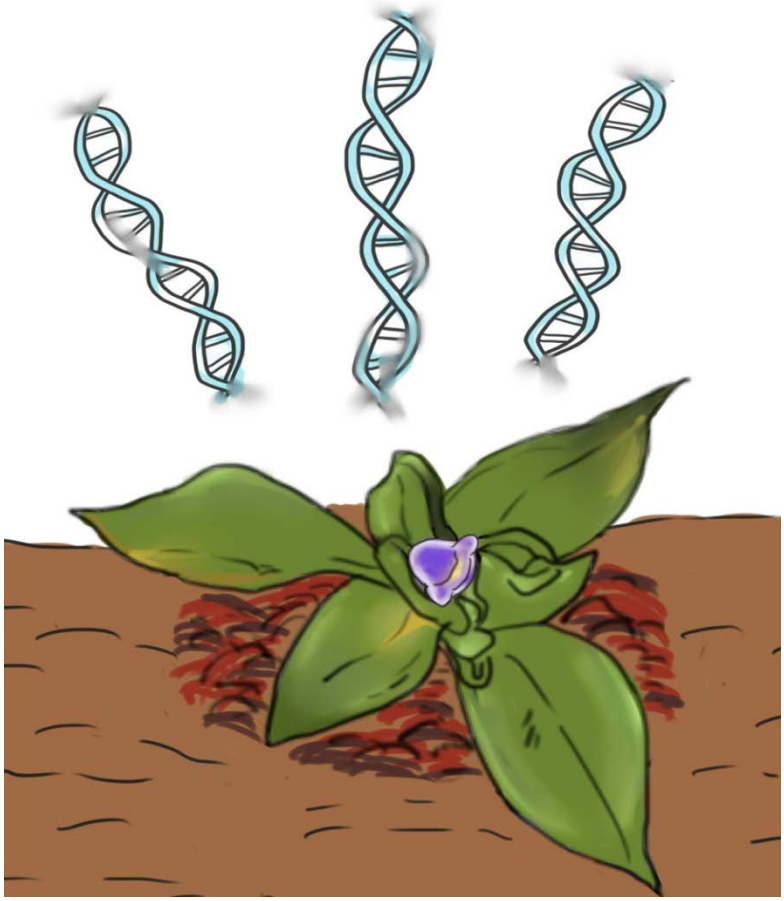
Analysis of two selected cell wall proteins and one lncRNA involved in desiccation tolerance of the resurrection plant *Craterostigma plantagineum*

Peilei Chen



Analysis of two selected cell wall proteins and one lncRNA involved in desiccation tolerance of the resurrection plant *Craterostigma plantagineum*

Dissertation

Zur Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Peilei Chen

aus Handan, Hebei, China

Bonn, 2019

Angefertigt mit Genehmigung der

Mathematisch-Naturwissenschaftlichen Fakultät der

Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Dorothea Bartels
2. Gutachter: Prof. Dr. Lukas Schreiber

Tag der Promotion: 19.12.2019

Erscheinungsjahr: 2020

CONTENTS

ABBREVIATIONS

A: adenine
aa: amino acid
ABA: abscisic acid
AGO: Argonaute
ANOVA: analysis of variance
APS: ammonium persulfate
β-ME: β-Mercaptoethanol
BiFC: bimolecular fluorescence complementation
BLAST: basic local alignment search tool
bp: base pair
BSA: bovine serum albumin
C: cytosine
CBB: coomassie brilliant blue
cDNA: complementary DNA
CDS: coding sequence
CDTA: 1, 2-cyclohexanediaminetetraacetic acid
CK: control
Col–0: Columbia–0
COLDAIR: cold assisted intronic noncoding RNA
COOLAIR: cold induced long antisense intragenic RNA
CRP1: cysteine-rich protein 1
CV: column volume
d: day
D: Dalton
DCL: Dicer-like protein
DMSO: dimethyl sulfoxide
DNA: deoxyribonucleic acid
DNase: deoxyribonuclease
dNTP: deoxyribonucleotide triphosphate
DTT: dithiothreitol
EDTA: ethylene diamintetraacetatic acid
EGF: epidermal growth factor
EL: Electrolyte leakage
ELISA: Enzyme Linked Immunosorbent Assay
FLC: flowering locus C
flg22: a conserved peptide motif in flagellar protein
FPKM: fragments per kilobase of exon per million fragments mapped
g: gram
g: acceleration
GFP: green fluorescent protein
GLP: germin-like protein
GO: Gene ontology
GRP: glycine-rich protein
GST: Gluthation-S-transferase
GTE: Glucose/Tris/EDTA
GUS: β-glucuronidase
h: hour
H_2O_2: hydrogen peroxide
HD-Zip: homeodomain-leucine zipper
His: histidine
IB: inclusion body
IgG: class G immunoglobulin
INA: 2,6-dichloroisonicotinic acid
IPS1: induced by phosphate starvation 1
IPTG: Isopropyl-β-D-thiogalactopyranoside
KAPP: kinase-associated protein phosphatase
kb: kilobase
kDa: kilodalton
KEGG: Kyoto Encyclopedia of Genes and Genomes
LB: Luria and Bertani medium
lincRNA: long intergenic ncRNA
lncRNA: long non-coding RNA
M: Molar
MeJA: methyl jasmonic acid
min: minute
miRNA: microRNA
ml: milliliter
MS: Murashige and Skoog
mRNA: messenger RNA
MW: molecular weight
NAA: naphthaleneacetic acid
NAT: natural antisense transcript
ncRNA: non-coding RNA
nm: nanometers
NPR: nonexpressor of pathogenesis-related genes
nt: nucleotide

OGs: oligo-galacturonides
ON: overnight
OX: *Arabidopsis thaliana* lines overexpressing the lncRNA 28852
OXO: oxalate oxidase
PBS: phosphate-buffered salt solution
PCR: polymerase chain reaction
PGA: polygalacturonic acid
PHO1;2: PHOSPHATE1;2
PHO2: ubiquitin-conjugating E2 enzyme
pI: theoretical isoelectric point
PME: pectin metyl esterase
PMSF: phenylmethanesulfonyl fluoride
Pol: polymerases
R: arginine
RdDM: RNA-directed DNA methylation
RDR: RNA-dependent RNA polymerases
RLKs: Receptor-like kinases
RNA: ribonucleic acid
RNase: ribonuclease
ROS: reactive oxygen species
rpm: rounds per minute
RT: room temperature
RT-PCR: reverse transcription-polymerase chain reaction
SA: salicylic acid/ salicylate
SDS: sodium dodecyl sulfate
sec: second
siRNA: small interfering RNA
SOD: superoxide dismutase
sRNAs: small RNAs
TAE: tris-acetate-EDTA
TARs: transcriptionally active regions
Taq: Thermophilus aquaticus
TBS: tris– buffered salt solution
TBST: tris– buffered salt solution with tween
TEMED: tetramethylethylendiamine
TMB: 3,3',5,5'-tetramethylbenzidine
Tris: tris-(hydroxymethyl)-aminomethane
Triton: X-100Poly(ethylenglycolether)n-octylphenol
Trp: tryptophan
V: volts
v/v: volume/volume
w/v: weight/volume
WAK: wall-associated kinase
WAK2cTAP: the dominant allele of WAK2
WAKL: WAK-like
WT: wild type
X: times
Y2H: yeast-two-hybrid
YEB: yeast extract broth
YFP: yellow fluorescent protein

List of Tables

List of Figures

SUMMARY

Water deficit is a severe stress negatively affecting plant growth and productivity. Extreme dehydration, termed desiccation, will lead to substantial water loss from cells and eventually results in less than 10% RWC (relative water content). Resurrection plants are a small group of desiccation-tolerant angiosperms, which are able to circumvent or overcome the desiccation stress in their unique ways. *Craterostigma plantagineum* Hochst. (Scrophulariaceae) is the model resurrection plant for dissecting the mechanisms underlying desiccation tolerance. It can revive after being desiccated for more than two years. The recovery requires cell wall remodeling during dehydration and rehydration. The cell wall protein CpWAK1 (*C. plantagineum* wall associated kinase 1) is supposed to be involved in cell wall remodeling by forming a CpWAK1-CpGRP1 (*C. plantagineum* glycine-rich protein 1) complex. This interaction between CpWAK1 and CpGRP1 was demonstrated *in yeast* and *in planta*. The activity of the complex was envisaged to be affected by cell wall pectin given that AtWAKs bind to pectins *in vivo* and *in vitro*. However, neither the binding of CpWAK1 to pectins nor the role of the CpWAK1-CpGRP1 complex has been investigated. In addition, there are other candidate genes implicated in the resistance to desiccation in *C.plantagineum*. The cell wall protein gene *CpCRP1* (*C.plantagineum* cysteine-rich protein 1) and the *lncRNA 28852* (long non-coding RNA 28852) were taxonomically restricted genes responsive to desiccation. The CpGLP1 (*C. plantagineum* germin-like protein) was identified as a putative interaction partner of CpCRP1 in a yeast-two-hybrid assay. Like many other GLPs, CpGLP1 also has SOD activity. The interaction between CpGLP1 and CpCRP1 *in planta* and whether CpGLP1 participates in the cross linking of cell walls under drought still need to be explored. A substantial amount of plant lncRNAs were discovered. Some of them are characterized as target mimicry of miRNAs, precursors of sRNAs or the scaffolds for chromatin modifiers. The newly identified taxonomically restricted gene *lncRNA 28852* is induced by dehydration, especially desiccation. However, its functions and the mechanisms it participates in are still obscure.

In this thesis, the main study is focused on CpWAKs. To understand the genome organization and the evolutionary patterns of CpWAKs, the full-length sequences and the upstream and downstream sequences of *CpWAK* genes were identified using genome-walking. The alignments of *CpWAK* genes and the phylogenetic analyses using CpWAK protein sequences suggest gene duplications for *CpWAK* genes after species divergence. The expression profiles of CpWAKs on the transcript and protein levels reveal that CpWAKs are involved in various biological processes, including regulation of the circadian clock, drought-induced responses and SA- and JA-related plant resistance to pathogens and wounding. The interactions of CpWAKs with pectins and CpGRP1 *in vitro* were demonstrated by ELISA assays. In the presence of DTT no band shift was seen on protein gels. This indicates that the formation of CpWAK multimers is the prerequisite for the CpWAK-pectin linkage. CpGRP1 shows a positive effect on the binding of CpWAKs to pectins in the ELISA assays. This effect was detected at different pH values as well. The CpWAK-pectin binding was determined using different pectin extracts. Different pectin extracts lead to the opposite trend of CpWAK-pectin binding in the presence of Ca^{2+} at pH 8. All of these observations collectively demonstrate that CpWAKs are potential sensors which can distinguish

the cell wall signals caused by diverse stimuli with the aid of other elements, such as CpGRPs, pH_{apo}, $Ca^{2+}{}_{[apo]}$ and the formation of CpWAK multimers or dimers.

Pull-down assays and BiFC assays did not confirm the interaction of CpCRP1 and CpGLP1 *in vitro* or *in planta*. But the detection of CpGLP1 in cell wall protein extracts and its accumulation in dehydrated and rehydrated leaves showed that it may also affect the cell wall remodeling during dehydration and rehydration.

The expression of the lncRNA 28852 was analyzed on total and polysomal RNA levels. The lncRNA 28852 is abundantly expressed during desiccation on the total RNA level, while the lncRNA 28852 from polysomal RNA only accumulated during dehydration and rehydration not during desiccation. The opposite accumulation of the lncRNA 28852 implies the interaction of lncRNA 28852 and ribosomes in the early stage of dehydration and rehydration. This interaction is considered an effective protection against RNA degradation and contributes to the accumulation of lncRNA 28852 under desiccation. The down-regulated lncRNA 28852 in polysomal RNA of desiccated leaves indicates that the lncRNA 28852 is presumably not implicated in protein translation during desiccation. Because no similarity to the sequences in the GenBank database was found, *Arabidopsis thaliana* lines overexpressing the lncRNA 28852 were established for functional analysis. To further uncover possible processes in which the lncRNA 28852 is involved, comparative RNA-seq analysis of wild-type and mutants was performed. The non-repeatable gene expression patterns and the non-conspicuous changes in the phenotypes of the mutants suggest the ineffective role or the intricate effect of *C.plantagineum* lncRNA 28852 in *Arabidopsis thaliana.*

1. INTRODUCTION

1.1 Plant cell wall proteins

Plant cell wall is a highly organized macromolecular gel-like structure (Vorwerk et al. 2004), including a dynamic primary cell wall, middle lamella and a relatively static secondary cell wall (Caffall and Mohnen 2009; Maureen C. McCanna et al. 2001; York. et al. 1986) (**Figure 1.1**). The primary cell wall, as a plastic and elastic cell layer, is implicated in cell expansion, cell adhesion and signal transduction (Caffall and Mohnen 2009; Fry 2004). The composition of the plant cell wall is complex. Apart from water which constitutes the majority of the growing cell wall mass, more than 90% of the primary wall consists of polysaccharides, proteins, aromatic and aliphatic compounds (Caffall and Mohnen 2009; Cosgrove 1997). Polysaccharides are the most abundant components in the cell wall, whereas proteins are a minor but important component (Cosgrove 1997). Cell wall proteins are ubiquitous and indispensable in the plant kingdom.

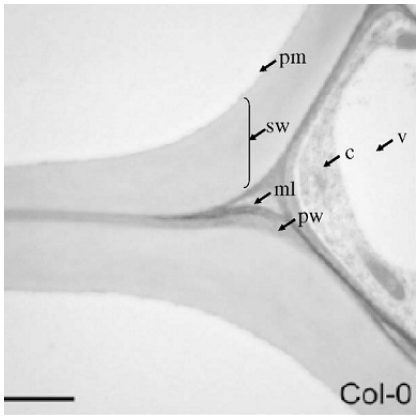


Figure 1.1 A transmission electron micrograph of root cell wall from *Arabidopsis thaliana*. pm: plasma membrane, sw: secondary wall, ml: middle lamella, pw: primary wall, c: cytosol, v: vacuole, Bar = 2 µM. (Caffall and Mohnen 2009)

Cell wall proteins are classified based on their structural characteristics like being rich in certain amino acids, containing highly repetitive sequence domains, or being glycosylated, e.g. glycine-rich protein (GRP), proline-rich proteins, arabinogalactan proteins and so on (Cassab 1998). However, with the development of cell wall protein studies more and more wall proteins with mixed structural characteristics are discovered (Carpita et al. 1996). The cell wall protein functions including both structural and physiological roles depend on the structural characteristics and localization. One class of cell wall modifying enzymes are expansins which are capable to loosen the cell wall during growth by breaking down the wall polysaccharides (Cosgrove 2000). The cell wall peroxidases are able to facilitate wall loosening and stiffening due to its dual hydroxylic and peroxidative cycles (Francoz et al. 2015). Giarola et al. (2015) identified a cell wall localized cysteine-rich protein (CpCRP1) which has a potential to be implicated in cell wall remodeling during rehydration. Wall-associated kinases (WAKs) are reported to be involved in cell wall signal transduction, cell expansion and stress responses because of their transmembrane localization and the cytoplasmic kinase domain (Kohorn 2015). Germins or germin-like proteins (GLPs) not only have a structural function, but possess an enzymatic activity (oxalate oxidase or superoxide dismutase) as well, which may be implicated in plant growth and responses to various stresses (Bernier and Berna 2001; Dunwell et al. 2008). This chapter will focus on the two cell wall proteins, WAKs and GLPs.

1.1.1 Cell Wall-associated kinases (WAKs)

Wall-associated kinases, or WAKs, are receptor protein kinases, which have a cytoplasmic protein kinase domain, and span the plasma membrane. The extracellular domains can bind to both the oligo-galacturonides (OGs) released from the cell wall and the cross-linked pectin. According to their pectin-binding capacity, WAKs are thought to be involved in cell expansion, responses to wounding and pathogen.

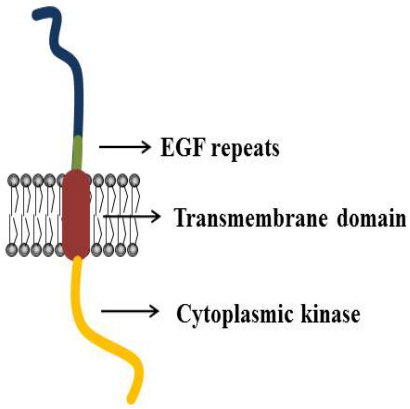
1.1.1.1 *WAK-like* (WAKL) genes

WAK genes were first identified in Arabidopsis. WAK proteins have also been detected in several other higher plants with AtWAK1 antiserum, such as pea, tobacco and maize, but not in algae (He et al. 1996). In Arabidopsis, five *WAKs* were identified by DNA sequences and PCR analysis (He et al. 1999), and twenty-two *WAKL* genes were identified through BLAST analysis (Verica and He 2002). Although *AtWAK/WAKL* genes are present on all five chromosomes, the majority of them are located on chromosome I, among which, *AtWAK1-5* and *AtWAKL1-7* are tandemly arranged in clusters (Verica et al. 2003; Verica and He 2002). Based on the genome analysis of Arabidopsis, the expansion of *AtWAK/WAKL* gene family is considered to be acquired *via* tandem duplication, segmental duplications or even a retrotransposon (Verica and He 2002). The gene structures of *AtWAK/WAKLs* are conserved with three exons and two introns except for five *AtWAKL* genes (*WAKL7*, *WAKL8*, *WAKL12*, *WAKL16*, and *WAKL19*) (Verica and He 2002). In other species, such as rice and apple, the number of *WAKL* genes expands to 125 (Zhang et al. 2005) and 44 (Zuo et al. 2018), respectively. Like the Arabidopsis *WAK/WAKLs* the *WAK/WAKL* genes in rice are distributed on all 12 chromosomes and most of the rice *WAKL*s are clustered (Zhang et al. 2005). These adjacent small *OsWAK* clusters suggest that localized gene duplications presumably are the main reason for the *OsWAK* gene family expansion (Shiu et al. 2004; Zhang et al. 2005). The conserved two-intron gene structure of *WAK/WAKL* also exists in rice, apple and barley despite the observation of extra/missed introns in *OsWAKs* and *MdWAK* genes (Kaur et al. 2013; Zhang et al. 2005; Zuo et al. 2018). Many *WAK/WAKL* genes were identified in other angiosperms, like tomato, *Craterostigma*, maize, wheat and *Brachypodium* (Giarola et al. 2016; Hurni et al. 2015; Liu et al. 2006; Rosli et al. 2013; Wu 2016).

1.1.1.2 Structure of WAK proteins

WAKs are a subset of the WAK-like superfamily which are characterized by a conserved cytoplasmic Ser/Thr kinase domain and the presence of EGF (epidermal growth factor) repeats in variable extracellular domains (Kohorn 2015; Kohorn and Kohorn 2012) (**Figure 1.2**). Among the 27 AtWAKs/WAKLs, the predicted protein sequences of 22 AtWAKs/AtWAKLs have the typical WAK traits and are divided into four groups (Verica and He 2002). The structural differences among the four groups mainly depend on the numbers and positions of the EGF or degenerate EGF domains (Verica and He 2002). The Group I contains AtWAK1 to 5 (He et al. 1999). The kinase domains and the extracellular domains of the five AtWAKs share 86% and 40-64% identity respectively (He et al. 1999). On the variable extracellular domains, the EGF repeats may

contribute to the calcium-mediated dimerization of proteins with the overlapping Asn sites and the six conserved cysteine residues are involved in the formation of disulfide-bridged complexes (Anderson et al. 2001; He et al. 1999; Verica and He 2002). Other regions in the extracellular domains are similar to the cell wall proteins which suggest the possibility of functioning in a carbohydrate rich environment (He et al. 1999). The remaining *AtWAKL* genes only encode the truncated WAK proteins, of which AtWAKL7, AtWAKL8, and AtWAKL19 may be secreted into the extracellular space to exert an impact on the formation of the active WAKL receptor complex as a result of lacking a transmembrane domain (Verica and He 2002). Apart from the AtWAKs/AtWAKLs, the WAKs in other species like HvWAK1 in barley (Kaur et al. 2013), Htn1 in maize (Hurni et al. 2015) and CpWAKs in *Craterostigma* (Giarola et al. 2016) also harbour a characteristic extracellular region and a conserved cytoplasmic kinase domain. In rice, the expanded OsWAK gene family includes 67 OsWAK-RLKs containing both EGF repeats and kinase domains, 41 abbreviated OsWAKs, 12 OsWAKs encoding a 300 aa peptide with an unknown domain and 5 pseudogenes (Zhang et al. 2005). The large number of OsWAK-RLKs raises the possibility that OsWAKs are implicated in multiple biological processes in rice.



In conclusion, the diversity of the extracellular domains of WAK proteins is the prerequisite for distinguishing the different signals in cell wall and the conserved kinase domain indicates that different WAK members may regulate diverse intermediate steps in the same biological process (Wagner and Kohorn 2001).

Figure 1.2 Cartoon delineating the general structures of WAKs (Modified from Anderson et al. 2001). The conserved cytoplasmic kinase, transmembrane domain and extracellular EGF repeats are depicted as yellow, dark red and light green respectively.

1.1.1.3 Expression of WAK genes and proteins

To provide more information on the analyses of WAK functions, many WAK expression profiles were explored. WAKs show distinct expression patterns with a wealth of overlap which implies the formation of WAK homo- or hetero-dimers (Wagner and Kohorn 2001). In Arabidopsis, the five AtWAKs are mainly expressed in the vegetative organs, leaves and stems, except for AtWAK4 which is primarily detected in siliques according to the RNA blot analysis (He et al. 1999; He et al. 1998). The histochemical GUS assay showed that AtWAK1, AtWAK2 and AtWAK3 were all expressed at organ junctions, in shoot and root apical meristems, and in expanding leaves although the expression in some other tissues was also detected (Wagner and Kohorn 2001). Of the 22 AtWAKLs, the expression patterns of seven AtWAKLs were analyzed using gene-specific RT-PCR. The AtWAKL1, AtWAKL3 and AtWAKL5 were expressed mainly in roots and flowers but not in vegetative organs (Verica et al. 2003). Similar to AtWAKs/WAKLs, other WAKs also show tissue-specific and developmentally regulated expression patterns (Kaur et al. 2013; Zhang et al. 2005; Zuo et al. 2018). In wheat, the TaWAKL1 and TaWAKL2 are mainly expressed in the juvenile stage while TaWAK1 and TaWAK3 show stronger expression in adult

stages (Liu et al. 2006). The expression of WAKs is also affected by a range of environmental stimuli. AtWAK1 is induced by pathogen, exogenous salicylate (SA) or its analog 2, 6-dichloroisonicotinic acid (INA) in a NPR (Nonexpressor of pathogenesis-related genes) dependent manner (He et al. 1998). Other AtWAKs are also induced by SA, except for AtWAK4 (He et al. 1999). Among the SA-inducible AtWAKL genes, only the expression of AtWAKL5 and AtWAKL7 is related to NPR1, and the SA-inducible AtWAKs/AtWAKLs are additionally responsive to wounding (Verica et al. 2003; Wagner and Kohorn 2001). Increasing numbers of WAKs in various species have been identified as SA-induced or pathogen-related genes (Czajkowska et al. 2019; Gadaleta et al. 2019; Hu et al. 2014; Hurni et al. 2015; Li et al. 2009; Liu et al. 2006; Meier et al. 2010; Saintenac et al. 2018; Shi et al. 2016; Zuo et al. 2018). Rosli et al. (2013) showed that one tomato gene encoding a cell wall-associated kinase (SIWAK1) was up-regulated by the flagellin microbe-associated molecular pattern flgII-28 but down-regulated by the *Pseudomonas syringae pv. tomato (Pst)* type III effectors AvrPto and AvrPtoB. Apart from the biotic stresses, WAKs also respond to abiotic stresses, such as cold (de Oliveira et al. 2014), heat (Wang et al. 2019) and dehydration (Giarola et al. 2016). The expression of AtWAK1 can be induced in roots by aluminum (Sivaguru et al. 2003) and the AtWAKL4 is responsive to many mineral ions, including Na^+, K^+, Cu^{2+}, Ni^{2+}, and Zn^{2+} (Hou et al. 2005). In rice, the expression of OsWAK11 is related to heavy metal, and it is significantly enhanced in the presence of Al^{3+} and Cu^{2+} (Hu et al. 2014). WAKs/WAKLs as a superfamily contain a substantial amount of members involved in various aspects of the plant life cycle and diverse responses of plants to stresses, which shows the diverse functions in growth and development of plants and the complexity of WAK expressions.

1.1.1.4 WAK proteins bind to pectins

WAK proteins as the name states are associated with the cell wall, which was demonstrated by immunohistochemistry (He et al. 1996). Recently, Giarola et al. (2016) showed that CpWAKs were secreted into the apoplastic space under the guidance of signal peptides. Binding of WAKs to pectin was first confirmed using different cell wall degrading enzymes. Among four cell wall degrading enzymes (hemicellulase, xyloglucanase, cellulase, and pectinase), only pectinase released WAKs from the cell wall, and the binding to WAKs was still detectable even on denaturing gels which implied covalent binding (Anderson et al. 2001; Kohorn and Kohorn 2012; Wagner and Kohorn 2001). The covalent binding was identified for two AtWAKs, WAK1 and WAKL6 (Verica et al. 2003). Subsequent experiments showed that the WAK-pectin association *in vitro* may partially rely on ionic bonds, as both the mutation of the positively charged residues (arginines and lysines) in WAK1 and the methyl esterification of the negatively charged oxygen groups in pectin led to a weaker linkage (Decreux and Messiaen 2005; Decreux et al. 2006; Kohorn et al. 2009). The purified extracellular domain of WAK1 showed a higher affinity for the egg box conformation of OGs with a dp (degree of polymerization) 9-15 (Cabrera et al. 2008; Decreux and Messiaen 2005). The binding between WAKs and various pectins including homogalacturonan, and rhamnogalacturonans I and II *in vitro* also implied that galacturonic acid may be the key element for WAK-pectin association (Decreux and Messiaen 2005; Kohorn et al. 2009; Kohorn and Kohorn 2012). The chemical modification of the reducing end of the oligogalacturonides did not affect the WAK-pectin binding (Cabrera et al. 2008). Despite the

validation of WAK-pectin binding, there is inadequate evidence to support that all WAKs bind to cell wall compounds.

Although both WAK and pectins are components of the cell wall matrix, their interactions take place at an early stage in a cytoplasmic compartment (Kohorn et al. 2006a). The observation of the green fluorescent protein (WAK1-GFP) expressed in leaf protoplasts showed that WAK1-GFP co-localized with the Golgi, where pectins are synthesized (Willats et al. 2001), and migrated slower than other proteins not associated with the cell wall. The migration of WAK1-GFP was inhibited by cellulose synthase inhibitor and influenced by fucosylated metabolites. The cellulose synthesis is correlated with pectin synthesis (His et al. 2001) despite the different sites of cellulose synthase on the plasmamembrane (Kohorn 2000). Therefore, the assembly and crosslinking of WAKs is presumably coordinated with cellulose synthesis by the indirect effect of pectin (Kohorn et al. 2006a).

1.1.1.5 The kinase activity, ligands and substrates of WAKs

Characterization of a kinase activity and identifying the ligands and substrates are efficient ways to find out how the WAK kinase functions. WAK as the typical tyrosine kinase is able to be autophosphorylated. The AtWAK1 extracted from leaves was detected by anti-phosphothreonine serum, but whether it is autophosphorylated or not in plants is still unknown (Anderson et al. 2001). OsWAK1 also has the ability of autophosphorylation and phosphorylating a zinc finger protein (Li et al. 2009). The AtGRP3 (Glycine-rich protein) was identified as an interaction partner of AtWAK1 in a two-hybrid assay using the extracellular domain of AtWAK1 with no EGF repeats as bait (Kohorn and Kohorn 2012; Park et al. 2001). The AtGRP3-AtWAK1 interaction was validated both by *in vitro* binding assays and *in vivo* immunoprecipitation assays, for which the cysteine-rich C-terminal domain of AtGRP3 is essential (Park et al. 2001). However, as both WAKs and GRPs are large gene families (Mousavi and Hotta 2005), the interaction does not happen with all isoforms. The AtGRP3 was the only one binding to AtWAK1 *in vitro* among the selected AtGRPs (AtGRP-2,3,4,6,7,8), and AtWAK2 was the only one not interacting with AtGRP3 *in vitro* among all five AtWAKs (Park et al. 2001). Nevertheless, the interaction between AtWAK2 and GRP1 in *C. plantagineum* was confirmed *in yeast*, and the CpGRP1 also binds to CpWAK1 *in yeast* and *in planta* (Giarola et al. 2016). In addition to GRPs, the OGs derived from pectins are the candidate for WAK ligands. Both the WAK-pectin-linkage *in vitro* and *in vivo* and the WAK2-dependent responses of some pectin-induced/repressed genes (Kohorn et al. 2009) suggest that OGs are key elements in the WAKs-related signaling pathway. Furthermore, the activation of the chimeric WAK-EFR kinase in a domain swap experiment also suggests that WAKs are the receptors for OGs (Brutus et al. 2010; Kohorn and Kohorn 2012). The suppression of the dominant WAK allele by *PME* (pectin metyl esterase) mutation and the enhanced responses to OGs treatment in *pme3* and *pme3*/WAK2cTAP mutants demonstrate that the de-esterified pectin is required for WAK activation (Kohorn et al. 2014), consistent with the features of WAK-pectin binding. The substrate of WAKs, KAPP (Kinase-associated protein phosphatase, a protein type 2C phosphatase) were first identified by screening a bacterial expression library with a radioactively labeled AtWAKI kinase and then detected in the GRP3-WAK1 complex extracted from Arabidopsis seedlings (Anderson et al. 2001; Park et al. 2001). The binding between KAPP and

the conserved kinase domain of AtWAK1 is independent of the active site of the WAK kinase domain which is different for the binding between KAPP and other receptor kinases (Anderson et al. 2001). Gramegna et al. (2016) discovered the prolonged expression of defense genes, increased H_2O_2 accumulation and enhanced callose deposition in both the overexpressing WAK1 and *kapp* and *grp-3* loss-of funtion mutants under the OGs, flg22 and wound treatments. The mutants were rescued by overexpressing GRP3 in *grp-3* mutants or in the KAPP overexpressing mutant lines. The observation indicated a positive function of AtWAK1 and a negative function of AtGRP3 and KAPP in the OG/flg22/wound-triggered defense responses (Gramegna et al. 2016). Although it is not known how the OGs/GRP3 initiates the defense responses and which specific role KAPP plays in the signaling pathway, it is certain that the putative ligands and substrates are involved in the defense responses.

1.1.1.6 The function of WAK proteins

WAK proteins play distinct but overlapping roles in various aspects of plant life, given their structures, localizations, expression patterns and the interaction partners. The reduced protein expression of AtWAKs caused by the AtWAK4 or AtWAK2 antisense expression impedes plant growth and development through disturbing the cell expansion and elongation (Lally et al. 2001; Wagner and Kohorn 2001). The disruption of cell expansion and elongation is not observed in the plants expressing gene-specific AtWAK1 or AtWAK2 antisense genes (Wagner and Kohorn 2001), which implies the redundant functions of WAKs. However, the growth retardation phenotype of seedlings is only observed under limiting sugar and salt conditions in the AtWAK2 null mutant *wak2-1* but not in other lines with individual loss of function alleles mutation in other WAKs (Kohorn et al. 2006b; Kohorn and Kohorn 2012). The *wak2-1* phenotype can be rescued both by exogenous sucrose, fructose or glucose and the expression of a sucrose phosphate synthase gene, suggesting the requirement of AtWAK2 for sugar metabolism. The transcription and activity of the vacuolar invertase genes involved in turgor maintenance are down regulated in *wak2-1*, which shows the possibility of linking the cell wall sensing to the solute metabolism and turgor maintenance via WAKs (Kohorn et al. 2006b). In *Craterostigma plantagineum*, the cell wall folding during dehydration and during rehydration presumably requires the involvement of WAKs as sensors. This is supported by the expressions of CpWAKs during dehydration and rehydration and the interaction of CpGRP1 and CpWAK1 (Giarola et al. 2016). The dominant allele of WAK2, WAK2cTAP, results in pathogen-induced ROS accumulation and stunted growth, which are rescued in the WAK2cTAP/*mapk6* mutant but not in WAK2cTAP/*mapk3* mutant (Kohorn et al. 2012). In the presence of de-esterified pectins the MAPK3 activity is enhanced in wild-type protoplasts, but not affected in the *wak2* mutant cell (Kohorn et al. 2009). These observations suggest that in the WAK-related processes MAPK3 is required for the cell expansion while MAPK6 is more associated with the stress responses (Kohorn and Kohorn 2012). Many WAKs are reported to be involved in distinct ways in stress responses. Meier et al. (2010) discovered that AtWAKL10 served as a twin-domain, kinase-GC signaling molecule in the biotic stress responses with the dependence on the second messenger cGMP (cyclic guanosine monophosphate). OsWAK11 may facilitate the expression of OsPME14 and thus modify the cell wall to immobilize the excess Cu and avoid its toxicity (Xia et al. 2018). CaWAKL20 isolated from pepper is inhibited by heat stress, and it modulates plant heat tolerance negatively by down-regulating the

expression of ABA-responsive genes in Arabidopsis CaWAKL20-overexpression lines (Wang et al. 2019).

1.1.1.7 The predicted WAK signaling pathway in the cell wall

WAK proteins as plasma membrane-localized kinases are key players in the signal transduction from cell wall to cytoplasm (He et al. 1999). The predicted signaling in which WAKs are involved in is depicted in **Figure 1.3** according to the relevant literatures. The onset of the assembly of pectin-linked WAKs is in the vesicles which are finally merged into the plasmamembrane and localize the pectin-linked WAKs (Kohorn et al. 2006a) (**Figure 1.3**). WAKs participate in cell-expansion-dependent growth and development and SA-inducible stress responses, which have connections to MPK3 and MPK6 respectively (Kohorn 2015; Kohorn and Kohorn 2012) (**Figure 1.3**). As for cell expansion, the vacuolar invertase is required. The vacuolar invertase controls the turgor by regulating the solute changes (Kohorn et al. 2006b). To identify the proteins implicated in the stress responses, both the reverse genetic approaches and the quantitative mass-spectrometric-based phosphoproteomic analysis were performed (Kohorn et al. 2016; Kohorn et al. 2014). A cytoplasmic receptor like kinase ROG2 and the pathogen-related transcription factors EDS1 and PAD4 positively modulate the stress responses, while a putative scaffold protein REM1.3 is likely to be the negative regulator (**Figure 1.3**). KAPP and GRP3 also show the negative effects on the stress responses (Gramegna et al. 2016) (**Figure 1.3**). The stress responses may depend on a balance between the negative and positive effects, which are probably triggered by the signaling recognition events in the cell wall. However, how WAKs distinguish the signals and switch on the different signaling pathway is still unknown. WAKs bind both covalently and electrovalently to pectins in the cell wall. They show more affinity for the OGs or the pectin fragments originated from the cell wall by wounding or pathogen invasion, where the egg-box model was determined, than the pectins cross-linked in the cell wall matrix. Therefore, the binding between WAKs and OGs may initiate the stress responses. GRP3 may suppress the defense responses by comparative binding affinity to WAKs (**Figure 1.3**). Nevertheless, the AtGRP3-overexpressing plants show enhanced Al tolerance which indicates the positive roles of GRP3 in the Al-induced stress responses and the complexity in WAK-associated elicitor recognition (Mangeon et al. 2017) (**Figure 1.3**). In addition to OGs, another elicitor flg22 induces the defense response which probably overlaps partially with the signaling pathway activated by OGs (Kohorn et al. 2016) (**Figure 1.3**).

Increasing studies show that WAKs have a function as linker connecting the external and internal cellular spaces. They play important roles in plant development and stress responses. Kohorn (2015) proposed a model that explains how WAKs switch from regulating cell expansion to stress response, but there are still a lot of questions remaining, like how one receptor activates different pathways, or how WAK recognizes other stresses, like cold, drought and so on. The calcium-binding EGF-like domains and the cysteine-rich regions of the WAKs can allow the formation of homo- and hetero-dimers. This can activate the downstream signaling along with the configuration changes. Furthermore, the expansions of the WAKs gene family also make it possible to distinguish the different signals.

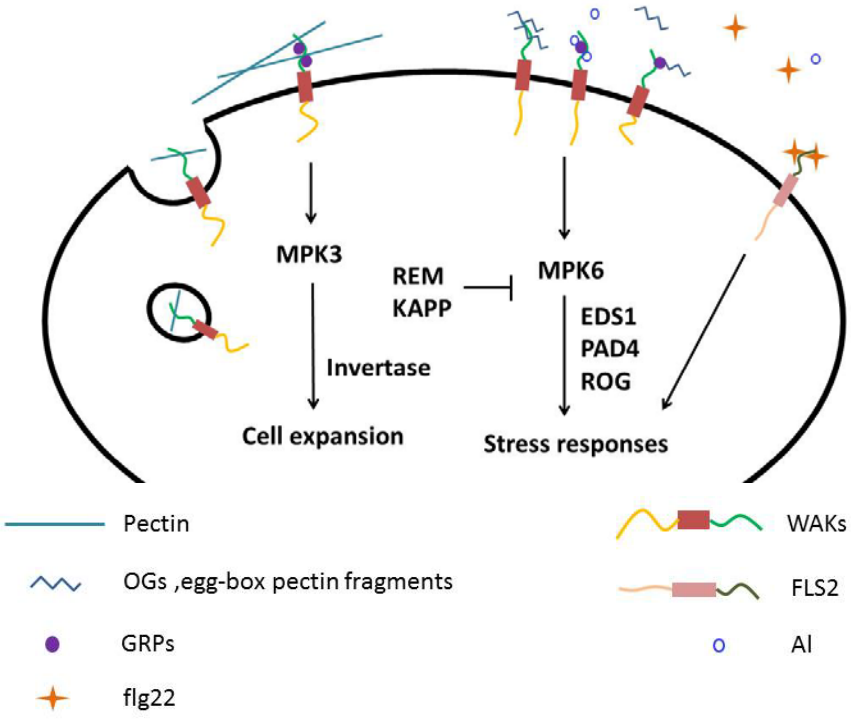


Figure 1.3 Involvement of WAK proteins in signaling pathways. The pectin-linked WAKs in the vesicles are transported to the plasmamembrane which are associated to both cell expansion-dependent growth and development and SA-inducible stress responses. The cell expansion requires MPK3 and vacuolar invertase which can adjust turgor and thus affect cell growth and development by mediating the solute changes. The OGs and pectin fragments derived from wounding or pathogen invasion can lead to the MPK6-dependent stress responses, which can be induced by the toxic metals, like Al, and GRPs binding resulted from certain stresses as well. The stress responses require the involvements of the transcription factors EDS1, PAD4 and cytoplasmic receptor like kinase ROG, and are repressed by the putative scaffold protein REM and kinase-associated protein phosphatase KAPP. The elicitor flg22 can activate the stress response partially overlapping with the one initiated by OGs via binding to another receptor kinase, FLS2.

1.1.2 Germins and Germin-like proteins (GLPs)

Germins and Germin-like proteins (GLPs) are reminiscent of WAKs, being ubiquitous in plants, associated with the extracellular matrix and participate in both, plant development and stress responses (Bernier and Berna 2001). However, the roles of GLPs are different from those of WAKs in various biological processes. In the light of their molecular and biochemical characteristics GLPs not only act as structural proteins and receptors, but also have enzymatic activities involving cell wall synthesis (Bernier and Berna 2001).

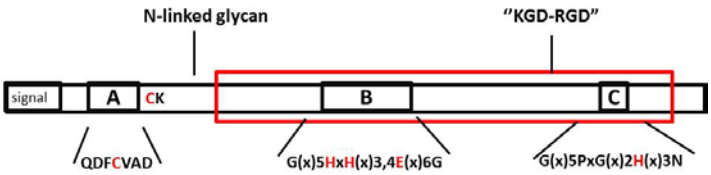
1.1.2.1 Germin and GLP genes and proteins

Germin was first discovered in wheat embryos and identified as a specific marker for the start of germination despite no homology found or the obscure function (Lane 1991; Thompson and Lane 1980). Then several years later, it was characterized as a glycosylated oxalate oxidase (OXO) after

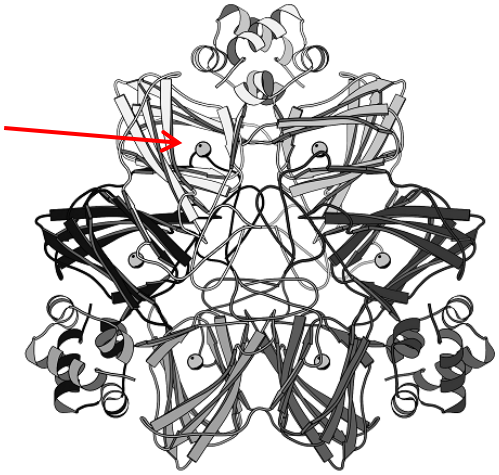
the purification of a barley germin-like protein (Dumas et al. 1993; Faye and Chrispeels 1988; Jaikaran et al. 1990). Meanwhile, two homologues were also identified in *Physarum polycephalum* and *Mesembryanthemum crystallinum* respectively and recognized to be related to water homeostasis (Bernier et al. 1987; Michalowski and Bohnert 1992). Afterwards, a strikingly growing number of genes encoding germin-like proteins were reported in a wide range of species, from cereals (Breen and Bellgard 2010; Druka et al. 2002; Membré and Bernier 1998; Saha et al. 2017) to other higher plants like Arabidopsis (Carter et al. 1998; Membre et al. 2000), peanut (Chen et al. 2011), tomato (Chattopadhyay 2014), soybean (Lu et al. 2010) or tea (Fu et al. 2018) and even including some lower plants (Nakata et al. 2004; Yamahara et al. 1999). To date, 37 *GLP* genes in Arabidopsis and 48 *GLP* genes in rice were identified and the distribution of *GLP* genes on chromosomes in Arabidopsis and rice demonstrated that the expansion of the *GLP* family may result from tandem duplications (Li et al. 2016). The genomic structure analysis of the *GLPs* showed that most of the *GLP* genes in *A. thaliana* possess the two exon-one intron genomic structures while the intron is lost in the wheat germin genes (Carter and Thornburg 1999).

Germins and Germin-like proteins are defined by the percentage of sequence identity (Bernier and Berna 2001). The amino acid identities in germins are usually more than 90%. This homogeneous group seems to be present only in Gramineae and the members of the group have the OXO activity (Breen and Bellgard 2010). GLPs as the heterogeneous group show identities ranging from 30% to 70% comparing with germins (Bernier and Berna 2001) and are found in various species, including angiosperm families, gymnosperms and mosses. The GLPs from Arabidopsis, wheat and barley were divided into five subgroups with germins as one subfamily of GLPs (Carter and Thornburg 1999). However, recently the phylogenetic analyses performed by Barman and Banerjee (2015) and Li et al. (2016) showed 9 clusters with the GLPs from 26 species and 6 major clades using the GLPs from Arabidopsis and rice and suggested the existence of GLP genes before the divergence of monocots and dicots and the species-specific expansions of some GLPs.

The typical structure of germins and GLPs is shown in **Figure 1.4 A**. The N-terminal signal peptide targets the proteins to the cell wall or extracellular matrix (Berna and Bernier 1997; Vallelian-Bindschedler et al. 1998). The conserved motifs of the mature germins and GLPs are box A, B and C. Box B and C are also known as germin box (Breen and Bellgard 2010; Bernier and Berna 2001). The conserved residues in the boxes are shown in red in **Figure 1.4 A**, of which the two cysteines form an internal disulphide bridge while the three histidines and one glutamic acid are related to the binding of manganese (Breen and Bellgard 2010; Bernier and Berna 2001). In addition, the putative N-glycosylation sites are also observed in some GLPs and germins, while the “KGD-RKD” tripeptides are only present in GLPs (Bernier and Berna 2001). Both germins and GLPs belong to the cupin superfamily, which was named according to the jellyroll β-barrel fold (cupa means barrel in Latin) (Dunwell 1998; Dunwell et al. 2000). The cupin domain in the germins and GLPs is indicated in **Figure 1.4 A** containing the germin box. Woo et al. (2000) confirmed the homohexamer (a trimer of dimers) structure of germin (**Figure 1.4 B**) based on the crystallographic evidence. They demonstrated that the germin dimer is similar to the monomer of the seed storage proteins (vicilins). The remarkable stability of the oligomers under treatments with proteases, heat and other denaturing agents arises from the structure, the extensive surface burial.



A



B

Figure 1.4 The structure of GLPs. **A**, The typical structure of germins and GLPs (modified from Bernier and Berna (2001)). The signal peptide and boxes A, B and C are indicated in rectangle. The conserved amino acids are shown in red letters. The red rectangle means the cupin domain. **B**, 3D-protein structure of germin. The small sphere indicated by arrow represents a manganese ion (Dunwell et al. 2008).

1.1.2.2 Expression of germins and GLPs

Germin was first identified in seeds, but increasing evidence shows that germins and GLPs are expressed in different tissues and various developmental stages. According to the statistics performed by Ilyas et al. (2016), the majority of germins and GLPs are discovered in leaves and some in stems and flowers, while only around 10% are reported in roots and fruits. The microarray data analysis using the rice and Arabidopsis *GLP* genes showed that the expressions of some *GLP* genes are restricted to certain tissues or development stages, like *OsGLP3-3* and *OsGLP8-2* in seed developmental stages and *OsGLP9-3* in stigma, while some keep high expression levels during all developmental stages, such as *OsGLP5-2*, *OsGLP2-4* and *AtGLP3-9*, *AtGLP3-8* (Li et al. 2016). Similar expression patterns of *AhGLP* family genes also have been determined in peanut (Wang et al. 2013). Expression of germins and GLPs are responsive to various abiotic and biotic stresses. Drought is able to induce the expressions of germin-like proteins which may be related to the lignification and oxidative cross linkage of cell walls in wild emmer (Krugman et al. 2010). Proteomic analysis also showed the increased accumulation of three germin-like proteins under drought stress both in sensitive and tolerant wheat roots (Faghani et al. 2015). However, the GLPs are not always up-regulated under drought stress. Norway spruce GLP1 and GLP2 are down-regulated during drought stress in the shoots (Fossdal et al. 2007). Except for drought, other abiotic stresses also affect the expressions of germins and GLPs (Barman and Banerjee 2015;

Dunwell et al. 2008), which implies the involvement of GLPs in stress responses. Germins and GLPs also respond to a wide range of biotic stimulus, including pathogen infections, nematode infections, chewing insect-feeding and so on (Dunwell et al. 2008). The heterologous overexpression of GLPs can enhance the stress tolerance in transgenic plants, which is summarized by Ilyas et al. (2016).

1.1.2.3 The roles of germins and GLPs

The expression patterns of germins and GLPs suggest that germins and GLPs are implicated in plant development and various stress responses, during which germins and GLPs may function as enzymes, structural proteins and receptors (Bernier and Berna 2001). Previous studies reported that germins and GLPs possess different enzyme activities, such as oxalate oxidase (OXO) (Dumas et al. 1993; Sakamoto et al. 2015), superoxide dismutase (SOD) (Rietz et al. 2012; Woo et al. 2000), polyphenol oxidase (Cheng et al. 2014), ADP-glucose pyrophosphatase (Fan et al. 2005) or proteolytic activity (da Cruz et al. 2019). Many germins and GLPs with OXO and SOD activities are localized in the apoplast and may participate in different biological processes by contributing to the production of apoplastic ROS and thus affecting cell wall cross-linking (Banerjee and Maiti 2010; Berna and Bernier 1997; Caliskan et al. 2004; Kim et al. 2004; Lane et al. 1992; Segarra et al. 2003). The produced apoplastic ROS exerts opposing actions *in vivo*. Lower levels of ROS can act as signals, while higher levels will lead to oxidative stresses (Miller et al. 2010; Petrov et al. 2015). In addition, some other GLPs without enzyme activity are also responsive to stresses. In barley, heat and H_2O_2 treatment as well as pathogen infection result in the stronger non-covalent binding of HvGLP1 to cell walls (Vallelian-Bindschedler et al. 1998). The transient expression of GLPs without oxalate oxidase activity in wheat cells also reduces the penetration efficiency of the fungus, which demonstrates that germins and GLPs may serve a structural role in cell wall stiffness (Schweizer et al. 1999). Furthermore, the conserved motif KGD-RGD of germins and GLPs (**Figure 1.4**) make it possible to interact with other proteins as receptors (Barman and Banerjee 2015). Swart et al. (1994) purified a putative rhicadhesin receptor containing an RGD attachment site from cell walls of pea roots. Afterwards, another putative rhicadhesin receptor PsGER1, a GLP associated with nodules, was isolated from *Pisum sativum* (Gucciardo et al. 2007). Membre et al. (2000) proposed that GLPs may be a class of receptors localized in cell wall matrix. However, the functions of germins and GLPs as receptors are not well understood, despite the binding between AtGLP4 and auxin *in vitro* (Yin et al. 2009) and their interactions with DING proteins, a group of secreted high-affinity phosphate-binding proteins in some *Pseudomonas* strains (Berna et al. 2009).

Cell walls require proteins acting roles of signal receptions and transductions. WAKs and GLPs are only a small fraction, albeit both of them belong to big protein families. Many proteins, polysaccharides, ions, and some other substances are involved in cell wall biogenesis during plant growth and development as well as cell wall remodeling under stresses. Therefore, the cell wall tends to be considered as an integrated network for cell wall studies (Kohorn and Kohorn 2012b).

1.2 Long non-coding RNA

Non-coding RNA (ncRNA), just as its name implies, refers to the transcripts with little or no protein-coding capacity. Cech and Steitz (2014) pointed out that many old rules on ncRNAs get challenged, refuted and finally rebuilt with the development of ncRNA studies. For example, the discoveries of ribozymes overturned the rule that only proteins can serve as enzymes, and the introns in the human genome revised the occurrence of RNA processing. The high-throughput studies in eukaryotic genomes uncovered strikingly large numbers of ncRNAs transcribed from more than 90% of the genome (Chekanova et al. 2007; Consortium et al. 2007; Kapranov et al. 2007; Ponting et al. 2009), which were considered as transcriptional “noise” (Struhl 2007) except for the housekeeping ncRNAs including transfer RNAs, ribosomal RNAs, small nuclear RNA, and small nucleolar RNA. The ncRNAs challenged the opinions on ncRNAs as well as the structure of genetic information in higher organisms (Mattick 2004). Life processes rely on proteins, but the number of protein-coding genes in human, covering less than 2% of the genome, is similar to that of the nematode *Caenorhabditis elegans*. Therefore, the number of protein-coding genes is not the only reason for the developmental and physiological complexity of higher organisms, which can be ascribed to the abundant ncRNAs as well (Wilusz et al. 2009).

The ncRNAs can be classified as small RNAs (less than 200nt), including microRNAs (miRNAs) and small interfering RNAs (siRNAs), and long non-coding RNAs (longer than 200nt, lncRNAs) sometimes encoding small peptides (Ng et al. 2013) according to the RNA length (Chitwood and Timmermans 2010). The lncRNAs can be termed as long intergenic ncRNAs (lincRNAs), intronic ncRNAs and natural antisense transcripts (NATs) respectively on the basis of their genomic origins (Chekanova 2015). Most lncRNAs are similar to mRNAs, which are transcribed by RNA polymerase II and have 5'-cap and 3’-tail structures (Andersson et al. 2014; Derrien et al. 2012). The non-polyadenylated lncRNA are generated by other RNA polymerases, such as RNA polymerase III (Derrien et al. 2012; Wu et al. 2012) and the plant-specific RNA polymerases, Pol IV and Pol V (Li et al. 2014; Wierzbicki et al. 2008). The lncRNAs can act as decoy molecules, signal molecules, backbone molecules, and guide molecules in different biological processes (Wang and Chang 2011). The tight regulation, cell-specific expression and subcellular-specific localizations of lncRNAs suggest that lncRNA may play important roles in diverse biological processes although they are less characterized and transcribed compared to mRNAs (Chekanova 2015; Wilusz et al. 2009).

1.2.1 Discoveries and expressions of lncRNAs in plants

In plants, both computational and experimental large-scale screening tools were used for identifying lncRNAs (Shafiq et al. 2016). The expressed sequence tags, high throughput sequencing data or tiling microarrays data are exploited to predict the lncRNAs by evaluating the coding potential (Wang et al. 2017). The chromatin signature-based approach is also able to identify new lncRNAs through searching the active transcribed regions marked by H3K4me3 (trimethylation of lysine 4 of histone H3) and H3K36me3 (trimethylation of lysine 36 of histone H3) (Zhu and Wang 2012). However, the method is still not suitable for plants due to the lack of the genome-wide H3K36me3 data in plants (Zhu and Wang 2012). So far thousands of plant lncRNAs have been identified in many species (Shafiq et al. 2016). Some useful plant lncRNA databases have been developed to organize and provide the information on lncRNAs and thus

facilitate relevant studies. These plant lncRNAs databases are summarized in Wang et al. (2017) and Nejat and Mantri (2018). In Arabidopsis, several plant-specific ncRNAs were first identified after filtering the expressed sequence tags (MacIntosh et al. 2001). Afterwards, Wang et al. (2005) and (2006b) predicted and identified thousands of NAT pairs in Arabidopsis, among which 957 NAT pairs were confirmed by the full-length cDNAs and public massively parallel signature sequencing data. Using 200 *A.thaliana* transcriptome data sets, Liu et al. (2012) identified 6480 lincRNAs and confirmed the expression of 2708 lincRNAs via RNA sequencing experiments. Wang et al. (2014b) identified 838 intermediate-sized ncRNAs (50-300nt) which are developmentally regulated in Arabidopsis using a modified RNA isolation strategy for deep-sequencing technology. Additionally, some novel lncRNAs with certain features in Arabidopsis were discovered using different strategies. The evolutionarily conserved ncRNAs were identified by comparative genomic approaches (Song et al. 2009). The analysis of strand-specific RNA-seq data from root cell-type-specific Arabidopsis libraries uncovered 918 *cis*-NAT pairs and the co-expression clusters of lncRNAs and protein-coding genes (Li et al. 2013; Li et al. 2016b). The opposite expression of some NAT pairs in neighboring cells and co-expression networks suggest the important roles of lncRNAs in mediating gene expression in specific cell types (Li et al. 2013; Li et al. 2016b). Novel transcriptionally active regions (TARs) were found by the transcriptome analysis of *Fusarium oxysporum*-infected Arabidopsis and some of the lncTARs directly interact with the pathogen-responsive transcription factors (Zhu et al. 2014). The discoveries of lncRNAs are instrumental to the examination of lncRNAs expression profiles. Many tissue-specific, cell-specific and organelle-specific lncRNAs are identified during the screening for lncRNAs (Li et al. 2013; Marker et al. 2002). The lncRNAs like other transcripts are also developmentally and environmentally regulated (Chekanova 2015; Wang et al. 2014a; Wang et al. 2014b). Abiotic stresses affect the accumulation of 22 lncRNAs in Arabidopsis (Ben Amor et al. 2009). Thereinto, one lncRNA is related to root growth during salt stress and another one is associated with leaf morphology (Ben Amor et al. 2009). 664 drought-responsive lncRNAs are identified in maize, with some acting as miRNA/siRNA/shRNA precursors and others functioning as longer molecules (Zhang et al. 2014). The expression of many lncRNAs is also affected by pathogen infections, such as a set of potato lincRNAs implicated in the defenses against *P. carotovorum subsp. brasiliense* (Kwenda et al. 2016) and the reduced expression of lncRNA S-slylnc0957 in the susceptible tomato following the yellow leaf curl virus infection (Wang et al. 2018). Di et al. (2014) proposed that non-polyadenylated lncRNAs have lower length and expression levels but show significantly responsive expressions under stress comparing to the polyadenylated lncRNAs in Arabidopsis. The comparative analysis for the expressions of some lncRNAs in wild-type plants and mutants indicate that the biogenesis and accumulation of the lncRNAs are modulated by some factors (Ben Amor et al. 2009; Hotto et al. 2011; Liu et al. 2012). The details concerning the regulation of lncRNAs expression are still not well understood.

1.2.2 Molecular functions of lncRNAs in plants

LncRNAs tend to exert functions as complexes by recruiting relevant proteins and precursors or sponges for some miRNAs/siRNAs by virtue of base pairing. In humans and animals, some molecular functions of lncRNAs are correlated with certain biological processes, such as X-chromosome inactivation and dosage compensation (Lee 2012; Rinn and Guttman 2014). But

the mechanisms of lncRNAs in plants are not well characterized. In this section, some molecular functions of lncRNAs in plants will be briefly reviewed.

1.2.2.1 LncRNAs as target mimicry of miRNAs

The role of lncRNAs as the target mimicry of miRNAs was first identified in plants (Franco-Zorrilla et al. 2007) and later the similar lncRNA function was discovered in human and animal cells (Cabili et al. 2011; Liu et al. 2015; Salmena et al. 2011). MiRNAs can trigger the site-specific cleavage and finally translational repression of the target mRNAs via complementary sequences (Chiou et al. 2006), whereas lncRNAs are able to compete with the authentic target and thus impede the miRNA-mRNA binding via partial complementary sequences (Franco-Zorrilla et al. 2007). The mismatch loop caused by partial complementary sequences is located at the cleavage site, which not only circumvents cleavage and sequesters the miRNAs but eventually prevents the degradation of target mRNAs as well (Franco-Zorrilla et al. 2007). *INDUCED BY PHOSPHATE STARVATION1 (IPS1)* is a lncRNA acting as the endogenous target mimic of a microRNA, miR399 in Arabidopsis (Franco-Zorrilla et al. 2007). Pi starvation leads to the accumulation of miR399 and suppression of its target ubiquitin-conjugating E2 enzyme/PHO2 mRNA (Aung et al. 2006; Bari et al. 2006; Chiou et al. 2006). Overexpression of IPS1 gives rise to the accumulation of PHO2 and reduced shoot Pi content by attracting miR399 (Franco-Zorrilla et al. 2007). Apart from IPS1, many other endogenous target mimics were identified by computational prediction in Arabidopsis, rice, cluster bean and other species (Deng et al. 2018; Ding et al. 2019; Meng et al. 2012; Sahu et al. 2018; Wang et al. 2016; Wang et al. 2019b; Wu et al. 2013). Thereinto, in Arabidopsis the endogenous target mimics of miR160 and miR166 are involved in the regulation of plant development, whose functions were confirmed with the overexpressions of endogenous or exogenous endogenous target mimics (Wu et al. 2013). The endogenous target mimics are also implicated in various defense responses. Shuai et al. (2014) showed that in *Populus trichocarpa* the drought-responsive lincRNA 20 is the endogenous target mimic of *Populus*-specific ptc-miR476, and lincRNA2752 as the endogenous target mimic of ptc-miR169 intervenes in the modulation of drought tolerance that miR169 and transcription factor NF-YA are involved in. In tomato, slylnc0195 are the endogenous target mimic of miR166, and the class III homeodomain-leucine zipper (class III HD-Zip) transcription factors are the predicted target of miRNA166a. The expression analyses of slylnc0195, miR166a and class III HD-Zip after *yellow leaf curl virus* inoculation or silencing of slylnc0195 suggest that slylnc0195 may compete with class III HD-Zip for binding miR166a in the pathogen resistance, which is functionally verified with the expression of exogenous slylnc0195 in *N. benthamiana* (Wang et al. 2016).

1.2.2.2 LncRNAs as NAT pairs and the precursors of small RNAs (sRNAs)

NATs, as mentioned above, refer to the natural antisense transcripts which can be divided into two groups: *cis*-NATs and *trans*-NATs (Yuan et al. 2015). The classification of the two categories depends on the original loci and the extent of sequence complementarity with their sense transcripts (Wang et al. 2006; Yuan et al. 2015). The *cis*-NATs are generated from the same genomic locus as their sense counterparts and show high degree of sequence complementarity, and

trans-NATs are transcribed from different genomic locus with partial complementarity (Wang et al. 2006; Yuan et al. 2015). The NATs and their corresponding sense transcripts with the same spatiotemporal expression patterns can form double-stranded RNAs, named NAT pairs. Wang et al. (2005) found 1,340 potential NAT pairs in Arabidopsis. But with different methods and criteria, Wang et al. (2014a) found 37,238 NATs, some of which are light-responsive and development-related NATs and predicted that 70% of the annotated mRNAs of Arabidopsis are involved in NATs. Yuan et al. (2015) discovered 4,080 potential *cis*-NAT pairs and 2,491 *trans*-NAT pairs from 23.9% genes of Arabidopsis. NATs are thought to be important regulators of gene expression (Wight and Werner 2013). Like the endogenous target mimic IPS1, a *cis*-NAT named *cis*-$NAT_{PHO1;2}$ in rice is also implicated in phosphate homeostasis (Jabnoune et al. 2013). PHOSPHATE1;2 (PHO1;2) is the sense transcript of *cis*-$NAT_{PHO1;2}$ and associated with phosphate loading in rice. Although the plant phosphate status cannot affect the transcription of *PHO1;2*, phosphate deficiency in rice will cause the induction of *cis*-$NAT_{PHO1;2}$, which reinforces the translation of PHO1;2 mRNA (Jabnoune et al. 2013). Besides strengthening the expressions of the corresponding protein-coding genes, some NATs show antagonistic effects. The antagonistic effect was termed as “Yin–Yang” regulation by Wunderlich et al. (2014) in the observation that the heat stress-induced NAT, asHSFB2a, retards the expression of HSFB2a, whose overexpression results in the down-regulation of asHSFB2a.

Many lncRNAs are the precursors of sRNAs. Li et al. (2014) proposed that in maize the overwhelming majority of lncRNAs are potential precursors of sRNAs. Almost simultaneously, eight drought responsive lncRNAs were identified as miRNA precursors, and 341 were recognized as sRNAs in maize (Zhang et al. 2014). Not only in maize, other lncRNAs as precursors in other species are also detected, such as 14 lncRNAs acting as precursors of 25 miRNAs in *Populus* under nitrogen deficiency (Chen et al. 2016) and 19 wheat stress responsive lncRNAs as precursors for two miRNAs and 97 siRNAs (Xin et al. 2011). The biogenesis of miRNAs (21-22 nt) and siRNAs (21-24 nt) are different: miRNAs are generated from the hairpin-structured single-strand RNAs transcribed from RNA polymerase II, and siRNAs are originated from the double stranded RNAs, requiring RNA polymerase II, or/an plant-specific RNA polymerase IV or V or RNA-dependent RNA polymerases (RDRs) (Medina et al. 2018; Zhang et al. 2013). The effects of the two sRNAs are similar, because both of them can result in the transcriptional or post-transcriptional regulation via guiding the Argonaute (AGO)-containing effector complex (Molnar et al. 2011). The endogenous siRNAs processed from the NAT pairs are named NAT-siRNAs (Yuan et al. 2015). In Arabidopsis and rice, deep sequencing data analysis showed that a substantial amount of siRNAs are likely to be processed from *cis*-NATs, with more siRNAs corresponding to overlapping regions of *cis*-NAT pairs (Zhang et al. 2012). In plants, the first NAT-siRNA was identified in Arabidopsis (Borsani et al. 2005). The 3’ region of the salt stress-induced SRO5 transcripts can form the NAT pairs with the 5’ region of the constitutively expressed P5CDH transcripts. The 24-nt nat-siRNA are formed by processing the the overlapping region of the SRO5 and P5CDH NAT pairs via DCL2 (Dicer-like protein 2), RDR6, SGS3 (Suppressor of gen silencing 3, required for RDR6-dependent dsRNA synthesis (Kumakura et al. 2009)), and plant-specific RNA polymerase IV. The formed 24-nt nat-siRNA will then lead to the formation of 21-nt nat-siRNA and the cleavage of P5CDH transcripts via DCL1, after the amplification by RNA polymerase IV and RDR6. The down-regulated expression of P5CDH

caused by the siRNA-directed mRNA cleavage promotes the accumulation of proline and ROS. Under salt stress, the increased ROS can then be counteracted by the SRO5 protein. Therefore, the antagonistic regulation of the ROS accumulation under salt stress by P5CDH and SRO5 requires the interfering of siRNAs.

1.2.2.3 LncRNAs as scaffolds or recruiters of chromatin modifiers

The involvements of lncRNAs as scaffolds in the epigenetic regulation have been studied in plants. The plant-specific RNA-directed DNA methylation (RdDM) as one epigenetic model gives rise to transcriptional silencing of repetitive DNA sequences by DNA methylation and even participates in the transmission of DNA methylation patterns to next generations (Matzke and Mosher 2014). During RdDM, 24-nt siRNA is an important factor in the RdDM effector complexes for targeting the homologous genomic regions. The lncRNAs produced by plant-specific RNA polymerases IV are the precursors of the 24-nt siRNAs, whose biogenesis is similar to the 24-nt nat-siRNA described before. The lncRNAs generated from RNA polymerases IV are processed to dsRNA by RNA-dependent RNA polymerase 2 (RDR2), then turned into 24-nt siRNAs by DCL3 (Dicer-like protein 3) and finally loaded to Argonaute 4 (AGO4). The lncRNAs transcribed by plant-specific RNA polymerase V and less RNA polymerase II act as the scaffolds or templates of siRNAs for guiding the siRNA-AGO4 complexes to the genomic targets, during which an increasing number of relevant proteins are recruited, such as methyltransferase 2 triggering the methylation of cytosine (Ariel et al. 2015; Chekanova 2015; Matzke and Mosher 2014; Wang et al. 2017). RdDM is also thought to be implicated in plant development and tolerance to various stresses. In rice, the RdDM directed by a si-RNA named Psi–LDMAR can result in the suppression of the *LDMAR* gene transcription, which is associated with normal male fertility of rice under long-day conditions (Ding et al. 2012). The observation of the DNA methylation during strawberry fruit ripening showed that DNA hypomethylation caused by the down-regulated RdDM activity is an important process during strawberry ripening (Cheng et al. 2018). In tomato, the down-regulated RdDM resulted from reduced expression of AGO4 and it can enhance plant tolerance to salt and drought stress (Huang et al. 2016). Except for the siRNA-AGO4, lncRNAs also tether other chromatin modifying complexes. In Arabidopsis, two classes of lncRNAs, *COOLAIR (COLD INDUCED LONG ANTISENSE INTRAGENIC RNA)* and *COLDAIR (COLD ASSISTED INTRONIC NONCODING RNA)* are transcribed in the antisense orientation related to *FLC* and from the intron of the *FLC* gene, respectively (Csorba et al. 2014; Heo and Sung 2011). FLC (FLOWERING LOCUS C) is a MADS-domain transcription factor which represses flowering of Arabidopsis and is modulated during vernalization (Michaels and Amasino 1999). Both *COOLAIR* and *COLDAIR* are induced by cold treatment and involved in the repression of *FLC* expression, but the two groups of lncRNAs exert the specific roles in different ways. The accumulation of COOLAIR is earlier than that of COLDAIR, and COOLAIR is related to the reduction in H3K36me3 or H3K4me2 at FLC without interacting with Polycomb-repressive complex 2 (PRC2) directly, while COLDAIR triggers the H3K27me3 accumulation at *FLC* by physically associating with a component of PRC2 (Csorba et al. 2014; Heo and Sung 2011). Auxin-controlled development is another example for lncRNAs implicated in chromatin modification. The expression of lincRNA APOLO is responsive to auxin, which mediates the expression patterns of polar auxin transport regulator gene (PID) by regulating the formation of a dynamic chromatin

loop encompassing the promoter of PID (Ariel et al. 2014).

The attitude towards lncRNAs is changing now with growing numbers of cases of confirmed functional lncRNAs. Compared to the earlier studies in animals, the investigations of lncRNAs in plants are still relatively scarce. A substantial amount of lncRNAs in plants have been discovered due to the rapid development of high-throughput RNA-seq and bioinformatics (Liu et al. 2015). However, the biogenesis and functional studies of lncRNAs are still challenges. In contrast to mRNAs, fewer lncRNA mutants are an obstacle for functional studies, which may be solved by the traditional reverse genetics (Liu et al. 2015). Co-expression analysis between lncRNAs and protein-coding genes is another way for identifying the roles of lncRNAs (Yamada 2017). Furthermore, as the function of lncRNAs in plants is similar to that of lncRNAs in animals, the functions of lncRNAs in aniamals is likely to provide hints for plant lncRNA analysis.

1.3 Aims of the study

The object of the study is a resurrection plant, named *Craterostigma plantagineum*. Resurrection plants belong to desiccation-tolerant plants, which can withstand virtually complete desiccation and revive once water available (Bartels 2005; Gaff 1971). Many species are desiccation tolerant, such as many lichens and bryophytes, a few ferns and angiosperms, except gymnosperms (Alpert 2000). The desiccation-tolerant angiosperms are designated resurrection plants. Most resurrection plants, whose habitat is linked to rock outcrops or inselbergs in the tropics, are found in southern and southeastern Africa, eastern South America and Western Australia (Gaff 1971; Gaff 1987; Porembski and Barthlott 2000). They tend to be herbaceous plants (Bartels, 2005), which were identified in both monocotyledon and less species in dicotyledonous (*Gesneriaceae*, *Myrothamnaceae*, and *Scrophulariaceae*) (Porembski & Barthlott, 2000). Craterostigma, capable of anabiosis after being dehydrated for more than two years (Scott 2000), is one of the representative genera in Scrophulariaceae according to the phylogenetic analysis (Rahmanzadeh et al. 2005). Since the first study on the mechanisms of desiccation tolerance in the African resurrection plant *Craterostigma plantagineum* Hochst.(Scrophulariaceae) (Bartels et al. 1990), this resurrection plant becomes an experimental model system for uncovering the secret of desiccation survival. In 2010, the transcriptomes of *C. plantagineum* leaves were characterized at four of hydraulic stages: fully hydration, partially dehydration, desiccation and rehydration (Rodriguez et al. 2010), which discloses the genes implicated in the dehydration stress response and provides directions for the future work about the desiccation tolerance of *C. plantagineum*. Eight transcripts were identified to be up-regulated during dehydration or desiccation with high covariation coefficients via the analysis of the transcriptome data (Giarola et al. 2015), which were then considered as the candidates related to desiccation tolerance. Further studies on the expression and functions of three candidates, CpGPR1 (Glycine-rich protein), CpCRP1 (Cysteine-rich rehydration responsive protein) and lncRNA 28852, were performed (Giarola et al. 2015; Giarola et al. 2016).

Giarola et al. (2015) identified the interaction between CpGRP1 and CpWAK1 *in yeast* and *in planta* and showed the down-regulated transcript expression of CpWAK1 and CpWAK2 under dehydration and up-regulated during rehydration, which demonstrated the involvement of

CpWAKs during dehydration. Given the pectin-WAK-linkage in Arabidopsis, CpWAKs were considered to be associated with cell wall remodeling. Thus, to investigate the roles of CpWAKs in *C.plantagineum*, the main tasks of CpWAK study are described below: 1. Using genome-walking technique to identify the full-length CpWAKs; 2. Studies of gene organization and phylogenetic analyses of CpWAK isoforms; 3. Expression profiles of CpWAKs on the transcript and protein levels; 4. The validation of the interaction of CpWAKs with pectins and CpGRP1 *in vitro*.

CpCRP1 is located in the cell wall and both the transcript and protein expression of CpCRP1 are repressed under dehydration (Giarola et al. 2015), which implies a possible role for CpCRP1 in the cell wall during dehydration. Subsequently, Dulitz (2016) identified the CpGLP1 is a putative interaction partner of CpCRP1 in a yeast-two-hybrid (Y2H) assay, and König (2017) confirmed the SOD activity of CpGLP. To confirm CpCRP1-CpGLP1 interaction probably means a novel mechanism in the cell wall responses under drought stress. So the section of the thesis mainly aims to confirm the interaction of CpCRP1 and CpGLP1 *in vitro* and *in planta*, and meanwhile identify the protein expression patterns of CpGLP1 under drought.

LncRNA 28852 is a taxonomically restricted gene, sharing no similarity to the sequences in the GenBank database (Giarola et al. 2015). The promoter analysis of lncRNA 28852 showed that it is responsive to ABA and mannitol treatment (Nogara 2011). Because of the difficulty in the functional analysis of lncRNAs, the reverse genetic tool was used. The *Arabidopsis thaliana* lines overexpressing the lncRNA 28852 were established by Dr. Giarola (IMBIO, Bonn, Germany) and the seeds were kindly provided to me. The study of lncRNA 28852 has the following objectives: 1, Analysis the transcript expressions of the lncRNA 28852 isolated from total RNA and polysomal RNA; 2, Comparative RNA-seq analysis of wild-type and mutants; 3, Screening the transgenic mutants and evaluation of the mutants.

2. MATERIALS AND METHODS

2.1 Materials

2.1.1 Plant materials

Craterostigma plantagineum Hochst. (Scrophulariaceae) was originally collected in South Africa and further cultivated in the Botanical Institute of the University of Bonn.

The wild type *Nicotiana benthamiana* seeds were provided by Dr. Giarola (IMBIO, Bonn, Germany) and cultivated for BiFC assay.

Arabidopsis thaliana ecotype Col-0 and T-DNA insertion mutant overexpressing lncRNA 28852 (provided by Dr. Giarola) were used for RNA-seq and phenotype observation.

2.1.2 Chemicals

The chemicals used in the study were from the following companies:
Amersham Bioscience (Freiburg, Germany); AppliChem (Darmstadt, Germany); Becton Dicknson and Company(Sparks, USA); BIOMOL (Hamburg, Germany); Bio-Rad (Munich, Germany); Clontech (Heidelberg, Germany); EURx (Berlin, Germany); Fermentas (St. Leon-Rot, Germany); GE Healthcare (Freiburg, Germany); Invitrogen (Karlsruhe, Germany); Macherey-Nagel (Düren, Germany); Merck (Darmstadt, Germany); Roth (Karlsruhe, Germany); Sigma-Aldrich Chemie GmbH (Munich, Germany); Thermo Fisher Scientific (Waltham, USA); Tokyo Chemical Industry Co., Ltd (Tokyo, Japan).

2.1.3 Enzymes and DNA-marker

Phusion Green High-Fidelity DNA polymerase, Restriction enzymes, T4 DNA ligase, Fermentas (St. Leon–Rot,Germany); Taq DNA–polymerase (Ampliqon, Skovlunde, Denmark or isolated and provided by Frederik Faden); Gateway® BP Clonase® II enzyme mix and LR Clonase® II enzyme mix (Invitrogen™, Karlsruhe, Germany); 1 kb DNA-ladder SM0311, MBI Fermentas, (St. Leon-Rot, Germany).

2.1.4 Kits

- GenomeWalker™ Universal Kit, Clontech (Heidelberg, Germany)
- CloneJET PCR Cloning Kit, Fermentas (St. Leon–Rot, Germany)
- NucleoSpin® Gel and PCR Clean-up, Macherey–Nagel (Düren, Germany)
- GeneMATRIX Universal RNA Purification Kit, EURx (Berlin, Germany)
- NucleoBond Xtra Maxi Plus Kit, Macherey-Nagel (Düren, Germany)

- RevertAid First Strand cDNA Synthesis Kit. Fermentas (St. Leon–Rot, Germany)
- NucleoSpin® RNA isolation kit, Macherey–Nagel (Düren, Germany)
- NucleoSpin® RNA Clean-up column, Macherey–Nagel (Düren, Germany)

All the kits were used according to manufacturer's instructions.

2.1.5 Primers

Table 2.1 Primers used in the study

Name	*Sequence (5' →3')*
CpWAK-Fi	CAT CTT GAA AGG ACT GAA GG
CpWAK1-U-Fi-1	GCCACAGTTTCGTAATTCTCA
CpWAK1-U-Fi-2	CCATCGAAATAGCATAGACGTA
CpWAK1-D-Fi	CTACTAACATTACCCTTGAAC
CpWAK2-U-Fi	GGAACGTGTTACTTGGCCAT
CpWAK1-U-CF	TCGCTCTCTTATTCTGTCGTGA
CpWAK1-U-CR	GAGTCTTGCGTGTAATTCTTTC
CpWAK1-D-CF	CATCAAAAGTGCTGAGCAGA
CpWAK1-D-CR	TAATTTTAAAACGTTTTGTGGTGT
CpWAK1_F	AAATGAGGACTTCTTCAATACTTGCT
CpWAK2_F	CACGGAAATGAGGACTTCTTC
CpWAK_R	TCAATATCCATAACCGATAATCCA
Oligo(dT)18 Primer	TTTTTTTTTTTTTTTTTT
p35S-pROK2	CACTGACGTAAGGGATGACGC
28852F2	TGCGTTGATCAAGGTCTTACA

Table 2.2 Gene-specific primers for genome walking

Name	*Sequence (5' →3')*
CpWAK1-Up-GSP1	TGTGCTAAGGCAGACGCAGAAAACGA
CpWAK1-Up-GSP2	TGCAGCAAGTATTGAAGAAGTCCTCA
CpWAK1-Down-GSP1	TTGACTTACGGAGGTATGGAGCGTGAC
CpWAK1-Down-GSP2	GCCGAATGCTCGAGGTAATATTAGTGGA
CpWAK2-Up-GSP1	GGATGATGATGATGATGATGACGATGATG
CpWAK2-Up-GSP2	AGTTGTGCAAAGGCAGACGCAGAAAAC
CpWAK2-Down-GSP1	CCAACGAATACGCTCGATGTAGTGATT
CpWAK2-Down-GSP2	TAATCCATATCAAGAGGGCTGAGCAGA
CpWAK3-Up-GSP1	AGTCTCTGACATCATAACCATTACATC
CpWAK3-Up-GSP2	ATGGGTATGGAAATGGATGATGATGGT
CpWAK3-Down-GSP1	TCGATATAGCGTGAGCGATCTTTCCCT
CpWAK3-Down-GSP2	TCGGCGTCGTCATGGTGGAACTCCTCA
AP1	GTAATACGACTCACTATAGGGC
AP2	ACTATAGGGCACGCGTGGT

Table 2.3 Vector–specific primers

Name	*Sequence (5' →3')*
pJET1.2 forward	CGACTCACTATAGGGAGAGCGGC
pJET1.2 reverse	AAGAACATCGATTTTCCATGGCAG
T7 promoter	TAATACGACTCACTATAGGG
T7 teminator	GCTAGTTATTGCTCAGCGG
M13 forward	GTAAAACGACGGCCAG
M13 reverse	CAGGAAACAGCTATGAC

Table 2.4 RT-PCR primers

Name	*Sequence (5' →3')*
CpWAK1_RT_F	GAGCTCAGAGTGCAGTATCGAC
CpWAK1_RT_R	GGGAAAGATCGCTAAATCGAG
CpWAK2_RT_F	AAGGTGCTGTGCAGACTCATC
CpWAK2_RT_R	AAGATTGCCCCAGTAATATGGAC
CpWAK3_RT_F	AGAGTGCAGTATCGGCTTACAT
CpWAK3_RT_R	CTCACGCTATATCGACTTGAGCA
CpGRP1_RT_F	GGTGGTTATGGTGGAAGAGG
CpGRP1_RT_R	TTGCATGAATGAAACGGAGAT
CpEF1 α_RT_F	AGTCAAGTCCGTCGAAATGC
CpEF1 α_RT_R	CACTTGGCACCCTTCTTAGC
28852CDS_F	GAGAATATTGAGTGCTTGTGAGTGC
28852CDS_R	GCTAAGACCATGATTTCATTTACTGG
AtEF1α_F	GTGATGCTGGTATGGTGAAG
AtEF1α_R	CCCAATTACGAGAACAACGC
AtActin_F	GGAATCCACGAGACAACCTATAAC
AtActin_R	GAAACATTTTCTGTGAACGATTCCT
AT5G05600_F	GATCCTGGCGGTATGACCATTCT
AT5G05600_R	GCCTTGTGGACCTTGAGTTCTGATA
CLH1_F	TTG ATC CAG TCG CAG GAA CTA AC
CLH1_R	CGA ATC TCC GCT TTT TCA CC
AOC1_F	GAAAACCCCAGACCAAGCAAAG
AOC1_R	TGTCCGTAGGCACCTTCAAAG
LHCB4.3_F	CTTTGAAGCTGGCTGAGATA
LHCB4.3_R	ATTTCCGTAAGCCACTAGAC
F3H_F	AATCACGTTTGCCGAGATGTAT
F3H_R	CGTCAAATACAGGTGGAGTATGAT
TT4_F	GCCATAGACGGACATTTGAGGGAAG
TT4_ R	ACCCCACTCCAACCCTTCTC
DFR_F	CGCCAAGGGACGTTATATTTG
DFR_R	TCCCCGTTTCTGTCTTGTTAC
OPR3_F	CGCTTACCTTCACGTTACAC

OPR3_R	AAACTCAGAGGCGGGAAAAAG
LOX4_F	TCCGTCATCACCACCATCAT
LOX4_R	CCTTCTCTATCCGTCCGATTTC
HPL1_F	ACCGAAGTCAGGAGAGTTTG
HPL1_R	GAAGCCGTGAGAGTGACAAT

Table 2.5 Protein overexpressing primers

Name	*Sequence (5' →3')*	*Restriction site*
CpWAK1-R-2	ATACTGCTCGAGCGTGTAATTCTTTCC	*XhoI*
CpWAK1-R-3-for	ACTACCATGGATCAAGACTCTTGCCA	*NcoI*
CpWAK1-R-3-rev	ATACTGCTCGAGACACGTGAAAGAGC	*XhoI*
CpWAK1-R-4	ATACTGCTCGAGCAAATAATCATCGCT	*XhoI*
CpWAK2-XhoI-R	ATACTGCTCGAGACATGTGGAAGAG	*XhoI*

Table 2.6 Primers for BiFC

Name	*Sequence (5' →3')*
attB1-adapter	GGGGACAAGTTTGTACAAAAAAGCAGGCT
attB2-adapter	GGGGACCACTTTGTACAAGAAAGCTGGGT
attB3-adapter	GGGGACAACTTTGTATAATAAAGTTGCT
attB4-adapter	GGGGACAACTTTGTATAGAAAAGTTGGGT
CpCRP1-attB1	AAAAAGCAGGCTTAACCATGGCCGCCCACGAAATGGAA
CpCRP1-attB4	GAAAAGTTGGGTGGCCTCTTATACCGCAGCA
CpGLP1-attB3	ATAATAAAGTTGCTACCATGTACGATCCCAGCCCATTG
CpGLP1-attB2	AGAAAGCTGGGTAAAAGGCAGCTTGAAGGCC
pBifc-nYFP-F	ACGACGGCAACTACAAGACC
pBifc-cYFP-F	CAAGATCAGGCACAACATCG
pBifc-cYFP-R	CGATGGGGGTGTTCTGCTGGTA
pBifc-nYFP-R	GCCGGACACGCTGAACTTGT

2.1.6 Media, buffers and solutions

Unless specially stated, all media, buffers and solutions were autoclaved for 20 min at 121 ºC at 1.5 bars.

2.1.6.1 Media

½ MS-medium (1L): 2,15 g MS-salt mixture; 10 g sucrose; 2 ml vitamin solution; adjust pH to 5.8 with KOH; 8 g bacto-agar for solid medium.
LB-medium (1L): 20 g Lennox LB powder (10 g Tryptone, 5 g yeast extract, 5 g NaCl) or 35 g Lennox LB Agar powder (10 g Tryptone, 5 g yeast extract, 5 g NaCl, 15 g agar) for solid medium.
YEB-medium (1L): 5 g sucrose; 5 g of meat extract; 5 g peptone; 1 g yeast extract; 0.493 g $MgSO_4$; adjust pH to 7.0; 15 g bacto-agar for solid medium.

2.1.6.2 Buffers and solutions

Vitamin solution: 2 mg/ml glycine; 0.5 mg/ml Niacin (Nicotine acid); 0.5 mg/ml pyridoxine-HCl; 0.1 mg/ml thiamine-HCl. Use 1:1,000 dilution of the autoclaved solution; store at 4 °C.
50× TAE buffer: 2 M Tris base; 100 mM EDTA; pH 8.0, adjust pH with glacial acetic acid.
6× DNA loading buffer: 10 mM Tris; 0.03% (w/v) bromophenol blue; 0.03% (w/v) xylene cyanol FF; 60% (w/v) glycerol; 60 mM EDTA; pH 7.6 adjust pH with NaOH.
1× TE buffer: 10 mM Tris-HCl; 1 mM EDTA; pH 8.0, store at room temperature.
RNase A (stock solution): 10 mg/ml RNase A in milli-Q sterile water; store in aliquots at -20 °C.
IPTG (stock solution): 100 mM IPTG in water; sterilized by filtration through 0.22 μm membrane and store at -20 °C
Ampicillin (stock solution): 100 mg/ml in water; sterilized by filtration through 0.22 μm membrane and store at -20 °C; working solution: 1:1,000 dilution.
Kanamycin (stock solution): 50 mg/ml in water; sterilized by filtration through 0.22 μm membrane and store at -20 °C; working solution: 1:1,000 dilution.
Rifampicin (stock solution): 50 mg/ml in methanol; dissolved in 10 N NaOH first to facilitate dissolving. Alternatively dissolve in DMSO; store at -20 °C; working solution: 1:500 dilution.
Spectinomycin (stock solution): 100 mg/ml in water; sterilized by filtration through 0.22 μm membrane and stored at -20 °C; working solution: 1:1,000 dilution.

2.1.7 Vectors, microorganisms

All the vectors and microorganisms used in the study are listed as below with the molecular details. The vectors were kept either as plasmids at -20 °C or in the bacteria strains as frozen glycerol stock cultures at −80 °C, like the way of storing microorganisms.

2.1.7.1 Vectors (vector maps are shown in the supplementary data)

- pJET1.2/blunt
  This vector is a linearized cloning vector from Fermentas (St. Leon-Rot, Germany), and always used for blunt-end cloning of PCR-fragments.

- pET28a
  This vector carries an N-terminal His-Tag/thrombin/T7-Tag and an optional Cterminal His-Tag sequence, which was used for His-tagged protein overexpression (Novagen, Darmstadt, Germany).

- pDONR221-P1P4 and pDONA221-P3P2
  The two Gateway donor vectors carry the positive selection gene *ccdB* and harbor the attP1-attP4 sites and attP3-attP2 sites respectively. The two vectors were used for the construction of entry clones in the study.

- pBiFCt-2in1-NN, pBiFCt-2in1-NC, pBiFCt-2in1-CN, pBiFCt-2in1-CC
  The four pBiFCt-2in1 vectors (Grefen and Blatt, 2012) contain the positive selection gene

ccdB which is flanked by attR1 and attR4 sites or attR3 and attR2 sites. The red fluorescent protein (RFP), N- and C-terminal yellow fluorescent protein (YFP) tagged with hemagglutinin (HA) and myc respectively are under the control of 35S promoter. The destination vectors for the BiFC assay are obtained with the help of the shuttle plasmids (pDONR221-P1P4 and pDONA221-P3P2).

2.1.7.2 Microorganisms

- *Escherichia coli* DH10B (Lorow and Jessee 1990)
  Genotype: F^- endA1 recA1 galE15 galK16 nupG rpsL ΔlacX74 Φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrAΔ(mrr-hsdRMS-mcrBC) λ^-.
  This strain was used as host strain for cloning.

- *Escherichia coli* BL21 (Pharmacia, Freiburg)
  Genotype: F^- ompT gal dcm lon hsdS$_B$(r_B^- m_B^-) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]).
  This strain was used for overexpressing recombinant proteins.

- *Escherichia coli* DB3.1 (Invitrogen)
  Genotype: F^- gyrA462 endA1 glnV44 Δ(sr1-recA) mcrB mrr hsdS20(r_B^-, m_B^-) ara14 galK2 lacY1 proA2 rpsL20(Sm_r) xyl5 Δleu mtl1.
  E. coli DB3.1 contains a specific mutation in the DNA gyrase (*gyrA462*) that makes it resistant to the lethal *ccdB* gene product. It was used as a host strain for propagation of the plasmind with *ccdB* gene.

- *Agrobacterium tumefaciens* GV3101/pmP90RK (Koncz and Schell 1986)
  Genotype: C58C1 pMK90RK, Rifr, Gmr, Kmr.
  This strain was used for transformation of wild-type *Nicotiana benthamiana*.

2.1.8 Software, programs and online tools

- Microsoft Office 2010
- Sequence Manipulation Suite (http://www.bioinformatics.org/sms2/index.html)
- MEGA software 5.1
- Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/)
- Gene Structure Display Server 2.0 (http://gsds.cbi.pku.edu.cn/)
- Blast programs in NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi)
- NCBI CD-Search tool (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi)
- ExPASy ScanProsite tool (http://prosite.expasy.org/scanprosite/)
- TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/)
- PLACE database (https://www.dna.affrc.go.jp/PLACE/?action=newplace)
- Plantcare (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/)
- Oligo 7 (http://www.oligo.net/tutorials.html)
- SnapGene software 1.1.3

- APE software
- GraphPad Prism 5.0
- INTAS GelDoc
- Vector NTI Advance Version 11.5.1
- ImageJ and Fiji ImageJ software
- Venn diagrams (http://bioinformatics.psb.ugent.be/webtools/Venn/)
- Cluster 3.0
- JAVA TreeView
- Tair website (https://www.arabidopsis.org)
- DAVID Bioinformatics Resources (https://david.ncifcrf.gov/home.jsp)
- KEGG pathway database (https://www.genome.jp/kegg/pathway.html)

2.1.9 Machines and other devices

- T3-Thermocycler, Biometra, Göttingen, Germany
- Mastercycler gradient, Eppendorf, Hamburg, Germany
- SmartSpec 3000, Bio-rad, Hercules, Canada
- Genesys 10-S, Thermo Spectronic, USA
- Electronic balance ABJ 120-4M, KERN & SOHN, Balingen, Germany
- pH-meter, SCHOTT GLAS, Mainz, Germany
- pH-meter, METTLER TOLEDO, Giessen, Germany
- Conductivity Meter Set, Qcond 2400, Darmstadt, Germany
- Electrophoresis power supply, Gibco BRL, Carlsbad, Canada
- EasyCast, Owl-Scientific, Portsmouth, USA
- Compact XS, Whatman Biometra, Göttingen, Germany
- SDS-PAGE Minigel system, Biometra, Göttingen, Germany
- Criterion Blotter, Biorad, Munich, Germany
- Nitrocellulose membrane 0.45 µm, Sartorius, Göttingen, Germany
- Protran BA-85 0.45 µm, Whatman, Maidstone, UK
- Electroporation system GenepulserII Electroporator, Bio-Rad, Hercules, USA
- UV-light table PeQlab, Vilber, Eberhardzell, Germany
- ImageScannerTM Ⅲ, GE Healthcare Life Science, Buckinghamshire, UK
- Chemiluminescence detector Intelligent Dark Box II, FUJIFILM Corporation, Tokyo,Japan
- c150 Geldokumentationssystem Imager, Biozym, Oldendorf, Germany
- Nanodrop BioSpec-Nano, Shimadzu Biotech, Chiyoda-ku, Japan
- Centrifuges 5415R; 5417R, 5810R, Eppendorf AG, Hamburg, Germany
- Ultracentrifuges L8-70M, Beckman, Krefeld, Germany
- Lyophilisation machine LDC–2, Christ, Osterode am Harz, Germany
- Clean bench, BIO-FLOW Teehnik, Meckenheim, Germany
- Ultrasonic Processor UP200S, Hielscher, Teltow, Germany
- Zeiss LSM-780, Carl Zeiss Microscopy, Jena, Germany
- Sonification water bath Sonorex Super RK102P, Bandelin electronics, Berlin, Germany
- Amicon® Ultra-15/ Ultra-4 10K Centrifugal Filter Devices, Millipore, Billerica, USA
- Consumables Pipette tips and centrifugal tubes, Sarstedt AG,Nümbrecht, Germany

- HIS-Select Nickel Affinity Gel, Sigma-Aldrich, St. Louis, USA
- HisPur™ cobalt resin columns, Thermo Scientific, USA
- Glutathione Sepharose 4B resin, GE Healthcare Life Science, Freiburg, Germany
- Desalting columns PD–10, GE Healthcare, Freiburg, Germany
- Nunc Maxisorp flat-bottom plates, Invitrogen, CA, USA
- Electroporation–cuvette, Bio-Rad, Germany
- GenePulser II, Bio-Rad, Germany
- Plant climatics, Percival, Wertingen, Germany

2.2 Methods

2.2.1 Cultivation and treatments

2.2.1.1 Plant cultivation

Craterostigma plantagineum

Plants of Craterostigma plantagineum (Hochst.) were grown and multiplied according to Bartels et al. (1990) with some modification. *C.plantagineum* seedlings were grown under sterile conditions in glass jars on MS agar. The sterile seedlings were cultivated in in a light intensity of 80 µE/m^2/s at 22°C in a day/night cycle of 16/8 hours, and after six weeks subcultivated into unsterile LamstedtTon® granulate (Leni, Bergneustadt DE), watered with a 0.1 % solution of Wuxal (Manna, Ammerbuch–Pfäffingen, DE) at 18°C in a day/night cycle of 13/11 hours.

Arabidopsis thaliana

Seeds of *A. thaliana* (ecotype Columbia-0 and other transgenic lines) were sterilized in 70% (v/v) ethanol for 2 min and then in 7% (v/v) NaClO + 0.1% (w/v) SDS for 10 min. After rinsed 3-4 times with sterile distilled water, the sterilized seeds were sown on 1/2 MS-agar plates (Murashige and Skoog 1962) with the help of 0.1% agar. Transgenic seeds were selected on 1/2 MS-agar plates containing 50µg/ml kanamycin. The sown seeds were kept in dark at 4℃ for 2~4 days for stratification, and then moved into a growth chamber with 120-150 µE m-2 s-1 light and a 8/ 16h photoperiod (eight hours of light at 22°C and sixteen hours of darkness at 18°C). The 8~10-day seedlings on the petri dish were finally transferred into wet, Lizetan® (Bayer, Leverkusen, Germany)-treated soil and subjected to different treatments. For flowering and collecting seeds 4~6-week-old *A. thaliana* were moved to a growth chamber with a 16/8 h photoperiod. *A. thaliana* seeds were stored at 4°C.

2.2.1.2 Growth of microorganisms

All *E. coli* strains were cultured at 37°C in LB medium (with or without agar). The cultures were supplemented with appropriate selective antibiotics if required.

2.2.1.3 Treatments of microorganisms

Craterostigma plantagineum

The adult unsterile *C. plantagineum* grown on LamstedtTon® granulate (Leni, Bergneustadt DE) were subjected to drought stress. Drought stress was imposed by withholding water for 7 days, and then on the 8th day all pots were watered until saturation (normally for 3 days). The intact *C.plantagineum* was divided into two groups: Light (L) and Dark (D). The samples in group L grew normally and collected every four hours from 7:00 am to 23:00 pm. And the samples in group D as control were in extended darkness from 7:00 am to 23:00 pm and collected as group L. The detached *C. plantagineum* leaves were soaked in water, 100 µM naphthaleneacetic acid (NAA, a synthetic auxin analogue), 1 mM salicylic acid (SA) and 100 µM methyl jasmonic acid (MeJA) for 1h, 3h, 6h, 24h and 48h. All samples were collected and frozen in liquid nitrogen and stored at -80 °C for further analyses.

Arabidopsis thaliana

3~4-week-old *A. thaliana* were subjected to drought and salt stress treatments. The pots were dried for 2~3 weeks and then rehydrated for 2~4 days. The *A. thaliana* seedlings in 1/2 MS-agar plates were supplemented with 50 µM MeJA, 20 µM ABA, 250 mM mannitol, and 100 mM NaCl respectively.

2.2.1.4 Germination rate

About 100 surface-sterilized *A. thaliana* seeds (wide type and mutant lines) were sown on 1/2 MS-agar plates with/without supplements and then vernalized for 2~4 days. Germination (fully emerged radicle) was scored at 6-hour intervals after the plates were transferred under light. The germination rate was recorded as a percentage of seeds sown rate for 3~5 days.

2.2.1.5 Root elongation

3~4-day seedlings were transferred onto vertical 1/2 MS-agar plates with/without supplements. The roots were kept pointing downward for around 5 days. The root length was then measured using ImageJ software after scanning the plates.

2.2.1.6 Electrolyte leakage

Electrolyte leakage (EL) was measured as described by Lutts et al. (1996) with a few modifications. The leaves in the same developmental stage and similar mass were chosen and washed with deionized water, placed in tubes with 10 ml of deionized water and incubated for 2h at 25℃. Subsequently, the electrical conductivity of the solution (L1) was determined. Samples were boiled at 100℃ for 30 min and final conductivity (L2) was measured after equilibration at 25℃.The EL was defined as follows: EL(%)= (L1/L2)*100

2.2.2 Nucleic acids

2.2.2.1 Extraction of genomic DNA from plants

Plant tissues (50-200 mg) were ground to a fine powder in liquid nitrogen. Then the powder was homogenized in 300 µl 2× lysis buffer, 300 µl 2 M urea and 600 µl of phenol/chloroform/isoamyl alcohol (25/24/1). The mixture was centrifuged for 10 min (14,000 rpm, RT) and the supernatant was collected. DNA was pelleted after centrifugation for 15-20 min (14,000 rpm, 4 °C) from the supernatant mixed with 0.7 volume of isopropanol. After washing twice with 70% ethanol, the air-dried DNA pellet was dissolved in TE buffer containing 20 µg/ml RNase A. RNAs in the DNA samples were removed after 5 min incubation at 37 °C.

2× lysis buffer: 0.6 M NaCl; 0.1 M Tris-HCl, pH 8.0; 40 mM EDTA, pH 8.0; 4% (w/v) Sacrosyl; 1% (w/v) SDS

2.2.2.2 Mini-prep plasmid DNA from *E.coli*

The alkaline-lysis mini-preparation of plasmid DNA from *E.coli* was done according to Birnboim and Doly (1979) with minor modifications. One bacterial colony was inoculated in 3 ml LB medium and cultured at 37 °C overnight with vigorous shaking (200 rpm). Cells were pelleted by centrifugation at 13000 rpm for 30 sec at room temperature. After removing the supernatant, the cell pellet was resuspended in 100 µl ice-cold GTE solution by vigorous vortex. Then 200 µl of freshly prepared NaOH/SDS solution was added. After mixing by tapping with fingers lightly and incubation on ice for 5 min, 150 µl of potassium acetate was added, mixed thoroughly and incubated on ice for another 5 min. The bacterial lysate was centrifuged for 6 min (13,000 rpm, RT). The supernatant was transferred into one volume (450 µl) of phenol/chloroform/isoamyl alcohol (25/24/1) and vortexed for 10 sec. The aqueous upper layer was carefully transferred to a new tube after a very short centrifugation (at maximum speed, RT), and two volumes of 95% ethanol were added. The mixture stayed for 2 min at room temperature. The plasmid DNA was precipitated after centrifugation at 13000 rpm for 5 min. The DNA pellet was washed with 70% (v/v) ethanol before air-drying and dissolving in TE buffer with 20 µg/ml RNase A. The RNAs were removed after incubation at 37 °C. The plasmid solution could be stored at -20 °C or directly used for analysis.

Glucose/Tris/EDTA (GTE): 50 mM glucose; 25 mM Tris-Cl, pH 8.0; 10 mM EDTA. Autoclave and store at 4 °C.
NaOH/SDS solution: 0.2 M NaOH; 1% (w/v) SDS. Prepare immediately before use.
Potassium acetate solution: 29.5 ml glacial acetic acid; KOH pellets to pH 4.8; bring to 100 ml with H_2O. Store at room temperature (do not autoclave).

Alternatively, a considerable amount of plasmid DNAs were extracted and purified using the plasmid DNA purification kit (NucleoBond® Xtra Midi) according to the manufacturer's instructions.

2.2.2.3 DNA extraction from agarose gels

DNA fragments from enzymatic digestions and PCR products were extracted from agarose gels after electrophoresis and purified using NucleoSpin® Gel and PCR clean-up kit according to the manufacturer's instructions.

2.2.2.4 Extraction of total plant RNA

RNA extraction

Total plant RNA extraction was based on Valenzuela-Avendaño *et al.* (2005) with minor modification. Plant tissue was ground to a fine powder under liquid nitrogen. 50-100 mg powder was added to 1.5 ml of RNA extraction buffer and homogenized prior to 10 min incubation at room temperature. The mixture was centrifuged for 10 min (10,000 *g*, RT), and the supernatant was decanted to a fresh tube and mixed thoroughly with 300 µl of chloroform/isoamyl alcohol (24/1) by vortexing for 10 sec. The suspension was centrifuged for 10 min (10,000 *g*, 4 °C) and the clear upper aqueous phase was removed into another fresh tube. 375 µl of isopropanol and 375 µl of 0.8 M sodium citrate/1 M sodium chloride were added, mixed thoroughly and the sample was allowed to stand at room temperature for 10 min. The sample was centrifuged for 10 min (12,000 *g*, 4 °C) and the pellet was washed with pre-chilled (-20 °C) 70% ethanol, air-dried and dissolved in 20 µl sterile H_2O.

Extraction buffer: 38% (v/v) buffer-saturated phenol; 0.8 M guanidine thiocyanate; 0.4 M ammonium thiocyanate; 0.1 M sodium acetate (pH 5.0); 5% (v/v) glycerol

RNA extraction for RNA-seq

Total plant RNA for RNA-seq was extracted and dissolved in RNase-free water using RNA isolation kit (NucleoSpin® RNA) according to manufacturer's instructions. RNA samples were then mixed thoroughly with 0.1 volume 3M sodium acetate and 2.5-3 volume ice cold 100% ethanol and precipitated at -80℃ overnight. After washing twice with 0.5 ml ice cold 75% ethanol, the RNA pellets were collected and dissolved in an appropriate volume of RNase-free water again. RNA samples were then digested with DNase I at 37℃ for 30 mins. Final clean-up was carried out using a commercial spin column (NucleoSpin® RNA Clean-up column). All the RNA samples were delivered in ethanol to GATC Biotech Company for sequencing.

2.2.2.5 Polysomal RNA isolation

Polysomal RNA was isolated as described by Juszczak and Bartels (2017) with minor modification. Leaf material was ground to powder in liquid nitrogen and then homogenized in extraction buffer. After centrifugation (13200g, 5 min, 4℃), the supernatant fraction was added with 1/20 vol of 10% (w/v) sodium deoxycholate and incubated on ice for 5 min. The mixture was centrifuged at 13200g, 4℃ for 15 min again and then loaded on a sucrose gradient (15/30/40/56%

(w/v) sucrose). The loaded sucrose gradient was ultracentrifugation at 4℃, 45000 rpm for 80 min (Beckman L8-70M) and fractionated into 10 fractions. In order to identify fractions with polysomes but without free ribosomes and free mRNAs, control gradients were supplemented with 20 mM puromycin. All fractions of both sample and control gradients were visualized on 1% (w/v) agarose gel and the fractions (normally fractions 6-10) which did not show nuclei acids were chosen for polysomal RNA extraction using phenol/chloroform/isoamyl alcohol (25:24:1). The isolated RNA from fractions 6-10 was then pooled and purified with 5 M LiCl and GeneMatrix Universal RNA Purification Kit (EURx, Gdansk, Poland).

Extraction buffer: 200 mM Tris-HCl, pH 9.0; 200 mM KCl; 35 mM MgCl2; 25 mM EGTA; 200 mM sucrose; 1% (v/v) Triton X-100; 2% (v/v) polyoxyethylen-10-tridecylether; add freshly 100 mM β-mercaptoethanol; 1 mg ml^{-1}heparin, 100 g ml^{-1} chloramphenicol and 25 g ml^{-1} cycloheximide

Sucrose gradient: 40 mM Tris-HCl pH 8.5; 20 mM KCl; 10 mM $MgCl_2$; 100 g ml^{-1} chloramphenicol; 500 g ml−1 heparin, 15/30/40/56% sucrose

2.2.2.6 Qualitative and quantitative estimation of nuclei acids

The qualitative analysis of nuclei acids samples was performed by electrophoresis on a 1% (w/v) agarose gel using the 1 kb ladder as size marker. The concentration of nucleic acids was measured using a NanoDrop microvolume spectrophotometer at 260 and 280 nm. The value of ratio OD_{260}/OD_{280} is important information about nuclei acids purity and protein contamination. The OD_{260}/OD_{280} ratio of pure nuclei acids should be between 1.8 and 2.0. In addition, the ratio of OD_{260}/OD_{230} also displays the nuclei acids purity, the lower of which indicates a higher amount of salt. As a guideline, the OD_{260}/OD_{230} ratio should be greater than 1.5.

2.2.2.7 Reverse transcription of RNA

The DNA in the RNA sample was digested after incubation at 37 °C for 30 min in the 10 µl reaction mixture containing 4 µg of total RNA, 1× DNase I buffer (20 mM Tris/HCl pH 8.4; 50 mM KCl and 2 mM $MgCl_2$) and 10 U RNase-free DNase I. 1 µl of 25 mM EDTA was added into the mixture, and the mixture was heated at 65 °C for 10 min to deactivate the DNase I. 1 µl of 100 pmol/µl oligo-dT primer and 1 µl H_2O were added into 10 µl DNase I-treated RNA. After incubation for 5 min at 65 °C, the reaction was cooled on ice and added with the following reagents: 4 µl 5 x First strand buffer, 1 µl RiboLockTM RNase Inhibitor, 2 µl 10 mM dNTP mix and 1 µl Reverse transcriptase. The mixture was incubated at 42 °C for 60 min. The reaction was terminated by heating at 70 °C for 5 min. The mixture was used directly for PCR or stored at -20 °C.

2.2.2.8 Polymerase chain reaction (PCR)

Primers

Primers were designed using the software Oligo 7 (http://www.oligo.net/tutorials.html). The sequences of primers are shown in **section 2.1.5.** All primers were synthesized by Eurofins MWG Operon (Ebersberg, Germany) and dissolved in water to a final concentration of 100 µM.

PCR with Taq polymerase

A large amount of specific DNA fragments can be amplified from various DNA sources, such as cDNA, plasmid and genomic DNA, using polymerase chain reaction (PCR).
The composition of a 20 µl PCR system was as follows:

15.6 µl H_2O (sterile double distilled)
2.0 µl 10× PCR-buffer with $MgCl_2$
0.4 µl 10 mM dNTPs
0.4 µl Forward-primer (10 pmol/µl)
0.4 µl Reverse-primer (10 pmol/ µl)
0.2 µl Taq-polymerase (5 U/µl)
1.0 µl DNA template: plasmid DNA (5-100 ng/µl) or genomic DNA (50-300 ng/µl) or a bacterial colony

The cycling condition of PCR was as follows:

1. 95 °C 3-5 min for initial denaturation
2. 95 °C 30 sec for cycling denaturation
3. 50-65 °C 45 sec for primer annealing
4. 72 °C for extension (1 min/kb), recycle from step 2 for 21-35 times.
   The annealing temperature and cycle numbers were determined empirically for each PCR.
5. 72 °C 5-10 min for final extension
6. 4 °C for holding the samples until they were collected

PCR with high fidelity DNA polymerase (Phusion)

The DNA amplification was performed with Phusion DNA polymerase, when the high fidelity DNA synthesis was required.
The composition of a 20 µl PCR system was as follows:

13.6 µl H_2O (sterile double distilled)
4.0 µl 5× Phusion HF buffer
0.4 µl 10 mM dNTPs
0.4 µl Forward-primer (10 pmol/µl)
0.4 µl Reverse-primer (10 pmol/ µl)
0.2 µl Phusion-polymerase
1.0 µl DNA template: plasmid DNA (5-100 ng/µl) or genomic DNA (50-300 ng/µl)

The cycling condition of PCR was as follows:

1. 98 °C 1 min for initial denaturation
2. 98 °C 5 sec for cycling denaturation
3. 50-65 °C 15 sec for primer annealing
4. 72 °C for extension (for plasmid 15s/kb, for genomic DNA 30s/kb, for cDNA 40s/kb), recycle from step 2 for 21-35 times.

The annealing temperature and cycle numbers were determined empirically for each PCR.

5. 72 °C 10 min for final extension
6. 4 °C for holding the samples until they were collected

Both PCR systems were performed in a TRIO-thermo block (Biometra, Göttingen, Germany).

2.2.2.9 Agarose gel electrophoresis

DNA and RNA can be monitored by agarose gel electrophoresis. Agarose gel (0.8-1.5% (w/v)) was prepared with 1 × TAE buffer and Ethidium bromide solution (final concentration approximately 0.5μg/mL). After loading the mixture of nuclei acid sample and loading buffer on the agarose gel in 1× TAE buffer, the electrophoresis was performed at 120-130V for 20-30 min. The nuclei acid on the gel was visualized under UV light and recorded by Intas gel iX imager

1X TAE buffer: 40mM Tris, 20mM Acetate, 1mM EDTA, pH around 8.6, do not adjust
Ethidium bromide solution: 1 mg/l ethidium bromide in 1× TAE buffer

2.2.2.10 Restriction endonuclease digestions

Plasmids and purified PCR products were digested with restriction enzymes according to manufacturer's instructions. The reactions were incubated at optimum enzyme temperature for 5h or overnight.

2.2.2.11 Ligation

The purified insert DNA and linearized vector were used for ligation with T4 DNA ligase to generate recombinant construct. The 10 μl ligation mixture contained 1.0 μl 10×T4 DNA Ligase buffer, x μl vector DNA, y μl insert DNA, nuclease-free water and 0.5 μl T4 DNA Ligase. The volumes of vector DNA and insert DNA were calculated based on insert to vector molar ratio (usually 3:1 or 5:1) using the website: https://nebiocalculator.neb.com/#!/ligation. The reaction was incubated at 22 °C for 10 min to 1h, or 16°C overnight.

2.2.2.12 DNA sequencing

Recombinant constructs were prepared as the requirements of eurofins/GATC company for sequencing.

2.2.2.13 Genome walking

Four genome walking libraries of *C.plantagineum* were built according to the the GenomeWalker^{TM} Universal kit (Clontech, Heidelberg, Germany) manufacture instruction. The upstream and downstream of CpWAK genes were sequenced using the adaptor primers AP1 and AP2 from the kit in combination with gene-specific primers (Table 2.1).

2.2.3 Bacterium

2.2.3.1 Preparation of calcium-competent *E. coli*

100 ml of bacterial culture (OD_{600}= 0.5) was centrifuged at 5000 rpm, 4 °C for 5 min. The pellet was collected and washed twice with 1 ml pre-chilled 0.1 M $CaCl_2$. Then the pellet was resuspended in 9.0 ml pre-chilled 0.1 M $CaCl_2$, and centrifuged as before. After removing the supernatant, the cells were resuspended with 1 ml pre-chilled 0.1 M $CaCl_2$, and stored at -80°C in 50 µl aliquots with 15% (v/v) glycerol.

2.2.3.2 Transformation of calcium-competent E. coli

50 µl of calcium-competent *E. coli* was thaw on ice and then mixed with 1µl plasmid DNA (5-10 ng/µl) or 1-5 µl of the ligation product. After incubation on ice for 30 min, the mixture was heat-shocked in a water bath at 42 °C for 45-50 seconds. Then 250 µl LB medium was added to the mixture and incubated at 37°C for 1 h. 100-200 µl culture was finally spread on selective agar-plates and incubated at 37ºC overnight.

2.2.3.3 Transformation of *A. tumefaciens* via electroporation (Tung and Chow 1995)

Competent *A. tumefaciens* cells (provided by Dr. Hou) were thawed on ice. 1µl of plasmid DNA (approximately 10-50 ng/µl) was mixed with the electro-competent cells in a pre-chilled 2 mm Electroporation–cuvette (Bio-Rad, Germany). The plasmid was introduced into the cells by applying a single electrical pulse of 3 to 5 sec (the parameters: 25 µF Capacity, 2.5 kV power, 400 Ω resistance) (GenePulser II, Bio-Rad). The cells were immediately transferred into 1 ml YEB-medium and incubated for 2-3 h at 28 ºC under agitation (250 rpm). 100-200 µl culture was finally spread on selective agar-plates and incubated at 28 ºC for 2-3 d.

2.2.3.4 Screening for transformed bacterial clones

The transformed bacterial clones can grow on selective agar-plates containing appropriate antibiotics after overnight (16-18 hours) incubation. Then the transformed colonies were verified by colony PCR. The plasmids of the positive clones were isolated for sequencing.

2.2.3.5 Preparation of bacterial glycerol stocks

Overnight culture of one bacterial clone picked from a selective plate was prepared for a glycerol stock. 700 µl of the bacterial culture was thoroughly mixed with 300 µl of autoclaved 100% (v/v) glycerol solution in a sterilized screw-cap tube. Then the tube was immediately frozen into liquid nitrogen and stored at -80 °C

2.2.3.6 *A. tumefaciens*-mediated transient transformation of *Nicotiana benthamiana* leaves (Schutze et al. 2009)

One single *A. tumefaciens* colony was inoculated into 5 mL of YEB medium with appropriate antibiotics (50 μg/ml kanamycin, 50 μg/ml rifampicin and 100μg/ml spectinomycin) and incubated at 28ºC overnight under agitation (250 rpm) (OD_{600} 0.8-1.0). Then the cells were washed and diluted with AS medium to OD_{600} 0.7-0.8. The suspensions were mixed with p19 plasmid at a 1:1 ratio and kept at room temperature for 2-4 h. Five-week-old *N.benthamiana* plants are suitable for infiltration. The plants need to be watered with excess of water before infiltration. Two or three leaves per plants were infiltrated with the same sample using a syringe. The fluorescence in the epidermal cell layer can be observed 1-3 days after infiltration.

AS medium: 10 mM $MgCl_2$; 150 μM acetosyringone; 10 mM MES-KOH pH 5.6

2.2.4 Protein

2.2.4.1 Induction of recombinant proteins from *E. coli* cells

The recombinant protein was strongly induced by adding IPTG to a concentration of 1 mM in *E. coli* BL21 (DE3) at 22-26°C for 3-5h in the dark. IPTG was added to the bacteria culture when its OD_{600} was approximately 0.5 at 37°C. Then the culture was incubated in 26°C for 15 min in the dark. 1 ml-culture sample was taken before and every one or two hours after addition of IPTG. Before sample collection, the OD_{600} of each sample was measured to determine the volume of the 1× Laemmli buffer needed. All samples were then centrifuged at 14,000 rpm, 4 °C for 5 min. The supernatants were discarded, and the pellets were dissolved in 1× Laemmli sample buffer and then treated according to the modified rapid protein extraction protocol (2.2.4.3) to confirm the successful induction of recombinant protein. The remaining of the culture was centrifuged at 8800 rpm, 4 °C for 20 min, and then pellets were kept at -20°C for further experiments.

2.2.4.2 Extraction and purification of His-tag fusion proteins by immobilized metal-ion affinity chromatography

Fusion proteins extraction and purification under native conditions

Soluble His-tagged recombinant proteins were purified on a nickel column under native conditions. The bacterial pellets from 100 ml IPTG-treated culture were lysed by sonication (6 X 20s) in 5 ml of lysis buffer at 4 °C. The crude lysate was centrifuged at 10,000 g, 4 °C for 20 min and then the supernatant and inclusion body were saved for purification using a nickel-affinity column.

The fusion protein in the supernatant was purified under native conditions according to the following protocol: after filtered through a 0.45 μm membrane, the supernatant was load on the nickel-affinity column which was equilibrated with 3 ml of buffer A, and then was washed with 10

ml of buffer A and 8 ml of buffer B. The fusion protein was finally eluted with 3.5 ml of buffer C, and 7 fractions were collected (each 500 µl). The qualitative and quantitative analysis of the protein fractions were performed using SDS-PAGE and the Bradford assay. The column was regenerated as described below.

Lysis buffer: 50 mM NaH_2PO_4, 300 mM NaCl, 5 mM imidazole, 10% (v/v) glycerol, 0.1% (v/v) Triton X-100, 1 mM β-mercaptoethanol, 1 mg ml^{-1} lysozyme (add freshly), pH 8.0
Buffer A: 50 mM NaH_2PO_4, 300 mM NaCl, 5 mM imidazole, 10% (v/v) glycerol, 0.1% (v/v) Triton X-100, 1 mM β-mercaptoethanol, pH 8.0
Buffer B: 50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, 10% (v/v) glycerol, 0.1% (v/v) Triton X-100, 1 mM β-mercaptoethanol, pH 8.0
Buffer C: 50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, 10% (v/v) glycerol, 0.1% (v/v) Triton X-100, 1 mM β-mercaptoethanol, pH 8.0

All the β-mercaptoethanol was added before use.

Fusion proteins extraction and purification from inclusion body

The target protein purification from inclusion body was under denaturing conditions. The inclusion body obtained after sonication and centrifugation was washed 3 times with IB-wash buffer, 2 times with IB-wash buffer + 0.5% (v/v) Triton X-100. The suspension was centrifuged at 10000 rpm for 30 min at 4°C after each wash. The pellet was then resuspended in 7 ml of equilibration buffer for 1h and then filtered through a 0.45 µm membrane before loading on the column. The extraction was loaded on the column equilibrated with 3 ml equilibration buffer. After the continuous washing with 10 ml equilibration buffer and 12 ml wash buffer, the recombinant fusion protein was finally eluted with 3.5 ml elution buffer. The qualitative and quantitative analysis of the protein fractions were performed using SDS-PAGE and the Bradford assay. The column was regenerated as below.

IB-wash buffer: 20% (w/v) sucrose, 3 mM EDTA, pH 7.3, stored at 4°C
Equilibration buffer: 0.1 M NaH_2PO_4, 8 M urea, pH 8.0
Wash buffer: 0.1 M NaH_2PO_4, 8 M urea, pH 6.3
Elution buffer: 6 M urea, 100 mM HEPES, 500 mM imidazole, pH 7.9

Nickel affinity column regeneration

After purification, the nickel column was washed with 3 ml of water and then 5 ml of 6 M guanidine-HCl (pH 7.5). The guanidine was removed with 3 ml of water. And the impure proteins in the column were removed with 5 ml of 0.1 M EDTA (pH 7 or 8) and 3 ml of water. The column was recharged with 3 ml of charge buffer and 3 ml of water. Finally the column was stored in 30 % ethanol at 4°C.

Charge buffer: 10 mg/ml nickel (II) sulfate hexahydrate

2.2.4.3 Extraction and purification of GST-CpCRP1 fusion proteins under native conditions using glutathione-affinity chromatography

The extraction of GST-CpCRP1 fusion proteins from crude lysate was conducted as described in **section 2.2.4.2.**

The fusion protein in the supernatant was purified under native conditions according to the following protocol: after filtration through a 0.45 µm membrane, the supernatant was loaded on the Glutathione Sepharose 4B resin (GE Healthcare Life Science) which was equilibrated with 5 ml of PBS, and then was washed with 10 ml of PBS/EDTA/PMSF and 10 ml of buffer PBS/EDTA. The fusion protein was finally eluted with 4 ml of glutathione buffer, and 8 fractions were collected (each 500 µl). The qualitative and quantitative analysis of the protein fractions were performed using SDS-PAGE and the Bradford assay. The column was regenerated and stored with PBS buffer.

Lysis buffer: 50 mM Tris; 50 mM NaCl; 5 mM EDTA; 1 µg/ml leupeptin; 0.15 mM PMSF; 1 mM β-mercaptoethanol (add freshly), pH 8.0
1X PBS: 20 mM sodium phosphate; 300 mM sodium chloride; pH 7.4
PBS/EDTA/PMSF: 1x PBS; 5 mM EDTA; 0.15 mM PMSF; pH 7.4 (fresh made)
PBS/EDTA: 1x PBS; 5 mM EDTA; pH 7.4
Glutathione buffer: 50 mM Tris; 10 mM reduced glutathione; pH 8.0 (fresh made)

The purified recombinant proteins were then desalted or renatured by disposable PD-10 columns packed with Sephadex G-25 Medium (GE Healthcare Life Science, Freiburg, Germany), and then lyophilized or concentrated using Amicon® Ultra-4 10K centrifugal filter devices (Merck Millipore).

2.2.4.4 Pull-down assay

The Pull-down assay was carried out using HisPur™ cobalt resin spin columns according to the Pierce™ Pull-Down PolyHis protein interaction protocol with some modification. The bait protein (His-CpGLP1) and prey protein (GST-CpCRP1) were purified in native conditions and then concentrated using Amicon® Ultra-4 10K centrifugal filter devices (Merck Millipore). Two resin spin columns were used and labeled as Sample (bait and prey) and Control-Prey (no bait, only prey) respectively. The columns were equilibrated with 400 µl of equilibration buffer and then centrifuged for 30 s at 1300 g. After being washed for five times, the sample column was loaded with about 800 µl of bait protein and incubated on the rotating platform for 1 h. The bait flow-through sample was collected. The sample column was washed with 400 µl of wash buffer for five times and the wash samples were also collected. Both the sample and control-prey columns were then loaded with 400-500 µl of prey protein and incubated on the rotating platform for 2 h. The prey flow-through samples from the two columns were collected. After being washed as before, the immobilized proteins on the two columns were eluted with 250 µl of elution buffer. The elution samples from two columns were collected. The pull-down assay result was visualized

using Coomassie blue staining and Western-blot. The columns were washed and regenerated with resin regeneration buffer. The experiment was performed at 4 °C or on ice.

Equilibration buffer: 1x PBS; 10 mM imidazole; pH 7.4
Wash buffer: 1x PBS; 25 mM imidazole; pH 7.4
Elution buffer: 1x PBS; 290 mM imidazole; pH 7.4
Resin regeneration buffer: 20 mM 2-(N-morpholine)-ethanesulfonic acid (MES), 0.1 M sodium chloride; pH 5.0

On-column pull-down assay

The pull-down assay was also carried out using the bait protein (His-CpGLP1) extracted from inclusion bodies. The on-column refolding was conducted according to Oganesyan et al. (2004) with some modification. Two gravity nickel affinity columns were used and labeled as sample (bait and prey) and control-bait (no prey, only bait) respectively. The two columns were loaded with the same amount of bait protein, and then washed with 10 CV (column volume) of Wash buffer 1, Wash buffer 2, Wash buffer 3 and Wash buffer 4 for protein refolding. Then samples were collected after final wash. The two columns then were equilibrated with 15 CV of equilibration buffer. The sample column was then loaded with concentrated prey protein and incubated on a shaker overnight. The flow-through sample was collected from sample column. The two columns were washed with 10 CV of wash buffer. The wash samples were collected. The bait and prey proteins were finally eluted with 1.5 CV of elution buffer. The elution samples from two columns were also collected. The pull-down assay result was visualized using silver staining. The columns were regenerated as before. The experiment was performed at 4 °C or on ice

Wash Buffer 1: 20 mM Tris; 0.1 M NaCl; pH 8.0
Wash Buffer 2: 20 mM Tris; 0.1 M NaCl; 0.1% Triton X-100; 10% glycerol; pH 8.0
Wash Buffer 3: 20 mM Tris; 0.1 M NaCl; 5 mM β-cyclodextrin; 10% glycerol; pH 8.0
Wash Buffer 4: 20 mM Tris; 0.1 M NaCl; 10% glycerol; pH 8.0
Equilibration buffer: 20 mM Tris; 0.1 M NaCl; 10 mM imidazole; pH 8.0
Wash buffer: 20 mM Tris; 0.1 M NaCl; 25 mM imidazole; pH 8.0
Elution buffer: 20 mM Tris; 0.1 M NaCl; 300 mM imidazole; pH 8.0

2.2.4.5 Rapid protein extraction

The modified rapid protein extraction from the Laemmli protein extraction protocol (Laemmli 1970) was used for rapidly extracting the total proteins from both bacteria and plant. Bacteria pellet or finely ground plant material was homogenized with 1× Laemmli sample buffer. Subsequently, the samples were boiled at 95 °C for 5 min and then centrifuged for 5 min (13,000 rpm, RT) to separate insoluble material. The supernatant was decanted to other tubes ready for SDS-PAGE or stored at -20 °C. The samples were boiled at 95 °C for 5 min each time before loading on the gel.

Laemmli buffer (1×): 50 mM Tris-Cl pH 6.8; 10% (v/v) glycerol; 2% SDS (w/v); 0.005%

(w/v) bromophenol blue and 0.1 M DTT (add freshly just before use)

2.2.4.6 Cell wall proteins extraction

Cell wall proteins extraction was carried out according to the protocol 3 in Printz et al (2015). 5 g of plant material was ground in liquid nitrogen for cell wall enrichment. The ground sample was homogenized with 20 mL of buffer A and incubated on a rocking platform at 4℃ overnight. The slurry was then centrifuged (1000 g, 15 min, 4℃) to collect the pellet. Subsequently, the pellet was resuspended with 10 mL buffer B and incubated on a rocking platform at 4℃ for 30 min. The pellet was collected by centrifugation as before. After that, the pellet was washed respectively with 10 mL of buffer C and twice with 10 mL of buffer D for cell wall material preparation.

The prepared cell wall fraction was resuspended with 7.5 mL of buffer AC and incubated on a rocking platform for 30 min at 4℃. The supernatant was saved as $CaCl_2$ fraction after centrifugation (10,000 g, 15 min, 4℃). This collection step was repeated again, and the supernatants were pooled. Then the remaining pellet was resuspended respectively thrice in 10 mL of buffer AE and shaken vigorously at 37℃ for 1 h, once in 15 mL of buffer AL and incubated on a rocking platform at 4℃ overnight. All the supernatants were pooled as EGTA fraction and LiCl fraction respectively after centrifuation as before. Every step was carried out on ice or at 4℃.

Buffer A: 5 mM Na acetate, 0.4 M sucrose, pH 4.6, 4℃
Buffer B: 5 mM Na acetate, 0.6 M sucrose, pH 4.6, 4℃
Buffer C: 5 mM Na acetate, 1M sucrose, pH 4.6, 4℃
Buffer D: 5 mM Na acetate, pH 4.6, 4℃
Buffer AC: 5 mM Na acetate,200 mM $CaCl_2$, pH 4.6, 4℃
Buffer AE: 5 mM Na acetate, 50 mM EGTA, pH 4.6
Buffer AL: 5 mM Na acetate, 3 M LiCl, pH 4.6, 4℃

The $CaCl_2$, EGTA and LiCl fractions were concentrated by Amicon® Ultra centrifugal filter, and then precipitated with 4 volumes of cold acetone at -20℃ for at least 30 min or overnight. Then protein was centrifuged at 15000g, 4℃ for 15 min. If additional cycles of precipitation are necessary to completely remove the interfering substance, then repeat it again. The pellet was dissolved in a minimal volume of 1X lammeli buffer, after being dried under a hood to eliminate acetone residue. The samples was analyzed using SDS-PAGE and western blot or stored at -20℃.

2.2.4.7 Determination of total protein concentration (Bradford assay)

Protein concentration was determined according to Bradford (1976). 1-5 µl of protein samples were diluted to 800 µl with distilled water and incubated with 200 µl of Bradford-reagent for 5 min at room temperature. Water was used as control. The optical density at 595 nm was determined using a spectrophotometer with the control as blank. The protein concentration was determined from a standard curve of bovine serum albumin (BSA).

The SDS in the protein sample dissolved in Laemmli buffer was removed for Bradford assay. 1-5

µl of protein samples were incubated with 100 µl of 0.1 M postassium phosphate buffer (pH 6.8) for 10 min at RT and centrifuged for 5 min at RT. The supernatant was diluted to 800 µl and then the Bradford assay was carried out as described before. The Laemmli buffer was used as control.

2.2.4.8 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE, a technique widely used for separating proteins based on the molecular weight, was performed according to Laemmli (1970). Two parts of the gel, one named stacking gel and the other referred to as separating gel, were made as described below. Before loading, proteins were heated at 95℃ for 5 minutes to break up the spatial structures. The gels were run within the 1× SDS running buffer for about 2 hours, with samples concentrated in the stacking gel at 200V (10-15 mA) and separated in the separating gel at 400V (20-25 mA). The protein markers contain: β-galactosidase (*E. coli*; 116.0 kDa), Bovine serum albumin (bovine plasma; 66.2 kDa), Ovalbumin (chicken egg white; 45.0 kDa), Lactate dehydrogenase (porcine muscle; 35.0 kDa), Restriction endonuclease Bsp98I (*E. coli*; 25.0 kDa), β -lactoglobulin (bovine milk; 18.4 kDa) and Lysozyme (chicken egg white; 14.4).

4% stacking gel (3 ml): 2.16 ml sterile H_2O; 375 µl 1 M Tris-HCl pH 6.8; 405µl 30% (v/v) acrylamide; 30 µl 10% (w/v) SDS; 3 µl TEMED; 30 µl 10% (w/v) APS
12% separating gel (6 ml): 1.92 ml sterile H_2O; 1.56 ml 1.5 M Tris-HCl pH 8.8; 2.4 ml 30% (v/v) acrylamide; 60 µl 10% (w/v) SDS; 2.4 µl TEMED; 60 µl 10% (w/v) APS
1× SDS running buffer: 25 mM Tris; 192 mM glycine; 0.1% (w/v) SDS. The pH was not adjusted.

2.2.4.9 Gel staining

Coomassie blue staining of SDS-PAGE (Zehr et al. 1989)

After electrophoresis, the gels were incubated into fixation solution for 1-2 hours with gentle shaking. Then the gels were washed three times with distilled water for 10 minutes, followed by an incubation in the staining solution on a shaker overnight. After that, the gels were rinsed in distilled water several times to destain background. Images were scanned by ImageScannerTM Ⅲ (GE Healthcare Life Science, Buckinghamshire, UK)

Fixation solution: 40% (v/v) methanol; 10% (v/v) acetic acid.
Staining stock solution: 100 g/l ammonium sulfate; 1% (v/v) phosphoric acid; 0.1% (w/v) Coomassie blue G-250.
Staining solution: 4 volumes staining stock solution + 1 volume methanol

Silver staining

The sensitivity of silver staining was much stronger than that of Coomassie blue staining. The SDS-PAGE gel was first incubated in the fixation solution for 1 h. Then the gel was soaked in 0.02% Na_2SO_3 for only 1 min, after being rinsed with water for at least 30 min. The gel was washed in

water for 3x20 sec and then incubated in cold staining solution for 20 min at 4℃. After staining, the gel was washed again for 3x20 sec and 1x1 min, and then soaked in the reduction solution until it turned yellow. The reaction was terminated by addition of 5% acetic acid for 5 min. The gel was kept in 1% acetic acid at 4℃.

Fixation solution: 40% (v/v) ethanol; 10% (v/v) acetic acid, 50% H_2O.
Staining solution:1g/l AgNO3, 0.002% formaldehyde, stored at 4℃ (add formaldehyde just before use)
Reduction solution: 3% sodium carbonate, 0.05% formaldehyde (add formaldehyde just before use)

2.2.4.10 Western blot

Western blot, also referred as to protein immunoblot, is always used for detecting a specific protein using antibodies. After SDS-PAGE, the separated proteins are transferred to the nitrocellulose Protran BA-85 membrane (Whatman), making it accessible to detect antibodies. The electro-blotting ran with the pre-chilled transfer buffer in an electro-blotting system at 70 V for 1-2 h (Towbin 1979). To control protein transfer efficiency, the membrane (protein-side up) was stained in Ponceau-Red solution for 5-10 min with gentle shaking. The position of protein markers were marked by pencil, followed by destaining the membrane with distilled water. Afterwards, the membrane was blocked by gently shaking in the blocking solution overnight at 4 °C or for 1 h at room temperature. The target proteins were probed at room temperature for 1 h by the corresponding antibodies diluted in the blocking solutions. The ratio of the antibody and blocking solution was determined empirically, ranging from 1:1000 (v/v) to 1:5000 (v/v). Subsequently, the membrane was washed in TBST solution (1× rinse, 1× 15 min and 3× 5 min). Then the secondary antibody blocking solution, anti-rabbit IgG coupled to horseradish peroxidase diluted in blocking solution as 1:5000, was used to detect the primary antibody on the membrane, being incubated for 45 min at room temperature. The membrane was then washed as described before. The ECL Plus Western blotting detection kit (Amersham, Braunschweig, Germany) was used to reveal the binding reactions among target proteins and antibodies. The chemiluminescence signal was detected by a CCD camera (Intelligent Dark Box II, Fujifilm Corporation).

Transfer buffer: 25 mM Tris; 192 mM glycine; 20% (v/v) methanol. Stored at 4℃
It is not necessary to adjust the pH.
Ponceau-Red solution: 0.2% (w/v) Ponceau S in 3% (w/v) Trichloroacetic acid (TCA)
10× TBS (pH 7.5): 200 mM Tris-HCl (24.2 g/l); 1.5 M NaCl (87.6 g/l); add H_2O to 1 L after adjusting the pH with HCl.
TBST solution: 1× TBS + 0.1% (v/v) Tween-20
Blocking solution: 4% (w/v) non-fat dry-milk powder dissolved in TBST

2.2.4.11 ELISA binding assay

Direct ELISA binding assay

The ELISA binding assay was performed according to Decreux and Messiaen (2005) with modifications. Nunc Maxisorp flat-bottom plates (Invitrogen, CA, USA) were coated with pectin solution (25 µg well–1) at 4°C overnight. Non-specific binding sites were blocked for 2 h at room temperature with 100 µl of blocking solution. The wells were incubated for 2 h at room temperature with 50 µl of purified His-tagged recombinant protein in binding buffer after removing the blocking solution. The wells were washed four times with wash buffer and incubated with 50 µl of anti-His-tag antibody (1:10,000) (Invitrogen) or anti-WAK antibody (1: 2500) or anti-GRP antibody (1:5,000) in incubation buffer for 1 h at room temperature. After washing wells four times, 50 µl of goat anti-rabbit IgG peroxidase antibody (1:10,000) (Sigma, A9169) prepared in incubation buffer were added and incubated for 1 h at room temperature. After washing the plates six times, the bound recombinant protein was visualized in the presence of the TMB substrate (Sigma, T2885). The absorbance was measured at 450 nm after sufficient color development in the dark and the reaction was stopped by adding 50 µl of 10% (v/v) phosphoric acid.

Competitive ELISA binding assay

The procedures of the competitive ELISA binding assay were similar to the direct ELISA binding assay. The recombinant protein was pre-incubated with another protein for 1h at RT. Then the mixture was loaded on the pectin-coated and blocked wells of the Nunc plates and incubated for 2h at RT as described in direct ELISA binding assay. The recombinant protein was detected by the pectin immoblized on the wells. All following steps were identical to the direct ELISA binding assay.

Wash buffer: 20 mM Tris-HCl, 150 mM NaCl, pH 4, 5, 6, 7, 8.0
Blocking solution: 3% (w/v) low fat dried milk dissolved in wash buffer
Binding buffer: 1% low fat dried milk, 20 mM Tris-HCl, 150 mM NaCl, with or without 2 mM $CaCl_2$, pH 4, 5, 6, 7, 8.
Incubation buffer: 1% low fat dried milk, 20 mM Tris-HCl, 150 mM NaCl, pH 4, 5, 6, 7, 8.

2.2.4.12 Bimolecular fluorescence complementation (BiFC) assay

The bimolecular flurescence complementation (BiFC) was carried out to confirm protein-protein interaction in planta according to Grefen and Blatt (2012) and Schutze et al. (2009) with some modifications. The vectors used in the experiment were constructed using both the pBiFCt-2in1 vector series (Grefen and Blatt, 2012) and MultiSite Gateway system. The CpGLP1 sequence encoding amino acids 29-226 (protein fragment without N-terminal signal peptide) was subcloned into pDONR221-P3P2 and the CpCRP1 sequences encoding amino acids 25-166 (protein fragment without N-terminal signal peptide) was subcloned into pDONR221-P1P4. The three entry clones (**Supplementary figure 9**) were constructed using BP reactions. These fragments were finally recombined from the entry clones into pBiFC-2in1 gateway vectors using LR reactions. The expression clones (**Supplementary figure 9**) were finally transformed into the *Agrobacterium tumefaciens* strain GV3101 (pMP90) respectively. The fluorescence was observed 1-3 days after infiltration using a confocal laser scanner microscope (Zeiss LSM-780; Carl Zeiss

Microscopy, Jena, Germany) with standardized excitation intensities and photomultiplier gains. The excitation wavelength for YFP was 514 and emitted light was detected at 521–553. In addition to the pBiFCt-2in1 vector series, the pBatTL-B-sYFPc or pBatTL-B-sYFPn gateway vectors were also used for BiFC assay. The full-length CDS of CpGLP1 and CpCRP1 were subcloned into pDONR201. The following steps were the same as before.

2.2.5 Pectin

2.2.5.1 Pectin extraction

Pectin was extracted according to Cornuault et al. (2014) with some modifications. Three mg freeze-dried plant material was homogenized with 1 ml of 50 mM CDTA (1, 2-cyclohexanediaminetetraacetic acid), pH 7.5. The mixture was incubated on a shaking platform for 1 h at room temperature and then centrifuged at 16,000 × g for 15 min. The supernatants were collected and stored at -20 °C.

2.2.5.2 Pectin estimation

The galacturonic acid content of the CDTA fraction was determined according to Verma et al. (2014). 2.4 ml of 75 mM sodium tetraborate in H_2SO_4 were mixed with 400 µl of the CDTA fraction, and heated in a water bath at 100°C for 15 min. The reaction mixture was cooled down in an ice bath for 10 min. Eighty µl of m-hydroxydiphenyl solution (80 µl of 0.5% NaOH as blank) was added to the mixture and vortexed thoroughly. After 5 to 10 min at room temperature, the absorbance was read at 525 nm. A galacturonic acid standard curve was obtained using commercial polygalacturoinc acid (Sigma 81325).

2.2.6 Bioinformatic analysis

2.2.6.1 Nucleic acid sequence analysis

The gene structures were constructed using the Gene Structure Display Server 2.0 (http://gsds.cbi.pku.edu.cn/) (Hu et al. 2015). The DNA sequences alignments were carried out using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). The putative cis-acting elements in promoters were identifed using PLACE database (Higo et al. 1999) and PlantCARE (Lescot et al. 2002).

2.2.6.2 Protein sequence analysis

The basic characteristics of proteins were obtained using Sequence Manipulation Suite (http://www.bioinformatics.org/sms2/index.html). The protein sequence alignments were carried out using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). Homologs were retrieved with BLAST algorithm from GenBank® (http://blast.ncbi.nlm.nih.gov/Blast.cgi) or from the transcriptomic databanks of *L.brevidens* and *L. subracemosa* (VanBuren et al. 2018) with the help

of Dr.Giarola. Protein domains were identified using NCBI CD-Search tool (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (Marchler-Bauer et al. 2011), ExPASy ScanProsite tool (http://prosite.expasy.org/scanprosite/) (de Castro et al. 2006) and TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/). The phylogenetic tree was generated using MEGA5.1 software.

2.2.6.3 RNA-seq analysis

The RNA samples were prepared as described in **section 2.2.2.4.** The single-end RNA-seq libraries were prepared, sequenced and partially analyzed by GATC Biotech Company (Germany). The RNA-seq raw reads of all samples were filtered and then mapped and assembled to *A. thaliana* genome using Bowtie and TopHat. The alignment results were provided as binary SAM (BAM) format and then processed with Cufflinks and Cuffmerge to identify and annotate the full-length transcripts. The differential expression levels between wild-type and mutants were determined with the transcripts estimated as FPKM (fragments per kilobase of exon per million fragments mapped) using Cuffdiff. All genes with $q < 0.05$ and an absolute value of the $\log_2$ (fold change) > 1 were filtered as significantly up- or down- regulated items and used for further analysis. The scatter plots and volcano plots were generated based on the expression data using the cummerbund package. The Venn diagram was obtained on the website http://bioinformatics.psb.ugent.be/webtools/Venn/. The heatmap was plotted and visualized using Cluster 3.0 (de Hoon et al. 2004) and JAVA TreeView (Saldanha 2004) respectively. DAVID bioinformatics resources 6.8 (https://david.ncifcrf.gov/home.jsp) was used for GO (Gene ontology) (p<0.05) and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment analyses (p<0.001) (Huang et al. 2008; 2009). The overrepresented KEGG pathways were identified in the KEGG pathway database (https://www.genome.jp/kegg/pathway.html).

2.2.6.4 Statistical analysis

All the statistical analyses in the work were performed with Excel and Graphpad prism 5.0

3 RESULTS

3.1 Study on *C.plantagineum* wall-associated protein kinases (CpWAKs)

In silico analysis of proteins is an effiecnt way to aquire the basic informations of the proteins, such as the amino acid composition and phylogenetic position, which will provide the directional cues for the experimentall studies *in vitro* and *in vivo* and thus lay a firm foundation for figuring out the molecular functions and biological processes the proteins involved in. In this study some bioinformatic tools, databases and large-scale studies were exploited for the characterization of CpWAK genes.

3.1.1 *In silico* analysis of CpWAK genes

3.1.1.1 Basic characterization of CpWAK genes

CpWAK1 and CpWAK2 were first identified and characterized by Giarola et al. (2016). In this thesis, CpWAK3 is identified as an additional wall-associated protein kinase homolog. Both full-length nucleic acid and amino acid sequences of the three CpWAKs are shown in the **Figure 3.1** and **Supplementary figure 1 and 2**. There is no intron in all three CpWAK genes according to the genomic sequencing results (**Figure 3.4**) (the coding sequences of CpWAK1: 2266 bp, CpWAK2: 2325 bp, CpWAK3: 2169 bp). The basic characteristics of the predicted CpWAK proteins were obtained using Sequence Manipulation Suite (http://www.bioinformatics.org/sms2/index.html). The number of amino acids, calculated molecular weight, theoretical isoelectric point (pI), and grand average of hydropathy are presented in **Table 3.1**. From knowledge of the amino acid composition and the hydropathy plots of CpWAKs (**Figure 3.1 and 3.2**), the distribution of polar and non-polar amino acids is roughly the same (**Table 3.1**).

The three CpWAK homologues are highly conserved with CpWAK3 being more similar to CpWAK1 according to the alignment analysis (**Figure 3.3**). The identification of protein domains was carried out based on Giarola et al. (2016). The three CpWAK proteins show the features of wall-associated kinases, which contain the extracellular galacturonide-binding domain, EGF (epidermal growth factor) repeats, transmembrane domain and the cytoplasmic protein kinase domain (**Figure 3.3**). Except for the predicted N-terminal signal position and the transmembrane domain, other domains are relatively weakly hydrophilic (**Figure 3.2**).

Table 3.1 Basic characteristics of CpWAKs

	Amino acids	Molecular weight (kDa)	Isoelectric point pI	Grand average of hydropathy
CpWAK1	708	78.71	6.35	-0.107
CpWAK2	726	81.59	6.75	-0.177
CpWAK3	722	80.42	5.78	-0.153

	Non-polar (%)	Polar (%)	Acidic and polar (%)	Basic and polar (%)
CpWAK1	38.98	38.14	11.02	11.86
CpWAK2	37.47	39.39	10.88	12.26
CpWAK3	37.81	38.78	11.63	11.77

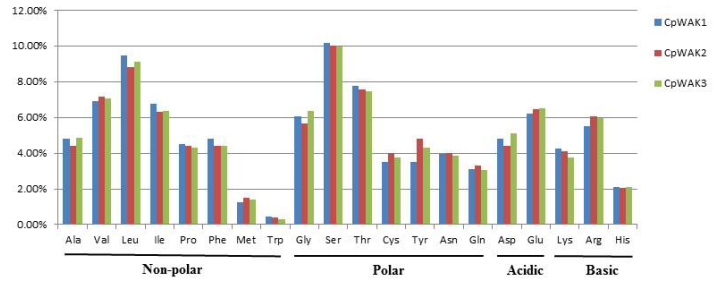
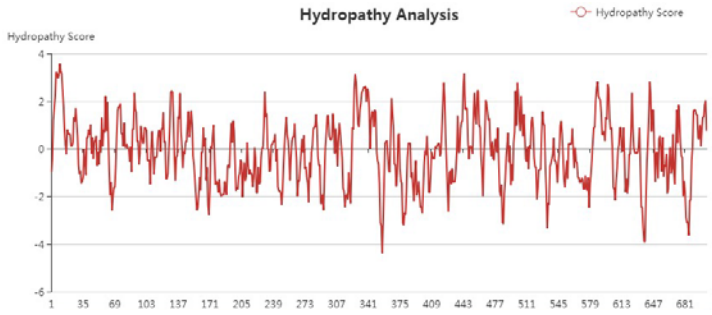
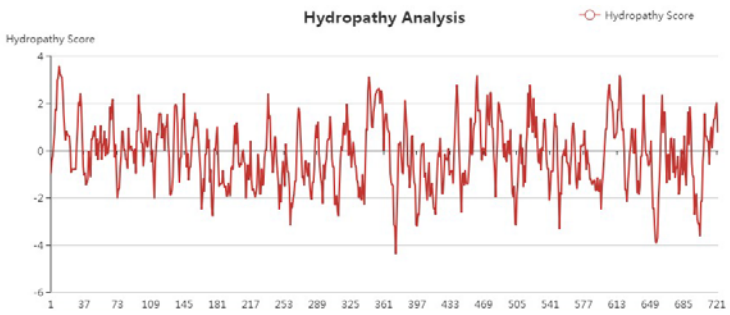


Figure 3.1 Amino acid compositions of proteins

CpWAK1



CpWAK2



CpWAK3

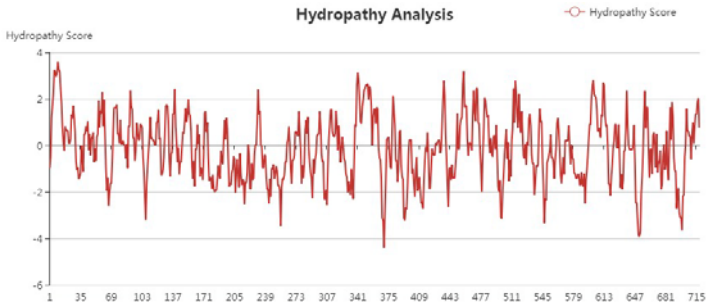


Figure 3.2 Hydropathy plots of CpWAKs

Figure 3.3 Alignment of the predicted CpWAK1, CpWAK2 and CpWAK3 protein sequences (modified from Giarola et al. 2016). The alignment analysis of the three CpWAK protein sequences was performed using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). The symbols in the figure indicate the fully conserved positions (asterisk "*"), conserved substitutions (colon ":"), and semi-conserved substitutions (period ".") respectively. Protein domains were identified according to Giarola et al. (2016). Different coloured strips highlight different domains. Red: predicted signal peptide; blue: galacturonan-binding domain; dark green: EGF-like domain; light green: calcium-binding EGF-like domain; orange: predicted transmembrane domain; yellow: predicted protein kinase domain. The red rectangle shows the invariant catalytic motif RDxxxxN, whose conserved residues are indicated by red arrows.

3.1.1.2 Phylogenetic analysis of CpWAK genes

In order to investigate the evolutionary relationship of CpWAK genes with other species, the phylogenetic analysis of WAK genes from *C.plantagineum* and other species (*Physcomitrella patens*, *Arabidopsis thaliana*, *Oryza sativa*, *Malus domestica*, *Lindernia subracemosa* and *Lindernia brevidens*) was carried out using the complete amino acid coding sequences with the Neighbor-Joining method in MEGA 5.1 software (**Figure 3.4**). The reasons for choosing the five species for the phylogenetic analysis are: 1.The genome-wide analyses of the WAK or WAKL (WAK-like) super family were performed in three species: Arabidopsis (Verica et al. 2003; Verica and He 2002), rice (Oryza sativa) (de Oliveira et al. 2014; Zhang et al. 2005) and apple (Malus domestica) (Zuo et al. 2018) 2. Both the genomic and transcriptomic databanks of *L.subracemosa* and *L.brevidens* are published (VanBuren et al. 2018). The seven species contain the lower plant (the moss *Physcomitrella patens*) and higher plants like monocotyledonous, dicotyledonous and Linderniaceae family members. The WAK homologs for phylogenetic analysis were retrieved using BLASTP analysis from the NCBI database (**Table 3.2**) or TBLASTN analysis in *L. brevidens* and *L. subracemosa* transcriptomic databanks with CpWAK1 protein sequence as query (expectation value *E*<10-10). The top two hits of each species were selected for further phylogenetic analysis. The lengths of all the selected WAKs are around 700 aa except for RXH_67760.1 being 1284 aa. The selected WAK protein structures were predicted using NCBI CD-Search tool (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (Marchler-Bauer et al. 2011), ExPASy ScanProsite tool (http://prosite.expasy.org/scanprosite/) (de Castro et al. 2006) and TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/). The gene structures of all selected WAK genes were displayed using Gene Structure Display Server 2.0 (http://gsds.cbi.pku.edu.cn/) (Hu et al. 2015).

Table 3.2 Selected homologs of CpWAK1 from NCBI database.

Accession number	Organism	Query coverage	Percent identity	E-value
RXH67760.1	*Malus domestica (apple)*	94%	44.85%	0.0
XP_008380329.1	*Malus domestica (apple)*	95%	44.44%	0.0
NP_001185009.1	*Arabidopsis thaliana*	95%	43.06%	3e-175
NP_001323321.1	*Arabidopsis thaliana*	95%	43.20%	6e-174
XP_015627146.1	*Oryza sativa (rice)*	94%	37.73%	2e-135
EAZ21467.1	*Oryza sativa (rice)*	91%	38.14%	1e-134
XP_024376490.1	*Physcomitrella patens*	51%	43.34%	2e-84
XP_024376488.1	*Physcomitrella patens*	51%	43.34%	3e-84

The first two items were selected from each species for further phylogenetic analysis after BLASTP analysis in NCBI database with CpWAK1 protein sequence as query.

The phylogenetic analysis manifested that all homologs are divided into four clusters which are in line with the genus classification (group I: moss, group II: monocots, group III: dicots, group IV: Linderniaceae family) and only WAK genes in group IV have no introns or fewer introns than others (**Figure 3.4**). However, the protein structures of the WAK homologs are highly similar,

especially in higher plants. All selected WAK proteins in higher plants contain the conserved kinase domain，the EGF-like domain and the extracellular galacturonan-binding domain, while a MATE-like domain (cd13132) is present in RXH_67760.1 and a PNPLA domain (PS51635) in Lbr_010788 (**Figure 3.4**). The MATE-like domain is related to iron homeostatis under osmotic stress, and the PNPLA domain is involved in cleavage of fatty acids from membrane lipids during defense against pathogens. No transmembrane domain locates in NP_001323321.1. The WAK-like proteins of *P.patens* showed similarity only with a conserved kinase domain and a special Cupredoxin motif (cl19115) located on the extracellular domain.

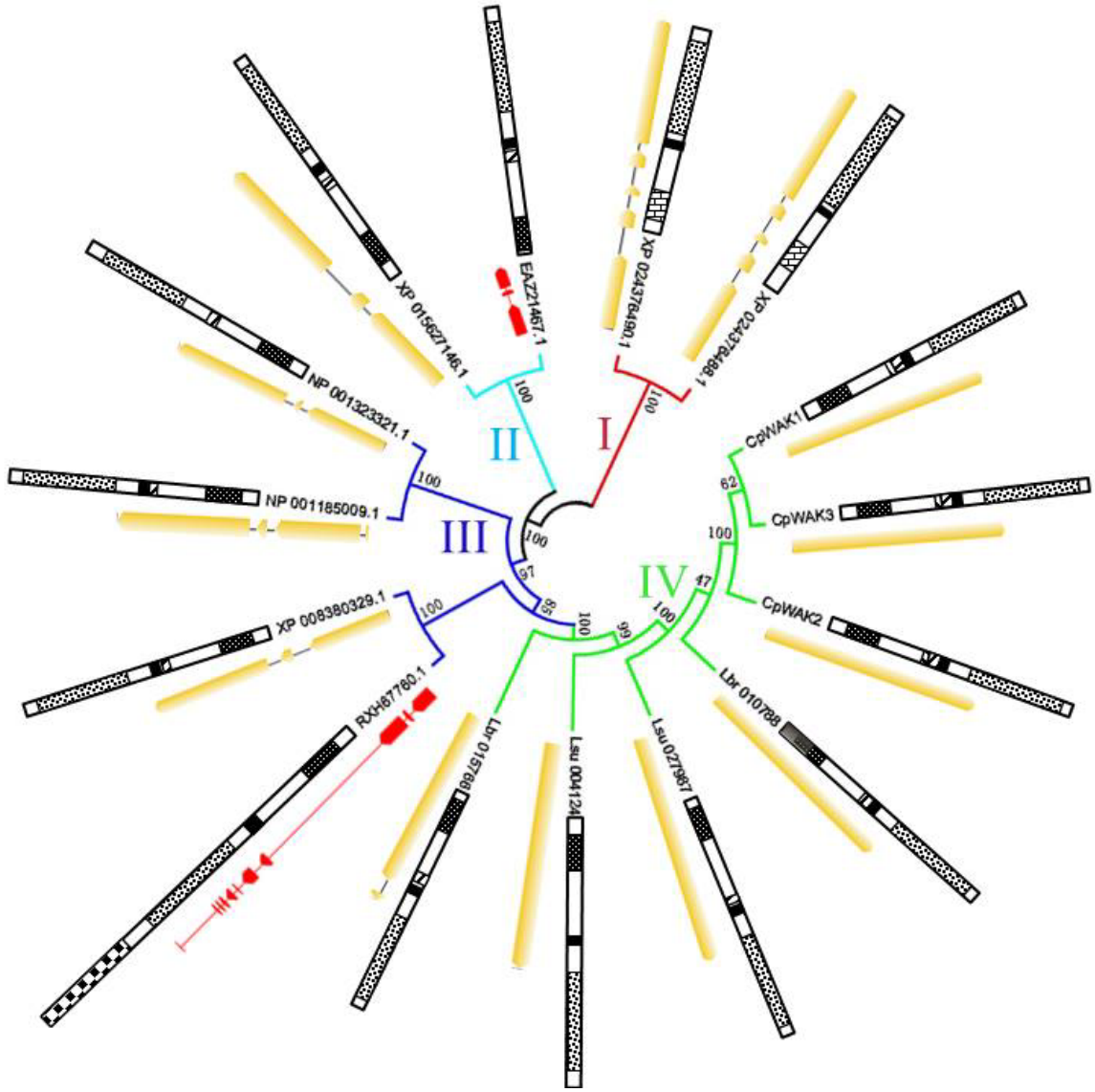


Figure 3.4 Phylogenetic analysis of WAK proteins. The 15 predicted WAK protein sequences of *C.plantagineum*, *L. subracemosa*, *L. brevidens*, *A. thaliana*, *Malus domestica*, *Oryza sativa* and *Physcomitrella patens* were subjected to a multiple sequence alignment using the ClustalW program of MEGA5.1 software. The phylogenetic tree was then constructed by neighbor-joining method with 1000 bootstrap replications in MEGA5.1. The bootstrap values displayed on branches indicate the reliable level to the nods of the tree. All homologs were divided into four clusters (group I, II, III, IV distinguished in different colors). The gene structures were generated using the Gene Structure Display Server 2.0 (http://gsds.cbi.pku.edu.cn/) (Hu et al. 2015). .

3.1.1.3 Analysis of the promoters of the *CpWAK* genes

Both the upstream and downstream sequences of the CpWAK coding sequences were successfully cloned using the genome walking technique. As there is no intron in *CpWAK* genes (**Figure 3.4**) and the three CpWAKs are highly similar, gene specific primers were designed using mainly the sequences coding for extracellular domains which show less identity (**Figure 3.3 and Supplementary figure 3**). Six fragments were finally isolated with the gene specific primers walking upstream or downstream of the CpWAK genes (2593 bp of 5' and 435 bp of 3'-terminal sequences for CpWAK1, 1902 bp 5' and 1730 bp of 3'-terminal sequences for CpWAK2, 987 bp of 5' and 225 bp of 3'-terminal sequences for CpWAK3) (**Supplementary figure 1**). In Arabidopsis, the majority of the WAK and WAK- like (WAKL) genes exist as gene clusters, where the distances between two genes range from around 600 bp to more than 3000 bp (Verica et al. 2003; Zheng-Hui He 1999). No gene cluster was identified among the three CpWAK genes using different primer combinations designed on the isolated 5' and 3'-terminal sequences of CpWAKs (data not shown).

The region from -937 (the translation start codon ATG indicated as +1) to +2353 of CpWAK1 shows more similarity with the regions of two other CpWAK genes (CpWAK2: -908 to +2411, CpWAK3: -927 to +2394) according to the multiple alignment (**Supplementary figure 3**). The 5'- upstream regions (-937 to -1) of the three CpWAK genes were regarded as promoters and analysed using PLACE database (Higo et al., 1999) and PlantCARE (Lescot et al. 2002). The putative *cis*-acting regulatory elements and positions are shown in **Table 3.3** and **Figure 3.5**. Most of the motifs are light-responsive elements and one is involved in circadian control. In addition, there are two hormone-related motifs, one related to auxin signaling and another one, W-box also detected on the promoters of AtWAKL genes (Verica et al. 2003) , is associated with salicylic acid (SA) and methyl jasmonate (MeJA) signaling.

Table 3.3 Putative *cis*-acting elements in the promoters of CpWAK genes were predicted using PLACE and PlantCARE

Site Name	Sequence	Numbers	Function
ACE	SHVACGTAYB	2 (CpWAK1)/ 1 (CpWAK2) 2 (CpWAK3)	Involved in light responsiveness
Box4	ATTAAT	1 (CpWAK1)/ 1 (CpWAK2) 1 (CpWAK3)	Involved in light responsiveness
circadian	CAAAGATATC	1 (CpWAK1)/ 1 (CpWAK2) 1 (CpWAK3)	Involved in circadian control

G-box	YACRWN(G)	4 (CpWAK1)/ 4 (CpWAK2) 4 (CpWAK3)	Involved in light responsiveness
I-box	(C)CWYWTMYRMT	2 (CpWAK1)/ 2 (CpWAK2) 2 (CpWAK3)	Part of a light responsive element
TGA-element	AACGAC	2 (CpWAK1)/ 2 (CpWAK2) 2 (CpWAK3)	Auxin-responsive element
W-box	(C)TGAC(Y)	2 (CpWAK1)/ 1 (CpWAK2) 2 (CpWAK3)	Pathogenesis-related and involved in SA and MeJA responsiveness

The letters in the sequence are defined as follows: A = adenine, C = cytosine, G = guanine, T = thymine, R = (G/A), Y = (T/C), K = (G/T), M = (A/C), S = (G/C), W = (A/T), B = (G/T/C), D = (G/A/T), H = (A/C/T), V = (G/C/A), N = (A/G/C/T).

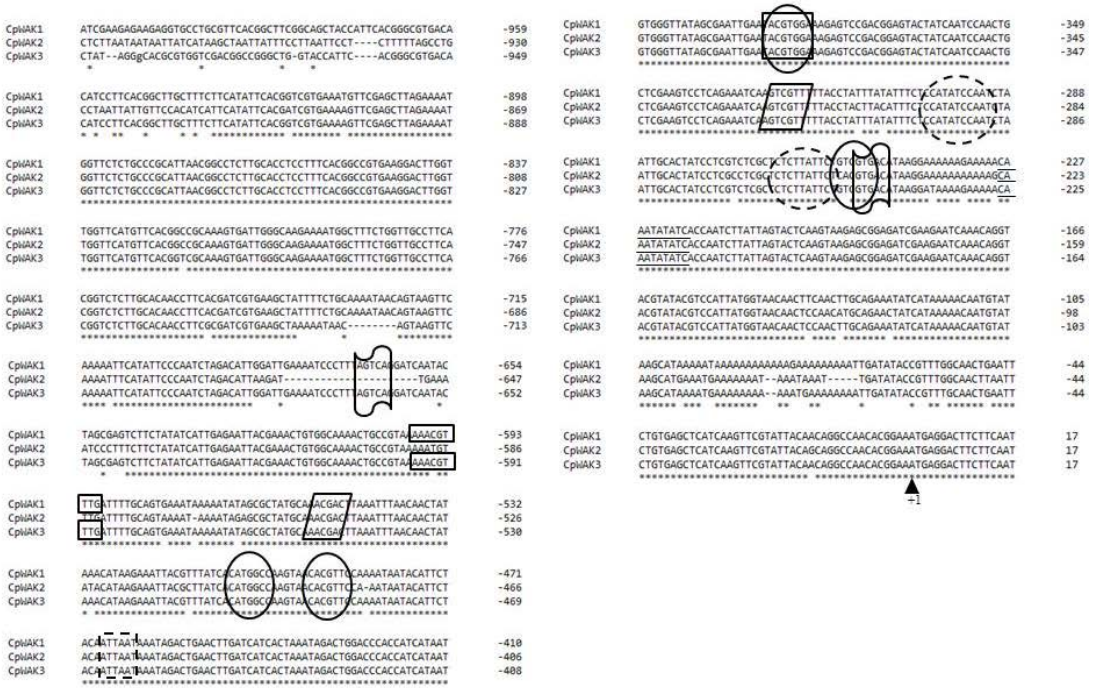


Figure 3.5 Putative *cis*-elements in the CpWAK promoters. The translation start site (triangle) is indicated as +1. The *cis*-acting elements tabulated in **Table 3.3** are highlighted with the following symbols. ▭: ACE, ▭ : Box4, — : Circadian ◯ :G-box; ◯: I-box, ▱: TGA-element, ▭: W-box.

3.1.2 Expression analyses of CpWAK genes

The expression analysis focused on *C.plantagineum* leaves, because CpWAK gene transcripts were mainly expressed in leaves (Giarola et al. 2016). To examine the expression profiles of CpWAK genes, *C.plantagineum* plants were subjected to various treatments. According to the promoter analysis (**section 3.1.3**), the transcript expression of CpWAKs was induced by light, auxin, and SA or MeJA. Therefore, the detached *C.plantagineum* leaves were exposed to 100 µM naphthaleneacetic acid (NAA, a synthetic auxin analogue), 1 mM SA and 100 µM MeJA separately for 48h. The intact *C.plantagineum* plant, cultivated in a day/night cycle of 16/8 hours (daytime: 7:00-23:00), was collected every four hours from 7:00 am to 23.00 pm with the ones in the dark as

control.

The semi-quantitative analysis by RT-PCR demonstrated that none of the three hormones induced CpWAKs expression after two-day treatment, while on the contrary SA and MeJA downregulated CpWAK transcripts remarkably and NAA had no significant effect on CpWAK genes expression (**Figure 3.6 B**). Regarding the effect of light on CpWAK gene expression, **Figure 3.6 A** show that the similar expression level and trends of CpWAKs in both light and dark groups with increasing transcript level from 7:00-11:00 am and 19:00-23:00 pm, decreasing transcript level from 11:00 am-19:00 pm. However, the curves observed in the dark group are more distinct and conserved among the three CpWAK genes than those in the light group (**Figure 3.6 A**). Despite a slight variation at the peaks of CpWAK1 and CpWAK3 in the light group, it still exhibited similar expression curves (**Figure 3.6 A**). The result indicates that the expression of CpWAKs is dictated by a circadian clock and extended darkness has no effect on the rhythm. However, some unknown factors in the light may affect the circadian clock because of the variation on the curves in the light group.

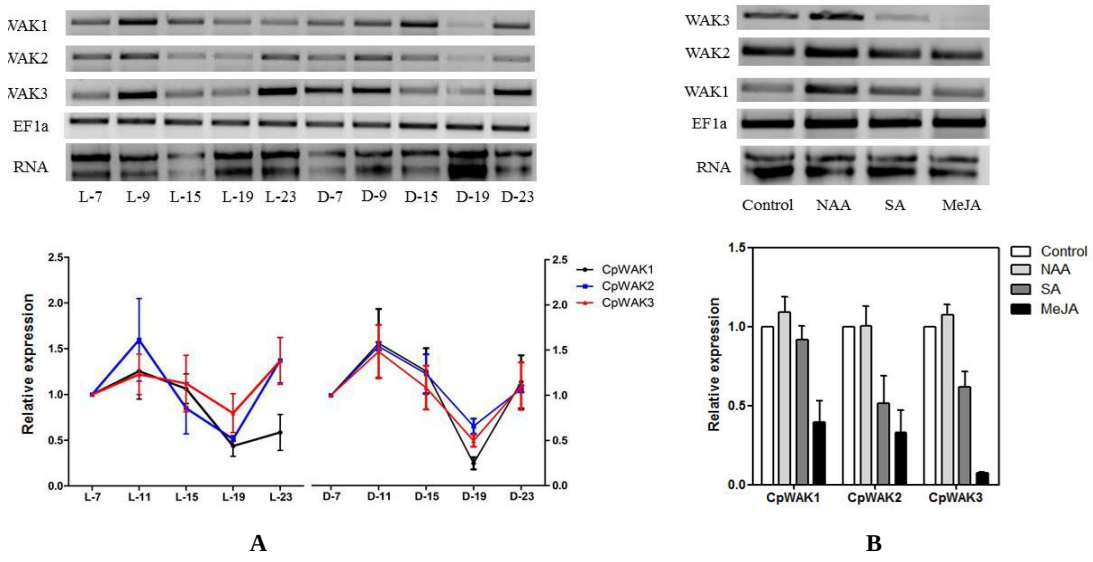


Figure 3.6 Transcript expression analyses of *CpWAK* genes. All the RNAs were isolated from *C.plantagineum* leaves and then reverse transcribed into cDNAs which were then amplified with specific primers: CpWAK1, CpWAK2, CpWAK3 and CpEF1α (reference gene). **A**, *C.plantagineum* plants were divided into two groups: Light (L) and Dark (D). The samples in group L grew normally and collected every four hours from 7:00 am. The samples in group D were put into darkness at 7:00 am and then collected as group L. **B**, *C.plantagineum* detached leaves were exposed to 100 μM NAA, 1 mM SA and 100 μM MeJA separately for 48h, while the leaves soaked in water served as control (CK). Both the line chart (**A**) and bar chart (**B**) represent the semi-quantitative RT-PCR analysis of CpWAKs using imageJ software. The relative expression levels of CpWAK genes in A were normalized to L-7 or D-7, while that of CpWAK genes in B were normalized to Control. n=6.

Giarola et al. (2016) demonstrated that the transcripts of CpWAK1 and CpWAK2 were down-regulated during desiccation. Because of the high similarity of CpWAK promoters and the different expression profiles between transcript and protein level, the protein expression of

CpWAKs under drought was investigated. The untreated, desiccated and rehydrated *C.plantagineum* leaves were collected for SDS-PAGE and Western-blot. As anti-CpWAK1 antibody (Sergeeva 2014) is polyclonal and cannot discriminate among the members of the CpWAK family, the protein expression analyses in this study reflect all CpWAK genes. CpWAK proteins are often degraded in total protein extracts even in the presence of protease inhibitors, so that the protein expression analyses were carried out with cell wall proteins. Cell wall proteins were extracted according to the protocol 3 of Printz et al (2015). The expression profiles were investigated in $CaCl_2$, EGTA and LiCl protein fractions. As shown in **Figure 3.7**, the protein expression of CpWAK genes is consistent with the transcript expression profile. The CpWAK proteins in all three fractions were down-regulated during dehydration. The CpWAKs were the most abundant in the LiCl fraction (**Figure 3.7**).

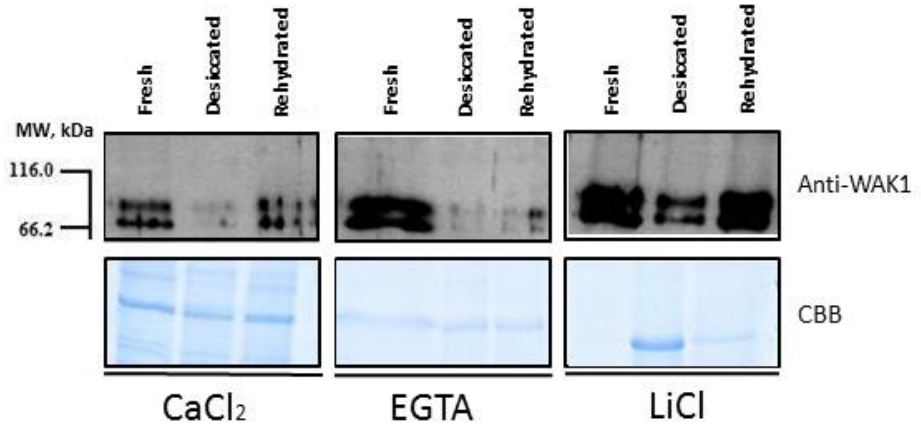
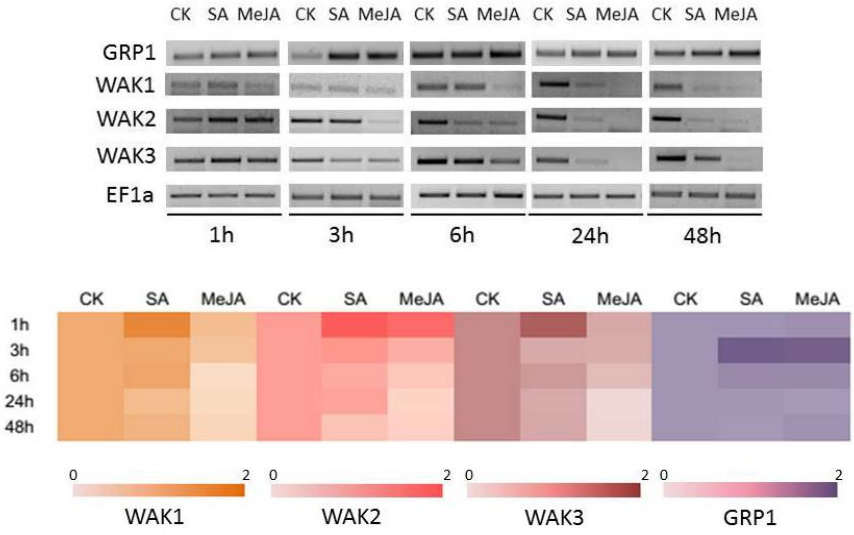


Figure 3.7 Protein expression analyses of CpWAK genes under drought. *C.plantagineum* leaf samples were collected from intact plants of *C.plantagineum* treated in three different ways: growing normally (Fresh), being dried in pots for 7 days (Desiccated), and being rehydrated in pots for 3 days (Rehydrated). All protein samples were analyzed using 12% SDS-PAGE, and subsequently transferred to a nitrocellulose membrane, and detected by polyclonal antibodies (anti-CpWAK1 antibody 1:5000). Protein expression analyses were investigated in cell wall proteins and total proteins with CBB (Coomassie brilliant blue-stained) gels as loading controls.

It was reported that WAK interacts with the glycine-rich protein (GRP) (Giarola et al. 2016; Park et al.2001). Both WAK and GRP are upregulated upon SA or the SA-analogue 2, 6-dichoroisonicotinic acid (INA) treatment (Park et al. 2001; Verica et al. 2003; Zheng-Hui He 1999) , which suggests that the two proteins are involved in defense-response pathways. Thus the transcript (**Figure 3.8 A**) and protein (**Figure 3.8 B**) expression of the CpWAK genes and CpGRP1 upon SA and MeJA treatment were analyzed, although the transcript expression analyses of CpWAK genes under SA and MeJA treatment for 48h were perfomed before (**Figure 3.6 B**) and the result was opposite to that observed in *A.thaliana* (Park et al. 2001; Verica et al. 2003; Zheng-Hui He. et al. 1999). The heatmap in **Figure 3.8 A** clearly showed the different transcript expression patterns between CpWAK genes and CpGRP1, with faster responses of CpWAKs after 1h of SA and MeJA (only CpWAK2 upregulated) treatment and a little slower of CpGRP1 after 3h of SA and MeJA treatment. All of transcripts were downregulated after reaching the peak, except for the CpWAK1 and CpWAK2 under MeJA treatment which were never upregulated and

decreased exceedingly after the treatment with MeJA. The upregulated protein expression of CpGRP1 was much weaker than that of the transcripts and detected after 6h-24h of SA and MeJA treatments which correspond to that of CpWAK genes in $CaCl_2$ and LiCl protein fractions (**Figure 3.8 B**).

A Transcript expression analysis of CpWAK1 and CpGRP1



B Protein expression analysis of CpWAK1 and CpGRP1

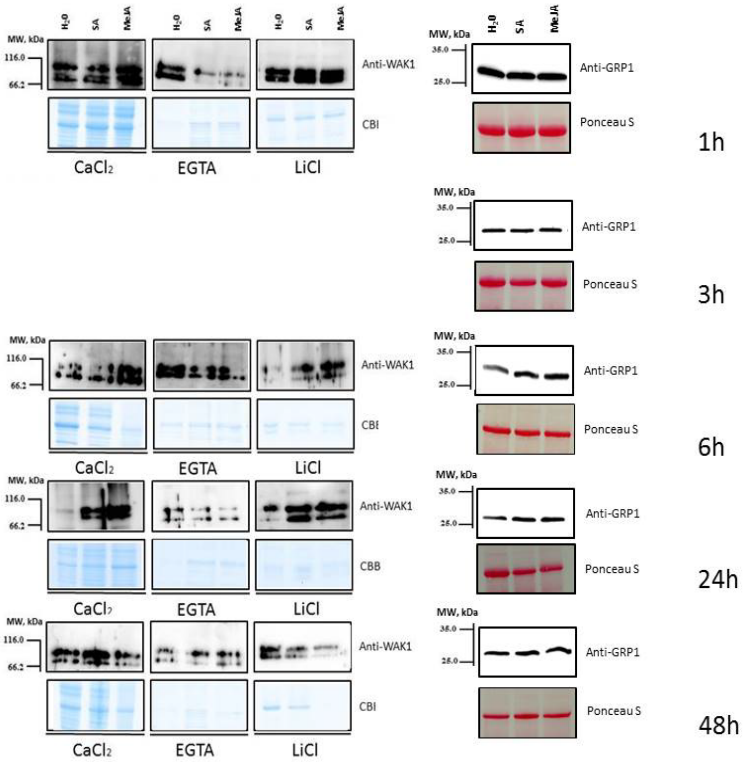


Figure 3.8 Effects of SA and MeJA treatments on transcript and protein expression of CpWAK genes and its interacting partner CpGRP1. *C.plantagineum* detached leaves were exposed to 1 mM SA and 100 µM MeJA separately for 1h, 3h, 6h, 24h and 48h, while the leaves soaked in water served as control (CK). Transcript (**A**) and

protein (**B**) expression patterns of CpWAK genes and CpGRP1 were identified using these leave samples. The heatmap in **A** showed the semi-quantitative RT-PCR analyses of CpWAK genes and CpGRP1 using imageJ software. The relative expression levels of CpWAKs and CpGRP1 in SA and MeJA were normalized to that in CK. n=9. All protein samples were analyzed by 12% SDS-PAGE, transferred to nitrocellulose membrane, and detected by polyclonal antibodies (anti-CpWAK1 antibody 1:5000, anti-CpGRP1 antibody 1:5000). Protein expression analyses of CpWAK genes and CpGRP1 were investigated in cell wall proteins and total proteins with CBB (Coomassie brilliant blue-stained) gels and Ponceau S stained membranes as loading controls respectively.

3.1.3 Production and purification of recombinant proteins

The extracellular domain of AtWAK1 was shown to bind to pectins *in vitro* by Decreux and Messiaen (2005). Therefore, to investigate the interaction among CpWAKs, its interaction partner CpGRP1 and pectin, the His-tagged recombinant proteins of CpWAK1-EX, CpWAK2-EX, CpWAK1-R-2, CpWAK1-R-3, CpWAK1-R-4 (abbreviated to CpWAK1, CpWAK2, R-2, R-3, R-4) and CpGRP1 were expressed (**Figure 3.9**). Both CpWAK1 and CpWAK2 are the extracellular domains of CpWAK proteins without signal peptides (amino acid 31-315 for CpWAK1 and 37-333 for CpWAK2) plus a six-His peptide, a recombinant protein (**Figure 3.9, Supplementary figure 2**). R-2, R-3 and R-4 are the truncated segments of CpWAK1 recombinant protein, containing the amino acids 31-160, 161-315 and 31-220, respectively (**Figure 3.9, Supplementary figure 2**). The putative molecular weight and pI of His-tagged recombinant proteins are CpWAK1 (32.2 kDa, 4.71), CpWAK2 (34.1 kDa, 5.31), R-2 (14.9 kDa, 7.74), R-3 (18.6 kDa, 4.31), R-4 (21.8 kDa, 6.22) and CpGRP1 (14.18 kDa, 8.35).

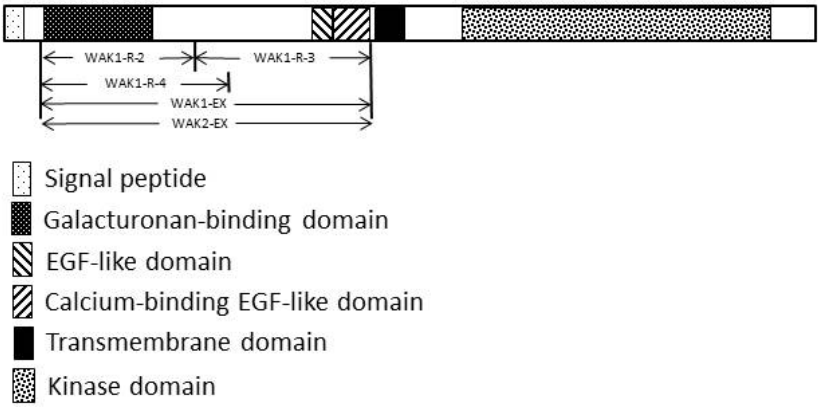


Figure 3.9 Domain structures of CpWAK proteins and the fragments for His-tagged recombinant proteins.

All the recombinant proteins were overexpressed in *E. coli* BL21 using the pET28a plasmid. The *E. coli* BL21 cells carrying the CpWAK1 and CpGRP1 overexpressing pET28a vetcors were provided by Anna Sergeeva and Dr. Valentino Giarola, respectively. Other pET28a vectors bearing the His-tagged fragments of CpWAK2, R-2, R-3 and R-4 were constructed and transformed into *E. coli* BL21 competent cells. The expression vector maps are shown in **Supplementary figure 4.** The induction of CpWAK2, R-2, R-3 and R-4 were performed in the presence of 1 mM IPTG for five hours, except for the overnight induction for R-4. **Figure 3.10** showed that the accumulations

of other recombinant proteins were significant after 5h of induction aside from R-4 and there is a considerable leaky expression in the induction of R-2 and R-4. The molecular weights of all the recombinant proteins conform to expectation (**Figure 3.10**). All CpWAK fusion proteins were accumulated in inclusion bodies. Only the CpGRP1 protein could be extracted from the soluble protein fraction.

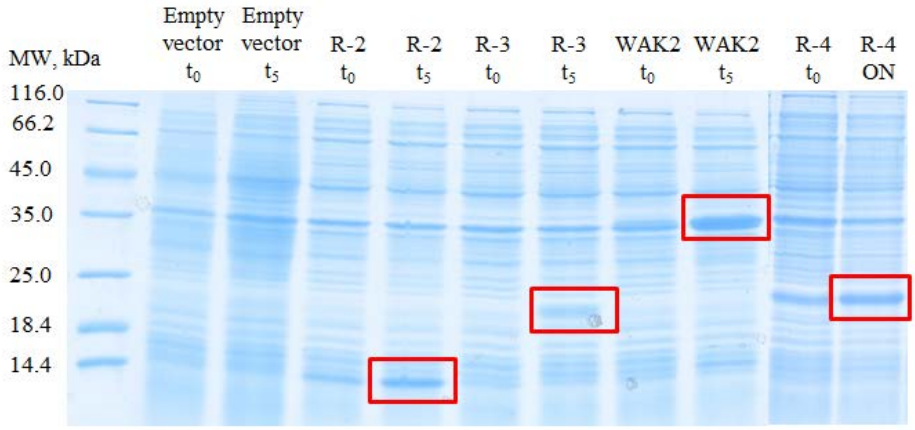


Figure 3.10 Induction of His-tagged recombinant proteins. The recombinant proteins CpWAK2, R-2, R-3 and R-4 were induced with 1 mM IPTG for 5 hours, except R-4 being induced overnight (ON). The bacterial pellets were collected and then loaded onto the SDS-PAGE gel after being homogenized with 1× Laemmli sample buffer. The induced proteins were encased in rectangle.

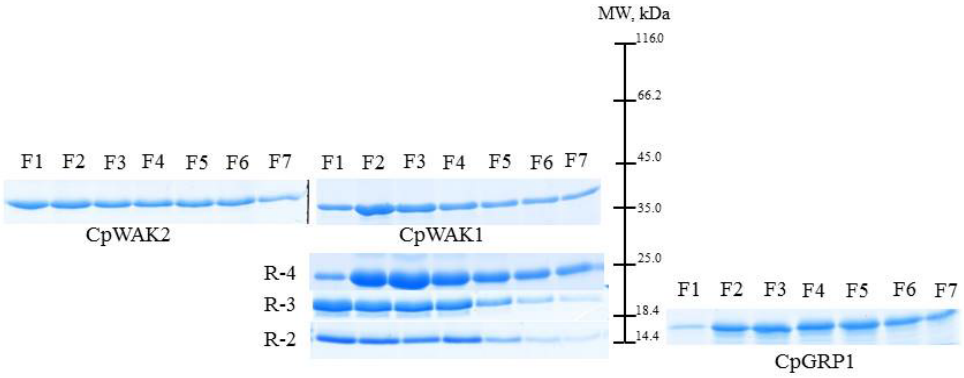


Figure 3.11 His–tag affinity chromatography of recombinant proteins. All the recombinant proteins were purified using a nickel affinity column. All CpWAK fusion proteins were obtained from inclusion bodies, and CpGRP1 was extracted from soluble protein fractions under native conditions. F1–F7: Eluted His-tag fractions. The His-tag CpWAK fusion proteins, CpWAK1 (32.2 kDa), CpWAK2 (34.1 kDa), R-2 (14.9 kDa), R-3 (18.6 kDa) and R-4 (21.8 kDa) are shown on the left of the protein ladder, and CpGRP1 (14.18 kDa) is shown on the right.

The purification of the fusion proteins was carried out as described in **section 2.2.4.2**. As depicted in **Figure 3.11,** the fractions F1-F6 of CpWAK1 and CpWAK2, F1-F5 of R-2 and R-3, F2-F7 of R-4 and CpGRP1 showed a strong protein band and thus were collected and pooled for desalination or renaturation and then lyophilization. The same amount of CpWAK fusion proteins were then separated by SDS-PAGE and finally blotted to a nitrocellulose membrane for protein immunodetection with anti-His-tag and anti-CpWAK1 antibodies. **Figure 3.12** indicates that all the

CpWAK fusion proteins can be detected by two antibodies and form multimers, especially R-2.

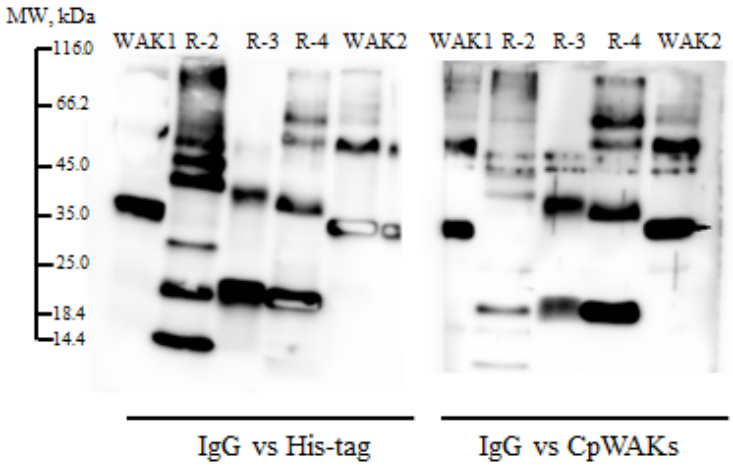


Figure 3.12 Western-blot analyses of CpWAK fusion proteins with anti-His-tag and anti-CpWAK1 antibodies. All CpWAK fusion proteins were detected by the polyclonal anti-His-tag antibody (1:1000) (Invitrogen) and the polyclonal anti-CpWAK1 antibody (1:5000).

3.1.4 CpWAK1 is precipitated with Ca^{2+} *in vitro*

The recombinant proteins were overexpressed in *E. coli* BL21 cells and were prepared for the protein-pectin binding assays. However, the role of Ca^{2+} in the precipitation of recombinant protein CpWAK1 was discovered by varying experimental parameters (**Figure 3.13**). The same amount of purified CpWAK1 and control protein BSA (bovine serum albumin) were incubated in Tris/NaCl buffer (20 mM Tris HCl/ 150 mM NaCl, pH 8.0) with/without Ca^{2+}, Mg^{2+} or EGTA for 1h at room temperature. After centrifugation, the proteins in the upper and bottom layers were collected separately and then subjected to SDS-PAGE gel and detected by silver staining. The lane 3 and 5 in **Figure 3.13** show that 2 mM Ca^{2+} is able to aggregate all CpWAK1 even with some Ca^{2+} chelator EGTA, however more EGTA can alleviate the precipitation effect of Ca^{2+} (Lane 6). On the other hand, Mg^{2+} has no effect on CpWAK1 aggregation as shown in lane 1 and 4 of **Figure 3.13**.

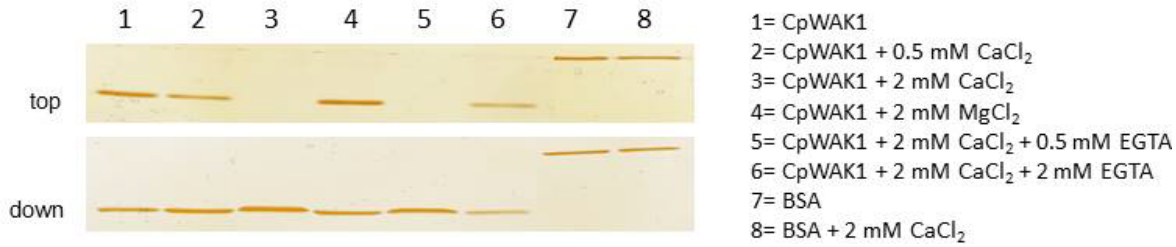


Figure 3.13 The recombinant protein CpWAK1 is precipitated with Ca^{2+} in Tris/NaCl buffer (pH 8.0). 50 ng of the purified recombinant protein CpWAK1 and BSA (as control) were incubated in 30 µl of Tris/NaCl buffer (pH 8.0) with/without Ca^{2+}, Mg^{2+} or EGTA for 1 hour at room temperature. After centrifugation for 15 min (13,000 rpm, RT), the top 10 µl and bottom 10 µl of the samples were pipetted out and proteins were visualized by silver staining.

3.1.5 Interaction of CpWAKs with pectin and the cell wall protein CpGRP1 *in vitro*

The protein-pectin binding assay was carried out as ELISA binding assay according to Decreux and Messiaen (2005) with some modifications. All the ELISA binding assays were executed in Tris/NaCl buffer (20 mM Tris/150 mM NaCl, Ca^{2+}/Mg^{2+}/EGTA:+/-) with the same amount of fusion proteins (amount to the molar number of 0.2 µg of purified recombinant CpWAK1). The CpWAK-pectin binding was confirmed with the pectin extracts from untreated *C.plantagineum* leaves, but not with polygalacturonic acid (PGA, Sigma) or commercial pectin (pectin from citrus peel, Sigma) **(Figure 3.14 A)**. The binding between CpWAK and commercial pectin were finally identified in Tris/NaCl/Ca^{2+} buffer after saponification of the commercial pectins **(Figure 3.14 D)**. In addition to CpWAK1, both CpGRP1 and CpWAK2 also show affinity for pectin extracts, with overwhelmingly higher affinity of CpGRP1 and about two-times more binding capacity of CpWAK2 (**Figure 3.14 B and C**). **Figure 3.14 C** also demonstrats that Ca^{2+} can facilitate CpWAKs binding to pectin extracts. Giarola et al. (2016) identified the interaction between CpWAK1 and CpGRP1 in both the yeast two-hybrid system and *planta*, so that CpWAK1 was

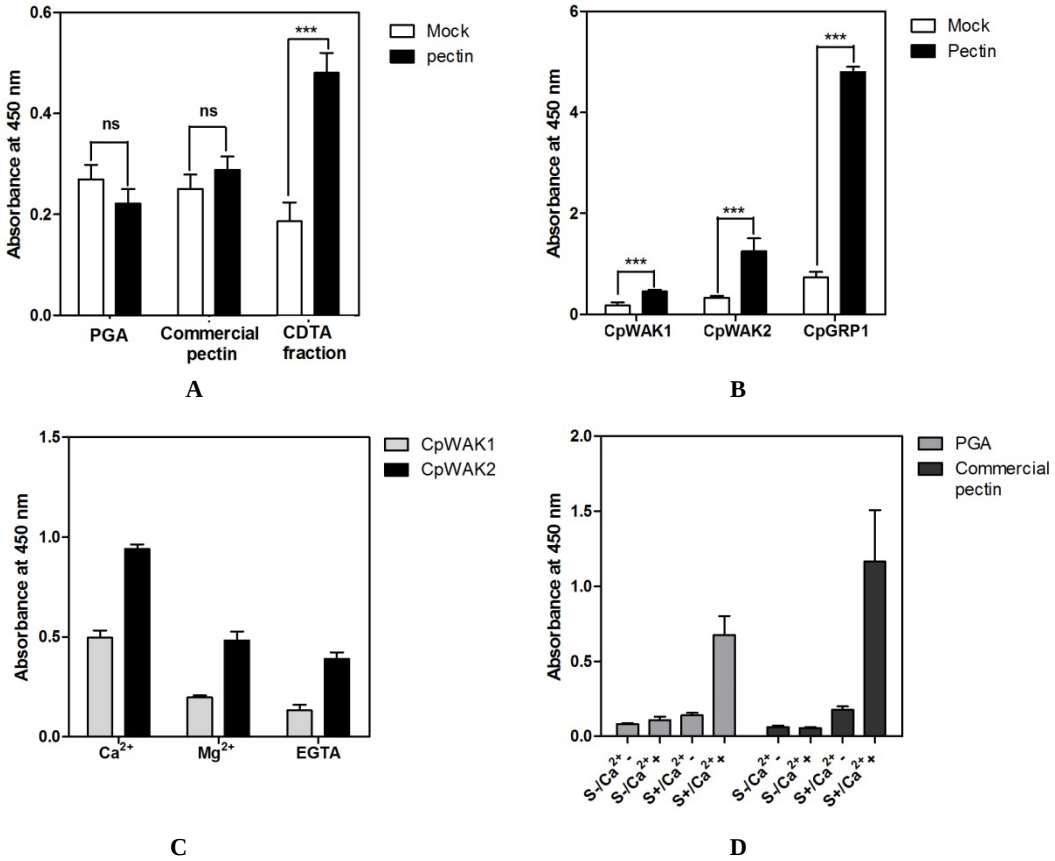


Figure 3.14 CpWAKs bind to egg-box conformation of pectin. Quantification of ELISA tests are shown in A, B, C, D. Polygalacturonic acid (PGA, Sigma), Commercial pectin (pectin from citrus peel, Sigma) and *C.plantagineum* pectin isolated from *C.plantagineum* leaves with CDTA (1, 2-cyclohexanediaminetetraacetic acid) solution were normalized by pectin estimation assay, and immobilized in ELISA plate wells, incubated with 0.2 µg of purified recombinant CpWAK1 or the same amount of other recombinant proteins. The bound recombinant proteins were

detected with His-tag antibody (1:10,000), n=9. Error bars indicate SEM and Mock indicates only buffer without pectin. (ns means no significant, *** $P < 0.001$, t-test compared to Mock) (**A)** The CpWAK1 extracellular domain only binds to the pectin extracts of *C.plantagineum* leaves. (**B**) The interaction of the pectin extracts of *C.plantagineum* leaves with CpGRP1 is much stronger than with CpWAKs, and CpWAK2 shows higher affinity for pectin extracts. (**C**) Ca^{2+} facilitates the binding between pectin extracts with the CpWAKs extracellular domain. The binding assay was conducted in Tris/NaCl buffer with Ca^{2+}, Mg^{2+} or EGTA. (**D**) Both saponification and Ca^{2+} are necessary for binding of the CpWAK1 extracellular domain to polygalacturonic acid and commercial pectin. S+/-: with or without saponification, Ca^{2+} +/-: with or without Ca^{2+}.

chosen to further investigate CpWAK1-CpGRP1-pectin interaction. In order to gain a better insight into CpWAK-pectin binding, the extracellular domain of CpWAK1 was subdivided into three segments, R-2 including the galacturonan-binding domain, R-3 containing EGF-like domains and R-4, a little longer than the R-2 segment but also without the EGF-like domains (**Figure 3.9**). Among the three segments, R-2 showed the strongest preference to pectin extracts. R-4, in spite of being inferior to R-2, also had higher affinity to pectin extracts than R-3 which had even weaker binding capacity to pectin extracts than CpWAK1 (**Figure 3.15)**. The function of CpGRP1 on the CpWAK1-pectin binding was studied using competitive ELISA assays, where the CpWAK1 and CpGRP1 was premixed before incubating with the pectin-immoblized ELISA wells. The heatmaps in **Figure 3.15** manifested that full-sized CpGRP1 contributed to more CpWAK1, R-2 and R-3 binding to pectin extracts, while the CpWAK1 segments decreased the number of CpGRP1 binding to pectin extracts, especially R-3. The His-tagged N-terminal of CpGRP (provided by Niklas Jung), the fragment without the putative pectin-binding domain, competed with the pectin extracts on the ELISA plate wells for interacting with CpWAK1 and thus led to less CpWAK1 bound on the ELISA plates (**Figure 3.15 B**).

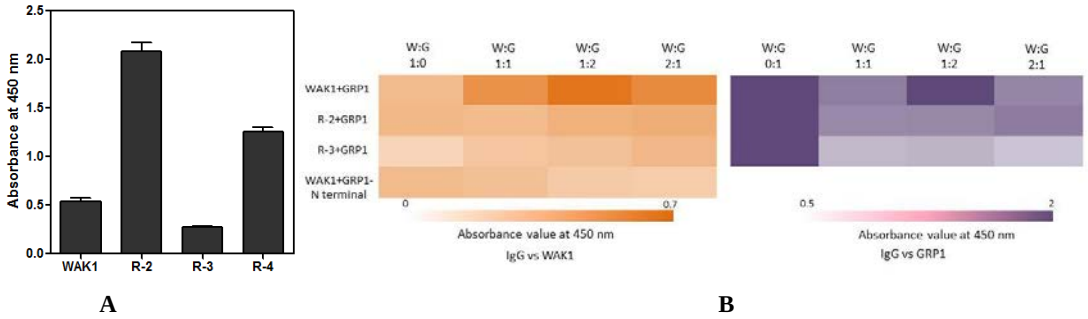


Figure 3.15 CpGRP1-CpWAK1-pectin interaction *in vitro*. (**A**) Different subdomains of CpWAK1 showed different pectin binding capacity. WAK1, WAK2, R-2, R-3 and R-4 represent different fragments as shown in **Figure 3.7**, and the recombinant proteins were detected with His-tag antibody (1:10,000). (**B**) Heatmaps reveal that CpGRP1 contributes to the binding activity of CpWAK1. Each heatmap depicts the ELISA absorbance values with respectively anti-CpWAK1 (orange, 1:2500) and anti-CpGRP1 (purple, 1:5000) antibody. Different protein combinations are visualized longitudinally, and the ratios of the amounts of CpWAK1 (W) and CpGRP1 (G) are shown horizontally. The proteins were pre-mixed for 1h before incubating with the pectin-immobilized ELISA plates.

The apoplastic pH is normally around five and oscillates between 4 to 7, which depends on the

states of the cells growing under stresses (Geilfus 2017). To investigate the effect of the apoplastic pH on the interaction among CpWAK1, CpGRP1 and pectin, the ELISA binding assays were performed at different pH values, from 4 to 8. As shown in **Figure 3.16,** CpWAK1 binding to pectins was not affected at different pH values, but CpGRP1 showed stronger binding capacity at pH 4, 6 and 7. And all the CpGRP1-involved bindings showed the same trend as CpGRP1. The pH also influences on the effect of Ca^{2+} on CpWAK1-pectin interaction. **Figure 3.17** A showed that Ca^{2+} exerted the impact on CpWAK1-pectin binding only when the pH value increased to 8.0, which is either instrumental or unfavourable for the binding, depending on the source of pectin extracts. The binding was strengthened when the pectin was extracted from untreated leaves, while weakened when pectin was prepared from detached water-soaked leaves. In addition, CpWAK1 did not show significantly different affinity for the pectin extracted from untreated, partially dried, desiccated and rehydrated *C.plantagineum* leaves although **Figure 3.17** B exhibited slightly weaker binding to pectin extracts from dried *C.plantagineum* leaves. Interestingly, the SA and MeJA treatments affected the CpWAK1-pectin binding, with less CpWAK1 being immobilized by pectin extracts from detached leaves subjected to SA and MeJA treatments (**Figure 3.17 C**).

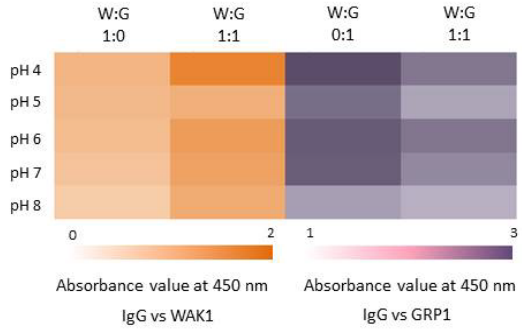


Figure 3.16 WAK1-GRP1 complex binding affinity for pectin extracts is affected by pH values. Different pH values are visualized longitudinally, and the ratios of the amounts of CpWAK1 (W) and CpGRP1 (G) are shown horizontally. The proteins were pre-mixed for 1h before incubation with the pectin in ELISA plates.

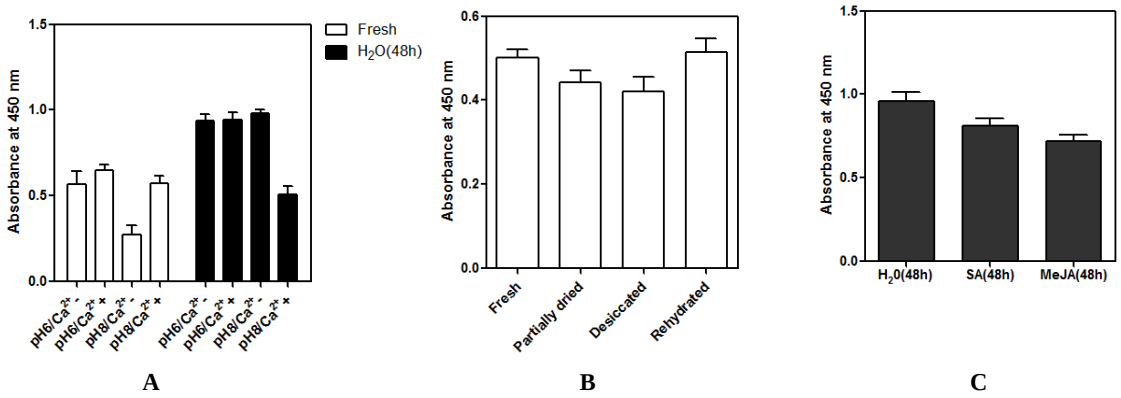


Figure 3.17 Analysis of CpWAK1-pectin binding**.** The pectin extracts for the binding assays were prepared from fresh, partially dried, desiccated and rehydrated *C.plantagineum* leaves or detached *C.plantagineum* leaves soaked in water, SA or MeJA for 48h. **(A)** The effect of Ca^{2+} on the CpWAK1-pectins binding is related to pH values and

the source of pectin extracts. **(B)** CpWAK1 showed similar binding capacity to the pectin extracts prepared from fresh, partially dried, desiccated and rehydrated *C.plantagineum* leaves. The experiment was carried out in Tris/NaCl buffer (pH 8.0 with Ca^{2+}). Statistical analysis was performed using one-way ANOVA (P = 0.097, not significant). **(C)** CpWAK1 showed different binding capacity to the pectin extracts prepared from detached *C.plantagineum* leaves exposed to water, SA or MeJA for 48h. Statistical analysis was performed using one-way ANOVA (P =0.0039, significant) .n=9. Error bars indicate SEM.

3.2 Interaction between the *C.plantagineum* germin-like protein 1 (CpGLP1) and the *C.plantagineum* cysteine-rich rehydration responsive protein 1 (CpCRP1)

C.plantagineum germin-like protein 1 (CpGLP1) was identified as the putative interaction partner of *C.plantagineum* cysteine-rich rehydration responsive protein 1 (CpCRP1) in yeast-two-hybrid (Y2H) assay with CpCRP1 as bait (Dulitz 2016). CpCRP1 is one of the taxonomically restricted genes discovered from the transcriptome analysis (Rodriguez et al. 2010), which is involved in the dehydration/rehydration cycle (Giarola et al. 2015). Germins and GLPs as glycoproteins associated with cell wall are ubiquitous in all plants and somhow function with their oxalate oxidase/superoxide dismutase activity during dehydration (Bernier and Berna 2001; Dunwell et al. 2008). As both CpCRP1 and GLPs are localized extracellular space and associated with the plant responses to drought stress, their interaction may provide new hints to understand the cell wall responses during dehydration. To further study the interaction between CpGLP1 and CpCRP1, the protein expression patterns of CpGLP1 during dehydration and rehydration were investigated and the pull-down assay and bimolecular fluorescence complementation (BiFC) assay were performed. Both the nucleic acids sequences and protein sequences of CpGLP1 and CpCRP1 are shown in **Supplementary figure 1** and **2**. The basic characteristics of CpGLP1 and CpCRP1 were deduced using Sequence Manipulation Suite (http://www.bioinformatics.org/sms2/index.html) and are presented in **Table 3.4**.

Table 3.4 Basic characteristics of CpGLP1 and CpCRP1

	Amino acids	Molecular weight (kDa)	Isoelectric point (pI)	Grand average of hydropathy
CpGLP1	226	24.03	9.54	0.155
CpCRP1	165	17.76	8.33	0.15

3.2.1 Analysis of CpGLP1 protein expression

Protein-protein interaction requires the simultaneous presence of the interaction partners, which means the similar spatial and temporal expression patterns of the interaction partners. CpCRP1 was reported to be secreted to the apoplast, and downregulated in both desiccated and rehydrated leaves, albeit the CpCRP1 transcript was accumulated considerably in rehydrated samples (Giarola et al. 2015). The CpGLP1 transcript showed the opposite trend to CpCRP1, being up-regulated in desiccated leaves and reduced in rehydrated levels (Dulitz 2016). However, the protein expression

is more important concerning to protein-protein interaction. Therefore, the analysis of CpGLP1 protein expression was performed using different samples.

To test the localization of CpGLP1, the cell wall tissues, $CaCl_2$ and LiCl extracts of cell wall proteins were prepared for Coomassie blue-stained SDS-PAGE and western blot (**Figure 3.18**). As we can see from **Figure 3.18**, CpGLP1 was detected in all samples, especially in the $CaCl_2$ and LiCl extracts of cell wall proteins, which indicates CpGLP1 localizes in the cell wall and thus lays the spatial foundation for CpCRP1-CpGLP1 interaction. The protein expression pattern of CpGLP1 in the dehydration/rehydration cycle was also analyzed (**Figure 3.19**). The accumulation of CpGLP1 protein in dehydrated and rehydrated leaves was observed in both total proteins and $CaCl_2$ and LiCl extracts of cell wall proteins (**Figure 3.19 B and C**). Although the protein expression patterns of CpCRP1 and CpGLP1 are different, both proteins occur in fresh and rehydrated leaves simultaneously (**Figure 3.19, C**). Therefore, the CpGLP1-CpCRP1 interaction is possible according to the protein expression.

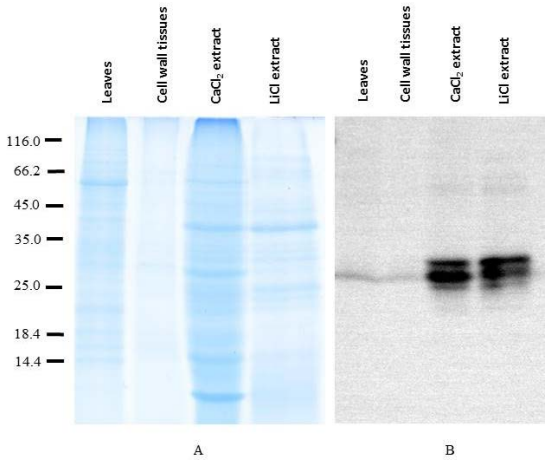


Figure 3.18 Protein extractions from desiccated *C.plantagineum* leaves. The proetein extraction were analyzed by Coomassie blue-stained SDS-PAGE (A) and western blot (B) with anti-CpGLP1 antibody (1: 5000). The proteins from desiccated leaves and cell wall tissues were dissolved directly with lx Laemmli buffer. The cell wall tissue preparations were done according to Feiz et al. (2006), while the preparation of $CaCl_2$ extracts and LiCl extracts were carried out according to Printz et al. (2015).

3.2.2 Analysis of CpGLP1-CpCRP1 interaction using pull-down assays

The interaction between CpGLP1 and CpCRP1 was analyzed *in vitro* using pull-down assays. The CpGLP1 fragment without signal peptide and the binding domain of CpCRP1 were subcloned into expressing vectors pET28a and pGEX-4T2, respectively. The *E.coli* BL21 (DE3) cells overexpressing His-CpGPL1 and GST-N-terminal CpCRP1 (abbreviated as GST-CpGRP1) were provided by Stefano Manduzio. The induction and purification of two fusion proteins were carried out as **section 2.2.4.1-2.2.4.3** (**Supplementary figure 5**). The pull-down assay using the bait and prey proteins extracted under native conditions is shown in **Figure 3.20**, which is visualized by Coomassie Blue staining and Western blot with anti-CpCRP1 antibody. However, the antibody can also detect the bait protein His-CpGLP1 (**Figure 3.20**). As only one band (His-CpGLP1) on the

Elution S lane was detected by both the stained gel and Western blot (**Figure 3.20**), the interaction between CpGLP1 and CpCRP1 cannot be confirmed. In addition, the bait protein, His-CpGLP1, extracted from inclusion bodies was also used for pull-down assays, which were refolded on the column (**section 2.2.4.4**). This pull-down assays were only analyzed by silver staining (**Supplementary figure 6**), which exhibited the same result as **Figure 3.20**.

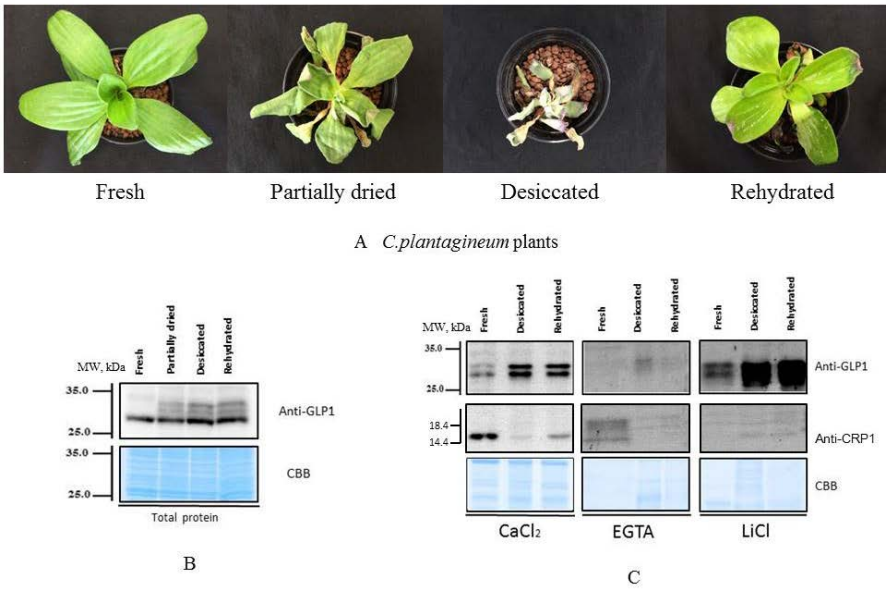


Figure 3.19 Protein expression analysis of CpGLP1 and CpCRP1. *C.plantagineum* leaf samples were collected from intact plants of *C.plantagineum* treated in four different ways: growing normally (Fresh), being dried in pots for 3-4 days (Partially dried), dried in pots for 7 days (Desiccated), and rehydrated in pots for 3 days (Rehydrated) (**A**). All protein samples were analyzed by 12% SDS-PAGE, transferred to nitrocellulose membranes, and detected by polyclonal antibodies (anti-CpGLP1 antibody 1:5000, anti-CpCRP1 antibody 1:5000). Protein expression analyses were investigated in total proteins (**B**), and cell wall proteins (**C**) prepared according to Printz et al. (2015) with CBB (Coomassie brilliant blue-stained) gels as loading controls.

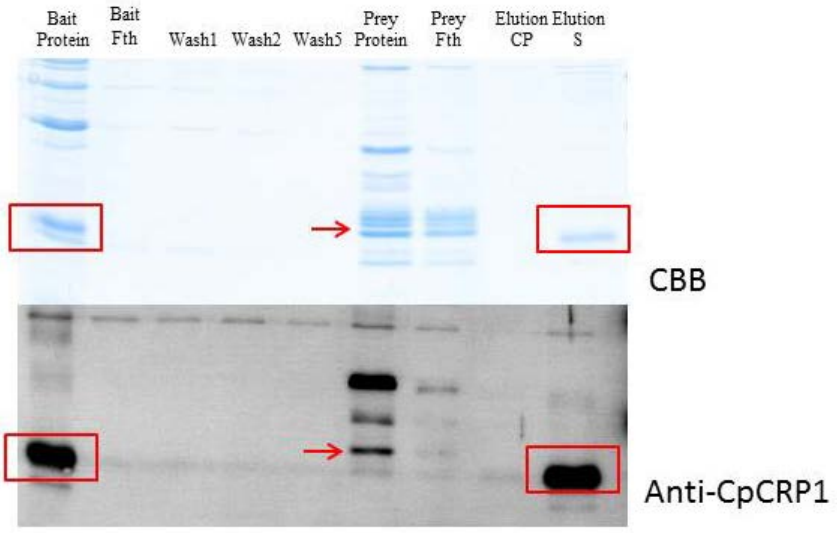
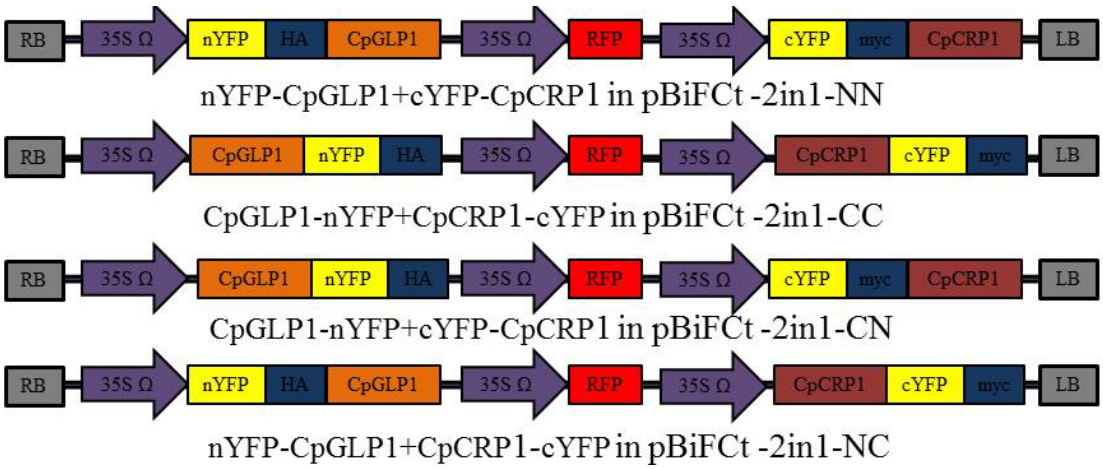


Figure 3.20 Analysis of interactions between CpGLP1 and CpCRP1 by pull-down assay. The pull-down assay was performed with fusion proteins purified under native conditions and checked with CBB (Coomassie brilliant blue-stained) gels and Western blot. The anti-CpCRP1 antibody (1:10000) was used for detecting CpCRP1. The proteins loaded were Bait protein (His-CpGLP1), Bait Fth (Bait protein flow-through sample), Wash1 (first wash sample), Wash2 (second wash sample), Wash5 (final wash sample), Prey protein (GST-CpGRP1), Prey Fth (Prey protein flow-through sample), Elution CP (Elution sample from Control-Prey column), Elution S (Elution sample from Sample column). The bait protein (His-CpGLP1) and prey protein (GST-CpCRP1) are indicated by red rectangle and arrow respectively.

3.2.3 Analysis of CpGLP1-CpCRP1 interaction using bimolecular fluorescence complementation (BiFC) *in planta*

The interaction between CpGLP1 and CpCRP1 was also analyzed in planta by bimolecular fluorescence complementation (BiFC). The pBiFCt-2in1 vector series contain four types (**Figure 3.21, A**) depending on the tagging orientations. The CpGLP1 and CpCRP1 were integrated into the pBiFCt-2in1 vector series (**Figure 3.21, A**) through recombination into a shuttle plasmid (CpGLP1 in pDONR221-P3P2 and CpCRP1 in pDONR221-P1P4, **Supplementary figure 7**). The fluorescence signals were generated in all *Nicotiana benthamiana* cells expressing nYFP-CpGLP1+cYFP-CpGRP1,CpGLP1-nYFP+CpGRP1-cYFP, CpGLP1-nYFP+cYFP-CpGRP1, nYFP-CpGLP1+CpGRP1-cYFP translational fusions and the corresponding fusions without target protein fragments, and the signals were even stronger in the *N.benthamiana* leaves infected with *Agrobacterium tumefaciens* carrying the pBiFCt-2in1-NN and pBiFCt-2in1-NC empty vectors (**Figure 3.21, B**). In addition, the BiFC assays using CpGLP1 and CpCRP1 segment (amino acids 25-138) or using different split BiFC vectors (pBatTL-BsYFPn, pBatTL-BsYFPc) were performed. However, the fluorescence intensity in the *N.benthamiana* cells expressing CpGLP1+CpCRP1 fusions and the cells expressing CpGLP1+CpCRP1-segment fusions showed no significant difference (data not shown). Very weak fluorescence signals were observed in both *N.benthamiana* cells expressing only the N- and C-terminal YFP fragments and the cells expressing nYFP-CpGLP1+cYFP-CpGRP1 in the split BiFC system (data not shown). Therefore, the interaction between CpGLP1 and CpCRP1 was not confirmed by the BiFC assays either.

A



B

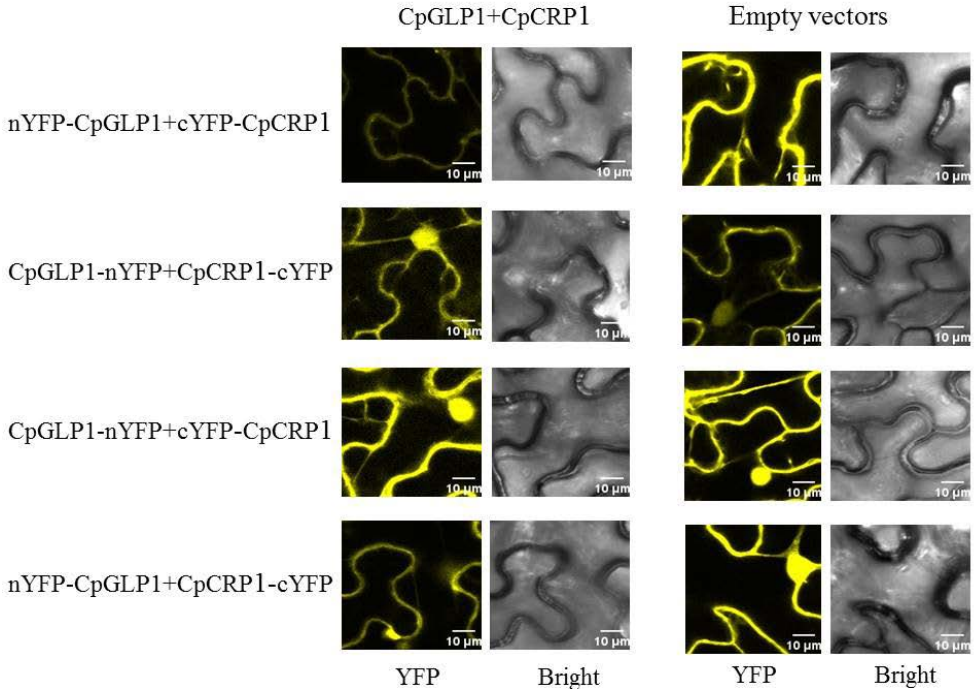


Figure 3.21 Analysis of the interaction between CpGLP1 and CpCRP1 in planta. (**A**) Schematic vector maps depicting the CpGLP1 and CpCRP1 in the pBiFC-2in1 vector series (modified from Grefen and Blatt, 2012). (**B**) The N-terminal yellow fluorescence protein (YFP)-CpGLP1 and C-terminal YFP-CpCRP1 translational fusions and their corresponding translational fusions were all expressed in tobacco leaves infiltrated with *Agrobacterium* carrying pBiFC-2in1 vectors. YFP fluorescence was analyzed 2 d after infiltration with a confocal laser scanner microscope. Bars, 10 μm.

3.3 Study on long non-coding RNA 28852

Many putative novel genes involved in drought stress tolerance were identified according to the

C.plantagineum transcriptome databank (Rodriguez et al. 2010). One contig numbered 28852 was highly induced under desiccation (Giarola et al. 2015). The full-length 28852 transcript is 631 bp (**Supplementary figure 1**). There is no similar sequence identified with the BLAST algorithm using the GenBank database (Giarola et al. 2015) and no long coding sequence identified *in silico* analysis (Dr. Giarola, IMBIO, Bonn, Germany) which indicates that 28852 is a not only a taxonomically restricted gene but also a long non-coding RNA (lncRNA).

3.3.1 Expression analyses of lncRNA 28852 in *C.plantagineum*

As the lncRNA does not encode proteins, only the transcript expression profiles of lncRNA 28852 were investigated. LncRNA 28852 were mainly expressed in the desiccated plant tissues including both roots and leaves, and its expression level was low (**Figure 3.22, A**). The expressions of lncRNA 28852 in total RNA and polysomal fractions were also compared (**Figure 3.22, B**). Although no protein coding sequence was predicted in the lncRNA 28852, the polysomes were shown to be associated with lncRNA 28852. The lncRNA 28852 in total RNA was induced in partially dehydrated leaves, accumulated in desiccated leaves and downregulated in rehydrated leaves, which was opposite to the trend in polysomal RNA. More lncRNA 28852 in the polysomal RNA was accumulated in partially dehydrated and rehydrated samples while less was in the desiccated leaves (**Figure 3.22, B**), which indicates that the lncRNA 28852 is preferentially associated with polysomes in the early stage of dehydration and rehydration.

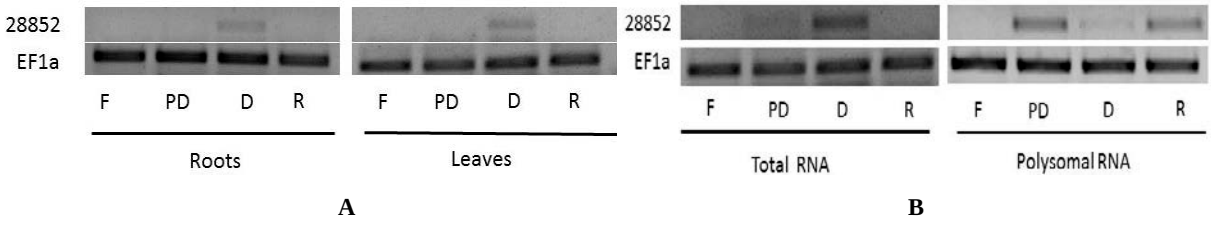


Figure 3.22 Transcript expression profiles of lncRNA 28852 in *C.plantagineum*. *C.plantagineum* tissues were collected from *C.plantagineum* plants treated in four different ways: growing normally (F), dried in pots for 3-4 days (PD), dried in pots for 7 days (D), and rehydrated in pots for 3 days (R). (**A)** LncRNA 28852 expression in leaves and roots from *C.plantagineum* exposed to dehydration and rehydration. (**B)** LncRNA 28852 expression in total and polysomal RNA isolated from *C.plantagineum* leaves exposed to dehydration and rehydration. All cDNAs were amplified using primers specific for lncRNA 28852 and elongation factor 1α (*EF 1α*) (as control).

3.3.2 Screening transgenic *Arabidopsis thaliana* lines overexpressing the LncRNA 28852

To investigate the function of the lncRNA 28852 during desiccation or rehydration, the transgenic *Arabidopsis thaliana* lines overexpressing the lncRNA 28852 were established (Dr. Giarola, IMBIO, Bonn, Germany) and the seeds were kindly provided to me. The transgenic lines were used for further experiments.

As depicted in **Figure 3.23. A,** the T-DNA vector used for the establishment of transgenic *A. thaliana* lines overexpressing the lncRNA 28852 contains the full-length lncRNA 28852 and CaMV 35S promoter. The T-DNA specific primer pair, p35S-pROK2 and 28852F2 and gene-specific primer pair, 28852CDS F and R were used for genotyping and transcript expression analysis (**Figure 3.23 B, C**). Both the T-DNA specific primer combination and the gene-specific primer combination amplified the target fragments using the genomic DNA extracted from all 30 transformants (**Figure 3.23 B**). After the selection of the first 10 lines of T_2 generation on the basis of kanamycin resistance, nine T_2 transgenic lines were regenerated (line 1, 2, 3, 4, 5, 7, 8, 9, 10) (**Figure 3.24**). The line 3, 5, and 9 accumulated the transcript at a higher level, although the lncRNA 28852 transcript was detectable in the leaves of all the 9 lines. Among the regenerated nine lines, only line 1, 3 and 5 grew normally on the 1/2 MS agar plates containing kanamycin and showed high morphological homogeneity after being transferred to soil (**Figure 3.24**). Therefore the transgenic line 3 and 5 were selected for the RNA-seq analysis.

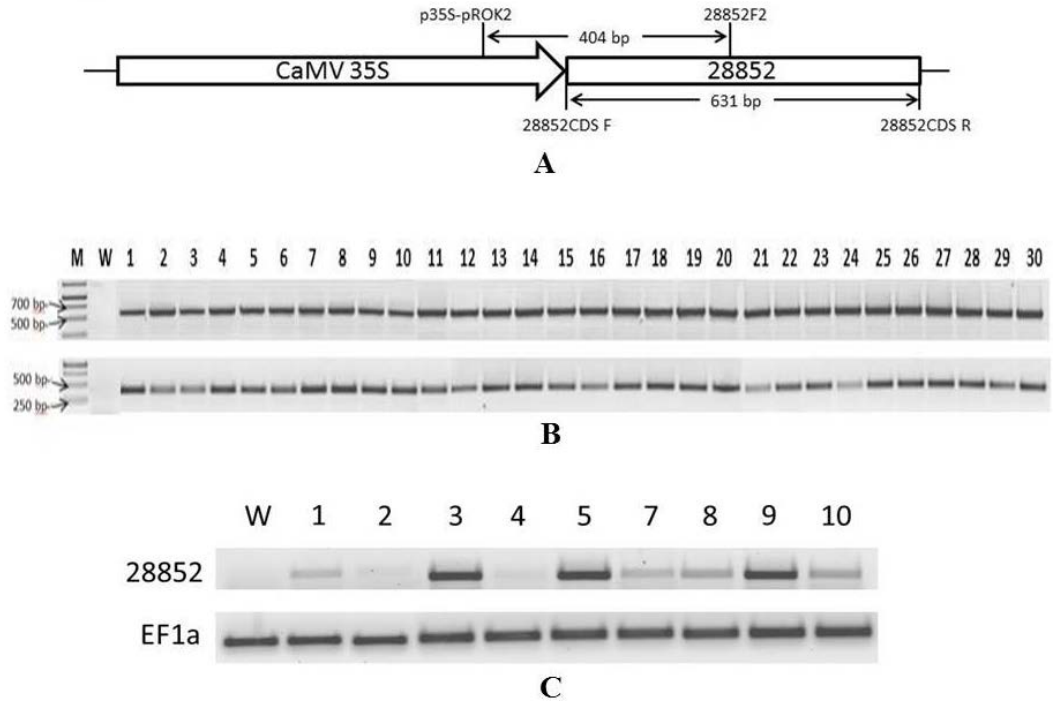
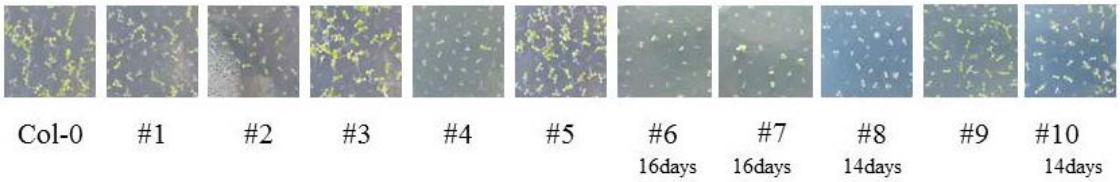


Figure 3.23 Genotyping and transcript expression analysis of transformants. (**A)** The schematic diagram of the T-DNA vector with 631-bp lncRNA 28852 gene and CaMV 35S as promoter. The primers used for genotyping and transcript expression analysis are indicated in the diagram. Primers: 28852CDS F and 28852CDS R. and p35S-pROK2 and 28852F2. (**B)** Genotyping of the transformants. 1-30 represents different mutant lines transformed with 35S-28852 T-DNA. The first and second lines are amplified with a gene-specific primer pair and a T-DNA specific primer pair, respectively. M, Marker. W: wild type. (**C)** Transcript expression of lncRNA 28852 under the CaMV 35S promoter in transformants. W: wild type. 1,-10 represent different lines of transgenic Arabidopsis. *EF1α* was used as a reference gene.

A

Ten days:



B

Four weeks:

Seven weeks:

Figure 3.24 Phenotype comparisons of wild type and transformants. Col-0: wild type *A. thaliana*, #1 - #10: transgenic *A. thaliana* lines. (**A**) Ten-day-old seedlings of wild type and mutants. (**B**) Four-week-old and seven-week-old plants of wild type and mutants.

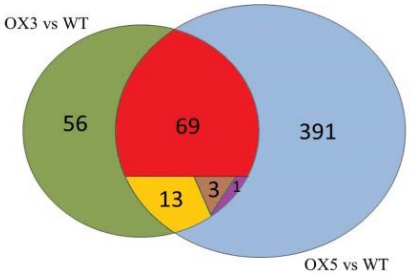
3.3.3 Comparative RNA-seq analysis of wild type and transgenic *A. thaliana* line 3 and 5

Considering the effect of lncRNA 28852 on the transcriptomes, the comparative RNA-seq analysis of wild type (WT) and T_2 transgenic *A.thaliana* overexpressing lncRNA 28852 line 3 and 5 (OX3 and OX5) were carried out. The RNA samples of WT, OX3 and OX5 were isolated from leaves with an RNA isolation kit (Macherey -Nagel, Germany). After DNase I treatment, all samples were purified with clean-up columns (Macherey -Nagel, Germany). The quality and quantity of RNA samples were checked by gel electrophoresis (**Supplymentary figure 8 A and B**) and Nanodrop with following results: WT (170 ng/µl, $OD_{260/280}$: 2.14, $OD_{260/230}$: 2.29), OX3: (165 ng/µl, $OD_{260/280}$: 2.11, $OD_{260/230}$: 1.93) and OX5: (159 ng/µl, $OD_{260/280}$: 2.12, $OD_{260/230}$: 2.02). To confirm that the 28852 gene expressed in the transgenic lines, the cDNAs prepared from the RNA samples for RNA-seq analysis were amplified with lncRNA 28852 primers (**Supplymentary figure 8 C**). The total RNA samples were then used for the single-end RNA-seq (GATC Biotech, Germany).

About fifty million raw reads of each sample were filtered and mapped to the reference genome, and the average mapping rates were around 97% (**Supplementary table 1**). High correlations were observed between the two samples (WT/OX3, WT/OX5 and OX3/OX5, **Supplementary figure 9**), which indicates the similar expression profiles among the three samples and high quality of RNA-seq data. The differentially expressed genes in wild type and mutants were manifested in the volcano plots (**Supplementary figure 10**), in which the significantly differentially expressed genes are indicated as red dots ($p < 0.001$). More genes were up-regulated among the differentially expressed genes in mutants, especially in OX3, and the absolute value of the $\log_2$ (fold change) among the significantly differentially expressed genes was less than 6 and more than 1 (**Supplementary figure 10**). There are 142 genes and 477 genes differentially expressed in OX3 and OX5 compared to WT according to the $q < 0.05$ and |Log2 (fold change)| > 1, among which 86 genes appeared in both mutants (**Figure 25, A**). The descriptions of 86 genes are shown in **Supplementary table 2**. These genes are involved in defense responses, transcription, plant cell wall organization and so on. Of the 86 transcripts, 69 were up-regulated and 1 was activated, while 13 were down-regulated and 3 were silent (**Figure 25, A**). The 86 transcripts were then clustered and visualized by a heatmap based on the FPKM (fragments per kilobase of transcript per million mapped reads) values (**Figure 25, B**). The red bands represent the up-regulated genes, and green ones indicated the down-regulated genes. Four clusters were identified in the 86 differentially expressed genes. Group I contained all the down-regulated genes. Among the up-regulated groups only the genes in group IV are up-regulated to a similar extent in the two mutants. While the genes in group II were up-regulated strongly in OX5 and the genes in group III were expressed at higher level in OX3. To explore the potential functions of the differentially expressed genes and thus figure out the effect of lncRNA 28852, the Gene Ontology (GO)-based annotation (**Supplementary table 2**) and enrichment (**Figure 3.26**) analyses were performed using the DAVID bioinformatics resources 6.8 (https://david.ncifcrf.gov/home.jsp) (Huang et al. 2008; 2009). 38 GO terms were significantly enriched ($p < 0.05$), among which only two GO terms not only contained more than 10% of the differentially expressed genes but the fold enrichment > 10 as well (**Figure 3.26**). The two GO terms were response to jasmonic acid (GO:0009753) and response

to wounding (GO:0009611) (**Figure 3.26**), The expression profiles of all the genes in the two GO terms are indicated in bold type in **Supplementary table 2**, one of which termed TT4 implicated in wounding response and flavonoid biosynthesis (**Figure 3.26, Supplementary table 2 and Figure 3.27 B**) were down-regulated while the others up-regulated. Besides, the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways analysis was also conducted and enriched in two pathways (p < 0.001), α-linolenic acid metabolism (ath00592) and flavonoid biosynthesis (ath00941) (**Figure 3.28**). The genes enriched in the two pathways are indicated by red font and red rectangles in **Figure 3.28** and represented by boxes in **Supplementary table 2**. All the genes in the α-linolenic acid metabolism pathway are up-regulated in the mutants, while those in the flavonoid biosynthesis pathway are down-regulated in the mutants.

A



B

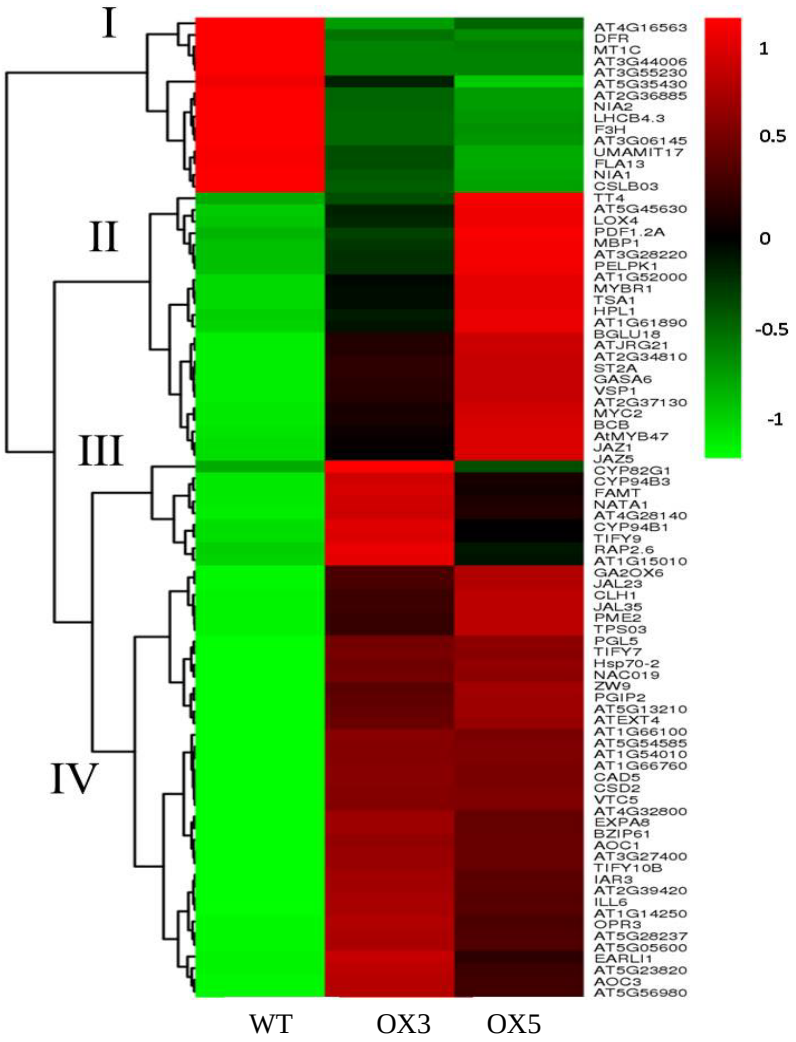


Figure 3.25 Modified Venn diagram and heatmap visulization of the genes differentially expressed in wild-type (WT) and T_2 generations of overexpression line 3 (OX3) and overexpression line 5 (OX5). (**A**) The modified Venn diagram indicates the number of differentially expressed genes ($q < 0.05$, $|\text{Log}_2$ (fold change)$|>1$). 86 genes are differentially expressed in both OX3 and OX5. 69 of these genes are up-regulated, 13 down-regulated, 3 silent and 1 activated. The Venn diagram was obtained using http://bioinformatics.psb.ugent.be/webtools/Venn/. (**B**) The expression profiles of 86 significantly expressed genes were illustrated by the heatmap. FPKM in each row are z-score normalized, and hierarchical clustering was based on the Euclidean distance.

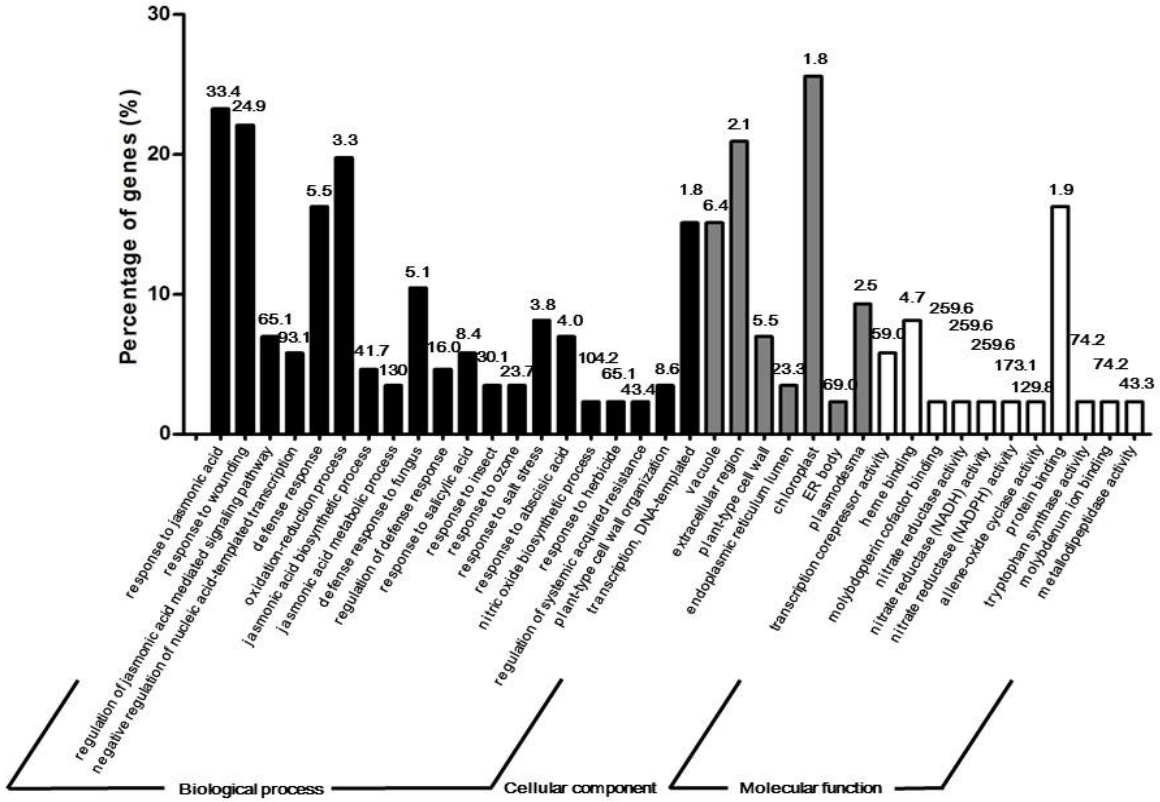
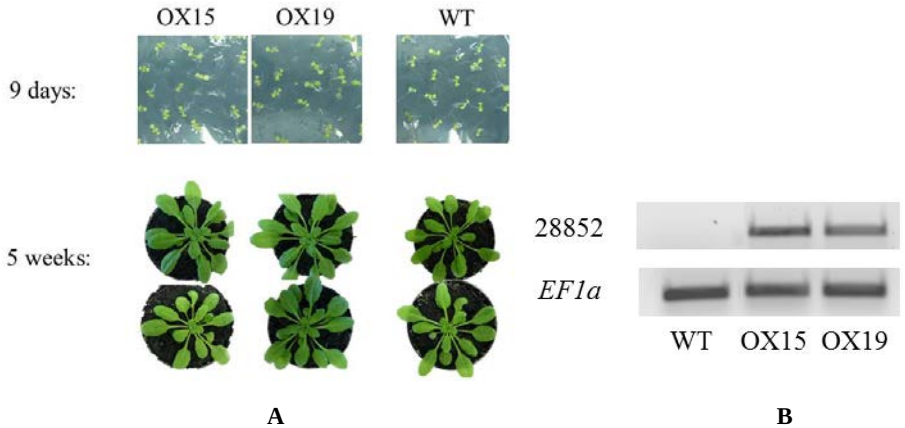


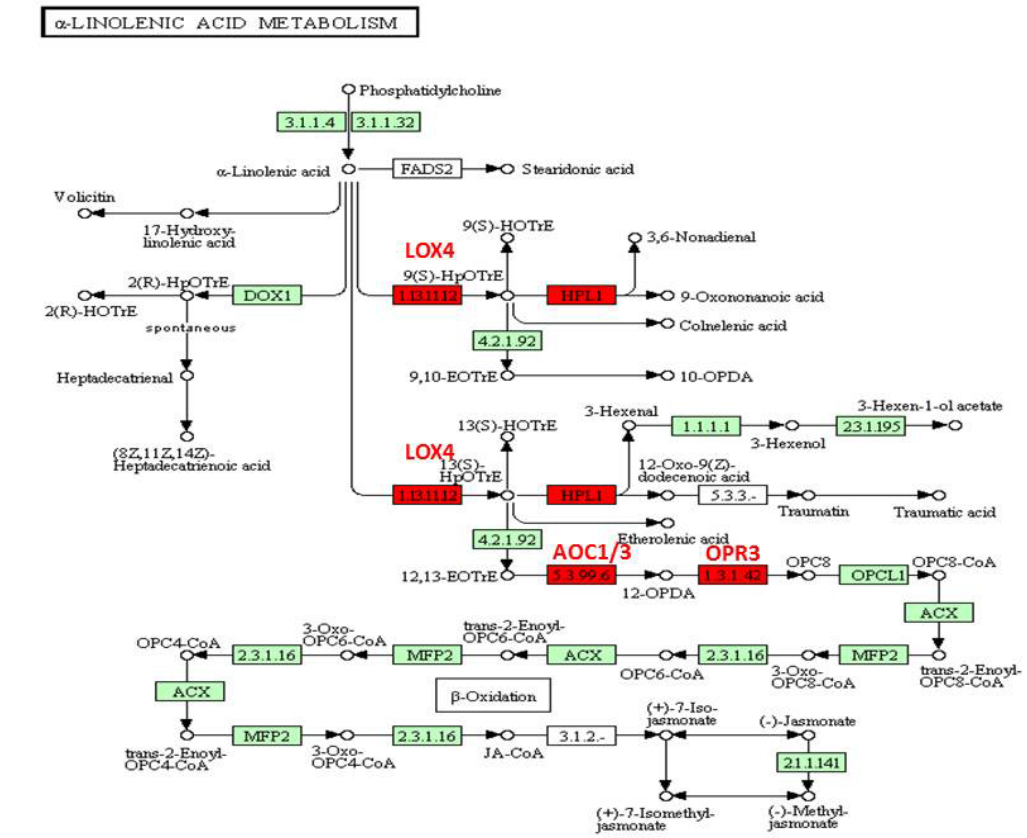
Figure 3.26 GO enrichment analysis of 86 genes differentially expressed in WT and T_2 generations of OX3 and OX5. (**A**) Overrepresented GO terms, involving biological process, cellular component and molecular function, were identified with a modified Fisher's exact test ($p < 0.05$). The y axis shows the percentage of genes in each GO term, and the x axis represents the GO terms. The number above each bar indicates the fold enrichment of each term.



A **B**

Figure 3.27 Phenotype observations of wild type and transformants OX15 and OX19. (**A**) The Phenotype of wild type and transformants OX15 and OX19. (**B**) The expression of lncRNA 28852 in wild type and mutants.

To validate the RNA-seq data and investigate the expression patterns of specific genes, the semi-quantitative PCR was performed with 10 genes using the RNAs isolated from WT, T_2 generations of OX3, OX5, OX15 and OX19. All the selected genes conformed to one or two of the following requirements: 1, the significant change in the expression levels between WT and mutants 2, FPKM > 50 (at least one sample) 3, being in one of the enriched GO terms (**Figure 3.26**) or KEGG pathways (ath00592 and ath00941). Of the 10 genes selected, 3 genes were enriched in the flavonoid biosynthesis (ath00941) pathway (TT4, F3H and DFR) and four genes were in α-linolenic acid metabolism (ath00592) pathway (AOC1, LOX4, HPL1 and OPR3). The the other three genes encoded light harvesting complex photosystem II (LHCB4.3), chlorophyllase 1 (CLH1) and 2-oxoglutarate and Fe (II)-dependent oxygenase superfamily protein (AT5G05600). The two mutants, OX15 and OX19, were selected according to their phenotype and the expression of the target transcript lncRNA28852 (**Figure 3.27**). The RT-PCR analysis (**Figure 3.29**) showed that four genes, LHCB 4.3, F3H, TT4 and DFR, were down-regulated and others were up-regulated in OX3 and OX5 which is consistent with the RNA-seq data. However, in the mutant OX15 the DFR showed the same expression level as WT and only DFR was down-regulated in OX19. The expression profiles of the ten genes showed similar trend in OX3, OX5 and OX15 when not comparing with those in WT. Nevertheless, both the large error bars in the **Figure 3.29** and the correlation (R^2=0.74, **Supplementary figure 11**) between RNA-seq data and RT-PCR data indicated unstable expression of the genes in the mutants.



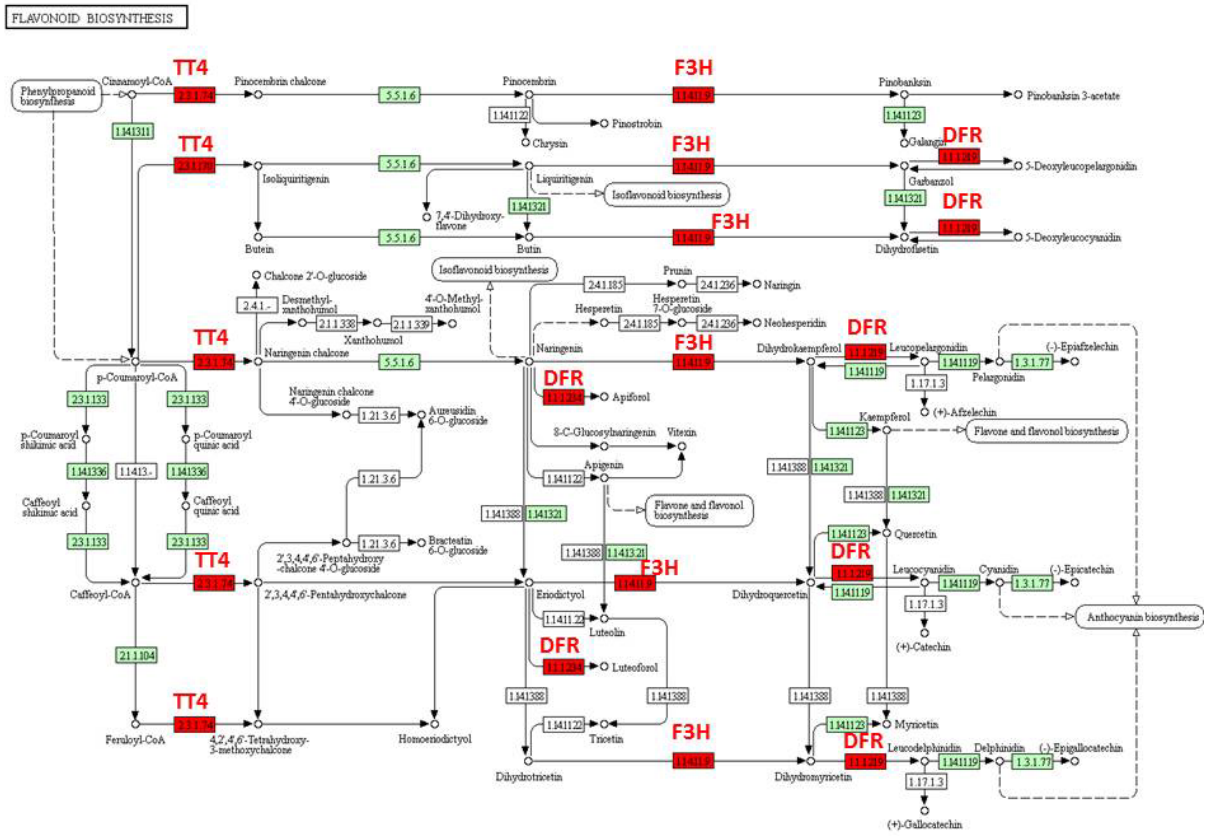


Figure 3.28 KEGG pathway enrichment analyses. The pathway enrichment analysis was performed using DAVID bioinformatics resources 6.8 (https://david.ncifcrf.gov/home.jsp) (p < 0.001) (Huang et al. 2008; 2009). The overrepresented KEGG pathways were identified in the KEGG pathway database (https://www.genome.jp/kegg/pathway.html). The significantly differentially expressed genes in mutants associated in the pathways are indicated by red fonts and rectangles.

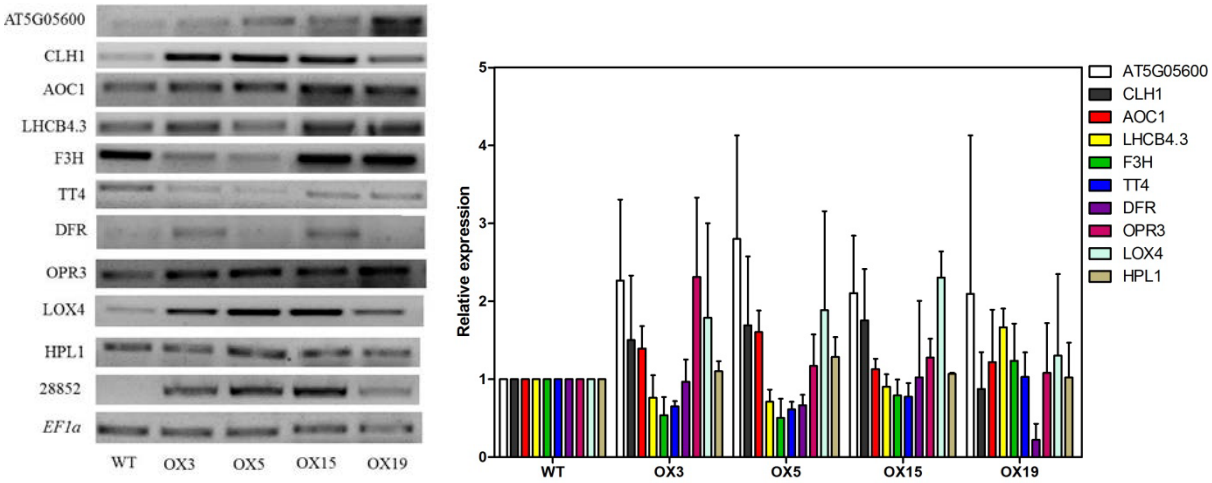


Figure 3.29 The expression patterns of the ten selected genes were confirmed by RT-PCR. The bar chart represents the semi-quantitative RT-PCR analysis using ImageJ software. The relative expression levels of the genes were normalized to that of WT and generated from three biological replicates. Error bar ± Sem.

3.3.4 Evaluation of T_3 transgenic *Arabidopsis thaliana* lines overexpressing the lncRNA 28852 (OX3 and OX5)

To investigate the effect of lncRNA 28852 on the phenotypes of transgenic *A.thaliana*, the germination rate, root length and ion-leakage of wild type (WT) and T_3 generation of transgenic lines (OX3 and OX5) under different conditions were determined. As shown in the **Figure 3.30 A and B,** there is no significant difference in the germination rates between wild type and mutants (OX3 and OX5), and both the WT and mutants (OX3 and OX5) germinated about at the same time frame. According to the GO and KEGG pathway enrichment analyses on the RNA-seq data, the effect of MeJA on the germination was also tested. The seeds were incubated on 1/2 MS medium with 50 µM MeJA. Compared to the mutants, WT seeds germinated a little earlier (**Figure 3.30 B**). **Figure 3.30 C** shows that the root length of mutants (OX3 and OX5) were significantly shorter than that of WT, but the the maximum difference between WT and mutants was only 0.3 cm. No significant difference was detected when the seedlings were transferred to plates with 20 µM ABA, 250 mM mannitol, or 100 mM NaCl, respectively (data not shown). Measurements of the ion-leakage after dehydration and rehydration showed that no serious damage to the cell membranes of the WT or the mutants (OX3 and OX5).

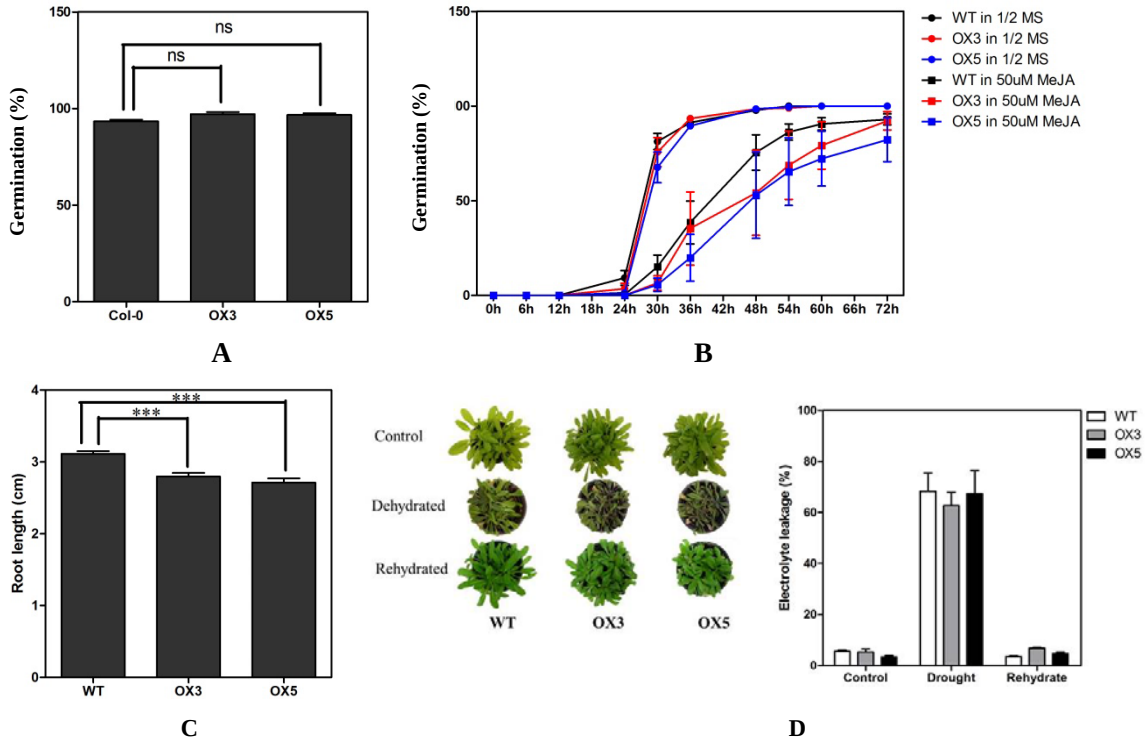


Figure 3.30 Phenotypes observations of WT, OX3 and OX5. Germination rates of WT and mutants (OX3 and OX5) were measured on 1/2 MS medium (**A, B**) and 1/2 medium containing 50 µM MeJA (**B**). The seeds were plated on 1/2 MS medium (with or without MeJA) and the germination (full radicle emergency) was determined after 3-days growing on petri dish (**A**, t-test compared to WT, ns means no significant, n ≥ 100) or at 6-hour intervals (**B**, one-way ANOVA, $P = 0.049$, n ≥ 100). The root length of WT and and mutants (OX3 and OX5) were determined using the ten-day seedlings grown on 1/2 MS medium (**C,** t-test compared to WT, *** $P < 0.001$, n ≥ 50). The electrolyte leakage of WT and mutants (OX3 and OX5) under dehydration and rehydration was determined (**D,** one-way ANOVA, $P = 0.9737$, n = 9). Error bar ± SEM.

4. DISCUSSION

4.1 *C.plantagineum* wall-associated protein kinases (CpWAKs)

CpWAK1 and CpWAK2 were first identified by Giarola et al (2016). The two CpWAKs showed the same expression trends that both are down-regulated by dehydration but induced by rehydration. CpWAK1 interacts with another cell wall protein CpGRP1 *in yeast* and *in planta.* These observations suggest the involvement of CpWAKs in the reversible cell wall folding during dehydration and rehydration (Giarola et al. 2016). In addition to the two CpWAKs, CpWAK3 was identified in this work. The newly identified CpWAK protein shows high similarity to the other two CpWAKs. To investigate the functions of CpWAKs in the cell wall matrix, the prediction of gene and protein structures and the evolutionary analysis of CpWAKs were first performed. And the expression patterns and the interaction among CpWAK-CpGRP-pectin *in vitro* were determined as well. In this part, possible roles of CpWAKs will be discussed based on the results obtained during the thesis work.

4.1.1 Evolution of CpWAKs

Rizzon et al. (2006) proposed that the genes encoding membrane proteins and associated with stress responses are often organized in tandem repeat gene clusters. The genome-wide analyses of WAKs in rice, Arabidopsis and apple confirmed the existence of WAK gene clusters and indicated that tandem duplication and segmental duplication contribute to the expansion of WAK gene families (Shiu et al. 2004; Zhang et al. 2005; Zuo et al. 2018). In *C.plantagineum*, the high sequence similarity among the three CpWAK DNA sequences (including the promoter regions, **Supplementary figure 5**) suggests the duplication of *CpWAK* genes. However, to date no evidence confirms the cluster localization of *CpWAK* genes, because the analysis of genome sequencing data is not yet available. The phylogenetic analysis displays that the WAK homologs from different species were clustered in distinct species-specific groups (**Figure 3.4**). The cluster analysis is consistent with that in rice, Arabidopsis and apple (Zhang et al. 2005; Zuo et al. 2018), which suggests that gene expansion/gene duplication takes place after species divergence. Among the selected homologs of CpWAK1, the non-intron gene structure is only conserved in the WAK genes from the Linderniaceae family (group IV) (**Figure 3.4**). This observation also supports the conclusion above. The protein structure prediction of the selected WAKs shows the conserved cytoplasmic kinase domain and variable extracellular regions. In the diverse extracellular regions, except for the conserved EGF-like domain and galacturonan-binding domain there are some domains localized in certain WAK homologs (**Figure 3.4**). These domains exert special functions in some biological processes, such as the membrane lipid-related PNPLA domain in Lbr 010788 (**Figure 3.4**). Besides the extracellular region, the conserved kinase domains of plant WAKs also have evolved into two classes: WAK-RD and WAK-nonRD after the monocot-dicot separation (de Oliveira et al. 2014). The classification of RD and nonRD classes depends on the presence of a conserved arginine (R) residue before the catalytic motif DxxxxN. All the three CpWAKs possess the RDxxxxN (**Figure 3.3**) and thus belong to the WAK-RD class. The nonRD WAKs only occur

in monocots. The different catalytic domains in the two WAK classes may lead to the different signaling pathways (Kohorn 2015). All in all, the miscellaneous extracellular region and the evolvement of the kinase domain collectively contribute to the involvements of WAKs in different biological processes.

4.1.2 The expression of CpWAKs under different conditions

The promoter analysis of the three highly conserved *CpWAK* genes identified some known functions of *cis*-acting elements. In these elements, there are four light-responsive elements, one circadian-related motif and two elements involved in hormone responsiveness (**Table 3.3**, **Figure 3.5**). These identified *cis*-elements imply that the transcript expression of CpWAKs is probably regulated by many different conditions and thus supporting the notion that CpWAKs may be involved in diverse biological processes.

The transcript expression analysis of CpWAKs under light and dark from 7:00 am to 23:00 pm showed that the expression of CpWAKs are controlled by a circadian clock which can be disturbed by unknown factors in the light (**Figure 3.6 A**). The identification of light and circadian-related *cis*-elements in the promoter region of MdWAK-RLKs (Zuo et al. 2018) and the authentication of other RLKs expression controlled by the circadian clock (Nguyen et al. 2015; Wang et al. 2011) make it tempting to speculate that WAKs as members of RLKs are likely to function in certain biological processes related to the circadian clock regulation. The auxin responsive cis-element TGA-element may be not effective due to insignificant up-regulated transcript expression of CpWAKs under NAA treatment (**Figure 3.6 B)**. The SA or MeJA related W-box motifs only show weak effect on the transcript expression of *CpWAK*s, because the transcript expression of *CpWAK*s was only slightly up-regulated after 1h of SA treatment and significantly reduced after 24h treatment (**Figure 3.6 B, Figure 3.8 A)**. The transcript expression of *CpWAK1* and *CpWAK3* were retard under MeJA treatment while *CpWAK2* was accumulated after 1h MeJA treatment (**Figure 3.8 A)**. This observation and the lower number of predicted W-boxes in the CpWAK2 promoter (**Table 3.3)** suggest that there are other *cis*-elements associated with the regulation of CpWAKs under MeJA treatment. In *Arabidopsis*, the transcript expression of WAKs is inducible by SA or its analog, INA (Park et al. 2001; Verica et al. 2003; Zheng-Hui He. et al. 1999). In *C.plantagineum* only CpWAK proteins accumulated after 3h of SA and MeJA treatments and down regulated after 48h of treatments while no significant up-regulated expression patterns were observed on the transcript level **(Figure 3.8)**. This means that the expression of CpWAKs under SA and MeJA are mainly controlled post transcriptionally. In contrast to the modulation of CpWAK expression under SA and MeJA treatment, gene expression of *CpWAK*s under drought stress are regulated on the transcriptional or post-transcriptional level, because the expression of CpWAKs shows the same trend on the transcript and protein level under dehydration and rehydration (Giarola et al. 2016) (**Figure 3.7)**. Therefore, the regulation of gene expression under different treatments indicates that CpWAKs participate in multiple responses through different signaling pathway.

4.1.3 CpWAKs can form multimers

Western blot analysis of purified recombinant proteins with both His-tag and WAK1 antibodies shows

two or more than two bands detected by the two antibodies in each lane (**Figure 3.10**). The multiple bands in the membrane suggest that CpWAKs can form dimers or multimers. The formation of the multimers is mainly in the R-2 segment because there were more bands in the R-2 lane. During the experiments, the recombinant proteins of truncated fragments of CpWAKs cannot migrate into SDS-PAGE gels without DTT, except for the recombinant protein R-3. The cysteine residues in the protein sequences are instrumental to the formation of disulfide bridges. In the recombinant extracellular domains of CpWAK proteins, there are many cysteine residues localized in the galcturonan-binding domains, EGF-like domains and other unnamed domains (**Figure 3.3, 3.9**). These cysteine residues, especially those located in the R-2 segment, presumably result in the formation of CpWAK dimers or multimers via intermolecular disulfide bonds (**Figure 4.1**). Besides the cysteine-rich domains, the EGF-repeats can also lead to the dimerization of proteins mediated by calcium (Anderson et al. 2001). The recombinant protein of the extracellular region of CpWAK1 is precipitated in the presence of Ca^{2+} (**Figure 3.14**), which may result from the calcium-mediated protein dimerization by EGF-repeats (**Figure 4.1**).

4.1.4 The CpWAKs bind to pectins

Decreux and Messiaen (2005) first identified the WAK-pectin binding *in vitro* via ELISA binding assay. This similar binding was also observed using the recombinant proteins of CpWAKs and the pectin extracts from untreated *C.plantagineum* (**Figure 3.14**). According to Decreux and Messiaen (2005), AtWAK1 shows higher affinity for the pectins in the "egg-box" model (**Figure 4.1**), a calcium-induced conformation of polyuronic acids (Grant et al., 1973). This observation is also consistent with that in CpWAKs. CpWAKs do not bind with the polygalacturonic acid (PGA, Sigma) or commercial pectin (pectin from citrus peel, Sigma) directly **(Figure 3.14 A**). However, the binding between CpWAK1 and commercial pectin was finally identified in Tris/NaCl/Ca^{2+} buffer after saponification of the commercial pectins **(Figure 3.14 D**). Saponification can break ester bonds. Then the pectic chains can form the "egg-box" structure in a calcium environment (Sedan et al. 2007). CpWAKs do not bind to the commercial pectins due to the high methyl ester concentration of commercial pectins ($\geqslant$74%). The binding assays using different segments of the extracellular CpWAK1 proteins exhibite that both R-2 and R-4 containing the galgcturonan-binding domain have a stronger preference to pectin extracts while R-3 with only EGF-repeats shows the weakest binding (**Figure 3.9, Figure 3.15 A**). This result demonstrates the importance of the galgcturonan-binding domain in the CpWAK-pectin binding. Protein band shift assays on PAGE gels were also carried out (data not shown). However, on band shift was observed in the presence of DTT. Therefore, the cleavage of the disulfide bonds with DTT not only disrupts the formation of multimers but prevents the binding of CpWAKs to pectins as well. This suggests that the CpWAKs bind to pectins as dimers or multimers caused by disulfide bonds and the affinity of CpWAKs to pectins increases as the CpWAK multimer formation (**Figure 4.1**).

4.1.5 The binding of CpWAK1 to pectins is modulated by CpGRP1

The interaction of CpWAK1 and CpGRP1 was demonstrated *in yeast* and *in planta* by Giarola et al. (2016). It was hypothesized that CpGRP1 may modulate the interactions of CpWAKs and other cell wall components based on its structural and chemical property (Giarola et al. 2016). The heatmap in

Figure 3.15 B shows that more CpGRP1 lead to more CpWAK1 binding to pectins but more CpWAK1 leads to less CpGRP1 binding. The N-terminal fragment of CpGRP1 without the pectin-binding domain competed with pectins for interacting with CpWAK1 **(Figure 3.15 B)**. These results indicate that CpGRP1 contributes to the binding of CpWAK1 to pectins by interacting with CpWAK1 (**Figure 4.1**) and support the hypothesis proposed by Giarola et al. (2016). GRPs interact with the extracellular domain of WAKs (Giarola et al. 2016; Park et al. 2001). The effect of CpGRP1 on the binding of CpWAK1 and pectins was also observed using the CpWAK1 fragments, R-2 and R-3 **(Figure 3.15 B)**. The R-3 fragment reduces the binding of CpGRP1 to pectins, which implies that CpGRP1 mainly interacts with the R-3 fragment of the CpWAK1.

4.1.6 The roles of CpWAK-CpGRP-pectin complex in defense responses to stresses

All the pectin binding assays discussed above were performed in buffered solutions with pH 8. However, the pH of the apoplast (pH_{apo}) varies in the range of 4 to 7 and generally is around pH 5 (Grignon and Sentenac 1991). The acidification and alkalinization of apoplast take place in the growing tissues and tissues under stress respectively (Geilfus 2017; Grignon and Sentenac 1991). The duration and the range of pH_{apo} changes are different under different stresses (Geilfus 2017). The pectin-binding capacity of CpWAK1 *per se* did not show significant differences at different pH values (from pH 4 to pH 8), but the CpWAK1-CpGRP1 complex was sensitive to the pH changes, showing stronger affinity for pectins at pH 4, 6 and 7 (**Figure 3.16**). The influence of CpGRP1 on the interaction of CpWAK1 and pectins at different pH values means that CpGRP may aid in distinguishing signals of CpWAK1 via perception of pH_{apo} changes. Thus CpWAK1-CpGRP1 can be involved in the cell wall loosening and cell expansion induced by the acidification of the apoplast as well, besides the stress responses (Giarola et al. 2016; Kohorn 2015). Ca^{2+} can enhance the binding between AtWAK1 and pectins (Decreux and Messiaen 2005). The similar phenomenon was observed for CpWAKs (**Figure 3.14 C**), but only at pH 8 using the pectins isolated from untreated *C.plantagineum* leaves (**Figure 3.17 A**). WAKs interact with pectins in calcium-induced "egg-box" structures formed with de-esterified pectins (Decreux and Messiaen 2005). Therefore the insignificant effect of Ca^{2+} on the WAK-pectin binding may arise from the saturation of the binding sites of the de-esterified pectins. The opposite effects of Ca^{2+} on the CpWAK1 binding to different pectin extracts at pH 8 may result from the effect of the pH on CpWAK1 conformation and/or different cell wall compositions in the pectin samples. No pH_{apo} at 8 is reported so far, and the pH_{apo} can go up to 7 at the most. The pH_{apo} may reach pH 8 when the plants are under extreme stresses. Therefore, it is likely that CpGRP1 prioritize the Ca^{2+} in modulating the CpWAK1-pectin binding in normal conditions when sufficient $Ca^{2+}_{[apo]}$ participates in the formation of "egg-box" pectin structures. Presumably Ca^{2+} influences on the CpWAK1-pectin linkage and thus triggers special signaling pathways under extreme conditions.

WAKs are always thought to be associated with the wounding or pathogenesis-related process partially due to the increased expression after wounding or pathogen infection (Kohorn 2015; Kohorn and Kohorn 2012; Park et al. 2001). Plant hormones SA and JA are involved in the responses of plants to pathogens and wounding (Dong 1998; Reymond and Farmer 1998). The protein expression of CpWAKs can be induced by both SA and MeJA. Treatments of

C.plantagineum leaves with the two hormones also resulted in increased expression of CpGRPs, the interaction partner of CpWAKs (**Figure 3.8**). The protein accumulation of CpWAKs and CpGRPs almost happened at the same time, after 6h of the treatments, and finished after 48h of the treatments (**Figure 3.8**). The simultaneous accumulation of CpWAKs and CpGRPs make it possible that CpGRP acts as a modulator in regulating the cell wall signal perception of CpWAKs after wounding or pathogen infection. The de-esterified pectins in the cell wall are necessary for the formation of "egg-box" gelatin. The pectins can be de-esterified by pectin methyl esterase, which is inhibited by the pectin methyl esterase inhibitor (Micheli 2001). Previous studies showed that SA and MeJA led to the up-regulated expression of pectin methyl esterase inhibitor and thus gave rise to the controlled activity of pectin methyl esterase and decreased de-esterified pectins (An et al. 2008; Meng et al. 2009). So the weaker affinity of CpWAK1 for the pectin extracts from SA and MeJA-treated *C.plantagineum* leaves (**Figure 3.16 C**) may result from the fewer "egg-box" structures in the cell wall pectins induced by the lower activity of pectin methyl esterase. To sum up, CpWAK decodes the cell wall signals in concert with CpGRP as the modulator of CpWAK-pectin interaction in responses to wounding or pathogen infections. The cell wall signals involve the changes of pH_{apo} and the state of cell wall pectins which may be affected by plant hormones SA and MeJA (**Figure 4.1**). Additionally, CpWAKs are also responsive to drought stress. The significantly decreased protein expression of CpWAKs was validated in the desiccated leaf tissues (**Figure 3.7**). Based on the responses of CpWAKs and CpGRP1 to drought stress, Giarola et al (2016) proposed that the CpGRP1-CpWAK1 complex is implicated in the cell wall remodeling during dehydration and rehydration. Although more pectins in "egg-box" conformation may be present in the desiccated tissues in the light of increased de-esterified pectins and Ca^{2+} in the dried Craterostigma plants (Jung et al. 2019; Vicré et al. 2004), the CpWAKs did not show significantly different affinities for the pectin extracts from hydrated or dehydrated *C. plantagineum* leaves (**Figure 3.17 B**). These observations imply that CpWAKs require the modulation of CpGRP1 in the early stage of drying (**Figure 4.1**). During desiccation the low level of CpWAKs presumably gives rise to less WAK homo- or hetero-dimers and multimers. Given that the CpWAK multimers appear to promote the CpWAK-pectin binding, CpWAKs may mediate their binding to pectins *per se* under desiccation (**Figure 4.1**).

WAKs, as receptor kinases should be capable of recognizing different signals from cell wall with the help of ions and protein ligands. It is a big challenge to understand how WAKs distinguish the different signals from cell walls. In this work, CpGRP, pH_{apo}, $Ca^{2+}_{[apo]}$ and the formation of CpWAK multimers or dimers are all considered as potential factors involved in the orchestrated processes. Each factor has a potential to play a prevailing role in the regulation of signal reception under stress (**Figure 4.1**). More work is required to figure out the mechanisms, and the protein structure information and the calcium- and ligand- induced conformation change may be the key for solving the puzzles.

Model for the role of CpWAK proteins

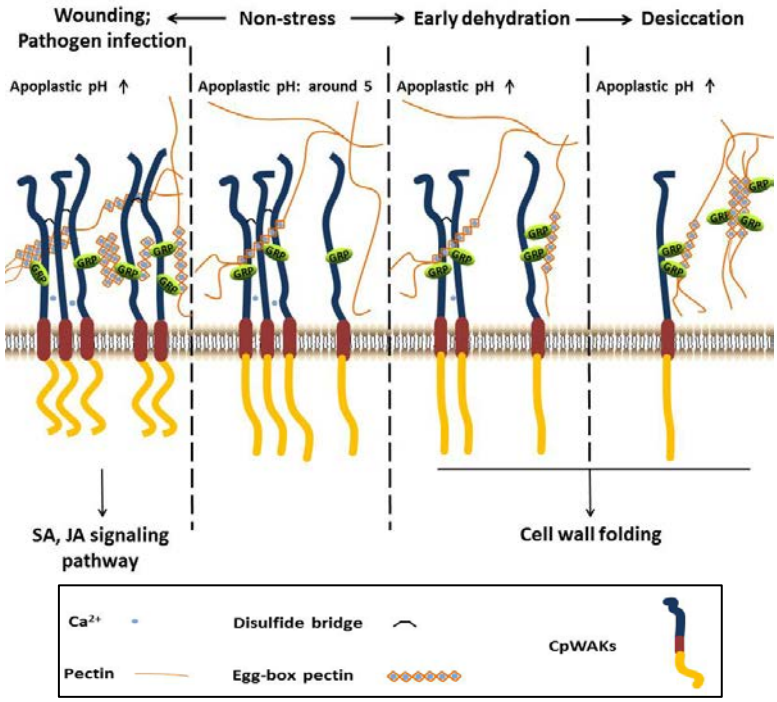


Figure 4.1 The model for the role of CpWAK proteins. CpWAKs are a group of cell wall-located receptor protein kinases involved in the cell wall folding during dehydration and resistance in response to wounding or pathogen infections. CpGRPs are interaction partners of CpWAKs in the cell wall matrix. Both CpWAKs and CpGRPs have a stronger preference to pectins in "egg-box" structure. The binding capacity of CpGRPs to pectins is sensitive to pH changes. In the early stage of dehydration, CpGRPs are up-regulated while less CpWAKs accumulate. The linkages of CpWAKs to pectins can be strengthened indirectly by the increased pH_{apo} with the aid of CpGRPs. Under desiccation, a considerable amount of CpWAKs are degraded, which may result in less CpWAK multimers or dimers. Although the increased CpGRPs contribute to the binding of CpWAKs to pectins, the decreased multimers or dimers may become the major factor negatively affecting the binding of CpWAKs to pectins. In contrast to the early stage of dehydration, wounding and pathogen infections also induce more CpWAKs accumulated and cell wall pectin fragmented except for the increased pH_{apo}, the accumulation of CpGRPs and the de-esterified pectins. More CpWAKs indicate more multimers or dimers formed which can facilitate the binding of CpWAKs and pectins. The pectin fragments have a higher affinity for the CpWAKs than the long pectin oligogalacturonides. The combined effects of the increased $pH_{apo,}$ the modulation of CpGRPs, the formation of CpWAK multimers and the state of cell wall pectins on the conformation of CpWAKs may finally cause the defense responses in which the SA, JA signaling are involved in.

4.2 *C.plantagineum* Germin-like protein 1 (CpGLP1)

Many GLPs have the OXO and SOD activities and are localized to the apoplast (Dunwell et al. 2008). *C.plantagineum* germin-like protein 1 (CpGLP1), as a putative interaction partner of *C.plantagineum* cysteine-rich rehydration responsive protein 1 (CpCRP1) in yeast-two-hybrid (Y2H) assay (Dulitz 2016), also possesses the SOD activity (König 2017). It can be detected in cell wall protein extracts of *C.plantagineum* leaves, especially in the $CaCl_2$ and LiCl fractions

(**Figure 3.18, 3.19**), which indicates that CpGLP1 is localized in the apoplast. GLPs are supposed to play structural roles in the cell wall matrix (Schweizer et al. 1999). The cell wall GLPs with the activity of SODs also participate in the production of apoplastic ROS, which play the dual roles *in vivo* acting as signals or triggering the oxidative stress and affect the cell wall cross-linking (Fry 1998; Miller et al. 2010; Passardi et al. 2004; Petrov et al. 2015). The protein expression profile analysis of CpGLP1 showed that both dehydration and rehydration can induce the accumulation of CpGLP1 (**Figure 3.19**). Therefore, the apoplastic CpGLP1 has a potential to affect the cell wall remodeling during dehydration and rehydration. The different mobilities of proteins on the SDS PAGE gels indicate that the proteins may have isoforms (Bernier and Berna 2001). The two bands of CpGLP1 in **Figure 3.19 C** suggest that the apoplastic CpGLP1 may be present in two isoforms. The multiple CpGLP bands on the gels loaded with the total proteins means that there are other CpGLP isoforms intracellularly. The interaction of CpGLP1 and CpCRP1 was confirmed *in yeast* (Dulitz 2016), but it cannot be identified in pull-down assays (**Figure 3.20**) or BiFC assays (**Figure 3.21**). There are many reasons why CpGLP1-CpCRP1 interaction could not demonstrated *in planta*, such as false positive interaction *in yeast*, the involvement of additional proteins in the interaction, and using unsuitable vectors or experimental systems.

4.3 Long non-coding RNA 28852

LncRNA 28852 was identified as a desiccation-induced taxonomically restricted gene in the *C.plantagineum* transcriptome analysis (Giarola et al. 2015). LncRNA 28852 can be amplified from the cDNA samples reverse transcribed with oligo-dT primer, so that the lncRNA 28852 is considered as polyadenylated RNA. The transcript expression analysis of lncRNA 28852 using total RNAs extracted from *C.plantagineum* roots and leaves (**Figure 3.22 A**) confirmed the the transcriptome analysis that the lncRNA 28852 is responsive to desiccation in both roots and leaves. However, the expression analysis using the polysomal RNAs showed the opposite expression trend during desiccation and rehydration with more accumulation in partially dehydrated leaves and rehydrated leaves but less in desiccated leaves (**Figure 3.22 B**). The polysomal RNAs refer to the RNAs associated with ribosomes. The polysome profiling analysis is favorable for identifying the RNAs in the translation elongation stage (Li et al. 2017). The expression analysis of the lncRNA 28852 using both total RNAs and polysomal RNAs indicated that lncRNA 28852 does not interact with the translation elongation complex during desiccation and thus may not function in the protein translation when it is strongly up-regulated. Guttman et al. (2013) proposed that the RNAs can be protected by the associated ribosomes without initiating the translation. Therefore, the interaction between lncRNA 28852 and ribosomes in the early stage of dehydration is likely to facilitate the accumulation of lncRNA 28852 during desiccation.

Because the biological functions of lncRNAs can be analyzed by the investigation of the relevant mutants (Liu et al. 2015), transgenic *Arabidopsis thaliana* lines overexpressing the lncRNA 28852 were established. However, the overexpressing lines showed diverse phenotypes (**Figure 3.24**), and the selected overexpressing lines OX3 and OX5, with the strongest lncRNA 28852 expression among ten lines (**Figure 3.23 C**), did not show improved drought tolerance (**Figure 3.30 D)**. The comparative RNA-seq analysis of wild type (WT) and OX3 and OX5 was performed. 86 genes were identified as significantly differentially expressed in WT and mutants OX3 and OX5 **(Figure**

3.25), based on which the GO and KEGG pathway enrichment analysis showed that the JA signaling pathway got affected in the mutants (**Figure 3.26, Figure 3.28**). But no significant change in germination rate was detected in WT and mutants in the absence or presence of MeJA (**Figure 3.30 A and B**). Only the root length of the mutants OX3 and OX5 showed significant differences, but the maximum difference between WT and mutants was only 0.3 cm (**Figure 3.30 C**). The overexpressing lines OX15 and OX19 also showed strong expression of lncRNA 28852 (**Figure 3.27)**. But the expression patterns of the selected genes were not repeated in OX15 and OX19 (**Figure 3.27, Figure 3.29**). The regulatory effects of lncRNAs on the transcription of the target genes in the close spatial proximity (Rinn and Guttman 2014) is likely to explain for the observations on the mutants given the random insertion of the foreign DNA. The disordered expression patterns of the selected genes and the unconspicuous change in the phenotypes indicate either the ineffective role or the intricate effects of *C.plantagineum* lncRNA 28852 on *Arabidopsis thaliana*.

5 SUPPLEMENTARY DATA

CpWAK1 (Genomic sequence):

CCTCAACCATACATATACAGCGACATTGCAATAATTCAGTGGGACGTGCATTAATTTGATATATCGTACT
GTTATACGTTACGTCGAAATCACTTCATCGTCACCAGGTCAAGTCAACGACCCTATTTGCCTTTCTTGCA
TTAAACTCGTCCATATTTTCCCAACCAAGCAAATAGCTATCTATTCTATACACTGCTCTCACCACAAAAG
ATCCTCTAAAGCTATTAACGGGATGTCTCATCTAAAGTGCGTCTGAATATTTGAAATCATATAACTACAA
ACATGGTGATCGAGCGAGTAATAAAATACAGCGAACATCAACAAACTATCATATGCTCAACAGGCAAG
TGTACCCATCGAAATAGCATAGACGTAGTAAGTATTAGGGTATCGAATACGAAGAGAACTATTTTGAAA
TTTTTTTGAAGTATCCTAGTATGAATGTTTGATTGTTGGTAAATAACAATACTTAGAAATACAACTTATCG
AGAGGAAGCCCTTATGAAAACAAACGTATAAGATTTCGAATCCCCCTAACTTGACCAAAACAAATATC
AAGTGTACGTAAACAAAGCTATCATTAGACCGTGAAAATTCCTATAACATGCATACGAGTTCGTATGGC
TGACGAAACTGCGCCTTACATACAATAATCGTAAACAACTAATCAACGAATTAGGTAGGTGGGCAAAC
TTCCCTAACTACTTAACCGAAGATCGCATTAAAACACGATGAATTTCGCTATTAGATTTGACTCTATCCG
ACGAGATAGTTCAACATATCTCAAGCGATGAGTTTATTCCGCTCGTTTACGAGTGGTCTAATCGTTCGC
CCCTCTCTCGAGCGTGGCTCCCGATTTCCGAAACTACGTTAACTTAATCACATAAACATGATCGAGCTA
ATTAGACAATTTCAATCACATAGACATAATCAAAACATATGATTCATAAACGAGTCATAACAGTAGCAA
ACATCAAATACACGCAAACATACGAACGATCGCGCTAGGGTTCATCTCATCTCAAATACAAATTATTTA
ACTACTCATAATGATCACAACAAAGCACATAAGTAACATTGAATGCATCATAATGCTTCCCTGAATTTGA
CTTCGAAGTTAATCCACGTTCTTCGGCGAGTACTTTTATCGAACACGTGGTGAACGAACTAATTTGTGT
AGAGAAGTGTTCTTGGACTTATGGAAGTAACGTGTATGTCTATGTAAGAGTGAATTGGCTTGTACCTTT
CTCACGTTCCTCCTCTTCTTTTTCTTTTCCTCACGTTGCTCTTTCTATAGAATTTTCCCTTCTTTCACTTG
TGCAATTTCTTCAACGTGAGTTTCAAGGAGAAGAAATTTTGAAAGATCTATAGCTCCAACACATTCTTT
GCAGCCCGAAAATCTGATTTCATGTCTGTTATTAAATTGAATGGCTTGTATTATGCAGAGATATGGTCCC
AAGCTCTTACGTTTGTGATCTTTCCTTTGACAAATGTGGCACATTTCGATGTAAGTAGCCTTGACTTTTA
ACATGGTCATGGAAACGTTCACGGCCTCGTATACATAGCTTCGCGGTCGTGACAACATTGGTATCGAAG
AGAAGAGGTGCCTGCGTTCACGGCTTCGGCAGCTACCATTCACGGGCGTGACACATCCTTCACGGCTT
GCTTTCTTCATATTCACGGTCGTGAAATGTTCGAGCTTAGAAAATGGTTCTCTGCCCGCATTAACGGCC
TCTTGCACCTCCTTTCACGGCCGTGAAGGACTTGGTTGGTTCATGTTCACGGCCGCAAAGTGATTGGG
CAAGAAAATGGCTTTCTGGTTGCCTTCACGGTCTCTTGCACAACCTTCACGATCGTGAAGCTATTTTCT
GCAAAATAACAGTAAGTTCAAAAATTCATATTCCCAATCTAGACATTGGATTGAAAATCCCTTTAGTCA
GGATCAATACTAGCGAGTCTTCTATATCATTGAGAATTACGAAACTGTGGCAAAACTGCCGTAAAAACG
TTTGATTTTGCAGTGAAATAAAAATATAGCGCTATGCAAACGACTTAAATTTAACAACTATAAACATAAG
AAATTACGTTTATCACATGGCCAAGTAACACGTTCCAAAATAATACATTCTACAATTAATAAATAGACTG
AACTTGATCATCACTAAATAGACTGGACCCACCATCATAATGTGGGTTATAGCGAATTGAATACGTGGA
AAGAGTCCGACGGAGTACTATCAATCCAACTGCTCGAAGTCCTCAGAAATCAAGTCGTTTTTACCTATT
TATATTTCTCCATATCCAATCTAATTGCACTATCCTCGTCTCGCTCTCTTATTCTGTCGTGACATAAGGAA
AAAAGAAAAACAAATATATCACCAATCTTATTAGTACTCAAGTAAGAGCGGAGATCGAAGAATCAAAC
AGGTACGTATACGTCCATTATGGTAACAACTTCAACTTGCAGAAATATCATAAAAACAATGTATAAGCAT
AAAAATAAAAAAAAAAAAGAAAAAAAAATTGATATACCGTTTGGCAACTGAATTCTGTGAGCTCATCA
AGTTCGTATTACAACAGGCCAACACGGAA**ATG**AGGACTTCTTCAATACTTGCTGCAGTCGTTTTCT
GCGTCTGCCTTAGCACAACTTCGGCTATACCATCATCATCCATTTCCATACCCATGGCGAAGCCT

AATTGTAGCGAAACATGCGGAAATATTACCGTTCCATATCCATTCGGCATGAGCTCAGAGTGCAG
TATCGACTTAGATTTTCTCGTCGAGTGCCGAAACTCCACAAATCCTCCGACGCTGTTCTTGTCCA
GCATCGACCTGCAAGTAACTGACATTCGCGTGAACGGCACAATCGTCGTAAAATATCCCGTCAC
TCCAATCAACTGTACCGCCGTCAAGAAAACGGCGTCCCTCGGGAGACAATTCTTGGGGAGCCC
TTTCACTATCTCAGCCGATGGAAACACCTTCGCCGTCTTAGGCTGCAGAAACTCCATCTTCTTGA
AAATCAATGGGACGGGACTCAGCGGGTGCTTCCCCGCTTGTGGAAAGAATTACACGCAAGACT
CTTGCCAAATAGATATTCCGCCAAGATCCCAGGAGCTTATTTATACATACCAAAGCACCACTGAT
ACTCAACCAGGCAATACCACGCAATACTGTGGTTATGCGTTTCCTGTGGAGAGGGAGTCCCTAA
GCAAGACTTATGACTTGTATAGAGGCCTCAGCGATGATTATTTGAACCCATTCAACGAGCAACTT
ACGCATGCGCCGCTGGTGCTTGATTGGGAGTTAACACACATAGATATCAGCCCCTACTGGGAAC
CGCCTCCCAATACTCTCTGCGCAACTCGATTTAGCGATCTTTCCCTGTCTACAACTAAGAAATGT
TCTTGTTCCTATGGTTTTCGAGGAAATCCATATTTAGATGCAGGCTGTGTTGATATCAACGAGTG
CGAGGAGGATCCAAGCTTATGCGGTGCTACTGGGGCGACTTGTGTCAATGAAATTGGCTCTTTC
ACGTGTCATGATCAGTATTCAAGCCATCGAGTAAGAAACATATTGCTAGTTACTTTCGGCAGCCT
CTTCGTTGCAGGGATTTTCATTCCGTGCCTGTCAAAAGTCATCTTGAAAGGACTGAAGGCCCGC
CGAAGAAGAAAATTCTTCGTGCGTAACGGCGGTCTGCTCCTCGAACAAAAGCTGTCCTCGATCG
ATAACGACTACAAGAAATCCAAGCTGTTCACATCCAAGGAGTTGAAGCAAGCCACCAATCATTA
CAGCGAGAATCGGGTGCTTGGTAGGGGCGGCCAAGGGACGGTTTACAAGGGAATGCTGACAGA
TGGGAGCATAGTAGCCGTGAAGAAATCCAAAACGGTCCAAGAAACCGACGTCGAGTCCTTCGT
AAACGAGGTTGTCATTCTGTCCCAAATCATCCATCGAAGCATAGTCAGGCTCCTCGGATGCTGT
CTGGAGACTGAAACCCCAATCCTCGTATACGAATTTGTTCCTAACGGCACGCTATTCGAGCACAT
CCATGACCGAAGCGAAGATTTCCCTCTTACCTGGCAAATGAGGCTTCGAATCGTTGTAGAAATA
GCCAGCGCCCTTTCTTATCTACACTCCTATGCTTCAGCCCCTATCTTCCACAGGGATATCAAGTC
GACAAACATACTGTTGGACGAGAAATACAGAGCGAAGGTCTCAGATTTTGGTACCTCAAGGTCG
CTCGCCATCGATCAAACTCATTTCACAACTAGGGTTTGCGGCACCTACGGCTACTTGGACCCAG
AGTACTTCCAATCGAATCAGTTCACGGAGAAGAGTGACGTCTACAGCTTCGGCGTCGTCATGGT
GGAGCTCCTCACCGGCGAGATAGCCGTGTCGTTGCTGAGGGCGGGGACGCGGAGGAGCTTAG
CTACGCATTTCTTACACTCGATGGAAGAGGGCAAGCTGTTCGAAATCGTTGATCCGAGAATCGT
GGAAGGGGGGGAGCGCGAGAGGGAAGAAGAGATCACGATGGTGGCTGAACTCGCCCGACGGT
GCCTCCAATTGAAAGGCAGTTTACGGCCGACCATGAGGCAAATCGCCAACGAATTGGAGAGCG
TAATCCACATCAAAAGTGCTGAGCAGAGCCATGATCGCGATGAAGTTGAGCTTTCTGTGATTGA
TTTCTCATCAATTTCACCTTTCGCAATCGCATCCAAT<u>TAG</u>GTTCACGATTTCTCCTTTTGTGTGATAT
TTTCTGATTATATTGACTTACGGAGGTATGGAGCGTGACATCACGCTTATATCATAACTTCGCCGAATGC
TCGAGGTAATATTAGTGGATTATCGGTTATGGATATTGAAGATGATGTCATGACTTTTCACTTTGCATGA
AAATGCCCGACCTTGATGCGCTTGATGGATATTTCTGTTTATGTACCTCTTTTAAAACGTTTGCTTCGTT
AGCGAAGATCAAATAATTCGTGTAAAAAATGATACTTTTGTAGCAACAGCATATGAGGAGAGAAGTAC
GTAAATAGTGTTATTTTGCTAAATAAAATTACAACACTAGTTGAGTTACTAAAACTATTACAAAACTAAT
CAAGTTTTTACTTTTTAACTAGTATTCATTTCATTTTACACCACAAAACGTTTTAAAAT

CpWAK2 (Genomic sequence):

ATCTCAAAGTTTTAGACGTGCATAGGGTGCTTATTTAAAGGAACTGAAAAGGACGTATTTGGGATCTTA
GGACTTGTTTAGGACATGTTTGGGACGTTGGATCATGTTTGGGGATGTGTCTGGGACGTGTTTGGGAC
GTGTAGAGACATGTCGGGACGTCTGGGACGTGTTGGGACGTAAATAAAGGGACACAAAGAATTACGT

GATTCGGATCTTTGCAGATCCTACGTCCACGGGCGCTCTAAAATCTCTTTTCACTATACTTAACAAATTA
CACACTCGAACTCGCTTTCACTATAATATCTAAGTATCGCGAGGAAAGTGCTTAACCAACTGAAAAGA
GAGAAGTTAGTGAGTATTTTCGTATGTGGAATGAATAATCTCGTGAAAATGGTTAAGTTCGGCTTTTAA
AGGAGCTTCCGGATATTGCCTGGAGGTACGTCCCTGGCTCCTGAATATTGCCCAGATTCCTGCTGATAT
GACCtcCTGGGACACTGTGATGCTACAGAATGGCAGGCCGTCTGTCAGAGGTCACTCCTGCAACAATTA
CAGCTTCTGTGCTGTCAGGGACGTCAGGCGGCTGGAGATGGGACTTGGCTCTGCGCTGTCAGGTACA
ACAGGTGGATGGGACATGTTTTGTCAGGACAGCTTTGCAGTGTTGTGAGACGTCCTTTCAGACTTCTG
ACCANTTGGTGAAGCTTCCTCACGACGCTTCCTCGGGACGACAGGACCTGCTTAGCACAATATGTCCC
AAAATAGTTCCCTAATATGGGCGTTAGGTCCCTAAAGTCTCTTGGAGGTCCCTAAAATCCTTATTTCCTG
TTGTTAGCTCCCTAAAGTGTGTTGGAGATCCATAAAACCCGTCTGAGGTCCCTAAAAGTCTCCCTTCAT
AATCTAACTTTGTATTTTTCAAGGAAACTCATATTACTAAGTTCTCTTAATAATAATTATCATAAGCTAATT
ATTTCCTTAATTCCTCTTTTTAGCCTGCCTAATTATTGTTCCACATCATTCATATTCACGATCGTGAAAAG
TTCGAGCTTAGAAAATGGTTCTCTGCCCGCATTAACGGCCTCTTGCACCTCCTTTCACGGCCGTGAAG
GACTTGGTTGGTTCATGTTCACGGCCGCAAAGTGATTGGGCAAGAAAATGGCTTTCTGGTTGCCTTCA
CGGTCTCTTGCACAACCTTCACGATCGTGAAGCTATTTTCTGCAAAATAACAGTAAGTTCAAAATTTCA
TATTCCCAATCTAGACATTAAGATTGAAAATCCCTTTCTTCTATATCATTGAGAATTACGAAACTGTGGC
AAAACTGCCGTAAAAATGTTTGATTTTGCAGTAAAATAAAATAGAGCGCTATGCAAACGACTTAAATTT
AACAACTATATACATAAGAAATTACGCTTATCACATGGCCAAGTAACACGTTCCAAATAATACATTCTAC
AATTAATAAATAGACTGAACTTGATCATCACTAAATAGACTGGACCCACCATCATAATGTGGGTTATAGC
GAATTGAATACGTGGAAAGAGTCCGACGGAGTACTATCAATCCAACTGCTCGAAGTCCTCAGAAATCA
AGTCGTTTTTACCTACTTACATTTCTCCATATCCAATCTAATTGCACTATCCTCGCCTCGCTCTCTTATTCT
CACGTGACATAAGGAAAAAAAAAAAGCAAATATATCACCAATCTTATTAGTACTCAAGTAAGAGCGGA
GATCGAAGAATCAAACAGGTACGTATACGTCCATTATGGTAACAACTCCAACATGCAGAACTATCATAA
AAACAATGTATAAGCATGAAATGAAAAAAATAAATAAATTGATATACCGTTTGGCAACTTAATTCTGTG
AGCTCATCAAGTTCGTATTACAGCAGGCCAACACGGAA**ATG**AGGACTTCTTCAATACAAGCCGCAG
TCGTTTTCTGCGTCTGCCtTTGCACAACTTCGGCTATACCATCATCGTCATCATCATCATCATCAT
CCATTTTCATACCCATGGCGAAGCCTAATTGTAGCCAAACATGCGGAAATATTACCGTTCCATATC
CATTCGGCATGAGCGCAAGGTGCTGTGCAGACTCATCTTTTCTCGTCGAGTGCCTAAACTCCAC
AAATCCTCCGACGCTGTACTTGCCAAGACTCGACCTGCAAGTAACTGACATTCGCGTGAACGGC
ACAATCGTCGTAAAATATCCCGTCACTCCAATCAAATGTACCGCTGTCAAGAAAATGGAGTCCCT
CGGGATAAAATTATTGGGGAGCTCTTTCACTATCTCAGCCGATGAAAACACCTTCACCGTATTAG
GCTGCAGAAACTCCATCTTCTTGCAAATCAATGGGACGGGATACAGCGGATGTTTCCCAGCTTG
TGGAGTGGATTACACGCAAGATTCTTGCCAAATAGATATTCCGCCAAGATCCCAGGAGCTTATTT
ATACATACCAAAGCACCACTTATACTCAACCAGGCAATACCACGCAATACTGTGGTTATGCGTTT
CCTGTGGAGAGGGCGTCGCTAAGCAAGACTTATGAGTTGTATAAAGGCCTTAGAGATGATTATTT
TAACCCATTCGACGAGCAACTTACGCATGCGCCGCTGGTGCTTGATTGGGAGTTAACATACATA
GATACCAACGAATACGCTCGATGTAGTGATTATATCAGCAACAACAGGGAACTGCGTCCCAATAC
TCTTTGCGTAAGTCCATATTACTGGGGCAATCTTTACCATTATACAACTAAGAAATGTTCTTGTTG
CTATGGTTTTCGAGGAAATCCTTATTTAGATGGAGGCTGTGTTGATATCAACGAGTGCGAGGAG
GATCCAAGCAGATGCGCTGCTCCTGGGGTGACTTGTGTCAATGAAATCGGCTCTTCCACATGTC
ATTATCAGTATTCAAGCCATCGAGTAAGAAACATATTGCTAGTTACTTTCGGCAGCCTCTTCGTT
GCAGGGATTTTCATTCCGTGCCTGTCAAAAGTCATCTTGAAAAGACTGAAGGCTCGCCGAAGAA
GAAAATTCTTCGTGCGTAACGGCGGTCTGCTCCTCGAACAAAAGCTGTCCTCGATCGACAACGA
CTACAAGAAATCCAAGCTGTTCACATCCGAGGAGTTGAAGCAAGCCACCAATCATTACAGCGAG

AATCGGGTGCTTGGTAGGGGCGGCCAAGGGACGGTTTACAAGGGAATGCTGACAGATGGGAGC
ATAGTAGCCGTGAAGAAATCCAAAACGGTCCAAGAAACCGACGTCGAGTCCTTCGTCAACGAG
GTTGTCATTCTGTCCCAAATCATCCATCGTAGCATAGTCAGGCTCCTCGGATGCTGTCTGGAGAC
TGAAACCCCAATCCTCGTATACGAATTTGTTCCTAACGGCACGCTATTCGAGCACATCCATGACC
GAAGCGAAGATTTCCCTCTTACCTGGCAAATGAGGCTTCGAATCGTTGTAGAAATAGCCAGCGC
CCTTTCTTATCTACACTCCTATGCTTCAGCCCCTATCTTCCACAGGGATATCAAGTCGACCAACAT
ACTGTTGGACGAGAAATACAGAGTGAAGGTCTCAGATTTTGGTACCTCAAGGTCGCTCGCCATC
GATCAAACTCATTTCACAACTAGGGTTTGCGGCACCTACGGCTACTTGGACCCAGAGTACTTCC
AATCGAATCAGTTCACGGAGAAGAGCGACGTCTACAGCTTCGGCGTCGTCATGGTGGAGCTCC
TCACCAGCGAGATAGTCGTGTCGTTGCTGAGGGCGGGGACGCGGAGGAGCTTAGCTACGCATT
TCTTACACTCGATGGAAGAGGGCAAGCTGTTCGAAATCGTTGATCCGAGAATCATGGAAGGGG
GGGAGCGCGAGAGGGAAGAAGAGATCACGATGGTGGCTGAACTCGCCCGACGGTGCCTCCAA
TTGAAAGGCAGTTTACGGCCGACCATGAGGCAAATCGCCAACGAGTTGGAGAGCGTAATCCATA
TCAAGAGGGCTGAGCAGAGCCATGATCGTGATGAAGTAGAGCTTTCTGTAACTGATTTCTCATC
AATTTCACCTTTCGCAATCGCATCCAAT<u>TAG</u>GTTCACGATTTCTCCTCTTGTGTGATATTTTCTGATTA
TATTGACTTACGGAGGTATGGAGCGTGACATCACGCTTATATCATAACTTCGCCGAATGCTCGAAGTAAT
ATTAGTGGATTATCGGTTATGGATATTGAAGATGATGTCATGACTTTTCACTTTGCATGAAAATGCCCGA
CCTTGATGCGCTTGATGGATATTTCTGTTTATGTATACCCCTCTTTTAAAACGTTTGCTTCGTTAGCGAAG
ATCAAATAATTCGTCTAAAAAATGATACTTCTGTAGCAACAGCATATGAGGAGAGAATTACGTAAATAG
TGTTGTTTTGCTAAATAAAATTACAACATTAGTTGAGTTACTAAAACTATTACAAAACTAATCAAGTTTT
TACTTTTTAACTAGTATTCTTTTCATTTTACACCACAAAACGTTTTAAAATTACAATTCACTACAATTATT
ATTATAAGTTAGTAAACAATTACTAATATATATTTCACCACAATTCATTTTTTATTACGAAAAAATAACCTA
TATAGAATACATTCGAAATATTATTTTAGAATACATTCAAAAATAACCTACTTTGACTTTAATATATGAAA
GCAATATTAACGCATAAATAAATAACAATAAAAGAAAATAATAATCCTCAAATCAAAATTTTCAAATAAT
ATGTTGTCTAACAAATATTAATTTCTGAAAAAGTTTCACAACTACTAACATTACCCTTGAACGTTTCTTG
TAAGACTTTCGATCTATTTCATCTGAATATCATCCTATTCGATTACGTCCAACTTTTATCATTGACGCAAA
TAAGGAGCGTAATGAAAAACAGGCTTTCGTCAAGACCCCTCTAATGAAAACTCAAACTGGTAAGAAT
GTTTTGTCTCCATTATTTCAATTCTAAATCATTTGAAACACATTAACATGGATCAACTTTCAAATACTTGC
AAACCTAAAGAAAATAGAGCAATCCAACAACAACGACAACAACATTAACAACGACAACAACAAAAA
CAATAATAATAATATACTTGCATTGCATGTGACGTGACATCTTGTATACTAATCACTTACCACGTAAATGA
ATGATATGTTTGACTAGCATGATTAACTCGAAAGGGACGATATGGTTTACCATGAGCAAGAACACTTGT
TGTCCAAGTAACTATGAGATTGTTTGATTTTTTATTTTTTTGTTGATTTAATTAATTGATGAATACAATAGT
AAAAGCATGTTTTTAAGAATAAATTATTGATTTATTTAATCACATGGGAATGTAGGCTATATTACATGTTA
GTTACTTTCTTTTTATTTATTTCATAATTAATATATTTATATTAAAATGGTGTTGGGTAAAAATATAATTATT
AACAAGTTTTATGTTGGTTTAAGATAATAATAGTGAAAATTTGTAATTTTATGTATTTTGTGGTGGAAATG
AAAATAATAATGAGAGTTAAAATTCAAAAAGTTCATTAGTTTTGTAATAGTTTTGAAAACTCGACTAGT
TTtGTAATTTAATTATCAAAATAATACTATTTACGTAATTCTCTCTACAGTTAAGCTGCATAAAATACAAG
AAGGCTGAAAATGCATCTGGCCTGGGAGCCCCATTTTACAGCTCTTTTAAAATTGCTGCCATTTTATTTT
CTGCTTTATGTTTTCCCATGCAATAATATCCTCCGGCGATGATATATATTCGTCGCAAAGACTTCTTCCCC
TCAGG

CpWAK3 (Genomic sequence):

ACTATAGGGCACGCGTGGTCGACGGCCGGGCTGGTACCATTCACGGGCGTGACACATCCTTCACGGCT

TGCTTTCTTCATATTCACGGTCGTGAAAAGTTCGAGCTTAGAAAATGGTTCTCTGCCCGCATTAACGGC
CTCTTGCACCTCCTTTCACGGCCGTGAAGGACTTGGTTGGTTCATGTTCACGGTCGCAAAGTGATTGG
GCAAGAAAATGGCTTTCTGGTTGCCTTCACGGTCTCTTGCACAACCTTCGCGATCGTGAAGCTAAAAA
TAACAGTAAGTTCAAAAATTCATATTCCCAATCTAGACATTGGATTGAAAATCCCTTTAGTCAGGATCAA
TACTAGCGAGTCTTCTATATCATTGAGAATTACGAAACTGTGGCAAAACTGCCGTAAAAACGTTTGATT
TTGCAGTGAAATAAAAATATAGCGCTATGCAAACGACTTAAATTTAACAACTATAAACATAAGAAATTA
CGTTTATCACATGGCCAAGTAACACGTTCCAAAATAATACATTCTACAATTAATAAATAGACTGAACTTG
ATCATCACTAAATAGACTGGACCCACCATCATAATGTGGGTTATAGCGAATTGAATACGTGGAAAGAGT
CCGACGGAGTACTATCAATCCAACTGCTCGAAGTCCTCAGAAATCAAGTCGTTTTTACCTATTTATATTT
CTCCATATCCAATCTAATTGCACTATCCTCGTCTCGCTCTCTTATTCTGTCGTGACATAAGGATAAAAGA
AAAACAAATATATCACCAATCTTATTAGTACTCAAGTAAGAGCGGAGATCGAAGAATCAAACAGGTAC
GTATACGTCCATTATGGTAACAACTCCAACTTGCAGAAATATCATAAAAACAATGTATAAGCATAAAATG
AAAAAAAAAAAATGAAAAAAAATTGATATACCGTTTGCAACTGAATTCTGTGAGCTCATCAAGTTCGTA
TTACAACAGGCCAACACGGAA**ATG**AGGACTTCTTCAATACTAGCTGCAGTCGTTTTCTGCGTCTG
CCTTAGCACAACTTCGGCTATACCATCATCATCCATTTCCATACCCATGGCGAAGCCTAATTGTAG
CGAAACATGCGGAAATATTACCGTTCCATATCCATTCGGCATGAGCTCAGAGTGCAGTATCGGCT
TACATTTTCTCGTCGAGTGCCGAAACTCCACAAATCCTCCGACGCTGTTCTTGCCCAGCATCGA
CCTGCAAGTAACTGACATTCGCGTGAACGGCACAATCGTCGTAAAATATCCCGTCACTCCAATC
AACTGTACCGCTGTCAAGAAAAAGGAGTCCCTCGGGGGAACATTATCGGGGAGCTCTTTCACTA
TCTCAGCCGATGGAAACACCTTCACCGTATTTGGCTGCAGAAACTCCATCTTCTTGCAAATCAAT
GGGACGGGATACAGCGGATGCTTCCCAGCTTGTGGAGTGGCTTACATGCAAGACTCTTGCCAAA
TAGATATTCCCCCAAGACCCCTGGAGCTTATTTATACATACCAAAGCACCACTGATACTCAACCA
GGCAATGCCACGCAATACTGTGGTTATGCGTTTCCTGTGGAGAGGGAGTCCCTAAGCAAGACTT
ATGACTTGTATAGAGACCTCAGCGACGATGATTATTTGAACCCATTCGACGAGCAACTTACGCAT
GCGCCGCTGGTGCTTGATTGGGAGTTAACAGACGCAGATACCAACGAATACGCTCGATGTAATG
GTTATGAtGTCAGAGACTACAGGGAACTGCCTCCCAATACTCTCTGCTCAAGTCGATATAGCGTG
AGCGATCTTTCCCTGTATGCAACTAGGCAATGTTCTTGTTGCTATGGTTTTCGAGGAAATCCATA
TTTAGATGGAGGCTGTGTTGATATCAACGAGTGCGAGGAGGATCCAAGCATATGCGGTGCTACT
GGGGCGACTTGTGTCAATGAAATCGGCTCTTCCACGTGTCATTATCAGTATTCAAGCCATCGAGT
AAGAAACATATTGCTAGTTACTTTCGGCAGCCTCTTCGTTGCAGGGATTTTCATTCCGTGCCTGT
CAAAAGTCATCTTGAAAGGACTGAAGGCCCGCCGACGAAGAAAATTCTTCGTGCGTAACGGCG
GTCTGCTCCTCGAACAAAAGCTGTCCTCGATCGACAACGACTACAAGAAATCCAAGCTGTTCAC
ATCCAAGGAGTTGAAGCAAGCCACCAATCATTACAGCGAGAATCGGGTGCTTGGTAGGGGCGG
CCAAGGGACGGTTTACaAGGGAATGCTGAcAGATGGGAGCATAGTAGCCGTGAAGAAATCCAAA
ACGGTCCAAGAAACCGACGTCGAGTCCTTCGTAAACCACCAATCCTCGTATACGAATTTGTTCC
TAACGGCACGCTATTCGAGCACATCCATGACCGAAGCGAAGATTTCCCTCTTACCTGGCAAATG
AGGCTTCGAATCGTTGTAGAAATAGCCAGCGCCCTTTCTTATCTACACTCCTATGCTTCAGCCCC
TATCTTCCACAGGGATATCAAGTCGACAAACATACTGTTGGACGAGAAATACAGAGCGAAGGTC
TCAGATTTTGGTACCTCAAGGTCGCTCGCCATCGATCAAACTCATTTCACAACTAGGGTTTGCG
GCACCTACGGCTACTTGGACCCAGAGTACTTCCAATCGAATCAGTTCACGGAGAAGAGCGACGT
CTACAGCTTCGGCGTCGTCATGGTGGAACTCCTCACCGGCGAGATAGCCGTGTCGTTGCTGAG
GGCGGGGACGCGGAGGAGCTTAGCTACGCATTTCTTACACTCGATGGAAGAGGGCAAGCTGTT
CGAAATCGTTGATCCGAGAATCGTGGAAGGGGGGGAGCGCGAGAGGGAAGAAGAGATCACGAT
GGTGGCTGAACTCGCCCGACGGTGCCTCCAATTGAAAGGCAGTTTACGGCCGACCATGAGGCA

AATCGCCAACGAGTTGGAGAGCGTAATCCATATCAAGAGGGCTGAGCAGAGCCACGATCGTGAT
GAAGTAGAGCTTTCTGTAACTGATTTCTCATCAATTTCACCTTTCGCAATCGCATCCAAT<u>TAG</u>GTT
CACGATTTCTCCTTTTGTGTGATATTTTCTGATTATATTGACTTACGGAGGTATGGAGCGTGACATCACG
CTTATATCATAACTTCGCCGAATGCTCGAAGCAATATTAGTGGATTATCGGtTATGGATATtGAAGATGATG
TCATGACTTTTCACTTTGCATGAAAATGCCCGACCTTGATGCGCTTGATGGATATTTCTGTTTATGTATAC
CCCTCTTTT

CpGLP1 (mRNA sequence):

<u>ATG</u>GCGGCGAACGGTAAAACTTCATCGATGTTCTTTCTGATCATTACAGCAAGCCTGCTGGCGA
TGACGAGGACGATCTTCGCATACGATCCCAGCCCATTGCAAGACATATGCGTAGCAGATCTAAA
CTCAACTGCAGTAAGAGTCAACGGCTTGCCATGCAAAGATCCTTCGACGGTCAAAGCCGACGA
CTTCTTCTTCTCAGGCATGGATAAACCTGGGAACACAACTAACCCTATCAAAGCCACCTTCAGC
CCAGTCAACGTAAGGCAAGTTCCAGGAGCTAACACGCTGGGGCTCACCATAGCTCGCCTGGAC
TTCGCCGCAGGGGGGTTTCTCCCGCCGCACTTCCACCCAAGGGCGTCGGAGTTCTTGATGGTT
CTAAAAGGTTCCATGGAAGTAGGGATGGTCATACCTAGCCAAGGGTACAAGCTTCTGAACAAAA
CCCTAAACAAGGGCGATGCTTTCGTCGTCCCCGTCGGCTTGGTTCATTATCAGAGGAACAAGGC
GGCGAAGGGGCGGAGCACGGTGGTCTTTGCGGCGCTCAACAGTCAGAATCCCGGACTAACCGT
ACTTGCCAACAGCGTCTTTGGGGCGACGCCCGAGATCGACGGCGGTTTGCTCGCCGAGGCTTT
TCGGTTGGATGAGAAGACTGTTCAAGGGCTTCAAGCTGCCTTT<u>TAA</u>TAAAAGATGATTATCATTC

CpCRP1 (mRNA sequence):

GTCATAACAATATATCACATTAAGCCTTTGAGACCCACG**<u>ATG</u>GCTCAAAGTAGCACTTCCAAAGTTG**
CCGTAGTTCTTCTTCTCCTCCTCTCCCTAGCGGCAGCTCTCTCCGCCGCCCACGAAATGGAAAA
CGTCGCCTCCCTCACCGTCGAGCCATCAGCCAAGCTGTCGTCTGAGGTCCTCGTGGTCAGCGA
CGCCTTCTCTTCCGCCGGGATAGAGAAGTTCCTAGAATACATCAAGCGCTCCGCCGACGTCGTC
GAAGAGGTGGCGGCGGAGATCTGCAAAAAGGCGGAGTCGTGCCCCTGCAGGGACGGCAGCTT
TGTGTGTCGGGCTTGCTGCAGCATAAAGCGCCAAGTTCCGCCGCCGAGAGCGCCGGCGAACAG
GCTGAAGATCCACCAGTTCATACAGAACTGCAAGAGATTCCGGACGGAATGTCAGCAGCCGCTC
GGCGAGATCTGTCGGAGGGCCGCCATTGGATGCCCTTGCAGAGCTGGTAACGACCTCAAGTGC
CTCGTTTGCTGCGGTATAAGAGGC<u>TGA</u>AGGTCGCCGTCGCCGTCGTGAAGGCGGCGGCGTCACGG
TGAGAAGTCGTTCTGTGACGACTTTTATTACTATTTAACTGCAATTTGTGTTGGTTTTGCTTTCCCACTG
TACTGAGTTTGCTGTGCTGGTAAACTCATCGCTAGTGTTGTCTCGGGTATTAATTATGTAATCCTTGGAA
TAAGACAAAGAAATTTGCAAT

Lnc RNA 28852 (mRNA sequence):

GAGAATATTGAGTGCTTGTGAGTGCATGTGCGTGTGTGTATATATAAAAAATAATATGTGTGGGTCTTGT
GTTGTCATCTAGTGTCCATCAAAATCCTTGTGAGTGTGATTATAATAGGGTGAGTTTTAGTGATTGTATA
CGCGAGGTGTCTATCATAATAATTGAGTGGTGCAATAGTTATAAAAATAGTTTATTATAATTATTACTTGT
GCCATTCAATTATTCATGTGCTATTCAATTATTCAACAATATATGTTTTATGTAGTATTCGACTGACGTTGT
AGTGTAAGACCTTGATCAACGCAATGACAATTCGTCCGTGAAAGAGGAGAATATAGACAGACAGAAA
TTATATATTATATATATCATTTTTGTCATACATTTGAACGCTCCGTTGCTGCAGTACAATTCAACGGACAG

TCATTGACAATATCTATCAAAGGGATAGCGCACCGATCACAGTACCAATCCCCAATGGCCTGAGCAATC
GTCTGCAAAGAAGAAGAAAATATTTTTATAAATGAAAATTTTCATTAGAAAATACAAGAATAGATATATA
CCAAATTTTCAGGGTAGGTGATCCCTTCCAATTATATAATTGATTCCAGTAAATGAAATCATGGTCTTAG
C

Supplementary figure 1 Overview of nucleic acids sequences. The coding sequences were written in bold and the translation start codon and stop codon were underlined.

CpWAK1:

MRTSSILAAVVFCVCLSTTSAIPSSSISIPMAKPNCSETCGNITVPYPFGMSSECSIDLDFLVECRNSTNPPTLF
LSSIDLQVTDIRVNGTIVVKYPVTPINCTAVKKTASLGRQFLGSPFTISADGNTFAVLGCRNSIFLKINGTGLS
GCFPACGKNYTQDSCQIDIPPRSQELIYTYQSTTDTQPGNTTQYCGYAFPVERESLSKTYDLYRGLSDDYLN
PFNEQLTHAPLVLDWELTHIDISPYWEPPPNTLCATRFSDLSLSTTKKCSCSYGFRGNPYLDAGCVDINECEE
DPSLCGATGATCVNEIGSFTCHDQYSSHRVRNILLVTFGSLFVAGIFIPCLSKVILKGLKARRRRKFFVRNGG
LLLEQKLSSIDNDYKKSKLFTSKELKQATNHYSENRVLGRGGQGTVYKGMLTDGSIVAVKKSKTVQETDV
ESFVNEVVILSQIIHRSIVRLLGCCLETETPILVYEFVPNGTLFEHIHDRSEDFPLTWQMRLRIVVEIASALSYL
HSYASAPIFHRDIKSTNILLDEKYRAKVSDFGTSRSLAIDQTHFTTRVCGTYGYLDPEYFQSNQFTEKSDVY
SFGVVMVELLTGEIAVSLLRAGTRRSLATHFLHSMEEGKLFEIVDPRIVEGGEREREEVITMVAELARRCLQ
LKGSLRPTMRQIANELESVIHIKSAEQSHDRDEVELSVIDFSSISPFAIASN

CpWAK2:

MRTSSIQAAVVFCVCLCTTSAIPSSSSSSSSSIFIPMAKPNCSQTCGNITVPYPFGMSARCCADSSFLVECLNS
TNPPTLYLPRLDLQVTDIRVNGTIVVKYPVTPIKCTAVKKMESLGIKLLGSSFTISADENTFTVLGCRNSIFLQ
INGTGYSGCFPACGVDYTQDSCQIDIPPRSQELIYTYQSTTYTQPGNTTQYCGYAFPVERASLSKTYELYKG
LRDDYFNPFDEQLTHAPLVLDWELTYIDTNEYARCSDYISNNRELRPNTLCVSPYYWGNLYHYTTKKCSCC
YGFRGNPYLDGGCVDINECEEDPSRCAAPGVTCVNEIGSSTCHYQYSSHRVRNILLVTFGSLFVAGIFIPCLS
KVILKRLKARRRRKFFVRNGGLLLEQKLSSIDNDYKKSKLFTSEELKQATNHYSENRVLGRGGQGTVYKG
MPTDGSIVAVKKSKTVQETDVESFVNEVVILSQIIHRSIVRLLGCCLETETPILVYEFVPNGTLFEHIHDRSED
FPLTWQMRLRIVVEIASALSYLHSYASAPIFHRDIKSTNILLDEKYRVKVSDFGTSRSLAIDQTHFTTRVCGT
YGYLDPEYFQSNQFTEKSDVYSFGVVMVELLTSEIVVSLLRAGTRRSLATHFLHSMEEGKLFEIVDPRIMEG
GEREREEEITMVAELARRCLQLKGSLRPTMRQIANELESVIHIKRAEQSHDRDEVELSVTDFSSISPFAIASN

CpWAK3:

MRTSSILAAVVFCVCLSTTSAIPSSSISIPMAKPNCSETCGNITVPYPFGMSSECSIGLHFLVECRNSTNPPTLF
LPSIDLQVTDIRVNGTIVVKYPVTPINCTAVKKKESLGGTLSGSSFTISADGNTFTVFGCRNSIFLQINGTGYS
GCFPACGVAYMQDSCQIDIPPRPLELIYTYQSTTDTQPGNATQYCGYAFPVERESLSKTYDLYRDLSDDDYL
NPFDEQLTHAPLVLDWELTDADTNEYARCNGYDVRDYRELPPNTLCSSRYSVSDLSLYATRQCSCCYGFRG
NPYLDGGCVDINECEEDPSICGATGATCVNEIGSSTCHYQYSSHRVRNILLVTFGSLFVAGIFIPCLSKVILKG
LKARRRRKFFVRNGGLLLEQKLSSIDNDYKKSKLFTSKELKQATNHYSENRVLGRGGQGTVYKGMLTDG
SIVAVKKSKTVQETDVESFVNEVVILSQIIHRSIVRLLGCCLETETPILVYEFVPNGTLFEHIHDRSEDFPLTW
QMRLRIVVEIASALSYLHSYASAPIFHRDIKSTNILLDEKYRAKVSDFGTSRSLAIDQTHFTTRVCGTYGYLD
PEYFQSNQFTEKSDVYSFGVVMVELLTGEIAVSLLRAGTRRSLATHFLHSMEEGKLFEIVDPRIVEGGERER
EEEITMVAELARRCLQLKGSLRPTMRQIANELESVIHIKRAEQSHDRDEVELSVTDFSSISPFAIASN

CpWAK1-EX:

MAKPNCSETCGNITVPYPFGMSSECSIDLDFLVECRNSTNPPTLFLSSIDLQVTDIRVNG
TIVVKYPVTPINCTAVKKTASLGRQFLGSPFTISADGNTFAVLGCRNSIFLKINGTGLSG
CFPACGKNYTQDSCQIDIPPRSQELIYTYQSTTDTQPGNTTQYCGYAFPVERESLSKTY
DLYRGLSDDYLNPFNEQLTHAPLVLDWELTHIDISPYWEPPPNTLCATRFSDLSLSTTK
KC SCSYGFRGNPYLDAGCVDINECEEDPSLCGATGATCVNEIGSFTCLEHHHHHH*

CpWAK1-R-2:

MAKPNCSETCGNITVPYPFGMSSECSIDLDFLVECRNSTNPPTLFLSSIDLQVTDIRVNG
TIVVKYPVTPINCTAVKKTASLGRQFLGSPFTISADGNTFAVLGCRNSIFLKINGTGLSG
CFPACGKNYTLEHHHHHH*

CpWAK1-R-3:

MDQDSCQIDIPPRSQELIYTYQSTTDTQPGNTTQYCGYAFPVERESLSKTYDLYRGLSD
DYLNPFNEQLTHAPLVLDWELTHIDISPYWEPPPNTLCATRFSDLSLSTTKKCSCSYGF
RG NPYLDAGCVDINECEEDPSLCGATGATCVNEIGSFTCLEHHHHHH*

CpWAK1-R-4:

MAKPNCSETCGNITVPYPFGMSSECSIDLDFLVECRNSTNPPTLFLSSIDLQVTDIRVNG
TIVVKYPVTPINCTAVKKTASLGRQFLGSPFTISADGNTFAVLGCRNSIFLKINGTGLSGC
FPACGKNYTQDSCQIDIPPRSQELIYTYQSTTDTQPGNTTQYCGYAFPVERESLSKTYD
LYRGLSDDYLLEHHHHHH*

CpWAK2-EX:

MAKPNCSQTCGNITVPYPFGMSARCCADSSFLVECLNSTNPPTLYLPRLDLQVTDIR
VNGTIVVKYPVTPIKCTAVKKMESLGIKLLGSSFTISADENTFTVLGCRNSIFLQING
TGYSGCFPACGVDYTQDSCQIDIPPRSQELIYTYQSTTYTQPGNTTQYCGYAFPVER
ASLSKTYELYKGLRDDYFNPFDEQLTHAPLVLDWELTYIDTNEYARCSDYISNNRE
LRPNTLCVSPYYWGNLYHYTTKKCSCCYGFRGNPYLDGGCVDINECEEDPSRCAA
PGVTCVNEIGSSTCLEHHHHHH*

CpGLP1:

MAANGKTSSMFFLIITASLLAMTRTIFAYDPSPLQDICVADLNSTAVRVNGLPCKDPSTVKAD
DFFFSGMDKPGNTTNPIKATFSPVNVRQVPGANTLGLTIARLDFAAGGFLPPHFHPRASEFLM
VLKGSMEVGMVIPSQGYKLLNKTLNKGDAFVVPVGLVHYQRNKAAKGRSTVVFAALNSQN
PGLTVLANSVFGATPEIDGGLLAEAFRLDEKTVQGLQAAF

CpCRP1:

MAQSSTSKVAVVLLLLLSLAAALSAAHEMENVASLTVEPSAKLSSEVLVVSDAFSSAGIEKFL
EYIKRSADVVEEVAAEICKKAESCPCRDGSFVCRACCSIKRQVPPPRAPANRLKIHQFIQNCK
RFRTECQQPLGEIC RRAAIGCPCRAGNDLKCLVCCGIRG

Supplementary figure 2 Overview of protein sequences

```
CpWAK1      ATCGAAGAGAAGAGGTGCCTGCGTTCACGGCTTCGGCAGCTACCATTCACGGGCGTGACA   -959
CpWAK2      CTCTTAATAATAATTATCATAAGCTAATTATTTCCTTAATTCCT----CTTTTTAGCCTG   -930
CpWAK3      CTAT--AGGgCACGCGTGGTCGACGGCCGGGCTG-GTACCATTC----ACGGGCGTGACA   -949
              *                 *           *    *
```

```
CpWAK1      CATCCTTCACGGCTTGCTTTCTTCATATTCACGGTCGTGAAATGTTCGAGCTTAGAAAAT   -898
CpWAK2      CCTAATTATTGTTCCACATCATTCATATTCACGATCGTGAAAAGTTCGAGCTTAGAAAAT   -869
CpWAK3      CATCCTTCACGGCTTGCTTTCTTCATATTCACGGTCGTGAAAAGTTCGAGCTTAGAAAAT   -888
            * *  **   *     * *  ********** ******** ****************

CpWAK1      GGTTCTCTGCCCGCATTAACGGCCTCTTGCACCTCCTTTCACGGCCGTGAAGGACTTGGT   -837
CpWAK2      GGTTCTCTGCCCGCATTAACGGCCTCTTGCACCTCCTTTCACGGCCGTGAAGGACTTGGT   -808
CpWAK3      GGTTCTCTGCCCGCATTAACGGCCTCTTGCACCTCCTTTCACGGCCGTGAAGGACTTGGT   -827
            ************************************************************

CpWAK1      TGGTTCATGTTCACGGCCGCAAAGTGATTGGGCAAGAAAATGGCTTTCTGGTTGCCTTCA   -776
CpWAK2      TGGTTCATGTTCACGGCCGCAAAGTGATTGGGCAAGAAAATGGCTTTCTGGTTGCCTTCA   -747
CpWAK3      TGGTTCATGTTCACGGTCGCAAAGTGATTGGGCAAGAAAATGGCTTTCTGGTTGCCTTCA   -766
            ********** ********************************************

CpWAK1      CGGTCTCTTGCACAACCTTCACGATCGTGAAGCTATTTTCTGCAAAATAACAGTAAGTTC   -715
CpWAK2      CGGTCTCTTGCACAACCTTCACGATCGTGAAGCTATTTTCTGCAAAATAACAGTAAGTTC   -686
CpWAK3      CGGTCTCTTGCACAACCTTCGCGATCGTGAAGCTAAAAATAAC--------AGTAAGTTC   -713
            ********** ********** *       *********

CpWAK1      AAAAATTCATATTCCCAATCTAGACATTGGATTGAAAATCCCTTTAGTCAGGATCAATAC   -654
CpWAK2      AAAATTTCATATTCCCAATCTAGACATTAAGAT----------TGAAA   -647
CpWAK3      AAAAATTCATATTCCCAATCTAGACATTGGATTGAAAATCCCTTTAGTCAGGATCAATAC   -652
            **** ********** *                 *

CpWAK1      TAGCGAGTCTTCTATATCATTGAGAATTACGAAACTGTGGCAAAACTGCCGTAAAAACGT   -593
CpWAK2      ATCCCTTTCTTCTATATCATTGAGAATTACGAAACTGTGGCAAAACTGCCGTAAAAATGT   -586
CpWAK3      TAGCGAGTCTTCTATATCATTGAGAATTACGAAACTGTGGCAAAACTGCCGTAAAAACGT   -591
               *  ********************************************** **

CpWAK1      TTGATTTTGCAGTGAAATAAAAATATAGCGCTATGCAAACGACTTAAATTTAACAACTAT   -532
CpWAK2      TTGATTTTGCAGTAAAAT-AAAATAGAGCGCTATGCAAACGACTTAAATTTAACAACTAT   -526
CpWAK3      TTGATTTTGCAGTGAAATAAAAATATAGCGCTATGCAAACGACTTAAATTTAACAACTAT   -530
            ********** **** ****** ***********************************

CpWAK1      AAACATAAGAAATTACGTTTATCACATGGCCAAGTAACACGTTCCAAAATAATACATTCT   -471
CpWAK2      ATACATAAGAAATTACGCTTATCACATGGCCAAGTAACACGTTCCA-AATAATACATTCT   -466
CpWAK3      AAACATAAGAAATTACGTTTATCACATGGCCAAGTAACACGTTCCAAAATAATACATTCT   -469
            * ********** ********************************** **********

CpWAK1      ACAATTAATAAATAGACTGAACTTGATCATCACTAAATAGACTGGACCCACCATCATAAT   -410
CpWAK2      ACAATTAATAAATAGACTGAACTTGATCATCACTAAATAGACTGGACCCACCATCATAAT   -406
CpWAK3      ACAATTAATAAATAGACTGAACTTGATCATCACTAAATAGACTGGACCCACCATCATAAT   -408
```

```
                ************************************************************

CpWAK1          GTGGGTTATAGCGAATTGAATACGTGGAAAGAGTCCGACGGAGTACTATCAATCCAACTG   -349
CpWAK2          GTGGGTTATAGCGAATTGAATACGTGGAAAGAGTCCGACGGAGTACTATCAATCCAACTG   -345
CpWAK3          GTGGGTTATAGCGAATTGAATACGTGGAAAGAGTCCGACGGAGTACTATCAATCCAACTG   -347
                ************************************************************

CpWAK1          CTCGAAGTCCTCAGAAATCAAGTCGTTTTTACCTATTTATATTTCTCCATATCCAATCTA   -288
CpWAK2          CTCGAAGTCCTCAGAAATCAAGTCGTTTTTACCTACTTACATTTCTCCATATCCAATCTA   -284
CpWAK3          CTCGAAGTCCTCAGAAATCAAGTCGTTTTTACCTATTTATATTTCTCCATATCCAATCTA   -286
                ********************************** *** ********************

CpWAK1          ATTGCACTATCCTCGTCTCGCTCTCTTATTCTGTCGTGACATAAGGAAAAAAGAAAAACA   -227
CpWAK2          ATTGCACTATCCTCGCCTCGCTCTCTTATTCTCACGTGACATAAGGAAAAAAAAAAAGCA   -223
CpWAK3          ATTGCACTATCCTCGTCTCGCTCTCTTATTCTGTCGTGACATAAGGATAAAAGAAAAACA   -225
                *************** **************  ************* **** **** **

CpWAK1          AATATATCACCAATCTTATTAGTACTCAAGTAAGAGCGGAGATCGAAGAATCAAACAGGT   -166
CpWAK2          AATATATCACCAATCTTATTAGTACTCAAGTAAGAGCGGAGATCGAAGAATCAAACAGGT   -159
CpWAK3          AATATATCACCAATCTTATTAGTACTCAAGTAAGAGCGGAGATCGAAGAATCAAACAGGT   -164
                ************************************************************

CpWAK1          ACGTATACGTCCATTATGGTAACAACTTCAACTTGCAGAAATATCATAAAAACAATGTAT    -105
CpWAK2          ACGTATACGTCCATTATGGTAACAACTCCAACATGCAGAACTATCATAAAAACAATGTAT    -98
CpWAK3          ACGTATACGTCCATTATGGTAACAACTCCAACTTGCAGAAATATCATAAAAACAATGTAT    -103
                *************************** **** ******* ******************

CpWAK1           AAGCATAAAAATAAAAAAAAAAAAGAAAAAAAAATTGATATACCGTTTGGCAACTGAATT    -44
CpWAK2           AAGCATGAAATGAAAAAAAT--AAATAAAT-----TGATATACCGTTTGGCAACTTAATT    -44
CpWAK3           AAGCATAAAATGAAAAAAAA--AAATGAAAAAAAATTGATATACCGTTTGCAACTGAATT    -44
                 ****** ***  *******   **   **      *       *  ** ****** ****

CpWAK1           CTGTGAGCTCATCAAGTTCGTATTACAACAGGCCAACACGGAAATGAGGACTTCTTCAAT    17
CpWAK2           CTGTGAGCTCATCAAGTTCGTATTACAGCAGGCCAACACGGAAATGAGGACTTCTTCAAT    17
CpWAK3           CTGTGAGCTCATCAAGTTCGTATTACAACAGGCCAACACGGAAATGAGGACTTCTTCAAT    17
                 ***************************** **********************************

CpWAK1           ACTTGCTGCAGTCGTTTTCTGCGTCTGCCTTAGCACAACTTCGGCTATACCATCA-----    72
CpWAK2           ACAAGCCGCAGTCGTTTTCTGCGTCTGCCTTTGCACAACTTCGGCTATACCATCATCGTC    77
CpWAK3           ACTAGCTGCAGTCGTTTTCTGCGTCTGCCTTAGCACAACTTCGGCTATACCATCA-----    72
                 **  ** ************************ ************************

CpWAK1           ------------TCATCCATTTCCATACCCATGGCGAAGCCTAATTGTAGCGAAACATG    119
CpWAK2           ATCATCATCATCATCATCCATTTTCATACCCATGGCGAAGCCTAATTGTAGCCAAACATG    137
```

```
CpWAK3      ------------TCATCCATTTCCATACCCATGGCGAAGCCTAATTGTAGCGAAACATG    119
                       ********** ****************************** *******

CpWAK1      CGGAAATATTACCGTTCCATATCCATTCGGCATGAGCTCAGAGTGCAGTATCGACTTAGA    179
CpWAK2      CGGAAATATTACCGTTCCATATCCATTCGGCATGAGCGCAAGGTGCTGTGCAGACTCATC    197
CpWAK3      CGGAAATATTACCGTTCCATATCCATTCGGCATGAGCTCAGAGTGCAGTATCGGCTTACA    179
            ************************************ **  **** **   * ** *

CpWAK1      TTTTCTCGTCGAGTGCCGAAACTCCACAAATCCTCCGACGCTGTTCTTGTCCAGCATCGA    239
CpWAK2      TTTTCTCGTCGAGTGCCTAAACTCCACAAATCCTCCGACGCTGTACTTGCCAAGACTCGA    257
CpWAK3      TTTTCTCGTCGAGTGCCGAAACTCCACAAATCCTCCGACGCTGTTCTTGCCCAGCATCGA    239
            ****************** ************************** **** * **  ****

CpWAK1      CCTGCAAGTAACTGACATTCGCGTGAACGGCACAATCGTCGTAAAATATCCCGTCACTCC    299
CpWAK2      CCTGCAAGTAACTGACATTCGCGTGAACGGCACAATCGTCGTAAAATATCCCGTCACTCC    317
CpWAK3      CCTGCAAGTAACTGACATTCGCGTGAACGGCACAATCGTCGTAAAATATCCCGTCACTCC    299
            ************************************************************

CpWAK1      AATCAACTGTACCGCCGTCAAGAAAACGGCGTCCCTCGGGAGACAATTCTTGGGGAGCCC    359
CpWAK2      AATCAAATGTACCGCTGTCAAGAAAATGGAGTCCCTCGGGATAAAATTATTGGGGAGCTC    377
CpWAK3      AATCAACTGTACCGCTGTCAAGAAAAAGGAGTCCCTCGGGGGAACATTATCGGGGAGCTC    359
            ****** ******** ********** ** **********  *  *** * ******* *

CpWAK1      TTTCACTATCTCAGCCGATGGAAACACCTTCGCCGTCTTAGGCTGCAGAAACTCCATCTT    419
CpWAK2      TTTCACTATCTCAGCCGATGAAAACACCTTCACCGTATTAGGCTGCAGAAACTCCATCTT    437
CpWAK3      TTTCACTATCTCAGCCGATGGAAACACCTTCACCGTATTTGGCTGCAGAAACTCCATCTT    419
            ******************** ********** **** ** ********************

CpWAK1      CTTGAAAATCAATGGGACGGGACTCAGCGGGTGCTTCCCCGCTTGTGGAAAGAATTACAC    479
CpWAK2      CTTGCAAATCAATGGGACGGGATACAGCGGATGTTTCCCAGCTTGTGGAGTGGATTACAC    497
CpWAK3      CTTGCAAATCAATGGGACGGGATACAGCGGATGCTTCCCAGCTTGTGGAGTGGCTTACAT    479
            **** ******************  ****** ** ***** *********  *  *****

CpWAK1      GCAAGACTCTTGCCAAATAGATATTCCGCCAAGATCCCAGGAGCTTATTTATACATACCA    539
CpWAK2      GCAAGATTCTTGCCAAATAGATATTCCGCCAAGATCCCAGGAGCTTATTTATACATACCA    557
CpWAK3      GCAAGACTCTTGCCAAATAGATATTCCCCCAAGACCCCTGGAGCTTATTTATACATACCA    539
            ****** ********************** ****** *** **********************

CpWAK1      AAGCACCACTGATACTCAACCAGGCAATACCACGCAATACTGTGGTTATGCGTTTCCTGT    599
CpWAK2      AAGCACCACTTATACTCAACCAGGCAATACCACGCAATACTGTGGTTATGCGTTTCCTGT    617
CpWAK3      AAGCACCACTGATACTCAACCAGGCAATGCCACGCAATACTGTGGTTATGCGTTTCCTGT    599
            ********** ****************** ********************************

CpWAK1      GGAGAGGGAGTCCCTAAGCAAGACTTATGACTTGTATAGAGGCCTCAGCGA---TGATTA    656
```

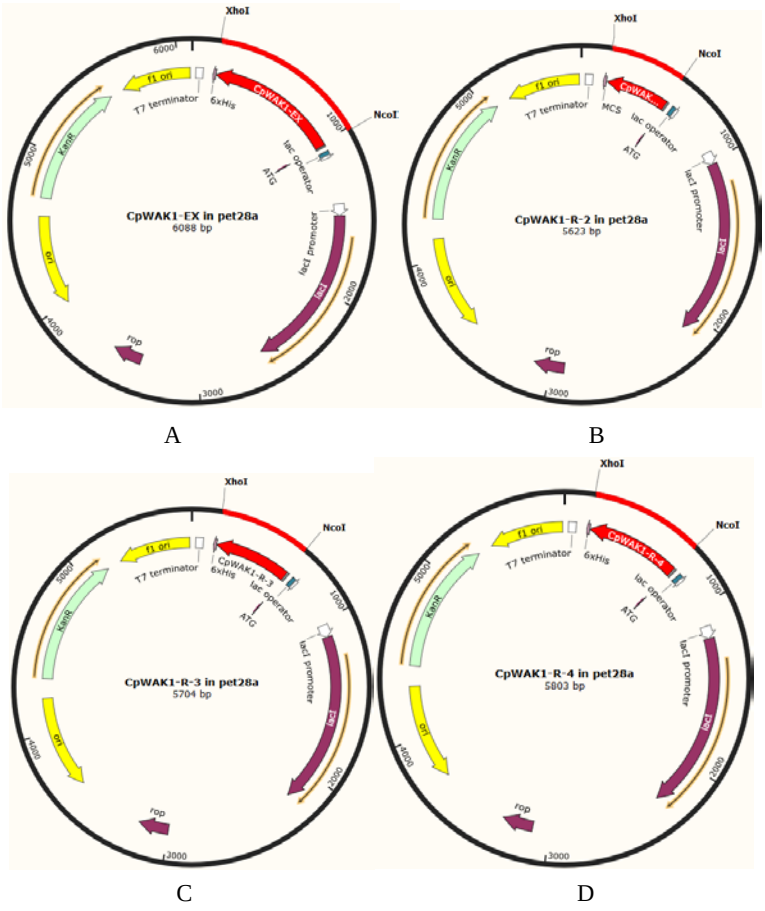
```
CpWAK2      GGAGAGGGCGTCGCTAAGCAAGACTTATGAGTTGTATAAAGGCCTTAGAG---ATGATTA   674
CpWAK3      GGAGAGGGAGTCCCTAAGCAAGACTTATGACTTGTATAGAGACCTCAGCGACGATGATTA   659
            ******** *** ********************* ******* ** *** ** *    ******

CpWAK1      TTTGAACCCATTCAACGAGCAACTTACGCATGCGCCGCTGGTGCTTGATTGGGAGTTAAC   716
CpWAK2      TTTTAACCCATTCGACGAGCAACTTACGCATGCGCCGCTGGTGCTTGATTGGGAGTTAAC   734
CpWAK3      TTTGAACCCATTCGACGAGCAACTTACGCATGCGCCGCTGGTGCTTGATTGGGAgTTAAC   719
            *** ********* ********************************************

CpWAK1      ACACATAGATATCAGCCCCTACTGGG--------------------------------A   743
CpWAK2      ATACATAGATACCAACGAATACGCTCGATGTAGTGATTA---TATCAGCAACAACAGGGA   791
CpWAK3      AGACGCAGATACCAACGAATACGCTCGATGTAATGGTTATGATGTCAGAgACTACAGGGA   779
            * **  ***** ** *   ***                                      *

CpWAK1      ACCGCCTCCCAATACTCTCTGCGCA------ACTCGATTTAGCGATCTTTCCCTGTCTAC   797
CpWAK2      ACTGCGTCCCAATACTCTTTGCGTAAGTCCATATTACTGGGGCAATCTTTACCATTATAC   851
CpWAK3      ACTGCCTCCCAATACTCTCTGCTCAAGTCGATATAGCGTGAGCGATCTTTCCCTGTATGC   839
            ** ** ************ ***  *        *       ** ****** **  * * *

CpWAK1      AACTAAGAAATGTTCTTGTTCCTATGGTTTTCGAGGAAATCCATATTTAGATGCAGGCTG   857
CpWAK2      AACTAAGAAATGTTCTTGTTGCTATGGTTTTCGAGGAAATCCTTATTTAGATGGAGGCTG   911
CpWAK3      AACTAGGCAATGTTCTTGTTGCTATGGTTTTCGAGGAAATCCATATTTAGATGGAGGCTG   899
            ***** * ************ ********************** ********** ******

CpWAK1      TGTTGATATCAACGAGTGCGAGGAGGATCCAAGCTTATGCGGTGCTACTGGGGCGACTTG   917
CpWAK2      TGTTGATATCAACGAGTGCGAGGAGGATCCAAGCAGATGCGCTGCTCCTGGGGTGACTTG   971
CpWAK3      TGTTGATATCAACGAGTGCGAGGAGGATCCAAGCATATGCGGTGCTACTGGGGCGACTTG   959
            **********************************  ***** **** ****** ******

CpWAK1      TGTCAATGAAATTGGCTCTTTCACGTGTCATGATCAGTATTCAAGCCATCGAGTAAGAAA   977
CpWAK2      TGTCAATGAAATCGGCTCTTCCACATGTCATTATCAGTATTCAAGCCATCGAGTAAGAAA  1031
CpWAK3      TGTCAATGAAATCGGCTCTTCCACGTGTCATTATCAGTATTCAAGCCATCGAGTAAGAAA  1019
            ************ ******* *** ****** ****************************

CpWAK1      CATATTGCTAGTTACTTTCGGCAGCCTCTTCGTTGCAGGGATTTTCATTCCGTGCCTGTC  1037
CpWAK2      CATATTGCTAGTTACTTTCGGCAGCCTCTTCGTTGCAGGGATTTTCATTCCGTGCCTGTC  1091
CpWAK3      CATATTGCTAGTTACTTTCGGCAGCCTCTTCGTTGCAGGGATTTTCATTCCGTGCCTGTC  1079
            ************************************************************

CpWAK1      AAAAGTCATCTTGAAAGGACTGAAGGCCCGCCGAAGAAGAAAATTCTTCGTGCGTAACGG  1097
CpWAK2      AAAAGTCATCTTGAAAAGACTGAAGGCTCGCCGAAGAAGAAAATTCTTCGTGCGTAACGG  1151
CpWAK3      AAAAGTCATCTTGAAAGGACTGAAGGCCCGCCGACGAAGAAAATTCTTCGTGCGTAACGG  1139
            ***************** ********** ****** ************************
```

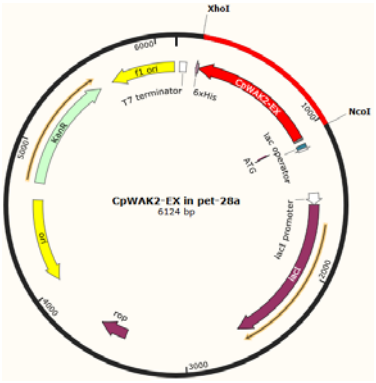
```
CpWAK1      CGGTCTGCTCCTCGAACAAAAGCTGTCCTCGATCGATAACGACTACAAGAAATCCAAGCT   1157
CpWAK2      CGGTCTGCTCCTCGAACAAAAGCTGTCCTCGATCGACAACGACTACAAGAAATCCAAGCT   1211
CpWAK3      CGGTCTGCTCCTCGAACAAAAGCTGTCCTCGATCGACAACGACTACAAGAAATCCAAGCT   1199
            ********************************** ***********************

CpWAK1      GTTCACATCCAAGGAGTTGAAGCAAGCCACCAATCATTACAGCGAGAATCGGGTGCTTGG   1217
CpWAK2      GTTCACATCCGAGGAGTTGAAGCAAGCCACCAATCATTACAGCGAGAATCGGGTGCTTGG   1271
CpWAK3      GTTCACATCCAAGGAGTTGAAGCAAGCCACCAATCATTACAGCGAGAATCGGGTGCTTGG   1259
            ********** ************************************************

CpWAK1      TAGGGGCGGCCAAGGGACGGTTTACAAGGGAATGCTGACAGATGGGAGCATAGTAGCCGT   1277
CpWAK2      TAGGGGCGGCCAAGGGACGGTTTACAAGGGAATGCTGACAGATGGGAGCATAGTAGCCGT   1331
CpWAK3      TAGGGGCGGCCAAGGGACGGTTTACaAGGGAATGCTGACAGATGGGAGCATAGTAGCCGT   1319
            ************************************************************

CpWAK1      GAAGAAATCCAAAACGGTCCAAGAAACCGACGTCGAGTCCTTCGTAAACGAGGTTGTCAT   1337
CpWAK2      GAAGAAATCCAAAACGGTCCAAGAAACCGACGTCGAGTCCTTCGTCAACGAGGTTGTCAT   1391
CpWAK3      GAAGAAATCCAAAACGGTCCAAGAAACCGACGTCGAGTCCTTCGTAAACGAGGTTGTCAT   1379
            ******************************************** **************

CpWAK1      TCTGTCCCAAATCATCCATCGAAGCATAGTCAGGCTCCTCGGATGCTGTCTGGAGACTGA   1397
CpWAK2      TCTGTCCCAAATCATCCATCGTAGCATAGTCAGGCTCCTCGGATGCTGTCTGGAGACTGA   1451
CpWAK3      TCTGTCCCAAATCATCCATCGAAGCATAGTCAGGCTCCTCGGATGCTGTCTGGAGACTGA   1439
            ********************* ************************************

CpWAK1      AACCCCAATCCTCGTATACGAATTTGTTCCTAACGGCACGCTATTCGAGCACATCCATGA   1457
CpWAK2      AACCCCAATCCTCGTATACGAATTTGTTCCTAACGGCACGCTATTCGAGCACATCCATGA   1511
CpWAK3      AACACCAATCCTCGTATACGAATTTGTTCCTAACGGCACGCTATTCGAGCACATCCATGA   1499
            *** ********************************************************

CpWAK1      CCGAAGCGAAGATTTCCCTCTTACCTGGCAAATGAGGCTTCGAATCGTTGTAGAAATAGC   1517
CpWAK2      CCGAAGCGAAGATTTCCCTCTTACCTGGCAAATGAGGCTTCGAATCGTTGTAGAAATAGC   1571
CpWAK3      CCGAAGCGAAGATTTCCCTCTTACCTGGCAAATGAGGCTTCGAATCGTTGTAGAAATAGC   1559
            ************************************************************

CpWAK1      CAGCGCCCTTTCTTATCTACACTCCTATGCTTCAGCCCCTATCTTCCACAGGGATATCAA   1577
CpWAK2      CAGCGCCCTTTCTTATCTACACTCCTATGCTTCAGCCCCTATCTTCCACAGGGATATCAA   1631
CpWAK3      CAGCGCCCTTTCTTATCTACACTCCTATGCTTCAGCCCCTATCTTCCACAGGGATATCAA   1619
            ************************************************************

CpWAK1      GTCGACAAACATACTGTTGGACGAGAAATACAGAGCGAAGGTCTCAGATTTTGGTACCTC   1637
CpWAK2      GTCGACCAACATACTGTTGGACGAGAAATACAGAGTGAAGGTCTCAGATTTTGGTACCTC   1691
CpWAK3      GTCGACAAACATACTGTTGGACGAGAAATACAGAGCGAAGGTCTCAGATTTTGGTACCTC   1679
            ****** **************************** ************************
```

```
CpWAK1      AAGGTCGCTCGCCATCGATCAAACTCATTTCACAACTAGGGTTTGCGGCACCTACGGCTA  1697
CpWAK2      AAGGTCGCTCGCCATCGATCAAACTCATTTCACAACTAGGGTTTGCGGCACCTACGGCTA  1751
CpWAK3      AAGGTCGCTCGCCATCGATCAAACTCATTTCACAACTAGGGTTTGCGGCACCTACGGCTA  1739
            ************************************************************

CpWAK1      CTTGGACCCAGAGTACTTCCAATCGAATCAGTTCACGGAGAAGAGTGACGTCTACAGCTT  1757
CpWAK2      CTTGGACCCAGAGTACTTCCAATCGAATCAGTTCACGGAGAAGAGCGACGTCTACAGCTT  1811
CpWAK3      CTTGGACCCAGAGTACTTCCAATCGAATCAGTTCACGGAGAAGAGCGACGTCTACAGCTT  1799
            ******************************************* **************

CpWAK1      CGGCGTCGTCATGGTGGAGCTCCTCACCGGCGAGATAGCCGTGTCGTTGCTGAGGGCGGG  1817
CpWAK2      CGGCGTCGTCATGGTGGAGCTCCTCACCAGCGAGATAGTCGTGTCGTTGCTGAGGGCGGG  1871
CpWAK3      CGGCGTCGTCATGGTGGAACTCCTCACCGGCGAGATAGCCGTGTCGTTGCTGAGGGCGGG  1859
            ******************* ********* ********* ********************
CpWAK1      GACGCGGAGGAGCTTAGCTACGCATTTCTTACACTCGATGGAAGAGGGCAAGCTGTTCGA  1877
CpWAK2      GACGCGGAGGAGCTTAGCTACGCATTTCTTACACTCGATGGAAGAGGGCAAGCTGTTCGA  1931
CpWAK3      GACGCGGAGGAGCTTAGCTACGCATTTCTTACACTCGATGGAAGAGGGCAAGCTGTTCGA  1919
            ************************************************************

CpWAK1      AATCGTTGATCCGAGAATCGTGGAAGGGGGGGAGCGCGAGAGGGAAGAAGAGATCACGAT  1937
CpWAK2      AATCGTTGATCCGAGAATCATGGAAGGGGGGGAGCGCGAGAGGGAAGAAGAGATCACGAT  1991
CpWAK3      AATCGTTGATCCGAGAATCGTGGAAGGGGGGGAGCGCGAGAGGGAaGAAGaGATCACGAT  1979
            ******************* ****************************************

CpWAK1      GGTGGCTGAACTCGCCCGACGGTGCCTCCAATTGAAAGGCAGTTTACGGCCGACCATGAG  1997
CpWAK2      GGTGGCTGAACTCGCCCGACGGTGCCTCCAATTGAAAGGCAGTTTACGGCCGACCATGAG  2051
CpWAK3      GGTGGCTGAACTCGCCCGACGGTGCCTCCAATTGAAAGGCAGTTTACGGCCGACCATGAG  2039
            ************************************************************

CpWAK1      GCAAATCGCCAACGAATTGGAGAGCGTAATCCACATCAAAAGTGCTGAGCAGAGCCATGA  2057
CpWAK2      GCAAATCGCCAACGAGTTGGAGAGCGTAATCCATATCAAGAGGGCTGAGCAGAGCCATGA  2111
CpWAK3      GCAAATCGCCAACGAGTTGGAGAGCGTAATCCATATCAAGAGGGCTGAGCAGAGCCACGA  2099
            ********************* ****************** ****** ** ********** **

CpWAK1      TCGCGATGAAGTTGAGCTTTCTGTGATTGATTTCTCATCAATTTCACCTTTCGCAATCGC  2117
CpWAK2      TCGTGATGAAGTAGAGCTTTCTGTAACTGATTTCTCATCAATTTCACCTTTCGCAATCGC  2171
CpWAK3      TCGTGATGAAGTAGAGCTTTCTGTAACTGATTTCTCATCAATTTCACCTTTCGCAATCGC  2159
            *** ******** *********** * ***********************************

CpWAK1      ATCCAATTAGGTTCACGATTTCTCCTTTTGTGTGATATTTTCTGATTATATTGACTTACG  2177
CpWAK2      ATCCAATTAGGTTCACGATTTCTCCTCTTGTGTGATATTTTCTGATTATATTGACTTACG  2231
CpWAK3      ATCCAATTAGGTTCACGATTTCTCCTTTTGTGTGATATTTTCTGATTATATTGACTTACG  2219
            ************************** **********************************
```

```
CpWAK1      GAGGTATGGAGCGTGACATCACGCTTATATCATAACTTCGCCGAATGCTCGAGGTAATAT   2237
CpWAK2      GAGGTATGGAGCGTGACATCACGCTTATATCATAACTTCGCCGAATGCTCGAAGTAATAT   2291
CpWAK3      GAGGTATGGAGCGTGACATCACGCTTATATCATAACTTCGCCGAATGCTCGAAGCAATAT   2279
            ************************************************** * *****

CpWAK1      TAGTGGATTATCGGTTATGGATATTGAAGATGATGTCATGACTTTTCACTTTGCATGAAA   2297
CpWAK2      TAGTGGATTATCGGTTATGGATATTGAAGATGATGTCATGACTTTTCACTTTGCATGAAA   2351
CpWAK3      TAGTGGATTATCGGtTATGGATATTGAAGATGATGTCATGACTTTTCACTTTGCATGAAA   2339
            ************************************************************

CpWAK1      ATGCCCGACCTTGATGCGCTTGATGGATATTTCTGTTTATG----TACCTCTTTTAAAAC   2353
CpWAK2      ATGCCCGACCTTGATGCGCTTGATGGATATTTCTGTTTATGTATACCCCTCTTTTAAAAC   2411
CpWAK3      ATGCCCGACCTTGATGCGCTTGATGGATATTTCTGTTTATGTATACCCCTCTTTT-----   2394
            ****************************************        ********
```

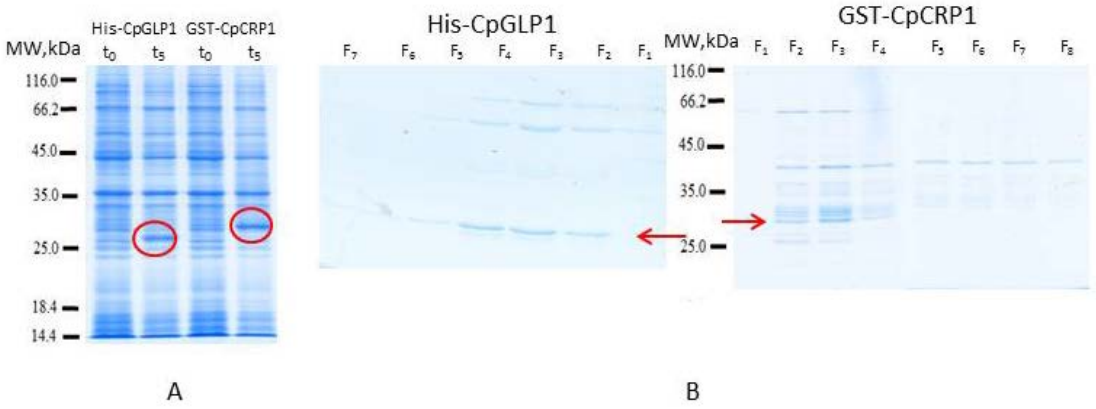
Supplementary figure 3 Multiple alignment of CpWAK DNA sequences. The translation start and stop condons (ATG and TAG) are indicated in bold. The translation start site ATG corresponds to+1 through +3. And the asterisks represent consensus residues.



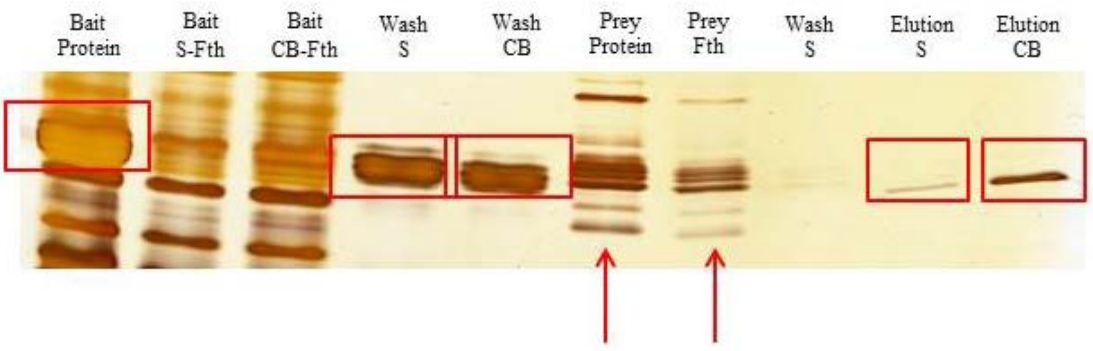


E

Supplementary figure 4 Expression vectors pET28a bearing the His-tagged fragments: CpWAK1-EX (A), CpWAK1-R-2 (B), CpWAK1-R-3 (C), CpWAK1-R-4 (D), CpWAK2-EX (E). The red arrows represent the extracellular fragments of CpWAK1 and CpWAK2.

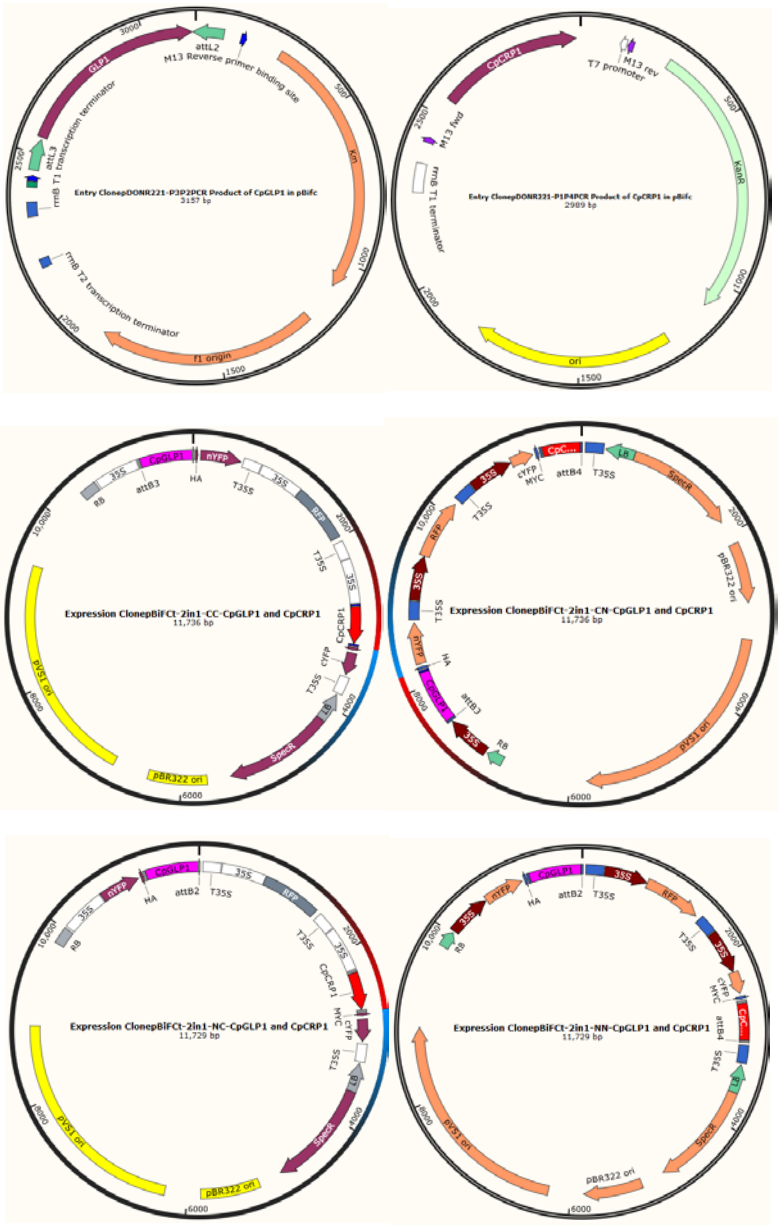


Supplementary figure 5 Inductions and purifications of His-CpGLP1 and GST-CpCRP1 under native conditions. (**A**) Inductions of the recombinant proteins His-CpGLP1 and GST-CpCRP1. The two proteins were induced with 1 mM IPTG for 5 hours. The bacterial pellets were collected and then loaded onto the SDS-PAGE gel after being homogenized with 1×Laemmli sample buffer. The induced proteins were encased in circle. (**B**) Purifications of His-CpGLP1 and GST-CpCRP1 under native conditions. The arrows represent the target recombinant proteins His-CpGLP1 and GST-CpCRP1. F1–F7: Eluted fractions

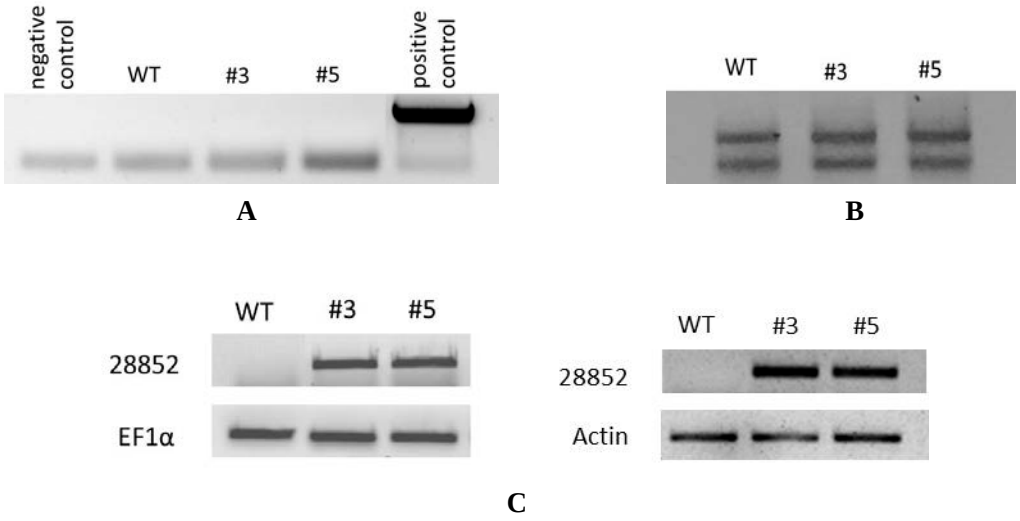


Supplementary figure 6 Analysis of interactions between CpGLP1 and CpCRP1 by pull-down assay with bait

proteins extracted from inclusion body. The bait protein, His-CpGLP1, extracted from inclusion body was loaded on the column directly and refolded on-column. The pull-down assay was analyzed with silver-stained gels. The proteins loaded were Bait protein (His-CpGLP1), Bait S-Fth (Bait protein flow-through sample from Sample column), Bait CB-Fth (Bait protein flow-through sample from Control-Bait column), Wash S (final wash sample from Sample column), Wash CB (final wash sample from Control-Bait column), Prey protein (GST-CpGRP1), Prey Fth (Prey flow-through sample), Wash S (final wash sample from Sample column after prey protein loaded), Elution S (Elution sample from Sample column), Elution CB (Elution sample from Control-Bait column). The bait protein (His-CpGLP1) and prey protein (GST-CpCRP1) were indicated by rectangle and arrow respectively.



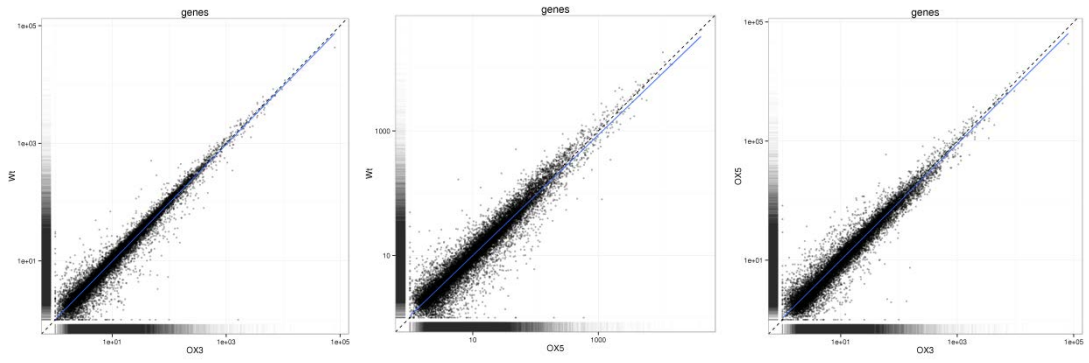
Supplementary figure 7 Entry clones and expression clones for BiFC assay.



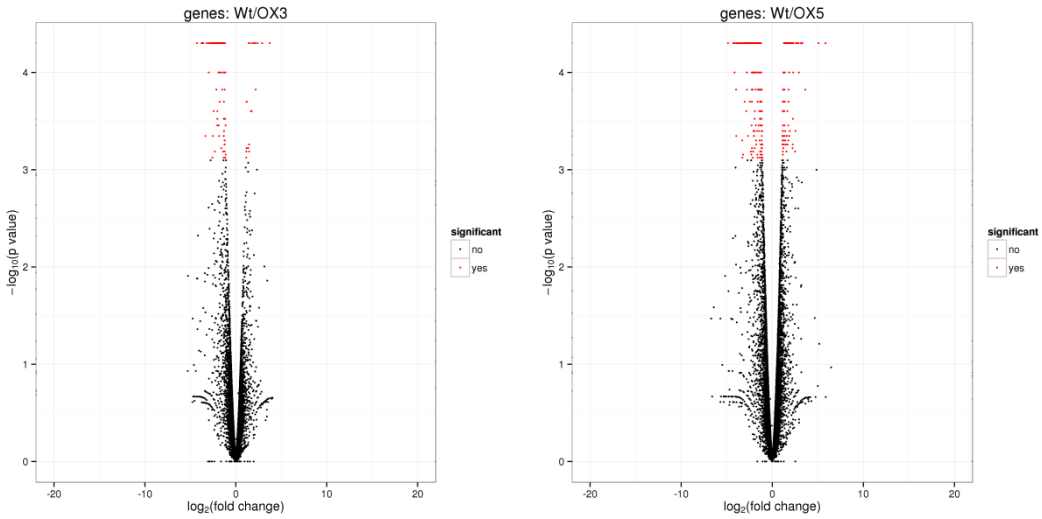
Supplementary figure 8 Preparations for RNA-seq. WT: wild type *A. thaliana*, #3**:** transgenic *A. thaliana* line 3, #5: transgenic *A. thaliana* line 5. (**A**) gDNA detection after DNase 1 treatment. Both the RNA and genomic DNA samples were amplified with *EF1a* primers. Negative control: wild type *A. thaliana* (old RNA sample whose gDNA was totally digested), Positive control: wild type *A. thaliana* genomic DNA. (**B**) RNA quality detection after DNase 1 treatment. (**C**) Expression of transcript 28852 in transgenic lines 3 and 5. The cDNAs were amplified using primers specific for lncRNA 28852, elongation factor 1α (*EF 1α*) (as reference) and actin (as reference).

Sample	Total reads	Clean reads	Mapped reads	% Mapped
WT	54950533	54881934	53580670	97.6
OX3	53454269	53388483	52146106	97.7
OX5	51016768	50899660	49726005	97.7

Supplementary table 1 Summary of mapped reads



Supplementary figure 9 Scatter plots of RNA-expression analysis was shown in A (WT and OX3), B (WT and OX5) and C (OX3 and OX5).



A **B**

Supplementary figure 10 Volcano plots of differentially expressed genes in WT and mutants.

Gene	Log_2(fold change) (OX3/WT)	Log_2(fold change) (OX5/WT)	Descriptions (Biological process, cellular component, molecular function)
CYP82G1	4.27205	2.27266	oxidation-reduction process, secondary metabolite biosynthetic process, terpene biosynthetic process, oxygen binding, (3E)-4,8-dimethyl-1,3,7-nonatriene synthase activity, 4,8,12-trimethyltrideca-1,3,7,11-tetraene synthase activity, heme binding, iron ion binding
RAP2.6	3.76747	2.66615	Gregulation of transcription, response to osmotic stress, cold, water deprivation, wounding, and salt stress, chloroplast organization, response to abscisic acid, salicylic acid and jasmonic acid, ethylene-activated signaling pathway, cellular response to heat
AT1G15010	3.69515	2.65236	defense response to fungus, integral component of membrane, mitochondrion
CYP94B1	3.67476	2.78624	oxidation-reduction process, response to wounding, oxygen binding，heme binding, iron ion binding, monooxygenase activity, oxidoreductase activity, chloroplast, integral component of membrane
EARLI1	3.58004	3.0334	defense response to fungus, induced systemic resistance, lipid transport, response to abscisic acid, response to cold, salt stress, lipid binding, chloroplast outer membrane, endoplasmic reticulum, extracellular region, plant-type cell wall, plasmodesma
CYP94B3	3.2342	2.56073	stigma and anther development, defense response to insect, fruit development, jasmonic acid metabolic process, oxidation-reduction process, pollen development, response to wounding, oxygen binding，heme binding, iron ion binding, jasmonoyl-isoleucine-12-hydroxylase activity, oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen extracellular region, integral component of membrane
NATA1	3.0151	2.39214	response to jasmonic acid, defense response, ornithine metabolic process, N-acetyltransferase activity, chloroplast, cytoplasm

GA2OX6	2.9805	3.4197	gibberellin biosynthetic process, gibberellin catabolic process, oxidation-reduction process, response to red or far red light, C-19 gibberellin 2-beta-dioxygenase activity, metal ion binding, cytoplasm
AOC3	2.81447	2.41838	jasmonic acid biosynthetic process, response to fungus, salt stress, allene-oxide cyclase activity, chloroplast, membrane, plasma membrane, vacuolar membrane
AT5G28237	2.7589	2.45011	tryptophan biosynthetic process, tryptophan synthase activity, mitochondrion
JAZ10	2.7062	1.94061	defense response, negative regulation of nucleic acid-templated transcription,regulation of jasmonic acid mediated signaling pathway, regulation of systemic acquired resistance, regulation of transcription, DNA-templated, response to jasmonic acid, response to wounding, transcription, DNA-templated, protein binding, transcription corepressor activity, nucleus
AT4G28140	2.68458	2.12432	regulation of transcription, ethylene-activated signaling pathway, response to chitin, nucleus, DNA binding
AT5G09530	2.65735	4.10071	post-embryonic development
EXT4	2.64521	2.82404	response to abscisic acid, response to jasmonic acid, response to salicylic acid, response to wounding, structural constituent of cell wall, nucleus
MBP1	2.59592	4.23596	defense response, extracellular region, nucleus, vacuole, carbohydrate binding
PME2	2.5668	2.86921	cell wall modification, pectin catabolic process, aspartyl esterase activity, pectinesterase activity, pectinesterase inhibitor activity, extracellular region, plant-type cell wall
ST2A	2.32635	2.84996	jasmonic acid metabolic process, cytoplasm, chloroplast, sulfotransferase activity, hydroxyjasmonate sulfotransferase activity,
AT1G54010	2.31173	2.29644	endoplasmic reticulum, extracellular region, membrane, vacuolar membrane, vacuole carboxylic ester hydrolase activity, hydrolase activity, acting on ester bonds
AT5G05600	2.26555	1.99032	cellular response to toxic substance, flavonoid biosynthetic process, oxidation-reduction process, response to karrikin, metal ion bindin, cytoplasm
VSP1	2.22088	2.76122	defense response, dephosphorylation, response to jasmonic acid, transcription factor binding，acid phosphatase activity, nutrient reservoir activity, chloroplast, vacuole
NAC019	2.14714	2.24807	multicellular organism development, regulation of transcription, DNA-templated,response to water deprivation, transcription, DNA-templated, transcription factor activity, nucleus
JAZ1	2.12405	2.90069	defense response to bacterium, flower development, jasmonic acid mediated signaling pathway, negative regulation of nucleic acid-templated transcription, pollen development, regulation of cellular response to alkaline pH, regulation of defense response, regulation of jasmonic acid mediated signaling pathway, regulation of transcription, DNA-templated, response to jasmonic acid, response to wounding, transcription, DNA-templated, protein binding, transcription corepressor activity, nucleus
ATJRG21	2.02912	2.596	flavonoid biosynthetic process, oxidation-reduction process, dioxygenase activity, metal ion binding, cytoplasm
AT1G66100	2.02458	2.16783	defense response, extracellular region, toxin receptor binding
CLH1	1.99769	2.41194	chlorophyll catabolic process, defense response to bacterium, defense response to fungus chloroplast, vacuole, chlorophyllase activity, pheophytinase b activity
JAZ5	1.95577	2.72285	defense response, regulation of jasmonic acid mediated signaling pathway, regulation of transcription, DNA-templated, response to wounding, nucleus, protein

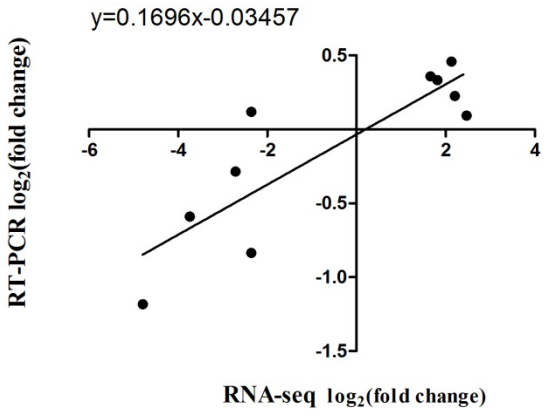
			binding, transcription corepressor activity
AT3G27400	1.94493	1.81128	pectin catabolic process, response to nematode, extracellular region, metal ion binding, pectate lyase activity
AT5G54585	1.91915	1.86505	---
LOX4	1.90939	3.02406	anther dehiscence and development, defense response, growth, jasmonic acid biosynthetic process, lipid oxidation, oxylipin biosynthetic process, pollen development, response to bacterium, response to ozone, response to wounding, stamen filament development, chloroplast, linoleate 13S-lipoxygenase activity, metal ion binding
AT1G66760	1.88773	1.87581	drug transmembrane transport, response to wounding, membrane, transporter activity
TPS03	1.87396	2.17407	response to herbivore, response to insect, response to jasmonic acid, response to wounding, sesquiterpenoid biosynthetic process, E)-beta-ocimene synthase activity, alpha-farnesene synthase activity,magnesium ion binding, myrcene synthase activity, tricyclene synthase activity, chloroplast, cytosol,
AOC1	1.86799	1.75655	jasmonic acid biosynthetic process, response to desiccation, chloroplast, chloroplast envelope, chloroplast thylakoid membrane, allene-oxide cyclase activity
JAL35/JR1	1.84009	2.23643	response to cold, salt stress and wounding, response to jasmonic acid, vegetative to reproductive phase transition of meristem, carbohydrate binding, protein binding chloroplast, cytoplasm, membrane, nucleus, vacuole
OPR3	1.79772	1.5218	stamen development，jasmonic acid biosynthetic process, oxidation-reduction process, oxylipin biosynthetic process, response to fungus, response to ozone, response to wounding, 12-oxophytodienoate reductase activity, FMN binding, NADPH dehydrogenase activity, peroxisome
AT5G23820	1.79632	1.44512	defense response, protein binding, ER body, extracellular region, plant-type vacuole
PGL5	1.78585	2.08806	carbohydrate metabolic process, pentose-phosphate shunt, 6-phosphogluconolactonase activity, chloroplast, cytosol
AT1G61890	1.76921	2.77832	response to karrikin, antiporter activity, transporter activity, chloroplast, membrane
ILL6	1.76657	1.55326	auxin metabolic process, jasmonic acid metabolic process, peptide catabolic process, proteolysis, regulation of systemic acquired resistance, extracellular region IAA-amino acid conjugate hydrolase activity, jasmonyl-Ile conjugate hydrolase activity, metallodipeptidase activity, metallopeptidase activity, zinc ion binding
JAL23	1.76359	2.148	cytoplasm, plasmodesma, vacuole
IAR3	1.74448	1.55104	proteolysis, metabolic process, response to wounding, peptide catabolic process, plasma membrane, chloroplast, metallo peptidase activity, zinc ion binding, IAA-Ala conjugate hydrolase
AT3G28220	1.74102	3.0745	chloroplast envelope, cytoplasm, integral component of membrane, vacuole
EXPA8	1.69985	1.56284	plant-type cell wall loosening, plant-type cell wall modification involved in multidimensional cell growth, syncytium formation, unidimensional cell growth cell wall, extracellular region, membrane
AT1G14250	1.67401	1.47012	vacuole, plasma membrane, ATP binding, hydrolase activity
FAMT	1.63565	1.13234	methylation, S-adenosylmethionine-dependent methyltransferase activity, farnesoic acid O-methyltransferase activity, chloroplast
BGLU18	1.6261	2.53638	abscisic acid metabolic process, abscisic acid-activated signaling, defense response to fungus, glucosinolate catabolic process, protein polymerization, regulation of stomatal movement, response to insect, salt stress and water deprivation, water homeostasis

			ER body, chloroplast, endoplasmic reticulum, endoplasmic reticulum lumen, extracellular region, nucleus, peroxisome, plasmodesma, vacuole
			abscisic acid glucose ester beta-glucosidase activity, beta-glucosidase activity,hydrolase activity, hydrolyzing O-glycosyl compounds
CAD5	1.62079	1.58774	lignin biosynthetic process, oxidation-reduction process, cytoplasm, cytosol cinnamyl-alcohol dehydrogenase activity, oxidoreductase activity, sinapyl alcohol dehydrogenase activity, zinc ion binding
AT1G52000	1.61925	2.8545	carbohydrate binding
MYC2	1.49207	2.04169	abscisic acid-activated signaling pathway, jasmonic acid mediated signaling pathway, positive regulation of flavonoid biosynthetic process, positive regulation of transcription, DNA-templated, regulation of defense response to insect, response to chitin, response to desiccation, response to jasmonic acid, response to wounding, transcription, DNA-templated, DNA binding, protein binding, nucleus
AT2G34810	1.4777	1.97654	oxidation-reduction process, response to jasmonic acid, response to wounding cytoplasm, extracellular region, electron carrier activity, flavin adenine dinucleotide binding, oxidoreductase activity, acting on CH-OH group of donors
CSD2	1.4746	1.44199	cellular response to ozone, cellular response to salt stress, cellular response to sucrose stimulus, gene silencing by miRNA, oxidation-reduction process, removal of superoxide radicals, response to copper ion, response to iron ion, response to light stimulus, response to oxidative stress, apoplast, chloroplast, chloroplast stroma, thylakoid, copper ion binding, superoxide dismutase activity, zinc ion binding
AT5G56980	1.40547	1.1384	chloroplast, integral component of membrane
PGIP2	1.381	1.54798	defense response, signal transduction, cell wall, extracellular region, membrane, plant-type cell wall, cytosol，polygalacturonase inhibitor activity
TSA1	1.34934	2.12712	defense response to fungus, photomorphogenesis
			chloroplast thylakoid membrane, endoplasmic reticulum lumen, nuclear, envelope, nucleus, peroxisome, vacuole, calcium ion binding, protein binding
AT2G39420	1.33969	1.17646	Golgi apparatus, endoplasmic reticulum, acylglycerol lipase activity, catalytic activity
AT5G13210	1.33511	1.4848	chloroplast
GASA6	1.33125	1.74235	gibberellic acid mediated signaling pathway, response to fructose, response to glucose, response to karrikin, response to sucrose, extracellular region
MYB47	1.3221	1.91991	cell differentiation, regulation of transcription, DNA-templated, response to jasmonic acid, response to salt stress, cytoplasm, nucleus, DNA binding
TIFY7	1.32097	1.37201	defense response, negative regulation of nucleic acid-templated transcription regulation of jasmonic acid mediated signaling pathway, DNA-templated, nucleus, protein binding, transcription corepressor activity
PDF1.2A	1.317	2.35128	defense response to fungus, jasmonic acid and ethylene-dependent systemic resistance, killing of cells of other organism, response to ethylene, response to insect, response to jasmonic acid, cell wall, extracellular region
BZIP61	1.31201	1.18631	DNA-templated, transcription, DNA-templated, Nucleus, DNA binding, protein binding,
VTC5	1.29853	1.27782	L-ascorbic acid biosynthetic process, glucose metabolic process, response to jasmonic acid, response to ozone, GDP-D-glucose phosphorylase activity, GDP-L-galactose phosphorylase activity,galactose-1-phosphate guanylyltransferase (GDP) activity, guanyl-nucleotide exchange factor activity, hydrolase activity, nucleotide

			binding, quercetin 4'-O-glucosyltransferase activity, cytoplasm, nucleus
Hsp70	1.27631	1.31313	transcription, cell wall, nucleus, Golgi apparatus, cytosol, plasma membrane, protein binding, ATP binding,
AT4G32800	1.22443	1.20295	transcription, ethylene-activated signaling pathway, nucleus, DNA binding, transcription factor activity
HPL1	1.21424	1.94574	fatty acid metabolic process, oxidation-reduction process, response to wounding, sterol metabolic process, chloroplast envelope, heme binding, iron ion binding, monooxygenase activity, oxidoreductase activity
TIFY10B	1.18736	1.09192	defense response, regulation of jasmonic acid mediated signaling pathway, regulation of transcription, DNA-templated, response to wounding, transcription, DNA-templated nucleus, protein binding, transcription corepressor activity
BCB	1.15187	1.64869	aluminum cation transport, cellular response to cold, defense response to fungus, oxidation-reduction process, regulation of lignin biosynthetic process, response to absence of light, response to wounding
AT2G37130	1.14785	1.53915	anchored component of membrane, anchored component of plasma membrane, vacuole copper ion binding, protein binding, electron carrier activity, metal ion binding response to oxidative stress, plant-type cell wall organization, hydrogen peroxide catabolic, extracellular region, cytosol, plant-type cell wall, peroxidase activity, heme binding, metal ion binding,
MYBR1	1.12355	1.80305	abscisic acid-activated signaling pathway, cell differentiation, defense response to bacterium, defense response to fungus, regulation of jasmonic acid and salicylic acid mediated signaling pathway, regulation of transcription from RNA polymerase II promoter, regulation of transcription, DNA-templated, response to auxin, response to cadmium ion, response to chitin, ethylene, gibberellin, salt stress, and water deprivation, transcription, DNA-templated, DNA binding, RNA polymerase II transcription factor activity, nucleus
ZW9	1.11474	1.17306	cytoplasm, integral component of membrane, plasmodesma
AT2G36885	-1.15241	-3.15186	chloroplast, integral component of membrane, plasma membrane
NIA2	-1.18646	-1.5139	nitrate assimilation, nitric oxide biosynthetic process, oxidation-reduction process, response to herbicide, light stimulus and symbiotic fungus, cytosol, mitochondrion, plasma membrane, vacuole, FAD binding, heme binding, molybdenum ion binding, molybdopterin cofactor binding, nitrate reductase (NADH) activity, nitrate reductase (NADPH) activity,nitrate reductase activity
FLA13	-1.19985	-1.83458	Golgi apparatus, anchored component of membrane, plant-type cell wall, plasma membrane, plasmodesma
NIA1	-1.41777	-2.27308	nitrate assimilation, nitric oxide biosynthetic process, oxidation-reduction process, response to herbicide and light stimulus, protein binding，FAD binding, heme binding, molybdenum ion binding, molybdopterin cofactor binding, nitrate reductase (NADH) activity, nitrate reductase (NADPH) activity, nitrate reductase activity, cytosol, mitochondrion
MT1C	-1.45812	-1.42065	response to copper ion, cytosolic ribosome, copper ion binding
AT3G06145	-1.46014	-1.78107	mitochondrion
CSLB3	-1.65587	-2.60932	polysaccharide biosynthetic process, plant-type primary cell wall biogenesis, cellulose, Golgi membrane, Golgi apparatus, plasma membrane, RNA polymerase II regulatory

			region sequence-specific DNA binding, transferase
AT4G08300	-1.74501	-2.22586	amino acid export, homeostasis and import, plasma membrane
AT4G16563	-1.76412	-1.35424	Proteolysis, extracellular region, plant-type cell wall
F3H	-1.97985	-2.73598	flavonoid biosynthetic process, oxidation-reduction process, response to UV-B cytoplasm, ATP binding, L-ascorbic acid binding, metal ion binding, naringenin 3-dioxygenase activity, oxidoreductase activity
LHCB4.3	-2.12211	-3.30442	chloroplast, chloroplast envelope, chloroplast thylakoid, chloroplast thylakoid membrane, integral component of membrane, light-harvesting complex, membrane, photosystem I, photosystem II, plastoglobule, thylakoid chlorophyll binding, metal ion binding, pigment binding
TT4	-2.39508	-5.07522	flavonoid biosynthetic process，auxin polar transport, chalcone biosynthetic process, regulation of anthocyanin biosynthetic process, response to UV-B, response to auxin, gravity, jasmonic acid, oxidative stress and wounding cytoplasm, endoplasmic reticulum, nucleus, plant-type vacuole membrane naringenin-chalcone synthase activity, protein binding
DFR	-3.7442	-5.83781	anthocyanin-containing compound biosynthetic process, oxidation-reduction process, extrinsic component of endoplasmic reticulum membrane, catalytic activity, dihydrokaempferol 4-reductase activity, flavanone 4-reductase activity, coenzyme binding,
AT3G44006	silent	silent	defense response, response to jasmonic acid, vacuole, chloroplast, acid phosphatase activity, transcription factor binding, nutrient reservoir activity
AT3G55230	silent	silent	defense response, phenylpropanoid biosynthetic process, extracellular region, apoplast, guiding stereospecific synthesis activity,
AT5G35430	silent	silent	---
AT5G45630	activated	activated	nucleus

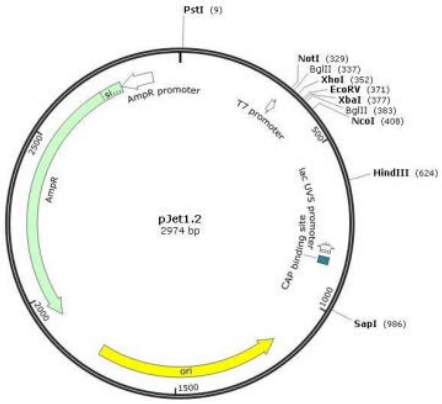
Supplementary table 2 86 genes significantly differentially expressed in both OX3 and OX5. The descriptions encompassing the biological process, cellular component and molecular function were written according to Tair (https://www.arabidopsis.org/) and DAVID (https://david.ncifcrf.gov/home.jsp). The genes enriched in GO:0009753 (response to jasmonic acid) and GO:0009611 (response to wounding) and KEGG pathways (ath00592 and ath00941) were indicated in bold and by boxes respectively.



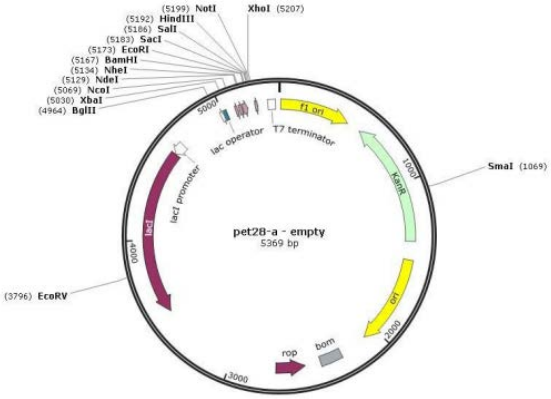
Supplementary figure 11 . Correlation of fold change values from RNA-seq and semi RT-PCR.

Vector maps

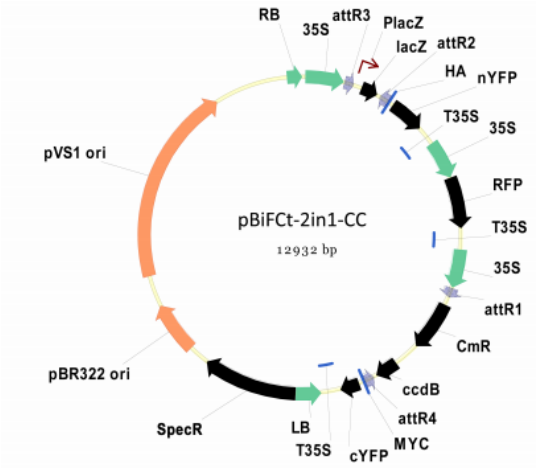
For more information see section 2.1.7



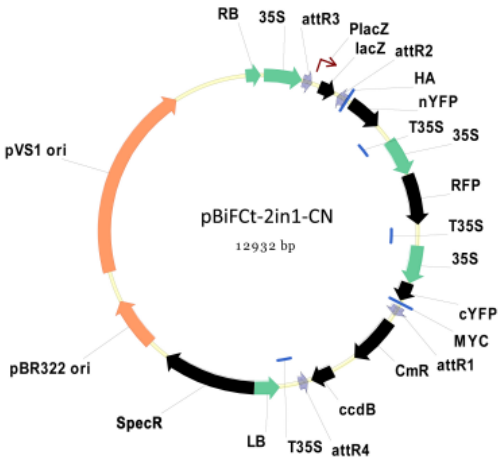
pJET1.2



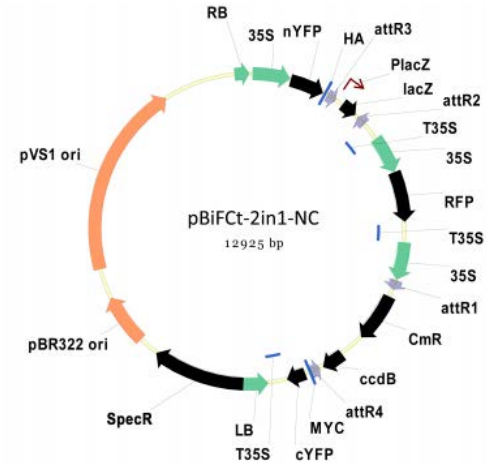
pET28a



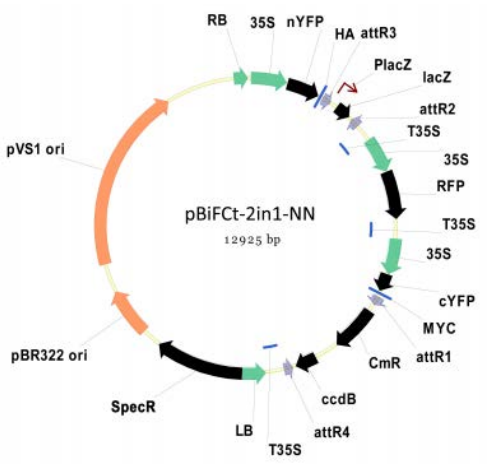
pBiFCt-2in1-CC



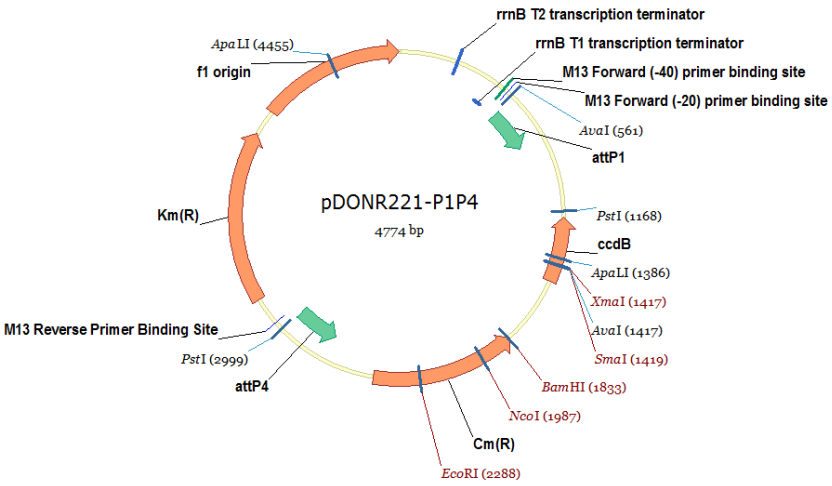
pBiFCt-2in1-CN



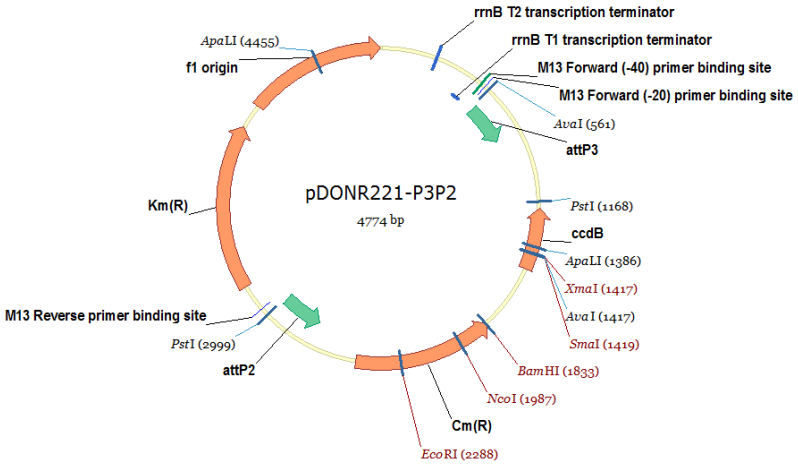
pBiFCt-2in1-NC



pBiFCt-2in1-N



pDONR221-P1P4



pDONR221-P3P2

6 REFERENCES

Alpert P (2000) The discovery, scope, and puzzle of desiccation tolerance in plants. Plant Ecol 151: 5-17

An SH, Sohn KH, Choi HW, Hwang IS, Lee SC, Hwang BK (2008) Pepper pectin methylesterase inhibitor protein CaPMEI1 is required for antifungal activity, basal disease resistance and abiotic stress tolerance. Planta 228: 61-78

Anderson CM, Wagner TA, Perret M, He ZH, He D, Kohorn BD (2001) WAKs: cell wall-associated kinases linking the cytoplasm to the extracellular matrix. Plant molecular biology 47: 197-206

Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, Ntini E, Arner E, Valen E, Li K, Schwarzfischer L, Glatz D, Raithel J, Lilje B, Rapin N, Bagger FO, Jorgensen M, Andersen PR, Bertin N, Rackham O, Burroughs AM, Baillie JK, Ishizu Y, Shimizu Y, Furuhata E, Maeda S, Negishi Y, Mungall CJ, Meehan TF, Lassmann T, Itoh M, Kawaji H, Kondo N, Kawai J, Lennartsson A, Daub CO, Heutink P, Hume DA, Jensen TH, Suzuki H, Hayashizaki Y, Muller F, Forrest ARR, Carninci P, Rehli M, Sandelin A, Consortium F (2014) An atlas of active enhancers across human cell types and tissues. Nature 507: 455

Ariel F, Jegu T, Latrasse D, Romero-Barrios N, Christ A, Benhamed M, Crespi M (2014) Noncoding Transcription by Alternative RNA Polymerases Dynamically Regulates an Auxin-Driven Chromatin Loop. Molecular Cell 55: 383-396

Ariel F, Romero-Barrios N, Jégu T, Benhamed M, Crespi M (2015) Battles and hijacks: noncoding transcription in plants. Trends Plant Sci 20: 362-371

Aung K, Lin SI, Wu CC, Huang YT, Su CL, Chiou TJ (2006) pho2, a phosphate overaccumulator, is caused by a nonsense mutation in a MicroRNA399 target gene. Plant physiology 141: 1000-1011

Banerjee J, Maiti MK (2010) Functional role of rice germin-like protein1 in regulation of plant height and disease resistance. Biochem Bioph Res Co 394: 178-183

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

Barman AR, Banerjee J (2015) Versatility of germin-like proteins in their sequences, expressions, and functions. Functional & integrative genomics 15: 533-548

Bartels D (2005) Desiccation tolerance studied in the resurrection plant Craterostigma plantagineum. Integr Comp Biol 45: 696-701

Bartels D, Schneider K, Terstappen G, Piatkowski D, Salamini F (1990) Molecular Cloning of Abscisic Acid-Modulated Genes Which Are Induced during Desiccation of the Resurrection Plant *Craterostigma Plantagineum*. Planta 181: 27-34

Ben Amor B, Wirth S, Merchan F, Laporte P, d'Aubenton-Carafa Y, Hirsch J, Maizel A, Mallory A, Lucas A, Deragon JM, Vaucheret H, Thermes C, Crespi M (2009) Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses. Genome Res 19: 57-69

Berna A, Bernier F (1997) Regulated expression of a wheat germin gene in tobacco: Oxalate oxidase activity and apoplastic localization of the heterologous protein. Plant molecular

biology 33: 417-429

Berna A, Scott K, Chabrière E, Bernier F (2009) The DING family of proteins: ubiquitous in eukaryotes, but where are the genes? BioEssays: news and reviews in molecular, cellular and developmental biology 31: 570-580

Bernier F, Berna A (2001) Germins and germin-like proteins: Plant do-all proteins. But what do they do exactly? Plant Physiol Bioch 39: 545-554

Bernier F, Lemieux G, Pallotta D (1987) Gene families encode the major encystment-specific proteins of Physarum polycephalum plasmodia. Gene 59: 265-277

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic acids research 7: 1513-1523

Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK (2005) Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. Cell 123: 1279-1291

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical biochemistry 72: 248-254

Breen J, Bellgard M (2010) Germin-like proteins (GLPs) in cereal genomes: gene clustering and dynamic roles in plant defence. Functional & integrative genomics 10: 463-476

Brutus A, Sicilia F, Macone A, Cervone F, De Lorenzo G (2010) A domain swap approach reveals a role of the plant wall-associated kinase 1 (WAK1) as a receptor of oligogalacturonides. Proceedings of the National Academy of Sciences 107: 9452-9457

Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Gene Dev 25: 1915-1927

Cabrera JC, Boland A, Messiaen J, Cambier P, Van Cutsem P (2008) Egg box conformation of oligogalacturonides: the time-dependent stabilization of the elicitor-active conformation increases its biological activity. Glycobiology 18: 473-482

Caffall KH, Mohnen D (2009) The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. Carbohydrate research 344: 1879-1900

Caliskan M, Turet M, Cuming AC (2004) Formation of wheat (Triticum aestivum L.) embryogenic callus involves peroxide-generating germin-like oxalate oxidase. Planta 219: 132-140

Carlevaro-Fita J, Rahim A, Guigó R, Vardy LA, Johnson R (2016) Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. Rna 22: 867-882

Carpita N, McCann M, Griffing LR (1996) The plant extracellular matrix: News from the cell's frontier. Plant Cell 8: 1451-1463

Carter C, Graham RA, Thornburg RW (1998) Arabidopsis thaliana contains a large family of germin-like proteins: characterization of cDNA and genomic sequences encoding 12 unique family members. Plant molecular biology 38: 929-943

Carter C, Thornburg RW (1999) Germin-like proteins: structure, phylogeny, and function. Journal of Plant Biology 42: 97-108

Cassab GI (1998) Plant cell wall proteins. Annual review of plant biology 49: 281-309

Cech TR, Steitz JA (2014) The noncoding RNA revolution-trashing old rules to forge new ones.

Cell 157: 77-94

Chattopadhyay T (2014) In silico analysis of the germinlike protein multigene family members of tomato with predicted oxalate oxidase activity. International Journal of Agriculture, Environment and Biotechnology 7: 669

Chekanova JA (2015) Long non-coding RNAs and their functions in plants. Current opinion in plant biology 27: 207-216

Chekanova JA, Gregory BD, Reverdatto SV, Chen H, Kumar R, Hooker T, Yazaki J, Li P, Skiba N, Peng Q (2007) Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the Arabidopsis transcriptome. Cell 131: 1340-1353

Chen M, Wang C, Bao H, Chen H, Wang Y (2016) Genome-wide identification and characterization of novel lncRNAs in *Populus* under nitrogen deficiency. Molecular genetics and genomics : MGG 291: 1663-1680

Chen XP, Wang ML, Holbrook C, Culbreath A, Liang XQ, Brenneman T, Guo BZ (2011) Identification and Characterization of a Multigene Family Encoding Germin-Like Proteins in Cultivated Peanut (*Arachis hypogaea* L.). Plant Mol Biol Rep 29: 389-403

Cheng JF, Niu QF, Zhang B, Chen KS, Yang RH, Zhu JK, Zhang YJ, Lang ZB (2018) Downregulation of RdDM during strawberry fruit ripening. Genome biology 19

Cheng X, Huang XJ, Liu SY, Tang M, Hu WF, Pan SY (2014) Characterization of germin-like protein with polyphenol oxidase activity from Satsuma mandarine. Biochem Bioph Res Co 449: 313-318

Chiou TJ, Aung K, Lin SI, Wu CC, Chiang SF, Su CL (2006) Regulation of phosphate homeostasis by microRNA in Arabidopsis. Plant Cell 18: 412-421

Chitwood DH, Timmermans MCP (2010) Small RNAs are on the move. Nature 467: 415-419

Consortium EP, Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PJ, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dhami P, Dillon SC, Dorschner MO, Fiegler H, Giresi PG, Goldy J, Hawrylycz M, Haydock A, Humbert R, James KD, Johnson BE, Johnson EM, Frum TT, Rosenzweig ER, Karnani N, Lee K, Lefebvre GC, Navas PA, Neri F, Parker SC, Sabo PJ, Sandstrom R, Shafer A, Vetrie D, Weaver M, Wilcox S, Yu M, Collins FS, Dekker J, Lieb JD, Tullius TD, Crawford GE, Sunyaev S, Noble WS, Dunham I, Denoeud F, Reymond A, Kapranov P, Rozowsky J, Zheng D, Castelo R, Frankish A, Harrow J, Ghosh S, Sandelin A, Hofacker IL, Baertsch R, Keefe D, Dike S, Cheng J, Hirsch HA, Sekinger EA, Lagarde J, Abril JF, Shahab A, Flamm C, Fried C, Hackermuller J, Hertel J, Lindemeyer M, Missal K, Tanzer A, Washietl S, Korbel J, Emanuelsson O, Pedersen JS, Holroyd N, Taylor R, Swarbreck D, Matthews N, Dickson MC, Thomas DJ, Weirauch MT, Gilbert J, Drenkow J, Bell I, Zhao X, Srinivasan KG, Sung WK, Ooi HS, Chiu KP, Foissac S, Alioto T, Brent M, Pachter L, Tress ML, Valencia A, Choo SW, Choo CY, Ucla C, Manzano C, Wyss C, Cheung E, Clark TG, Brown JB, Ganesh M, Patel S, Tammana H, Chrast J, Henrichsen CN, Kai C, Kawai J, Nagalakshmi U, Wu J, Lian Z, Lian J, Newburger P, Zhang X, Bickel P, Mattick JS, Carninci P, Hayashizaki Y, Weissman S, Hubbard T, Myers RM, Rogers J, Stadler PF, Lowe TM, Wei CL, Ruan Y, Struhl K, Gerstein M, Antonarakis SE, Fu Y, Green ED, Karaoz U, Siepel A, Taylor J, Liefer LA, Wetterstrand KA, Good PJ,

Feingold EA, Guyer MS, Cooper GM, Asimenos G, Dewey CN, Hou M, Nikolaev S, Montoya-Burgos JI, Loytynoja A, Whelan S, Pardi F, Massingham T, Huang H, Zhang NR, Holmes I, Mullikin JC, Ureta-Vidal A, Paten B, Seringhaus M, Church D, Rosenbloom K, Kent WJ, Stone EA, Program NCS, Baylor College of Medicine Human Genome Sequencing C, Washington University Genome Sequencing C, Broad I, Children's Hospital Oakland Research I, Batzoglou S, Goldman N, Hardison RC, Haussler D, Miller W, Sidow A, Trinklein ND, Zhang ZD, Barrera L, Stuart R, King DC, Ameur A, Enroth S, Bieda MC, Kim J, Bhinge AA, Jiang N, Liu J, Yao F, Vega VB, Lee CW, Ng P, Shahab A, Yang A, Moqtaderi Z, Zhu Z, Xu X, Squazzo S, Oberley MJ, Inman D, Singer MA, Richmond TA, Munn KJ, Rada-Iglesias A, Wallerman O, Komorowski J, Fowler JC, Couttet P, Bruce AW, Dovey OM, Ellis PD, Langford CF, Nix DA, Euskirchen G, Hartman S, Urban AE, Kraus P, Van Calcar S, Heintzman N, Kim TH, Wang K, Qu C, Hon G, Luna R, Glass CK, Rosenfeld MG, Aldred SF, Cooper SJ, Halees A, Lin JM, Shulha HP, Zhang X, Xu M, Haidar JN, Yu Y, Ruan Y, Iyer VR, Green RD, Wadelius C, Farnham PJ, Ren B, Harte RA, Hinrichs AS, Trumbower H, Clawson H, Hillman-Jackson J, Zweig AS, Smith K, Thakkapallayil A, Barber G, Kuhn RM, Karolchik D, Armengol L, Bird CP, de Bakker PI, Kern AD, Lopez-Bigas N, Martin JD, Stranger BE, Woodroffe A, Davydov E, Dimas A, Eyras E, Hallgrimsdottir IB, Huppert J, Zody MC, Abecasis GR, Estivill X, Bouffard GG, Guan X, Hansen NF, Idol JR, Maduro VV, Maskeri B, McDowell JC, Park M, Thomas PJ, Young AC, Blakesley RW, Muzny DM, Sodergren E, Wheeler DA, Worley KC, Jiang H, Weinstock GM, Gibbs RA, Graves T, Fulton R, Mardis ER, Wilson RK, Clamp M, Cuff J, Gnerre S, Jaffe DB, Chang JL, Lindblad-Toh K, Lander ES, Koriabine M, Nefedov M, Osoegawa K, Yoshinaga Y, Zhu B, de Jong PJ (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447: 799-816

Cornuault V, Manfield IW, Ralet MC, Knox JP (2014) Epitope detection chromatography: a method to dissect the structural heterogeneity and inter-connections of plant cell-wall matrix glycans. Plant J 78: 715-722

Cosgrove DJ (1997) Assembly and enlargement of the primary cell wall in plants. Annual review of cell and developmental biology 13: 171-201

Cosgrove DJ (2000) Loosening of plant cell walls by expansins. Nature 407: 321-326

Csorba T, Questa JI, Sun QW, Dean C (2014) Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization. P Natl Acad Sci USA 111: 16160-16165

Czajkowska BI, Jones G, Brown TA (2019) Diversity of a wall-associated kinase gene in wild and cultivated barley. Plos One 14: e0218526

da Cruz WT, Bezerra EH, Grangeiro TB, Lopes JL, Silva MZ, Ramos MV, Rocha BA, Oliveira JS, Freitas DC, Freitas CD (2019) Structural and enzymatic characterization of Peruvianin-I, the first germin-like protein with proteolytic activity. International journal of biological macromolecules 126: 1167-1176

de Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res 34:

W362-W365

de Hoon MJ, Imoto S, Nolan J, Miyano S (2004) Open source clustering software. Bioinformatics 20: 1453-1454

de Oliveira LFV, Christoff AP, de Lima JC, de Ross BCF, Sachetto-Martins G, Margis-Pinheiro M, Margis R (2014) The Wall-associated Kinase gene family in rice genomes. Plant Sci 229: 181-192

Decreux A, Messiaen J (2005) Wall-associated kinase WAK1 interacts with cell wall pectins in a calcium-induced conformation. Plant & cell physiology 46: 268-278

Decreux A, Thomas A, Spies B, Brasseur R, Van Cutsem P, Messiaen J (2006) In vitro characterization of the homogalacturonan-binding domain of the wall-associated kinase WAK1 using site-directed mutagenesis. Phytochemistry 67: 1068-1079

Deng FN, Zhang XP, Wang W, Yuan R, Shen FF (2018) Identification of Gossypium hirsutum long non-coding RNAs (lncRNAs) under salt stress. Bmc Plant Biol 18

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J, Veeravalli L, Ruan XA, Ruan YJ, Lassmann T, Carninci P, Brown JB, Lipovich L, Gonzalez JM, Thomas M, Davis CA, Shiekhattar R, Gingeras TR, Hubbard TJ, Notredame C, Harrow J, Guigo R (2012) The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. Genome Res 22: 1775-1789

Di C, Yuan JP, Wu Y, Li JR, Lin HX, Hu L, Zhang T, Qi YJ, Gerstein MB, Guo Y, Lu ZJ (2014) Characterization of stress-responsive lncRNAs in Arabidopsis thaliana by integrating expression, epigenetic and structural features. Plant Journal 80: 848-861

Ding JH, Shen JQ, Mao HL, Xie WB, Li XH, Zhang QF (2012) RNA-Directed DNA Methylation Is Involved in Regulating Photoperiod-Sensitive Male Sterility in Rice. Molecular plant 5: 1210-1216

Ding ZH, Tie WW, Fu LL, Yan Y, Liu GH, Yan W, Li YA, Wu CL, Zhang JM, Hu W (2019) Strand-specific RNA-seq based identification and functional prediction of drought-responsive lncRNAs in cassava. Bmc Genomics 20

Dong XN (1998) SA, JA, ethylene, and disease resistance in plants. Curr Opin Plant Biol 1: 316-323

Druka A, Kudrna D, Kannangara CG, Von Wettstein D, Kleinhofs A (2002) Physical and genetic mapping of barley (Hordeum vulgare) germin-like cDNAs. Proceedings of the National Academy of Sciences 99: 850-855

Dulitz S-J (2016) Isolation and characterization of the *Craterostigma plantagineum* germin-like protein 1 CpGLP1. Mathematics and Natural Science. Rheinische Friedrich-Wilhelms-Universität, Bonn

Dumas B, Sailland A, Cheviet JP, Freyssinet G, Pallett K (1993) Identification of Barley Oxalate Oxidase as a Germin-Like Protein. Cr Acad Sci Iii-Vie 316: 793-798

Dunwell JM (1998) Cupins: A new superfamily of functionally diverse proteins that include germins and plant storage proteins. Biotechnol Genet Eng 15: 1-32

Dunwell JM, Gibbings JG, Mahmood T, Naqvi SMS (2008) Germin and germin-like proteins: Evolution, structure, and function. Crit Rev Plant Sci 27: 342-375

Dunwell JM, Khuri S, Gane PJ (2000) Microbial relatives of the seed storage proteins of higher plants: Conservation of structure and diversification of function during evolution of the

cupin superfamily. Microbiol Mol Biol R 64: 153-+

Faghani E, Gharechahi J, Komatsu S, Mirzaei M, Khavarinejad RA, Najafi F, Farsad LK, Salekdeh GH (2015) Comparative physiology and proteomic analysis of two wheat genotypes contrasting in drought tolerance. Journal of proteomics 114: 1-15

Fan Z, Gu H, Chen X, Song H, Wang Q, Liu M, Qu LJ, Chen Z (2005) Cloning and expression analysis of Zmglp1, a new germin-like protein gene in maize. Biochem Biophys Res Commun 331: 1257-1263

Faye L, Chrispeels MJ (1988) Common antigenic determinants in the glycoproteins of plants, molluscs and insects. Glycoconjugate Journal 5: 245-256

Fossdal CG, Nagy NE, Johnsen Ø, Dalen LS (2007) Local and systemic stress responses in Norway spruce: Similarities in gene expression between a compatible pathogen interaction and drought stress. Physiological and Molecular Plant Pathology 70: 161-173

Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-Somoza I, Leyva A, Weigel D, Garcia JA, Paz-Ares J (2007) Target mimicry provides a new mechanism for regulation of microRNA activity. Nat Genet 39: 1033-1037

Francoz E, Ranocha P, Nguyen-Kim H, Jamet E, Burlat V, Dunand C (2015) Roles of cell wall peroxidases in plant development. Phytochemistry 112: 15-21

Fry SC (2004) Primary cell wall metabolism: tracking the careers of wall polymers in living plant cells. New Phytologist 161: 641–675

Fu J-Y, Wang X-C, Mao T-F, Cheng H, Chen F, Yang Y-J (2018) Identification and functional analysis of germin-like protein Gene family in tea plant (Camellia sinensis). Scientia horticulturae 234: 166-175

Gadaleta A, Colasuonno P, Giove SL, Blanco A, Giancaspro A (2019) Map-based cloning of QFhb.mgb-2A identifies a WAK2 gene responsible for Fusarium Head Blight resistance in wheat. Sci Rep-Uk 9

Gaff D (1971) Desiccation-tolerant flowering plants in southern Africa. Science 174: 1033-1034

Gaff DF (1987) Desiccation Tolerant Plants in South-America. Oecologia 74: 133-136

Geilfus CM (2017) The pH of the Apoplast: Dynamic Factor with Functional Impact Under Stress. Mol Plant 10: 1371-1386

Giarola V, Krey S, Frerichs A, Bartels D (2015) Taxonomically restricted genes of Craterostigma plantagineum are modulated in their expression during dehydration and rehydration. Planta 241: 193-208

Giarola V, Krey S, von den Driesch B, Bartels D (2016) The Craterostigma plantagineum glycine-rich protein CpGRP1 interacts with a cell wall-associated protein kinase 1 (CpWAK1) and accumulates in leaf cell walls during dehydration. The New phytologist 210: 535-550

Gramegna G, Modesti V, Savatin DV, Sicilia F, Cervone F, De Lorenzo G (2016) GRP-3 and KAPP, encoding interactors of WAK1, negatively affect defense responses induced by oligogalacturonides and local response to wounding. Journal of Experimental Botany 67: 1715-1729

Grant GT, Morris ER, Rees DA, Smith PJ, Thom D (1973) Biological interactions between polysaccharides and divalent cations: the egg‐box model. FEBS letters 32: 195-198

Grefen C, Blatt MR (2012) A 2in1 cloning system enables ratiometric bimolecular fluorescence

complementation (rBiFC). Biotechniques 53: 311-314

Grignon C, Sentenac H (1991) pH and Ionic Conditions in the Apoplast. Annu Rev Plant Phys 42: 103-128

Gucciardo S, Wisniewski JP, Brewin NJ, Bornemann S (2007) A germin-like protein with superoxide dismutase activity in pea nodules with high protein sequence identity to a putative rhicadhesin receptor. Journal of Experimental Botany 58: 1161-1171

Guttman M, Russell P, Ingolia NT, Weissman JS, Lander ES (2013) Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. Cell 154: 240-251

He ZH, Cheeseman I, He D, Kohorn BD (1999) A cluster of five cell wall-associated receptor kinase genes, Wak1-5, are expressed in specific organs of Arabidopsis. Plant molecular biology 39: 1189-1196

He ZH, Fujiki M, Kohorn BD (1996) A cell wall-associated, receptor-like protein kinase. The Journal of biological chemistry 271: 19789-19793

He ZH, He DZ, Kohorn BD (1998) Requirement for the induced expression of a cell wall associated receptor kinase for survival during the pathogen response. Plant Journal 14: 55-63

Heo JB, Sung S (2011) Vernalization-Mediated Epigenetic Silencing by a Long Intronic Noncoding RNA. Science 331: 76-79

Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic acids research 27: 297-300

His I, Driouich A, Nicol F, Jauneau A, Hofte H (2001) Altered pectin composition in primary cell walls of korrigan, a dwarf mutant of Arabidopsis deficient in a membrane-bound endo-1,4-beta-glucanase. Planta 212: 348-358

Hotto AM, Schmitz RJ, Fei ZJ, Ecker JR, Stern DB (2011) Unexpected Diversity of Chloroplast Noncoding RNAs as Revealed by Deep Sequencing of the Arabidopsis Transcriptome. G3-Genes Genom Genet 1: 559-570

Hou XW, Tong HY, Selby J, DeWitt J, Peng XX, He ZH (2005) Involvement of a cell wall-associated kinase, WAKL4, in Arabidopsis mineral responses. Plant physiology 139: 1704-1716

Hu B, Jin JP, Guo AY, Zhang H, Luo JC, Gao G (2015) GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics 31: 1296-1297

Hu W, Lv YY, Lei WR, Li X, Chen YH, Zheng LQ, Xia Y, Shen ZG (2014) Cloning and characterization of the Oryza sativa wall-associated kinase gene OsWAK11 and its transcriptional response to abiotic stresses. Plant Soil 384: 335-346

Huang DW, Sherman BT, Lempicki RA (2008) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic acids research 37: 1-13

Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols 4: 44

Huang W, Xian ZQ, Hu GJ, Li ZG (2016) SlAGO4A, a core factor of RNA-directed DNA methylation (RdDM) pathway, plays an important role under salt and drought stress in tomato. Mol Breeding 36

Hurni S, Scheuermann D, Krattinger SG, Kessel B, Wicker T, Herren G, Fitze MN, Breen J, Presterl T, Ouzunova M (2015) The maize disease resistance gene Htn1 against northern corn leaf blight encodes a wall-associated receptor-like kinase. Proceedings of the

National Academy of Sciences: 201502522

Ilyas M, Rasheed A, Mahmood T (2016) Functional characterization of germin and germin-like protein genes in various plant species using transgenic approaches. Biotechnol Lett 38: 1405-1421

Jabnoune M, Secco D, Lecampion C, Robaglia C, Shu Q, Poirier Y (2013) A rice cis-natural antisense RNA acts as a translational enhancer for its cognate mRNA and contributes to phosphate homeostasis and plant fitness. The Plant Cell 25: 4166-4182

Jaikaran ASI, Kennedy TD, Dratewkakos E, Lane BG (1990) Covalently Bonded and Adventitious Glycans in Germin. Journal of Biological Chemistry 265: 12503-12512

Jung NU, Giarola V, Chen P, Paul Knox J, Bartels D (2019) *Craterostigma plantagineum* cell wall composition is remodelled during desiccation and the glycin-rich protein Cp GRP 1 interacts with pectins through clustered arginines. The Plant Journal

Juszczak I, Bartels D (2017) LEA gene expression, RNA stability and pigment accumulation in three closely related Linderniaceae species differing in desiccation tolerance. Plant Sci 255: 59-71

König MF (2017) Untersuchung des GLP1-Protein aus Craterostigma plantagineum auf Superoxiddismutaseaktivität. Mathematisch-Naturwissenschaftlichen Fakultät. Rheinischen Friedrich-Wilhelms-Universität Bonn

Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermuller J, Hofacker IL, Bell I, Cheung E, Drenkow J, Dumais E, Patel S, Helt G, Ganesh M, Ghosh S, Piccolboni A, Sementchenko V, Tammana H, Gingeras TR (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 316: 1484-1488

Kaur R, Singh K, Singh J (2013) A root-specific wall-associated kinase gene, HvWAK1, regulates root growth and is highly divergent in barley and other cereals. Functional & integrative genomics 13: 167-177

Ke YQ, Han GQ, He HQ, Li JX (2009) Differential regulation of proteins and phosphoproteins in rice under drought stress. Biochem Bioph Res Co 379: 133-138

Kim HJ, Pesacreta TC, Triplett BA (2004) Cotton-fiber germin-like protein. II: Immunolocalization, purification, and functional analysis. Planta 218: 525-535

Kohorn BD (2000) Plasma membrane-cell wall contacts. Plant physiology 124: 31-38

Kohorn BD (2015) Cell wall-associated kinases and pectin perception. Journal of experimental botany 67: 489-494

Kohorn BD, Hoon D, Minkoff BB, Sussman MR, Kohorn SL (2016) Rapid Oligo-Galacturonide Induced Changes in Protein Phosphorylation in Arabidopsis. Mol Cell Proteomics 15: 1351-1359

Kohorn BD, Johansen S, Shishido A, Todorova T, Martinez R, Defeo E, Obregon P (2009) Pectin activation of MAP kinase and gene expression is WAK2 dependent. The Plant journal: for cell and molecular biology 60: 974-982

Kohorn BD, Kobayashi M, Johansen S, Friedman HP, Fischer A, Byers N (2006a) Wall-associated kinase 1 (WAK1) is crosslinked in endomembranes, and transport to the cell surface requires correct cell-wall synthesis. Journal of cell science 119: 2282-2290

Kohorn BD, Kobayashi M, Johansen S, Riese J, Huang LF, Koch K, Fu S, Dotson A, Byers N (2006b) An Arabidopsis cell wall-associated kinase required for invertase activity and

cell growth. Plant Journal 46: 307-316

Kohorn BD, Kohorn SL (2012a) The cell wall-associated kinases, WAKs, as pectin receptors. Frontiers in plant science 3: 88

Kohorn BD, Kohorn SL (2012b) The Cell Wall-Associated Kinases, WAKs, Regulate Cell Expansion and the Stress Response. Receptor-like Kinases in Plants. Springer, pp 109-124

Kohorn BD, Kohorn SL, Saba NJ, Martinez VM (2014) Requirement for Pectin Methyl Esterase and Preference for Fragmented over Native Pectins for Wall-associated Kinase-activated, EDS1/PAD4-dependent Stress Response in Arabidopsis. Journal of Biological Chemistry 289: 18978-18986

Kohorn BD, Kohorn SL, Todorova T, Baptiste G, Stansky K, McCullough M (2012) A dominant allele of Arabidopsis pectin-binding wall-associated kinase induces a stress response suppressed by MPK6 but not MPK3 mutations. Molecular plant 5: 841-851

Krugman T, Chague V, Peleg Z, Balzergue S, Just J, Korol AB, Nevo E, Saranga Y, Chalhoub B, Fahima T (2010) Multilevel regulation and signalling processes associated with adaptation to terminal drought in wild emmer wheat. Functional & integrative genomics 10: 167-186

Kumakura N, Takeda A, Fujioka Y, Motose H, Takano R, Watanabe Y (2009) SGS3 and RDR6 interact and colocalize in cytoplasmic SGS3/RDR6-bodies. FEBS letters 583: 1261-1266

Kwenda S, Birch PR, Moleleki LN (2016) Genome-wide identification of potato long intergenic noncoding RNAs responsive to Pectobacterium carotovorum subspecies brasiliense infection. Bmc Genomics 17: 614

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. nature 227: 680

Lally D, Ingmire P, Tong HY, He ZH (2001) Antisense expression of a cell wall-associated protein kinase, WAK4, inhibits cell elongation and alters morphology. Plant Cell 13: 1317-1331

Lane BG (1991) Cellular desiccation and hydration: developmentally regulated proteins, and the maturation and germination of seed embryos. The FASEB journal 5: 2893-2901

Lane BG, Cuming AC, Fregeau J, Carpita NC, Hurkman WJ, Bernier F, Dratewkakos E, Kennedy TD (1992) Germin Isoforms Are Discrete Temporal Markers of Wheat Development - Pseudogermin Is a Uniquely Thermostable Water-Soluble Oligomeric Protein in Ungerminated Embryos and Like Germin in Germinated Embryos, It Is Incorporated into Cell-Walls. Eur J Biochem 209: 961-969

Lee JT (2012) Epigenetic Regulation by Long Noncoding RNAs. Science 338: 1435-1439

Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombauts S (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res 30: 325-327

Li H, Zhou SY, Zhao WS, Su SC, Peng YL (2009) A novel wall-associated receptor-like protein kinase gene, OsWAK1, plays important roles in rice blast disease resistance. Plant molecular biology 69: 337-346

Li L, Eichten SR, Shimizu R, Petsch K, Yeh CT, Wu W, Chettoor AM, Givan SA, Cole RA, Fowler JE, Evans MMS, Scanlon MJ, Yu JM, Schnable PS, Timmermans MCP, Springer NM, Muehlbauer GJ (2014) Genome-wide discovery and characterization of maize long

non-coding RNAs. Genome biology 15

Li L, Xu XH, Chen C, Shen ZG (2016a) Genome-Wide Characterization and Expression Analysis of the Germin-Like Protein Family in Rice and Arabidopsis. Int J Mol Sci 17

Li S, Liberman LM, Mukherjee N, Benfey PN, Ohler U (2013) Integrated detection of natural antisense transcripts using strand-specific RNA sequencing data. Genome Res 23: 1730-1739

Li S, Yamada M, Hang XW, Ohler U, Benfey PN (2016b) High-Resolution Expression Map of the Arabidopsis Root Reveals Alternative Splicing and lincRNA Regulation. Dev Cell 39: 508-522

Li Y-F, Mahalingam R, Sunkar R (2017) Isolation of Polysomal RNA for analyzing stress-responsive genes regulated at the translational level in plants. Plant Stress Tolerance. Springer, pp 151-161

Liu J, Jung C, Xu J, Wang H, Deng SL, Bernad L, Arenas-Huertero C, Chua NH (2012) Genome-Wide Analysis Uncovers Regulation of Long Intergenic Noncoding RNAs in Arabidopsis. Plant Cell 24: 4333-4345

Liu X, Hao L, Li D, Zhu L, Hu S (2015) Long non-coding RNAs and their biological roles in plants. Genomics, proteomics & bioinformatics 13: 137-147

Liu Y, Liu D, Zhang H, Gao H, Guo X, Fu X, Zhang A (2006) Isolation and characterisation of six putative wheat cell wall-associated kinases. Functional plant biology 33: 811-821

Lorow D, Jessee J (1990) Max efficiency DH10B™: a host for cloning methylated DNA. Focus 12: 19

Lu M, Han YP, Gao JG, Wang XJ, Li WB (2010) Identification and analysis of the germin-like gene family in soybean. Bmc Genomics 11

Lutts S, Kinet J, Bouharmont J (1996) NaCl-induced senescence in leaves of rice (Oryza sativaL.) cultivars differing in salinity resistance. Annals of botany 78: 389-398

MacIntosh GC, Wilkerson C, Green PJ (2001) Identification and analysis of Arabidopsis expressed sequence tags characteristic of non-coding RNAs. Plant physiology 127: 765-776

Mangeon A, Menezes-Salgueiro AD, Sachetto-Martins G (2017) Start me up: Revision of evidences that AtGRP3 acts as a potential switch for AtWAK1. Plant Signal Behav 12

Marchler-Bauer A, Lu SN, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke ZX, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang DC, Zhang NG, Zheng CJ, Bryant SH (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res 39: D225-D229

Marker C, Zemann A, Terhörst T, Kiefmann M, Kastenmayer JP, Green P, Bachellerie J-P, Brosius J, Hüttenhofer A (2002) Experimental RNomics: identification of 140 candidates for small non-messenger RNAs in the plant Arabidopsis thaliana. Current Biology 12

Mattick JS (2004) RNA regulation: a new genetics? Nat Rev Genet 5: 316-323

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

Maureen C. McCanna, Max Busha, Dimitra Milionia, Pierre Sadoa, Nicola J. Staceya, Gareth Catchpolea, Marianne Defernezb, Nicholas C. Carpitac, Herman Hofted, Peter Ulvskove,

Reginald H. Wilsonb, Roberts K (2001) Approaches to understanding the functional architecture of the plant cell wall. Phytochemistry 57 811-821

Medina C, da Rocha M, Magliano M, Raptopoulo A, Marteu N, Lebrigand K, Abad P, Favery B, Jaubert-Possamai S (2018) Characterization of siRNAs clusters in Arabidopsis thaliana galls induced by the root-knot nematode Meloidogyne incognita. Bmc Genomics 19: 943

Meier S, Ruzvidzo O, Morse M, Donaldson L, Kwezi L, Gehring C (2010) The Arabidopsis Wall Associated Kinase-Like 10 Gene Encodes a Functional Guanylyl Cyclase and Is Co-Expressed with Pathogen Defense Related Genes. Plos One 5

Membré N, Bernier F (1998) The rice genome expresses at least six different genes for oxalate oxidase/germinlike proteins (GenBank AF032971, AF032972, AF032973, AF032974, AF032975, AF032976)(PGR98-021). Plant physiology 116: 868

Membre N, Bernier F, Staiger D, Berna A (2000) Arabidopsis thaliana germin-like proteins: common and specific features point to a variety of functions. Planta 211: 345-354

Meng XH, Han J, Wang Q, Tian SP (2009) Changes in physiology and quality of peach fruits treated by methyl jasmonate under low temperature stress. Food Chem 114: 1028-1035

Meng YJ, Shao CG, Wang HZ, Jin YF (2012) Target mimics: an embedded layer of microRNA-involved gene regulatory networks in plants. Bmc Genomics 13

Michaels SD, Amasino RM (1999) FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. Plant Cell 11: 949-956

Michalowski CB, Bohnert HJ (1992) Nucleotide-Sequence of a Root-Specific Transcript Encoding a Germin-Like Protein from the Halophyte Mesembryanthemum-Crystallinum. Plant physiology 100: 537-538

Micheli F (2001) Pectin methylesterases: cell wall enzymes with important roles in plant physiology. Trends Plant Sci 6: 414-419

Miller G, Suzuki N, Ciftci-Yilmaz S, Mittler R (2010) Reactive oxygen species homeostasis and signalling during drought and salinity stresses. Plant Cell Environ 33: 453-467

Molnar A, Melnyk C, Baulcombe DC (2011) Silencing signals in plants: a long journey for small RNAs. Genome biology 12: 215

Mousavi A, Hotta Y (2005) Glycine-rich proteins: a class of novel proteins. Applied biochemistry and biotechnology 120: 169-174

Murashige T, Skoog F (1962) A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. Physiol Plantarum 15: 473-497

Nakata M, Watanabe Y, Sakurai Y, Hashimoto Y, Matsuzaki M, Takahashi Y, Satoh T (2004) Germin-like protein gene family of a moss, Physcomitrella patens, phylogenetically falls into two characteristic new clades. Plant molecular biology 56: 381-395

Nejat N, Mantri N (2018) Emerging roles of long non-coding RNAs in plant response to biotic and abiotic stresses. Crit Rev Biotechnol 38: 93-105

Ng SY, Lin L, Soh BS, Stanton LW (2013) Long noncoding RNAs in development and disease of the central nervous system. Trends Genet 29: 461-468

Nguyen QN, Lee YS, Cho LH, Jeong HJ, An G, Jung KH (2015) Genome-wide identification and analysis of Catharanthus roseus RLK1-like kinases in rice. Planta 241: 603-613

Nogara M (2011) Functional promoter analysis of the desiccation-related gene 28852 isolated from the resurrection plant Craterostigma plantagineum. Università degli studi di Verona

Oganesyan N, Kim S-H, Kim R (2004) On-column chemical refolding of proteins. PharmaGenomics 6: 177-182

Park AR, Cho SK, Yun UJ, Jin MY, Lee SH, Sachetto-Martins G, Park OK (2001) Interaction of the arabidopsis receptor protein kinase Wak1 with a glycine-rich protein, AtGRP-3. Journal of Biological Chemistry 276: 26688-26693

Park HC, Kim ML, Kang YH, Jeon JM, Yoo JH, Kim MC, Park CY, Jeong JC, Moon BC, Lee JH, Yoon HW, Lee SH, Chung WS, Lim CO, Lee SY, Hong JC, Cho MJ (2004) Pathogen- and NaCl-induced expression of the SCaM-4 promoter is mediated in part by a GT-1 box that interacts with a GT-1-like transcription factor. Plant physiology 135: 2150-2161

Petrov V, Hille J, Mueller-Roeber B, Gechev TS (2015) ROS-mediated abiotic stress-induced programmed cell death in plants. Frontiers in plant science 6

Ponting CP, Oliver PL, Reik W (2009) Evolution and Functions of Long Noncoding RNAs. Cell 136: 629-641

Porembski S, Barthlott W (2000) Granitic and gneissic outcrops (inselbergs) as centers of diversity for desiccation-tolerant vascular plants. Plant Ecol 151: 19-28

Printz B, Dos Santos Morais R, Wienkoop S, Sergeant K, Lutts S, Hausman JF, Renaut J (2015) An improved protocol to study the plant cell wall proteome. Frontiers in plant science 6: 237

Rahmanzadeh R, Muller K, Fischer E, Bartels D, Borsch T (2005) The Linderniaceae and Gratiolaceae are further lineages distinct from the Scrophulariaceae (Lamiales). Plant Biology 7: 67-78

Reymond P, Farmer EE (1998) Jasmonate and salicylate as global signals for defense gene expression. Curr Opin Plant Biol 1: 404-411

Rietz S, Bernsdorff FEM, Cai DG (2012) Members of the germin-like protein family in Brassica napus are candidates for the initiation of an oxidative burst that impedes pathogenesis of Sclerotinia sclerotiorum. Journal of Experimental Botany 63: 5507-5519

Rinn J, Guttman M (2014) RNA Function. RNA and dynamic nuclear organization. Science 345: 1240-1241

Rizzon C, Ponger L, Gaut BS (2006) Striking similarities in the genomic distribution of tandemly arrayed genes in Arabidopsis and rice. Plos Comput Biol 2: 989-1000

Rodriguez MC, Edsgard D, Hussain SS, Alquezar D, Rasmussen M, Gilbert T, Nielsen BH, Bartels D, Mundy J (2010) Transcriptomes of the desiccation-tolerant resurrection plant Craterostigma plantagineum. The Plant journal: for cell and molecular biology 63: 212-228

Rosli HG, Zheng Y, Pombo MA, Zhong S, Bombarely A, Fei Z, Collmer A, Martin GB (2013) Transcriptomics-based screen for genes induced by flagellin and repressed by pathogen effectors identifies a cell wall-associated kinase involved in plant immunity. Genome biology 14: R139

Saha D, Rana RS, Arya L, Mondal TK (2017) Genomic organization and structural diversity of germin-like protein coding genes in foxtail millet (Setaria italica L.). Agri Gene 3: 87-98

Sahu S, Rao AR, Pandey J, Gaikwad K, Ghoshal S, Mohapatra T (2018) Genome-wide identification and characterization of lncRNAs and miRNAs in cluster bean (Cyamopsis tetragonoloba). Gene 667: 112-121

Saintenac C, Lee WS, Cambon F, Rudd JJ, King RC, Marande W, Powers SJ, Berges H, Phillips

AL, Uauy C, Hammond-Kosack KE, Langin T, Kanyuka K (2018) Wheat receptor-kinase-like protein Stb6 controls gene-for-gene resistance to fungal pathogen Zymoseptoria tritici. Nat Genet 50: 368-+

Sakamoto A, Nishimura T, Miyaki Y, Watanabe S, Takagi H, Izumi S, Shimada H (2015) In vitro and in vivo evidence for oxalate oxidase activity of a germin-like protein from azalea. Biochem Bioph Res Co 458: 536-542

Saldanha AJ (2004) Java Treeview-extensible visualization of microarray data. Bioinformatics 20: 3246-3248

Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP (2011) A ceRNA Hypothesis: The Rosetta Stone of a Hidden RNA Language? Cell 146: 353-358

Schutze K, Harter K, Chaban C (2009) Bimolecular fluorescence complementation (BiFC) to study protein-protein interactions in living plant cells. Methods in molecular biology 479: 189-202

Schweizer P, Christoffel A, Dudler R (1999) Transient expression of members of the germin-like gene family in epidermal cells of wheat confers disease resistance. The Plant journal : for cell and molecular biology 20: 541-552

Scott P (2000) Resurrection plants and the secrets of eternal leaf. Ann Bot-London 85: 159-166

Sedan D, Pagnoux C, Chotard T, Smith A, Lejolly D, Gloaguen V, Krausz P (2007) Effect of calcium rich and alkaline solutions on the chemical behaviour of hemp fibres. J Mater Sci 42: 9336-9342

Segarra CI, Casalongue CA, Pinedo ML, Ronchi VP, Conde RD (2003) A germin-like protein of wheat leaf apoplast inhibits serine proteases. Journal of Experimental Botany 54: 1335-1341

Shafiq S, Li JR, Sun QW (2016) Functions of plants long non-coding RNAs. Bba-Gene Regul Mech 1859: 155-162

Shi G, Zhang Z, Friesen TL, Raats D, Fahima T, Brueggeman RS, Lu S, Trick HN, Liu Z, Chao W (2016) The hijacking of a receptor kinase–driven pathway by a wheat fungal pathogen leads to disease. Science advances 2: e1600822

Shiu S-H, Karlowski WM, Pan R, Tzeng Y-H, Mayer KF, Li W-H (2004) Comparative analysis of the receptor-like kinase family in Arabidopsis and rice. The plant cell 16: 1220-1234

Shuai P, Liang D, Tang S, Zhang ZJ, Ye CY, Su YY, Xia XL, Yin WL (2014) Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in Populus trichocarpa. Journal of Experimental Botany 65: 4975-4983

Sivaguru M, Ezaki B, He ZH, Tong HY, Osawa H, Baluska F, Volkmann D, Matsumoto H (2003) Aluminum-induced gene expression and protein localization of a cell wall-associated receptor kinase in Arabidopsis. Plant physiology 132: 2256-2266

Song DD, Yang Y, Yu B, Zheng BL, Deng ZD, Lu BL, Chen XM, Jiang T (2009) Computational prediction of novel non-coding RNAs in Arabidopsis thaliana. Bmc Bioinformatics 10

Struhl K (2007) Transcriptional noise and the fidelity of initiation by RNA polymerase II. Nat Struct Mol Biol 14: 103-105

Swart S, Logman TJJ, Smit G, Lugtenberg BJJ, Kijne JW (1994) Purification and Partial Characterization of a Glycoprotein from Pea (Pisum-Sativum) with Receptor Activity for Rhicadhesin, an Attachment Protein of Rhizobiaceae. Plant molecular biology 24: 171-183

Thompson E, Lane B (1980) Relation of protein synthesis in imbibing wheat embryos to the cell-free translational capacities of bulk mRNA from dry and imbibing embryos. Journal of Biological Chemistry 255: 5965-5970

Valenzuela-Avendaño JP, Mota IAE, Uc GL, Perera RS, Valenzuela-Soto EM, Aguilar JJZ (2005) Use of a simple method to isolate intact RNA from partially hydratedSelaginella lepidophylla plants. Plant Molecular Biology Reporter 23: 199-200

Vallelian-Bindschedler L, Mosinger E, Metraux JP, Schweizer P (1998) Structure, expression and localization of a germin-like protein in barley (Hordeum vulgare L.) that is insolubilized in stressed leaves. Plant molecular biology 37: 297-308

VanBuren R, Wai CM, Pardo J, Giarola V, Ambrosini S, Song X, Bartels D (2018) Desiccation tolerance evolved through gene duplication and network rewiring in Lindernia. The Plant Cell 30: 2943-2958

Verica JA, Chae L, Tong H, Ingmire P, He ZH (2003) Tissue-specific and developmentally regulated expression of a cluster of tandemly arrayed cell wall-associated kinase-like kinase genes in Arabidopsis. Plant physiology 133: 1732-1746

Verica JA, He ZH (2002) The cell wall-associated kinase (WAK) and WAK-like kinase gene family. Plant physiology 129: 455-459

Verma C, Singh R, Mishra S (2014) Biochemical characterization of pectin methylesterase from Musa acuminata referring to delayed ripening. Journal of Pharmacy and Biological Sciences 9: 42-47

Vicré M, Farrant JM, Driouich A (2004) Insights into the cellular mechanisms of desiccation tolerance among angiosperm resurrection plant species. Plant Cell Environ 27: 1329-1340

Vorwerk S, Somerville S, Somerville C (2004) The role of plant cell wall polysaccharide composition in disease resistance. Trends Plant Sci 9: 203-209

Wagner TA, Kohorn BD (2001) Wall-associated kinases are expressed throughout plant development and are required for cell expansion. Plant Cell 13: 303-318

Wang H, Chua NH, Wang XJ (2006a) Prediction of trans-antisense transcripts in Arabidopsis thaliana. Genome biology 7

Wang H, Chung PJ, Liu J, Jang I-C, Kean MJ, Xu J, Chua N-H (2014a) Genome-wide I dentification of long noncoding natural antisense transcripts and their responses to light in Arabidopsis. Genome Res 24: 444-453

Wang H, Niu HH, Liang MM, Zhai YF, Huang W, Ding Q, Du Y, Lu MH (2019a) A Wall-Associated Kinase Gene CaWAKL20 From Pepper Negatively Modulates Plant Thermotolerance by Reducing the Expression of ABA-Responsive Genes. Frontiers in plant science 10

Wang JJ, Meng XW, Dobrovolskaya OB, Orlov YL, Chen M (2017) Non-coding RNAs and Their Roles in Stress Response in Plants. Genom Proteom Bioinf 15: 301-312

Wang JY, Yang YW, Jin LM, Ling XT, Liu TL, Chen TZ, Ji YH, Yu WG, Zhang BL (2018) Re-analysis of long non-coding RNAs and prediction of circRNAs reveal their novel roles in susceptible tomato following TYLCV infection. Bmc Plant Biol 18

Wang JY, Yu WG, Yang YW, Li X, Chen TZ, Liu TL, Ma N, Yang X, Liu RY, Zhang BL (2016) Genome-wide analysis of tomato long non-coding RNAs and identification as endogenous target mimic for microRNA in response to TYLCV infection Sci Rep-Uk 6:

32828.

Wang KC, Chang HY (2011) Molecular mechanisms of long noncoding RNAs. Molecular cell 43: 904-914

Wang PF, Dai LM, Ai J, Wang YM, Ren FS (2019b) Identification and functional prediction of cold-related long non-coding RNA (lncRNA) in grapevine. Sci Rep-Uk 9

Wang T, Chen XP, Zhu FH, Li HF, Li L, Yang QL, Chi XY, Yu SL, Liang XQ (2013) Characterization of Peanut Germin-Like Proteins, AhGLPs in Plant Development and Defense. Plos One 8

Wang W, Barnaby JY, Tada Y, Li H, Tor M, Caldelari D, Lee DU, Fu XD, Dong XN (2011) Timing of plant immune responses by a central circadian regulator. Nature 470: 110-U126

Wang X-J, Gaasterland T, Chua N-H (2005) Genome-wide prediction and identification of cis-natural antisense transcripts in Arabidopsis thaliana. Genome biology 6: R30

Wang Y, Wang X, Deng W, Fan X, Liu T-T, He G, Chen R, Terzaghi W, Zhu D, Deng XW (2014b) Genomic features and regulatory roles of intermediate-sized non-coding RNAs in Arabidopsis. Molecular plant 7: 514-527

Wierzbicki AT, Haag JR, Pikaard CS (2008) Noncoding Transcription by RNA Polymerase Pol IVb/Pol V Mediates Transcriptional Silencing of Overlapping and Adjacent Genes. Cell 135: 635-648

Wight M, Werner A (2013) The functions of natural antisense transcripts. Essays in biochemistry 54: 91-101

Willats WG, McCartney L, Mackie W, Knox JP (2001) Pectin: cell biology and prospects for functional analysis. Plant molecular biology 47: 9-27

Wilusz JE, Sunwoo H, Spector DL (2009) Long noncoding RNAs: functional surprises from the RNA world. Genes Dev 23: 1494-1504

Woo EJ, Dunwell JM, Goodenough PW, Marvier AC, Pickersgill RW (2000) Germin is a manganese containing homohexamer with oxalate oxidase and superoxide dismutase activities. Nat Struct Biol 7: 1036-1040

Wu HJ, Wang ZM, Wang M, Wang XJ (2013) Widespread Long Noncoding RNAs as Endogenous Target Mimics for MicroRNAs in Plants. Plant physiology 161: 1875-1884

Wu J, Okada T, Fukushima T, Tsudzuki T, Sugiura M, Yukawa Y (2012) A novel hypoxic stress-responsive long non-coding RNA transcribed by RNA polymerase III in Arabidopsis. RNA biology 9: 302-313

Wu X (2016) Characterisation of wall-associated kinases (WAKs) in grasses. The university of Melbourne

Wunderlich M, Groß-Hardt R, Schöffl F (2014) Heat shock factor HSFB2a involved in gametophyte development of Arabidopsis thaliana and its expression is controlled by a heat-inducible long non-coding antisense RNA. Plant molecular biology 85: 541-550

Xia Y, Yin SJ, Zhang KL, Shi XT, Lian CL, Zhang HS, Hu ZB, Shen ZG (2018) OsWAK11, a rice wall-associated kinase, regulates Cu detoxification by alteration the immobilization of Cu in cell walls. Environ Exp Bot 150: 99-105

Xin M, Wang Y, Yao Y, Song N, Hu Z, Qin D, Xie C, Peng H, Ni Z, Sun Q (2011) Identification and characterization of wheat long non-protein coding RNAs responsive to powdery mildew infection and heat stress by using microarray analysis and SBS sequencing. Bmc

Plant Biol 11: 61

Yamada M (2017) Functions of long intergenic non-coding (linc) RNAs in plants. J Plant Res 130: 67-73

Yamahara T, Shiono T, Suzuki T, Tanaka K, Takio S, Sato K, Yamazaki S, Satoh T (1999) Isolation of a germin-like protein with manganese superoxide dismutase activity from cells of a moss, Barbula unguiculata. Journal of Biological Chemistry 274: 33274-33278

Yin K, Han XX, Xu ZH, Xue HW (2009) Arabidopsis GLP4 is localized to the Golgi and binds auxin in vitro. Acta Bioch Bioph Sin 41: 478-487

York. WS, Darvill. AG, McNeil. M, T.Stevenson. T, Albersheim P (1986) Isolation and characterization of plant cell walls and cell wall components. Methods in Enzymology 118: 3-40

Yuan C, Wang J, Harrison AP, Meng X, Chen D, Chen M (2015) Genome-wide view of natural antisense transcripts in Arabidopsis thaliana. DNA research 22: 233-243

Zehr BD, Savin TJ, Hall RE (1989) A one-step, low background Coomassie staining procedure for polyacrylamide gels. Analytical biochemistry 182: 157-159

Zhang S, Chen C, Li L, Meng L, Singh J, Jiang N, Deng XW, He ZH, Lemaux PG (2005) Evolutionary expansion, gene structure, and expression of the rice wall-associated kinase gene family. Plant physiology 139: 1107-1124

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

Zhang X, Lii Y, Wu Z, Polishko A, Zhang H, Chinnusamy V, Lonardi S, Zhu JK, Liu R, Jin H (2013) Mechanisms of small RNA generation from cis-NATs in response to environmental and developmental cues. Molecular plant 6: 704-715

Zhang X, Xia J, Lii YE, Barrera-Figueroa BE, Zhou X, Gao S, Lu L, Niu D, Chen Z, Leung C (2012) Genome-wide analysis of plant nat-siRNAs reveals insights into their distribution, biogenesis and function. Genome biology 13: R20

Zhu QH, Stephen S, Taylor J, Helliwell CA, Wang MB (2014) Long noncoding RNAs responsive to Fusarium oxysporum infection in Arabidopsis thaliana. New Phytologist 201: 574-584

Zhu QH, Wang MB (2012) Molecular Functions of Long Non-Coding RNAs in Plants. Genes-Basel 3: 176-190

Zuo C, Liu Y, Guo Z, Mao J, Chu M, Chen B (2018) Genome-wide annotation and expression responses to biotic stresses of the WALL-ASSOCIATED KINASE - RECEPTOR-LIKE KINASE (WAK-RLK) gene family in Apple (Malus domestica). European Journal of Plant Pathology

ACKNOWLEDGEMENTS

The dissertation took me an immense effort to finish, but I know that I cannot manage it without the supports and assistance from my supervisors, my friends and my family. I would like to express my sincere gratitude to them.

Firstly, the sincere thanks go to my conscientious and enthusiastic supervisor, Prof. Dr. Dorothea Bartels. She gave me the valuable opportunity to study in her lab in Germany. The amazing experience studying abroad will be unforgettable and beneficial to my academic career. During the four years in her lab, I learned a lot. Not only are my lab skills and academic English improved, but also my outlook on life and values are affected profoundly by her patience on guiding students, her scientific attitude to research and her immense knowledge. And I also greatly appreciate for her efforts on my thesis manuscript, which is very important for the completion of my PhD thesis. Besides, I would like to thank Prof. Dr. Lukas Schreiber, Prof. Dr. Peter Dörmann, and Prof. Dr. Barbara Reichert for being the supervisors of my doctoral committee.

Then my thanks also go to my colleagues. Dr. Valentino Giarola gave me many helpful suggestions for my research and provided the technical assistance to my experiments. Without his contribution, I cannot make big progress on my research. And my oral English was greatly improved thanks to the daily communication with my nice colleagues, Aishwarya Singh, Selvakumar Sukumaran, Ahmad Mollazadeh Taghipour, Dr. Juszczak, Ilona, Nasr Aziz and Niklas Jung, who also gave me a lot of assistance to my life. Furthermore, I am very grateful to my Chinese friends here, Dr. Junyi Zhao, Xiaomin Song, Zhenzhen Chu, Mengmeng Huang, Dr. Qingwei Zhang and Dr. Quancan Hou. It is difficult to study and live abroad. I cannot imagine how I can overcome the culture shock and the obstacles in life without their companionship and help.

I am also deeply indebted to my family. Whenever and wherever I am, the boundless love and supports from my parents give me the courage to strive for a better future. And I would like to offer my special thanks to Pei Xie for his encouragement, understanding and waiting. My great expectations for the future come from his love.

Finally I would like to acknowledge the scholarship from China Scholarship Council (CSC). Without the financial support from Chinese government, my dream studying abroad cannot come true.