Localization and possible functions of Arabidopsis HOTHEAD protein

by

Pei-Chun Chang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Biology

Waterloo, Ontario, Canada, 2016

© Pei-Chun Chang 2016

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Molecular analyses using sixteen insertion-deletion polymorphic markers revealed that somatic sectoring occurs relatively frequently during the course of normal vegetative development. This is the first report that documents the spontaneous but targeted appearance of unique genomic insertions at multiple discreet loci in single plants. These sectors hosted genetic variation attributed to single nucleotide changes, insertions, or sequence loss. The most important finding is the appearance of a 54 base-pair insertion in the progeny that resulted in an identical sequence match with the corresponding allele of the grand-parental genome, rather than the genome of the immediate parent.

Because somatic sectoring was observed five times more frequently in *hothead* (*hth*) mutant plants, studies were initiated to determine HTH protein localization with a view toward elucidating its possible functions. Here, I present the results of HTH protein localization using a fluorescent protein-tagged HTH fusion protein generated by a native promoter-driven construct (annotated as HTH_{pro}*:HTH-FP)*. The HTH-FP protein was predominantly localized to the epidermis of seedling and mature tissues; moreover, it was also present in the seed coat outer integument that is of epidermal origin. Most interestingly, in seedlings the HTH-FP protein was localized to the endoplasmic reticulum (ER) and ER-derived structures called ER bodies. Since ER bodies have been previously associated with stress response, the ER body localization suggests a role of HTH in stress responses. This notion is supported by the effect of the wounding hormone methyl jasmonate which elevated

HTH expression in wildtype plants and induced ER bodies in rosette leaf epidermis of HTH_{pro}*:HTH-FP* plants.

Previously, HTH has been proposed to function either as a mandelonitrile lyase involved in cyanogenesis or as a fatty alcohol dehydrogenase involved in the biosynthesis of cutin monomers (fatty acids). To determine whether HTH has any of these catalytic activities, a maltose binding protein (MBP)-HTH fusion protein was generated in bacteria and the recombinant protein used for *in vitro* assays. Although results of the enzymatic assays were inconclusive, bioinformatics analyses of putative catalytic residues favor functional involvement in fatty alcohol dehydrogenation, rather than in nitrile lyation. This prediction suggests that HTH might be functionally distinct from the closely-related enzyme, mandelonitrile lyase. In addition, coexpression analysis showed that *HTH* is coexpressed not only with genes involved in cutin synthesis but also with those modulated by pathogens and stress. Although results also pointed towards an association between HTH and defense/stress response, how this association might be linked to genome instability observed in *hth* mutants is discussed.

Acknowledgements

I thank my supervisor Dr. Susan Lolle who guided me through the winding path of scientific research that is full of challenges and adventure. She sets an example of how vision and determination can drive a scientist to persevere.

I would like to thank Dr. Matthew Smith, Dr. Simon Chuong, and Dr. Merrin Macrae for serving on the committee and providing advice. Invaluable suggestions by Dr. Sonia Gazzarrini have pointed new directions and expanded the scope of this thesis. Special thanks are dedicated to Dr. Andrew Doxey for his help with bioinformatics and Dr. Todd Holyoak for technical assistance in enzymatic assays. I am indebted to Dr. Barbara Moffatt and her lab members who have generously shared resources and offered insights. Gratitude is due to Dr. John Heikkila and Dr. Brian Dixon for permission to use lab equipment. I am also greatly appreciative of the Associate Chair Dr. Kirsten Müller and the staff in the Department of Biology who went extra miles to be helpful.

My dearest thanks to Dr. Marianne Hopkins, Daryl Enstone and Nimhani Perera for their expertise in molecular techniques and microscopy. I am grateful to many friends who have kindly given me professional and personal support: Maye Saechao, Lital Sever, Chris Meyer, Jesseline Gough, Duan Jin, Shanqi Zhang, Nathan Secord, Nina Bailly, Aaron Khalid, Johanna Halbauer and Chris Morcos. My pursuit of graduate school in Canada would have been impossible without the support from my family in Taiwan that always makes me feel accompanied regardless of the distance. I thank my husband Christopher Peterson for bringing fresh spirits and new perspectives into my life. I cannot express enough gratitude to

Drs. Larry and Carol Peterson for their generosity and kindness. I also would like to thank my friend Dr. Heidi Hoernig and the Hoernig's family who made Canada my home when I first arrived. Special feline meowks to Freya, Cleo and Oskar for their "help" during the write-up. Last, I would like to thank CBC radio for nourishing my soul with Canadian culture and values. Financial support was provided by Alexander Graham Bell Canada Graduate Scholarship (Natural Sciences and Engineering Research Council) and President's Graduate Scholarship (University of Waterloo).

Table of Contents

List of Figures

List of Tables

List of Appendices

List of Abbreviations & Acronyms

3'UTR: three prime untranslated region

5'UTR: five prime untranslated region

A. tumefaciens: *Agrobacterium tumefaciens*

Ab: antibody

ABA: abscisic acid

ADH: alcohol dehydrogenase

ATP: adenosine triphosphate

BioNJ: bio-neighbour-joining algorithm

bp: base pair

BSA: bovine serum albumin

CArG: DNA sequences of high similarity to the motif CC[A/T-rich]6GG

cDNA: complementary DNA

CHOX: cholesterol oxidase

COPII: coat protein complex II

DCA: dicarboxylic acids

DCA: dicarboxylic fatty acid

DMSO: dimethyl sulfoxide

DNA: deoxyribonucleic acid

dNTP: deoxynucleotide phosphate

DTT: dithiothreitol

E. coli: *Escherichia coli*

EDTA: ethylenediaminetetraacetic acid

EF-1: elongation factor-1

EMS: ethyl methanesulfonate

ER: endoplasmic reticulum

EV: empty vector

FAD: flavin adenine dinucleotide

FAE: fatty acid elongase

FAH: fatty acyl ω-hydroxylase

FAO: fatty acid oxidase

g: grams

×g: measures of relative centrifugal force

GDSL: motif consensus amino acid sequence of glycine, aspartate, serine, and leucine around the active site serine

GMC oxidoreductase: glucose-methanol-choline oxidoreductase

GMC_oxred_C: glucose-methanol-choline oxidoreductase family protein C-terminal domain

GMC_oxred_N: glucose-methanol-choline oxidoreductase family protein N-terminal domain

GOX: glucose oxidase

GPAT: glycerol 3-phosphate acyltransferase

GPI: glycosylphosphatidyl-inositol

HFADH: hydroxy fatty acyl dehydrogenase

HNL: hydroxynitrile lyase

IgG: immunoglobulin G

IPTG: isopropyl β-D-1-thiogalactopyranoside

kDa: kilo Dalton

KDEL: ER-retention signal: lysine (K), aspartic acid (D), glutamic acid (E), and leucine (L)

LB: Luria Broth

LIS-1: linum insertion sequence 1

LTP: lipid transfer protein

MDL: mandelonitrile lyase

MeJA: methyl jasmonate

mg: milligram

min: minutes

ml: millilitre

mRNA: messenger ribonucleic acid

MS: Murashige and Skoog basal medium

MW: molecular weight

MYC: myelocytomatosis transcription factor

NAC domain: NAC (NAM, ATAF1/2 and CUC2) domain; a domain of transcription factors

NAD: nicotinamide adenine dinucleotide

ng: nanograms

NNI: nearest-neighbour interchange

OFADH: oxo fatty acyl dehydrogenase

PCR: polymerase chain reaction

PMSF: phenylmethylsulphonyl fluoride

PS: polyester synthase

PYK10: a β-thioglucoside glucohydrolase

RFLP: restriction fragment length polymorphism

rpm: rotations per minute

RT-qPCR: reverse transcription quantitative polymerase chain reaction

SDS: sodium dodecyl sulphate

SDS-PAGE: SDS polyacrylamide gel electrophoresis

T1: plants of the first transformed generation

T2: plants of the second transformed generation

T3: plants of the third transformed generation

TBS-T: Tris-buffered saline with Tween 20.

UPR: unfolded protein response

UV: ultraviolet

VLCFA: very long chain fatty acid

Ws: *Arabidopsis thaliana* ecotype Wassilewskija

ΔA_{275}: Increase/decrease of absorbance of light measured at wavelength 275nm

ε: millimolar extinction coefficient

μg: microgram

μl: microliter

μm: micrometer

μmole: micromole

ω: omega

DNA code

A: adenine

C: cytosine

G: guanine

T: thymine

Amino acid code

Arg (R): arginine

Asn (N): asparagine

Asp (D): aspartic acid

Gln (Q): glutamine

Glu (E): glutamic acid

Gly (G): glycine

His (H): histidine

Ile (I): isoleucine

Lys (K): lysine

Phe (F): phenylalanine

Pro (P): proline

Ser (S): serine

Tyr (Y): tyrosine

Val (V): valine

Contributions to Knowledge

1. Documentation of individual Arabidopsis plants that are capable of producing somatic sectors during the course of normal vegetative development.

2. These somatic sectors have distinct genetic profiles, and variation is attributed to either single nucleotide composition, small DNA insertions or sequence loss.

3. The genomic insertions have characteristics that are consistent with the previously proposed template-driven mechanism (Lolle et al., 2005).

4. Cuticle-specific HTH protein function is retained by C-terminal fluorescent protein-tagged HTH proteins (HTH-FPs). HTH-FPs are localized to the epidermis of seedling and mature tissues.

5. Seeds of mutant *hth* plants are often misshapen and have enhanced seed coat permeability. Furthermore, HTH-FPs are present in the seed coat outer integument, a tissue that is of epidermal origin. This is the first report implicating HTH in seed coat development.

6. HTH-FPs derived in plants are glycosylated. The protein is localized to the endoplasmic reticulum (ER) network and ER bodies which have been previously associated with stress responses. In accordance, *HTH* expression was elevated by methyl jasmonate, a plant hormone involved in the response to wounding. These represent the first evidence implicating HTH in plant stress response pathways.

7. Predictions of catalytic residues support the postulated function of fatty alcohol dehydrogenase although HTH is phylogenetically closer to mandelonitrile lyases.

8. *HTH* is part of a network in which genes encode proteins associated with both fatty acid processing and stress responses. Moreover, these proteins are predominantly specific to land plants.

Overview of chapters

Arabidopsis *hothead* (*hth*) mutants typically exhibit floral organ fusion and increased cuticle permeability (Lolle et al., 1998). An unusual case of genome instability was observed in *hth* mutants wherein progeny appeared to acquire novel DNA sequences that were absent in the immediate parent plant but present in an earlier ancestor (Lolle et al., 2005). Little is known about how mutation of the *HTH* gene is associated with genome instability and the biochemical function of the protein. This thesis aims to 1) investigate genetic instability of molecular markers distributed across all five chromosomes, 2) examine *HTH* expression patterns, and 3) further investigate previously proposed protein functions of HTH using *in vitro* assays and bioinformatics. Based on the experimental results and bioinformatics analyses, the possible link between *hth* mutation and genome instability through plant defense responses is discussed.

In **Chapter 1**, a review of literature relevant to 1) the plant cuticle structure and its constituents, 2) genes involved in cutin and wax biosynthesis, 3) seed coat structure, 4) plant stress responses, and 5) the *hth* mutant phenotype and putative functions of the HTH protein is given; special attention is given to cutin biosynthesis and stress-related genome instability. In **Chapter 2**, we used genetic and molecular approaches to test whether the inherited genomic changes initially discovered in 2005 could be explained by outcrossing. In this chapter we showed that DNA sequence changes occur in somatic tissues and that individual plants are, in fact, genetic mosaics. The following experiments led to this key finding. First, to establish how susceptible *hth* plants were to out-crossing, mutant plants were grown

together with a pollen donor harboring a dominant gene conferring resistance to the herbicide glufosinate. Outcrossing frequencies were compared to those observed in plants isolated from exogenous pollen sources. Second, to investigate the extent of changes that occur, we used sixteen small insertion/deletion (indel) polymorphic molecular markers distributed across all five chromosomes. Using these polymorphic lines we demonstrated that these markers were not stably inherited in progeny derived from F2 parent lines with known indel marker profiles. In addition, we showed that the observed genetic discordance between parent and offspring reflected sporophytic as opposed to gametophytic events by collecting multiple tissue samples from individual soil-grown adult plants and from shoots and roots of single seedlings grown under sterile conditions. Finally, tissue samples that were subjected to quantitative assays revealed these sectors to be relatively small, assuming the copy number was proportionally related to the number of cells harbouring the reverted insertion within a fixed tissue area.

Genetic analyses have shown that the *HTH* gene is important for proper cuticle function and that its mutation leads to a floral organ fusion phenotype (Lolle et al., 1998). In **Chapter 3,** I analyzed transgenic plants expressing fluorescent protein-tagged constructs (HTH_{pro}*:HTH-FP*) that allowed direct visualization of protein localization. To minimize artifacts due to over-expression these HTH reporter constructs were driven by the native *HTH* gene promoter. Using these transgenic lines, I was able to show that HTH-FP fusion protein was predominantly localized to the epidermis of seedling and mature tissues; moreover, its expression was also present in the seed coat outer integument that is of epidermal origin. Interestingly, within epidermal cells HTH-FP was found to localize to the

endoplasmic reticulum (ER) network and ER-derived bodies that have been associated with plant stress responses. In agreement, the RT-qPCR results showed that the *HTH* expression in Ws wildtype plants was elevated by exposure to the wounding hormone methyl jasmonate (MeJA) which, in turn, also induced ER bodies in epidermal cells.

Based on sequence homology, the HTH protein belongs to the glucose-methanol-choline (GMC) oxidoreductase family. To date no direct biochemical evidence has been published to address the question of HTH protein function although two different enzymatic activities have been proposed. In 2003, Krolikowski et al. proposed that *HTH* encodes a mandelonitrile lyase (MDL) based on sequence similarity and the genetic identification of functional residues, whereas Kurdyukov et al. (2006b) proposed that *HTH* is involved in cutin monomer biosynthesis and encodes an ω-fatty alcohol dehydrogenase based on a shift in the cutin monomer profile of mutant plants (Kurdyukov et al., 2006b). In **Chapter 4**, these two possible catalytic functions were investigated using *in vitro* enzymatic assays with maltose binding protein (MBP)-tagged recombinant HTH protein generated in *Escherichia coli*. No activities were detected in these assays. Later analyses revealed that when expressed in plants HTH-FP is a glycoprotein offering one possible explanation for why expression in *E. coli* may have produced a non-functional protein. With a view towards understanding the protein function, I constructed a phylogenetic tree and compared functional residues of HTH and other GMC oxidoreductase proteins. Moreover, multiple sequence alignment was used to identify HTH's putative catalytic residues. However, predicted catalytic residues suggest that HTH shares common active sites with a bacterial medium chain fatty alcohol dehydrogenase, supporting the possibility of an enzymatic function distinct from the MDLs. Furthermore,

genes involved in cutin synthesis and also those in defense or stress responses were found to be coexpressed with HTH.

In **Chapter 5**, I summarize results from previous chapters and discuss possible mechanisms by which a putative fatty alcohol dehydrogenase such as HTH, could be associated with plant defense/stress response. Lastly, genome instability in response to stress is reviewed.

Chapter 1 General Introduction

Evolution from aquatic multicellular green algae to land plants required new strategies to cope with the terrestrial environment. The development of a cuticle, the cutin-based layer sealing the epidermis of aerial organs, allowed for the colonization and spread of land plants culminating in the gymnosperms and angiosperms. In addition to enabling plant growth on land, the cuticle layer has important roles in plant biology beyond regulating water status; it also acts as a selectively permeable barrier to control the movement of gases, solutes, small signaling molecules and charged large molecules such as herbicides (Kerstiens, 1996; Lolle and Pruitt, 1999; Pruitt et al., 2000; Schreiber, 2002; Schreiber, 2005). From a practical perspective, understanding the cuticle is important to the agriculture industry for generating more drought tolerant crops in face of global climate change and for improving the effectiveness of herbicide absorption.

The cuticle is implicated in normal plant development and is essential for achieving organ partitioning after inception of a meristem. In some cases during formation of the female reproductive organ, some epidermal cells, upon growing into direct contact, participate in cell wall fusion (Lolle et al., 1992). Regulation of organ fusion is achieved by cell-cell signaling that is mediated by small, water-soluble molecules as demonstrated by classic experiments in which barriers that block exchange of water-soluble molecules between carpel primordia have been shown to prevent fusion of *Catharanthus roseus* carpels. Using barriers of known pore size, it was demonstrated that unidentified water-soluble agents of a molecular weight less than 1000 daltons can diffuse through the fusion zone (two cell

walls and the cuticle) and allow dedifferentiation of the contacting epidermal cells (Verbeke and Walker, 1986). Accordingly, the concept of a “morphogen” molecule as the factor initiating cell dedifferentiation was proposed. Siegel and Verbeke (1989) showed that these diffusible factors could be trapped in agar, and could stimulate redifferentiation of epidermal cells to parenchymal cells by exposing the non-fusing carpel region to the factor-loaded agar barriers. These results serve as evidence that organ separation is at least in part regulated by cuticle permeability.

In Arabidopsis, defects in the cuticle can result in abnormal organ fusion as observed in cuticle mutants such as *fiddlehead* (*fdh*) and *hothead* (*hth*) (Lolle et al., 1992; Lolle et al., 1998). The cuticle also mediates cell-cell signaling influencing the spatial distribution of trichomes and stomata. Furthermore, it plays an important role in sexual reproduction. The cuticle of the stigma papillary cells serves as a diffusion barrier between the pollen grains and papillary cells since water and other regulatory factors are transported through it (Lolle and Pruitt, 1999), and as a result the cuticle has a determining effect on pollen adhesion, compatibility recognition, and pollen tube growth (Hulskamp et al., 1995).

1.1.1 **Cuticle structure**

The epidermal cuticle is a matrix consisting of cutin polymers, polysaccharide microfibrils and waxes. It is found external to the epidermal cell wall and its development has been detected as early as the late globular stage of embryogenesis. Cuticle thickness can vary greatly across plant species ranging between 0.02 to 32 µm, with mature Arabidopsis leaves generally being covered by a cuticle that is 22 - 45 nm in thickness (Franke et al., 2005;

Schreiber and Riederer, 1996; Vogg et al., 2004). Although the composition also varies across plant species, generally the cuticle found on mature tissue is a composite structure made of three layers (Figure 1.1).

Outermost are the epicuticular waxes and subjacent to this is the cuticle proper that mainly consists of cutin polymers embedded in intracuticular waxes. The third and final layer is found between the cell wall and the cuticle proper. This layer contains polysaccharides in addition to cutin polymers and waxes. The epicuticular and intracuticular waxes are hydrophobic compounds that are composed predominantly of aliphatic lipids, such as very long chain fatty acids (VLCFAs) (C24 - C34) and their derivatives (Samuels et al., 2008). When deposited on the outermost surface, waxes form a film or crystals that constitute the epicuticular wax layer. In contrast, waxes deposited in the cuticle proper are surrounded tightly by cuticle polymers and form dense, well-packed domains. It should be noted that cuticles of different parts of a plant can be highly heterogeneous as shown by Schreiber (2005) who observed that the cuticle covering *Vicia faba* stomata and trichomes is less lipophilic and thus forms the preferential site of ion penetration.

1.1.2 Cutin monomers and polymers

The three-dimensional structure of the cuticle polymer is not clear, yet the monomer compositions can readily be identified by gas chromatography and mass spectrometry. The common constituents of cutin polymers are monomers such as C16 and C18 unsubstituted fatty acids, ω-hydroxy fatty acids and dicarboxylic fatty acids; some of these monomers might contain mid-chain functional groups (Table 1.1). To a lesser extent, fatty alcohols,

glycerol and phenolics have been identified as components of cutin polymer domains (Pollard et al., 2008).

Whether cutins polymerize as branching molecules of a certain range of molecular weights, or as a greatly cross-linked network is unclear. Several theoretical three-dimensional structures of cutin (Figure 1.2) have been proposed by Pollard et al. (2008). The polymerization of ω-hydroxy fatty acids results in a linear polyester chain with primary ester linkages. The linear structure can develop branches at the sites of mid-chain oxygen-containing functional groups (epoxy, oxo, hydroxy or vicinal diol). For example, the mid-chain hydroxyls may be esterified to the carboxyl group of other monomers, forming a secondary ester linkage and a local branching structure. Alternatively, branching structures can also be achieved with the presence of glycerol (Graca et al., 2002). Any of the three hydroxyl groups of glycerol can be esterified with the carboxyl group of fatty acids. With glycerol acting as a "linker" between dicarboxylic fatty acids (DCA), a much larger branching structure can be achieved. Glycerol-DCA structures can also form extensively cross-linked network structures. However, these large branching and cross-linked polymer domains might only account for a small portion of total cutin polymer domains since dicarboxylic fatty acids normally exist as a minor monomer component (< 5%) (Pollard et al., 2008).

One exception is found in the cuticle of *Arabidopsis thaliana*, which contains high levels (> 50%) of C18:2 α,ω-dicarboxylic acid monomers (Bonaventure et al., 2004). High dicarboxylic acid content is usually diagnostic of suberin, a polyester also made of long chain

fatty acids (Bonaventure et al., 2004; Matzke and Riederer, 1991). Provided that the cuticle is found to be ten times thinner in *Arabidopsis thaliana* than many other plants, a glycerol-DCA cross-linked polymer with higher strength might be the dominating structure in *Arabidopsis thaliana* (Kurdyukov et al., 2006b; Pollard et al., 2008).

1.1.3 Cutin monomer biosynthesis

Cutin monomers are likely synthesized from fatty acids made in plastids. Figure 1.3 shows one possible order for cutin monomer synthesis. The pathway starts with pyruvate or acetate that is transformed into acetyl-CoA by the pyruvate dehydrogenase complex (PDC) or acetyl-CoA synthetase (ACS), respectively. The acetyl-CoA then goes through ATP-dependent carboxylation by acetyl CoA-carboxylase (ACCase), resulting in malonyl-CoA. With the stepwise addition of two-carbon acetyl groups, malonyl-CoA is elongated by fatty acid synthase (FAS) to fatty acids of various lengths. These molecules are either exported and integrated into membranes and other cellular components, or further elongated into C16 or C18 fatty acids that can be further processed into cutin building blocks. Plastid-derived fatty acids in turn are transported into the endoplasmic reticulum (ER). In the ER the acyl chain is first activated by long chain acyl-CoA synthase (LACS) which is then hydroxylated by fatty acyl ω-hydroxylase (FAH). The following step by ω-hydroxy fatty acyl dehydrogenase (HFADH) transforms hydroxy fatty acids to oxo products which are processed into dicarboxylic fatty acids by ω-oxo fatty acyl dehydrogenase (OFADH). Alternatively, hydroxyl acids can also be modified directly by FAH to give rise to dicarboxylic acids. These modified fatty acids may be esterified to glycerol-3-phosphate by glycerol 3-phosphate acyltransferase (GPAT) before being exported. It is possible that monomers can be esterified

by polyester synthase (PS) to form oligomers or polymer domains in the ER (Gronwald, 1991; Pollard et al., 2008).

These synthesized cutin monomers or building blocks are first exported from the ER to the cell wall and then subsequently to the cuticle. Although the key routes of cutin building block export are known, the sites for cutin polymer assembly and the presence of polyester synthases remain hypothetical. Possible mechanisms and putative cellular locations of cutin assembly are shown in Figure 1.4 (Pollard et al., 2008). In the first scenario, a plasma membrane-anchoring ER domain is in direct contact with the plasma membrane allowing monomer synthesis and transport across the membrane at one location. ATP binding cassette (ABC) transporters may be required for direct transport to occur. The second pathway is thought to be cytoplasmic, requiring soluble carrier proteins to move cutin monomers or small oligomers in concert with an ABC transporter and/or glycosylphosphatidyl-inositol (GPI)-anchored lipid transfer protein (LTPG) (DeBono et al., 2009). For larger oligomers or polyester domains, especially highly branched ones, different mechanisms are likely needed. Larger oligomers could be shuttled by oleophilic droplets whose genesis is similar to the budding process of seed oil bodies, and exocytosis by Golgi-mediated secretory vesicles.

After being exported out of the cytoplasm, how do hydrophobic cutin monomers pass through a predominantly hydrophilic cell wall to reach the cuticle? Possible mechanisms include binding to lipid transfer proteins (Kader, 1997) or to a polysaccharide in the cell wall. Another mechanism is unchaperoned movement of oleophilic droplets across the cell wall as

suggested by the observation of oleophilic droplets found in rice internode epidermal cells under rapid expansion (Hoffmannbenning et al., 1994).

1.1.4 **Wax biosynthesis**

Cuticular waxes predominantly consist of hydrophobic very-long-chain aliphatic compounds, such as straight-chain C25-35 alkanes and alcohols, aldehydes and fatty acids but also include cyclic compounds such as triterpenoids, sterols and flavonoids. Starting with hexadecanoic acid (C16), very-long-chain fatty acids of an even carbon number (C24 - C34) are produced by the fatty acid elongase (FAE) complex in the ER (Haslam and Kunst, 2013). These fatty acids are reduced to fatty aldehyde and primary alcohols or reduced and decarbonylated to alkanes, which can be further converted to secondary alcohols and ketones. These wax components are secreted to the cuticle in a variety of different ways. Some are Golgi-independent while others are Golgi-mediated. Waxes can also be transported from the ER directly to the plasma membrane as droplets, or exocytosis through the Golgi apparatus. ABC transporters and non-specific lipid transfer proteins might also be involved (Kunst and Samuels, 2003).

When the wax components are deposited on the surface of the cuticle, they can self-assemble into crystalline structures such as rods, tubes, or plates (Koch and Ensikat, 2008). The hydrophobicity of wax provides water repellency, and additionally the roughness created by the deposition patterns can further prevent water from adhering to the surface. Free fatty acids and alkanes in many cases accumulate in the epicuticular layer, whereas wax components, such as triterpenoids and very-long-chain aliphatic primary alcohols,

preferentially accumulate in the intracuticular layer instead. Given that wax components of similar chain lengths are typically distributed evenly between the layers, it has been proposed that partitioning occurs spontaneously due to the physicochemical properties of the wax compounds and interactions with the intracuticular polymers (Buschhaus and Jetter, 2011).

1.1.5 Cuticular polysaccharides

Polysaccharides such as cellulose, hemicelluloses and pectins have been isolated from cuticles, and they are important for the rheological properties of the cuticle (Domínguez et al., 2011; Lopez-Casado et al., 2007). The polysaccharides that are associated with the cuticle are thought to originate from the epidermal cell wall as the polysaccharide composition ratio of the cuticular layer was found to be similar to that of a primary cell wall (Guzman et al., 2014; Lopez-Casado et al., 2007).

Polysaccharides such as pectin may be excreted shortly after cell division but prior to the formation of the pro-cuticle, providing a structural framework for subsequent cuticle assembly (Guzman et al., 2014). Consistent with this notion, Fourier-transformed infrared (FTIR) spectrometric studies revealed that polysaccharides are enriched in the inner layer (cuticular layer) (Heredia-Guerrero et al., 2014), whereas in the outer layer (cuticle proper) waxes and cutin predominate. Although the cuticle proper was originally defined to be a region free of polysaccharides, cellulose and pectins have been detected in enzymatically-isolated cuticles, occasionally found just underneath the outermost epicuticular wax layer (Guzman et al., 2014).

In addition to improving the elastic strength of the plant cuticle, polysaccharides are important for the ionic exchange capacity of the cuticle with the cutin matrix itself thought to play only a minor role (Schonherr and Bukovac, 1973). An asymmetric charge gradient is established across the cuticle wherein the cuticular layer carries a net negative charge, likely ascribed to the polysaccharides while the outer wax-rich layer is mainly uncharged (Heredia and Benavente, 1991). This charge gradient is an important property that influences absorption, uptake and transport of ions and charged molecules. The presence of cellulose and pectins has been speculated to have characteristics that contribute to the bi-directional transport of water and solutes (Zwieniecki et al., 2001). Furthermore, polysaccharides are particularly important for water retention at low moisture. Water retained in the cuticle can be categorized as one of two types, either “volatile” or “embedded”. Volatile water molecules are in equilibrium with the ambient moisture and are held by one hydrogen bond with the hydroxyl groups of polysaccharides. By contrast, the embedded water molecules are held by two or three hydrogen bonds with the cutin and the polysaccharides simultaneously. This type of water cannot escape even at temperatures higher than 100°C (Heredia-Guerrero et al., 2014)

1.2 Cutin/wax biosynthetic genes in *Arabidopsis thaliana*

Many of the key Arabidopsis genes involved in cuticle formation have been identified by forward genetic screens. The majority of identified genes are involved in fatty acid and/or wax biosynthesis, modification, transport, and polymerization of components, as well as the regulation of pathways that are involved in these processes (Javelle et al., 2011). Mutant

phenotypes include postgenital organ fusion, changes in the load, quantity or composition of cutin/wax, changes in cuticle permeability and/or ultrastructure, and altered resistance to pathogens. Among those, genes associated with cutin biosynthesis will be further discussed below, and presented in the order of the cutin biosynthesis steps (Figure 1.3).

1.2.1 Long-chain fatty acyl-CoA synthetase (LACS)

The *LACS* gene family encodes enzymes required for long-chain fatty acyl-CoA formation as the first step in cutin and wax monomer synthesis. Characterization of *lacs1*, *lacs2* and *lacs1 lacs2* double mutants revealed that *LACS1* plays a role in the biosynthesis of cuticular wax and *LACS2* in cutin monomer biosynthesis. Double mutant analysis indicated that deficiencies in both cutin and wax synthesis has a compounding effect on the functional integrity of the cuticle, including altering transpiration, water-soluble molecule movement, and organ fusion (Lue et al., 2009; Schnurr et al., 2004; Tang et al., 2007b; Weng et al., 2010; Xiao et al., 2004).

Similar to the function of *LACS*, *FDH* encodes a protein related to β-keto acyl-CoA synthase (KCS) that is associated with wax and suberin biosynthesis as a part of the fatty acid elongation complex (FAE) found in the ER (Pruitt et al., 2000; Yephremov et al., 1999). Results obtained by *in situ* hybridization of mRNA revealed that *FDH* is expressed predominantly in epidermal cells, and this finding is consistent with the highly permeable cuticle that is characteristic of these mutants (Lolle et al., 1998). Additionally, the detection of *FDH* transcripts in ovules suggested its role in ovule development (Pruitt et al., 2000).

1.2.2 **Fatty acyl ω-hydroxylase (FAH), a cytochrome P450 (CYP) family protein**

For cutin synthesis, fatty acyl-CoA can be transformed to hydroxy-fatty acids by reactions catalyzed by fatty acid hydroxylases (FAH), a group of cytochrome P450 (CYP) proteins (Kandel et al., 2006; Pinot and Beisson, 2011). A few examples are CYP86A1, a fatty acid ω-hydroxylase (Benveniste et al., 1998); CYP96A15, a mid-chain alkane hydroxylase responsible for cuticular wax formation (Greer et al., 2007); and CYP86B1, putatively a very long chain fatty acid hydroxylase for polyester biosynthesis (Compagnon et al., 2009). Two better-characterized genes of the CYP86 family are *LCR* (*LACERATA*) and *ATT1* (*ABERRANT INDUCTION OF TYPE THREE 1*). The *LCR* and *ATT1* genes encode CYP86A8 and CYP86A2, respectively; both are putative monooxygenases with ω-hydroxylase activity that catalyze ω-hydroxylation of fatty acids ranging from C12 to C18:1 (Bak et al., 2011; Wellesen et al., 2001).

1.2.3 **Hydroxy fatty acyl dehydrogenase (FADH) and oxo-fatty acyl dehydrogenase (OFADH)**

The two putative step transforming hydroxy fatty acids to oxo products are catalyzed by ω-hydroxy fatty acyl dehydrogenase (HFADH), while ω-oxo fatty acyl dehydrogenase (OFADH) acts on the oxo products to produce dicarboxylic fatty acids (Pollard et al., 2008). Little is known about genes encoding proteins involved in these steps although a wound-inducible ω-hydroxy fatty acid dehydrogenase has been purified from potato (*Solanum tuberosum* L.) and is postulated to be involved in oxidation of hydroxy fatty acids in the synthesis of the suberin lamella (Agrawal and Kolattukudy, 1977; Agrawal and Kolattukudy, 1978a; Agrawal and Kolattukudy, 1978b). In Arabidopsis, five putative HFADH genes

(*AT1G12570*, *AT1G72970*, *AT5G51950*, *AT5G51930*, *AT1G14185*, *AT3G56060*) are predicted to function as FADHs in the cutin biosynthesis pathway (Plant Metabolic Network, http://www.plantcyc.org/). However, little experimental evidence validating this prediction is available. Among the five listed previously, *HOTHEAD* (*HTH*; *AT1G72970*) is the only gene that has been investigated and will be discussed later in the section dedicated to this gene.

1.2.4 **Acyltransferase and polyester synthase for polymerization**

Glycerol-3-phosphate acyltransferases (GPAT) are known for their ability to create the ester bond between fatty acids and glycerol. With glycerol acting as a "linker" between dicarboxylic fatty acids, larger branching or cross-linked cutin structures can be achieved. Nine *GPAT*-like genes (*GPAT1* to *GPAT9*) have been identified in Arabidopsis (TAIR, www.arabidopsis.org) by forward genetics. Overexpression of *GPAT4* and *GPAT8* increased the cuticular permeability and resulted in a more structurally diffuse cuticle as indicated by TEM results (Li et al., 2007), even though cutin monomer load became elevated. These results point to the importance of GPATs for polymerization of cutin monomers.

Another two known acyltransferases required for incorporating monomers into a polymeric structure are encoded by *DEFECTIVE IN CUTICULAR RIDGES* (*DCR*) and *BODYGUARD 1* (*BDG1*). Mutant *dcr* plants manifest postgenital organ fusion as well as significant reduction of a major cutin monomer (Panikashvili et al., 2009). These mutant plants also exhibited more susceptibility to abiotic stress such as water deprivation due to a defective cuticle that is unable to serve its function as a protective barrier. The BDG1 protein is localized to the extracellular space of the cell wall and has been proposed as an

extracellular polyester synthase (PS) that produces cutin polymers (Kurdyukov et al., 2006a). *bdg1* mutant plants exhibit increased cuticle permeability and share phenotypes reminiscent of transgenic Arabidopsis expressing an extracellular fungal cutinase (Sieber et al., 2000).

1.2.5 **ATP binding cassette (ABC) transporter**

Arabidopsis plants harbouring a mutation in the *ABCG11*/*WBC11* (*ATP BINDING CASSETTE G11/ WHITE-BROWN COMPLEX HOMOLOG PROTEIN 11*; also known as *PERMEABLE LEAVES 1*) show a reduction of cutin load on the leaf surface, stunted growth, and leaf fusions. Studies using the recombinant fluorescent fusion protein, YFP-WBC11, showed that this protein is localized to the plasma membrane while T-DNA knock-out mutants exhibited lipidic inclusion bodies in the cytoplasm of epidermal cells (Bird et al., 2007). Similar inclusions were also observed in mutants that exhibited reduced stem cuticular wax loads such as *eceriferum5/abcg12* (Pighin et al., 2004). The expression of *ABCG13*, on the other hand, is restricted mainly to petals and carpels, and its mutant displayed significant reduction in flower cutin monomers and inter-organ postgenital fusion (Panikashvili et al., 2011). According to these findings, *ABCG11, 12* and *13* appear to encode proteins that secrete the building blocks of cutin and waxes (Bessire et al., 2011; Panikashvili et al., 2007; Panikashvili et al., 2011).

LTPG1 and *LTPG2* encodes proteins categorized to the class of glycosylphosphatidylinositol (GPI)-anchored lipid transfer proteins (LTPs). *LTPG1* is expressed in the epidermis and is primarily localized in the plasma membrane (transmembrane protein) but is also present in the extracellular matrix (DeBono et al., 2009).

It has the highest expression in regions of rapid expansion, such as inflorescence stems. The fact that LTPG1 is capable of binding to lipids was experimentally determined by incubating *Escherichia coli*-expressed LTPG1 with the fluorescent lipophilic probe 2-p-toluidinonaphthalene-6-sulfonate (TNS). Mutant *ltpg1* plants showed a great reduction (> 50%) in the C29 alkane, a major component of cuticular waxes of the stems and siliques, a defect that can be rescued by native promoter–driven *LTPG1* expression (DeBono et al., 2009; Lee et al., 2009b). Based on these properties, LTPGs are thought to be carriers of cutin and wax constituents to the plant surface.

1.2.6 **Regulators of cutin/wax biosynthesis**

The Arabidopsis *SHINE1* (*SHN1*)/*WAX INDUCER 1* (*WIN1*) was first associated with the cuticle for the glossy appearance of the leaf surface and increase cuticle permeability of the mutant plants (Aharoni et al., 2004). *SHN1* overexpression results in an increase in cutin and wax production in vegetative and reproductive organs, and such changes are preceded by induction of several genes known or likely to be involved in cutin biosynthesis. (Broun et al., 2004; Kannangara et al., 2007). It has been shown that at least one of such cutin pathway genes is *LACS2* as its promoter sequence is a direct target of SHN1 (Kannangara et al., 2007). Interestingly, the expression of *SHN1* is under control by another group of transcription factors, including MYB106 and MYB16, known regulators of epidermal cell differentiation (Jakoby et al., 2008).

CER3 encodes a transmembrane protein that is implicated in wax alkane synthesis (Bernard et al., 2012). The *cer7* mutant exhibits reduced cuticular wax accumulation, a

finding consistent with reverse transcription polymerase chain reaction (RT-PCR) studies that show considerably lower *CER3* expression levels in *cer7* plants. It was proposed that *CER7* encodes a putative 3'-5' exoribonuclease that acts by degrading an mRNA species encoding a negative regulator of *CER3* (Hooker et al., 2007).

Furthermore, Voisin et al. (2009) proposed that the cuticle mutants can alleviate the functional disorder of the cuticle by reinforcing different cell integrity pathways. Using an *in silico* screening method, the author identified a gene *SERRATE* (*SE*) that encodes a protein involved in RNA-processing. It was demonstrated that the *se lcr* and *se bdg* double mutation eradicated severe leaf deformations as well as the organ fusions that are typical of *lcr* and *bdg*, suggesting that plants are capable of controlling the integrity of the cuticle by regulating small-RNA signaling.

1.3 Arabidopsis seed coat development

1.3.1 Seed coat structure

An Arabidopsis seed consists of three main components, the embryo, the endosperm, and the seed coat. The seed coat, which is the outer most layer, constitutes about 20% of mature dry seed weight (Li et al., 2006). Distinct from the embryo and endosperm whose genetic makeup is a combination of both female and male counterparts, the seed coat is maternally derived and arises from the ovule integuments. The seed coat controls endosperm and embryo expansion during seed maturation, and for fully developed seeds, it maintains

dormancy and serves as a protective barrier shielding the embryo from adverse conditions (Haughn and Chaudhury, 2005).

During early female gametophyte development, the megaspore mother cell resides within the ovule and is surrounded by the outer and inner integuments, both of which are of epidermal origin (Beeckman et al., 2000). The outer integument (oi) consists of an inner (oi1) and an outer layer (oi2). The inner layer is defined as the adaxial layer to the ovule axis and the outer layer is the abaxial layer (Truernit and Haseloff, 2008). Similarly, the inner integument (ii) also has an inner (ii1, also known as the endothelium) and an outer layer (ii2) but with an extra internal layer (ii') between ii1 and ii2 (Figure 1.5).

In the outer integument, vacuoles appear in the cells at the onset of embryogenesis while amyloplasts start to form at the globular stage. At the torpedo stage, mucilage production is initiated in the outer most oi2 layer. These cells gradually mature into specialized cells designed for seed rupture and mucilage release. When the embryo starts to expand (walking stick stage), the enlargement of the mucilage compartments pushes the starch grain-containing amyloplast to form a small column, i.e. columella, in the center of the oi2 cell. While the embryo continues to enlarge, oi1, ii1' and ii1 layers compress against the enlarged oi2. At the desiccation stage, ii1' and ii2 (and sometime also oi1) collapse to form the brown pigment layer (bpl) that gives the brown colour of mature seeds (Beeckman et al., 2000; Creff et al., 2015; Windsor et al., 2000).

1.3.2 **Cuticle layers in the seed coat**

Cutin and suberin monomers have been identified in the seed coat of plants such as *Arabidopsis thaliana*, *Brassica napus* and *Glycine max* (Espelie et al., 1979; Molina et al., 2006; Molina et al., 2008; Shao et al., 2007). Studies of cuticle mutants have demonstrated that deposition of these fatty acid monomers and their polymers are essential for the seed coat to attain proper permeability (Beisson et al., 2007; Compagnon et al., 2009; De Giorgi et al., 2015). Recently in a study of the *bdg1* mutant, De Giorgi et al. (2015) demonstrated that this mutation also leads to an increase in permeability of an endosperm-associated cuticle layer, i.e. the cuticle on the inner surface of ii1. This layer was first reported in a histological study by Beeckman et al. (2000) who observed an electron-dense layer and considered it as the original cuticle of the inner integument. In addition to this endosperm-associated cuticle layer, a layer rich in cutin-like material situated in the thickened inner periclinal cell wall (also called wall 3) between the inner and outer integuments has also been described (Beeckman et al., 2000; Creff et al., 2015). Wall 3 represents a fusion zone at the boundary of the two integuments. By examining wall 3 at later developmental stages, Creff et al. (2015) revealed that wall 3 material was predominantly laid down by oi1, and that wall 3 thickening was regulated by oi1's response to mechanical stress. It was further postulated that oi1 senses the mechanical pressure on the seed coat due to endosperm expansion, and oi1 thickens its outer cell wall (wall 3) to restrict seed growth, in essence serving as a corset around the developing seed. Whether or not wall 3 also regulates the seed coat permeability is unknown.

Given the epidermal origin of integuments and the existence of cuticle-like structures in the seed coat, it is not surprising that some genes essential for shoot epidermal cuticle formation are also important for seed coat development. In addition to the aforementioned *bdg1* mutant that showed inability to restrict toluidine blue penetration into the endosperm as a result of higher permeability (De Giorgi et al., 2015), the *dcr* mutant also has been shown to have a more permeable seed coat. In addition, *dcr* mutant seeds were often deformed, showed evidence of seed fusion, and had limited mucilage release following seed imbibition (Panikashvili et al., 2009). Furthermore, many *ltpg* mutants discussed previously have decreased levels of ω-hydroxy fatty acids in seed coats and permit tetrazolium salt uptake into seeds. These observations demonstrate that some genes essential for cuticle formation are also important for seed coat development.

1.4 **Plant stress responses**

Plants are sessile organisms that are exposed to a diversity of environmental challenges including water stress, soil salinity, temperature fluctuations, freezing, exposure to toxic metals, variable light intensity and mechanical wounding. In addition to abiotic stresses, plants also face the hazard of pathogens (including bacteria, fungi and viruses) and attack by herbivores or pests. Thus, plants have had to evolve mechanisms for sensing potentially harmful conditions to improve their chances of survival (Suzuki et al., 2014).

1.4.1 **Stress responses associated with an ER-derived structure, the ER body**

About one-third of all proteins are assembled in the ER (Deng et al., 2013). The capacity of the ER to fold, modify, assemble and route proteins, however, can be compromised under conditions of stress. Oxidative stress caused by reactive oxygen species (ROS), for example, can lead to protein misfolding (Malhotra and Kaufman, 2007) and can trigger the unfolded protein response (UPR). The UPR enables plants to sense and respond to adverse environmental conditions. Accumulation of unfolded or misfolded proteins activates the UPR pathway which leads to proteolysis of undesired proteins, changes in mRNA splicing, or cell death (Deng et al., 2013; Ruberti and Brandizzi, 2014).

The ER forms highly organized network structures composed of tubules and cisternae. Environmental stresses can induce ER-derived organelles including protein bodies (PBs; 1-2 μm), precursor-accumulating vesicles (PACs; 0.3-0.5 μm), KDEL-tailed protease-accumulating vesicles (KVs; 0.2-0.5 μm), ricinosomes (0.2-0.5 μm), and coat protein complex (COP) II vesicles (0.05-0.1 μm); many of these structures function as repositories of proteases (Hara-Nishimura et al., 1998; Hara-Nishimura and Matsushima, 2003; Nakano et al., 2014). Among members of the *Brassica* family, a type of ER-derived structure called the ER body has been shown to be distinct from the aforementioned ER-derived bodies; ER bodies not only are longer and larger (5∼10 μm long and ∼1 μm wide) but also accumulate different kinds of proteins, mainly β-glucosidases (Matsushima et al., 2003b; Sherameti et al., 2008). The ER body was first discovered in radish root epidermal and cortical cells and was initially thought to be dilated ER cisternae (Bonnett and Newcomb, 1965). Decades after, these large dilated ER domains were also found in Arabidopsis but were initially

described as “mystery organelles” (Gunning, 1998). These so-called mystery organelles later were confirmed to be ER-derived using the ER-targeted green fluorescent protein (Hawes et al., 2001). Hayashi et al. (2001) proposed the term ‘ER body’ be used to describe these distinctive ER-derived structures.

ER bodies are constitutively present in epidermal cells of cotyledons, hypocotyls and roots of young Arabidopsis seedlings. Although the density and distribution of these “constitutive ER bodies” varies with tissue type and developmental stage, these constitutive ER bodies are normally absent in rosette leaves (Hayashi et al., 2001). Nonetheless, it was discovered that methyl jasmonate (MeJA) treatment and wounding could induce ER bodies (“induced ER bodies”) in rosette leaves, and in MeJA-insensitive *coronatine*-insensitive *1* (*coi1*) mutant plants, ER body induction was suppressed (Matsushima et al., 2002). This result is consistent with the notion that ER bodies are associated with plant stress responses since MeJA is a plant hormone involved in plant defense mechanisms. Furthermore, ER body formation appears to be a systemic response as when one of the two cotyledons was wounded, both the damaged and the intact cotyledon had increased numbers of ER bodies (Ogasawara et al., 2009).

Electron microscopy analysis revealed a relatively high electron density in the ER body lumen, suggesting that the ER body contains a large amount of proteins (Nakano et al., 2014). Studies of a mutant with no constitutive ER bodies, the *nai1* mutant, revealed that a myrosinase called PYK10 is a major protein component of ER bodies in *A. thaliana* (Matsushima et al., 2003b). Myrosinase, a type of β-glucosidase, is known for its ability to

catalyze the hydrolysis of glucosinolates (Rask et al., 2000). The myrosinase activity of PYK10 has been directly demonstrated by hydrolysis of the fluorogenic substrate 4-methylumbelliferyl β-glucopyranoside (Matsushima et al., 2003a). Based on analyses of mutants that lack ER bodies or form abnormal ones, a model for the formation of ER bodies in *A. thaliana* seedlings has been proposed. In this model, *NAI1* encodes a basic-helix-loop-helix type putative transcription factor that regulates the expression of four key genes for ER body formation: *PYK10*, *NAI2*, *MEB1* (*MEMBRANE OF ER BODY 1*), and *MEB2*. In the ER, PYK10 and NAI2 first interact to initiate the budding process, and then NAI2 forms a complex with MEB1 and MEB2 that are later integrated to the ER body-specific membrane (Matsushima et al., 2004; Nakano et al., 2014) (Figure 1.6).

1.4.2 Glucosinolate pathways as defense mechanisms

Myrosinase-catalyzed hydrolysis of glucosinolates gives rise to products that are components of a defense mechanism against herbivores and fungal infection (Hopkins et al., 2009; MacLeod and Rossiter, 1986; Sherameti et al., 2008). The reaction can give rise to a variety of derivatives, depending on reaction conditions such as cofactors, pH and facilitating proteins. Isothiocyanates are the most common product at neutral pH, whereas nitrile products are favoured when Fe^{2+} concentration or acidity (<pH 5.0) is elevated. Among derivatives, hydroxynitrile products can be catalyzed by hydroxynitrile lyase to form hydrogen cyanide (HCN), and this type of cyanogenesis is common in higher plants, particularly *Brassicaceae* plants (Figure 1.7) (Brabban and Edwards, 1995; Kissen and Bones, 2009; Nakano et al., 2014).

The substrates and enzymes of the glucosinolate-myrosinase system are sequestered in separated subcellular compartments or different tissues preventing undesired production of toxic compounds. For example, glucosinolates are stored in vacuoles while myrosinase resides in the cytosol or other organelles. Cell damage can bring these compounds into direct contact to yield hydrolytic products, but similar results can also be achieved by translocation of enzymes into vacuoles (Grob and Matile, 1979; Nakano et al., 2014; Poulton, 1990). For instance, three β-glucosidases that contain ER retention signals at their respective C termini (At1g52400, REEL; At1g66270, RDEL; At3g09260, KDEL) have been identified in the vacuole (Carter et al., 2004), indicating vacuolar sorting of β-glucosidases originating from the ER. Moreover, Hayashi et al. (2001) demonstrated fusion of ER bodies to each other and to the vacuole in the hypocotyl epidermal tissues under stress, an example of a plant stress response via Golgi-independent pathway of protein targeting to vacuoles (Xiang et al., 2013). In addition, ER body fusion with the plasma membrane has recently been proposed (Nakano et al., 2014).

1.4.3 **Genome instability induced by stress**

Genetic variation is fundamental to the survival of a species, and it allows adaptation to changing environments. For crop plants the reduction of genetic variability, for example, is now a pressing issue for plant breeders around the world because of the need for intensifying food production and the predicted negative impact of climate change on crop productivity. There are many known mechanisms that drive genetic variation and include homologous recombination, polyploidy, DNA mutation, gene duplication, transposable element movement, chromosomal rearrangements, and epigenetics (Feng et al., 2010; Gbadegesin,

2012; Kaeppler and Phillips, 1993; Kovalchuk et al., 2003; Lippman et al., 2003; Wheeler, 2013; Yao and Kovalchuk, 2011).

Biotic or abiotic environmental factors such as pathogen attack and changes in growth conditions can trigger genetic instability and can in turn, result in genetic variation that provides plants with greater adaptive versatility (Boyko et al., 2005; Boyko et al., 2006; Boyko et al., 2010; Chen et al., 2005; Choi and Sano, 2007; Kovalchuk et al., 2003; Lucht et al., 2002; Madlung and Comai, 2004). In addition to increasing the fitness of the individual plant that hosts the novel genetic variation, there is evidence that in some cases environment-induced genome instability can persist and is inheritable by its progeny and future generations (Agrawal et al., 1999; Boyko et al., 2007; Galloway and Etterson, 2007).

1.4.4 **Environment-induced genetic instability**

Flax (*Linum usitatissimum*) has a genome that can be induced to undergo changes in response to specific growth conditions, with the most notable occurrences in the variety Stormont Cirrus, also known as the “plastic” line. Under inducing growth conditions, heritable genomic changes in the sequences encoding the ribosomal RNAs and particular repetitive sequence families have been reported, and most recently have come to include the acquisition of a relatively large insertion sequence (Chen et al., 2009; Chen et al., 2005; Cullis et al., 2004; Schneeberger and Cullis, 1991). The acquisition of this single copy insertion termed Linum Insertion Sequence 1 (LIS-1) comprises a 5.7 kilobase (kb) DNA fragment, and it is identical to a sequence found in other flax varieties. As the sequence of LIS-1 fragment was not detected in the progenitor plant, it was proposed that this novel insertion arose through a

series of reproducible, targeted and complex rearrangements or insertion events that occur naturally (Chen et al., 2005). The environmental condition required for the appearance of LIS-1 is highly specific. Only plants treated with either solely nitrogen or no fertilizers hosted the LIS-1 insertion, whereas plants treated with three-component fertilizers (with nitrogen, potassium and phosphate) showed no such insertion. Furthermore, the LIS-1 insertion required the appropriate inducing conditions to be maintained, otherwise it was subsequently lost in the offspring (Chen et al., 2009). These results agree with the postulation that the mechanism for the LIS-1 insertion event is specifically regulated and is not part of normal developmental processes in flax.

The phenomenon of naturally-occurring environment-induced genome instability has been adopted as a strategy in plant breeding to produce novel genetic variation in highly homogeneous agriculture crops such as soybean. For example, Fasoula and Boerma (2005) discovered that growth condition such as ultra-low density (one-row plots with a row spacing of 0.76 m and a row length of 3.5 m) was effective in producing significant variation of seed protein and oil for three soybean cultivars. Using simple sequence repeat (SSR) markers as an indication for genetic variation, Yates et al. (2012) provided evidence that some of these induced phenotypic variations were likely due to genetic variation, rather than epigenetics or biological regulation.

1.4.5 **Somatic variation and tissue culture-induced somatic mosaicism**

Unlike animals, plants consist of a series of repeating units (modules) that typically have identical genetic makeup. Modules that differ genetically from other modules can naturally

occur via somatic mutation, allowing the introduction of genetic variation into the gene pool of an individual without sexual reproduction. These genetically different units offer unique adaptive advantages as the increased diversity contributes to the fitness of the entire individual plant. Furthermore, in some cases, mutations arising somatically have a greater probability of being transmitted than mutations that arise in the gametes (Whitham and Slobodchikoff, 1981) because germ line cells are derived from somatic tissues that arise late in the developmental stage of the plant (Satina and Blakeslee, 1941; Youngson and Whitelaw, 2008).

Somatic sectoring can be induced by biotic and abiotic stresses such as ionizing radiation, heavy metals, temperature and water (Yao and Kovalchuk, 2011). For instance, the occurrence of somatic sectors increased by a factor up to 56 when Arabidopsis plants were exposed to DNA-damaging agents such as UV-C, X-ray and methyl methanesulfonate (Kovalchuk et al., 2000). Mechanistically, somatic variation can arise from homologous recombination, microsatellite instability and DNA rearrangement, with somatic homologous recombination being the most common mechanism (Boyko et al., 2006; Boyko et al., 2010; Boyko and Kovalchuk, 2011; Kovalchuk et al., 2003; Lucht et al., 2002).

Micropropagation techniques, such as tissue culture, are extensively used for maintaining highly desirable traits and to mass-produce certain economically important crop plants. Nonetheless, genetic changes can occur when plants are regenerated from dedifferentiated callus produced by tissue cultures (Phillips et al., 1994). These sporadic occurrences of somatic variation pose a great challenge to the commercialization of elite

clones. On the other hand, the resultant novel genotypes can also be useful in crop improvement, especially for highly homogeneous varieties (Jain, 2001).

Tissue culture-induced somatic variation has been observed in a number of crop plants, including banana (*Musa* spp.), soybean (*Glycine max*), and rice (*Oryza sativa* ssp. *japonica*). The inequality in susceptibility to genetic variation among different genomic regions was demonstrated in banana whose genome contains one particularly labile portion especially susceptible to higher rearrangement than other portions of the genome during tissue culture (Oh et al., 2007). Using restriction fragment length polymorphism (RFLP) markers, Roth et al. (1989) reported that root tissue obtained from individual soybean plants developed novel RFLP allelic differences at various loci following *in vitro* culturing. What was more interesting is that these newly arisen alleles were almost always the same as ones previously found in other varieties of cultivated soybean. Although the genetic mechanisms driving such somaclonal variation are not well understood, Roth et al. (1989) postulated that the reappearance of these specific alleles resulted from precisely controlled recombination events. In rice tissue culture, Gao et al. (2011) observed a gain-of-function mutation that gave rise to the inheritable, dominant purple sheath trait. Regenerated plants with this trait harbour a 34-bp insertion in a gene encoding a putative transcription factor for anthocyanin pigmentation. Interestingly, the tissue culture-induced insertion-containing allele ("functional allele") had been previously identified in another rice variety, similar to the cases of flax and soybean discussed previously (Chen et al., 2005; Roth et al., 1989). Hence, insertion of an extra-genomic sequence was one hypothesized mechanism for this mutation, aside from homologous recombination and transposition events.

1.4.6 **Pathogen-induced somatic mosaicism**

Pathogens apply one of the strongest selective pressures on plant populations and can bring about genetic variation in plants that is proposed as a strategy to improve the population fitness (Karasov et al., 2014). It was found that somatic recombination in Arabidopsis was elevated upon the infection of the water mould pathogen *Peronospora parasitica*, and the same effect was observed when plants were exposed to 2,6-dichloroisonicotinic acid and benzothiadiazole, chemicals that are known to trigger plant pathogen-defense mechanisms (Lucht et al., 2002). These results suggest that pathogen-induced genome instability might be activated via defense signaling pathways.

Pathogens, such as viruses, can induce a systemic signal that leads to an increase in genetic variation. Such a signal has been demonstrated in *Nicotiana tabacum* (tobacco) plants infected with tobacco mosaic virus. When upper, virus-free leaves from an infected plant (the 'signal-carrying' leaves) were grafted onto healthy plants (Kovalchuk et al., 2003), the uninfected plants serving as scions experienced a 2.3 times increase in somatic homologous recombination. This finding suggests that the recombination-inducing signal can be transmitted between different tissues of an individual plant and also between plants through grafting, independent of the presence of the virus.

Earlier research has shown that elevated genetic variation caused by this systemic signal is heritable. Experiments conducted by Brakke (1984) demonstrated that plants infected with barley stripe mosaic viruses were able to give rise to an increased number of mutations in non-infected progeny. Furthermore, Boyko et al. (2007) found that the progeny of tobacco

plants to which the signal was transmitted by grafting also exhibited an increased frequency of homologous recombination. Moreover, in the progeny of plants that received this systemic signal, their genomes were found to be considerably hypermethylated. However, substantial hypomethylation was observed in several specific leucine-rich repeat (LRR)-containing loci that are associated with pathogen recognition (Diévart and Clark, 2004). Since methylation of DNA and/or histones is thought to stabilize the genome, the loss of methyl groups may increase the susceptibility of the genome to rearrangement and mobilization of elements. Hence, Boyko et al. (2007) postulated that genome-wide hypermethylation of the progeny is part of a general protection mechanism incited by the stress signal, whereas locus-specific hypomethylation, such as that at the LRR-containing loci, is a consequence of a higher frequency of rearrangements.

1.5 The *HOTHEAD* gene

1.5.1 The *HOTHEAD* gene and the hypothetical protein model

The *HOTHEAD* (*HTH*) gene maps to chromosome 1 (locus: AT1G72970) and is also known as *EMBRYO SAC DEVELOPMENT ARREST 17* (*EDA17*) or *ADHESION OF CALYX EDGES* (*ACE*) (Araki et al., 1998; Krolikowski et al., 2003; Lolle et al., 1998; Pagnussat et al., 2005). In the Columbia ecoptype background, its coding region is 2834 basepairs (bp) in length, and the putative 5'-upstream promoter region from the ATG start codon to the stop codon of the neighboring upstream gene is 2009 bp long (Figure 1.8). Based on transcript data, the gene consists of six exons encoding a putative 594 amino acid (aa) protein that is 65.3 kilodaltons (kDa) in size with the isoelectric point of 10.2 (based on protein coding gene

model AT1G72970.1; www.arabidopsis.org). The transcript anlyses also indicate putative alternative splicing at the fifth exon, giving rise to a protein isoform that is 567 aa-long and 62.2 kDa in size (based on protein coding gene model AT1G72970.2).

Over the past decade numerous *hth* mutant alleles have been generated by ethyl methanesulfonate (EMS), T-DNA insertion, and transposon mutagenesis. Among these, *hth-1* to *hth-11* were identified in the Landsberg *erecta* background (except *hth-9* in Wassilewskija), and they all harbour single nucleotide mutations (Figure 1.8) (Krolikowski et al., 2003; Lolle et al., 1998). In the *hth-1* mutant, for example, the mutation introduces a stop codon, while in the *hth-9* mutant the point mutation is predicted to alter a splice junction sequence. Thus, *hth-1* and *hth-9* likely encode truncated polypeptides. The other nine mutant alleles are predicted to encode proteins with single amino acid substitutions that may change the folding or catalytic properties of the HTH protein. The *hth-12* mutant allele was generated by *En/Spm* transposon insertion in the 5'-untranslated region (UTR) of *HTH* in the Columbia ecotype background (Kurdyukov et al., 2006b). *eda17* was generated by *Ac/Ds* transposon mutagenesis with the insertion site in the first intron (personal communication).

1.5.2 The *hth* mutant phenotype

Mutations in the *HTH* locus result in promiscuous interactions between contacting epidermal cells that leads to fusion between organs (Lolle et al, 1998). The *hth* postgenital fusion phenotype is mostly restricted to the floral tissue but can occasionally be observed on rosette leaves. Similar to other organ fusion mutants such as *deadhead*, *thunderhead*, and *fiddlehead*, the floral organ fusion of *hth* alters the inflorescence configuration and blocks the emergence

of petals and anthers with the pistil generally protruding out from individual floral buds (Figure 1.10) (Lolle et al., 1997; Lolle et al., 1998).

The organ fusion phenotype was previously associated with the increase in cuticle permeability. Lolle et al. (1998) described the isolation and characterization of 29 independently derived mutations that led to organ fusion in Arabidopsis. Using complementation analyses, nine putative genes, including *HTH*, were identified. These mutants most frequently showed interorgan fusions within the flowers with a great range of severity. Occasionally fusion events between vegetative tissues (e.g., *fiddlehead* and *thunderhead*) were observed. Results of the chlorophyll extraction assay revealed that the cuticle of most mutants, including *hth*, were more permeable, suggesting the organ fusion phenotype was caused by a defective cuticle.

In addition to organ fusion, *hth* mutants support pollen adhesion, germination and growth on epidermal cells other than stigmatic papillary cells. Pollen germination normally occurs only on the stigmatic surface of a receptive flower, and this response requires specific recognition interactions between the pollen grain and the stigma papillary cells. On the stigma, the cuticle functions as a selective and semi-permeable diffusion barrier and thereby acts in the identification of compatible pollen grains permitting germination and growth of pollen tubes (Hulskamp et al., 1995). For that reason, on vegetative tissues pollen grains normally do not hydrate or germinate. Mutant *hth* plants, however, allow the adhesion and germination of pollen grains on vegetative tissues, a phenomenon observed in at least five other mutants that display floral organ fusion and elevated cuticle permeability (Krolikowski

et al., 2003; Lolle and Cheung, 1993; Lolle et al., 1998; Lolle and Pruitt, 1999; Sieber et al., 2000). Hence, the ectopic pollen germination on vegetative tissues of the *hth* mutant is likely a consequence of altered cuticle permeability.

In addition to floral organ fusion, seed set is greatly reduced in most *hth* mutant lines, likely caused by the disfiguration of reproductive organs and ovule defects. Ovule abnormalities have been documented in *hth-8*, *hth-10*, and *eda17* mutants (Lolle et al., 1998; Pagnussat et al., 2005). Unlike other *hth* mutants, *eda17* was identified in a large-scale mutant screen aimed at identifying *A. thaliana* plants with defects in female gametophyte development. This mutant manifests abnormal embryo sac development, arresting at the two-nuclear stage of gametophyte development. Collectively, these findings suggest that the HTH plays some role, directly or indirectly, in embryo sac development at the early mitotic phase in addition to serving important functions in epidermal development post-embryonically.

1.5.3 HTH protein localization and function

HTH expression has been studied previously by two research groups using indirect methods that included the analysis of transgenic plants expressing promoter-driven reporter constructs, *in situ* mRNA hybridization, and RT-PCR assays (Krolikowski et al., 2003; Kurdyukov et al., 2006b). Tissue-specific expression was firstly demonstrated by Krolikowski et al. (2003) who used RT-PCR and *in situ* mRNA hybridization to show that *HTH* mRNA was expressed in all organs tested including the leaf, root, inflorescence and silique. This finding is consistent with results using promoter-driven reporter constructs (HTH_{pro}*:GUS* and HTH_{pro}*:GFP*). In four-leaf seedlings, GUS expression was detected in a

region surrounding the shoot apical meristem, in emerging leaves and lateral root initials. Additionally, GFP expression was detected in all floral organs including the sepal, petal, anther, ovule, and ovary septum (Kurdyukov et al., 2006b).

Cell layer-specific HTH protein localization extrapolated from results generated by these two research groups are, however, at variance. Using *in situ* mRNA hybridization, Krolikowski et al. (2003) showed uniform distribution of *HTH* mRNA in both the epidermis and subepidermal cells of floral tissues, whereas Kurdyukov et al. (2006b) showed that, for a young apex and 4-week old stem, *HTH* mRNA was detected exclusively in the epidermal cells. Epidermis-specific expression of *HTH* was also observed in the anther, pedicle, and ovary wall of HTH_{pro}*:GFP* plants, which also showed expression in individual ovules, specifically in the integument and embryo sac (Kurdyukov et al., 2006b).

Although possible functions of HTH have been proposed based on its gene expression pattern and properties of the mutant cuticle, it is unclear how the mutations at the *HTH* gene contribute to the mutant phenotype. Based on the elevated cuticle permeability, pollen germination on vegetative tissue, and ubiquitous presence of the mRNA, Krolikowski et al. (2003) suggested that *HTH* encodes a product that is involved in a fundamental metabolic process required for cell function as well as cuticle formation. Alternatively, Kurdyukov et al. (2006b) proposed that HTH serves a catalytic function in cutin monomer biosynthesis based on the epidermal expression of *HTH* and fatty acid composition of the mutant *hth-12* cuticle. The fatty acid profile of the cuticle from the *hth-12* mutant (transposon insertion in the 5’ upstream region) demonstrated that the mutant had lower than normal levels of α,ω-

dicarboxylic fatty acids and elevated levels of ω-hydroxy fatty acids. This deviation in cutin composition led to the hypothesis that HTH is a ω-hydroxy fatty acyl dehydrogenase (HFADH) that oxidizes long chain ω-hydroxy fatty acids to ω-oxo products, precursors of the cutin monomer α,ω-dicarboxylic fatty acids (Table 1.1 and Figure 1.11). Therefore, the disruption of HTH function might alter the dicarboxylic acid ratio to other monomers and cause perturbation in cuticular polyester structures and, consequently, affect cuticle permeability. Assigning HFADH function to HTH, however, provides no obvious explanation for the ovule abnormalities observed in *hth*-4, *hth-8,* and *eda17* mutants (Lolle et al., 1998; Pagnussat et al., 2005).

The HTH protein shares sequence similarity with long-chain ω-fatty alcohol dehydrogenases from *Candida* species (Kurdyukov et al., 2006b) but sequence analyses also suggest that the HTH protein belongs to the glucose-methanol-choline (GMC) oxidoreductase family and may function as a mandelonitrile lyase (MDL) (Krolikowski et al., 2003). MDL is a hydroxynitrile lyase that catalyzes hydroxynitriles to hydrogen cyanide and aldehydes or ketones (Figure 1.12) (Sharma et al., 2005; Yemm and Poulton, 1986). The substrate of this cyanogenesis reaction, hydroxynitrile, can derive from glucosinolate hydrolysis, a plant defense compound discussed earlier that gives rise to a variety of derivatives functioning as a pest deterrent. Were HTH to function as an MDL, this would lend support to the idea that HTH may be involved in plant defense pathways.

1.5.4 **Genome instability in *hth* mutants**

One of the most unusual phenomena documented in *hth* mutants is the non-Mendelian inheritance of wildtype *HTH* alleles. When self-fertilized, homozygous *hth* mutants gave rise to wildtype progeny that were genotypically heterozygous (*HTH/hth*) at a frequency up to 10% (Figure 1.13) (Lolle et al., 2005). The wildtype progeny appeared to acquire novel DNA sequences that did not exist in the parents, but rather existed in an earlier ancestor. Although such individuals could have arisen from cross-pollination with neighbouring wildtype plants, two lines of evidence demonstrated that *hth*/*hth* homozygous mutant plants were capable of producing wildtype (*HTH*) gametes. First, it was shown that homozygous wildtype embryos could be isolated from homozygous mutant plants, and second it was determined that mutant plants could produce wildtype pollen. Although the source from which these wildtype alleles arose was not clear, these authors proposed the existence of an extra-genomic RNA cache that harboured these alleles (Lolle et al., 2005).

Extra-genomic RNA-based heredity has been documented in several other organisms. In *Caenorhabditis elegans*, for example, RNA has been shown to be able to drive genomic sequence changes (Fire et al., 1998). Experiments where RNA interference (RNAi) was used have shown that interference was evident not only in the individuals injected with the exogenous double stranded RNA but also in their progeny and the effect persisted for several generations (Chandler et al., 2000; Fire et al., 1998; Rassoulzadegan et al., 2006). In mice (*Mus musculus*), epigenetic inheritance associated with the zygotic transfer of exogenous RNA molecules has been reported (Rassoulzadegan et al., 2006), while in *Oxytricha trifallix*, injected RNA was shown to drive genome rearrangement (Nowacki et al., 2008).

Evidence for the presence of cached maternal RNAs that guide rearrangement of genomic sequence in subsequent generations was demonstrated in ciliates by Nowacki et al. (2008). They showed that exogenous RNA injected directly into cells could reprogram genome rearrangements and direct specific DNA sequence changes. Moreover, the influence of injected templates extended to various alleles of the same gene despite template mismatches. RNA has also been shown to serve as a template for DNA synthesis during repair of a chromosome double-strand break *in vivo* (Storici et al., 2007).

Together with the aforementioned genome instability induced by stress, pathogens and varying growth conditions, a view of genomes is emerging that sees genomes as much more fluid and much more responsive to extrinsic factors than previously thought. In plants, this fluidity may be even more pronounced wherein genome structure and function may be dynamically influenced both on an acute, as well as, a multi-generational scale by the environment. It may be that the genome changes seen in the *hth* mutants, in effect, reveal an inherent and completely novel mechanism that drives a form of selective genetic variation. In considering the type of instability manifested by Arabidopsis *hth* mutant plants, however, it is clear that there remain many more questions than answers.

1.6 Experimental objectives

This thesis aims to investigate genetic instability of molecular markers in the *hth* background, examine HTH protein localization and investigate previously proposed HTH protein functions. In **Chapter 2**, we used molecular approaches to examine the outcrossing

frequency of mutant *hth* plants. Expanding from previous studies, non-Mendelian inheritance was investigated using insertion/deletion molecular markers, and the restored novel DNA sequence in the progeny was compared to its parental ancestor. Furthermore, we also determined whether the observed genetic discordance between the parent and offspring was due to somatic sectoring.

Arabidopsis hothead (*hth*) mutants typically exhibit floral organ fusion, and genetic analyses have shown that the *HTH* gene is important for proper cuticle function. In **Chapter 3**, I examined *hth* mutant phenotypes in terms of floral fusion and cuticle permeability. To investigate HTH protein localization pattern in tissue and cells, I developed transgenic plants harbouring fluorescent protein tagged recombinant HTH protein and tested whether the recombinant protein complements the mutant phenotype. To examine the possible association with stress, I measured *HTH* expression changes upon methyl jasmonate treatment.

To date, no direct biochemical evidence has been published to address the question of HTH protein function although two different enzymatic activities have been proposed, one being a mandelonitrile lyase and another being an alcohol dehydrogenase (Krolikowski et al., 2003; Kurdyukov et al., 2006b). These two possible catalytic functions were investigated in **Chapter 4** using enzymatic assays and bioinformatics. *In vitro* enzymatic assays were conducted using recombinant HTH protein that was expressed and purified in a prokaryotic system. In addition, I conducted sequence, phylogenetic, and coexpression analyses of the HTH protein as well as constructed its candidate tertiary structures. Based on these results,

putative catalytic sites of HTH were identified and compared to those of related mandelonitrile lyases and alcohol dehydrogenases.

Figure 1.1 Schematic representation of a leaf cuticle. Cutin covers the outer cell wall of epidermal cells. The cuticular layer contains cutin, cell wall carbohydrates and waxes. The cuticle proper mainly comprises cutin embedded in waxes. Epicuticular waxes cover the cuticle proper. The middle lamella consists of suberin and pectin. The thickness, composition and existence of these layers can vary significantly among tissues, organs, developmental stages, and species. (Modified from Pollard et al. (2008))

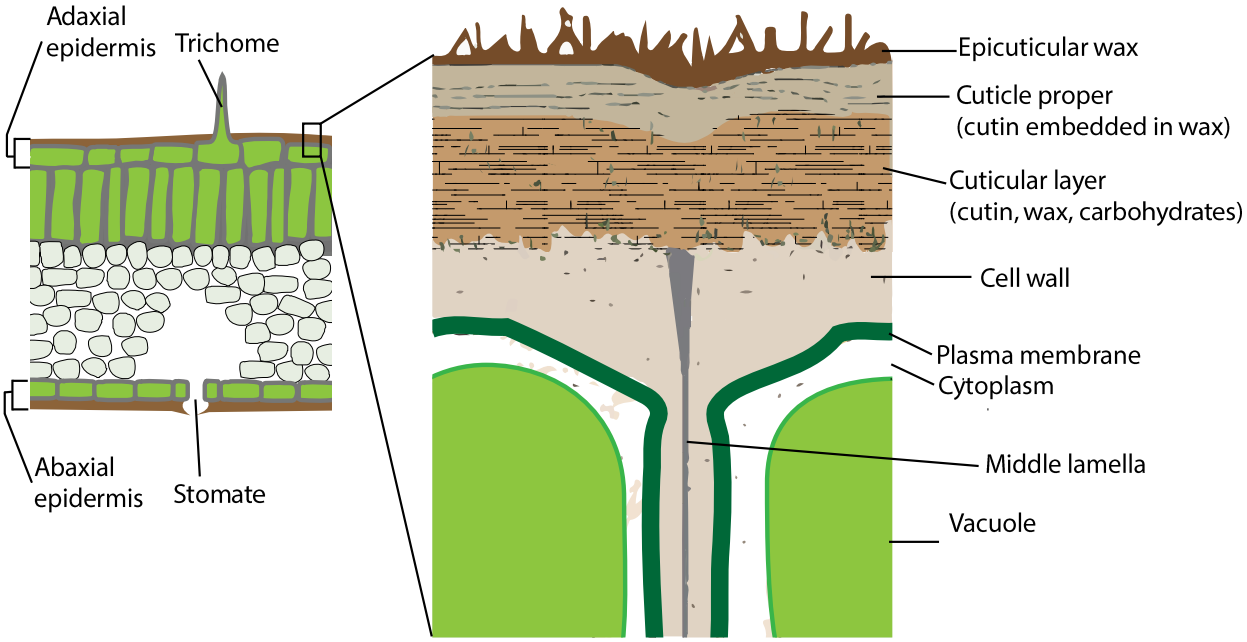
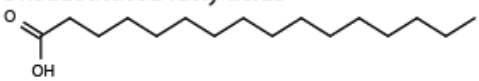
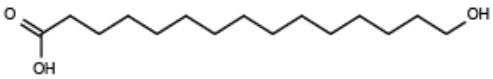
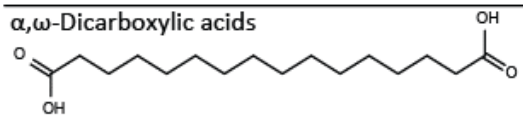
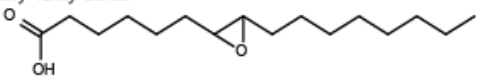
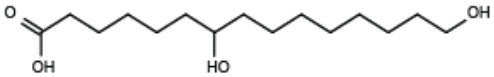
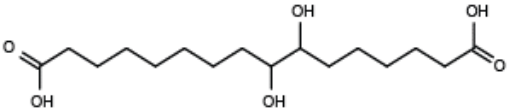
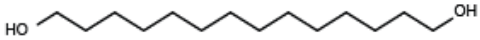
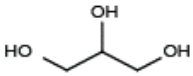
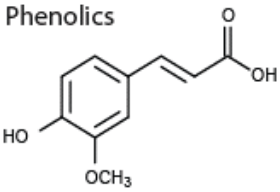

Adaxial epidermis
Trichome
Abaxial epidermis
Stomate
Epicuticular wax
Cuticle proper (cutin embedded in wax)
Cuticular layer (cutin, wax, carbohydrates)
Cell wall
Plasma membrane
Cytoplasm
Middle lamella
Vacuole

Table 1.1 Structures of seven types of common cutin monomers and typical ranges of composition values. Monomer types were assigned with individual symbols, which are referred to throughout this chapter. Representative structures within each monomer type are shown. The mid-chain functional groups including epoxy, hydroxy, vicinal dihydroxy and oxo groups can be part of normal fatty acids, ω-hydroxy fatty acids and α,ω-dicarboxylic acids. (Modified from Pollard et al. (2008))

Examples of cutin monomer	Abundance (%) and common monomers	Symbol
Unsubstituted fatty acids 	1-25% C16:0, C18:0, C18:1, C18:2	
ω-Hydroxy fatty acids 	1-32% C16:0, C18:1, C18:2	
α,ω-Dicarboxylic acids 	Usually < 5% but >50% in Arabidopsis C16:0, C18:0, C18:1, C18:2	
Mid-chain functionalized monomers		
Epoxy-fatty acids 	0-34% C18:0 (9,10-epoxy) C18:1(9,10-epoxy)	
Polyhydroxy-fatty acids 	16-92% C16:0 (10,16-dihydroxy) C18:0 (9,10,18-trihydroxy)	
Polyhydroxy α,ω-dicarboxylic acids 	Trace	
Fatty alcohols		
Alkan-1-ols and alken-1-ols 	0-8% C16:0, 18:1	
α,ω-Alkanediols and alkenediols 	0-5% C18:1	
Glycerol 	1-14%	
Phenolics 	1-14% Ferulate	

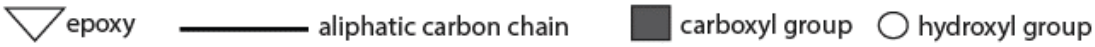


Figure 1.2 Hypothetical monomer linkage patterns of cutin polymers. (A) A small segment of a glycerol-linked cutin polyester is shown to illustrate the dominant primary ester bonds and a secondary ester bond that enables a branch point. (B-D) Representations of possible monomer polymerization patterns found in cutin structures. Mid-chain functional groups other than OH groups are omitted. (B) A branching domain made of fatty acid and ω-hydroxy fatty acid monomers that are connected by primary and secondary ester bonds. (C) A branching domain made of α,ω-dicarboxylic acids and glycerol monomers with free OH groups on some of the glycerol monomers. (D) With less free OH groups on the glycerol, α,ω-dicarboxylic acids and glycerol monomers can also form a cross-linked network domain. See Table 1.1 for symbols. (Modified from Pollard et al. (2008))

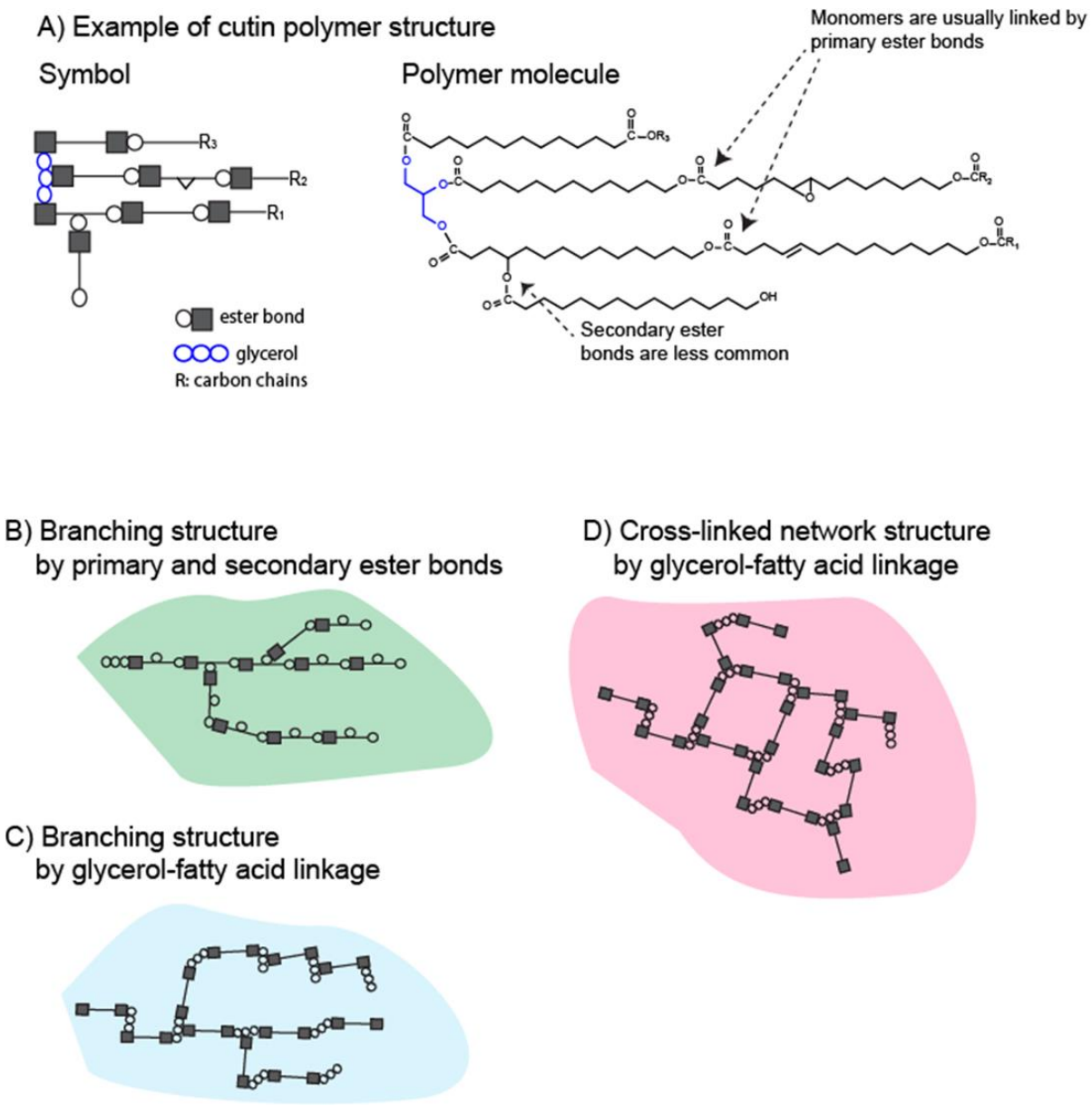

A) Example of cutin polymer structure
Symbol
Polymer molecule
Monomers are usually linked by primary ester bonds
R3
R2
R1
ester bond
glycerol
R: carbon chains
Secondary ester bonds are less common
B) Branching structure by primary and secondary ester bonds
C) Branching structure by glycerol-fatty acid linkage
D) Cross-linked network structure by glycerol-fatty acid linkage

Figure 1.3 A simplified biosynthetic scheme representing the steps for the synthesis of cutin building blocks. Acetate or pyruvate is processed to acetyl-CoA by acetyl-CoA synthase (ACS) and pyruvate dehydrogenase complex (PDC). The acetyl-CoA is then processed into malonyl-CoA by acetyl-CoA carboxylase (ACCase), which is transformed to fatty acids by fatty acid synthase (FAS). Fatty acids are transported to the endoplasmic reticulum for further modification. Acyl chains of fatty acids are activated to CoA by long-chain fatty acyl-CoA synthase (LACS) and hydroxylated by fatty acyl ω-hydroxylase (FAH). Alternatively, hydroxyl acids can be modified directly to dicarboxyl acids. The following step involving ω-hydroxy fatty acyl dehydrogenase (HFADH) transforms hydroxy fatty acids to oxo products, which are processed into dicarboxylic fatty acids by ω-oxo fatty acyl dehydrogenase (OFADH). These modified fatty acids may be esterified to a glycerol-3-phosphate by glycerol-3-phosphate acyltransferase (GPAT). Monomers may also be esterified by polyester synthase (PS) to form oligomers before being exported. The acyl glycerol synthesis step is shown for dicarboxylic acids only. X, position of a C-C double bound; solid arrow, steps within the biosynthesis pathway; dotted arrow, transport of molecules. (Modified from Pollard et al. (2008), Gronwald (1991), and Plant Metabolic Network, http://www.plantcyc.org)

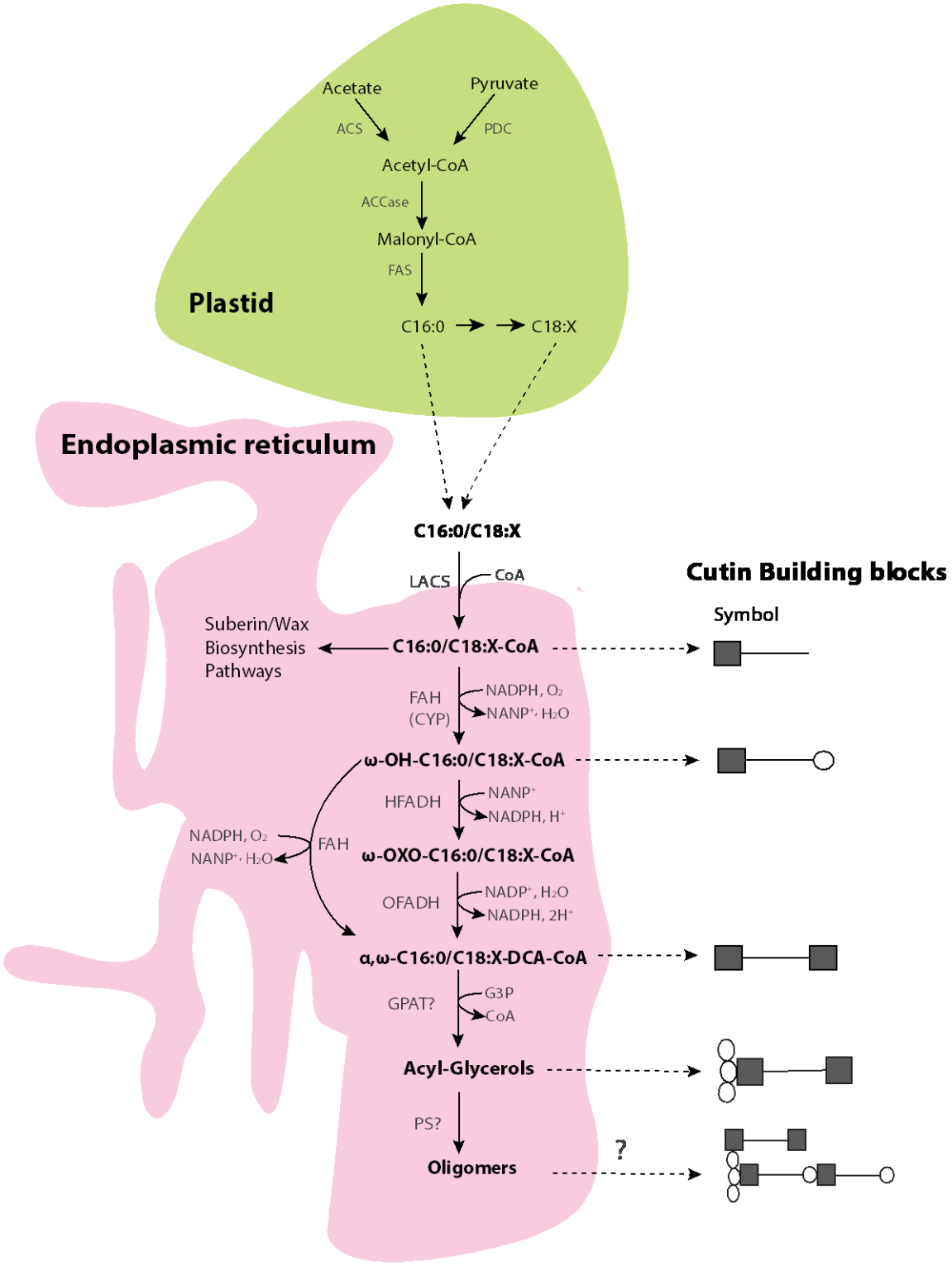

Acetate
Pyruvate
ACS
PDC
Acetyl-CoA
ACCase
Malonyl-CoA
FAS
Plastid
C16:0
C18:X
Endoplasmic reticulum
C16:0/C18:X
LACS
CoA
Cutin Building blocks
Symbol
Suberin/Wax
Biosynthesis
Pathways
C16:0/C18:X-CoA
FAH
(CYP)
NADPH, O2
NANP+, H2O
ω-OH-C16:0/C18:X-CoA
HFADH
NANP+
NADPH, H+
NADPH, O2
NANP+, H2O
FAH
ω-OXO-C16:0/C18:X-CoA
OFADH
NADP+, H2O
NADPH, 2H+
α,ω-C16:0/C18:X-DCA-CoA
GPAT?
G3P
CoA
Acyl-Glycerols
PS?
Oligomers
?

Figure 1.4 Putative mechanisms and subcellular locations of cutin assembly. Monomer biosynthesis is thought to localize to the endoplasmic reticulum (ER). Four major routes for the movement of cutin building blocks from the ER through the cell wall to their final destination in the cuticle are illustrated (a-d). Monomers and oligomers are likely transported via route (a) and (b), whereas route (c) and (d) are likely for oligomers. (a) Plasma membrane-anchoring ER domain facilitates the spatial coupling of monomer synthesis and transport across the plasma membrane. (b) Cytoplasmic soluble carrier protein. An ABC transporter and/or glycosylphosphatidyl-inositol (GPI)-anchored lipid transfer protein (LTPG) may be required for (a) and (b). (c) Oleophilic droplets. (d) Golgi-mediated vesicular secretion is a possible major route, especially for polymer domains or polymers attached to polysaccharides. In addition, two proposed mechanisms where lipophilic precursors pass through the cell wall to the cutin polymer assembly site are depicted (i and ii). (i) Unchaperoned movement of oleophilic droplets across the cell wall. (ii) Movement of monomers, oligomers or polymers bound to a protein carrier (e.g. lipid transfer proteins, LTP) or after attachment to a carrier such as a cell wall polysaccharide. Polyester synthases (PS) are enzymes catalyzing the polymerization between monomers, and also between putative polyester oligomers or domains; their cellular localization remains to be identified. (Modified from Pollard et al. (2008))

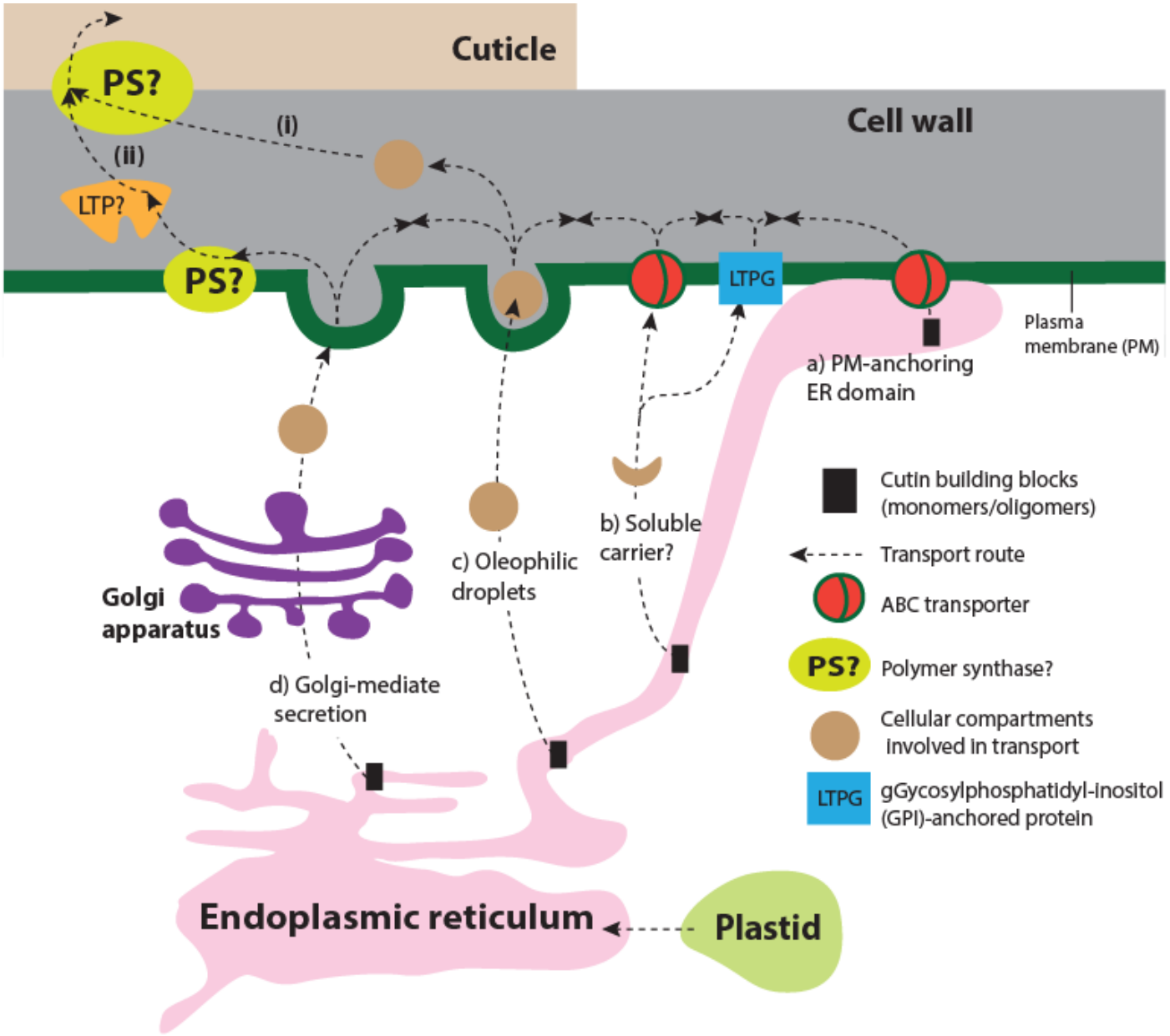

Cuticle
Cell wall
PS?
(i)
(ii)
LTP?
PS?
LTPG
Plasma membrane (PM)
a) PM-anchoring ER domain
b) Soluble carrier?
c) Oleophilic droplets
Golgi apparatus
d) Golgi-mediate secretion
Endoplasmic reticulum
Plastid
Cutin building blocks (monomers/oligomers)
Transport route
ABC transporter
PS?
Polymer synthase?
Cellular compartments involved in transport
LTPG
gGycosylphosphatidyl-inositol (GPI)-anchored protein

Table 1.2 Genes involved in cuticle cutin and wax formation. Modified from Javelle et al. (2011)

Class/Plant	Cutin and／or waxes	Gene name or mutant	Protein family or possible function	Organs with cuticle phenotype in mutant or transgenic plant	Reference
Signaling					
Arabidopsis thaliana	Not determined	*ZOU／RGE*	bHLH TF	Cotyledons	Yang et al. (2008); Kondou et al. (2008)
Arabidopsis thaliana	Not determined	*ALE1*	Subtilisin-like serine protease	Cotyledons and leaves	Tanaka et al. (2001)
Arabidopsis thaliana	Not determined	*ALE2*	RLK	Ovules, cotyledons and leaves	Tanaka et al. (2007)
Arabidopsis thaliana	Not determined	*ACR4*	RLK	Ovules and leaves	Watanabe et al. (2004)
Maize	Not determined	*CR4*	RLK	Leaves	Jin et al. (2000)
Arabidopsis thaliana	Not determined	*GSO1*	LRR kinase	Cotyledons	Tsuwamoto et al. (2008)
Arabidopsis thaliana	Not determined	*GSO2*	LRR kinase	Cotyledons	
Biosynthesis/polymerization					
Arabidopsis thaliana	Cutin	*LCR*	Cytochrome P450	Leaves and floral organs	Wellesen et al. (2001)
Arabidopsis thaliana	Cutin	*ATT1*	Cytochrome P450	Leaves and inflorescence stem	Xiao et al. (2004)
Arabidopsis thaliana	Cutin	*ACE／HTH*	Long-chain ω-fatty alcohol dehydrogenases	Floral organs	Kurdyukov et al. (2006b)
Arabidopsis thaliana	Cutin	*GPAT4*	Glycerol-3-phosphate acyltransferase	Seedlings	Li et al. (2007)
Arabidopsis thaliana	Cutin	*GPAT8*	Glycerol-3-phosphate acyltransferase	Cuticular edges of stomata	Li et al. (2007)
Arabidopsis thaliana	Cutin	*DCR*	Glycerol-3-phosphate acyltransferase	Seeds, vegetative organs and floral organs, trichomes, and seed coat	Panikashvili et al. (2009);(Marks et al., 2009)
Arabidopsis thaliana	Waxes	*KCS1*	Long-chain acyl-CoA synthetase		(Todd et al., 1999)
Arabidopsis thaliana	Waxes	*KCS2*	Long-chain acyl-CoA synthetase	Roots and seeds	(Franke et al., 2009)
Arabidopsis thaliana	Waxes	*KCS5*	Long-chain acyl-CoA synthetase	Stems	(Trenkamp et al., 2004)
Solanum lycopersicum	Cutin	*CD1*	Acyltransferase		Yeats et al. (2012)
Arabidopsis thaliana	Cutin and waxes	*BDG*	α／β hydrolase	Leaves, trichomes and seed coat	Kurdyukov et al. (2006a)
Arabidopsis thaliana	Waxes	*LACS1／CER8*	Long-chain acyl-CoA synthetase	Floral organs	Lue et al. (2009)
Arabidopsis thaliana	Cutin	*LACS2*	Long-chain acyl-CoA synthetase	Vegetative organs	Schnurr et al. (2004)
Arabidopsis thaliana	Waxes	*CER9*	ubiquitin-protein ligase activity	cotyledons, leaves, roots, stems, inflorescences and siliques	Lu et al. (2012)
Arabidopsis thaliana	Waxes	*CER10*	ECR	Vegetative and floral organs, siliques	Gable et al. (2004)
Arabidopsis thaliana	Waxes	*PAS2*	HCD	Seeds and vegetative organs	Bach et al. (2008)
Arabidopsis thaliana	Cutin and waxes	*CER3/WAX2*	Sterol desaturase	Vegetative organs and floral organs, siliques, lateral root primordia	Chen et al. (2003)
Arabidopsis thaliana	Waxes	*CER4*	FAR	Inflorescence stem	Rowland et al. (2006)
Arabidopsis thaliana	Waxes	*WSD1*	Acyl-CoA: diacylglycerol acyltransferase	Inflorescence stem	Li et al. (2008)
Arabidopsis thaliana	Waxes	*MAH1*	Cytochrome P450	Inflorescence stem	Greer et al. (2007)

Table 1.2 (Continued)

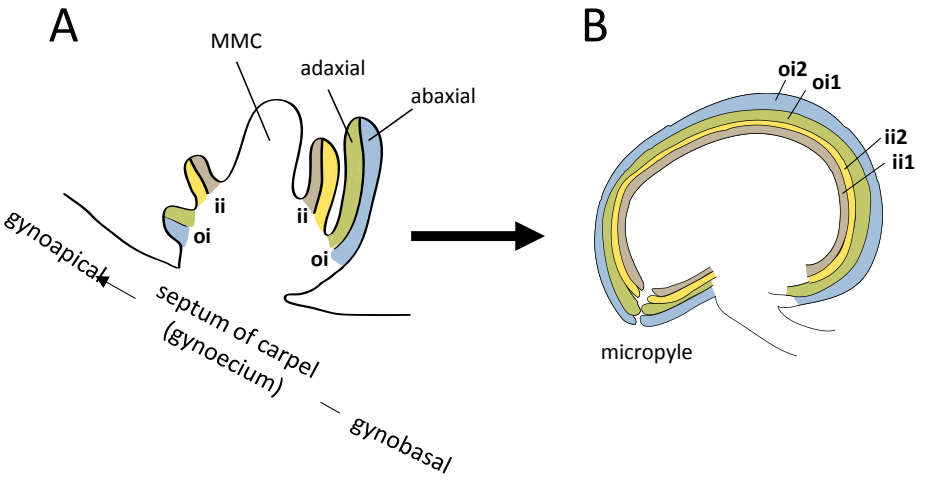
Class/Plant	Cutin and／or waxes	Gene name or mutant	Protein family or possible function	Organs with cuticle phenotype in mutant or transgenic plant	Reference
Arabidopsis thaliana	Waxes	*HIC*	KCS	Stomata	Gray et al. (2000)
Arabidopsis thaliana	Waxes	*CER6／CUT1*	KCS	Inflorescence stem, siliques and pollen	Fiebig et al. (2000); Millar et al. (1999)
Arabidopsis thaliana	Waxes	*FAE1*	KCS	Seeds	James et al. (1995)
Arabidopsis thaliana	Wax	*FDH*	KCS	Leaves and floral organs	Yephremov et al. (1999)
Rice	Waxes	*WSL1*	KCS	Leaves and sheath	Yu et al. (2008)
Arabidopsis thaliana	Waxes	*KCR1*	KCR	Seeds, vegetative and floral organs	Beaudoin et al. (2009)
Maize	Waxes	*GLOSSY8*	KCR	Juvenile leaves	Dietrich et al. (2005)
Arabidopsis thaliana	Waxes	*CER1*	Fatty acid hydrolase／putative decarbonylase	Inflorescence stem and pollen	Aarts et al. (1995)
Maize	Cutin and waxes	*GLOSSY1*	Desaturase／hydroxylase	Juvenile leaves	Sturaro et al. (2005)
Transport					
Arabidopsis thaliana	Waxes	*ATABCG12/CER5*	ABC transporter	Inflorescence stem	Pighin et al. (2004)
Arabidopsis thaliana	Cutin and waxes	*WBC/ ATABCG11*	ABC transporter	Vegetative organs, trichomes and floral organs	Bird et al. (2007); Panikashvili et al. (2007)
Arabidopsis thaliana	Cutin	*ATABCG13*	ABC transporter		Panikashvili et al. (2011)
Arabidopsis thaliana	Cutin	*ATABCG32*	ABC transporter	Vegetative organs, trichomes and floral organs, siliques	Bessire et al. (2011)
Arabidopsis thaliana	Cutin and waxes	*LTPG1*	LTPG	Inflorescence stem, siliques and seed coat	Lee et al. (2009a); DeBono et al. (2009)
Arabidopsis thaliana	Cutin and waxes	*LTPG2*	LTPG	Inflorescence stem	(Kim et al., 2012)
Regulation					
Arabidopsis thaliana	Cutin and waxes	*WIN／SHN1*	AP2／EREBP TF	Vegetative and floral organs	Aharoni et al. (2004)
Arabidopsis thaliana	Not determined	*WIN／SHN2*	AP2／EREBP TF	Vegetative and floral organs Leaves and siliques	Aharoni et al. (2004); Kannangara et al. (2007)
Arabidopsis thaliana	Not determined	*AtMYB41*	MYB R2R3 TF		Cominelli et al. (2008)
Arabidopsis thaliana	Waxes	*AtMYB30*	MYB R2R3 TF	Leaves	Raffaele et al. (2008)
Arabidopsis thaliana	Waxes	*CER7*	RRP45 3'exoribonuclease	Inflorescence stem and siliques	Hooker et al. (2007)
Medicago sativa	Waxes	*WXP1*	AP2／EREBP TF	Leaves	Zhang et al. (2007)
Arabidopsis thaliana	Cutin and waxes	*ACP4*	Acyl carrier protein	Leaves	Xia et al. (2009)

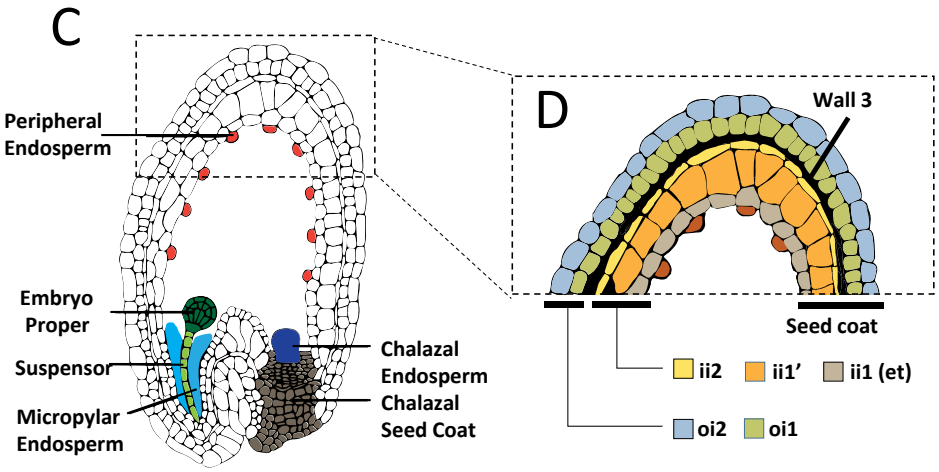
Table 1.2 (Continued)

Class/Plant	Cutin and / or waxes	Gene name or mutant	Protein family or possible function	Organs with cuticle phenotype in mutant or transgenic plant	Reference
Other					
Arabidopsis thaliana	Cutin	*CDEF1*	GDSL lipase/esterase	Lateral root emergence	Takahashi et al. (2010)
Maize	Waxes	*GLOSSY2*	Transferase similar to CER2	Juvenile leaves	Tacke et al. (1995)
Maize	Waxes	*glossy13*	ABC transporter	Seedlings	Li et al. (2013)
Maize	Waxes	*glossy3*	Elongation step C28-C30	Juvenile leaves	Bianchi et al. (1985); Avato et al. (1987)
Maize	Waxes	*glossy4*	Elongation step C30-C32		
Maize	Waxes	*glossy5*	Reductase producing C32 alcohols		
Maize	Waxes	*glossy7*	Production of fatty acids acting downstream of *GLOSSY1*		
Maize	Waxes	*glossy11*	Reductase producing aldehydes		
Maize	Waxes	*glossy16*	Elongation step C30-C32		
Maize	Waxes	*glossy18*	Production of fatty acids		

ACE, ADHESION OF CALIX EDGE; ACR, ARABIDOPSIS CRINKLY; ALE, ABNORMAL LEAF SHAPE; AP2, ACTIVATOR PROTEIN 2; ATT, ABERRANT INDUCTION OF TYPE THREE GENES; BDG, BODYGUARD; bHLH, basic helix-loop-helix; *CDEF1, CUTICLE DESTRUCTING FACTOR 1; CR, CRINKLY; DCR, DEFECTIVE IN CUTICULAR RIDGES*; ECR, enoyl-CoA reductase; *EREBP, ETHYLENE RESPONSE ELEMENT BINDING PROTEIN; FAE, FATTY ACID ELONGATION;* FAR, fatty acid reductase*; FDH, FIDDLEHEAD; GPAT, GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE; GSO, GASSHO;* HCD, β-hydroxyacyl-CoA dehydratase; *HIC, HIGH CARBON DIOXIDE; HTH, HOTHEAD; KCR, β-KETO ACYL REDUCTASE*; KCS, β-ketoacyl-CoA synthase; *LACS, LONG-CHAIN ACYL-COA SYNTHETASE; LCR, LACERATA*; LRR, leucine-rich repeat; *LTPG, LIPID TRANSFER PROTEIN G; MAH, MIDCHAIN ALKANES HYDROXYLASE; PAS, PASTICCHINO; RGE, RETARDED GROWTH OF EMBRYO*; RLK, receptor-like kinase; RRP, ribosomal RNA processing*; SHN, SHINE*; TF, transcription factor; *WBC, WHITE BROWN COMPLEX; WIN, WAX INDUCER; WSD, WAX ESTER SYNTHASE / ACYL-COA:DIACYLGLYCEROL ACYLTRANSFERASE; WSL, WAX CRYSTAL-SPARSE LEAF; WXP, WAX PRODUCTION; ZOU, ZHOUPI; DCR, DEFECTIVE IN CUTICULAR RIDGES.*

Figure 1.5 Schematic diagrams of the *Arabidopsis* seed coat development. (A) During early female gametophyte development, integuments on the gynobasal side elongate first. Each of the outer (oi) and inner (ii) integuments consists of an abaxial (‘2’) and an adaxial (‘1’) layer. (B) At the mature stage, the inner and outer layers of gynobasal integuments have grown around the embryo sac. The gynoapical integuments elongate to a lesser extent. Inner and outer integuments from both sides meet and form the micropyle. (C-D) Illustration of the seed and seed coat structure. After fertilization, ovule integuments develop into the seed coat. Between ii2 and ii1 (also known as endothelium, et), an extra internal cell layer (ii1’) is present towards the chalazal zones of the seed coat. The outer integuments are separated from the inner integuments by an electron-dense cell wall layer (‘wall 3’) that is rich in cutin-like material. The vast majority of the wall material deposited in wall 3 is produced by the oi1 layer. (E) Shortly before the embryo is fully expanded, mucilage formation is completed in oi2, with the presence of the amyloplast-containing columella in the middle and two mucilage pockets on the sides. Cells of ii2 and ii1’ (sometimes also oi1) collapse and form the brown pigment layer (bpl). MMC, megaspore mother cell; ii1, inner (adaxial) layer of inner integument; ii2: outer (abaxial) layer of inner integument; oi1: inner (adaxial) layer of outer integument; oi2: outer (abaxial) layer of outer integument. (derived from Truernit and Haseloff (2008) and www.seedgenenetwork.net)


A
MMC
adaxial
abaxial
ii
oi
ii
oi
gynoapical
septum of carpel
(gynoecium)
gynobasal
B
oi2
oi1
ii2
ii1
micropyle


C
Peripheral
Endosperm
Embryo
Proper
Suspensor
Micropylar
Endosperm
Chalazal
Endosperm
Chalazal
Seed Coat
D
Wall 3
Seed coat
ii2
ii1'
ii1 (et)
oi2
oi1

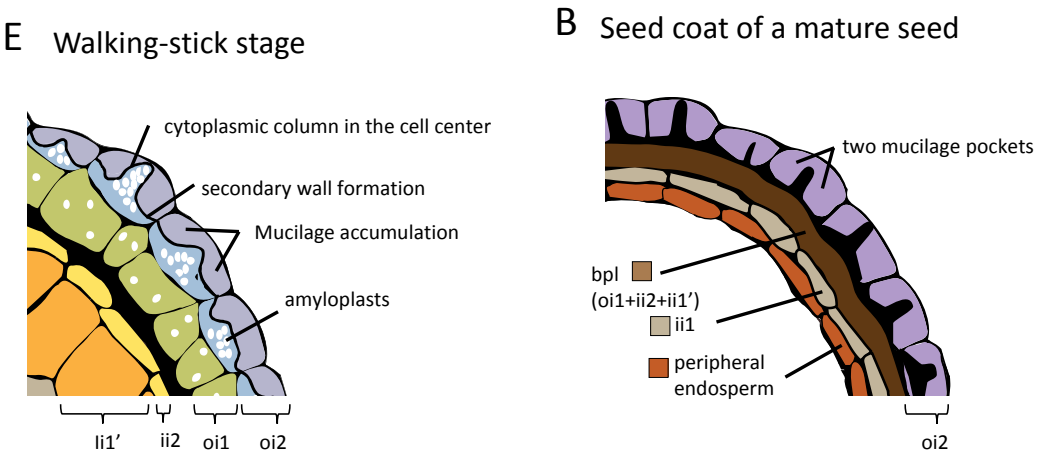

E Walking-stick stage
cytoplasmic column in the cell center
secondary wall formation
Mucilage accumulation
amyloplasts
Ii1'
ii2
oi1
oi2
B Seed coat of a mature seed
two mucilage pockets
bpl
(oi1+ii2+ii1')
ii1
peripheral
endosperm
oi2

Figure 1.6 A model of ER body formation in *A. thaliana*. *NAI1* encodes a putative transcription factor that regulates the expression of four key genes for ER body formation: *PYK10*, *NAI2*, *MEB1* (*MEMBRANE OF ER BODY 1*), and *MEB2*. In the ER cisternae, PYK10 and NAI2 first interact to form a core that continues to enlarge. NAI2 forms a complex with MEB1 and MEB2 that are later integrated to the ER body-specific membrane. At maturation, the spindle-shaped ER subdomain breaks off from the ER network and form a separate body. (Derived from Nakano et al. (2014))

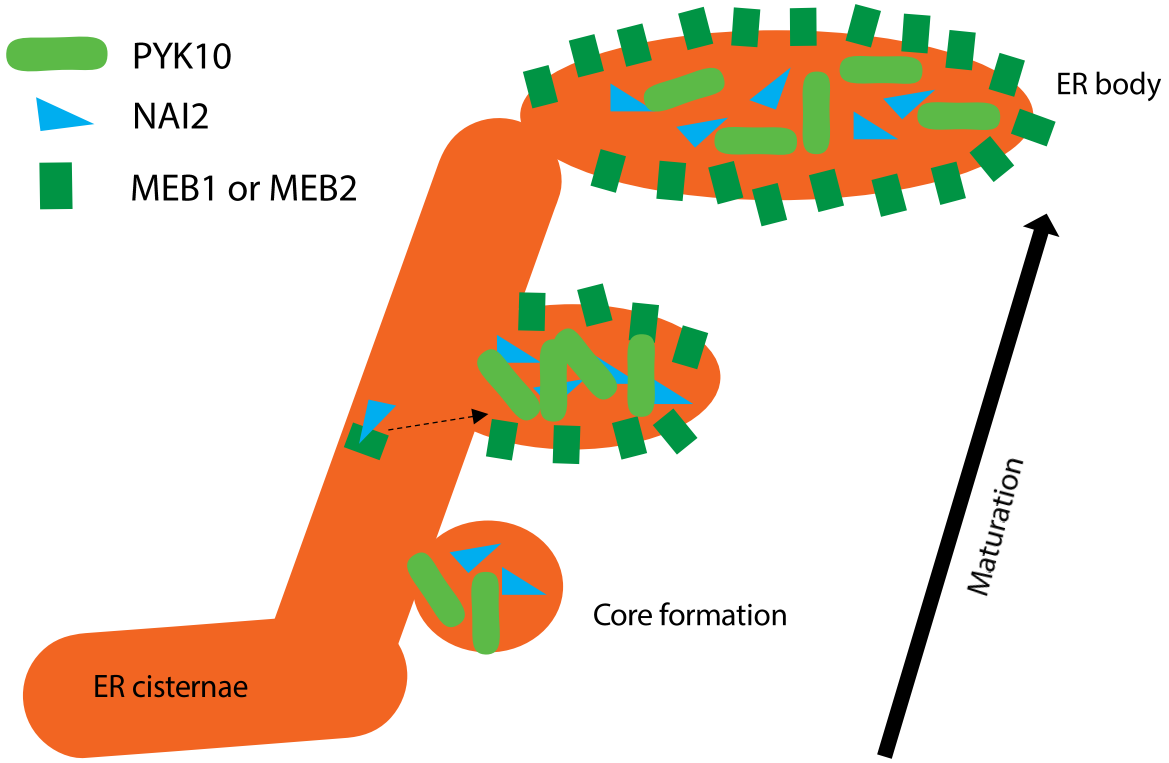

PYK10
NAI2
MEB1 or MEB2
ER body
Core formation
ER cisternae
Maturation

Figure 1.7 Simplified scheme of glucosinolate hydrolysis . Myrosinase acts on glucosinolates to form an unstable aglycone intermediate that spontaneously forms an isothiocyanate or thiocyanate by default. Under certain conditions (e.g. the presence of Fe2+ or at pH < 5), the aglycone can give rise to a corresponding nitrile. Nitriles can be metabolized to produce hydrogen cyanide for cyanogenesis. NSP is required for the nitrile formation, whereas ESP is required for epithionitriles. ESP, epithio-specifier protein; NSP nitrile-specifier protein; R, variable side chain; n = 1 or 2. (Modified from Kissen and Bones (2009) and Lambrix et al. (2001))

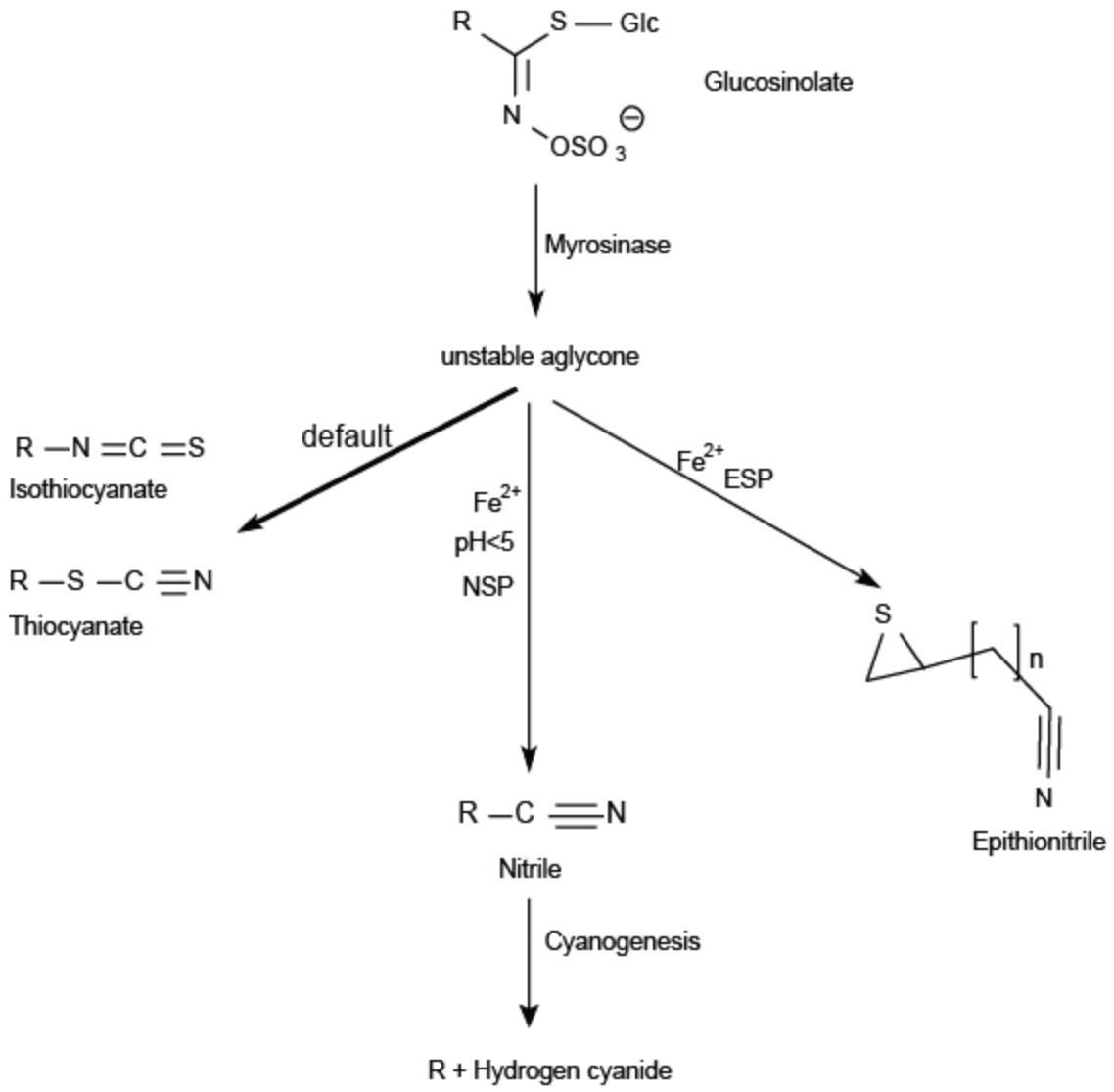

R
S — Glc
N
OSO 3
Glucosinolate
Myrosinase
unstable aglycone
default
R —N =C =S
Isothiocyanate
R —S —C ≡N
Thiocyanate
Fe2+
pH<5
NSP
Fe2+
ESP
S
n
N
Epithionitrile
R —C ≡N
Nitrile
Cyanogenesis
R + Hydrogen cyanide

Figure 1.8 The *HTH* gene model. The rectangular boxes represent the six exons and the lines connecting the rectangles represent the introns. The relative positions of single nucleotide point mutations (*hth-1* to *hth-11*), transposon insertion sites (*hth-12* and *eda17*) and T-DNA insertion sites (*hth-13, hth-14* and *hth-15*) are indicated (Krolikowski et al., 2003; Kurdyukov et al., 2006b). The length of the 5' upstream region is 2009 bp and the coding region is 2834 bp. Mutant alleles generated by T-DNA insertion are in bold. Genes labelled with an asterisk (*) putatively encode a truncated protein.

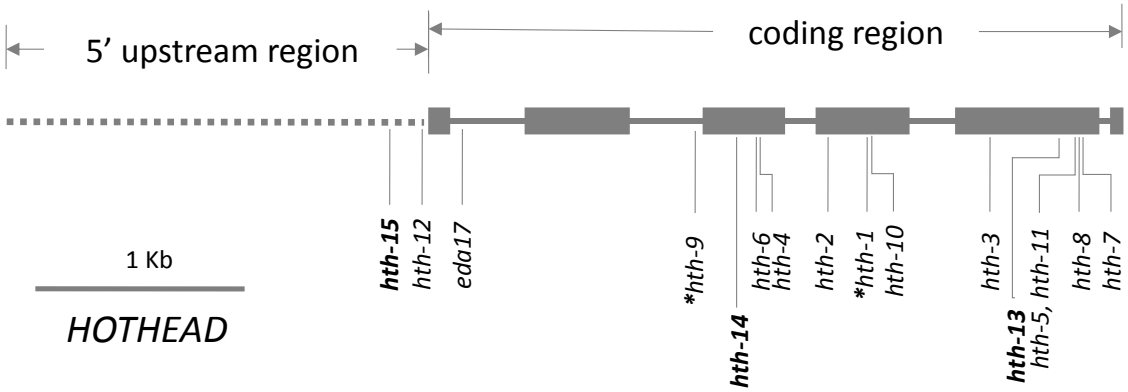

5' upstream region
coding region
1 Kb
HOTHEAD
hth-15
hth-12
eda17
*hth-9
hth-14
hth-6
hth-4
hth-2
*hth-1
hth-10
hth-3
hth-13
hth-5, hth-11
hth-8
hth-7

Figure 1.9 Single nucleotide changes found in mutant *hth* alleles. The change in DNA sequence and the position relative to the start of the coding sequence are indicated. (Modified from Krolikowski et al. (2003))

Allele	Genetic Background	DNA mutation[a]	Amino acid change[b]
Truncation/Splicing			
hth-1	Landsberg *erecta*	C_{1937} > T	Gln_{353} > stop
hth-9	Wassilewskija	G_{1257} > A	Splice site
Amino acid substitution			
hth-2	Landsberg *erecta*	G_{1761} > A	Gly_{294} > Glu
hth-3	Landsberg *erecta*	G_{2267} > A	Gly_{435} > Arg
hth-4	Landsberg *erecta*	C_{1472} > T	Arg_{227} > Cys
hth-5,11	Landsberg *erecta*	C_{2654} > T	Pro_{564} > Ser
hth-6	Landsberg *erecta*	G_{1445} > A	Gly_{218} > Ser
hth-7	Landsberg *erecta*	C_{2661} > A	Thr_{566} > Ile
hth-8	Landsberg *erecta*	G_{2657} > A	Gly_{565} > Arg
hth-10	Landsberg *erecta*	G_{1947} > A	Gly_{356} > Glu

a, the change in DNA sequence and the position relative to the start of the coding sequence
b, the change in the corresponding theoretical protein sequence and the position of amino acid

Figure 1.10 Floral phenotypes of wildtype and *hth* mutant plants. (A-C) wildtype and (D-F) *hth* mutant Arabidopsis plants. (A, D) Images of Arabidopsis inflorescences. (B-F) Schematic drawings of longitudinal (B, E) and cross-sectional (C, F) views of an individual flower. The mutant phenotype of closed flowers and protruding pistils were illustrated. Scale bar: 0.5 cm.

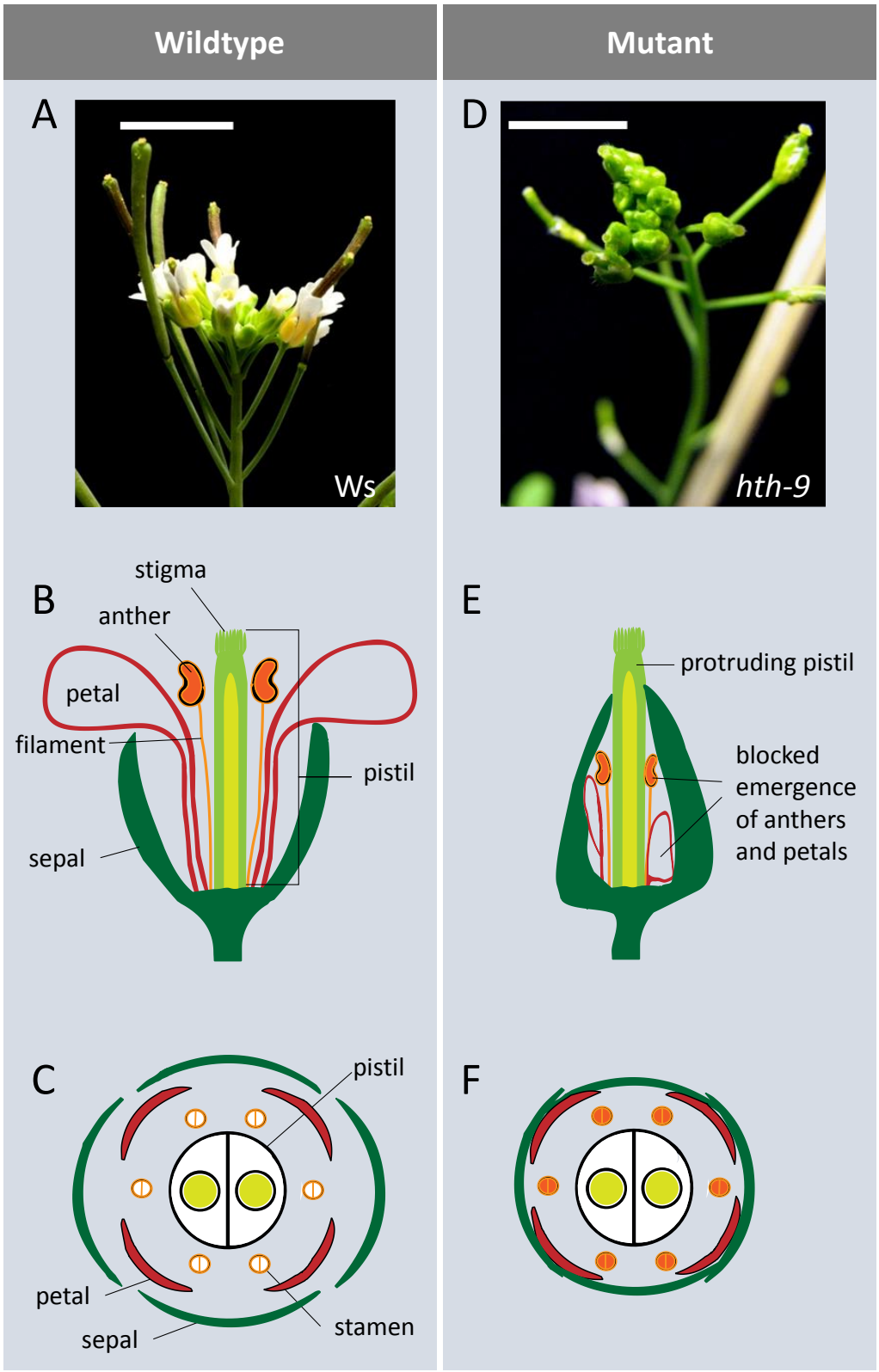

Wildtype
Mutant
A
Ws
D
hth-9
B
stigma
anther
petal
filament
pistil
sepal
E
protruding pistil
blocked
emergence
of anthers
and petals
C
pistil
petal
sepal
stamen
F

Figure 1.11 A proposed ω-oxidation pathway of fatty acids in cutin monomer biosynthesis in Arabidopsis. LCR and ATT1 have been identified as fatty acyl ω-hydroxylases that give rise to hydroxypalmitate. HTH is proposed to convert the hydroxyl fatty acid to an oxo product, which later is oxidized to a dicarboxylic acid. Oxidation steps for palmitate (C16:0) are shown, but HTH may also act on other substrates. FAH, fatty acyl ω-hydroxylase; HFADH, ω-hydroxy fatty acyl dehydrogenase; OFADH, ω-oxo fatty acyl dehydrogenase. (Modified from Kurdyukov et al. (2006b))

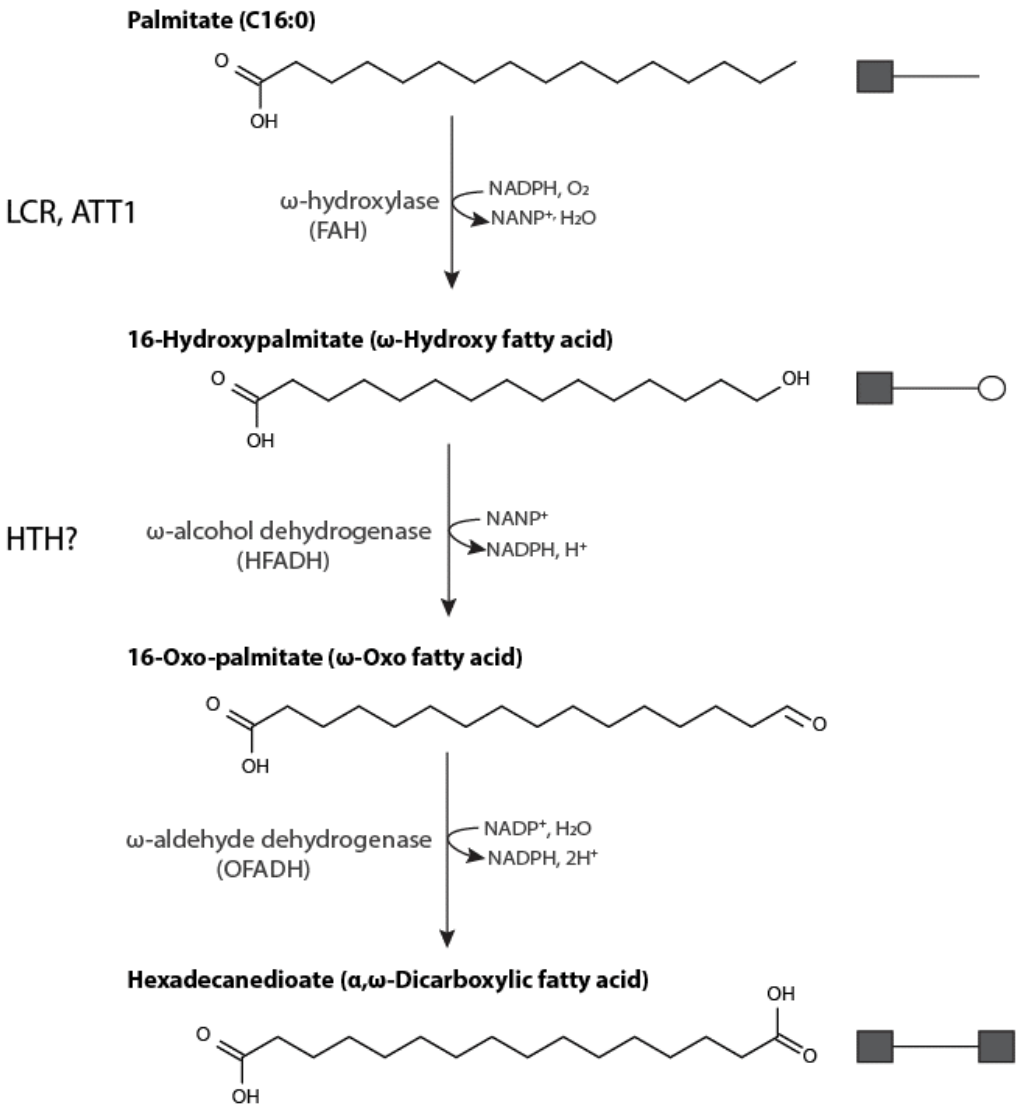

Palmitate (C16:0)
O
OH
LCR, ATT1
ω-hydroxylase
(FAH)
NADPH, O2
NANP+, H2O
16-Hydroxypalmitate (ω-Hydroxy fatty acid)
O
OH
OH
HTH?
ω-alcohol dehydrogenase
(HFADH)
NANP+
NADPH, H+
16-Oxo-palmitate (ω-Oxo fatty acid)
O
OH
O
ω-aldehyde dehydrogenase
(OFADH)
NADP+, H2O
NADPH, 2H+
Hexadecanedioate (α,ω-Dicarboxylic fatty acid)
OH
O
OH
O

Figure 1.12 Cyanogenesis from mandelonitrile, a hydroxyl nitrile. Mandelonitrile lyase catalyzes the chemical reaction that yields hydrogen cyanide and benzaldehyde (Modified from Yemm and Poulton (1986)).

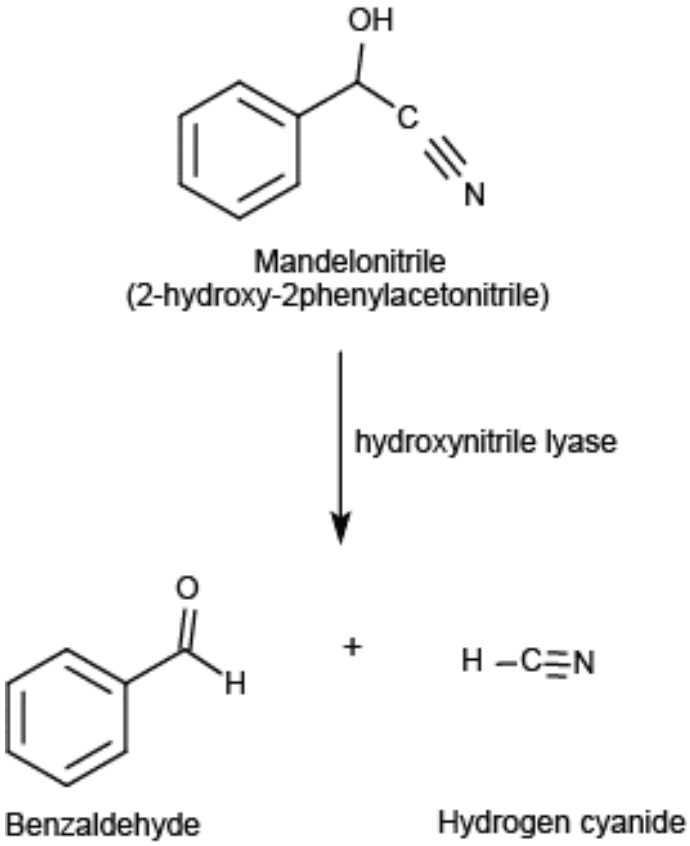
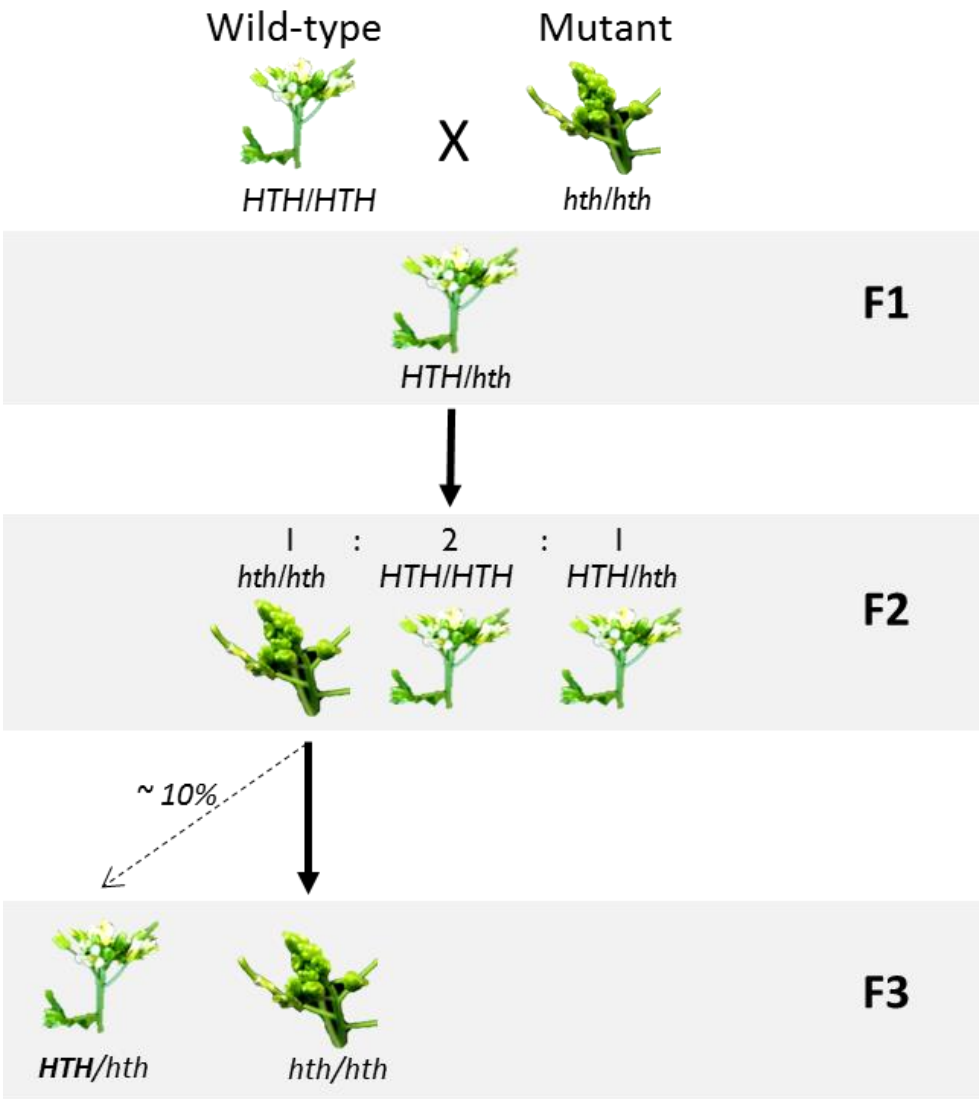

OH
C
N
Mandelonitrile
(2-hydroxy-2phenylacetonitrile)
hydroxynitrile lyase
O
H
+
H –C≡N
Benzaldehyde
Hydrogen cyanide

Figure 1.13 The schematic diagram depicting non-Mendelian inheritance observed by Lolle et al. (2005). Homozygous mutant F2 plants give rise to genotypically heterozygous F3 progeny at a frequency up to 10%. The F3 progeny harbours the HTH allele that is absent in the F2 parent but present in the F1 grandparental generation. The authors proposed the existence of an extra-genomic mechanism that involves a template-directed process. HTH: wildtype HOTHEAD allele; hth: mutant allele

.


Wild-type
Mutant
X
HTH/HTH
hth/hth
F1
HTH/hth
1 : 2 : 1
hth/hth
HTH/HTH
HTH/hth
F2
~ 10%
F3
HTH/hth
hth/hth

Chapter 2 *De novo* genetic variation revealed in somatic sectors of single Arabidopsis plants

This research was published in F1000Research. Hopkins MT, Khalid AM, **Chang PC**, Vanderhoek KC, Lai D, Doerr MD and Lolle SJ. *De novo* genetic variation revealed in somatic sectors of single Arabidopsis plants. F1000Research 2013, 2:5.

This was a collaborative effort and my contributions are specified below:

- I participated in the design of experiments.
- I was intensely involved in sampling, DNA extraction and molecular genotyping. I carried out a great amount of Indel marker genotyping work for F2 and F3 progeny (Section 2.3.2) and sectoring experiments for the adult and seedlings (Figure 2.3).
- I assisted with the DNA sequencing work (Figure 2.4) in terms of cloning, sample preparation and alignment.
- I assisted with sample preparation for qPCR experiments (Figure 2.5 and Figure 2.6).
- I participated in manuscript preparation. I was involved in revisions and image modifications for the final manuscript (original images provided by Hopkins MT and Khalid AM).

2.1 Introduction

Plants live in ever changing environments and must adapt using strategies that fundamentally differ from those employed by animals. Developmental plasticity is at the core of those strategies allowing plants to modify their growth in response to different environmental signals. This type of open-ended modular development enhances survival because damaged or diseased units can readily be discarded without compromising viability. Furthermore, because plants are constrained to sessile life styles, a modular growth habit affords greater versatility allowing phenotypic and genetic variation between modules to be used to the plant's advantage, aiding adaption to pathogen life cycles (Todesco et al., 2010) or to longer-term environmental perturbations such as climate change. As a consequence of this profound developmental versatility, even individuals composed of cell populations derived from different plant species are viable and can coordinate the growth and development of chimeric organs (Szymkowiak and Sussex, 1996). It was proposed that mosaicism offers a unique adaptive advantage for plants by allowing introduction of genetic variants into the gene pool either through vegetative propagation or through sexual reproduction (Whitham and Slobodchikoff, 1981). The authors further proposed that mutations arising somatically have a greater probability of being incorporated into the gene pool than mutations that arise in the gametes precisely because germ line cells are derived from somatic tissues that arise late in the developmental history of the plant (Satina and Blakeslee, 1941; Youngson and Whitelaw, 2008).

The relatively frequent occurrence of mosaics among various plant species has been extensively utilized in the development of novel ornamentals and for the selection and maintenance of desirable traits in many cultivated crops. Any desirable cultivars that have arisen in this manner have been maintained through vegetative propagation and, to date, are responsible for a significant fraction of agriculturally important perennial plants. On the other hand, desirable traits in many important annual crops, such as rice, soybean, maize and wheat, have been introduced through classical genetic manipulations using directed breeding strategies. Once generated, annuals with good agronomic performance are usually maintained by inbreeding.

In recent years, concern has grown over the presumed loss of genetic diversity resulting from the application of modern horticultural and breeding practices. Therefore, the benefit of excellent performance may come with a significant cost (Hopkin, 2008; Walck and Dixon, 2009). However, recent and surprising results suggest that even highly inbred species harbor unanticipated sources of intrinsic genetic variation. For example, highly inbred soybean cultivars have been shown to manifest significant phenotypic and genetic variation in the absence of sexual manipulation (Fasoula and Boerma, 2005; Fasoula and Boerma, 2007; Yates et al., 2012). Such high intrinsic genetic variation has also been demonstrated for a number of other crop plants (Rasmusson and Phillips, 1997).

In the natural world, inbreeding occurs in many highly successful flowering plant species including wild relatives of *Arabidopsis thaliana* (Tang et al., 2007a). Therefore, in nature species that are highly inbred have persisted despite their predicted reduction in

genetic diversity. Why would such inbreeding strategies be successful and what are the implications from an adaptive perspective? One possibility put forward by Barrett (2002) is that such populations are very successful in their particular niches and benefit from producing large numbers of genetically identical offspring. Nevertheless, selection should favor plant species that can co-evolve on time scales reflecting particular environmental challenges such as fluctuations and variations in pathogen populations. In keeping with this view, it has been shown that sequence variation in 20 diverse strains of Arabidopsis is highly non-random. In gene families mediating biotic interactions, such as those implicated in pathogen defense, variation far exceeds that seen in families involved in basic biological processes (Clark et al., 2007).

The underlying mechanisms driving phenotypic variation in highly inbred lines, whether domesticated or wild, have often been inferred and have had limited experimental verification. Nevertheless, relatively simple molecular approaches have provided insight into some of the genomic events coinciding with visible changes in phenotype. In flax, for example, molecular assays have demonstrated that heritable phenotypic changes induced by environmental shifts are accompanied by reproducible changes in genomic DNA including changes in total DNA content, non-random changes in DNA sequences or sequence rearrangements (Chen et al., 2009; Chen et al., 2005; Cullis et al., 2004; Schneeberger and Cullis, 1991). In soybean, reproducible non-random DNA sequence changes induced by in vitro culturing of root explants have also been demonstrated using restriction fragment length polymorphic markers (Roth et al., 1989). Genomic changes manifesting similar hallmarks of

biased sequence alterations have also been described for banana (Oh et al., 2007) and in rice hybrids (Xu et al., 2007).

In the work described by Roth et al. (1989) soybean root explants were shown to repeatedly give rise to particular alleles that were absent in the donor plants but had previously been found and characterized in other varieties of cultivated soybean. To account for the appearance of these particular allelic variants the authors proposed that these organisms had evolved “internal generators of genetic variation” that mediated genome changes through some type of recombination process. Later, Lolle et al. (2005) described a genome-wide phenomenon in Arabidopsis hothead (*hth*) mutants that was very reminiscent of that described by Roth et al. (1989). Based on the nature and genome-wide locations of the sequence changes detected, it was proposed that a template-directed process mediated these changes and that these cryptic but stable extra-genomic templates themselves had persisted since at least the grandparental generation. Not surprisingly, this proposal met with considerable skepticism and numerous alternative explanations for these data have since been published (Chaudhury, 2005; Comai and Cartwright, 2005; Krishnaswamy and Peterson, 2007; Mercier et al., 2008; Peng et al., 2006; Ray, 2005).

In this study we have employed presence-absence molecular markers to test for non-Mendelian inheritance and found that Arabidopsis plants can inherit novel insertion sequences that were absent in their immediate parents. Furthermore, we show that discordant DNA-based marker profiles can be found between tissues isolated from different parts of an individual plant. These experiments demonstrate that individual plants spontaneously

produce somatic sectors and are genetic mosaics. Since genetic variation can occur in the same plant in the absence of sexual reproduction, we propose that these novel insertion sequences must originate from cryptic reserves intrinsic to the host plant itself. The data presented here support the original contention that a previously unknown template-directed mechanism exists (Lolle et al., 2005) and raise the encouraging possibility that other inbreeding species, including crop plants, may also harbor a cryptic reserve of genetic variation.

2.2 Methods

2.2.1 Plant material and growth conditions

All genetic stocks of *Arabidopsis thaliana* used for these experiments have been described previously (Lolle et al., 1998). Arabidopsis seeds derived from these stocks were sown onto moistened potting mix (1:1 mixture of LC1:LG3 Sungro Sunshine potting mixes, Sungro Horticulture, Seba Beach, AB) and stratified at 4°C for 2-5 days. Plants were maintained in growth chambers (Econoair AC60, Ecological Chambers Inc., Winnipeg, MB; GC8-VH/GCB-B, Environmental Growth Chambers, Chagrin Falls, Ohio; Conviron PGW36/E15, Controlled Environments Ltd., Winnipeg, MB) and illuminated with a mixture of incandescent and fluorescent lights (140 – 170 μmol m^{-2} sec^{-1} at pot level) with a 24-hour photoperiod. Growth chambers were maintained at 20 ± 4°C at 40 - 60% relative humidity. Plants were grown in flats or in 3- or 6-inch pots and watered as needed. Seeds used for seedling root-shoot comparison were surface sterilized using bleach and plated on agar medium containing half strength MS basal salts (Sigma-Aldrich, St. Louis, Missouri, USA).

Seedlings were harvested approximately 5 days post-germination. Hybrid lines were generated between wildtype Landsberg *erecta* plants or homozygous *hth* mutant lines in the Landsberg *erecta* background and Columbia accessions by manual pollination. All crosses were done reciprocally. F2 seed was obtained from self-fertilized F1 plants. Individual F2 plants were reared in plastic tubes (Johnston Industrial Plastics, Ontario, Canada) and F3 seed collected from each F2 plant individually. Tissue samples were collected from individual F2 and F3 plants, and genotypic profiles were determined using insertion-deletion polymorphic molecular markers (see Figure 2.1).

2.2.2 **Out-crossing experiments**

Experimental set ups were replicated twice and the net out-crossing frequencies determined. Herbicide-resistant transgenic Arabidopsis pollen donors previously transformed with the pCB302 mini binary vector only (Xiang et al., 1999) and mutant test plants were grown in a 1:1 ratio and arranged in randomized positions (www.random.org). Out-crossing frequencies were also compared to plants under the same conditions but reared within plastic tubes. Progeny were sprayed with glufosinate (40 micrograms ml^{-1} active ingredient: WipeOut, Nu-Gro IP Inc., Ontario) to test for herbicide resistance and resistant plants tested for segregation of *hth* mutant progeny plants.

2.2.3 **DNA extraction and molecular genotyping**

For DNA extraction, rosette or cauline leaf tissue was collected and DNA extracted according to the method of Edwards *et al.*(1991). Samples not processed immediately were stored at -20°C. To distinguish the mutant *hth-4* allele from the wildtype, genomic DNA was

amplified using oligonucleotide primers immediately flanking the *hth-4* point mutation (GAAGCTGGTGAAGGAGTCGT, CTCCGCCGCGGTGTGTC). The resulting 205 base pair (bp) PCR product was then digested with *Sal*I restriction endonuclease (New England Biolabs, Ipswich, Massachusetts, USA) and endonuclease treated PCR products size separated by agarose gel electrophoresis. Sixteen sets of DNA oligonucleotide primers were designed to amplify approximately 150-300 bp genomic regions by polymerase chain reactions (PCR), each containing one 45-94 bp marker which is present in the Columbia but absent in the Landsberg accession (Table 2.1). PCR amplicon products were size separated by agarose gel electrophoresis.

2.2.4 **Isolation, cloning and sequencing of PCR products**

Portions of genomic DNA were PCR amplified and sequenced directly or products cloned into standard pGEM TA vectors (Promega). Amplified or cloned PCR products were sequenced at the Centre for Applied Genomics (http://www.tcag.ca/, Toronto, Ontario). Sequence alignments were generated using CLC Sequence Viewer 6.4 software (www.clcbio.com).

2.2.5 **Quantitative PCR methods**

Quantitative PCR (qPCR) was performed on a Real-Time thermal cycler CFX96 attached to a computer running CFX Manager (Bio-Rad Laboratories, Hercules, California, USA). SsoFast EvaGreen Supermix (Bio-Rad) was used according to manufacturer's instructions. A series of primers either flanking or internal to the insertion sequences were used to generate control and experimental amplicons. The positive control was a PCR product amplified from

the Columbia accession, spanning the indel sequence of interest by ~700-900 bp. The positive control was gel purified and used to generate a standard curve for conversion of C_t value to copy number of the insertion sequence and the external reference sequence. External reference primers immediately flanked the indel markers. Insertion sequences were detected using one external reference primer paired with a primer homologous to sequences within the insertion itself. Primer sequences and amplicon product sizes are listed in Table 2.2. The colours indicated in the first column (insertion-deletion marker) correspond to the colours used for the qPCR-generated bar graphs.

2.3 Results

2.3.1 Mutant *hth* plants are susceptible to higher rates of out-crossing

Homozygous *hth* mutant Arabidopsis plants were previously shown to give rise to wildtype (wt) progeny at relatively high frequencies (Lolle et al., 1998; Lolle et al., 2005). Although an intrinsic mechanism was proposed (Lolle et al., 2005), cross-pollination with neighboring plants was subsequently put forward as the more likely explanation for the appearance of these wildtype revertant offspring (Mercier et al., 2008; Peng et al., 2006). To test the susceptibility of *hth* plants to out-crossing under our growth conditions, experiments were conducted using a pollen donor harboring a dominant gene conferring resistance to the herbicide glufosinate. Herbicide-resistant transgenic lines were grown together with *hth* and *eceriferum-10* (*cer-10*) (Koornneef et al., 1989) floral fusion mutants and wildtype Landsberg plants. These analyses confirmed that the majority of *hth* mutant plants did not cross-pollinate. However, when cross-pollination occurred, frequencies varied considerably

between individual *hth* mutant plants. Mutants with floral fusion phenotypes were predisposed to higher pollen capture than wildtype plants (0.02-0.43% for *hth-4*, *hth-8* and *hth-10* mutants, 0.89% for *cer10* mutants, 0.01% for wildtype plants). In addition, factors such as donor-recipient proximity, the severity of the floral fusion phenotype, growth chamber airflow patterns and plant handling influenced the propensity to cross-pollinate. Nevertheless, growing *hth* mutant F2 plants in the complete absence of *HTH* pollen donors did not eliminate wildtype progeny from F3 progeny pools and, on average, 1.53% of F3 progeny were phenotypically wildtype for *HTH* despite being derived from self-fertilized homozygous F2 *hth* mutant parent plants (2/133 *hth-4*, 2/131 *hth-8* and 2/127 *hth-10* gave rise to wildtype F3 progeny). Under our laboratory conditions, out-crossing could not be completely eliminated within *hth* mutant populations if mutants were grown together with wildtype plants, even if every *hth* mutant plant was shielded in transparent plastic tubes.

While conducting segregation analyses and scoring offspring for herbicide resistance, a single *hth* mutant plant with a large phenotypically wildtype floral sector was identified (Figure 2.2). Sampling of shoot tissues confirmed that phenotype corresponded to genotype and that both mutant *hth-4* and wildtype *HTH* alleles could be detected in tissue derived from this large wildtype sector (Figure 2.2B).

The identification of this sectored individual provided the first phenotypic evidence that *hth* plants were capable of producing somatic sectors. This finding suggested that perhaps some of the wildtype revertants originally found among *hth* mutant progeny might have arisen from genetically heterozygous sectors on the parent plant (Lolle et al., 2005). Since

well over 300,000 mutant plants were screened in the course of our out-crossing experiments and only one plant with a very large phenotypically wildtype sector found such as that shown in Figure 2.2B, we reasoned that if sectoring does occur, the vast majority of sectors would be too small to result in a visible phenotype. This possibility prompted us to test whether novel genotypes could be detected in tissue samples obtained from single *hth* plants.

2.3.2 **Single plants can have multiple genotypes**

For these experiments we chose to focus exclusively on molecular markers consisting of genomic DNA sequence tracts between 45-94 nucleotides in length that are either present or absent in the Columbia and Landsberg Arabidopsis accessions (insertion-deletion polymorphic indel markers or; Figure 2.1). In choosing to use indel markers we reasoned that deletions would be recalcitrant to enzyme repair or modification and therefore would help differentiate between enzyme-based mechanisms such as the one put forth by Comai and Cartwright (Comai and Cartwright, 2005) and a template-directed mechanism like the one previously proposed (Lolle et al., 2005). Hybrid F1 plants were constructed between Columbia and Landsberg accessions by manual cross-pollination, F1 plants allowed to self-seed and F2 and F3 descendants used as experimental material. The Columbia accession was always wildtype for *HTH* while *hth* mutant alleles, when introduced in hybrid lines, originated from the Landsberg genetic background. For all of the indel markers used in this study, Columbia is homozygous for the insertion.

Initially, F3 seed progeny derived from hybrid F2 parent lines with known indel marker profiles were screened to test whether or not these markers were stable. All F2 parent plants

were reared in plastic tubes to minimize outcrossing. When marker profiles were compared between *hth-4* parent plants and their F3 adult offspring, 2.16% [6/277] deviated from the expected profile. This frequency is approximately 5 times higher than baseline rates (0.02-0.43%) seen in outcrossing experiments described above. When F3 progeny were assayed as seedlings, similar frequencies were seen, with 2.5% [15/600] of the F3 seedlings showing discordant marker profiles. Altogether 600 seedlings were tested using a total of 30 seedlings per F2 plant (eleven *hth-4*, five *hth-7*, two *hth-8* and two *hth-10* F2 plants). Of the 15 F3 seedlings that tested positive for at least one non-parental marker, 7 had acquired insertions.

To test whether the observed genetic discordance between parent and offspring was due to sectoring, multiple tissue samples were collected from individual adult plants and indel marker profiles compared between these different samples. Molecular analyses confirmed that some tissue samples taken from individual *hth* mutant plants had novel marker profiles. For the plant shown in Figure 2.3A, seven out of eight samples scored homozygous for the Landsberg deletion marker as expected, however, one sample produced two amplicon products, one of which co-migrated with the Landsberg deletion allele while a second larger amplicon co-migrated with Columbia insertion allele.

To test whether sectors could be detected earlier in development, the molecular genotype of shoots and roots of single seedlings grown under sterile conditions were compared to one another. On the assumption that wildtype plants would not produce sectors, identical tests were also conducted on wildtype hybrid lines as negative controls. In the majority of cases, as expected, there was a perfect correspondence between the molecular

profiles of root and shoot. However, in some cases, individual seedlings were found to have molecular signatures that differed between the two organ systems (10/44 *hth-3*; 1/50 *hth-4;* 9/76 *hth-7*; Figure 2.3B). Surprisingly, wildtype hybrid seedlings also showed novel genotypes when roots and shoots from the same seedling were compared (10/184 wildtype hybrids; Figure 2.3B).

2.3.3 **Markers are discordant with parental DNA sequences**

A subset of amplicon samples were subjected to DNA sequence analyses in order to determine their molecular features. Sequence analyses of DNA clones derived from individuals where the non-parental amplicon co-migrated with the smaller deletion allele showed identity with the Landsberg deletion marker (Figure 2.4). In two instances, polymorphisms immediately upstream of the deletion were also detected (Figure 2.4A). As indicated, the Landsberg accession differs from Columbia at these exact three nucleotides. DNA sequence analysis of novel amplicons that co-migrated with the larger insertion allele showed that this seedling shoot had acquired a 54-nucleotide insertion that shares identity with the Columbia reference genome (Figure 2.4B). This same insertion was absent in the F2 parent plant. These particular seedlings descended from the same wildtype hybrid parent plant as the F3 progeny whose profiles are shown in Figure 2.3B.

2.3.4 **Sectors have complex genotypes**

To obtain an estimate of sector size, tissue samples were subjected to quantitative assays where the copy number of a genomic reference sequence immediately flanking the marker of interest was compared to the copy number of a sequence internal to that particular insertion

marker (Figure 2.5 and Figure 2.6). Hybrid plants verified to be homozygous for a deletion at specific indel markers were subjected to quantitative assays. The quantitative polymerase chain reaction (qPCR) data reveal two remarkable findings. First, the majority of tissue samples collected from individual *hth* mutant plants tested positive for the presence of at least one insertion marker (Figure 2.5). In addition, multiple insertion sequences could be detected in many of the tissue samples tested (Figure 2.5B). In most instances the copy number of any given insertion sequence, relative to the reference, was very low (less than one copy per 1000). Second, wildtype hybrid plants also showed evidence of sectors with novel genotypes (Figure 2.6). Only two out of four wildtype plants tested, however, showed evidence of novel insertions.

Figure 2.1 Haploid representation of the 5 Arabidopsis chromosomes indicating the relative locations of the 16 insertion-deletion polymorphic markers used in this study. Nine of the markers are intergenic (*). Marker names reflect clone designations. The size of the insertion sequence is indicated in base pairs (bp). The relative location of *HOTHEAD* (*HTH*) is shown at the bottom of chromosome 1.

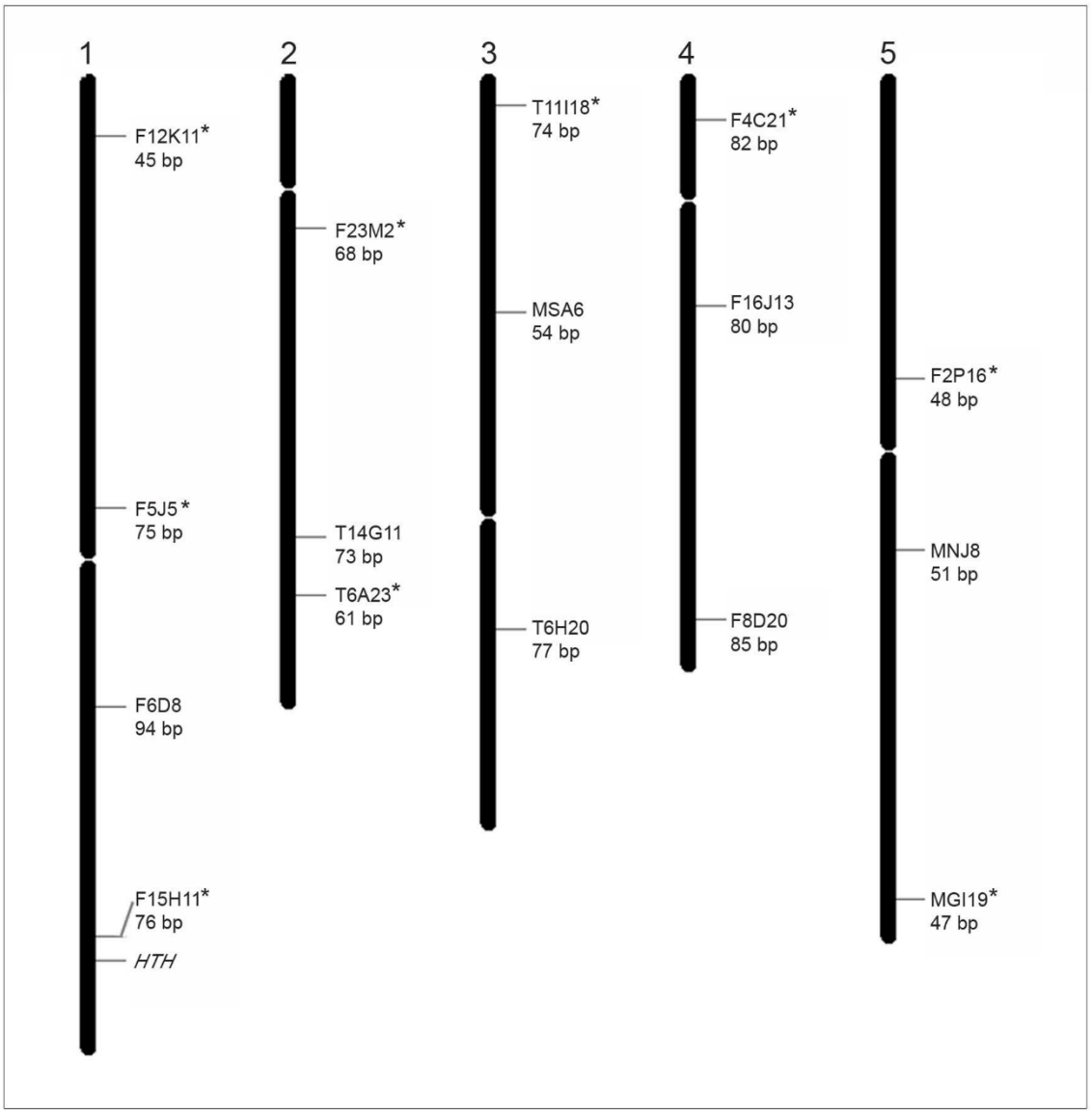

1
2
3
4
5
F12K11*
45 bp
F5J5*
75 bp
F6D8
94 bp
F15H11*
76 bp
HTH
F23M2*
68 bp
T14G11
73 bp
T6A23*
61 bp
T11I18*
74 bp
MSA6
54 bp
T6H20
77 bp
F4C21*
82 bp
F16J13
80 bp
F8D20
85 bp
F2P16*
48 bp
MNJ8
51 bp
MGI19*
47 bp

Table 2.1 List of primer pairs used for PCR-based molecular genotyping. Expected amplicon product sizes for the Columbia and Landsberg accessions are shown in adjacent columns.

Insertion-deletion Marker	Primer pairs	Columbia product size in base pairs	Landsberg product size in base pairs
F12K11	ccatatcttggagttggcaga tgtcttcaggaacacaacca	166	121
F5J5	tgaagatttcgtggaagcaa ctcatggatgcctaataccg	275	200
F6D8	ctccgtcttccagagtttga ttcgggtgattagtacggaaa	211	107
F15H11	atttgcggctgaaagacaag tgagtgtgtcatgagtgtttgttt	229	153
F23M2	taaagttgttggccgaggag tcggagatacccgagctaaa	231	163
T14G11	cctatgtgtcaagagagatttcca tttgttccatttataagcgtttctc	286	213
T6A23	aacaccaagtcaactgtttttgtt tcaaaataaacacccccaact	241	180
T11I18	ccccaattcgaaatgtaagg cgctccttgacagttttcct	203	129
MSA6	ctggggtgttctcacaggat cgttggaggtggtcttaggt	199	145
T6H20	tgcattggtttctctgcttg gggaaacctccatactcgaa	231	154
F4C21	tggttagggttctggtcagg agtggctcatcgttcgagat	195	113
F16J13	gaagcatgttttgtgtatcttgc ccgcatctccacatttcatt	224	144
F8D20	caccagacggtgatgaagag cattcgcgcatttattgttg	202	117
F2P16	aaaatggtttaccacatggaca tcccaaatcaattcaaggaaa	223	175
MNJ8	catggatcaaagatgatctcca ttcgcttttcgtgtttctga	184	133
MGI19	tgcacatgacttcaacagaaaa atgtgggtgggtgttgattt	203	156

Table 2.2 List of primer sets used for qPCR analyses. Primer positions, left and right primer sequences and expected amplicon sizes are indicated for each marker. Colours correspond to those used in Figure 2.5B and Figure 2.6B.

Insertion-deletion Marker	Primer Position	Left primer sequence	Right primer sequence	Product size in base pairs
F6D8	Positive control	ctgaccagcaaattctcaagg	tgagcaggtgaaacagatgg	766
	External reference	aagtttaaaacgaaaactttataaaatacc	tttcgtgttcgtggttttca	214
	Within insertion	aaacaagtgcatgttgcg	tttcgtgttcgtggttttca	266
F15H11	Positive control	ctccactaactcccgttattcc	gaacaatcgggccacatatag	701
	External reference	tttcgtcacttttcaaaactaac	gtgtgtgtgtgtgtgtgtgctc	151
	Within insertion	tgatgattttggattgaacgtc	gtgtgtgtgtgtgtgtgtgctc	201
T14G11	Positive control	gagttgtgttccagggccta	tttgttgtgcgaattcattg	897
	External reference	cacaaaaattaaggaataataaatgttctc	tttgttccatttataagcgtttctc	143
	Within insertion	ttgtcccattttatttgatgtttg	tttgttccatttataagcgtttctc	176
T6H20	Positive Control	tttcctgtttgggatctgag	tcaggagatagtccaccatgc	839
	External reference	tgggcttaccctgttcatggag	gcagagaaaccaatgcattttca	151
	Within insertion	tgggcttaccctgttcatggag	ccagaaaccgagtctctaagatttca	259
MGI19	Positive control	atatgcttgtcagtgagggaag	gaattcgacaggagcgtgaag	800
	External reference	gaacaatttgtggaaaaatggaa	cctagtttcatgtgcatatatgtc	181
	Within insertion	gaacaatttgtggaaaaatggaa	tgacatgtactcaccgcaatg	212

Figure 2.2 Molecular analysis of a mutant *hth-4* plant showing a large wildtype sector. (A) Two mutant branches (white boxes) flank a phenotypically wildtype flower branch (magenta box). Examples of normal wildtype (*HTH/HTH*) and mutant (*hth/hth*) flowers are shown on the right. (B) DNA was extracted from tissue samples and allele-specific PCR-based molecular markers used to determine genotype. The wildtype branch scored as heterozygous (*hth-4/HTH*), while mutant branches scored as homozygous for the *hth*-4 allele.

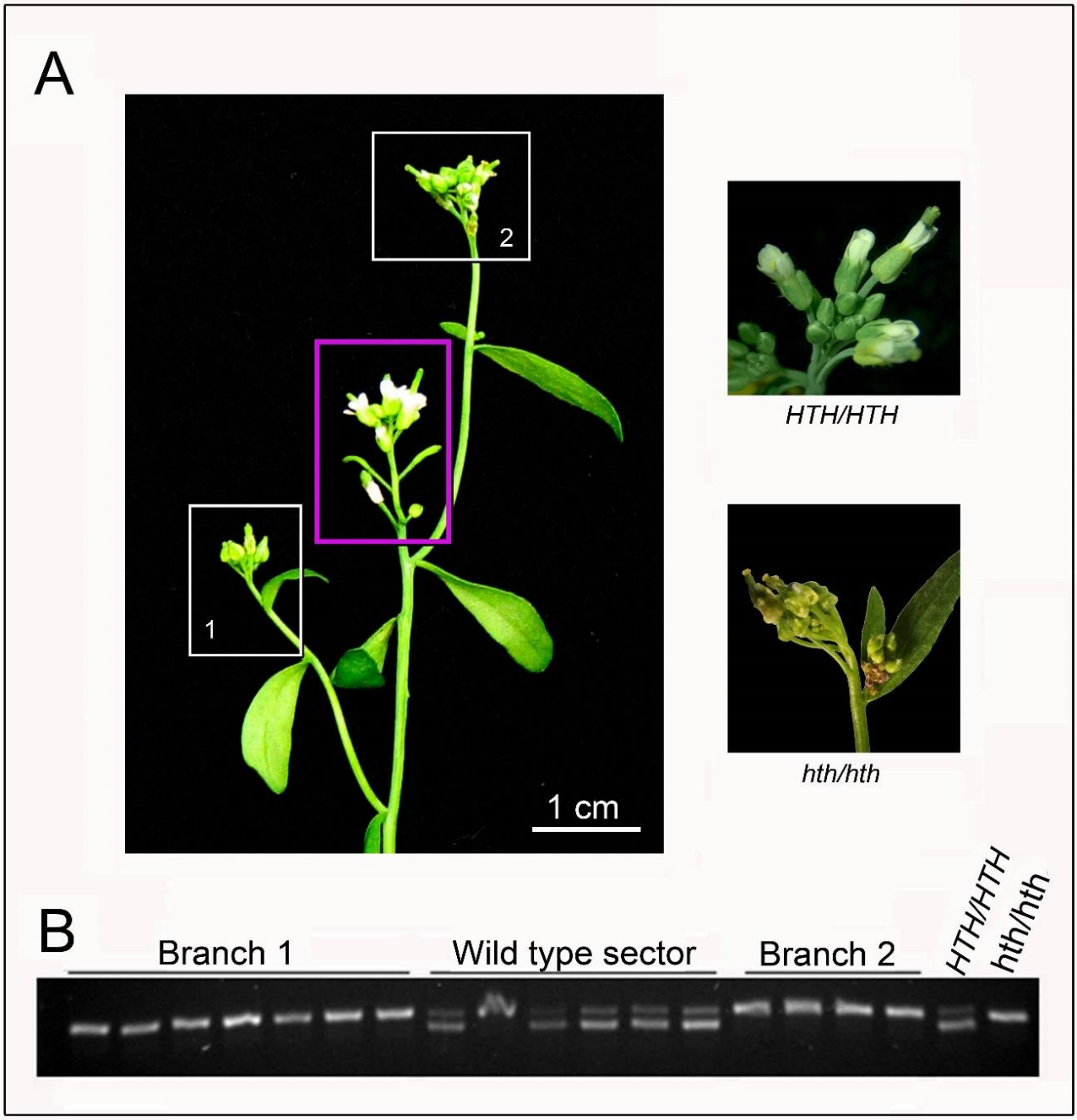

A
2
1
1 cm
HTH/HTH
hth/hth
B
Branch 1
Wild type sector
Branch 2
HTH/HTH
hth/hth

Figure 2.3 Molecular analysis of an adult mutant plant and bisected mutant and wildtype seedlings. (A) DNA was extracted from multiple tissue samples and PCR-amplified using F8D20 primers. A novel PCR amplicon product corresponding in size to the insertion allele (C) was detected in *hth-7* tissue sample 3 (arrow). (B) Sterile seeds were sown onto petri plates (top left) and 5-day old seedlings cut at the root-shoot junction (illustrated in the top right panel) and genotyped individually. DNA extracted from shoot (S) and root (R) samples derived from individual *hth-3* or wildtype seedlings were PCR-amplified using F12K11 and F4C21 primers, respectively. Samples were loaded in pairs (indicated by horizontal bars). Novel amplicon bands were detected in five seedling samples (arrows) that correspond in size to the insertion allele (C). In one *hth-3* sample, both organs (S, R) had a novel band, while a novel amplicon was detected only in the root in a second sample. In three cases, DNA extracted from wildtype seedlings gave rise to novel bands corresponding in size to the insertion allele (C) (arrows, S). In both cases, the parent plant was homozygous for the deletion allele (L) at the corresponding marker. Heterozygote (H), no DNA control sample (ND).

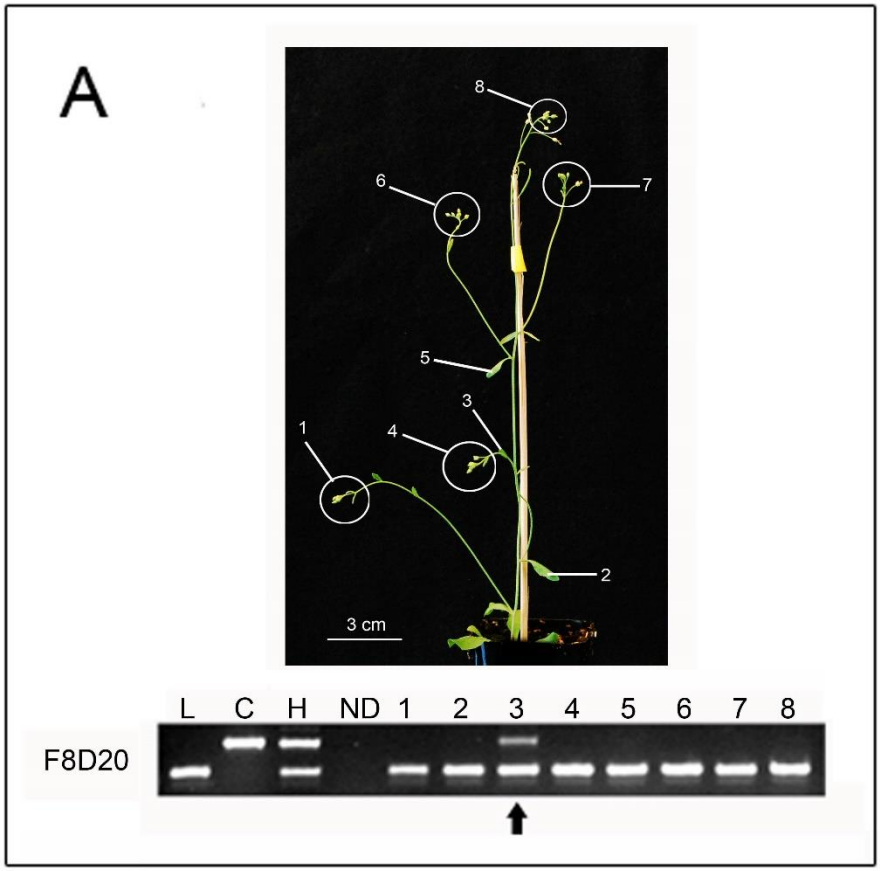

A
8
6
7
5
3
1
4
2
3 cm
L C H ND 1 2 3 4 5 6 7 8
F8D20
B
2 cm
2 mm
ND H C L S R S R S R S R S R S R S R S R
(hth-3)
F12K11
C L S R S R S R S R S R S R S R S R S R
(wt)
F4C21

Figure 2.4 DNA sequence alignments showing F8D20 and MSA6 indel loci. (A) The F2 *hth-3* parent (F2 *hth*) shares sequence identity with 2 of 3 DNA clones isolated from this single *hth-3* seedling (F3 R2 and F3 S1). DNA sequence data obtained from a root clone (F3 R1) shares identity with the Landsberg erecta sequence (Ler), including 3 flanking sequence polymorphisms (arrows) and a corresponding 85 base-pair deletion. The Columbia reference sequence (Col) is shown on the top line of the alignment. (B) The *HTH* wildtype hybrid parent (F2 wt) shares sequence identity with 2 of 3 DNA clones isolated from this single seedling (F3 S2 and F3 R1). DNA sequence data obtained from one shoot clone (F3 S1), however, reveals a 54 base-pair insertion sequence (junctions shown by arrows) and shares identity with the Columbia reference sequence (Col).

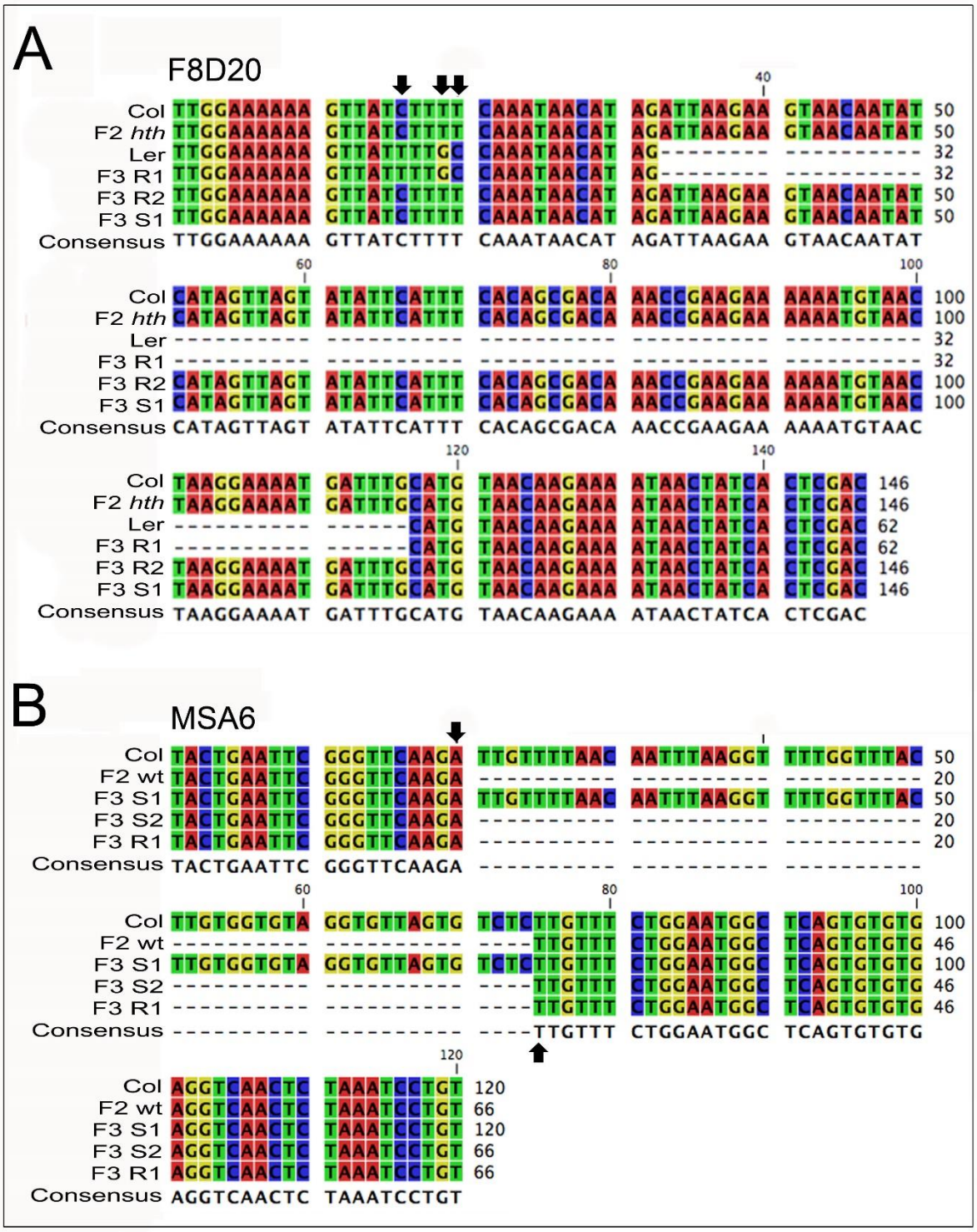

A
F8D20
40
Col TTGGAAAAAA GTTATCTTTT CAAATAACAT AGATTAAGAA GTAACAATAT 50
F2 hth TTGGAAAAAA GTTATCTTTT CAAATAACAT AGATTAAGAA GTAACAATAT 50
Ler TTGGAAAAAA GTTATTTTGC CAAATAACAT AG-------- ---------- 32
F3 R1 TTGGAAAAAA GTTATTTTGC CAAATAACAT AG-------- ---------- 32
F3 R2 TTGGAAAAAA GTTATCTTTT CAAATAACAT AGATTAAGAA GTAACAATAT 50
F3 S1 TTGGAAAAAA GTTATCTTTT CAAATAACAT AGATTAAGAA GTAACAATAT 50
Consensus TTGGAAAAAA GTTATCTTTT CAAATAACAT AGATTAAGAA GTAACAATAT
60 80 100
Col CATAGTTAGT ATATTCATTT CACAGCGACA AACCGAAGAA AAAATGTAAC 100
F2 hth CATAGTTAGT ATATTCATTT CACAGCGACA AACCGAAGAA AAAATGTAAC 100
Ler ---------- ---------- ---------- ---------- ---------- 32
F3 R1 ---------- ---------- ---------- ---------- ---------- 32
F3 R2 CATAGTTAGT ATATTCATTT CACAGCGACA AACCGAAGAA AAAATGTAAC 100
F3 S1 CATAGTTAGT ATATTCATTT CACAGCGACA AACCGAAGAA AAAATGTAAC 100
Consensus CATAGTTAGT ATATTCATTT CACAGCGACA AACCGAAGAA AAAATGTAAC
120 140
Col TAAGGAAAAT GATTTGCATG TAACAAGAAA ATAACTATCA CTCGAC 146
F2 hth TAAGGAAAAT GATTTGCATG TAACAAGAAA ATAACTATCA CTCGAC 146
Ler ---------- ------CATG TAACAAGAAA ATAACTATCA CTCGAC 62
F3 R1 ---------- ------CATG TAACAAGAAA ATAACTATCA CTCGAC 62
F3 R2 TAAGGAAAAT GATTTGCATG TAACAAGAAA ATAACTATCA CTCGAC 146
F3 S1 TAAGGAAAAT GATTTGCATG TAACAAGAAA ATAACTATCA CTCGAC 146
Consensus TAAGGAAAAT GATTTGCATG TAACAAGAAA ATAACTATCA CTCGAC
B
MSA6
Col TACTGAATTC GGGTTCAAGA TTGTTTTAAC AATTTAAGGT TTTGGTTTAC 50
F2 wt TACTGAATTC GGGTTCAAGA ---------- ---------- ---------- 20
F3 S1 TACTGAATTC GGGTTCAAGA TTGTTTTAAC AATTTAAGGT TTTGGTTTAC 50
F3 S2 TACTGAATTC GGGTTCAAGA ---------- ---------- ---------- 20
F3 R1 TACTGAATTC GGGTTCAAGA ---------- ---------- ---------- 20
Consensus TACTGAATTC GGGTTCAAGA ---------- ---------- ----------
60 80 100
Col TTGTGGTGTA GGTGTTAGTG TCTCTTGTTT CTGGAATGGC TCAGTGTGTG 100
F2 wt ---------- ---------- ----TTGTTT CTGGAATGGC TCAGTGTGTG 46
F3 S1 TTGTGGTGTA GGTGTTAGTG TCTCTTGTTT CTGGAATGGC TCAGTGTGTG 100
F3 S2 ---------- ---------- ----TTGTTT CTGGAATGGC TCAGTGTGTG 46
F3 R1 ---------- ---------- ----TTGTTT CTGGAATGGC TCAGTGTGTG 46
Consensus ---------- ---------- ----TTGTTT CTGGAATGGC TCAGTGTGTG
120
Col AGGTCAACTC TAAATCCTGT 120
F2 wt AGGTCAACTC TAAATCCTGT 66
F3 S1 AGGTCAACTC TAAATCCTGT 120
F3 S2 AGGTCAACTC TAAATCCTGT 66
F3 R1 AGGTCAACTC TAAATCCTGT 66
Consensus AGGTCAACTC TAAATCCTGT

Figure 2.5 Relative genomic copy number of insertion sequences in a *hth-7* mutant plant. (A) DNA was extracted from branches 1-7 of this *hth-7* mutant plant and amplified using qPCR or standard PCR reactions. (B) Graphical representation of qPCR results using four different indel markers (F8D6 (red), F15H11 (yellow), T14G11 (blue), and T6H20 (green). Coloured bars show the number of insertion sequences per 1000 copies of the reference sequence (lines indicate standard error of the mean, n = 3). All 7 samples showed novel insertion sequences. (C) Standard PCR-amplification using T6H20 primers showed amplicons that corresponded exclusively to the deletion allele (L). Primer positions (arrows) relative to the T6H20 indel (green box) are depicted to the right of the gel image. (D) Pooled amplicon product from T6H20 reference primers demonstrate that this region was amplified equally in all samples, as was the positive control (+). The reference sequence is upstream of the T6H20 insertion marker, as depicted on the right. (E) Quantitative PCR using a primer anchored within the T6H20 indel gave rise to amplicons that corresponded in size to the positive control (+). No product was amplified from sample six. T6H20 indel (green box), Columbia (C), Landsberg (L), heterozygote (H), no DNA control sample (ND).

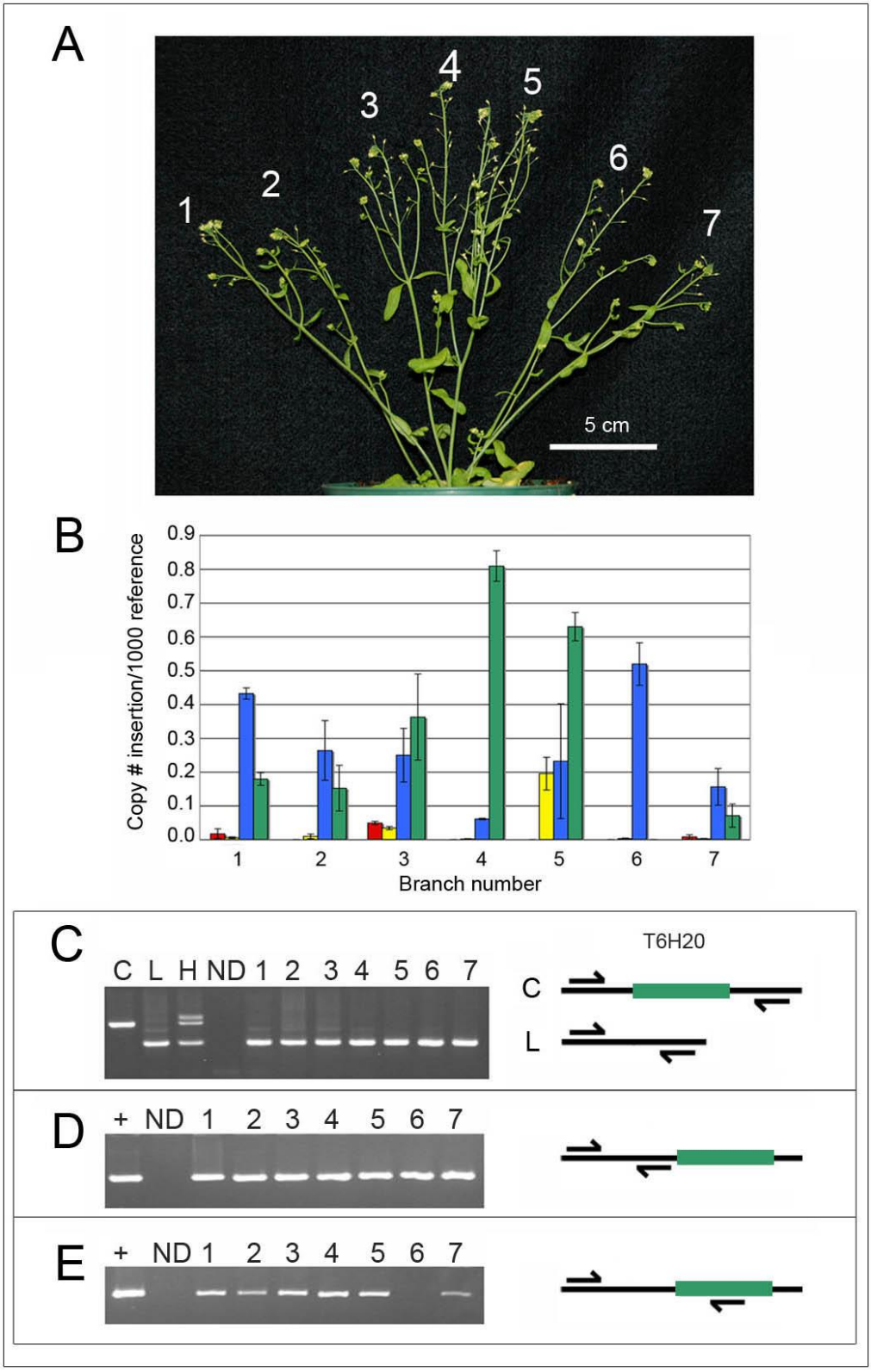
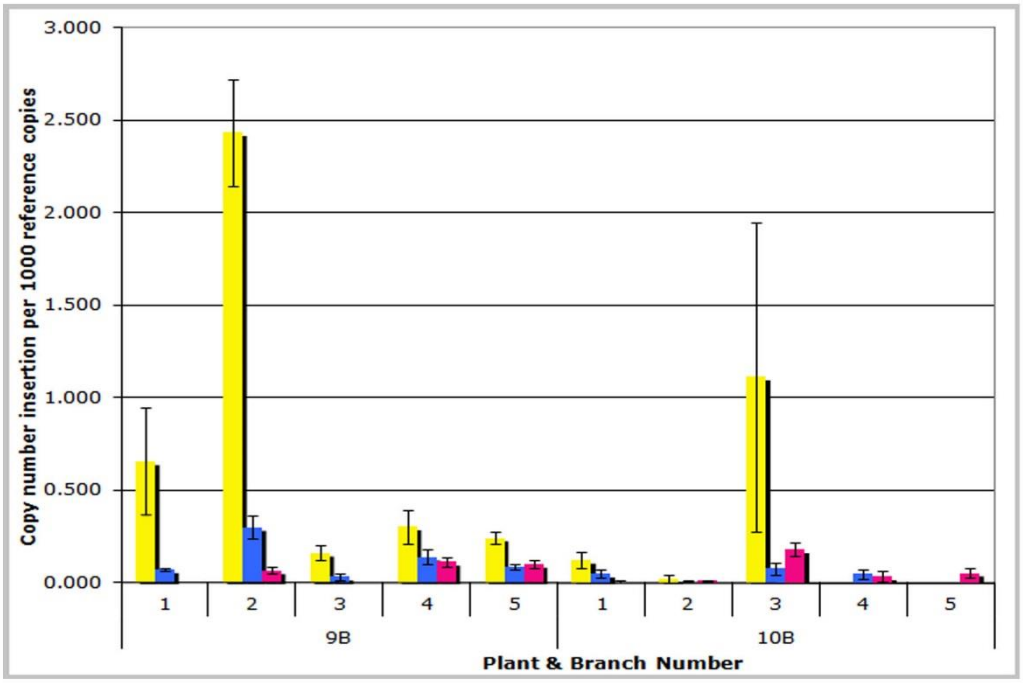

A
1
2
3
4
5
6
7
5 cm
B
Copy # insertion/1000 reference
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
1
2
3
4
5
6
7
Branch number
C
C L H ND 1 2 3 4 5 6 7
T6H20
C
L
D
+ ND 1 2 3 4 5 6 7
E
+ ND 1 2 3 4 5 6 7

Figure 2.6 Relative genomic copy number of insertion sequences in two wildtype plants. (A) DNA was extracted from branches 1-5 of two wildtype hybrid plants (9B and 10B) and amplified using qPCR. (B) Graphical representation of qPCR results using three different indel markers ((F15H11 (yellow), T14G11 (blue), and MGI19 (pink)). Coloured bars show the number of insertion sequences per 1000 copies of the reference sequence (lines indicate standard error of the mean, n = 3). Novel insertion sequences could be detected in all 10 samples.


A
1
2
3
4
5
9B
1
2
3
4
5
10B
5 cm

B


Copy number insertion per 1000 reference copies
3.000
2.500
2.000
1.500
1.000
0.500
0.000
1
2
3
4
5
9B
1
2
3
4
5
10B
Plant & Branch Number

2.4 **Discussion**

By employing classical genetic approaches in conjunction with low and high-resolution molecular methods, we show that one Arabidopsis plant can have multiple genotypes. We have found instances of intra-organismal variation in different genetic backgrounds, in plants reared in different growth chambers, at different developmental stages and under sterile growth conditions. Furthermore, the incidence of sectoring and genetic discordance appears to be in some way conditioned by the *hth* mutant background as we found a consistently higher frequency of genetic discordance within single *hth* plants as compared to *HTH* wildtype plants. This was also true for shoot and root systems compared between aseptically grown seedlings and for tissue samples taken from adult plants and subjected to qPCR. Of critical importance, in showing that single Arabidopsis plants are genetic mosaics, experimental error due to cross-pollination and seed contamination can be completely discounted. To the best of our knowledge, this is the first report that documents the spontaneous but targeted appearance of unique genomic insertions at multiple discreet loci in single plants.

Only two other cases of spontaneous genomic insertions have been reported in plants that similarly could not be explained by any previously known mechanism. In both cases the insertion was non-random and targeted a specific locus. In the case of flax, the insertion sequence was 5.7 kilobase (kb) pairs in size (Chen et al., 2005) while in rice the insertion was comparatively small, being only 34 base pairs in size (Gao et al., 2011). Our data suggest that these reported cases of spontaneous genomic insertion events, like the sequence changes

reported here, occur by a process intrinsic to the plant. As before, we propose the possibility that Arabidopsis plants harbor a cryptic store of sequence templates that can overwrite the parentally contributed genomes by a template-directed mechanism (Lolle et al., 2005).

If intrinsic drivers of genetic variation exist in inbreeding plant species, have additional incidents of cryptic genetic variation been documented in other systems? We believe that in soybean and cauliflower such events have indeed been reported and presented as cases of enigmatic phenotypic variation (Chable et al., 2008; Fasoula and Boerma, 2005; Fasoula and Boerma, 2007). In other studies, molecular data have been featured. Again in flax, for example, molecular assays have demonstrated that heritable phenotypic changes induced by environmental shifts are accompanied by reproducible locus-specific copy number changes in genomic DNA (Chen et al., 2009; Chen et al., 2005; Schneeberger and Cullis, 1991). In soybean, reproducible non-random changes in restriction length polymorphic markers induced by *in vitro* (Folse and Roughgarden, 2012; Thomson et al., 1991) culturing of root explants have also been documented (Roth et al., 1989). Genomic changes manifesting similar hallmarks of biased sequence alterations have also been described in rice (Gao et al., 2011; Xu et al., 2007) and corn (Tracy et al., 2000) hybrids, as well as in Arabidopsis (Jiang et al., 2011; Yi and Richards, 2008; Yi and Richards, 2009).

In long-lived arborescent plants, intra-organism genetic variation has been demonstrated in a variety of systems (Lopez et al., 2010; Thomson et al., 1991; Whitham and Slobodchikoff, 1981). The fitness benefits have also been validated using models that test whether the production of genetically divergent modules is an effective strategy for achieving

adaptive co-evolution with organisms that feed on or infect the plant (Boyko and Kovalchuk, 2011; Folse and Roughgarden, 2012; Pineda-Krch and Lehtila, 2004). Models testing fitness benefits of module-level selection show that this is an effective strategy for achieving adaptive co-evolution between long-lived trees and short-lived herbivores when individual tree branches diverge genetically (Folse and Roughgarden, 2012). Furthermore, this held true across a range of assumptions, even when reproduction was predominantly asexual. However, the fitness benefits were only fully realized for sufficiently long-lived trees that experienced strong selection (Folse and Roughgarden, 2012). This fitness paradox is not exclusive to plants but also is relevant to organisms outside of the plant kingdom that have remained evolutionarily robust even though reproduction is predominantly asexual (Pineda-Krch and Lehtila, 2004).

For a short-lived organism such as Arabidopsis, what adaptive value would within-organism genetic variation have? One possibility is that this heterogeneity offsets the predicted decline in genetic variation that should result from inbreeding. Plant development is open-ended and reiterative, allowing for the continuous output of repetitive units or modules that function to support the growth and reproduction of the individual. When combined with developmental plasticity and the absence of a sequestered germ line, modular development may actually drive plants toward becoming genetically heterogeneous (Fagerstrom et al., 1998; Lopez et al., 2010; Pineda-Krch and Fagerstrom, 1999; Pineda-Krch and Lehtila, 2004). As posited by Whitham and Slobodchikoff (1981), somatic sector formation permits the introduction of genetic variants into the gene pool either through vegetative propagation or through sexual reproduction. As these authors point out, germ line

cells are derived from somatic tissues that arise late in the developmental history of the plant and therefore somatic mutations are more likely to introduce genetic variation than mutations that arise in the gametes (Sangster et al., 2008; Satina and Blakeslee, 1941; Whitham and Slobodchikoff, 1981). By expanding the window of tolerance for genetic variation, plants may be afforded a better adaptive strategy given lifestyle constraints. The versatility of modular development combined with tolerance for genetic variation may allow plants to adapt at rates tailored to pathogen life cycles (Todesco et al., 2010) or to relatively expanded time scales, such as those affecting climate change. Even though self-fertilization is thought to have evolved approximately one million years ago (Tang et al., 2007a), Arabidopsis plants have not suffered the consequential genetic erosion but have continued to thrive.

In addition to benefiting from a natural tendency toward genetic heterogeneity, the plant genome itself is thought to buffer the cost of having limited genetic diversity. In wild relatives of Arabidopsis the genome is thought to be highly dynamic and to respond to changes in environmental conditions or other extrinsic factors (Boyko and Kovalchuk, 2011; Yao and Kovalchuk, 2011). Genome responses include elevated rates of homologous recombination that persist for multiple generations (Molinier et al., 2006), changes in copy number (DeBolt, 2010) and modulation of epigenetic gene regulation (Lang-Mladek et al., 2010). Pervasive genetic buffering (Queitsch et al., 2002; Sangster et al., 2008) ensures that phenotypes with potentially deleterious consequences are attenuated. In addition to the genome responses listed above, our findings suggest that an intrinsic source of genetic variation can be leveraged to enhance the diversity in genetic output achieved by Arabidopsis plants.

In considering alternate template-dependent mechanisms, such as gene conversion or homologous recombination, none can account for the *de novo* appearance of unique sequence insertions. Nevertheless, it is possible that the insertion or deletion of small DNA sequence tracts, as described here, could reflect the activity of transposable elements (Lisch, 2009; Tenaillon et al., 2010). However, numerous lines of evidence argue against this possibility. For instance, when novel amplicons were detected, they co-migrated with their corresponding insertion or deletion allele and did not show size heterogeneity, as would have been expected for transposon-driven excision or insertion events. Sequence data confirm that deletion events reproducibly eliminate a fixed length of sequence while insertion events reproducibly introduce a fixed sequence tract and both events repeatedly target precise genomic sites. Insertion and deletion events do not appear to produce obvious junction sites with altered nucleotides. Similarly, insertion events introduce sequences that share identity with the Columbia reference genome and do not appear to be chimeric gene or genome fragments. Furthermore, transposable element-mediated events cannot account for the fact that these insertion sequences appear to be generated *de novo* since no comparable conserved region of homology exists elsewhere in the host genome, as demonstrated by our qPCR data. Lastly, as determined by DNA database searches, none of the indel markers used in this study share significant sequence homology with annotated Arabidopsis transposable elements.

If the genome of an intensely studied model organism such as Arabidopsis is subject to modification by the template-directed mechanism we propose, why has this phenomenon not been described previously? Our research shows that target choice and methodological

approach are critical in differentiating these genomic events from other processes that also modify DNA sequences. Based on our findings, the only genomic targets that are truly diagnostic of this phenomenon are deletions. To the best of our knowledge, deletions alleles have been used in genetic studies precisely because they are known to be stable and not to revert but have not been used to study phenomena related to epigenetic inheritance. There is no generalized precedent for genetic instability of deletions and assuming otherwise would go against an established biological paradigm. Polymorphic molecular markers such as single nucleotides, simple sequence repeats, or insertions that are subject to alterations by other processes will not provide sufficient resolution to differentiate mechanism, even though they are also likely targets for this process. In particular, our findings may explain why genome sequencing efforts have failed to register these sequence deviations or, if detected, why they may have been attributed to sequencing error and eliminated during curation. One possibility that immediately emerges from this prediction is that raw sequence data contained in existing genome database archives may already contain evidence of extra-genomic sequence information, revealed by features such as highly biased loci-specific “errors”.

Collectively, our genetic and molecular data show that many, and perhaps most, insertion events occur somatically in both seedlings and adult plants. Sectoring may therefore be a constitutive process that takes place throughout development but may be limited such that, at any given time, only a few cells host these genetic changes. Importantly, this may explain why sequence changes seen in revertant *hth* progeny have rarely been found to affect both alleles. Although sexual transmission of non-parental markers clearly does occur (Lolle et al., 2005), the fact that we have not found *HTH/HTH* progeny among seed-derived

offspring suggests that sectors populating the gamete forming lineages are unstable or very rare. The qPCR data are consistent with this supposition. However, it is also possible that mechanistic differences exist between somatic and germ line tissues or that insertion events remain dynamic, limiting sexually transmitted changes to those that stabilize. It is also possible that certain genetic backgrounds condition this process as suggested by the greater number of events detected in *hth* mutants.

In addition to validating our genetic and molecular data, the qPCR results extend those findings and suggest that the genetic makeup of individuals can be surprisingly complex. Our data show that each plant can produce multiple discreet sectors, at many different growing points and each with unique marker profiles. This finding implies that sectoring may be a relatively common occurrence, even in wildtype genetic backgrounds. Since the adult plants used for these experiments were left largely intact and only a small proportion of the plant sampled, many more sectors may have been present than quantified. As such, it is possible that our current census underestimates the frequency with which these smaller islands of genetic variation arise. Although sectors are more readily detected using qPCR, this method cannot distinguish, for example, between copy number variation within a small cluster of cells versus multiple cells that remain strictly diploid and are clonally related. Similarly, it is not possible to distinguish whether one sector hosts the full complement of genetic sequence changes, whether independent events occur in multiple discreet sectors, or if sectors overlap. Visualization of sectors in living tissue or tissue sections should help distinguish between these possibilities.

In addition to models demonstrating the fitness benefits of module-level selection (Folse and Roughgarden, 2012), computational models provide surprisingly strong support for an ancestrally based “error-correcting” mechanism such as the one we propose to exist in Arabidopsis plants (FitzGerald et al., 2010). In these constrained-optimization simulations, the evolutionary benefit of “genetic repair” strategies was compared between populations that access repair templates derived either from parents, grandparents or great-grandparents. Interestingly, a grandparent- or great grandparent-based genetic repair strategy is strongly favored over parental repair strategies. Furthermore, simulation results show that using a randomly selected template consistently gave superior results to those achieved using templates from the fittest parent or grandparent. From a biological perspective, such a strategy has considerable merit. Retaining a cache of templates derived from grandparental lineages would guarantee greater allele diversity precisely because the reservoir of allele variants would be deeper and allele redundancy would be less likely to occur. Random selection of templates would be the most parsimonious strategy to affect genome repair, again because it would promote diversity across alleles and between individuals. Since only those individuals that survived in previous generations would contribute to these cached templates, represented alleles would be biased to those that have proven robust under a spectrum of selective pressures.

2.5 **Conclusions**

The research presented here brings to light five striking findings. First, individual Arabidopsis plants are capable of producing somatic sectors during the course of normal

vegetative development. Second, those sectors can have distinct and unique marker profiles and can differ in single nucleotide composition, can acquire small DNA insertions or can experience DNA sequence loss. Third, the *de novo* appearance of genomic insertions supports our original contention that cryptic sequence templates drive some of these changes (Lolle et al., 2005). Fourth, this phenomenon can be detected in wildtype genetic backgrounds raising the possibility that many Arabidopsis lab strains may be genetic mosaics. Finally, this process is genome-wide, impacting all 5 chromosomes, whether or not the target loci reside within genes or between genes.

Our data expand on the ideas put forth by Whitham and Slobodchikoff (1981) and suggest that sector formation, even in a short-lived organism like Arabidopsis, may be a normal part of development and, furthermore, that the formation of sectors serves to capture novel genetic variation, irrespective of the source of that variation. Models testing the benefit of within organism genetic heterogeneity suggest that the average fitness of the population increases if some individuals within that population are genetic mosaics (Folse and Roughgarden, 2012). As our data show, not all individuals in the populations we tested showed evidence of genetically distinct sectors but for those individuals that did, the number of sectors varied greatly. Our findings raise the possibility that inbreeding plants and, perhaps other organisms that predominantly propagate asexually, may sequester cryptic sources of genetic variation that can be harnessed to promote greater genetic diversity.

Chapter 3 Localization of the Arabidopsis HOTHEAD protein: insights into the protein function

3.1 Introduction

The plant epidermal cuticle is key to plant-environment interactions and protects plants from adverse environmental factors such as dehydration, excessive radiation, heat or cold stress, and attacks by herbivores (Kerstiens, 1996). The plant cuticle is a heterogeneous layer consisting of cutin, polysaccharide microfibrils, and waxes (Figure 1.1). Cutin is an insoluble biopolymer that is mainly composed of C16 and C18 ω-hydroxylated fatty acids interlinked via ester bonds (Table 1.2 and Figure 1.2). Waxes, which are a mixture of very-long-chain fatty acids (VLCFA) and their derivatives, can be deposited on the surface (epicuticular waxes) of or embedded (intracuticular waxes) within the cutin matrix (Domínguez et al., 2011; Koch and Ensikat, 2008; Kolattukudy, 2001; Nawrath, 2006).

Forward genetic screens have identified a number of genes that are important to cuticular function. Altered cuticle integrity such as that seen in the Arabidopsis *eceriferum* (*cer*), *lacerata* (*lcr*), *fiddlehead* (*fdh*), *wax, long chain fatty acid-CoA synthetase* (*lacs*) and *hothead* (*hth*) mutants, reveals that the cuticle also serves to maintain organ integrity and separation during normal development (Chen et al., 2003; Lolle et al., 1998; Lolle and Pruitt, 1999; McNevin et al., 1993; Pruitt et al., 2000; Schnurr et al., 2004; Wellesen et al., 2001). Perhaps not surprisingly, many of these genes are involved in the biosynthesis of cuticle components. For example, *FDH* (Pruitt et al., 2000) and *CER10* (Zheng et al., 2005) are both known to encode enzymes required for the biosynthesis of VLCFA that are constituents of epicuticular waxes. On the other hand, *LCR* and *LACS2* encode enzymes that specifically catalyze reactions in cutin biosynthesis. The LCR protein is a CYP86A type of cytochrome

P450 monooxygenase that catalyzes ω-hydroxylation of fatty acids needed for cutin precursor production (Wellesen et al., 2001), whereas the LACS2 is likely a protein required for fatty acyl-CoA formation in the first step in cutin monomer synthesis (Schnurr et al., 2004). In addition, many of these cuticle mutants also exhibit elevated permeability in the seed coat. For example, the *defective in cuticular ridges* (*dcr*) and *bdg1*mutant seed coats were more permeable allowing more toluidine blue staining. In addition, mutant *dcr* seeds were often deformed and occasionally fused showing diminished release of mucilage upon imbibition (De Giorgi et al., 2015; Panikashvili et al., 2009).

The focus of this study, the *HTH* gene, is important to the cuticular function as mutations in the gene result in plants that have a more permeable cuticle and undergo floral organ fusions. Fusion severity varies among different alleles and is dependent upon the ecotype background in which they are expressed (Krolikowski et al., 2003; Lolle et al., 1998). Regardless of the fusion phenotype, all *hth* mutants have the capacity to self-fertilize, and therefore the mutations can be maintained in homozygous state although seed yield varies with the severity of floral organ fusion. Unlike *lacs2* and *lcr*, *hth* mutant plants rarely display fusion of rosette leaves, and the integrity of the cuticle layer is only moderately impaired (Bessire et al., 2007). Fatty acid analysis of mature rosette leaves from *hth-12* mutant plants showed a reduction in some types of dicarboxylic acids, and this altered cuticle monomer profile led to the supposition that HTH is involved in cutin monomer synthesis and functions as an ω-alcohol dehydrogenase (Kurdyukov et al., 2006b).

In addition to the protective role, the cuticle layer also serves as a differentially permeable barrier that plays multiple regulatory roles in plant reproduction and in plant-pathogen interactions. For instance, the cuticle regulates pollen-pistil interactions by providing a receptive surface for adhesion, hydration and germination of compatible pollen. For mutants with altered cuticular permeability, pollen germination has been shown to occur on organs other than the stigma. Wildtype Arabidopsis pollen can hydrate and grow pollen tubes on vegetative organs of plants with higher cuticle permeability; these include mutant *cer*, *fdh*, *hth*, and fungal cutinase-expressing transgenic plants (Lolle and Cheung, 1993; Lolle et al., 1998; Sieber et al., 2000; Takahashi et al., 2010). In many cases, a more permeable cuticle layer increases susceptibility to biotic and abiotic stress as demonstrated in *ltpg1*, *cer4* and *dcr* (Jenks et al., 1995; Lee et al., 2009b; Panikashvili et al., 2009). However, the contrary has been observed. For example, Tang et al. (2007b) showed that loss of *LACS2* function increased Arabidopsis plants' sensitivity to a virulent *Pseudomonas syringae* strain as well as water and salt stresses. On the other hand, these *lac2* mutants also showed improved resistance to a virulent strain of the necrotrophic fungus *Botrytis cinerea*. It was hypothesized that certain cutin-related fatty acids may function as signal molecules, and disrupting their synthesis could lead to changes in recognition of and interactions with pathogens (Tanaka et al., 2001; Xiao et al., 2004).

Stress can lead to changes in plant physiology and metabolism that cause protein misfolding and degradation in the endoplasmic reticulum (ER) (Deng et al., 2013), a highly organized network composed of tubules and cisternae. Environmental stresses can induce the formation of ER-derived organelles such as protein bodies, precursor-accumulating vesicles

and ER bodies (Matsushima et al., 2003b; Sherameti et al., 2008). Methyl jasmonate (MeJA) treatment and wounding can similarly induce ER body formation or increase their number in Arabidopsis, therefore linking this organelle with plant stress responses (Hayashi et al., 2001; Ogasawara et al., 2009).

Although two previous studies have investigated *HTH* expression patterns, results were somewhat inconsistent (Krolikowski et al., 2003; Kurdyukov et al., 2006b). To the best of my knowledge, this current study of HTH protein localization is the first to use a native *HTH* promoter to drive expression of a fluorescent protein-tagged HTH fusion protein (HTH-FP). By analyzing these transgenic reporter lines, I determined the tissue and cellular localization profile of the HTH-FP protein. Based on this work, the HTH protein appears to be present in seedlings, floral tissues, ovules and developing seeds. HTH-FP localization to the integument has not been reported previously, and led to a more in-depth analysis of the *hth* mutant seed phenotype that in turn revealed changes to seed morphology and seed coat permeability. At the cellular level, HTH-FP appeared to be localized to the ER network and stress-associated ER-derived bodies. To test whether stress regulates *HTH* gene expression, *HTH* expression levels were also investigated in non-transgenic lines using quantitative RT-qPCR following treatment with methyl jasmonate.

3.2 **Materials and Methods**

3.2.1 **Plant material and growth conditions**

Wildtype and mutant lines used in this study include Columbia (Col), Landsberg *erecta* (L*er*), Wassilewskija (Ws), *hth-1, hth-4, hth-5, hth-7, hth-9, hth-13* (SALK_019460), *hth-14* (SALK_024611) and *hth-15* (SALK_141882). Arabidopsis seeds of homozygous Col, L*er*, *hth-13, hth-14* and *hth-15* were acquired from the Arabidopsis Biological Resource Center (Columbus, Ohio, USA), and the other mutant alleles were developed by Lolle et al. (1998). Arabidopsis seeds were sown onto moistened potting mix (1:1 mixture of LC1:LG3 Sungro Sunshine potting mixes, Sungro Horticulture, Alberta, Canada) either in flats or 5 cm pots and stratified at 4°C for two to five days before being moved to the growth chamber. Growth chambers (Econoair AC60, Ecological Chambers Inc., Winnipeg, Canada, MB; GC8-VH/GCB-B, Environmental Growth Chambers, Chagrin Falls, Ohio, USA; Conviron PGW36/E15, Controlled Environments Ltd., Winnipeg, Manitoba, Canada) were illuminated with a mixture of incandescent and fluorescent lights (140-170 μmol m^{-2} sec^{-1} at pot level) with a 24-hour photoperiod and maintained at 20 ± 4°C at 40 - 60% relative humidity. Plants were watered as needed. Reproductive tissues of mature plants were used for microscopy, and juvenile plants were harvested 11 days post-germination for the methyl jasmonate MeJA treatment.

To grow plants in a sterile condition, seeds were placed in open microcentrifuge tubes 5-6 layers thick and exposed to Cl_2 gas for 1.5 hours in an air-tight chamber. The Cl_2 gas was generated by mixing 100 mL of bleach (Javax, 5.25% NaOCl) with 4 mL of 1N HCl in a

beaker. Sterilized seeds were sprinkled on an agar medium containing half strength MS basal salts (Sigma-Aldrich, St. Louis, Missouri, USA) at a density of 10-15 seeds/plate and stratified at 4°C before being transferred to growth chambers. Four-day-old seedlings and two-week-old plant grown in the petri plates were used for microscopy. For colocalization studies, seedlings containing the *erRFP* construct were selected on half strength MS agar medium containing 20 µg/ml hygromycin B (BS725; Bio Basic Canada, Markham, Ontario, Canada). After stratification, the seeds were exposed to light for 12 hours to promote germination prior to growth in the dark for five days. Seedlings exhibited elongated hypocotyl indicated hygromycin resistance and were observed for erRFP localization.

3.2.2 **Bioinformatics analyses**

Putative *HTH* expression patterns were analyzed using the microarray-based expression data (www.bar.utoronto.ca). Homologues of HTH were identified using the Basic Local Alignment Search Tool (BLAST) (www.ncbi.nlm.nih.gov). The COBALT tool (www.ncbi.gov) was used to look for similar protein sequences and to create multiple alignments. Basic characteristics of the HTH protein were acquired from UniProtKB/Swiss-Prot at Expasy (www.expasy.org). Protein composition and structure were predicted using the PROFsec tool (www.predictprotein.org). Globularity was analyzed by GLOBPLOT (www.globlot.eml.de). Promoter analysis was performed using the CISTOME tool of the BAR database (Bio-Array Resource database; http://bar.utoronto.ca). SignalP (www.cbs.dtu.dk/services/SignalP) was used for predicting subcellular localization.

3.2.3 *HTH_{pro}:HTH-FP* transgene constructs and generation of transgenic plants

The 4.8 kilobase (kb) genomic sequence containing the 2009 base pair (bp) fragment upstream of the start codon of *HTH* and full length genomic *HTH* sequence excluding the stop codon was amplified from purified genomic DNA (Appendix A) of *Arabidopsis thaliana* (Landsberg) using Phusion Hot Start II DNA Polymerase (Thermo Scientific, Waltham, Massachusetts, USA) using the forward 5'-AGAGGAGAGAAACAAAGAATCTTCTTACT-3' and reverse 5'-AACACCAGCTTTGTTTCCAAGT-3' primers. The resulting target PCR product was integrated by topiosomerase-mediated cloning into the pCR8/GW/TOPO vector (Invitrogen, Carlsbad, California, USA). Entry clones containing the *HTH* coding sequence, pENTR-*HTH*, were digested with NheI, and the complete attL-flanked fragment recombined into the pGWB640 and pGWB650 vectors (Nakagawa et al., 2007) (Appendix B) using LR clonase (Invitrogen). Resulting expression constructs, *HTH_{pro}:HTH-EYFP* and *HTH_{pro}:HTH-G3GFP*, were transformed into *Escherichia coli* DH5-α cells. Plasmids containing the Nhe1 *HTH* genomic fragment were selected on spectinomycin Luria broth (LB) plates, and the *HTH*-containing plasmids subjected to DNA sequencing. Expression constructs were subsequently transformed into *Agrobacterium tumefaciens* strain GV3101.

A. tumefaciens-mediated transformation of *A. thaliana* plants was accomplished by using the floral dip technique (Bechtold and Pelletier, 1998). The amplified *HTH* wildtype genomic sequence with the 5’ upstream region were cloned into pGWB640 and pGWB650 (Nakamura et al., 2010) to generate recombinant constructs *HTH_{pro}:HTH-EYFP* and *HTH_{pro}:HTH-G3GFP*, respectively. The empty vectors (referred to as ‘EV’) were used as

negative controls. Constructs, and corresponding empty vectors were transformed into homozygous *hth-9* mutant (in the Ws background). Two to five independent T1 plants were isolated and characterized for each construct. Homozygous T2 plants were identified by testing for segregation of glufosinate resistance in their T3 progeny. These homozygous transgenic lines were used as the material for microscopy and immunoblotting experiments.

The *erRFP* construct was provided by Dr. Jaideep Mathur at the University of Guelph. Sequences of monomeric RFP sequence with an N-terminal Arabidopsis chitinase signal peptide sequence and C-terminal HDEL ER retrieval signal were cloned in the binary vector pCAMBIA and expressed under a 35S CaMV promoter (Sinclair et al., 2009). The *erRFP* construct was transformed using *A. tumefaciens* into Ws wildtype plants, HTH_{pro}*:HTH-EYFP*, and HTH_{pro}*:HTH-G3GFP* lines to generate the double transgenic lines for co-localization studies. The erRFP protein targets the ER network and ER-derived organelles called ER bodies. Transgenic plants containing the *erRFP* construct were selected by hygromycin B resistance. Two to five independent T1 plants of each transgenic line were evaluated.

3.2.4 **Permeability Assays**

Cuticle permeability was quantified by monitoring the rate of chlorophyll diffusion as described in Lolle et al. (1997). Whole, undamaged cauline leaves from approximately four-week old plants were collected, immersed in 80% ethanol and gently agitated. Aliquots of the ethanol solution were removed at 20, 40, 60, 90, 120, 160 minutes as well as at 24 hours following tissue immersion and absorption determined spectrophotometrically. The

chlorophyll extraction rate was determined by standardizing to the concentration of the *hth-9* mutant sample after 24 hours (maximum extraction). Chlorophyll content in each sample was determined using absorption readings at 647 and 664 nm using a Cary 100 UV-Vis spectrophotometer (Agilent Technologies, Santa Clara, California, USA). The experiment was repeated three times. Data were processed and graphed using Sigma Plot (Systat Software, San Jose, California, USA).

Toluidine blue staining was used to visualize the difference of cuticle permeability. Whole, fresh flowers were incubated in a solution of 0.025% (w/v) toluidine blue (89640; Sigma-Aldrich) in ¼ Luria Broth (LB; (w/v) 1.0% tryptone, 0.5% yeast extract, 1.0% NaCl at pH 7.5) for 1 hour followed by rinsing with water for 10 minutes. Rosette leaves of 10-day-old plants were incubated with 5-µl droplets of a 0.025% (w/v) solution of toluidine blue for 2 hours and rinsed with water (Bessire et al., 2007).

Seed coat permeability was assessed by the tetrazolium assay. An aliquot of 50 mg of dried Arabidopsis seeds was incubated in 1 ml of 1% (w/v) tetrazolium red (Triphenyltetrazolium chloride; T8877, Sigma-Aldrich) solution at 30°C for 48 hours in darkness. Red-coloured formazans are produced if tetrazolium permeates the seed coat and comes into contact with living tissue.

3.2.5 SDS-polyacrylamide gel electrophoresis and protein immuno-detection

Arabidopsis tissue used for protein extraction was flash frozen in liquid nitrogen immediately after harvest. The tissue was ground to a fine powder using a mortar and pestle, or pulverized by vortexing frozen tissue in sealed 2 mL tubes containing 2 mm stainless steel beads (1/8”

diameter; Abbott Ball Company, West Hartford, Connecticut, USA). Extraction buffer (100 mM Tris-HCl pH 8.0, 8M urea, 5mM EDTA, 2.5% (w/v) SDS, 10% (v/v) glycerol, 1mM PMSF, 100 mM DTT and protease inhibitor cocktail (P9599; Sigma-Aldrich, St. Louis, Missouri, USA) was added and samples vortexed for 2 minutes, followed by centrifugation to pellet cell debris. The supernatant was collected and the total protein concentration determined using Bio-Rad Quick Start™ Bradford 1x Dye Reagent (Bio-Rad Laboratories, Hercules, California, USA), according to the manufacturer's instructions. Protein was solubilized in Laemmli Sample Buffer (60 mM Tris-Cl pH 6.8, 2% (w/v) SDS, 10% (v/v) glycerol, 5% (v/v) β-mercaptoethanol, 0.01% (w/v) bromophenol blue) and size separated using a 10% (w/v) SDS-polyacrylamide gel. Following electrophoresis, proteins were transferred onto 2 μm nitrocellulose membrane (Bio-Rad) using Trans-Blot® SD Semi-Dry Electrophoretic Transfer Cell (Bio-Rad). Membranes were stained with Ponceau-S to verify protein transfer, washed and then blocked with 1 pg/mL polyvinyl alcohol (P8136; Sigma-Aldrich, molecular weight: 30K-70K) in Tris-buffered saline with Tween-20 (TBS-T; 20mM Tris pH 7.5, 300 mM NaCl, 0.1% (v/v) Tween-20). Membranes were incubated overnight at 4°C with anti-GFP antibody (1:2500; Abcam, ab6556) in 5% (w/v) skimmed milk in TBS-T. Membranes were then washed with TBS-T five times for 5 minutes each and incubated with a 1:10,000 dilution of anti-rabbit IgG antibody conjugated to horseradish peroxidase (Sigma-Aldrich, A0545) for 1 hour before washing. After washing steps with TBS-T, membranes were treated with ECL Prime Western Blotting Detection Reagents (GE Healthcare, Little Chalfont, Buckinghamshire, United Kingdom), and exposed to CL-Xposure films (PI34093;

Thermo Fisher Scientific, Waltham, Massachusetts, USA) for 1-10 minutes. X-ray films were developed using a CP1000 film processor (Agfa-Gevaert N.V., Mortsel, Belgium).

3.2.6 **Dyes and microscopy imaging**

Seedling and plant tissues were mounted on GoldLine microscope slides (VWR International, Pennsylvania, USA) in water and examined using either Zeiss Axiophot epifluorescence microscope (Carl Zeiss Inc., Germany) or Zeiss LSM 510 META laser scanning confocal microscope (Carl Zeiss Inc., Germany). Wildtype ecotype Ws, *hth-9* plants transformed with pGWB640 and pGWB650, or Ws transformed with erRFP construct was prepared as needed in parallel as negative controls. Plant material was examined using either a fluorescence microscope (Zeiss Axio Imager D1 microscope equipped with a Zeiss AxioCam MRm camera controlled by Axio software) or a confocal microscope (Zeiss LSM 510 META laser scanning confocal microscope controlled by Zen software). To label nuclei, samples were equilibrated in 0.1% (w/v) Hoechst 33258 nucleic acid stain (Invitrogen) in water for 5 minutes. For staining mitochondria, whole seedlings were equilibrated with 50 nM TMRM (tetramethylrhodamine; T-668, Thermo Fisher Scientific) in half MS medium for 30 minutes. Stained samples were rinsed three times with water before microscopy imaging. The epifluorescence microscope was equipped with a Q-Imaging digital camera (Quorum Technologies Inc., Guelph, Ontario, Canada) controlled by the manufacturer's Axio software. Under UV illumination, filters of different excitation (ex) and emission (em) wavelengths were selected for different target signals as follows (ex/em): Hoechst, 365/395 nm; EYFP, 500/515 nm; G3GFP, 470/525 nm; erRFP; autofluorescence, 470/525 nm. Note, when exposure setting was optimized for RFP in *erRFP*-expressing transgenic lines,

chloroplast autofluorescence was undetectable, especially in etiolated hypocotyl cells. For confocal scanning microscopy, the specimens were excited with an argon laser using the following excitation and emission wavelengths (ex/em): RFP, 543/560-615 nm; HTH-FP, 488/505-530 nm; TMRM, 543/550-600 nm; autofluorescence, 543/642-749 nm. Controlled by Zen 2009 software, various pinhole and frame sizes were selected to minimize light exposure, fluorophore fading, and tissue damage. Representative images were chosen after similar results were obtained from at least three independent transgenic lines. For ER colocalization studies, images acquired by either epifluorescence or confocal microscopes were analyzed and the colocalization coefficient was determined by the Coloc 2 function with or without specific region of interest (ROI) settings on the Fiji/ImageJ platform (Schindelin et al., 2012; Schneider et al., 2012). Representative images (n = 10) were chosen after similar results were obtained from at least three independent transgenic lines. Images for transgenic lines harbouring the recombinant HTH protein, either tagged with EYFP or G3GFP, are all labelled as HTH_{pro}*:HTH-FP*, as the results were similar regardless of the fluorescent tag used.

3.2.7 **Methyl jasmonate (MeJA) treatment**

To make a 50 µM MeJA solution, 95% MeJA (Cat# 392707, Sigma-Aldrich, St. Louis, Missouri, USA) was first diluted in 95% ethanol, and the MeJA-ethanol mix was added in Milli-Q water, followed by stirring for 30 minutes at room temperature. Eleven-day-old wildtype (Ws) plants, including the root, were collected from the growth media. Soil was gently washed off the root by dipping the root in water repeatedly. Plants were transferred to the MeJA solution, placing the root in the solution and floating rosette leaves on the surface

and incubated at 22°C under continuous light. As a control, plants reared identically were removed from growth media, washed and floated in Milli-Q water. Rosette leaves were inspected 36 hours after the treatments using an epifluorescence microscope and the tissue subsequently stored at -80°C. MeJA treatments were repeated four times. Tissue collected from each replicate was considered one biological sample, resulting in four biological samples for each treatment. Samples were assayed using quantitative RT-qPCR.

3.2.8 **Quantitative RT-PCR**

Tissue collection, RNA isolation and cDNA synthesis

MeJA- and water-treated wildtype plants (four biological samples in total) were flash frozen in liquid nitrogen and pulverized by vortexing frozen tissue with stainless steel beads (1/8" diameter; Abbott Ball Company, West Hartford, Connecticut, USA) in 2 mL tubes. To prevent thawing, tubes were dipped in liquid nitrogen intermittently. Total RNA was extracted from 100 mg of tissue using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany). The RNA quality was assessed using RNA agarose gel electrophoresis. The RNA gel consisted of 1.5% (w/v) agarose, 1× MOPS buffer (20 mM 3-(N-Morpholino) propanesulfonic acid, 5 mM sodium acetate, 1 mM EDTA), 1.2% (v/v) formaldehyde and DEPC-treated Milli-Q water. Samples that showed no smearing and a discreet 28S to 18S RNA bands were selected for DNase treatment. DNAase treatment was done using Turbo DNA-free Kit (Am1907; Ambion, Naugatuck, Connecticut, USA) and following the manufacturer's instructions. The total RNA concentration of each DNase-treated RNA sample was determined using the NanoDrop 2000 spectrometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA). An aliquot of each RNA sample was transferred into a

separate tube and diluted with Milli-Q water until all samples reached the same total RNA concentration. These RNA samples were then reverse transcribed with random hexamer primers using SuperScript III or IV Reverse Transcriptase (Invitrogen, Carlsbad, California, USA).

Quantitative RT-PCR experimental setup and data analysis

Quantitative RT-PCR (RT-qPCR) was performed on a Real-Time Thermal Cycler CFX96 (Bio-Rad Laboratories, Hercules, California, USA). The PCR program consisted of an initial denaturing step at 98°C for 30 seconds, followed by 39 cycles at 98°C for 5 seconds, 60°C for 3 seconds and a plate read. The primer sequences were designed based on gene structure models obtained at the Arabidopsis Information Resource (TAIR; http://www.arabidop-sis.org) using QuantPrime (http://www.quantprime.de) (Arvidsson et al., 2008) or Beacon Designer™ Free Edition (Premier Biosoft International, Ltd., Palo Alto, California, USA) and by setting the primer melting temperature at 60°C.

Standard curves were generated using different template cDNA concentrations to determine the reaction efficiency. To ensure the standard curve covered all potential template concentrations that might be encountered in the study, the eight biological cDNA samples of both MeJA treated and water treated samples were pooled, and a tenfold dilution series was generated over six points, starting from the most concentrated cDNA samples. For each dilution, a standard qPCR protocol was performed in triplicate for all the primer pairs. The standard curve was constructed by plotting the log of the starting quantity of the template against the C_T values obtained by the CFX manager software 1.0 package (Bio-Rad

Laboratories). Primers that yielded 90-110% amplification efficiency and coefficient of determination (r^2) values > 0.980 were selected.

HTH cDNA was amplified using primers forwards 5'-GAGAGGTGGCGTTCCGTTTA-3' and reverse 5'-TTCACGAACGCAGCATCGG -3'. To verify that MeJA treatment was effective in triggering stress responses in plants, the transcript level of *VEGETATIVE STORAGE PROTEIN 2* (*VSP2*; AT5G24770) was measured using forward primer 5'-CCGTTGGAAGTTGTGGAAGAAT-3' and reverse primer 5'-TCTTCACGAGACTCTTCCTC-3'. Three house-keeping genes, *ACTIN* 7 (ACT7; AT5G09810), *GLYCERALDEHYDE*-3-*PHOSPHATE DEHYDROGENASE C-2* (*GAPC2*; AT1G13440) and *TUBULIN* 6 (*TUB6*; AT5G12250), were included for normalization of *HTH* and *VSP2* transcript levels. The primer sequences were: ACT7, forward 5'-TGGAACTGGAATGGTGAAGG-3' and reverse 5'-GACTGAGCTTCATCACCAACG-3'; GAPC2, forwards 5'-GGTGACAACAGGTCAAGCATT-3' and reverse 5'-CAACCACACACAAACTCTCGC-3'; TUB6, forward 5'-GGATTCTCCTCTGCACCATAAAA-3' and reverse 5'-CATTGACACGCTCCAACTGC-3'. The amplicon sizes ranged between 144 and 201 bp. Standard melting-curve analysis provided by the instrument (Real-Time Thermal Cycler CFX96, Bio-Rad Laboratories) was performed between 65°C to 95°C with an increment of 0.5°C after each PCR run to determine whether a single PCR product was amplified in each reaction. For each primer, four biological samples and three technical replicates were included for MeJA- and water-treated plants. Therefore, one RT-qPCR run consisted of 24 15-µl reactions containing 300 or 500 nM of primers, cDNAs, nuclease-free water, and the SsoFast EvaGreen Supermix (Bio-

Rad Laboratories) as per manufacturer's instructions. In addition to experimental samples, each qPCR run also included control reactions, performed in triplets, that contained no cDNA templates (‘no template control’) and reactions that contained pooled RNA sample from all biological samples of both treatments that had not been subjected to reverse transcription (‘no RT control’). Expression levels of house-keeping genes were analyzed for stability (Coefficient Variance and M value) using CFX Manager 3.1 software (Bio-Rad Laboratories). The expression of target genes (*HTH* and *VSPS2*) were then normalized to the house-keeping genes. SigmaPlot 11.0 (Systat Software) was used to plot graphs and perform statistics

3.3 Results

3.3.1 Mutant phenotypes

HTH gene was previously identified by its mutant floral fusion phenotype. Other mutant phenotypes including increased cuticle permeability, ovule defects, changes in cuticular lipid composition and ectopic pollen germination have been described (Krolikowski et al., 2003; Lolle et al., 1998; Lolle and Pruitt, 1999; Pagnussat et al., 2005). These reports mostly include descriptions of organ fusion observed in mutants generated by single point mutations in the Landsberg *erecta* (L*er*) background (*hth-1* to *hth-8, hth-10* to *hth-11*) although mutations in other ecotype backgrounds have been identified. Here, *hth* mutant flowers were examined from three T-DNA insertion lines (*hth-13, hth-14* and *hth-15*) generated in the Columbia (Col) background, one line harboring the *hth-9* allele in the Wassilewskija (Ws) background and multiple lines harboring single point mutations in the L*er* background

(Figure 3.1). Figure 1.8 shows the point mutations and T-DNA insertion sites found in these mutant alleles.

In this study, plants of the L*er* ecotype are referred to as 'wildtype' despite harboring a mutant *erecta* allele as many *hth* mutants were generated in this background. The flowers of L*er* plants display fully opened and expanded petals regardless of shortened internodes due to the mutation. L*er* siliques were straight or slightly curved. Mutant *hth-1*, *hth-4*, *hth-5*, *hth-7* and *hth-8* flowers showed organ fusion of different severity (Figure 3.1B-G); for example, *hth-8* flowers were completely closed preventing petals from being visible, whereas sepals of *hth-5* were less fused and hence did not completely block petal emergence. Although the flower buds stayed partially or entirely closed, the pistil generally would protrude out of individual floral buds. If fertilized, developing mutant siliques tended to be shorter, bent or tangled. In *hth-4*, flowers were not entirely closed with petals exposed due to the opening of overlapping sepals on the side although sepals near the distal region remained fused.

Ws wildtype plants showed a floral morphology similar to L*er* having fully opened flowers. In contrast, mutant *hth-9* flowers were mostly closed and had an undulating surface (Figure 3.1J-L). As the pistil elongated, occasionally the flower opened slightly allowing sepals to become fully separated. The petals became visible but remained furrowed. Sometimes, small tears on inner sepals were observed, usually at the edge of the overlap towards the tip.

The floral phenotype of the T-DNA insertion lines *hth-13* and *hth-14* resembled mutants in the two other backgrounds. In contrast, *hth-15* only showed moderate organ fusion (Figure

3.1N-P). Similar to *hth-9*, siliques of *hth-13* were short and contained few seeds. Within these siliques, the empty spaces were predominantly occupied by aborted seeds (darker coloured) rather than unfertilized ovules (lighter coloured).

3.3.2 **Preliminary bioinformatics analyses of the putative HTH protein**

Based on the nucleic acid sequence, the *HTH* gene encodes a protein consisting of 594 amino acids with a predicted molecular mass of 65.3 kilodaltons (kDa). The HTH protein belongs to the glucose-methanol-choline (GMC) oxidoreductase family, a protein group that exhibits diverse protein functions. Characterized GMC oxidoreductase enzymes exhibit multiple functions, including choline dehydrogenase, methanol oxidase and cellobiose dehydrogenase as well as a hydroxynitrile lyase (Dreveny et al., 2001). To elucidate some of the key structural features that characterize the HTH protein, bioinformatics analyses using different tools were employed. Using the PROFsec tool a compositional ratio of 22% helix, 21% extended sheet structure and 57% loop was predicted, and GLOBPLOT predicted that HTH is likley a globular protein. UniProtKB/Swiss-Prot at Expasy (www.expasy.org) revealed a FAD binding site in the corresponding GMC_oxred_N conserved domain. SignalP identified a putative 19 amino acid (aa)-long signal peptide at the N-terminus with the cleavage site located between the 19th and 20th aa (confidence: 0.606). The only predicted transmembrane helix motif is located in this region, indicating that the cleaved mature HTH (theoretical molecular weight 62.2 kDa) is likely a non-transmembrane protein.

Putative *HTH* expression patterns and transcriptional regulators were analyzed using various prediction tools (Figure 3.2). According to available microarray-based expression

data, *HTH* is expressed predominantly in apical meristem tissue, young floral buds, and young siliques. Among floral organs, petals and ovaries exhibited the strongest expression. The analysis of the 500 bp upstream region of the *HTH* gene using the CISTOME tool revealed two types of putative transcription factor binding sequences, CArG and MYC. CArG is a target of MADS-domain transcription factors that are involved in plant development, whereas MYC recognition sites are found in the promoter region of many stress-responsive genes. Two CArG sites were identified at -82 and -321 bp, and three MYC sites at -71, -290 and -303 bp upstream of the ATG codon.

3.3.3 Phenotypes of HTH_{pro}:*HTH-FP* transgenic plants in the *hth-9* background

HTH protein localization has not been directly determined but rather only inferred using indirect methods such as *in situ* mRNA hybridization or promoter-reporter fusion constructs (Krolikowski et al., 2003; Kurdyukov et al., 2006b). To further investigate HTH protein localization, transgenic plants were generated to allow direct visualization of a HTH protein that was fused with either a yellow (YFP) or green fluorescent protein (GFP). To minimize artifacts due to over-expression, the expression of *HTH-FP* was driven by the full-length 5' upstream 2009 bp-long region flanking the *HTH* gene. Two reporter constructs HTH_{pro}:*HTH-EYFP* and HTH_{pro}:*HTH-G3GFP* were generated from host vectors pGWB640 and pGWB650, respectively (Nakagawa et al., 2007). The two translational reporter constructs were transformed into *hth-9* mutant plants and tested for rescue of the mutant phenotype (Figure 3.3). Meanwhile, the host vectors were also transformed into *hth-9* plants, and the resultant plants are referred to as empty vector (EV) lines.

Multiple independently isolated transgenic T1 plants were tested, and in every case *hth-9* plants transformed with the translational reporter constructs exhibited wildtype phenotypes. Mutant plants transformed with the empty vector pGWB640 retained the mutant phenotype, whereas *hth-9* mutant plants harbouring the HTH_{pro}*:HTH-EYFP* construct gave rise to phenotypically wildtype flowers (Figure 3.3B-E). Mutant *hth-9* plants transformed with either translational reporter construct were phenotypically wildtype, and the observed expression patterns were also identical. Therefore, the resulted transgenic reporter plants will henceforth be collectively referred to as HTH_{pro}*:HTH-FP* transgenic lines unless otherwise specified.

The expression of HTH_{pro}*:HTH-FP* was further verified by immunoblotting (Figure 3.3F). While no immunoreactive protein species were detected in flowers of empty vector lines, an immunoreactive protein species approximately 135 kDa in size was detected in seedlings, flower buds and siliques using an anti-GFP antibody. However, HTH-FP protein was not detected in rosette leaves using this method.

To determine whether HTH_{pro}*:HTH-FP* transgenic plants also had cuticle permeability restored to wildtype levels, a chlorophyll extraction assay was conducted. Using this assay changes in cuticle permeability can be quantified; the more permeable the cuticle is, the faster chlorophyll can be extracted from leaves (Figure 3.4A). Similar to the floral phenotypes, the rates of chlorophyll extraction from HTH_{pro}*:HTH-EYFP* transgenic plants and HTH_{pro}*:HTH-G3GFP* were comparable to that of Ws wildtype. In contrast, the permeability of EV plants was comparable to that of mutant *hth-9* plants.

By staining inflorescences and rosette leaves with toluene blue (TBO) it is possible to visualize differences in cuticle permeability between wildtype, *hth*-9 and transgenic plants (Figure 3.4B-G). Based on TBO staining, *hth-9* tissues were more permeable than wildtype and transgenic plant tissues. While Ws inflorescence was only slightly TBO stained, *hth-9* younger bud pedicels and sepal edges were stained dark blue. Similarly, partially exposed *hth-9* petal tissues were also stained. TBO staining of 10-day-old rosette leaves isolated from wildtype differed from the *hth-9* mutant, but was similar to HTH_{pro}*:HTH-EYFP* transgenic plants.

In summary, expression of the HTH_{pro}:HTH-FP in *hth-9* mutants appeared to render transgenic plants phenotypically wildtype and restored cuticle permeability to levels comparable to those found in Ws wildtype plants. These results suggest that the HTH-FP fusion protein functionally complements the *hth-9* mutant.

3.3.4 HTH-FP localization in seedlings and juvenile plants

HTH-FP was detected in the primary and lateral roots, the hypocotyl, the shoot apical meristem (SAM) region, the cotyledons, and trichomes (Figure 3.5). The protein localization was predominantly detected in the vasculature of the hypocotyl and cotyledon as well as the hydathode (Figure 3.5B and D). Fluorescence was also detected in two-week-old plants in the leaf trichomes. As shown in Figure 3.5F, both EV and HTH_{pro}*:HTH-FP* rosette leaves had trichomes, but only HTH_{pro}*:HTH-FP* trichomes fluoresced. In the root, fluorescence was localized to the stele of both primary and lateral roots, but no fluorescence was detected in root hairs (Figure 3.6). In the cotyledons, hypocotyl, and emerging leaves of seedlings, HTH-

FP appeared to be localized to the epidermis (Figure 3.7E and F). Vasculature localization was also evident (Figure 3.7C, E and F). Interestingly, punctates were observed in epidermal cells of the hypocotyl. Strong fluorescence was observed in the stipules of two-week old plants. At higher resolution, both epifluorescence and confocal laser scanning microscopy confirmed fluorescence in the cotyledon epidermis, including pavement and guard cells (Figure 3.8A, B, H and L). Fluorescence confined to small punctates or aggregates within these epidermal cells was also observed (Figure 3.8H).

3.3.5 HTH-FP localization in floral and reproductive tissues

HTH-FP in floral tissues is shown in Figure 3.9. In young flower buds, green fluorescence was detected in the vasculature, floral receptacle and pistil (Figure 3.9A and C). In isolated sepals and petals, expression was prominent in the veins and epidermis (Figure 3.9F and I). In the pistil, fluorescence was confined to the ovary wall (Figure 3.9K). For anthers, fluorescence was observed in the stamen epidermis, vascular bundle and the cells between adjoining locules (Figure 3.9N and O).

To further examine the temporal expression pattern of HTH, the ovules/developing seeds were examined by microscopies (Figure 3.10 and Figure 3.11). In a dissected ovary, HTH-FP was present in the chalazal region of individual ovules (Figure 3.10E). To observe HTH-FP localization over time, ovules were removed from the ovary at various time points prior to and following anthesis ((Figure 3.10F-M). Two days before anthesis (FG3-4), HTH-FP became detectable at the chalazal end of the embryo sac and was confined to a relatively small area. At FG6, fluorescent protein signal was highly polarized to the chalazal end of the

embryo sac, coincident with the antipodal cells. At anthesis, the chalazal fluorescence became more diffuse, and fluorescence outside of the embryo sac was also observed

After fertilization, integument expression became apparent at 3 day post-anthesis (+3 DPA), and it persisted at later stages when HTH-FP was more pronounced in the chalazal and micropylar seed coat (Figure 3.11). In Figure 3.12F-H, epifluorescence microscopy images of the developing seed coat at three different time points after anthesis (+7, +8 and +10 DPA) indicate that HTH-FP was localized to the inner layer of the outer integument (oi1) and possibly also other integument layers underneath since it is possible that fluorescence from multiple cell layers was incorporated in those images. At the post mature stage (+20 DPA), the confocal images showed that in addition to localization to oi1, the HTH-FP was also detected in the cytoplasm of the columella structures in the oi2 layer (Figure 3.13I and K).

3.3.6 **Seed mutant phenotypes**

Detecting HTH-FP in the integument prompted further investigation of seed phenotypes among various *hth* mutant lines (Figure 3.14). Although homogenous in size for each ecotype, the seed size varied greatly between ecotypes, with Ws seeds being the largest and L*er* seeds the smallest. Interestingly, regardless of the ecotype background, all *hth* mutant seeds were visually larger and more variant in size than their respective wildtype counterparts. Seed weight differences corroborated these initial observations (Table 3.1). Ws seeds weighed twice that of L*er* seeds. Furthermore, there was approximately a 1.5 - 2 times increase in weight for mutant seeds relative to the corresponding wildtype. In addition,

mutant seeds were misshapen and lacked the more regular oval shape typical of wildtype seeds (insets in Figure 3.14). Although only observed infrequently, intact seeds adhering superficially to one another were also found (Figure 3.14G-inset). Importantly, among of *HTH_{pro}:HTH-G3GFP* and *HTH_{pro}:HTH-EYFP* plants, relative seed size and weight was comparable to that of the wildtype.

To examine whether the mutant seed coat also has elevated permeability, and more importantly whether *HTH_{pro}:HTH-FP* expression restores it, the permeability properties of mutant and wildtype seed coats, were tested using tetrazolium red. Tetrazolium red is a cationic dye that is largely excluded by a normal Arabidopsis seed coat. However, if this dye permeates the seed coat and comes into contact with embryonic tissue, it is reduced to red-coloured formazans by NADPH-dependent reductases offering a simple visual assay for seed coat permeability (Beisson et al., 2007). In the Col and Ws backgrounds, exposure of mutant seeds to tetrazolium red consistently stained embryos red while wildtype seeds were stained minimally (Figure 3.15A and B). This difference was less apparent for mutants in the L*er* background where wildtype seeds were also stained to some extent (Figure 3.15C). Among T-DNA insertion lines *hth-13* and *hth-14* that showed severe flower fusion produced seeds allowing high levels of tetrazolium staining, whereas *hth-15* that showed a moderate floral fusion phenotype only exhibited moderate tetrazolium staining of seeds (Figure 3.15A). For transgenic *HTH_{pro}:HTH-FP* plants, the staining pattern resembled that of wildtype Ws plants (Figure 3.15B); mutant embryos that had been removed from tetrazolium treated seeds were larger and misshapen relative to their wildtype counterpart. Overall, these results are

consistent with the mutant seed coat having altered and probably elevated seed coat permeability.

3.3.7 **Subcellular localization of HTH-FP**

As shown above in Figure 3.7 and Figure 3.8, HTH-FP was detected in the cotyledon petiole, emerging leaf primordia and the hypocotyl. Within cells HTH-FP is localized to a reticular network and occasionally to discreet spindle-shaped bodies (see Appendix C for images of homozygous T2 individuals that descended from four independent T1 lines). To further resolve HTH-FP's subcellular localization, hypocotyl and pedicel cells of young seedlings were examined using confocal laser scanning microscopy (Figure 3.16F-M). Confocal imaging revealed that the punctates were predominantly spindle shaped and 5-10 µm in size. These bodies were motile, trafficking at a speed of 0.5-1 µm/sec in the cell.

To establish the etiology of these cellular bodies, organelle-specific stains or lines co-expressing organelle-specific reporter constructs were used and co-localization assessed. For mitochondria labeling (Figure 3.16N-P), a fluorescent dye TMRM (tetramethylrhodamine) that is readily sequestered by functional mitochondria was used, and its fluorescent emission can be separated from HTH-FP. TMRM-stained mitochondria, however, did not colocalize with the fluorescent bodies observed in HTH_{pro}*:HTH-FP* seedlings. Similarly, colocalization to chloroplasts and nuclei was not detected (Appendix D).

To test for localization to the ER, a more comprehensive analysis of expression in seedling hypocotyl cells was undertaken using an epifluorescence and confocal microscope (Figure 3.17). A Cauliflower Mosaic Virus 35S promoter-driven construct (Sinclair et al.,

2009) that leads to the expression of an ER-targeted red fluorescent protein (*erRFP*) was transformed into wildtype Ws plants. As previously reported, the pattern of erRFP localization in wildtype plants was restricted to ER-derived structures (i.e. ER bodies) and the network, comparable to that of HTH-FP in the HTH_{pro}:*HTH-FP* plant (Figure 3.17A-B). Furthermore, doubly transformed lines coexpressing HTH_{pro}:*HTH-FP* and *erRFP* were generated. In most cases, HTH-FP was detected in both the ER network and bodies, while erRFP predominantly in bodies (Figure 3.17C-I). This difference between erRFP and HTH-FP in distribution to ER domains was reflected on Pearson's coefficients. When colocalization to the spherical bodies only was considered using region of interest (ROI) analyses, the coefficient is higher (0.79 ± 0.063) than the coefficient (0.25 ± 0.051) by using the entire region of images (n = 10). In summary, the colocalization of HTH-FP and erRFP in the spindle-shaped structures further verified the ER origin of these highly motile, elliptical bodies.

3.3.8 **The effect of MeJA on the expression level of *HTH***

ER bodies are commonly present in seedlings but rarely seen in rosette leaves. Yet, Hayashi et al. (2001) showed that methyl jasmonate (MeJA) and wounding can induce ER bodies in rosette leaves. HTH-FP's localization to this stress-associated organelle suggests a possible role in stress response pathways. Experiments were carried out to determine whether exposure of plants to MeJA elevates *HTH* expression and whether in turn, MeJA exposure induces the formation of HTH-FP containing ER bodies. Ws wildtype were exposed to MeJA and RT-qPCR used to quantify changes in *HTH* expression in wildtype Ws rosette leaves. The RT-qPCR analysis showed that *HTH* expression was elevated in MeJA-treated wildtype

Ws plants (Figure 3.18A). Plants treated with MeJA showed dark purple colouration in the petioles of treated plants (Figure 3.18C) and an increase in expression of the MeJA responsive gene *VEGETATIVE STORAGE PROTEIN 2* (*VSP2*) (Dombrecht et al., 2007). When HTH_{pro}:*HTH-FP* epidermal cells along the rosette leaf midrib were examined using epifluorescence microscopy, HTH-FP was detected in the ER network and not ER bodies for water-treated control plants (Figure 3.18D). Following MeJA-treatment HTH-FP fluorescence was localized to structures reminiscent of ER bodies, suggesting that MeJA induces HTH relocalization to ER bodies (Figure 3.18E). Future experiments with the erRFP marker need to be conducted to confirm these preliminary data.

Figure 3.1 Phenotypes of *hothead* mutants in different ecotype backgrounds. (A) L*er* ecotype flowers were open and had fully expanded petals. Individual petals (p) and sepals (s) were well separated. Siliques were straight or with a slight curve. (B-H) Mutant flowers in the L*er* background. (B) *hth-5* flowers were not fully open and petals failed to emerge. (C-F) *hth-4*, *hth-7* and *hth-8* exhibit severe floral organ fusion. Flowers were completely closed, and petals were enclosed within the unopen bud. Siliques were often bent in severe mutants harboring these alleles. (G) *hth-1* siliques were short and tangled. (H) A pistil protruded through the tip of a closed flower. Petals exposed due to opening of overlapping sepals (asterisk) although sepal edges remained fused near the distal end (arrow). (I) Wildtype Wassilewskija (Ws) ecotype flowers. (J-M) Flowers and siliques of *hth-9*, a mutant in the Ws background. (J) Mutant flowers were mostly closed and had an uneven undulating surface. Flower buds remained enclosed by the fused sepals. (K) Occasionally, tissue rupture (arrow) was observed at the edge of overlap. (L) In some mature flowers sepal separation did occur but the petals remain furrowed. (M) In general, mutant plants had smaller siliques and fewer fertilized carpels. (N) Flowers of wildtype Columbia (Col) and three T-DNA insertion mutants. *hth-13* and *hth-14* exhibit more severe floral fusion than *hth-15*. (O-R) Mutant *hth-13* plants produce shorter siliques that contain fewer mature seeds (ms). The majority of fertilized ovules failed to mature. These aborted seeds (as) are clearly distinguishable from unfertilized ovules (uo). Scale bar: (A-L, N) 2 mm; (M) 10 mm; (O) 5 mm; (P-R) 400 µm.

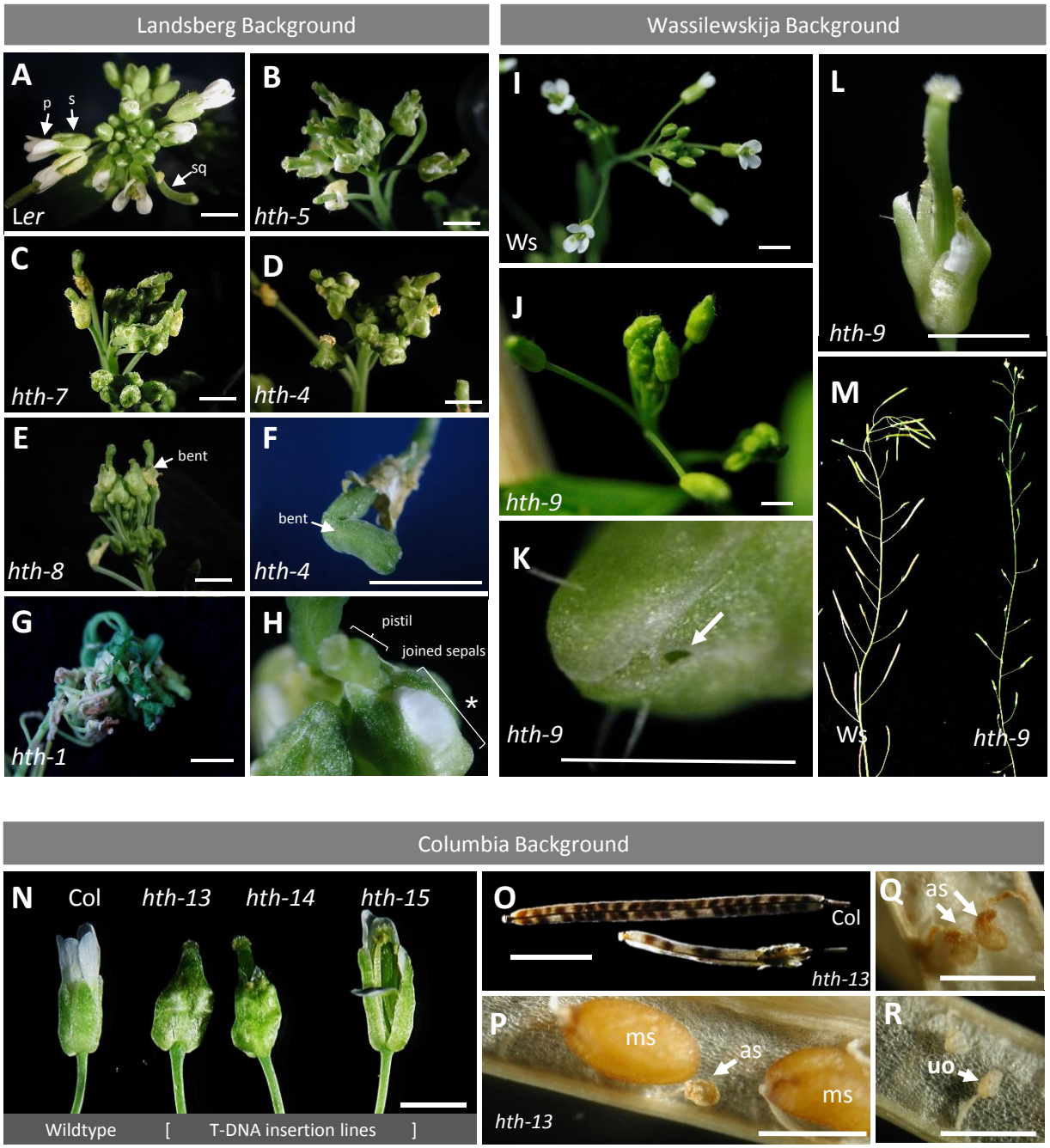
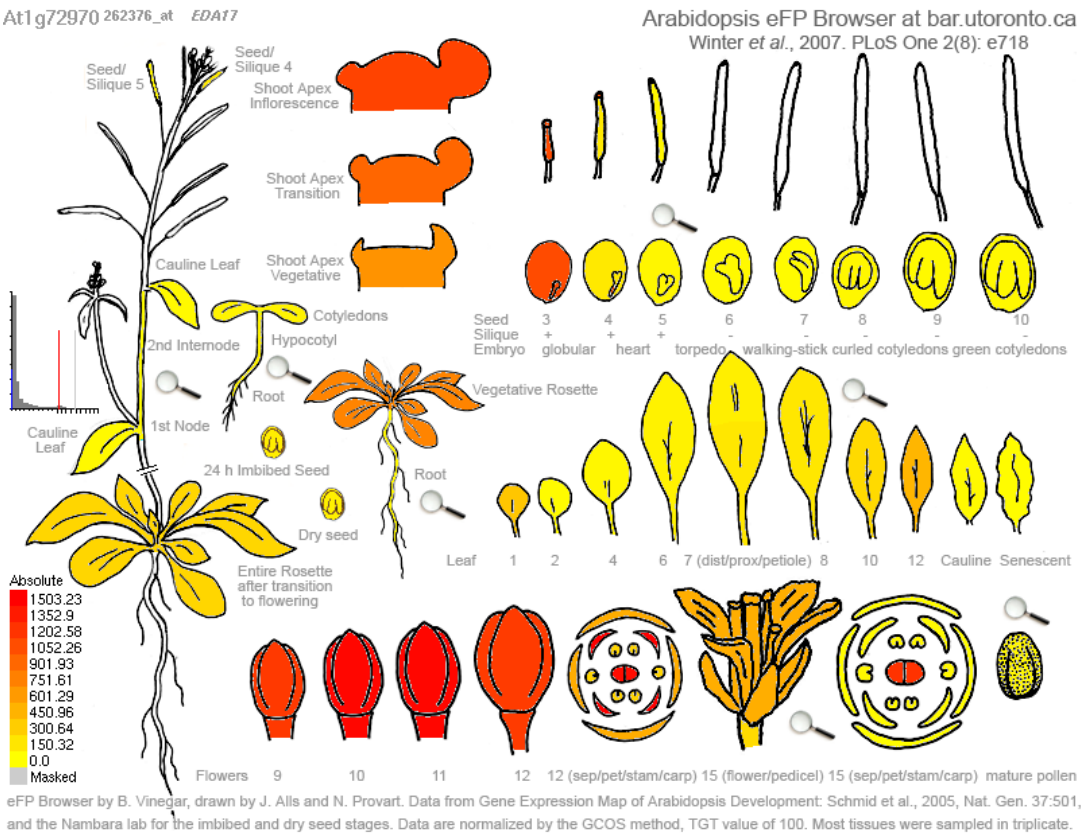

Landsberg Background
Wassilewskija Background
A
p
s
sq
Ler
B
hth-5
C
hth-7
D
hth-4
E
bent
hth-8
F
bent
hth-4
G
hth-1
H
pistil
joined sepals
*
I
Ws
J
hth-9
K
hth-9
L
hth-9
M
Ws
hth-9
Columbia Background
N
Col
hth-13
hth-14
hth-15
Wildtype
[
T-DNA insertion lines
]
O
Col
hth-13
Q
as
P
ms
as
ms
hth-13
R
uo

Figure 3.2 Predicted tissue expression patterns and promoter elements in the upstream 500 bp region of *HTH*. (A). A graphic representation of microarray-based expression patterns of the wildtype *HTH* gene. This illustration was generated by the electronic fluorescent pictograph browser (eFP Browser) at BAR (http://bar.utoronto.ca). HTH-FP was most prominent in apical meristem, young flower buds, young siliques and immature ovules. (B) Analysis of the 500 bp upstream region of the *HTH* sequence. Two types of promoter elements, CArG and MYC, were identified. The CArG type element is a target of MADS-domain containing transcription factors that are involved in plant development, whereas MYC recognition sites are found in the promoter region of many stress-responsive genes. The numbers indicate the positions of elements relative to the ATG initiation codon. Promoter analysis was performed by the CISTOME tool of the Bio-Analytic Resource (www.bar.utoronto.ca) with the setting of Ze cutoff of 3.5, functional depth cutoff of 0.9 and proportion of genes of 0.5.

A


At1g72970 262376_at EDA17
Arabidopsis eFP Browser at bar.utoronto.ca
Winter et al., 2007. PLoS One 2(8): e718
Seed/ Silique 5
Seed/ Silique 4
Shoot Apex Inflorescence
Shoot Apex Transition
Shoot Apex Vegetative
Cauline Leaf
Cotyledons
Hypocotyl
2nd Internode
Root
1st Node
Cauline Leaf
24 h Imbibed Seed
Dry seed
Entire Rosette after transition to flowering
Vegetative Rosette
Root
Seed 3 4 5 6 7 8 9 10
Silique + + + - - - - -
Embryo globular heart torpedo walking-stick curled cotyledons green cotyledons
Leaf 1 2 4 6 7 (dist/prox/petiole) 8 10 12 Cauline Senescent
Absolute
1503.23
1352.9
1202.58
1052.26
901.93
751.61
601.29
450.96
300.64
150.32
0.0
Masked
Flowers 9 10 11 12 12 (sep/pet/stam/carp) 15 (flower/pedicel) 15 (sep/pet/stam/carp) mature pollen
eFP Browser by B. Vinegar, drawn by J. Alls and N. Provart. Data from Gene Expression Map of Arabidopsis Development: Schmid et al., 2005, Nat. Gen. 37:501, and the Nambara lab for the imbibed and dry seed stages. Data are normalized by the GCOS method, TGT value of 100. Most tissues were sampled in triplicate.

B

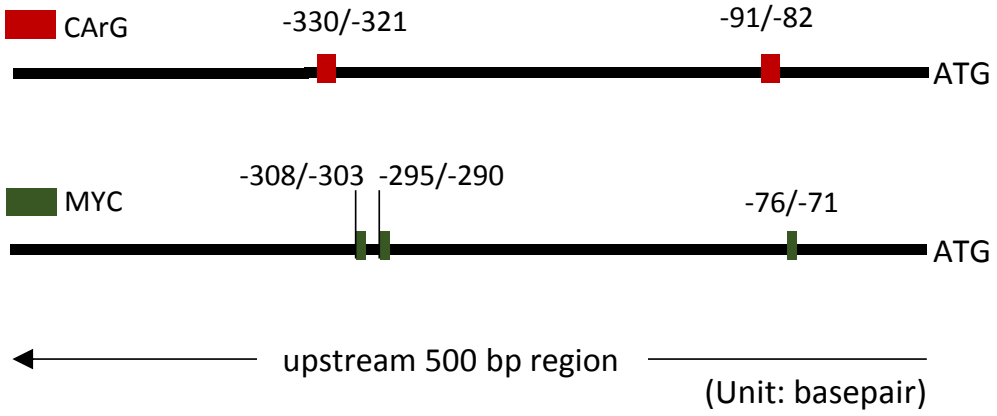
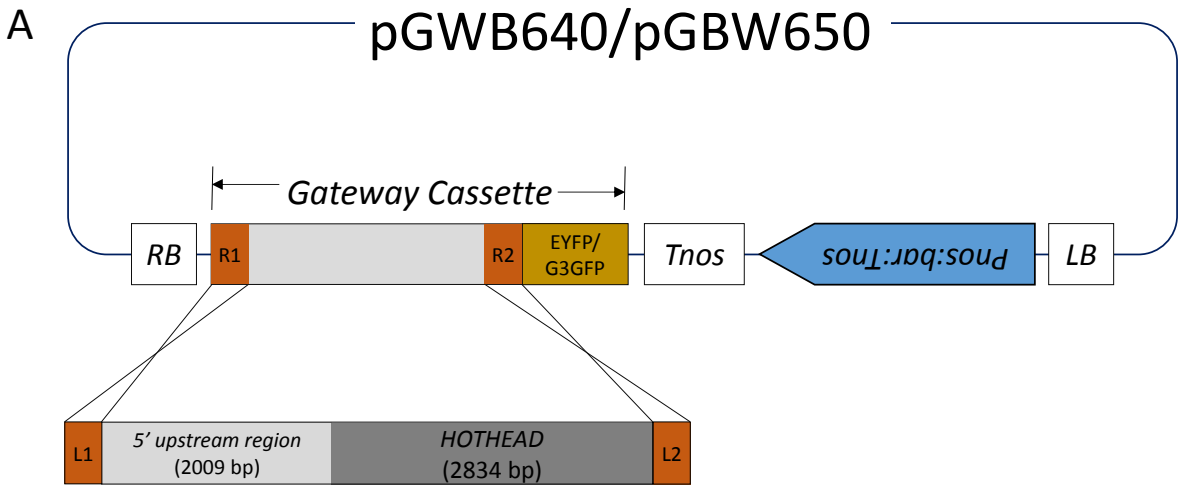
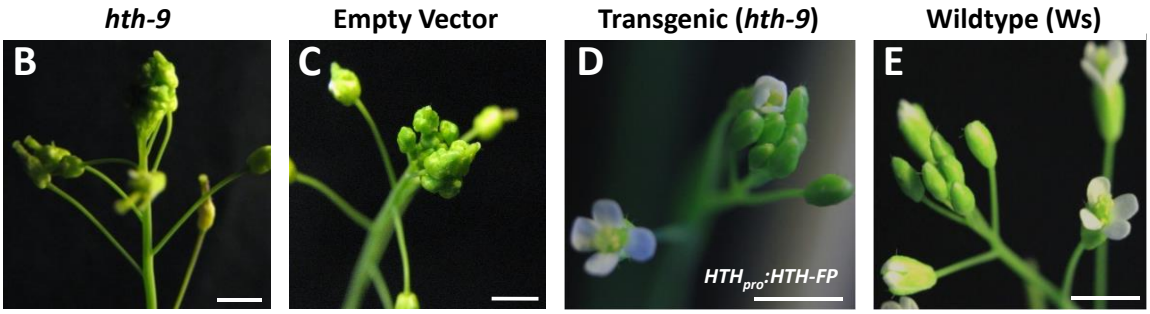

CArG
-330/-321
-91/-82
ATG
MYC
-308/-303
-295/-290
-76/-71
ATG
upstream 500 bp region
(Unit: basepair)

Figure 3.3 *HTH_{pro}:HTH-FP* constructs and floral phenotype of transgenic plants. (A) Gateway pGWB640 and pGWB650 destination vectors contain a C-terminal tagged Gateway cassette that is flanked by attR sites (R1 and R2) (Nakagawa et al., 2007). The *HTH* gene with its putative promoter (the 5' upstream region) is flanked by attL sites (L1 and L2) and subsequently swapped in place of the Gateway cassette, resulting in expression vectors *HTH_{pro}:HTH-EYFP* and *HTH_{pro}:HTH-G3GFP* (collectively referred to as *HTH_{pro}:HTH-FP*). (B) Mutant flowers of untransformed *hth-9* plants are fused. (C) Mature flowers of *hth-9* plants resemble those of mutant plants when transformed with the vector alone (empty vector). (D-E) Mutant *hth-9* plants transformed with the *HTH_{pro}:HTH-FP* vector showed a wildtype floral phenotype that is indistinguishable from the Ws wildtype. (F) An anti-GFP antibody cross-reacts with protein bands when protein extracts are electrophoresed, transferred to a supporting membrane and probed using immunoblotting techniques. The lower panel is a membrane that has been Ponceau stained showing relative protein loading. *RB*, right boarder; *LB*, left boarder; *Pnos*, promoter of the nopaline synthase gene*; Tnos*, terminator of nopaline synthase*; bar*, bialaphos resistance gene*;* EYFP, enhanced yellow fluorescent protein; G3GFP, G3 green fluorescent protein; L1, L2, R1, R2: Gateway attL and attR recombination sites for sequence exchange. M: protein marker. Scale bar = 0.5 cm.


A
pGWB640/pGBW650
Gateway Cassette
RB
R1
R2
EYFP/
G3GFP
Tnos
Pnos:bar:Tnos
LB
L1
5' upstream region
(2009 bp)
HOTHEAD
(2834 bp)
L2


hth-9
Empty Vector
Transgenic (hth-9)
Wildtype (Ws)
B
C
D
E
HTHpro:HTH-FP

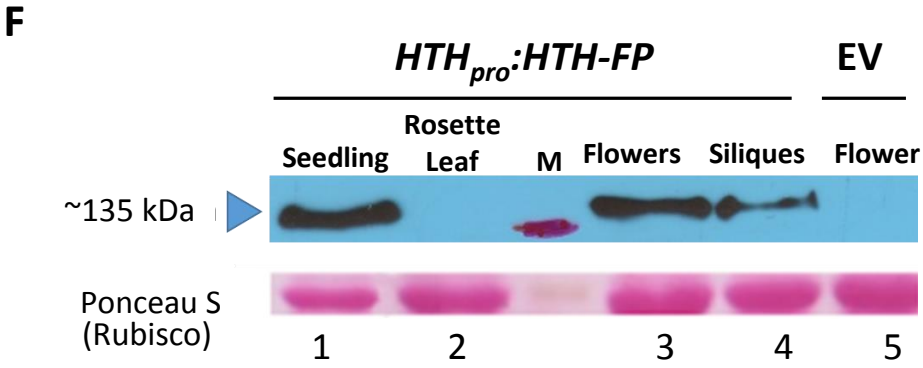
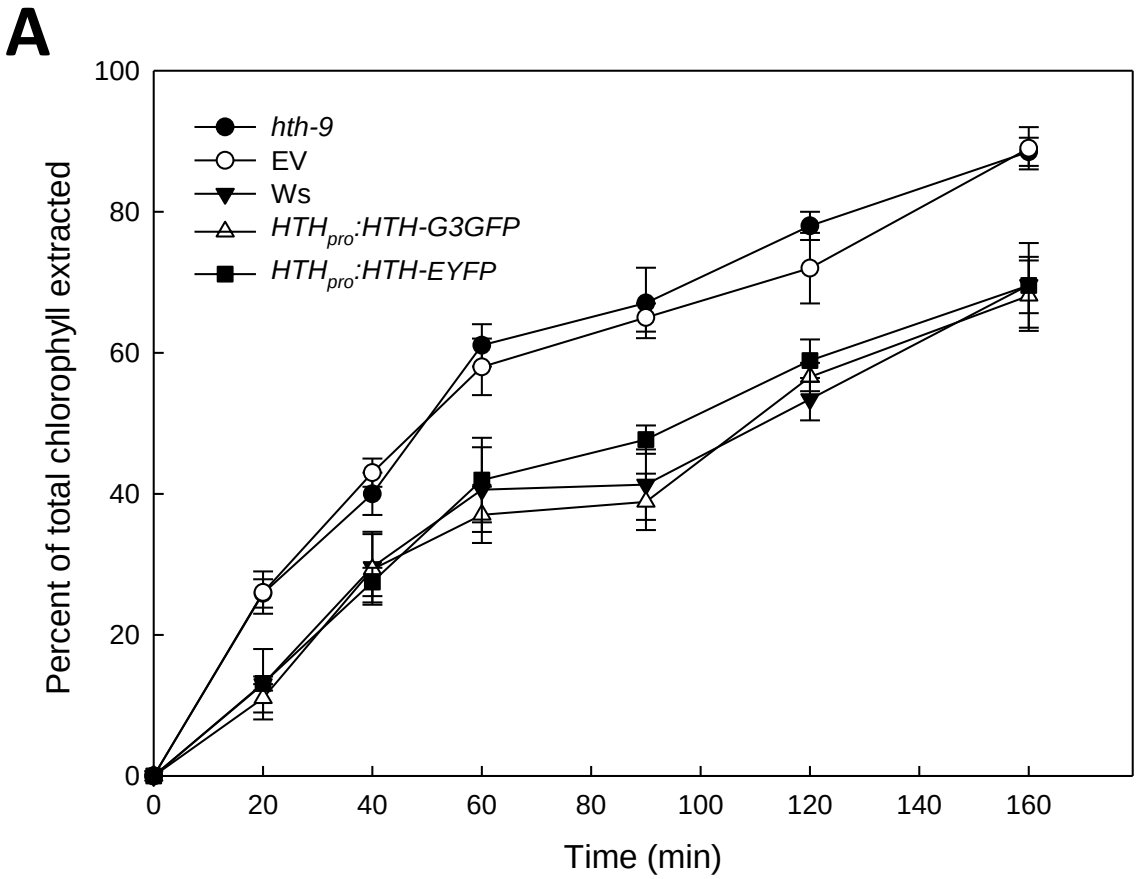

F
HTHpro:HTH-FP
EV
Seedling
Rosette
Leaf
M
Flowers
Siliques
Flower
~135 kDa
Ponceau S
(Rubisco)
1
2
3
4
5

Figure 3.4 Chlorophyll extraction rates for wildtype, *hth-9* and transgenic lines. (A) A graph showing the rate of chlorophyll extraction from cauline leaves submerged in 80% ethanol over a period of 160 minutes. The chlorophyll concentration after 24 hours was set to be 100% as the extraction maximum. Error bar = ± 2 standard error (n = 4). (B-G) Flowers and rosette leaves stained with toluidine blue (TBO). (B, F) Ws wildtype flowers. Petals and sepals were only lightly stained. Anthers were heavily stained (F). For *hth-9* mutants (C), in addition to anthers, pedicels (pd) and sepal edges (se) were clearly stained. (D, E) The *hth-9* petals (p) and 10-day-old rosette leaves were more readily stained. Relative staining of transgenic flower tissues (*HTH_{pro}: HTH-EYFP/-G3GFP*) resembled that of wildtype flowers. (G) A close-up showing stained sepal edges and petals of *hth-9*. EV: empty vector transgenic plants (*pGWB640*); *HTH_{pro}:HTH-G3GFP* and *HTH_{pro}:HTH-EYFP*: transgenic plants in the *hth-9* background transformed with respective vectors. Scale bar: 2 mm.


A
Percent of total chlorophyll extracted
Time (min)
hth-9
EV
Ws
HTH$_{pro}$:HTH-G3GFP
HTH$_{pro}$:HTH-EYFP
0
20
40
60
80
100
120
140
160

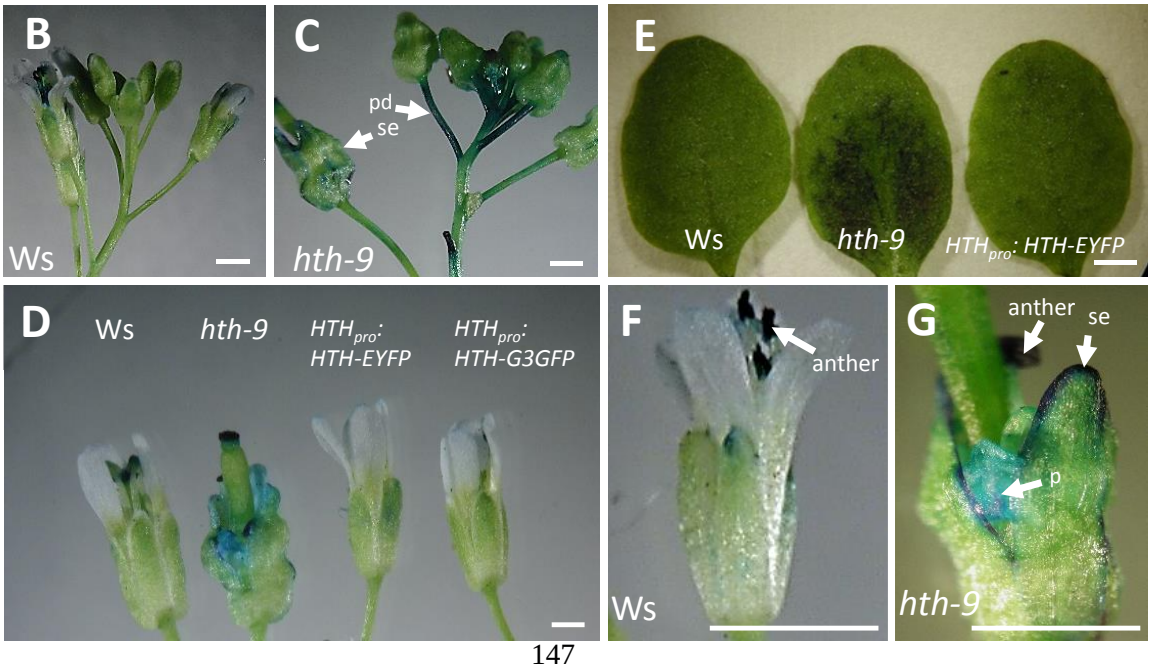

B
Ws
C
pd
se
hth-9
E
Ws
hth-9
HTH$_{pro}$: HTH-EYFP
D
Ws
hth-9
HTH$_{pro}$:
HTH-EYFP
HTH$_{pro}$:
HTH-G3GFP
F
anther
Ws
G
anther
se
p
hth-9

Figure 3.5 Epifluorescence micrographs showing HTH-FP tissue localization in *HTH_{pro}:HTH-FP* transgenic plants. Side-by-side comparisons of *HTH_{pro}:HTH-FP* and empty vector (EV) seedlings. (A-C) For four-day-old seedlings, fluorescence was detected the shoot apical meristem (SAM) region, within the hypocotyl (hp), cotyledons (co), lateral roots (lr) and vasculature (v). No fluorescence was present in the root hair (rh). (D) In cotyledons, fluorescence was detected in the veins (v) and hydathode (h). (E-F) Fluorescence was detected in the trichomes (t) of true leaves from two-week-old plants. Fluorescence+Auto, merge of fluorescence (green) and autofluorescence (red). Scale bars: 500 µm.

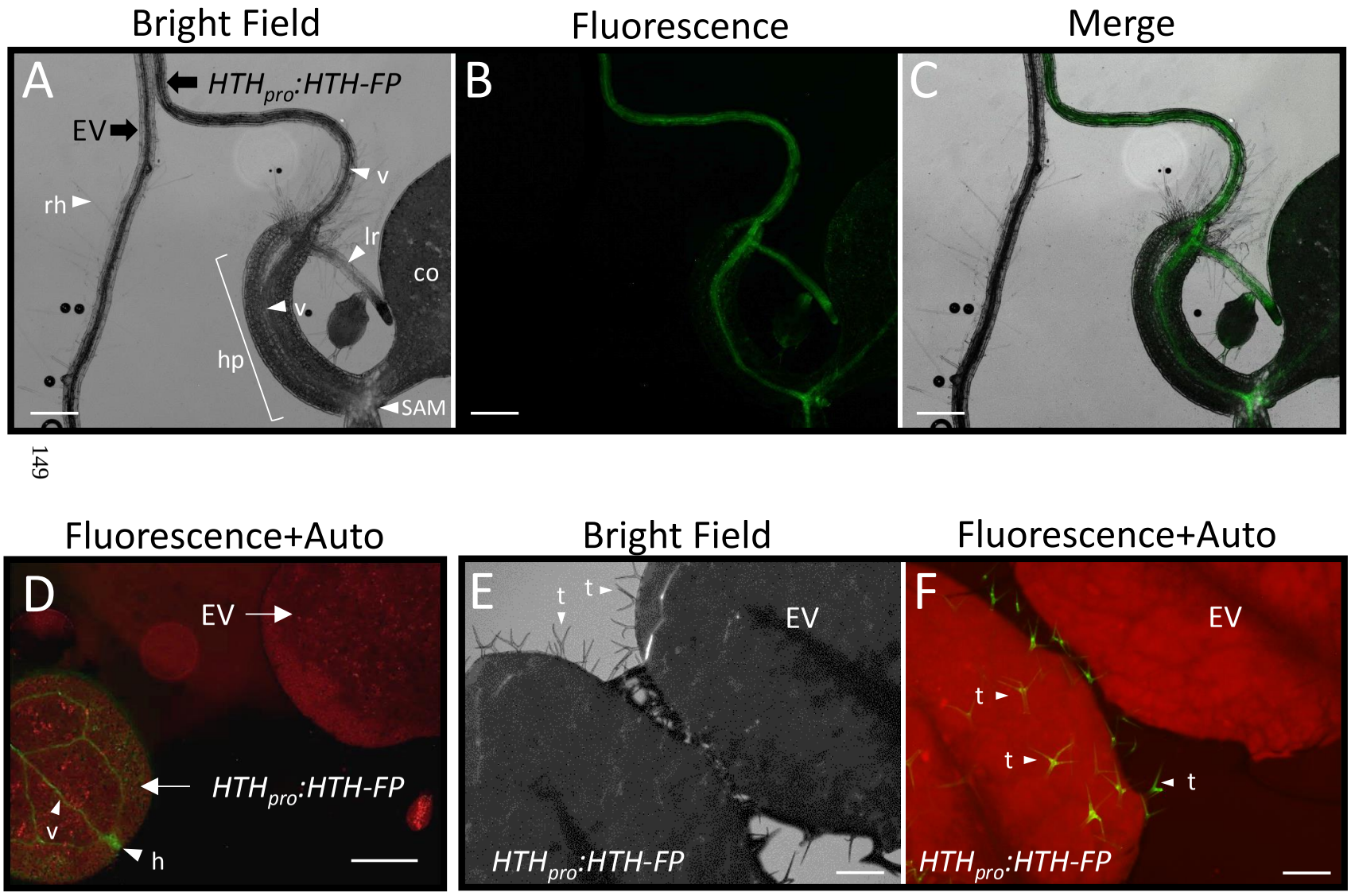

Bright Field
Fluorescence
Merge
A
B
C
HTH$_{pro}$:HTH-FP
EV
rh
v
lr
co
v
hp
SAM
Fluorescence+Auto
Bright Field
Fluorescence+Auto
D
E
F
EV
HTH$_{pro}$:HTH-FP
v
h
t
t
EV
HTH$_{pro}$:HTH-FP
EV
t
t
t
HTH$_{pro}$:HTH-FP

Figure 3.6 Epifluorescence micrographs showing HTH-FP localization in the root of 4-day-old *HTH_{pro}:HTH-FP* seedlings. (A-D) HTH-FP was localized to the stele (s) of the primary root and emerging lateral root (lr). (E-H) The HTH-FP localization in the lateral root continued to show in a more developed lateral root. No fluorescence was detected in epidermal cells and root hairs (rh). (I-L) No expression was observed in empty vector controls. Fluorescence+Auto, merge of fluorescence (green) and autofluorescence (blue). Scale bar: 100 µm.

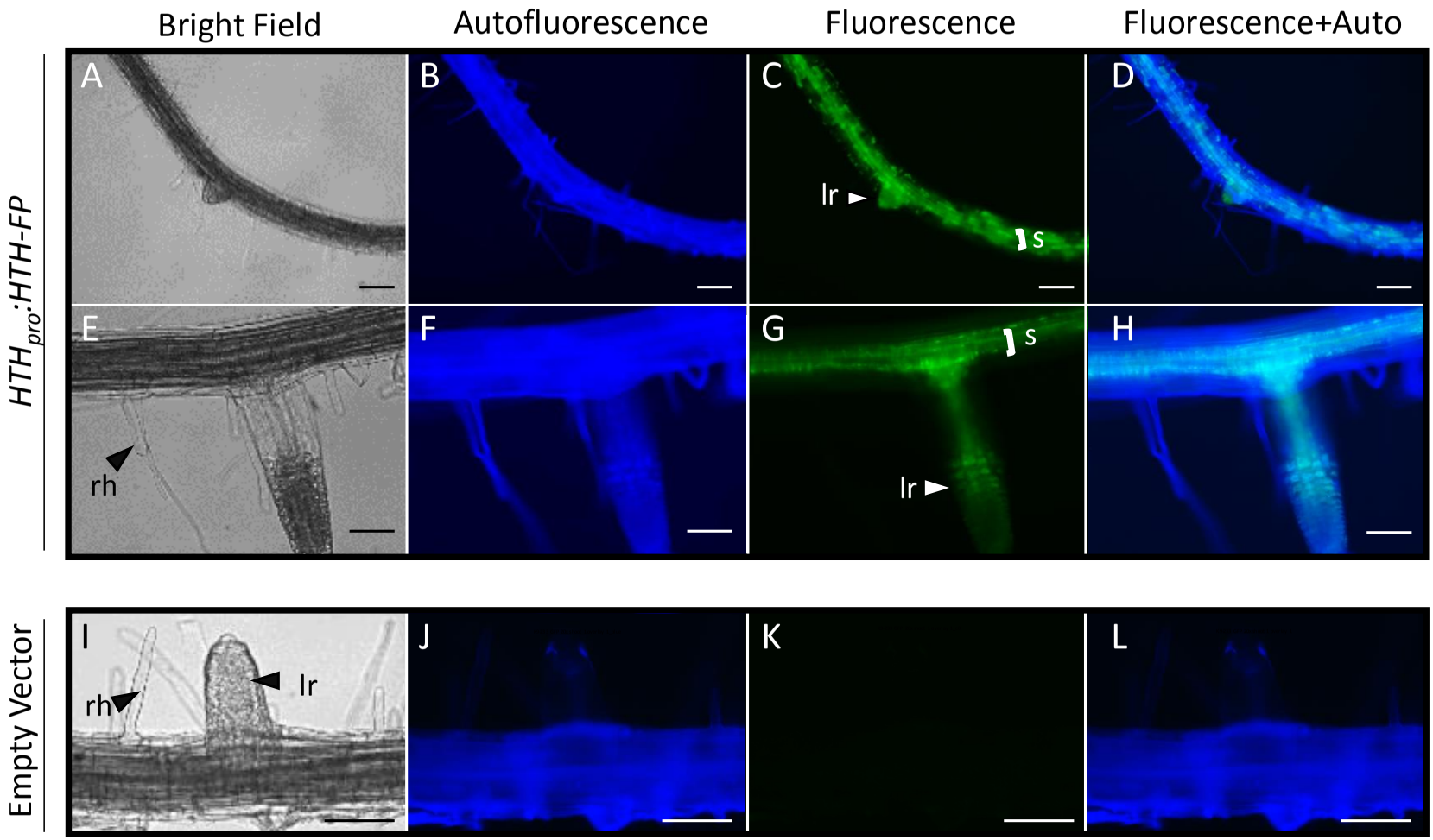

Bright Field
Autofluorescence
Fluorescence
Fluorescence+Auto
HTH$_{pro}$:HTH-FP
Empty Vector
A
B
C
D
E
F
G
H
I
J
K
L
lr
s
rh

Figure 3.7 Epifluorescence micrographs showing HTH-FP localization in above-ground tissues of *HTH*$_{pro}$*:HTH-FP* transgenic plants. (A-F) Four-day-old seedlings. (G-H) Two-week-old plants. HTH-FP fluorescence was observed in young *HTH*$_{pro}$*:HTH-FP* seedlings (A-C), whereas no fluorescence in empty vector seedlings (D). (E) The fluorescence was apparent in vasculature (v) of cotyledons (co) and hypocotyl (hp), as well as trichomes (t). (F) At higher magnification, HTH-FP fluorescence was observed in the shoot apical meristem (SAM) region and the epidermal cells (ep) of cotyledon petioles (cp) and emerging leaves (el). Fluorescent punctates (p) were observed in some epidermal cells. (G) In two-week-old juvenile plants, HTH-FP was prominent in the stipules (st, arrows). (H) A stipule shown at higher magnification. Scale bar: (A-F) 500 µm; (G-H) 300 µm.

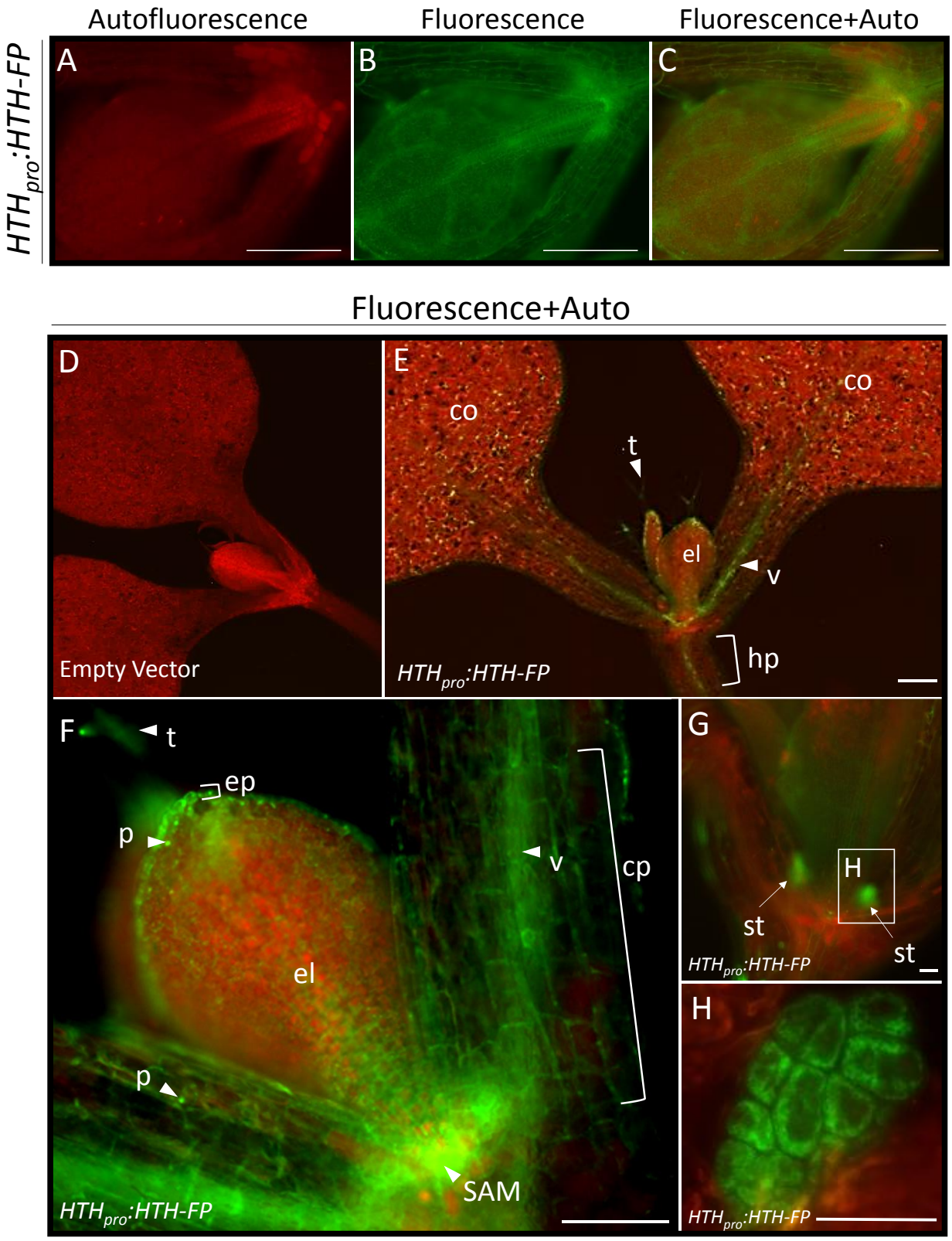

Autofluorescence
Fluorescence
Fluorescence+Auto
HTH_pro:HTH-FP
A
B
C
Fluorescence+Auto
D
Empty Vector
E
co
co
t
el
v
hp
HTH_pro:HTH-FP
F
t
ep
p
v
cp
el
p
SAM
HTH_pro:HTH-FP
G
H
st
st
HTH_pro:HTH-FP
H
HTH_pro:HTH-FP

Figure 3.8 Micrographs showing HTH-FP in cotyledon epidermal cells. (A-H) Images captured using an epifluorescence microscope. (I-L) Images captured using a confocal laser scanning microscope. (A) Detection of HTH-FP in cotyledon epidermal cells (ep), veins (v) and hydathodes (h). (B) HTH-FP was observed in pavement cells and guard cells (arrow). (B-inset) A merged image of HTH-FP fluorescence and nuclei staining (blue) showing a guard cell at higher magnification. (C-E) No expression in the empty vector plants was detected. (F-H) HTH-FP is restricted to the epidermis, and fluorescent punctates (p) were observed. (I-L) Confocal images also indicated epidermal localization of HTH-FP. Fluorescence+Auto, merge of fluorescence (green) and autofluorescence (red). Hoechst nuclei staining is shown in blue. Scale bar: (A-B) 100 µm; (C-L) 50 µm.

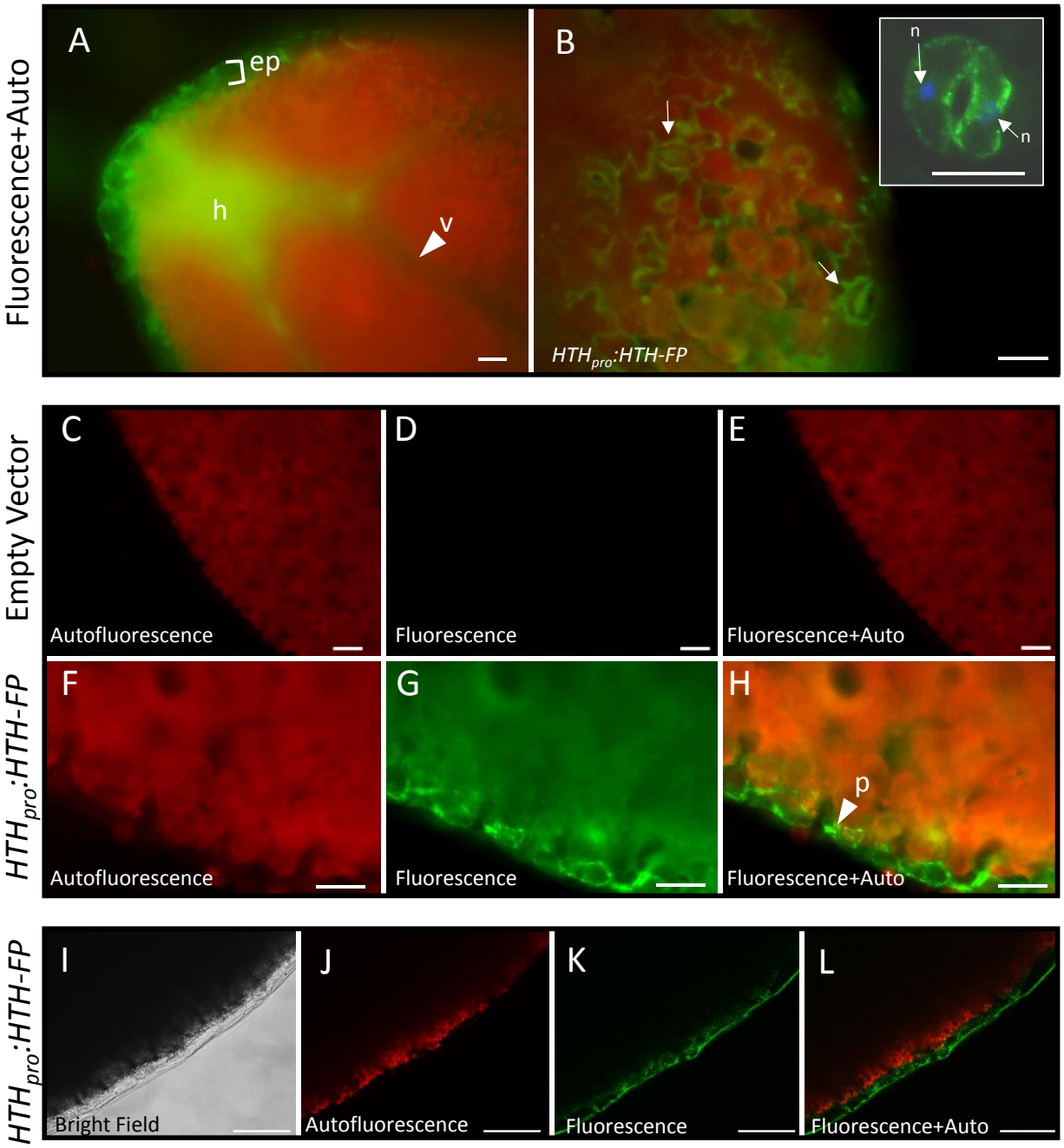

Fluorescence+Auto
A
ep
h
v
B
n
n
HTHpro:HTH-FP
Empty Vector
C
Autofluorescence
D
Fluorescence
E
Fluorescence+Auto
HTHpro:HTH-FP
F
Autofluorescence
G
Fluorescence
H
p
Fluorescence+Auto
HTHpro:HTH-FP
I
Bright Field
J
Autofluorescence
K
Fluorescence
L
Fluorescence+Auto

Figure 3.9 Epifluorescence micrographs showing HTH-FP localization in floral tissues. (A-B) young flower buds, (C) pistils, (D-I) sepals and petals, (J-M) pistil cross-sections, and (N-P) anthers. (C-D, G) Side-by-side comparison of HTH-FP fluorescence in *HTH_{pro}:HTH-FP* and empty vector (EV) transgenic lines. HTH-FP was detected in flower buds (A), the ovary wall (K), vasculature (E, H) and epidermis (F, I) of sepals and petals. (J-M) In the pistil, fluorescence was confined to the ovary wall. (N-P) Whole mount and cross section images of anthers showing HTH-FP fluorescence in the epidermis (ep), vascular bundle (vb) and cells between adjoining locules (lc, arrow). Only autofluorescence and no HTH-FP was detected for pollen grains (pg). Ws, Wassilewskija wildtype plant. Fluorescence+Auto, merge of fluorescence (green) and autofluorescence (red). Scale bar: (A-C) 100 µm; (D-P) 50 µm.

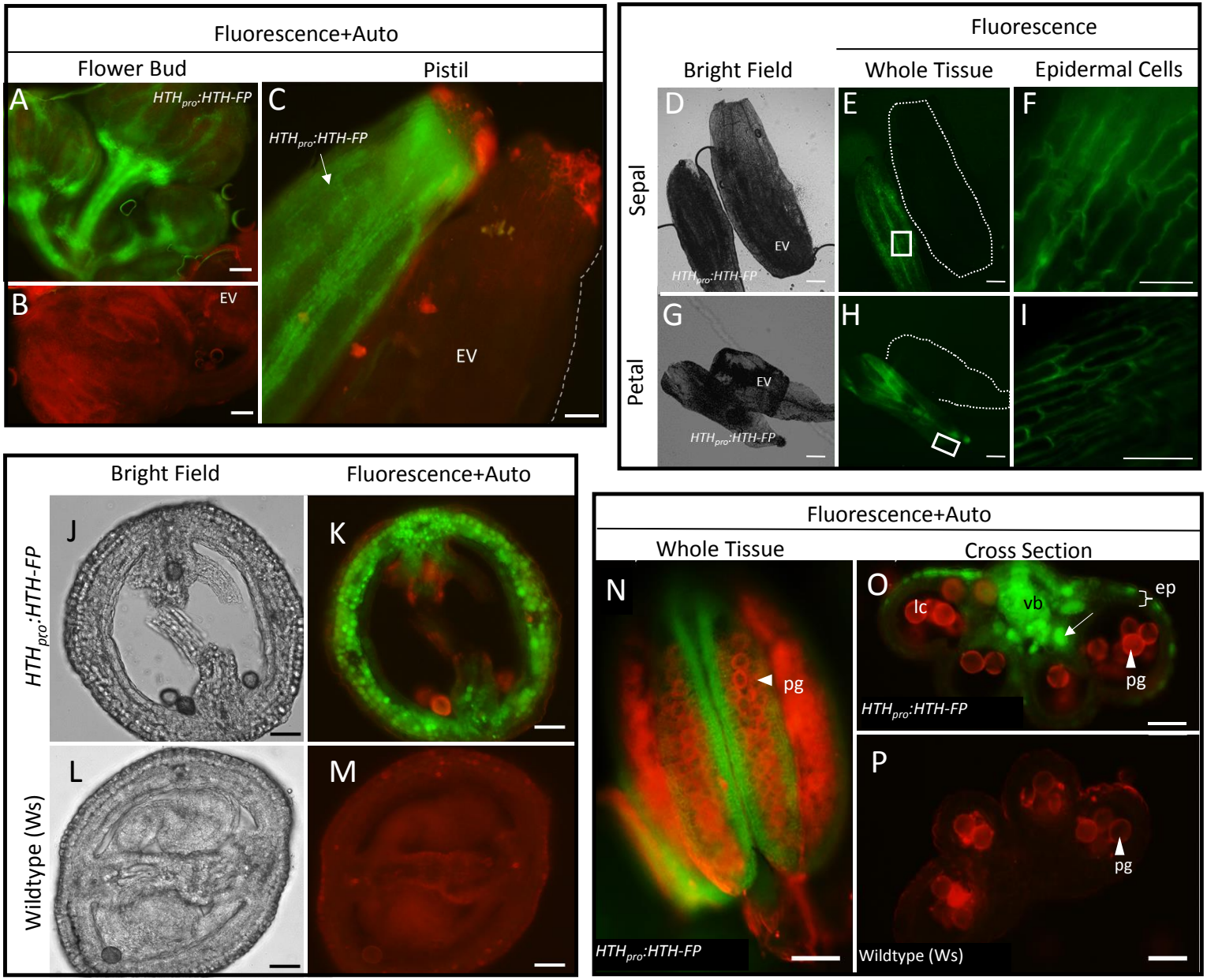

Fluorescence+Auto
Flower Bud
Pistil
A
HTH$_{pro}$:HTH-FP
B
EV
C
HTH$_{pro}$:HTH-FP
EV
Fluorescence
Bright Field
Whole Tissue
Epidermal Cells
Sepal
Petal
D
E
F
G
H
I
EV
HTH$_{pro}$:HTH-FP
Bright Field
Fluorescence+Auto
HTH$_{pro}$:HTH-FP
Wildtype (Ws)
J
K
L
M
Fluorescence+Auto
Whole Tissue
Cross Section
N
pg
HTH$_{pro}$:HTH-FP
O
lc
vb
ep
pg
HTH$_{pro}$:HTH-FP
P
pg
Wildtype (Ws)

Figure 3.10 Micrographs of HTH-FP localization in ovules prior to fertilization. (A) Illustrations of the seven female gametophyte developmental stages (FG1-FG7). (B) An illustration of an Arabidopsis ovule showing the seven cells that make up the mature female gametophyte. (C) A differential interference contrast (DIC) image of a fully developed ovule. In the dissected ovary at stage FG6-7, no fluorescence was observed in the ovules isolated from empty vector plants (D, in the orientation as panel C). (E) Fluorescence was detected at the chalazal end of embryo sacs of *HTH_{pro}:HTH-FP* plants. An ovule is outlined and is shown in the same orientation as the ovule in panel B. (F-M) HTH-FP localization in unfertilized ovules at different stages. The female gametophyte developmental stages are matched with the ovule sampling time relative to anthesis. Two days before anthesis, HTH-FP became detectable at the chalazal end of the ovule and it became more diffuse at anthesis. (N-O) In this ovule, fluorescence was less diffuse and was detected in discreet entities in the chalazal region, coincident with the antipodal cells. The embryo sac is outlined. The boxed region is shown at higher magnification panel P. (P) The arrowheads point to distinct entities. ac: antipodal cell; cc, central cell; sc, synergid cell; ec, egg cell; f, funiculus; mp, micropyle. Fluorescence+Auto, merge of fluorescence (green) and autofluorescence (red). Fluorescence+Auto, merge of fluorescence (green) and autofluorescence (red); Fluorescence+Auto+BF, merge of fluorescence, autofluorescence and bright field images (black and white). Scale bar: 100 µm.

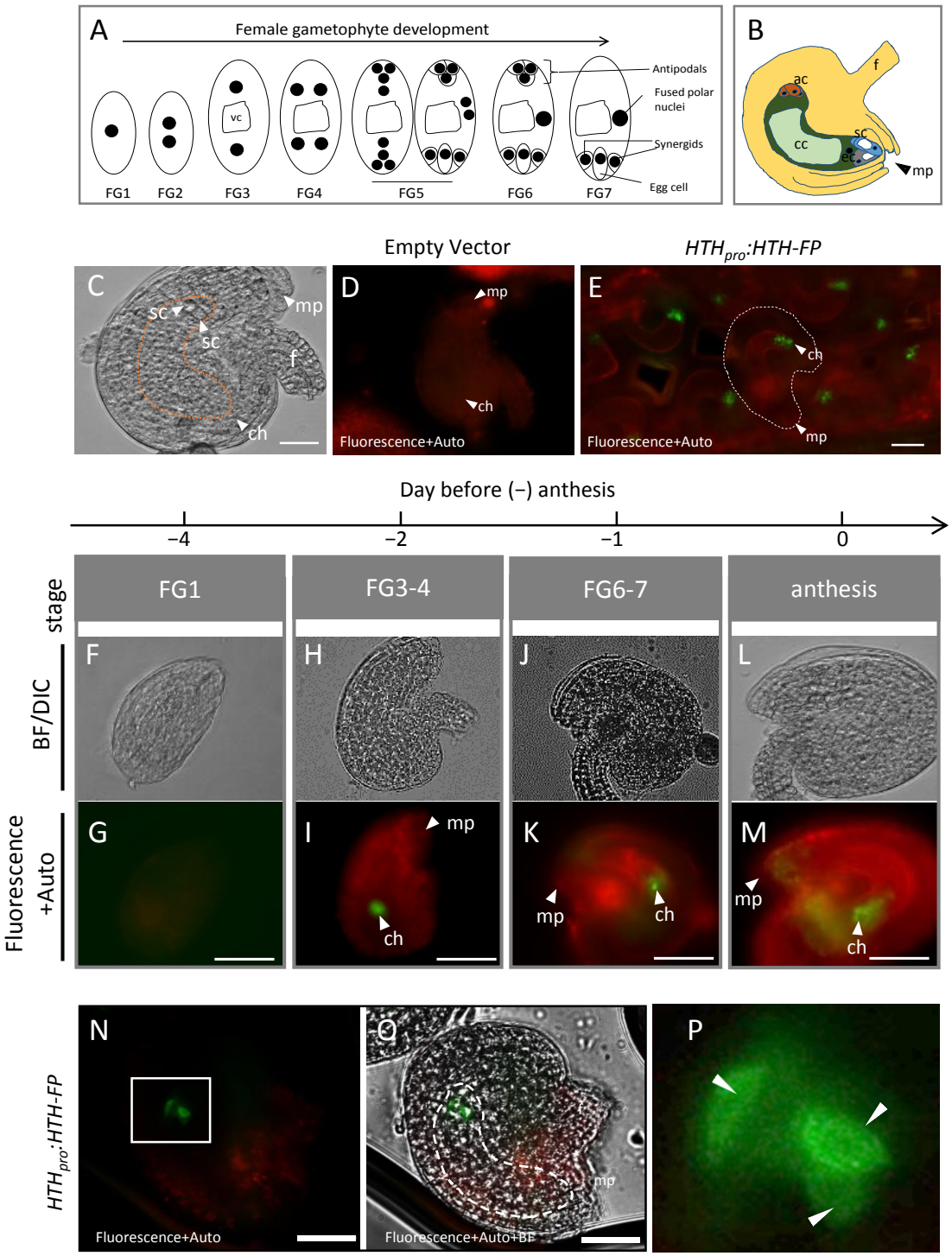

A
Female gametophyte development
vc
Antipodals
Fused polar nuclei
Synergids
Egg cell
FG1
FG2
FG3
FG4
FG5
FG6
FG7
B
f
ac
cc
sc
mp
Empty Vector
HTH_pro:HTH-FP
C
sc
mp
f
ch
D
mp
ch
Fluorescence+Auto
E
ch
mp
Fluorescence+Auto
Day before (−) anthesis
−4
−2
−1
0
stage
FG1
FG3-4
FG6-7
anthesis
BF/DIC
F
H
J
L
Fluorescence +Auto
G
I
mp
ch
K
mp
ch
M
mp
ch
HTH_pro:HTH-FP
N
Fluorescence+Auto
O
mp
Fluorescence+Auto+BF
P

Figure 3.11 Micrographs showing HTH-FP localization in seeds at different developmental stages. Three days after anthesis, the integumental expression became apparent. The expression persisted to later developmental stages, particularly in the micropylar (mp) and chalazal (ch) seed coat. BF, bright field; Fluorescence+Auto, merge of fluorescence (green) and autofluorescence (red). Scale bar: 100 µm.

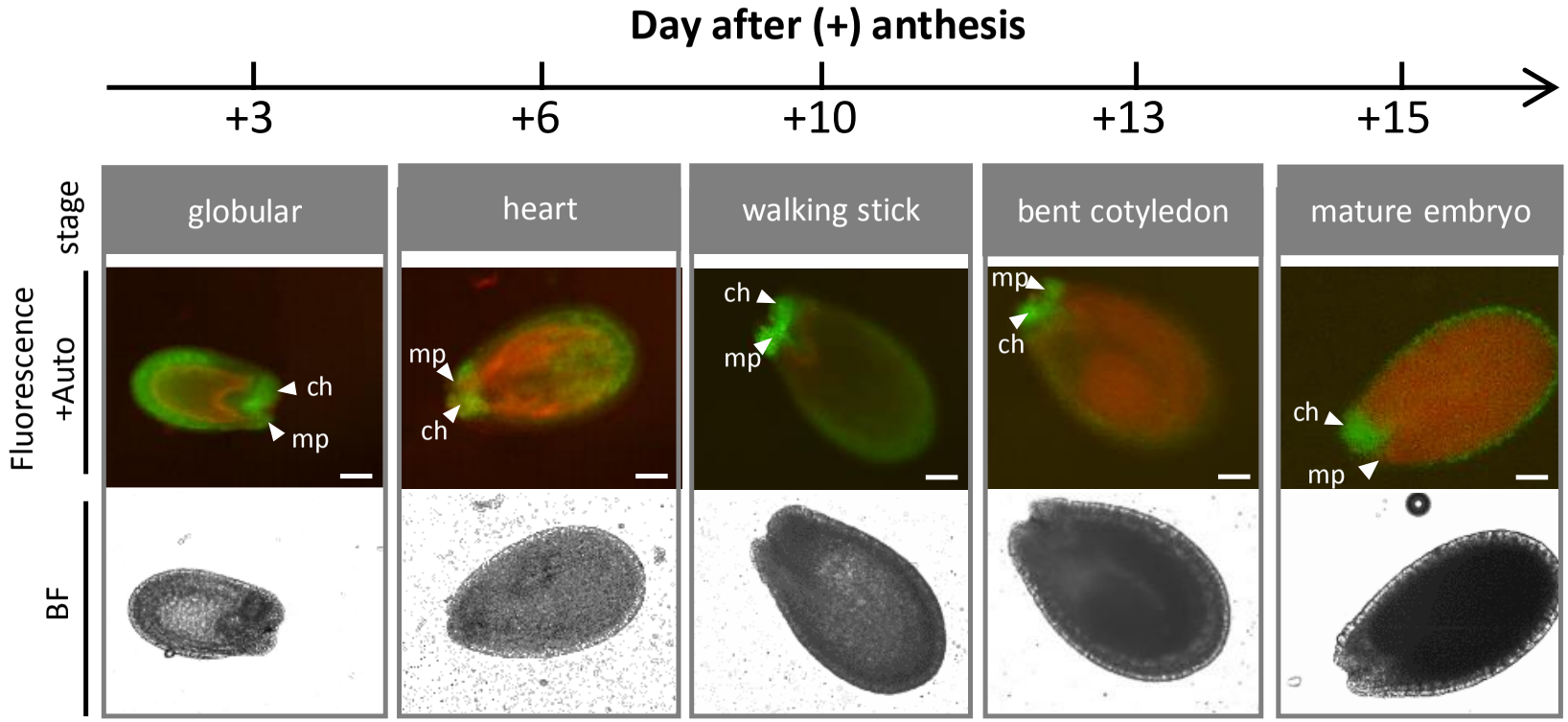
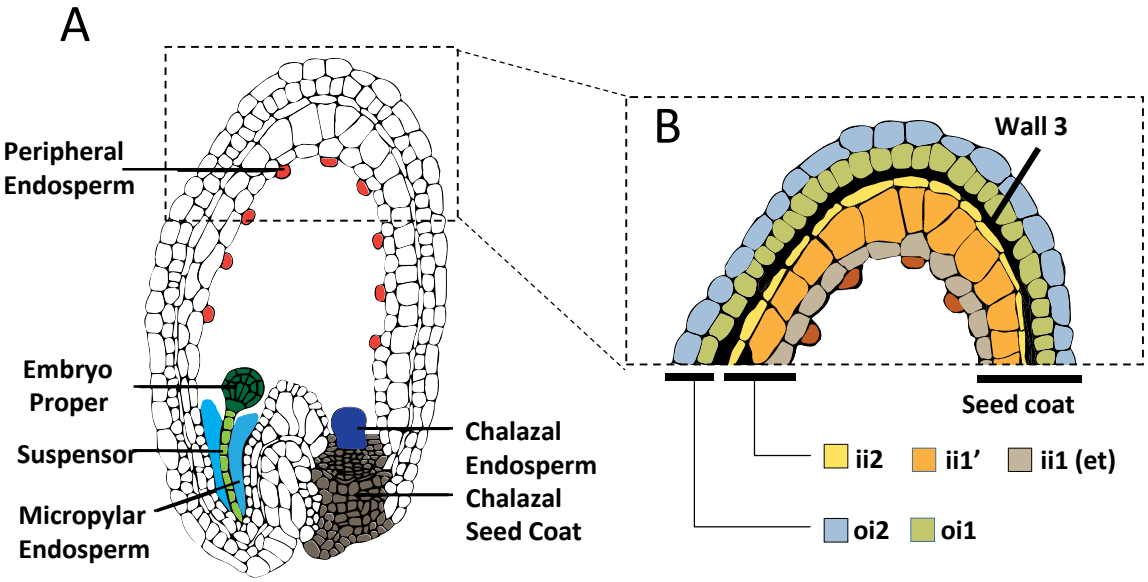

Day after (+) anthesis
+3
+6
+10
+13
+15
stage
globular
heart
walking stick
bent cotyledon
mature embryo
Fluorescence
+Auto
BF
ch
mp

Figure 3.12 Seed structure and HTH-FP localization. Schematic diagrams showing the seed (A) and seed coat structure (B; modified from www.seedgenenetwork.net). The seed coat is composed of the outer integument (oi) and the inner integument (ii). Each integument is comprised of an outer ('2') and an inner ('1') layer. Between ii2 and ii1 (also known as endothelium, et), an internal cell layer (ii1') is present towards the chalazal zones of the seed coat. The outer integument is separated from the inner integument by an electron-dense cell wall layer ('wall 3') that is rich in cutin-like material. The vast majority of the wall material deposited in wall 3 is produced by the oi1 layer. (C-E) Epifluorescence micrographs showing HTH-FP localization in +7 DPA maturing seeds. HTH-FP (fluorescence, green) is localized to the integument that is external to the embryo sac (autofluorescence, red). (F-H) Images at higher magnification indicates that HTH-FP was not detected in the outermost oi2 layer but the inner layer of the outer integument (oi1) and possibly also inner integumentary layers. (H) Cross-section through the ovary showing an ovule ten days after anthesis. DPA, day post anthesis; BF, bright field; Fluorescence+Auto, merge of fluorescence (green) and autofluorescence (red); Fluorescence+BF, merge of fluorescence and bright field; Fluorescence+Auto+BF, merge of fluorescence, autofluorescence and bright field. Scale bar: 100 µm.


A
Peripheral Endosperm
Embryo Proper
Suspensor
Micropylar Endosperm
Chalazal Endosperm
Chalazal Seed Coat
B
Wall 3
Seed coat
ii2
ii1'
ii1 (et)
oi2
oi1

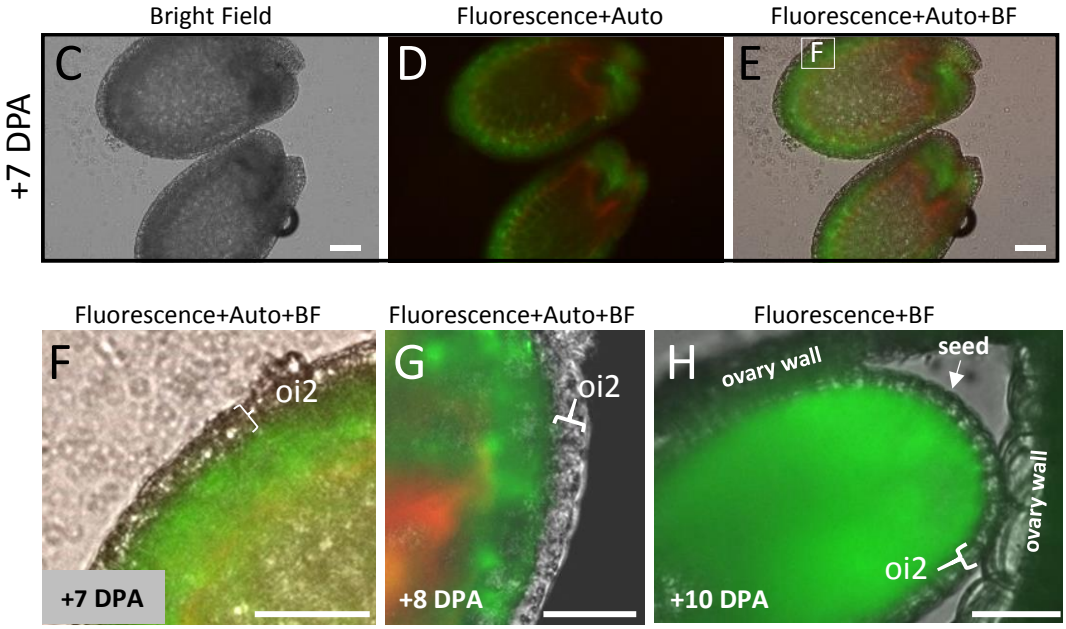

Bright Field
Fluorescence+Auto
Fluorescence+Auto+BF
+7 DPA
C
D
E
F
Fluorescence+Auto+BF
Fluorescence+Auto+BF
Fluorescence+BF
F
oi2
+7 DPA
G
oi2
+8 DPA
H
ovary wall
seed
ovary wall
oi2
+10 DPA

Figure 3.13 Confocal images of HTH-FP in the seed coat of a developing seed (+20 DPA) at the post mature stage. (A-C) No fluorescence was detected in the Ws wildtype seed coat. (D-F) At this later developmental stage, HTH-FP was detected in the outer layer (oi2) in addition to the inner layer (oi1) of the outer integument. (G) An illustration of the outer integuments of a developing seed at the post mature stage. Amyloplast-containing columella are present in the oi2 cells. (H-I) The inset in panel F at higher magnification. Two layers of the outer integument, oi2 and oi1, are indicated. Within a single oi2 cell, the columella has developed in the center and two mucilage compartments are on the side. (J-L) Top view of the oi2 layer. No HTH-FP was not detected in amyloplasts (arrowhead) within the columella but rather in the cytosol. mu, mucilage; co, columella. Scale bar: (A-F, H-I) 50 µm; (J-L) 20 µm.

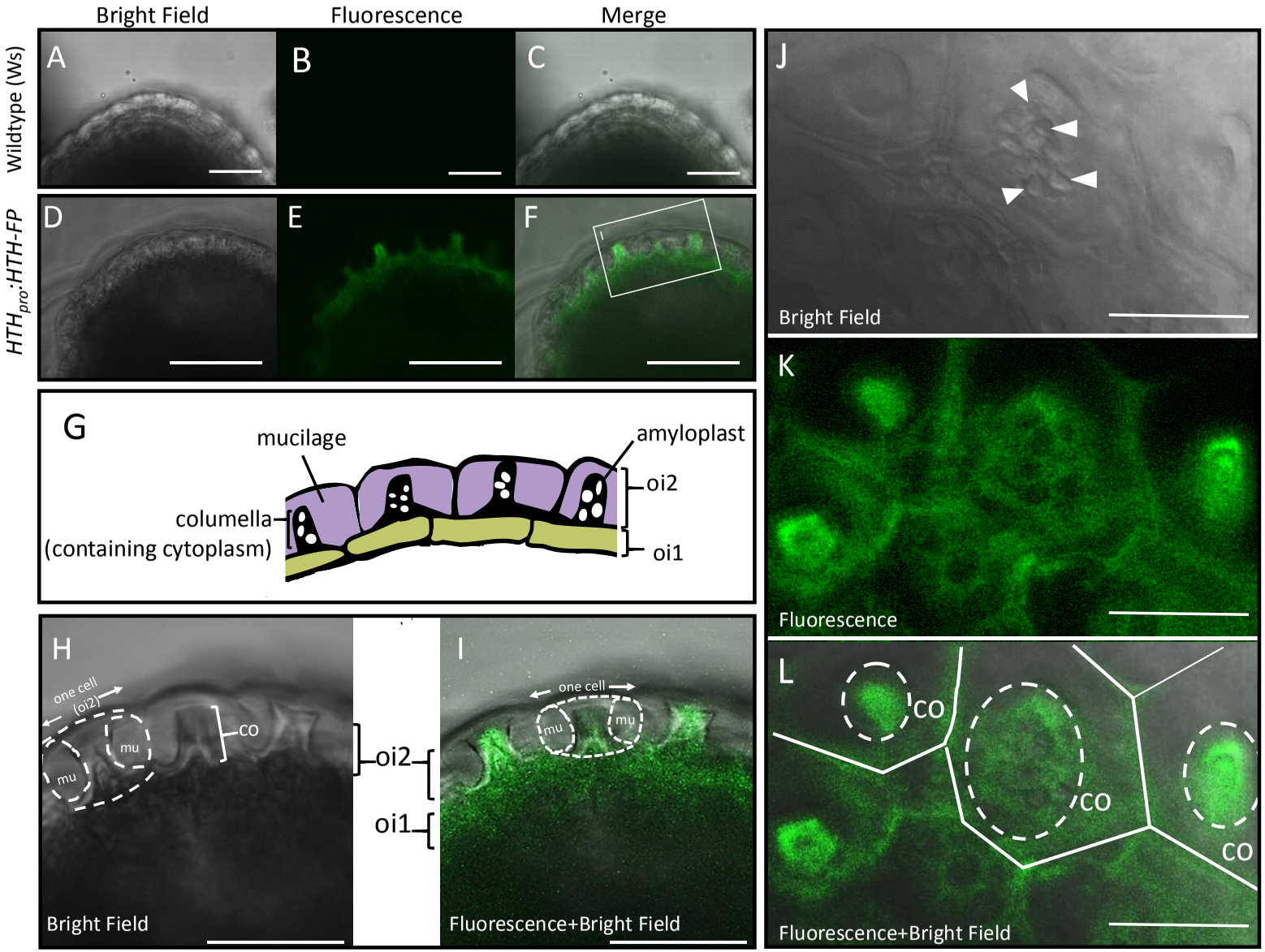

Bright Field
Fluorescence
Merge
Wildtype (Ws)
HTH_pro:HTH-FP
A
B
C
D
E
F
G
mucilage
amyloplast
oi2
columella
(containing cytoplasm)
oi1
H
one cell
(oi2)
mu
mu
co
oi2
oi1
Bright Field
I
one cell
mu
mu
Fluorescence+Bright Field
J
Bright Field
K
Fluorescence
L
co
co
co
Fluorescence+Bright Field

Figure 3.14 Seed images of wildtype and mutant plants. (A-D) *hth*-13, *hth*-14 and *hth*-15 are T-DNA insertion mutants in the Col background. (E-H) *hth*-1, *hth*-4 and *hth*-8 are mutants isolated in the L*er* background harboring single point mutations. (I-J) The *hth*-9 mutant harbours a single point mutation and was isolated in the Ws background. (K-L) Seed derived from two independent transgenic lines in the *hth*-9 background. (B, C, G, H-inset) Misshaped seeds (arrow) are shown at higher magnification. (G-inset) Two of the misshaped seeds are attached (arrowheads). Ecotype: L*er*, Landsberg *erecta*; Ws, Wassilewskija; Col, Columbia. Scale bar: 1 mm.

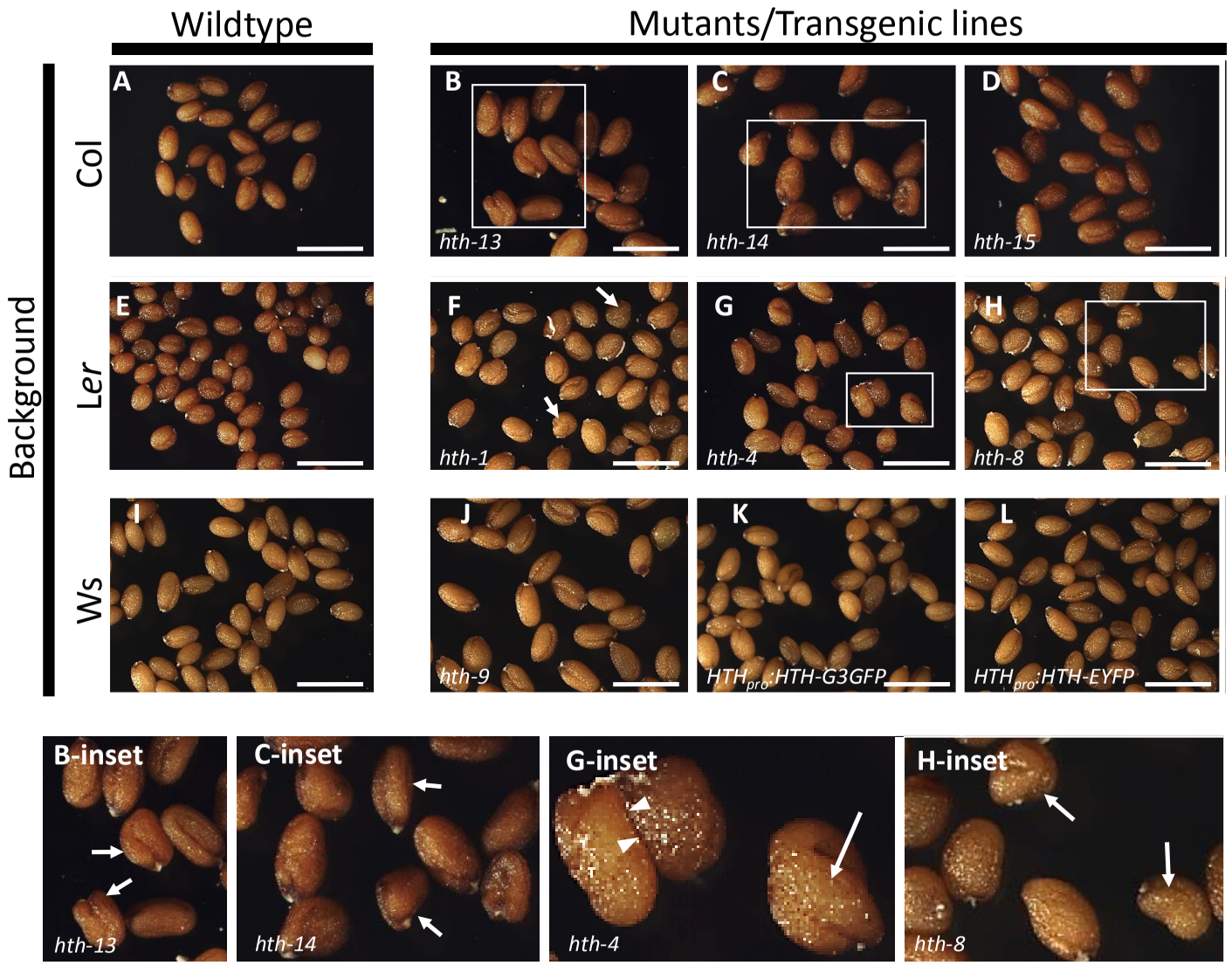

Wildtype
Mutants/Transgenic lines
Background
Col
Ler
Ws
A
B
C
D
E
F
G
H
I
J
K
L
hth-13
hth-14
hth-15
hth-1
hth-4
hth-8
hth-9
HTH pro:HTH-G3GFP
HTH pro:HTH-EYFP
B-inset
C-inset
G-inset
H-inset
hth-13
hth-14
hth-4
hth-8

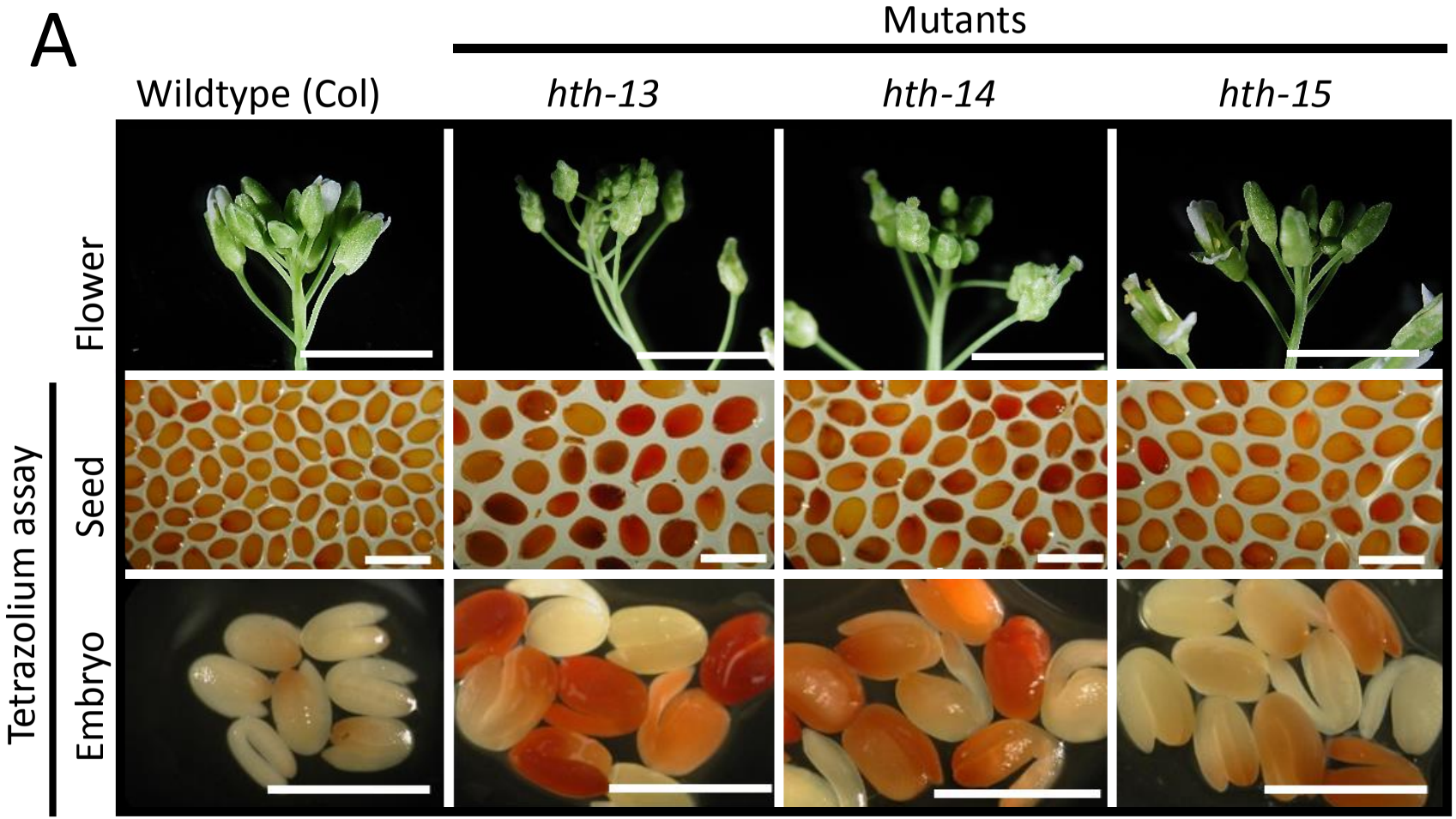
Table 3.1 A table summarizing the weight of wildtype and mutant seeds. The relative seed weight for each wildtype ecotype background was normalized to 100%. Seed weight differs among ecotype backgrounds (Ws> Col> L*er*). The seed weight of mutants was consistently greater than seed derived from the corresponding wildtype lines. The weight of seeds from two transgenic plant lines (HTH_{pro}*:HTH-G3GFP*/EYFP) resembles that of Ws. Ecotype: Col, Columbia; L*er*, Landsberg *erecta*; Ws, Wassilewskija.

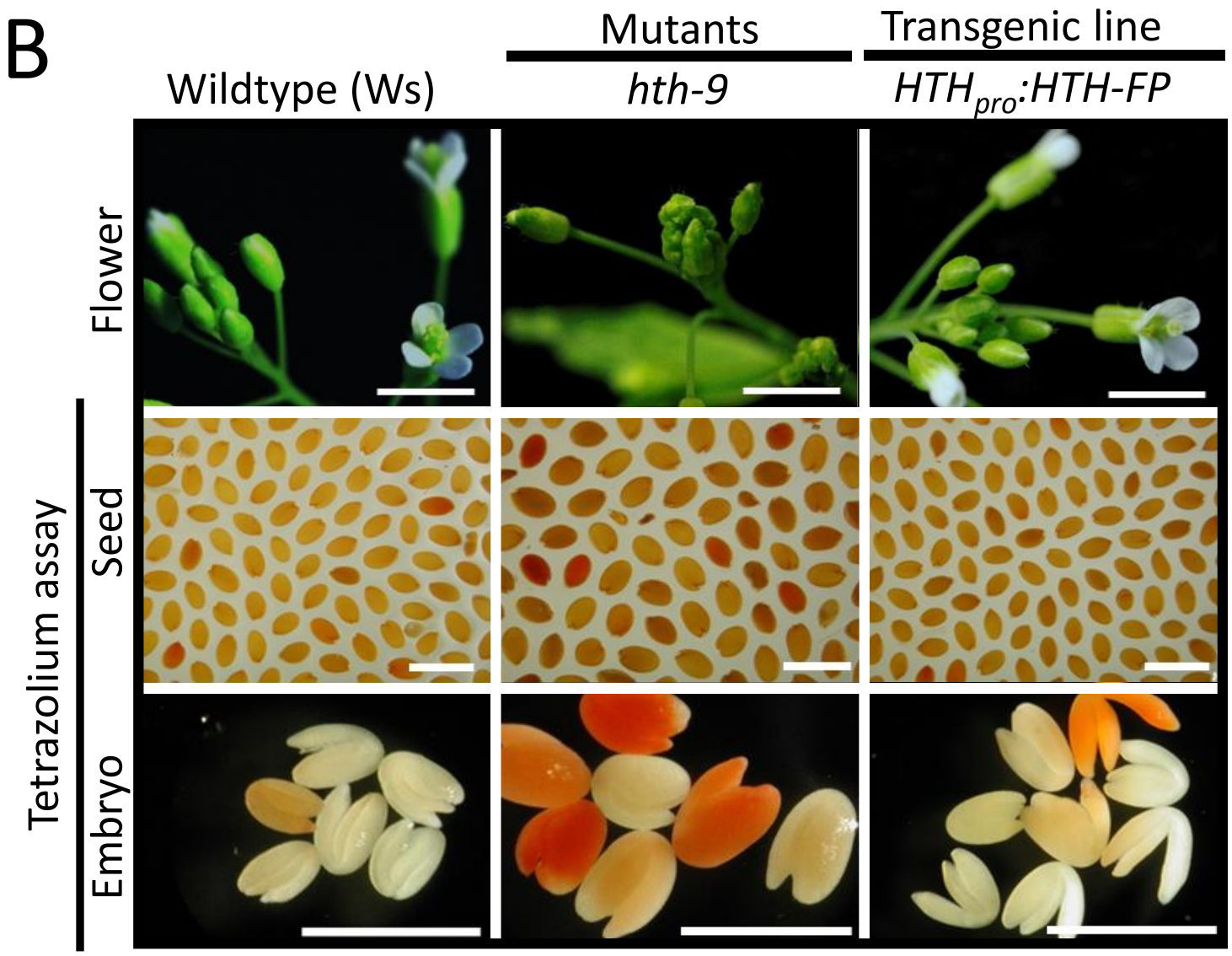
Weight increase for mutant seeds

	Weight (g)/2000 seeds)	% of Increase*
Col	52.7±3.3	100%
hth-13	123.2±6.7	234%
hth-14	106.2±6	202%
hth-15	77±1.5	146%
L*er*	40.7±2.5	100%
hth-1	71.3±8.2	175%
hth-4	72.3±7.6	178%
hth-8	67.6±8.9	166%
Ws	79.3±12.4	100%
hth-9	143.3±17	181%
HTH_{pro}:*HTH-G3GFP*	80.5±6.6	101%
HTH_{pro}:*HTH-EYFP*	87.1±5.8	110%

* the wildtype plants in each background are set at 100%.

Figure 3.15 Floral phenotypes and seed coat permeability of plants. Flowers, tetrazolium-treated seeds and dissected embryos of isolated wildtype, mutants and transgenic plants in the Col (A), Ws (B) and L*er* (C) backgrounds. Without exception, mutant seeds displayed more prominent staining. The transgenic HTH_{pro}*:HTH-FP* plants showed fully open and unfused flowers, normal seed size and seed coat permeability. L*er* wildtype seeds were more readily stained than those of Ws and Col. The ecotype background is indicated in each case. Scale bar: flower, 5 mm; seed and embryo, 1 mm.


A
Mutants
Wildtype (Col)
hth-13
hth-14
hth-15
Flower
Tetrazolium assay
Seed
Embryo


B
Wildtype (Ws)
Mutants
hth-9
Transgenic line
HTH pro:HTH-FP
Flower
Tetrazolium assay
Seed
Embryo

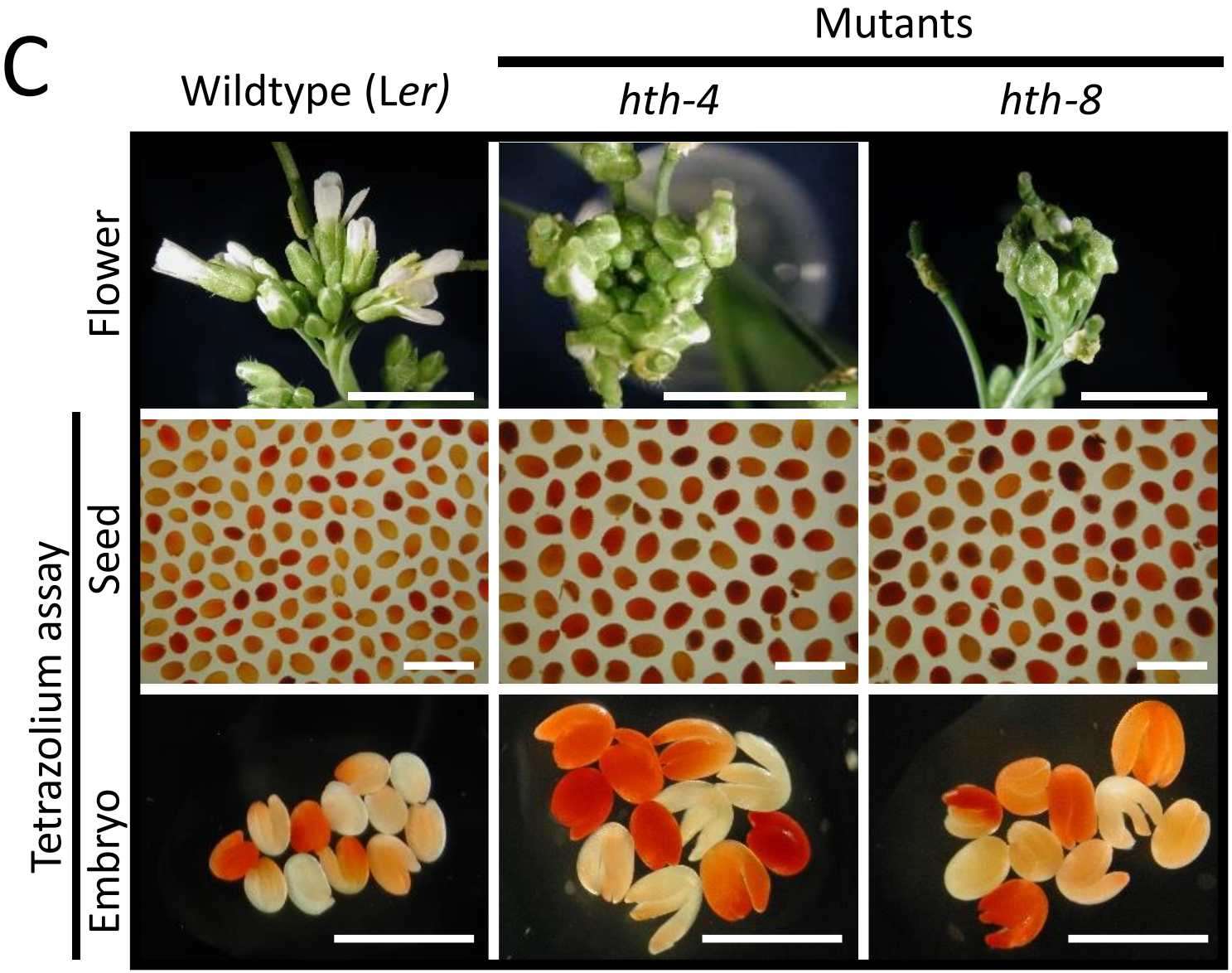

C
Mutants
Wildtype (Ler)
hth-4
hth-8
Flower
Tetrazolium assay
Seed
Embryo

Figure 3.16 Micrographs showing HTH-FP localization in hypocotyl cells of four-day-old *HTH_{pro}:HTH-FP* transgenic seedlings. (A) HTH-FP was observed in discreet bodies (arrowhead) in young seedlings. (B-E) Images of hypocotyl epidermal cells. HTH-FP was sometime predominantly localized to in the bodies and sometime also in a reticular network (bracket) (D). The bodies were spindle shaped and are typically 5-10 µm in size. (F-M) Images of time-series showing movement of HTH-FP-containing bodies in cotyledon petiole epidermal cells. These spindle-shaped bodies (arrowhead) moved approximately 0.5-1 µm/sec. The direction of movement is indicated (dashed arrow). (N-P) Images of *HTH_{pro}:HTH-FP* transgenic cotyledon petiole epidermal cells stained with the mitochondrial dye, TMRM. HTH-FP fluorescence does not colocalize with the TMRM-stained mitochondria. HTH-FP (G3GFP; 505-530 nm) and TMRM (550-600 nm) were detected in two emission wavelength ranges. A-E, epifluorescence images; F-P, confocal images. Scar bar: (A-E) 50 µm; (F-P) 20 µm.

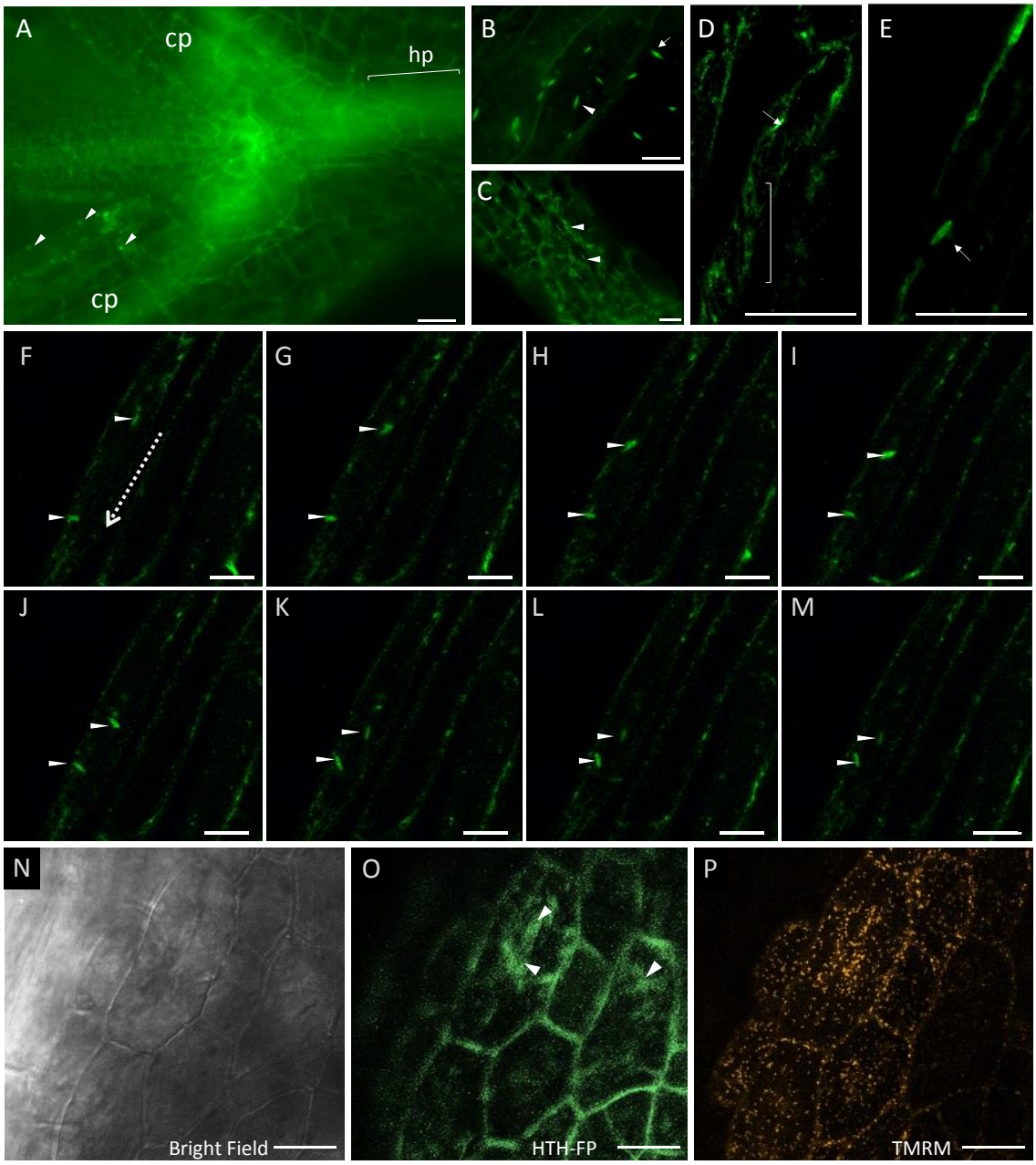

A
cp
hp
cp
B
C
D
E
F
G
H
I
J
K
L
M
N
Bright Field
O
HTH-FP
P
TMRM

Figure 3.17 Colocalization of HTH-FP and erRFP in hypocotyl cells of four-day-old seedlings. (A) Ws wildtype harbouring the *erRFP* construct. (B) HTH_{pro}*:HTH-FP* transgenic plants. (C-I) HTH_{pro}*:HTH-FP* transformed with the *erRFP* construct. A-E, epifluorescence micrographs; F-I, confocal micrographs. Pearson's colocalization coefficient based on region of interest (ROI) analysis: 0.79 ± 0.063 (n = 10). Merge, overlapping images of HTH-FP (green) and erRFP (magenta). Scar bar: 20 µm.

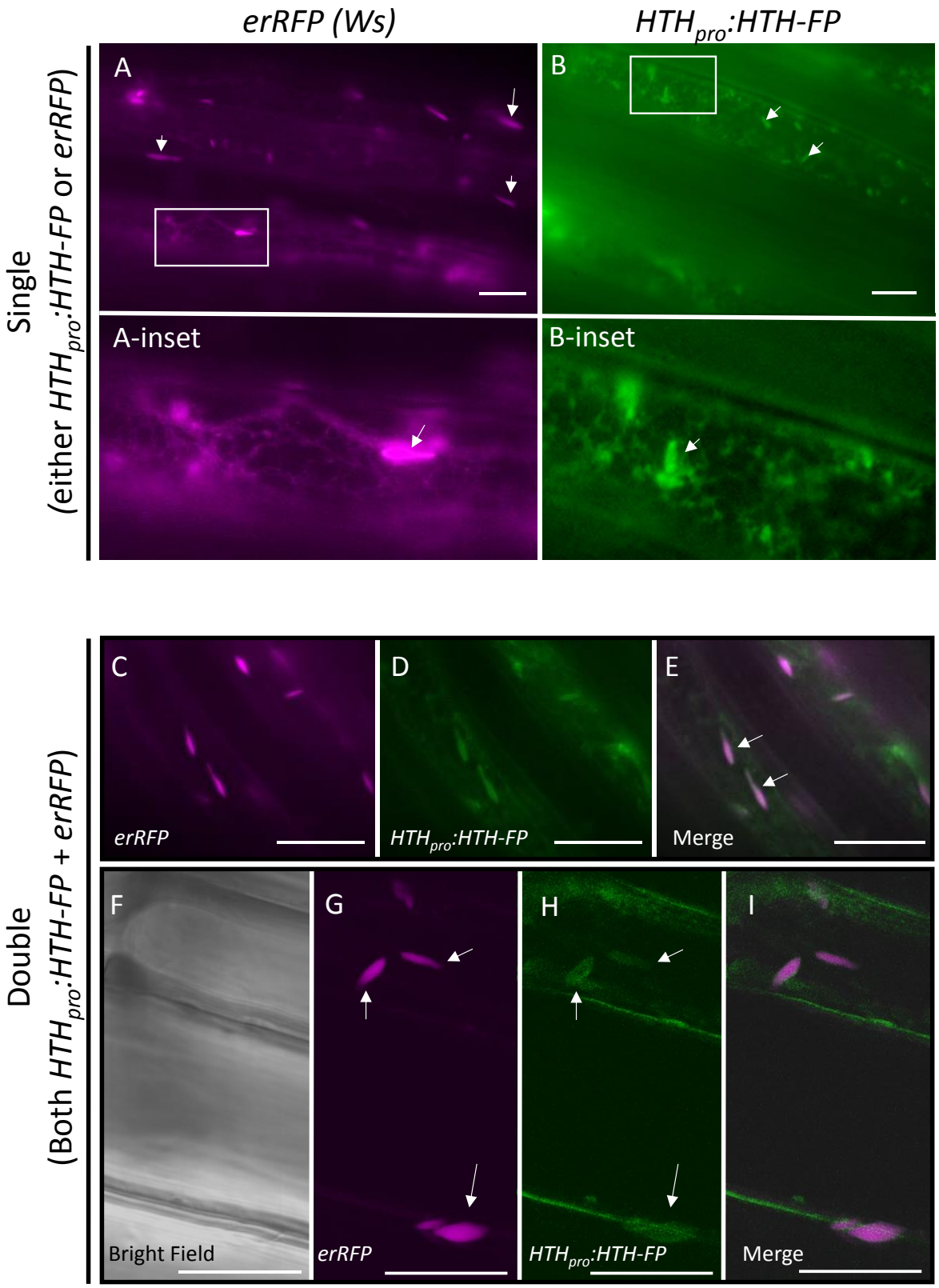
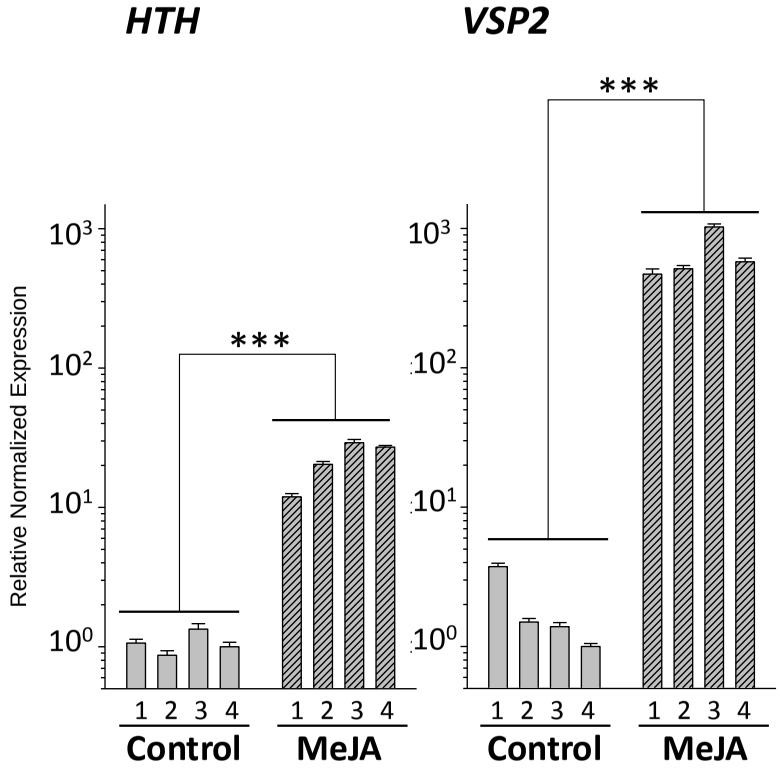

erRFP (Ws)
HTH$_{pro}$:HTH-FP
Single
(either HTH$_{pro}$:HTH-FP or erRFP)
A
B
A-inset
B-inset
Double
(Both HTH$_{pro}$:HTH-FP + erRFP)
C
D
E
erRFP
HTH$_{pro}$:HTH-FP
Merge
F
G
H
I
Bright Field
erRFP
HTH$_{pro}$:HTH-FP
Merge

Figure 3.18 MeJA-induced changes in *HTH* expression and HTH-FP localization . (A-B) RT-qPCR analysis of *HTH* expression in wildtype Ws plants. (A) *HTH* expression was elevated in 11-day old MeJA-treated Ws. Exposure to MeJA increased *VSP2* (*VEGETATIVE STORAGE PROTEIN 2*) expression. Numbers 1-4 indicate four biological replicates. Three technical repeats were performed for each biological sample. Error bar: 1 standard error of the technical repeats. **$p < 0.001$ (*t*-test). (B) The coefficient variances and M values of housekeeping genes used to normalize *HTH* and *VSP2* expression. (C) The dark purple colouration is due to anthocyanin deposition. Only MeJA-treated Ws plants gave rise to purple petioles. (D-E) HTH_{pro}:HTH-FP plants were observed. Leaf midrib epidermal cells showing HTH-FP localization with and without MeJA treatment. Fluorescence was detected in the ER network for both control and MeJA-treated plants. More ER bodies (arrow) containing HTH-FP were observed in the MeJA-treated samples. Scale bar: 20 µm.

A

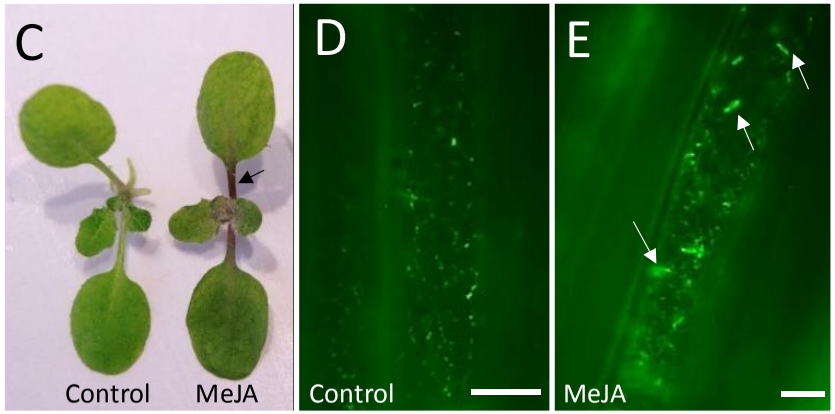


B

House keeping genes used for normalization

	Coeffient Variance	M Value
Tubulin 6	0.1897	0.6273
Actin 7	0.3134	0.7401
GAPC 2	0.3268	0.8040
Average	**0.2766 (< 0.5*)**	**0.7238 (< 1*)**

*Acceptable stability values for heterogeneous samples.



3.4 **Discussion**

3.4.1 **Floral fusion phenotypes and cuticle permeability**

The organ fusion of *hth* mutants is predominantly restricted to flowers, as seen in many other organ fusion mutants described previously. Lolle et al. (1998) isolated and characterized 29 independently derived mutations that led to organ fusion in Arabidopsis. Using complementation analyses, nine putative genes, including *HTH*, were identified. These mutants most commonly showed interorgan fusions within the flowers ranging in severity. Occasionally fusion events between vegetative tissues (e.g., *fiddlehead* and *thunderhead*) and abnormal ovule morphology (e.g., *hth* and *deadhead*) were observed. Results of the chlorophyll extraction assay revealed that the cuticle of most mutants were more permeable, suggesting the organ fusion phenotype was caused by a defective cuticle. Studies on these mutants have led to the characterization of genes such as *FIDDLEHEAD* involved in very long chain fatty acids (VLCFA; longer than 18 carbons) elongation reactions required for cuticular wax synthesis, and fatty acid analyses of *hth-12* mutants have suggested the involvement of *HTH* in cutin biosynthesis (Kurdyukov et al., 2006b; Pruitt et al., 2000; Yephremov et al., 1999).

Numerous *hth* mutant alleles have been generated by single point mutations, transposon mutagenesis, and T-DNA insertion. Floral fusion of mutant lines in the L*er* background (*hth-1* to *hth-11*, except *hth-9)* has been described previously (Krolikowski et al., 2003; Lolle et al., 1998). Mutants in the Col and Ws backgrounds were examined in the current study and have not been extensively evaluated previously. For mutants in L*er*, fusion propensity varied

greatly, with *hth-5* mutants showing weak fusion and *hth-1* mutants showing stronger fusion that completely blocked petal emergence. For *hth-4*, petal emergence was not completely prevented. Often the sepals remained joined at the distal end of the closed flower (Figure 3.1), revealing the "adhesion of calyx edge" phenotype. The *HTH* gene was originally identified as *ADHESION OF CALYX EDGES* (*ACE*) because of fusion between sepal epidermal cells (Araki et al., 1998). In *hth-9* mutants, tissue rupture along sepal edges was also evident. Interestingly, TBO staining was apparent along *hth-9* sepals margins (Figure 3.4), consistent with *hth-9* plants potentially having higher cuticle permeability at calyx edges.

In the Col background, T-DNA insertions in *HTH* also resulted in a floral organ fusion phenotype. Like *hth-9* mutants, *hth-13* and *hth-14* gave rise to shorter siliques that generally contained fewer seeds than wildtype (approximately a 90% decrease). For *hth-15* plants, however, floral fusion was less severe and silique length intermediate. This difference in fusion severity might be due to the position effect since the insertion sites of *hth-13* and *hth-14* are in the exons of the HTH gene, and the insertion of *hth-15* is located in the 5' upstream putative promoter region (Figure 1.8).

3.4.2 *HTH* tissue expression

To date, *HTH* expression has been studied using methods that include promoter-reporter constructs (HTH_{pro}:*GFP*), *in situ* mRNA hybridization and reverse transcription-polymerase chain reaction (RT-PCR) assays. Krolikowski et al. (2003) used RT-PCR and *in situ* mRNA hybridization to show that *HTH* mRNA is expressed in all organs tested including the leaf,

root, inflorescence and siliques; moreover, the expression was found not only in the epidermis but also in subepidermal cells. In contrast, results reported by Kurdyukov et al. (2006b) showed that *HTH* expression was exclusive to the epidermis, as shown by results of both HTH_{pro}*:GFP* and *in situ* mRNA hybridization**.**

To expand on these previous works, transgenic lines harbouring YFP or GFP tagged-HTH proteins were generated by expressing HTH_{pro}*:HTH-FP* constructs in both wildtype and *hth* mutant plants. To minimize possible expression artifacts, the entire 2009 bp-long 5'upstream region of the *HTH* gene was cloned as the promoter to drive the transgene expression. Results of a preliminary RT-qPCR experiment (Appendix E) indicated that the expression levels of wildtype *HTH* in Ws and the transgene (*HTH-FP*) in a HTH_{pro}*:HTH-FP* line were quite comparable, suggesting that the observed HTH-FP localization was unlikely an artifact of excessive overexpression. Even so, the fluorescent tagging itself may be sufficient to alter function or localization of the target protein by masking function motifs, changing conformation or interfering with binding partners (DeBlasio et al., 2010; Tanz et al., 2013). Therefore, a complementation test was conducted to confirm normal function of the tagged protein. The fact that the HTH_{pro}*:HTH-FP* construct rescued all *hth-9* mutant phenotypes, including organ fusion, cuticle and seed coat permeability and seed size, provides evidence that the C-terminal fluorescence tag did not disrupt HTH function, and HTH fusion protein localization was likely reflective of the native HTH protein.

Analyses of seedlings and young plants in this study revealed that HTH-FP was localized to shoot epidermal cells, seedling vascular tissue, the hypocotyl, cotyledons,

emerging true leaves, trichomes, the apical meristem region and the stele of primary and lateral roots. However, fluorescence was absent in the root epidermis, root hairs, and subepidermal mesophyll tissue (Figure 3.5 to Figure 3.8). These localization patterns in seedling are similar to those found by Kurdyukov et al. (2006b) who also reported *HTH* promoter (1.9 kb upstream region of the gene)-driven GUS (β-glucuronidase) expression in the shoot apical meristem region and in emerging leaves. In addition, HTH-FP in emerging lateral roots (Figure 3.6) is also in keeping with observations made by Kurdyukov et al. (2006b), although Kurdyukov and colleagues did not report *HTH-FP* expression in the stele of primary roots.

In addition, to the best of my knowledge, this study provides the first evidence of HTH-FP in the trichomes and guard cells. Fluorescence detected in trichomes, as shown in Figure 3.5 and Figure 3.7, suggests that the presence of HTH-FP in trichomes initiates early (4-day-old seedling) and continues to the later stage of development (2-week-old plants). Trichomal expression has been reported for another cuticle-associated gene, *DCR*. Mutant *dcr* plants have a defective cuticle that results in postgenital organ fusion with significant reduction of a hydroxylated 16-carbon fatty acid (Panikashvili et al., 2009). Its encoded protein is localized to the epidermal cells and trichomes, and occasional tangling of expanding trichomes has been reported in the *dcr* mutant (Marks et al., 2009). Likewise, reduced trichome numbers or morphological abnormalities were also reported for mutants of *FDH* (wax biosynthesis) and *LCR* (cutin biosynthesis) that exhibit organ fusion, although no trichomal expression has been directly observed (Pruitt et al., 2000; Wellesen et al., 2001).

Expression of genes like *HTH* in the vasculature is less readily explained. However, other genes known to be involved in cuticle biosynthesis are expressed in vascular tissue. For example, *FDH* mRNA was detected in the phloem tissues by *in situ* RNA hybridization and *LACS1* (long-chain fatty acid synthesis) in the primary and lateral root, and vasculature bundle of young leaves (Weng et al., 2010).

HTH-FP in floral buds, sepals, petals, stamens and ovaries (Figure 3.9) is consistent with what observed by Kurdyukov et al. (2006b) and was not unexpected given the diagnostic floral fusion phenotype of *hth* mutants. These localization patterns also are in accordance with the microarray-based profiles (Figure 3.2) that showed little expression in rosette leaves and high levels of expression in young floral buds, petals, carpels, siliques and their ovules, a result consistent with the floral organ fusion phenotype seen in these mutants. These transcript-based expression patterns are also consistent with profiles seen in the immunoblotting results of HTH_{pro}*:HTH-FP* transgenic lines (Figure 3.3).

In this study we corroborate earlier findings showing *HTH* expression in maternally derived ovule tissues but, as discussed below, extend this to include expression in accessory cells in the mature embryo sac. HTH-FP ovule localization was polarized and specifically localized to the chalazal end of the embryo sac (Figure 3.10). Although Kurdyukov et al. (2006b) also demonstrated expression in the embryo sac, it was not restricted to the chalazal end. These authors also showed expression in the integuments which are sporophytically derived and share a common L1 origin with epidermal cells.

The subepidermal localization of HTH-FP in the ovary wall (Figure 3.9C) contradicts the results reported by Krolikowski et al. (2003) and Kurdyukov et al. (2006b). These differences might be attributed to methodology. In this study the entire genomic region encoding the HTH protein was translationally fused to fluorescent reporter genes and the constructs driven by the full-length *HTH* promoter, whereas Kurdyukov et al. (2006b) did not use a translational fusion construct.

The method employed by Krolikowski et al. (2003) who use *in situ* RNA hybridization to detect *HTH* transcripts instead of the HTH protein might also contribute to the different expression profile observed in the current study. Krolikowski et al. (2003) reported that *HTH* expression appeared to be present uniformly in epidermal and subepidermal tissues. As is evident from a growing body of literature, cells and tissues to which a protein localizes may not correspond directly to those synthesizing the mRNAs; that is, the mRNAs or proteins may be non-cell autonomous (Lee et al., 2011; Zhou et al., 2014). A classic example of non-cell-autonomous regulation can be seen in the regulation of root hair (H) and non-hair (N) cell fates in root epidermis. *In situ* hybridization and promoter fusion studies revealed that *GLABRA 3* (*GL3*) mRNAs are specifically expressed in H cells and not N cells, but the GL3 protein was found in both H cells and N. This GL3 protein distribution is achieved by transport of the protein itself through the plasmodesmata between the two adjacent cell types (Bernhardt et al., 2005). In some cases, the RNA transcripts and not the proteins can travel long distance to other organs. For instance, mRNA transcribed by the *NACP* gene *CmNACP*, a member of the NAC domain gene family of transcription factors, was found in the phloem sap of mature pumpkin (*Cucurbita maxima*) leaves. Heterograft studies furthermore showed

that this mRNA could accumulate in cucumber (*Cucumis sativus*) scion phloem and apical tissues, suggesting that specific mRNA transcripts are transported from the body of the plant to the shoot apex (Ruiz-Medrano et al., 1999).

3.4.3 **A possible role for HTH in female gametophyte development**

Female gametophyte development consists of two phases: megasporogenesis followed by megagametogenesis. During Arabidopsis megasporogenesis, the diploid megaspore mother cell gives rise to four haploid megaspores via meiosis. Three of these megaspores degenerate, and the one that survives becomes the functional megaspore. During Arabidopsis megagametogenesis, the functional megaspore develops into the mature female gametophyte in seven stages. At the first stage (FG1, female gametophyte stage 1), the functional megaspore contains a single nucleus. This nucleus undergoes mitosis without cell division, and this is the double-nucleate FG2 stage. The two nuclei move to the opposite ends of the embryo sac and a vacuole forms in the center, defining the FG3 stage. With one more round of mitosis, four nuclei are present at FG4. At FG5, one more nuclear division occurs and are followed by cellularization. During cellularization, two polar nuclei migrate toward the center. At FG6, the mature megagametophyte consists of seven cells and eight nuclei. consisting of an egg cell, two synergids, a central cell and three antipodal cells. At FG7, degeneration of the three antipodal cells occurs and the final four-celled female gametophyte (i.e. also known as the embryo sac) is ready for fertilization. After fertilization, gametic cells, the egg cell and the bi-nucleate central cell, form the embryo and the endosperm respectively (Christensen et al., 1997; Drews et al., 1998; Drews and Koltunow, 2011).

A role for the HTH protein in the female gametophyte development is supported by genetic evidence (Pagnussatt et al, 2005) and by *HTH_{pro}:HTH-FP* temporal expression patterns shown here (Figure 3.10). HTH-FP was detected early in developing ovules (at about the FG3 stage), a stage when the two nuclei migrate to the opposite ends of the embryo sac symplasm and a large vacuole forms at the center. As the ovules mature, HTH-FP appears to resolve to distinct entities at the chalazal pole of the embryo sac. Following anthesis and fertilization, fluorescence becomes markedly diffuse as would be expected if cellular integrity were lost.

Pagnussat et al. (2005) analyzed the *eda17 mutant*, a mutant *hth* allele generated by *Ds* transposon insertion, and showed developmental arrest at the FG3/two-nuclear stage. Ovule abnormalities of two *hth* mutants (*hth-8* and *hth-10*) have also been reported previously (Lolle et al., 1998). This timeline of *HTH-FP* expression corresponds to the stage at which mutant ovules arrest and further supports a role for HTH in embryo sac development. Interestingly, no *hth* mutants with floral fusion phenotypes have mutations within a 1 kb genomic region at the 5' end of the gene, a region that includes the first two introns and exons. These genetic studies suggest that mutations falling within this 1 kb region either do not result in a visible phenotype or cause lethality. Identification of *eda17*, a mutant that harbors a *Ds* transposon insertion in the first intron suggests that this region (and perhaps the intron itself), is essential for female gametophyte development. Both genetic and expression data provide evidence for sporophytic and gametophytic *HTH* functions.

Antipodal cells are metabolically active and one proposed function is nourishing the embryo by transporting metabolites to the central cell via plasmodesmata (Vijayaraghavan et al., 1988). Secretory activity linking antipodal cells and the endosperm syncytium has been reported in wheat (*Triticum aestivum* L.). In wheat plants, large vacuoles of the endospermal syncytium contained fragments of the nucleolus and chromatin were extruded from adjacent antipodal cells (Chaban et al., 2011). However, what role HTH serves in accessory cell development and/or function remains unclear.

3.4.4 **A novel role for *HTH* in seed coat development**

Mutant *hth* seeds were found to be 1.5 - 2 times larger than their wildtype counterpart, and were often misshapen and occasionally adhered to each other (Figure 3.14). In addition, the mutant seed coats showed enhanced permeability as demonstrated by tetrazolium assays (Table 3.1). These seed phenotypes were rescued by expressing HTH_{pro}*:HTH-FP* constructs in the *hth-9* background, demonstrating a previously unknown role for *HTH* in seed development.

The question of how seed coat permeability is related to the normal function of the post-embryonic epidermal cuticle remains. Some clues can be found in considering the relative fusion phenotypes of various *hth* alleles. For example, in comparing the relative severity of floral organ fusion, it is clear that *hth-15* shows a milder fusion phenotype then either *hth-13* or *hth-14*. This same pattern is reflected in seed coat permeability, with *hth-15* showing much decreased tetrazolium red staining (Figure 3.15). Similarly, *hth-15* mutants tend to produce

fewer morphologically abnormal seeds. Whether or not similar trends exists among *hth* alleles in different ecotype backgrounds is an area for future investigation.

The Arabidopsis seed coat is maternally derived and is composed of four to five cell layers that develop from the ovule integument. During early female gametophyte development, the megaspore mother cell is surrounded by the outer and inner integuments, which are both of epidermal origin (Beeckman et al., 2000). Each integument consists of two layers; however the inner integuments have an extra internal layer (ii') between ii1 and ii2 (Figure 3.12B). Vacuoles appear in the cells of the outer integument at the onset of embryogenesis, and amyloplasts start to form at the globular stage. In the outermost integument layer (oi2) mucilage production initiates at the torpedo stage. These cells gradually mature into specialized cells designed for seed rupture and mucilage release. When the embryo reaches the walking stick stage (expanding embryo), the enlargement of the mucilage compartments pushes cytoplasm and amyloplasts towards to the middle and form a small column, i.e. columella, that line up in the center of the oi2 cell (Figure 3.13G). At the desiccation stage, oi1, ii1' and ii2 collapse to form the brown pigment layer that gives the brown colour of mature seeds (Beeckman et al., 2000; Creff et al., 2015; Windsor et al., 2000).

At the torpedo stage, in addition to the initiation of mucilage development, deposition of a thickened periclinal cell wall commences between the outer and inner integuments, also known as "wall 3". Thickening of the primary wall rather than the formation of a secondary wall is responsible for increasing the width of wall 3 (Beeckman et al., 2000). Wall 3 is

deposited between the two L1-dervived integuments and represents a zone of fusion between oi1 and ii2. Similar to the leaf epidermal cuticle, an electron dense layer of cutin-like polyester material has been identified within wall 3 as well as a cuticle on the oi2 and ii1/endothelium layer (Beeckman et al., 2000; Creff et al., 2015; De Giorgi et al., 2015; Molina et al., 2008; Watanabe et al., 2004). Creff et al. (2015) showed that the innermost cell layer of the outer integument (oi1) is mainly responsible for wall 3 deposition. Interestingly, wall 3 thickening was shown to be regulated by mechanical stress. Wall 3 has been proposed to limit seed size by serving as a corset around the developing seed. It was suggested that mechanical pressure on the seed coat due to embryo and endosperm expansion is sensed by oi1, and in response oi1 thickens wall 3.

Examination of HTH_{pro}:*HTH-FP* developing seeds using both epifluorescence and confocal microscopy show HTH-FP localization in the ovule integument which later becomes the seed coat (Figure 3.11 to Figure 3.13). HTH-FP was initially detected in the mechanosensitive oi1 integument layer (Figure 3.12) and later in both oi1 and oi2 layers in the post mature stage (Figure 3.13) (Creff et al., 2015; Western et al., 2000; Windsor et al., 2000). This HTH-FP localization to the seed coat is in accordance with the transcriptome-based analyses reported previously (Appendix F). Taken together with the observed increase in seed size and changes in seed coat permeability, these data suggest a possible role for *HTH* in the production of the cutin-like polyester materials of the oi2 layer and wall 3 which is positioned at the boundary of two integumentary layers. Increased seed coat permeability may be a consequence of changes to the composition and integrity of the cuticle-like

structures of the *hth* mutant seed coat, since changes in the cuticle of vegetative tissues can lead to similar consequences for differential permeability (Lolle et al., 1998).

Like the shoot epidermis, the integumentary cells of epidermal origin likely take part in the formation of a selectively permeable barrier. One way to achieve this is to synthesize an extracellular matrix, such as the wall 3 or the endothelium cuticle layer, that shares properties with the cuticle typically found on shoot epidermal cells. Therefore, common genes might be shared for cuticle (or cuticle-like layers) synthesis of the shoot and seed coat. Genes essential for shoot cuticle formation have also been reported to take part in seed coat development. The seed coat of *dcr* mutants, for example, has been shown to be more permeable resulting in greater toluidine blue staining than seen in Columbia wildtype seeds. Additionally, mutant *dcr* seeds were often deformed and occasionally fused and failed to release mucilage upon imbibition (Panikashvili et al., 2009). In *bdg1* mutants, toluidine blue could permeate seeds and stain the endosperm, and possibly as a consequence of this enhanced permeability, showed reduced seed viability and dormancy (De Giorgi et al., 2015). *GPAT5*, a member of the glycerol-3-phosphate acyltransferase group, like HTH_{pro}:*HTH-FP*, has been shown to be expressed in the oi1 layer. Seeds harbouring mutations in the *GPAT5* gene showed altered permeability to tetrazolium dyes (Beisson et al., 2007; Molina et al., 2008). Although it is unknown whether a direct causal link exists between the observed changes in seed coat permeability and changes in wall 3 it remains a possibility that the epidermal cuticle and seed coat share overlapping biosynthetic pathways.

The L*er* ecotype harbours a mutant allele of *ERECTA* and conditions some developmental aspects of the seed coat, including seed coat permeability. The *ERECTA* (*ER*) gene encodes a receptor kinase and its mutation has pleiotropic effects. As with other *ER*-like genes, ER regulates plant architecture, such as internode and pedicel elongation, axial polarity and stomata patterning, likely through modulating cell division and expansion (van Zanten et al., 2009). Although the *er* mutation itself did not result in ovule abnormities, it has been shown to enhance the severity of mutations that target female gametophyte development as demonstrated in *short integuments1* (*sin1*) (Lang et al., 1994) The effect of *hth-4* and *hth-8* on seed coat permeability in this genetic background may therefore be less obvious, although increases in seed size and morphological abnormalities were still evident.

3.4.5 HTH is associated with stress responses

ER bodies were originally described as 'mystery organelles' by Gunning (1998) because their size and shape are distinct from other ER-derived subcellular structures, such as coat protein vesicles that are responsible for protein export from the ER, and precursor-accumulating vesicles that mediate the direct protein transport from the ER into vacuoles (Hara-Nishimura et al., 1998). A few years later, ER-targeted green fluorescent protein was used to confirm that these fusiform structures had an ER origin (Hawes et al., 2001), while ultrastructure studies provided evidence for these bodies being surrounded by membranes with ribosomes (Hawes et al., 2001). Hayashi et al. (2001) proposed the term 'ER bodies' to describe these distinctive ER-derived structures. Similar structures have since been reported in 46 other species of *Brassicaceae*, seven species of *Capparaceae* and four species from other families (Hara-Nishimura and Matsushima, 2003).

Several structural features predict that the HTH protein resides in the ER lumen. Based on bioinformatics analysis, HTH is predicted to be a globular protein without transmembrane helix motifs. It has a putative N-terminal signal peptide (1-19 a.a), although it lacks a C-terminal ER-retention motif. ER-resident proteins often have C-terminal ER retention signals such as KDEL, HDEL, or REEL. These motifs allow for selective retrieval to the ER from the ER-Golgi intermediate compartment or the Golgi complex via a recycling pathway. For example, in Arabidopsis a major protein component of ER bodies is a β-glucosidase called PYK10 that has a KDEL retention signal at the C-terminus (Matsushima et al., 2003b). Although evidence presented here putatively localizes the HTH-FP protein to the ER and ER bodies, HTH does not contain a canonical C-terminal ER retention signal (i.e. KDEL, HDEL, or REEL) but instead has three KDEL-like sequences ([KRHQSA]-[DENQ]-E-L) at amino acid positions 270 (KDEK), 310 (KKEL), and 387 (KNEL). These signal motifs are predicted to position at or near exposed protein surfaces potentially allowing interactions with other proteins such as ER-sorting receptors. HTH ER retention could also be achieved by motifs other than those KDEL and KDEL-like signals; for example, the N-terminal tandem repeat PPPVHL and C-terminal cysteine-rich motif of maize γ-zein are essential for retention in the ER and ER-derived protein bodies (Saumonneau et al., 2011). Moreover, the retention motifs can be present in either the C- or N-termini as is the case for a rice prolamin, whose ER-retention motif resides in the middle of the protein sequence (Masumura et al., 2015).

Protein retention in ER bodies can also be achieved by other means such as protein aggregation. For instance, some Arabidopsis proteins known to reside in ER bodies such as

the vacuolar processing enzyme RESPONSIVE-TO-DESICCATION 21 (RD21) and the vacuolar invertase FRUCTOSIDASE 4 have no ER retention signals (Hayashi et al., 2001; Rojo et al., 2003), but are thought to be retained forming aggregates with β-glucosidase PYK10 (Hara-Nishimura et al., 2004). Instances of β-glucosidase aggregates have also been demonstrated in oats (Kim et al., 2000) and flax seedlings (Fieldes and Gerhardt, 1994).

One well known response to wide-ranging adverse environmental factors such as drought, pest or pathogen attacks is the formation of ER bodies (Hara-Nishimura and Matsushima, 2003; Matsushima et al., 2002; Matsushima et al., 2003a; Matsushima et al., 2003b). ER bodies are commonly present in seedling cotyledon epidermal cells, hypocotyls and roots of young Arabidopsis seedlings. These bodies, however, are rare in rosette leaves but can be induced by wounding, salt stress or stress signaling compounds (Hara-Nishimura and Matsushima, 2003; Hayashi et al., 2001; Matsushima et al., 2002; Ogasawara et al., 2009). For example, treatment with the wound hormone, methyl jasmonate (MeJA), has been shown to induce ER body formation in rosette leaves, and this induction did not occur in MeJA-insensitive *coronatine-insensitive 1* (*coi1*) mutant plants (Matsushima et al., 2002). In addition, Ogasawara et al. (2009) demonstrated that ER body formation is a systemic defense response, showing that when one of two seedling cotyledons was wounded, both the damaged and the intact cotyledon had increased the number of ER bodies.

Epifluorescent and confocal microscopy imaging revealed that the HTH-FP were localized to 5-10 µm-long, spindle-shaped motile bodies that move at a maximum speed of 0.5-1 µm/sec in the hypocotyl cell (Figure 3.16A-M), comparable to the in vivo ER tubule

growth rates (Sparkes et al., 2009). The properties described above are diagnostic characteristics of ER bodies. Furthermore, TMRM staining excluded HTH-FP localization to the mitochondria (Figure 3.16N-P), and crosses with a transgenic line (Kaleidocell) containing fluorescent protein labelled organelles ruled out nucleus and chloroplast localization (Appendix D). Last, colocalization studies using an ER-targeting RFP (erRFP) verified that HTH-FP was localized to the ER network and its derived ER bodies (Figure 3.17). These findings support HTH-FP localization in ER bodies in epidermal cells.

The ER body localization seen in our HTH-FP transgenic lines associate HTH with cellular structures known to be induced following exposure to stress. To directly test whether HTH expression was regulated by stress, HTH expression levels were quantified in Ws wildtype plants treated with MeJA (Figure 3.18) and shown to increase following these treatments. In addition, preliminary data suggest that MeJA treatment can induce ER bodies in the midrib epidermal cells of HTH_{pro}*:HTH-FP* rosette leaves. On the other hand, no ER bodies were observed in water-treated (control) plants. Although ER markers need to be used in future research to further verify the ER body formation, this result is in keeping with previous work showing that ER body formation can be triggered by mechanical wounding or by exposure to the wound hormone MeJA (Hara-Nishimura and Matsushima, 2003; Hayashi et al., 2001; Matsushima et al., 2002; Ogasawara et al., 2009). Although HTH-FP containing ER bodies were not seen under control conditions, *HTH-FP* expression was detected in rosette leaves despite not been detected using immunoblotting approaches (Figure 3.3). This may reflect detection limits of immunoblotting compared with fluorescent protein detection (Swenson et al., 2007).

3.5 Conclusions

Epidermal expression of HTH-FP in vegetative and reproductive tissues suggests that HTH protein is involved with cutin biosynthetic pathways that ultimately determine cuticle structure. HTH-FP expression in the integument and elevated seed coat permeability in *hth* mutants suggest that HTH might serve a previously unknown function in seed coat development. Although embryo sac expression was observed, it is less clear what function HTH serves in this context. At the subcellular level, HTH-FP protein was found to reside in ER-derived structures and to colocalize with eRFP to ER bodies. By extending this and showing elevated *HTH* expression and possible ER body formation in response to MeJA treatment, results in this chapter also suggests a role of *HTH* in stress responses. What specific role HTH plays in stress responses, however, awaits further investigation.

3.6 Future Research

The HTH-FP localization to the female gametophyte and seed coat suggests possible alternative functions for HTH in addition to the one involved in cuticle synthesis. To verify the importance in gametophyte development, the *eda17* mutant plants that exhibit embryo sac development arrest serve as a suitable material. Functional complementation of *eda17* mutant plants by introducing HTH_{pro}*:HTH-FP* would corroborate a role for HTH in female gametophyte development. To determine whether elevated cuticle permeability observed in mutants is a direct consequence of altered plant cuticle ultrastructure, ultrastructural examination by transmission electron microscopy (TEM) could be conducted. Furthermore,

with the increasing accuracy of genome editing techniques such as the type II clustered regularly interspaced short palindromic repeat (CRISPR)/Cas9 (CRISPR-associated) system, motifs essential for protein function can be verified by site-directed insertion/deletion on the native *HTH* gene (Barrangou et al., 2007; Jinek et al., 2012; Schiml and Puchta, 2016).

Results showed that MeJA induced formation of structures similar to ER bodies in HTH_{pro}-*HTH-FP* transgenic plants. To confirm preliminary results described here the ER body localization, two more control samples could be included. One is the *erRFP* transgenic plant as a positive control for its labeled ER bodies; the other is wildtype plants because MeJA may induce production of secondary compounds that emit fluorescence at the wavelength range of the reporter fluorescent protein. In addition, identifying other biotic or abiotic stress factors that also increase *HTH* expression or induce ER body formation in rosette leaves might point to novel connections between *HTH* expression and stress response pathways.

HTH_{pro}:*HTH-FP* transgenic plants developed in this research also serve as a material for studying protein-protein interactions. Co-immunoprecipitation can now be performed using commercially available antibodies that specifically bind to the fluorescent protein tags described here. Partner proteins that bind to HTH as part of a protein complex can be identified by techniques such as matrix-assisted laser desorption/ionization time-of-flight mass spectrometer (MALDI-TOF). Results might help elucidate not only HTH's biochemical function but may also provide further insight into localization. In addition to proteins, HTH's possible interaction with DNA could be investigated by chromatin immunoprecipitation.

Chapter 4 Bioinformatics analyses and enzymatic assays of the Arabidopsis HOTHEAD protein

4.1 Introduction

The epidermal cuticle is an extracellular hydrophobic layer that serves as an interface between the plant and the environment, preventing water loss and protecting the plant from harmful pathogens and chemicals. The composition of this external layer is complex, and the specific functional role of its individual constituents is relatively poorly understood. A major component of the cuticle is cutin, a polyester composed of fatty acids, phenolic compounds, and glycerol (Table 1.1) (Nawrath, 2006; Pollard et al., 2008). The composition of these cuticular monomers is a major determinant of the architecture of cutin polymers and cuticle properties (Pollard et al., 2008). Perturbation of biochemical pathways involved in fatty acid monomer synthesis can result in a number of phenotypes including changes in cuticle permeability and post-genital organ fusion (Chen et al., 2003; Lolle et al., 1998; McNevin et al., 1993; Pruitt et al., 2000; Schnurr et al., 2004; Wellesen et al., 2001).

The Arabidopsis *HOTHEAD* (*HTH*) gene is among the numerous identified genes that regulate cuticle formation. Plants harbouring various mutant *hth* alleles display floral organ fusion although fusion of vegetative tissue is rarely observed (Lolle et al., 1998). As is the case for other mutants in this class, there is evidence that the cuticular properties of *hth* mutants are altered as demonstrated by increased cuticular permeability (Kurdyukov et al., 2006a; Lolle and Cheung, 1993; Lolle et al., 1998).

Typical cutin monomers are C16 and C18 ω-hydroxy fatty acids that can be esterified at the primary hydroxyl groups to produce a linear polyester chain. Branched structures can be produced by esterification between the carboxyl group of one fatty acid and a glycerol

hydroxyl group or a secondary hydroxyl group of another fatty acid. The majority of cutin monomers in *Arabidopsis thaliana* and *Brassica napus*, however, are α,ω-dicarboxylic acids (DCAs), a typical type of suberin monomer. DCAs can be esterified to glycerol and give rise to a cross-linked cutin architecture characteristic in the *Brassicaceae* species (Pollard et al., 2008). The reduction in the relative percentage of C16 and C18 DCAs in *hth-12* mutants (Kurdyukov et al., 2006b) points to a possible cause for the changes in cuticle properties observed in other *hth* mutants. Based on these compositional changes the HTH protein has been proposed to function as a ω-fatty alcohol dehydrogenase required for DCA biosynthesis, and mutations that cause loss of function would therefore disrupt normal cutin monomer genesis and consequently lead to changes in cuticle properties.

On the other hand, the HTH protein has also been proposed to be a mandelonitrile lyase (MDL) (Krolikowski et al., 2003). Both HTH and mandelonitrile lyase belong to the glucose-methanol-choline-oxidoreductase (GMC) family that includes proteins with a variety of catalytic activities including dehydrogenases, oxidases and lyases (Cavener, 1992). Hydroxynitrile lyases catalyze cyanogenesis in which dissociation of α-hydroxynitriles results in the production of hydrogen cyanide (HCN) and an aldehyde or ketone (Figure 1.12) (Poulton, 1990). Production of HCN is commonly considered to serve as a means to defend against herbivores and pathogen attack and is not exclusive to higher plants but rather common to a spectrum of organisms including ferns, bacteria, fungi and insects (Conn, 1981).

One of the most commonly used approach to function prediction is established on the idea of inheritance through homology. This method is based on the premise that proteins with similar sequences are more closely related and likely perform similar functions. The activity of many enzymatic proteins is largely dependent of its overall tertiary structure and the catalytic triad, a group of three amino acids involved in catalysis. Hence, structure and catalytic site predictions often provide insight into the function of a protein (Edwards and Cottage, 2003; Lee et al., 2007). In addition, in functional genomics studies, mRNA expression data are often used to discover regulatory networks with the assumption that genes with similar mRNA expression profiles are likely to be regulated via the same mechanisms or involved in similar functions (Heyndrickx and Vandepoele, 2012).

In this chapter, enzymatic assays were conducted to experimentally test for these two possible protein functions using a bacterial recombinant HTH protein. In addition, bioinformatics tools were employed to mine for evidence supporting either a ω-fatty alcohol dehydrogenase or hydroxynitrile lyase function for the HTH protein. Although neither enzymatic activity was detected in the *in vitro* assay, bioinformatics analyses suggest that HTH protein is more likely to function as a ω-fatty alcohol dehydrogenase involved in the biosynthesis of components integral to the epidermal cuticle.

4.2 Materials and Methods

4.2.1 cDNA isolation and cloning

Total RNA was extracted from inflorescence tissue of 6-week-old *Arabidopsis thaliana* plants in the Landsberg *erecta* (L*er*) background using the RNeasy Plant Mini Kit (Qiagen, Venlo, Limburg, Netherlands). Reverse transcription to generate first strand cDNAs was performed using the SuperScript® III First-Strand Synthesis System Kit (Invitrogen, Carlsbad, California, USA) and 50 µM oligo (dT) primers.

Synthesized cDNAs were used as the template to amplify the cDNA lacking the 5' terminal fragment that encodes the predicted signal peptide. The polymerase chain reaction (PCR) was conducted using Phusion Hot Start II High-Fidelity DNA Polymerase (Cat. F549S, Thermo Fisher Scientific, Waltham, Massachusetts, USA) that produces amplicons with blunt ends. To amplify the cDNA of *HTH* without the region encoding the signal peptide, the forward primer MBPHTHNS (5'-TCCACTGCCTCTAAAGGTAAAGAGAAG-3') and reverse primer MBPEcoRI_R1 (5'-TATTGAATTCTTATTAAACACCAGCTTTGTTTCC-3') were used. An EcoRI restriction site was engineered at the 3' end of the reverse primer.

Approximately 3 ng of first strand cDNA was added to a 20 µl reaction mix (12.8 µl dH_2O, 4.0 µl 5x Phusion HF Buffer, 0.4 µl dNTPs (10 mM), 0.5 µl forward primer (20 µM), 0.5 µl reverse primer (20 µM), 0.2 µl Phusion DNA Polymerase (2 U/µl), 0.6 µl DMSO). Amplification was carried out using the following conditions: (a) 98°C for 30 seconds, (b) 98°C for 10 seconds, (c) 70.1°C for 10 seconds, (d) 72°C for 30 seconds, repeat steps (b) -

(d) 25 times, (e) 72°C for 5 minutes. Blunt-ended PCR products were size separated using agarose gel electrophoresis, and amplicons of target size gel-purified using QIAquick Gel Extraction Kit (Cat. 28704, Qiagen).

The *HTH* cDNA was cloned into the pMAL-c4x vector for cytoplasmic expression using the pMAL Protein Fusion and Purification System (New England Biolabs, Ipswich, Massachusetts, USA). The cloned cDNA was inserted downstream from the *malE* gene, resulting in the expression of a fusion protein with an N-terminal maltose-binding protein (MBP) tag that has a high affinity to amylose (Appendix G). The pMAL-c4x vector was digested with XmnI and EcoRI at 37°C for 1 hour and enzymes heat inactivated at 65°C for 20 minutes to generate linearized vectors and the cDNA digested with the same restriction enzymes. Ligation reactions were conducted by mixing 2 µl of linearized vector (40 ng), 2 µl of digested cDNA (80 ng), 6 µl H2O, 10 µl 2X Quick Ligation Reaction Buffer and 1 µl Quick T4 DNA ligase (Cat. M2200, New England Biolabs) followed by a 10-minute incubation at room temperature. The ligation mix was transformed into DH5-α competent *Escherichia coli* cells using standard $CaCl_2$-heat-shock techniques and transformed colonies selected for ampicillin resistance and the Lac^- phenotype (i.e. white colonies).

4.2.2 Recombinant MBP-HTH protein purification

Bacterial cultures were grown in rich broth media with ampicillin (10 g/L trypton, 5 g/L yeast extract, 5 g/L NaCl, 2 g/L D-glucose, 100 µg/mL ampicillin) at 37°C at 200 rpm overnight. A 15 mL aliquot of the overnight culture was used to inoculate 1.5 L of rich broth supplemented with ampicillin, and cultures incubated at 37°C on a shaker at 200 rpm until

OD600 reached 0.4 - 0.6. Protein expression was then induced by the addition of isopropylthiogalactoside (IPTG) to a final concentration of 0.1 mM, followed by overnight incubation at room temperature at 200 rpm. Afterwards, cells were harvested by centrifugation at 4000 ×g for 25 minutes at 4°C. The supernatant was discarded and the cells re-suspended in 100 mL column buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, 1 mM DTT, 1 mM sodium azid, 2 mM phenylmethanesulfonylfluoride (PMSF)) with 0.1% (v/v) Triton X-100 and frozen at -20°C for later use.

Frozen cells were thawed in an ice-water bath for approximately one hour and then placed on ice. The cell suspension was sonicated in short pulses of approximately 15 seconds for a total of 2 minutes. Cell lysate was centrifuged at 4000 ×*g* at 4°C for 30 minutes to separate the soluble and insoluble fractions. The soluble fraction was diluted at a 1:1 ratio with column buffer and for every 50 mL, 2 mL of washed amylose agarose beads were added. Amylose agarose beads (New England Biolabs) were rinsed thoroughly with column buffer at a ratio of 1:10 before use. The protein-bead mixture was incubated at 4°C overnight on a rocker to prevent beads from settling.

The MBP-HTH recombinant protein was purified using an affinity column. A 60 mL syringe was packed with compact glass wool to the 10-mL mark and rinsed with 40 mL of column buffer before pouring in the protein-bead mixture. The flow-through was collected at a flow rate of 1 mL/minute. For every 50 mL of the protein-bead mixture applied, an aliquot of 100 mL of column buffer was used to wash the column. The wash was collected to allow for detection of any protein loss.

The recombinant protein was eluted with 30 mL of 10 mM maltose dissolved in column buffer at a flow rate of 1mL/minute. The eluted fraction was further concentrated using Amicon Ultra-15 Centrifugal Filter Unit with Ultracel-50 membrane (UFC905024; EMD Millipore, Billerica, Massachusetts, USA) by centrifugation at 4000 ×*g* for 30 minutes to eliminate proteins smaller than 50 kDa. The flow-through was discarded and the concentrated fraction retained. The total protein concentration was determined using Quick Start™ Bradford 1x Dye Reagent (Bio-Rad Laboratories, Hercules, California, USA) according to the manufacturer's instructions. For MBP cleavage from the fusion protein, 1 µg of factor Xa (P8010S, 1 mg/mL; New England Biolabs) was added in every 100 µg of fusion protein. The reaction mixture was incubated for 16 hours at 4°C.

4.2.3 **Enzymatic assays**

Hydroxynitrile lyase activity was determined by monitoring cleavage of mandelonitrile into benzaldehyde and HCN at 25°C as described by Jorns (1979) with modifications. In the final assay solution, the concentrations are 95 mM sodium acetate, 2.8 mM mandelonitrile, 0.003% (w/v) bovine serum albumin, and 1.6% (v/v) ethanol. Affinity purified MBP-HTH was treated with the protease factor Xa to remove the MBP tag, and post-cleavage mixture (MBP+HTH) used for the enzymatic assay. The absorbance of the assay solution was first monitored at 275 nm in a quartz cuvette with 1 cm light path for 5 minutes at 25°C until the readings stabilized. The reaction was then initiated by addition of either mandelonitrile lyase from almonds (Cat. M6782, Sigma-Aldrich) or MBP+HTH samples. Immediately after protein addition, the OD_{275} of the reaction was continuously monitored by Cary 100 UV-Vis spectrometer (Agilent Technologies, California, USA) over the duration of 45 minutes. The

slope of ΔOD_{275nm} was determined at the linear range of the curve and corrected using the blank. By using the millimolar extinction coefficient of benzaldehyde, ε_{275} = 1.2 $mM^{-1}cm^{-1}$, the enzyme activity (nmol/mg protein/min) was calculated using the slope. One unit of enzyme activity is defined as the formation of 1.0 μmole of benzaldehyde and HCN from mandelonitrile per minute at pH 5.4.

Alcohol dehydrogenase activity was assayed by spectrophotometry using Cary 100 UV-Vis spectrometer. Alcohol dehydrogenase activity was determined by monitoring the absorbance at 340 nm resulting from reduction of NAD as described by Kagi and Vallee (1960). The final concentrations of the assay mixture are 22 mM sodium pyrophosphate, 7.5 mM β-nicotinamide adenine dinucleotide, 0.3 mM sodium phosphate, 0.003% (w/v) bovine serum albumin and alcohol substrates, including ethanol (3.2%, v/v), 1-hexanol (3.2%, v/v) and benzyl alcohol ((v/v) 0.32%, with 0.1% Triton-X). The absorbance of the assay solution was first monitored at 340 nm at 25°C to achieve temperature equilibrium before initiating the 1-mL reaction by mixing in various amounts of yeast (*Saccharomyces cerevisiae*) alcohol dehydrogenase (A7011, Sigma-Aldrich) or post-cleavage purified protein (MBP+HTH). The enzyme activity was calculated using the slope of ΔOD_{340nm} in the linear range with the millimolar extinction coefficient of β-NADH, ε_{340} = 6.22 $mM^{-1}cm^{-1}$.

4.2.4 **Deglycosylation and protein immuno-detection**

HTH_{pro}:HTH-EYFP seedlings (see Chapter 3) were flash frozen in liquid nitrogen immediately after harvest. The tissue was pulverized by vortexing frozen tissue in sealed 2 mL tubes containing ten stainless steel beads (1/8” diameter; Abbott Ball Company, West

Hartford, Connecticut, USA). Before thawing, pulverized tissue was mixed with the extraction buffer (100 mM Tris-HCl pH 8.0, 8 M urea, 5 mM EDTA, 2.5% (w/v) SDS, 10% (v/v) glycerol, 1 mM PMSF, 100 mM DTT) and protease inhibitor cocktail (1:50 (v/v) ratio; P9599; Sigma-Aldrich, St. Louis, Missouri, USA). The mix was vortexed for two minute, followed by centrifugation at 6000 ×g at 4°C to pellet cell debris. The supernatant was collected as the crude extract. To determine the total protein concentration, protein was acetone precipitated to remove interferences in the crude extract. One volume of crude protein solution was mixed with four volumes of cold acetone before incubation at -20ºC overnight. Samples were centrifuged for 5 minutes at 4°C at maximum speed (13000 ×g). The supernatant was carefully discharged and the pellet was dried by inverting the tube on tissue paper. The pellet was then resuspended with 100 mM Tris-HCl pH8.0. Protein concentration was determined using Bio-Rad Quick Start™ Bradford 1x Dye Reagent (Bio-Rad Laboratories, Hercules, California, USA), according to the manufacturer's instructions.

For protein deglycosylation, Protein Deglycosylation Mix (V4931; Promega, Madison, Wisconsin, USA) was used to remove glycans from both O-linked and N-linked glycosylation sites, and PNGase F (G1549, Sigma-Aldrich) to specifically deglycosylate N-linked glycoproteins. To first denature proteins, an aliquot of 18 µl crude extract was mixed with 10× denaturing solution (0.5% (w/v) SDS, 40 mM DTT, 1× concentration) to give a denaturing reaction volume of 20 µl, followed by incubation for 10 minutes at 95°C and 5 minutes on ice. For both O-linked and N-linked deglycosylation, 5 µl of 10× Deglycosylation Reaction Buffer, 5 µl of 10% (v/v) NP-40, 15 µl of water, and 5 µl of Protein Deglycosylation Mix were mixed into the tube to give a final reaction volume of 50 µl. For

specific N-linked deglycosylation, 5 µl (1.5 unit) of PNGase F was added instead. The reactions were incubated overnight at 37°C. The negative control mock reactions were conducted the same way without adding any glycosidases.

4.2.5 **SDS-polyacrylamide gel electrophoresis and protein immuno-detection**

Protein samples were mixed in 5x Laemmli sample buffer (60 mM Tris-Cl pH 6.8, 2% (w/v) SDS, 10% (v/v) glycerol, 5% (v/v) β-mercaptoethanol, 0.01% (w/v) bromophenol blue), boiled for 10 minutes, and size-separated using a 10% (w/v) SDS-polyacrylamide gel (for each gel: 5 mL Tris pH 8.8, 9.8 mL H_2O, 200 µL 10% (w/v) SDS, 5.2 mL 30% (v/v) acrylamide, 100 µL 10% (w/v) ammonium persulfate (APS), 30 µL tetramethylethylenediamine (TEMED)) with a 4% stacking gel (for each gel, 2.5 mL Tris pH 6.8, 6.2 mL H_2O, 100 µL 10% (w/v) SDS, 1.3 mL 30% acrylamide, 50 µL 10% (w/v) APS, 20 µL TEMED). Spectra Multicolour Broad Range Protein Ladder (Cat. 26623; Thermo Fisher Scientific) was loaded as size markers. Following electrophoresis using a Tris-glycine buffer (250 mM Tris, 1.92 M glycine, 1% (w/v) SDS), proteins were transferred onto membranes. Nitrocellulose or PVDF membranes were equilibrated in transfer buffer (25 mM Tris, 192 mM glycine, 10% (v/v) methanol) for 30 minutes while the gel was equilibrated for 10 minutes prior to transfer. When PVDF used, membranes were pre-wetted in 100% methanol prior to equilibration. Proteins in the acrylamide gel were then transferred onto membranes using Trans-Blot® SD Semi-Dry Electrophoretic Transfer Cell (Bio-Rad Laboratories) at 20 volts for 25 minutes Membranes were stained with 0.1% (w/v) Ponceau-S in 5.0% (v/v) acetic acid; P3504, Sigma-Aldrich) to verify protein transfer, washed and then blocked with 5% (w/v) skimmed milk for 1 hour, or 1 pg/mL polyvinyl alcohol (P8136;

molecular weight: 30K-70K; Sigma-Aldrich, St. Louis, Missouri, USA) for 10 seconds in Tris-buffered saline with Tween-20 (TBS-T; 20 mM Tris pH 7.5, 300 mM NaCl, 0.1% (v/v) Tween-20).

Membranes were incubated overnight at 4°C with anti-MBP monoclonal antibody (E8032S; New England Biolabs) using a 1:10,000 dilution, anti-GFP antibody (1:2500; ab6556; Abcam, Cambridge, UK) or anti-HTH antibody (1:500 dilution) in 5% (w/v) skimmed milk in TBS-T. The anti-HTH antibody was generated using the oligo peptide VIFKDEKGNQHQAL, an epitope sequence of the HTH protein, as the antigen. Membranes were washed with TBS-T 5 times for 5 minutes each and incubated in a 1:10,000 dilution of horseradish peroxidase-conjugated anti-rabbit IgG antibody (Cat. A0545, Sigma-Aldrich,) or alkaline phosphatase-conjugated anti-rabbit IgG antibody (Cat. A3687, Sigma-Aldrich) for 1 hour. Membranes were washed with TBS-T, and then treated with either chemiluminescent reagents or alkaline phosphatase colourimetric solution for detection. For chemiluminescent detection, Clarity™ Western ECL Substrate and ChemiDoc™ (Bio-Read Laboratories) were used to detect signals for the deglycosylation assays. Otherwise, ECL Prime Western Blotting Detection Reagents (Cat. RPN2232, GE Healthcare, Little Chalfont, Buckinghamshire, United Kingdom) were used and the chemiluminescent signals detected by exposure of CL-Xposure X-ray films (Cat. PI34093, Thermo Fisher Scientific). X-ray films were developed using a CP1000 film processor (Agfa-Gevaert N.V., Mortsel, Belgium). For colourimetric detection, membranes were incubated in alkaline phosphatase buffer pH 9.5 (100 mM Tris, 100 mM NaCl, 50 mM $MgCl_2$) with 0.03% (w/v) NBT (nitro blue tetrazolium) and 0.02% mM BCIP (w/v) 5-bromo-4-chloro-3-indolyl-phosphate).

4.2.6 **Bioinformatics analyses**

Features of the theoretical protein sequence of HTH (Appendix H) were predicted using bioinformatics tools available at the Expasy server (www.expasy.org) and based on automated UniProtKB/Swiss-Prot annotations. Protein domains were identified using the Pfam database (http://pfam.xfam.org). Posttranslational modification sites such as N-glycosylation and phosphorylation were determined using PROSITE (http://prosite.expasy.org). A coexpression gene network was constructed using CytoScape version 3.3.0 (Shannon et al., 2003) through the GeneMania plugin (Montojo et al., 2010) with the *Arabidopsis thaliana* dataset. The phylogenetic distribution pattern of genes coexpressed with HTH was performed by String database and analysis tools (www.string-db.org). Alignments using (predicted) full-length protein sequences were generated using ClustalO as implemented in the SeaView package (version: 4.5.0). Based on this alignment, the putative active sites of HTH and other GMC oxidoreductase family proteins were assigned by identifying the aligned sequences' residues corresponding to the active sites of mandelonitrile lyase PdMDL2 (UniProt: Q945K2). According to the same alignment, phylogenetic trees were computed using PHYML within the SeaView package. The following parameters were used: LG model, bootstrap with 100 replicates, model-given amino acid equilibrium frequency, nearest-neighbour interchange (NNI) for tree searching operation, and neighbour-joining algorithm BioNJ for starting tree topology.

The X-ray resolved three-dimensional protein structures of mandelonitrile lyase PdMDL2 (PDB: 1JU2) and pyranose dehydrogenase AmPDH (PDB: 4H7U) were used as the templates to derive protein models of HTH by the Investigator function of PHYRE

(Protein Homology/analogY Recognition Engine; www.sbg.bio.ic.ac.uk /phyre2). PYMOL (www.pymol.org) was used to generate the images. The quality of modeling was estimated by ProQ2 (Ray et al., 2012) and Ramachandran plot analyses. A HTH sequence-to-secondary structure alignment was obtained by threading the HTH sequence onto the known template secondary structure of PdMDL2 and AmPDH using the PHYRE server.

4.3 Results

4.3.1 The HTH protein model

The predicted full-length HTH protein is 594 amino acids (aa)-long, contains a 19-aa signal peptide, and both N- and C-terminal glucose-methanol-choline (GMC) oxidoreductase protein family domains (based on the protein coding gene model AT1G72970.1; www.arabidopsis.org) (Figure 4.1). Its splice variant has a 27-aa deletion in the 5th exon, upstream of the putative catalytic site residues, resulting in a 567 aa-long isoform (based on the protein coding gene model AT1G72970.2; see Appendix H for the HTH protein sequences). The protein is predicted to contain two domains; the N-terminal domain (GMC_oxred_N) corresponds to the cofactor-binding domain, and the C-terminal domain (GMC_oxred_C) is thought to function in substrate binding and catalysis.

Eight amino acid residues (blue bar) that are important for HTH protein function were previously identified by mutant screens (Krolikowski et al., 2003; Lolle et al., 1998) and are hereafter referred to as functional residues. These functional amino acid residues, together with predicted posttranslational modifications and putative catalytic sites, are indicated on

the HTH protein sequence in Figure 4.1. The majority of functional residues are located downstream of the third exon and within the GMC domains, with the exception of Gly-356. Of the five N-linked glycosylation and four phosphorylation sites predicted, none correspond to the positions of the eight functional residues identified previously. Three putative catalytic active site residues, Ile-527, His-529, and Asn-567, identified by multiple sequence alignment (See Section 4.3.5) situate near the C-terminal end of the GMC_oxred_C domain. Interestingly, the putative catalytic site residue Asn-567 clusters with three functional residues (Pro-564, Gly-565 and Thr-566) identified genetically.

4.3.2 **Sequence and phylogenetic analyses**

To gain insight into HTH protein function, a phylogenetic tree was constructed from GMC oxidoreductases that share similarity to the 594-aa long theoretical HTH sequence (including the predicted signal peptide) in terms of their sequence identity and/or theoretical protein folding structure (Table 4.1; http://blast.ncbi.nlm.nih.gov; http://www.sbg.bio.ic.ac.uk). Five unknown Arabidopsis proteins, AtGMC1- AtGMC5, have the highest sequence similarity to HTH (approximately 90% coverage and 51% identity), followed by a group of putative mandelonitrile lyases (MDLs; approximately 88% coverage and 39% identity) from *Prunus serotina* (black cherry) and *Prunus dulcis* (almond). Herein, coverage indicates the percentage of the sequence that is included for the comparison. Among these MDLs, only PdMDL2 has been characterized (Dreveny et al., 2001; Dreveny et al., 2002; Dreveny et al., 2009). The remaining HTH-related GMC family proteins include alcohol dehydrogenases (PpADH, PoADH, ToADH, NrADH and AmADH), an aryl alcohol oxidase (PeAAO), a pyranose dehydrogenase (AmPDH), a cholesterol oxidase (SsCHOX), glucose oxidases

(PaGOX and AnGOX), fatty acid oxidases from Arabidopsis (AtFAO4A, AtFAO1, AtFAO4B, and AtFAO3) and from *Candida* species (CcFAO1, CcFAO2, and CtFAO3).

The phylogenetic relationship among HTH and related GMC oxidoreductases is depicted in Figure 4.2 (See Appendix I) for the sequence alignment used for tree construction). These proteins were grouped into seven groups (bootstrap value > 91), including AtFAO (fatty acid oxidases of *Arabidopsis thaliana*), CcFAO (fatty acid oxidase of *Candida* species), MDL (mandelonitrile lyases), HTH-AtGMC (HTH and HTH-like GMC oxidoreductases in *Arabidopsis thaliana*), ADH (alcohol dehydrogenases), OXDH (an aryl alcohol oxidase and a pyranose dehydrogenase), and GOX (glucose oxidases). Arabidopsis HTH-like GMC proteins (AtGMC1 to AtGMC4) are clustered with HTH in Group HTH-AtGMC (bootstrap value = 91). The clade of HTH-AtGMC has an immediate common ancestor with the proteins in Group MDL, where PdMDL2 and all other putative mandelonitrile lyases are closely clustered (bootstrap value = 100). These two sister clades (HTH-AtGMC and MDL) form a higher order clade. On the same level, Group ADH, GOX and OXDH come together into a second higher order clade containing GMC oxidoreductase family proteins of diverse functions. The third higher order clade includes fatty acid oxidases of yeast (Group CcFAO) and Arabidopsis (Group AtFAO).

4.3.3 Conservation of functional residues

DNA lesions that underlie the mutant phenotype can lead to the identification of amino acid residues important for the biological function of a protein, and the conservation of these residues could be indicative of a similar protein function. Eight functional amino acid

residues have been previously identified by mutant screens (Krolikowski et al., 2003; Lolle et al., 1998). Residues corresponding to the eight functional amino acids of HTH were investigated in closely related MDLs (Table 4.2) and other GMC oxidoreductases (Appendix J).

Based on full sequence alignment all eight residues are conserved in the closely related but uncharacterized HTH-like proteins (AtGMC1-AtGMC4), except AtGMC5 where Gly-356 was substituted by alanine. Six of these eight residues are invariant in the immediate MDL sister clade, except residues corresponding to Gly-565 and Thr-566 of the HTH sequence. Rather, these two residues are mostly invariant within the MDL clade, being alanine/asparagine and threonine respectively.

4.3.4 **Hydroxynitrile lyase assay**

The proposed mandelonitrile lyase function for HTH is supported by the close phylogenetic relationship and matching functional residues (Figure 4.2 and Table 4.2) identified by multiple sequence alignment (Appendix I). To determine whether HTH has mandelonitrile lyase activity, a MBP-tagged recombinant HTH protein was generated and tested for activity using an *in vitro* enzyme assay. Figure 4.3 shows SDS-PAGE profiles of *E. coli* cell lysates obtained from strains transformed with the pMAL-c4x:*HTH* vector. Proteins of higher molecular mass, approximately the predicted size of the MBP-HTH fusion protein (~MW 114.1 kDa), are more abundant in the lysate of IPTG-induced (+) cells relative to un-induced cells (-).

To remove the MBP tag, affinity purified MBP-HTH was treated with factor Xa protease. Immunoblotting analysis shows that the MBP tag was cleaved in the factor Xa-treated samples (+). HTH (~MW 63.3 kDa; without the signal peptide) was detected using anti-HTH antibody, whereas the disassociated MBP tag (~MW 50.8 kDa) was only detected using the anti-MBP antibody.

Cleaved MBP-HTH samples that constitute a mixture of the HTH recombinant protein and MBP tag (labeled as MBP+HTH) as well as the almond MDL were used for the hydroxynitrile lyase activity assays. A commercially available MDL extracted from almond was used as a positive control. Reaction kinetics are shown in Figure 4.4. The ΔA_{275} /min was determined by absorbance detected in the linear range in the period of 7 to 15 minutes. The specific activity of almond MDL was 208 nmol/mg/min, in the range of activities indicated by the manufacture (80-240 nmol/mg/min). The column buffer (CB) in which the protease-treated recombinant HTH protein was solubilized showed little background effect on the enzymatic assay (d in Figure 4.4A). In contrast, no hydroxynitrile lyase activity was detected when HTH+MBP samples were used in the reactions.

4.3.5 **Alcohol dehydrogenase assay**

To determine whether HTH functions as an alcohol dehydrogenase, the MBP-tagged recombinant HTH protein used for the hydroxynitrile lyase assay was tested in an *in vitro* alcohol dehydrogenase (ADH) assay. The affinity purified MBP-HTH was cleaved with factor Xa protease to remove the MBP tag prior to performing the assay. The treated sample MBP+HTH was used in the dehydrogenation reaction. A commercially available alcohol

dehydrogenase isolated from yeast was used as a positive control. Alcohols of different carbon length and structure were used as the substrate, including ethanol (C_2H_5OH, 2 carbon), 1-hexanol ($CH_3(CH_2)_5OH$, 6-carbon aliphatic fatty alcohol) and benzyl alcohol ($C_6H_5CH_2OH$, aromatic alcohol). The reaction kinetics were determined by measuring ΔA_{340} /min in the linear range, and the calculated specific activities (nmol/mg/min) are listed (Figure 4.5). With this particular reaction setting, the yeast ADH reacted with ethanol at a specific activity of 216 ± 52 nmol/mg/min. When the carbon number of the substrate increased to six, yeast ADH activity dramatically diminished to 0.60 ± 0.23 nmol/mg/min, and no activity was detected when benzyl alcohol was used as the substrate. Benzyl alcohol was selected as a substrate because its chemical structure is similar to mandelonitrile lyase (Figure 4.5D), which has an additional −C≡N group. No such alcohol dehydrogenase enzymatic activity was detected in the *in vitro* assay using the above substrates.

4.3.6 **Glycosylation analyses**

Glycosylation, the attachment of sugar moieties to proteins, is a post-translational modification that is critical for a wide range of biological processes (Rayon et al., 1998; Strasser, 2014). When a plant gene is heterologously expressed in a prokaryotic organism, the recombinant protein is likely not properly glycosylated, and this can lead to a loss of its protein function. If glycosylation is important for HTH protein folding and function, the absence of this type of post-translational modification could lead to a loss of enzymatic activity.

To determine if the native HTH protein is glycosylated in Arabidopsis, protein extracted from *HTH_{pro}:HTH-EYFP* plants (See Chapter 3) was deglycosylated and analyzed for size shift. The crude extract was treated with a mix containing glycosidases that removes glycans from both N-link and O-linked (N+O) glycosylation sites. Protein samples were also treated with PNGase F which removes only N-linked glycans. Western blot analyses of HTH-EYFP protein using anti-GFP antibody revealed that HTH-EYFP in untreated crude protein extracts was detected at ~ 135 kDa and following glycosidase treatments migrated at ~124 kDa (Figure 4.6), showing an 11 kDa size shift in both types of deglycosylation reactions.

4.3.7 **Predicted HTH catalytic sites**

Amino acids that directly participate in the catalytic reaction mechanism, catalytic site residues, are often highly conserved in functionally related proteins as a result of common ancestry or convergent evolutionary processes (Torrance et al., 2005). To further explore the possible function of HTH, putative active sites were identified by a multiple sequence alignment with mandelonitrile lyase PdMDL2, a protein that is closely related to HTH and has been characterized by crystallography and biochemical assays. PdMDL2 has mandelonitrile lyase activity and three catalytic sites of PdMDL2, Tyr-484, His-486 and His-524, have been identified (Dreveny et al., 2001; Dreveny et al., 2002; Dreveny et al., 2009). The cladogram that accompanies the active site table and grouping was made according to the radial phylogenetic tree shown in Figure 4.2. Except Group HTH-AtGMC, each group is represented by at least one protein that has been experimentally validated (Cheng et al., 2004; Dickinson and Wadforth, 1992; Dreveny et al., 2001; Dreveny et al., 2002; Hecht et al.,

1993; Kalisz et al., 1997; van Beilen et al., 1992; Vanhanen et al., 2000; Whittington et al., 1990; Yue et al., 1999).

For all of these GMC oxidoreductases, the second active site histidine (His) is strictly conserved. The first active site residue is slightly more variable while the third active site residue is either an asparagine (N) or a histidine (H). Identified putative catalytic residues are generally invariant within each phylogenetic group, as demonstrated in the CsFAO, MDL, and GOX clades where the conservation of putative active sites is stringent. Moreover, the fatty acid oxidase catalytic residues of two functionally similar groups, AtFAO (Arabidopsis) and CsFAO (yeast), are nearly identical. Importantly, the catalytic residues identified by multiple sequence alignment match known catalytic sites of SsCHOX, PeAAO, AmPDH and AnGOX (residues in bold, Figure 4.7). In spite of having a higher number of matching functional residues, HTH shares only the conserved histidine with the closely related MDLs. However, HTH and PoADH, a bacterial alcohol dehydrogenase that was experimentally verified to have mid-chain fatty acid alcohol dehydrogenase activity (van Beilen et al., 1992) have identical putative active site compositions.

4.3.8 **Structural models of HTH**

To examine the relative spatial positions of functional residues and predicted catalytic sites, putative tertiary structures for HTH were constructed based on mandelonitrile lyase PdMDL2 and pyranose dehydrogenase AmPDH (Figure 4.8). These two templates are among the best fits identified by the Phyre 2 server. Out of the 594 amino acid residues making up the HTH protein, 510 and 505 amino acids were modeled with 100% confidence on the PdMDL2 (PDB: 1JU2) and AmPDH (PDB: 4H7U) crystal structure, respectively. The predicted 19-aa

signal peptide sequence did not match with either model and hence was not included in the predicted protein structure. The confidence of tertiary structure modeling is shown in Appendix K. The protein sequence alignments showing the secondary structure of the two templates and the predicted secondary structure of HTH are included in Appendix L and Appendix M. The tertiary structure modeling for the 567 aa-long HTH isoform predicted by the splice variant was also modeled onto the PdMDL2 strucure. The protein sequence alignments showing the secondary structure and predicted tertiary isoform structure are shown in Appendix N and Appendix O repsectively.

The crystal structure of PdMDL2 (Figure 4.8A) shows that three active site residues (shown in red) are positioned in the catalytic pocket (grey) and are in close proximity to the FAD (flavin adenine dinucleotide) cofactor. The three-dimensional structure of AmPDH (Figure 4.8C) shares great overall similarity to PdMDL2 and consists of one alignment-based (Tyr-535) and two experimentally validated (His-537 and His-581) active sites (Tan et al., 2013). As expected, the predicted active site residues are situated in the catalytic pocket of the PdMDL2- and AmPDH-templated HTH tertiary structural models (Figure 4.8B, D). Moreover, among seven of the eight functional residues (except Gly-356) that are included in the models, five of these functional residues are either within (Pro-564, Gly-565, and Thr-566) or close to (Gly-218 and Arg-227) the catalytic pocket that is adjacent to the cofactor FAD/FED riboflavin ring. It is interesting to note that the peptide sequence absent in the smaller HTH isoform forms a helix on the protein surface according to the PdMDL2 model (Figure 4.8B).

4.3.9 **Coexpression gene network**

The coexpression profile analysis was conducted with the aim of shedding light on HTH's possible biological functions, focusing on evidence associated with its mutant cuticle phenotype and stress response (see Chapter 3). The top twenty genes coexpressed with *HTH* and their (putative) functions, as determined by GeneMania analysis, are listed in Table 4.3. The complete network and function categories of selected genes are shown in Figure 4.9. The degree of correlation is proportional to the node/circle size depicted in the network.

Genes that are involved, or likely so, in lipid metabolism or transport (highlighted in blue) include *GLYCOSYLPHOSPHATIDYLINOSITOL-ANCHORED LIPID PROTEIN TRANSFER* (*LTPG1*), *LACERATA* (*LCR*), *SUBTILASE 1.3* (*SBT1.3*), *LTPG2*, *SEED FATTY ACID REDUER 5* (*SFAR5*), *BODYGUARD 1* (*BDG1*), *3-KETOACYL-COA SYNTHASE 5/ECERIFERUM 60* (*KCS5/CER60*), *CYTOCHROME P450 FAMILY 86 PROTEIN CYP86A4*, *DEFECTIVE IN CUTICULAR RIDGES/PERMEABLE LEAVES 3* (*DCR/PEL3*). Among these genes, some have been shown to play an important role in cuticle formation, specifically in either cutin/wax monomer synthesis (*LCR*, *KCS5*/*CER60* and *CYP86A4*), lipid transport (*LTPG1*, *LTPG2* and *AT4G16140*), or cutin/wax monomer/oligomer polymerization (*SBT1.3*, *AT5G45670*, *AT5G45950*, *BDG1* and *DCR*/*PEL3*). Genes that are essential for normal seed coat development (Figure 4.9, asterisk) include *LTPG1*, *LTPG2*, *BDG1* and *DCR*/*PEL3*. Interestingly, the majority of coexpressed genes essential for cuticle development are also associated with plant stress/defense response to biotic and abiotic stress (highlighted in pink). Genes grouped to this category either exhibit increased stress-induced expression (*SBT1*.3 and *LTL1*), have been implicated in stress signaling (*LTPG1* and *LCR*),

or confer increased pathogen resistance when mutated (*LCR*, *LTPG2*, *BDG1* and *DCR*/*PEL3*). Overall, HTH appears to be networked with genes whose functions are involved in lipid processing or stress response, and in some cases both functions.

4.3.10 **Phylogenetic distribution of proteins encoded by coexpressed genes in the evolutionary tree**

One omics-based method to infer biological function is phylogenetic profiling. By correlating the phylogenetic distribution of target genes with phenotypic characteristics or with a set of genes with known function, possible biological function can be deduced (Kensche et al., 2008). A phylogenetic profile of the coexpression gene network of *HTH* was generated using the String database (Figure 4.10) and included protein sequences encoded by *HTH* and top twenty coexpressed genes. Four housekeeping genes included as controls were *FUMARASE 1* (*FUM1*), *TUBULIN ALPHA-4* (*TUA4*), *ACTIN 7* and *RIBULOSE BISPHOSPHATE CAROXYLASE SMALL CHAIN 1A* (*RBCS1A*). FUM1 is a mitochondrial-localized essential protein that plays a role in in the tricarboxylic acid cycle (Pracharoenwattana et al., 2010). TUA4 and ACTIN7 are components of eukaryotic cytoskeletal microtubules and microfilaments, respectively. RBCS1A belongs to the RuBisCO small subunit protein family and is important in the carbon fixation process.

Based on the phylogenetic profile, homologous proteins of FUM1 occur in most organisms across prokaryotes and eukaryotes and those of TUA4 and ACTIN7 only in eukaryotes (including metazoa and fungi). In contrast, proteins with high homology to RBCS1A occur only in land plants and algae. This phylogenetic profile of known proteins

matches with evolutionary characteristics of different groups of organisms. In contrast, the phylogenetic distribution of proteins encoded by *HTH* and its coexpressed genes are predominant in land plants but not in algae. Among them, homologous sequences of LTPG1, LTPG2 and MYC1 are restricted to land plants. No homologues proteins (cut off: 4%) were identified out of *Brassicaceae* for two uncharacterized proteins AT4G16140 and AT4G29020. Proteins with less than 15% homology outside of the land plants are LTL1, AT5G45670, AT5G45950, AT4G18970, DRB5 and DCR/PEL3. Proteins that have less than 25% homology outside of the land plant group consist of LCR, AT5G13400, AT1G10640, BDG1, OCT3, CYP86A4 and HTH. KCS5/CER60 and SBT1.3 however, have greater than 37% homology outside of the land plants.

Figure 4.1 Schematic representation of the HTH protein showing the corresponding six exons, predicted glucose-methanol-choline (GMC) oxidoreductase domains, N-terminal signal peptide, posttranslational modification sites, active sites, and mutant phenotype-determining residues. The signal peptide and GMC domains were predicted by UniProtKB/Swiss-Prot at Expasy (http://www.expasy.org). Posttranslational modification site recognition was performed by PROSITE, and the motifs are labelled with the composition residues and their positions. Amino acid residues identified by genetic analysis (Krolikowski et al., 2003) showing sequence positions and corresponding mutant designations. Active site predictions were generated by protein sequence alignment to a mandelonitrile lyase (see Section 4.3.5). Signal: putative signal peptide; GMC_oxred_N: Glucose-Methanol- Choline oxidoreductase family protein N-terminal domain; GMC_oxred_C: Glucose-Methanol-Choline oxidoreductase family protein C-terminal domain. Sequences for posttranslational modification sites: N-glycosylation site, N-{P}-[ST]-{P}, Protein kinase C phosphorylation site, [ST]-x-[RK]. The eight genetically identified functional amino acid residues are designated with blue bars, and the three putative catalytic active sites with red bars.

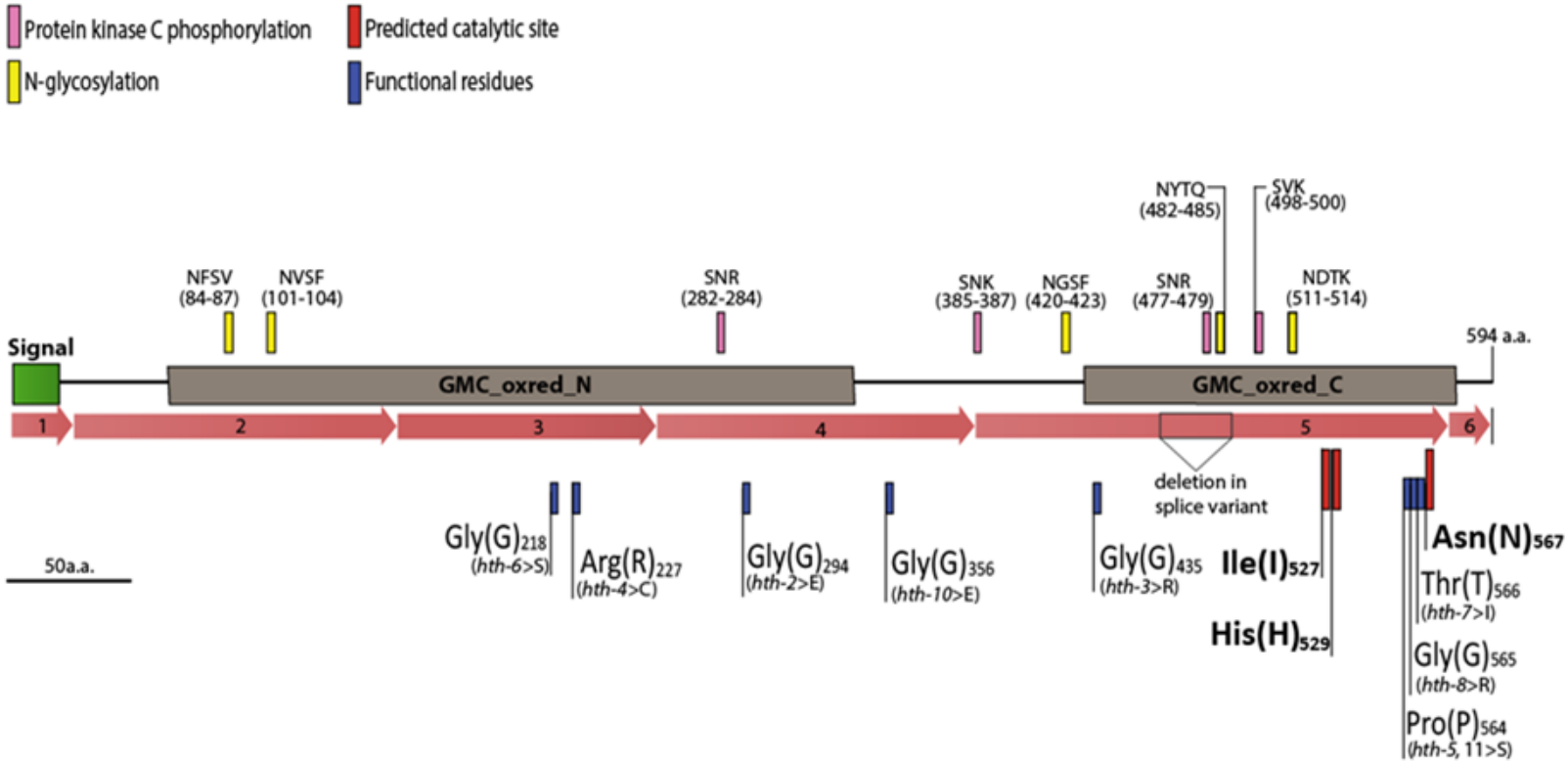

Protein kinase C phosphorylation
Predicted catalytic site
N-glycosylation
Functional residues
Signal
NFSV (84-87)
NVSF (101-104)
SNR (282-284)
SNK (385-387)
NGSF (420-423)
SNR (477-479)
NYTQ (482-485)
SVK (498-500)
NDTK (511-514)
594 a.a.
GMC_oxred_N
GMC_oxred_C
1
2
3
4
5
6
deletion in splice variant
50a.a.
Gly(G)218 (hth-6>S)
Arg(R)227 (hth-4>C)
Gly(G)294 (hth-2>E)
Gly(G)356 (hth-10>E)
Gly(G)435 (hth-3>R)
Ile(I)527
His(H)529
Asn(N)567
Thr(T)566 (hth-7>I)
Gly(G)565 (hth-8>R)
Pro(P)564 (hth-5, 11>S)

Table 4.1 List of HTH-related proteins investigated in this chapter. These proteins were selected based on similarity to the HTH protein with respect to sequence identity and/or theoretical protein folding structure (http://blast.ncbi.nlm.nih.gov; http://www.sbg.bio.ic.ac.uk/phyre2). The first two letters of the sequence names represent the organism of origin. AAO: aryl alcohol oxidase; ADH: alcohol dehydrogenase; CHOX: cholesterol oxidase; FAO: fatty acid oxidase; GOX: glucose oxidase; MDL: mandelonitrile lyase; PDH: pyranose dehydrogenase. UniProt codes are referenced from http://www.uniprot.org/. NCBI accession numbers are referred from http://www.ncbi.nlm.nih.gov.

Sequence Name.	UniProt/NCBI Accession	Sequence Identity	Sequence Coverage	E-value	Organism
AtGMC2	Q94KD2	52%	92%	~0	*Arabidopsis thaliana*
AtGMC4	Q66GI5	51%	92%	~0	*Arabidopsis thaliana*
AtGMC3	Q93ZK1	51%	88%	~0	*Arabidopsis thaliana*
AtGMC1	F4KEQ5	50%	89%	$3.00e^{-175}$	*Arabidopsis thaliana*
AtGMC5	Q9XI68	50%	80%	$5.00e^{-94}$	*Arabidopsis thaliana*
PdMDL2	Q945K2	40%	88%	$5.00e^{-128}$	*Prunus dulcis*
PsMDL3	P52707	39%	90%	$4.00e^{-128}$	*Prunus serotina*
PsMDL2	O50048	39%	90%	$7.00e^{-128}$	*Prunus serotina*
PsMDL1	P52706	39%	88%	$2.00e^{-128}$	*Prunus serotina*
PsMDL4	O82784	39%	88%	$6.00e^{-125}$	*Prunus serotina*
PdMDL1	O24243	38%	88%	$5.00e^{-118}$	*Prunus dulcis*
PsMDL5	O82435	37%	88%	$2.00e^{-117}$	*Prunus serotina*
SsCHOX	P12676	32%	20%	$3.00e^{-06}$	*Streptomyces sp.*
AtFAO1	Q9ZWB9	31%	13%	$1.20e^{+00}$	*Arabidopsis thaliana*
AtFAO4A	O65709	30%	12%	$7.00e^{-05}$	*Arabidopsis thaliana*
CcFAO2	Q9P8D7	28%	35%	$5.00e^{-07}$	*Candida cloacae*
AtFAO4B	Q94BP3	28%	24%	$1.50e^{-02}$	*Arabidopsis thaliana*
CtFAOT	Q9P8D9	28%	23%	$3.00e^{-04}$	*Candida tropicalis*
CcFAO1	Q9P8D8	28%	17%	$1.00e^{-03}$	*Candida cloacae*
NrADH	WP_022978378.1	26%	88%	$4.00e^{-25}$	*Nevskia ramosa*
ToADH	5DPH3	26%	87%	$6.00e^{-32}$	*Thalassolituus oleivorans*
AmADH	WP_020743879.1	26%	86%	$1.00e^{-23}$	*Alteromonas mediterranea*
AmPDH	Q3L245	25%	93%	$2.00e^{-26}$	*Agaricus meleagris*
PpADH	Q9WWW2	25%	88%	$4.00e^{-29}$	*Pseudomonas putida*
PoADH	Q00593	24%	88%	$1.00e^{-25}$	*Pseudomonas oleovorans*
PeAAO	O94219	24%	87%	$6.00e^{-28}$	*Pleurotus eryingii*
AnGOX	P13006	24%	87%	$4.00e^{-18}$	*Aspergillus niger*
AtFAO3	Q9LW56	24%	24%	$2.00e^{-03}$	*Arabidopsis thaliana*

Figure 4.2 Phylogenetic relationships of HTH-related proteins. An unrooted radial phylogenetic tree constructed using the alignment of GMC (glucose-methanol-choline) oxidoreductases related to HTH (arrow head). These proteins were selected in terms of their sequence identity and/or theoretical protein folding structure (http://blast.ncbi.nlm.nih.gov; http://www.sbg.bio.ic.ac.uk/phyre2). Protein sequence alignments were generated using ClustalO as implemented in the SeaView package (version: 4.5.0), and phylogenetic trees were computed using PHYML within the SeaView package. The bootstrap values are indicated for the higher order clades. Seven clades (bootstrap value > 91) were identified. Group AtFAO proteins are closely related to corresponding genes in the *Candida* species in Group CsFAO, which comprises long-chain fatty acid alcohol oxidases capable of oxidizing α,ω-diols and probably, ω-hydroxy fatty acids. Group MDL is represented by mandelonitrile lyase PdMDL2. Group HTH-AtGMC (HTH and HTH-like GMC oxidoreductases) comprises five proteins including HTH. Group ADH includes five bacterial fatty acid dehydrogenases. Group OXDH include a fungal aryl alcohol oxidase and a fungal pyranose dehydrogenase. Group GOX contains two closely related glucose oxidases. HTH and the mandelonitrile lyase PdMDL2 are labeled with an arrow. Parameters for PHYML: bootstrap with 100 replicates, model-given amino acid equilibrium frequency, NNI for tree searching operation, and BioNJ for starting tree topology. The first two letters of the sequence names represent the organism of origin. AAO: aryl alcohol oxidase; ADH: alcohol dehydrogenase; CHOX: cholesterol oxidase; FAO: fatty acid oxidase; GOX: glucose oxidase; MDL: mandelonitrile lyase; PDH: pyranose dehydrogenase. UniProt codes are referenced from http://www.uniprot.org/. NCBI accession numbers are referred from http://www.ncbi.nlm.nih.gov.

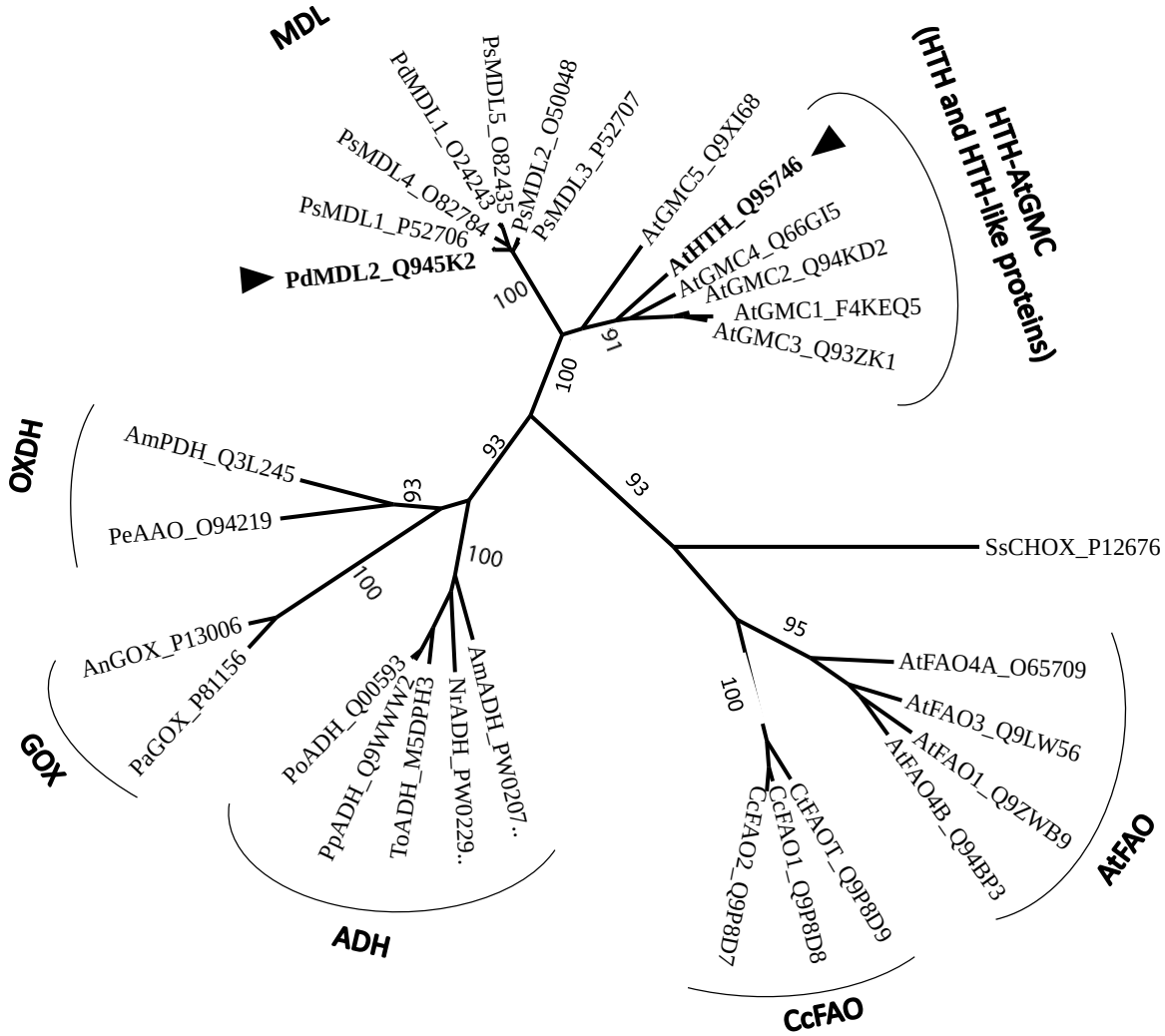

MDL
PdMDL1_O24243
PsMDL5_O82435
PsMDL2_O50048
PsMDL3_P52707
PsMDL4_O82784
PsMDL1_P52706
PdMDL2_Q945K2
100
AtGMC5_Q9XI68
AtHTH_Q9S746
AtGMC4_Q66GI5
AtGMC2_Q94KD2
AtGMC1_F4KEQ5
AtGMC3_Q93ZK1
91
HTH-AtGMC
(HTH and HTH-like proteins)
100
93
OXDH
AmPDH_Q3L245
PeAAO_O94219
93
100
AnGOX_P13006
PaGOX_P81156
GOX
100
PoADH_Q00593
PpADH_Q9WWW2
ToADH_M5DPH3
NrADH_PW0229..
AmADH_PW0207..
ADH
93
SsCHOX_P12676
95
AtFAO4A_O65709
AtFAO3_Q9LW56
AtFAO1_Q9ZWB9
AtFAO4B_Q94BP3
AtFAO
100
CtFAOT_Q9P8D9
CcFAO1_Q9P8D8
CcFAO2_Q9P8D7
CcFAO

Table 4.2 Eight functional HTH amino acids and the corresponding residues in four HTH-like GMC oxidoreductases (AtGMC1 - AtGMC4) and mandelonitrile lyase PdMDL2. These eight HTH residues were identified by genetic analysis (Krolikowski et al., 2003) and are labeled with the corresponding HTH protein sequence position. The corresponding residues in other GMC proteins were identified based on sequence alignment (see Appendix I). Amino acid conservation is highlighted.

	Protein	UniProt/NCBI Accession	Functional Residues of HTH / Corresponding Residues in other GMC Proteins								Match
	HTH	**Q9S746**	**G_{218}**	**R_{227}**	**G_{294}**	**G_{356}**	**G_{435}**	**P_{564}**	**G_{565}**	**T_{566}**	Self
HTH-like proteins	AtGMC1	F4KEQ5	G	R	G	G	G	P	G	T	
	AtGMC2	Q94KD2	G	R	G	G	G	P	G	T	
	AtGMC3	Q93ZK1	G	R	G	G	G	P	G	T	
	AtGMC4	Q66GI5	G	R	G	G	G	P	G	T	8/8
Mandelonitrile lyase	PdMDL2	Q945K2	G	R	G	G	G	P	A	S	
	PsMDL1	P52706	G	R	G	G	G	P	A	S	
	PdMDL1	O24243	G	R	G	G	G	P	N	S	
	PsMDL5	O82435	G	R	G	G	G	P	N	S	
	PsMDL4	O82784	G	R	G	G	G	P	A	S	
	PsMDL2	O50048	G	R	G	G	G	P	A	S	
	PsMDL3	P52707	G	R	G	G	G	P	A	S	6/8

Figure 4.3 Protein profiles obtained from bacterial cell lysates and immunoblots of affinity purified MBP-HTH recombinant protein probed with an anti-HTH or anti-MBP antibody. (A) Cell lysate SDS-PAGE profiles of *E. coli* proteins obtained from cells transformed with the pMAL-c4x:*HTH* construct. IPTG-induced cells expressed more MBP-HTH protein (~MW 114.1 kDa; arrow head) than uninduced cells (Control). (B-C) Immunoblot analysis of affinity purified MBP-HTH (5 µg) that was treated with (+) or without (-) the protease factor Xa that cleaves off the MBP tag. The MBP-HTH (~MW 114.1 kDa) was detected by anti-HTH and anti-MBP antibodies in the uncleaved (-) sample. In the Factor Xa treated sample, HTH (~MW 63.3 kDa) was detected by the anti-HTH antibody, whereas the disassociated MBP tag (~MW 50.8 kDa) was detected by the anti-MBP antibody. M: protein molecular mass marker.

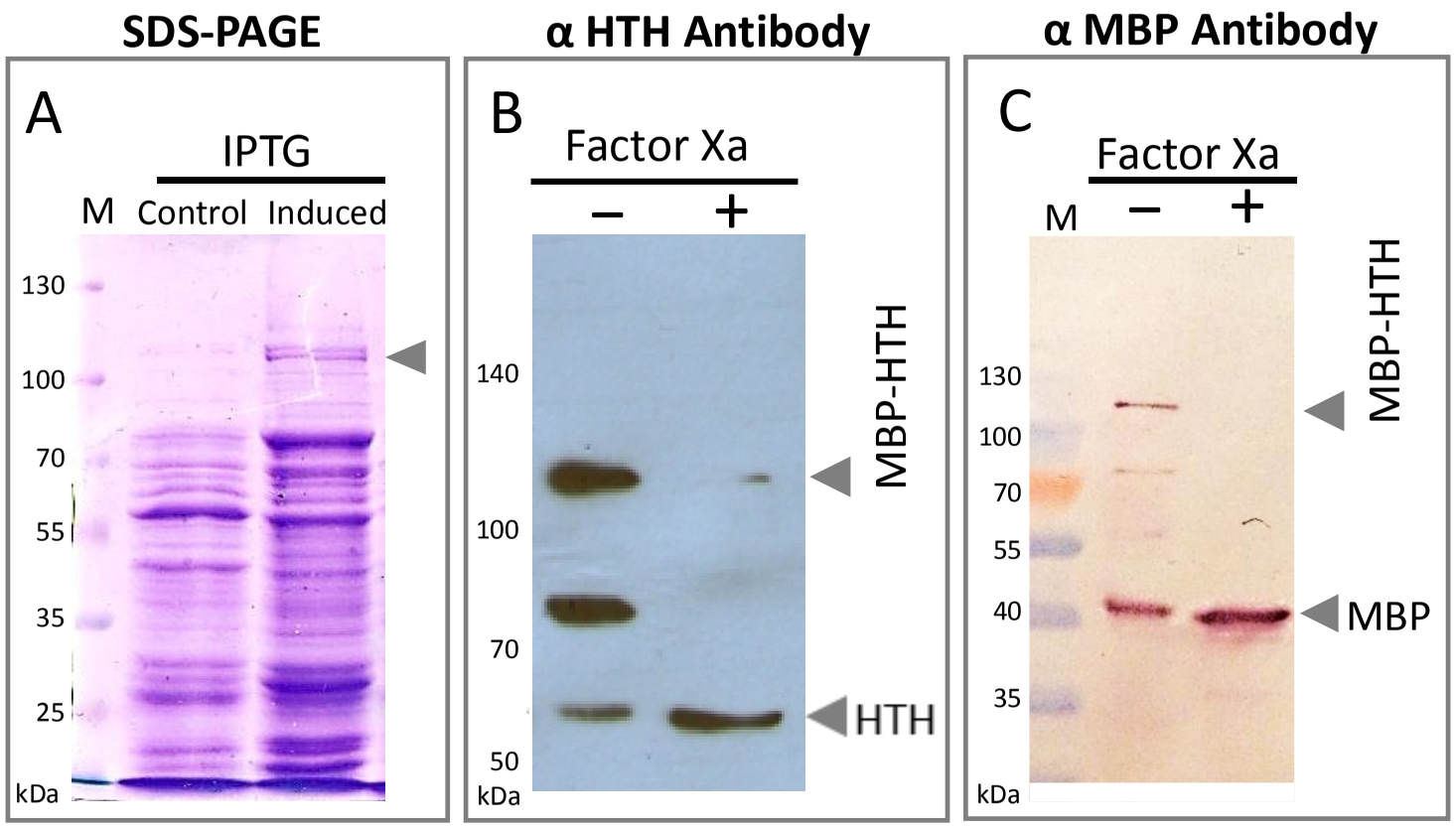
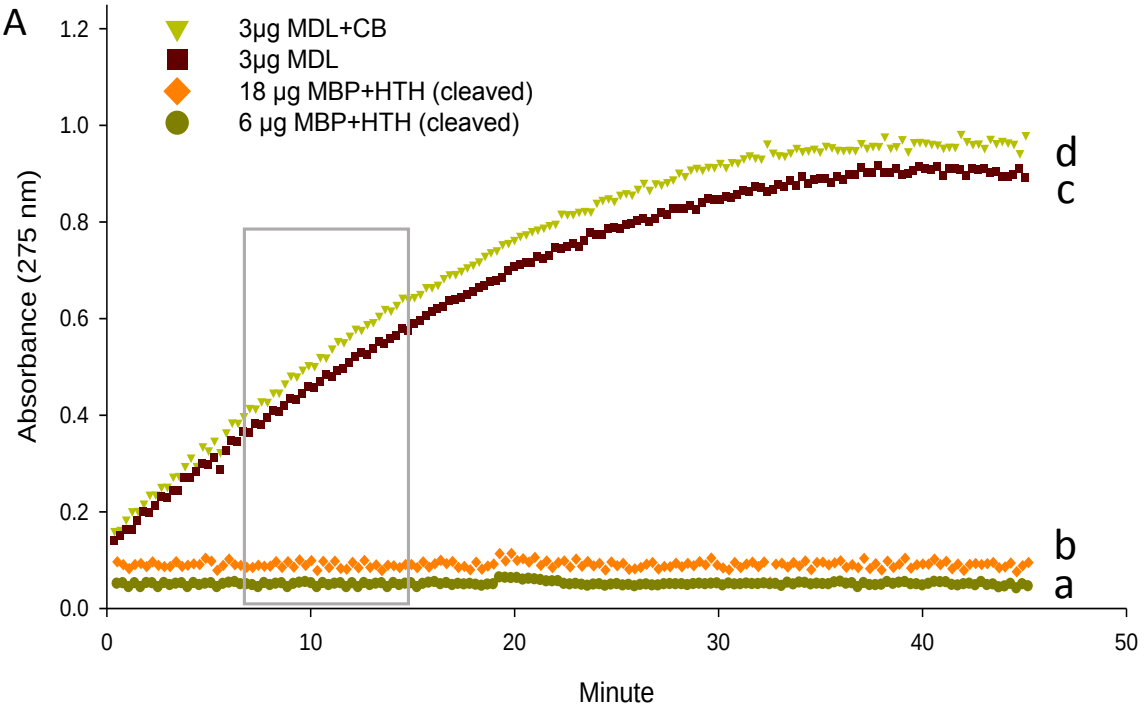

SDS-PAGE
A
IPTG
M
Control
Induced
130
100
70
55
35
25
kDa
α HTH Antibody
B
Factor Xa
−
+
140
100
70
50
kDa
MBP-HTH
HTH
α MBP Antibody
C
Factor Xa
M
−
+
130
100
70
55
40
35
kDa
MBP-HTH
MBP

Figure 4.4 Spectrophotometric assays of hydroxynitrile lyase activity. (A) Reaction kinetics of four protein samples. Affinity purified MBP-HTH protein was cleaved with Factor Xa to remove the MBP tag prior to conducting the assay. 6µg (a) and 18µg (b) of the HTH and MBP tag protein (HTH+MBP) mix was directly used in the reaction. Absorbance at 275 nm was continuously monitored for the production of benzaldehyde to determine enzymatic activities. (c) Mandelonitrile lyase (MDL) extracted from almond was used as positive control. (d) The effect of Column Buffer (CB) in which MBP+HTH was solubilized is shown. (B) Calculated specific activities (nmol/mg protein/min). The ΔA275 /min was determined by absorbance detected in the period of 7 to 15 minutes (boxed linear range). All absorbance readings were corrected against a blank sample. Values are the mean ± 1 standard error. Experiments were repeated three times.

A



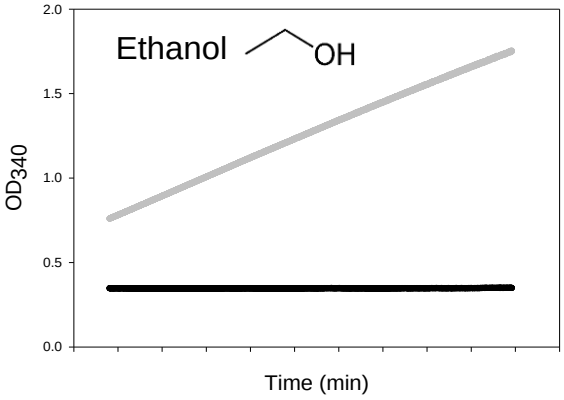
B

Specific Activities (nmoles/mg protein/min) of MDL and cleaved MBP-HTH

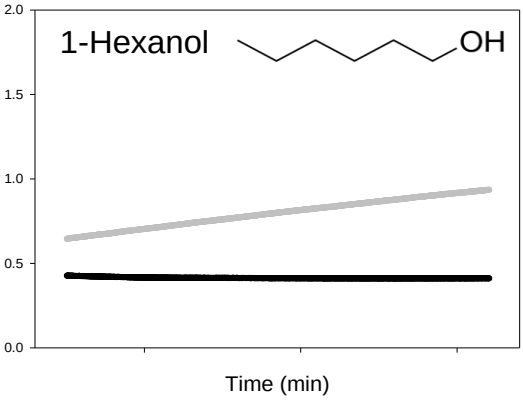
	6 µg cleaved MBP+HTH (a)	18 µg cleaved MBP+HTH (b)	MDL (c)	MDL+CB (d)
ΔA_{275} /min	0.00003	0.00003	0.02531	0.02662
Specific Activity	0.2±0.7	0.2±0.8	198±3	208±5

Figure 4.5 Spectrophotometric assays for alcohol dehydrogenase activity of recombinant HTH protein.. (A-C) Reaction kinetics of alcohol dehydrogenase activities using different alcohol substrates. Affinity purified MBP-HTH protein was cleaved with Factor Xa to remove the MBP tag, and this digest (MBP+HTH, back line) and alcohol dehydrogenase from yeast (positive control, grey line) were used for assays. (D) Benzyl alcohol was selected as a substrate due to its similar structure to mandelonitrile. (E) Calculated specific activities (nmol/mg protein/min). The ΔA_{340} /min was determined by absorbance detected in the linear range. All absorbance readings were corrected against a blank sample. Values are the mean ± 1 standard error. Experiments were repeated three times.

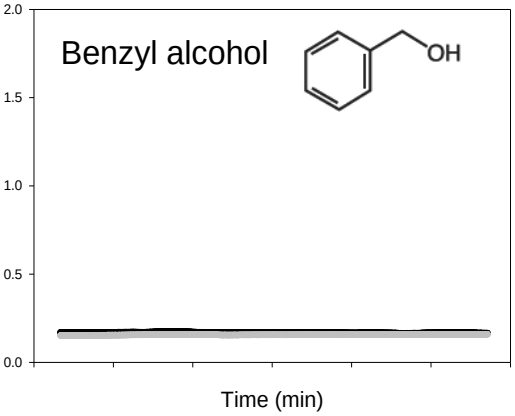
A



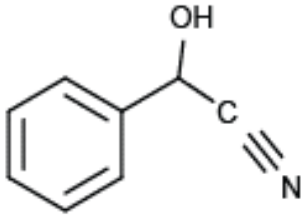
B



C



D



Mandelonitrile
(2-hydroxy-2phenylacetonitrile)

E

Substrate	**Specific activity (nmole/mg protein/min)**	
	Yeast ADH	HTH
Ethanol (2C)	**216 ± 52**	NOT DETECTABLE
1-Hexanol (6C)	0.60 ± 0.23	NOT DETECTABLE
Benzyl alcohol (RING)	NOT DETECTABLE	NOT DETECTABLE

Figure 4.6 Western blot analyses of HTH protein glycosylation using protein extract from *HTH_{pro}:HTH-EYFP* plants. Probed with anti-GFP antibody, HTH-EYFP in the crude protein extract (lane 1) was detected at ~ 135 kDa. When the crude extract was subjected to a deglycosylation mix containing glycosidases that remove glycans from both N- and O-linked (N+O) glycosylation sites, HTH-EYFP (lane 3) was detected at the theoretical size ~124 kDa. When the protein extract was treated with PNGase F (lane 5), a glycosidase that removes only N-linked glycans, the decrease in size was similar to that between lane 2 and lane 3. Each mock reaction (negative control) contained all the components except glycosidases. Molecular masses are indicated in kDa.

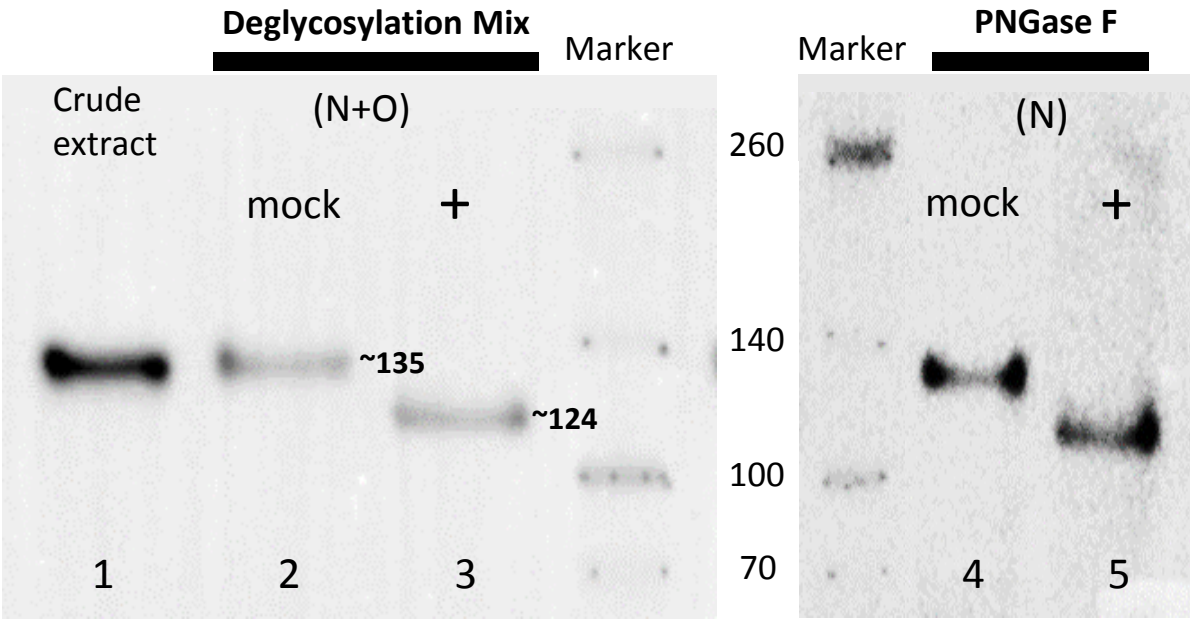
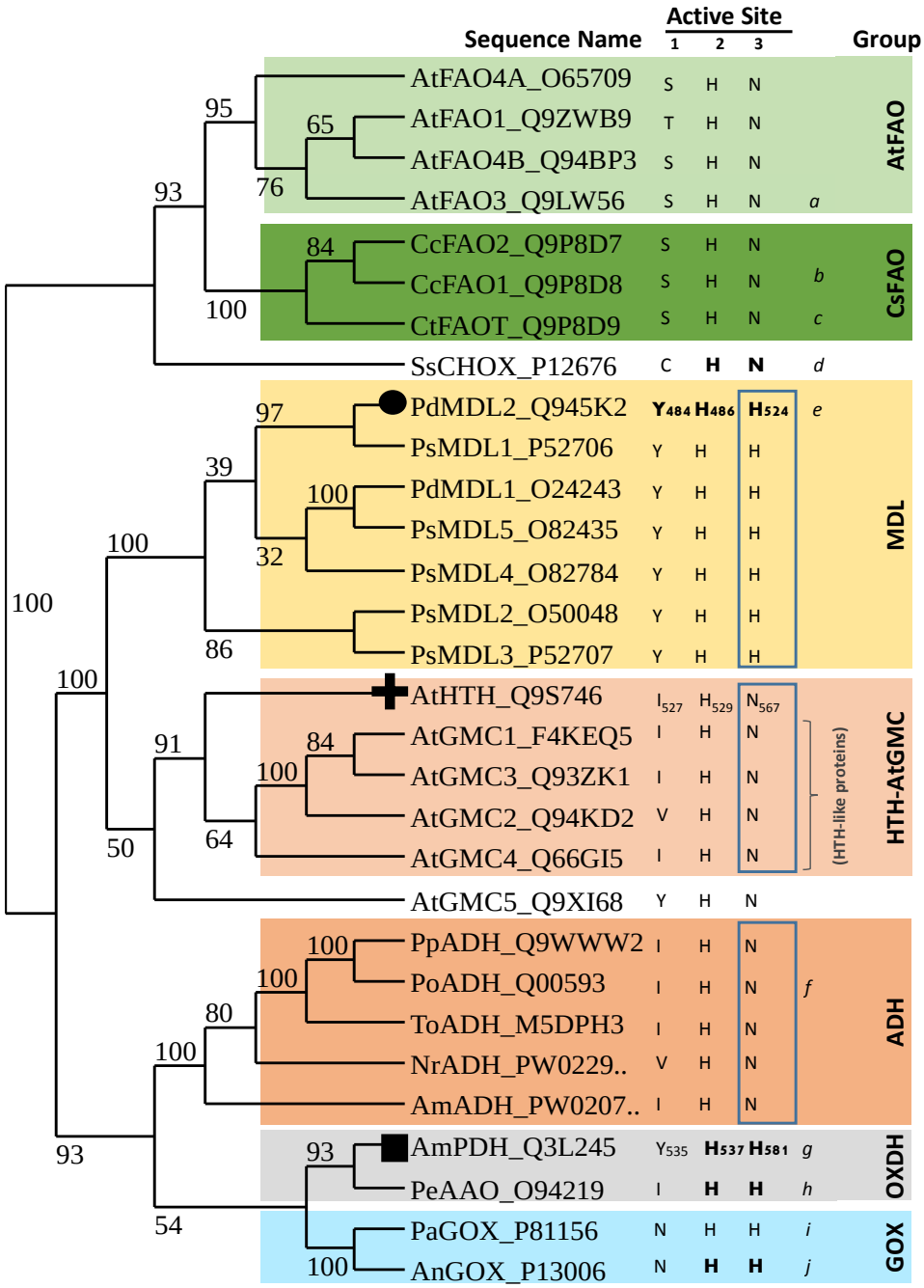

Deglycosylation Mix
Marker
Marker
PNGase F
Crude extract
(N+O)
mock
+
~135
~124
260
140
100
70
1
2
3
(N)
mock
+
4
5

Figure 4.7 Putative key catalytic sites of HTH and other glucose-methanol-choline (GMC) oxidoreductases. The three putative protein active sites were identified by protein sequence alignment with PdMDL2 and its crystallography-verified sites. Active sites that have been identified by crystal structure are in bold. The previously identified catalytic sites PdMDL2 (circle) and AmPDH (square) as well as the predicted sites of HTH (cross) are labelled with the corresponding positions of their own sequences. The third active site (boxed) of HTH and HTH-like proteins (AtGMC1-AtGMC4) matches with that of the ADH group rather than the MDL group. The cladogram is derived from the phylogenetic tree shown in Figure 4.2, and the bootstrap values indicated for each branch. For information about sequence names and grouping, refer to Table 4.1. AAO: aryl alcohol oxidase; ADH: alcohol dehydrogenase; CHOX: cholesterol oxidase; FAO: fatty acid oxidase; GOX: glucose oxidase; MDL: mandelonitrile lyase; OXDH: a group of oxidase and dehydrogenase; PDH: pyranose dehydrogenase.



Reference

a, long-chain fatty acid oxidase activity (Cheng et al., 2004)

b, long-chain fatty acid oxidase activity (Vanhanen et al., 2000)

c, long-chain fatty acid oxidase activity (Dickinson et al., 1992)

d, active sites identified by mutants and crystal structure (Yue, Kass et al., 1999)

e, mandelonitrile lyase activity; active sites determined by crystal structure and substrate binding (Dreveny et al., 2001; Dreveny et al., 2002; Dreveny et al., 2009)

f, alcohol dehydrogenase activity by complementation of mutants (van Beilen et al., 1992)

g, pyranose dehydrogenase activity, active sites (Tan et al., 2013)

h, aryl alcohol oxidase activity, active sites (Fernandez et al., 2009)

i, glucose oxidase activity (Kalisz et al., 1997)

j, glucose oxidase activity, active sites (Hecht et al., 1993; Whittington et al., 1990)

Figure 4.8 Ribbon diagrams showing two hypothetical three-dimensional structures of HTH. Models were generated by the PHYRE2 server and graphed by the PYMOL tool. The predicted enzymatic pocket is shown in grey, and the putative active site residues in red. Seven functional residues identified by mutant screens are shown in blue. The position of the cofactor was superimposed onto the model. 510 (86%) and 505 HTH residues (85%) were modeled with 100.0% confidence on the PdMDL2 and AmPDH crystal structure, respectively. (A) The tertiary structure of the mandelonitrile lyase PdMDL2 (PDB:1JU2) with the cofactor FAD. (B) Hypothetical structure of full length (594 aa) HTH modeled on PdMDL2. The 27-aa deletion sequence of the isoform resulting from transcript alternative splicing is indicated on the modeled structure. (C) The tertiary structure of pyranose dehydrogenase AmPDH (PDB: 4H7U) with the cofactor FED. (D) Hypothetical structure of HTH modeled on AmPDH. The isoform deletion sequence was not included in modeling by the PHYRE2 server. FAD, flavin adenine dinucleotide; FED, [(2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl]methyl (2R,3S,4S)-2,3,4-trihydroxy-5-[(4aR)-4a-hydroxy-7,8-dimethyl-2,4-dioxo-3,4,4a,5-tetrahydrobenzo[g]pteridin-10(2H)-yl]pentyl dihydrogen diphosphate.

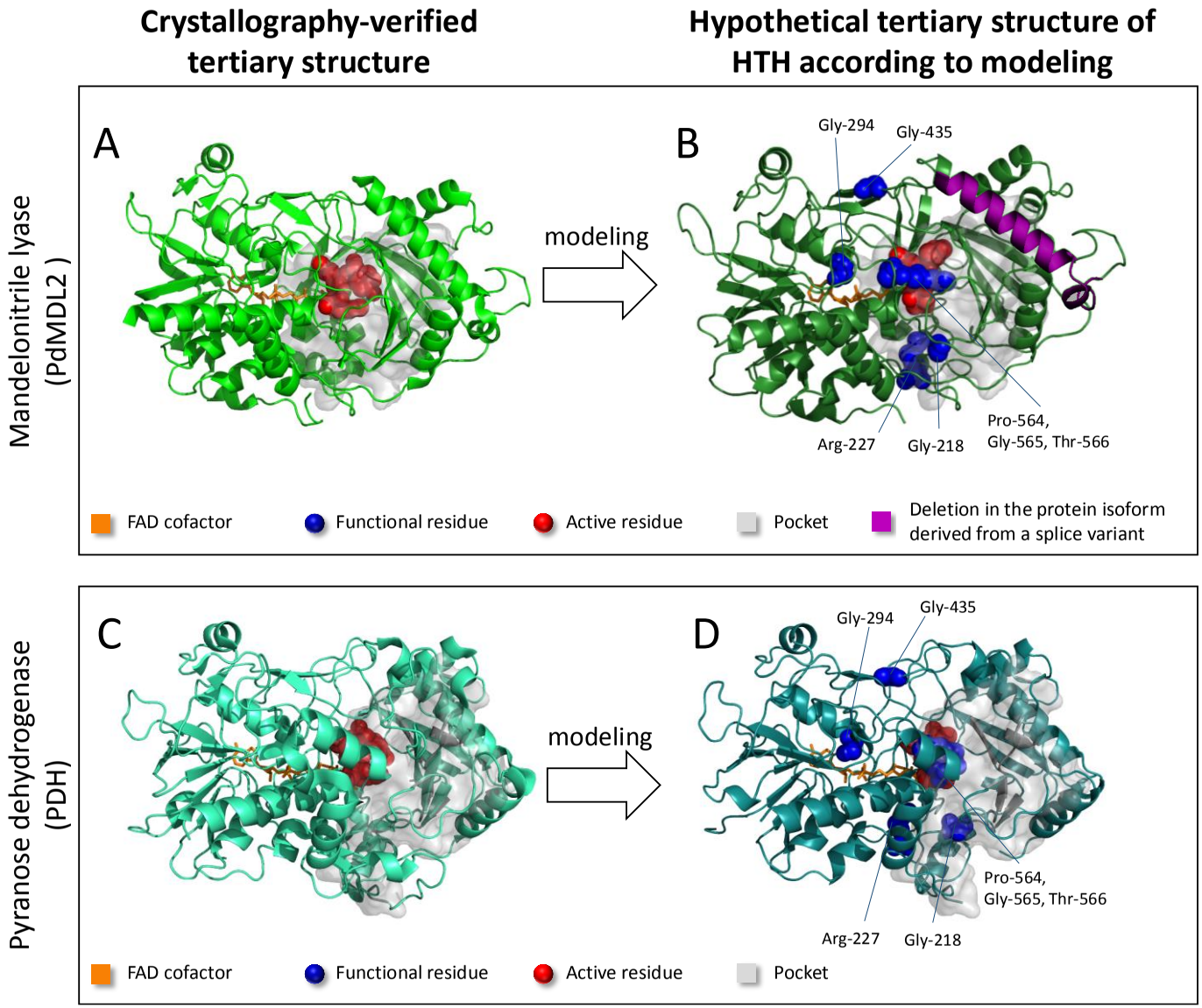
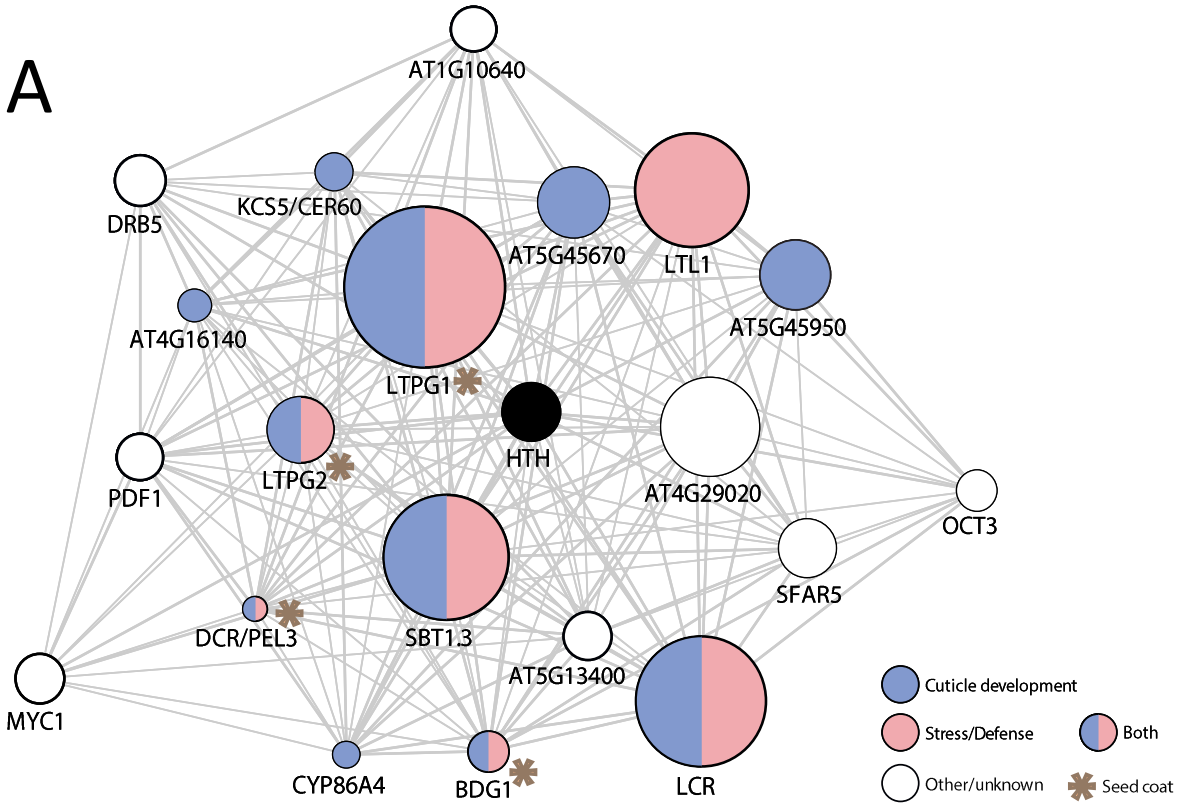

Crystallography-verified tertiary structure
Hypothetical tertiary structure of HTH according to modeling
Mandelonitrile lyase (PdMDL2)
A
modeling
B
Gly-294
Gly-435
Arg-227
Gly-218
Pro-564, Gly-565, Thr-566
FAD cofactor
Functional residue
Active residue
Pocket
Deletion in the protein isoform derived from a splice variant
Pyranose dehydrogenase (PDH)
C
modeling
D
Gly-294
Gly-435
Arg-227
Gly-218
Pro-564, Gly-565, Thr-566
FAD cofactor
Functional residue
Active residue
Pocket

Table 4.3 The list of top 20 coexpressed genes and their encoded protein functions. Genes are ranked according to the correlation coefficient. The coexpression profile was generated using the GeneMania plugin of Cytoscape with the *Arabidopsis thaliana* dataset. Abbreviation: *LTPG, GLYCOSYLPHOSPHATIDYLINOSITOL-ANCHORED LIPID PROTEIN TRANSFER; LCR, LACERATA; SBT, SUBTILASE; LTL, LI-TOLERANT LIPASE; SFAR, SEED FATTY ACID REDUCER; DRB, DOUBLE STRANDED RNA-BINDING PROTEIN; MYC1, TRANSCRIPTION FACTOR MYC1; PDF, PROTODERMAL FACTOR; BDG, BODYGUARD; OCT3, ORGANIC CATION/CARNITINE TRANSPORTER; CER, ECERIFERUM; KCS, 3-KETOACYL-COA SYNTHASE; CYP86A4, CYTOCHROME P450 FAMILY 86 PROTEIN; DCR, DEFECTIVE IN CUTICULAR RIDGES; PEL, PERMEABLE LEAVES;* GPI, phosphatidyl-inositol; GDSL, glycine, aspartic acid, serine, and leucine domain. *a:* inferred from subtilase *ABNORMAL LEAF-SHAPE 1*. *b:* inferred from *subtilase SBT1*.7. *c:* inferred from *OCT1* (71% similarity)

Rank	Gene ID	Gene Name	(Putative) Protein function or family
1	AT1G27950	*LTPG1*	Cuticle development (Lee et al., 2009b), lipid transport (DeBono et al., 2009), seed coat permeability (Edstam and Edqvist, 2014), possible defense response signaling (Maldonado et al., 2002)
2	AT2G45970	*LCR*	Fatty acid hydrolase, cuticle development (Wellesen et al., 2001)
3	AT5G51750	*SBT1.3*	Cuticle formation (Tanaka et al., 2001)[a], response to stress (Golldack et al., 2003)[b]
4	AT3G04290	*LTL1*	Overexpression increases salt tolerance (Naranjo et al., 2006),
5	AT4G29020	-	-
6	AT5G45670	-	GDSL lipase
7	AT5G45950	-	GDSL lipase
8	AT3G43720	*LTPG2*	Cuticle development, lipid transport (Kim et al., 2012); seed coat development (Edstam and Edqvist, 2014); possible defense response signaling (Maldonado et al., 2002)
9	AT4G18970	*SFAR5*	Seed fatty acid content regulation (Chen et al., 2012); GDSL lipase
10	AT5G41070	*DRB5*	MicroRNA pathway (Eamens et al., 2012)
11	AT4G00480	*MYC1*	Trichome density (Symonds et al., 2011)
12	AT5G13400	-	-
13	AT2G42840	*PDF1*	L1 layer-specific expression (Abe et al., 1999)
14	AT1G10640	-	Pectin lyase-like superfamily protein
15	AT1G64670	*BDG1*	Cuticle development, mutants have higher resistance to a fungal pathogen (Kurdyukov et al., 2006a); seed coat development (De Giorgi et al., 2015)
16	AT1G16390	*OCT3*	Lateral root development (Lelandais-Briere et al., 2007) [c]
17	AT1G25450	*KCS5/CER60*	Cuticle development, synthesis of very long chain fatty acids (Trenkamp et al., 2004)
18	AT4G16140	-	Predicted GPI-anchored protein (Borner et al., 2003)
19	AT1G01600	*CYP86A4*	Fatty acid hydrolase (Rupasinghe et al., 2007), cuticle development (Li-Beisson et al., 2009)
20	AT5G23940	*DCR/PEL3*	Cuticle development, diacylglycerol acyltransferase (Rani et al., 2010); seed coat development, extensive root branching (Panikashvili et al., 2009)

Figure 4.9. The *HTH* coexpression network. (A) The coexpression gene network. The size of a node is proportional to the correlation coefficient. Each gene is colour-coded in blue (cuticle development or lipid processing) and/or pink (stress related) according to its functions. (B) A table including genes associated with cuticle development, seed coat development, and/or stress. Genes are ranked according to the correlation coefficient. Function codes: Cuticle formation: I, Cutin/wax monomer synthesis; II, Lipid transport, III, Cutin/wax monomer/oligomer polymerization. Stress response: I, Increased expression by stress; II, Stress singling; III, Increased resistance to pathogen in the mutant. The number of + signs is proportional to the strength of evidence in the literature. Refer to Figure 4.2 for gene name abbreviations.

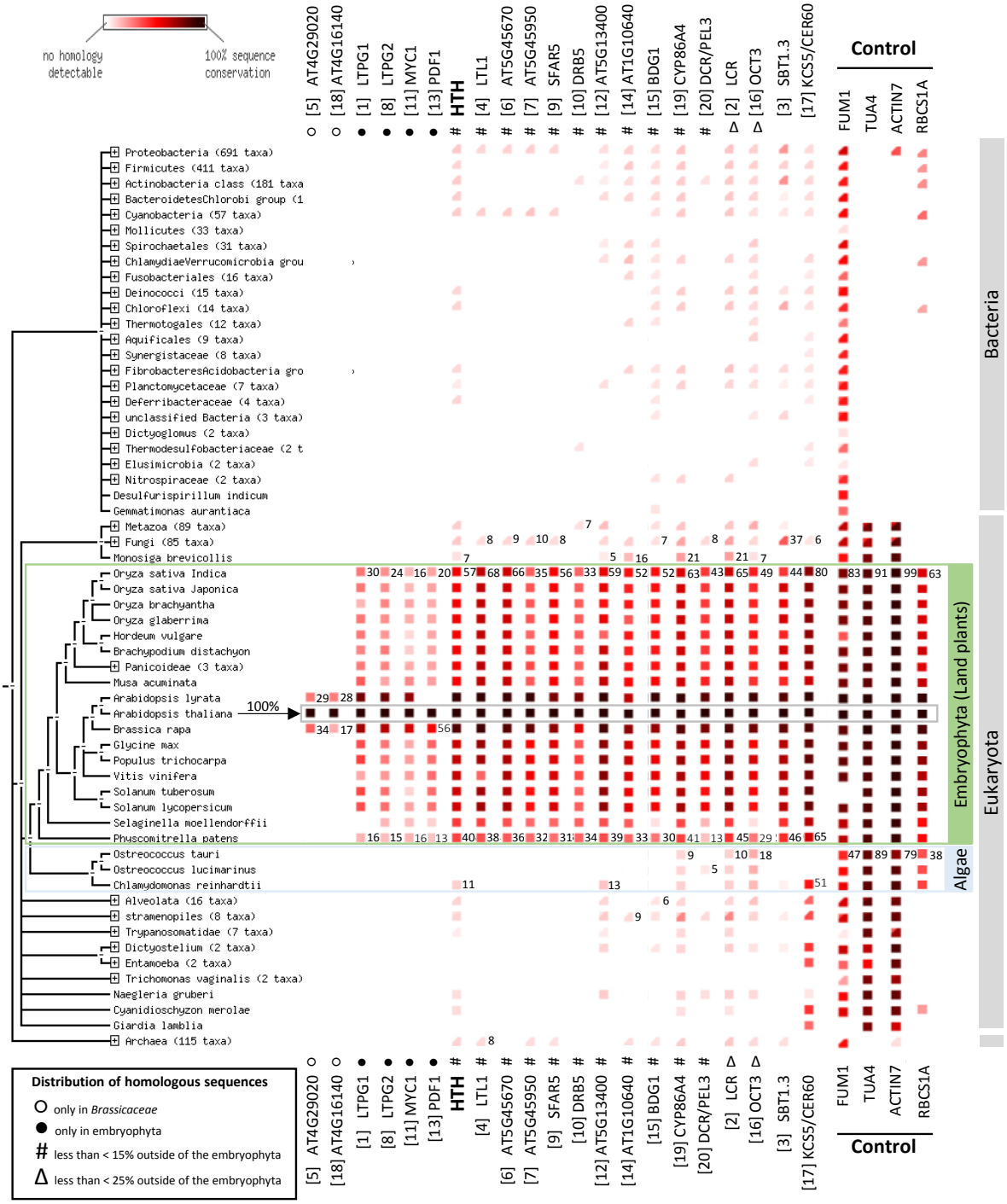
A



B

			Cuticle Development			Seed Coat		Stress Response	
			I	II	III		I	II	III
Rank	Gene ID	Gene Name							
1	AT1G27950	*LTPG1*		++		+		+	
2	AT2G45970	*LCR*	++					+	+++
3	AT5G51750	*SBT1.3*		+	+		+		
4	AT3G04290	*LTL1*					+++		
6	AT5G45670	-			+				
7	AT5G45950	-			+				
8	AT3G43720	*LTPG2*		+		+++			+
15	AT1G64670	*BDG1*			++	+++			+++
17	AT1G25450	*KCS5/CER60*	++						
18	AT4G16140	-		+					
19	AT1G01600	*CYP86A4*	+						
20	AT5G23940	*DCR/PEL3*			+++	+++			+++

Figure 4.10. Phylogenetic distribution of proteins in the HTH co-expression network. Each protein is named after the gene and is labelled with the co-expression rank shown in Table 4.3. The presence of the protein in a species is marked with a red square and absence with a white space. The intensity of the colour reflects the amount of conservation of the homologous protein in other species (100% in *Arabidopsis thaliana*). The number to the right of the intensity square indicates the protein sequence homology to the respective protein in Arabidopsis. Proteins of higher homology to HTH and its coexpressed proteins are predominantly specific to land plants, except for KCS5/CER60 and SBT1.3. More precisely, proteins are grouped according to the occurrence of their homologous proteins: only in Brassicaceae (open circle), only in land plants (closed circle; cut off: 4%), homology less than 15% (pound sign) and less than 25% (open triangle) detected outside of the land plant group. Controls include FUMARASE 1 (FUM1; AT2G47510), TUBULIN ALPHA-4 CHAIN (TUA4; AT1G04820), ACTIN7 (AT5G09810) and RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL CHAIN 1A (RBCS1A; AT1G67090). The distribution pattern was generated using the String database (www.string-db.org).


no homology detectable
100% sequence conservation
[5] AT4G29020
[18] AT4G16140
[1] LTPG1
[8] LTPG2
[11] MYC1
[13] PDF1
HTH
[4] LTL1
[6] AT5G45670
[7] AT5G45950
[9] SFAR5
[10] DRB5
[12] AT5G13400
[14] AT1G10640
[15] BDG1
[19] CYP86A4
[20] DCR/PEL3
[2] LCR
[16] OCT3
[3] SBT1.3
[17] KCS5/CER60
Control
FUM1
TUA4
ACTIN7
RBCS1A
Proteobacteria (691 taxa)
Firmicutes (411 taxa)
Actinobacteria class (181 taxa
BacteroidetesChlorobi group (1
Cyanobacteria (57 taxa)
Mollicutes (33 taxa)
Spirochaetales (31 taxa)
ChlamydiaeVerrucomicrobia grou
Fusobacteriales (16 taxa)
Deinococci (15 taxa)
Chloroflexi (14 taxa)
Thermotogales (12 taxa)
Aquificales (9 taxa)
Synergistaceae (8 taxa)
FibrobacteresAcidobacteria gro
Planctomycetaceae (7 taxa)
Deferribacteraceae (4 taxa)
unclassified Bacteria (3 taxa)
Dictyoglomus (2 taxa)
Thermodesulfobacteriaceae (2 t
Elusimicrobia (2 taxa)
Nitrospiraceae (2 taxa)
Desulfurispirillum indicum
Gemmatimonas aurantiaca
Metazoa (89 taxa)
Fungi (85 taxa)
Monosiga brevicollis
Oryza sativa Indica
Oryza sativa Japonica
Oryza brachyantha
Oryza glaberrima
Hordeum vulgare
Brachypodium distachyon
Panicoideae (3 taxa)
Musa acuminata
Arabidopsis lyrata
Arabidopsis thaliana
100%
Brassica rapa
Glycine max
Populus trichocarpa
Vitis vinifera
Solanum tuberosum
Solanum lycopersicum
Selaginella moellendorffii
Physcomitrella patens
Ostreococcus tauri
Ostreococcus lucimarinus
Chlamydomonas reinhardtii
Alveolata (16 taxa)
stramenopiles (8 taxa)
Trypanosomatidae (7 taxa)
Dictyostelium (2 taxa)
Entamoeba (2 taxa)
Trichomonas vaginalis (2 taxa)
Naegleria gruberi
Cyanidioschyzon merolae
Giardia lamblia
Archaea (115 taxa)
Bacteria
Eukaryota
Embryophyta (Land plants)
Algae
Distribution of homologous sequences
O only in Brassicaceae
● only in embryophyta
less than < 15% outside of the embryophyta
Δ less than < 25% outside of the embryophyta

4.4 Discussion

4.4.1 The predicted tertiary structure of HTH has characteristics of an enzyme

The two putative tertiary structural models shown in Figure 4.8 revealed that the majority of residues known to be important to HTH protein function based on genetic analyses are positioned within or in close proximity to the predicted catalytic pocket where putative active residues reside, corroborating the prediction that the main function of the HTH protein is enzymatic. This prediction, however, has not yet been experimentally verified and other functions mediated through protein-protein interaction cannot be ruled out. The following sections will discuss how the two previously proposed functions, i.e. mandelonitrile lyase and ω-fatty alcohol dehydrogenase, are substantiated by other lines of evidence.

4.4.2 HTH is closely related to mandelonitrile lyase

HTH shares 40% sequence identity with the FAD-binding mandelonitrile lyase PdMDL2 from almond (*Prunus dulcis*) and has two conserved GMC oxidoreductase domains (Dreveny et al., 2001; Hu and Poulton, 1997). In addition, HTH and PdMDL2 share considerable structural similarity (Appendix L) and close phylogenetic relationships (Figure 4.2), raising the possibility that they may have a similar function. Furthermore, six of the eight functional amino acid residues of HTH are conserved in four HTH-like proteins (AtGMC1 - AtGMC4) PdMDL2 (Table 4.2), suggesting functional similarity. However, this conservation also occurs among other GMC oxidoreductases that are functionally distinct from MDL though to a leser extent (Appendix J; three to five out of the eight residues). Also, it is to be noted that six of the eight functional residues identified are glycine; this bias may be an artifact of EMS

(ethyl methanesulfonate) mutagenesis, the method used to generate the original *hth* organ fusion mutants (Lolle et al., 1998; Perry et al., 2009). As such, other functional residues may not have been identified using this approach.

PdMDL2 is a hydroxynitrile lyase that catalyzes the dissociation of α-hydroxynitriles to HCN and aldehydes or ketones (Figure 1.12) (Dreveny et al., 2001; Sharma et al., 2005; Yemm and Poulton, 1986). Its substrate hydroxynitriles can be derived from the glucosinolate metabolism pathway (Frisch and Moller, 2012) where hydrolysis of glucosinolates by myrosinase, a type of β-glucosidase, gives rise to a variety of derivatives (Figure 1.7). Although isothiocyanates are the most common product at neutral pH, nitrile products are favoured when Fe^{2+} concentration or acidity is elevated (<pH 5.0) (Brabban and Edwards, 1995; Kissen and Bones, 2009; Nakano et al., 2014). Some hydroxylated nitrile products can subsequently be catalyzed by a hydroxynitrile lyase to form hydrogen cyanide, and this type of cyanogenesis is a common defense strategy employed in higher plants, particularly the *Brassicaceae*. For instance, PYK10, a β-glucosidase, is the major protein component of the stress-inducible ER body (Matsushima et al., 2003b), and therefore HTH localization to this organelle (see Chapter 3) raises the possibility of HTH participating in the cyanogenesis defense response via glucosinolate catabolism.

4.4.3 **Prokaryotically derived HTH showed neither mandelonitrile lyase activity nor alcohol dehydrogenase activity**

To experimentally determine whether HTH has a hydroxynitrile lyase activity, a recombinant MBP-HTH protein was used to conduct enzymatic assays. The putative HTH signal peptide domain was excluded from the pMAL-c4x:*HTH* construct because inclusion

of the signal peptide has been previously shown to impede enzyme activity. For example, Padham et al. (2007) showed that a MBP-tagged triacylglyceride lipase exhibited higher activity when the putative transit peptide was excluded from the recombinant protein. Affinity purification using a MBP tag was selected because of its demonstrated ability to increase the solubility of over-expressed eukaryotic fusion proteins in bacteria. Kapust and Waugh (1999), for example, showed that MBP can promote proper folding of the attached protein and was to a great extent the most effective solubilizing agent in comparison with glutathione S-transferase (GST) and thioredoxin (TRX). As large affinity tags such as MBP may interfere with the protein function and structure (Bucher et al., 2002; Smyth et al., 2003; Terpe, 2003), the N-terminal MBP tag was cleaved from the HTH protein prior to the hydroxynitrile lyase assay. Immunoblotting analysis (Figure 4.3) verified that the recombinant protein MBP-HTH was successfully produced in *E. coli*, and the MBP tag was effectively cleaved although non-specific cleavage or proteolysis at the C-terminus might have occurred as a ~70-80 kDa product was detected by both the anti-HTH and anti-MBP antibodies.

Results from the *in vitro* lyase enzymatic assays (Figure 4.4A) showed that the prokaryotically expressed HTH protein had negligible hydroxynitrile lyase activity, whereas the almond mandelonitrile lyase showed a specific activity of around 200 nmol/mg protein/min with minimal inhibition by the column buffer. Because the HTH protein itself only accounted for roughly half of the total protein in the mixture (MBP+HTH), twice the amount of protein (6 µg) was used relative to the commercial MDL samples. An assay using a triple amount of cleaved recombinant protein (18 µg) was also performed, to compensate

for possible confounding effects of non-specifically proteolysis (Figure 4.4B) but detectable activity remained negligible.

Similar to the mandelonitrile lyase assay, no alcohol dehydrogenase activity was detected using the cleaved recombinant MBP-HTH protein. Three substrates were tested and included ethanol (2 carbon), aliphatic alcohol 1-hexanol (6 carbon), and an aromatic alcohol benzyl alcohol. Benzyl alcohol was chosen as a substrate because benzyl alcohol has a similar chemical structure to the substrate of MDL, mandelonitrile. Both compounds consist of a benzene ring and a single hydroxymethyl substituent, but mandelonitrile bears the additional functional C≡N group (Figure 4.5C and D). Given the similar tertiary structures yet different catalytic residues of PdMDL2 and HTH (Figure 4.8), benzyl alcohol might fit into the pocket and hence interact with HTH protein.

Based on fatty acid metabolite profiling, HTH as a putative alcohol dehydrogenase likely catalyzes long-chain (> 16 carbon) fatty alcohols into oxo acids (Kurdyukov et al., 2006b). Therefore, if HTH has strict specificity for the alcohol carbon length, its enzymatic activity wouldn't be detected in the *in vitro* assays conducted in this study. This substrate specificity was in fact observed for the yeast alcohol dehydrogenase that was used as a positive control. The yeast ADH reacted with ethanol at a specific activity of 216 ± 52 nmol/mg/min. When the carbon number of the substrate increased to six, its activity decreased to 0.60 ± 0.23 nmol/mg/min.

4.4.4 **HTH-FP is glycosylated in plants**

Protein glycosylation has multiple functions; for instance, in the ER glycosylation is a way to control protein quality and to increase protein stability (Lodish et al., 2000; Rayon et al., 1998; Strasser, 2014). In the absence of glycans many proteins misfold and even aggregate (Parodi, 2000; Shental-Bechor and Levy, 2008). To determine whether the HTH protein is glycosylated in Arabidopsis, protein extracted from HTH_{pro}*:HTH-EYFP* plants was treated with glycosidases and analyzed for mobility shifts. The result showed that HTH-FP size was reduced by 11 kDa following treatments. This size shift suggested the recombinant HTH-FP is glycosylated, and hence the native HTH protein is likely a glycoprotein. Moreover, HTH-FP is glycosylated with mostly, if not exclusively, N-linked rather than O-linked glycans (Figure 4.6). Since N-linked glycosylation predominantly occurs in the endoplasmic reticulum (ER) and O-linked glycosylation in the Golgi apparatus, this result corroborates HTH-FP localization in organelles that are directly derived from the ER (see Chapter 3).

The N-linked glycosylation process involves glycans binding to the amino group of asparagine by oligosaccharide transferases. Hence, amino acid substitution can lead to altered glycosylation state and disrupt protein function, resulting in abnormal phenotypes. The relative positions of predicted glycosylation sites of the HTH protein and eight amino acids essential to protein function were shown in Figure 4.1. These amino acids were altered by single point mutations induced by EMS and were identified as functional residues because these substitutions resulted in floral fusion phenotype (Lolle et al., 1998). As the figure shows, none of these known functional residues are asparagine and therefore cannot be N-linked glycosylated, suggesting that the mutant phenotype was not caused by changes in

protein glycosylation. This result, however, does not exclude the likelihood that glycosylation is important for HTH protein function because glycan-binding residues might yet to be identified with more screening with different methods such as site-directed mutagenesis. Furthermore, if the lack of accurate glycosylation of HTH is lethal to plants, no functional residues accountable for glycosylation could have been identified.

Whether the absence of glycosylation affects HTH folding and hence its enzymatic activity remains to be determined. Strategies to bypass this issue include using eukaryotic expression systems for *in vitro* assays, conducting complementation assays in yeast fatty alcohol dehydrogenase mutants, or even coexpressing genes encoding enzymes that catalyze required glycosylation reactions in prokaryotic systems (Geisse et al., 1996; Khow and Suntrarachun, 2012; Laage and Langosch, 2001).

4.4.5 **Putative catalytic residues of HTH are similar to those of fatty acid dehydrogenases**

With a view towards understaning HTH's biochemical function, catalytic sites of these related GMC oxidoreductases were predicted using bioinformatics tools. Although close phylogenetic relationships often indicate similar protein functions, closely related proteins can evolve different protein functions through the substitution of a few key residues such as those at catalytic sites (Greenhagen et al., 2006), while otherwise showing high sequence or conformational similarity. For instance, two crotonase family proteins, enoyl-CoA hydratase and 4-chlorobenzoyl-CoA dehalogenase catalyze very different reactions regardless of their significant similarity in sequence and structure (Murzin, 1998). Enzyme active sites,

however, are the parts of an enzyme that directly interact with a substrate, and for that reason they are often highly conserved in functionally related proteins (Torrance et al., 2005) making these sites useful to identifying putative enzymatic function. To investigate the possible enzymatic function, putative catalytic sites of HTH as well as other GMC oxidoreductases were identified using multiple sequence alignment with PdMDL2.

Putative GMC oxidoreductases catalytic site compositions uncovered by full length protein sequence alignment further supports the phylogenetic relationship among them (Figure 4.2 and Figure 4.7). That is, catalytic residues are mostly conserved within each phylogenetic group, suggesting the closely related proteins within each group might have similar functions. For instance, in group MDL, alignment indicates high conservation of putative catalytic site residues, consistent with the strict conservation of Tyr-484 and His-524 observed among the majority of FAD-dependent hydroxynitrile lyases (Dreveny et al., 2002). Similarly, identified putative active sites are highly conserved in both CsFAO and AtFAO that have similar functions. The enzymatic activities of three proteins, including CtFAOT, CcFAO1 and AtFAO3, in these two groups have been experimentally verified. Yeast CtFAOT and CcFAO1 in *Candida* species are capable of catalyzing oxidation of C12 and C14 alkan-1-ols as long chain fatty alcohols (Dickinson and Wadforth, 1992; Vanhanen et al., 2000). Arabidopsis AtFAO3 is a homologue of CcFAO1, and purified AtFAO3 overexpressed in *E. coli* showed long chain fatty acid oxidase activity for 1-dodecanol (C12), 1-hexadecanol (C16), and 16-hexadecandiol (C16) (Cheng et al., 2004). The common function of AtFAO3, CcFAO1 and CtFAOT from two different clades and their nearly

invariant (putative) catalytic residues suggest that the active sites identified by sequence alignment are related to protein function.

Furthermore, the catalytic sites predicted by multiple sequence alignment match with the crystallography-verified active sites in the case of the cholesterol oxidase SsCHOX, pyranose dehydrogenase AmPDH, aryl alcohol oxidase PeAAO and glucose oxidase AnGOX (Figure 4.7, residues in bold). These results illustrate the usefulness of identifying putative catalytic sites using sequence alignment strategies. Therefore, if HTH functions as a mandelonitrile lyase, active site residues in PdMDL2 and HTH should show relatively high levels of conservation.

The results shown in Figure 4.7, however, indicate that important putative catalytic residues identified in HTH deviate from those of the MDLs. For PdMDL2, the substrate binding position is close to the FAD cofactor, with interactions occurring between the hydroxyl group of the substrate and the side chains of residues Tyr-484, His-486, and His-524. Cys-355 was previously thought to directly interact with the substrate (Dreveny et al., 2001), but since isoleucine or valine can also be present at this position, this possibility was later ruled out (Dreveny et al., 2009). Therefore, Cys-355 was not included in the analysis. Based on a proposed mechanism of cyanohydrin cleavage, His-524 likely acts as the general base, and Tyr-484 probably acts as a hydrogen bond donor to the mandelonitrile-OH (Dreveny et al., 2009), while the strictly conserved His-486 is involved in interaction with the cleaved cyanide product. In HTH the nonpolar aliphatic Ile-527 is found in the position corresponding to the aromatic amino acid Tyr-484 in PdMDL2, a substitution that would

eliminate the OH side group that is thought to form a hydrogen bond with the substrate mandelonitrile, making it less likely that HTH functions as a mandelonitrile lyase.

His-486 and His-524 are strictly conserved in most known FAD-dependent hydroxynitrile lyases, a finding supported by site-directed mutagenesis showing that both residues are essential for the cleavage reaction. When His-524 (the third active site position) is substituted with an asparagine (Asn), the mutant protein showed less than 5% activity compared to wildtype, indicating that His-524 is essential for mandelonitrile lyase to cleave the cyanide (Dreveny et al., 2009). Accordingly, the fact that HTH has an Asn-567 in the position corresponding to His-524 in PdMDL2 strongly suggests that HTH does not have a lyase activity. The genetic identification of three functional residues (Pro-564; Gly-565, Thr-566; Figure 4.1) surrounding HTH Asn-567 lends further support to the importance of this residue because amino acid substitutions near a critical catalytic residue are likely to hinder substrate interaction and catalytic function.

Interestingly, based on active site predictions, HTH and AtGMC1 to AtGMC4 share greater commonality with enzymes in Group ADH. Predicted HTH active site residues (Ile-527, His-529, and Asn-567) share identity with four alcohol dehydrogenases including PoADH. PoADH functions as a fatty alcohol dehydrogenase that converts aliphatic medium chain alcohols (C6-12) into aldehydes. This enzymatic activity has been confirmed by complementing *Pseudomonas putida* alcohol dehydrogenase mutants for growth on alkanes (van Beilen et al., 1992). This active site composition match to alcohol dehydrogenases suggests that HTH might have a similar function. It also corroborates the proposed ω-fatty alcohol dehydrogenase function (Kurdyukov et al., 2006b) wherein a fatty acid metabolite

analysis of *hth-12* mutant plants revealed that the mutants were defective in the biosynthesis of major long α,ω-dicarboxylic fatty acids, the predominant type of cutin monomers in Arabidopsis.

In terms of catalytic site prediction, HTH's higher similarity to ADHs instead of the more closely related MDLs raises the question about the evolutionary relationships between these proteins. The fact that the FAD cofactor does not participate in the catalysis of mandelonitrile as an electron donor or acceptor suggests that the cofactor is an evolutionary remnant from an alcohol-oxidizing ancestor (Dreveny et al., 2001). Hence, it is possible that PdMDL2 has evolved a new enzymatic activity, while the ancestral function is retained in HTH. Together with the match of active site composition of HTH and ADHs (alcohol-oxidizing enzymes), this proposed alcohol-oxidizing catalytic activity of an ancestor supports the notion that HTH retains the activity of the common alcohol-oxidizing ancestor of these three groups. That is to say, MDLs, HTH and ADHs share an ancestor, and the ancestral alcohol oxidase/dehydrogenase function is conserved in HTH and ADHs rather than in the MDL clade where a lyase function evolved specifically within.

A postulated ω-fatty alcohol dehydrogenase activity of HTH is also consistent with ER-localization demonstrated in Chapter 3 as many enzymes involved in fatty acid monomer synthesis are known to reside in the ER. ER-localized enzymes participating in the pathway of wax and cutin monomer synthesis include, for example, long chain acyl-CoA synthetase (LACS) family proteins that catalyze the synthesis of ω-hydroxy fatty acyl-CoA intermediates (Pulsifer et al., 2012; Zhao et al., 2010) and cytochrome P450 (CYP)

CYP86A1 and CYPB1 that catalyze hydroxylation of fatty acids (Compagnon et al., 2009; Kandel et al., 2006; Pinot and Beisson, 2011). Another CYP450 member, *LCR* has been proposed to catalyze the hydroxylation step (Wellesen et al., 2001; Duan et al., 2005), followed by the dehydrogenation step putatively carried out by HTH. In this scenario, HTH would function as a ω-alcohol dehydrogenase that converts the ω-fatty alcohol substrate into a ω-aldehyde product, the precursor of α,ω-dicarboxylic fatty acid monomers (Figure 1.11). The ER localization of many fatty acid cutin monomer processing steps (Figure 1.3) and the finding that HTH localizes to this organelle, makes a stronger case for *HTH* encoding a ω-fatty alcohol dehydrogenase involved in cutin monomer biosynthesis.

Genes coexpressed with *HTH* are involved in both lipid processing and stress response

Genes coding for components of a biosynthetic or response pathway are likely to have similar expression patterns (Eisen et al., 1998; Spellman et al., 1998). For example, genes that encode enzymes involved in fatty acid biosynthetic pathways have been shown to coexpress (Williams and Bowles, 2004). Furthermore, coexpression profiles can be informative about regulatory systems, as in some cases coexpressed genes are co-regulated by the same elements (Allocco et al., 2004). For *HTH*, many of the highly coexpressed genes fall into three categories involved in 1) cuticle formation, 2) seed coat development and 3) stress responses (Figure 4.9). Details about these genes are discussed below in this order.

4.4.5.1 **Cuticle development**

Genes associated with cuticle development were further categorized into three groups according to their (potential) functions: cutin/wax biosynthesis (*LCR*, *CYP86A4* and *KCS5*),

fatty acid monomer transport (*LTPG1*, *LTPG2*, *AT4G16140*) and monomer polymerization (*DCR*, *BDG*, *AT5G45670* and *AT5G45950*). Each group is discussed below.

Cutin/wax monomer synthesis

The common constituents of cutin polymers are monomers such as C16 and C18 unsubstituted fatty acids, ω-hydroxy fatty acids and dicarboxylic fatty acids; some of these monomers might contain mid-chain functional groups (Table 1.1). Cutin monomers are synthesized from fatty acids made in plastids, and these molecules are either exported and integrated into membranes and other cellular components, or further elongated into C16 or C18 fatty acids that can be further processed into cutin building blocks. Plastid-derived fatty acids in turn are transported into the ER (Wang and Benning, 2012). In the ER the acyl chain is first activated by long chain acyl-CoA synthase (LACS) which is then hydroxylated by fatty acyl ω-hydroxylase (FAH). The following step by ω-hydroxy fatty acyl dehydrogenase (HFADH) transforms hydroxy fatty acids to oxo products which are processed into dicarboxylic fatty acids by ω-oxo fatty acyl dehydrogenase (OFADH). Alternatively, hydroxyl acids can also be modified directly by FAH to give rise to dicarboxylic acids.

The *LCR* gene encodes cytochrome P450 protein CYP86A8, a ω-hydroxylase that catalyzes ω-hydroxylation of fatty acids ranging from C12 to C18:1 (Wellesen et al., 2001). Its encoded protein is ER-localized and is likely involved in cutin monomer synthesis (Li-Beisson et al., 2013; Pollard et al., 2008; Tang et al., 2007b; Wellesen et al., 2001). LCR and HTH are postulated to catalyze consecutive steps in the cutin monomer biosynthetic pathway (Figure 1.11); LCR is the ω-hydroxylase that converts fatty acids into hydroxyl acids, which

are subsequently turned into oxo-acids by ω-alcohol dehydrogenase, an enzymatic function proposed for HTH (Kurdyukov et al., 2006b). It should be pointed out that among coexpressed genes is the closely related cytochrome P450 protein CYP86A4 whose ω–hydroxylase activity was confirmed by *in vitro* assays (Rupasinghe et al., 2007). Its expression is regulated by a cutin biosynthesis transcription factor *WAX INDUCER 1* (*WIN1*), and the fatty acid profile of *cyp86a4* mutant plants showed a 50% reduction in cutin monomers such as 16-hydroxypalmitate, 10,16-dihydroxypalmitate, and 1,16-hexadecanedioic acid (Kannangara et al., 2007; Li-Beisson et al., 2009). Furthermore, a *HTH*-like gene AtGMC4 (AT1G12570; UniProt: Q66GI5; Figure 4.7) is coexpressed with CYP86A7, another putative ω-hydroxylase that catalyzes the same reaction as LCR. The coexpression of *HTH* or *HTH*-like genes with ω-hydroxylases further supports HTH's role in the cutin biosynthetic pathway.

Very long chain fatty acids (VLCFAs; longer than 18 carbons) are the basic building components for cuticular waxes. Long chain fatty acids are elongated to form VLCFAs by an ER-localized fatty acid elongation complex (FAE) where ketoacyl-CoA synthases (KCSs) condense the acyl-CoA with malonyl-CoA and produce a β-ketoacyl-CoA to initiate the elongation cycle. Twenty one KCSs have been identified in Arabidopsis (Haslam and Kunst, 2013). The coexpressed KCS5/CER60 is an ER-localized ketoacyl-CoA synthase. It is highly homologous to KCS6/CER6 (89% protein sequence identity under 100% coverage) whose mutations lead to reduction in long-chain lipids in the pollen coat and on the stem surface; moreover, complementation studies showed restored fertility and stem cuticle phenotype (Fiebig et al., 2000). When KCS5 was expressed in yeast, it was enzymatically active and

catalyzed endogenous yeast VLCFA elongation by cooperating with the yeast elongase complexes (Trenkamp et al., 2004).

Lipid transport

Synthesized cutin monomers or building blocks are first exported from the ER to the cell wall and then subsequently to the cuticle. Possible mechanisms and putative cellular locations of cutin assembly are shown in Figure 1.4 (Pollard et al., 2008). One possible pathway is thought to be cytoplasmic, requiring soluble carrier proteins to move cutin monomers or small oligomers in concert with an ABC transporter and/or glycosylphosphatidyl-inositol (GPI)-anchored lipid transfer protein (LTPG) (DeBono et al., 2009).

Proteins encoded by *LTPG1*, *LTPG2* and *AT4G16140* belong to the class of glycosylphosphatidylinositol (GPI)-anchored lipid transfer proteins (LTP). *LTPG1* is expressed in the epidermis and is primarily localized to the plasma membrane (transmembrane protein) but is also found in the extracellular matrix (DeBono et al., 2009). Its lipid binding capacity was experimentally verified using *E. coli*-expressed LTPG1 in combination with the fluorescent lipophilic probe 2-p-toluidinonaphthalene-6-sulfonate (TNS). LTPG1 is thought to be a carrier of cutin and wax constituents to the plant surface. Mutant *ltpg1* plants showed a great reduction (>50%) in the C29 alkane, a major component of cuticular waxes of the stems and siliques, and this defect in alkane load can be rescued by native promoter–driven *LTPG1* expression (DeBono et al., 2009; Lee et al., 2009b).

LTPG1 has the highest expression in regions of rapid expansion growth, such as inflorescence stems. Spatial expression, as determined using *LTPG1* promoter*::*GUS transgenic plants, showed a pattern similar to HTH-FP (see Chapter 3); that is, expression was detected in young seedlings, emerging lateral roots, the seed coat, and seedling vasculature. *LTPG2* was identified later, and was found to be functionally redundant with *LTPG1*. *ltpg2* mutants showed reduced load of wax, particularly the C29 alkane and have an expression pattern that overlaps with *LTPG1* (Kim et al., 2012). *AT4G16140* is an uncharacterized proline-rich family protein with a 23-aa long signal peptide.

Cutin/wax monomer/oligomer polymerization

Cutin monomers polymerize to form bigger oligomers in the cell or branching network in the extracellular space. With different monomer compositions, larger branching or cross-linked cutin structures can be achieved, and the different size and tertiary structure of these cutin polymers determines the cuticle properties such as its permeability.

BDG1 is required for normal cuticle formation and encodes an extracellular protein with an α/β hydrolase domain. It is expressed exclusively in epidermal cells and is localized to the outermost cell wall of the epidermis. The *bdg1* loss-of-function mutants display increased leaf surface permeability, have a thinner cuticular membrane with pockets in the cuticular zone and are devoid of an intervening cuticular wax layer between fused leaves, despite increased cutin and wax loads (Kurdyukov et al., 2006a). The *bdg1* phenotype is similar to that of the fungal cutinase-expressing transgenic plants (Sieber et al., 2000), suggesting that

BDG1 may not be directly involved in cutin monomer/oligomer synthesis. Rather BDG1 may function as a polyester synthase for cutin polymer formation (Pollard et al., 2008).

DCR encodes a BAHD acyltransferase required for monomer incorporation into the cuticular polymeric structure. Mutant *dcr* plants exhibit postgenital organ fusions, have a significant reduction of a major cutin monomer and manifest excessive root branching. These mutants show altered fatty acid profiles such that a major flower-specific cutin monomer, 9(10),16-dihydroxy-hexadecanoic acid, is reduced to near undetectable levels, while C16 dicarboxylic fatty acid 1,16-hexadecanedioic acid levels are greatly elevated. Unlike BDG1, DCR is localized to the cytoplasm and is thought to be involved in acyl transfer of cutin monomers leading to formation of precursor intermediates or oligomeric structures (Panikashvili et al., 2009). Interestingly, both *dcr* and *bdg1* mutants can give rise to misshapen trichomes (Marks et al., 2009; Panikashvili et al., 2009).

The GDSL motif lipase family proteins have consensus amino acid sequence of glycine, aspartic acid, serine, and leucine around the active site, and they exhibit diverse functions with broad substrate specificities (Akoh et al., 2004). In plants, they are thought to play a role in cuticle biosynthesis (Irshad et al., 2008; Matas et al., 2011). Coexpressed gene *AT5G45670* and *AT5G45950* belong to the GDSL-motif esterase/acyltransferase/lipase family that share a SGNH (serine-glycine-asparagine-histidine) domain. The *Agave americana* (AgaSGNH) protein has been characterized as a plant SGNH-motif hydrolase; like the α/β hydrolase BDG1, it is localized to the epidermis outer cell wall, and *AgaSGNH* is mostly expressed in regions where cutin biosynthesis is active, such as rapidly expanding

leaves (Reina et al., 2007). The SGNH-motif protein encoded by *AT5G45670* and *AT5G45950* might also have similar functions. In fact, putative proteins encoded by these genes have notable homology (up to 53%) to a tomato extracellular acyltransferase *TOMATO CUTIN DEFICIENT 1* (*CD1*), also a GDSL-like lipase family protein. The CD1 protein has been shown to have polyester synthesis activity and is required for cutin accumulation *in vivo* (Yeats et al., 2012).

Others

PROTODERMAL FACTOR 1 (*PDF1*) is thought to be involved in cuticle development because *PDF1* expression is exclusive to the L1 layer of vegetative, inflorescence and floral meristems and to the protoderm of organ primordia. Yet, its biochemical function and whether it takes part in cuticle development is yet to be determined. *SUBTILASE 1.3* (*SBT1.3*) encodes a protein that has serine-type endopeptidase activity, but the biological function of SBT1.3 is unclear. Based on protein sequence similarity, SBT1.3 is related to *ALE1* (*ABNORMAL LEAF SHAPE 1*; 44% identity) that plays a role in non-cell autonomous peptide signaling and, in embryos and juvenile tissues in Arabidopsis, is required for cuticle formation and epidermal differentiation (Tanaka et al., 2001).

4.4.5.2 **Seed coat development**

Cuticle formation and seed coat development show genetic overlap such that some mutants with a cuticle phenotype also have a seed coat phenotype. In a study where nine *LTPG1-6* T-DNA insertion lines were examined, mutant *ltpg2*-6 seeds showed elevated permeability to tetrazolium salts. Lipid analysis of *ltpg6* mutant lines revealed an increase of C20:0, C22:0 and C24:0 and a decrease in ω-hydroxy fatty acids, suggesting altered suberin/cutin

deposition in the seed coat. Analyses using scanning electron microscopy revealed seed-specific morphological changes with hair-like outgrowths in *ltpg4* and *ltpg5* seeds (Edstam and Edqvist, 2014). In cotton (*Gossypium hirsutum*), LTPs have been implicated in fiber development through cutin deposition (Orford and Timmis, 2000). In *dcr* mutants, the seed coat was more porous than the Columbia wildtype as determined by toluidine blue staining. Mutant *dcr* seeds were often deformed, occasionally fused to one another and following imbibition, failed to release mucilage (Panikashvili et al., 2009). The *bdg1* mutant also showed increased toluidine blue permeability resulting in endosperm staining (De Giorgi et al., 2015).

The notion of a "seed cuticle" emerged from the work above (De Giorgi et al., 2015; Panikashvili et al., 2009), wherein a cutin-containing layer surrounds the entire outer side of the endosperm, in effect encasing all living seed tissues. Likely as a result of higher permeability, the *bdg1* mutant seeds suffer low seed viability and dormancy. Furthermore, *dcr* and *bdg1* seed were more sensitive to salinity, osmotic, and water deprivation stress conditions, possibly as a consequence of the elevated level of oxidative stress conferred by increased porosity of the seed coat. The reported association of these cutin biosynthesis or polymerization genes with seed coat development is consistent with the reported increase in seed coat permeability of *hth* mutants described in Chapter 3.

4.4.5.3 **Stress response**

Research has shown that cuticle defects can lead to increased susceptibility to pathogens or abiotic stress. For example, *ltpg1* mutant plants that exhibit disorganized and diffuse cuticle

showed lower resistance to the necrotropic fungal pathogen *Alternaria brassicicola* (Lee et al., 2009b). This type of change in interaction with stress/pathogens is likely due to lack of protection provided by a normal cuticle barrier. Genes whose mutations caused cuticle defects and lowered resistance to stress, however, are not included in the group of stress/defense associated genes in Figure 4.9. Rather, included genes must show either 1) increased expression by stress, 2) increased resistance to pathogen in the mutant or 3) potential role in stress signaling. Genes included in this category are discussed below.

LTL1 encodes a GDSL-like lipase family protein that upon exposure to LiCl or NaCl is rapidly induced and can confer higher salt tolerance if overexpressed (Naranjo et al., 2006). Moreover, its gene expression can be activated by salicylic acid, a known mediator in the response to pathogen attack, suggesting a role of LTL1 in plant defense responses. Subtilases (SBTs) constitute a large family of serine peptidases with diverse functions. Subtilase P69 from *Solanum lycopersicum* was the first member of this family identified and like its related subtilases P69A and P69B, has been shown to play a role in responses to pathogens (Granell et al., 1987; Jorda et al., 1999; Tornero et al., 1997). In Arabidopsis thaliana, subtilases comprise six distinct families AtSBT1 to AtSBT6 (Rupasinghe et al., 2007; Schaller et al., 2012). The function of Arabidopsis *SBT1.3* is unknown, but its encoded protein is categorized into the AtSBT1 group and is closely related to SBT1.7 (also known as ARA12 or AtSLP1) (53% protein sequence identify; At1G04110) (Rautengarten et al., 2005; Schaller et al., 2012). The expression of *SBT1.7* and two other subtilases genes *AtSLP2* and *AtSLP3* can be altered by environmental stress and were shown to be elevated by jasmonate treatment in juvenile plants. Although the specific function is yet to be determined, research on closely

related subtilases have provided evidence of a possible association of Arabidopsis SBT1.3 with plant response to environmental stress.

Mutants such as *bdg1*, *lcr* and *lacs2* (*long-chain acyl-CoA synthetase 2*) have also been shown to increase the resistance to a virulent fungal pathogen *Botrytis cinerea* (Kurdyukov et al., 2006a; Tang et al., 2007b; Wellesen et al., 2001). Likewise, transgenic plants expressing a fungal cutinase gene also showed elevated resistance to *B. cinerea* (Chassot et al., 2007; Sieber et al., 2000). These results demonstrate that the cuticle not only serves as a protective barrier but also as a component of the defense response signaling cascade that involves many components.

In a study of *lcr*, *fdh* and *bdg1* mutants, Voisin et al. (2009) proposed that the mutants alleviate the functional disorder of the cuticle by reinforcing different cell integrity pathways. Using an *in silico* screening method, the authors identified a gene that encodes a protein involved in small-RNA signaling, *SERRATE* (*SE*), that is essential for the elevated resistance of *lcr* and *bdg1* mutants. In the *se* mutant background increased resistance to *B. cinerea* exhibited by *lcr*, *fdh* and *bdg1* mutants is lost. The interconnection of a micro-RNA associated protein with cutin synthesis genes for disease resistance raises the possibility the coexpressed gene *DRB5* which is associated with microRNA pathway might interact with HTH in a similar manner. Moreover, evidence has shown that a permeable cuticle is associated with the production of reactive oxygen species (ROS) (L'Haridon et al., 2011). It has been suggested that higher cuticle permeability allows early sensing and response to *B. cinerea* by the host, resulting in greater resistance to this pathogen (Reina-Pinto and

Yephremov, 2009). ROS production could be part of the early pathogen responses triggered by rapid penetration of fungus-secreted elicitors as ROS can potentially activate a cascade of stress-associated pathways (Schmitt et al., 2014). This notion is supported by studies showing that plant leaves treated with cutinase prior to the inoculation accumulated more ROS and also exhibited greater resistance to *B. cinerea* (Kauss et al., 1999; L'Haridon et al., 2011).

In addition to aforementioned mechanisms, proteins encoded by genes involved in cutin production have been reported to play a role in plant defense signaling of plant-pathogen interactions. CYP86A2 encoded by *ATT1* (*aberrant induction of type three genes 1*) is a cytochrome P450 protein that catalyzes fatty acid hydroxylation (Bak et al., 2011). The cutin content is reduced to 30% in *att1*, indicating that CYP86A2 plays a major role in cuticle formation. *att1* has a diffuse cuticle of elevated permeability and a higher transpiration rate. In addition to these phenotypes, it has been demonstrated that the *att* mutation represses the expression of a *Pseudomonas syringae* gene essential for its virulence (Xiao et al., 2004). As many gram-negative bacterial pathogens, *P. syringae* employs the type III secretion system to deliver effector proteins into the host to initiate infection. When plants were incubated with *P. syringae*, the bacterial type III gene *avrPto* expression in the intercellular space was higher for *att1* than for wildtype plants, suggesting the type III gene was negatively regulated in *ATT1*. Xiao et al. (2004) proposed that cutin monomers or other lipids derived from cutin monomers encoded by *ATT1* may repress type III gene expression. This gene repression cannot be solely attributed to the diffuse cuticle structure because this enhanced *avrPto* expression was not observed in *wax2* that exhibits thick but translucent cuticle membrane

and postgenital fusion (Chen et al., 2003). Similarly, Lee et al. (2009b) reported that *ltpg1* knockout mutant had increased vulnerability to the fungal pathogen *Alternaria brassicicola*, and suggested that GPI-anchored lipid transfer LTPG1 might also have functions in lipid signaling for plant defense against fungal pathogen attack as many lipid transfer proteins are involved in long-distance signaling during acquisition of systemic resistance in Arabidopsis (Lee et al., 2009b; Maldonado et al., 2002).

In summary, the connection to a gene network that is associated with both functions points to a possibility that HTH might also play a role in both biological processes. The coexpression network shows that more than half of the top genes coexpressed with *HTH* are associated with lipid processing and seed coat development. The analysis also revealed that many of the cuticle development related genes are associated with plant defense or stress response.

4.4.6 *HTH* and coexpressed genes are predominantly specific to land plants and not algae

The evolution of the cutin-based layer sealing the epidermis of aerial plant organs allowed for the colonization and spread of land plants from their aquatic ancestors, the green algae, by regulating water status to enabling plant growth on land (Ligrone et al., 2012). Whereas proteins of high homology to the small subunit of the RuBisCO protein RBCS1A occur in both land plants and algae, those coexpressed with *HTH* are mainly present in land plants (embryophyta) and not in algal species (Figure 4.10). It is noteworthy that uncharacterized AT4G16140 and AT4G29020, proteins encoded by neighboring genes, are strictly specific to *Brassicaceae*, indicating that they might be newly evolved proteins in this genus. In contrast,

genes that are involved in very-long-chain fatty acid elongation such as KCS5/CER60 have homologous sequences in organisms such as protists and slime molds suggesting that these are relatively ancient and conserved proteins. Although the function of endodermis-specific PDF1 has not been elucidated, it may play a role in cuticle formation given that this gene is found only among plants.

4.5 **Conclusions**

Enzymatic assays and bioinformatics analyses were conducted to mine for evidence supporting either a hydroxynitrile lyase or ω-fatty alcohol dehydrogenase function for HTH. The apparent conservation of genetically identified functional residues between HTH and other known MDLs, the similarity in predicted structures, and phylogeny indicate that HTH and MDLs are closely related. However, comparison of putative active sites suggests that HTH might have a function distinct from MDLs. Although no definitive conclusions could be drawn from the results of *in vitro* enzymatic assays, other lines of evidence favour a fatty acid alcohol dehydrogenase function. First, HTH's putative active site residues are identical to an alcohol hydrogenase that is capable of converting medium chain fatty acid alcohols in to aldehydes, a function comparable to the one put forward by Kurdyukov et al. (2006b). Second, *HTH* is coexpressed with genes that are directly or indirectly involved in cutin monomer biosynthesis or cutin polymerization/remodeling. Third, the coexpression profile analysis indicates that the majority of proteins encoded by coexpressed genes are specific to land plants. Last, the subcellular localization to the ER network and its derived bodies (see

Chapter 3), although it does not exclude a role in cyanogenesis, favours a function involved in cutin monomer biosynthesis which occurs predominantly in the ER.

4.6 Future Research

To resolve the question of enzymatic activity, further experimentation is required. One approach is to determine the tertiary structure by protein X-ray crystallography. The derived electron-density map can be used to confirm the predicted catalytic residues that are indicative of the protein function (Wlodawer et al., 2013). However, the function suggested by catalytic sites still needs experimetnal confirmation by *in vitro* enzymatic assays. Given that the native HTH protein appears to be a glycoprotein in plants and glycosylation can be essential for protein function, the *E. coli*-based prokaryotic expression system employed in this study might be improved by coexpressing genes encoding enzymes required for glycosylation reactions (Geisse et al., 1996; Khow and Suntrarachun, 2012; Laage and Langosch, 2001). Additionally, codon bias in *E. coli* could also be taken into account to prevent translational errors. Although these measures might increase the likelihood of generating functional eukaryotic proteins in a prokaryotic system, it is optimal to use a eukaryotic organism such as yeast to produce the recombinant protein.

Production of recombinant proteins in *Pichia pastoris*, a methylotrophic yeast, is an established system for biopharmaceuticals and industrial enzymes (Looser et al., 2015). Due to its increasing relevance, various strategies and host strains have been developed, including strains engineered to achieve complex N-glycosylation (De Pourcq et al., 2010; Hamilton and

Zha, 2015; Jacobs et al., 2009). Expression in glycosylation-competent *P. pastoris* may facilitate expression of a functional recombinant HTH protein for enzymatic assays. Additionally, the tag-removed functional recombinant protein would serve as an ideal antigen to generate polyclonal antibodies against the native HTH protein. Prior to setting up the *P. pastoris* expression system, the proposed alcohol dehydrogenase activity can also be tested in a complementation assay using yeast fatty alcohol dehydrogenase mutants such as the null deletion mutant stain *saf1* (Achkor et al., 2003). The yeast *SFA1/YDL168W* encodes a class III alcohol dehydrogenase bifunctional protein that has both alcohol dehydrogenase and formaldehyde dehydrogenase activities (Wehner et al., 1993). SFA1 can act on a variety of alcohol substrates including the long chain alcohol 10-hydroxydecanoic acid and 12-hydroxydodecanoic acid (Dickinson et al., 2003).

Chapter 5 General Discussion

Arabidopsis thaliana propagates mainly by self-fertilizing, and therefore, like many crop plants, theoretically has a limited potential for producing genetically diverse offspring. Despite this, inbreeding has persisted in Arabidopsis for over a million years suggesting that some underlying adaptive mechanism such as somatic variation buffers the deleterious consequences of this reproductive strategy. In Chapter 2, we used presence-absence molecular markers to demonstrate that individual Arabidopsis plants are capable of producing somatic sectors during the course of normal vegetative development. Although genetically heterogeneous sectors have been detected in wildtype genetic backgrounds, *hothead* (*hth*) mutant plants give rise to genetically discordant somatic sectors and progeny more frequently (Hopkins et al., 2013). This finding suggests that the HTH protein might play a role in mediating genome instability.

Sequence analyses reveal that these genetically discordant sectors contained single nucleotide changes, loss of sequences and, surprisingly, acquisition of unique genomic insertions. Estimates based on quantitative analyses suggest that these sectors are very small but can have a complex genetic makeup. In ruling out more trivial explanations for these data, our findings raise the possibility that intrinsic drivers of genetic variation are responsible for the targeted sequence changes we detect. The *de novo* appearance of genomic insertions supports our original contention that cryptic sequence templates drive some of these changes (Lolle et al., 2005). This process is genome-wide, impacting all five chromosomes, whether or not the target loci reside within genes or between genes. Given the

evolutionary advantage afforded to populations with greater genetic diversity, we hypothesize that organisms that primarily self-fertilize or propagate clonally counteract the genetic cost of such reproductive strategies by leveraging a cryptic reserve of extra-genomic information.

While the connection between the *hth* mutant background and elevated frequencies of genome instability remains to be elucidated, previous genetic analyses have clearly demonstrated a role for the Arabidopsis *HTH* gene in achieving proper cuticle function as perturbation of the *HTH* gene leads to a floral organ fusion phenotype (Lolle et al., 1998). Little is certain about what type of biochemical pathways the HTH protein is involved and how its mutation would lead to changes in cuticle integrity. Two possible enzymatic functions have been proposed for HTH. In accordance with the fusion phenotype, Kurdyukov et al. (2006b) proposed that HTH is an alcohol dehydrogenase involved in the biosynthesis of cutin monomer α,ω-dicarboxylic fatty acids, and the resulted cutin monomer composition would cause perturbation in cuticular polyester structures. Alternatively, based on protein sequence analyses, Krolikowski et al. (2003) suggested that HTH may function as a mandelonitrile lyase, a hydroxynitrile lyase that catalyzes hydroxynitriles to hydrogen cyanide and aldehydes or ketones.

In Chapter 3, the localization of HTH was examined to gain insight into HTH's function. I showed the results of the first study of HTH protein localization using a fluorescent protein-tagged HTH fusion protein generated by native promoter-driven construct (HTH_{pro}*:HTH-FP)*. HTH-FP was predominantly localized to the epidermis of seedling and

mature tissues; moreover, it was also present in the ovule outer integument that is of epidermal origin. HTH-FP's presence in the L1 layer is in accordance with its mutant organ fusion phenotype and the proposed fatty alcohol dehydrogenase function associated with cutin monomer biosynthesis.

On the subcellular level, the HTH protein was observed, for the first time, to reside in a subcellular structure that likely corresponds to the so-called "ER body". ER bodies are ER-derived compartments that have characteristics distinctive from other vesicles in the secretory pathway (Hara-Nishimura and Matsushima, 2003). ER bodies are constitutively present in the epidermal cells of cotyledons, hypocotyls and roots of young Arabidopsis seedlings. Although ER bodies generally are not found in rosette leaves, their formation can be induce by mechanical wounding and other biotic/abiotic stress (Ogasawara et al., 2009). In agreement, the RT-qPCR results showed that the *HTH* expression in Ws wildtype plants was elevated by the wounding hormone methyl jasmonate (MeJA) which also induced ER bodies in epidermal cells. Together, the localization of HTH-FP in the ER bodies and its elevated expression by MeJA suggests that one function of the HTH protein may be responsive to various types of plant stress.

In Chapter 4, the two previously proposed functions were further examined by *in vitro* assays and bioinformatics analyses. Although results of the assays were inconclusive, analyses of amino acid conservation, protein structure and phylogeny indicated that HTH shares great similarity and close relationship with the mandelonitrile lyases PdMDL2 and many MDL-like proteins. However, investigation of putative active sites that are directly

involved in the catalytic reaction revealed that HTH shares the same catalytic sites of several fatty alcohol dehydrogenases, one of which was experimentally verified to be capable of converting medium chain fatty alcohols into aldehydes, a function comparable to the one put forward by Kurdyukov et al. (2006b).

For cutin monomer synthesis, plastid-derived fatty acids at first are transported from plastids into the ER where various metabolic pathways give rise to a variety of cutin monomers. After a fatty acid is imported into the ER, the acyl chain is first activated and then hydroxylated by fatty acyl ω-hydroxylase (Figure 1.3). The hydroxy fatty acids can then in turn be transformed into oxo products (aldehydes) and subsequently dicarboxylic acids by dehydrogenases. Together with other constituents, different compositions of these monomers give rise to cutin polymers of various physical and chemical properties. Kurdyukov et al. (2006b) proposed that HTH is a ω-hydroxy fatty acyl dehydrogenase that oxidizes long chain ω-hydroxy fatty acids to ω-oxo products, precursors of the cutin monomer α,ω-dicarboxylic fatty acids (Table 1.1 and Figure 1.11). Since this step of fatty acid modification occurs in the ER network, HTH localization in the ER is in agreement with the alcohol dehydrogenase enzymatic function. HTH's putative role in the cutin monomer biosynthesis was further supported by the fact that *HTH* is coexpressed with genes that are involved in cutin/wax monomer biosynthesis, polymerization and transport. In addition, the ER localization is also consistent with the existence of N-linked sugar to the protein as shown by the glycosylation analysis since this type of glycosylation predominantly occurs in the ER.

These synthesized cutin monomers or building blocks need to be exported from the ER to the cell wall and then subsequently to the cuticle where they polymerize. Specific carrier proteins, transmembrane transporters or oleophilic droplets have been known or proposed to be involved in the transport of cutin monomer/oligomers (Figure 1.4). HTH localization to ER bodies raise the possibility of ER bodies being part of the monomer exportation system. However, no fusion of ER bodies with the cytoplasmic membrane has been observed in the current study nor by other research groups. The association between ER body localization and the possible alcohol dehydrogenase activity of HTH remains to be determined.

The ER body location also provides a clue for HTH's association with stress responses. In addition to elevated expression by exposure to MeJA, bioinformatics analyses revealed that many genes of the coexpression profile are involved in stress/defense response. Even more interesting, many of the cuticle development related genes are associated with plant defense or stress response. The connection to a gene network that is associated with both functions points to a possibility that HTH might also play a role in both biological processes. This findings brings up the question of how the putative fatty acid processing function of HTH can be also associated with stress response.

Fatty acid metabolites have been shown to play a part in a wide spectrum of stress/defense related biological processes. For example, elevated production of ω-fatty alcohol dehydrogenase is known to be a stress response to wounding in potato tubers (*Solarium tuberosum* L) and is thought to promote suberin biosynthesis for wound healing (Agrawal and Kolattukudy, 1977; Agrawal and Kolattukudy, 1978a; Agrawal and

Kolattukudy, 1978b; Yang and Bernards, 2006). Furthermore, studies on altered pathogen resistance of cuticle mutants point to the role of cuticle permeability in regulating plant defense responses. The importance of cuticle integrity to plant defense has been shown in various cuticle mutants that present higher resistance to the necrotrophic fungus *Botrytis cinerea* (Bessire et al., 2007; Tang et al., 2007b). In these cases, it was proposed that higher cuticle permeability allowed early sensing and responses to *B. cinerea* by the host and eventually rendered the plants more resistant (Reina-Pinto and Yephremov, 2009). Rapid penetration of fungus-secreted elicitors can also elevate the production of reactive oxygen species (ROS) that potentially activate a cascade of stress-associated pathways (Schmitt et al., 2014). This notion is supported by studies showing that plant leaves treated with cutinase prior to pathogen inoculation accumulated more ROS and exhibited greater resistance to *B. cinerea* (Kauss et al., 1999; L'Haridon et al., 2011).

In addition to serving as structural components, cutin and wax monomers and their derivatives can function as modulators of a variety of signal transduction pathways triggered by environmental stimuli (Kandel et al., 2006; Walley et al., 2013). An example is *ATT1* (*ABERRANT INDUCTION OF TYPE THREE GENES 1*) whose mutation results in a diffuse cuticle layer with increased sensitivity to water deprivation. Fatty acid profiling has suggested a role of ATT1 in cutin monomer synthesis (Xiao et al., 2004). Moreover, ATT1 may also play a potential role in mediating the bacterial type III secretion system for delivering effector proteins into the host to initiate infection. When plants were incubated with *Pseudomonas syringae*, the bacterial type III gene *avrPto* expression in the intercellular space was higher for *att1* than for wildtype plants, suggesting that *ATT1* negatively regulates

avrPto that is important for bacterial virulence. Xiao et al. (2004) further proposed that certain cutin-related fatty acids synthesized by *ATT1/CYP86A2* may function as signal molecules that repress the expression of bacterial type III genes.

The imbalance among fatty acids has also been shown to elicit plant defense responses. For instance, elevated levels of palmitoleic acid (C16:1) in eggplants (*Solanum melongena)* resulted in improved resistance to the fungal plant pathogen *Verticillium dahlia* (Xing and Chin, 2000). Similarly, a change in equilibrium between saturated and unsaturated fatty acids was also observed in the *ssi* (*suppressor of salicylic acid-insensitive*) Arabidopsis mutant that exhibits high levels of C18:0 fatty acids and decreased levels of C18:1 fatty acids (Shah et al., 2001). The *ssi* mutant plants were resistant to the oomycete *Hyaloperonospora arabidopsidis* and a virulent bacterial strain of *Pseudomonas syringae*; this elevated resistance is likely attributed to constitutive activation of the resistance (*R*) genes in this mutant background by the change in the composition of fatty acids (Nandi et al., 2003; Rojas et al., 2014).

In addition to the possibility that HTH mediates plant defense responses through the control of cuticle permeability or fatty acid composition, evidence for serving as an enzyme involved in cutin monomer synthesis that is localized in the stress-inducible organelle (ER body) raises the question of whether HTH is a multifunctional protein that can function in different pathways (Bunz, 2008; Huberts and van der Klei, 2010). Such proteins can arise by alternative splicing, posttranslational modification, or association with other partners (Moore, 2004) and are known as multifunctional or moonlighting proteins. These proteins can act

enzymatically to recognize multiple substrates or perform independent non-enzymatic functions (Copley, 2003).

In past decades, multifunctional proteins have been found in many species including plants, animals, yeast and prokaryotes (Copley, 2012; Huberts and van der Klei, 2010). A well-known example is the eukaryotic polypeptide elongation factor EF-1. EF-1 is a major translational factor but also contributes to signal transduction, cytoskeletal organization, apoptosis, nutrition, and nuclear processes such as RNA synthesis (Ejiri, 2002). Another example of a multifunctional protein is the plant peroxisomal multifunctional protein (MFP) that catalyzes multiple steps of fatty acid β-oxidation in the peroxisome matrix. In a microtubule-binding protein fraction extracted from rice seeds, MFP was found to cross link to mRNA, and this result was confirmed by expressing histidine-tagged MFP that showed mRNA and microtubule binding activities in addition to the enzyme activity involved in the β-oxidation of fatty acids (Chuong et al., 2005). Based on these studies, MFP is thought to associate with microtubules at the periphery of the peroxisome to enrich mRNA coded for peroxisome-destined proteins.

Another example of a plant multifunctional protein is the currant tomato (*Lycopersicon esculentum*) LeCp protein, whose dual function was proposed by Matarasso et al. (2005). LeCP is an orthologue of known Arabidopsis vacuolar processing enzymes that typically act as cysteine proteases in the cytoplasm. When a small ubiquitin-like modifier binds to these proteins, they are transported to the nucleus and act as transcription factors for the gene that encodes 1-aminocyclopropane-1-caboxylic acid synthase, leading to ethylene production

(Rosin et al., 2005). The detection of a *HTH* splice variant transcript points to the possible existence of a smaller HTH isoform (Appendix H). The theoretical tertiary structure model (Appendix O) indicated that the peptide absent in the isoform is situated on the surface of the HTH protein. The protein surface provides binding sites for different ligands, allowing the protein's function to be regulated or even modified (Kristiansen, 2004). Therefore, changes in amino acids positioned on the protein surface can greatly affect protein functions by altering interactions with other elements, as aforementioned cases of MFP and LeCP. Whether this HTH peptide sequence interacts with a partner important for protein function or regulation remains to be elucidated, but the existence of an isoform with a surface peptide deletion speaks to a possibility of HTH being a moonlighting protein whose isoforms serve different functions determined by the partnering element.

The research described herein has brought new information and perspectives to the localization and possible functions of HTH. The epidermal localization and putative catalytic sites prediction point to an alcohol dehydrogenase function in cutin monomer biosynthesis. Since the modification of these monomer fatty acids predominantly occurs in the ER network, HTH's glycosylation state and localization to the ER network are also in agreement with this enzymatic function. On the other hand, detection of HTH-FP in ER bodies and expression elevated by MeJA associate HTH with stress responses. In accordance, coexpression profile analyses indicate that *HTH* is coexpressed with genes involved in both lipid processing and stress/defense response. Further work needs to be done to verify whether HTH has a fatty alcohol dehydrogenase activity and to elucidate its role in stress responses.

References

Aarts, M.G.M., Keijzer, C.J., Stiekema, W.J. and Pereira, A. (1995). Molecular characterization of the *CER1* gene of *Arabidopsis* involved in epicuticular wax biosynthesis and pollen fertility. Plant Cell, 7(12):2115-2127.

Abe, M., Takahashi, T. and Komeda, Y. (1999). Cloning and characterization of an L1 layer-specific gene in *Arabidopsis thaliana*. Plant and Cell Physiology, 40(6):571-580.

Achkor, H., Díaz, M., Fernández, M.R., Biosca, J.A., Parés, X. and Martínez, M.C. (2003). Enhanced formaldehyde detoxification by overexpression of glutathione-dependent formaldehyde dehydrogenase from Arabidopsis. Plant Physiology, 132(4):2248-2255.

Agrawal, A.A., Laforsch, C. and Tollrian, R. (1999). Transgenerational induction of defences in animals and plants. Nature, 401(6748):60-63.

Agrawal, V.P. and Kolattukudy, P.E. (1977). Biochemistry of suberization: ω-hydroxyacid oxidation in enzyme preparations from suberizing potato tuber disks. Plant Physiology, 59(4):667-672.

Agrawal, V.P. and Kolattukudy, P.E. (1978a). Purification and characterization of a wound-induced ω-hydroxyfatty acid:NADP oxidoreductase from potato tuber disks (*Solanum tuberosum* L.). Archives of Biochemistry and Biophysics, 191(2):452-465.

Agrawal, V.P. and Kolattukudy, P.E. (1978b). Mechanism of action of a wound-induced ω-hydroxy-fatty acid:NADP oxidoreductase isolated from potato tubers (*Solanum tuberosum* L). Archives of Biochemistry and Biophysics, 191(2):466-478.

Aharoni, A., Dixit, S., Jetter, R., Thoenes, E., van Arkel, G. and Pereira, A. (2004). The SHINE clade of AP2 domain transcription factors activates wax biosynthesis, alters cuticle properties, and confers drought tolerance when overexpressed in *Arabidopsis*. Plant Cell, 16(9):2463-2480.

Akoh, C.C., Lee, G.C., Liaw, Y.C., Huang, T.H. and Shaw, J.F. (2004). GDSL family of serine esterases/lipases. Progress in Lipid Research, 43(6):534-552.

Allocco, D., Kohane, I. and Butte, A. (2004). Quantifying the relationship between co-expression, co-regulation and gene function. BMC Bioinformatics, 5(1):18-27.

Araki, T., Nakatani, M. and Iwabuchi, M. (1998). *ADHESION OF CALYX EDGES*, a gene involved in the regulation of postgenital fusion in *Arabidopsis*. Paper presented at the 9th International Conference on Arabidopsis Research

Arvidsson, S., Kwasniewski, M., Riano-Pachon, D.M. and Mueller-Roeber, B. (2008). QuantPrime- A flexible tool for reliable high-throughput primer design for quantitative PCR. BMC Bioinformatics, 9:465-479.

Avato, P., Bianchi, G., Nayak, A., Salamini, F. and Gentinetta, E. (1987). Epicuticular waxes of maize as affected by the interaction of mutant *gl8* with *gl3*, *gl4* and *gl15*. Lipids, 22(1):11-16.

Bach, L., Michaelson, L.V., Haslam, R., Bellec, Y., Gissot, L., Marion, J., Da Costa, M., Boutin, J.-P., Miquel, M., Tellier, F., Domergue, F., Markham, J.E., Beaudoin, F., Napier, J.A. and Faure, J.-D. (2008). The very-long-chain hydroxy fatty acyl-CoA dehydratase PASTICCINO2 is essential and limiting for plant development. Proceedings of the National Academy of Sciences of the United States of America, 105(38):14727-14731.

Bak, S., Beisson, F., Bishop, G., Hamberger, B., Hofer, R., Paquette, S. and Werck-Reichhart, D. (2011). Cytochromes P450 Vol. 9. Arabidopsis Book (pp. e0144).

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science, 315(5819):1709-1712.

Barrett, S.C.H. (2002). The evolution of plant sexual diversity. Nature Reviews Genetics, 3(4):274-284.

Beaudoin, F., Wu, X., Li, F., Haslam, R.P., Markham, J.E., Zheng, H., Napier, J.A. and Kunst, L. (2009). Functional characterization of the *Arabidopsis* β-ketoacyl-coenzyme a reductase candidates of the fatty acid elongase. Plant Physiology, 150(3):1174-1191.

Bechtold, N. and Pelletier, G. (1998). *In planta Agrobacterium*-mediated transformation of adult *Arabidopsis thaliana* plants by vacuum infiltration. In Martinez-Zapater, J.M. and Salinas, J. (Series Eds.), Vol. 82. Arabidopsis Protocols (pp. 259-266). Retrieved from http://dx.doi.org/10.1385/0-89603-391-0:259 doi:10.1385/0896033910

Beeckman, T., De Rycke, R., Viane, R. and Inzé, D. (2000). Histological Study of Seed Coat Development in *Arabidopsis thaliana*. Journal of Plant Research, 113(2):139-148.

Beisson, F., Li, Y., Bonaventure, G., Pollard, M. and Ohlrogge, J.B. (2007). The acyltransferase GPAT5 is required for the synthesis of suberin in seed coat and root of *Arabidopsis*. The Plant Cell, 19(1):351-368.

Benveniste, I., Tijet, N., Adas, F., Philipps, G., Salaun, J.P. and Durst, F. (1998). *CYP86A1* from *Arabidopsis thaliana* encodes a cytochrome P450-dependent fatty acid ω-hydroxylase. Biochemical and Biophysical Research Communications, 243(3):688-693.

Bernard, A., Domergue, F., Pascal, S., Jetter, R., Renne, C., Faure, J.-D., Haslam, R.P., Napier, J.A., Lessire, R. and Joubès, J. (2012). Reconstitution of plant alkane biosynthesis in yeast demonstrates that *Arabidopsis* ECERIFERUM1 and ECERIFERUM3 are core components of a very-long-chain alkane synthesis complex. Plant Cell, 24(7):3106-3118.

Bernhardt, C., Zhao, M., Gonzalez, A., Lloyd, A. and Schiefelbein, J. (2005). The bHLH genes *GL3* and *EGL3* participate in an intercellular regulatory circuit that controls cell patterning in the *Arabidopsis* root epidermis. Development, 132(2):291-298.

Bessire, M., Chassot, C., Jacquat, A.C., Humphry, M., Borel, S., Petetot, J.M., Metraux, J.P. and Nawrath, C. (2007). A permeable cuticle in *Arabidopsis* leads to a strong resistance to *Botrytis cinerea*. The EMBO Journal, 26(8):2158-2168.

Bessire, M., Borel, S., Fabre, G., Carraca, L., Efremova, N., Yephremov, A., Cao, Y., Jetter, R., Jacquat, A.C., Metraux, J.P. and Nawrath, C. (2011). A member of the PLEIOTROPIC DRUG RESISTANCE family of ATP binding cassette transporters is required for the formation of a functional cuticle in *Arabidopsis*. Plant Cell, 23(5):1958-1970.

Bianchi, A., Bianchi, G., Avato, P. and Salamini, F. (1985). Biosynthetic pathways of epicuticular wax of maize as assessed by mutation, light, plant-age and inhibitor studies. Maydica, 30(2):179-198.

Bird, D., Beisson, F., Brigham, A., Shin, J., Greer, S., Jetter, R., Kunst, L., Wu, X., Yephremov, A. and Samuels, L. (2007). Characterization of *Arabidopsis* ABCG11/WBC11, an ATP binding cassette (ABC) transporter that is required for cuticular lipid secretion. Plant Journal, 52(3):485-498.

Bonaventure, G., Beisson, F., Ohlrogge, J. and Pollard, M. (2004). Analysis of the aliphatic monomer composition of polyesters associated with *Arabidopsis* epidermis: occurrence of octadeca-*cis*-6, *cis*-9-diene-1,18-dioate as the major component. Plant Journal, 40(6):920-930.

Bonnett, H.T., Jr. and Newcomb, E.H. (1965). Polyribosomes and cisternal accumulations in root cells of radish. Journal of Cell Biology, 27(2):423-432.

Borner, G.H., Lilley, K.S., Stevens, T.J. and Dupree, P. (2003). Identification of glycosylphosphatidylinositol-anchored proteins in *Arabidopsis*. A proteomic and genomic analysis. Plant Physiology, 132(2):568-577.

Boyko, A., Filkowski, J. and Kovalchuk, I. (2005). Homologous recombination in plants is temperature and day-length dependent. Mutation Research, 572(1-2):73-83.

Boyko, A., Hudson, D., Bhomkar, P., Kathiria, P. and Kovalchuk, I. (2006). Increase of homologous recombination frequency in vascular tissue of Arabidopsis plants exposed to salt stress. Plant and Cell Physiology, 47(6):736-742.

Boyko, A., Kathiria, P., Zemp, F.J., Yao, Y., Pogribny, I. and Kovalchuk, I. (2007). Transgenerational changes in the genome stability and methylation in pathogen-infected plants (virus-induced plant genome instability). Nucleic Acids Research, 35(5):1714-1725.

Boyko, A., Golubov, A., Bilichak, A. and Kovalchuk, I. (2010). Chlorine ions but not sodium ions alter genome stability of *Arabidopsis thaliana*. Plant and Cell Physiology, 51(6):1066-1078.

Boyko, A. and Kovalchuk, I. (2011). Genome instability and epigenetic modification: Heritable responses to environmental stress? Current Opinion in Plant Biology, 14(3):260-266.

Brabban, A.D. and Edwards, C. (1995). The effects of glucosinolates and their hydrolysis products on microbial growth. Journal of Applied Bacteriology, 79(2):171-177.

Brakke, M.K. (1984). Mutations, the abberant ratio phenomenon, and virus infection of maize. Papers in Plant Pathology:163.

Broun, P., Poindexter, P., Osborne, E., Jiang, C.Z. and Riechmann, J.L. (2004). WIN1, a transcriptional activator of epidermal wax accumulation in *Arabidopsis*. Proceedings

of the National Academy of Sciences of the United States of America, 101(13):4706-4711.

Bucher, M.H., Evdokimov, A.G. and Waugh, D.S. (2002). Differential effects of short affinity tags on the crystallization of *Pyrococcus furiosus* maltodextrin-binding protein. Biological Crystallography, 58(3):392-397.

Bunz, F. (2008). Principles of Cancer Genetics. New York City, NY: Springer Publishing.

Buschhaus, C. and Jetter, R. (2011). Composition differences between epicuticular and intracuticular wax substructures: How do plants seal their epidermal surfaces? Journal of Experimental Botany, 62(3):841-853.

Carter, C., Pan, S., Zouhar, J., Avila, E.L., Girke, T. and Raikhel, N.V. (2004). The vegetative vacuole proteome of *Arabidopsis thaliana* reveals predicted and unexpected proteins. Plant Cell, 16(12):3285-3303.

Cavener, D.R. (1992). GMC oxidoreductases. A newly defined family of homologous proteins with diverse catalytic activities. Journal of Molecular Biology, 223(3):811-814.

Chaban, I.A., Lazareva, E.M., Kononenko, N.V. and Polyakov, V.Y. (2011). Antipodal complex development in the embryo sac of wheat. Russian Journal of Developmental Biology, 42(2):79-91.

Chable, V., Rival, A., Cadot, V., Boulineau, F., Salmon, A., Bellis, H. and Manzanares-Dauleux, M.J. (2008). Aberrant plants in cauliflower: 1. Phenotype and heredity. Euphytica, 164(2):325-337.

Chandler, V.L., Eggleston, W.B. and Dorweiler, J.E. (2000). Paramutation in maize. Plant Molecular Biology, 43(2-3):121-145.

Chassot, C., Nawrath, C. and Metraux, J.-P. (2007). Cuticular defects lead to full immunity to a major plant pathogen. Plant Journal, 49(6):972-980.

Chaudhury, A. (2005). Plant genetics: Hothead healer and extragenomic information. Nature, 437(7055):E1.

Chen, M., Du, X., Zhu, Y., Wang, Z., Hua, S., Li, Z., Guo, W., Zhang, G., Peng, J. and Jiang, L. (2012). Seed Fatty Acid Reducer acts downstream of gibberellin signalling

pathway to lower seed fatty acid storage in *Arabidopsis*. Plant Cell Environ, 35(12):2155-2169.

Chen, X., Goodwin, S.M., Boroff, V.L., Liu, X. and Jenks, M.A. (2003). Cloning and characterization of the *WAX2* gene of *Arabidopsis* involved in cuticle membrane and wax production. The Plant Cell, 15(5):1170-1185.

Chen, Y., Lowenfeld, R. and Cullis, C. (2009). An environmentally induced adaptive (?) insertion in flax. International Journal of Genetics and Molecular Biology, 1(3):38-47.

Chen, Y.M., Schneeberger, R.G. and Cullis, C.A. (2005). A site-specific insertion sequence in flax genotrophs induced by environment. New Phytologist, 167(1):171-180.

Cheng, Q., Liu, H.T., Bombelli, P., Smith, A. and Slabas, A.R. (2004). Functional identification of AtFao3, a membrane bound long chain alcohol oxidase in *Arabidopsis thaliana*. FEBS Letters, 574:62-68.

Choi, C.-S. and Sano, H. (2007). Abiotic-stress induces demethylation and transcriptional activation of a gene encoding a glycerophosphodiesterase-like protein in tobacco plants. Molecular Genetics and Genomics, 277(5):589-600.

Christensen, A.C., King, J.E., Jordan, R.J. and Drews, N.G. (1997). Megagametogenesis in *Arabidopsis* wild type and the *Gf* mutant. Sexual Plant Reproduction, 10(1):49-64.

Chuong, S.D.X., Park, N.I., Freeman, M.C., Mullen, R.T. and Muench, D.G. (2005). The peroxisomal multifunctional protein interacts with cortical microtubules in plant cells. BMC Cell Biology, 6:13.

Clark, R.M., Schweikert, G., Toomajian, C., Ossowski, S., Zeller, G., Shinn, P., Warthmann, N., Hu, T.T., Fu, G., Hinds, D.A., Chen, H.M., Frazer, K.A., Huson, D.H., Schoelkopf, B., Nordborg, M., Raetsch, G., Ecker, J.R. and Weigel, D. (2007). Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. Science, 317(5836):338-342.

Comai, L. and Cartwright, R.A. (2005). A toxic mutator and selection alternative to the non-Mendelian RNA cache hypothesis for hothead reversion. Plant Cell, 17(11):2856-2858.

Cominelli, E., Sala, T., Calvi, D., Gusmaroli, G. and Tonelli, C. (2008). Over-expression of the *Arabidopsis AtMYB41* gene alters cell expansion and leaf surface permeability. Plant Journal, 53(1):53-64.

Compagnon, V., Diehl, P., Benveniste, I., Meyer, D., Schaller, H., Schreiber, L., Franke, R. and Pinot, F. (2009). CYP86B1 is required for very long chain ω-hydroxyacid and α,ω-dicarboxylic acid synthesis in root and seed suberin polyester. Plant Physiology, 150(4):1831-1843.

Conn, E.E. (1981). Secondary plant products. In Stumpf, P.K. and Conn, E.E. (Eds.), The Biochemistry of Plants: A Comprehensive Treatise (Vol. 7, pp. 479–500). New York: Academic Press.

Copley, S.D. (2003). Enzymes with extra talents: Moonlighting functions and catalytic promiscuity. Current Opinion in Chemical Biology, 7(2):265-272.

Copley, S.D. (2012). Moonlighting is mainstream: Paradigm adjustment required. Bioessays, 34(7):578-588.

Creff, A., Brocard, L. and Ingram, G. (2015). A mechanically sensitive cell layer regulates the physical properties of the *Arabidopsis* seed coat. Nature Communications, 6:6382-6389.

Cullis, C., Swami, S. and Song, Y. (2004). RAPD polymorphisms detected among the flax genotrophs. Plant Molecular Biology, 41(6):795-800.

De Giorgi, J., Piskurewicz, U., Loubery, S., Utz-Pugin, A., Bailly, C., Mene-Saffrane, L. and Lopez-Molina, L. (2015). An endosperm-associated cuticle is required for *Arabidopsis* seed viability, dormancy and early control of germination. PLoS Genetics, 11(12):e1005708.

De Pourcq, K., De Schutter, K. and Callewaert, N. (2010). Engineering of glycosylation in yeast and other fungi: current state and perspectives. Applied Microbiology and Biotechnology, 87(5):1617-1631.

DeBlasio, S.L., Sylvester, A.W. and Jackson, D. (2010). Illuminating plant biology: Using fluorescent proteins for high-throughput analysis of protein localization and function in plants. Briefings in Functional Genomics, 9(2):129-138.

DeBolt, S. (2010). Copy number variation shapes genome diversity in *Arabidopsis* over immediate family generational scales. Genome Biology and Evolution, 2:441-453.

DeBono, A., Yeats, T.H., Rose, J.K.C., Bird, D., Jetter, R., Kunst, L. and Samuelsa, L. (2009). *Arabidopsis* LTPG Is a glycosylphosphatidylinositol-anchored lipid transfer protein required for export of lipids to the plant surface. Plant Cell, 21(4):1230-1238.

Deng, Y., Srivastava, R. and Howell, S.H. (2013). Endoplasmic reticulum (ER) stress response and its physiological roles in plants. International Journal of Molecular Sciences, 14(4):8188-8212.

Dickinson, F.M. and Wadforth, C. (1992). Purification and some properties of alcohol oxidase from alkane-grown *Candida tropicalis*. Biochemical Journal, 282:325-331.

Dickinson, J.R., Salgado, L.E. and Hewlins, M.J. (2003). The catabolism of amino acids to long chain and complex alcohols in *Saccharomyces cerevisiae*. Journal of Biological Chemistry, 278(10):8028-8034.

Dietrich, C.R., Perera, M., Yandeau-Nelson, M.D., Meeley, R.B., Nikolau, B.J. and Schnable, P.S. (2005). Characterization of two GL8 paralogs reveals that the 3-ketoacyl reductase component of fatty acid elongase is essential for maize (*Zea mays* L.) development. Plant Journal, 42(6):844-861.

Diévart, A. and Clark, S.E. (2004). LRR-containing receptors regulating plant development and defense. Development, 131(2):251-261.

Dombrecht, B., Xue, G.P., Sprague, S.J., Kirkegaard, J.A., Ross, J.J., Reid, J.B., Fitt, G.P., Sewelam, N., Schenk, P.M., Manners, J.M. and Kazan, K. (2007). MYC2 differentially modulates diverse jasmonate-dependent functions in *Arabidopsis*. Plant Cell, 19(7):2225-2245.

Domínguez, E., Cuartero, J. and Heredia, A. (2011). An overview on plant cuticle biomechanics. Plant Science, 181(2):77-84.

Dreveny, I., Gruber, K., Glieder, A., Thompson, A. and Kratky, C. (2001). The hydroxynitrile lyase from almond: A lyase that looks like an oxidoreductase. Structure, 9(9):803-815.

Dreveny, I., Kratky, C. and Gruber, K. (2002). The active site of hydroxynitrile lyase from *Prunus amygdalus*: Modeling studies provide new insights into the mechanism of cyanogenesis. Protein Science, 11(2):292-300.

Dreveny, I., Andryushkova, A.S., Glieder, A., Gruber, K. and Kratky, C. (2009). Substrate binding in the FAD-dependent hydroxynitrile lyase from almond provides insight into the mechanism of cyanohydrin formation and explains the absence of dehydrogenation activity. Biochemistry, 48(15):3370-3377.

Drews, G.N., Lee, D. and Christensen, C.A. (1998). Genetic analysis of female gametophyte development and function. The Plant Cell, 10(1):5-17.

Drews, G.N. and Koltunow, A.M.G. (2011). The Female Gametophyte Vol. 9. The Arabidopsis Book / American Society of Plant Biologists (pp. e0155). Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3268550/pdf/tab.0155.pdf

Eamens, A.L., Wook Kim, K. and Waterhouse, P.M. (2012). DRB2, DRB3 and DRB5 function in a non-canonical microRNA pathway in *Arabidopsis thaliana*. Plant Signaling & Behavior, 7(10):1224-1229.

Edstam, M.M. and Edqvist, J. (2014). Involvement of GPI-anchored lipid transfer proteins in the development of seed coats and pollen in *Arabidopsis thaliana*. Physiologia Plantarum, 152(1):32-42.

Edwards, K., Johnstone, C. and Thompson, C. (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Research, 19(6):1349-1349.

Edwards, Y.J. and Cottage, A. (2003). Bioinformatics methods to predict protein structure and function. A practical approach. Molecular Biotechnology, 23(2):139-166.

Eisen, M.B., Spellman, P.T., Brown, P.O. and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proceedings of the National Academy of Sciences of the United States of America, 95(25):14863-14868.

Ejiri, S. (2002). Moonlighting functions of polypeptide Elongation Factor 1: From actin bundling to zinc finger protein R1-associated nuclear localization. Bioscience Biotechnology and Biochemistry, 66(1):1-21.

Espelie, K.E., Dean, B.B. and Kolattukudy, P.E. (1979). Composition of lipid-derived polymers from different anatomical regions of several plant species. Plant Physiology, 64(6):1089-1093.

Fagerstrom, T., Briscoe, D.A. and Sunnucks, P. (1998). Evolution of mitotic cell-lineages in multicellular organisms. Trends in Ecology and Evolution, 13(3):117-120.

Fasoula, V.A. and Boerma, H.R. (2005). Divergent selection at ultra-low plant density for seed protein and oil content within soybean cultivars. Field Crops Research, 91(2-3):217-229.

Fasoula, V.A. and Boerma, H.R. (2007). Intra-cultivar variation for seed weight and other agronomic traits within three elite soybean cultivars. Crop Science, 47(1):367-373.

Feng, S., Jacobsen, S.E. and Reik, W. (2010). Epigenetic reprogramming in plant and animal development. Science, 330(6004):622-627.

Fiebig, A., Mayfield, J.A., Miley, N.L., Chau, S., Fischer, R.L. and Preuss, D. (2000). Alterations in *CER6*, a gene identical to *CUT1*, differentially affect long-chain lipid content on the surface of pollen and stems. Plant Cell, 12(10):2001-2008.

Fieldes, M.A. and Gerhardt, K.E. (1994). An examination of the β-glucosidase (linamarase) banding pattern in flax seedlings using Ferguson plots and sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Electrophoresis, 15(5):654-661.

Fire, A., Xu, S.Q., Montgomery, M.K., Kostas, S.A., Driver, S.E. and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature, 391(6669):806-811.

FitzGerald, A., O'Donoghue, D.P. and Liu, X.Y. (2010). Genetic repair strategies inspired by *Arabidopsis thaliana*. In Coyle, L. and Freyne, J. (Eds.), Artificial Intelligence and Cognitive Science (Vol. 6206, pp. 61-71).

Folse, H.J. and Roughgarden, J. (2012). Direct benefits of genetic mosaicism and intraorganismal selection: Modeling coevolution between a long-lived tree and a short-lived herbivore. Evolution, 66(4):1091-1113.

Franke, R., Briesen, I., Wojciechowski, T., Faust, A., Yephremov, A., Nawrath, C. and Schreiber, L. (2005). Apoplastic polyesters in *Arabidopsis* surface tissues- A typical suberin and a particular cutin. Phytochemistry, 66(22):2643-2658.

Franke, R., Hofer, R., Briesen, I., Emsermann, M., Efremova, N., Yephremov, A. and Schreiber, L. (2009). The *DAISY* gene from *Arabidopsis* encodes a fatty acid elongase condensing enzyme involved in the biosynthesis of aliphatic suberin in roots and the chalaza-micropyle region of seeds. Plant Journal, 57(1):80-95.

Frisch, T. and Moller, B.L. (2012). Possible evolution of alliarinoside biosynthesis from the glucosinolate pathway in *Alliaria petiolata*. FEBS Journal, 279(9):1545-1562.

Gable, K., Garton, S., Napier, J.A. and Dunn, T.M. (2004). Functional characterization of the *Arabidopsis thaliana* orthologue of Tsc13p, the enoyl reductase of the yeast microsomal fatty acid elongating system. Journal of Experimental Botany, 55(396):543-545.

Galloway, L.F. and Etterson, J.R. (2007). Transgenerational plasticity is adaptive in the wild. Science, 318(5853):1134-1136.

Gao, D.Y., He, B., Zhou, Y.H. and Sun, L.H. (2011). Genetic and molecular analysis of a purple sheath somaclonal mutant in *Japonica* rice. Plant Cell Reports, 30(5):901-911.

Gbadegesin, M.A. (2012). Transposable elements in the genomes: parasites, junks or drivers of evolution? African journal of medicine and medical sciences, 41 Suppl:13-25.

Geisse, S., Gram, H., Kleuser, B. and Kocher, H.P. (1996). Eukaryotic expression systems: A comparison. Protein Expression and Purification, 8(3):271-282.

Golldack, D., Vera, P. and Dietz, K.-J. (2003). Expression of subtilisin-like serine proteases in *Arabidopsis thaliana* is cell-specific and responds to jasmonic acid and heavy metals with developmental differences. Physiologia Plantarum, 118(1):64-73.

Graca, J., Schreiber, L., Rodrigues, J. and Pereira, H. (2002). Glycerol and glyceryl esters of ω-hydroxyacids in cutins. Phytochemistry, 61(2):205-215.

Granell, A., Bellés, J.M. and Conejero, V. (1987). Induction of pathogenesis-related proteins in tomato by citrus exocortis viroid, silver ion and ethephon. Physiological and Molecular Plant Pathology, 31(1):83-90.

Gray, J.E., Holroyd, G.H., van der Lee, F.M., Bahrami, A.R., Sijmons, P.C., Woodward, F.I., Schuch, W. and Heterington, A.M. (2000). The HIC signalling pathway links CO_2 perception to stomatal development. Nature, 408(6813):713-716.

Greenhagen, B.T., O'Maille, P.E., Noel, J.P. and Chappell, J. (2006). Identifying and manipulating structural determinates linking catalytic specificities in terpene synthases. Proceedings of the National Academy of Sciences of the United States of America, 103(26):9826-9831.

Greer, S., Wen, M., Bird, D., Wu, X., Samuels, L., Kunst, L. and Jetter, R. (2007). The cytochrome P450 enzyme CYP96A15 is the midchain alkane hydroxylase responsible for formation of secondary alcohols and ketones in stem cuticular wax of *Arabidopsis*. Plant Physiology, 145(3):653-667.

Grob, K. and Matile, P.H. (1979). Vacuolar location of glucosinolates in horseradish root cells. Plant Science Letters, 14(4):327-335.

Gronwald, J.W. (1991). Lipid biosynthesis inhibitors. Weed Science, 39(3):435-449.

Gunning, B.E.S. (1998). The identity of mystery organelles in *Arabidopsis* plants expressing GFP. Trends in Plant Science, 3(11):417-417.

Guzman, P., Fernandez, V., Garcia, M.L., Khayet, M., Fernandez, A. and Gil, L. (2014). Localization of polysaccharides in isolated and intact cuticles of eucalypt, poplar and pear leaves by enzyme-gold labelling. Plant Physiology and Biochemistry, 76:1-6.

Hamilton, S.R. and Zha, D. (2015). Progress in yeast glycosylation engineering. Methods in Molecular Biology, 1321:73-90.

Hara-Nishimura, I., Shimada, T., Hatano, K., Takeuchi, Y. and Nishimura, M. (1998). Transport of storage proteins to protein storage vacuoles is mediated by large precursor-accumulating vesicle. Plant Cell, 10(5):825-836.

Hara-Nishimura, I. and Matsushima, R. (2003). A wound-inducible organelle derived from endoplasmic reticulum: A plant strategy against environmental stresses? Current Opinion in Plant Biology, 6(6):583-588.

Hara-Nishimura, I., Matsushima, R., Shimada, T. and Nishimura, M. (2004). Diversity and formation of endoplasmic reticulum-derived compartments in plants. Are these compartments specific to plant cells? Plant Physiology, 136(3):3435-3439.

Haslam, T.M. and Kunst, L. (2013). Extending the story of very-long-chain fatty acid elongation. Plant Science, 210:93-107.

Haughn, G. and Chaudhury, A. (2005). Genetic analysis of seed coat development in *Arabidopsis*. Trends in Plant Science, 10(10):472-477.

Hawes, C., Saint-Jore, C., Martin, B. and Zheng, H.Q. (2001). ER confirmed as the location of mystery organelles in *Arabidopsis* plants expressing GFP! Trends in Plant Science, 6(6):245-246.

Hayashi, Y., Yamada, K., Shimada, T., Matsushima, R., Nishizawa, N.K., Nishimura, M. and Hara-Nishimura, I. (2001). A proteinase-storing body that prepares for cell death or stresses in the epidermal cells of *Arabidopsis*. Plant and Cell Physiology, 42(9):894-899.

Hecht, H.J., Kalisz, H.M., Hendle, J., Schmid, R.D. and Schomburg, D. (1993). Crystal structure of glucose oxidase from *Aspergillus niger* refined at 2.3 A resolution. Journal of Molecular Biology, 229(1):153-172.

Heredia-Guerrero, J.A., Benitez, J.J., Dominguez, E., Bayer, I.S., Cingolani, R., Athanassiou, A. and Heredia, A. (2014). Infrared and Raman spectroscopic features of plant cuticles: A review. Frontiers in Plant Science, 5:305.

Heredia, A. and Benavente, J. (1991). A study of membrane potential across isolated fruit cuticles for NaCl and $CaCl_2$ solutions. Biochimica et Biophysica Acta, 1062(2):239-244.

Heyndrickx, K.S. and Vandepoele, K. (2012). Systematic identification of functional plant modules through the integration of complementary data sources. Plant Physiology, 159(3):884-901.

Hoffmannbenning, S., Klomparens, K.L. and Kende, H. (1994). Characterization of growth-related osmiophilic particles in corn coleoptiles and deep-water rice internodes. Annals of Botany, 74(6):563-572.

Hooker, T.S., Lam, P., Zheng, H. and Kunst, L. (2007). A core subunit of the RNA-processing/degrading exosome specifically influences cuticular wax biosynthesis in *Arabidopsis*. Plant Cell, 19(3):904-913.

Hopkin, M. (2008). Biodiversity: Frozen futures. Nature, 452(7186):404-405.

Hopkins, M., Khalid, A., Chang, P.-C., Vanderhoek, K., Lai, D., Doerr, M. and Lolle, S. (2013). *De novo* genetic variation revealed in somatic sectors of single *Arabidopsis* plants. F1000Research, 2:5-15.

Hopkins, R.J., van Dam, N.M. and van Loon, J.J. (2009). Role of glucosinolates in insect-plant relationships and multitrophic interactions. Annual Review of Entomology, 54:57-83.

Hu, Z. and Poulton, J.E. (1997). Sequencing, genomic organization, and preliminary promoter analysis of a black cherry (*R*)-(+)-mandelonitrile lyase gene. Plant Physiology, 115(4):1359-1369.

Huberts, D.H. and van der Klei, I.J. (2010). Moonlighting proteins: An intriguing mode of multitasking. Biochimica et Biophysica Acta, 1803(4):520-525.

Hulskamp, M., Kopczak, S.D., Horejsi, T.F., Kihl, B.K. and Pruitt, R.E. (1995). Identification of genes required for pollen-stigma recognition in *Arabidopsis thaliana*. Plant Journal, 8(5):703-714.

Irshad, M., Canut, H., Borderies, G., Pont-Lezica, R. and Jamet, E. (2008). A new picture of cell wall protein dynamics in elongating cells of *Arabidopsis thaliana*: Confirmed actors and newcomers. BMC Plant Biology, 8:94.

Jacobs, P.P., Geysens, S., Vervecken, W., Contreras, R. and Callewaert, N. (2009). Engineering complex-type N-glycosylation in *Pichia pastoris* using GlycoSwitch technology. Nature Protocols, 4(1):58-70.

Jain, S.M. (2001). Tissue culture-derived variation in crop improvement. Euphytica, 118(2):153-166.

Jakoby, M.J., Falkenhan, D., Mader, M.T., Brininstool, G., Wischnitzki, E., Platz, N., Hudson, A., Hulskamp, M., Larkin, J. and Schnittger, A. (2008). Transcriptional profiling of mature Arabidopsis trichomes reveals that *NOECK* encodes the MIXTA-like transcriptional regulator MYB106. Plant Physiology, 148(3):1583-1602.

James, D.W., Lim, E., Keller, J., Plooy, I., Ralston, E. and Dooner, H.K. (1995). Directed tagging of the *Arabidopsis FATTY-ACID ELONGATION-1* (*FAE1*) gene with the maize transposon activator. Plant Cell, 7(3):309-319.

Javelle, M., Vernoud, V., Rogowsky, P.M. and Ingram, G.C. (2011). Epidermis: The formation and functions of a fundamental plant tissue. New Phytologist, 189(1):17-39.

Jenks, M.A., Tuttle, H.A., Eigenbrode, S.D. and Feldmann, K.A. (1995). Leaf epicuticular waxes of the *eceriferum* mutants in *Arabidopsis*. Plant Physiology, 108(1):369-377.

Jiang, C.F., Mithani, A., Gan, X.C., Belfield, E.J., Klingler, J.P., Zhu, J.K., Ragoussis, J., Mott, R. and Harberd, N.P. (2011). Regenerant *Arabidopsis* lineages display a distinct genome-wide spectrum of mutations conferring variant phenotypes. Current Biology, 21(16):1385-1390.

Jin, P., Guo, T. and Becraft, P.W. (2000). The maize CR4 receptor-like kinase mediates a growth factor-like differentiation response. Genesis, 27(3):104-116.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science, 337(6096):816-821.

Jorda, L., Coego, A., Conejero, V. and Vera, P. (1999). A genomic cluster containing four differentially regulated subtilisin-like processing protease genes is in tomato plants. Journal of Biological Chemistry, 274(4):2360-2365.

Jorns, M.S. (1979). Mechanism of catalysis by the flavoenzyme oxynitrilase. Journal of Biological Chemistry, 254(23):12145-12152.

Kader, J.-C. (1997). Lipid-transfer proteins: A puzzling family of plant proteins. Trends in Plant Science, 2(2):66-70.

Kaeppler, S.M. and Phillips, R.L. (1993). Tissue culture-induced DNA methylation variation in maize. Proceedings of the National Academy of Sciences, 90(19):8773-8776.

Kagi, J.H. and Vallee, B.L. (1960). The role of zinc in alcohol dehydrogenase. [V. The effect of metal-binding agents on thestructure of the yeast alcohol dehydrogenase molecule]. Journal of Biological Chemistry, 235(11):3188-3192.

Kalisz, H.M., Hendle, J. and Schmid, R.D. (1997). Structural and biochemical properties of glycosylated and deglycosylated glucose oxidase from *Penicillium amagasakiense*. Applied Microbiology and Biotechnology, 47(5):502-507.

Kandel, S., Sauveplane, V., Olry, A., Diss, L., Benveniste, I. and Pinot, F. (2006). Cytochrome P450-dependent fatty acid hydroxylases in plants. Phytochemistry Reviews, 5(2-3):359-372.

Kannangara, R., Branigan, C., Liu, Y., Penfield, T., Rao, V., Mouille, G., Hofte, H., Pauly, M., Riechmann, J.L. and Broun, P. (2007). The transcription factor WIN1/SHN1 regulates cutin biosynthesis in *Arabidopsis thaliana*. Plant Cell, 19(4):1278-1294.

Kapust, R.B. and Waugh, D.S. (1999). *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Science, 8(8):1668-1674.

Karasov, T.L., Horton, M.W. and Bergelson, J. (2014). Genomic variability as a driver of plant-pathogen coevolution? Current Opinion in Plant Biology, 18:24-30.

Kato, N., Reynolds, D., Brown, M.L., Boisdore, M., Fujikawa, Y., Morales, A. and Meisel, L.A. (2008). Multidimensional fluorescence microscopy of multiple organelles in Arabidopsis seedlings. Plant Methods, 4:9-22.

Kauss, H., Fauth, M., Merten, A. and Jeblick, W. (1999). Cucumber hypocotyls respond to cutin monomers via both an inducible and a constitutive H_2O_2-generating system. Plant Physiology, 120(4):1175-1182.

Kensche, P.R., van Noort, V., Dutilh, B.E. and Huynen, M.A. (2008). Practical and theoretical advances in predicting the function of a protein by its phylogenetic distribution. Journal of the Royal Society Interface, 5(19):151-170.

Kerstiens, G. (1996). Signalling across the divide: A wider perspective of cuticular structure function relationships. Trends in Plant Science, 1(4):125-129.

Khow, O. and Suntrarachun, S. (2012). Strategies for production of active eukaryotic proteins in bacterial expression system. Asian Pacific Journal of Tropical Biomedicine, 2(2):159-162.

Kim, H., Lee, S.B., Kim, H.J., Min, M.K., Hwang, I. and Suh, M.C. (2012). Characterization of glycosylphosphatidylinositol-anchored lipid transfer protein 2 (LTPG2) and overlapping function between LTPG/LTPG1 and LTPG2 in cuticular wax export or accumulation in *Arabidopsis thaliana*. Plant and Cell Physiology, 53(8):1391-1403.

Kim, Y.W., Kang, K.S., Kim, S.Y. and Kim, I.S. (2000). Formation of fibrillar multimers of oat β-glucosidase isoenzymes is mediated by the As-Glu1 monomer. Journal of Molecular Biology, 303(5):831-842.

Kissen, R. and Bones, A.M. (2009). Nitrile-specifier proteins involved in glucosinolate hydrolysis in *Arabidopsis thaliana*. Journal of Biological Chemistry, 284(18):12057-12070.

Koch, K. and Ensikat, H.-J. (2008). The hydrophobic coatings of plant surfaces: Epicuticular wax crystals and their morphologies, crystallinity and molecular self-assembly. Micron, 39(7):759-772.

Kolattukudy, P.E. (2001). Polyesters in higher plants. Advances in Biochemical Engineering Biotechnology, 71:1-49.

Kondou, Y., Nakazawa, M., Kawashima, M., Ichikawa, T., Yoshizumi, T., Suzuki, K., Ishikawa, A., Koshi, T., Matsui, R., Muto, S. and Matsui, M. (2008). RETARDED GROWTH OF EMBRYO1, a new basic helix-loop-helix protein, expresses in endosperm to control embryo growth. Plant Physiology, 147(4):1924-1935.

Koornneef, M., Hanhart, C. and Thiel, F. (1989). A genetic and phenotypic description of e*ceriferum (cer)* mutants in *Arabidopsis thaliana*. Journal of Heredity, 80(2):118-122.

Kovalchuk, I., Kovalchuk, O. and Hohn, B. (2000). Genome-wide variation of the somatic mutation frequency in transgenic plants. EMBO Journal, 19(17):4431-4438.

Kovalchuk, I., Kovalchuk, O., Kalck, V., Boyko, V., Filkowski, J., Heinlein, M. and Hohn, B. (2003). Pathogen-induced systemic plant signal triggers DNA rearrangements. Nature, 423(6941):760-762.

Krishnaswamy, L. and Peterson, T. (2007). An alternate hypothesis to explain the high frequency of "revertants" in hothead mutants in *Arabidopsis*. Plant Biology, 9(1):30-31.

Kristiansen, K. (2004). Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function. Pharmacology & Therapeutics, 103(1):21-80.

Krolikowski, K.A., Victor, J.L., Wagler, T.N., Lolle, S.J. and Pruitt, R.E. (2003). Isolation and characterization of the *Arabidopsis* organ fusion gene *HOTHEAD*. Plant Journal, 35(4):501-511.

Kunst, L. and Samuels, A.L. (2003). Biosynthesis and secretion of plant cuticular wax. Progress in Lipid Research, 42(1):51-80.

Kurdyukov, S., Faust, A., Nawrath, C., Bar, S., Voisin, D., Efremova, N., Franke, R., Schreiber, L., Saedler, H., Metraux, J.P. and Yephremov, A. (2006a). The epidermis-specific extracellular *BODYGUARD* controls cuticle development and morphogenesis in *Arabidopsis*. Plant Cell, 18(2):321-339.

Kurdyukov, S., Faust, A., Trenkamp, S., Bar, S., Franke, R., Efremova, N., Tietjen, K., Schreiber, L., Saedler, H. and Yephremov, A. (2006b). Genetic and biochemical evidence for involvement of *HOTHEAD* in the biosynthesis of long-chain α-,ω-dicarboxylic fatty acids and formation of extracellular matrix. Planta, 224(2):315-329.

L'Haridon, F., Besson-Bard, A., Binda, M., Serrano, M., Abou-Mansour, E., Balet, F., Schoonbeek, H.J., Hess, S., Mir, R., Leon, J., Lamotte, O. and Metraux, J.P. (2011). A permeable cuticle is associated with the release of reactive oxygen species and induction of innate immunity. PLoS Pathogens, 7(7):e1002148.

Laage, R. and Langosch, D. (2001). Strategies for prokaryotic expression of eukaryotic membrane proteins. Traffic, 2(2):99-104.

Lambrix, V., Reichelt, M., Mitchell-Olds, T., Kliebenstein, D.J. and Gershenzon, J. (2001). The *Arabidopsis* epithiospecifier protein promotes the hydrolysis of glucosinolates to nitriles and influences *Trichoplusia ni* herbivory. Plant Cell, 13(12):2793-2807.

Lang-Mladek, C., Popova, O., Kiok, K., Berlinger, M., Rakic, B., Aufsatz, W., Jonak, C., Hauser, M.T. and Luschnig, C. (2010). Transgenerational inheritance and resetting of stress-induced loss of epigenetic gene silencing in *Arabidopsis*. Molecular Plant, 3(3):594-602.

Lang, J.D., Ray, S. and Ray, A. (1994). *sin1*, a mutation affecting female fertility in *Arabidopsis*, interacts with *mod1*, its recessive modifier. Genetics, 137(4):1101-1110.

Lee, D., Redfern, O. and Orengo, C. (2007). Predicting protein function from sequence and structure. Nature Reviews Molecular Cell Biology, 8(12):995-1005.

Lee, H.O., Davidson, J.M. and Duronio, R.J. (2009a). Endoreplication: polyploidy with purpose. Genes and Development, 23(21):2461-2477.

Lee, J.-Y., Cho, S.-K. and Sager, R. (2011). Plasmodesmata and non-cell-autonomous signaling in plants. In Murphy, A., Peer, W. and Schulz, B. (Eds.), The Plant Plasma Membrane (pp. 87-107). New York: Springer.

Lee, S.B., Go, Y.S., Bae, H.-J., Park, J.H., Cho, S.H., Cho, H.J., Lee, D.S., Park, O.K., Hwang, I. and Suh, M.C. (2009b). Disruption of glycosylphosphatidylinositol-anchored lipid transfer protein gene altered cuticular lipid composition, increased plastoglobules, and enhanced susceptibility to infection by the fungal pathogen *Alternaria brassicicola*. Plant Physiology, 150(1):42-54.

Lelandais-Briere, C., Jovanovic, M., Torres, G.A., Perrin, Y., Lemoine, R., Corre-Menguy, F. and Hartmann, C. (2007). Disruption of AtOCT1, an organic cation transporter gene, affects root development and carnitine-related responses in *Arabidopsis*. Plant Journal, 51(2):154-164.

Li-Beisson, Y., Pollard, M., Sauveplane, V., Pinot, F., Ohlrogge, J. and Beisson, F. (2009). Nanoridges that characterize the surface morphology of flowers require the synthesis of cutin polyester. Proceedings of the National Academy of Sciences, 106(51):22008-22013.

Li-Beisson, Y., Shorrosh, B., Beisson, F., Andersson, M.X., Arondel, V., Bates, P.D., Baud, S., Bird, D., Debono, A., Durrett, T.P., Franke, R.B., Graham, I.A., Katayama, K., Kelly, A.A., Larson, T., Markham, J.E., Miquel, M., Molina, I., Nishida, I., Rowland, O., Samuels, L., Schmid, K.M., Wada, H., Welti, R., Xu, C., Zallot, R. and Ohlrogge, J. (2013). Acyl-lipid metabolism. Arabidopsis Book, 11:e0161.

Li, F., Wu, X., Lam, P., Bird, D., Zheng, H., Samuels, L., Jetter, R. and Kunst, L. (2008). Identification of the wax ester synthase/acyl-coenzyme A: Diacylglycerol acyltransferase WSD1 required for stem wax ester biosynthesis in *Arabidopsis*. Plant Physiology, 148(1):97-107.

Li, L., Li, D., Liu, S., Ma, X., Dietrich, C.R., Hu, H.C., Zhang, G., Liu, Z., Zheng, J., Wang, G. and Schnable, P.S. (2013). The maize *glossy13* gene, cloned via BSR-Seq and Seq-walking encodes a putative ABC transporter required for the normal accumulation of epicuticular waxes. PLoS One, 8(12):e82333.

Li, Y., Beisson, F., Pollard, M. and Ohlrogge, J. (2006). Oil content of Arabidopsis seeds: the influence of seed anatomy, light and plant-to-plant variation. Phytochemistry, 67(9):904-915.

Li, Y., Beisson, F., Koo, A.J., Molina, I., Pollard, M. and Ohlrogge, J. (2007). Identification of acyltransferases required for cutin biosynthesis and production of cutin with suberin-like monomers. Proceedings of the National Academy of Sciences of the United States of America, 104(46):18339-18344.

Ligrone, R., Duckett, J.G. and Renzaglia, K.S. (2012). Major transitions in the evolution of early land plants: a bryological perspective. Annals of Botany, 109(5):851-871.

Lippman, Z., May, B., Yordan, C., Singer, T. and Martienssen, R. (2003). Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. PLoS Biology, 1(3):e67.

Lisch, D. (2009). Epigenetic regulation of transposable elements in plants. Annual Review of Plant Biology, 60:43-66.

Lodish, H., Berk, A. and Zipursky, S. (2000). Protein Glycosylation in the ER and Golgi Complex. In Freeman, W.H. (Ed.), Molecular Cell Biology (4 ed.). New York.

Lolle, S.J., Cheung, A.Y. and Sussex, I.M. (1992). *Fiddlehead*: An *Arabidopsis* mutant constitutively expressing an organ fusion program that involves interactions between epidermal cells. Developmental Biology, 152(2):383-392.

Lolle, S.J. and Cheung, A.Y. (1993). Promiscuous germination and growth of wildtype pollen from *Arabidopsis* and related species on the shoot of the *Arabidopsis* mutant, *fiddlehead*. Developmental Biology, 155(1):250-258.

Lolle, S.J., Berlyn, G.P., Engstrom, E.M., Krolikowski, K.A., Reiter, W.D. and Pruitt, R.E. (1997). Developmental regulation of cell interactions in the *Arabidopsis fiddlehead-1* mutant: A role for the epidermal cell wall and cuticle. Developmental Biology, 189(2):311-321.

Lolle, S.J., Hsu, W. and Pruitt, R.E. (1998). Genetic analysis of organ fusion in *Arabidopsis thaliana*. Genetics, 149(2):607-619.

Lolle, S.J. and Pruitt, R.E. (1999). Epidermal cell interactions: A case for local talk. Trends in Plant Science, 4(1):14-20.

Lolle, S.J., Victor, J.L., Young, J.M. and Pruitt, R.E. (2005). Genome-wide non-mendelian inheritance of extra-genomic information in *Arabidopsis*. Nature, 434(7032):505-509.

Looser, V., Bruhlmann, B., Bumbak, F., Stenger, C., Costa, M., Camattari, A., Fotiadis, D. and Kovar, K. (2015). Cultivation strategies to enhance productivity of *Pichia pastoris*: A review. Biotechnology Advances, 33(6 Pt 2):1177-1193.

Lopez-Casado, G., Matas, A.J., Dominguez, E., Cuartero, J. and Heredia, A. (2007). Biomechanics of isolated tomato (*Solanum lycopersicum* L.) fruit cuticles: the role of the cutin matrix and polysaccharides. Journal of Experimental Botany, 58(14):3875-3883.

Lopez, C.M.R., Wetten, A.C. and Wilkinson, M.J. (2010). Progressive erosion of genetic and epigenetic variation in callus-derived cocoa (*Theobroma cacao*) plants. New Phytologist, 186(4):856-868.

Lu, S., Zhao, H., Des Marais, D.L., Parsons, E.P., Wen, X., Xu, X., Bangarusamy, D.K., Wang, G., Rowland, O., Juenger, T., Bressan, R.A. and Jenks, M.A. (2012). *Arabidopsis* ECERIFERUM9 involvement in cuticle formation and maintenance of plant water status. Plant Physiology, 159(3):930-944.

Lucht, J.M., Mauch-Mani, B., Steiner, H.Y., Metraux, J.P., Ryals, J. and Hohn, B. (2002). Pathogen stress increases somatic recombination frequency in *Arabidopsis*. Nature Genetics, 30(3):311-314.

Lue, S., Song, T., Kosma, D.K., Parsons, E.P., Rowland, O. and Jenks, M.A. (2009). *Arabidopsis CER8* encodes LONG-CHAIN ACYL-COA SYNTHETASE 1 (LACS1) that has overlapping functions with LACS2 in plant wax and cutin synthesis. Plant Journal, 59(4):553-564.

MacLeod, A.J. and Rossiter, J.T. (1986). Isolation and examination of thioglucoside glucohydrolase from seeds of *Brassica napus*. Phytochemistry, 25(5):1047-1051.

Madlung, A. and Comai, L. (2004). The effect of stress on genome regulation and structure. Annals of Botany, 94(4):481-495.

Maldonado, A.M., Doerner, P., Dixon, R.A., Lamb, C.J. and Cameron, R.K. (2002). A putative lipid transfer protein involved in systemic resistance signalling in *Arabidopsis*. Nature, 419(6905):399-403.

Malhotra, J.D. and Kaufman, R.J. (2007). Endoplasmic reticulum stress and oxidative stress: A vicious cycle or a double-edged sword? Antioxidants & Redox Signaling, 9(12):2277-2293.

Marks, M.D., Wenger, J.P., Gilding, E., Jilk, R. and Dixon, R.A. (2009). Transcriptome analysis of *Arabidopsis* wild-type and gl3-sst sim trichomes identifies four additional genes required for trichome development. Molecular Plant, 2(4):803-822.

Masumura, T., Shigemitsu, T., Morita, S. and Satoh, S. (2015). Identification of the region of rice 13 kDa prolamin essential for the formation of ER-derived protein bodies using a heterologous expression system. Bioscience Biotechnology and Biochemistry, 79(4):566-573.

Matarasso, N., Schuster, S. and Avni, A. (2005). A novel plant cysteine protease has a dual function as a regulator of 1-aminocyclopropane-1-carboxylic acid synthase gene expression. Plant Cell, 17(4):1205-1216.

Matas, A.J., Yeats, T.H., Buda, G.J., Zheng, Y., Chatterjee, S., Tohge, T., Ponnala, L., Adato, A., Aharoni, A., Stark, R., Fernie, A.R., Fei, Z., Giovannoni, J.J. and Rose, J.K. (2011). Tissue- and cell-type specific transcriptome profiling of expanding tomato fruit provides insights into metabolic and regulatory specialization and cuticle formation. Plant Cell, 23(11):3893-3910.

Matsushima, R., Hayashi, Y., Kondo, M., Shimada, T., Nishimura, M. and Hara-Nishimura, I. (2002). An endoplasmic reticulum-derived structure that is induced under stress conditions in *Arabidopsis*. Plant Physiology, 130(4):1807-1814.

Matsushima, R., Hayashi, Y., Yamada, K., Shimada, T., Nishimura, M. and Hara-Nishimura, I. (2003a). The ER body, a novel endoplasmic reticulum-derived structure in *Arabidopsis*. Plant and Cell Physiology, 44(7):661-666.

Matsushima, R., Kondo, M., Nishimura, M. and Hara-Nishimura, I. (2003b). A novel ER-derived compartment, the ER body, selectively accumulates a β-glucosidase with an ER-retention signal in *Arabidopsis*. Plant Journal, 33(3):493-502.

Matsushima, R., Fukao, Y., Nishimura, M. and Hara-Nishimura, I. (2004). *NAI1* gene encodes a basic-helix-loop-helix-type putative transcription factor that regulates the formation of an endoplasmic reticulum-derived structure, the ER body. Plant Cell, 16(6):1536-1549.

Matzke, K. and Riederer, M. (1991). A comparative study into the chemical constitution of cutins and suberins from *Picea abies* (L.) Karst., *Quercus robur* L., and *Fagus sylvatica* L. Planta, 185(2):233-245.

McNevin, J.P., Woodward, W., Hannoufa, A., Feldmann, K.A. and Lemieux, B. (1993). Isolation and characterization of *eceriferum* (*cer*) mutants induced by T-DNA insertions in *Arabidopsis thaliana*. Genome, 36(3):610-618.

Mercier, R., Jolivet, S., Vignard, J., Durand, S., Drouaud, J., Pelletier, G. and Nogué, F. (2008). Outcrossing as an explanation of the apparent unconventional genetic behavior of *Arabidopsis thaliana hth* mutants. Genetics, 180(4):2295-2297.

Millar, A.A., Clemens, S., Zachgo, S., Giblin, E.M., Taylor, D.C. and Kunst, L. (1999). *CUT1*, an *Arabidopsis* gene required for cuticular wax biosynthesis and pollen fertility, encodes a very-long-chain fatty acid condensing enzyme. Plant Cell, 11(5):825-838.

Molina, I., Bonaventure, G., Ohlrogge, J. and Pollard, M. (2006). The lipid polyester composition of *Arabidopsis thaliana* and Brassica napus seeds. Phytochemistry, 67(23):2597-2610.

Molina, I., Ohlrogge, J.B. and Pollard, M. (2008). Deposition and localization of lipid polyester in developing seeds of *Brassica napus* and *Arabidopsis thaliana*. Plant Journal, 53(3):437-449.

Molinier, J., Ries, G., Zipfel, C. and Hohn, B. (2006). Transgeneration memory of stress in plants. Nature, 442(7106):1046-1049.

Montojo, J., Zuberi, K., Rodriguez, H., Kazi, F., Wright, G., Donaldson, S.L., Morris, Q. and Bader, G.D. (2010). GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. Bioinformatics, 26(22):2927-2928.

Moore, B.D. (2004). Bifunctional and moonlighting enzymes: Lighting the way to regulatory control. Trends in Plant Science, 9(5):221-228.

Murzin, A.G. (1998). How far divergent evolution goes in proteins. Current Opinion in Structural Biology, 8(3):380-387.

Nakagawa, T., Suzuki, T., Murata, S., Nakamura, S., Hino, T., Maeo, K., Tabata, R., Kawai, T., Tanaka, K., Niwa, Y., Watanabe, Y., Nakamura, K., Kimura, T. and Ishiguro, S.

(2007). Improved gateway binary vectors: High-performance vectors for creation of fusion constructs in transgenic analysis of plants. Bioscience Biotechnology and Biochemistry, 71(8):2095-2100.

Nakamura, S., Mano, S., Tanaka, Y., Ohnishi, M., Nakamori, C., Araki, M., Niwa, T., Nishimura, M., Kaminaka, H., Nakagawa, T., Sato, Y. and Ishiguro, S. (2010). Gateway binary vectors with the bialaphos resistance gene, *bar*, as a selection marker for plant transformation. Bioscience Biotechnology and Biochemistry, 74(6):1315-1319.

Nakano, R.T., Yamada, K., Bednarek, P., Nishimura, M. and Hara-Nishimura, I. (2014). ER bodies in plants of the *Brassicales* order: Biogenesis and association with innate immunity. Frontiers in Plant Science, 5:e73.

Nandi, A., Krothapalli, K., Buseman, C.M., Li, M., Welti, R., Enyedi, A. and Shah, J. (2003). *Arabidopsis sfd* mutants affect plastidic lipid composition and suppress dwarfing, cell death, and the enhanced disease resistance phenotypes resulting from the deficiency of a fatty acid desaturase. Plant Cell, 15(10):2383-2398.

Naranjo, M.A., Forment, J., Roldan, M., Serrano, R. and Vicente, O. (2006). Overexpression of *Arabidopsis thaliana LTL1*, a salt-induced gene encoding a GDSL-motif lipase, increases salt tolerance in yeast and transgenic plants. Plant, Cell and Environment, 29(10):1890-1900.

Nawrath, C. (2006). Unraveling the complex network of cuticular structure and function. Current Opinion in Plant Biology, 9(3):281-287.

Nowacki, M., Vijayan, V., Zhou, Y., Schotanus, K., Doak, T.G. and Landweber, L.F. (2008). RNA-mediated epigenetic programming of a genome-rearrangement pathway. Nature, 451(7175):153-158.

Ogasawara, K., Yamada, K., Christeller, J.T., Kondo, M., Hatsugai, N., Hara-Nishimura, I. and Nishimura, M. (2009). Constitutive and inducible ER bodies of *Arabidopsis thaliana* accumulate distinct β-glucosidases. Plant and Cell Physiology, 50(3):480-488.

Oh, T.J., Cullis, M.A., Kunert, K., Engelborghs, I., Swennen, R. and Cullis, C.A. (2007). Genomic changes associated with somaclonal variation in banana (*Musa* spp.). Physiologia Plantarum, 129(4):766-774.

Orford, S.J. and Timmis, J.N. (2000). Expression of a lipid transfer protein gene family during cotton fibre development. Biochimica et Biophysica Acta, 1483(2):275-284.

Padham, A.K., Hopkins, M.T., Wang, T.W., McNamara, L.M., Lo, M., Richardson, L.G., Smith, M.D., Taylor, C.A. and Thompson, J.E. (2007). Characterization of a plastid triacylglycerol lipase from *Arabidopsis*. Plant Physiology, 143(3):1372-1384.

Pagnussat, G.C., Yu, H.J., Ngo, Q.A., Rajani, S., Mayalagu, S., Johnson, C.S., Capron, A., Xie, L.F., Ye, D. and Sundaresan, V. (2005). Genetic and molecular identification of genes required for female gametophyte development and function in *Arabidopsis*. Development, 132(3):603-614.

Panikashvili, D., Savaldi-Goldstein, S., Mandel, T., Yifhar, T., Franke, R.B., Hoefer, R., Schreiber, L., Chory, J. and Aharoni, A. (2007). The *Arabidopsis* DESPERADO/AtWBC11 transporter is required for cutin and wax secretion. Plant Physiology, 145(4):1345-1360.

Panikashvili, D., Shi, J.X., Schreiber, L. and Aharoni, A. (2009). The *Arabidopsis DCR* encoding a soluble BAHD acyltransferase is required for cutin polyester formation and seed hydration properties. Plant Physiology, 151(4):1773-1789.

Panikashvili, D., Shi, J.X., Schreiber, L. and Aharoni, A. (2011). The *Arabidopsis* ABCG13 transporter is required for flower cuticle secretion and patterning of the petal epidermis. New Phytologist, 190(1):113-124.

Parodi, A.J. (2000). Role of N-oligosaccharide endoplasmic reticulum processing reactions in glycoprotein folding and degradation. Biochemical Journal, 348:1-13.

Peng, P., Chan, S.W.L., Shah, G.A. and Jacobsen, S.E. (2006). Plant genetics: Increased outcrossing in hothead mutants. Nature, 443(7110):E8-E8.

Perry, J., Brachmann, A., Welham, T., Binder, A., Charpentier, M., Groth, M., Haage, K., Markmann, K., Wang, T.L. and Parniske, M. (2009). TILLING in lotus japonicus identified large allelic series for symbiosis genes and revealed a bias in functionally defective ethyl methanesulfonate alleles toward glycine replacements. Plant Physiology, 151(3):1281-1291.

Phillips, R.L., Kaeppler, S.M. and Olhoft, P. (1994). Genetic instability of plant tissue cultures: Breakdown of normal controls. Proceedings of the National Academy of Sciences, 91(12):5222-5226.

Pighin, J.A., Zheng, H., Balakshin, L.J., Goodman, I.P., Western, T.L., Jetter, R., Kunst, L. and Samuels, A.L. (2004). Plant cuticular lipid export requires an ABC transporter. Science, 306(5696):702-704.

Pineda-Krch, M. and Fagerstrom, T. (1999). On the potential for evolutionary change in meristematic cell lineages through intraorganismal selection. Journal of Evolutionary Biology, 12(4):681-688.

Pineda-Krch, M. and Lehtila, K. (2004). Costs and benefits of genetic heterogeneity within organisms. Journal of Evolutionary Biology, 17(6):1167-1177.

Pinot, F. and Beisson, F. (2011). Cytochrome P450 metabolizing fatty acids in plants: Characterization and physiological roles. FEBS Journal, 278(2):195-205.

Pollard, M., Beisson, F., Li, Y. and Ohlrogge, J.B. (2008). Building lipid barriers: Biosynthesis of cutin and suberin. Trends in Plant Science, 13(5):236-246.

Poulton, J.E. (1990). Cyanogenesis in plants. Plant Physiology, 94(2):401-405.

Pracharoenwattana, I., Zhou, W., Keech, O., Francisco, P.B., Udomchalothorn, T., Tschoep, H., Stitt, M., Gibon, Y. and Smith, S.M. (2010). *Arabidopsis* has a cytosolic fumarase required for the massive allocation of photosynthate into fumaric acid and for rapid plant growth on high nitrogen. Plant Journal, 62(5):785-795.

Pruitt, R.E., Vielle-Calzada, J.P., Ploense, S.E., Grossniklaus, U. and Lolle, S.J. (2000). *FIDDLEHEAD*, a gene required to suppress epidermal cell interactions in *Arabidopsis*, encodes a putative lipid biosynthetic enzyme. Proceedings of the National Academy of Sciences of the United States of America, 97(3):1311-1316.

Pulsifer, I.P., Kluge, S. and Rowland, O. (2012). *Arabidopsis* LONG-CHAIN ACYL-COA SYNTHETASE 1 (LACS1), LACS2, and LACS3 facilitate fatty acid uptake in yeast. Plant Physiology and Biochemistry, 51:31-39.

Queitsch, C., Sangster, T.A. and Lindquist, S. (2002). Hsp90 as a capacitor of phenotypic variation. Nature, 417(6889):618-624.

Raffaele, S., Vailleau, F., Leger, A., Joubes, J., Miersch, O., Huard, C., Blee, E., Mongrand, S., Domergue, F. and Roby, D. (2008). A MYB transcription factor regulates very-long-chain fatty acid biosynthesis for activation of the hypersensitive cell death response in *Arabidopsis*. Plant Cell, 20(3):752-767.

Rani, S.H., Krishna, T.H.A., Saha, S., Negi, A.S. and Rajasekharan, R. (2010). Defective in Cuticular Ridges (DCR) of *Arabidopsis thaliana*, a Gene Associated with Surface Cutin Formation, Encodes a Soluble Diacylglycerol Acyltransferase. Journal of Biological Chemistry, 285(49):38337-38347.

Rask, L., Andreasson, E., Ekbom, B., Eriksson, S., Pontoppidan, B. and Meijer, J. (2000). Myrosinase: Gene family evolution and herbivore defense in *Brassicaceae*. Plant Molecular Biology, 42(1):93-113.

Rasmusson, D.C. and Phillips, R.L. (1997). Plant breeding progress and genetic diversity from *de novo* variation and elevated epistasis. Crop Science, 37(2):303-310.

Rassoulzadegan, M., Grandjean, V., Gounon, P., Vincent, S., Gillot, I. and Cuzin, F. (2006). RNA-mediated non-Mendelian inheritance of an epigenetic change in the mouse. Nature, 441(7092):469-474.

Rautengarten, C., Steinhauser, D., Bussis, D., Stintzi, A., Schaller, A., Kopka, J. and Altmann, T. (2005). Inferring hypotheses on functional relationships of genes: Analysis of the *Arabidopsis thaliana* subtilase gene family. PLoS Comput Biol, 1(4):e40.

Ray, A. (2005). Plant genetics: RNA cache genome trash? Nature, 437(7055):E1-E2.

Ray, A., Lindahl, E. and Wallner, B. (2012). Improved model quality assessment using ProQ2. BMC Bioinformatics, 13:224-234.

Rayon, C., Lerouge, P. and Faye, L. (1998). The protein N-glycosylation in plants. Journal of Experimental Botany, 49(326):1463-1472.

Reina-Pinto, J.J. and Yephremov, A. (2009). Surface lipids and plant defenses. Plant Physiology and Biochemistry, 47(6):540-549.

Reina, J.J., Guerrero, C. and Heredia, A. (2007). Isolation, characterization, and localization of AgaSGNH cDNA: a new SGNH-motif plant hydrolase specific to *Agave americana* L. leaf epidermis. Journal of Experimental Botany, 58(11):2717-2731.

Rojas, C.M., Senthil-Kumar, M., Tzin, V. and Mysore, K.S. (2014). Regulation of primary plant metabolism during plant-pathogen interactions and its contribution to plant defense. Frontiers in Plant Science, 5:17-28.

Rojo, E., Zouhar, J., Carter, C., Kovaleva, V. and Raikhel, N.V. (2003). A unique mechanism for protein processing and degradation in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences of the United States of America, 100(12):7389-7394.

Rosin, F.M., Watanabe, N. and Lam, E. (2005). Moonlighting vacuolar protease: Multiple jobs for a busy protein. Trends in Plant Science, 10(11):516-518.

Roth, E.J., Frazier, B.L., Apuya, N.R. and Lark, K.G. (1989). Genetic variation in an inbred plant: Variation in tissue cultures of soybean [*Glycine max* (L.) Merill] Genetics, 121(2):359-368.

Rowland, O., Zheng, H., Hepworth, S.R., Lam, P., Jetter, R. and Kunst, L. (2006). *CER4* encodes an alcohol-forming fatty acyl-coenzyme A reductase involved in cuticular wax production in *Arabidopsis*. Plant Physiology, 142(3):866-877.

Ruberti, C. and Brandizzi, F. (2014). Conserved and plant-unique strategies for overcoming endoplasmic reticulum stress. Frontiers in Plant Science, 5:69-76.

Ruiz-Medrano, R., Xoconostle-Cazares, B. and Lucas, W.J. (1999). Phloem long-distance transport of *CmNACP* mRNA: Implications for supracellular regulation in plants. Development, 126(20):4405-4419.

Rupasinghe, S.G., Duan, H. and Schuler, M.A. (2007). Molecular definitions of fatty acid hydroxylases in *Arabidopsis thaliana*. Proteins, 68(1):279-293.

Samuels, L., Kunst, L. and Jetter, R. (2008). Sealing plant surfaces: cuticular wax formation by epidermal cells. Annual Review of Plant Biology, 59:683-707.

Sangster, T.A., Salathia, N., Lee, H.N., Watanabe, E., Schellenberg, K., Morneau, K., Wang, H., Undurraga, S., Queitsch, C. and Lindquist, S. (2008). HSP90-buffered genetic variation is common in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences of the United States of America, 105(8):2969-2974.

Satina, S. and Blakeslee, A.F. (1941). Periclinal chimeras in *Datura stramonium* in relation to development of leaf and flower. American Journal of Botany, 28(10):862-871.

Saumonneau, A., Rottier, K., Conrad, U., Popineau, Y., Gueguen, J. and Francin-Allami, M. (2011). Expression of a new chimeric protein with a highly repeated sequence in tobacco cells. Plant Cell Reports, 30(7):1289-1302.

Schaller, A., Stintzi, A. and Graff, L. (2012). Subtilases: Versatile tools for protein turnover, plant development, and interactions with the environment. Physiologia Plantarum, 145(1):52-66.

Schiml, S. and Puchta, H. (2016). Revolutionizing plant biology: multiple ways of genome engineering by CRISPR/Cas. Plant Methods, 12:8.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S. and Schmid, B. (2012). Fiji: an open-source platform for biological-image analysis. Nature Methods, 9(7):676-682.

Schmitt, F.J., Renger, G., Friedrich, T., Kreslavski, V.D., Zharmukhamedov, S.K., Los, D.A., Kuznetsov, V.V. and Allakhverdiev, S.I. (2014). Reactive oxygen species: Re-evaluation of generation, monitoring and role in stress-signaling in phototrophic organisms. Biochimica et Biophysica Acta, 1837(6):835-848.

Schneeberger, R.G. and Cullis, C.A. (1991). Specific DNA alterations associated with the environmental induction of heritable changes in flax. Genetics, 128(3):619-630.

Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nature Methods, 9(7):671-675.

Schnurr, J., Shockey, J. and Browse, J. (2004). The acyl-CoA synthetase encoded by *LACS2* is essential for normal cuticle development in *Arabidopsis*. Plant Cell, 16(3):629-642.

Schonherr, J. and Bukovac, M.J. (1973). Ion exchange properties of isolated tomato fruit cuticular membrane: Exchange capacity, nature of fixed charges and cation selectivity. Planta, 109(1):73-93.

Schreiber, L. and Riederer, M. (1996). Ecophysiology of cuticular transpiration: comparative investigation of cuticular water permeability of plant species from different habitats. Oecologia, 107(4):426-432.

Schreiber, L. (2002). Co-permeability of ^{3}H-labelled water and ^{14}C-labelled organic acids across isolated *Prunus laurocerasus* cuticles: Effect of temperature on cuticular paths of diffusion. Plant, Cell and Environment, 25(9):1087-1094.

Schreiber, L. (2005). Polar paths of diffusion across plant cuticles: New evidence for an old hypothesis. Annals of Botany, 95(7):1069-1073.

Shah, J., Kachroo, P., Nandi, A. and Klessig, D.F. (2001). A recessive mutation in the *Arabidopsis SSI2* gene confers SA- and NPR1-independent expression of PR genes and resistance against bacterial and oomycete pathogens. Plant Journal, 25(5):563-574.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research, 13(11):2498-2504.

Shao, S., Meyer, C.J., Ma, F., Peterson, C.A. and Bernards, M.A. (2007). The outermost cuticle of soybean seeds: chemical composition and function during imbibition. Journal of Experimental Botany, 58(5):1071-1082.

Sharma, M., Sharma, N.N. and Bhalla, T.C. (2005). Hydroxynitrile lyases: At the interface of biology and chemistry. Enzyme and Microbial Technology, 37(3):279-294.

Shental-Bechor, D. and Levy, Y. (2008). Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. Proceedings of the National Academy of Sciences of the United States of America, 105(24):8256-8261.

Sherameti, I., Venus, Y., Drzewiecki, C., Tripathi, S., Dan, V.M., Nitz, I., Varma, A., Grundler, F.M. and Oelmuller, R. (2008). PYK10, a β-glucosidase located in the endoplasmatic reticulum, is crucial for the beneficial interaction between *Arabidopsis thaliana* and the endophytic fungus *Piriformospora indica*. Plant Journal, 54(3):428-439.

Sieber, P., Schorderet, M., Ryser, U., Buchala, A., Kolattukudy, P., Metraux, J.P. and Nawrath, C. (2000). Transgenic *Arabidopsis* plants expressing a fungal cutinase show alterations in the structure and properties of the cuticle and postgenital organ fusions. Plant Cell, 12(5):721-738.

Siegel, B.A. and Verbeke, J.A. (1989). Diffusible factors essential for epidermal cell redifferentiaion in *Catharanthus roseus*. Science, 244(4904):580-582.

Sinclair, A.M., Trobacher, C.P., Mathur, N., Greenwood, J.S. and Mathur, J. (2009). Peroxule extension over ER-defined paths constitutes a rapid subcellular response to hydroxyl stress. Plant Journal, 59(2):231-242.

Smyth, D.R., Mrozkiewicz, M.K., McGrath, W.J., Listwan, P. and Kobe, B. (2003). Crystal structures of fusion proteins with large-affinity tags. Protein Science, 12(7):1313-1322.

Sparkes, I.A., Ketelaar, T., de Ruijter, N.C. and Hawes, C. (2009). Grab a Golgi: laser trapping of Golgi bodies reveals in vivo interactions with the endoplasmic reticulum. Traffic, 10(5):567-571.

Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D. and Futcher, B. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. Molecular Biology of the Cell, 9(12):3273-3297.

Storici, F., Bebenek, K., Kunkel, T.A., Gordenin, D.A. and Resnick, M.A. (2007). RNA-templated DNA repair. Nature, 447(7142):338-341.

Strasser, R. (2014). Biological significance of complex N-glycans in plants and their impact on plant physiology. Frontiers in Plant Science, 5:363-368.

Sturaro, M., Hartings, H., Schmelzer, E., Velasco, R., Salamini, F. and Motto, M. (2005). Cloning and characterization of *GLOSSY1*, a maize gene involved in cuticle membrane and wax production. Plant Physiology, 138(1):478-489.

Suzuki, N., Rivero, R.M., Shulaev, V., Blumwald, E. and Mittler, R. (2014). Abiotic and biotic stress combinations. New Phytologist, 203(1):32-43.

Swenson, E.S., Price, J.G., Brazelton, T. and Krause, D.S. (2007). Limitations of green fluorescent protein as a cell lineage marker. Stem Cells, 25(10):2593-2600.

Symonds, V.V., Hatlestad, G. and Lloyd, A.M. (2011). Natural allelic variation defines a role for ATMYC1: trichome cell fate determination. PLoS Genetics, 7(6):e1002069.

Szymkowiak, E.J. and Sussex, I.M. (1996). What chimeras can tell us about plant development. Annual Review of Plant Physiology and Plant Molecular Biology, 47:351-376.

Tacke, E., Korfhage, C., Michel, D., Maddaloni, M., Motto, M., Lanzini, S., Salamini, F. and Doring, H.P. (1995). Transposon tagging of the maize *GLOSSY2* locus with the transposable element *En/Spm*. Plant Journal, 8(6):907-917.

Takahashi, K., Shimada, T., Kondo, M., Tamai, A., Mori, M., Nishimura, M. and Hara-Nishimura, I. (2010). Ectopic expression of an esterase, which is a candidate for the unidentified plant cutinase, causes cuticular defects in *Arabidopsis thaliana*. Plant and Cell Physiology, 51(1):123-131.

Tan, T.C., Spadiut, O., Wongnate, T., Sucharitakul, J., Krondorfer, I., Sygmund, C., Haltrich, D., Chaiyen, P., Peterbauer, C.K. and Divne, C. (2013). The 1.6 Å crystal structure of pyranose dehydrogenase from *Agaricus meleagris* rationalizes substrate specificity and reveals a flavin intermediate. PLoS One, 8(1):e53567.

Tanaka, H., Onouchi, H., Kondo, M., Hara-Nishimura, I., Nishimura, M., Machida, C. and Machida, Y. (2001). A subtilisin-like serine protease is required for epidermal surface formation in *Arabidopsis* embryos and juvenile plants. Development, 128(23):4681-4689.

Tanaka, H., Watanabe, M., Sasabe, M., Hiroe, T., Tanaka, T., Tsukaya, H., Ikezaki, M., Machida, C. and Machida, Y. (2007). Novel receptor-like kinase ALE2 controls shoot development by specifying epidermis in *Arabidopsis*. Development, 134(9):1643-1652.

Tang, C.L., Toomajian, C., Sherman-Broyles, S., Plagnol, V., Guo, Y.L., Hu, T.T., Clark, R.M., Nasrallah, J.B., Weigel, D. and Nordborg, M. (2007a). The evolution of selfing in *Arabidopsis thaliana*. Science, 317(5841):1070-1072.

Tang, D., Simonich, M.T. and Innes, R.W. (2007b). Mutations in *LACS2*, a long-chain acyl-coenzyme a synthetase, enhance susceptibility to avirulent *Pseudomonas syringae* but confer resistance to *Botrytis cinerea* in *Arabidopsis*. Plant Physiology, 144(2):1093-1103.

Tanz, S.K., Castleden, I., Small, I.D. and Millar, A.H. (2013). Fluorescent protein tagging as a tool to define the subcellular distribution of proteins in plants. Frontiers in Plant Science, 4:214-222.

Tenaillon, M.I., Hollister, J.D. and Gaut, B.S. (2010). A triptych of the evolution of plant transposable elements. Trends in Plant Science, 15(8):471-478.

Terpe, K. (2003). Overview of tag protein fusions: From molecular and biochemical fundamentals to commercial systems. Applied Microbiology and Biotechnology, 60(5):523-533.

Thomson, J.D., Herre, E.A., Hamrick, J.L. and Stone, J.L. (1991). Genetic mosaics in strangler fig trees: Implications for tropical conservation. Science, 254(5035):1214-1216.

Todd, J., Post-Beittenmiller, D. and Jaworski, J.G. (1999). KCS1 encodes a fatty acid elongase 3-ketoacyl-CoA synthase affecting wax biosynthesis in *Arabidopsis thaliana*. Plant Journal, 17(2):119-130.

Todesco, M., Balasubramanian, S., Hu, T.T., Traw, M.B., Horton, M., Epple, P., Kuhns, C., Sureshkumar, S., Schwartz, C., Lanz, C., Laitinen, R.A.E., Huang, Y., Chory, J., Lipka, V., Borevitz, J.O., Dangl, J.L., Bergelson, J., Nordborg, M. and Weigel, D. (2010). Natural allelic variation underlying a major fitness tradeoff in *Arabidopsis thaliana*. Nature, 465(7298):632-636.

Tornero, P., Conejero, V. and Vera, P. (1997). Identification of a new pathogen-induced member of the subtilisin-like processing protease family from plants. Journal of Biological Chemistry, 272(22):14412-14419.

Torrance, J.W., Bartlett, G.J., Porter, C.T. and Thornton, J.M. (2005). Using a library of structural templates to recognise catalytic sites and explore their evolution in homologous families. Journal of Molecular Biology, 347:565-581.

Tracy, W.F., Talbert, L.E. and Gerdes, J.T. (2000). Molecular variation and F1 performance among strains of the sweet corn inbred P39. Crop Science, 40(6):1763-1768.

Trenkamp, S., Martin, W. and Tietjen, K. (2004). Specific and differential inhibition of very-long-chain fatty acid elongases from *Arabidopsis thaliana* by different herbicides. Proceedings of the National Academy of Sciences of the United States of America, 101(32):11903-11908.

Truernit, E. and Haseloff, J. (2008). *Arabidopsis thaliana* outer ovule integument morphogenesis: ectopic expression of *KNAT1* reveals a compensation mechanism. BMC Plant Biology, 8:35-49.

Tsuwamoto, R., Fukuoka, H. and Takahata, Y. (2008). *GASSHO1* and *GASSHO2* encoding a putative leucine-rich repeat transmembrane-type receptor kinase are essential for the normal development of the epidermal surface in *Arabidopsis* embryos. Plant Journal, 54(1):30-42.

van Beilen, J.B., Eggink, G., Enequist, H., Bos, R. and Witholt, B. (1992). DNA sequence determination and functional characterization of the OCT-plasmid-encoded *alkJKL* genes of *Pseudomonas oleovorans*. Molecular Microbiology, 6(21):3121-3136.

van Zanten, M., Snoek, L.B., Proveniers, M.C. and Peeters, A.J. (2009). The many functions of ERECTA. Trends in Plant Science, 14(4):214-218.

Vanhanen, S., West, M., Kroon, J.T., Lindner, N., Casey, J., Cheng, Q., Elborough, K.M. and Slabas, A.R. (2000). A consensus sequence for long-chain fatty-acid alcohol oxidases from *Candida* identifies a family of genes involved in lipid ω-oxidation in yeast with homologues in plants and bacteria. Journal of Biological Chemistry, 275(6):4445-4452.

Verbeke, J.A. and Walker, D.B. (1986). Morphogenetic factors controlling differentiation and dedifferentiation of epidermal cells in the gynoecium of *Catharanthus roseus*: II. Diffusible morphogens. Planta, 168(1):43-49.

Vijayaraghavan, M.R., Seth, N. and Jan, A. (1988). Ultrastructure and histochemistry of angiosperm embryosac: An overview. Proceedings of the Indian Natural Science Academy, 54(1):93-110.

Vogg, G., Fischer, S., Leide, J., Emmanuel, E., Jetter, R., Levy, A.A. and Riederer, M. (2004). Tomato fruit cuticular waxes and their effects on transpiration barrier properties: functional characterization of a mutant deficient in a very-long-chain fatty acid β-ketoacyl-CoA synthase. Journal of Experimental Botany, 55(401):1401-1410.

Voisin, D., Nawrath, C., Kurdyukov, S., Franke, R.B., Reina-Pinto, J.J., Efremova, N., Will, I., Schreiber, L. and Yephremov, A. (2009). Dissection of the complex phenotype in cuticular mutants of *Arabidopsis* reveals a role of SERRATE as a mediator. PLoS Genetics, 5(10):e1000703.

Walck, J. and Dixon, K. (2009). Time to future-proof plants in storage. Nature, 462(7274):721-721.

Walley, J.W., Kliebenstein, D.J., Bostock, R.M. and Dehesh, K. (2013). Fatty acids and early detection of pathogens. Current Opinion in Plant Biology, 16(4):520-526.

Wang, Z. and Benning, C. (2012). Chloroplast lipid synthesis and lipid trafficking through ER–plastid membrane contact sites. Biochemical Society Transactions, 40(2):457-463.

Watanabe, M., Tanaka, H., Watanabe, D., Machida, C. and Machida, Y. (2004). The ACR4 receptor-like kinase is required for surface formation of epidermis-related tissues in *Arabidopsis thaliana*. Plant Journal, 39(3):298-308.

Wehner, E.P., Rao, E. and Brendel, M. (1993). Molecular structure and genetic regulation of *SFA*, a gene responsible for resistance to formaldehyde in *Saccharomyces cerevisiae*, and characterization of its protein product. Molecular and General Genetics, 237(3):351-358.

Wellesen, K., Durst, F., Pinot, F., Benveniste, I., Nettesheim, K., Wisman, E., Steiner-Lange, S., Saedler, H. and Yephremov, A. (2001). Functional analysis of the *LACERATA* gene of *Arabidopsis* provides evidence for different roles of fatty acid ω-hydroxylation in development. Proceedings of the National Academy of Sciences of the United States of America, 98(17):9694-9699.

Weng, H., Molina, I., Shockey, J. and Browse, J. (2010). Organ fusion and defective cuticle function in a *lacs1 lacs2* double mutant of *Arabidopsis*. Planta, 231(5):1089-1100.

Western, T.L., Skinner, D.J. and Haughn, G.W. (2000). Differentiation of mucilage secretory cells of the *Arabidopsis* seed coat. Plant Physiology, 122(2):345-356.

Wheeler, B.S. (2013). Small RNAs, big impact: small RNA pathways in transposon control and their effect on the host stress response. Chromosome Research, 21(6-7):587-600.

Whitham, T.G. and Slobodchikoff, C.N. (1981). Evolution by individuals, plant-herbivore interactions, and mosaics of genetic-variability: The adaptive significance of somatic mutations in plants. Oecologia, 49(3):287-292.

Whittington, H., Kerry-Williams, S., Bidgood, K., Dodsworth, N., Peberdy, J., Dobson, M., Hinchliffe, E. and Ballance, D.J. (1990). Expression of the *Aspergillus niger* glucose oxidase gene in *A. niger*, *A. nidulans* and *Saccharomyces cerevisiae*. Current Genetics, 18(6):531-536.

Williams, E.J. and Bowles, D.J. (2004). Coexpression of neighboring genes in the genome of *Arabidopsis thaliana*. Genome Research, 14(6):1060-1067.

Windsor, J.B., Symonds, V.V., Mendenhall, J. and Lloyd, A.M. (2000). *Arabidopsis* seed coat development: morphological differentiation of the outer integument. Plant Journal, 22(6):483-493.

Wlodawer, A., Minor, W., Dauter, Z. and Jaskolski, M. (2013). Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. The FEBS journal, 280(22):5705-5736.

Xia, Y., Gao, Q.-M., Yu, K., Lapchyk, L., Navarre, D., Hildebrand, D., Kachroo, A. and Kachroo, P. (2009). An intact cuticle in distal tissues is essential for the induction of systemic acquired resistance in plants. Cell Host and Microbe, 5(2):151-165.

Xiang, C.B., Han, P., Lutziger, I., Wang, K. and Oliver, D.J. (1999). A mini binary vector series for plant transformation. Plant Molecular Biology, 40(4):711-717.

Xiang, L., Etxeberria, E. and Van den Ende, W. (2013). Vacuolar protein sorting mechanisms in plants. FEBS Journal, 280(4):979-993.

Xiao, F.M., Goodwin, S.M., Xiao, Y.M., Sun, Z.Y., Baker, D., Tang, X.Y., Jenks, M.A. and Zhou, J.M. (2004). *Arabidopsis* CYP86A2 represses *Pseudomonas syringae* type III genes and is required for cuticle development. EMBO Journal, 23(14):2903-2913.

Xing, J. and Chin, C.-K. (2000). Modification of fatty acids in eggplant affects its resistance to *Verticilliumdahliae*. Physiological and Molecular Plant Pathology, 56(5):217-225.

Xu, P.Z., Yuan, S., Li, Y., Zhang, H.Y., Wang, X.D., Lin, H.H. and Wu, X.J. (2007). Genome-wide high-frequency non-Mendelian loss of heterozygosity in rice. Genome, 50(3):297-302.

Yang, S., Johnston, N., Talideh, E., Mitchell, S., Jeffree, C., Goodrich, J. and Ingram, G. (2008). The endosperm-specific *ZHOUPI* gene of *Arabidopsis thaliana* regulates endosperm breakdown and embryonic epidermal development. Development, 135(21):3501-3509.

Yang, W.L. and Bernards, M.A. (2006). Wound-induced metabolism in potato (*Solanum tuberosum*) tubers: Biosynthesis of aliphatic domain monomers. Plant Signaling & Behavior, 1(2):59-66.

Yao, Y. and Kovalchuk, I. (2011). Abiotic stress leads to somatic and heritable changes in homologous recombination frequency, point mutation frequency and microsatellite stability in *Arabidopsis* plants. Mutation Research, 707(1-2):61-66.

Yates, J.L., Boerma, H.R. and Fasoula, V.A. (2012). SSR-marker analysis of the intracultivar phenotypic variation discovered within 3 soybean cultivars. Journal of Heredity, 103(4):570-578.

Yeats, T.H., Martin, L.B., Viart, H.M., Isaacson, T., He, Y., Zhao, L., Matas, A.J., Buda, G.J., Domozych, D.S., Clausen, M.H. and Rose, J.K. (2012). The identification of cutin synthase: Formation of the plant polyester cutin. Nature Chemical Biology, 8(7):609-611.

Yemm, R.S. and Poulton, J.E. (1986). Isolation and characterization of multiple forms of mandelonitrile lyase from mature black cherry (*Prunus serotina* Ehrh.) seeds. Archives of Biochemistry and Biophysics, 247(2):440-445.

Yephremov, A., Wisman, E., Huijser, P., Huijser, C., Wellesen, K. and Saedler, H. (1999). Characterization of the *FIDDLEHEAD* gene of *Arabidopsis* reveals a link between adhesion response and cell differentiation in the epidermis. Plant Cell, 11(11):2187-2201.

Yi, H. and Richards, E. (2008). Phenotypic instability of *Arabidopsis* alleles affecting a disease *Resistance* gene cluster. BMC Plant Biology, 8:36.

Yi, H. and Richards, E. (2009). Gene duplication and hypermutation of the pathogen *Resistance* gene *SNC1* in the *Arabidopsis bal* variant. Genetics, 183(4):1227-1234.

Youngson, N.A. and Whitelaw, E. (2008). Transgenerational epigenetic effects. Annual Review of Genomics and Human Genetics, 9:233-257.

Yu, D., Ranathunge, K., Huang, H., Pei, Z., Franke, R., Schreiber, L. and He, C. (2008). *WAX CRYSTAL-SPARSE LEAF1* encodes a β-ketoacyl CoA synthase involved in biosynthesis of cuticular waxes on rice leaf. Planta, 228(4):675-685.

Yue, Q.K., Kass, I.J., Sampson, N.S. and Vrielink, A. (1999). Crystal structure determination of cholesterol oxidase from *Streptomyces* and structural characterization of key active site mutants. Biochemistry, 38(14):4277-4286.

Zhang, J.-Y., Broeckling, C.D., Sumner, L.W. and Wang, Z.-Y. (2007). Heterologous expression of two *Medicago truncatula* putative ERF transcription factor genes, *WXP1* and *WXP2*, in *Arabidopsis* led to increased leaf wax accumulation and improved drought tolerance, but differential response in freezing tolerance. Plant Molecular Biology, 64(3):265-278.

Zhao, L., Katavic, V., Li, F., Haughn, G.W. and Kunst, L. (2010). Insertional mutant analysis reveals that long-chain acyl-CoA synthetase 1 (LACS1), but not LACS8, functionally overlaps with LACS9 in *Arabidopsis* seed oil biosynthesis. The Plant Journal, 64(6):1048-1058.

Zheng, H., Rowland, O. and Kunst, L. (2005). Disruptions of the *Arabidopsis* enoyl-CoA reductase gene reveal an essential role for very-long-chain fatty acid synthesis in cell expansion during plant morphogenesis. Plant Cell, 17(5):1467-1481.

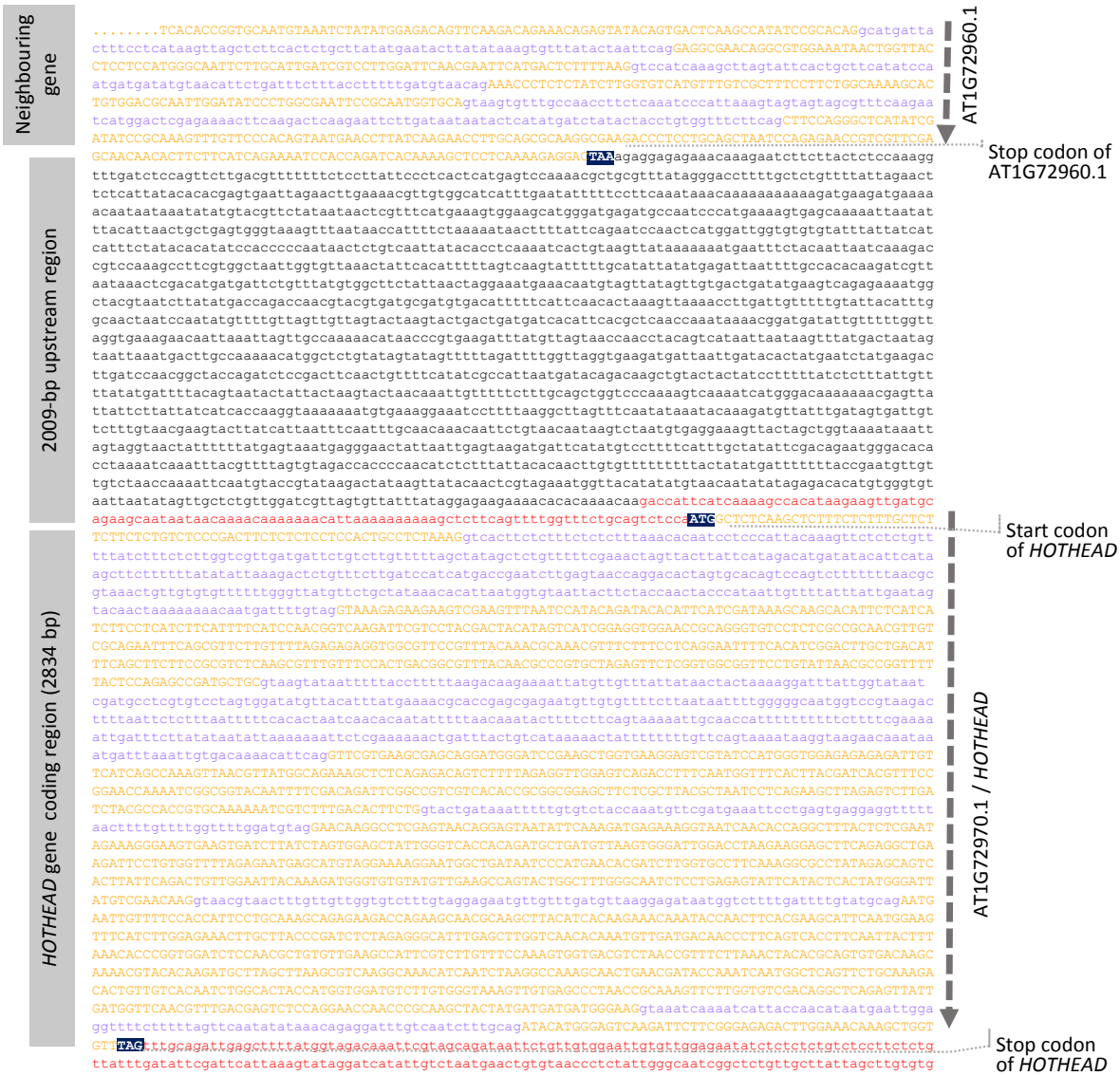
Zhou, F., Roy, B., Dunlap, J.R., Enganti, R. and von Arnim, A.G. (2014). Translational control of *Arabidopsis* meristem stability and organogenesis by the eukaryotic translation factor eIF3h. PLoS One, 9(4):e95396.

Zwieniecki, M.A., Melcher, P.J. and Michele-Holbrook, N.M. (2001). Hydrogel control of xylem hydraulic resistance in plants. Science, 291(5506):1059-1062.

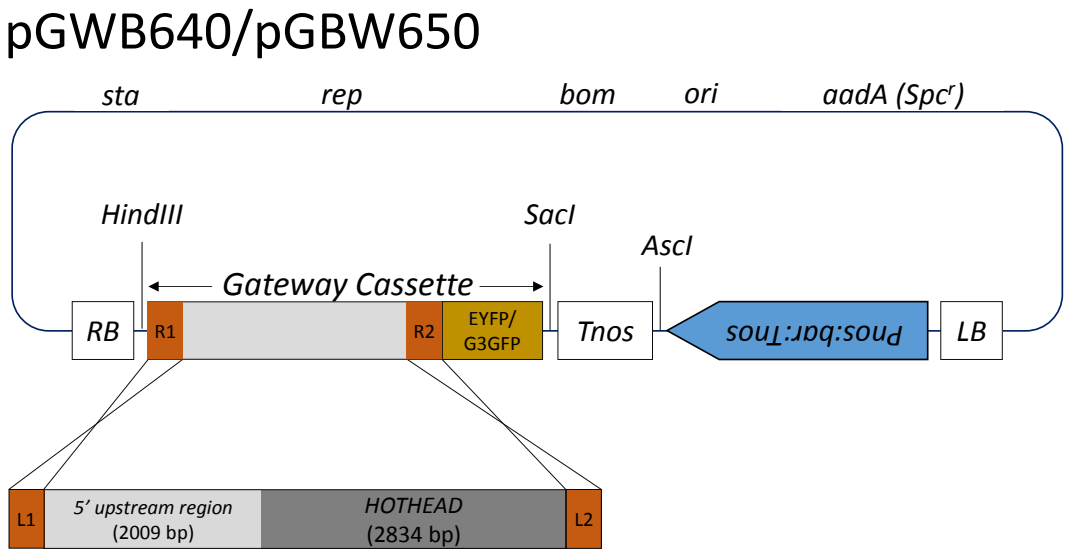
Appendices

Appendix A. The 5’ upstream region and the genomic sequence of *HOTHEAD* (Chapter 3 Materials and Methods). Source: http://www.arabidopsis.org

ATG = Translational Start/Stop	atgc = UTR
ATGC = Exon	atgc = Intron



Appendix B. Gateway binary vectors pGWB640 and pGWB650(Chapter 3 Materials and Methods).

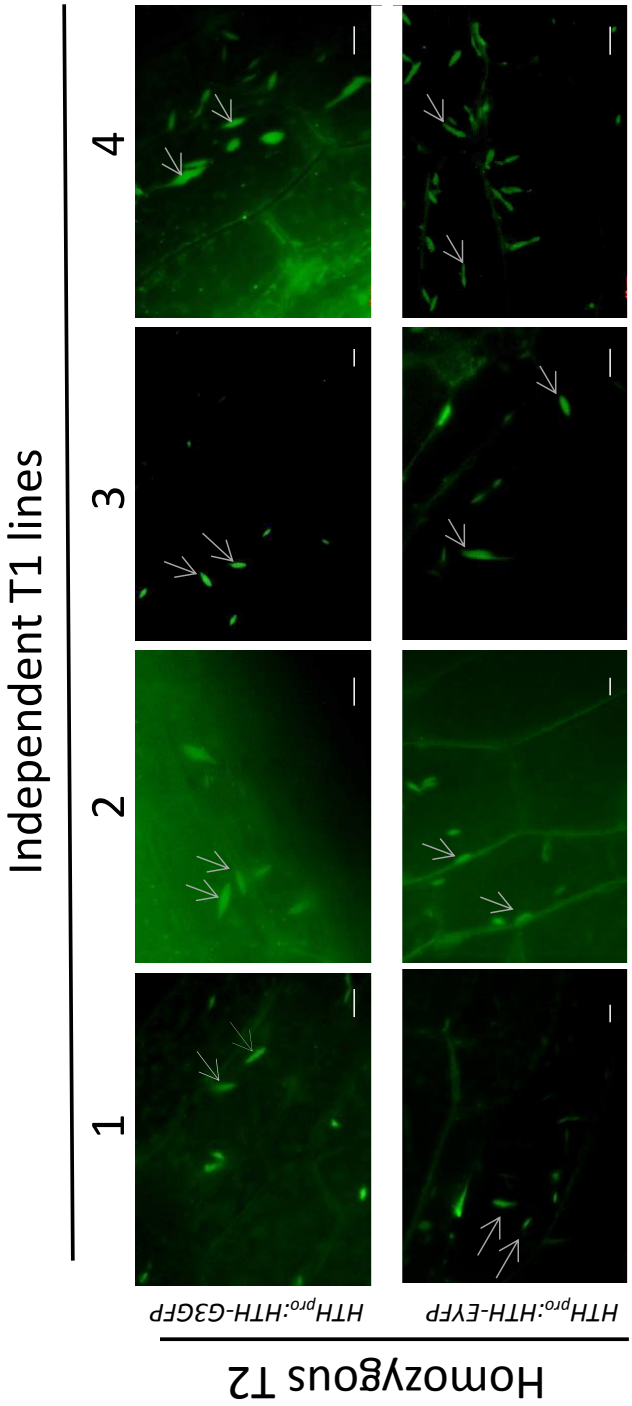


Vector name	Bacterial selection	Gateway cassette	Markers for plant	Type	Accession Number
pGWB640	Spc^r	*att*R1-Cm^r-*ccd*B-*att*R2-EYFP-T_{NOS}	P_{NOS}:bar ($BASTA^r$)	no pro, C-EYFP	AB543141
pGWB650	Spc^r	*att*R1-Cm^r-*ccd*B-*att*R2-G3GFP-T_{NOS}	P_{NOS}:bar ($BASTA^r$)	no pro, C-G3GFP	AB543147

RB, right boarder; *LB*, left boarder; *Pnos*, promoter of the nopaline synthase gene*; Tnos*, terminator of nopaline synthase*; bar*, bialaphos resistance gene*; sta*, region conferring stability in *Agrobacterium tumefaciens*; *rep*, broad host-range replication origin; *bom*, *cis*-acting element for conjugational transfer; *ori*, ColE1 recoplication origin; *addA*, gene for spectinomycin resistance (Spc^r) used for selection in the bacteria; EYFP, enhanced yellow fluorescent protein; G3GFP, G3 green fluorescent protein; Cmr, chloramphenicol resistance gene, ; ccdB, a lethal gene that targets DNA gyrase; L1, L2, R1, R2: Gateway attL and attR recombination sites for sequence exchange (derived from Nakagawa et al. (2007).

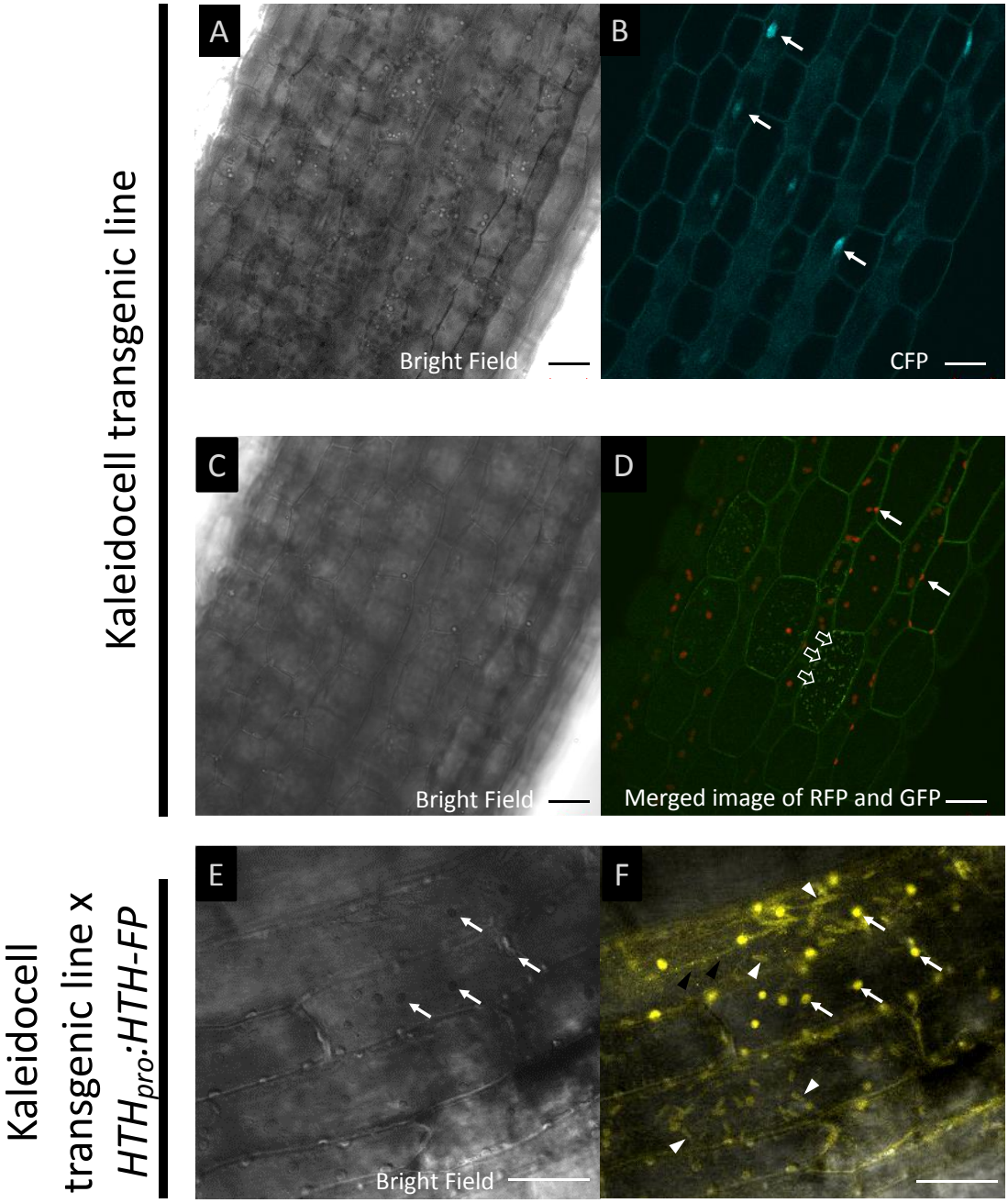
Source: Nakagawa, T., Suzuki, T., Murata, S., Nakamura, S., Hino, T., Maeo, K., Tabata, R., Kawai, T., Tanaka, K., Niwa, Y., Watanabe, Y., Nakamura, K., Kimura, T. and Ishiguro, S. (2007). "Improved gateway binary vectors: High-performance vectors for creation of fusion constructs in Transgenic analysis of plants." Bioscience Biotechnology and Biochemistry 71(8): 2095-2100.

Appendix C. Epifluorescence micrographs showing cellular localization of HTH-FP in seedlings descended from four independent T1 lines. Plant materials are cotyledons of T3 seedlings derived from homozygous (transgene) *HTH_{pro}:HTH-EYFP* and *HTH_{pro}:HTH-G3GFP* T2 plants. Epidermal cells of the hypocotyl were examined. These transgenic lines are in the *hth-9* background, and all these transgenic T2 plants showed rescued floral phenotypes that resemble wildtype. ER bodies (arrow) were observed in all of these independent transgenic rescued lines. Scale bar: 10 µm.


Independent T1 lines
1
2
3
4
Homozygous T2
HTHpro:HTH-EYFP
HTHpro:HTH-G3GFP

Appendix D. Confocal micrographs showing organelles labelled by fluorescent proteins in hypocotyls of 4-day-old seedlings of transgenic plants. In the Kaleidocell transgenic line, nuclei, plastids, mitochondria are genetically tagged with different fluorescent proteins; cyan fluorescent protein (CFP) for nuclei (A-B, arrow), red (RFP) for plastids (arrow) and green (GFP) for mitochondria (open arrow) (C-D). To examine if HTH-FP is colocalized with any of these organelles, crosses were conducted to generate plants harbouring transgenes of both Kaleidocell and HTH_{pro}:*HTH-FP* lines (E-F). Due to difficulties separating the signal wavelengths, plastids, mitochondria, endoplasmic reticulum (ER) and ER bodies were all detected at the same time using a setting that allows detection for both RFP and GFP. Based on images A-D, the bright round-shaped organelles were identified as plastids (arrow) which can also be seen in E. Spindle-shaped ER bodies were also detected (arrowhead). Small speckles were mitochondria (solid arrowhead). Excitation (ex) and emission (em) wavelengths (nm): (B, D) CFP, ex 458, em 475 - 525; RFP, ex 543, em 560 - 615; GFP, ex 488, em 505 - 530. (F) ex 514, em 530 - 560. Scale bar: 20 µm. For information about the Kaleidocell transgenic line, see Kato et al. (2008).

These results showed that the spindle-shaped bodies detected in *HTH*$_{pro}$*:HTH-FP* plants are not nuclei or plastids according to the shape and size.



Appendix E. RT-qPCR analyses of *HTH* and *HTH-FP* expression in Ws wildtype and HTH_{pro}*:HTH-FP* 4-day-old seedlings. (A) The detected *HTH* expression in Ws and the calculated *HTH-FP* level of a transgenic line (HTH_{pro}*:HTH-EYFP*). The primers used to quantify expression do not distinguish the transgene from *hth-9* or wildtype *HTH*. Numbers 1-3 indicate three biological replicates, each sampled at a different time. Three technical repeats were performed. Error bar: 1 standard error of the technical repeats. (B) The coefficient variances and M values of housekeeping genes used to normalize expression.

Materials and methods:

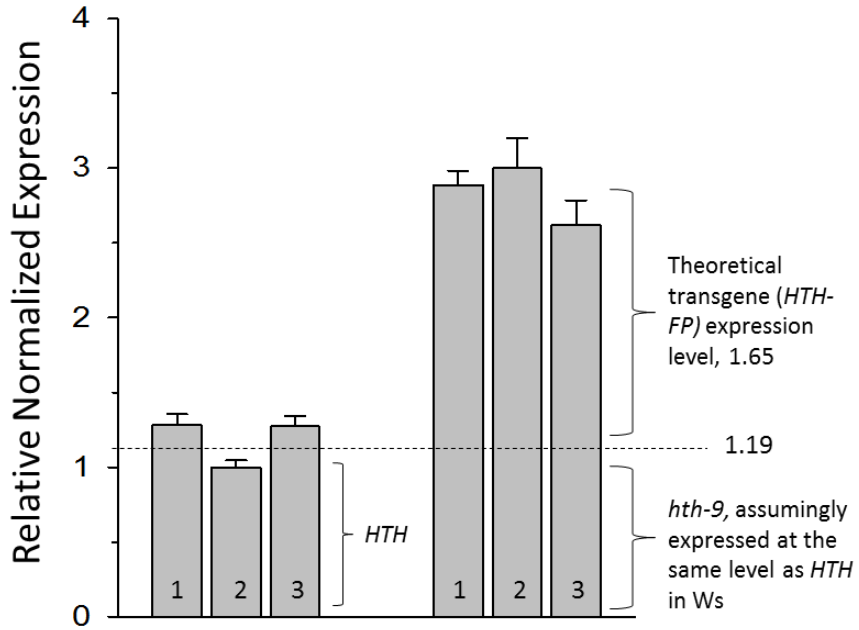
To amplify the cDNA of *HTH* transcripts, the forward primer QHTH-201_F (5'-GAGAGGTGGCGTTCCGTTTA-3') and reverse primer QHTH-201_R (5'-TTCACGAACGCAGCATCGG-3') were used. Procedures described in Section 3.2.8 were used to perform reverse transcription and quantitative PCR for this experiment.

Note:

This experiment was carried out to determine whether the *HTH* transcript level is comparable between the wild type and the native promoter driven transgenic line (HTH_{pro}*:HTH-FP*, homozygous). Similar expression levels would lend strength to the accuracy of HTH-FP localization results (such as in ER bodies, see Chapter 3). Since the primers used in RT-qPCR detect both *hth-9* and the transgene transcripts, the *HTH* expression level in homozygous HTH_{pro}*:HTH-FP* (in the *hth-9* background) is attributed to both kinds of transcripts.

Assuming *hth-9* and *HTH* (Ws) were expressed at the same level (dash line), the calculated expression level of the transgene alone was 1.65 (2.84 - 1.19 = 1.65), being 1.4 times of the wildtype gene. This RT-qPCR result suggests that the native promoter-driven transgene was not excessively overexpressed and therefore HTH-FP fusion protein localization reported in Chapter 3 is reflective of the native HTH protein.

A



Plant Material		**Ws**	***HTH$_{pro}$:HTH-FP***
Transcript(s) detected		*HTH*	*hth9* and *HTH-FP*
		Expression Level	
Biological Replicates	1	1.28	2.89
	2	1.00	3.00
	3	1.27	2.62
Average		1.19 ± 0.09	2.84 ± 0.11

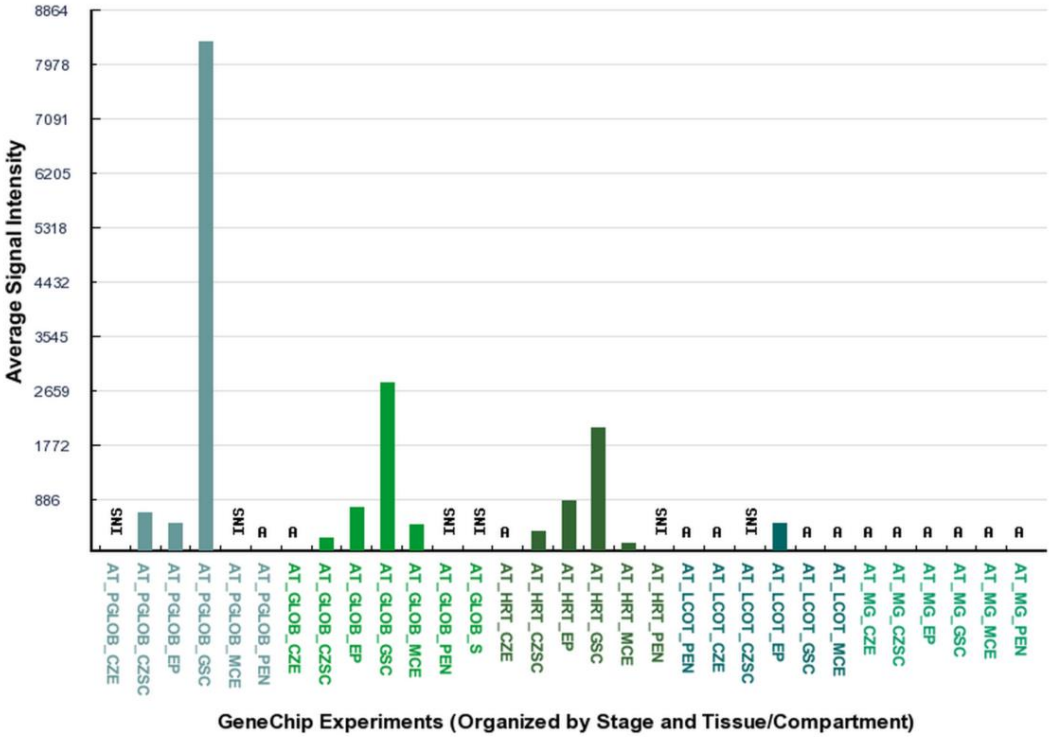
B

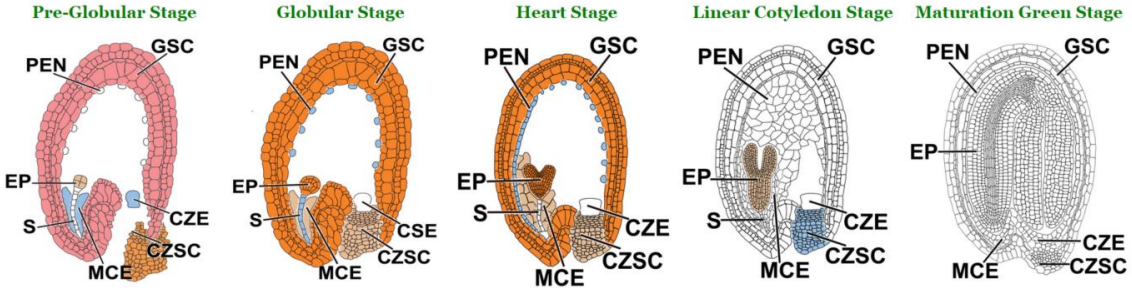
House keeping genes used for normalization

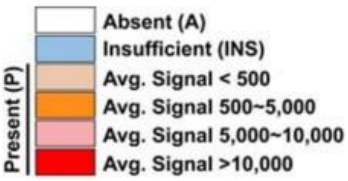
	Coeffient Variance	M Value
Tubulin 6	0.2722	0.6322
Actin 7	0.1379	0.3818
GAPC 2	0.1816	0.4133
Average	0.1972 **(< 0.5*)**	0.4758 **(< 1*)**

*Acceptable stability values for heterogeneous samples.

Appendix F. The *HOTHEAD* expression in the seed predicted by GeneChip Expression Profile (www.seedgenenetwork.net). CZE, chalazal endosperm; CZSC, chalazal seed coat; EP, embryo proper; GSC, general seed coat; MCE, micropylar endosperm; PEN, peripheral endosperm; S, suspensor.

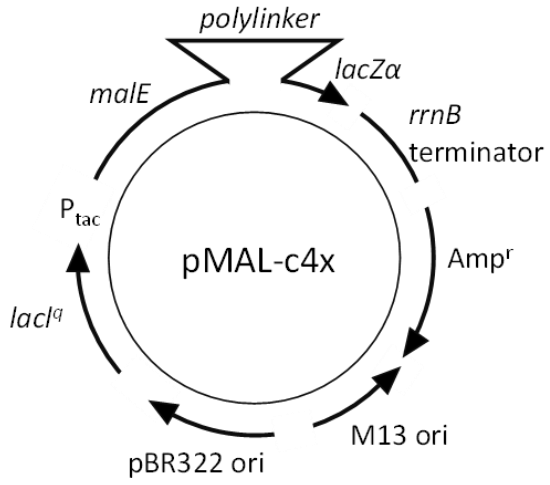

Average Signal Intensity
8864
7978
7091
6205
5318
4432
3545
2659
1772
886
INS
A
AT_PGLOB_CZE
AT_PGLOB_CZSC
AT_PGLOB_EP
AT_PGLOB_GSC
AT_PGLOB_MCE
AT_PGLOB_PEN
AT_GLOB_CZE
AT_GLOB_CZSC
AT_GLOB_EP
AT_GLOB_GSC
AT_GLOB_MCE
AT_GLOB_PEN
AT_GLOB_S
AT_HRT_CZE
AT_HRT_CZSC
AT_HRT_EP
AT_HRT_GSC
AT_HRT_MCE
AT_HRT_PEN
AT_LCOT_PEN
AT_LCOT_CZE
AT_LCOT_CZSC
AT_LCOT_EP
AT_LCOT_GSC
AT_LCOT_MCE
AT_MG_CZE
AT_MG_CZSC
AT_MG_EP
AT_MG_GSC
AT_MG_MCE
AT_MG_PEN
GeneChip Experiments (Organized by Stage and Tissue/Compartment)


Pre-Globular Stage
Globular Stage
Heart Stage
Linear Cotyledon Stage
Maturation Green Stage
PEN
GSC
EP
S
CZE
CSE
CZSC
MCE


Absent (A)
Insufficient (INS)
Avg. Signal < 500
Avg. Signal 500~5,000
Avg. Signal 5,000~10,000
Avg. Signal >10,000
Present (P)

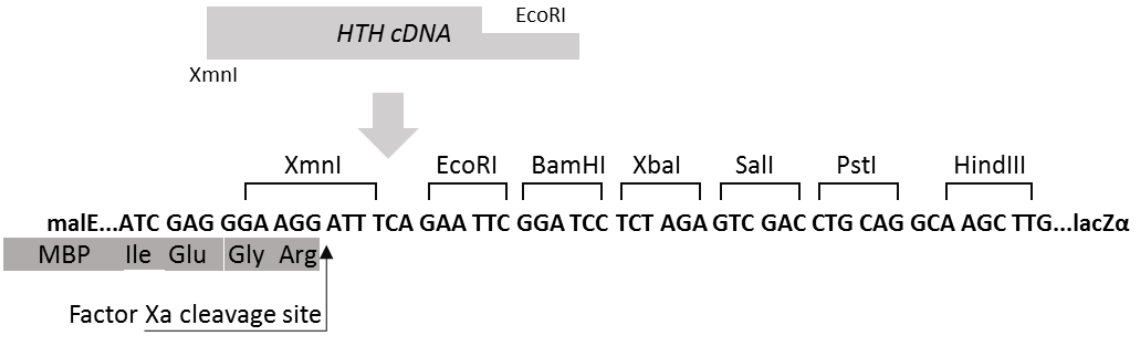
Appendix G. The pMAL-c4x vector and insertion site of *HTH* cDNA (Chapter 4 Materials and Methods). (A) The pMAL-c4x vector. lac^q, transcription promoter; P_{tac}, transcription promoter; *malE*, a gene encoding maltose binding protein binding; polylinker, multiple cloning site; *lacZα*, *rrnB* terminator, rrnB T1T2 transcriptional terminator; the β-galactosidase gene; Amp^r, β-lactamase gene confering ampicillin resistance; M13 ori, M13 origin of replication; pBR322 ori, pMB1 origin of replication. (B) The *HTH* cDNA contains not sequence coding for the putative signal peptide. The cDNA product was digested with XmnI and EcoRI for directional insertion into the linearized pMAL-c4x vector that was digested with XmnI and EcoRI. Factor Xa cleaves after its four amino acid recognition sequence (Ile-Glu-Glu-Arg). Figures are derived from www.neb.com.

A



B

HTH cDNA cloned into the polylinker of a pMAL-c4x vector



Appendix H. Predicted protein sequence of HOTHEAD (HTH). Source:

http://www.arabidopsis.org

>HOTHEAD (AT1G72970.1), 594 aa, 65.3 kDa
MALKLFLFALLLCLPTSLSSTASKGKEKKSKFNPYRYTFIDKASTFSSSSSSSFSSNGQDSSYD
YIVIGGGTAGCPLAATLSQNFSVLVLERGGVPFTNANVSFLRNFHIGLADISASSASQAFVSTD
GVYNARARVLGGGSCINAGFYSRADAAFVKRAGWDPKLVKESYPWVEREIVHQPKLTLWQ
KALRDSLLEVGVRPFNGFTYDHVSGTKIGGTIFDRFGRRHTAAELLAYANPQKLRVLIYATV
QKIVFDTSGTRPRVTGVIFKDEKGNQHQALLSNRKGSEVILSSGAIGSPQMLMLSGIGPKKEL
QRLKIPVVLENEHVGKGMADNPMNTILVPSKAPIEQSLIQTVGITKMGVYVEASTGFGQSPES
IHTHYGIMSNKNELFSTIPAKQRRPEATQAYITRNKYQLHEAFNGSFILEKLAYPISRGHLSLV
NTNVDDNPSVTFNYFKHPVDLQRCVEAIRLVSKVVTSNRFLNYTQCDKQNVHKMLSLSVKA
NINLRPKQLNDTKSMAQFCKDTVVTIWHYHGGCLVGKVVSPNRKVLGVDRLRVIDGSTFDE
SPGTNPQATMMMMGRYMGVKILRERLGNKAGV

>HOTHEAD (AT1G72970.2), splice variant, 567 aa, 62.2 kDa
MALKLFLFALLLCLPTSLSSTASKGKEKKSKFNPYRYTFIDKASTFSSSSSSSFSSNGQDSSYD
YIVIGGGTAGCPLAATLSQNFSVLVLERGGVPFTNANVSFLRNFHIGLADISASSASQAFVSTD
GVYNARARVLGGGSCINAGFYSRADAAFVKRAGWDPKLVKESYPWVEREIVHQPKLTLWQ
KALRDSLLEVGVRPFNGFTYDHVSGTKIGGTIFDRFGRRHTAAELLAYANPQKLRVLIYATV
QKIVFDTSGTRPRVTGVIFKDEKGNQHQALLSNRKGSEVILSSGAIGSPQMLMLSGIGPKKEL
QRLKIPVVLENEHVGKGMADNPMNTILVPSKAPIEQSLIQTVGITKMGVYVEASTGFGQSPES
IHTHYGIMSNKNELFSTIPAKQRRPEATQAYITRNKYQLHEAFNGSFILEKLAYPISRGHLSLV
NTNVDDNPSVTFNYFKHPCDKQNVHKMLSLSVKANINLRPKQLNDTKSMAQFCKDTVVTI
WHYHGGCLVGKVVSPNRKVLGVDRLRVIDGSTFDESPGTNPQATMMMMGRYMGVKILRE
RLGNKAGV

*VDLQRCVEAIRLVSKVVTSNRFLNYTQ: missing in AT1G72970.2

Appendix I. Sequence alignment used to construct the phylogenetic tree, and to compare putative functional residues and identify putative active sites among glucose-methanol-choline (GMC) oxidoreductases. The putative active sites are indicated by blue boxes (on page 4/4).

(page 1/4)

```
                  1
AtHTH_Q9S746      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
AtGMC1_F4KEQ5     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
AtGMC2_Q94KD2     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
AtGMC3_Q93ZK1     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
AtGMC4_Q66GI5     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
AtGMC5_Q9XI68     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
PdMDL2_Q945K2     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
PsMDL1_P52706     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
PdMDL1_O24243     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
PsMDL5_O82435     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
PsMDL4_O82784     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
PsMDL2_O50048     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
PsMDL3_P52707     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
PpADH_Q9WWW2      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
PoADH_Q00593      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
ToADH_M5DPH3      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
NrADH_PW0229..    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
AmADH_PW0207..    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
PeAAO_O94219      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
AmPDH_Q3L245      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
AtFAO3_Q9LW56     ---------- ---MDKYKVA GKFGLPDITV AEMESLASFC EAVLPSVQPP PEELSGE--G DNHRNKEALR SFYSTSGSKT PVLRQSIELV
AtFAO1_Q9ZWB9     MVGGRRLGKR GSPLLRWSVK QESFSHGFSK SDLQALSSIC DAIMPPVPLE SLNLEMK--L KVLRNDALLS FFKSSSSESH VRPDEVAELL
AtFAO4B_Q94BP3    -MEDVRRRNR GHPLLRSKKR GEGYNHGFSP SQIQSLAVIC QTFLPPETT- ---------- ---SEQQAVN SFHVASSTQP PFTDEVAEMI
AtFAO4A_O65709    ---------- ---------- ---------- --MESLVAIC DTFISSIDDS GV-------- --GHVDDCVA GYFSASASQT GTPDRVARLM
PaGOX_P81156      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
AnGOX_P13006      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
CcFAO1_Q9P8D8     ---------- ---------- ---MSHQVED HDLDVFCLLA DAVLHEIPPS EIVEYLHPDF PKDKIEEYLT GFSR--PSAV PQFRQCAKKL
CcFAO2_Q9P8D7     ---------- ---------- ---MNPVVED SHLDVFCLLA DAVVHEIPPS EIVEYLHPDF PKDKVEEYLA EFSH--PSAI PEFREVAKRI
CtFAOT_Q9P8D9     ---------- ---------M ASFLPDKVDY KNVDTLLLLC DGIIHETTVD EIRNVIDPNF PEDKYEEYVK TFTK--PSQT PGFRETIYDI
SsCHOX_P12676     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

                 91
AtHTH_Q9S746      ---------- ---------- ---------- ---------- ---------- ---------- ------MALK LFLFALLLCL PTSLS-----
AtGMC1_F4KEQ5     -----MTI-- ---------- ---------- ---------- ---------- ---------- -----ISCIT IHIFRNIL-F AILIF-----
AtGMC2_Q94KD2     ---------- ---------- ---------- ---------- ---------- ---------- -----MGLQI LWLFRKIF-I ITYIF-----
AtGMC3_Q93ZK1     ---------- ---------- ---------- ---------- ---------- ---------- -----MNFHI FRFFQFIL-V AVFIF-----
AtGMC4_Q66GI5     ---------- ---------- ---------- ---------- ---------- ---------- ------MMDR FWSWRLFVAL SLFLH-----
AtGMC5_Q9XI68     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
PdMDL2_Q945K2     MEKSTMSA-- ---------- ---------- ---------- ---------- ---------- -----IL-LV LYIFVLHLQY SEVHS-----
PsMDL1_P52706     MEKSTMSA-- ---------- ---------- ---------- ---------- ---------- -----IL-LV LHLFVLLLQY SEVHS-----
PdMDL1_O24243     MEKSTMSV-- ---------- ---------- ---------- ---------- ---------- -----IL-FV LHLLVLHLQY SEVHS-----
PsMDL5_O82435     MEKSTMSV-- ---------- ---------- ---------- ---------- ---------- -----IL-FV LHLLVLHLQY SRVHS-----
PsMDL4_O82784     MEKSTMSA-- ---------- ---------- ---------- ---------- ---------- -----VV-LV LNLLVLHLQY SEVHS-----
PsMDL2_O50048     MVKSTMSA-- ---------- ---------- ---------- ---------- ---------- -----ILVLA LHLFVLHLQY SEVQS-----
PsMDL3_P52707     MVKSTMSA-- ---------- ---------- ---------- ---------- ---------- -----VL-LV LHIFVLHLQY SEVQS-----
PpADH_Q9WWW2      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
PoADH_Q00593      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
ToADH_M5DPH3      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
NrADH_PW0229..    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
AmADH_PW0207..    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
PeAAO_O94219      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
AmPDH_Q3L245      ---------- ---------- ---------- ---------- ---------- ---------- ---------- --------ML PRV-------
AtFAO3_Q9LW56     TKRGTIEAYI ATRLILFLLA TRLGTLLICG TECLVS-RWP FVEKFSELSL EKRERVLQKQ F-KNWILTPI RAAFVYIKVA FLFCFFSRVN
AtFAO1_Q9ZWB9     ATKAIPLTVL VVRIVLRILT FRLGTLLLCG LVCLDKKHWP FLLKFSEMSL EKREKVLQRW N-TQWYNPLA RIGFMMIKAI FLFYYFTWTN
AtFAO4B_Q94BP3    VKNGRSEAVK VLRIILMILS FRFGTLLLCG SLCLDK-SWP FVLKFSQLPL DKREAILRNW SRQSGFLLPF RITFFLAKFY TLFYFFSQTD
AtFAO4A_O65709    SERLHHPKKW ILRAGLWLLS TWIGSLVLCG WRSFTG-EFP YFRRFCRLPE KRREEILLNW S--SSYFSLL RMLFRTIKLI SALVFFTQVD
PaGOX_P81156      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
AnGOX_P13006      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
CcFAO1_Q9P8D8     INRGSELSIK LF----LYLT TALDSRILAP ALT------N SLTLIRDMDL SQREELLRSW R--DSPLTAK RRLFRVYASF TLSTF----N
CcFAO2_Q9P8D7     INKGTVLSIK LF----LLLA TALDSRILAP ALT------N STTLIRDMDL SQREELLRSW R--DSPFTTK RKLFRVYNSF TLNAF----S
CtFAOT_Q9P8D9     INANTTDAIH SF----IILT TALDSRILAP TLT------N SLTPIREMSL PERERLLASW R--DSPIAAK RRLFRLVSSL TLSTF----V
SsCHOX_P12676     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

                181
AtHTH_Q9S746      -------STA SKGK------ ---------- ----EKKSKF NPYRYTFIDK A-------ST F----SS--- -----SSSSS FSSNGQDSSY
AtGMC1_F4KEQ5     -------YG- SCYC------ ---------- ----DKGTIL ISYKYNLILV YIS---IFFF FLVKLET--- -----TFMKD ATLAPKNASF
AtGMC2_Q94KD2     -------HA- SCSP------ ---------- ----DKAG-- ---------- ---------- -----YY--- -----SFMKD ATLAPMFARF
AtGMC3_Q93ZK1     -------HD- SCYS------ ---------- ----DEAG-- ---------- ---------- -----NY--- -----RFMKD ATLAPKLSHF
AtGMC4_Q66GI5     -------SP- ICSS------ ---------- ----DKAP-- ---------- ---------- -----NY--- -----SFMRD ATGSPTTSYY
AtGMC5_Q9XI68     ---------- ---------- ---------- ---------- ---------- ---------- ------M--- -----PYMTT DAKEVSGKSF
PdMDL2_Q945K2     -------LA- TTS------- ---------- ----DHDF-- ---------- ---------- ----SYL--- -----SFAYD ATDLELEGSY
PsMDL1_P52706     -------LA- TTS------- ---------- ----NHDF-- ---------- ---------- ----SYL--- -----RFAYD ATDLELEGSY
PdMDL1_O24243     -------LA- NTS------- ---------- ----AHDF-- ---------- ---------- ----SYL--- -----KFVYN ATDTSLEGSY
PsMDL5_O82435     -------LA- NTS------- ---------- ----DHDF-- ---------- ---------- ----SYL--- -----KFVYN ATNSSLEGSY
PsMDL4_O82784     -------LA- NTSS------ ---------- ----EHDF-- ---------- ---------- ----GYL--- -----KFVYN AVDLELEGSY
PsMDL2_O50048     -------LA- NTS------- ---------- ----AHDF-- ---------- ---------- ----SYL--- -----EFVYD ANDTELEGTY
PsMDL3_P52707     -------LA- NTS------- ---------- ----SHDF-- ---------- ---------- ----SYL--- -----SFVYD ATDPELEGSY
PpADH_Q9WWW2      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- --------MY
PoADH_Q00593      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- --------MY
ToADH_M5DPH3      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- --------MY
NrADH_PW0229..    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---MSDQKTF
AmADH_PW0207..    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- --MSEVLSKY
PeAAO_O94219      ---------- ---------- ---------- ---------- ----MSFGAL RQL------- LLIACLA--- -----LPSLA ATN-LPTADF
AmPDH_Q3L245      ---------- ---------- ---------- -------TKL NSRLLSLALL GIQ------- --IARGA--- -----ITYQH PDDLPSGVDY
AtFAO3_Q9LW56     PNGENPAWEA IGYRVNPDEN K--------- ---PSETHNE RPLEKGIVET MEETEQTLLE SLAHKGLEAV L--------D TEHDAIRIKC
AtFAO1_Q9ZWB9     ENSENPVWDA INYSVEIGEN E--------- ----DMEQKE RPLDEGIIET AKEDEMTIKQ RMINKGLKVT E--------D RERDTYKIEC
AtFAO4B_Q94BP3    ENLKNPALEA IGYCIDGTER SSN------- -KKSEADEKR RPLEKGIIET MHESDVTITQ SLTEKGVHVA R--------D DGDNVYRIRC
AtFAO4A_O65709    EKGRNLAWKA IGYNGPSPDH SDHEVELNEE KKKKKPEEIF GPLYNGIVDL KSPR-EAVEK KLAGRGFAVS NQKRNTNGSS ISDPVMKIQC
PaGOX_P81156      ---------- ---------- ---------- ---------- -------YLP AQQ------- IDVQ------ -----SSLLS DPSKVAGKTY
AnGOX_P13006      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------T DPKDVSGRTV
CcFAO1_Q9P8D8     KLGTDLHFKA LGYPGRELRT QI-------- -----QDYEV DPFRY----- ------SFME KLK------- ---------H EGHELFLPDI
CcFAO2_Q9P8D7     KTATDLHFKA LGYPGRELRT QI-------- -----QDYEV DPFRY----- ------TFLE KPQ------- ---------Q DGQELHFPDI
CtFAOT_Q9P8D9     RLASDLHLKA IHYPERDLRE KA-------- ----YESQEI DPFKY----- ------SFEE KPK------- ---------F DGAESYLPDI
SsCHOX_P12676     ---------- ---------- ---------- --MTAQQHLS RRRMLGMA-- ------AFGA AALAGGTTIA APRAAA--AA KSAADNGGYV
```

(continued, page 2/4)

```
                271
AtHTH_Q9S746    DYIVIGGGTA GCPLAATLSQ N--FSVLVLE RGGVPFTN-A NVSFLRNFHI GLADISA--- ---------- ---------- -SSASQAFVS
AtGMC1_F4KEQ5   DYIIIGGGTA GCALAATLSQ N--ASVLVLE RGGSPYEN-P TATDMGNSVN TLLNNTP--- ---------- ---------- -NSWSQLFIS
AtGMC2_Q94KD2   DYIIIGGGTS GCALAATLSQ N--ASVLVLE RGGAPYDN-P TATDIENFAT TLSNTSP--- ---------- ---------- -KSWSQLFIS
AtGMC3_Q93ZK1   DYIIIGGGTA GCALAATLSQ N--ATVLVLE RGGSPYDD-P AATDIGNFAN TLLNITP--- ---------- ---------- -NSWSQLFIS
AtGMC4_Q66GI5   DYIIIGGGTA GCPLAATLSQ N--ASVLLLE RGDSPYNN-P NITRLSAFGA ALSDLSE--- ---------- ---------- -SSPSQRFVS
AtGMC5_Q9XI68   DYIVVGGGTA GCSLAATLSE K--YSVLVIE RGGSPFGD-P LVEERKYFGY SLLNTDE--- ---------- ---------- YSSVAQSFTS
PdMDL2_Q945K2   DYVIVGGGTS GCPLAATLSE K--YKVLVLE RGSLPTAY-P NVLTADGFVY NLQQEDD--- ---------- ---------- GKTPVERFVS
PsMDL1_P52706   DYVIVGGGTS GCPLAATLSE K--YKVLVLE RGSLPTAY-P NVLTADGFVY NLQQEDD--- ---------- ---------- GKTPVERFVS
PdMDL1_O24243   DYIVIGGGTS GCPLAATLSE K--YKVLLLE RGTIATEY-P NTLTADGFAY NLQQQDD--- ---------- ---------- GKTPVERFVS
PsMDL5_O82435   DYIVVGGGTS GCPLAATLSE K--YKVLLLE RGTIATEY-P NTLTADGFAY NLQQQDD--- ---------- ---------- GKTPVERFVS
PsMDL4_O82784   DYIIVGGGTS GCPLAATLSA N--YSVLVLE RGTIATEY-P NTLTVDGFAY NLQQQDD--- ---------- ---------- GKTPVERFVS
PsMDL2_O50048   DYIIVGGGTA GCPLAATLSA N--YSVLVLE RGTLPTEY-P NLLTSDGFIY NLQQEDD--- ---------- ---------- GQTPVERFVS
PsMDL3_P52707   DYIIVGGGTA GCPLAATLSA N--YSVLVLE RGSLPTEY-P NLLISDGFVY NLQQEDD--- ---------- ---------- GKTPVERFVS
PpADH_Q9WWW2    DYIIVGAGSA GCVLANRLSA DPSKRVCLLE AGPRDTN--P LIHMPLGI-A LLSNSKKL-- ---------N WA-------- FQTAPQQHLN
PoADH_Q00593    DYIIVGAGSA GCVLANRLSA DPSKRVCLLE AGPRDTN--P LIHMPLGI-A LLSNSKKL-- ---------N WA-------- FQTAPQQNLN
ToADH_M5DPH3    DYIIVGAGSA GCVLANRLSA DSSKRVCLLE AGPADKN--P FIHMPIGI-A LLANNKTL-- ---------N WA-------- FNTEKQNKLN
NrADH_PW0229..  DFIIAGGGSA GCVLANRLSA SGKWKVCLIE AGPADSS--P FIHMPAGI-I PVVRSKIL-- ---------N WN-------- FWTAPQANCG
AmADH_PW0207..  DYIIIGGGSA GAVLATRLSE NPALDILLLE AGSKDTN--P LIHIPFGL-S VLSRFEGI-- ---------G WG-------- YHTAPQKELY
PeAAO_O94219    DYVVVGAGNA GNVVAARLTE DPDVSVLVLE AGVSDEN--V LGAEAPLLAP GLVPNSIF-- ---------D WN-------- YTTTAQAGYN
AmPDH_Q3L245    DFIVAGGGTA GLVVASRLSE NSNWKVLVIE AGPSNKD--A FVTRVPGLAS TLGAGSPI-- ---------D WN-------- YTTIPQDGLD
AtFAO3_Q9LW56   DVVVVGSGSG GGVAASVLAK S-GLKVVVLE KGSYFTPSEH RPFEGPGLDK ------LY-- ---------- ---------- ENGGILP-SV
AtFAO1_Q9ZWB9   DAVVVGSGCG GGVAAAILAK S-GLRVVVIE KGNYFAPRDY SALEGPSMFE ------LF-- ---------- ---------- ESNSLMM-TH
AtFAO4B_Q94BP3  DAVVVGSGSG GGVAAANLAK A-GLKVLVLE KGNYFTAHDY SGLEVPSMLE ------LY-- ---------- ---------- EKGGLLT-TV
AtFAO4A_O65709  DAVVVGSGSG GGVAAGVLAK A-GYKVLVIE SGNYYARSKL SLLEGQAMDD ------MY-- ---------- ---------- LSGGLLA-TS
PaGOX_P81156    DYIIAGGGLT GLTVAAKLTE NPKIKVLVIE KGFYESNDGA IIEDPNA--Y GQIFGTTV-- ---------D QN-------- YLTVPLIN--
AnGOX_P13006    DYIIAGGGLT GLTTAARLTE NPNISVLVIE SGSYESDRGP IIEDLNA--Y GDIFGSSV-- ---------D HA-------- YETVELAT-N
CcFAO1_Q9P8D8   DVLIIGSGSG AGVVAQTLTE S-GLKSLVLE KGKYFASEEL CMTDLDGNEA ------LF-- ---------- ---------- ESGGTIP-ST
CcFAO2_Q9P8D7   DVLIIGSGSG AGVVAQTLSE N-GLKSLVLE KGKYFSNDEL TMNDLEGSEA ------LF-- ---------- ---------- ENGGALS-ST
CtFAOT_Q9P8D9   DVIIIGSGAG AGVVAHTLAN D-GYKTLVLE KGRYFSNLEL NFNDKDGVQE ------LY-- ---------- ---------- QGGGALT-TT
SsCHOX_P12676   PAVVIGTGYG AAVSALRLGE A-GVQTLMLE MGQLWNQPGP DGNIFCGMLN PDKRSSWFKN RTEAPLGSFL WLDVVNRNID PYAGVLDRVN

                361
AtHTH_Q9S746    TDGVYNARAR VLGGGSCINA GFYSRADAAF VKR-----A- -GWDPKLVKE S--------- -----YPWVE REIVHQ---- -------P--
AtGMC1_F4KEQ5   EDGVYNTRPR VLGGGSVING GFYSRAGNDY VEE-----A- -EWEMEEVEA A--------- -----YEWVE KKLVFE---- -------P--
AtGMC2_Q94KD2   EDGVYNTRAR VLGGGSVLNA GFYTRAGDEY VKE-----T- -EWKTDEVEA A--------- -----YEWVE KKVAFQ---- -------P--
AtGMC3_Q93ZK1   EDGVFNSRAR VLGGGTVINA GFYSRAEEDF VAE-----A- -GWERDEVEA A--------- -----YEWVE KKVVFE---- -------P--
AtGMC4_Q66GI5   EDGVINARAR VLGGGSALNA GFYTRAGTKY VRN-----M- -GWDGALANE S--------- -----YQWVE AKVAFQ---- -------P--
AtGMC5_Q9XI68   VDGIENYRGR VLGGSSAING GFYSRASDEF VKK-----A- -GWDKGLVQE S--------- -----YKWVE SKVVFM---- -------P--
PdMDL2_Q945K2   EDGIDNVRGR VLGGTSIINA GVYARANTSI YSA-----SG VDWDMDLVNQ T--------- -----YEWVE DTIVYK---- -------P--
PsMDL1_P52706   EDGIDNVRGR VLGGTSMINA GVYARANTSI YSA-----SG VDWDMDLVNK T--------- -----YEWVE DTIVFK---- -------P--
PdMDL1_O24243   EDGIDNVRAR ILGGTTIINA GVYARANISF YSQ-----TG IEWDLDLVNK T--------- -----YEWVE DAIVVK---- -------P--
PsMDL5_O82435   EDGIDNVRGR ILGGTTIINA GVYARANISY YNQ-----TG IEWDLDLVNK T--------- -----YEWVE DTIVVK---- -------P--
PsMDL4_O82784   EDGIDNVRSR ILGGTTIINA GVYARANESF YNN-----SG VEWDLDLVNE A--------- -----YEWVE DAIVYK---- -------P--
PsMDL2_O50048   GDGIDNVRGR VLGGTSMINA GVYVRANTSF FNQ-----TG IEWDMDLVNK T--------- -----YDWVE DTIVFK---- -------P--
PsMDL3_P52707   EDGIDNVRGR VLGGTSMINA GVYVRANTSF FNQ-----TG IEWDMDLVNQ T--------- -----YEWVE DTIVFE---- -------P--
PpADH_Q9WWW2    ERSLFWPRGK TLGGSSSINA MVYIRGHEED YQAWEQA-GG EYWGWKRAFA L--------- -----FKKLE HNQRFD---- ------KSNY
PoADH_Q00593    GRSLFWPRGK TLGGSSSINA MVYIRGHEDD YHAWEQA-AG RYWGWYRALE L--------- -----FKRLE CNQRFD---- ------KSEH
ToADH_M5DPH3    NRELFWPRGK TLGGSSSINA MVYIRGHEKD FDEWQKS-AG NNWSWDRALT L--------- -----FKRLE KNERFG---- ------DSQY
NrADH_PW0229..  NRQMYWPRGR TLGGSSSINA MCYVRGHAWD YDHWASL-GN EGWSYKDVLP Y--------- -----FRRME NFEAIDK--- --VD-DGRSY
AmADH_PW0207..  DRELFWPRGK TLGGSSSVNA MCYIRGQKED YDRWASEEGA EGWSFDDVLP Y--------- -----FKRSE NFEEG----- ------ADEF
PeAAO_O94219    GRSIAYPRGR MLGGSSSVHY MVMMRGSTED FDRYAAVTGD EGWNWDNIQQ F--------- -----VRKNE MVVPPADNHN TSGE-FIPAV
AmPDH_Q3L245    GRSLDYPRAK ILGGCSTHNG MVYTRGSKDD WNSWAGIIGD QGLGWDSILP A--------- -----IKKAE KFTQDFTDQS VKGH-IDPSV
AtFAO3_Q9LW56   DGSFMVLAGA TVGGGSAVNW SACIKTPKSV LQEWSEDQNI PLFGTKEYLT A--------- ---MEV---- --VWKRMG-V TEK----CEL
AtFAO1_Q9ZWB9   DGRFRFMAGS TVGGGSVVNW AASLKTPDAI IEEWSVHRGI SIYSSEKYKA A--------- ---MGI---- --VCKRLG-V TEK----IIR
AtFAO4B_Q94BP3  DGKFMLLAGS AVGGGTAVNW SASIRTPDHV LQEWSEGSKI KFFGSQEYQS A--------- ---MDE---- --VTIRIG-V TER----CVK
AtFAO4A_O65709  DTNVVILAGS TVGGGSTINW SASIKTPEHV MKEWAEKSKL EMFGSDLYRE A--------- ---MDV---- --VCKRMG-V QCG----FVE
PaGOX_P81156    NRTNNIKAGK GLGGSTLING DSWTRPDKVQ IDSWEKVFGM EGWNWDNMFE Y--------- -----MKKAE AARTPTAAQL AAGHSFNATC
AnGOX_P13006    NQTALIRSGN GLGGSTLVNG GTWTRPHKAQ VDSWETVFGN EGWNWDNVAA Y--------- -----SLQAE RARAPNAKQI AAGHYFNASC
CcFAO1_Q9P8D8   NQQLFMIAGS TFGGGSTVNW SACLKTPFKV RKEWYDDFGL DFVATQQYDD C--------- ---MDY---- --VWKKMG-A STE----HIE
CcFAO2_Q9P8D7   NQQIFIIAGS TFGGGSTVNW SACLKTPFKV RKEWYDNFGL DFVATQYYED C--------- ---MDY---- --VWKKMG-A SNE----NID
CtFAOT_Q9P8D9   NQQMFILAGS TFGGGTTVNW SACLKTPFKV RKEWYVDYGV EFAADETYDM A--------- ---QDY---- --VWKQMG-A STE----GIT
SsCHOX_P12676   YDQMSVYVGR GVGGGSLVNG GMAVEPKRSY FEEILPR--- -VDSSEMYDR YFPRANSMLR VNHIDTKWFE DTEWYKFARV SRE----QAG

                                                                          G218       R227
                451
AtHTH_Q9S746    ---------- --KLTLWQKA LRDSLLEVGV R------PFN GFTYDHVSGT KIGGTIFDRF GRRHTAA--E LLAYANPQKL RVLIYATVQK
AtGMC1_F4KEQ5   ---------- --QVIEWQKA FKDGLLEAGE S------PDN GFTYDHIYGT KIGGTIFDRA GHRHTAA--N LLEYANPNRI VVYLHASVHK
AtGMC2_Q94KD2   ---------- --PVLGWQTA FKDGLLEAGE F------PYN GFTYDHIYGT KIGGTIFDRA GHRHTAA--D LLEYANPGNI VVYLHASVHK
AtGMC3_Q93ZK1   ---------- --PVNKWQSA FRDGLLEAGV T------PYN GFTYEHIVGT KFGGTIFDRD GHRHTAA--N LLEYANPNMI VVYLHASVHK
AtGMC4_Q66GI5   ---------- --PMGRWQTA VRDGLLEAGI V------PNN GFTYDHINGT KFGGTIFDRN GNRHTAA--D LLEYADPKGI TVLLHATVHR
AtGMC5_Q9XI68   ---------- --ELTQWQSV VQFGFLEAGF Y------PYN GYSLEHTQGT KIGGSIYDQC GKRHTSA--D LLGFGKPNCI TVLLNTTVKS
PdMDL2_Q945K2   ---------- --NSQSWQSV TKTAFLEAGV H------PNH GFSLDHEEGT RITGSTFDNK GTRHAAD--E LLNKGNSNNL RVGVHASVEK
PsMDL1_P52706   ---------- --NYQPWQSV TGTAFLEAGV D------PNH GFSLDHEAGT RITGSTFDNK GTRHAAD--E LLNKGNSNNL RVGVHASVEK
PdMDL1_O24243   ---------- --NNQSWQSV IGEGFLEAGI L------PDN GFSLDHEAGT RLTGSTFDNN GTRHAAD--E LLNKGDPNNL LVAVQASVEK
PsMDL5_O82435   ---------- --NNQSWQSV IGEAFLEAGI L------PDN GFSLDHEAGT RLTGSTFDNN GTRHAAD--E LLNKGDPNNL LVAVQASVEK
PsMDL4_O82784   ---------- --SNQSWQSI TGTAFLEAGV H------PDN GFGLVHEEGT RLTGSTFDNS GTRHASD--E LLNKGDPDNL KVAVEAAVQK
PsMDL2_O50048   ---------- --DFQFWQNL TGTAFLEVGI L------PDN GFSLDHLEGT RLTGSTFDNN GTRHASD--E LLNKGDPNNL RVAVHAAVEK
PsMDL3_P52707   ---------- --DSQTWQTV IGTAYLEAGI L------PNN GFSVDHLAGT RLTGSTFDNN GTRHASD--E LLNKGDPNNL RVAVQAAVEK
PpADH_Q9WWW2    HGTDGELAVS DLK---DLNP LSKSFVQAGM E--AKISFNG DFNGAHQEGV GFYQVTQKHG Q-RWSSARAF LHDVIDRPNL DIITEAHATK
PoADH_Q00593    HGVDGELAVS DLK---YINP LSKAFVQAGM E--ANINFNG DFNGEYQDGV GFYQVTQKNG Q-RWSSARAF LHGVLSRPNL DIITDAHASK
ToADH_M5DPH3    HGNQGELSVS ELQ---TTNP LSKAFIEAGQ Q--TGLPLND DFNGASQEGV GFYQVTQKKG R-RWSSAKAF LTGALDRPNL DIITNARATR
NrADH_PW0229..  HGSGGPLNVA EPR---YLNP LMKAFVAAGR E--AGHAVTD DFNGAQQEGV GYYHTMQKNG Q-RCSNARGF LDPVKGRDNL TVITRAHVAR
AmADH_PW0207..  HGTGGPLNVS KLR---HTSV LSETFVNSAS F--AGYKQLT DFNRDDREGI GFYHVTQANG Q-RCSTAKGY LTQAKHRNNL TVLTKVAAEK
PeAAO_O94219    HGTNGSVSIS LPG---FPTP LDDRVLATTQ EQSEEFFFNP DMGTGHPLGI SWSIASVGNG Q-RSSSSTAY LRPAQSRPNL SVLINAQVTK
AmPDH_Q3L245    HGFDGKLSVS AAY---SNIS FNDLLFETTK ELNAEFPFKL DMNDGKPIGL GWTQYTIDNH AERSSSATSY LESTG--DNV HVLVNTLVTR
AtFAO3_Q9LW56   ESFQ------ -------NQI LRKGCENLGF N-VENVPRNS SE--SHYCGS CGYGCRQGD- --KKGSDRTW LVDAVGH-GA VILTGCKAER
AtFAO1_Q9ZWB9   EGFQ------ -------NQI LRKGCEKLGL D-VTIVPRNS TE--KHYCGS CSYGCPTGE- --KRGTDSTW LVDAVNN-NA VILTQCKAEK
AtFAO4B_Q94BP3  HGFQ------ -------NQV LRKGCERLGL Q-VESVPRNS PE--DHYCGL CGYGCRAGA- --KNGTDQTW LVDAVEN-GA VILTGIKAER
AtFAO4A_O65709  EGFN------ -------NEV LRKGCEKLGL P-VKNIPRNA PS--DHYCGF CCLGCKKGQ- --KQGTSETW LVDLVESDNG LILPGCQATE
PaGOX_P81156    HGTNGTVQSG ARDNGQPWSP IMKALMNTVS A--LGVPVQQ DFLCGHPRGV SMIMNNLDEN QVRVDAARAW LLPNYQRSNL EILTGQMVGK
AnGOX_P13006    HGVNGTVHAG PRDTGDDYSP IVKALMSAVE D--RGVPTKK DFGCGDPHGV SMFPNTLHED QVRSDAAREW LLPNYQRPNL QVLTGQYVGK
CcFAO1_Q9P8D8   HSAA------ -------NAV IMDGAAKLGY A-HRALEQNT GGHV-HDCGM CHLGCRFGI- --KQGGVNCW FREPSEK-GS KFMEQVVVEK
CcFAO2_Q9P8D7   HSGA------ -------NSV ILEGSKKLGY P-HRAVEQNN GGKI-HDCGM CHLGCRFGI- --KQGSVNCW FRGPSEN-GS KFMQQVLVDK
CtFAOT_Q9P8D9   HSLA------ -------NNV VIEGAKKLGY K-SKPIDQNS GGHPHHPCGF CHLGCKYGI- --KQGSVNNW FRDAAAN-GS KFMQQVRVLQ
SsCHOX_P12676   KAGL------ -------GTV FVPNVYDFGY M-QREAAGEV PK--SALATE VIYGNNH-G- --KQSLDKTY LAAALGT-GK VTIQTLHQVK
```

(continued, page 3/4)

```
                541                                                 G294
AtHTH_Q9S746    IVFDTS---- ---G-TRPRV TGVIFKDEKG NQHQALLS-- NRKGSEVILS SGAIGSPQML MLSGIGPKKE LQRLKI-PVV LENEHVGKGM
AtGMC1_F4KEQ5   VLFTT----- --------EA YEVLFEDANG VFHKANLA-- NKATNEVILS AGALGSPQLL MLSGVGPAVH LEAHGVNPLV LDQPMVGQGM
AtGMC2_Q94KD2   ILFTTK---- ---GRPRPKA YGVIFQDANG VLHKAELE-- KNSMNEVILS AGAIGSPQLL MLSGIGPAAH LAAHGIKPLV LDHPMVGQGM
AtGMC3_Q93ZK1   ILFTIK---- ---GNQRPKA YGVIFLDANG VSYKAELATQ DSTMSEVILS AGAIASPQLL MLSGVGPAAH LAAYRVNPVI VDQPMVGQGM
AtGMC4_Q66GI5   ILFRTR---- ---GTTKPIA NGVVYRDRTG QAHRAYLK-- EGALSEIILS AGTLGSPQLL MLSGVGPSAQ LQAQNIT-VV MDQPHVGQGM
AtGMC5_Q9XI68   IIFDSS---- ---NKT--RA VGVRFMESDG NSSKSYKVHV EQHRGEVILA AGALGSPQIL LLSGIGPENH LKDFDI-PVI VNLKEVGRKM
PdMDL2_Q945K2   IIFS-N---- ---A-PGLTA TGVIYRDSNG TPHQAFVR-- --SKGEVIVS AGTIGTPQLL LLSGVGPESY LSSLNI-PVV LSHPYVGQFL
PsMDL1_P52706   IIFS-N---- ---A-PGLTA TGVIYRDSNG TPHRAFVR-- --SKGEVIVS AGTIGTPQLL LLSGVGPESY LSSLNI-PVV LSHPYVGQFL
PdMDL1_O24243   ILFSSN---- ---T-SNLSA IGVIYTDSDG NSHQAFVR-- --GNGEVIVS AGTIGTPQLL LLSGVGPESY LSSLNI-TVV QPNPYVGQFL
PsMDL5_O82435   ILFSSN---- ---T-SNLSA IGVMYTDSDG NSHEAFVR-- --GNGEVIIS AGTIGTPQLL LLSGIGPESY LSSLNI-TVV YPHPYVGQFV
PsMDL4_O82784   IIFSTE---- ---S-SGLTA VGVVYTDSNG TSHRALVS-- --GKGEVILS AGTLGTPQLL LLSGVGPESY LTSLNI-SVV ASHPYVGQYV
PsMDL2_O50048   IIFSSD---- ---S-SGVTA IGVIYTDSNG TTHQAFVR-- --GDGEVILS AGPIGSPQLL LLSGVGLESY LTSLNI-SVV ASHPYVGQYI
PsMDL3_P52707   IIFSSN---- ---T-SGVTA IGVIYTDSNG TTHQAFVR-- --GEGEVILS AGPIGSPQLL LLSGVGPESY LTSLNI-SVV ASHPYVGQYI
PpADH_Q9WWW2    VLFEDR---- ---KAVGVS- ----YIQKNM HQQ-VKTT-- --DSGEVILS LGAVNTPQLL MLSGVGAAAE LKEHGI-ALV HDLPEVGKNL
PoADH_Q00593    ILFEDR---- ---KAVGVS- ----YIKKNM HHQ-VKTT-- --SGGEVLLS LGAVGTPHLL MLSGVGAAAE LKEHGV-SLV HDLPEVGKNL
ToADH_M5DPH3    VIFDST---- ---TASGVA- ----YQQGNN SEQ-VMLN-- --QGGEVLLS GGAVNSPQLL MLSGIGAADE LKKHSI-SVI HNLPEVGKNL
NrADH_PW0229..  VLFEGK---- ---RAIGVR- ----YLGGGR IHD-VR---- --ASQEVILS AGAIGSPQVL LLSGVGPKAE LSKHAI-PQI HELPGVGENL
AmADH_PW0207..  VLLKEG---- ---RAIGVQ- ----VREKGA VSR-Y--F-- --AKSEVILC GGAINSPQLL MLSGVGPRNE LEDKGI-FVH QDLPGVGQNL
PeAAO_O94219    LVNSGT---- ---TNGLPAF RCVEYAEQEG APT-TTVC-- --AKKEVVLS AGSVGTPILL QLSGIGDEND LSSVGI-DTI VNNPSVGRNL
AmPDH_Q3L245    VLSAS----- ----GNGTDF RKVEFAVDAN SPK-KQLE-- --AKKEVIVA GGVIASPQIL MNSGIGERKV LQAVGI-DTL IDNPSVGKNL
AtFAO3_Q9LW56   FILEKNGSNK ---GGKQMKC LGVMAKSLNG NIAKMLKI-- --EAKVTVSA GGALLTPPLM ISSGLRNRNI GKNLHL---- --HP-VLMAW
AtFAO1_Q9ZWB9   LILADNDANK REESGRRKRC LGVAA-SLSH QTRKKLQI-- --NAKVTIVA CGSLKTPGLL ASSGLKNSNI SRGLHI---- --HP-IMMAW
AtFAO4B_Q94BP3  FVLVDNTSSS ---NERKKRC VGVFASSVGG KIGKKFII-- --EARVTVSS AGSLLTPPLM LSSGLKNPNI GRNLKL---- --HP-VLMTW
AtFAO4A_O65709  VMYDCE---- ---QGKKKKA TGVAF-AF-- -GEEIYVV-- --ESRVTIVA CGALRTPHLL KRSGLKNSNI GRNLCL---- --HP-VVMAW
PaGOX_P81156    VLFKQT---- ---A-SGPQA VGVNFGT-NK AVN-FDVF-- --AKHEVLLA AGSAISPLIL EYSGIGLKSV LDQANV-TQL LDLP-VGINM
AnGOX_P13006    VLLSQN---- ---G-TTPRA VGVEFGT-HK GNT-HNVY-- --AKHEVLLA AGSAVSPTIL EYSGIGMKSI LEPLGI-DTV VDLP-VGLNL
CcFAO1_Q9P8D8   ILQH------ ------KGKA TGILCRDTE- SGIKFKIT-- --GPKKYVVS GGSLQTPVLL QKSGFKNKHI GANLKL---- --HP-VSVAL
CcFAO2_Q9P8D7   ILQR------ ------DGKA VGVLCRDVV- TGVKFKIT-- --GPKKIVVF WWFFANSGFV TKSGFKNKHI GANLKL---- --HP-VSLTL
CtFAOT_Q9P8D9   IINK------ ------KGIA SGILCEDVA- TGVKFTIT-- --GPKKYVVA AGALNTPIVL NNSGFKNKHI GKNLTL---- --HP-VSTVF
SsCHOX_P12676   TIRQTK---- -----DGGYA LTVEQKDTDG KLLATKEI-- --SCRYLFLG AGSLGSTELL VRARDT---- GTLPNL---- -NSE-VGAGW

                631                 G356
AtHTH_Q9S746    ADNPMNTILV P--SKAPIEQ -SLIQTVGIT KMGVYVEAST GFGQSPE--- -----SIHTH YGIMSN-K-- NELFSTIP-- --AKQRRPEA
AtGMC1_F4KEQ5   ADNPMNFVAI P--SPQPVEL -SLIQAVGIT KFDSYIEGLS GLSLSF---- ----DITRRF FDGVLNLL-- NETSH----- TTSRKILTQS
AtGMC2_Q94KD2   GDNPMNAIFI P--SPTPVEV -SLIQVVGIT KFESYIEGAS GVIFSY---- ----SWTRRF FDGVLNYL-- NEMQTSRTTS TTSPTLSTQS
AtGMC3_Q93ZK1   GDNPMNPVFI P--SPEPVEV -SLVQAVGIT KFGSYIEGGS ALSLSI---- ----SLTRSF FDGVLNLL-- KKT------- ----KLPTQS
AtGMC4_Q66GI5   YDNPMNAVFV P--SPVPVEV -SLIEVVGIT GEGTYVEAAG GENFGGGGGG SSGSSSTRDY YAMFSP---- ------RA-- ---TLLESNS
AtGMC5_Q9XI68   SDNPAISLLV D--RFSQNRT LEPPQVAAIA --------E GYKFI----- ---------- ---------- ---------- ----------
PdMDL2_Q945K2   HDNPRNFINI L--PPNPIEP -TIVTVLGIS NDF-YQCSFS SLPFTT---- ---------- ---------- ---------- --------PP
PsMDL1_P52706   HDNPRNFINI L--PPNPIEP -TIVTVLGIS NDF-YQCSFS SLPFTT---- ---------- ---------- ---------- --------PP
PdMDL1_O24243   YNNPRNFINN F--PPNPIEA -SVVTVLGIR SDY-YQVSLS SLPFST---- ---------- ---------- ---------- --------PP
PsMDL5_O82435   YDNPRNFINI L--PPNPIEA -SVVTVLGIS SYY-YQISLS SLPFST---- ---------- ---------- ---------- --------PP
PsMDL4_O82784   NDNPRNFINI L--PPNPIEP -STVTVLGIT SDF-YQCSLS SLPFDT---- ---------- ---------- ---------- --------PP
PsMDL2_O50048   YDNPRNFINI L--PPNPIEA -STVTVLGIT SDF-YQCSIS SLPFST---- ---------- ---------- ---------- --------AP
PsMDL3_P52707   YDNPRNFINI L--PPNPIEA -STVTVLGIT SDF-YQCSIS SLPFDT---- ---------- ---------- ---------- --------PP
PpADH_Q9WWW2    QDHLDITLMC AANSRTPIGV AFSFIPRGLV GLFSYIFKRK GFLTSN---- ---------- -------VAE SGGFVKSSPE RDRPNLQFHF
PoADH_Q00593    QDHLDITLMC AANSREPIGV ALSFIPRGVS GLFSYVFKRE GFLTSN---- ---------- -------VAE SGGFVKSSPD RDRPNLQFHF
ToADH_M5DPH3    QDHLDITLMN TANSSLPIGV AFGAIPKSIG AVFSYIFKSK GFLTSN---- ---------- -------VAE SGGFVKSDET RDRPNVQFHF
NrADH_PW0229..  QDHLDIHITM RERTRHSISL RPLGLLKGAW GVIRYLFGRR GELTSN---- ---------- -------FAQ AGGFICSNAE QTVPDLQWHL
AmADH_PW0207..  QDHLDAIVQY TCKAREGYAV ALGALPSYVK ATADYAFKRN GIFSSN---- ---------- -------IAE AGGFVSSSLA SQGPDIQFHF
PeAAO_O94219    SDHLLLPAAF FVNSNQTFDN IFRDSSEFNV DLDQWTNTRT GPLTA----- ---LIANHLA WLRLPSNSSI FQTFPDPAAG PNSAHWETIF
AmPDH_Q3L245    SDQGATSVMF DTT------- LPSTDFDVDA ALTEWTNSHT GPLARG---- ---ARLNHLT FVRLPDDKLN ---GQDPSSG KNSPHIEFQF
AtFAO3_Q9LW56   GYFPDKESSN ISFKGNSYEG -------GII T------SVS KVLSE--DSE VR-AIIETPQ L--------G PGSF------ --SVLTPWTS
AtFAO1_Q9ZWB9   GYFPEKNS-- -ELEGAAHE- -------GEI VT-----SLH YVHPM--DST TPNITLETPA I--------G PGTF------ --AALTPWVS
AtFAO4B_Q94BP3  GYFPEKDS-- -EFSGKMYEG -------GII T------SVH HMNDT--ESG CK-AILENPL I--------G PASY------ --AGLSPWVS
AtFAO4A_O65709  GWFPEEDKWP -EKKKKSYEG -------GIM T------AMS SVVIEETHSS YGEMVIQTPA L--------H PGMF------ --SGIIPWTS
PaGOX_P81156    QDQTTTTVSS RASSAGAGQG QAVFFANFTE TFGDYAPQAR DLLNTK---- --------L- -DQWAEETVA RGGFHNVT-- --ALKVQYEN
AnGOX_P13006    QDQTTATVRS RITSAGAGQG QAAWFATFNE TFGDYSEKAH ELLNTK---- --------L- -EQWAEEAVA RGGFHNTT-- --ALLIQYEN
CcFAO1_Q9P8D8   GDFGNEVD-- ----FEAYKR -------PLM T------AVC NAVDDLD-GK AHGTRIEAIL H--------A PYVT------ --APFYPWQS
CcFAO2_Q9P8D7   GDFGNNVD-- ----FEAYRK -------PIM T------SIC NKVEDLD-GK AHGTRIEAML N--------A PYGV------ --APFFPWKS
CtFAOT_Q9P8D9   GDFGKEVQ-- ----ADHFHK -------SIM T------SLC YEVADLD-GK GHGCRIETIL N--------A PFIQ------ --ASLLPWRG
SsCHOX_P12676   GPN------- ---------G -------NIM T------ARA NHMWN--PTG AHQSSIPALG IDAWDNS--D SSVFAEIAP- MPAGLETWVS

                721                                         G435
AtHTH_Q9S746    TQAYITRNK- -------YQL HEAFN----G SFILEKLAYP ISRGHLSLVN -TNVDDNP-S VTFNYFKHPV DLQRCVEAIR LVSKVVTSNR
AtGMC1_F4KEQ5   IAVLLKSFD- -------VKL EVRMN----G GLIFQKVDGP ASKGHMKLRN -TNPRDNP-S VTFNYYQEPE DLNKCVKGLN TIIRMINSKA
AtGMC2_Q94KD2   ITDFFNPIN- -------PLL NATTR----A GLILQKIAGP ISRGHLELRN -TNPDDNP-S VRFNYYQEPE DLQICVEGIN TIIKVINSKA
AtGMC3_Q93ZK1   ISKFFKSLD- -------LTL NVTTK----A GVIIQKVNGP LSRGHLELRN -TNPDDNP-S VTFNYFKDPE DLNKCVEGLS TIIKVIDSKG
AtGMC4_Q66GI5   MT------K- -------LSS AQPFQ----G GFLLEKVMGP LSTGHLELKT -RNPKDNP-I VTFNYFQHPD DLKRCVRGIQ TIERVVQSKA
AtGMC5_Q9XI68   ----LESEV- -------LPT DITTT----R ISIAAKIAFP KSKGRLKLNS -TNPRENP-S VKFNYLENKA DLDACLEMVL HLQHVARSET
PdMDL2_Q945K2   -FGFFPSAS- -------YPL PNS-T----F AHFASKVAGP LSYGSLTLKS SSNVRVSP-N VKFNYYSNLT DLSHCVSGMK KIGELLSTDA
PsMDL1_P52706   -FSFFPSTS- -------YPL PNS-T----F AHFASKVAGP LSYGSLTLKS SSNVRVSP-N VKFNYYSNPT DLSHCVSGMK KIGELLSTDA
PdMDL1_O24243   -FSLFPTTS- -------YPL PNS-T----F AHIVSQVPGP LSHGSVTLNS SSDVRIAP-N IKFNYYSNST DLANCVSGMK KLGDLLRTKA
PsMDL5_O82435   -FSLFPTPS- -------YPL PNS-T----F AHVVSQVPGP LSYGSVTLNS SSDVRIAP-N VKFNYYSNST DLANCVSGMK KLGDVLRTKA
PsMDL4_O82784   -FSLFPTTS- -------YPL PNQ-T----F AHIVSKVPGP LSAGSLTLQS SSNVSVAP-N VKFNYCSDPV DLTHCVSGMK KIGVFLSTDA
PsMDL2_O50048   -FGFFPNPT- -------YPL PNT-T----F AHIVNKVPGP LSHGTVLLQS TSDVRVAP-N VTFNYYSNTT DLAHCVSGMK KIGEFLSSDA
PsMDL3_P52707   -FSFFPTTS- -------YPL PNQ-T----F AHIVNKVPGP LSHGTVTLNS SSDVRVGP-N VKFNYYSNLT DLSHCVSGMK KLGEVLSTDA
PpADH_Q9WWW2    LPTYLKDH-- ---GRKIAV- -----GY--G YTLHICDLLP KSRGRIGLKS -ANPMDDP-L IDPNYLSDPE DIKTMIAGIK IGRAIFDAPS
PoADH_Q00593    LPTYLKDH-- ---GRKIAG- -----GY--G YTLHICDLLP KSRGRIGLKS -ANPLQPP-L IDPNYLSDHE DIKTMIAGIK IGRAILQAPS
ToADH_M5DPH3    LPVYLNDH-- ---GRKITP- -----GY--G FTLHICDLLP KSRGFIGLKS -ADPLADP-L IDPCYLEHPD DLKTMMDAVK IARKIFRSPA
NrADH_PW0229..  VPFVYANH-- ---GQTLAPL F-R--HY--A YTLMTCFLRP ESRGRLSLRD -ADPMSAP-L IDANYLATER DMQALVEGFH KARELLAQPA
AmADH_PW0207..  LPAILNDH-- ---GRQLAF- -----GY--G YGLHVCCLYP KSRGTISLQS -NHPADQA-L IDPNYLSAAE DQQVMIEGVR IARKLLSAPD
PeAAO_O94219    SNQWFHPAI- ---------PR P-DTGSF--- MSVTNALISP VARGDIKLAT -SNPFDKP-L INPQYLSTEF DIFTMIQAVK SNLRFLSGQA
AmPDH_Q3L245    AQI--TPQVP TLGVPKQAPL P-AANSYRLL LQLAVVNLYS ISRGSISLSD -NNPFTYP-L IDLNMFKEDI DIAILREGIR SAGRMFSSKA
AtFAO3_Q9LW56   GLDM------ ----KKRMAR YSRTAS---- LI---TIVRD RGSGEVKTE- ---------- GRINYTVDKT DRDNLKAGLR ESLRILIAAG
AtFAO1_Q9ZWB9   GSDM------ ----KERMAK YARTAH---- IF---AMVRD EGVGEVKG-- ---------- DIVKYRLTKA DEENLTIGLK QALRILVAAG
AtFAO4B_Q94BP3  GPDL------ ----KERMIK YGRTAH---- LF---ALVRD LGSGEVMME- ---------- NEVTYRTTKK DRENLRAGLR QALRVSVAAG
AtFAO4A_O65709  SKDF------ ----KTRMLK FSRTAH---- IF---ALLRD KGTGTIDSK- ---------- TYIDYNLNDE DEESLKNGLE RVLKILAAAG
PaGOX_P81156    YRNWLLDED- ----VAFAEL FMDTEG---K INFDLWDLIP FTRGSVHILS -SDPYLWQFA NDPKFFLNEF DLLGQAAASK LARDLTSQGA
AnGOX_P13006    YRDWIVNHN- ----VAYSEL FLDTAG---V ASFDVWDLLP FTRGYVHILD -KDPYLHHFA YDPQYFLNEL DLLGQAAATQ LARNISNSGA
CcFAO1_Q9P8D8   GAQA------ ----RKNLLK YKQTVP---- LL---LLSRD TSSGTVTYDK -QKPD----V LVVDYTVNKF DRNSILQGFL VASDILYIEG
CcFAO2_Q9P8D7   GAES------ ----RKDLLR YKQTVP---- IL---LLSRD TTSGSVTYDK -QKPD----A LVIDYLLNKF DRNSILQGFL IASDLLYIEG
CtFAOT_Q9P8D9   SDEV------ ----RRDLLR YNNMVA---- ML---LTTRD TTSGSVSADP -KKPE----A LIVDYDVSKF DKNAILAQAFL ITSDMLHIEG
SsCHOX_P12676   L--------- ---------- ---------- YL---AITKN PQRGTFVYDA -ATDR----- AKLNWTRD-- ---------- ----------
```

(continued, page 4/4)

Active site 1 2

```
AtHTH_Q9S746     FLNYTQCDKQ NVHKMLSLSV KANINLRPKQ LNDT----KS ---MAQFCKD ---------- ---------- -TVVTIWHYH GGCLVG----
AtGMC1_F4KEQ5    FSKYKYPG-V TARELLNLML ALPINLRPRH VTSA----FN ---LKQFCID ---------- ---------- -TVTSVWHYH GGCQVG----
AtGMC2_Q94KD2    FSKFKYPD-A TIHGLLDLML SVPTNLRPRH ITSM----FN ---LRQFCID ---------- ---------- -TVMTIWHYH GGCQVG----
AtGMC3_Q93ZK1    YSKYKYPL-A SARGLLNLIL ALPTNLRPRH ITST----FD ---LEQYCID ---------- ---------- -TVMTIYHYH GGCQVG----
AtGMC4_Q66GI5    FSRYKYAD-V SFEYLLNLTA STPVNLRPPR SGPGASLPPS ---AEEFCQH ---------- ---------- -TVTTIWHYH GGCVVG----
AtGMC5_Q9XI68    VTFFLGTQ-- ---------- -----AHDKL VA----GDEE ---LKSFCIK ---------- ---------- -NVRTYYHYH GGCVVG----
PdMDL2_Q945K2    LKPYKVEDLP GVEGFNILGI PLPKDQTD-- -------DAA ---FETFCRE ---------- ---------- -SVASYWHYH GGCLVG----
PsMDL1_P52706    LKPYKVEDLP GIEGFNILGI PLPKDQTD-- -------DAA ---FETFCRE ---------- ---------- -SVASYWHYH GGCLVG----
PdMDL1_O24243    LEPYKARDVL GIDGFNYLGV PLPENQTD-- -------DAS ---FETFCLD ---------- ---------- -NVASYWHYH GGSLVG----
PsMDL5_O82435    LEPYKARDVP GIDGFNYLGV PLPENQTD-- -------DAA ---FETFCQD ---------- ---------- -NVASYWHYH GGSLVG----
PsMDL4_O82784    LKPYKVDDLP GIDGFNILGT PLPENQTD-- -------DAA ---FEKFCRD ---------- ---------- -TVASYWHYH GGAIVG----
PsMDL2_O50048    LKPYKVEDLP GIEGFDILGI PLPENQTD-- -------DAA ---FETFCRE ---------- ---------- -AVASYWHYH GGCLVG----
PsMDL3_P52707    LEPYKVEDLP GIDGFNILGI PLPENQTD-- -------DAA ---FETFCRE ---------- ---------- -SVASYWHYH GGCLVG----
PpADH_Q9WWW2     MAKHFKRE-- ---------- -----IVPG- PA--VTSDDE ---IVADIRS ---------- ---------- -RAETIYHPV GTCRMGKD--
PoADH_Q00593     MAKHFKHE-- ---------- -----VVPG- QA--VKTDDE ---IIEDIRR ---------- ---------- -RAETIYHPV GTCRMGKD--
ToADH_M5DPH3     MSAHTKCE-- ---------- -----LLPG- LD--VQTDAQ ---LEADIRK ---------- ---------- -RAETIYHPV GTCRMGAD--
NrADH_PW0229..   IARHAVDE-- ---------- -----LEPG- RA--VQSDDD ---IRAYIRA ---------- ---------- -KSETVYHPI GTCKMGRD--
AmADH_PW0207..   FDRFQGSE-- ---------- -----LYPG- EE--AQTDDE ---ILEFLRE ---------- ---------- -RAETIYHPI GTCKMGSDDD
PeAAO_O94219     WADFVIRP-- ---------- -----FDPRL RD--PTDDAA ---IESYIRD ---------- ---------- -NANTIFHPV GTASMSPRGA
AmPDH_Q3L245     FKNSVNKF-- ---------- -----VY-PP AD--ATSDED ---LDAFLRS ---------- ---------- -STFSYVHGV GTLSMSPKGA
AtFAO3_Q9LW56    AEEVGTHRSD G--------- ------QRLI CK--GVNENS ---IQEFLDS VSTEEGA--- ----KGMTEK WNVYSSAHQM GSCRI-GENE
AtFAO1_Q9ZWB9    AAEVGTYRSD G--------- ------QRMK CD--GIKQKD ---LEAFLDT VNAPPGV--- ----VSMSKH WTQSFTAHQI GCCRM-GATE
AtFAO4B_Q94BP3   AVEVGTYRSD G--------- ------QKMK CE--AITKEA ---MEEFLDE VDAVGGV--- ----GTKGEY WTTYFSAHQM GSCRM-GVTA
AtFAO4A_O65709   AEEIGTHHSE G--------- ------RSLN VR--TASSLE ---IERFVRE ESSKP----- ------LKDL SGQICSAHQM GSCRM-GIRP
PaGOX_P81156     MKEYFAGE-- ---------- -----TLPGY N---LVQNAT LSQWSDYVLQ ---------- ---------- -NFRPNWHAV SSCSMMSRE-
AnGOX_P13006     MQTYFAGE-- ---------- -----TIPGD N---LAYDAD LSAWTEYIPY ---------- ---------- -HFRPNYHGV GTCSMMPKE-
CcFAO1_Q9P8D8    AKEILS-PQA WVPTF----- ---KSNKPKH AR--SIKDED ---YVKWRET VAKIP----- ------FDSY GSPYGSAHQM SSCRMSGKGP
CcFAO2_Q9P8D7    ASRDHVTYKL GYQWF----- ---KSSKPKH AR--SIEDED ---YVNWRAK VAKIP----- ------FDSY GSPYGSAHQM STCRMSGKGP
CtFAOT_Q9P8D9    AKRILS-PQA WVPIF----- ---ESSKPRD ER--SIDDKD ---YVEWRAK AAKIP----- ------FDSY GSAYGSAHQM STCRMSGKGP
SsCHOX_P12676    ---------Q NAPAV----- ---NAAK--- ---------- -----ALFDR INKANGTIYR YDLFGTQLKA FADDFCYHPL GGCVLGKA--
```

P_{564}, G_{565}, T_{566} Active site 3

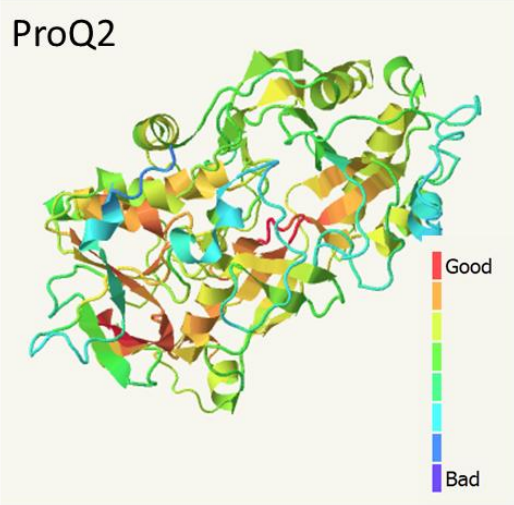
```
                 901
AtHTH_Q9S746     --KVVSPNRK VLGVDRLRVI DGSTFDESPG TNPQATMMMM GRYMGVKILR ERLGNKAGV- ---------- ---------- --
AtGMC1_F4KEQ5    --KVVDKNYK VLGIDGLRVI DGSTFLKSPG TNPQATVMML GR-------- ---------- ---------- ---------- --
AtGMC2_Q94KD2    --RVVDKNYR VLGIDSLRVI DGSTFLKSPG TNPQATVMML GRYMGQRILQ EREIYNKPDK EA-------- ---------- --
AtGMC3_Q93ZK1    --KVVDNNYK VLGVDALRII DGSTFLKSPG TNPQATIMML GRYMGQKILR ERMAFRGKEE ET-------- ---------- --
AtGMC4_Q66GI5    --RVVDGDYK VIGIDRLRVI DMSTVGYCPG TNPQATVMML GRYMGVKILR ERLTKK---- ---------- ---------- --
AtGMC5_Q9XI68    --SVVDEEYK VNGVKRLRVV DGSTFEESPG TNPMATVLML GRYQGIKILK EREEQEDTFL SPQGSPQPQP ---------- --
PdMDL2_Q945K2    --KVLDGDFR VTGINALRVV DGSTFPYTPA SHPQGFYLML GRYVGIKILQ ERSASDLKI- --------LD SLKSAASLVL --
PsMDL1_P52706    --KVLDGDFR VTGIDALRVV DGSTFPYTPA SHPQGFYLML GRYVGIKILQ ERSASDLKI- --------LD SLKSAASLVL --
PdMDL1_O24243    --KVLDDSFR VMGIKALRVV DASTFPYEPN SHPQGFYLML GRYVGLQILQ ERSIRLEAIH NIQESM---- ---------- --
PsMDL5_O82435    --KVLDDSFS VAGIKALRVV DASTFPYEPN SHPQGFYLML GRYVGLQILQ ERSIRLEAIH NIHDSK---- ---------- --
PsMDL4_O82784    --KVIDGNFR VTGINALRVV DGSTFPATPA SHPQGFYLML GRYVGTKIVQ ERSASGEAIH TSTFKPKLMD SLKSALSFAF --
PsMDL2_O50048    --EVLDDDFR VTGINALRVV DGSTFPSTPA SHPQGFYLML GRYMGTKILQ ERLASEEALH KSTFEPKILE SLESALSFAF ES
PsMDL3_P52707    --KVLDDGFR VTGINALRVV DGSTFPSTPA SHPQGFYLML GRYMGIQILQ ERSASEDAIR NLGFQENILD SPKSTSSFAF --
PpADH_Q9WWW2     PASVVDPCLQ VRGLRNIRVV DASIMPNLVA GNTNAPTIMI AENAAEIIVR KVDMASLD-A SIGFTRQNLE PELL------ --
PoADH_Q00593     PASVVDPCLK IRGLANIRVV DASIMPHLVA GNTNAPTIMI AENAAEIIMR NLDVEALE-A SAEFAREGAE LELAMIAVCM --
ToADH_M5DPH3     ENSVVDPELK VRGVQGLRVI DASIMPTLVA GNTNAPTMMI AENAADMILG NI-------- ---------- ---------- --
NrADH_PW0229..   AMAVVDPRLR VHGLTGLRVV DASIMPTLIG GNTNAPTTMI AERAADLILQ DAALARSV-P NYEGATAAAV ---A------ --
AmADH_PW0207..   EMAVVDTQLR VRGIAGLRVV DASVMPSLIG GNTNAPTVMI AERAAEFIKA AHEGQPVSLA KAESA----- ---------- --
PeAAO_O94219     SWGVVDPDLK VKGVDGLRIV DGSILPFAPN AHTQGPIYLV GKQGADLIKA DQ-------- ---------- ---------- --
AmPDH_Q3L245     SWGVVNPDFK VKGTSGLRVV DASVIPHAPA AHTQLPVYAF AEYASALIAK SYN------- ---------- ---------- --
AtFAO3_Q9LW56    KEGAIDLNGE SWEAEKLFVC DASALPSAVG VNPMITVMST AYCISTRIAK SMTTGLSH-- ---------- ---------- --
AtFAO1_Q9ZWB9    KEGAIDGKGE SWEAEDLYVC DASVLPTALG VNPMITVQST AYCISNRIAE LMKKRKKD-- ---------- ---------- --
AtFAO4B_Q94BP3   EEGALDENGE SWEAEGLFVC DGSILPSAVG VNPMITIQST AYCISSKIVD SLQNKTKV-- ---------- ---------- --
AtFAO4A_O65709   EESAVRPTGE TWEVERLFVA DTSVFPTALG VNPMVTVQSI AYCIGLNVVD VLKKKK---- ---------- ---------- --
PaGOX_P81156     LGGVVDATAK VYGTQGLRVI DGSIPPTQVS SHVMTIFYGM ALKVADAILD DYAKSA---- ---------- ---------- --
AnGOX_P13006     MGGVVDNAAR VYGVQGLRVI DGSIPPTQMS SHVMTVFYAM ALKISDAILE DY-------- ---------- ---------- --
CcFAO1_Q9P8D8    GYGACDTKGR LFECNNVYVA DASVMPTASG VNPMITTMAF ARHVALCLAK DLQPQTKL-- ---------- ---------- --
CcFAO2_Q9P8D7    GYGACDTKGK LFECSNVYVA DASTLPTASG ANPMVSTMSF ARHVSLGIVK ELQQSKL--- ---------- ---------- --
CtFAOT_Q9P8D9    KYGAVDTDGR LFECSNVYVA DASVLPTASG ANPMITTMTF ARNIALGLAE DLKSKAKL-- ---------- ---------- --
SsCHOX_P12676    ----TDDYGR VAGYKNLYVT DGSLIPGSVG VNPFVTITAL AERNVERIIK QDVTAS---- ---------- ---------- --
```

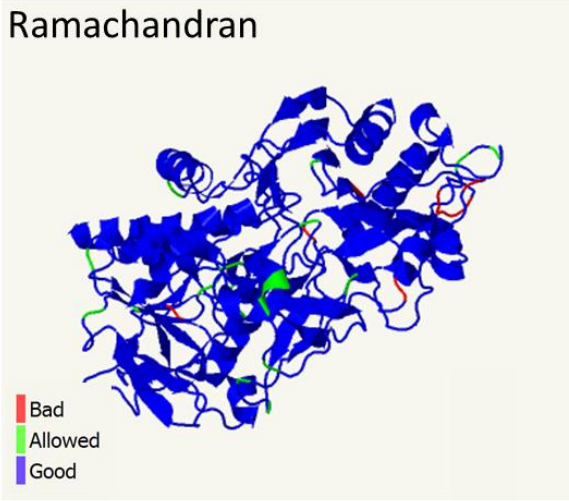
Appendix J. Eight functional HTH amino acid residues and their corresponding residues in related GMC proteins. (A) These eight residues were identified by genetic analysis (Krolikowski et al., 2003) and are labeled with the corresponding HTH protein sequence position. (B) The corresponding residues in other GMC proteins were identified based on sequence alignment (see Appendix H). Amino acid conservation of HTH functional residues in other GMC proteins is highlighted.

Group	Protein	UniProt/NCBI Accession	Functional Residues of HTH / Corresponding Residues in other GMC Proteins								Match
HTH-AtGMC	**HTH**	**Q9S746**	**G_{218}**	**R_{227}**	**G_{294}**	**G_{356}**	**G_{435}**	**P_{564}**	**G_{565}**	**T_{566}**	Self
	AtGMC1	F4KEQ5	G	R	G	G	G	P	G	T	
	AtGMC2	Q94KD2	G	R	G	G	G	P	G	T	
	AtGMC3	Q93ZK1	G	R	G	G	G	P	G	T	
	AtGMC4	Q66GI5	G	R	G	G	G	P	G	T	8/8
	AtGMC5	Q9X168	G	R	G	A	G	P	G	T	7/8
MDL	PdMDL2	Q945K2	G	R	G	G	G	P	A	S	
	PsMDL1	P52706	G	R	G	G	G	P	A	S	
	PdMDL1	O24243	G	R	G	G	G	P	N	S	
	PsMDL5	O82435	G	R	G	G	G	P	N	S	
	PsMDL4	O82784	G	R	G	G	G	P	A	S	
	PsMDL2	O50048	G	R	G	G	G	P	A	S	
	PsMDL3	P52707	G	R	G	G	G	P	A	S	6/8
ADH	PpADH	Q9WWW2	Q	R	G	G	G	V	A	G	
	PoADH	Q00593	Q	R	G	G	G	V	A	G	
	ToADH	5DPH3	Q	R	G	S	G	V	A	G	
	NrADH	WP_022978378.1	H	R	G	G	G	I	G	G	
	AmADH	WP_020743879.1	H	R	G	Y	G	I	G	G	3-5/8
OXDH	PeAAO	O94219	I	R	G	F	G	P	N	A	
	AmPDH	Q3L245	Q	R	G	V	G	P	A	A	4/8
AtFAO	AtFAO3	Q9LW56	G	K	G	G	G	V	G	V	
	AtFAO1	Q9ZWB9	G	K	G	G	G	L	G	V	
	AtFAO4B	Q94BP3	G	K	G	G	G	V	G	V	
	AtFAO4A	O65709	G	K	G	G	G	L	G	V	5/8
GOX	PaGOX	P81156	M	R	G	F	G	V	S	S	
	AnGOX	P13006	P	R	G	F	G	M	S	S	3/8
CsFAO	CcFAO1	Q9P8D8	G	K	G	P	G	S	G	V	
	CcFAO2	Q9P8D7	G	K	W	P	G	S	G	A	
	CtFAOT	Q9P8D9	G	K	G	S	G	S	G	A	3-4/8
Other	SsCHOX	P12676	G	K	G	N	G	V	G	V	4/8

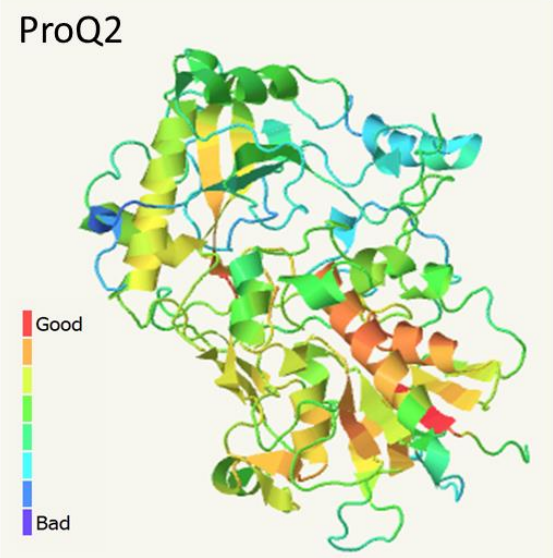
Appendix K. The confidence of the tertiary structure modelling of full-length HTH derived by ProQ2 and Ramachandran analyses. (A) HTH onto the mandelonitrile lyase PdMDL2 (PDB:1JU2 and (B) the pyranose dehydrogenase AmPDH (PDB:4H7U). Both estimation methods indicate that the quality of modelling is average to good. ProQ2 (Ray et al., 2012) is a model quality assessment algorithm that uses support vector machines to predict local as well as global quality of protein models. Ramachandran plot analysis indicates the likelihood of the predicted structure based on the backbone angles. Both analyses show that the majority of the modelled three-dimensional protein structure of HTH is of good quality. Analyses were executed by Phyre2 (http://www.sbg.bio.ic.ac.uk/~phyre2).

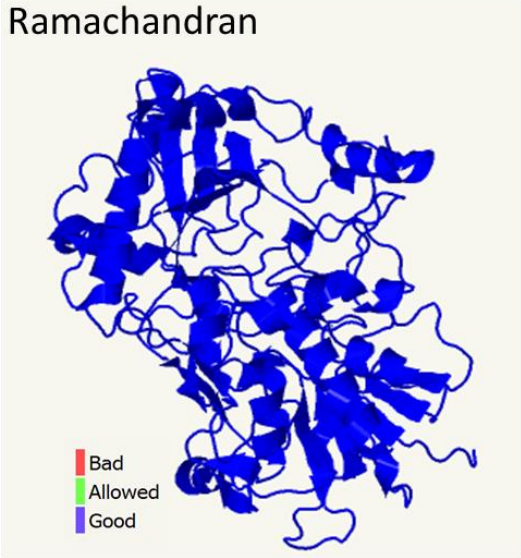
A. Model based on PdMDL2



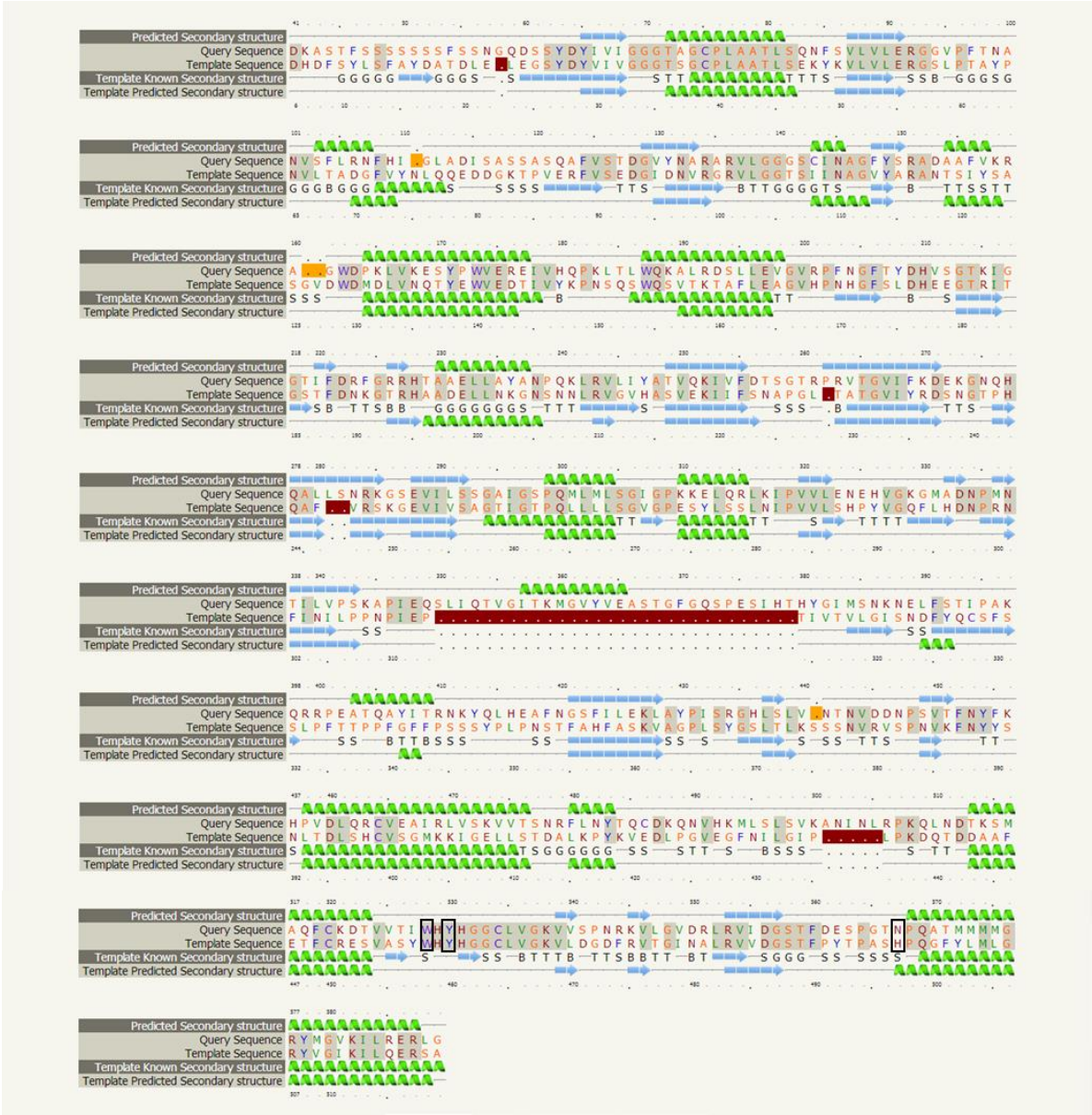


B. Model based on AmPDH

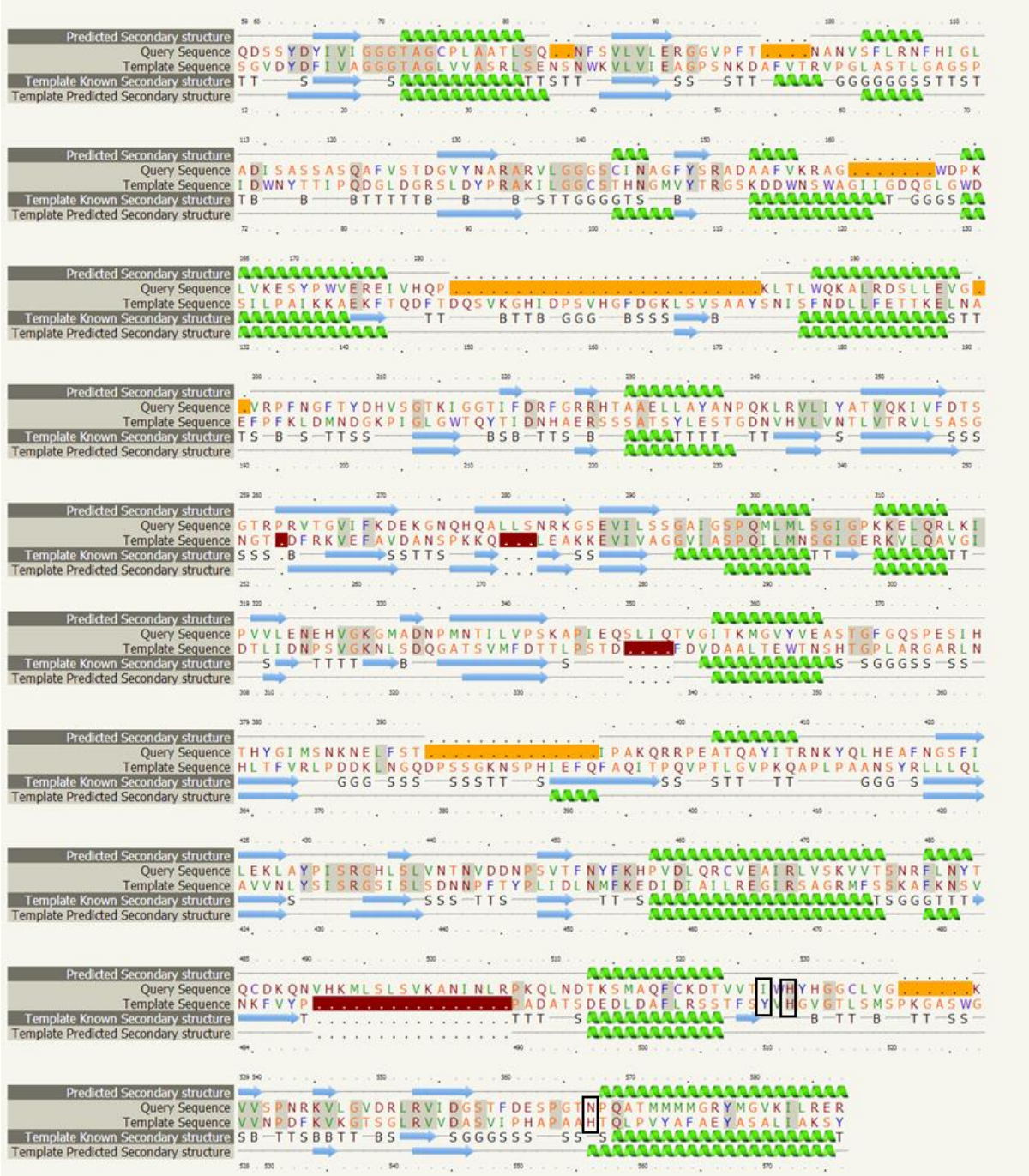




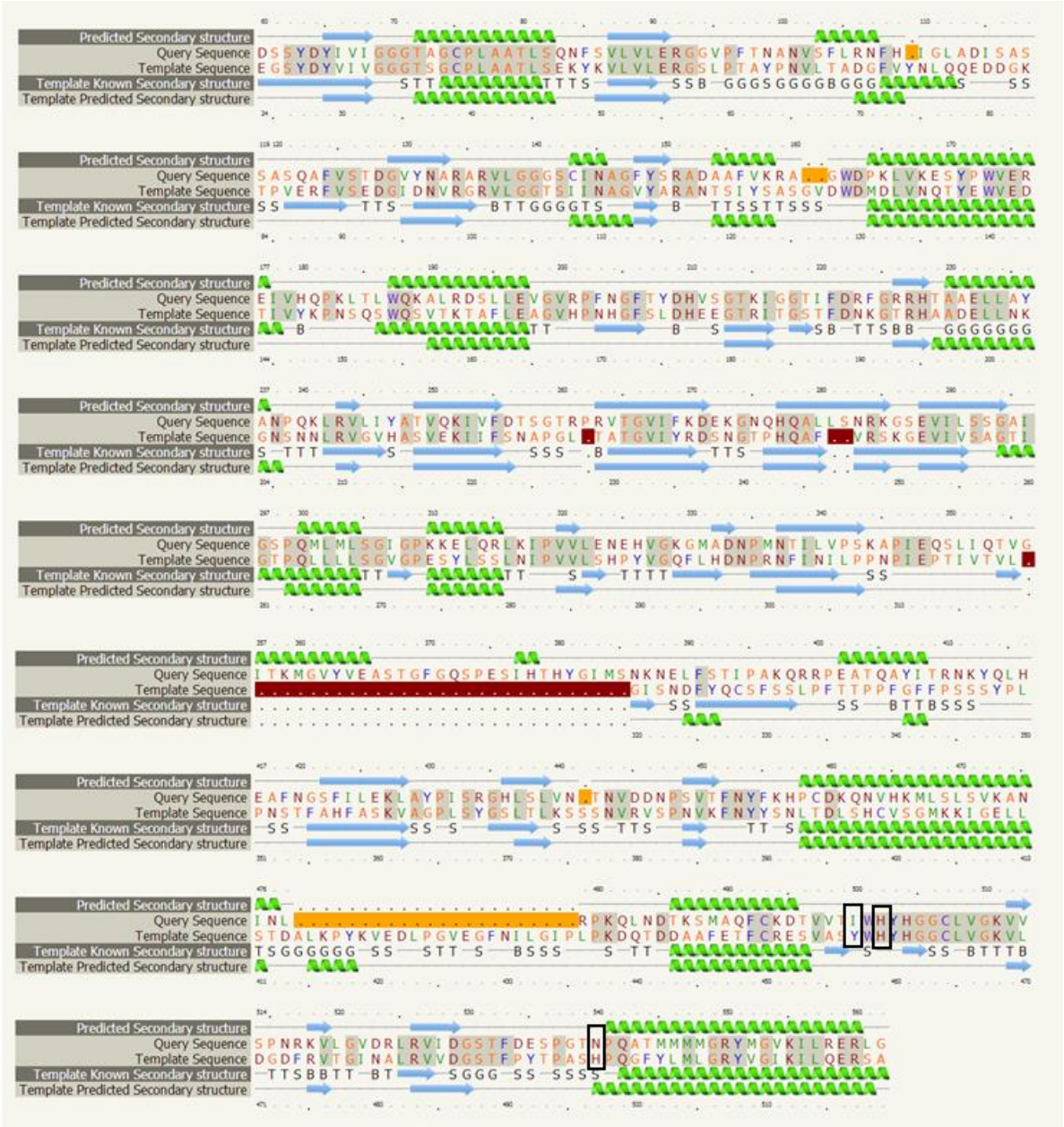
Appendix L. Protein sequence alignments showing the secondary structure of mandelonitrile lyase PdMDL2 and the predicted secondary structure of full-length HTH. The sequence alignment was derived from threading the HTH sequence onto the known template structure of the mandelonitrile lyase PdMDL2 (PDB ID: 1JU2) using the PHYRE server. Green helices represent α-helices, blue arrows indicate β-strands, faint lines indicate coil, and red and yellow blocks indicate alignment gaps. The key catalytic residues are boxed. 510 residues (86%) of the query sequence have been modelled with 100.0% confidence by the single highest scoring template. The query and template sequences share 39% of sequence identity. Source: Protein Homology/analogY Recognition Engine V 2.0 (Phyre 2; http://www.sbg.bio.ic.ac.uk).


Predicted Secondary structure
Query Sequence
Template Sequence
Template Known Secondary structure
Template Predicted Secondary structure

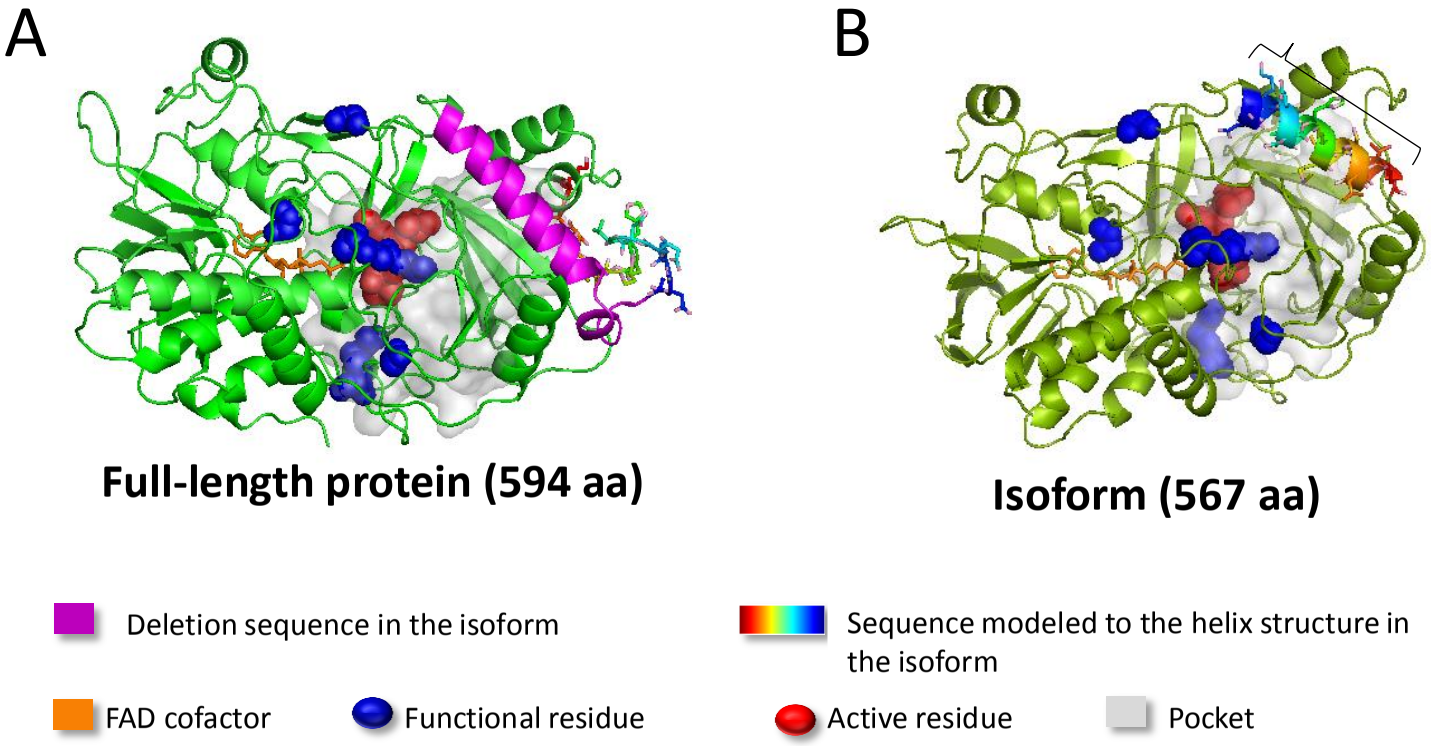
Appendix M. Protein sequence alignments showing the secondary structure of pyranose dehydrogenase AmPDH and the predicted secondary structure of full-length HTH. The sequence alignment was derived from threading the HTH sequence onto the known template structure of mandelonitrile lyase (PDB ID: 4H7U) using the PHYRE server. Green helices represent α-helices, blue arrows indicate β-strands, faint lines indicate coil, and red blocks and dots indicate alignment gaps. The key catalytic residues are boxed. 505 residues (85%) of the query sequence have been modelled with 100.0% confidence by the single highest scoring template. The query and template sequences share 22% of sequence identity. Source: Protein Homology/analogY Recognition Engine V 2.0 (Phyre 2; http://www.sbg.bio.ic.ac.uk).


Predicted Secondary structure
Query Sequence
Template Sequence
Template Known Secondary structure
Template Predicted Secondary structure

Appendix N. Protein sequence alignments showing the secondary structure of the mandelonitrile lyase PdMDL2 and the predicted secondary structure of the predicted HTH isoform derived from a splice variant. The sequence alignment was derived from threading the HTH sequence onto the known template structure of the mandelonitrile lyase PdMDL2 (PDB ID: 1JU2) using the PHYRE server. Green helices represent α-helices, blue arrows indicate β-strands, faint lines indicate coil, and red and yellow blocks indicate alignment gaps. The key catalytic residues are boxed. 470 residues (83%) of the query sequence have been modelled with 100.0% confidence by the single highest scoring template. The query and template sequences share 41% of sequence identity. Source:
Protein Homology/analogY Recognition Engine V 2.0 (Phyre 2;
http://www.sbg.bio.ic.ac.uk).


Predicted Secondary structure
Query Sequence
Template Sequence
Template Known Secondary structure
Template Predicted Secondary structure

Appendix O. Ribbon diagrams showing hypothetical three-dimensional structures of the full-length and a predicted smaller isoform of HTH. Protein sequences were modeled on the mandelonitrile lyase PdMDL2 three-dimensional crystal structure. The predicted tertiary structures of two HTH variants were generated by the PHYRE2 server and graphed by the PYMOL tool. The predicted enzymatic pocket is shown in grey, and the putative active site residues in red. Seven functional residues identified by mutant screens are shown in blue. (A) The theoretical structure of the 594-aa full-length HTH. The sequence absent in the isoform was highlighted in purple. (B) The theoretical structure of the predicted 567-aa long HTH isoform (with a 27-aa deletion: VDLQRCVEAIRLVSKVVTSNRFLNYTQ). Although helix structure-forming deletion sequence (purple) is absent in the smaller isoform, a similar helix structure (rainbow) consisted of neighbouring residues (CDKQNVHKMLSLSVK) is predicted, these residues are also highlighted in rainbow in panel A. Refer to Figure 4.8 for modeling parameters


A
Full-length protein (594 aa)
B
Isoform (567 aa)
Deletion sequence in the isoform
Sequence modeled to the helix structure in the isoform
FAD cofactor
Functional residue
Active residue
Pocket