70, HSP60 and small heat shock proteins (sHSP) (Vierling, 1991, Kregel, 2002). This last group exists of heat shock proteins with a molecular weight ranging from 15 to 30 kDa within plants (Mansfield and Key, 1987, Vierling, 1991). The group of small heat shock proteins is very diverse, especially in their sequences. However, the small heat shock proteins are conserved with respct to their structural properties (Lindquist and Craig, 1988).

All three genes encoding heat shock proteins that were cloned within this research belong to the class of small heat shock proteins. More specifically, according to the HMM-scan, all *R. communis* candidate genes selected in this research belong to the same HSP-20 like superfamily with (~20 kDa) proteins containing an alpha crystallin domain. Like all small heat shock proteins, HSP20 proteins share the property of forming highly polymeric (multi-meric) structures: heat-shock granules (Lindquist and Craig, 1988, Horwitz, 2003). It is known that these multimeric structures can function as molecular chaperones to protect other proteins from denaturation and aggregation (Groenen et al., 1994). Interaction of these heat shock proteins with their substrate occurs by binding to its hydrophobic sites, to prevent denaturation of the substrate because of heat (Scharf et al., 2001). This is verificated by the GO-terms associated with the HSP20/alpha crystalline that include ‘response to heat’ (biological process), ‘unfolded protein binding’ (molecular function) (Madera et al., 2004, Van Aken et al., 2009).

HSP functions in acquired thermotolerance

The three selected heat shock proteins could all be represented in the same phylogenetic tree because they were all HSP20 proteins (figure 10). The three genes are devided over separate branches of the phylogenetic tree, each with their own distinct subcellular location prediction pattern. RcHSP1 is located in a branch with HSP20 proteins that are mostly located in the chloroplasts or mitochondria, whereas RcHSP2 is located in a branch predominantly located in the cytosol, but with a lot of other localization domains that could place these proteins in other locations such as the nucleus, the mitochondria, chloroplasts or even extracellular. RcHSP3 in its turn is located in a branch that is almost exclusively predicted to be in the cytoplasm.

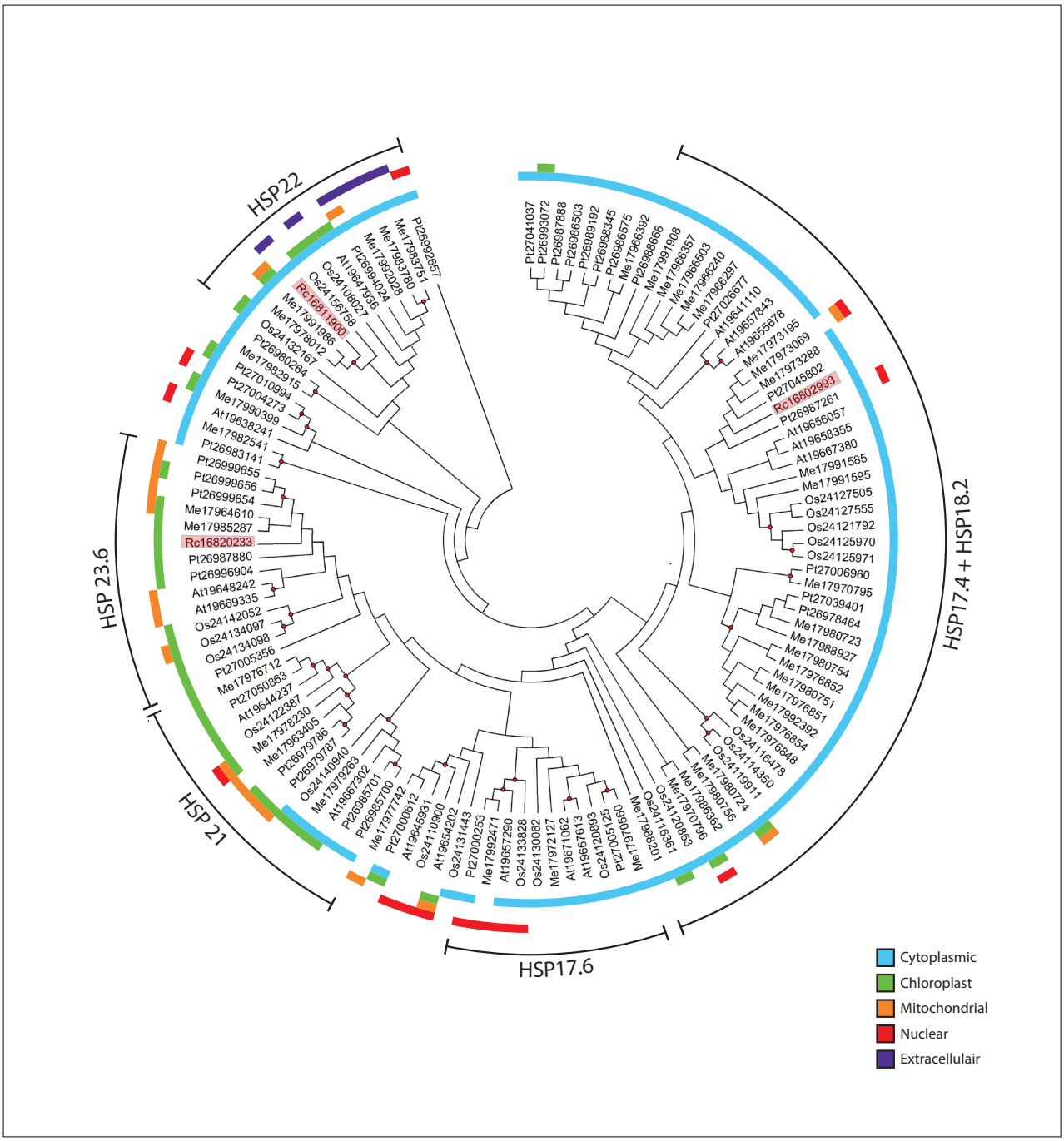

Figure 10: Phylogenetic tree constructed for the RcHSP proteins. The phylogenetic tree depicts the homology and the high certainty branches (>95% after 500x bootstrapping) by the red dots. The subcellular prediction for each gene is represented by the coloured rings. Homology groups are depicted in the outer ring.

Within this phylogenetic tree several homolog groups with sub-families of the HSP20 super family could be discovered. The three RcHSP proteins each fall within a separate sub-family. Firstly there is the Rc16820233 (RcHSP1) protein and its homolgs which are most similar to the mitochondrion-localized small heat shock protein 23.6 (AT4G25200) or an Arabidopsis protein very closely related to AtHSP23.6; AtHSP23.5 (AT5G51440). This findings are in agreement with the predicted localization for this branch since these proteins are dual targeted and can be found within the mitochondria and chloroplasts (Macherel et al., 2007, Scharf et al., 2001, Van Aken et al., 2009, Wang et al., 2004).

The next heat shock protein, Rc16811900 (RcHSP2), and its closest relatives have the highest similarity to the *A. thaliana* heat shock protein 22 (*AT4G10250*). Rc16802993 (*Rc*HSP3), which is located within the branch that exists of homologs of AtHSP17.4 (*AT3G46230*), AtHSP18.2 (*AT5G59720*) and an undescribed *A. thaliana* heat shock protein (*AT1G53540*). The closests *A. thaliana* homolog of *Rc*HSP3 is the AtHSP17.4 protein.

Significanlty elevated levels for genes encoding heat shock proteins 17, 18, 21, 22, 60, 70, 100 and 101 were found in *A. thaliana* under heat stress (Rizhsky et al., 2004). This validates the elevated levels of expression of *RcHSP2* and *RcHSP3* because this group includes their closest homolog. However this is not the case for *RcHSP1 (*homologous to the

AtHSP23.6). More specifically, the expression of *AtHSP17.4, AtHSP18.2* and *AtHSP17*.6 were found to be induced after a heatshock of 30°C and highly elevated at 37°C. For *AtHSP21* expression was only observed in plants subjected to 37°C. None of these genes was expressed at the control temperature (Prändl et al., 1998). From these results it can be postulated that these close *A. thaliana* homologs of the RcHSPs are indeed involved in the response to heat.

Nevertheless, AtHSP23.5 and AtHSP23.6 as closest homologs to RcHSP1 are shown to be involved in heat tolerance as well (Van Aken et al., 2009, Macherel et al., 2007). *AtHSP23.5 (At5g51440)* and *AtHSP23.6 (At4g25200)* have a very similar expression pattern and are assumed to form a functional pair (Van Aken et al., 2009). It could even be argued, since heat shock proteins are shown to hybridize, that they need to dimerize to be functional (Stamler et al., 2005, Van Aken et al., 2009). AtHSP23.5, AtHSP23.6 and AtHSP22 are shown to be involved in the mitochondrial response to abiotic stress (Van Aken et al., 2009, Macherel et al., 2007). More specifically, they are predicted to function in seed mitochondria (Macherel et al., 2007). Seed mitochondria are descibed as mitochondria that withstand dessication and that are functional at the onset of imbibition and their integrity and performance systematically improves during germination. Oxidative damage as a result of heat can strongly affect the mitochondrial membranes inhibiting enzymes important for the functioning of the mitochondria including their ATP production, and the carbon and nitrogen metabolism. Inhibitation of these important processes by denaturation or degradation of involved enzymes can therefore decrease cell viability (Sanmiya et al., 2004, Van Aken et al., 2009, Macherel et al., 2007). An effective mitochondrial stress response is therefore essential for proper plant development. This stress response is shown to involve accumulation of HSP22, HSP23.5, HSP23.6 and a putative LEA protein (*At5g44310* or *At4g21020)* (Macherel et al., 2007, Bardel et al., 2002). Expected is that HSP22 (homolog group of RcHSP2) is involved in the heat stress as molecular chaperone as well as independently for developmental functions during seed maturation (Macherel et al., 2007, Bardel et al., 2002, Wehmeyer and Vierling, 2000).

The third *R. communis* heat shock protein, RcHSP3, is predicted to be a cytosolic heat shock protein (Löw et al., 2000). It is the gene most homologous to three different *A. thaliana* proteins; AtHSP17.4, AtHSP18.2 and AtHSP17.6, which all are grouped into one class of small heat shock proteins (C1) (Scharf et al., 2001, Löw et al., 2000) (figure 10). Proteins within this group have very high sequence similarity. Their chaperone function was established in vivo (Löw et al., 2000).

Overexpression of genes encoding sHSP from tobacco (Sanmiya et al., 2004) and maize (*ZmHSP22)* in *A. thaliana* (Rhoads et al., 2005) improved acquired thermotolerance. Furthermore overexpression of *ZmHSP22* in *A. thaliana* caused super-induction of expression of the endogenous *AtHSP23.6* gene and caused alterations in the expression of *AtHSP17.4*.

In conclusion, overproduction of heat shock proteins can improve acquired thermotolerance. Its constitutive overexpression could allow the plants to survive in otherwise lethal temperatures without the necessity of a period of sub-lethal stress. The question is raised if a sudden occurance of lethal temperatures would be true to nature, since temperatures change more gradually than in lab conditions. Furthermore, the energy and nutrient costs of producing the heat shock proteins could decrease the yield.

Transcription factors: TF1, TF2 and ZFP2

Three of the genes that show down-regulated expression in *R. communis* seeds under heat stress are classified as transcription factors (*RcTF1*, *RcTF2* and *RcZFP2*). The proteins encoded by these genes are hypothesized to interact with DNA to regulate gene expression. Lowered expression of these genes under heat could be key factors in the explanation for the disturbed development of the *R. communis* seedlings.

Focusing on the two genes annotated as transcription factors (RcTF1 and RcTF2), it is clear that they are fairly different. This is depicted in the phylogenetic tree of the transcription factors by the separation of the two genes into separate branches (figure 11). The RcTF1 protein belongs to the No Apical Meristem (NAM) transcription factors falling under the superfamily of NAC proteins whereas the RcTF2 protein belongs to the family of Nuclear Factor Y (NFY) proteins. By the large amount of proteins in this tree it can be seen that these transcription factors belong to large families of transcription factors with highly similar proteins. Despite the large number of genes, the phylogenetic three does show many high certainty branches (over 95% certainty).

As expected because of their DNA-related function, most of the transcription factors are predicted to be located within the nucleus (figure 11). Besides in the nucleus, some transcription factors are also predicted to be located in the cytoplasm, extracellular, in mitochondria or in chloroplasts. The structure of the tree can be explained best by looking at the homology groups for annotated *A. thaliana* genes. Especially for the NFY proteins, the different subunits are grouped together. For the NAC proteins, homology groups that are formed are less specific with parts of branches homologous to many different *A. thaliana* NAC proteins. Most importantly, the *R. communis* NAC protein homolog (RcTF1*)* falls in a quite well defined homology group of NAC25 (*AT1G61110)* and NAC2 (*AT3G15510*) proteins (figure 11). Of these two proteins, RcTF1 is most homologous to AtNAC25 according to the Plant Transcription Factor Database (Jin et al., 2013).

RcTF1: a NAC protein

Proteins of the NAC family all have the characteristic NAC domain of approximately 160 amino acids near the N-terminus of the protein. NAC genes are plant specific and are abundantly present in the plant genome. To date, more than 130 different NAC genes have been identified within the Arabidopsis genome of which only a proportion has been studied (Riechmann et al., 2000, Olsen et al., 2005). Within this superfamily of NAC proteins, three distinguishable families of transcription factors exists: the No Apical Meristem (NAM), Arabidopsis Transcription Activation Factors (ATAF) and the Cup-Shaped Cotyledon (CUC) families (Olsen et al., 2005). Literature reports the NAC transcription factors to be

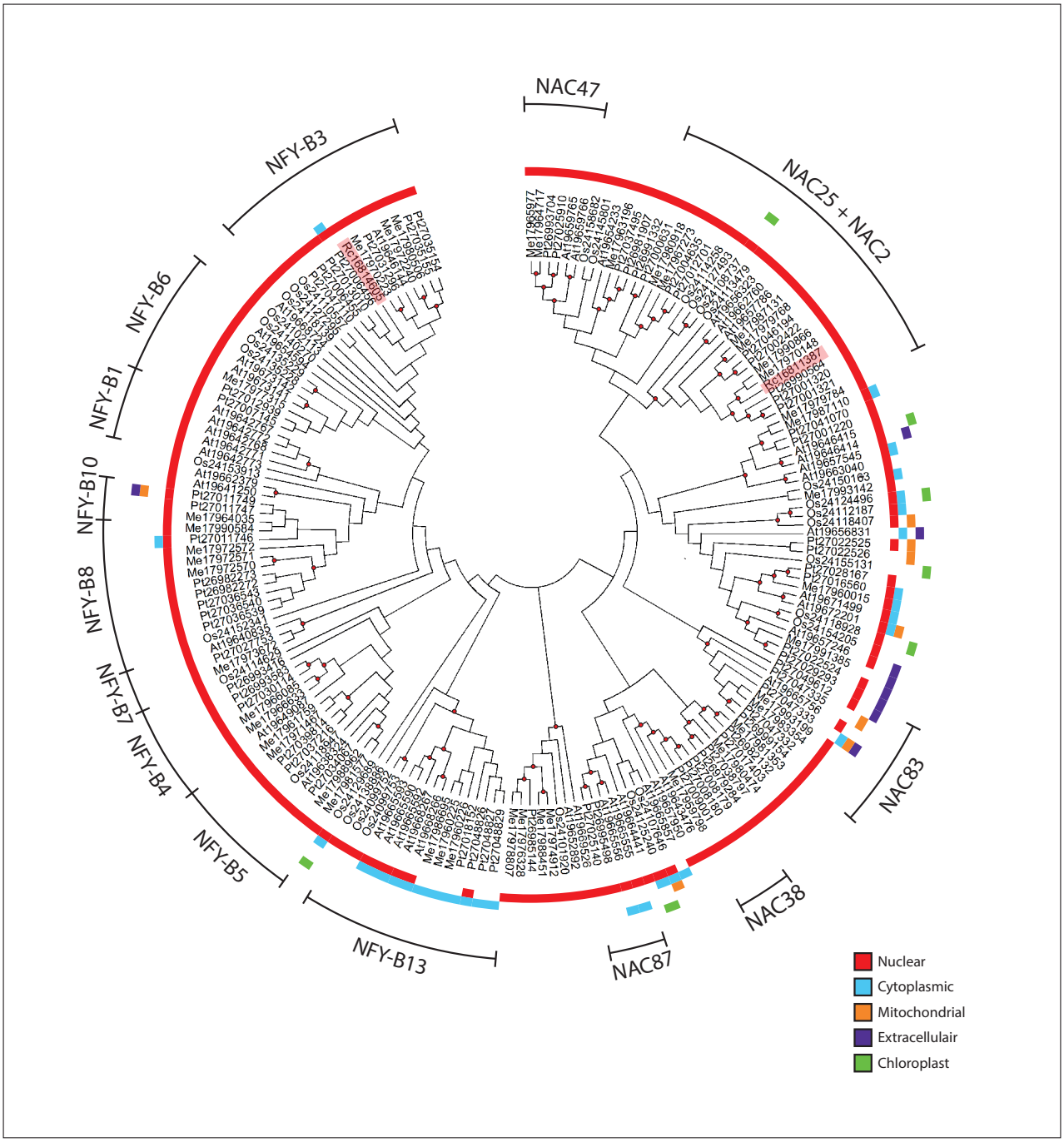

Figure 11: Phylogenetic tree constructed for the RcTF proteins. The phylogenetic tree depicts the homology and the high certainty branches (>95% after 500x bootstrapping) by the red dots. The subcellular prediction for each gene is represented by the coloured rings. Homology groups are depicted in the outer ring.

functional in the embryonic, floral and vegetative development, including regulation of formation of the shoot apical meristem, floral organs and lateral roots. Additionally, they are reported to be involved in pathogen defence and abiotic stress responses (Olsen et al., 2005). However, due to these large numbers of NAC proteins there is functional redundancy of the genes, increasing the difficulty of functional analysis.

The NAC protein homologs of RcTF1 presented in the phylogenetic tree belong to the subfamily of NAM transcription factors (PF02365). This type of transcription factor was first discovered in Petunia (Souer et al., 1996). Petunia seedlings with mutated NAM genes (nam-) showed a phenotype with no shoot apical meristem (SAM) (Souer et al., 1996). The shoot apical meristem generates vegetative structures such as leaves, internodes and axillary meristems and is therefore essential for development and survival of the seedlings. Furthermore the NAM proteins turned out to be required for normal flower development (Souer et al., 1996). Arabidopsis *AtNAMs* (homologs of the petunia *NAM* genes) are preferentially expressed in developing seeds, predominantly at the location of the embryonic SAM and in dry seeds (Duval et al., 2002). It is hypothesized that AtNAMs are also involved in the SAM development of Arabidopsis and that it has additional roles in embryo development (Duval et al., 2002).

More distantly related species including monocots like rice (*O. sativa)* and even moss species *(Physcomitrella patens)* show proteins containing NAC domains with very high sequence similarity to the *AtNAM* genes (Ooka et al., 2003, Olsen et al., 2005). This means that the NAC domain is well conserved. All in all, the function of this protein seems to be shared by different plant species and is therefore expected to be similar in *R. communis*.

For more specific information about the function of RcTF1, the closest *A. thaliana* homologs AtNAC25 and AtNAC2 proteins were researched. *AtNAC* gene expression profiles are very complex (Chavez, 2007). *AtNAM* genes are being expressed in many different tissues dependent on different developmental stages (Chavez, 2007). Nevertheless, an exception on this complicated expression pattern was found for one gene: *AtNAC25* (*At1g61110*), which was only expressed in the floral tissue. More specifically, it was only expressed in the anther tapetum (Chavez, 2007). *AtNAC25* was therefore re-named; *TAPNAC*. TAPNAC is expected to be an important factor in the regulatory mechanisms involved in the production, storage and secretion of compounds required for the development of pollen (Chavez, 2007). Loss of function or disturbance of this gene is likely to result in male sterility of *A. thaliana* which was indeed observed in the *A. thaliana* plants overexpressing the *RcTF1* gene.

No *TAPNAC* gene expression was found in germinating seeds according to the eFP browser (Winter et al., 2007) or in different stages of *A. thaliana* embryos (Chavez, 2007). Therefore, this gene seems to have a different expression pattern than the *RcTF2* gene which is expressed in germinating seeds under control temperatures. This difference in expression pattern between *AtTAPNAC* and *RcTF2* also suggest that the RcTF2 gene has different (or additional) functions. For future research it would be interesting to test *RcTF2* expression levels in different tissues and developmental phases of *R. communis.*

*At*NAC2 is the other close homolog to *Rc*TF1. *At*NAC2 regulates the development at the torpedo-shaped embryo stage (Kunieda et al., 2008). The *At*NAC2 protein is hypothesized to enhance the growth of the embryo by stimulating production of growth regulatory substances such as phytohormones (Kunieda et al., 2008). Furthermore, *At*NAC2 is likely to play a role in secretion of regulatory factors from the integuments to the embryo (Kunieda et al., 2008). Since most of the development regulatory functions of *At*NAC2 have been described for the embryonic phase, the question remains if this gene also exerts its function in post-germinative seedling growth.

In conclusion, for both these Arabidopsis homologs their described functions do not give a clear explanation for their involvement in the poor establishment of *R. communis* seedlings germinating under heat stress (35°C). It is hypothesized that the RcTF1 protein either has different functions, or that the functions of these homologs in the seedling heat stress responses are still undiscovered. However, the protein function assigned to the whole NAM protein family, regulation of formation of the shoot apical meristem, would possibly contribute to the arrested growth and reduced seedling establishment of *R. communis* seedlings germinating under stress. Therefore it is advised for future research to conduct a more precise expression and functional analysis of the *RcTF1* gene in *R. communis,* which could maybe clarify the function of the RcTF1 protein in germination and seedling establishment (under abiotic stress conditions).

RcTF2: a NF-YB protein

RcTF2 was originally annotated being a 'CCAAT-binding transcription factor subunit A'. According to the HMM-scan performed on this transcription factor protein and its homologs, all belong to the CBFD_NFYB_HMF histone like transcription factors and archaeal histone family (PF00808). This family is a subgroup within the family of histone-fold proteins. Histones mediate DNA organization and they play an important role in regulation of eukaryotic transcription. The functional Nuclear Factor Y (NF-Y) molecule is a heterotrimeric complex existing of NF-YA, NF-YB and NF-YC subunits. The NF-Y protein can recognize and bind to CCAAT-boxes, which are part of approximately 30% of all promoters (Mantovani, 1999, Bucher, 1990). The phylogenetic tree strongly supports the finding that the *Rc*TF2 gene is a Nuclear Factor YB subunit (NF-YB) (figure 11). NF-YB is also known as Heme Activator protein 3 (HAP3) or CCAAT-Binding Factor A (CBF-A).

In plants, the NF-Y transcription factors have been described to have approximately ten different genes encoding each of the subunit of this heterotrimerc transcription factor complex (Nelson et al., 2007, Gusmaroli et al., 2002). For *A. thaliana* 36 NF-Y subunit genes are identified; 10 NF-YA, 13 NF-YB, and 13 NF-YC subunits (Siefers et al., 2009). Of these 13 NF-YB subunits described for *A. thaliana,* nine are represented in the RcTF phylogenetic three as homologous genes, including AtNF-YB1, -4 until 8, -10 and -13. The RcTF2 protein falls in a very distinguishable group with only homologs of the Arabidopsis B3 subunit (AtNF-YB3 - *AT4G14540*). Its predicted subcellular location is solely in the nucleus as is described for histone proteins (Yu et al., 2006). Only the NF-YB13, which is least homologous to *Rc*TF, is besides the nucleus also predicted to locate in the cytoplasm.

Most interesting would be to define the function of the whole heterotrimeric Nuclear Factor Y complexes containing the B3 domain (as closest RcTF2 homolog). Nevertheless, recent findings show that this is too complex since AtNF-Y domains have the potential to form many slightly different heterotrimers (combining different homologs of the different subunits). The 36 NF-Y subunits of Arabidopsis could theoretically combine to form 1690 unique complexes, increasing the possibilities of combining with many promoters (Gusmaroli et al., 2002, Siefers et al., 2009). This form of combinatorial transcription factors can allow plants, as sessile organisms, to modulate growth and differentiation in their post-embryonic developmental program to meet the environmental requirements to survive (Gusmaroli et al., 2002, Siefers et al., 2009). NF-Y-regulated gene expression can be positive and negative, tissue specific, developmentally regulated or constitutive (Maity and de Crombrugghe, 1998, Ceribelli et

al., 2008). The MADS box transcription factors are a well-studied example of this type of multiple-protein transcription factors that participate in higher order complexes to control multiple genes (Messenguy and Dubois, 2003). To start the research into these genes, expression patterns of the different genes coding for the subunits were analysed, but the expression patterns turned out to be very diverse and complex for all different subunits (Edwards et al., 1998, Gusmaroli et al., 2002, Siefers et al., 2009).

For some of the separate subunits a functional analysis has already been performed. The *LEC1/NF-YB* gene has a specific expression pattern; its expression is strictly confined to seeds. Its function is described as a central regulating factor in embryonic developmental processes including the cotyledon identity (Lee et al., 2003, Lotan et al., 1998, Kwong et al., 2003). Additionally, overexpression of the *AtNF-YB1* and the maize *ZmNF-YB2* genes improved plant performance under drought conditions (Nelson et al., 2007). In soybean the NF-Y factors were also found to have a potential function in plant development and tolerance to drought responses (Quach et al., 2014). These findings show that NF-YB stimulates abiotic stress tolerance responses.

More specifically, the separate subunit AtNF-YB3 (as closest homolog to RcTF2), controls flowering time (Kumimoto et al., 2010, Kumimoto et al., 2008, Wenkel et al., 2006, Cai et al., 2007), although this was not found to be the case in Barley (Liang et al., 2012). In Arabidopsis, the AtNF-YB3 regulation of flowering time is shown to function by activation of the key floral regulator Flowering Locus T (FT) (Samach et al., 2000). In Wheat, overexpressing *Ta-NF-YB3* has shown to significantly increase the leaf chlorophyll content, photosynthesis rate and early growth rate (Stephenson et al., 2011, Petroni et al., 2012).

If *RcNF-YB3* overexpressing *A. thaliana* transformants also show increased early growth rate like wheat, this would be an indication that the RcTF2 protein has indeed the same function as the Arabidopsis NF-YB3 subunit. Furthermore this involvement of the NF-YB3 subunit in vigorous seedling growth supports the hypothesis that reduced transcription of the gene encoding this subunit (as a result of heat stress) could contribute to the arrested and aberrant seedling growth of *R. communis* seeds germinating at 35°C. Even more, because of its proposed function as subunit within a group of very important plant developmental transcription factors, which disturbance can have major impact on plant development and therefore seedling establishment.

_Rc_ZFP2: a zinc finger transcription factor

As third protein within the group of transcription factors, the *R. communis* zinc finger protein (RcZFP) is addressed. Zinc fingers form a group of special transcription factors. Two genes were selected that were putative *R. communis* zinc fingers. As described in the section 'HMM-scan results and prediction of subcellular domains' the RcZFP1 was rather a GDSL lipase protein than a true zinc finger like RcZFP2. Alignment of the protein sequences of RcZFP1 and RcZFP2 showed no clear conserved domains at all (Appendix 6).

Additionally, as expected for a transcription factor, RcZFP2 and its homologs were predicted to be located in the nucleus (figure 12), but RcZFP1 and its homologs not. The RcZFP1 protein and its homologs were predicted to be located in the plasma membrane, extracellular and in the vacuole (figure 13). Due to this lack of homology between the two zinc finger proteins it was not found useful or possible to produce a phylogenetic three combining both genes. So, two separate trees were generated. One for RcZFP1 (figure 13) and one for RcZFP2 (figure 12).

RcZFP2 was shown to belong to the family of C2H2 type of zinc finger proteins. More than 150 different zinc finger C2H2 transcription factors (or TF3a) in A. thaliana have been defined (Riechmann et al., 2000, Englbrecht et al., 2004). This group of zinc finger C2H2 transcription factors is characterized by two cysteine and two histidine amino acids of the conserved sequence, CX2_4CX3FXsHX3_5H. These four molecules coordinate the position of a single zinc atom that is used to form a compact structure of the zinc finger protein consisting of two beta sheets and an alpha helix (a ββα structure) (Pavletich and Pabo, 1991, Wolfe et al., 2000). This structure (and more specifically the amino acids on the surface of the α-helix) allows interaction with the major groove of DNA for gene expression regulation (Wolfe et al., 2000).

Four ZFP homolog groups for AtZFP proteins were discovered, of which AtZFP1 (AT1G80730) and AtZFP3 (AT5G25160) homologs are grouped together whereas AtZFP4 (AT1G66140) and AtZFP2 (AT5G57520) show distinct groups (figure 12). The RcZFP2 protein falls within the distinct group of AtZFP2 homologs.

All AtZFP homolog groups displayed in this tree are single zinc finger proteins (Tague and Goodman, 1995), meaning that they have only one zinc finger domain. Although zinc fingers are important developmental regulators for all eukaryotes, these single zinc finger transcription factors are unique for the plant taxa (Yanagisawa and Schmidt, 2000, Dinkins et al., 2012). Furthermore, proteins of the four depicted homolog groups were all classified in the C1-li subset (Englbrecht et al., 2004). This subset is defined by zinc fingers either containing a single zinc finger or a varying number (of 2 to 5) of dispersed zinc fingers. Another characteristic for this group is the conserved QALGGH sequence in positions 2 to 7 of the amino acid sequence coding for the DNA interacting alpha-helix (Englbrecht et al., 2004). Since the DNA sequence that can be recognized by single finger zinc finger proteins (for interaction) is short, different ways of increasing DNA specificity are proposed, including an additional nucleic acid binding domain somewhere in zinc finger structure or multiple zinc fingers that are expressed in the same tissue interact for increased specificity (Tague et al., 1996, Takatsuji and Matsumoto, 1996).

Looking at the expression profiles of these genes, the *AtZFP2* gene (the closest homolog of *RcZFP2*) shows a different expression than the other four described genes within the tree. Whereas *AtZFP1, AtZFP3, AtZFP4* and *AtZFP7* are expressed in roots and the inflorescence stem, the *AtZFP2* gene is expressed in the siliques, the inflorescence stem

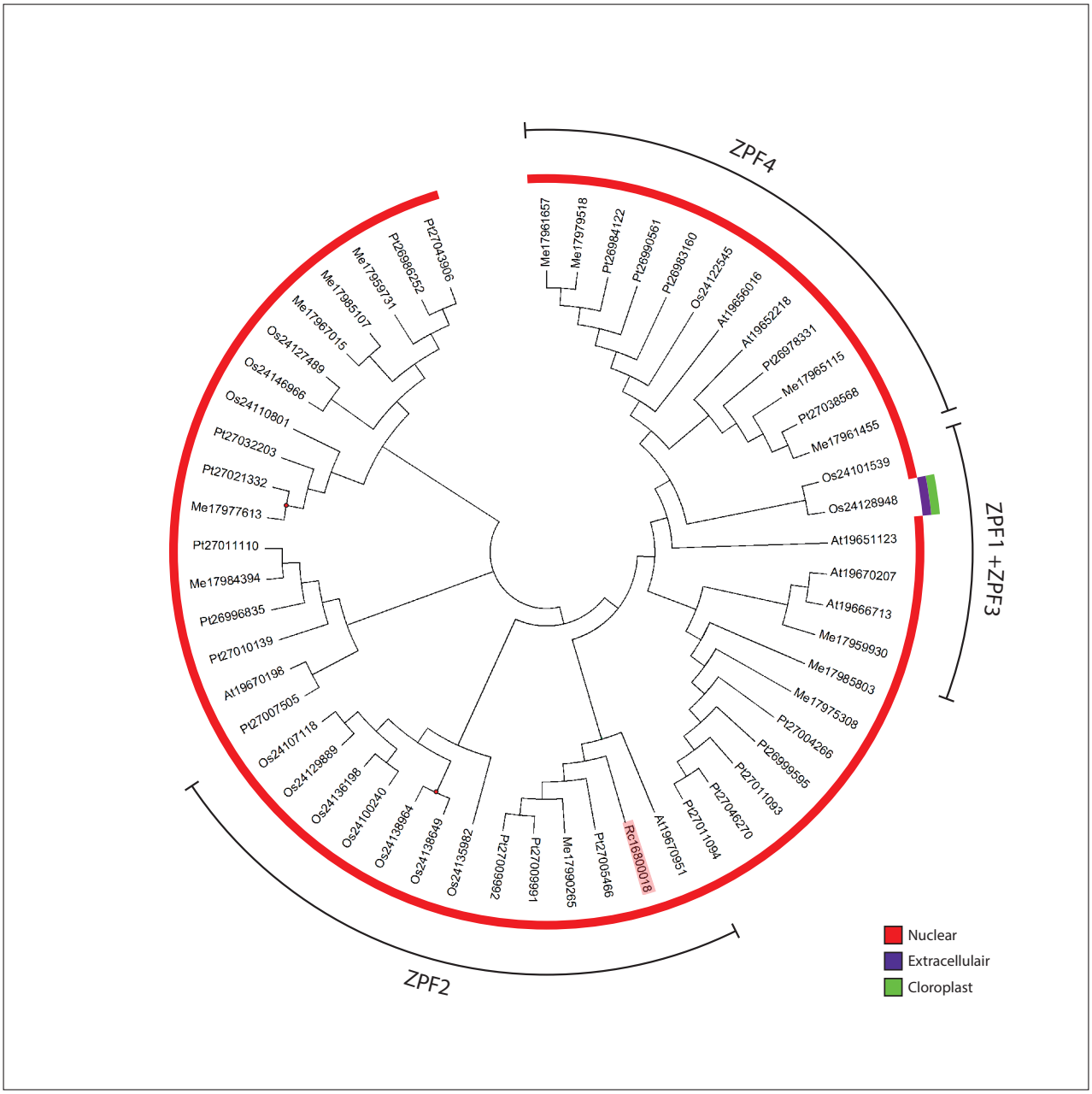

Figure 12: Phylogenetic tree constructed for the RcZFP2 proteins. The phylogenetic tree depicts the homology and the high certainty branches (>95% after 500x bootstrapping) by the red dots. The subcellular prediction for each gene is represented by the coloured rings. Homology groups are depicted in the outer ring.

(Tague and Goodman, 1995) and two weeks old seedlings (Cai and Lashbrook, 2008).

Use of ectopic overexpression as means for functional analysis has been applied to several different single zinc finger proteins including *AtZFP1, AtZFP2, AtZFP9 (SUPERMAN), AtZFP10* and *AtZFP11* (Dinkins et al., 2002, Tague et al., 1996, Sendon et al., 2014, Cai and Lashbrook, 2008, Dinkins et al., 2012). These analysis showed that these factors are important factors in regulation of plant growth and development both vegetative as well as reproductive. Disturbance of the expression (homeostasis) of these genes leads in several occasions to dwarf phenotypes.

Experiments for functional characterization of AtZFP2 were predominantly conducted in the context of regulation of abscission. It was found that AtZFP2 actively represses organ shed in combination with another type of zinc finger; the C_2C_2 DOF 4.7 *(AT4G38000*) (Wei et al., 2010, Cai and Lashbrook, 2008). Other phenotypic observations for these overexpression lines (of *AtZFP2)* were that flowers were aberrant with shorter stamen filaments, curved carpels and green petals. These deformed flowers lead to sterility of the plants. For the leaves, an increase in trichome number was observed (Cai and Lashbrook, 2008). In poplar trees *(Populus tremula × P. alba)* functional analysis by overexpression of the *PtaZFP2* within the poplar tree has shown to play a role in the trees regulation of stem growth and acclimation to mechanical load (Martin et al., 2014).

In conclusion the *RcZFP2* gene is a proper candidate to explain the arrested growth in the *R. communis* seedlings. The disturbance of the RcZFP2 homeostasis, caused by heat, can have detrimental effects on plant developmental programs.

With continuous overexpression (by the CaMV35 promoter), the heat-caused reduction in expression of *RcZFP2* could be counteracted. The question remains if the overexpression does not cause too much disturbance of the normal expression pattern of this gene by itself. Phenotyping of the *A. thaliana* plants overexpressing this gene will give more information about the possibility of using this gene for improving heat tolerance in *R. communis.*

Oil catabolism: lipolysis by ZFP1 and starch biosynthesis by GK and MLS

Within seeds, there are three types of nutrients that can be stored as reserves; carbohydrates, proteins and fats. In the seeds of *A. thaliana* as well as in *R. communis,* seed reserves are predominantly stored as fat molecules in the form of oil, which accumulates during seed maturation (Li et al., 2006, Bewley and Black, 1994). This oil, consisting of triacylglycerol (TAG) molecules, is a highly efficient form of nutrient storage (Graham, 2008, Baud and Lepiniec, 2010). Upon seed germination the reserves need to be converted into soluble metabolites that can support the plant during germination and early seedling establishment until photosynthesis becomes efficient. This means that reserves have to be sufficient, and well mobilized to be able to support seedling growth up to photoautotropism before reserves are exhausted (Graham, 2008).

Members of the third group of candidate genes, of which expression is down regulated by heat stress, are all involved in such nutrient mobilizing pathways. All three of these candidate genes are involved in the catabolism of triacylglycerols. The first step of oil reserve mobilization is lipolysis of the triacylglycerol (Graham, 2008). This lipolysis reaction is catalysed by the lipase enzyme which separates the three fatty acids groups of the backbone glycerol molecule. It is yet to be discovered which specific TAG-lipases are involved in lipolysis for seed storage oil mobilization (Quettier and Eastmond, 2009, Graham, 2008).

RcZFP1: a GDSL lipase

Lipases are involved in cleaving of mono-, di-, and triacylglycerol at the interfaces of oil and water. This reaction always results in production of free fatty acids and glycerol. The *R. communis* protein 'putative zinc finger 1 (RcZFP1) turned out to have high homology to such a lipase enzyme, more specifically a GDSL lipase. This was concluded from the data of the HMM-scan as well as the results for the homolog groups of the phylogenetic tree (figure 13). Therefore, from now on this protein will be referred to as *R. communis* GDSL lipase (RcGLIP).

GDSL lipases are shown to have broad substrate and region-specificity (Akoh et al., 2004). The key characteristic for lipases of this subclass is that they have a flexible active site that changes dependent on the substrate it binds to (Akoh et al., 2004). The RcGLIP protein is therefore probably involved in catalyzing lipolysis of a wide array of substrates. For a more specific description of its function, homolog groups were identified within the phylogenetic tree of the RcGLIP protein. The closest *A. thaliana* homolog to the *Rc*GLIP protein has been shown to be AtGLIP1 *(AT5G40990); GDSL lipase 1.* Many homologs of this protein were found in *P. trichocarpa* with two separate branches consisting almost solely of (putative) PtGLIP1 proteins. In one of the branches the PtGLIP1 proteins are grouped together with two *P. trichocarpa* genes homologous to AtGLIP3 *(AT1G53990)*. The RcGLIP protein is located in this particular branch (figure 13). The predicted subcellular localization in the vacuole and mitochondria for this branch is different from the predominantly predicted plasma membrane. Oil bodies in plants (where lipases can catalyse lipolysis) are structures of triacylglycerol surrounded by a layer of phospholipids (like the plasma membrane) and oleosins (Huang, 1992). Therefore the predicted membrane localization could be in agreement with the proposed function of this gene.

The AtGLIP1 protein is shown to be functional in the response to biotic stresses (Kwon et al., 2009). This is in agreement with the GO-terms for the biological process associated with the AtGLIP1 protein, which are predominantly defense responses to biotic stress such as 'defense response to bacterium', 'defense response to fungus' and 'jasmonic acid and ethylene-dependent systemic resistance' (TAIR, 2014). Additionally, its involvement in the triacylglycerol catabolism was also established by the GO-term 'Lipid catabolic process'.

Because of the abundance of *P. trichocarpa* genes, and the observed clustering of the *A. thaliana* proteins separately from these *P. trichocarpa* branches, the two PtGLIP proteins closest to *R. communis;* Pt26998715 (Potri.003G071400) and Pt26993577 (Potri.013G065100) were analysed as well by searching for their described function in the PlantCAZyme database (Ekstrom et al., 2014). Both proteins were described as PtGLIP1 proteins involved in triacylglycerol degradation, which therefore strongly supports the hypothesis of triacylglycerol degradation as the main function of RcGLIP1 too.

All in all, the function of this gene is not yet completely specified but this GDSL lipase is expected to be involved in (biotic) stress tolerance and could be involved in seed storage oil mobilization. To draw founded conclusions about the lypolytic function in oil mobilization for germination it would be necessary to test the seeds of *A. thaliana* lines overexpressing *RcGLIP* for increased glycerol content.

*Rc*GK and *Rc*MLS: enzymes involved in starch biosynthesis

After the liposysis of triacylglycerol is completed two new products are available; free fatty acids and glycerol. These two molecules will be catalysed further, to fuel seedling growth. The two enzymes cloned in the research of Mutimawurugo (2014), glycerol kinase (RcGK) and malate synthase (RcMLS), code for enzymes involved in the starch biosynthesis of the triacylglycerol mobilization pathway (Mahan, 2000, Graham, 2008). They function as enzymes catalyzing specific reactions that are necessary for oil conversion to carbohydrate molecules as energy supply. More specifically, RcGK is involved in the conversion of glycerol to starch and RcMLS is involved in the metabolic pathway that converts fatty acids into starch (Graham, 2008).

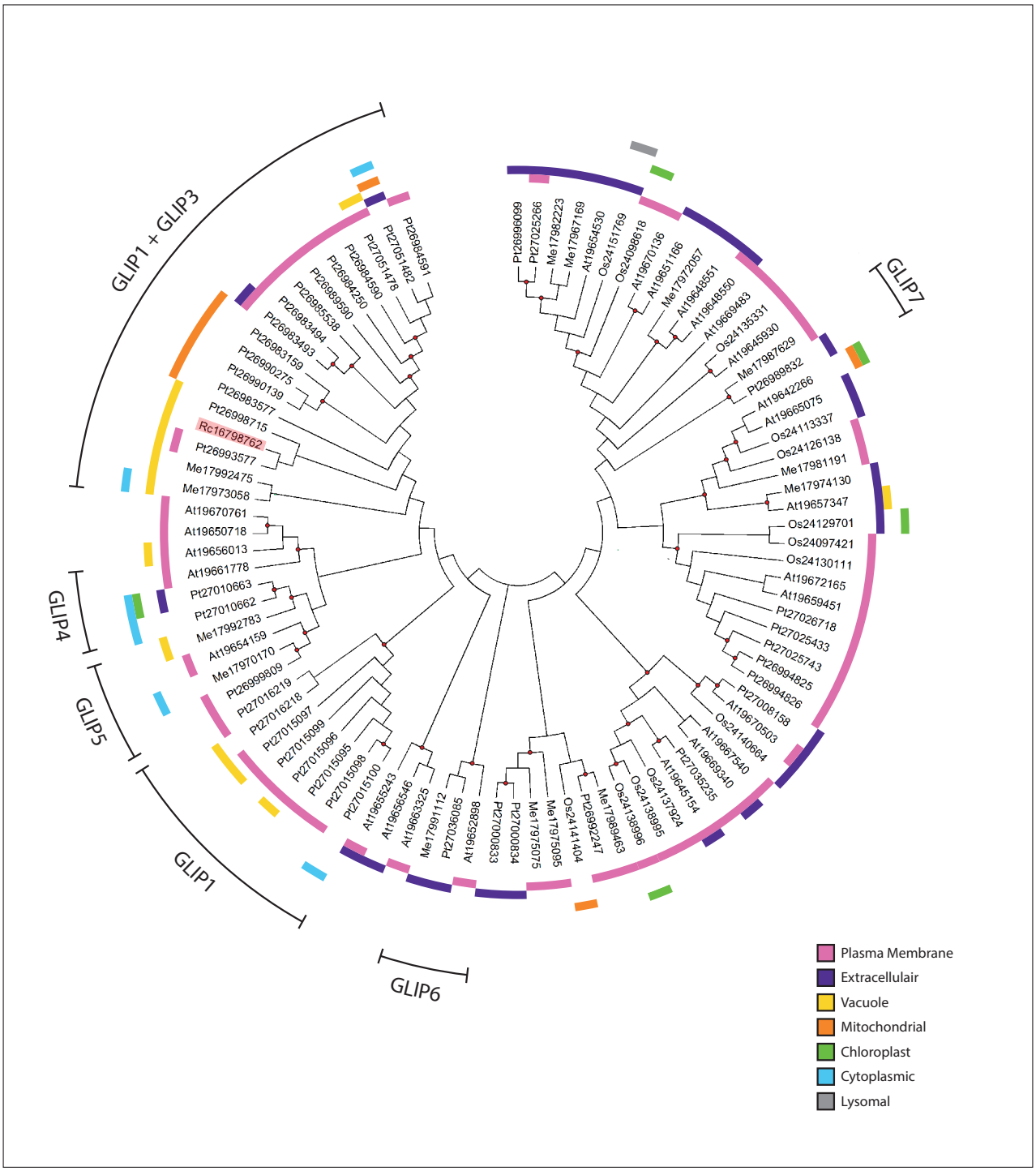

Figure 13: Phylogenetic tree constructed for the RcZFP1 protein. The phylogenetic tree depicts the homology and the high certainty branches (>95% after 500x bootstrapping) by the red dots. The subcellular prediction for each gene is represented by the coloured rings. Homology groups are depicted in the outer ring.

The annotation of the *R. communis* genes and their putative functions were confirmed by the HMM-scan results. The RcGK protein and its homologs all had as highest hit the domain of the FGGY family of carbohydrate kinases Nterminal domain (PF00370). This domain is the functional domain of glycerol kinase and other carbohydrate kinases according to the Interpro database (Hunter et al., 2009). The phylogenetic tree confirmed the identity of the gene to be glycerol kinase and not another carbohydrate kinase protein because all its homologs were described as glycerol kinases. RcMLS and its homologs all displayed the malate_synthase domain as top hit. Proteins belonging to the family of Malate synthase (PF012470) are involved as enzymes catalyzing Acetyl-CoA conversion to Malate as was described in the PFAM database (Punta et al., 2011). Both, acetyl-CoA and malate are intermediates of the pathway of converting fatty acids to starch.

Furthermore, the function of these cloned *R. communis* genes (*RcGK* and *RcMLS*) in the starch biosynthesis pathway was validated using a transient expression assay in tobacco leaves (Mutimawurugo, 2014, Ribeiro de Jesus et al., 2014c).

The results of this experiment showed that overexpression of these two *R. communis* genes in tobacco leaves caused higher levels of starch in those leaves, as expected.

Phylogenetic trees for RcGK and RcMLS are given in Appendix 7, but both are little informative, since they contain very few homologs. For RcGK, one homolog for this protein per species (some in different splicing forms) and no paralogs are found. Except for the *M. esculenta* glycerol kinases (MeGK), for which three different genes are present in this three. Of these, MeGK the 'cassava4.1_027307m' protein is most homologous to RcGK. Finding only one homolog per species and no paralogs was expected because FGGY kinases, with different substrate specificities, have only emerged once in evolutionary history (Zhang et al., 2011). This means that after duplication, proteins subsequently specified within this family, but the same (specialized) function forming paralogs did not occur except for *M. esculentum.* For this plant species, a paleo-genomic duplication was discovered which could explain the paralogs found for MeGK within this species (Prochnik et al., 2012).

For the RcMLS tree, the whole tree truly exists of one homolog (some with 2 splicing forms) for each species.

The structure of both threes is defined by the inter-specific differences. It is clear that the glycerol kinase protein sequences of different organisms display very different domains besides their GK domain, since all are predicted to be at different subcellular locations. For the RcGK protein, its closest MeGK homolog and the AtGK homologs (*AT1G80460.1* and *AT1G80460.2*), the predicted subcellular location are the cytoplasm and mitochondria. For RcMLS the tree shows that the PtMLS and MeMLS proteins are closest related to this *R. communis* protein. This is consistent with the evolutionary distances between these species. The Arabidopsis homolog of this protein is AtMLS (*AT5G03860.1)*. All the MLS proteins are predicted to be located in peroxisomes (and some also in the mitochondria). The peroxisomes are the microbodies where catabolic pathways unfold, including the triacylglycerol catabolism by malate synthase (Tolbert, 1981).

In conclusion, it is expected that heat-related disturbance of the expression of genes involved in the seed storage oil mobilization could have detrimental effects on seedling establishment. Without proper functioning of this catabolism pathway, seedlings will not be provided with enough energy and nutrients to properly establish themselves.

One thing to be taken into account when phenotyping the *A. thaliana* plants overexpressing these genes, is that differences are reported between the Arabidopsis and *R. communis* oil storage. According to Bewley and Black (1994) the *R. communis* endosperm is composed of 64% of oil and 36% of other storage molecules including 18% proteins. For *R. communis* the total oil content of a seed is approximately 60%, which is localized in the endosperm (Baud and Lepiniec, 2010). For *A. thaliana* it is reported that ~30% of the seed consists of oil which is predominantly localized in the cotyledons of the embryo (Katavic et al., 1995, Periappuram et al., 2000, Li et al., 2006).

Transmembrane transporter: *Rc*AAT

The last gene that was cloned and used to transform *A. thaliana* was *RcAAT*. This putative amino acid transporter is expected to be an integral component of the plasma membrane and is involved in the trans-membrane transport of amino acids. This function is supported by the subcellular location prediction that places RcAAT and all its homologs at the plasma membrane except for Pt27022717 (*Potri.002G226500*) (figure 14). More support for the amino acid transporter function of the gene was found in the results of the HMM-scan that showed that all proteins contained the amino acid transporter trans-membrane domain, except for one (Pt27039353 *(Potri.008G075500*)). The GO-terms associated with proteins containing this domain are 'amino acid transport' (biological process), 'integral component of membrane' (cellular component) and 'amino acid transmembrane transporter activity' (molecular function). All these analysis support the annotated function for RcAAT.

Different expression of genes located at the plasma membrane was expected in our experiments with germination of *R. communis* at high temperature, since the plasma membrane is very heavily affected by heat in its fluidity. Additionally, the permeability of the membrane is known to be affected (Kotak et al., 2007). This all causes membrane processes like trans-membrane transport (of amino acids) to be disrupted too (Larkindale et al., 2005d). Damage to the membrane or differences in the membrane fluidity might induce regulation or inhibition of genes encoding membrane related proteins.

Within the AAT phylogenetic tree, groups of genes that are homologous to specific *A. thaliana* amino acid transporter proteins can be found. The closest homolog of Arabidopsis to RcAAT is the AT3G30390.1 protein. However, this AtAAT protein and the other AtAAT proteins within this tree are not yet described or renamed. Checking the classification of the AtAAT proteins, it was discovered that all Arabidopsis homologs (and their homolog groups) belong to the same class of Aromatic and Neutral amino acid Transporters (ANT) (Rentsch et al., 2007, Svennerstam, 2008). The only *A. thaliana* amino acid transporter described for this class is AtANT1 (*AT3G11900.1*) (Chen et al., 2001). AtANT is an integral membrane protein that transports aromatic (like tyrosine and tryptophan) and neutral amino acids (such alanine and threonine), as well as arginine (which is a polar amino acid). According to the ePF browser this gene is only expressed in the seeds of *A. thaliana* (Winter et al., 2007). However, this *A. thaliana* gene or its homologs were not found to be a homolog of RcAAT with similarity above the threshold for including it in the phylogenetic tree. The biochemical properties of the AtANT proteins remain unknown (Rentsch et al., 2007).

Contradictory to the found *A. thaliana* homolog groups of proteins all belonging to the AtANT class is the classification of the rice amino acids transporters that are found in this phylogenetic tree. The rice amino acid transporters homologous to RcAAT were all classified as amino acid like transporters (OsALTa) amino acid transporters (Zhao et al., 2012). OsANT genes were also identified in rice but were

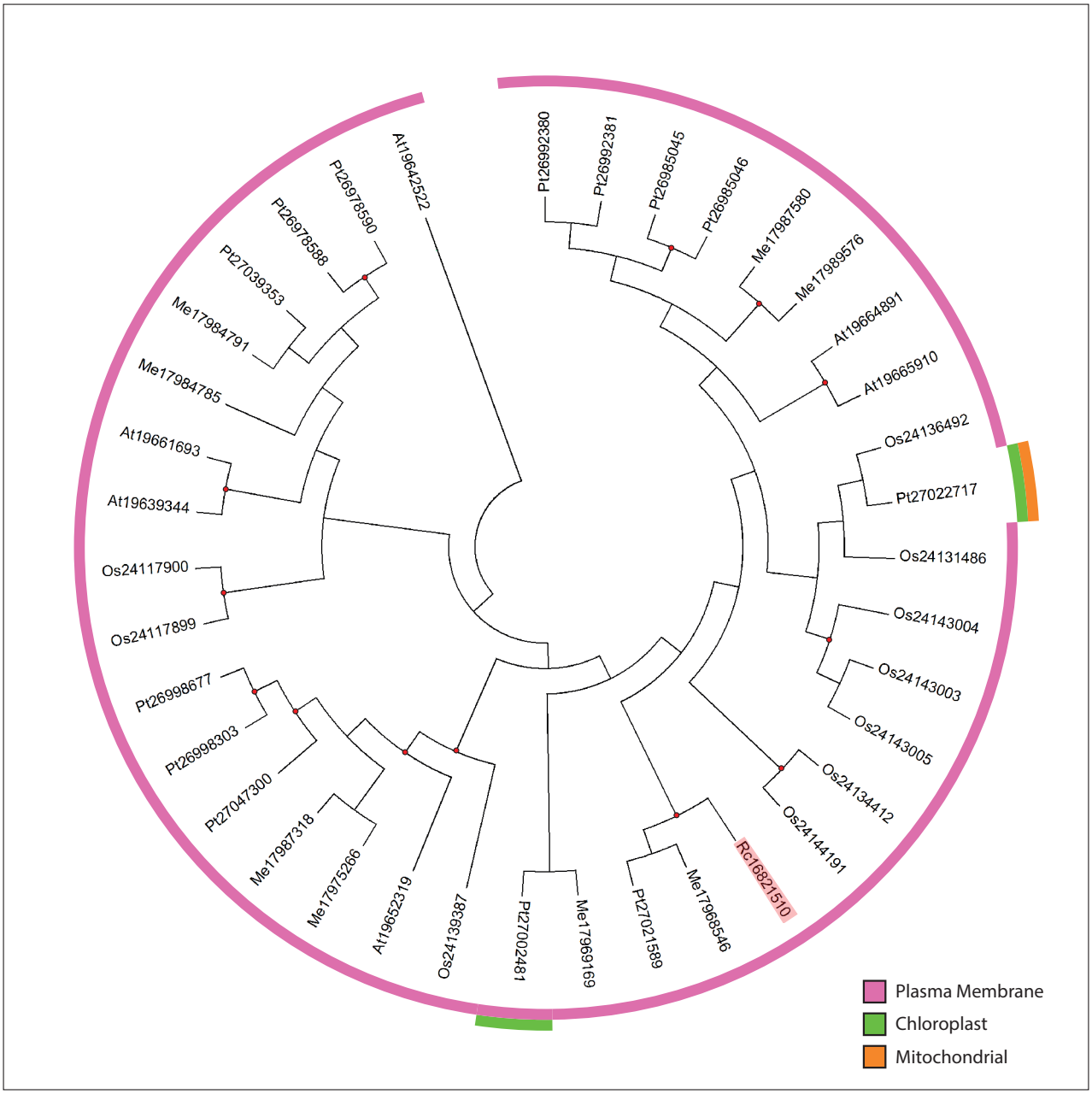

Figure 14: Phylogenetic tree constructed for the RcAAT protein. The phylogenetic tree depicts the homology and the high certainty branches (>95% after 500x bootstrapping) by the red dots. The subcellular prediction for each gene is represented by the coloured rings. Homology groups are depicted in the outer ring.

classified to be in a separate group. However, it was shown that these groups were closely related in rice because of their high sequence homology (Zhao et al., 2012). *OsALT* genes were shown to be mainly expressed in the panicle and the leaves, whereas *OsANT* genes where shown to be predominantly expressed in 7 days old seedlings (Zhao et al., 2012). However different *OsANT* and *OsALT* genes show different expression patterns and therefore different functions are expected, but no clear functions are yet described. In addition, for other plant species ALT amino acid transporters have not been reported (Rentsch et al., 2007, Svennerstam, 2008, TAIR, 2014, Zhao et al., 2012). Since both the *A. thaliana* ANT transporters and the *Oryza sativa* ALT transporters are equally close related to the *RcAAT* protein according to this phylogenetic tree, more research is required to establish its molecular specificity.

Independent of the precise molecular function and specificity of this particular amino acid transporter, it is clear that amino acid transporters play an indispensable role within cellular homeostasis. Amino acids are the building blocks for enzymes and proteins and their availability within the cell is essential for plant metabolism and structure. Amino acids can serve as nitrogen donors for the synthesis of a variety of compounds which are critical for plant development, including nucleotides, chlorophyll, hormones and secondary metabolites (Tegeder, 2012). Furthermore, amino acids are found to be involved in signal transduction pathways of biological processes that are of major importance to the plants such as photosynthesis, glycolysis and mitochondrial functioning (Olson et al., 2000, Ooka et al., 2003, Paul et al., 2008). Without proper expression of amino acid transporters depletion of amino acids can arise in the cell because the amino acids cannot be transported into the cell anymore.

Without a functional amino acid transport system all the processes in which amino acids are involved could also be disturbed. All in all, it is clear that having a functional amino acid transportation system is necessary for proper seedlings establishment in *R. communis*.

Promising candidate genes

The functions of all *R. communis* candidate genes are very divergent. Nonetheless, all are predicted to be of essential importance for development of well-functioning plants. For some of the genes, their predicted function is expected to be more involved in the observed *R. communis* seedling damage by heat stress than others. Therefore some genes are more promising candidates to be able to improve thermotolerance than other candidate genes. One of the most promising candidates according to the in-silico analysis is the *RcMBF* gene, because of its possible key regulator function in basal thermotolerance. For improvement of the acquired thermotolerance the *RcHSP1* gene is of most interest because it has been shown to be involved in the protection of mitochondria specifically for seeds. Furthermore, the *RcTF2* gene promised to be interesting because of its involvement in major developmental processes and its recorded influence on seedling growth (in wheat). Lastly, genes involved in the seed oil storage mobilization pathways were selected as being promising candidates *(RcGK, RcMLS and RcZFP1)* for improving thermotolerance of *R. communis*. Disruption of this pathway for example by down-regulation of (one of) these key enzymes could lead to insufficient nutrients to support proper seedling growth up to its establishment with efficient photosynthesis.

Future research

For future research, the first step is to phenotype the stable transformed *A. thaliana* lines overexpressing the *R. communis* candidate genes. These lines should be tested for phenotypes with regard to thermotolerance and seedling development. With the heat stress threshold experiment, as performed in this research, lines can be tested for basal thermotolerance. Besides that, it will also be necessary to test for acquired thermotolerance in an experiment where the tested plants are subjected to a short period of sub-lethal heat stress, followed by a recovery period, after which the plants are placed in (normally) lethal temperatures (Clarke et al., 2004, Larkindale and Huang, 2004). This will be of special interest for phenotyping (improved) thermotolerance when overexpressing the heat shock proteins.

Findings of this phenotyping experiment could be validated, testing Arabidopsis (T-DNA) knock-out lines of the homologs of the *R. communis* genes. Furthermore this additional research in which knock-out phenotypes are observed, will provide supportive data for function prediction of the genes (Bouché and Bouchez, 2001, Krysan et al., 1999).

Phenotyping *A. thaliana* lines overexpressing their closest homologs to the *R. communis* candidate genes is another form of proposed validation for the phenotyping experiment results. This is proposed to minimize the risk of misinterpretation of the influence of the *R. communis* genes. This could, for example, arise because of the high levels of ricinoleic acids present in castor bean, whereas this compound is not produced in Arabidopsis and therefore the enzymes or pathways might be slightly different in function (Chan et al., 2010).

After establishing the functionality of the genes in *A. thaliana*, the next step is to transform *R. communis* for overexpression of its heat tolerance improving genes. Research into stable transformation of *R. communis* is already initiated and could be applied once further developed (Sujatha et al., 2008, Malathi et al., 2006, Sailaja et al., 2008, Sujatha and Sailaja, 2005). Furthermore, the understanding of thermotolerance that is obtained could be extrapolated to other crops for improvement of their thermotolerance during germination and seedling establishment.

Acknowledgements

I would like to express my gratitude to Paulo Ribeiro for all of his help, daily supervision and assistance. His experience, help and feedback has allowed me to learn so many different aspects of plant biotechnological research and has increased the quality of my project significantly. My thanks goes also out to Wilco Ligterink for his supervision and feedback. Furthermore I would like to thank Mariana Silva Arthur who kindly shared her knowledge of phylogenetic research and also Anderson Tadeu Silva and Juriaan Rienstra who helped me a great deal by shared their knowledge concerning *A. thaliana* transformation and genotyping. Lastly, I would like to thank all (PhD) students and employees of the Plant Physiology Department for making me feel welcome and for their cooperation.

References

AGARWAL, S., SAIRAM, R., SRIVASTAVA, G., TYAGI, A. & MEENA, R. 2005. Role of ABA, salicylic acid, calcium and hydrogen peroxide on antioxidant enzymes induction in wheat seedlings. Plant Science, 169, 559-570.

AHRAZEM, O., RUBIO-MORAGA, A., TRAPERO-MOZOS, A., CLIMENT, M. F. L., GÓMEZ-CADENAS, A. & GÓMEZ-GÓMEZ, L. 2015. Ectopic expression of a stress-inducible glycosyltransferase from saffron enhances salt and oxidative stress tolerance in Arabidopsis while alters anchor root formation. Plant Science, 234, 60-73.

AKOH, C. C., LEE, G.-C., LIAW, Y.-C., HUANG, T.-H. & SHAW, J.-F. 2004. GDSL family of serine esterases/lipases. Progress in lipid research, 43, 534-552.

ALVARADO, V. & BRADFORD, K. 2002. A hydrothermal time model explains the cardinal temperatures for seed germination. Plant, Cell & Environment, 25, 1061-1069.

ANJANI, K. 2012. Castor genetic resources: A primary gene pool for exploitation. Industrial Crops and Products, 35, 1-14.

ARCE, D. P., GODOY, A. V., TSUDA, K., YAMAZAKI, K.-I., VALLE, E. M., IGLESIAS, M. J., DI MAURO, M. F. & CASALONGUÉ, C. A. 2010. The analysis of an Arabidopsis triple knock-down mutant reveals functions for MBF1 genes under oxidative stress conditions. Journal of plant physiology, 167, 194-200.

AVONCE, N., LEYMAN, B., MASCORRO-GALLARDO, J. O., VAN DIJCK, P., THEVELEIN, J. M. & ITURRIAGA, G. 2004. The Arabidopsis trehalose-6-P synthase AtTPS1 gene is a regulator of glucose, abscisic acid, and stress signaling. Plant Physiology, 136, 3649-3659.

BARDEL, J., LOUWAGIE, M., JAQUINOD, M., JOURDAIN, A., LUCHE, S., RABILLOUD, T., MACHEREL, D., GARIN, J. & BOURGUIGNON, J. 2002. A survey of the plant mitochondrial proteome in relation to development. Proteomics, 2, 880-898.

BAUD, S. & LEPINIEC, L. 2010. Physiological and developmental regulation of seed oil production. Progress in lipid research, 49, 235-249.

BEWLEY, J. D. & BLACK, M. 1994. Seeds, Springer.

BOUCHÉ, N. & BOUCHEZ, D. 2001. Arabidopsis gene knockout: phenotypes wanted. Current opinion in plant biology, 4, 111-117.

BRENNAN, R. & MATTHEWS, B. W. 1989. The helix-turn-helix DNA binding motif. J. biol. Chem, 264, 1903-1906.

BUCHER, P. 1990. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. Journal of molecular biology, 212, 563-578.

BURKE, J. J., O'MAHONY, P. J. & OLIVER, M. J. 2000. Isolation of Arabidopsis mutants lacking components of acquired thermotolerance. Plant Physiology, 123, 575-588.

CAI, S. & LASHBROOK, C. C. 2008. Stamen abscission zone transcriptome profiling reveals new candidates for abscission control: enhanced retention of floral organs in transgenic plants overexpressing Arabidopsis ZINC FINGER PROTEIN2. Plant physiology, 146, 1305-1321.

CAI, X., BALLIF, J., ENDO, S., DAVIS, E., LIANG, M., CHEN, D., DEWALD, D., KREPS, J., ZHU, T. & WU, Y. 2007. A putative CCAAT-binding transcription factor is a regulator of flowering timing in Arabidopsis. Plant physiology, 145, 98-105.

CAPELLA-GUTIÉRREZ, S., SILLA-MARTÍNEZ, J. M. & GABALDÓN, T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics, 25, 1972-1973.

CERIBELLI, M., DOLFINI, D., MERICO, D., GATTA, R., VIGANÒ, A. M., PAVESI, G. & MANTOVANI, R. 2008. The histone-like NF-Y is a bifunctional transcription factor. Molecular and cellular biology, 28, 2047-2058.

CHAN, A. P., CRABTREE, J., ZHAO, Q., LORENZI, H., ORVIS, J., PUIU, D., MELAKE-BERHAN, A., JONES, K. M., REDMAN, J. & CHEN, G. 2010. Draft genome sequence of the oilseed species Ricinus communis. Nature biotechnology, 28, 951-956.

CHAVEZ, V. A. 2007. Analysis of the Arabidopsis NAC gene superfamily in plant development. Texas A&M University.

CHEEMA, N. M., MALIK, M. A., QADIR, G., RAFIQUE, M. Z. & NAWAZ, N. 2010. Influence of temperature and osmotic stress on germination induction of different castor bean cultivars. Pakistan J. Bot, 42, 4035-4041.

CHEN, L., ORTIZ-LOPEZ, A., JUNG, A. & BUSH, D. R. 2001. ANT1, an aromatic and neutral amino acid transporter in Arabidopsis. Plant Physiology, 125, 1813-1820.

CLARKE, S. M., MUR, L. A., WOOD, J. E. & SCOTT, I. M. 2004. Salicylic acid dependent signaling promotes basal thermotolerance but is not essential for acquired thermotolerance in Arabidopsis thaliana. The Plant Journal, 38, 432-447.

CLOUGH, S. J. & BENT, A. F. 1998. Floral dip: a simplified method forAgrobacterium□mediated transformation ofArabidopsis thaliana. The plant journal, 16, 735-743.

DA SILVA CÉSAR, A. & OTÁVIO BATALHA, M. 2010. Biodiesel production from castor oil in Brazil: A difficult reality. Energy Policy, 38, 4031-4039.

DE SOUZA COSTA, E. T., GUILHERME, L. R. G., DE MELO, É. E. C., RIBEIRO, B. T., EUZELINA DOS SANTOS, B. I., DA COSTA SEVERIANO, E., FAQUIN, V. & HALE, B. A. 2012. Assessing the tolerance of castor bean to Cd and Pb for phytoremediation purposes. Biological trace element research, 145, 93-100.

DINKINS, R., PFLIPSEN, C., THOMPSON, A. & COLLINS, G. B. 2002. Ectopic expression of an Arabidopsis single zinc finger gene in tobacco results in dwarf plants. Plant and cell physiology, 43, 743-750.

DINKINS, R. D., TAVVA, V. S., PALLI, S. R. & COLLINS, G. B. 2012. Mutant and overexpression analysis of a C2H2

single zinc finger gene of Arabidopsis. Plant Molecular Biology Reporter, 30, 99-110.
DUVAL, M., HSIEH, T.-F., KIM, S. Y. & THOMAS, T. L. 2002. Molecular characterization of AtNAM: A member of theArabidopsis NAC domain superfamily. Plant molecular biology, 50, 237-248.
EDWARDS, D., MURRAY, J. A. & SMITH, A. G. 1998. Multiple genes encoding the conserved CCAAT-box transcription factor complex are expressed in Arabidopsis. Plant physiology, 117, 1015-1022.
EKSTROM, A., TAUJALE, R., MCGINN, N. & YIN, Y. 2014. PlantCAZyme: a database for plant carbohydrate-active enzymes. Database, 2014, bau079.
ENGLBRECHT, C. C., SCHOOF, H. & BÖHM, S. 2004. Conservation, diversification and expansion of C2H2 zinc finger proteins in the Arabidopsis thaliana genome. BMC genomics, 5, 39.
FAOSTAT, F. 2014. Statistical Databases. Food and Agriculture Organization of the United Nations.
FINN, R. D., CLEMENTS, J. & EDDY, S. R. 2011. HMMER web server: interactive sequence similarity searching. Nucleic acids research, gkr367.
FOSTER, J., ALLAN, G., CHAN, A., RABINOWICZ, P., RAVEL, J., JACKSON, P. & KEIM, P. 2010. Single nucleotide polymorphisms for assessing genetic diversity in castor bean (Ricinus communis). BMC plant biology, 10, 13.
FUJITA, M., FUJITA, Y., NOUTOSHI, Y., TAKAHASHI, F., NARUSAKA, Y., YAMAGUCHI-SHINOZAKI, K. & SHINOZAKI, K. 2006. Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. Current opinion in plant biology, 9, 436-442.
GAO, G., ZHONG, Y., GUO, A., ZHU, Q., TANG, W., ZHENG, W., GU, X., WEI, L. & LUO, J. 2006. DRTF: a database of rice transcription factors. Bioinformatics, 22, 1286-1287.
GODOY, A. V., ZANETTI, M. E., SAN SEGUNDO, B. & CASALONGUÉ, C. A. 2001. Identification of a putative Solanum tuberosum transcriptional coactivator up-regulated in potato tubers by Fusarium solani f. sp. eumartii infection and wounding. Physiologia Plantarum, 112, 217-222.
GOODSTEIN, D. M., SHU, S., HOWSON, R., NEUPANE, R., HAYES, R. D., FAZO, J., MITROS, T., DIRKS, W., HELLSTEN, U. & PUTNAM, N. 2012. Phytozome: a comparative platform for green plant genomics. Nucleic acids research, 40, D1178-D1186.
GRAHAM, I. A. 2008. Seed storage oil mobilization. Annu. Rev. Plant Biol., 59, 115-142.
GRIFFITHS, A. J. 2005. An introduction to genetic analysis, Macmillan.
GROENEN, P. J., MERCK, K. B., JONG, W. W. & BLOEMENDAL, H. 1994. Structure and Modifications of the Junior Chaperone α□Crystallin. European Journal of Biochemistry, 225, 1-19.
GUSMAROLI, G., TONELLI, C. & MANTOVANI, R. 2002. Regulation of novel members of the Arabidopsis thaliana CCAAT-binding nuclear factor Y subunits. Gene, 283, 41-48.
HALL, A. E. 2000. Crop responses to environment, CRC press.
HARLAN, J. R. 1976. Origins of African plant domestication, Walter de Gruyter.
HORVÁTH, E., SZALAI, G. & JANDA, T. 2007. Induction of abiotic stress tolerance by salicylic acid signaling. Journal of Plant Growth Regulation, 26, 290-300.
HORWITZ, J. 2003. Alpha-crystallin. Experimental eye research, 76, 145-153.
HUANG, A. H. 1992. Oil bodies and oleosins in seeds. Annual review of plant biology, 43, 177-200.
HUNTER, S., APWEILER, R., ATTWOOD, T. K., BAIROCH, A., BATEMAN, A., BINNS, D., BORK, P., DAS, U., DAUGHERTY, L. & DUQUENNE, L. 2009. InterPro: the integrative protein signature database. Nucleic acids research, 37, D211-D215.
JAWDAT, D. & KARAJOLI, I. 2012. A modified and inexpensive protocol for the rapid isolation of RNA from diverse plant species. JOURNAL OF HORTICULTURAL SCIENCE & BIOTECHNOLOGY, 87, 317-324.
JIN, J., ZHANG, H., KONG, L., GAO, G. & LUO, J. 2013. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. Nucleic acids research, gkt1016.
KATAVIC, V., REED, D. W., TAYLOR, D. C., GIBLIN, E. M., BARTON, D. L., ZOU, J., MACKENZIE, S. L., COVELLO, P. S. & KUNST, L. 1995. Alteration of seed fatty acid composition by an ethyl methanesulfonate-induced mutation in Arabidopsis thaliana affecting diacylglycerol acyltransferase activity. Plant Physiology, 108, 399-409.
KIM, H. U., JUNG, S.-J., LEE, K.-R., KIM, E. H., LEE, S.-M., ROH, K. H. & KIM, J.-B. 2014. Ectopic overexpression of castor bean LEAFY COTYLEDON2 (LEC2) in Arabidopsis triggers the expression of genes that encode regulators of seed maturation and oil body proteins in vegetative tissues. FEBS open bio, 4, 25-32.
KOTAK, S., LARKINDALE, J., LEE, U., VON KOSKULL-DÖRING, P., VIERLING, E. & SCHARF, K.-D. 2007. Complexity of the heat stress response in plants. Current opinion in plant biology, 10, 310-316.
KREGEL, K. C. 2002. Invited review: heat shock proteins: modifying factors in physiological stress responses and acquired thermotolerance. Journal of applied physiology, 92, 2177-2186.
KRYSAN, P. J., YOUNG, J. C. & SUSSMAN, M. R. 1999. T-DNA as an insertional mutagen in Arabidopsis. The Plant Cell Online, 11, 2283-2290.
KUMAR, S., NEI, M., DUDLEY, J. & TAMURA, K. 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Briefings in bioinformatics, 9, 299-306.
KUMIMOTO, R. W., ADAM, L., HYMUS, G. J., REPETTI, P. P., REUBER, T. L., MARION, C. M., HEMPEL, F. D. & RATCLIFFE, O. J. 2008. The Nuclear Factor Y subunits NF-YB2 and NF-YB3 play additive roles in the promotion of flowering by inductive long-day photoperiods in Arabidopsis. Planta, 228, 709-723.

KUMIMOTO, R. W., ZHANG, Y., SIEFERS, N. & HOLT, B. F. 2010. NF–YC3, NF–YC4 and NF–YC9 are required for CONSTANS-mediated, photoperiod□dependent flowering in Arabidopsis thaliana. The Plant Journal, 63, 379-391.
KUNIEDA, T., MITSUDA, N., OHME-TAKAGI, M., TAKEDA, S., AIDA, M., TASAKA, M., KONDO, M., NISHIMURA, M. & HARA-NISHIMURA, I. 2008. NAC family proteins NARS1/NAC2 and NARS2/NAM in the outer integument regulate embryogenesis in Arabidopsis. The Plant Cell Online, 20, 2631-2642.
KWON, S. J., JIN, H. C., LEE, S., NAM, M. H., CHUNG, J. H., KWON, S. I., RYU, C. M. & PARK, O. K. 2009. GDSL lipase□like 1 regulates systemic resistance associated with ethylene signaling in Arabidopsis. The Plant Journal, 58, 235-245.
KWONG, R. W., BUI, A. Q., LEE, H., KWONG, L. W., FISCHER, R. L., GOLDBERG, R. B. & HARADA, J. J. 2003. LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. The Plant Cell Online, 15, 5-18.
LANGEVIN, M. 2010. The Brazilian Biodiesel Program. Journal of Energy Security.
LARKINDALE, J., HALL, J., KNIGHT, M. & VIERLING, E. 2005a. Heat stress phenotypes of Arabidopsis mutants implicate multiple signaling pathways in the acquisition of thermotolerance. Plant Physiology, 138, 882-897.
LARKINDALE, J. & HUANG, B. 2004. Thermotolerance and antioxidant systems in Agrostis stolonifera: involvement of salicylic acid, abscisic acid, calcium, hydrogen peroxide, and ethylene. Journal of plant physiology, 161, 405-413.
LARKINDALE, J., MISHKIND, M. & VIERLING, E. 2005f. Plant responses to high temperature. Plant abiotic stress, 100-144.
LARKINDALE, J. & VIERLING, E. 2008. Core genome responses involved in acclimation to high temperature. Plant Physiology, 146, 748-761.
LEE, H., FISCHER, R. L., GOLDBERG, R. B. & HARADA, J. J. 2003. Arabidopsis LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. Proceedings of the National Academy of Sciences, 100, 2152-2156.
LI, Y., BEISSON, F., POLLARD, M. & OHLROGGE, J. 2006. Oil content of Arabidopsis seeds: the influence of seed anatomy, light and plant-to-plant variation. Phytochemistry, 67, 904-915.
LIANG, M., HOLE, D., WU, J., BLAKE, T. & WU, Y. 2012. Expression and functional analysis of NUCLEAR FACTOR-Y, subunit B genes in barley. Planta, 235, 779-791.
LINDQUIST, S. & CRAIG, E. 1988. The heat-shock proteins. Annual review of genetics, 22, 631-677.
LIU, X., GAO, Y., KHAN, S., DUAN, G., CHEN, A., LING, L., ZHAO, L., LIU, Z. & WU, X. 2008. Accumulation of Pb, Cu, and Zn in native plants growing on contaminated sites and their potential accumulation capacity in Heqing, Yunnan. Journal of Environmental Sciences, 20, 1469-1474.
LOGEMANN, E., BIRKENBIHL, R. P., ÜLKER, B. & SOMSSICH, I. E. 2006. An improved method for preparing Agrobacterium cells that simplifies the Arabidopsis transformation protocol. Plant Methods, 2, 16.
LOTAN, T., OHTO, M.-A., YEE, K. M., WEST, M. A., LO, R., KWONG, R. W., YAMAGISHI, K., FISCHER, R. L., GOLDBERG, R. B. & HARADA, J. J. 1998. Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. Cell, 93, 1195-1205.
LÖW, D., BRÄNDLE, K., NOVER, L. & FORREITER, C. 2000. Cytosolic heat-stress proteins Hsp17. 7 class I and Hsp17. 3 class II of tomato act as molecular chaperones in vivo. Planta, 211, 575-582.
MACHEREL, D., BENAMAR, A., AVELANGE-MACHEREL, M. H. & TOLLETER, D. 2007. Function and stress tolerance of seed mitochondria. Physiologia Plantarum, 129, 233-241.
MADERA, M., VOGEL, C., KUMMERFELD, S. K., CHOTHIA, C. & GOUGH, J. 2004. The SUPERFAMILY database in 2004: additions and improvements. Nucleic acids research, 32, D235-D239.
MAHAN, J. R. 2000. Thermal dependence of malate synthase activity and its relationship to the thermal dependence of seedling emergence. Journal of agricultural and food chemistry, 48, 4544-4549.
MAITY, S. N. & DE CROMBRUGGHE, B. 1998. Role of the CCAAT-binding protein CBF/NF-Y in transcription. Trends in biochemical sciences, 23, 174-178.
MALATHI, B., RAMESH, S., RAO, K. V. & REDDY, V. D. 2006. Agrobacterium-mediated genetic transformation and production of semilooper resistant transgenic castor (Ricinus communis L.). Euphytica, 147, 441-449.
MANSFIELD, M. A. & KEY, J. L. 1987. Synthesis of the low molecular weight heat shock proteins in plants. Plant physiology, 84, 1007-1017.
MANTOVANI, R. 1999. The molecular biology of the CCAAT-binding factor NF-Y. Gene, 239, 15-27.
MARTIN, L., DECOURTEIX, M., BADEL, E., HUGUET, S., MOULIA, B., JULIEN, J. L. & LEBLANC□FOURNIER, N. 2014. The zinc finger protein PtaZFP2 negatively controls stem growth and gene expression responsiveness to external mechanical loads in poplar. New Phytologist, 203, 168-181.
MCADAMS, H. H. & ARKIN, A. 1997. Stochastic mechanisms in gene expression. Proceedings of the National Academy of Sciences, 94, 814-819.
MESSENGUY, F. & DUBOIS, E. 2003. Role of MADS box proteins and their cofactors in combinatorial control of gene expression and cell development. Gene, 316, 1-21.
MILLER, G., SHULAEV, V. & MITTLER, R. 2008. Reactive oxygen signaling and abiotic stress. Physiologia Plantarum, 133, 481-489.
MITTLER, R., VANDERAUWERA, S., GOLLERY, M. & VAN BREUSEGEM, F. 2004. Reactive oxygen gene network

of plants. Trends in plant science, 9, 490-498.
MITTLER, R., VANDERAUWERA, S., SUZUKI, N., MILLER, G., TOGNETTI, V. B., VANDEPOELE, K., GOLLERY, M., SHULAEV, V. & VAN BREUSEGEM, F. 2011. ROS signaling: the new wave? Trends in plant science, 16, 300-309.
MOREY, J. S., RYAN, J. C. & VAN DOLAH, F. M. 2006. Microarray validation: factors influencing correlation between oligonucleotide microarrays and real-time PCR. Biological procedures online, 8, 175-193.
MOSHKIN, V. 1986. Castor, AA Balkema.
MUTIMAWURUGO, M. 2014. Cloning and function characterization of temperature responsive genes in Ricinus communis L. MSc Plant Sciences MSc thesis, Wageningen University.
NELSON, D. E., REPETTI, P. P., ADAMS, T. R., CREELMAN, R. A., WU, J., WARNER, D. C., ANSTROM, D. C., BENSEN, R. J., CASTIGLIONI, P. P. & DONNARUMMO, M. G. 2007. Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance and lead to improved corn yields on water-limited acres. Proceedings of the National Academy of Sciences, 104, 16450-16455.
OGUNNIYI, D. 2006. Castor oil: A vital industrial raw material. Bioresource technology, 97, 1086-1091.
OLIVARES, A. R., CARRILLO-GONZÁLEZ, R., GONZÁLEZ-CHÁVEZ, M. D. C. A. & HERNÁNDEZ, R. M. S. 2013. Potential of castor bean (Ricinus communis L.) for phytoremediation of mine tailings and oil production. Journal of environmental management, 114, 316-323.
OLSEN, A. N., ERNST, H. A., LEGGIO, L. L. & SKRIVER, K. 2005. NAC transcription factors: structurally distinct, functionally diverse. Trends in plant science, 10, 79-87.
OLSON, M. O., DUNDR, M. & SZEBENI, A. 2000. The nucleolus: an old factory with unexpected capabilities. Trends in cell biology, 10, 189-196.
OOKA, H., SATOH, K., DOI, K., NAGATA, T., OTOMO, Y., MURAKAMI, K., MATSUBARA, K., OSATO, N., KAWAI, J. & CARNINCI, P. 2003. Comprehensive analysis of NAC family genes in Oryza sativa and Arabidopsis thaliana. DNA research, 10, 239-247.
PANDEY, G. K., GRANT, J. J., CHEONG, Y. H., KIM, B. G., LI, L. & LUAN, S. 2005. ABR1, an APETALA2-domain transcription factor that functions as a repressor of ABA response in Arabidopsis. Plant Physiology, 139, 1185-1193.
PAUL, M. J., PRIMAVESI, L. F., JHURREEA, D. & ZHANG, Y. 2008. Trehalose metabolism and signaling. Annu. Rev. Plant Biol., 59, 417-441.
PAVLETICH, N. P. & PABO, C. O. 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science, 252, 809-817.
PERIAPPURAM, C., STEINHAUER, L., BARTON, D. L., TAYLOR, D. C., CHATSON, B. & ZOU, J. 2000. The plastidic phosphoglucomutase from Arabidopsis. A reversible enzyme reaction with an important role in metabolic control. Plant Physiology, 122, 1193-1200.
PETRONI, K., KUMIMOTO, R. W., GNESUTTA, N., CALVENZANI, V., FORNARI, M., TONELLI, C., HOLT, B. F. & MANTOVANI, R. 2012. The promiscuous life of plant NUCLEAR FACTOR Y transcription factors. The Plant Cell Online, 24, 4777-4792.
PNUELI, L., HALLAK□HERR, E., ROZENBERG, M., COHEN, M., GOLOUBINOFF, P., KAPLAN, A. & MITTLER, R. 2002. Molecular and biochemical mechanisms associated with dormancy and drought tolerance in the desert legume Retama raetam. The Plant Journal, 31, 319-330.
PRÄNDL, R., HINDERHOFER, K., EGGERS-SCHUMACHER, G. & SCHÖFFL, F. 1998. HSF3, a new heat shock factor from Arabidopsis thaliana, derepresses the heat shock response and confers thermotolerance when overexpressed in transgenic plants. Molecular and General Genetics MGG, 258, 269-278.
PROCHNIK, S., MARRI, P. R., DESANY, B., RABINOWICZ, P. D., KODIRA, C., MOHIUDDIN, M., RODRIGUEZ, F., FAUQUET, C., TOHME, J. & HARKINS, T. 2012. The cassava genome: current progress, future directions. Tropical plant biology, 5, 88-94.
PRUITT, K. D., TATUSOVA, T. & MAGLOTT, D. R. 2005. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic acids research, 33, D501-D504.
PUNTA, M., COGGILL, P. C., EBERHARDT, R. Y., MISTRY, J., TATE, J., BOURSNELL, C., PANG, N., FORSLUND, K., CERIC, G. & CLEMENTS, J. 2011. The Pfam protein families database. Nucleic acids research, gkr1065.
PURSEGLOVE, J. 1976. The origins and migrations of crops in tropical Africa. Origins of African Plant Domestication. Mouton, The Hague, 291-309.
QUACH, T. N., NGUYEN, H. T., VALLIYODAN, B., JOSHI, T., XU, D. & NGUYEN, H. T. 2014. Genome-wide expression analysis of soybean NF-Y genes reveals potential function in development and drought response. Molecular Genetics and Genomics, 1-21.
QUETTIER, A.-L. & EASTMOND, P. J. 2009. Storage oil hydrolysis during early seedling growth. Plant Physiology and Biochemistry, 47, 485-490.
RENTSCH, D., SCHMIDT, S. & TEGEDER, M. 2007. Transporters for uptake and allocation of organic nitrogen compounds in plants. Febs Letters, 581, 2281-2289.
RHOADS, D. M., WHITE, S. J., ZHOU, Y., MURALIDHARAN, M. & ELTHON, T. E. 2005. Altered gene expression in plants with constitutive expression of a mitochondrial small heat shock protein suggests the involvement of retrograde regulation in the heat stress response. Physiologia Plantarum, 123, 435-444.

RIBEIRO DE JESUS, P. 2014. RE: Personal communication.
RIBEIRO DE JESUS, P., WILLEMS, L., MUTIMAWURUGO, M., FERNANDEZ, L., DE CASTRO, R., LIGTERINK, W. & HILHORST, H. 2014. Biochemical and molecular requirements of Ricinus commnis germination for successful seedling establishment at different temperatures.
RIBEIRO DE JESUS, P., FERNANDEZ, L., DE CASTRO, R., LIGTERINK, W. & HILHORST, H. 2014a. Physiological and biochemical resposnses of Ricinus cummunis seedlings to different temperatures: a metabolomics approach. BMC plant biology, 14.
RIECHMANN, J., HEARD, J., MARTIN, G., REUBER, L., JIANG, C.-Z., KEDDIE, J., ADAM, L., PINEDA, O., RATCLIFFE, O. & SAMAHA, R. 2000. Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science, 290, 2105-2110.
RIZHSKY, L., LIANG, H. & MITTLER, R. 2002. The combined effect of drought stress and heat shock on gene expression in tobacco. Plant Physiology, 130, 1143-1151.
RIZHSKY, L., LIANG, H., SHUMAN, J., SHULAEV, V., DAVLETOVA, S. & MITTLER, R. 2004. When defense pathways collide. The response of Arabidopsis to a combination of drought and heat stress. Plant physiology, 134, 1683-1696.
ROLLETSCHEK, H., HOSEIN, F., MIRANDA, M., HEIM, U., GÖTZ, K.-P., SCHLERETH, A., BORISJUK, L., SAALBACH, I., WOBUS, U. & WEBER, H. 2005. Ectopic expression of an amino acid transporter (VfAAP1) in seeds of Vicia narbonensis and pea increases storage proteins. Plant Physiology, 137, 1236-1249.
ROMEIRO, S., LAGÔA, A. M., FURLANI, P. R., ABREU, C. A. D., ABREU, M. F. D. & ERISMANN, N. M. 2006. Lead uptake and tolerance of Ricinus communis L. Brazilian Journal of Plant Physiology, 18, 483-489.
ROUSSEFF, D. 2004. Biodiesel, O novo combustível do Brasil. Programa Nacional de Produção e uso do Biodiesel. Ministério de Minas e Energia.
RUBBI, C. P. & MILNER, J. 2003. Disruption of the nucleolus mediates stabilization of p53 in response to DNA damage and other stresses. The EMBO journal, 22, 6068-6077.
SAILAJA, M., TARAKESWARI, M. & SUJATHA, M. 2008. Stable genetic transformation of castor (Ricinus communis L.) via particle gun-mediated gene transfer using embryo axes from mature seeds. Plant cell reports, 27, 1509-1519.
SAKAMOTO, H., MARUYAMA, K., SAKUMA, Y., MESHI, T., IWABUCHI, M., SHINOZAKI, K. & YAMAGUCHI-SHINOZAKI, K. 2004. Arabidopsis Cys2/His2-type zinc-finger proteins function as transcription repressors under drought, cold, and high-salinity stress conditions. Plant physiology, 136, 2734-2746.
SAMACH, A., ONOUCHI, H., GOLD, S. E., DITTA, G. S., SCHWARZ-SOMMER, Z., YANOFSKY, M. F. & COUPLAND, G. 2000. Distinct roles of CONSTANS target genes in reproductive development of Arabidopsis. Science, 288, 1613-1616.
SÁNCHEZ, R., SERRA, F., TÁRRAGA, J., MEDINA, I., CARBONELL, J., PULIDO, L., DE MARÍA, A., CAPELLA-GUTÍERREZ, S., HUERTA-CEPAS, J. & GABALDÓN, T. 2011. Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. Nucleic acids research, 39, W470-W474.
SANGWAN, V., ÖRVAR, B. L., BEYERLY, J., HIRT, H. & DHINDSA, R. S. 2002. Opposite changes in membrane fluidity mimic cold and heat stress activation of distinct plant MAP kinase pathways. The Plant Journal, 31, 629-638.
SANMIYA, K., SUZUKI, K., EGAWA, Y. & SHONO, M. 2004. Mitochondrial small heat-shock protein enhances thermotolerance in tobacco plants. FEBS letters, 557, 265-268.
SAUSEN, T. L. & ROSA, L. M. G. 2010. Growth and carbon assimilation limitations in Ricinus communis (Euphorbiaceae) under soil water stress conditions. Acta Botanica Brasilica, 24, 648-654.
SCARPA, A. & GUERCI, A. 1982. Various uses of the castor oil plant (Ricinus communis L.) a review. Journal of ethnopharmacology, 5, 117-137.
SCHARF, K.-D., SIDDIQUE, M. & VIERLING, E. 2001. The expanding family of Arabidopsis thaliana small heat stress proteins and a new family of proteins containing α-crystallin domains (Acd proteins). Cell stress & chaperones, 6, 225.
SENDON, P. M., OO, M. M., PARK, J.-B., LEE, B. H., KIM, J. H., SEO, H. S., PARK, S.-K. & SONG, J. T. 2014. Activation of C2H2-type zinc finger genes induces dwarfism in Arabidopsis thaliana. Journal of the Korean Society for Applied Biological Chemistry, 57, 35-41.
SEVERINO, L. S. & AULD, D. L. 2013. Seed yield and yield components of castor influenced by irrigation. Industrial Crops and Products, 49, 52-60.
SEVERINO, L. S., AULD, D. L., BALDANZI, M., CÂNDIDO, M. J., CHEN, G., CROSBY, W., TAN, D., HE, X., LAKSHMAMMA, P. & LAVANYA, C. 2012. A review on the challenges for increased production of castor. Agronomy Journal, 104, 853-880.
SIEFERS, N., DANG, K. K., KUMIMOTO, R. W., BYNUM, W. E., TAYROSE, G. & HOLT, B. F. 2009. Tissue-specific expression patterns of Arabidopsis NF-Y transcription factors suggest potential for extensive combinatorial complexity. Plant physiology, 149, 625-641.
SILLITOE, I., CUFF, A. L., DESSAILLY, B. H., DAWSON, N. L., FURNHAM, N., LEE, D., LEES, J. G., LEWIS, T. E., STUDER, R. A. & RENTZSCH, R. 2012. New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures. Nucleic acids research, gks1211.
SOUER, E., VAN HOUWELINGEN, A., KLOOS, D., MOL, J. & KOES, R. 1996. The no apical meristem gene of Petunia is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. Cell, 85, 159-170.

STAMLER, R., KAPPÉ, G., BOELENS, W. & SLINGSBY, C. 2005. Wrapping the α-crystallin domain fold in a chaperone assembly. Journal of molecular biology, 353, 68-79.
STEPHENSON, T. J., MCINTYRE, C. L., COLLET, C. & XUE, G.-P. 2011. TaNF-YB3 is involved in the regulation of photosynthesis genes in Triticum aestivum. Functional & integrative genomics, 11, 327-340.
SUGIKAWA, Y., EBIHARA, S., TSUDA, K., NIWA, Y. & YAMAZAKI, K.-I. 2005. Transcriptional coactivator MBF1s from Arabidopsis predominantly localize in nucleolus. Journal of plant research, 118, 431-437.
SUJATHA, M., REDDY, T. P. & MAHASI, M. 2008. Role of biotechnological interventions in the improvement of castor (Ricinus communis L.) and Jatropha curcas L. Biotechnology Advances, 26, 424-435.
SUJATHA, M. & SAILAJA, M. 2005. Stable genetic transformation of castor (Ricinus communis L.) via Agrobacterium tumefaciens-mediated gene transfer using embryo axes from mature seeds. Plant cell reports, 23, 803-810.
SUZUKI, N., BAJAD, S., SHUMAN, J., SHULAEV, V. & MITTLER, R. 2008. The transcriptional co-activator MBF1c is a key regulator of thermotolerance in Arabidopsis thaliana. Journal of Biological Chemistry, 283, 9269-9275.
SUZUKI, N. & MITTLER, R. 2006. Reactive oxygen species and temperature stresses: a delicate balance between signaling and destruction. Physiologia Plantarum, 126, 45-51.
SUZUKI, N., RIZHSKY, L., LIANG, H., SHUMAN, J., SHULAEV, V. & MITTLER, R. 2005. Enhanced tolerance to environmental stress in transgenic plants expressing the transcriptional coactivator multiprotein bridging factor 1c. Plant Physiology, 139, 1313-1322.
SVENNERSTAM, H. 2008. Amino acid uptake in Arabidopsis.
TAGUE, B. W., GALLANT, P. & GOODMAN, H. M. 1996. Expression analysis of an Arabidopsis C2H2 zinc finger protein gene. Plant molecular biology, 32, 785-796.
TAGUE, B. W. & GOODMAN, H. M. 1995. Characterization of a family of Arabidopsis zinc finger protein cDNAs. Plant molecular biology, 28, 267-279.
TAIR 2014. The Arabidopsis Information Resource. www.arabidopsis.org: Phoenix Bioinformatics.
TAKATSUJI, H. & MATSUMOTO, T. 1996. Target-sequence recognition by separate-type Cys2/His2 zinc finger proteins in plants. Journal of Biological Chemistry, 271, 23368-23373.
TAKEMARU, K.-I., LI, F.-Q., UEDA, H. & HIROSE, S. 1997. Multiprotein bridging factor 1 (MBF1) is an evolutionarily conserved transcriptional coactivator that connects a regulatory factor and TATA element-binding protein. Proceedings of the National Academy of Sciences, 94, 7251-7256.
TEGEDER, M. 2012. Transporters for amino acids in plant cells: some functions and many unknowns. Current opinion in plant biology, 15, 315-321.
THOMPSON, J. D., GIBSON, T. & HIGGINS, D. G. 2002. Multiple sequence alignment using ClustalW and ClustalX. Current protocols in bioinformatics, 2.3. 1-2.3. 22.
TOJO, T., TSUDA, K., YOSHIZUMI, T., IKEDA, A., YAMAGUCHI, J., MATSUI, M. & YAMAZAKI, K.-I. 2009. Arabidopsis MBF1s control leaf cell cycle and its expansion. Plant and cell physiology, 50, 254-264.
TOLBERT, N. 1981. Metabolic pathways in peroxisomes and glyoxysomes. Annual review of biochemistry, 50, 133-157.
TSUDA, K., TSUJI, T., HIROSE, S. & YAMAZAKI, K.-I. 2004. Three Arabidopsis MBF1 homologs with distinct expression profiles play roles as transcriptional co-activators. Plant and cell physiology, 45, 225-231.
TSUDA, K. & YAMAZAKI, K.-I. 2004. Structure and expression analysis of three subtypes of Arabidopsis MBF1 genes. Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression, 1680, 1-10.
VAN AKEN, O., ZHANG, B., CARRIE, C., UGGALLA, V., PAYNTER, E., GIRAUD, E. & WHELAN, J. 2009. Defining the mitochondrial stress response in Arabidopsis thaliana. Molecular Plant, 2, 1310-1324.
VIERLING, E. 1991. The roles of heat shock proteins in plants. Annual review of plant biology, 42, 579-620.
WAHID, A., GELANI, S., ASHRAF, M. & FOOLAD, M. 2007. Heat tolerance in plants: an overview. Environmental and Experimental Botany, 61, 199-223.
WANG, W., VINOCUR, B., SHOSEYOV, O. & ALTMAN, A. 2004. Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. Trends in plant science, 9, 244-252.
WEHMEYER, N. & VIERLING, E. 2000. The expression of small heat shock proteins in seeds responds to discrete developmental signals and suggests a general protective role in desiccation tolerance. Plant Physiology, 122, 1099-1108.
WEI, P.-C., TAN, F., GAO, X.-Q., ZHANG, X.-Q., WANG, G.-Q., XU, H., LI, L.-J., CHEN, J. & WANG, X.-C. 2010. Overexpression of AtDOF4. 7, an Arabidopsis DOF family transcription factor, induces floral organ abscission deficiency in Arabidopsis. Plant physiology, 153, 1031-1045.
WENKEL, S., TURCK, F., SINGER, K., GISSOT, L., LE GOURRIEREC, J., SAMACH, A. & COUPLAND, G. 2006. CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of Arabidopsis. The Plant Cell Online, 18, 2971-2984.
WINTER, D., VINEGAR, B., NAHAL, H., AMMAR, R., WILSON, G. V. & PROVART, N. J. 2007. An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. PloS one, 2, e718.
WOLFE, S. A., NEKLUDOVA, L. & PABO, C. O. 2000. DNA recognition by Cys2His2 zinc finger proteins. Annual review of biophysics and biomolecular structure, 29, 183-212.
YANAGISAWA, S. & SCHMIDT, R. J. 2000. Diversity and similarity among recognition sequences of Dof transcription factors. The Plant Journal, 17, 209-214.
YU, C. S., CHEN, Y. C., LU, C. H. & HWANG, J. K. 2006. Prediction of protein subcellular localization. Proteins: Struc-

ture, Function, and Bioinformatics, 64, 643-651.
ZHANG, Y., ZAGNITKO, O., RODIONOVA, I., OSTERMAN, A. & GODZIK, A. 2011. The FGGY carbohydrate kinase family: insights into the evolution of functional specificities. PLoS computational biology, 7, e1002318.
ZHAO, H., MA, H., YU, L., WANG, X. & ZHAO, J. 2012. Genome-wide survey and expression analysis of amino acid transporter gene family in rice (Oryza sativa L.). PloS one, 7, e49210.