Doctoral thesis: 0481017

Evolution of a chloroplast-resident basic helix-loop-helix protein in tobacco plants

Yutaka Kodama

Graduate School of Biological Sciences
Nara Institute of Science and Technology, JAPAN

Submitted on 29th January, 2007

TABLES OF CONTENTS

ABBREVIATIONS

bHLH	basic helix-loop-helix
BY2	bright yellow 2
bZIP	basic domain plus leucine zipper
CMV	cauliflower mosaic virus
CND41	41 kDa chloroplast nucleoid DNA binding protein
DAPI	4',6-Diamidino-2-phenylindole dihydrochloride
GFP	green fluorescence protein
GUS	*β*-glucuronidase
HR	hypersensitive response
LRP	light-responsive promoter
MeJA	methyl jasmonic acid
NsWIN4	*Nicotiana sylvestris* wound-induced clone 4
NtWIN4	*Nicotiana tabacum* wound-induced clone 4
PCR	polymerase chain reaction
PEND	plastid envelope DNA binding protein
PTF1	plastid transcription factor 1
R-Luciferase	*Renilla* luciferase
RT-PCR	reverse transcriptase-polymerase chain reaction
TMV	tobacco mosaic virus
UTR	untranslated region
5' RACE	5' rapid amplification of cDNA ends

INTRODUCTION

Since emerging on earth about 500 million years ago, plants have continuously been exposed to biotic and abiotic environmental stresses, including pathogens, herbivores, competing other plants, wounding, low nutrients, water deficiency, extreme temperatures, and low and high light exposure. All these stresses can cause serious damage to individual plants, and occasionally cause extinction of species. To maintain life, therefore, plants have continually evolved to best fit diverse environmental conditions.

One process in plant evolution is considered to be diversification of proteins, and fine genomic analyses constitute a powerful approach to elucidate underlying mechanisms. Information is already available from three plant species, Arabidopsis, rice and poplar, whose genomic sequences have been completely decoded (The Arabidopsis Genome Initiative 2000; Goff et al. 2002; Tuskan et al. 2006). The numbers of genes encoding proteins greatly differ among the three plants; 25,554 for Arabidopsis, 37,544 for rice, and 45,555 for poplar. Such diversity in gene numbers may suggest that plants have independently evolved functional proteins to fit environmental conditions by increasing copy numbers of genes and/or chance protein modification. Data bases also indicate that relative frequencies of protein motifs are similar in each genome (Tuskan et al. 2006), suggesting that positions and/or combinations of motifs are important for protein evolution. However, it remains to be determined how functional proteins have actually evolved, and data on which proteins have actually converted to other functions during plant evolution are also limited.

Plastidic proteins

The plastids are unique organelles of plants, comprising proplastids, chloroplasts, amyloplasts, leucoplasts, etioplasts and chromoplasts. They originated from prokaryotic cyanobacteria, and function in various regulatory systems including biosynthesis of chlorophylls, carotenoids and fatty acids (Buchanan et al. 2000). Proteomic analysis has revealed that chloroplasts contains 3,500 - 4,000 polypeptides, the majority of them considered to be transported into chloroplasts from the cytoplasm, since the maximal number of chloroplast genes is estimated to be about 200 (Race et al. 1999; Leister 2003). In general, transported proteins are encoded by genes residing in the nucleus, and their N-terminal regions are critical for transportation into plastids (Buchanan et al. 2000). Termed transit peptides, they are characterized by a lack of positively-charged amino acids in the N-terminal proximal region and acidic residues in the central region (Soll and Tien 1998). The central region is also rich in hydroxylated amino acids (Soll and Tien 1998) but detailed features of transit peptides remain to be determined.

Eukaryotic ancestors

Among transported proteins, several DNA binding proteins have been identified, and their origins shown to be not only from prokaryotes but also in a few cases from eukaryotes (Sato 2001). Proteins of eukaryotic ancestors possess motifs of nuclear transcription factors, suggesting them to be converted from the latter during evolution. To date, five such proteins have been identified from plastids (Nakano et al. 1997; Sato et al. 1993, 1995, 1998; review in Sato 2001; Baba et al. 2001). One chloroplast nucleoid-DNA binding protein was isolated from *Nicotiana tabacum*. This protein of 41 kDa was designated as CND41, and shown to contain a zinc finger motif (Nakano et al.

1993). A plastid envelope DNA binding protein was isolated from *Pisum sativum*, named PEND (Sato et al. 1993) and found to possess a basic domain plus leucine zipper (bZIP) motif. Two additional proteins, PD1 and PD3, were also isolated from *P. sativum*, and shown to have two and five AT-hook motifs, respectively (Sato et al. 1995). A plastid transcription factor 1 (PTF1) equipped with a basic helix-loop-helix (bHLH) motif was similarly identified from *Arabidopsis thaliana* (Baba et al. 2001). These five plastid-resident proteins are thought to have been functionally converted from transcription factors during evolution (review in Sato 2001; in Kodama 2007). The following section describes current knowledge of the individual proteins.

CND41

During purification of chloroplasts-nucleoids from *N. tabacum*, a 41 kDa nucleoid DNA binding protein, CND41, was isolated by two-dimensional polyacrylamide gel electrophoresis and southwestern analysis (Nakano et al. 1993) and found to be present in both chloroplasts and etioplasts. The calculated isoelectric point was basic (Nakano et al. 1993) and two distinct domains were identified, a helix-turn-helix motif in the lysine-rich N-terminal region of mature protein, and an aspartyl protease active site (Nakano et al. 1997). *In vitro* assays showed CND41 to bind to DNA non-specifically through its lysine-rich region (Table 1) (Nakano et al. 1997). CND41 has strong proteolytic activity at acidic pH (pH2-4) *in vitro* (Murakami et al. 2000), and transgenic tobacco plants expressing antisense *CND41* clearly showed a role in controlling transcripts for *psbA*, *psbD/C*, *rbcL*, *16SrDNA* and *23SrDNA*, all encoded by chloroplast DNA (Kato et al. 2004; Kato et al. 2005). CND41 is also involved in degradation of ribose-1,5-bisphosphate carboxylase/oxygenase through post-translational regulation in

senescent leaves (Kato et al. 2004; Kato et al. 2005). In addition, CND41 appears to regulate plastid development and gibberellin biosynthesis (Nakano et al. 2003). These results strongly suggested that CND41 plays a critical role in transcription of the chloroplast genome, and in regulation of senescence, plastid development and gibberellin biosynthesis. Structural analysis indicated that CND41 possesses a zinc finger motif (Cys-X4-Cys-X2-Leu-X9-Cys-X4-Cys-X-Tyr) at amino acid positions 210-235 (Figure 1, Table 1). Zinc finger motifs have been identified in many eukaryotic proteins, and shown to interact with nuclear DNA (O'Halloran 1993). The motif found in CND41 closely resembles that of transcription factor IIIA from yeast (His-X4-His-X2-Leu/Met-X9-Cys-X4-Cys-X-Tyr/Phe) (Archambault et al. 1992; Nakano et al. 1997), albeit with histidine residues in transcription factor IIIA substituted with cysteine residues in CND41. These observations suggest that a zinc finger protein initially located in the nucleus was converted to a chloroplast-localized DNA binding protease.

PEND

Basic domain plus leucine zipper (bZIP) proteins are eukaryotic transcription factors, having both DNA binding and dimerization motifs of approximately 60 amino acids (Bohmann et al. 1987; Landschulz et al. 1988; Struhl 1987; Vogt et al. 1987). The plastid envelope DNA binding protein (PEND) from *P. sativum* contains a distinct bZIP motif, an apparently unique sextuple repeat region and a membrane-spanning region (Figure 1, Table 1) (Sato et al. 1993, 1998). PEND binds to a specific DNA sequence, TAAGAAGT, and is localized to the inner envelope membranes of chloroplasts (Table 1) (Sato et al. 1993, 1998; Sato and Ohta 2001), indicating a function in binding of

chloroplast DNA to envelope membranes (Sato et al. 1993, 1998). PEND homologs were recently identified in angiosperms, and found to localize in chloroplasts, suggesting a common function among PEND homologs (Terasawa and Sato 2005). Biological activity of a PEND homolog, BnPEND, from *Brassica napus* was examined using transgenic over-expressing tobacco plants under control of the cauliflower mosaic virus (CaMV) 35*S* promoter (Wycliffe et al. 2005). Leaves of resulting transgenic tobacco were shown to develop a chlorophyll-deficit, and few palisade cells and chloroplasts. These observations point to a PEND evolution from a nuclear bZIP transcription factor to an envelope-resident DNA binding protein, and that it interferes with plastid development leading to a distorted leaf composition.

PD1 and PD3

During research into the PEND protein (Sato et al. 1993, 1998), two cDNAs for other plastidic DNA binding proteins were identified from *P. sativum* by southwestern analysis using the psbM region of chloroplast DNA as the probe (Table 1) (Sato et al. 1995). These were designated as *PD1* and *PD3* genes and PD1 and PD3 proteins possess AT-hook motifs, originally found in high mobility group I (Y) proteins, and could be shown to interact with the minor groove AT-rich regions of nuclear DNA (Grasser 1995) (Table 1). AT-hook is a small motif, which has a typical sequence pattern of centered glycine-arginine-proline (GRP) tri-peptides (Reeves and Nissen 1990). PD1 has two AT-hook motifs at the N-terminus (Sato et al. 1995) (Figure 1). PD3 was previously reported to possess five AT-hook motifs (Sato et al. 1995), but scanning the latest PD3 amino acid sequence identified only four GRP tri-peptides (Figure 1). PD3 was also shown to contain a jmjC domain at the C-terminus, this possibly being

involved in chromatin organization by modulating heterochromatisation (described in the NCBI database; accession number CAA67296). Despite limited functional analyses, the structural information suggests that these proteins were converted to plastid DNA binding proteins from nuclear transcription factors containing AT-hook motifs.

PTF1

Basic helix-loop-helix (bHLH) proteins possess a motif, consisting of a basic region to bind to the E-box sequence (CANNTG) and a helix-loop-helix region to function to form homo- and/or hetero-dimers (Murre et al. 1989; Ferre-D'Amare et al. 1994). A number of bHLH proteins have been identified in both plants and mammals, and these are thought to constitute a family of transcription factors located within nuclei (Garrell and Modolell 1990; Quail 2000). An exceptional case is a plastid-localized bHLH protein from *A. thaliana.* This protein was isolated by yeast one-hybrid screening methods using a region derived from *psbD* light-responsive promoter (LRP) from chloroplast DNA (Baba et al. 2001) (Figure 1), and designated as plastid transcription factor 1 (PTF1). PTF1 was found to bind to the ACC repeat region of the *psbD* LRP sequence (Baba et al. 2001) (Table 1), and green fluorescent protein (GFP)-tagged PTF1 was shown to localize into chloroplasts of tobacco guard cells. A PTF1-deficient mutant, *ptf1,* showed several abnormalities, such as reduced activity of *psbD* LRP under continuous light conditions, and early bleaching, late flowering and dwarfism under short-day conditions. Based on these findings, PTF1 was proposed to regulate *psbD* LRP through its transcription in plastids, and was considered that a nuclear bHLH type transcription factor converted to a plastidic transcription factor.

A novel plastidic ancestor and aim of this thesis

During a past decade, we have screened tobacco genes, of which transcripts accumulated immediately after wound and pathogen stresses, and identified more than 100 clones (Hara et al. 2001; Yoda et al. 2002; Sugimoto et al. 2003; unpublished observations). Their identities are diverse, and the majority of transcripts was found to accumulate within one hour after mechanical injury and infection by tobacco mosaic virus (TMV), suggesting there critical role in defense responses. In this study, one particular clone, designated as *NtWIN4* (*N. tabacum* wound-induced clone 4), was subjected to further characterization, because of its specific feature encoding a novel bHLH protein localized to chloroplasts of tobacco leaves. This thesis describes novel finding concerning with its physiological function and plastid localization mechanism. A series of experiments lead me to propose an idea on evolution of plastidic ancestors.

Composition of this thesis

This thesis consists of three chapters. In chapter I, plastid localization and physiological function of NtWIN4 in *N. tabacum* are described. NtWIN4 possessed a nuclear transcriptional repression activity under Gal4-system with dual-luciferase assay, but it localized to chloroplasts of tobacco leaves. *NtWIN4* was up-regulated by wounding and pathogen attack, and regulated formation of hypersensitive cell death through chloroplast disruption. In chapter II, comparative analysis as to plastid localization between NtWIN4 and PTF1 is described. Although PTF1 is previously reported as a plastid-localized bHLH protein from *A. thaliana*, present observation suggested that it functions as a nuclear transcriptional repressor, but not as a plastid-localized protein. In chapter III, functional diversification of NtWIN4 due to alternative transcription during

generation of amphidiploidy is described. Results suggested natural selection of NtWIN4 after amphiploidization of tobacco and diversity of mRNA length in tobacco and its ancestral plant. These two events were considered to develop a plastid-type NtWIN4 during evolution of *N. tabacum*. In concluding remarks, a common structure among plastidic ancestors is proposed, and its involvement in changing localization from nuclei to plastids is discussed.

Table 1. Properties of plastid-resident eukaryotic ancestors

Name	Motif	Bound sequence	Presequence	Recognition sequence
CND41	Zinc finger	Non-specific	120 aa [c]	nd
PEND	bZIP	TAAGAAGT	15 aa	~265 aa
PD1	AT-hook	psbM in cpDNA[a]	nd	nd
PD3	AT-hook	psbM in cpDNA[a]	nd	nd
PTF1	bHLH	ACC repeat	nd	nd
NtWIN4	bHLH	nd [b]	nd	67 aa

The presequence indicates the number of amino acids that are cleaved out and the recognition sequence gives those that are essential for plastid localization.

[a] cpDNA; chloroplast DNA, [b] nd.; not determined, [c] aa; amino acids.

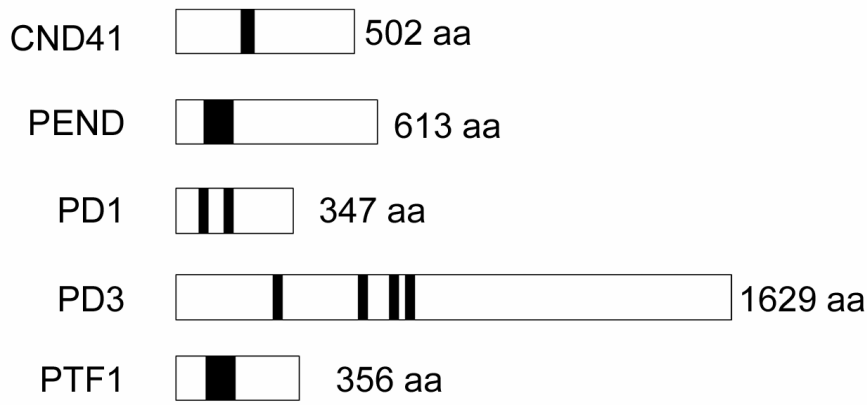

Figure 1. Schematic illustration of eukaryotic ancestor proteins in plastids. Closed boxes indicate motifs of nuclear transcription factors. Motifs for zinc finger and bZIP are shown in CND41 and in PEND, respectively. In PD1 and PD3, AT-hook motifs are indicated. bHLH motif is present within PTF1 protein. Lengths of amino acids are shown on the right side.

CHAPTER I

A tobacco chloroplast-resident bHLH protein involved in defense response

Introduction

Higher plants are constantly exposed to diverse environmental stresses, including physical wounding. Under natural conditions, the major cause is grazing by herbivores, which continuously damages tissues, occasionally resulting in whole plant death after secondary infection by pathogens. To cope with this problem, plants have evolved a variety of self-defense mechanisms, including production of proteins such as proteinase inhibitors, and of signaling components such as jasmonate and systemin (Green and Ryan 1972; Farmer and Ryan 1990; Pearce et al. 1991). Wound signals are quickly transmitted from damaged cells to the whole plant, allowing elaboration of a defense reaction by transcriptionally up- and/or down-regulating a number of genes (Schilmiller, and Howe 2005). Examples are proteinase inhibitors against insect feeding, extensions for reinforcement of cell walls, phenylalanine ammonia-lyase (EC 4.3.1.5) and chalcone synthase (EC 2.3.1.74) for phenylpropanoid synthesis, 1-aminocyclopropane-1-carboxylate (ACC) synthase and ACC oxidase for ethylene synthesis (Green and Ryan 1972; Corbin et al. 1987; Kende 1993). Their functions, however, have not been completely clarified.

One of phenotypical hallmarks of wounding is tissue necrosis. Cells at damaged areas rapidly die, thereby forming a physical barrier to secondary injury such as pathogen attack. Biochemically, chlorophyll degradation is one characteristic symptom of wounding, this being called chlorosis (Buchanan et al. 2000). Recent studies have shown that this occurs not only by wounding but also during pathogen responses,

suggesting that chlorophyll degradation is a common response in stressed plants. For example, the photosystem II complex was reported to be disrupted during the hypersensitive response (HR) upon inoculation of TMV (Culver 2002). Accumulation of TMV coat proteins in chloroplasts has been suggested to correlate with severe chlorosis (Culver 2002). Another example of a chloroplast protein involved in HR is chloroplast-resident FtsH, a member of AAA group of proteins, whose reduction results in acceleration of the HR in tobacco plants (Seo et al. 2002). These observations indicate a close correlation between disruption of chloroplast function, including chlorosis, and HR, although the underlying molecular mechanisms remain to be determined.

During a past decade, we have screened genes for which transcripts accumulate immediately after wound stress by fluorescence differential display and so far isolated more than 50 clones (Hara et al. 2000; unpublished observations). Although their identities are diverse, the majority of transcripts were found to accumulate within one hour of mechanical injury, suggesting critical roles in defense responses. In this study, one particular clone was subjected to further characterization, because of its specific feature encoding a novel bHLH protein localized to chloroplasts and contributing to HR through chlorosis.

Materials and Methods

Plant Materials and Treatment

Tobacco (*Nicotiana tabacum* cv Xanthi nc) plants were grown in soil in a growth cabinet at 23^{o}C under a 14-h-light/10-dark photo-cycle. Protoplasts of above-descibed plants, tobacco bright yellow 2 (BY2) and Arabidopsis MM2d cells were maintained in

suspension culture using modified Linsmaier and Skoog medium (Nagata et al. 1981) at 23^{o}C continuously in the dark. Wound stress was applied by cutting mature leaves of *N. tabacum* with a pair of scissors or a leaf punch. Wounded (local) and adjacent upper un-wounded (systemic) leaves were harvested at appropriate time points. For TMV infection, leaves of 2-month-old wild-type plants were inoculated with TMV (10 μg/ml) and incubated at 30^{o}C under continuous light for 48 h and then at 20^{o}C (temperature shift) as previously described (Yoda et al. 2002). Samples were harvested at appropriate time points. For treatment with stress-signaling molecules, leaf disc samples were transferred into a solution containing 500 μM salicylic acid (SA), 50 μM methyl jasmonic acid (MeJA), 100 μM ethephone (for ethylene production) or 10 mM hydrogen peroxide and harvested after 12 h. For time course analyses, healthy leaves were treated with 50 μM MeJA or 100μM paraquat and were harvested at appropriate time points. Bacterial infection was performed as previously reported (Sugimoto et al. 2004), with inoculation of *P. syringae* by infiltration and observation after 48 h.

Isolation of NtWIN4 cDNA

A DNA fragment encoding NtWIN4 was initially obtained from a cDNA population derived from wounded tobacco leaves by fluorescence differential display (Hara et al. 2000). The full length cDNA sequence was successively isolated by 5'RACE using a primer (5'-AGAAGAGACGAAGGTGCTTTTTCCAG-3') and a Marathon cDNA Amplification kit (Clontech, Palo Alto, CA) and by the thermal asymmetric interlaced (TAIL) polymerase chain reaction (PCR) method (Liu and Whittier 1995) using three TAIL-primers (5'-NGTCGASWGANAWGA-3', 5'-GTNCGASWCANAWGT-T-3', and 5'-WGTGNAGWANCANAGA-3') and three *NtWIN4* primers

(5’-CCATTGG-GAGCAATTTATGCAAATCCAAAT-3’; 5’-CTCCATCATCCTCTTG TTCATGTTC-CGATG-3’; 5’-TCGCCGATACATTTCCATAGCCAGAAACTC-3’). The nucleotide sequence of isolated DNA was determined for both strands with a Big Dye terminator sequencing kit (Applied Biosystems, Foster City, CA). The 5’terminus was determined by 5’ rapid amplification of cDNA ends (5’RACE) with the primer, 5’-ATGATTCGCCGATACATTTCCA-3’ (nucleotide position from the 5’ end; 161-182).

DNA and RNA Analyses

Genomic DNA was isolated from green leaves by the cetyl-trimethyl-ammonium bromide method (Murray and Thompson 1980) with modifications, and a 20 μg aliquot was digested with one of a series of restriction enzymes, BamHI, EcoRV or HindIII. Total RNA was isolated by the aurintricarboxylic acid method (Gonzalez et al. 1980) and DNA and RNA gel-blot hybridization was performed as previously described (Tamura et al. 2003) using a *NtWIN4* cDNA fragment as the probe. For PCR and reverse transcriptase-polymerase chain reaction (RT-PCR) analyses, *NtWIN4* genomic and cDNA fragments were amplified by PCR with *ExTaq* DNA polymerase (Takara, Ohtsu, Japan) using forward (5’-GCTCTCGAGATGAATCAAAGTGCTTTTGC-3’) and reverse (5’-TCACCCATGGGTTCGTTCTGTGGCTGAAG-3’) primers.

Protein Analyses – Tobacco leaves were ground in liquid nitrogen in a mortar, mixed with isolation buffer containing 50 mM Tris-HCl (pH7.5), 100 mM NaCl, 0.05% (w/v) Tween 20, 1 mM EDTA, 1 mM phenylmethylsulfonyl fluoride, 5 μg/ml leupepin, 5 μg/ml aprotinin and homogenized. After centrifugation at 15,000 rpm for 10 min to remove insoluble debris, crude extracts were directly used for western blot hybridization.

After fractionation by SDS-PAGE, proteins were electronically transferred onto cellulose membranes (Immobilon-N, Millipore, Bedford, MA), and NtWIN4 was detected with rabbit antibodies raised against a synthetic peptide (MIKERIRREKQKQSY) and anti-rabbit HRP conjugate antibodies (Bio-Rad Laboratories, Hercules, CA). Fractionation of cytoplasmic, mitochondria and chloroplastic fractions was performed as described (Nakamura et al. 2000; Yasuno and Wada 1998; Fan et al. 1999; Benkova et al. 1999).

Nuclear transcription assays

The *Gal4BD:NtWIN4* and *Gal4BD:At5g43650* effector plasmids were constructed by fusing a cDNA encoding NtWIN4 or At5g43650 with the Gal4 DNA binding domain in a yy64 vector, a derivative of pMA560 (Yamamoto and Deng 1998), respectively. *NtWIN4* and *At5g43650* were thus subcloned into the BglII/SalI sites of yy64, respectively. As an effector control, the yy64 vector alone was used. The reporter plasmid, yy96, contained a luciferase gene placed under control of the Gal4 binding site (Yamamoto and Deng 1998). An internal control plasmid, containing an R-luciferase gene placed under control of the CaMV35*S* promoter, was used to normalize for differences in bombardment efficiency. Five-day old BY2 and MM2d cells were plated on 1/2x Murashige-Skoog agar and bombarded with plasmids (effecter:reporter:reference=2:2:1) coated on a 1.0 μm microcarrier, under a vacuum of 28 inches of mercury using a helium pressure of 1,100 psi (PDS 1000, Bio-Rad). Cells were placed 6 cm from the stopping screen. After bombardment, they were incubated in the dark at 28^{o}C for 24 h. Luciferase and *Renilla* luciferase (R-luciferase) activities were assayed using a dual-luciferase assay kit (Promega, Madison, WI) according to the

manufacturer's instruction. Chemical luminescence was measured using a luminometer (Lumat LB9507, Berthold Technologies, Bad Wildbad, Germany).

Epifluorescence Analyses

Full-length or partial cDNA regions of *NtWIN4* were subcloned into the SalI/NcoI sites of a CaMV35S-sGFP(S65T)-nos vector, harboring a synthetic gene for improved green fluorescent protein (GFP) sGFP(S65T) driven by the CaMV35*S* promoter and *NOS* terminator (Chiu et al. 1996). Onion epidermal cell layers were bombarded with gold particles (Bio-Rad) coated with appropriate vector constructs, and after incubation at 28^{o}C for 6 - 12 h in the dark, samples were viewed under a microscope (Olympus PROVIS AX70 and Leica MZ FLIII) equipped with a fluorescence module.

Transgenic and Agro-infiltration Analyses

A cDNA fragment encoding NtWIN4$^{27\text{-}247}$ was amplified by PCR and subcloned into BamHI/SacI sites of pBI121 (Clontech). Resulting plasmid, pBI-NtWIN4$^{27\text{-}247}$, was transformed into *Agrobacterium tumefaciens* strain LBA4404 (Hiei et al. 1994), and used for stable transformation of tobacco as previously described (Uefuji et al. 2005). For agro-infiltration, *A. tumefaciens* LBA4404 transformed with pBI-NtWIN4$^{27\text{-}247}$ was grown on a LB agar plate for 2 days under dark, and suspended in a solution containing 10 mM morpholineethanesulfonic acid-NaOH (pH 5.5) and 200 μM acetosyringone to give an optical density of 0.5 at 600 nm. Bacterial solution was directly infiltrated into a leaf segment as described (Ueda et al. 2006). The chlorophyll contents in infiltrated regions were measured by the chlorophyll meter (SPAD-502, Minolta, Tokyo, Japan). Chlorophyll fluorescence was measured with a pulse amplitude-modulated fluorometer

(MINI-PAM, Heinz-Walz, Effeltrich, Germany). The minimum chlorophyll fluorescence at the open photosystem II center (Fo) was determined by measuring light at a light intensity of 0.05 - 0.1 μmol photons $m^{-2}sec^{-1}$. A saturating pulse of white light (800 msec, 8000 μmol photons $m^{-2}sec^{-1}$) was applied to determine the maximum chlorophyll fluorescence at the closed photosystem II center in the dark (Fm). The maximum quantum yield of photosystem II was calculated by equation (Fm-Fo)/Fm.

Results

Isolation of NtWIN4

One clone of over 50 genes identified by fluorescent differential display activated immediately after mechanical wounding in tobacco leaves (Hara et al. 2000), tentatively named *NtWIN4* (*Nicotiana tabacum* wound-induced clone 4), was further characterized in this study. Determination of its full length sequence by 5'RACE, yielded a 949 bp cDNA. *NtWIN4* transcripts transiently accumulated only in local leaves within 30 min after wounding, reaching maximal levels at 1 h and declining thereafter (Figure 2A). RT-PCR indicated *NtWIN4* to be constitutively expressed at a low level even in unwounded leaves, suggesting a basic function (Figure 2B). The copy number of *NtWIN4* in the genome of *N. tabacum* was estimated by Southern blot analysis by digesting genomic DNA with BamHI or HindIII, for which sites are not present in the *NtWIN4* cDNA probe sequence. Single fragments of 10 kb and 7 kb, respectively, were yielded (Figure 2C). Genomic DNA was also digested with EcoRV, for which two sites exist 132 bp apart in the center of the second intron of *NtWIN4* genome, and two fragments of 9 kb and 3.2 kb were clearly identified (Figure 2C). These results all pointed to a single copy of the *NtWIN4* gene in the *N. tabacum* genome.

Transcriptional Repression Activity

Full length *NtWIN4* was found to encode a protein of 247 amino acids with an apparent molecular mass of 28.7 kDa (Figure 3A). A data base search demonstrated possession of a bHLH motif, characteristic of a transcriptional factor. A PHI-BLAST homology search indicated close similarity at the bHLH motif and the C-terminal region with proteins from Arabidopsis (At5g43650) and rice (OJ1365_D05) (Figure 3A). To test whether or not NtWIN4 and At5g43650 proteins exhibit transcriptional activity, nuclear transcription assays were carried out *in vivo* by the dual-luciferase assay method. Effector constructs of *Gal4BD:NtWIN4* and *Gal4BD:At5g43650* were co-bombarded with a reporter plasmid and a reference plasmid into tobacco BY2 and Arabidopsis MM2d cells, respectively. Results showed *Gal4BD:NtWIN4* and *Gal4BD:At5g43650* effector constructs to suppress the luciferase activity to a level of 1/3 that of the *Gal4BD* control (Figure 3B). The results indicated that NtWIN4 and At5g43650 proteins possess transcriptional repression activity in nucleus, suggesting their structures to confer common functions.

Cellular Localization

In order to confirm nuclear localization, GFP was fused to the C-terminus of NtWIN4 or At5g43650 and assayed in onion epidermal cell layers after particle bombardment. GFP fluorescence from both NtWIN4:GFP and At5g43650:GFP resembled that from the control GFP alone, suggesting both to be localized in the cytoplasm and/or nucleus (Figure 4A). To verify this result, the exact localization of NtWIN4 was biochemically examined using antibodies raised against synthetic peptides. First, total proteins were extracted from tobacco leaf tissues, fractionated by SDS-PAGE and subjected to

immuno-blot staining. A clear single signal was found at the 17 kDa position, suggesting the native form of NtWIN4 to be 17 kDa (Figure 4B). This was confirmed by increase of the 17 kDa fraction upon wounding of leaf tissues, showing maximal levels after 1 h (Figure 4C), the pattern being consistent with that of transcript accumulation (Figure 2A). Second, cellular components were fractionated and subjected to immuno-assay. When cytoplasm, mitochondria and chloroplasts were mechanically separated and analyzed with antibodies, 17 kDa NtWIN4 was detected in fractions containing total proteins and chloroplasts, but not in cytoplasmic and mitochondrial fractions (Figure 4D). The chloroplast fraction was further examined for purity with control antibodies against compartment-specific proteins: chloroplast-resident large subunit of ribulose-1,5-bisphosphate carboxylase/oxgenase, nuclear-specific domains rearranged methyltransferase (Wada et al. 2003) and cytoplasmic-specific phosphoenol-pyruvate carboxylase. It was evident that the chloroplast fraction did not contain any contaminating nuclear fractions (Figure 4E). These results showed the native NtWIN4 to be 17 kDa and exclusively localized in chloroplasts. No molecular species of 28 kDa NtWIN4 was detectable in any fraction of tobacco plants, and therefore GFP analyses with an artificially constructed full length 28 kDa protein (Figure 4A) could not show the exact localization of the native NtWIN4. However, since the computer predicted molecular mass of NtWIN4 is 28.7 kDa, processing must occur *in planta*, facilitating movement into chloroplasts. Subsequently, the N-terminus of NtWIN4 was successively dissected to identify the signal peptide for chloroplast import by GFP assay. When started from the second Met at the position 27, the protein was distinctly incorporated into plastid, but not into mitochondria or peroxisomes (Figure 4F). Thus the chloroplast-transit signal is located in peptides lacking the first 26

amino acids, and therefore further analyses were performed with the protein starting from the second Met designated as NtWIN4$^{27\text{-}247}$ (Table 2). To determine whether the Arabidopsis counterpart contains such a transit signal, a mutated protein lacking the first 30 amino acids (At5g43650$^{31\text{-}114}$) was constructed for GFP fusion, although the native At5g43650 protein does not possess the second Met at the corresponding position to NtWIN4. The results clearly showed At5g43650$^{31\text{-}114}$ to be localized in plastids, suggesting that amino acid sequences which are potentially able to serve as signals for plastid localization are conserved among NtWIN4 homolog proteins (Figure 4F).

Transit Peptides for Chloroplast Localization

No apparent chloroplast-transit signal peptide in NtWIN4 and NtWIN4$^{27\text{-}247}$ was predicted from silico analysis with the transit peptide prediction website, ChloroP 1.1 (http://www.cbs.dtu.dk/services/ChloroP/), or the localization prediction websites, TargetP 1.1 (http://www.cbs.dtu.dk/services/TargetP/) and WoLF PSORT (http://wolfpsort.seq.cbrc.jp/) (Emanuelsson et al. 1999; Emanuelsson et al. 2000; Horton et al. 2006). However, when sequence only at the N-terminus region (from position 27 to positions between 73 and 148) was subjected to prediction by TargetP 1.1, putative chloroplast transit peptide was identified (data not shown). Similarly the N-terminus region of At5g43650 (from position 42 for the second Met to positions between 97 and 150) was predicted to contain a possible transit peptide, which showed a quite high but below the threshold probability of chloroplast localization (data not shown). Hence it was probable that these proteins originally possessed non-canonical chloroplast transit peptides. In order to substantially identify the N-terminal signal sequence, a set of truncated N-terminal regions was fused to the N-terminus of GFP

(Table 2), and transiently expressed in onion epidermal cell layers by particle bombardment. Results showed NtWIN4$^{27-78}$:GFP localized in the cytoplasm, whereas NtWIN4$^{27-111}$:GFP and NtWIN4$^{27-135}$:GFP proteins were clearly located in plastids (Figure 5A). The region between 78 and 111 amino acids contains the basic region and the first helix domain of bHLH motif, suggesting a critical role. Subsequently, fine dissection assay was carried out within these regions. NtWIN4$^{27-95}$:GFP was distinctly localized to plastids, while NtWIN4$^{27-87}$:GFP was located in the cytoplasm (Figure 5B). This indicated that eight amino acids between the positions 87 and 95 were critical for signaling. Further assays revealed NtWIN4$^{27-91}$:GFP and NtWIN4$^{27-92}$:GFP in cytoplasm, but NtWIN4$^{27-93}$:GFP and NtWIN4$^{27-94}$:GFP in plastids (Figure 5B). It was thus concluded that the chloroplast localization signal was 93 amino acids from the N-terminus of NtWIN4 (Figure 5B, Table I). This was confirmed with tobacco cultured cells, in which NtWIN4$^{27-87}$:GFP and NtWIN4$^{27-95}$:GFP were localized to the cytoplasm and chloroplasts, respectively (Figure 5C). The molecular mass of NtWIN4 lacking the first 93 amino acids is computer predicted to be approximately 17.7 kDa, similar to the native NtWIN4 protein. These results indicated that the transit peptide extends into the basic region, which usually functions as the DNA binding domain of bHLH motif, and therefore suggested that native NtWIN4 might be unable to bind to DNA in chloroplasts. Since Arabidopsis At5g43650$^{31-114}$:GFP was clearly shown to localize to plastids (Figure 4F), its DNA binding domain presumably also overlaps with a chloroplast localization signal, suggesting that the basic region of NtWIN4 homologs can potentially serve as a transit signal, whenever such a mutated structure lacking the first 30 or so amino acids is available.

Minimal Requirement for a Transit Peptide

To determine the minimal size which is required for function as a transit peptide, dissection assays were carried out by constructing peptides, initiated from variable positions at the N-terminus and terminating at position 111 (Figure 6A). First, upstream regions from position 27 (Met) were examined. NtWIN4$^{7\text{-}111}$:GFP was found localized in cytoplasm, while NtWIN4$^{14\text{-}111}$:GFP and NtWIN4$^{21\text{-}111}$:GFP were localized in plastids, suggesting the minimal peptide between 8 and 13 amino acids from the N-terminal (Figure 6B). Subsequent fine dissection assays indicated NtWIN4$^{8\text{-}111}$:GFP, NtWIN4$^{9\text{-}111}$:GFP, NtWIN4$^{10\text{-}111}$:GFP, NtWIN4$^{11\text{-}111}$:GFP and NtWIN4$^{12\text{-}111}$:GFP to be localized in cytoplasm, and NtWIN4$^{13\text{-}111}$:GFP in plastids (Figure 6B). Downstream regions from the position 27 (Met) were then examined. NtWIN4$^{31\text{-}111}$:GFP and NtWIN4$^{36\text{-}111}$:GFP were localized in plastids, while NtWIN4$^{40\text{-}111}$:GFP were again found in cytoplasm, suggesting the minimal C-terminus at positions between 36 and 39 amino acids (Figure 6C). Fine dissection assays within this region showed that NtWIN4$^{37\text{-}111}$:GFP, NtWIN4$^{38\text{-}111}$:GFP and NtWIN4$^{39\text{-}111}$:GFP were distinctly localized into plastids (Figure 6C). These results revealed that, in order to function as the transit peptide, amino acid residues initiating from positions between 13 and 39 are prerequisite. The second Met locates within this frame at position 27. It was concluded that the region between 39 and 95 amino acids is essential as transit peptide.

Expression Profile

The above observations implied that the native NtWIN4 does not function as a nuclear transcriptional regulator. The question then arises as to its biological role in chloroplasts. *NtWIN4* was initially identified as a wound-induced gene (Figure 2A), suggesting a

contribution to stress responses. Subsequently, transcript induction by stress-signaling molecules was examined by treating leaves of wild type tobacco plants with salicylic acid, methyl jasmonic acid (MeJA), ethephone (for ethylene production) or hydrogen peroxide. RT-PCR showed transcripts to be apparently induced by MeJA and hydrogen peroxide (Figure 7A), and time course analyses confirmed this, exhibiting transient accumulation with MeJA at 3 h (Figure 7B), and with paraquat (a hydrogen peroxide generator) after between 1 and 3 h (Figure 7C). The latter result suggested NtWIN4 to respond to reactive oxygen species, which are generated with various stresses and play critical roles as inducers of defense reaction. One such activity is the induction of programmed cell death, which typically occurs during pathogen responses (Yoda et al. 2002). Accordingly the effect of biotic stress was examined by challenging leaves with TMV. When tobacco plants carrying the resistant gene (*N* gene) are inoculated with TMV and incubated at 30^{o}C, at which temperature the *N* gene does not function, viral particles multiply. Upon transfer to 20^{o}C, the *N* gene is activated, resulting in a resistance called the HR (Gianinazzi 1970). Results using this experimental system showed that *NtWIN4* transcripts were induced 3 h after the onset of this HR and accumulated thereafter up to 9 h (Figure 7D). These observations suggest that NtWIN4 might play a common role in defense system against biotic and abiotic environmental stresses.

Chloroplast Disruption

In order to obtain a more precise idea of its role, transgenic tobacco plants were generated in which *NtWIN4*$^{27\text{-}247}$ was constitutively expressed under control of the CaMV35*S* promoter. Among several transgenic lines, two lines were finally selected

due to high levels of NtWIN4 production as identified by immuno-blot hybridization (Figure 8A top panel). However, both plantlets grew slowly and abnormally (Figure 8A bottom panel), and ultimately died at very young stages. When stained with trypan-blue, many cells were found to be dead (data not shown), indicating the transgene to be highly detrimental to growth. For further analysis, *NtWIN4*$^{27\text{-}247}$ was transiently expressed in tobacco leaves by agro-infiltration, which expresses the transgene in only infected regions. The *β-glucuronidase* (*GUS*) gene was used as the control. Infiltrated samples were incubated under light or dark conditions for several days, and the phenotype was observed. Under light conditions, the infiltrated regions with *NtWIN4*$^{27\text{-}247}$ were apparently damaged, but the control was also injured although the extent was less (data not shown). This could due to light stress, additively damaging the infiltrated tissues. In the dark, however, regions where *NtWIN4*$^{27\text{-}247}$ was expressed clearly showed chlorosis 5 days after infiltration and distinct tissue necrosis 7 days later (Figure 8B). Control infiltration with *GUS* did not cause any damage during this time period (Figure 8B). The relative chlorophyll contents of the affected regions were found to have decreased to less than half and 1/3 that of the controls at days 4 and 5, respectively (Figure 8C). This resulted in a clear reduction of maximum activity of photosystem II photochemistry, to less than half that of the controls 4 days after agro-infiltration (Figure 8D). Thus NtWIN4 appears to be involved in chloroplast disruption.

Hypersensitive Cell Death

To further examine the gene function, NtWIN4-silenced transgenic tobacco lines were produced by the RNAi method. Four lines were finally selected, among which R11 and

R131 plants, which showed almost complete silencing of *NtWIN4,* were used for biological assays (Figure 9A). Since *NtWIN4* transcripts were induced upon pathogen infection (Figure 7D), effects of silencing on necrotic lesion formation were first examined. When leaves were inoculated with *Pseudomonas syringae* pv. *glycinea* 801, which causes HR-like cell death in tobacco plants (Sugimoto et al. 2004), necrotic cell death occurred within 48 h in the controls, but not in R11 plants, leaving alive tissues within the inoculated areas (Figure 9B). Similarly, when leaves from wild type control plants were inoculated with TMV, lesions began to develop 24 h after temperature shift, and reached the final mature necrotic stage 36 h later (Figure 9C). In the R11 transgenic line, lesions were visible at 24 h, but did not develop to the final stage even after 48 h, with dead tissues only around the lesion forming ring-shape spots (Figure 9C) (Waller et al. 2006). Direct measurement of cell death by chloroplast auto-fluorescence indicated that, in R131 and R71 transgenic plants, HR was much delayed in comparison with wild type controls (Figure 9D). Taking these results overall, it is conceivable that one NtWIN4 function might be to enhance hypersensitive cell death possibly through disruption of chloroplast activity.

Discussion

This paper documents evidence that a bHLH protein, originally possessing transcriptional repression activity in the nucleus, is converted to a regulatory factor for chlorosis after deletion of a basic region at the N-terminus. The findings reveal two novel features of protein function: first, that a protein can be flexibly utilized for diverse objectives after intensive modification, and second, that a bHLH factor may regulate hypersensitive cell death probably through chloroplast disruption.

Conversion of bHLH Protein

Since the bHLH motif was first identified in transcriptional factors named E12 and E47 in murine cells (Murre et al. 1989), a set of bHLH proteins has been identified in both plants and animals (Garrell and Campuzano 1991; Quail 2000). The number of corresponding genes has continuously increased, for example, to 35 in nematodes, 56 in the fly, and 147 in Arabidopsis (Ledent and Vervoort 2001, Toledo-Ortiz et al. 2003). Biological functions of bHLH proteins are currently considered to be due to their activities as transcriptional factor by forming homo and/or hetero dimers, but not all have been biochemically characterized (Murre et al. 1989, Ferre-D'Amare et al. 1994). The novel bHLH protein, NtWIN4, identified in the present study shows particularly high similarity to Arabidopsis At5g43650 protein. When *NtWIN4* and *At5g43650*, each encoding a 28 kDa protein, were introduced in tobacco and Arabidopsis cells, respectively, clear transcriptional repression activity was observed in nucleus with both. Thus NtWIN4 can function as a transcriptional factor as do most bHLH proteins. Nevertheless, further analysis revealed that NtWIN4 was exclusively localized in chloroplasts. This raised two questions; first, how might it be translocated? Second, what is its function?

Proteins that are translocated to plastids usually possess a transit peptide of 40 to 50 amino acids at the N-terminus (Buchanan et al. 2000) which is cleaved off during translocation, resulting in plastid-resident mature proteins with smaller size than the precursors. Translocation of NtWIN4 into chloroplasts appears to be a typical such case, with deletion of the first 93 amino acids, if the protein is synthesized from the first AUG on the full length mRNA sequence. However, for translocation into chloroplasts, fine dissection analyses revealed that the protein should begin from an amino acid at

positions between 13 and 39, suggesting the first 12 amino acids interfere or mask the signal peptide activity. Two cases are conceivable to account for this phenomenon. First, at least the first 12 amino acids are processed after translation. In this case, amino acids from the N-terminal might be removed immediately after synthesis, resulting in a truncated polypeptide of lower than 28 kDa up to 24 kDa, if the first 38 amino acids are processed. Although this possibility can not be excluded, it appears of low probability, as no such protein species were detected in any cellular fractions under my experimental conditions. Second, the protein is synthesized not from the first Met, but from the second Met, which corresponds to the amino acid at position 27. The latter was found to be the case and is described in chapter III. Overall, I concluded that NtWIN4 homologs commonly possess potential chloroplast-transit peptides, and that whether or not they are active *in vivo* depends on the presence of preceding peptides, which usually prevent them to function.

However, a unique feature is that this transit peptide extends into the basic region, which is considered to be the DNA binding domain. DNA binding proteins in chloroplasts mostly originated from prokaryotes (Sato 2001), and, to date, only few have been reported with a eukaryote origin (Figure 1). Amino acid sequence analyses have suggested that these proteins might be originally nuclear transcriptional factors (Kodama 2007), providing grounds for speculation that nuclear transcriptional factors might have been reutilized in chloroplasts. In this context, NtWIN4 is of clear interest. Its specific feature is that, in contrast to the proteins which retain DNA binding capacity, this has been presumably lost with NtWIN4, so that it functions solely as a novel regulatory factor in chloroplasts.

Whatever the mechanism, the present findings indicate that not all bHLH proteins

are nuclear transcriptional factors, and that some could be converted to directly perform regulatory roles. Whether this feature is common among plants and/or organisms is currently not known, but it is attractive to speculate that a protein may evolve depending upon necessity.

Biological Function

Transgenic tobacco plants constitutively expressing *NtWIN4*$^{27\text{-}247}$ under the control by CaMV 35*S* promoter showed abnormal phenotypes with slow growth and albinism, and ultimately died before maturation. Even when *NtWIN4*$^{27\text{-}247}$ was spatially and temporarily expressed after agroinfiltration, leaf color turned yellow within several days, a typical symptom of chlorosis (Buchanan et al. 2000). In contrast, transgenic RNAi plants appeared normal under non-stressed conditions and showed delayed cell death when inoculated with pathogens, TMV and *P. syringae*, suggesting that NtWIN4 is directly involved in the process of cell death by enhancing disruption of chloroplasts. One indicator of chlorosis is a reduction of chlorophyll contents, which is caused by degradation due to external factors (Hendry et al. 1987). The first enzyme in the relevant pathway has been designated as chlorophyllase (CLH) (chlorophyll-chlorophyllido hydrolase, EC 3.1.1.14), catalyzing ester bond hydrolysis of chlorophyll to yield chlorophyllide and phytol (Matile et al. 1999; Takamiya et al. 2000). Encoding genes have been isolated from multiple plant species (Jacob-Wilk et al. 1999; Tsuchiya et al. 1999; Arkus et al. 2005; Tang et al. 2004), among which *AtCLH1* from Arabidopsis is best characterized (Tsuchiya et al. 1999). *AtCLH1* is up-regulated by wounding, methyl jasmonate and coronatine (a chlorosis-inducing phytotoxin produced by various plant pathogenic bacteria), and the encoded protein functions not only in

defense but also during development and senescence (Benedetti et al. 1998; Kariola et al. 2005; Hörtensteiner 2006; Takamiya et al. 2000; Tsuchiya et al. 1999). Recent surveys have identified two additional enzymes, pheophorbide *a* oxygenase and red chlorophyll catabolite reductase, both involved in light-dependent cell death, which catalyze the breakdown of phototoxic pigments produced during chloroplast disruption (Hörtensteiner 2006). Despite detailed biochemical studies, molecular mechanisms of chlorosis and regulation of chlorophyll breakdown are still poorly understood (Hörtensteiner 2006). The present study showed that *NtWIN4* is up-regulated by wounding, methyl jasmonate and pathogen, the pattern being similar to that of *AtCLH1*, and that the resultant over-expression is lethal due to chloroplast decay. Considering these results together with other available information, I speculate that NtWIN4 functions as a novel chlorosis-inducing factor in response to wounding and pathogen attack, thereby enhancing cell death, which is prerequisite to cope with such stresses.

NtWIN4 does not possess any enzymatic domains, suggesting it to be functional in combination with other protein factors. Since it loses the basic region or DNA binding domain during chloroplast translocation, it may be active as an HLH protein, like protein Id (inhibitor of DNA binding) and extramacrochaete (Benezra et al. 1990; Ellis et al. 1990; Garrell and Modolell 1990), which may act as antagonists of other bHLH proteins by forming nonfunctional heterodimer complexes (Van Doren et al. 1992). However, NtWIN4 and At5g43650 did not bind to a reported plastid-resident bHLH protein, PTF1 (unpublished observations). Other chloroplast-resident bHLH proteins have yet to be identified in plants, making it improbable that NtWIN4 forms heterodimers. Nevertheless structural analyses suggested that NtWIN4 possesses a high potential for dimerization and/or trimerization

(http://multicoil.lcs.mit.edu/cgi-bin/multicoil) (Wolf et al. 1997), affording a capacity to interact with protein(s) other than bHLH in chloroplasts. Identification of such proteins may provide clear-cut information as to its biological roles.

Our present work indicates that a bHLH protein has changed its cellular localization from the nucleus to the chloroplast through conversion of the DNA binding domain to the transit peptide for chloroplast localization. How did this occur during evolution? Considering the demonstrated nuclear activity, NtWIN4 might have originally been a nuclear transcription factor, but might have become susceptible to processing by a mutation of for example transcription and/or translation machinery, resulting in acquisition of translocation ability. In chloroplasts, NtWIN4 perhaps then activated chlorosis by chance, which fitted with survival of tobacco plants. Thus, my finding provides a good example of protein evolution, realized by maximal use of the limited genome information, thereby conferring on plants the best features adapting them to diverse environmental conditions.

Table 2. NtWIN4 constructs used in chapter I and summary of cellular localization with GFP

	Amino acids		
Designation	Position	No. of residues	Cellular localization
NtWIN4	1 - 247	247	Cytoplasm and Nucleus
NtWIN4$^{27\text{-}78}$	27 - 78	52	Cytoplasm and Nucleus
NtWIN4$^{27\text{-}87}$	27 - 87	61	Cytoplasm and Nucleus
NtWIN4$^{27\text{-}91}$	27 - 91	65	Cytoplasm and Nucleus
NtWIN4$^{27\text{-}92}$	27 - 92	66	Cytoplasm and Nucleus
NtWIN4$^{27\text{-}93}$	27 - 93	67	Chloroplast
NtWIN4$^{27\text{-}94}$	27 - 94	68	Chloroplast
NtWIN4$^{27\text{-}95}$	27 - 95	69	Chloroplast
NtWIN4$^{27\text{-}111}$	27 - 111	85	Chloroplast
NtWIN4$^{27\text{-}135}$	27 - 135	109	Chloroplast
NtWIN4$^{27\text{-}247}$	27 – 247	221	Chloroplast
NtWIN4$^{7\text{-}111}$	7– 111	105	Cytoplasm and Nucleus
NtWIN4$^{8\text{-}111}$	8– 111	104	Cytoplasm and Nucleus
NtWIN4$^{9\text{-}111}$	9 – 111	103	Cytoplasm and Nucleus
NtWIN4$^{10\text{-}111}$	10– 111	102	Cytoplasm and Nucleus
NtWIN4$^{11\text{-}111}$	11 – 111	101	Cytoplasm and Nucleus
NtWIN4$^{12\text{-}111}$	12 – 111	100	Cytoplasm and Nucleus
NtWIN4$^{13\text{-}111}$	13 – 111	99	Chloroplast
NtWIN4$^{14\text{-}111}$	14– 111	98	Chloroplast
NtWIN4$^{21\text{-}111}$	21– 111	91	Chloroplast
NtWIN4$^{31\text{-}111}$	31 – 111	81	Chloroplast
NtWIN4$^{36\text{-}111}$	36 – 111	76	Chloroplast
NtWIN4$^{37\text{-}111}$	37 – 111	75	Chloroplast
NtWIN4$^{38\text{-}111}$	38– 111	74	Chloroplast
NtWIN4$^{39\text{-}111}$	39 – 111	73	Chloroplast
NtWIN4$^{40\text{-}111}$	40 – 111	72	Cytoplasm and Nucleus

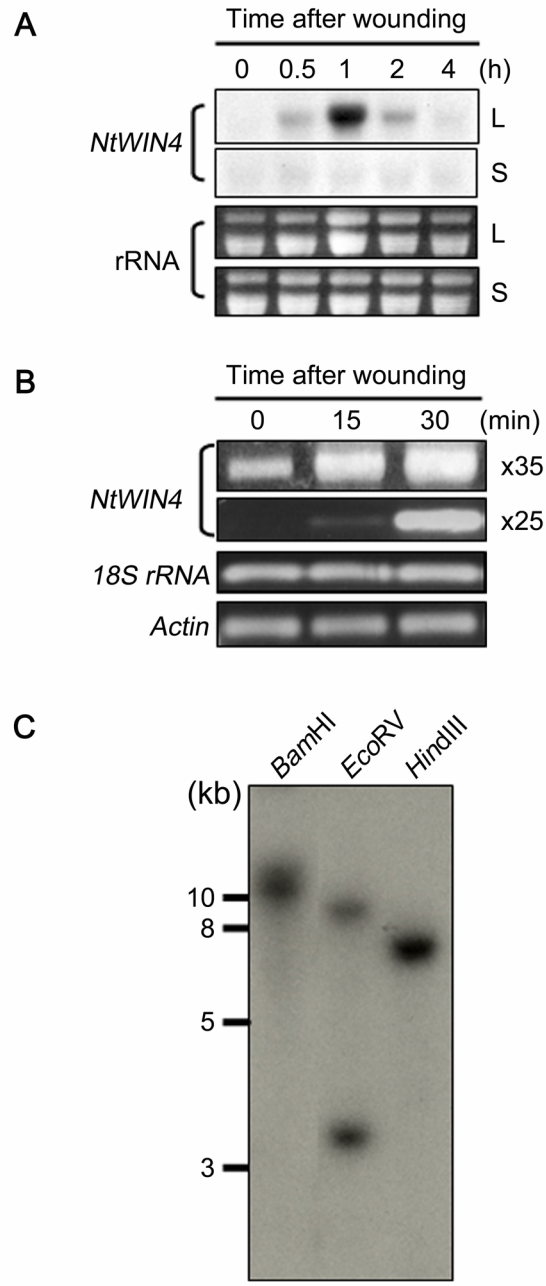

Figure 2. Identification of *NtWIN4*. (**A**), (**B**) Time course analysis of *NtWIN4* transcript accumulation. Total RNA was isolated from wounded (local) and un-wounded (systemic) leaves at the indicated time points after wounding. A 10-mg aliquot per lane was fractionated on a 0.8% agarose gel, and subjected to RNA blot hybridization with the 32P-labeled *NtWIN4* cDNA probe. As the loading control, rRNA was applied (A). Transcript accumulation was also examined by RT-PCR. *NtWIN4* was amplified by 35- or 25-cycle PCR. As internal controls, *18S rRNA* and *Actin* were amplified for 25 cycles (B). (**C**) DNA blot hybridization. A 10-mg aliquot of total DNA was digested with the indicated restriction enzyme and subjected to hybridization with the 32P-labeled *NtWIN4* cDNA probe. The size of DNA is indicated on the left in kb.

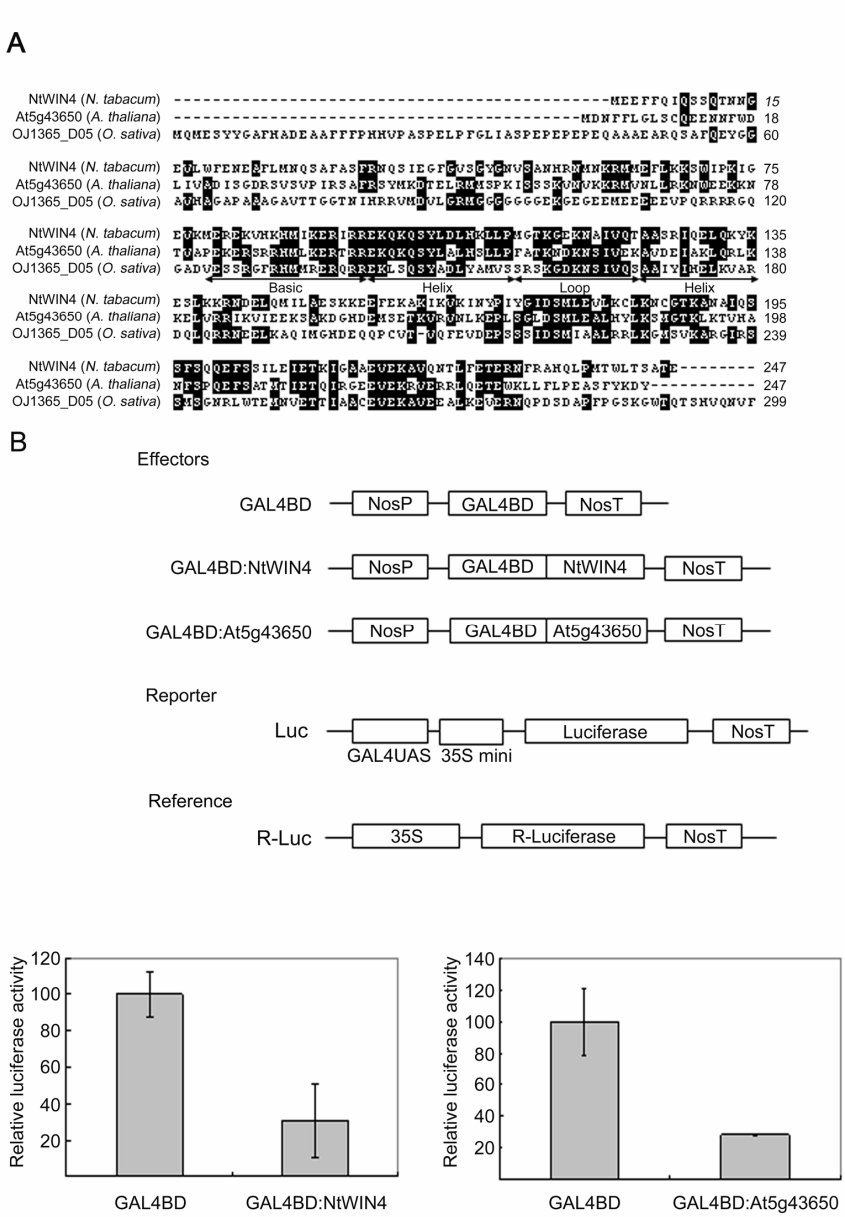

Figure 3. Properties of NtWIN4. (**A**) Comparison of amino acid sequences. Alignment of NtWIN4, At5g43650 (*Arabidopsis thaliana*) and OJ1365_D05.18 (*Oryza sativa*) was performed with the ClustalW program. Underlined arrows indicate the bHLH motif and the black shading is for identical residues shared by at least two samples. (**B**) Schemtic illustration of plasmids used in the bHLH-mediated transcription assay. NosP, nopaline synthase promoter; NosT, nopaline synthase terminator; 35*S*, cauliflower mosaic virus (CaMV) 35*S* promoter; Gal4UAS, Gal4 binding sequence; 35*S*mini, CaMV35*S* minimal promoter. (**C**) Transcription activity assay results. Effector constructs encoding Gal4BD:NtWIN4 or Gal4BD:At5g43650 were co-bombarded with the reporter and reference plasmids into 7-day old BY2 or MM2d cells, respectively. Gal4BD was applied as the effector control. Luciferase activity of each type of transformants was normalized to the respective R-luciferase activity. Relative luciferase activity was calculated by dividing the luciferase activity of each clone by that of the clone containing the Gal4BD effector construct (triplicate experiments).

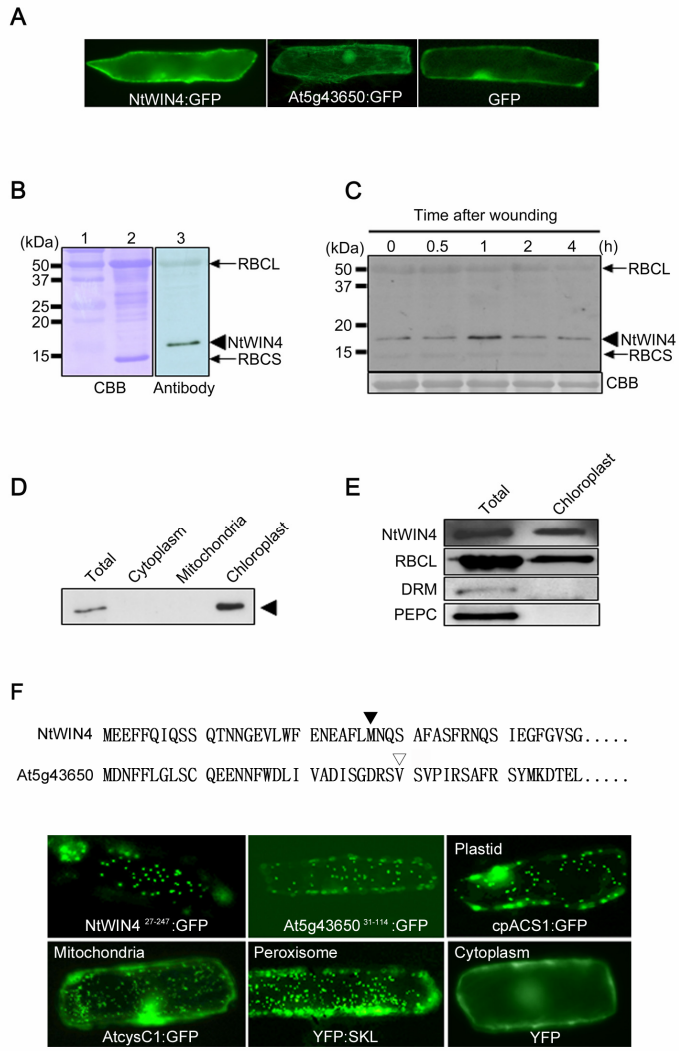

Figure 4. Cellular localization. (**A**) Localization of full length NtWIN4. NtWIN41-247 and At5g436501-247 fusion proteins with GFP were transiently expressed in onion epidermal cells (left and middle panels, respectively). The control was GFP alone (right panel). (**B**) Identification of NtWIN4 protein. Crude extracts of wounded tobacco leaves (lanes 2 and 3) were fractionated on SDS-PAGE and stained with coomassie brilliant blue (CBB) (left panel), and subjected to immuno-blot hybridization using an antibody raised against an NtWIN4 peptide (right panel). The protein size marker was fractionated in parallel (lane 1). Positions of the large subunit of ribulose-1,5-bisphosphate carboxylase/oxgenase (RBCL) and small subunit of ribulose-1,5-bisphosphate carboxylase/oxgenase (RBCS) are indicated by arrows. The arrow head indicates NtWIN4. (**C**) Time course analysis of NtWIN4 protein accumulation. Healthy leaves were wounded and harvested at the indicated time points to extract crude proteins, which were fractionated with SDS-PAGE and subjected to immuno-blot hybridization. Positions of the RBCL and RBCS are indicated by arrows, and NtWIN4 is indicated by an arrow head. Equal loading of proteins was monitored with RBCL stained using coomassie brilliant blue. (**D**) Fractionation of cellular components. Crude extracts were subjected to gradient centrifugation, and isolated cytoplasm, mitochondria and chloroplast fractions were fractionated with SDS-PAGE and subjected to immuno-blot hybridization with NtWIN4 antibodies. The arrow head indicates the position of NtWIN4. (**E**) Purity verification. Total cellular and chloroplast fractions were analyzed by immuno-blot hybridization using antibodies against compartment-specific proteins, RBCL (chloroplast), domains-rearranged DNA methyltransferase (DRM) (nucleus) and phosphoenol-pyruvate carboxylase (PEPC) (cytoplasm). (**F**) Plastid localization of NtWIN427-247 and At5g4365031-114. Forty seven amino acids in the N-terminus of NtWIN4 and At5g43650 are shown with indication of the second Met at position 27 in NtWIN4 (closed arrowhead), and Val at position 30 in At5g43650 (open arrowhead) (upper panel). Onion epidermal cell layers were bombarded with NtWIN427-247 or At5g4365031-114 fused GFP, and fluorescence was observed. The controls were AtCysC1:GFP (mitochondrial) (63), cpACS1:GFP (plastid) (64), YFP:SKL (peroxisome) (65) and YFP alone (lower panel).

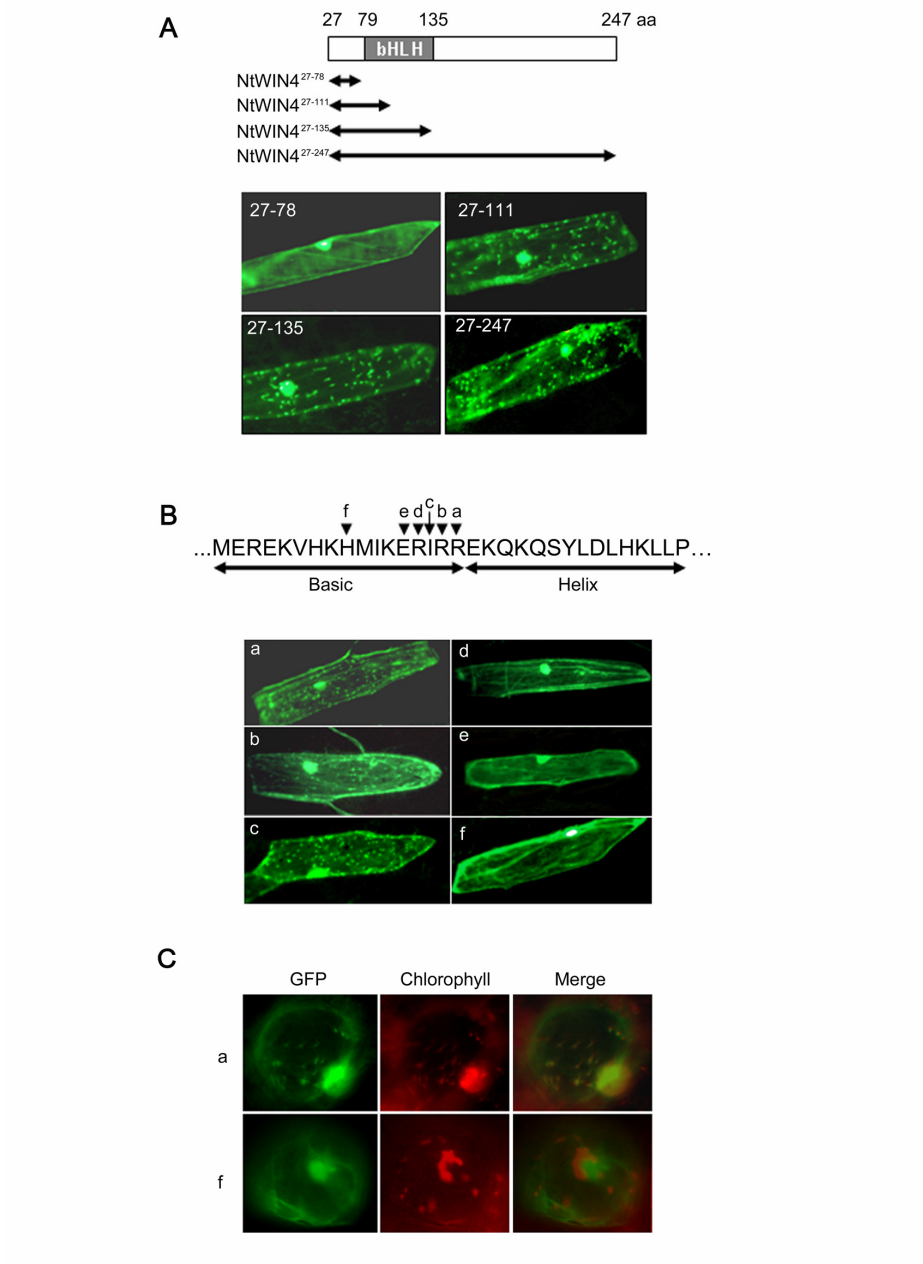

Figure 5. Identification of a transit peptide. (**A**) Schematic illustration of the NtWIN427-247 structure. Arrows indicate the size of dissected NtWIN4, to whose C terminus GFP was fused. Onion epidermal cell layers were bombarded with plasmids containing the indicated constructs and fluorescence was observed. (**B**) Illustration of the dissection approach. The amino acid sequence shows the basic and first helix regions of NtWIN4 (upper panel). Dissected positions are indicated by arrow-heads with small letters. The position 93 is indicated by the arrow with c. Onion epidermal cell layers were bombarded with plasmids containing the indicated constructs and fluorescence was observed (lower panel). Samples are NtWIN427-95:GFP (a), NtWIN427-94:GFP (b), NtWIN427-93:GFP (c), NtWIN427-92:GFP (d), NtWIN427-91:GFP (e) and NtWIN427-87:GFP (f). (**C**) Tobacco protoplast-derived cultured cells expressing NtWIN427-95:GFP (a) or NtWIN427-87:GFP (f). GFP fluorescence (right panel) and chlorophyll auto-fluorescence (middle panel) are merged (right panel).

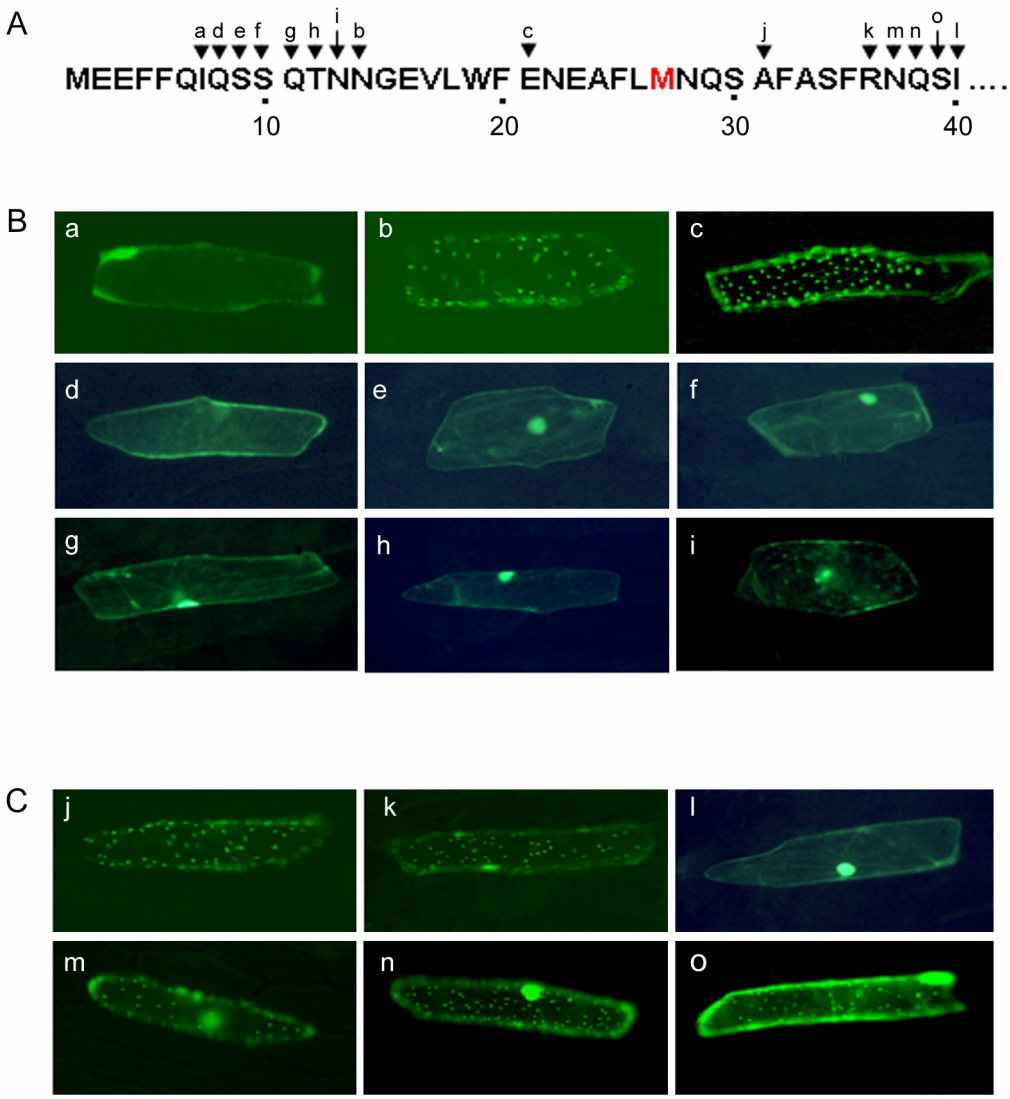

Figure 6. Determination of the transit peptide initiation site. (**A**) Amino acid sequence of the first 47 amino acids in the N-terminus of NtWIN4. Dissection positions are indicated by lower case letters. (**B**), (**C**) GFP epifluorescence analysis. Onion epidermal cell layers were bombarded with plasmids containing GFP fused to the indicated fragment in *A*, and fluorescence was observed. Dissected samples upstream of the 27th Met are NtWIN47-111:GFP (a), NtWIN414-111:GFP (b), NtWIN421-111:GFP (c), NtWIN48-111:GFP (d), NtWIN49-111:GFP (e), NtWIN410-111GFP (f), NtWIN411-111GFP (g), NtWIN412-111GFP (h), NtWIN413-111GFP (i) (B). Dissected samples down-stream of the 27th Met are NtWIN431-111GFP (j), NtWIN436-111GFP (k), NtWIN440-111GFP (l), NtWIN437-111GFP (m), NtWIN438-111GFP (n), NtWIN439-111GFP (o) (C).

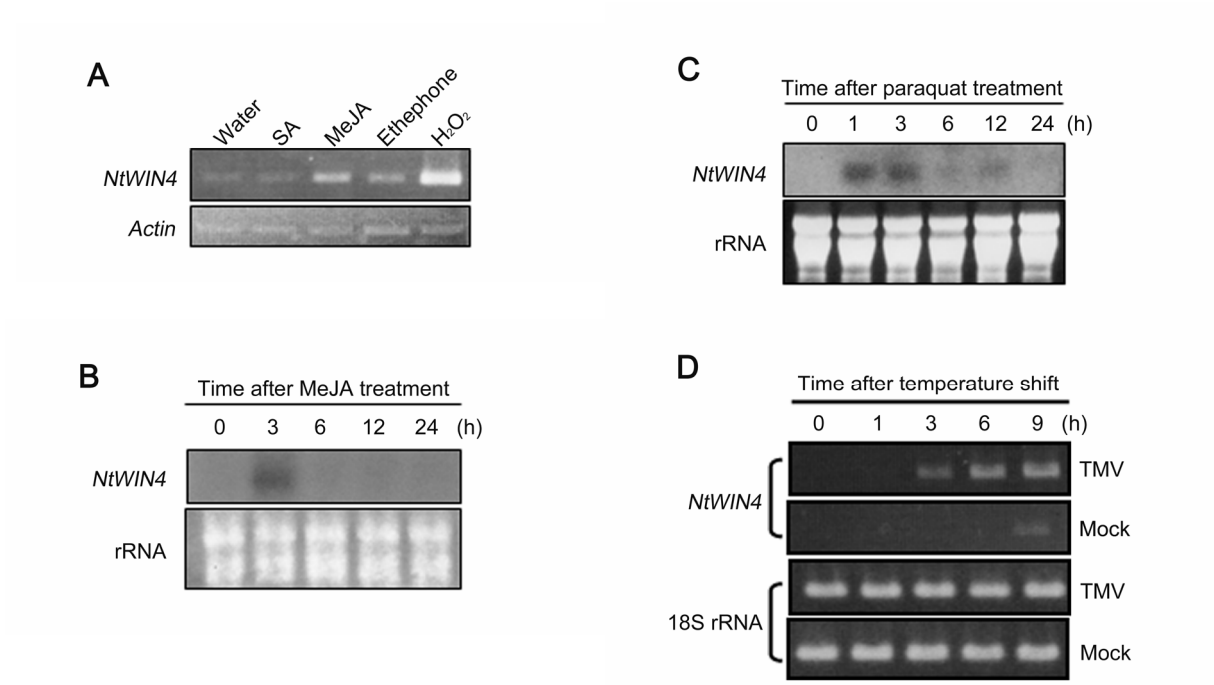

Figure 7. Expression profile. (**A**) Effects of phytohormones. Healthy leaves of wild-type tobacco were detached and treated for 12 h with water, 500 mM salicylic acid (SA), 50 mM methyl jasmonic acid (MeJA), 100 mM ethephone or 10 mM hydrogen peroxide (H2O2). *NtWIN4* transcripts were subjected to RT-PCR with a 30 cycle amplification program. As an internal control, *Actin* was used with an 18-cycle PCR. (**B**), (**C**), (**D**), Time course analyses of transcript accumulation. Leaves were treated with 50 mM MeJA (B) or 10 mM paraquat (C) and sampled at the indicated time points. Total RNA was isolated and subjected to northern blot hybridization with the 32P-labeled *NtWIN4* cDNA probe. As a loading control, rRNA was visualized by ethidium bromide staining. (**D**) Healthy leaves were inoculated with TMV (TMV) or buffer alone (Mock), maintained at 30oC for 2 days, then cultured at 20oC for the indicated time period. *NtWIN4* transcript accumulation was estimated with RT-PCR using 25-cycle amplification. As the control, *18S rRNA* DNA was amplified by 20-cycle PCR.

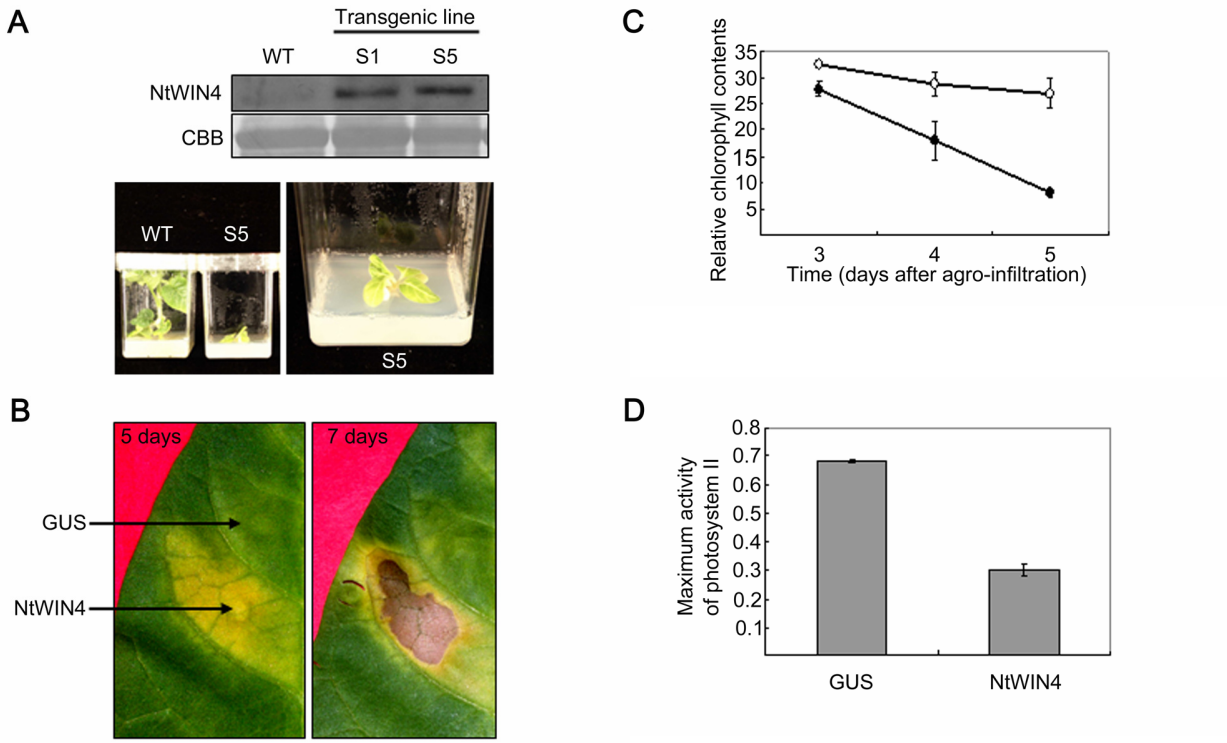

Figure 8. Properties of *NtWIN4* over-expressing plants. (**A**) Phenotypes of sense transgenic plants. Expression of introduced NtWIN4 in transgenic tissues (S1, S5) was confirmed by immuno-blot analysis using NtWIN4 antibodies. Wild type (WT) was used as the control. As a loading control, RBCL stained with CBB is shown (upper panel). Phenotypes of transgenic line S5 (S5) and wild-type (WT) plantlets (left panel) are shown together with an enlarged view of S5 (right panel). (**B**) Lesion formation in a region transiently expressing NtWIN424-247. A leaf sample was infiltrated with *A. tumefaciens* harboring a plasmid containing NtWIN424-247, and photographed 5 (left) and 7 (right) days later. As a control, *GUS* gene was used. (**C**) Estimation of chlorophyll contents. Relative chlorophyll contents at regions infiltrated with *NtWIN4* (closed circle) or *GUS* (open circle) in the above described samples were measured at the indicated time points by SPAD-502. (**D**) Chlorophyll fluorescence analysis. The maximum activity of photosystem II photochemistry in regions expressing *GUS* or *NtWIN4* was measured using MINI-PAM as described in the text.

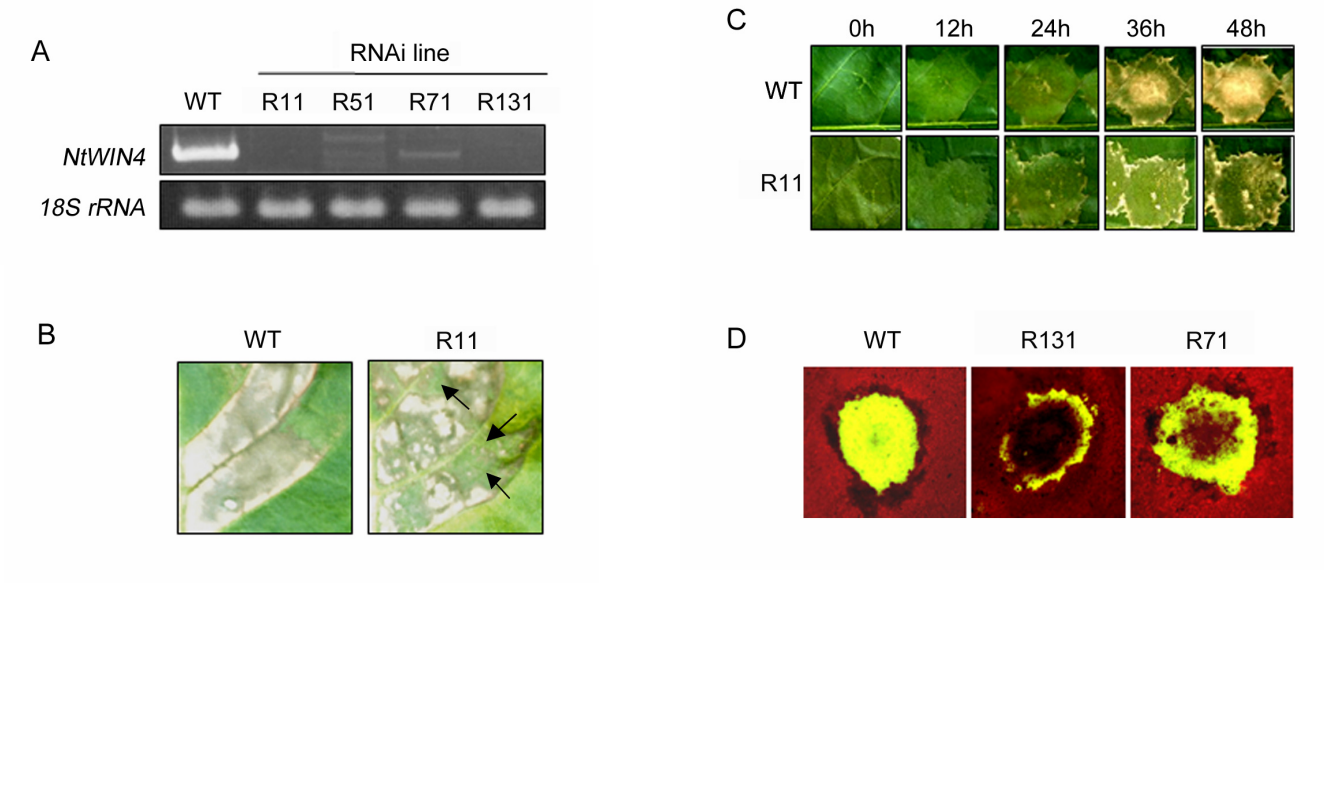

Figure 9. Properties of *NtWIN4* silenced plants. (**A**) Confirmation of silencing. Healthy leaves from wild type and RNAi transgenic plants were examined for NtWIN4 expression by RT-PCR with a 35-cycle amplification, which allows detection of constitutive, low level expression, as shown in FIG. 1*B*. As an internal control, expression of 18*S* rDNA was assessed. (**B**) Effects of *Pseudomonas syringae* pv. *glycinea* 801 infection. Healthy leaves from wild type (WT) and transgenic R11 (R11) plants were inoculated with *P. syringae* by infiltration and observed 48 h thereafter. In the wild type sample, hypersensitive response-like cell death occurred throughout the inoculated region, while in R11 tissues remained alive showing green color (indicated by arrows). (**C**) Effects of TMV infection. Healthy leaves from wild type (WT) and transgenic R11 (R11) plants were inoculated with TMV, maintained at 30oC for 2 days, then cultured at 20oC for the indicated time periods. Hypersensitive cell death is clearly apparent in the wild type, with necrotic lesions after 36 h, while it is not complete in the R11 case, as evidenced by live tissue inside of ring-shaped lesions. (**D**) Reduction of chloroplast auto-fluorescence. Wild type and transgenic R131 and R71 plants were inoculated with TMV as described above, and lesions were viewed 48 h after temperature shift under a microscope (Olympus PROVIS AX70 and Leica MZ FLIII) equipped with a fluorescence module. Since live chloroplasts emit auto-fluorescence (red color), they can be clearly distinguished from dead ones in necrotic regions (yellow color).

CHAPTER II

A comparative analysis of two bHLH proteins with reference to plastid localization

Introduction

In chapter I, I identified a novel bHLH protein that is translocated into plastids and participates in hypersensitive cell death in tobacco plants. This protein, designated as NtWIN4, was proposed to be translated from the second Met, which facilitates the N-terminus as a plastid transit signal, resulting in formation of a 17 kDa plastid-localized protein (Kodama and Sano 2006). In this context, NtWIN4 resembles another plastid-resident bHLH protein, PTF1 (Figure 1, Table 1), from Arabidopsis, although their amino acid sequences are not similar. In this chapter, I performed comparative studies of plastid localization mechanisms, and found that PTF1 was localized to the nucleus, and functioned as a transcription repressor.

Materials and Methods

Epifluorescence analyses

The full-length *PTF1* and plastid-localized type *NtWIN4*$^{27\text{-}247}$ (Table 2) were subcloned into the SalI/NcoI sites of a CaMV35S-sGFP(S65T)-nos vector, harboring a synthetic gene for improved GFP, sGFP(S65T) driven by the CMV35*S* promoter and an *NOS* terminator (Chiu et al. 1996). Onion epidermal cell layers were bombarded with gold particles (Bio-Rad Laboratories, Hercules, CA) coated with appropriate vector constructs, and after incubation at 28^{o}C for 6 - 12 h in the dark, samples were viewed under a microscope (Olympus PROVIS AX70) equipped with a fluorescence module. For DNA staining, samples were incubated with 1 mg/ml 4',6-diamino-2-phenylindole

(DAPI) solution before observation.

Nuclear transcription assays

The GAL4BD-PTF1 effector plasmid was constructed by fusing a cDNA encoding PTF1 with the GAL4 DNA binding domain in a yy64 vector, a derivative of pMA560 (Yamamoto and Deng 1998). *PTF1* was subcloned into the BglII/SalI sites and yy64 vector alone was used as a control. The reporter plasmid, yy96, contained a luciferase gene placed under control of the Gal4 binding site (Yamamoto and Deng 1998). An internal control plasmid, containing an R-luciferase gene placed under control of the 35*S*-CaMV promoter was used to normalize for differences in bombardment efficiency. Seven-day old MM2d cells were plated on MS agar and bombarded with plasmids (effecter:reporter:reference=2:2:1) coated on a 1.0 μm microcarrier, under a vacuum of 28 inches of mercury using a helium pressure of 1,100 psi (PDS 1000, Bio-Rad). Cells were placed 6 cm from the stopping screen. After bombardment, they were incubated in the dark at 28^{o}C for 22 h. Luciferase and R-luciferase activities were assayed using a dual-luciferase assay kit (Promega, Madison, WI) according to the manufacturer's instructions and chemical luminescence was measured using a luminometer (Lumat LB9507, Berthold Technologies, Bad Wildbad, Germany).

Results and Discussion

Plastid localization was visually examined for PTF1 and NtWIN4 by using GFP-fused proteins. Vectors with *PTF1-GFP* or *NtWIN4-GFP* were introduced into onion epidermal cell layers and cells were observed under light interference contrast and epifluorescence. Nuclei were identified by 4',6-diamidino-2-phenylindole,

dihydrochloride (DAPI) staining. In contrast to the control GFP localization in both the nucleus and cytoplasm (Figure 10, left panel), PTF1-GFP was solely localized only in nucleus and not in any organelles, including plastids (Figure 10, second panel from the left). This was confirmed by merging GFP with DAPI images, showing a complete match. NtWIN4-GFP, in contrast, was clearly localized in plastids (Figure 10, second panel from right), giving a similar fluorescence pattern to the plastid-resident control protein cysteine synthase, cpACS1-GFP (Figure 10, right panel). In chapter I, *in silico* analysis with the localization prediction program, TargetP 1.1 (http://www.cbs.dtu.dk/services/TargetP/) (Emanuelsson et al. 2000) indicated that NtWIN4 beginning from the second Met at position 26 contained a putative plastid transit peptide, although the program filed to predict the transit signal in full length NtWIN4 beginning from the first Met (Table 3). When PTF1 was examined by the same program, no plastid transit peptides were found in either full length or truncated peptide beginning from the second Met at position 58 (Table 3). Signals to other compartments, such as mitochondria and secretory pathways, were not found either (Table 3). Overall, biochemical and computer analyses indicated the low probability of PTF1 to localize in plastid.

Functions of PTF1 were further analyzed using nuclear transcription assays *in vivo* with the dual-luciferase assay. The constructs for this assay are shown as schematic illustrations in Figure 11A. When effector constructs of *GAL4BD* control and *GAL4BD-PTF1* were co-bombarded with a reporter plasmid and a reference plasmid into Arabidopsis MM2d cells, the *GAL4BD-PTF1* effector construct caused down-regulation of the luciferase activity to a level only 1/10th that of *GAL4BD* (Figure 11B). Thus PTF1 protein possesses transcriptional repression activity in nucleus,

suggesting function as a nuclear transcriptional repressor *in vivo*.

The full length polypeptide of NtWIN4 was predicted to be 28 kDa, but immuno-blot analysis of intact proteins in plastids revealed a value of 17 kDa (Figure 4). Since the DNA binding domain in the bHLH motif was deleted in this form, it is likely that loss of the capacity for binding to DNA occurs. In contrast, PTF1 was initially identified as a DNA binding protein by the yeast one-hybrid screening method, and shown to recognize the ACC repeat region of LRP sequence to regulate LRP transcription in plastids (Baba et al. 2001). An PTF1-deficient mutant, *ptf1,* showed reduced activity of *psbD* LRP under continuous light conditions, and early bleaching, late flowering and dwarfism under short-day conditions. Based on these findings, PTF1 is proposed to regulate *psbD* LRP through its transcription in plastids.

Due to its unique features, PTF1 has often been referred to as an example of non-nuclear localized bHLH protein (Nagashima et al. 2004; Tsunoyama et al. 2004; Hanaoka et al. 2003; Sekine et al. 2002; Thum et al. 2001; Kanamaru et al. 2001). However, findings have been rather controversial as to its function and localization. For example, *psbD* LRP was shown to be transcribed by a nuclear-encoded transcription factor, AtSig5 (Tsunoyama et al. 2004). This was confirmed with a mutant, *sig5,* in which *psbD* LRP activity was completely eliminated, whereas in *ptf1* mutants the activity was not affected (Baba et al. 2001). PTF1 is unlikely to be involved in light signaling based on the finding that light-dependent *psbD* transcription was not diminished in PTF1-deficient mutants (Tsunoyama et al. 2004). Considering these observations together with my present demonstration of typical transcription repression in the nucleus, I propose careful reevaluation of the cellular localization and biological function of PTF1 protein.

Table 3. Prediction of plastid localization within NtWIN4 and PTF1

Name	Position	Length	cTP	mTP	SP	Other	Localization
NtWIN4	1-247	247	0.183	0.107	0.121	0.671	Any other
	27-135	109	0.376	0.200	0.112	0.313	Chloroplast
PTF1	1-355	355	0.061	0.104	0.019	0.906	Any other
	1-122	122	0.071	0.092	0.021	0.901	Any other
	58-355	298	0.135	0.439	0.013	0.686	Any other
	58-122	65	0.207	0.524	0.032	0.541	Any other

NtWIN4 and PTF1 were examined by the TargetP 1.1 program (http://www.cbs.dtu.dk/services/TargetP/) using PLANT network without cutoffs. Examined samples were full length NtWIN4 (1-247), N-terminal-truncated NtWIN4 beginning from the second Met up to the basic domain (27-135) (Table 2), full length PTF1 (1-355), N-terminal region of PTF1 up to the basic domain (1-122), N-terminal-truncated full length PTF1 beginning from the second Met (58-355) and N-terminal-truncated PTF1 beginning from the second Met up to the basic domain (58-122). The probability scores of chloroplast (plastid) transit peptide (cTP), mitochondrial targeting peptide (mTP) and signal peptide for secretory pathway (SP) were estimated. Localization is predicted from the scores above.

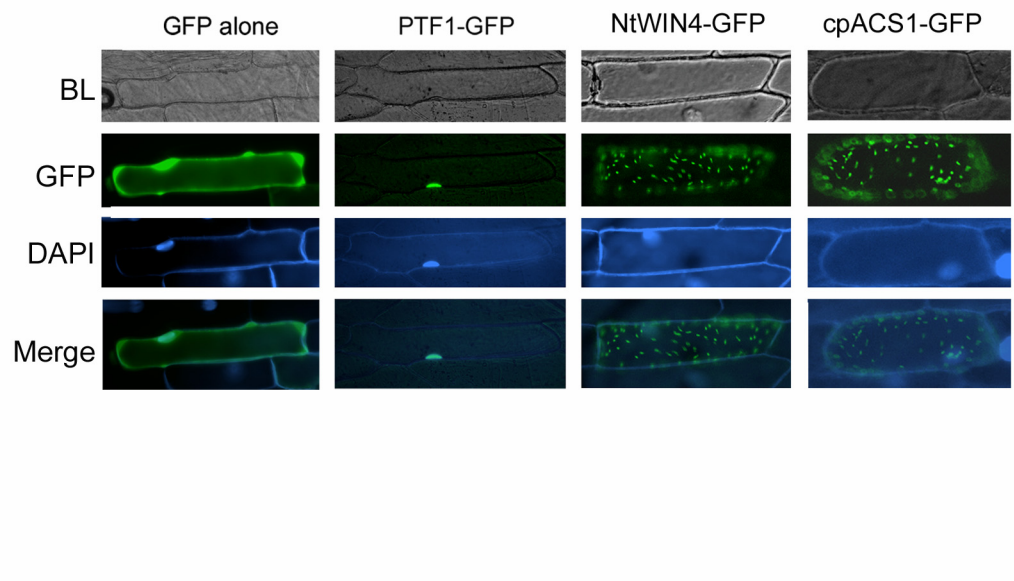

Figure 10. Comparison of sub-cellular localization. Onion epidermal cell layers were subjected to particle bombardment to introduce *GFP*, *PTF1-GFP*, *NtWIN4-GFP* or *cpACS1-GFP*. After incubation at 28oC for 6 - 12 h in the dark, samples were viewed under a microscope equipped with a fluorescence module. For DAPI staining, samples were incubated with 1mg/ml DAPI solution before observation. Cells were observed for images under bright light (BL), for epifluorescence (GFP), for chromosomes (DAPI), and DAPI and GFP images were merged (Merge). Onion epidermal cell layers expressed GFP-tagged proteins (GFP), PTF1-GFP (PTF1-GFP), NtWIN4-GFP (NtWIN4-GFP). Cells expressing cpACS1-GFP (cpACS1-GFP) was used for the plastid control.

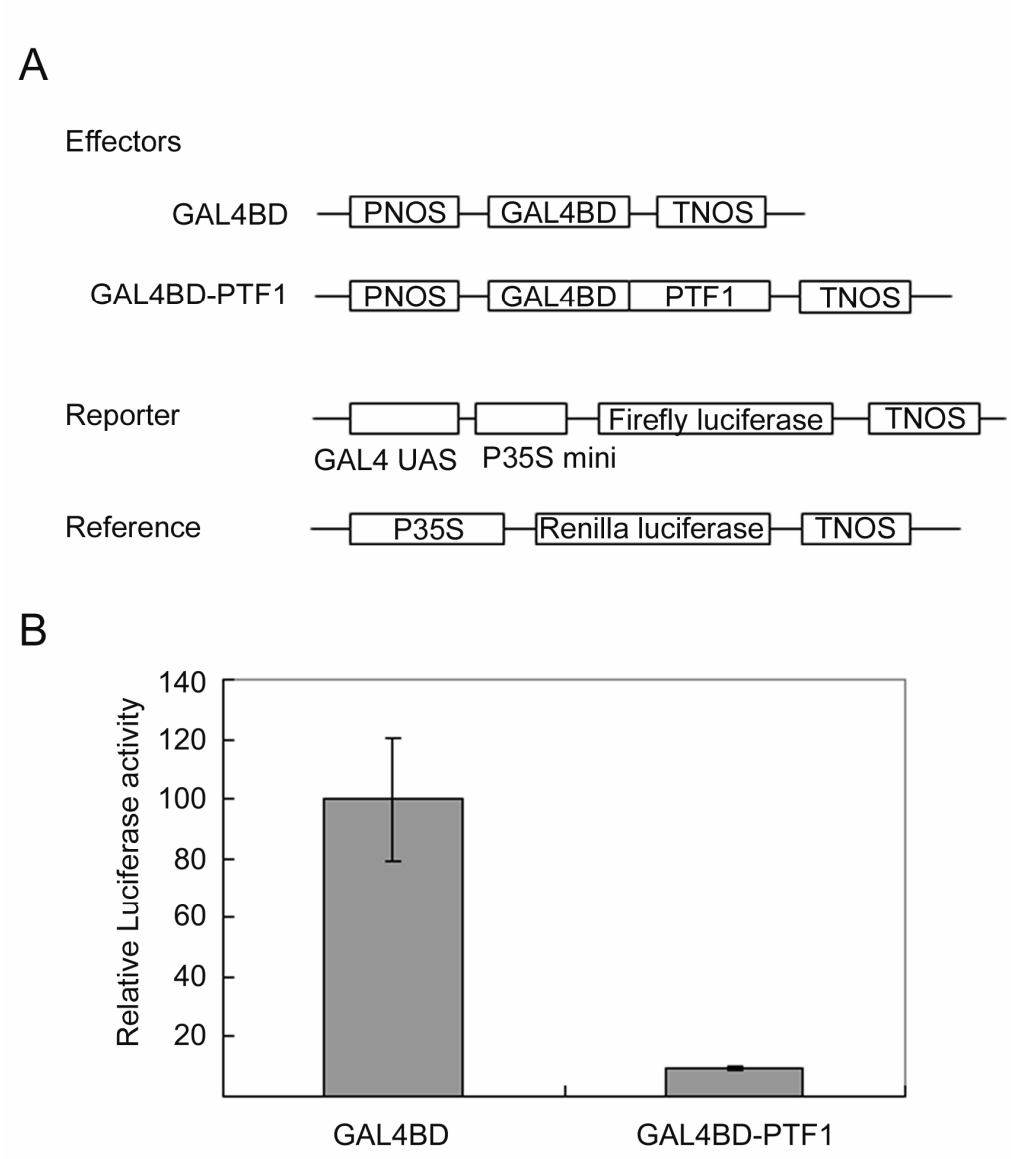

Figure 11. Nuclear transcription activity of PTF1. (**A**) Schematic illustration of plasmids used in the transcription assay. NosP, nopaline synthase promoter; NosT, nopaline synthase terminator; 35S, cauliflower mosaic virus (CaMV) 35*S* promoter; GAL4UAS, GAL4 binding sequence; 35Smini, CaMV35S minimal promoter. (**B**) Nuclear transcription activity assay. Effector constructs encoding GAL4BD-PTF1 were co-bombarded with the reporter and reference plasmids into 7-day old Arabidopsis MM2d cells. GAL4BD was used as the effector control. Luciferase activity of each transformant was normalized to the respective R-luciferase activity, calculated by dividing the luciferase activity of each clone by that of the clone containing the GAL4BD effector construct. Data are from triplicate experiments with standard deviations.

CHAPTER III

Functional diversification of a tobacco bHLH protein due to alternative transcription during generation of amphidiploidy

Introduction

Polyploidization of plants by genome doubling is an ubiquitous event in nature (Wendel 2000), constituting one of the major factors inducing evolutionary changes such as mutations, genomic rearrangements and addition of new components (Soltis 2005). Amphidiploidy is defined as a natural combining of two plant species, resulting in formation of a new species with two sets of chromosomes. *N. tabacum* has been proposed as one such example, based on morphological and cytological observation, biochemical characteristics, composition of gene families and of repeated sequences (Goodspeed 1954; Gray et al. 1974; Bland et al. 1985; Okamuro and Goldberg 1985; Matassi et al. 1991; Skalická et al. 2003, 2005) (see Figure 12A). Genomic in situ hybridization has also provided consistent evidence for the amphidiploidy, the *N. tabacum* genome comprising of two subgenomes from *N. sylvestris* and *N. tomentosiformis* (Kenton et al. 1993; Kitamura et al. 2000; Lim et al. 2000). These morphological and genomic studies thus strongly suggest that *N. tabacum* is a natural amphidiploid which arose approximately 6 million years ago. However, the influence of polyploidization on protein function is as yet poorly understood.

In preceding two chapters, I described a novel bHLH protein from *N. tabacum*, designated as NtWIN4, localized in chloroplasts. Functional and structural analyses indicated involvement of NtWIN4 in stress responses, and that it has been converted from a nuclear transcriptional repressor to a plastid-resident regulatory factor during

plant evolution. Amino acid dissection experiments showed that full length NtWIN4, synthesized from the first AUG codon of the mRNA, localizes in the nucleus and cytoplasm. In contrast, the protein synthesized from the in-frame second AUG codon lacks the first 26 amino acids from the N-terminus and localizes into plastids after removal of a transit peptide of 67 amino acids (Figures 4, 5 and 6). These observations indicate that not all bHLH proteins are nuclear transcription factors.

In this chapter, I attempted to elucidate the differential localization mechanism of NtWIN4, and found that size diversity of mRNAs and differing translation efficiencies might be responsible for the observed phenomena. I here present an argument that altered protein localization may arise at the time of polyploidization during evolution, perhaps due to natural selection forces.

Materials and Methods

Plant materials and wound treatment

Tobacco (*N. sylvestris*, *N. tomentosiformis* and *N. tabacum* cv Xanthi NC) plants were grown on soil in a growth cabinet at 23^{o}C under a 14-h-light/10-dark photo-cycle. Tobacco cultured cells, BY2, were maintained as suspension cultures in modified Linsmaier and Skoog medium (Nagata et al. 1981) at 23^{o}C continuously in the dark. Wound stress was applied by cutting mature leaves of *N. sylvestris* and *N. tabacum* with a pair of scissors, and wounded leaves were harvested at appropriate time points.

PCR and RT-PCR analyses

Genomic DNA was isolated from green leaves by the cetyl-trimethyl-ammonium bromide method (Murray and Thompson 1980) with modifications. Total RNA was

isolated by the aurintricarboxylic acid method (Gonzalez et al. 1980). For PCR and RT-PCR analyses, *NtWIN4* genomic and cDNA fragments were amplified by PCR with *ExTaq* DNA polymerase (Takara, Ohtsu, Japan) using *NtWIN4* forward (5'-GCTCTCGAGATGAATCAAAGTGCTTTTGC-3') and *NtWIN4* reverse (5'-TCACCCATGGGTTCGTTCTGTGGCTGAAG-3') primer sets. The 5'-terminus sequence was determined by the 5' RACE method using a 5'RACE core set (Takara, Ohtsu, Japan) with a primer, 5'-ATGATTCGCCGATACATTTCCA-3' (nucleotide position from the 5' end; 161-182).

Epifluorescence analyses

Full-length or partial cDNA regions of *NsWIN4* and *NtWIN4* were subcloned into the SalI/NcoI sites of a CMV35S-sGFP(S65T)-nos vector, harboring a synthetic gene for improved GFP, sGFP(S65T) driven by the CaMV35*S* promoter and *NOS* terminator (Chiu et al. 1996). Onion epidermal cell layers were bombarded with gold particles (Bio-Rad) coated with appropriate vector constructs, and after incubation at 28^{o}C for 6 - 12 h in the dark, samples were viewed under a microscope (Olympus PROVIS AX70) equipped with a fluorescence module.

Nuclear transcription assays

Gal4BD-NsWIN4 and *Gal4BD-NtWIN4* effector plasmids were constructed by fusing cDNAs encoding NsWIN4 or NtWIWN4 with the Gal4 DNA binding domain in a yy64 vector, a derivative of pMA560 (Yamamoto and Deng 1998). *NsWIN4* and *NtWIWN4* were then subcloned into the BglII and SalI sites of yy64, respectively. As an effector control, the yy64 vector alone was used. The reporter plasmid, yy96, contained a

luciferase gene placed under control of the Gal4 binding site (Yamamoto and Deng 1998). An internal control plasmid, containing an R-luciferase gene, under control of the CaMV35*S* promoter, was used to normalize for differences in bombardment efficiency. Five-day old BY2 cells were plated on 1/2x Murashige-Skoog agar and bombarded with plasmids (effecter:reporter:reference=2:2:1) coated on a 1.0 μm microcarrier, under a vacuum of 28 inches of mercury using a helium pressure of 1,100 psi (PDS 1000, Bio-Rad Laboratories, Hercules, CA). Cells were placed 6 cm from the stopping screen. After bombardment, they were incubated in the dark at 28^{o}C for 24 h. Luciferase and *Renilla* luciferase activities were assayed using a dual-luciferase assay kit (Promega, Madison, WI) according to the manufacturer's instructions. Chemical luminescence was measured using a luminometer (Lumat LB9507, Berthold Technologies, Bad Wildbad, Germany).

Protein analyses

For in vitro transcription/translation analysis, the TNT® SP6 high-yield protein expression system (Promega, Madison, WI) was used according to the manufacturer's instructions with modifications. Sequences containing the SP6 promoter (5'-AGAATTGGACTATTTAGGTGACACTATAGAAA-3') were located upstream of the cDNA sequences and template cDNA fragments amplified by PCR were directly subjected to expression reactions for 2 h. *N. tabacum* and *N. sylvestris* leaves were ground in liquid nitrogen in a mortar, and homogenized in isolation buffer containing 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 0.05% (w/v) Tween 20, 1 mM EDTA, 1 mM phenylmethylsulfonyl fluoride, 5 μg/ml leupepin and 5 μg/ml aprotinin. After centrifugation at 20,000 *g* for 10 min to remove insoluble debris, crude extracts were

directly used for immuno-blot staining. After fractionation by SDS-PAGE, proteins were electronically transferred onto cellulose membranes (Immobilon-N, Millipore, Bedford, MA), and NtWIN4/NsWIN4 was detected using rabbit antibodies raised against a synthetic peptide (MIKERIRREKQKQSY) or aHis-tagged recombinant protein (amino acid positions 27-247) and horse-radish peroxidase conjugated anti-rabbit antibodies (Bio-Rad Laboratories, Hercules, CA).

Results

Origin of NtWIN4

NtWIN4 was reported to exist as a single copy in the *N. tabacum* genome (Figure 2). Since *N. tabacum* is an amphidiploid (2n = 48), derived from *N. sylvestris* (2n = 24) and *N. tomentosiformis* (2n = 24) (Figure 12A), its origin was first examined by PCR. Results showed effective amplification of *NtWIN4* sequences with genomic DNA from *N. sylvestris*, but not from *N. tomentosiformis* (Figure 12B). This indicates that *NtWIN4* originated from *N. sylvestris*, in which the corresponding gene was designated as *NsWIN4* (*N. sylvestris WIN4)*. To examine whether both genes function similarly, their transcript accumulation profiles upon wounding were examined. In unstressed healthy leaves, neither gene was expressed (Figure 12C). When leaves were mechanically injured, transcripts began to accumulate, reaching the maximal level after 1 h, and gradually declined up to 3 h (Figure 12C). These results indicated that *NtWIN4* was transmitted from *N. sylvestris* to *N. tabacum* through amphidiploidization, and that the gene is equally involved in wound-stress responses in both species.

Structural properties

An amino acid sequence comparison between NtWIN4 and NsWIN4 revealed the two to be almost identical, except that tryptophan at position 70 in the former is substituted with glycine in the latter (Figure 13A). On DNA sequencing, this mutation corresponded to "GGG" in *N. sylvestris* instead of "TGG" in *N. tabacum* (Figure 13A, middle panel). When this region was compared among other plant species, "TGG" was found to be predominant (Figure 13A, bottom panel), indicating that the "GGG" mutation in *N. sylvestris* might have occurred after amphidiploid formation. To check whether or not this point mutation affects protein properties, localization and transcriptional repression activities of NsWIN4 were compared with those of NtWIN4. Cellular localization was first examined with GFP-tagged fusion proteins (Figure 13B). Since the N-terminus of NtWIN4 was found to be critical for localization (Figures 4, 5 and 6), two different fusion constructs were prepared; one a polypeptide consisting of 111 amino acids from the N-terminus (1-111), and the other a similar polypeptide, but with the first 26 amino acids deleted (27-111). Results showed that constructs of $NsWIN4^{1-111}$ and $NtWIN4^{1-111}$ equally localized in nucleus and cytoplasm, whereas $NsWIN4^{27-111}$ and $NtWIN4^{27-111}$ proteins were localized to plastids (Figure 13B), consistent with above observations (Figures 4 and 5). Nuclear transcription assays *in vivo* were then performed using the Luciferase-reporter system. *Gal4BD-NsWIN4* and *Gal4BD-NtWIN4* effector constructs equally suppressed the luciferase activity to a level 1/3 that of the *Gal4BD* control (Figure 13C). These results indicate that the observed point mutation does not influence protein properties, which are shared in common between NsWIN4 and NtWIN4.

Natural selection

The molecular mass of full length NtWIN4 was calculated to be 28 kDa. However, native NtWIN4 was shown to be a 17 kDa protein by immuno-blot staining, and exclusively located in plastids (Figure 4). Peptide dissection analysis suggested that a polypeptide with 67 amino acids starting at position 27 from the N-terminus serves as a chloroplast transit signal, and is cleaved off after translocation to yield the mature plastidic 17 kDa protein *in planta*. To see whether this is also the case for NsWIN4, immuno-blot analysis was carried out using total proteins extracted from leaf tissues of both *N. sylvestris* and *N. tabacum* (Figure 14). Results showed that, in the sample from *N. tabacum,* a single major signal was detected at the position of 17 kDa, while in the sample from *N. sylvestris,* two signals at positions of 17 and 26 kDa were distinct (Figure 14). Thus NsWIN4 simultaneously takes two molecular forms *in planta*, one of 17 kDa localizing in plastids, and the other of 26 kDa localizing in the cytoplasm and nucleus. This implies that NsWIN4 could be bifunctional in plastids and the nucleus, while NtWIN4 could be mono-functional only in plastids, and that such a conversion might be the result of natural selection associated with the amphidiploidization in the distant past.

Variation in mRNA length

When translated from the first AUG codon on mRNA, NtWIN4 was found to be localized in the cytoplasm and nucleus. When translated from the in-frame second AUG, the product is localized in plastids (Figure 4). It is thus conceivable that diversification in size and localization between NtWIN4 and NsWIN4 could have arisen from diversification of transcription and/or translation systems. To verify this hypothesis,

diversity of mRNA length was first examined by isolating as many mRNAs as possible from *N. sylvestris* and *N. tabacum* by the 5'RACE method (Figure 15A). Resulting clones were classified into three groups for *N. sylvestris*, and four groups for *N. tabacum*, and tentatively named as S1, S2 and S3 for the former and T1, T2, T3 and T4 for the latter (Figure 15A). In S1, S2, T1, T2 and T3 sequences, a distinct 5' untranslated region (UTR) was found to exist upstream of the first AUG codon. In contrast, in S3 and T4 sequences, the 5' terminal nucleotide was guanine located in the first AUG codon set, indicating that protein synthesis from the first AUG codon is improbable (Figure 15A). To test whether or not these sequences encode proteins, *in vitro* transcription/translation coupled assays were performed. S2, T2, T3 sequences were found to serve as templates for the 26 kDa protein, which is the full length protein starting from the first AUG codon (Figure 15B). In contrast, S3/T4 sequences encoded a 24 kDa protein, which was possibly synthesized from the in-frame second AUG codon as a precursor of the 17 kDa protein (Figure 15B). These experiments revealed various sizes of mRNA, which could serve as translation templates. However, translation efficiency greatly differed among the mRNA populations. The efficiency was highest with S3/T4 and lowest with S1, T1 and T2. The efficiency of S2 was lower than that of S3/T4, but appeared to be compatible with that of T3. Densitometric estimation of the signal intensity showed the translation efficiency of each sequence relative to the S3/T4 level to be less than 1/10 for T1, S1 and T2, 1/4 for T3 and 1/2 for S2 (Figure 15C). The results showed that a 24 kDa protein is predominantly produced in both species, and that a small amount of 26 kDa protein is also produced at a higher level in *N. sylvestris* than in *N. tabacum*. Thus differential translation efficiency between two species may partly account for the presence of 26 kDa protein in *N. sylvestris*. However, this idea

alone does not explain the absence of the 26 kDa protein in *N. tabacum*.

mRNA populations

The above-described results suggested that the major protein product with 17 kDa, which is the mature form of the 24 kDa protein *in planta*, is preferentially synthesized from T4 mRNA (tentatively referred as short mRNA) in *N. tabacum*. In contrast, two proteins of 17 kDa and 26 kDa are synthesized from S3 mRNA (short mRNA) and S2 mRNA (long mRNA), respectively, in *N. sylvestris*. However, in *N. tabacum*, the long mRNA is present and is translated in vitro, although the efficiency was half that of *N. sylvestris*. Since no 26 kDa protein was detected in *N. tabacum,* there must be some mechanism which interferes with production of the 26 kDa protein. One possibility concerns mRNA populations and therefore I examined the frequency of each mRNA in *N. tabacum* and *N. sylvestris* by RT-PCR using cDNA libraries from to two species. The forward primer sets 1 and 2 were designed to position upstream and downstream of the first AUG codon, respectively, thus facilitating amplification of long and short mRNAs (Figure 15D). The results revealed three distinct features. First, the short mRNA was efficiently amplified in both species, indicating its dominance. Second, the amplification efficiency of the long mRNA was much lower than that of the short mRNA in both species, suggesting the long mRNA to be the minor form. Third, its amplification was nevertheless over 2-fold higher in *N. sylvestris* than in *N. tabacum* (Figure 15D). Based on these observations, I conclude that alternative transcription might be responsible for variable mRNA population in both plant species, and that the production of 26 kDa protein in *N. sylvestris* is partly attributable to a higher frequency of the long mRNA in comparison with *N. tabacum*. Indeed, the ratio of the long mRNA

to the short mRNA was very low in *N. tabacum,* supporting almost no production of 26 kDa protein *in planta*.

Discussion

This paper documents compelling evidence of molecular mechanisms for differential expression and cellular localization of a bHLH protein between two tobacco species. In preceding chapters, I found that NtWIN4, a typical bHLH protein from *N. tabacum*, was translocated into plastids due to conversion of its DNA binding domain to a plastid transit peptide. In chapter III, I identified its counterpart, NsWIN4 from the ancestral *N. sylvestris*, and found both mRNA populations and translation efficiencies to have diverged between two species upon polyploidization.

Variable mRNAs

The size diversity of mRNA can be best explained by the idea of multiple transcription start sites, which was recently proposed based on studies of mammalian cells (Carninci et al. 2006). In mammals, promoters can be classified into two types, those enriched with TATA-box motifs and those enriched with CpG dinucleotides. The former possesses one transcription start site, whereas the latter possesses multiple transcription start sites. CpG-rich promoters are much more abundant than TATA-box enriched promoters, and are considered to have rapidly evolved through epigenetic control (Carninci et al. 2006). In the case of *NtWIN4/NsWIN4*, however, no apparent CpG islands were found within their approximately 1 kb promoter regions (data not shown). This may suggest the presence of a so far unknown mechanism for multiple transcription starting in plants.

The question then arises as to why transcription start sites differ between *N. sylvestris* and *N. tabacum*. Three causes are conceivable: mutation in genomic DNA, genomic rearrangement and addition of new components. The first probability was examined with approximately 1 kb promoter regions of *NtWIN4* and *NsWIN4,* but only a few differences in nucleotide sequence were identified between the two (data not shown). This finding makes the first idea unlikely, and also the second idea of genomic rearrangement unlikely as well. The third possibility that new regulatory components have been added, however, is worthy of consideration. *N. sylvestris* and *N. tomentosiformis* had independently evolved until the amphidiploid *N. tabacum* was formed. This means that each had developed it own transcription system, and that *N. tabacum* inherited both. Although *N. tomentosiformis* does not possess a *NtWIN4*-like gene, it is highly probable that the transcription system derived from *N. tomentosiformis* partly and differently might control expression of *N. sylvestris*-derived *NtWIN4*, resulting in production of unique mRNA sizes in *N. tabacum*.

In addition to size diversity, change in translation efficiency of each mRNA also appeared to be responsible for preferential production of certain protein species. Previously I proposed that the nucleotide length of the 5' UTR might be critical for translation initiation from the first AUG codon (Kodama and Sano 2006). This idea was based on the finding that *NtWIN4* mRNA possesses only a 17-base UTR, considered to be insufficient for recognition of the first AUG. The present finding is partly consistent with this speculation in regard to translation initiation being undetectable with the T1 mRNA (17-base UTR), but partly inconsistent with reference to translation efficiency, showing that shorter mRNAs such as S2 (7-base UTR) and T3 (2-base UTR) than the T1 mRNA were relatively efficiently translated from the first AUG. These observations

suggest that translation efficiency is not necessarily determined by nucleotide length of the 5'UTR, but perhaps by higher conformation of the 5' terminal regions. Structural analysis of these mRNA species is necessary to determine the mechanism.

Biological significance

Another question to be addressed concerns the biological functions of NtWIN4 and particularly of NsWIN4. *NtWIN4* was initially identified as a wound-responsive gene, and shown to be up-regulated not only by wounding but also by pathogen attack, possibly mediated through jasmonate and hydrogen peroxide (Figures 2 and 7). Transgenic approaches indicated induction of hypersensitive cell death by disrupting chloroplasts and it was concluded that NtWIN4, initially a nuclear-resident transcription regulator, evolved into a plastid-resident protein, resulting in acquisition of novel function(s) other than transcriptional regulation. In this context, NtWIN4 can be regarded as a bifunctional protein on the evolutionary time scale, although it is currently mono-functional in *N. tabacum*. In contrast, NsWIN4 is apparently bifunctional at present, serving as a transcription repressor in nucleus and as a chlorosis-inducer in *N. sylvestris*. To verify the activity as a transcription repressor, I performed gel-shift experiments with bacterially expressed NtWIN4 for binding to the G-box (CAGCTC) sequence, a common cis-element for bHLH proteins. However, results were negative, and subsequent yeast two hybrid assays indicated that the observed failure to bind was due to its inability to form homodimers, which are generally required for bHLH proteins to function (unpublished observation). It is conceivable that NtWIN4/NsWIN4 might be a negative regulator of expression of so far unidentified wound/pathogen responsive genes by forming heterodimers with other bHLH protein(s).

Overall, my work points to one series of molecular events which have occurred during plant evolution. Three steps took place to change protein function in this case: first, diversification of mRNA population caused by varied translation initiation; second, conversion of polypeptide function from DNA binding to plastid transit signaling; and third, formation of amphidiploid species from two independent ancestral species. NsWIN4 might have originally been bifunctional due to long and short mRNA populations, respectively encoding large (26 kDa) and small (24 kDa) proteins. The N-terminus of the small protein was by chance suited to function as a transit peptide after translocation into plastids while that of the large protein was not due to masking effects by the first 26 amino acid residues (Kodama and Sano 2006). Upon amphidiploidization, *NsWIN4* was transmitted to *N. tabacum*, as a consequence of which the frequency of long mRNA production and efficiency of its translation were reduced. This resulted in decline of the 26 kDa form, leaving NtWIN4 as a mono-functional plastid protein. The absence of NtWIN4 as a transcription regulator might have not been a major disadvantage for *N. tabacum*, possibly due to functional redundancy with similar proteins derived from the other ancestor, *N. tomentosiformis*. On the contrary, such a functional conversion might have been of advantage to better cope with environmental stresses, so that it was stably fixed as a new trait in *N. tabacum*. This model proposes the hypothesis that functional alteration of a protein is one of factors that induce fitness to survive severe environmental conditions.

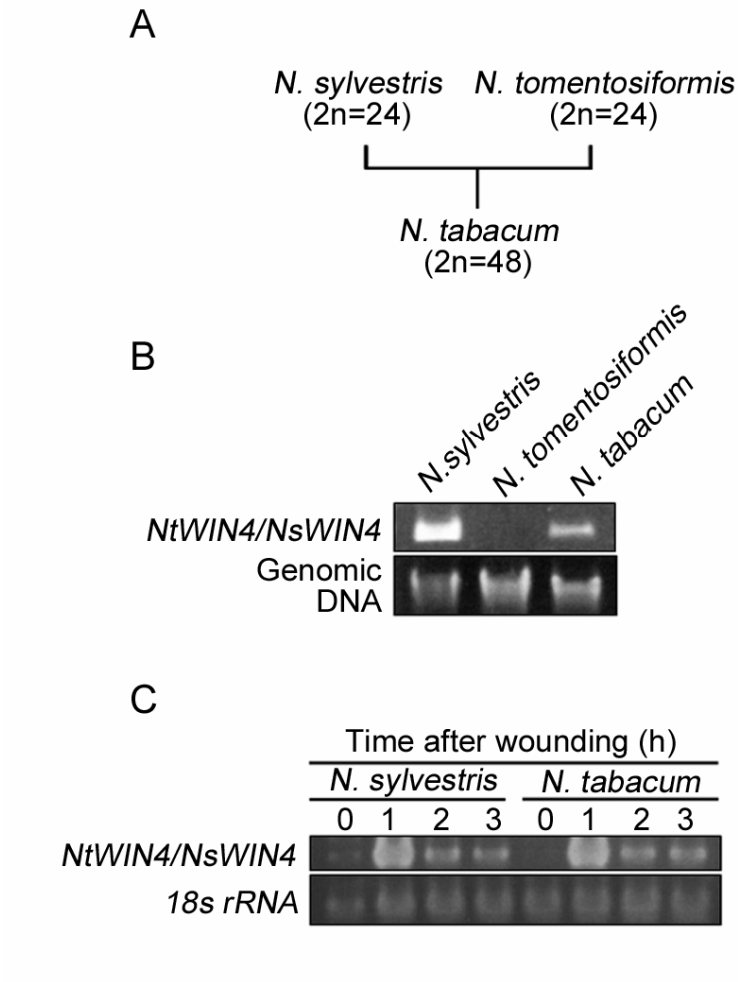

Figure 12. Properties of NtWIN4/NsWIN4. (**A**) Schematic illustration of amphiploidization. *N. tabacum* (2n=48) was formed from two ancestors, *N. sylvestris* (2n=24) and *N. tomentosiformis* (2n=24). (**B**) Determination of the *NtWIN4* ancestor. Genomic DNA was isolated from leaves of *N. sylvestris*, *N. tomentosiformis* and *N. tabacum*, and PCR analyses were performed with *NtWIN4* primer sets as described in the Materials and Methods. Amplified DNA fragments and 500 ng of intact genomic DNA as the control were loaded. (**C**) Wound-responses of *NtWIN4* and NsWIN4. Healthy leaves from *N. sylvestris* or *N. tabacum* were mechanically wounded, and harvested at the indicated time points. Total RNA was isolated from these leaves, and cDNA was synthesized by the reverse-transcriptase reaction with olido-dT primers. *NtWIN4/NsWIN4* and *18S rRNA* genes were amplified by 35- and 20-cycles of PCR, respectively.

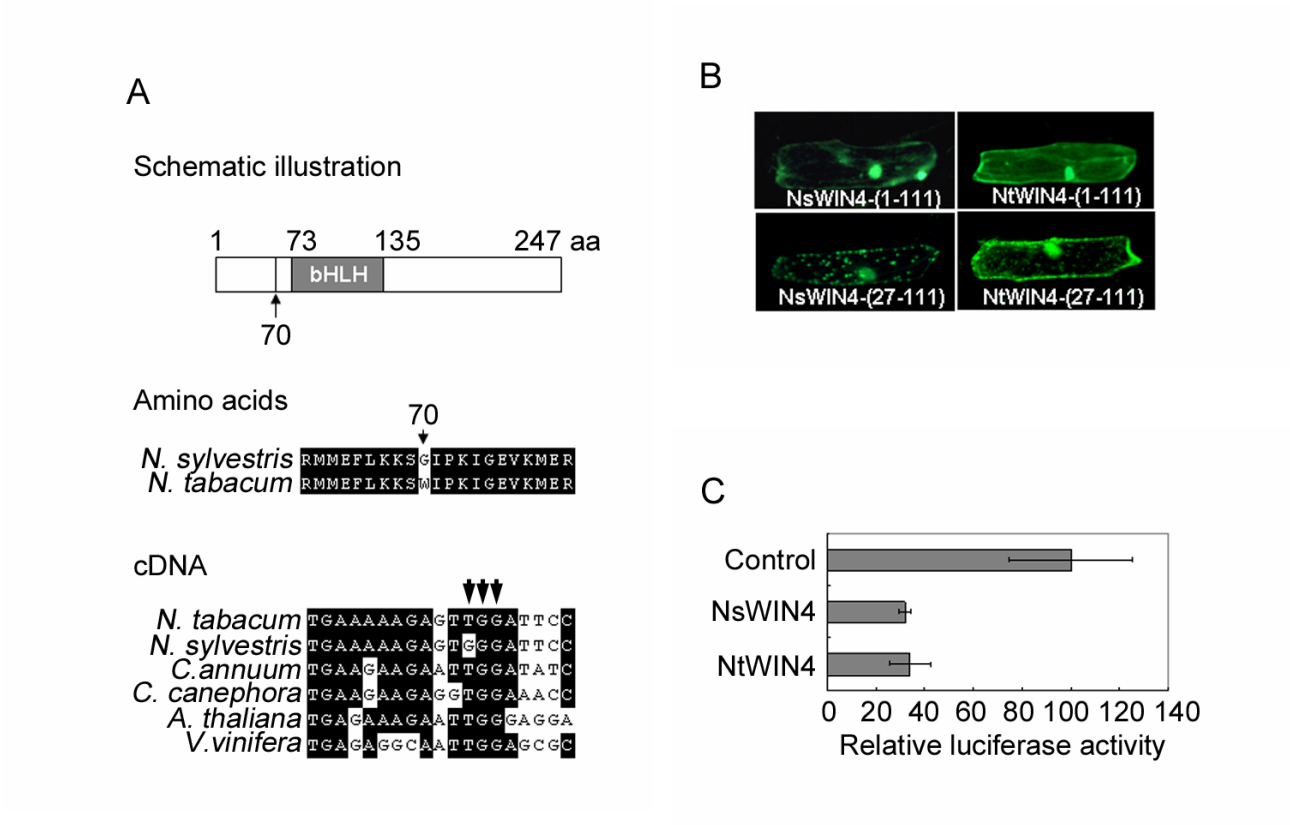

Figure 13. Cellular localization and transcriptional regulation. (**A**) Upper panel: Schematic illustration of full length NsWIN4/NtWIN4 consisting of 247 amino acids. A bHLH motif is located at amino acid positions 79-135. The arrow indicates the mutation at position 70. Middle panel: Amino acid sequences around the mutation at position 70. Bottom panel: cDNA sequence comparison among *N. tabacum, N. sylvestris* and other plants, *Capsicum annuum*, *Coffea canephora*, *Arabidopsis thaliana* and *Vitis vinifera*. Arrows indicate the triplet corresponding to the mutated amino acid at position 70. (**B**) Cellular localization. NsWIN4-(1-111), NtWIN4-(1-111), NsWIN4-(27-111) and NtWIN4-(27-111) were fused to the N-terminus of GFP, and fusion proteins were transiently expressed in onion epidermal cells after particle bombardment. Cells were examined by microscopy with the specified filter. (**C**) Dual-luciferase assays. *GBD-NsWIN4* and *GBD-NtWIN4* effector plasmids were bombarded with reporter and internal control plasmids into five-day old BY2 cells. As the effector control, a yy64 vector alone was used.

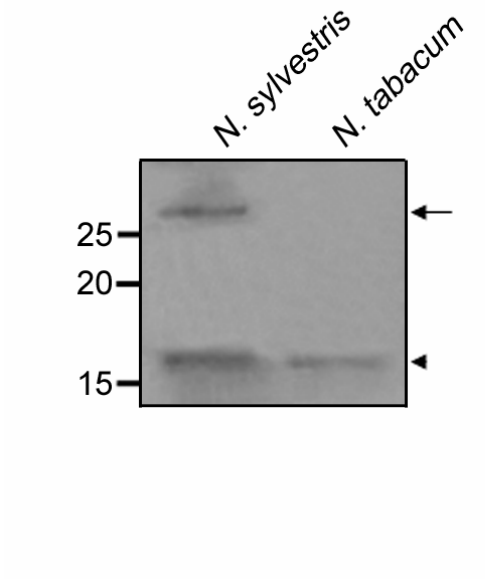

Figure 14. Immuno-blot staining. Total proteins were extracted from leaves of *N. tabacum* or *N. sylvestris*, and crude extracts were directly used for immuno-blotting. After fractionation by SDS-PAGE, proteins were electronically transferred onto cellulose membranes, and NtWIN4/NsWIN4 was detected with rabbit antibodies raised against a synthetic NtWIN4 peptide and horse-radish peroxidase conjugated anti-rabbit antibodies. The arrow and arrowhead indicate positions in 26 kDa and 17 kDa proteins, respectively.

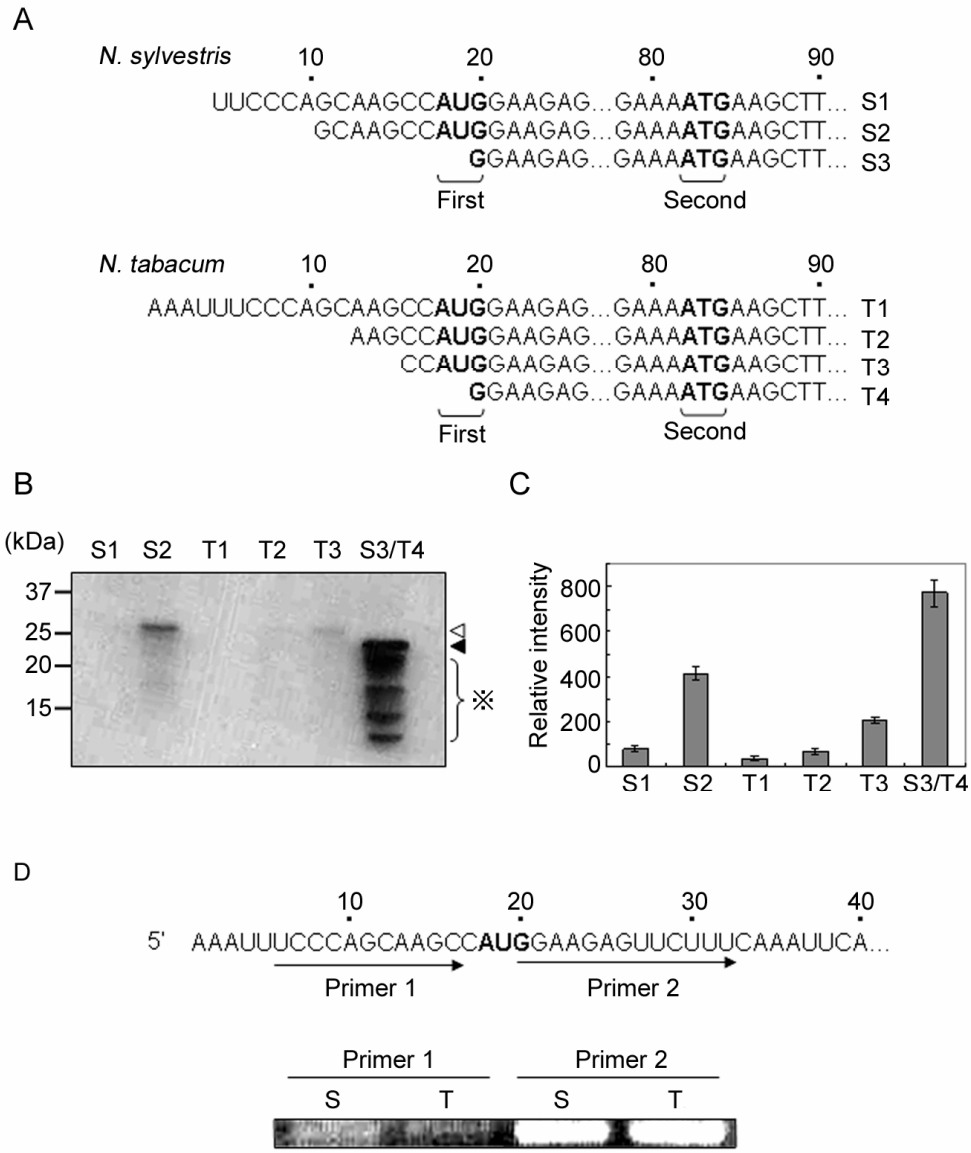

Figure 15. Properties of mRNAs. (**A**) Diversity in mRNA length. cDNAs corresponding to *NsWIN4* and *NtWIN4* were isolated from *N. sylvestris* and *N. tabacum*, respectively, and classified according to length; S1 through S3 are from *N. sylvestris* and T1 through T4 are from *N. tabacum.* The first and in-frame second AUG codons are indicated by bold characters. (**B**) *In vitro* transcription/translation coupled assays. Each cDNA clone shown above was subjected to assays using the TNT® SP6 high-yield protein expression system according to the manufacturer's instructions as described in the text. Resulting crude extracts were fractionated by SDS-PAGE and products visualized by immuno-blot staining with rabbit antibodies raised against a recombinant NtWIN4 protein and horse radish conjugated anti-rabbit antibodies. Molecular masses of migrating proteins are indicated on the left. The position of the 26 kDa protein is indicated by an open arrowhead, and that of 24 kDa protein by a closed arrowhead. The asterisk indicates non-specific signals. (**C**) Densitometric estimation of signal intensities shown in (B). Values are relative to that of S3/T4. (**D**) Population analysis of long and short mRNAs. Primers were designed to position before (Primer 1) and after the first AUG (upper panle), and RT-PCR was performed (lower panel) using cDNA templates prepared from *N. sylvestris* (S) or from *N. tabacum* (T).

CONCLUDING REMARKS

This thesis dealt with evolution of NtWIN4. NtWIN4 is a novel plastid-resident protein, containing a motif of nuclear transcriptional factor, and converted from a nuclear transcriptional repressor to a plastid-resident regulatory factor involved in hypersensitive cell death in tobacco plants during evolution. For such a conversion, a nuclear protein essentially requires localization change from nucleus to plastids. Although five such cases previously have been described before NtWIN4, a key structure for localization change has never been proposed. In this section, I discuss a possible common structure, which may play a key role for plastid localization among plastid-resident eukaryotic ancestors.

Plastid proteins are initially synthesized as precursors containing the N-terminal transit peptide, which is generally thought to be a removable polypeptide essential for plastid localization (Buchanan et al. 2000). In this section, transit peptide is defined to possess two features; the recognition sequence as the plastid localization signal, and the presequence as the removable sequence.

The presequence of CND41 was predicted by direct amino acid sequencing of the N-terminus of mature protein, revealing that 120 amino acids at the N-terminus are cleaved out when CND41 is incorporated in chloroplasts (Nakano et al. 1997) (Table 1). It is not clear whether the presequence alone is sufficient for plastid localization. In the case of NtWIN4, a clear recognition sequence could be identified although a presequence was not determined (Figures 5 and 6, Table 1). The C-terminal region with 15 amino acids of the recognition sequence was found to partly overlap with a basic region of the bHLH motif, which serves as a DNA binding domain. This indicates that

the overlapping sequence is critical for DNA binding and plastid localization. However, since the native molecular mass of NtWIN4 is 17 kDa, it is highly probable that this region is cleaved out, resulting in NtWIN4 that lacks DNA binding activity *in planta*. PEND has an N-terminal presequence of 15 amino acids, which is cleaved out when PEND is translocated into the chloroplast envelope (Sato et al. 1998; Sato and Ohta 2001) (Table 1). However, a GFP fusion protein containing only the presequence was not translocated into plastids, in contrast to a fusion protein containing the basic region of bZIP (Ohki and Sato 2000) (Table 1), suggesting this latter to be prerequisite for the recognition sequence of PEND. PEND homologs were also shown to contain a basic region at the N-terminus and were localized to plastids, suggesting a common localization system for this family (Terasawa and Sato 2005). A basic region was also found in the N-terminus of CND41 at amino acid positions 121-138 in its precursor protein (Nakano et al. 1997). Within N-termini of PD1 and PD3, basic regions could also be identified at around amino acid positions 100 and 35, respectively (Figure 16). Thus, a basic region at the N-terminus might be a key structure, which enables a protein to convert from a nuclear transcription factor to a plastid-resident protein.

DNA binding proteins such as transcription factors might originally possess the basic region at high probability. However, whether or not the basic regions are recognized as transit peptides may depend on properties of individual proteins. Whatever the mechanism, the present survey points to two novel features concerning protein localization. First, proteins containing a transcription factor motif are not always nuclear transcription factors. Particularly, careful examination is needed if a basic region exists at the N-terminus of a given protein. Second, studies on such proteins may provide evidence for the hypothesis that plants have developed the ability to convert

protein motifs, thereby acquiring a variety of functional proteins to best respond to severe environmental stresses.

In summary, this thesis revealed physiological function and plastid localization mechanisms of NtWIN4, and suggested a key structure for localization change from nuclei to plastids among plastid-resident eukaryotic ancestors. NtWIN4 functions in wound- and pathogen-responses, and participates in hypersensitive cell death through chloroplast disruption. NtWIN4 changes its localization from nucleus to plastids by motif conversion from bHLH to transit peptide, and by diversity of mRNA during amphidiploid formation. Based on structural features of other plastid eukaryotic ancestors, I found that the N-terminal basic region is a possible key structure for localization change, and propose a model of protein evolution with localization change from nuclei to plastids (Figure 17). Nuclear proteins, which originally possess the basic region at N-termini, are converted to plastidic proteins by various events (Figure 17). These events are perhaps specified to each protein, and might occur under different environmental conditions. This study demonstrated one of molecular mechanisms, by which a protein evolved to fit various environments.

Figure 16. Basic regions within N-terminuses of PD1 and PD3. The N-terminal 240 amino acids of PD1 (upper panel) and PD3 (lower panel) are shown. Arrowheads indicate basic residues (H, K and R) and underlining indicates the region of clustered basic residues.

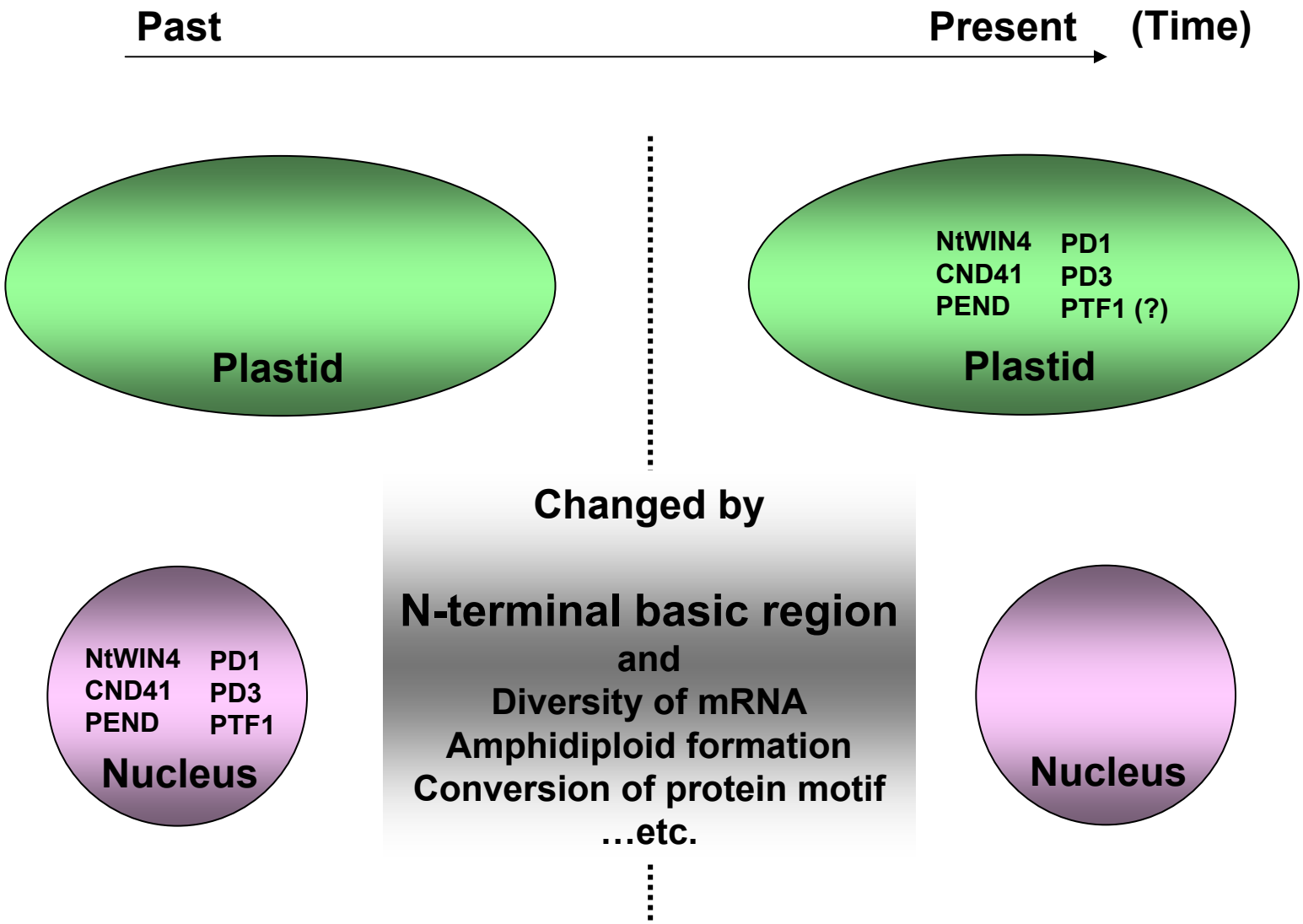

Figure 17. The model of protein evolution with localization change from nuclei to plastids.

REFERENCES

Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408, 796-815.

Archambault, J., Milne, C. A., Schappert, K. T., Baum, B., Friesen, J. D., and Segall, J. (1992). The deduced sequence of the transcription factor TFIIIA from Saccharomyces cerevisiae reveals extensive divergence from Xenopus TFIIIA. J. Biol. Chem. 267, 3282-3288.

Arkus, K. A. J., Cahoon, E. B., and Jez, J. M. (2005). Mechanistic analysis of wheat chlorophyllase. Arch. Biochem. Biophys. 438, 146-155.

Baba, K., Nakano, T., Yamagishi, K., and Yoshida, S. (2001). Involvement of a nuclear-encoded basic helix-loop-helix protein in transcription of the light-responsive promoter of *psb*D. Plant Physiol. 125, 595-603.

Benedetti, C. E., Costa, C. L., Turcinelli, S. R., and Arruda, P. (1998). Differential expression of a novel gene in response to coronatine, methyl jasmonate, and wounding in the Coi1 mutant of Arabidopsis. Plant Physiol. 116, 1037-1042.

Benkova, E., Witters, E., Van Dongen, W., Kolar, J., Motyka, V., Brzobohaty, B., Van Onckelen, H. A., and Machackova, I. (1999). Cytokinins in tobacco and wheat chloroplasts. Occurrence and changes due to light/dark treatment. Plant Physiol. 121, 245-252.

Benezra, R., Davis, R. L., Lockshon, D., Turner, D. L., and Weintraub, H. (1990). The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. Cell 61, 49-59.

Bland, M. M., Matzinger, D. F., and Levings III, C. S. (1985). Comparison of the

mitochondrial genome of Nicotiana tabacum with its progenitor species. Theor. Appl. Genet. 69, 535–541.

Bohmann, D., Bos, T. J., Admon, A., Nishimura, T., Vogt, P. K., and Tjian, R. (1987). Human proto-oncogene c-jun encodes a DNA binding protein with structural and functional properties of transcription factor AP-1. Science 238, 1386-1392.

Buchanan, B. B., Gruissem, W., and Jones, R. L. (2000). *Biochemistry & Molecular Biology of Plants*. American Society of Plant Physiologists, Rockville, Maryland.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C. A., Taylor, M. S., Engstrom, P. G., Frith, M. C. *et al.* (2006).Genome-wide analysis of mammalian promoter architecture and evolution. Nat. Genet. 38, 626-635.

Chiu, W., Niwa, Y., Zeng, W., Hirano, T., Kobayashi, H., and Sheen, J. (1996). Engineered GFP as a vital reporter in plants. Curr. Biol. 6, 325-330.

Corbin, D. R., Sauer, N., and Lamb, C. J. (1987). Differential regulation of a hydroxyproline-rich glycoprotein gene family in wounded and infected plants. Mol. Cell. Biol. 7, 4337-4344.

Culver, J. N. (2002). Tobacco mosaic virus assembly and disassembly: determinants in pathogenicity and resistance. Annu. Rev. Phytopathol. 40, 287-308.

Ellis, H. M., Spann, D. R., and Posakony, J. W. (1990). extramacrochaetae, a negative regulator of sensory organ development in Drosophila, defines a new class of helix-loop-helix proteins. Cell 61, 27-38.

Emanuelsson, O., Nielsen, H., and von Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci. 8, 978-984.

Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. 300, 1005-1016.

Fan, L., Zheng, S., Cui, D., and Wang, X. (1999). Subcellular distribution and tissue expression of phospholipase Dα, Dβ and Dγ in *Arabidopsis*. Plant Physiol. 119, 1371-1378.

Farmer, E. E., and Ryan, C. A. (1990). Interplant communication: airborne methyl jasmonate induces synthesis of proteinase inhibitors in plant leaves. Proc. Natl. Acad. Sci. U. S. A. 87, 7713-7716.

Ferre-D'Amare, A. R., Pognonec, P., Roeder, R. G., and Burley, S. K. (1994). Structure and function of the b/HLH/Z domain of USF. EMBO J. 13, 180–189.

Garrell, J., and Campuzano, S. (1991). The helix-loop-helix domain: a common motif for bristles, muscles and sex. Bioessays 13, 493-498.

Garrell, J., and Modolell, J. (1990). The Drosophila extramacrochaetae locus, an antagonist of proneural genes that, like these genes, encodes a helix-loop-helix protein. Cell 61, 39-48.

Gianinazzi, S. (1970). Hypersensibilite aux virus, temperatures et proteins solubles chez le Nicotiana tabacum cv. xanthi-nc. C. R Acad. Paris, 270, 2382-2386.

Goff, S. A., Ricke, D., Lan, T. H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B. M. et al. (2002). A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science 296, 92-100.

Gonzalez, R. G., Haxo, R. S., and Schleich, T. (1980). Mechanism of action of polymeric aurintricarboxylic acid, a potent inhibitor of protein–nucleic acid

interactions. Biochemistry 19, 4299–4303.

Goodspeed, T. H. (1954). *The genus Nicotiana.* Chronica Botanica Company, Waltham, Massachusetts.

Grasser, K. D. (1995). Plant chromosomal high mobility group (HMG) proteins. Plant J. 7, 185-92.

Gray, J. C., Kung, S. D., Wildman, S. G., and Sheen, S. J. (1974). Origin of Nicotiana tabacum L. detected by polypeptide composition of Fraction I protein. Nature 252, 226–227.

Green, T. R., and Ryan, C. A. (1972). Wound-induced proteinase inhibitor in plant leaves: a possible defense mechanism against insects. Science 175, 776-777.

Hara, K., Yagi, M., Koizumi, N., Kusano, T., and Sano, H. (2000). Screening of wound-responsive genes identifies an immediate-early expressed gene encoding a highly charged protein in tobacco plants. Plant Cell Physiol. 41, 684-691.

Hanaoka, M., Kanamaru, K., Takahashi, H., and Tanaka, K. (2003). Molecular genetic analysis of chloroplast gene promoters dependent on SIG2, a nucleus-encoded sigma factor for the plastid-encoded RNA polymerase, in *Arabidopsis thaliana*. Nucleic Acids Res. 31, 7090-7098.

Heim, M. A., Jacoby, M., Werber, M., Martin, C., Weisshaar, B., and Bailey, P. C. (2003). The basic helix-loop-helix transcription factor family in plants: A genome-wide study of protein structure and functional diversity. Mol. Biol. Evol. 20, 735–747.

Hendry, G. A. F., Houghton, J. D., and Brown, S. B. (1987). Chlorophyll degradation. A biological enigma. New phytol. 107, 255-302.

Hesse, H., and Altmann, T. (1995). Molecular cloning of a cysteine synthase cDNA

from Arabidopsis thaliana. Plant Physiol. 108, 851-852.

Hiei, Y., Ohta, S., Komari, T., and Kumashiro, T. (1994). Efficient transformation of rice (*Oryza sativa* L.) mediated by Agrobacterium and sequence analysis of the boundaries of the T-DNA. Plant J. 6, 271–282.

Horton, P., Park, K., Obayashi, T., and Nakai, K. (2006). Protein subcellular localization prediction with WoLF PSORT. Proc. 4th Asia-Pacific BIOINFORMATICS Conference APBC06. 39-48.

Hörtensteiner, S. (2006). Chlorophyll degradation during senescence. Annu. Rev. Plant Biol. 57, 55-77.

Jacob-Wilk, D., Holland, D., Goldschmidt, E. E., Riov, J., and Eyal, Y. (1999). Chlorophyll breakdown by chlorophyllase: isolation and functional expression of the Chlase1 gene from ethylene-treated Citrus fruit and its regulation during development. Plant J. 20, 653-661.

Kanamaru, K., Nagashima, A., Fujiwara, M., Shimada, H., Shirano, Y., Nakabayashi, K., Shibata, D., Tanaka, K., and Takahashi, H. (2001). An Arabidopsis sigma factor (SIG2)-dependent expression of plastid-encoded tRNAs in chloroplasts. Plant Cell Physiol. 42, 1034-1043.

Kariola, T., Brader, G., Li, J., and Palva, E. T. (2005). Chlorophyllase 1, a damage control enzyme, affects the balance between defense pathways in plants. Plant Cell 17, 282-294.

Kato, Y., Murakami, S., Yamamoto, Y., Chatani, H., Kondo, Y., Nakano, T., Yokota, A., and Sato, F. (2004). The DNA-binding protease, CND41, and the degradation of ribulose-1,5-bisphosphate carboxylase/oxygenase in senescent leaves of tobacco. Planta 220, 97-104.

Kato, Y., Yamamoto, Y., Murakami, S., and Sato, F. (2005). Post-translational regulation of CND41 protease activity in senescent tobacco leaves. Planta 222, 643-651.

Kende, H. (1993). Ethylene biosynthesis. Annu. Rev. Plant Physiol. Plant Mol. Biol. 44, 283-307.

Kenton, A., Parokonny, A. S., Gleba, Y. Y., and Bennett, M. D. (1993). Characterization of the Nicotiana tabacum L. genome by molecular cytogenetics. Mol. Gen. Genet. 240, 159–169.

Kitamura, S., Inoue, M., Ohmido, N., and Fukui, K. (2000). Quantitative chromosome maps and rDNA localization in the T subgenome of Nicotiana tabacum L. and its putative progenitors. Theor. Appl. Genet. 101, 1180–1188.

Kodama, Y. (2007). Plastidic proteins containing motifs of nuclear transcription factors. Plant Biotechnol. 24, 165-170.

Kodama, Y., and Sano, H. (2006). Evolution of a basic helix-loop-helix protein from the transcriptional repressor to a plastid-resident regulatory factor - involvement in hypersensitive cell death in tobacco plants. J. Biol. Chem. 281, 35369-35380.

Landschulz, W. H., Johnson, P. F. and McKnight, S. L. (1988). The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science 240, 1759-1764.

Ledent, V., and Vervoort, M. (2001). The basic helix-loop-helix protein family: comparative genomics and phylogenetic analysis. Genome Res. 11, 754-770.

Leister, D. (2003). Chloroplast research in the genomic age. Trends Genet. 19, 47-56.

Lim, K. Y., Matyásek, R., Lichtenstein, C. P., and Leitch, A. R. (2000). Molecular cytogenetic analyses and phylogenetic studies in the Nicotiana section Tomentosae. Chromosoma 109, 245–258.

Liu, Y. G., and Whittier, R. F. (1995). Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics 253, 674-681.

Ludwig, S. R., Habera, L. F., Dellaporta, S. L., and Wessler, S. R. (1989). Lc, a member of the maize R gene family responsible for tissue specific anthocyanin production, encodes a protein similar to transcriptional activators and contains a myc-homology region. Proc. Natl. Acad. Sci. U. S. A. 86, 7092–7096.

Matassi, G., Melis, R., Macaya, G., and Bernardi, G. (1991). Compositional bimodality of the nuclear genome of tobacco. Nucleic Acids Res. 19, 5561–5567.

Matile, P., and Hörtensteiner, S. (1999). Chlorophyll degradation. Annu. Rev. Plant Physiol. Plant Mol. Biol. 50, 67–95.

Monosov, E. Z., Wenzel, T. J., Luers, G. H., Heyman, J. A., and Subramani, S. (1996). Labeling of peroxisomes with green fluorescent protein in living P. pastoris cells. J. Histochem. Cytochem. 44, 581-589.

Murakami, S., Kondo, Y., Nakano, T., and Sato, F. (2000). Protease activity of CND41, a chloroplast nucleoid DNA-binding protein, isolated from cultured tobacco cells. FEBS Lett. 468, 15-18.

Murray, M. G., and Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. 8, 4321-4325.

Murre, C., McCaw, P. S., and Baltimore, D. (1989). A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD and myc proteins. Cell 56, 777–783.

Nagashima, A., Hanaoka, M., Shikanai, T., Fujiwara, M., Kanamaru, K., Takahashi, H., and Tanaka, K. (2004). The multiple-stress responsive plastid sigma factor, SIG5,

directs activation of the *psb*D blue light-responsive promoter (BLRP) in *Arabidopsis thaliana*. Plant Cell Physiol. 45, 357-368.

Nagata, T., Okada, K., Takabe, I., and Matsui, C. (1981). Delivery of tobacco mosaic virus RNA into plant protoplasts mediated by reverse-phase evaporation vesicles (Liposomes). Mol. Gen. Genet. 184, 161-165.

Nakamura, T., Yamaguchi, Y., and Sano, H. (2000). Plant mercaptopyruvate sulfurtransferase: molecular cloning, subcellular localization and kinetic properties. Eur. J. Biochem. 267, 5621-5630.

Nakano, T., Murakami, S., Shoji, T., Yoshida, S., Yamada, Y., and Sato, F. (1997). A novel protein with DNA binding activity from tobacco chloroplast nucleoids. Plant Cell 9, 1673-1682.

Nakano, T., Sato, F., and Yamada, Y. (1993). Analysis of nucleoid-proteins in tobacco chloroplasts. Plant Cell Physiol. 34, 873-880.

O'Halloran, T. V. (1993). Transition metals in control of gene expression. Science 261, 715-725.

Ohki, Y., and Sato, N. (2000). Analysis of chloroplast targeting sygnals of the PEND protein. Plant Cell Physiol. (supplements) 40, s222.

Okamuro, J. K., and Goldberg, R. B. (1985). Tobacco single-copy DNA is highly homologous to sequences present in the genomes of its diploid progenitors. Mol. Gen. Genet. 198, 290–298.

Pearce, G., Strydom, D., Johnson, S., and Ryan, C. A. (1991). A polypeptide from tomato leaves induces wound-inducible proteinase inhibitor proteins. Science 253, 895-898.

Quail, P. H. (2000). Phytochrome interacting factors. Semin. Cell Dev. Biol. 11,

457–466.

Race, H. L., Herrmann, R. G., and Martin, W. (1999). Why have organelles retained genomes? Trends Genet. 15, 364-370.

Reeves, R., and Nissen, M. S. (1990). The A·T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. J. Biol. Chem. 265, 8573-8582.

Sato, N. (2001). Was the evolution of plastid genetic machinery discontinuous? Trends Plant Sci. 6, 151-155.

Sato, N., Albrieux, C., Joyard, J., Douce, R., and Kuroiwa, T. (1993) Detection and characterization of a plastid envelope DNA-binding protein which may anchor plastid nucleoids. EMBO J. 12, 555-561.

Sato, N., Mitsumi, O., Joyard, J., and Douce, R. (1995). DNA-binding proteins mediate interaction of nucleoids with envelope membrane in developing plastids. *In Photosynthesis: from Light to Biosphere* (Mathis, P. ed.) Vol. III, Kluwer Academic Publishers, Dordrecht, pp 635-638.

Sato, N., Ohshima, K., Watanabe, A., Ohta, N., Nishiyama, Y., Joyard, J., and Douce, R. (1998). Molecular characterization of the PEND protein, a novel bZIP protein present in the envelope membrane that is the site of nucleoid replication in developing plastids. Plant Cell 10, 859-872.

Sato, N., and Ohta, N. (2001). DNA-binding specificity and dimerization of the DNA-binding domain of the PEND protein in the chloroplast envelope membrane. Nucleic Acids Res. 29, 2244-2250.

Schilmiller, A. L., and Howe, G. A. (2005). Systemic signaling in the wound response. Curr. Opin. Plant Biol. 8, 369-377.

Sekine, K., Hase, T., and Sato, N. (2002). Reversible DNA compaction by sulfite reductase regulates transcriptional activity of chloroplast nucleoids. J. Biol. Chem. 277, 24399-24404.

Seo, S., Okamoto, M., Iwai, T., Iwano, M., Fukui, K., Isogai, A., Nakajima, N., and Ohashi, Y. (2002). Reduced levels of chloroplast FtsH protein in tobacco mosaic virus-infected tobacco leaves accelerate the hypersensitive reaction. Plant Cell 12, 917-932.

Skalická, K., Lim, K. Y., Matyásek, R., Koukalová, B., Leitch, A. R., and Kovarik, A. (2003). Rapid evolution of parental rDNA in a synthetic tobacco allotetraploid line. Am. J. Bot. 90, 988–996.

Skalická, K., Lim, K. Y., Matyásek, R., Matzke, M., Leitch, A. R., and Kovarik, A. (2005). Preferential elimination of repeated DNA sequences from the paternal, Nicotiana tomentosiformis genome donor of a synthetic, allotetraploid tobacco. New Phytol. 166, 291–303.

Soll. J, and Tien, R. (1998). Protein translocation into and across the chloroplastic envelope membranes. Plant Mol. Biol. 38, 191-207.

Soltis, P. S. (2005). Ancient and recent polyploidy in angiosperms. New Phytol. 166, 87-110.

Struhl, K. (1987). The DNA-binding domains of the jun oncoprotein and the yeast GCN4 transcriptional activator protein are functionally homologous. Cell 50, 841-846.

Sugimoto, M., Yamaguchi, Y., Nakamura, K., Tatsumi, Y., and Sano, H. (2004). Hypersensitive response-induced ATPase associated with various cellular activities (AAA) protein from tobacco plants. Plant Mol. Biol. 56, 973-985.

Takamiya, K., Tsuchiya, T., and Ohta H. (2000). Degradation pathway (s) of chlorophyll: what has gene cloning revealed? Trends Plant Sci. 5, 426-431.

Tamura, T., Hara, K., Yamaguchi, Y., Koizumi, N. and Sano, H. (2003). Osmotic stress tolerance of transgenic tobacco expressing a gene encoding a membrane-located receptor-like protein from tobacco plants. Plant Physiol. 131, 454-462.

Tang, L., Okazawa, A., Itoh, Y., Fukusaki, E., and Kobayashi, A. (2004). Expression of chlorophyllase is not induced during autumnal yellowing in *Ginkgo biloba*. Z. Naturforsch. C 59, 415-120.

Terasawa, K., and Sato, N. (2005). Occurrence and characterization of PEND proteins in angiosperms. J. Plant Res. 118, 111-119.

Thum, K. E., Kim, M., Christopher, D. A., and Mullet, J. E. (2001). Cryptochrome 1, cryptochrome 2, and phytochrome a co-activate the chloroplast *psbD* blue light-responsive promoter. Plant Cell 13, 2747-2760.

Toledo-Ortiz, G., Huq, E., and Quail, PH. (2003). The Arabidopsis basic/helix-loop-helix transcription factor family. Plant Cell 15, 1749-1770.

Tsuchiya, T., Ohta, H., Okawa, K., Iwamatsu, A., Shimada, H., Masuda, T., and Takamiya, K. (1999). Cloning of chlorophyllase, the key enzyme in chlorophyll degradation: finding of a lipase motif and the induction by methyl jasmonate. Proc. Natl. Acad. Sci. U. S. A. 96, 15362-15367.

Tsunoyama, Y., Ishizaki, Y., Morikawa, K., Kobori, M., Nakahira, Y., Takeba, G., Toyoshima, Y., and Shiina, T. (2004). Blue light-induced transcription of plastid-encoded *psbD* gene is mediated by a nuclear-encoded transcription initiation factor, AtSig5. Proc. Natl. Acad. Sci. U. S. A. 101, 3304-3309.

Tuskan, G. A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam,

N., Ralph, S., Rombauts, S., Salamov, A. et al. (2006). The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science 313, 1596-1604.

Ueda, H., Yamaguchi, Y., and Sano, H. (2006). Direct interaction between the helicase domain of tobacco mosaic virus and ATP-bound N factor during the hypersensitive response in tobacco plants. Plant Mol. Biol. 61, 31-45.

Uefuji, H., Tatsumi, Y., Morimoto, M., Kaothien-Nakayama, P., Ogita, S., and Sano, H. (2005). Caffeine production in tobacco plants by simultaneous expression of three coffee N-methyltrasferases and its potential as a pest repellant. Plant Mol. Biol. 59, 221-227.

Van Doren, M., Powell, P. A., Pasternak, D., Singson, A., and Posakony, J. W. (1992). Spatial regulation of proneural gene activity: auto- and cross-activation of achaete is antagonized by extramacrochaetae. Genes Dev. 6, 2592-2605.

Vogt, P. K., Bos, T. J., and Doolittle, R. F. (1987). Homology between the DNA-binding domain of the GCN4 regulatory protein of yeast and the carboxyl-terminal region of a protein coded for by the oncogene jun. Proc. Natl. Acad. Sci. U. S. A. 84, 3316-3319.

Wada, Y., Ohya, H., Yamaguchi, Y., Koizumi, N., and Sano, H. (2003). Preferential *de novo* methylation of cytosine residues in non-CpG sequences by a domains rearranged DNA methyltransferase from tobacco plants. J. Biol. Chem. 278, 42386-42393.

Waller, F., Muller, A., Chung, K. M., Yap, Y. K., Nakamura, K., Weiler, E., and Sano, H. (2006). Expression of a WIPK-activated transcription factor results in increase of endogenous salicylic acid and pathogen resistance in tobacco plants. Plant Cell Physiol. 47, 1169-1174.

Wendel, J. F. (2000). Genomic evolution in polyploids. Plant Mol. Biol. 42, 225-249.

Wolf, E. Kim, P. S., and Berger, B. (1997). MultiCoil: a program for predicting two- and three-stranded coiled coils. Protein Sci. 6, 1179-1189.

Wycliffe, P., Sitbon, F., Wernersson, J., Ezcurra, I., Ellerstrom, M., and Rask, L. (2005). Continuous expression in tobacco leaves of a Brassica napus PEND homologue blocks differentiation of plastids and development of palisade cells. Plant J. 44, 1-15.

Yamaguchi, Y., Nakamura, T., Kusano, T., and Sano, H. (2000). Three Arabidopsis genes encoding proteins with differential activitiesfor cysteine synthase and b-cyanoalanine synthase. Plant Cell Physiol. 41, 465-476.

Yamamoto, Y. Y., and Deng, X. Y. (1998). A new vector set for GAL-4-dependent transactivation assay in plants. Plant Biotechnol. 15, 217-220.

Yasuno, R., and Wada, H. (1998). Biosynthesis of lipoic acid in *Arabidopsis*: cloning and characterization of the cDNA for lipoic acid synthase. Plant Physiol. 118, 935-943.

Yoda, H., Ogawa, M., Yamaguchi, Y., Koizumi, N., and Sano, H. (2002). Identification of early responsive genes associated with the hypersensitive response and properties of a WRKY-type transcription factor in tobacco plants upon tobacco mosaic virus infection. Mol. Genet. Genomics 267, 154-161.

PUBLICATION LIST

Kodama, Y., and Sano, H. (2006). Evolution of a basic helix-loop-helix protein from the transcriptional repressor to a plastid-resident regulatory factor – involvement in hypersensitive cell death in tobacco plants. J. Biol. Chem. 281, 35369-35380.

Kodama, Y. (2007). Plastidic proteins containing motifs of nuclear transcription factors. Plant Biotechnol. 24, 165-170.

Kodama, Y., and Sano, H. (2007). A comparative analysis of basic helix-loop-helix proteins, AtPTF1 and NtWIN4, with reference to plastid localization. Plant Biotechnol. (in press).

Kodama, Y., and Sano, H. (2007). Functional diversification of a basic helix-loop-helix protein due to alternative transcription during generation of amphidiploidy in tobacco plants. Biochem. J. (in press).

Yap, Y. K., Kodama, Y., Waller, F., Chung, K. M., Ueda, H., Nakamura, K., Oldsen, M., Yoda, H., Yamaguchi, Y., and Sano, H. (2005). Activation of a novel transcription factor through phosphorylation by WIPK, a wound-induced mitogen-activated protein kinase in tobacco plants. Plant Physiol. 139, 127-137.

Yano, A., Kodama, Y., Koike, A., Shinya, T., Kim, H. J., Matsumoto, M., Ogita, S., Wada, Y., Ohad, N., and Sano, H. (2006). Interaction between methylated-DNA binding protein and Ran GTPase during cell division in tobacco cultured cells. Ann. Bot. 98, 1179-1187.

SUMMARY IN JAPANESE

Evolution of a chloroplast-resident basic helix-loop-helix protein in tobacco plants
（タバコの葉緑体局在型 bHLH 蛋白質の分子進化）

植物は、長い時間をかけた進化によって様々な環境ストレスに適応してきた。これには、植物が持つ多様な機能や現象の発達が必要であり、それを促す原動力の一つは、蛋白質の分子進化である。その現象は広範にわたるが、一例として、細胞内局在性の変化がある。特に、核から葉緑体への細胞内局在性の変化に関しては、共通の特徴を持った幾つかの蛋白質が報告されている。これらの蛋白質は、葉緑体に局在するにも関わらず、核の転写因子モチーフを持っており、進化の過程で、核から葉緑体に局在性を変えたと推察されている。しかし、どのような機構を介して局在性を変化したかは、全く明らかとなっていない。

本研究では、タバコ（*Nicotiana tabacum*）の病害および傷害に対する未知の応答機構を探索する過程で単離した遺伝子である *NtWIN4* (*N. tabacum* wound-induced clone 4) に注目した。これがコードする蛋白質は、核の転写因子モチーフを持つ、新規の葉緑体局在型因子であった。そこで、蛋白質の分子進化の一端を知るため、NtWIN4 の葉緑体局在化の分子機構と生理機能について詳細な解析を行った。

タバコは、約 600 万年前に、二つの祖先植物（*N. sylvestris*、*N. tomentosiformis*）（共に、2n=24）が自然倍数化した複二倍体植物（2n=48）である。DNA ブロットの結果、*NtWIN4* は、タバコのゲノム中に 1 コピーしか存在しないことが示された。祖先植物のゲノム DNA を鋳型にして PCR を行ったところ、*N. sylvestris* に祖先遺伝子（*NsWIN4*）が存在した。*NtWIN4/NsWIN4* は、共に、傷処理によって発現したため、倍数化の際に傷害応答性も受け継がれたことがわかった。*NtWIN4* は、タバコモザイクウイルス感染によって発現が上昇したため、傷害だけでなく病害応答機構にも関与することが示された。

NtWIN4 は、247 残基で構成されており、N 末端側に、核の転写因子のモチーフである basic helix-loop-helix (bHLH)をコードしていた。核の転写因子と予測されたため、人工の転写装置を用いて転写活性を測定したところ、確かに、転写を抑制する活性を持っていた。

NtWIN4 の核局在性を確かめるため、GFP 融合蛋白質をタマネギ表皮細胞に発現させたところ、核・細胞質に局在した。しかし、タバコの葉肉細胞を分画後、NtWIN4 抗体を用いて免疫ブロットを行ったところ、実際は、核・細胞質ではなく葉緑体に局在していた。この奇妙な局在化機構には、長さが異なる *NtWIN4* mRNA が関与していた。これらの mRNA は、長さと関係して翻訳効率が異なった。そのうち最も高い翻訳効率を持っていた最短 mRNA は、一番目の AUG コドンを欠失しており、*in vitro* で二番目のインフレーム AUG コドンから蛋白質が翻訳された。これに GFP を融合して細胞に発現させたところ、葉緑体に局在した。したがって、葉緑体型 NtWIN4 は、この最短 mRNA から翻訳されているこ

とがわかった。祖先植物 *N. sylvestris* では、核・細胞質型と葉緑体型の両方の NsWIN4 が発現していた。しかし *N. tabacum* では、葉緑体型 NtWIN4 のみ発現する。これは、約 600 万年前の倍数化後に NtWIN4 の自然淘汰が起こった可能性を示している。

NtWIN4 の N 末端には葉緑体移行配列（cTP）の存在が予測されたため、様々な長さに切断した NtWIN4 に GFP を連結した融合蛋白質を細胞に発現させ、その挙動を観察した。その結果、N 末端の 67 残基が cTP として機能した。cTP の C 末端 15 残基は、bHLH の DNA 結合領域と完全に重複していた。これは、NtWIN4 を葉緑体に局在させるため、DNA 結合領域が cTP の一部に転用された可能性を示している。

NtWIN4 の生理機能を明らかにするため、形質転換タバコを用いて解析した。過剰発現タバコは、茎、根や花が形成されず、薄緑の葉を持つ異常な表現型を示した。葉の一部に NtWIN4 を発現させたところ、葉が黄化して葉緑体の機能が崩壊した。RNAi 発現抑制タバコに病原体を感染したところ、野生タバコに比べて、過敏感細胞死が抑制された。これらは、NtWIN4 が葉緑体の機能崩壊を介して過敏感細胞死に関与することを示唆している。

以上の結果から、NtWIN4 は、核の転写抑制因子から葉緑体局在型の過敏感細胞死関連因子に分子進化した、と結論付けた。タバコへ進化する過程で、植物のストレス応答機構には、核・細胞質型より葉緑体型の NtWIN4 が必要だったと推察される。これに助力した局在性変化の分子機構は、選択的転写制御、自然淘汰および蛋白質モチーフの転用であった。本研究より、植物が限られたゲノム情報を最大限に利用することによって多様な環境に適応してきた可能性を示す、新規の蛋白質分子進化モデルを提唱した。

ACKNOWLEDGEMENTS

The author greatly thanks Prof. Hiroshi Sano (Nara Institute of Science and Technology) for valuable advice throughout the work and preparation of the thesis. The author thanks Profs. Akiho Yokota, Ko Shimamoto and Takashi Hashimoto (Nara Institute of Science and Technology) for critical reading of the thesis. The author also thanks Dr. Malcolm Moore (Intermal, Nagoya) for critical reading of the manuscripts for publication, all my colleagues in the laboratories of Plant Molecular Breeding and Plant Protein Biology (Nara Institute of Science and Technology) for valuable discussion, Mr. Hitoshi Ohya and Mrs. Chisato Ueda (Nara Institute of Science and Technology) for experimental collaboration, Dr. Yuri Munekage (Nara Institute of Science and Technology) for technical advice regarding chlorophyll fluorescence analysis, and Mr. Koji Inai (Nara Institute of Science and Technology) for providing plant materials. This work was supported by a grant from the Scientific Research for Plant Graduate Students from Nara Institute of Science and Technology, sponsored by the Ministry of Education, Culture, Sports, Science and Technology, Japan.