University of Stavanger

Master thesis in Biological Chemistry

Prepared by: Chimuka Mwaanga

Year: 2011

Identification and expression analysis of peroxisome-targeted defence proteins mediating innate immunity in the model plant *Arabidopsis thaliana.*

University of Stavanger

Faculty of Science and Technology

MASTER'S THESIS

Study program/ Specialization: MSc in Biological Chemistry	Spring semester, 2011 Open access
Writer: Chimuka Mwaanga	…… (Writer's signature)
Faculty supervisor: Professor Sigrun Reumann External supervisor(s):	
Title of thesis: Identification and expression analysis of peroxisome-targeted defence proteins mediating innate immunity in the model plant *Arabidopsis thaliana.*	
Credits (ECTS): 60	
Key words:	Pages: 97 + enclosure: Nil Stavanger, 29.06.2011

Acknowledgements

First and foremost, I want to thank my supervisor professor Sigrun Reumann for making me a proper scientific researcher. Secondly, I would like to extend my heartfelt gratitude to professor Lutz Eichacker for his mentorship, support and guidance during my study.

I also take this chance to thank the PhD fellows Amr Kataya and Gopal Chowdhary for their euthesiastic co-supervision of some parts of the thesis. My gratitude also goes to Aline Benichou, Pradeep Soni and Altynai Adilbayeva for making sure that this study was a success. Quantitative PCR method was established by me in the Reumann laboratory with some help from Dr Kristine Marie Olsen and my sincere gratitude also goes to her.

Finally, I want to thank my wife, Sophie and my family for their love, support, encouragements, and care. Additionally, this master thesis is dedicated to the baby angel my daughter (Luyando Anna Mwaanga), who was born back home, whilst I was abroad pursuing this MSc degree in Biological Chemistry.

Abstract

Peroxisomes are single-membrane organelles that have oxidative metabolic functions. Peroxisomes carry out major functions such as lipid degradation, photorespiration and glyoxylate cycle. However, new functions have been recently reported such as peroxisome mediation in plant innate immunity. To elucidate more on peroxisomal roles in pathogen defence in plants, identification and expression analyses of both new and established peroxisome-targeted pathogen defence proteins in plants was investigated in this study. Subcellular localization analysis of four PTS1 carrying proteins with a pathogen-defence annotation was done. In addition, gene expression analysis of established peroxisomal pathogen defence was carried out using Real-Time Quantitative PCR (qPCR). The four PTS1 carry proteins that whose subcellular localization was studied are NUDT7, NUDT15, CHAT homolog and ATP-BP. NUDT15 and CHAT homolog targeted punctuate subcellular structures, which were later confirmed to be peroxisomes in double labelling experiment with a peroxisomal marker. NUDT7 and ATP-BP failed to target any subcellular structures, were therefore, putatively reported to be cytosolic in this study. Expression analyses were done on three NHL proteins (NHL4, NHL6 and NHL25) and also on three IAN proteins (IAN8, IAN11 and IAN12) using wild type Arabidopsis *Col* 0 plants, by mimicking pathogen attack with exogenously applied defence hormone-salicylic acid. All the NHL and IAN genes were induced after salicylic acid treatment. In addition, co-expression analyses were done on the aforementioned NHL and IAN proteins (except for NHL25). NHL6 and IAN8 were co-expressed with other Arabidopsis defence proteins. Whereas NHL4, IAN11 and IAN12 were found not be co-expressed in the dataset generated. In conclusion, in this study, two new peroxisomal pathogen defence proteins were identified namely NUDT15 and CHAT homolog, and also NHL6 and NHL25 were induced by salicylic acid treatment.

Key words: Peroxisomes PTS1 Immunity Salicylic acid Pathogens

---

CONTENTS

Acknowledgements. I

Abstract II

1. Introduction 1

1.1 Biogenesis and protein import of peroxisomes 1

1.2 Metabolic functions of plant peroxisomes 3

1.3 Role of peroxisomes in plant innate immunity 4

1.4 Pathogen defence proteins of interest 8

1.5 Thesis goals 11

2. Materials and Methods 12

2.1 Molecular cloning 12

2.1.1 cDNA constructs 14

2.1.2 Oligonucleotide primers 14

2.1.3 Polymerase chain reaction (PCR) 16

2.1.4 Agarose electrophoresis 19

2.1.5 Extraction of PCR fragments and restriction double digests 20

2.1.6 Ligation of cDNA fragments into destination vectors 21

2.1.7 Transformation of competent *E. coli* cells 22

2.1.8 Isolation of plasmid DNA 23

2.1.8.1 Illustra^{TM} plasmidPrep Mini Spin Kit (GE Healthcare) 23

2.1.8.2 Wizard® Plus SV Minipreps DNA Purification System (Promega) 24

2.1.9 Restriction digestion 25

2.1.10 Sequencing 25

2.2 Subcellular localization studies in *Allium cepa L*. 27

2.2.1 DNA precipitation onto gold particles 27

2.2.2 Transformation of onion epidermal cells by bombardment 27

2.2.3 Fluorescence microscopy 29

2.3 Gene expression analysis with qPCR 30

2.3.1 Plant growth conditions 30

2.3.2 RNA isolation 30
2.3.3 cDNA synthesis 31
2.3.4 Standard qPCR reaction 31
2.3.5 Method establishment 32
2.3.5.1 Primer testing and optimization 33
2.3.5.2 CT validation experiment 33
2.3.6 SA treatment of wild type Arabidopsis plants 34
3. Results 35
3.1 cDNA subcloning and subcellular localization analysis of four PTS1 protein candidates 35
3.1.1 Bioinformatic analysis of the four PTS1 candidates 35
3.1.2 Cloning of the four PTS1 protein candidates into pCAT via pGEM-T Easy vectors 38
3.1.2.1 cDNA verification and amplification 38
3.1.2.2 PCR fragment isolation 38
3.1.2.3 Directional cloning into pCAT via pGEM-T Easy vectors 40
3.1.3 *In vivo* validation of predicted peroxisome targeting 43
3.2 Gene expression analysis of selected NHL and IAN proteins 45
3.2.1 Bioinformatics analyses of NHL proteins 45
3.2.2 RNA isolation and cDNA synthesis 47
3.2.3 Validation experiments 48
3.2.4 Gene expression analysis with qPCR 54
3.2.4.1 SA treatment of plants 54
3.2.4.2 Relative quantitation of peroxisomal pathogen defence genes 55
3.3 Co-expression analysis 56
3.3.1 Dataset generation for co-expression analysis 56
3.3.2 Co-expression groups 56
3.3.3 Co-expression patterns of the 28 Arabidopsis pathogen defence proteins 58
4. Discussion 59
4.1 Subcellular localization studies 59
4.1.1 NUDT7 is a cytosolic pathogen defence protein in Arabidopsis 60

4.1.2 NUDT15 is localized in peroxisomes 62
4.1.3 The CHAT homolog is a peroxisomal protein 64
4.1.4 ATP-BP is a cytosolic R protein 65
4.2 Gene expression analysis 67
4.2.1 Plant growth and SA treatment 67
4.2.2 RNA extraction 67
4.2.3 cDNA synthesis and selection of endogenous control gene 69
4.2.4 Validation of gene expression analysis 70
4.2.5 Relative quantification of NHL and IAN genes by the comparative CT method 72
4.3 Signal transduction pathway analysis for two NHL and three Ian proteins 73
5. Conclusion 76
6. References 77
7. Appendix 80
7.1 cDNA sequence analysis in pGEM-T Easy and pCAT vectors 80
7.1.1 NUDT7 80
7.1.1.1 NUDT7 sequence in pGEM-T Easy vector 80
7.1.1.2 NUDT7 sequence in pCAT vector 80
7.1.2 NUDT15 82
7.1.2.1 NUDT15 sequence in pGEM-T Easy vector 83
7.1.2.2 NUDT15 sequence in pCAT vector 84
7.1.3 ATP-BP 85
7.1.3.1 ATP-BP full-length sequence in pGEM-T Easy vector 85
7.1.3.2 ATP-BP 420aa C-terminal sequence in pGEM-T Easy vector 86
7.1.3.3 ATP-BP 420aa C-terminal sequence in pCAT vector 87
7.1.4 The CHAT homolog 88
7.1.4.1 The CHAT homolog sequence in pGEM-T Easy vector 88
7.1.4.2 The CHAT homolog sequence in pCAT vector 89
7.2 Gene Expression analysis supplementary data 91

7.2.1. Micro-array data of NHL proteins from Genevestigator and NCBI databases 91

7.2.2 Relative gene quantification results 93

7.3 Co-expression analysis supplementary data 94

7.4 Abbreviations 96

1. Introduction

1.1 Biogenesis and protein import of peroxisomes

Peroxisomes are single-membrane bound organelles found in most of eukaryotic cells. They belong to a class of microbodies (Kagawa and Beevers, 1975). Phylogenetically, they are proposed to have an endosymbiotic origin. Ontogenetically, they have been proposed to have either an endoplasmic reticulum or de novo biogenesis (Gabaldon et al., 2006).

Peroxisome biogenesis in plants is induced by a number of factors such as; change in cellular redox (reduction-oxidation) state, pathogen and herbivore attack, and many more other abiotic and biotic stresses (Lopez-Huertas et al., 2000; Nyathi and Baker, 2006). The peroxisomal numbers are up-regulated through stress responses and down-regulated through a peroxisome degradation process called pexophagy. Pexophagy is a type of autophagy selectively meant for peroxisome degradation. Macropexophagy and micropexophagy are the two modes of pexophagy employed (Sakai et al., 2006). Peroxisome biogenesis involves import of proteins to the peroxisome. Once brought to the peroxisome proximity, proteins are incorporated into the single lipid-bilayer membrane as integral or peripheral proteins, whilst other proteins are imported into the matrix as soluble peroxisomal matrix proteins.

Proteins are termed as `working horses` for the cell. They carry out various critical and vital duties such as enzymatic, regulatory, metabolic, structural and mechanical activities. Organelles (subcellular compartments within the cell) equally need proteins. Peroxisomes being one of the eukaryotic cell organelle are not any exceptional to this profound need. Nonetheless, peroxisomes lack the critical material-DNA that is needed for protein synthesis. However, most peroxisomal proteins are nuclear encoded and synthesized on either endoplasmic reticulum-bound ribosomes or free ribosomes in the cytosol and imported post-translationally into the peroxisomes (Gabaldon et al., 2006).

Most of the peroxisomal proteome is of eukaryotic origin and a reasonable fraction of it comes from alpha-proteobacterial (Gabaldon et al., 2006). Peroxisomal proteins can be grouped roughly into peroxisomal membrane proteins and peroxisomal matrix proteins. Matrix proteins for the peroxisomes are much more characterized than the membrane ones. After being synthesized either on ribosomes in the cytosol or endoplasmic reticulum, some proteins target the peroxisomes distinctively. They do this through targeting signals called peroxisomal targeting signals (PTSs). PTSs found on matrix proteins are of two types, that is

peroxisomal targeting signal type 1 or type 2 (PTS1/2)(Reumann, 2004). Nonetheless, not all peroxisomal proteins carry these PTSs. The majority of matrix proteins carry PTS1s that are usually located at the carboxyl-terminus (C-terminus) of the protein. In some rare instances, PTS1 is internally located, however, it still exhibits a C-terminus bias (Kamigaki et al., 2003). PTS1 is a tripeptide sequence that is derived from a combination of position-specific amino acid residues such as ([SAPCFVGTLKI] [RKNMSLHGETFPQCY] [LMIVYF] >) (Reumann, 2004; Ma et al., 2006; Reumann et al., 2009; Lingner et al., 2011). There are 11 amino acid residues at position -3, 15 residues at position -2, and six residues at position -1 that allowed, in order to have a functional plant PTS1 tripeptides (Lingner et al., 2011). PTS1s can be either canonical-major or noncanonical-minor depending on their pattern and/or targeting efficiency. Tripeptide sequences such as SKL> and SSL> represent major and minor PTS1s respectively (Gould et al., 1987; Gould et al., 1989; Kamigaki et al., 2003; Reumann, 2004; Lingner et al., 2011),. There difference between major and minor PTS1s is that major PTS1s can target the peroxisomes entirely by themselves whereas minor ones need help from upstream residues such as Proline and basic amino acid residues for them to target the protein to the peroxisome (Reumann, 2004). Plant PTS1 tripeptides follow a distinct pattern, in which at least two high-abundance residues of presumably strong targeting strength ([SA][KR][LMI] >) are combined with one low-abundance PTS1 residue to yield functional plant PTS1 tripeptides; (x[KR] [LMI]>, [SA]y[LMI]>, and [SA][KR]z>) (Lingner et al., 2011). Unlike PTS1, PTS2 is an amino-terminus located nonapeptide such as RLx5HL and RIx5HL (Reumann, 2004; Reumann et al., 2009). RLx5HL and RIx5HL are examples of major PTS2 nonapeptides found in plants. Not so many matrix proteins carry PTS2, nevertheless, a significant number of matrix proteins with PTS2 have been studied.

Peroxisomal matrix proteins need to be taken to the matrix post-translationally. However, to help them achieve this goal, there are two receptors found in the cytosol scavenging for proteins carrying PTSs. These receptors are encoded by PEX5 and PEX7 genes are called Peroxin5 and Peroxin7 (Pex5p and Pex7p respectively). Pex5p is needed for transportation of both PTS1 and PTS2 proteins in plants and mammals, and not required for transportation of PTS2 proteins in fungi (Bonsegna et al., 2005; Hayashi et al., 2005). Pex7p is required for PTS2 targeting of the peroxisome (Hayashi et al., 2005). The import system of both PTS1 and 2 proteins can be summarized as a four step process; recognition of PTSs by the respective receptor, then cargo and receptor docking and translocation, then cargo offloading, receptor recycling (Dammai and Subramani, 2001; Brown and Baker, 2008).

1.2 Metabolic functions of plant peroxisomes

Peroxisomes play crucial roles in seed germination, seedling development, overall plant growth, hormone biosynthesis and disease resistance (Queval et al., 2007; Reumann et al., 2007; Reumann, 2011). All these important tasks are done through their involvement in lipid degradation, production and detoxification of Hydrogen peroxide (H2O2), hormone synthesis and signalling molecule production (Nyathi and Baker, 2006; Queval et al., 2007). Hydrogen peroxide is produced during Photorespiration by plants in the presence of light and air, and it plays an instrumental role as a signal molecule (Queval et al., 2007). H2O2 catabolism to water and oxygen is through catalase activities. Unlike in mammals, lipid catabolism (break down of fatty acids to a two carbon compound-acetyl CoA) in plants takes place solely in the peroxisome (Poirier et al., 2006). Lipid degradation process is called β-Oxidation. β-oxidation is a source of various metabolites for the cell. Several pathways derive metabolites from β-oxidation, which includes jasmonic acid biosynthesis pathway, glyoxylate cycle and indole-3-acetic acid synthesis (Nyathi and Baker, 2006; Poirier et al., 2006).

Plant developmental phase, anatomy, cell type and environment tend to dictate peroxisome type and concentration. This phenomenon is termed plasticity. The `plastic` nature of peroxisomes makes possible for them to carry out such dynamic and diverse functions mentioned in the preceding paragraphs. Peroxisomes are referred to by different names according to the main function they are carrying out in a particular part of plant and environment. They are referred to as glyoxysomes during seedling development and senescence where they help in synthesizing carbohydrates from lipids. During greening they are called leaf peroxisomes. It has been reported that glyoxysomes are capable of changing into leaf peroxisomes during greening and back to glyoxysomes from leaf peroxisomes during senescence. The evident inter-conversion between types of peroxisomes is regulated at various levels, such as at gene expression, splicing of the mRNA and degradation of proteins (Nishimura et al., 1986; Mano et al., 1996; Nishimura et al., 1996).

1.3 Role of peroxisomes in plant innate immunity

Plants growing in nature frequently encounter a wide range of environmental conditions comprising not only beneficial but also adverse conditions. Therefore, plants have to respond and cope with these dynamic adverse conditions in- and on-time to ensure their survival. Plant responses are both at cellular and physiological levels. Adverse conditions plants are frequently exposed to include biotic stresses such as pathogen and herbivore attack, and also abiotic stresses such as drought, heat, salinity and cold. Pathogens are more problematic to plants than the other stresses.

Plant pathogens use diverse life strategies to survive in and/or on their hosts. Pathogenic bacteria proliferate in intercellular spaces (the apoplast) after entering through gas or water pores (stomata and hydathodes, respectively), or gain access via wounds. Nematodes and aphids feed by inserting a stylet directly into a plant cell. Fungi can directly enter plant epidermal cells, or extend hyphae on top of, between, or through plant cells. Pathogenic and symbiotic fungi and oomycetes can invaginate feeding structures (haustoria), into the host cell plasma membrane. Pathogens are classified as biotrophs, hemi-biotrophs and necrotrophs (Panstruga et al., 2009). Biotrophic pathogens live and complete their whole life cycle inside the host. Hemi-biotrophs spend part of their life cycle stage in the host. Necrotrophs do not depend on the host for any part of their life cycle stage. These aforementioned diverse pathogen classes all deliver effector molecules (virulence factors) into the plant cell to block defence reactions by the host.

Plants, unlike animals, lack mobile defender cells and a somatic adaptive immune system. Instead, they rely on the innate immunity of each cell and systemic signals emanating from infection sites for resistance against invaders. Plant innate immunity operates at various levels. The first line of defence is called basal disease resistance. Basal disease resistance is activated by virulent pathogens on susceptible hosts. Thus, basal disease resistance is, at first glance, PAMP-triggered immunity (PTI) minus the effects of effector triggered susceptibility (ETS) (see Fig.1). However, there is also likely to be weak effector triggered immunity (ETI) triggered by weak recognition of effectors (Jones and Dangl, 2006; Panstruga et al., 2009). PTI is triggered by PAMPs or MAMPs (Pathogen/Microbe associated molecule patterns) such

as Flagellin (flg22), bacterial cold shock proteins and elongation factor Tu (EF-Tu)(Jones and Dangl, 2006; Panstruga et al., 2009).

The plant immune system can be represented as a four phased 'zigzag' model (Fig.1; (Jones and Dangl, 2006). In stage 1, PAMPs are recognized by PRRs (Pathogen recognition receptors), resulting in a PAMP-triggered immunity that can halt further colonization. In stage 2, successful pathogens inject effectors that contribute to pathogen virulence. Effector-triggered susceptibility (ETS) results when effectors interfere with the PAMP-triggered immunity (PTI). Effectors that enable pathogens to overcome PTI are sometimes recognized by specific disease resistance (R) proteins. The recognition is either indirect or direct. R proteins are encoded by genes. However, most R genes encode NB-LRR (Nucleotide binding leucine rich repeats) proteins. NB-LRR proteins are grouped into two classes; Toll-Interleukin-1 receptor (TIR) and Coiled coil (CC-NB-LRR) (Wiermer et al., 2005; Ge et al., 2007; Panstruga et al., 2009; Knepper et al., 2011).

TIR- NB-LRR proteins are regulated by ENHANCED DISEASE SUSCEPTIBILITY 1 (EDS1) and together with its interacting partner, PHYTOALEXIN DEFICIENT 4 (PAD4). EDS1 and PAD4 constitute a regulatory hub that is essential for basal resistance to invasive biotrophic and hemi-biotrophic pathogens (Wiermer et al., 2005; Panstruga et al., 2009). EDS1 is negatively regulated by Nudix hydrolase homolog 7 (NUDT7) (Ge et al., 2007)
The second type of R proteins i.e. CC-NB-LRR are regulated by *Arabidopsis thaliana* NDR1 (NON-RACE SPECIFIC DISEASE RESISTANCE-1), a plasma membrane localized protein. NDR1 activates the R proteins by monitoring modification of RIN4 (RPM1-INTERACTING PROTEIN-4). RIN4 modifications include phosphorylation and cleavage(Knepper et al., 2011). RIN4 is found to be always guarded by some specific R proteins. Therefore, this association keeps the R proteins involved, inactivated until some RIN4 modification which are perceived by NDR1 occur (Panstruga et al., 2009; Knepper et al., 2011).

In stage 3 of the zig zag model, a given effector is 'specifically recognized' by one of the NB-LRR proteins, resulting in effector-triggered immunity (ETI). Recognition is again either indirect, or direct. (Jones and Dangl, 2006; Knepper et al., 2011). ETI is an accelerated and amplified kind of PTI response, resulting in disease resistance and, sometimes a hypersensitive cell death response (HR) at the infection site may happen. In phase 4 of zig zag

model, some pathogens evolve so that they can elude the mounted ETI either by shedding or diversifying the recognized effector gene, or by acquiring additional effectors that suppresses ETI. Equally the host improves its defence mechanism through acquisition of new R gene specific to the new effectors so that ETI can be triggered again.

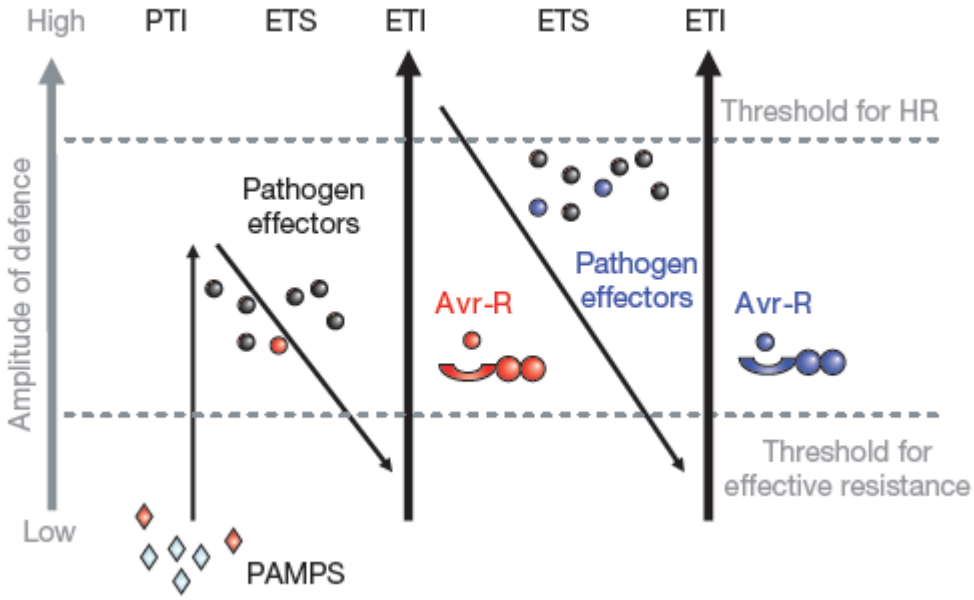


Figure 1.1: Schematic presentation of the plant innate immune system as a four phased model called `Zig Zag` model (Jones and Dangl, 2006). This model illustrates the quantitative output of the plant innate immune system. The ultimate amplitude of disease resistance or susceptibility is determined by the balance between PTI plus ETI`s amplitude and that of ETS. When PTI and/or ETI is greater than ETS the host becomes resistant to the infection or attack. In a situation where ETS is greater than PTI and/or ETI then the host becomes susceptible to infection. PTI is usually the first stage in plant innate immunity, then stage 2 is ETS where pathogens develop strategies to elude the PTI mounted against them by the host. Stage 3, host mounts a much stronger defensive mechanism-ETI. All the effectors that get recognized by R protein are termed avirulent (Avr) and the complex formed is Avr-R. In some cases ETI ends in HR.

R protein activation results in a network of cross-talk between response pathways deployed, in part, to differentiate between biotrophic and/or hemi-biotrophic from necrotrophic pathogen attack. This is maintained by the balance between salicylic acid and a combination of jasmonic acid (JA) and ethylene. SA is a local and systemic signal for resistance against many biotrophs, whilst the combination of JA and ethylene accumulation are signals that promote defence against necrotrophs (Bruinsma et al., 2009).

Once the pathogen type has been identified or differentiated, various defensive mechanisms are put up by the plant under attack. For intracellular pathogens like bacteria, SA mediated responses such as HR PCD (Hypersensitive response programmed cell death), autophagy or exocytosis are possible options. Actions against herbivory which are JA mediated include production of phytotoxins such as camalexins and nitriles. The major chemical defence system is the glucosinolate-myrosinase system (Rask et al., 2000). In this system an amino acid derivative glucosinolate is metabolised to compounds such as nitriles, isocyanates and cyanates by myrosinases such as thioglucosinolase (Rask et al., 2000). These toxins are either toxic to herbivores or attract predators to the herbivores (Bruinsma et al., 2009). In case of a fungal infection, plant responses include polarized toxin production at the site of attempted entry site and also callose formation (Panstruga et al., 2009; Knepper et al., 2011).

Peroxisomes contribute to all or most of the above mentioned responses through production of reactive oxygen species (ROS) that are needed for autophagy induction subsequently HR PCD (Liu et al., 2005; Scherz-Shouval et al., 2007; Hofius et al., 2009). They also produce signalling molecules such as reactive nitric species. JA biosynthesis takes place in the peroxisomes and JA is a very important defence hormone against necrotrophic attack as well as in cross-checking with SA actions. Arabidopsis penetration (PEN2) is a peroxisomal protein. PEN2 together with PEN3 (plasma membrane ABC transporter) are recruited to attempted fungal entry sites, apparently to mediate the polarized delivery of a toxin to the apoplast (Panstruga et al., 2009; Knepper et al., 2011).

Plants unlike animals they lack defence cells like B- and T-cells that are transported in the blood and are able to render pathogen resistance remotely. Nonetheless, when plants mount a local innate immunity against pathogens they simultaneously mount systemic acquired resistance (SAR) in remote parts of the plant (Schilmiller and Howe, 2005). SAR is mediated by salicylic acid-SA. SA accumulation in distal parts of the plant is by de novo synthesis via isochorismate synthase (Yasuda et al., 2008; Attaran et al., 2009).

1.4 Pathogen defence proteins of interest

Proteins with a pathogen defence function that are established as localized in the peroxisomes and also those with a putative defence annotation and carrying PTS1s were of much interest in this study. The established peroxisomal defence proteins from two families i.e. NDR1/HIN1-like (NHL) and immune-associated nucleotide-binding (IAN) protein family were of particular interest in this study. Additionally, four proteins carrying PTS1s and with a pathogen defence annotation were also of interest.

The Arabidopsis genome contains a family of NHL genes that are homologous to both the non race-specific disease resistance (NDR1)(Knepper et al., 2011) and the tobacco (*Nicotiana tabacum*) hairpin-induced (HIN1) genes. There are about 28 genes that encode NHL proteins. Therefore, NHL proteins are designated as NHL1-28. Some NHL proteins like NHL3 are pathogen-responsive hence, have a potential involvement in pathogen defence. Nonetheless, this study focuses on NHL4 (At1g54540 AKL>); NHL6 (At1g65690 LRL>); and NHL25 (At5g36970 FRL>). In red prints are c-terminal tripeptides (PTS1s) and stop codon of the proteins (Lingner et al., 2011). These three NHL proteins are peroxisomal proteins (Unpublished data by A. Kataya and S. Reumann; (Lingner et al., 2011). Collectively, these NHL genes (NHL4/6 and 25) can be induced by conditions like SA treatment, FLG22, drought and salt only to mention but a few as shown from figures below obtained from publicly available microarray data from Genevestigator. NHL4/6 and 25 have a conserved domain (LEA superfamily domain) similar to that of NDR1. LEA (Late embrogenesis abundact) proteins have been reported to be involved in plant innate immunity (Knepper et al., 2011).

IAN proteins belong to a family of AIG1-like GTPases. IAN proteins are also known as GTPase of immunity-associated proteins (GIMAP). All the IAN proteins have specific conserved amino acid domains: a AIG1 domain and a coiled-coil motif. The first IAN protein was found in Arabidopsis and designated as AIG (avrRpt2-induced gene). The Arabidopsis AIG1 (IAN8) and AIG2, which are the first identified IAN proteins, are involved in plant resistance to bacteria. Recent analysis of the expression patterns of Arabidopsis *IANs* suggests that these IAN proteins may play regulatory roles during plant development and response to both biotic and abiotic stress(Wang and Li, 2009). 3 IAN proteins of interest in this study are

IAN8 (At1g33960), IAN11 (At4g09930), IAN12 (At4g09940). IAN12 is peroxisomal protein where as IAN8 and IAN11 target yet unkown punctuate subcellular structures (Unpublished Data by A. Kataya and S. Reumann).

Publicly available microarray data from Genevestigator showed that NHL4 and NHL6 and also IAN8/11 and IAN12 proteins can be up- and /or down regulated by a number of biotic and abiotic stimuli. Stimuli which up-regulates gene expression, include SA treatment and Pseudomonas infection. Regulation of gene expression is time dependant (see results section). Using new PST1 prediction models four genes carrying PTS1s and with a pathogen defence related annotation were selected (Table.1). The plant PTS1 prediction model called position-specific weight matrices (PWM) model was used to select the four genes from the whole Arabidopsis genome ((Lingner et al., 2011). These four PTS1 carry proteins are NUDT (for nucleoside diphosphates linked to some moiety X) hydrolases 7 and 15, acetyl CoA: (Z)-3-hexen-1-ol acetyltransferase (CHAT; At5g17540) homolog and ATP binding protein (ATP-BP; At1g72840).

NUDT7 (At4g12720) and NUDT15 (At1g28960) belong to a gene family which hydrolyze ribonucleoside and deoxyribonucleoside triphosphates, nucleotide sugars, coenzymes, or dinucleoside polyphosphates (Ogawa et al., 2008). NUDT7 has four gene models (transcriptional and translational variants) and all the four have the same noncanonical PTS1-ASL>. NUDT7 was predicted to be a non peroxisomal protein by PWM model despite having a known PTS1 (Ogawa et al., 2008; Lingner et al., 2011). NUDT7 is a negative regulator of basal immunity in Arabidopsis, modulates two distinct defense response pathways and is involved in maintaining redox homeostasis (Ge et al., 2007). NUDT7 negatively regulates EDS1 which controls defence activation and programmed cell death conditioned by intracellular Toll-related immune receptors that recognize specific pathogen effectors. EDS1 is also needed for basal resistance to invasive pathogens by restricting the progression of disease (Bartsch et al., 2006). NUDT15 has five gene models. Two of the five NUDT15 gene models contain the same noncanonical PTS1-PKM>. The other three NUDT15 gene models also have the same c-terminal tripeptide (CMP>) though it is not a known PTS1. Despite having some of its gene model with known PTS1, NUDT 15 was predicted to be a non peroxisomal protein (Ogawa et al., 2008; Lingner et al., 2011).CHAT homolog is a PTS1 carrying protein with enzymatic activities and unlike the NUDT proteins it does not have more than one gene models (Lingner et al., 2011). CHAT homolog has a noncanonical PTS1-

SSL>. It was predicted to be a peroxisomal protein (Lingner et al., 2011). ATP-BP is a PTS1 carrying R protein and has two gene models (Lingner et al., 2011). The two gene models of ATP-BP carry each a different c-terminal tripeptides. One of the two has a noncanonical PTS1-PKM>, whereas the other carries a non PTS1 tripeptide-CMP>. ATP-BP has the NB-LRR conserved domain just like other R proteins.

Table 1.1: PWM model-based PTS1 protein predictions for NUDT7, NUDT15, CHAT homolog and ATP-BP. All the gene models for our four proteins of interest were predicted and shown. The C-terminal tripeptide, allows the gene models to be sorted by their predicted PTS1 tripeptide. The thresholds of the prediction scores for predicted peroxisome targeting are 0.412. Gene models predictions results are shown in the peroxisome prediction column as; 1 for PTS1 protein and 0 for non-PTS1 protein. Posterior probability (Post. Prob.) shows the chances of the gene model of targeting the peroxisomes. The highest for posterior probability is 1.

AGI code (TAIR9 ID)	Acronym	C-terminal 14 aa residues	C-term. Tripep.	Peroxisome prediction	Post. prob.	Pred. score
AT5G17540.1	CHAT homolog	RGSKSSNKLIMSSL	SSL	1	0,850842	0,483865
AT4G12720.1	NUDT7	KRLKVSRDQASASL	ASL	0	0,069248	0,315453
AT4G12720.2	NUDT7	KRLKVSRDQASASL	ASL	0	0,069248	0,315453
AT4G12720.3	NUDT7	KRLKVSRDQASASL	ASL	0	0,069248	0,315453
AT4G12720.4	NUDT7	KRLKVSRDQASASL	ASL	0	0,069248	0,315453
AT1G28960.4	NUDT15	AFIEQCPKFKYPKM	PKM	0	0,026734	0,280752
AT1G28960.2	NUDT15	AFIEQCPKFKYPKM	PKM	0	0,026734	0,280752
AT1G28960.1	NUDT15	FKYPKMVEKHTCMP	CMP	0	0	-1,06092
AT1G28960.3	NUDT15	FKYPKMVEKHTCMP	CMP	0	0	-1,06092
AT1G28960.5	NUDT15	FKYPKMVEKHTCMP	CMP	0	0	-1,06092
AT1G72840.1	ATP-BP	MNEEYSQEVRLSSL	SSL	0	0,037061	0,292363
AT1G72840.2	ATP-BP	IILCGVEHVGFVLK	VLK	0	0	-1,33342

1.5 Thesis goals

Recently, peroxisomes have been implicated to having to play a role in disease resistance (Nyathi and Baker, 2006). It was due to this interesting new peroxisomal function (not so much explored) that the Reumann research group carried out a pathogen defence research study on peroxisomes. This study aimed at identifying new peroxisomal-targeting pathogen defence proteins and also analyse gene expression of already established peroxisomal defence proteins.

Arabidopsis thaliana (Arabidopsis) was the model plant of study used to be used. Using, this model plant, four genes carry PTS1s and homologous to established defence proteins were identified and followed with subcellular localization analyses (Lingner et al., 2011). The four PTS1 carrying proteins were NUDT7, NUDT15, CHAT and ATP-BP.

Again using our model plant (Arabidopsis), gene expression analyses of established peroxisomal proteins was to be carried out. The proteins investigated for expression were NHL 4, NHL6, NHL25, IAN 8, IAN11 and IAN12. RT-qPCR (Reverse Transcription Quantitative Polymerase Chain Reaction) was to be used in gene expression analyses (Livak and Schmittgen, 2001; Schmittgen and Livak, 2008). Furthermore, Co-expression analysis on the aforementioned NHL and IAN proteins were to be carried out. Co-expression analyses were *in silico* studies were to be carried out using bioinformatics tool; the Expression Angler (http://bar.utoronto.ca/ntools/cgi-bin/ntools_expression_angler.cgi) and AtGenExpress Pathogen Set.

2 Materials and methods

2.1 Molecular cloning

Overview of subcloning steps of full-length cDNA of PTS1 protein candidates in the back of EYFP (Reporter protein)

Plasmid isolation from Bacterial stock

↓

Analytical PCR (to verify correct insert in purchased INRA and RIKEN plasmids**)**

↓

Analytical agarose gel electrophoresis (to check fragment size and PCR specificity**)**

↓

Preparative PCR (to amplify the cDNA insert in purchased plasmids**)**

↓

Analytical agarose gel electrophoresis (to verify successful insert amplification**)**

↓

Preparative agarose gel electrophoresis (to isolate preparative PCR fragment**)**

↓

Preparative PCR fragment gel extraction (GE Healthcare GFX PCR DNA and Gel band purification kit**)**

↓

PCR fragment subcloning into pGEMT-Easy (TA cloning**)**

↓

***E.coli* transformation and pGEMT-Easy vector isolation**

↓

Analytical double restriction digestion

↓

Analytical agarose gel electrophoresis (to verify correct insert size and determine the DNA concentration**)**

↓

Analytical PCR (to verify the presence of the insert in pGEMT-Easy**)**

↓

Insert sequencing

↓

Preparative double restriction digestion

↓

Preparative agarose gel electrophoresis (to isolate double restriction digestion fragment**)**

↓

Preparation of the destination pCAT-EYFP vector

↓

Insert subcloning into the pCAT-EYFP vector

↓

***E.coli* transformation and colony PCR**

↓

Plasmid miniprep. and insert sequencing

↓

DNA precipitation onto gold particles and onion epidermal cell bombardment

↓

Subcellular localization analysis with fluorescence microscopy

2.1.1 cDNA constructs

The full-length cDNA of the Arabidopsis proteins were obtained from public stock centres such as the French INRA and the RIKEN Biological Resource Centre (BRC, Japan). The stocks obtained from INRA were shipped as bacterial stabs while the stocks from RIKEN were shipped as plasmid DNA. All the purchased plasmids were ampicillin resistant.

2.1.2 Oligonucleotide primers

A pair of gene-specific oligonucleotide primers (forward and reverse) flanked with desired restriction endonuclease sites had been designed prior to the start of the thesis. The forward primers introduced a NotI site at the 5' end and the reverse primers at XbaI for NUDT7, NUDT15 and CHAT homolog. SacII sites at the 3' end of ATP-BP constructs were used instead of XbaI. Oligonucleotide primers were used for amplification of the full-length cDNA by polymerase chain reaction (PCR). The primer pairs that were used are shown in Table 2.2. The annealing temperature of the primers (i.e., the temperature at which the primers bind at highest specificity to the templates) was calculated according to the following formula:

$$T_m = 69.3\ ^\circ C + 41 \cdot \%GC - 650/n$$

$$T_a = T_m - 3\ ^\circ C$$

where T_m is the melting temperature at which the primers separate from the template, %GC is the ratio of the bases guanine (G) and cytosine (C) in the primer to the total number of bases, n is the number of bases, and T_a is the annealing temperature.

Table 2.1: Primer pairs used in PCR.

cDNAs acronym	**Primers**	***Sequences (5'→ 3')***
NUDT7 (At4g12720)	CM1f	ACTGCGGCCGCTATGGGTACTAGAGCTCAGAAG
	CM2r	CAAGTCTAGAGTCAGAGAGAAGCAGAGGCTTG
NUDT15(At1g28960)	CM3f	AAGACTGCGGCCGCTATGTTTTTGCTTTATCGT
	CM4r	CAAGTCTAGAGTCACATTTTAGGGTACTT
ATP-BP full length (At1g72840)	CM5f	ACTGCGGCCGCTATGGCTTCCTCGTCATCAAAG
	CM7r	GTACCGCGGTTATAGAGAAGAGAGCCT
ATP-BP 420aa C-terminal (At1g72840)	CM6f	AAGACTGCGGCCGCTCTGCTTCCAAACCTACGGATA
	CM7r	GTACCGCGGTTATAGAGAAGAGAGCCT
CHAT homolog (At5g17540)	CM8f	AAGACTGCGGCCGCTATGTCCGGGTCACTCACG
	CM9r	CAAGTCTAGAGTCACAGAGAAGACATGATCAA

2.1.3 Polymerase chain reaction (PCR)

DNA was amplified by using a thermocycler to induce an enzymatic elongation of primers complementary to a template DNA. Both analytical and preparative PCR, as well as colony PCR, were performed. In both analytical and colony PCR Dream Taq DNA polymerase (MBI Fermentas) was used; in preparative PCR a proof-reading DNA polymerase was used (Expand High Fidelity ^{PLUS} PCR System from Roche Applied Science). Colony PCR was used to quickly screen for plasmid inserts directly from *E. coli* colonies. Analytical PCR was always done as a pilot PCR experiment prior to preparative PCR. Analytical PCR was used to verify correct insert in purchased INRA and RIKEN plasmids and also to verify the presence of the inserts in pGEMT-Easy after TA-cloning. Preparative PCR was always used after analytical PCR**.**

The components in Tables 2.3 and 2.4 were added to PCR tubes (NOTE: the enzyme was added last, after the reaction mix had been cooled on ice). Table 2.3 shows the components for a typical analytical PCR with homemade DNA polymerase, whereas Table 2.4 shows the components needed for a preparative PCR with proof-reading DNA polymerase.

Table 2.2: Components of an analytical PCR

Component	**Volume**
Sterile double-distilled H_2O	to 50 µl
10x Taq buffer (final conc. 1x)	5 µl
25 mM $MgCl_2$ (final conc. 2.5 mM)	5 µl
10 mM dNTP (final conc. 0.2 mM)	1 µl
10 µM forward primer (final conc. 0.2 µM)	1 µl
10 µM reverse primer (final conc. 0.2 µM)	1 µl
Template DNA	0.5 µl
Taq polymerase (ca. 1.5 U/µl)	1.5 µl

In colony PCR, instead of adding template DNA from a solution, a fractional amount of *E. coli* colony was added. This was done by touching a colony on a Luria Bertani (LB) plate with a fine pipette tip, and then stirring the PCR mix with the pipette tip. Table 2.5 shows PCR machine settings, in order to successful amplify cDNA and/or DNA.

Table 2.3: Components of a preparative PCR

Component	**Volume**
Sterile double-distilled H_2O	to 50 µl
5x Expand HF buffer with 15 mM $MgCl_2$ (final conc. 1x)	10 µl
10 mM dNTP (final conc. 0.2 mM)	1 µl
10 µM forward primer (final conc. 0.4 µM)	2 µl
10 µM reverse primer (final conc. 0.4 µM)	2 µl
Template DNA (0.1-10 ng plasmid)	1 µl
Expand High Fidelity PCR system enzyme mix (5 U/µl)	0.5 µl

Table 2.4: Standard PCR program for preparative and analytical PCR

Step	**Cycle**	**Preparative PCR**		**Analytical PCR**	
		Temperature (°C)	**Time**	**Temperature (°C)**	**Time**
Initial denaturation	1	96	2 min	96	5 min
Denaturation	1-5	96	30 sec	96	45 sec
Annealing	1-5	T_a (1/2 primer)	30 sec	T_a (1/2 primer)	45 sec
Elongation	1-5	72	30 sec-4 min*	72	2 min
Denaturation	1-5	96	30 sec	96	45 sec
Annealing	6-25	T_a (full primer)	30 sec	T_a (full primer)	45 sec
Elongation	6-25	72	30 sec-4 min*	72	2 min
Final elongation		72	10 min	72	10 min
Cooling		12	∞	12	∞

* For preparative PCR, the elongation time was adapted to the length of the PCR product; ca. 1 minute per 1 kb was used.

2.1.4 Agarose electrophoresis

The two types of agarose gel electrophoresis were carried out after every PCR experiment, and also after restriction endonuclease digestions, namely analytical and preparative agarose gel electrophoresis. Analytical agarose gel electrophoresis was used to check fragment size and PCR specificity and also to verify successful insert amplification. Correct insert size and DNA concentration determination was also done by running analytical agarose gel electrophoresis. Preparative agarose gel electrophoresis was used to isolate preparative PCR and double restriction digestion fragments, respectively.

To make 1% (w/v) agarose gel, powdered agarose 1% (w/v) was melted in 1x TAE buffer (e.g. 0.5 g/50 ml) and then casted into a plate with the comb(s) for well-making.

Table 2.5: 50x TAE buffer composition

50x TAE buffer:	g/l
2 M Tris-Base	242 g
Acetic acid (glacial), pH 8.3	57.1 ml
EDTA	100 ml 0.5 M (pH 8.0)
H_2O	1000 ml

The gel with its gel plate was placed in the electrophoresis apparatus containing 1x running buffer (TAE) which covered the wells. To keep the loadings in the wells, 6x Fermentas Orange loading buffer was added (1:6 dilution). For an analytical electrophoresis the concentration of DNA in the loading mixture was significantly lower (10-70 ng, 1-3 µl) than for a preparative electrophoresis (0.2-1.5 µg, 15-25 µl). A 1:5 diluted size marker, GeneRuler^{TM}1 kb DNA Ladder (Fermentas), was loaded (0.1 µg/µl, 5 µl) into the first well, followed by loading of the samples. The gel was electrophoresed at 70 V (ca. 150 mA) for 45-60 minutes.

The gel was exposed to UV light to obtain a photograph for result documentation. For preparative electrophoresis low intensity UV light (365 nm) and a short exposure time were

used while the band was cut out to minimize DNA damage. The intensity and position of the band were compared to the size marker, and the concentration and size were determined.

Table 2.6: 1Kb DNA ladder preparation

1 kb standard:	Volume
sterile water	4 µl
Fermentas GeneRuler™ 1 kb DNA Ladder 0.5 µg/µl	1µl
Fermentas 6x Loading Buffer Orange	1 µl
Gel Red	1 µl

NOTE: Gel Red made DNA bands visible under UV light

2.1.5 Extraction of PCR fragments and restriction double digests

Preparative PCR products or restriction endonuclease digests were separated on 1% agarose gels (2.1.4). By using low intensity UV light (365 nm), the DNA band of interest was cut out from the gel and purified using illustra GFX™ PCR DNA and Gel Band Purification Kit (GE Healthcare). A DNase-free 1.5 ml microcentrifuge tube was weighed before and after adding the cut-out agarose gel band. The weight of the agarose gel slice was calculated and 10 µl capture buffer type 3 added for each 10 mg agarose gel slice. The sample was mixed and incubated at 60°C until the agarose gel was completely dissolved. The capture buffer type 3 plus sample mixture (600 µl) was then transferred to a GFX Microspin column placed inside a collection tube and incubated for 60 seconds at room temperature. The sample was centrifuged at 16000×g for 30 seconds and the flow-through discarded. This DNA binding step was repeated until the entire sample was loaded. The membrane bound DNA was then washed by adding 500 µl washing buffer type 1 and centrifuging at 16000×g for 30 seconds. The collection tube was discarded and the GFX microspin column transferred to a new 1.5ml DNase-free microcentrifuge tube. The DNA was then eluted from the filter by adding 15 µl elution buffer type 4, incubating for 60 seconds at room temperature and centrifuging at 16000×g for 60 seconds.

2.1.6 Ligation of cDNA fragments into destination vectors

Two different types of ligations were carried out; TA-ligation of cDNAs into the pGEMT-Easy vector and "sticky end" ligations of cDNAs into the pCAT-EYFP plant expression vector. The insert in the TA-ligation was a double-stranded PCR product containing an A-overhang at the 3' ends. pGEM-T Easy is a linearized vector with a T-overhang at the 3'ends. For the pCAT-EYFP-cDNA ligation the insert was released from the pGEMT-Easy vector by restriction endonuclease cleavage. The back bone of the destination vector was cut with the same restriction endonuclease enzymes as the cDNA resulting in complementary overhangs (sticky ends) in each DNA molecule.

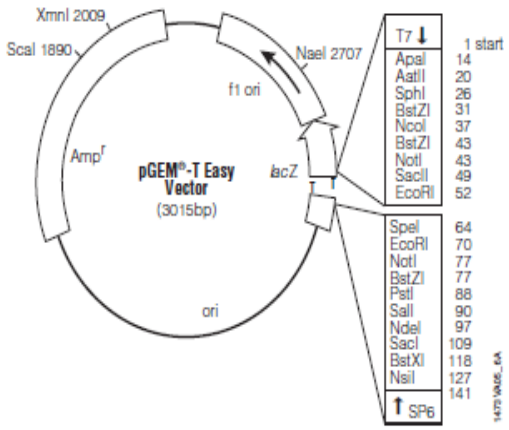


Figure 2.1: pGEM-T Easy vector map.

A stoichiometric ratio of vector-to-insert of approximately 1:3 was used in the reaction mix for DNA ligation. The size of pGEMT-Easy and pCAT vectors are 3 kbp and 4.5 kbp respectively. If the concentration of the cDNA was too low, the fragment was concentrated in a heating block (50 °C). The reaction was incubated at 4°C overnight. T4 DNA Ligase was inactivated by incubating at 65°C for 20minutes.

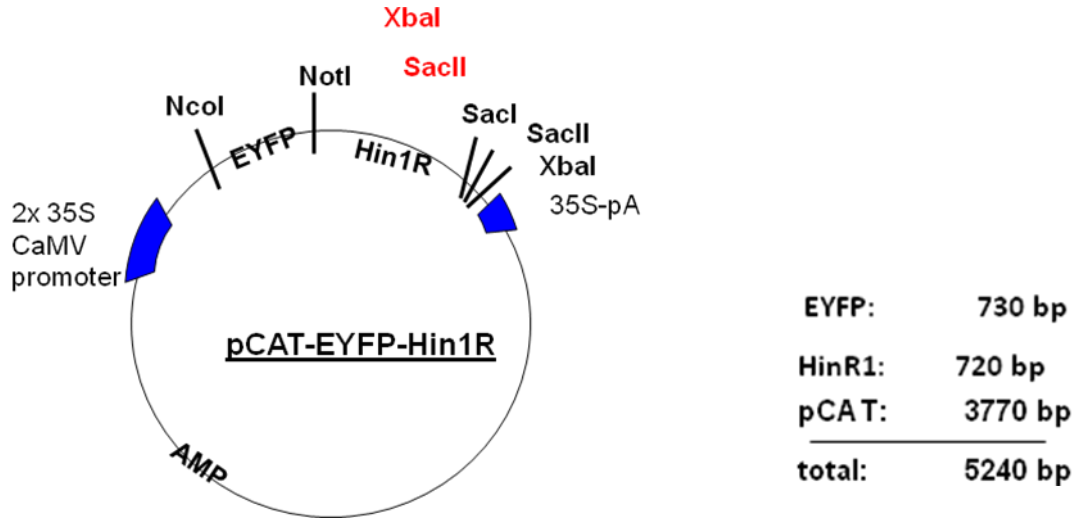


Figure 2.2: pCAT vector map. Hind1R (At1g54540, 720 bp) had been previously subcloned via NotI/SacI into a modified version of the original pCAT vector (containing the extended MCS NotI-SacI-SacII-XbaI downstream of EYFP). This vector was to be used for subcloning of ATP-BP, NUDT7, NUDT15 and CHAT homolog in the back of EYFP.

2.1.7 Transformation of competent *E. coli* cells

An appropriate aliqout of competent cells of *Escherichia coli* (*E. coli)* was thawed on ice. 1 µl of plasmid DNA was added to the competent cells and mixed carefully by pipeting up and down. The tube was then incubated on ice for 30 minutes. To perform a heat shock, the tube was placed in a water bath at a temperature of 42 °C for 50 seconds and then quickly placed back on ice for two minutes. LB medium (500µl) was added to the tube which was then incubated at 37 °C for 1 hour in a roller drum to allow the cells to express the antibiotic resistance gene. The cells were then plated on LB plates with the ampicillin, since all the plasmid carried this resistance gene.

When carrying out blue white screening, a preparation of 40 µl 100 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) and 40 µl 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside(X-Gal) were spread on top of the LB plates prior to inoculation. The plates were then incubated at 37 °C for 12-16 hours.

Table 2.7: LB medium and LB agar composition

Component (final conc.)	LB medium	LB agar
1% (w/v) Tryptone	10 g/l	10 g/l
0.5% (w/v) Yeast extract	5 g/l	5 g/l
1% (w/v) NaCl	10 g/l	10 g/l
	0.2 ml/l NaOH (5 N)	1ml/l NaOH (1 N)
1.5% (w/v) Agar		15 g/l

2.1.8 Isolation of plasmid DNA

The plasmid DNA was isolated from the bacterial cells using the Illustra™ plasmidPrep Mini Spin Kit (GE Healthcare; 2.1.8.1) or the Wizard® Plus SV Minipreps DNA Purification System (Promega; 2.1.8.2). The Promega purification kit was generally used to isolate to isolate pCAT plasmids due to the fact that it yields much more plasmid DNA than the GE Healthcare.

2.1.8.1 Illustra™ plasmidPrep Mini Spin Kit (GE Healthcare)

A single bacterial colony with the cDNA of interest was inoculated in LB medium (5 ml). Prior to the inoculation ampicillin was added to LB medium. The culture was then incubated at 37°C overnight to allow bacterial amplification of the plasmid containing our cDNA of interest to take place. The bacterial cells were first harvested by transferring 1.5 ml of the bacterial culture to an Eppendorf tube and centrifugation at 16000×g for 30 seconds. The supernatant was poured off and discarded. The same procedure was then repeated for another 1.5 ml culture volume. The pellet of bacterial cells was re-suspended in 200 µl lysis buffer type 7. Buffer type 8 (200 µl) was added and the tube was gently inverted to mix. This alkali treatment was done to lyse the cells and to denature genomic DNA and proteins. Lysis buffer type 9 (400 µl) was added right after. Type 9 buffer is an acetate buffered solution, containing a chaotropic salt and neutralized the pH of the lysate. The lysate was centrifuged at

16000×g for 4 minutes. The plasmid DNA remained in the supernant whilst cellular debris including genomic DNA, proteins and lipids were precipitated. The supernatant was transferred to a plasmid mini column inside a collection tube, centrifuged at 16000×g for 30 seconds and the flow-through discarded. During this step the plasmid was bound to the membrane due to the presence of chaotrope. 400 µl washing buffer type 1 was added and centrifuged at 16000×g for one minute. The ethanolic washing buffer removed residual salts and other contaminants. The flow-through and collection tube were discarded. The plasmid mini column was then transferred to a DNase free microcentrifuge tube and the plasmid DNA was eluted from the plasmid mini column by adding 50 µl water, followed by incubation for 30 seconds at room temperature and centrifugation at 16000×g for 30 seconds.

2.1.8.2 Wizard® Plus SV Minipreps DNA Purification System (Promega)

10 ml of LB medium with ampicillin was inoculated with a single bacterial colony that was transformed with the cDNA of interest. Overnight incubation as above in section 2.1.8.1 but 10 ml of bacterial culture was harvested by centrifugation at 10000×g for 5 minutes. The supernatant was poured off and the bacterial pellet thoroughly resuspended in 250 µl cell resuspension solution. Resuspension was done by pipetting up and down repeatedly. 250 µl of cell lysis solution was added. The tube containing the sample was mixed by inverting it four times. Then 10 µl of Alkaline Protease Solution was added inorder to inactivating endonucleases and proteins. The tube was again inverted four times and incubated for five minutes at room temperature. After the incubation 350 µl Neutralization Solution were added and mixed by inverting the tube four times. The bacterial lysate was centrifuged at maximum speed (20000×g) for 10 minutes. The supernatant was transferred to a spin column inside a collection tube and centrifuged at maximum speed for one minute. The flow-through was discarded. 750 µl Column Wash Solution was added and centrifuged at maximum speed for one minute. The flow-through was discarded and the washing step repeated with 250 µl column wash solution. The sample was later on centrifuged at maximum speed for two minutes. The spin column was then transferred to a DNase free microcentrifuge tube and the plasmid eluted by adding 50 µl of Nuclease-free water and centrifuging at maximum speed for one minute.

2.1.9 Restriction digestion

Restriction digestion was performed for both analytical and preparative purposes. The preparative restriction digestions with single enzymes about 1 µg plasmid DNA, 1 µl of 10x reaction buffer, 0.2µl restriction enzyme (10 U/µl) and water (to add up to 10 µl) were added to one tube. The mixture was incubated at 37°C overnight. It is important that the total volume of enzymes does not exceed 10% (v/v) of the reaction mix to avoid unspecific star activity. To achieve quantitative digest, the mixture was incubated for another hour and analyzed by agarose electrophoresis. For analytical restriction digest 1 µl DNA (ca. 0.1-0.6 µg), 1 µl of the appropriate 10x reaction buffer, 1 µl of each restriction enzyme (1-5 U) and water (to add up to 10 µl) were mixed together and incubated (37°C, 1 h). For double digestions the plasmid DNA or PCR product was digested with two restriction enzymes at the same time, and an appropriate reaction buffer ensuring the highest activity of both enzymes was used. Depending on the enzymes' activities in the chosen buffer the number of units of enzyme was adjusted to compensate for reduced activity. For example, for an analytical double digest with NotI and XbaI, buffer O (MBI Fermentas) was used. In this buffer NotI has a restriction enzyme activity of 100% while XbaI has an activity of 20-50%. Routinely 1 U NotI was used, and the reduced activity of XbaI was compensated by using 5 U XbaI.

2.1.10 Sequencing

Sequencing was done every time the cDNA had been ligated into a new vector and successfully amplified by transformation of *E. coli* cell. Sequencing was done by Seqlab (Goettingen, Germany). The purpose was to verify the correct nucleotide sequence of the cDNA as mutations could have occurred during template amplification, due to primer errors, or by the cDNA exposure to UV light. Two extended HotShot sequence runs (covering ca. 800 bp) were done on the cDNA of interest in pGEM-T Easy and the pCAT expression vector. T7 primer (5'-TAATACGACTCACTATAGGG -3') was used to sequence inserts in pGEM-T Easy from the forward end. Inserts in pCAT were sequenced using SR321f primer (5'- ACT ACC TGA GCT ACC AGT CC- 3') designed according to the specific region that had to be sequenced. If the cDNA size was longer than 800 bp, two separate sequencing runs were required, one with a reverse primer and one with a forward primer. It was very important to only use a single primer in each tube; two primers would result in two overlapping sequences that cannot be read. The samples to be sequenced were prepared adding the components in Table 2.10 to a PCR tube.

Table 2.8: Components of samples sent out for sequencing at seqlab

Component	Amount
Plasmid DNA	600-700 ng
Tris buffer pH 8.0	final concentration of 7.1 mM (e.g. 1 µl 50 mM)
Primer	20 pmol (e.g. 2 µl 10 pmol/µl)
dH_2O to a final volume of	7 µl

2.2 Subcellular localization studies in *Allium cepa L.* (Onion)

2.2.1 DNA precipitation onto gold particles

An aliquot of 50 mg gold particles was resuspended in 1 ml ethanol and vortexed for 3-5 minutes. The gold particles were then sedimented by centrifugation at 10000×g for three seconds and the supernatant was discarded. The washing step was repeated three times. After the last washing the gold particles were resuspended in 1 ml water and then vortexed and centrifuged as before. The supernatant was discarded and resuspended the gold particles in 1 ml water. The suspension was aliquoted in 50 µl aliquots. The next steps were performed on ice. The following components were added one after the other in given order and vortexed thoroughly for two minutes after each addition: 5 to 7 µl plasmid DNA (1 µg/µl) (final conc.: about 40 ng/µl), 50 µl 2.5 M $CaCl_2$ (final conc.: about 1 M) and 20 µl 0.1 M Spermidine (final conc.: about 10 mM). The DNA was then precipitated onto the gold particles by centrifugation at 10000×g for 3 seconds. The supernatant was removed. The particles were resuspended in 250 µl ethanol, vortexed and sedimented by centrifugation at 10000×g for 3 seconds. The supernatant was removed. This washing step was repeated three times. The particles were finally resuspended in 60 µl ethanol and ready for use for the transformation of onion epidermal cell by bombardment.

2.2.2 Transformation of onion epidermal cells by bombardment

A biolistic system (PDS-1000/He Particle Delievery system, Biorad) was used in the transformation of onion epidermal cells *(Allium cepa L.)*. The aforementioned biolistic system uses highly pressurized helium, which builds up above a rupture disk that bursts at a predefined pressure. When the rupture disk bursts, a helium shock wave is generated into the cell bombardment chamber. The helium shock wave propels the macrocarrier loaded with DNA coated gold particles toward the target cells at high velocity. The macrocarrier is stopped by a stopping screen while the DNA coated particles continue to the target and transform the onion epidermal cells.

Onion has to be made ready for bombardment before putting it in the biolistic system. A healthy onion was peeled and cut into well-sized slices. A quarter of a slice with the epidermal cell layer still attached was placed in a Petri dish on a wet piece of paper. The gun chamber was sterilized with 70% ethanol. The helium bottle was opened and the pressure

adjusted to 1400 PSI. The pressure should be adjusted a little higher than where the disc ruptures. The suspension of gold particles coated with the desired plasmids was vortexed thoroughly and 5 µl were loaded onto the macrocarrier holder in the shooting device. The gold particles were spread with the side of a pipette tip over an area of about 1 cm^2. A rupture disk was sterilized in ethanol and loaded into the retaining cap. The retaining cap was secured to the end of the gas acceleration tube and tightened with a torque wrench. The macrocarrier containing the DNA and the stopping screen were loaded into the microcarrier launch assembly. The microcarrier launch assembly was placed into the top shelf and the targeted onion cells placed into the third shelf. The chamber room was closed, the vacuum pump turned and the power switch on the bombardment device turned on. The vacuum was lowered to about 270 inches Hg and then held. The fire button was pushed until rupture of the rupture disc. The pump was turned off and the vacuum slowly released. The vacuum pump was turned off, the chamber door opened and the dish with the onion removed. The macrocarrier and stopping screen from the microcarrier launch assembly were unloaded and discarded as well as the spent rupture disk. When all the experiments were completed the helium bottle was closed. The helium pressure was released from the tubing by applying vacuum and shooting a couple of times. The onion was left in the Petri dish and incubated in the dark for about 18 hours to allow for transient expression of the cDNA of interest using the onion epidermal cells. After incubation, analysis by fluorescence microscopy followed. NUDT7, NUDT15 and ATP-BP were also incubated at 4°C for three to six days (Lingner et al., 2011).

2.2.3 Fluorescence microscopy

After being transiently expressed in onion epidermal layer cells, microscopic analysis was done on the PTS1 protein candidates. An inverted fluorescence microscope was used to in subcellular analysis of the transformed onion epidermal cells. The epidermal cell layer was peeled off, put onto a microscope slide and covered with a cover slide. The prepared sample was placed on the stage with the cover slide upside down. The cells were either single- or double-labelled. In single-labelled cells, only the vector containing EYFP was transformed into onion cells. Double-labelled cells contained one extra plasmid, the DsRed-SKL (in pWEN vector) that targeted to peroxisomes using SKL - major PST1 (Matre et al., 2009).

Fluorescence image acquisition was performed on a Nikon TE-2000U inverted fluorescence microscope and filters for YFP (exciter HQ500/20, emitter S535/30) and CFP (exciter S436/10, emitter S470/30), equipped with a Hamamatsu Orca ER 1394 cooled CCD camera. Volocity II software (Improvision) was used for picture capture. The different fluorescent proteins were distinguished from each other by using different filters on the microscope. A Macintosh computer was connected to the microscope which was used for picture storage. Emitted fluorescent light was changed to green (EYFP) and red (Ds-Red-SKL) to allow the detection of double-labelled peroxisomes. Two pictures of one cell (one observed with EYFP filter and one with DsRed-SKL filter) were overlaid in PhotoShop.

2.3 Gene expression analysis with qPCR

2.3.1 Plant growth conditions

Standard healthy Arabidopsis thaliana ecotype Columbia-0 plants were grown on a mixture of soil and vermiculite in the ratio 3:1 for 6 to 7 weeks. The growth chambers, where a 16-h-light/8-h-dark cycle at 22°C under a light intensity of 100 to 150 µE m^{-2} s^{-1} were used. After sowing the seeds, they were covered with a plastic dome for the first week to maintain humidity until germination.

2.3.2 RNA isolation

RNA isolation was carried out with RNeasy Plant Kit (Qiagen Kit). All the steps were carried out at room temperature, including centrifugation.

Previously harvested leaves that were stored at -80°C were thoroughly and quickly crushed pestle and mortar. Pestle and mortar were maintained at cryogenic temperature by liquid nitrogen. The thoroughly crushed leaf powder along with liquid nitrogen was transferred to an RNase free, liquid Nitrogen cooled 2 ml Microcentrifuge tube. Liquid nitrogen was left to evaporate but not allowing the leaf tissue to thaw. RLT buffer (450 µl) was added. Prior to use, an appropriate amount of mercaptoethanol was added to RLT buffer. Vigorously vortexing for 30 sec was done. The lysate was transferred to a QIAshredder spin column and centrifuged at 14,000 g for 2 minutes. Supernatant was carefully transferred to a new tube and 96 -100 % ethanol (225 µl) was added and mixed immediately by pippeting. The mixture was transferred to an RNeasy spin column and centrifuged at 8000 g for 15 sec. The flow through was discarded after centrifugation and reinserted the collection tube, to which, 700 µl of RW1 buffer was added. This step was repeated twice. RPE buffer (500 µl) was added to the same RNeasy spin column in two separate steps and centrifuged at 8000 g for 15 seconds and 2 minutes respectively before discarding the collection tube along with the flow through. To ensure that no carry-over solutions was extracted with the isolated RNA, the RNeasy column was placed in a new empty 2 ml collection tube and centrifuged at 14000 g for 1 minute 30 seconds. RNA was eluted into a new 1.5 ml collection tube by adding 50 µl of RNase free water directly to the spin column, and centrifuging at 8000 g for 1 minute.

Nano-Drop Spectrophotometer was used to determine RNA concentration and purity. Prior to determining the concentration and purity, 1 µl of extracted RNA was diluted in 2 µl of sterile water. 1.5% (w/v) non denaturing agarose gel was used to determine the extracted RNA integrity. Use only 1 µl of extracted RNA for this purpose. 45 µl of RNA from the total eluted

amount (ca. 250 ng/ µl) was used in cDNA synthesis after diluting it to a concentration of 100 ng/ µl. The remaining 3 µl of RNA from the total volume eluted was aliquoted into 1 µl and stored for future use such as reassessment of concentration, purity and integrity.

2.3.3 cDNA synthesis

cDNA was synthesized using the High Capacity cDNA Reverse Transcription kits. Three major steps were followed in order to synthesize single-stranded cDNA from total RNA.

1. 2X Reverse Transcription master mix was prepared as shown in the table below using the kit components

Table 2.9: Reverse Transcription master mix preparation

Component	Volume/Reaction (µL)
10X RT Buffer	2.0
25X dNTP mix (100mM)	0.8
10X RT Random Primers	2.0
Multiscribe™ Reverse Transcriptase	1.0
Nuclease-free water	4.2
Total per Reaction	**10**

Note: Allow the kit components to thaw on ice and prepare the RT master mix on ice

2. Addition of 10 µl of total RNA (100ng/ µl) to the 2X Reverse Transcription master mix to create a 1X mix
3. Performing Reverse Transcription in a thermal cycler (traditional PCR machine). The thermal cycler settings were; 25°C for 10 min (step 1), then 37°C for 2 hours (step 2) then 85°C for 5minutes (step 3) and last step at 10°C for indefinite.

2.3.4 Standard qPCR reaction

96 plate wells were used in all qPCR experiments. The qPCR reaction cocktail per well was as shown below in Table. 2X PCR master mix containing SYBR Green I dye, water, and cDNA were almost always mixed to form a master mix before adding gene-specific primer pairs. 20 µl of qPCR reaction components was dispatched into appropriate wells in order to

increase the triplicate precision by lessening the number of pipeting steps. The thermal cycle profile of the qPCR machine was set-up as shown in table .

Table 2.10: Real time PCR reaction mix

Component	Volume/Reaction (µL)
2X PCR master mix	10
cDNA (10ng/ µl)	1
Gene-specific Forward Primer (300nM/ µl)	1.2
Gene-specific Reverse Primer (300nM/ µl)	1.2
Nuclease-free water	6.6
Total volume/well	**20**

Table 2.11: qPCR Thermal Cycler Profile

Stage	Repetitions	Temperature	Time	Ramp Rate
1	1	50.0 °C	02:00	100
2	1	95.0 °C	10:00	100
3	40	95.0 °C	00:15	100
		60.0 °C	01:00	100
4 (Dissociation)	1	95.0 °C	00:15	Auto
		60.0 °C	00:20	Auto
		95.0 °C	00:15	Auto
		60.0 °C	00:15	Auto

2.3.5 Method establishment

Real time PCR had never been used in the Reumann laboratory as a gene expression analysis method. Therefore, a major goal of this master thesis was to establish the method in the group. To this end, before it qPCR for gene a number of validation experiments had to be carried out in order to establish the method. Primer testing, primer optimization, dynamic test and PCR amplification efficiencies of target and endogenous gene validation experiments were

performed prior to relative quantitation of gene by qPCR with the 7300 Applied Biosystem PCR machine.

2.3.5.1 Primer testing and optimization

The primer pairs used in this present study (e.g., for endogenous control, SA positive markers and genes of interest) in gene expression analysis are listed in the table below.

Table 2.12: qPCR primer pairs

Gene	AGI	Forward primer	Reverse primer
ACT2	AT3g18780	TGCCAATCTACGAGGGTTTC	CAGTAAGGTCACGTCCAGCA
PR2	At3g57260	AGCTTCCTTCTTCAACCACACAGC	TGGCAAGGTATCGCCTAGCATC
PR5	At1g75040	ATCACCCACAGCACAGAGACAC	AGCAATGCCGCTTGTGATGAAC
Ubi10		CGGATCAGCCAGAGGCTTATT	AGCCTGAGGACCAAGTGGAG
NHL4	At1g54540	TGCAGCAGCAACAACAAACAGG	TTCCGAGTTTGATGGCGACAGG
NHL6	At1g65690	TGGGAGCAAGATTACCGTGTGG	TTTGGCAACGACCCATTGCTTA
NHL25	At5g36970	CCAGAATCAGTAATGGGTCGTT	CCTGTTAACCGTTGTTGCTCTT
IAN8	At1g33960	TCAATGTGATTGACACTCCTG	ACTAAGAGCACAGCGTGTAGC
IAN11	At4g09930	TGGCCAAGAAGGTAGAGAAGG	TCTTCGCTGGATTCTTCGTGG
IAN12	At4g09940	AGAGTTCAACGCTACCCAATG	TGGCGACAGACTAAACAGACC

Each primer pair for the genes listed in the table was tested at various concentrations of forward and reverse primers. First, the forward primer concentration was kept constant i.e 150 nM/150 nM, 150 nM/300 nM and150 nM/600 nM. Next, the forward primer concentrations were varied as well as the reverse primer concentrations; 300 nM/150 nM, 300 nM/300 nM, 300 nM/600 nM, 600 nM/150 nM, 600 nM/300 nM and 600 nM/600 nM. The qPCR components that were used through the qPCR reactions are given in Table

Primer testing for dimer formation and non specific amplification was done simultaneously with primer optimization. No Template Control (NTC) for each primer pair optimized was run. The dissociation curve that formed after amplification was used in the primer testing analysis (refer to results, section 3.2.).

2.3.5.2 ΔΔCT validation experiment

A specific validation experiment was done in order to determine if the downstream ΔΔCt calculation would be valid. A summary of guidelines that were followed in this very vital experiment are provided here:

i. The input of cDNA shall span 5 logs (i.e 20 ng to 0.001 ng) and the expression levels of all targets

ii. Run triplicates for each standard curve point in order to determine the precision of the assay.

iii. The target and endogenous control reactions shall run in separate wells

iv. The primer concentration shall be adjusted to 300nM for both forward and reverse primers.

2.3.6 SA treatment of wild type Arabidopsis plants

Arabidopsis thaliana Col-0 were grown under standard conditions aforementioned in section and treated with SA (100 µM) as a way of mimicking pathogen infection. The SA was exogenously administered to plants by spraying using a simple laboratory hand sprayer of approximately 500 ml. Three biological replicates were treated with the same solution of SA. Leaves from each of the respective biological replicate were harvest at 0 hour (calibrator/untreated), 24hours (treated) and 72 hours (treated) post SA treatment and stored at -80°C.

3 Results

Research on peroxisome-mediated functions of plant innate immunity is a new research focus in the Reumann group, initiated by the PhD student, Amr Kataya. The goal of the present M. Sc. thesis was to perform subcellular localization analysis, relative quantification of gene expression by qPCR and co-expression analyses. Subcellular localization studies were done for four newly predicted PTS1 protein candidates, namely NUDT7, NUDT15, CHAT homolog and ATP-BP (see section 1.4). The methodology of qPCR was established in the group and used to determined relative SA-inducible expression of three putative pathogen defence genes of the family of NHL proteins (NHL4, NHL6, and NHL25) and three genes of the family of IAN proteins (IAN8, IAN11 and IAN12). The comparative Ct method was used as the relative quantitative method of choice with β-actin (ACT2), PR2 and PR5 as endogenous control and positive marker genes for SA treatment, respectively. Lastly, about 30 Arabidopsis defence proteins (including the proteins of interest, i.e., NHL4/6 and IAN8/11/12) were subjected to co-expression analyses.

3.1 cDNA subcloning and subcellular localization analysis of four PTS1 protein candidates

3.1.1 Bioinformatic analysis of the four PTS1 candidates

Bioinformatic analyses of the four PTS1 protein candidates of interest were carried out in order to obtain an overview of the different possible protein variants that could be synthesized from each gene of interest. One gene can result in transcriptional and translational protein variants that often differ in subcellular localization. Therefore, bioinformatics analysis was a crucial step prior to cDNA subcloning in expression vectors during subcellular localization studies.

NUDT7 (AT4G12720) has four splice variants according to data obtained from publicly available database, The Arabidopsis Information Resource (TAIR; http://www.arabidopsis.org/) (data not shown). In general, eukaryotic gene models consist of three parts; the untranslated region (UTR), introns and exons. All four gene models of NUDT7 have the same last four exons. Three gene models have the same total number of exons but differ in their respective 5`- and 3`- UTR. The fourth gene model has one additional

exon compared to the others and, additionally, has also a different and shortened UTR. cDNA from any of the four gene models can be used in subcellular localization analysis of NUDT7.

NUDT15 (AT1G28960) has five gene models (Fig. 3.1). Two gene models (AT1G28960.2 and AT1G28960.4) have the same exons with respect to number and size. The other three gene models are clearly different from the first 2 but identical. Apparently they differ in their respective 3` UTR. However, UTR differences among gene models do not affect the transcriptional and translational protein variants that are synthesized by the gene in question. Furthermore, NUDT15 gene models were analysed at amino acid sequence level (Fig 3.2). This analysis was done so that similarities and differences in PTS1 domains could properly be understood. Gene models AT1G28960.2 and AT1G28960.4 have the same predicted PTS1 domain and PTS1, whilst AT1G28960.1, AT1G28960.3 and AT1G28960.5 also have a different but common C-terminus that lacks the prediction of a PTS1 domain (see Table 1.1). Taken together, gene models AT1G28960.2 and AT1G28960.4 were likely to target peroxisomes, whereas, AT1G28960.1, AT1G28960.3 and AT1G28960.5 are likely to be cytosolic protein variants. Gene models AT1G28960.2 and AT1G28960.4 were selected for subcloning because of their higher probability of being localized in peroxisomes.

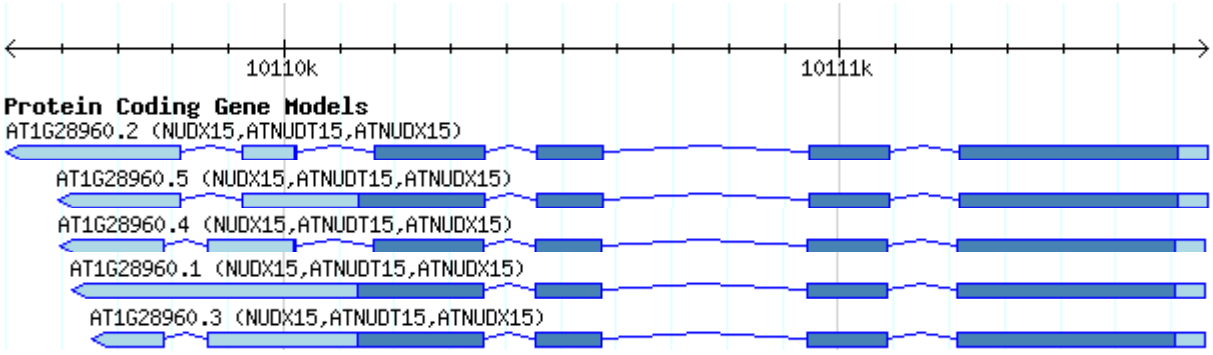


Figure 3.1: NUDT15 gene models. NUDT15 has five gene models. Gene models (AT1G28960.2 and AT1G28960.4) are identical. The other remaining gene models (AT1G28960.1, AT1G28960.1 and AT1G28960.5) are also identical.

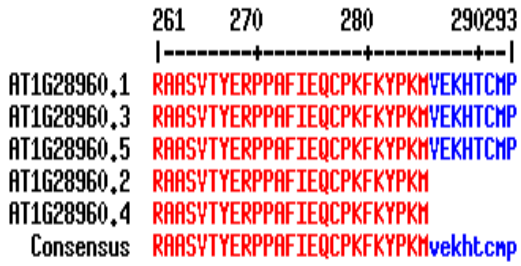


Figure 3.2: PTS1 domain analysis of NUDT15. The sequence alignment of five NUDT15 protein variants at the amino acid level was made in order to compare the different domains. The PTS1 domains of AT1G28960.2 and AT1G28960.4 have the same PTS1 tripeptide sequence (PKM>) and also AT1G28960.1, AT1G28960.3 and AT1G28960.5 have the same C-terminal tripeptide (CMP>), however, this domain is not predicted as a PTS1 domain (see Table 1.1).

The CHAT homolog (AT5G17540) has only one and a very simple gene model; one intron, and two exons (data not shown). The second exon is larger than the first one. The CHAT homolog was predicted with high probability to be localized in peroxisomes (see Table 1.1). Unlike the CHAT homolog, ATP-BP has two gene models (AT1G72840.1 and AT1G72840.2) (Fig. 3.3). Gene model AT1G72840.1 was predicted to be likely peroxisomal more than AT1G72840.2 (see Table 1.1). Nevertheless, prediction scores for both gene models were below peroxisomal localization threshold score. Gene model AT1G72840.1 has a predicted PTS1 (SSL>) whilst AT1G72840.2 lacks a predicted PTS1 tripeptide at the C-terminal. In conclusion, AT1G72840.1 was selected for use in subcellular targeting studies of ATP-BP since it was predicted to be peroxisome than AT1G72840.2.

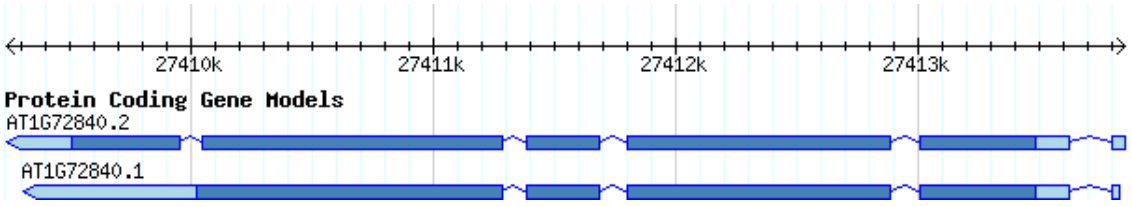


Figure 3.3: ATP-BP gene models. ATP-BP has two gene models which differ at the 3` end of their models. ATP-BP full length protein is encoded by AT1G72840.2 which has an extra exon and intron 3` end AT1G72840.1.

3.1.2 Cloning of the four PTS1 protein candidates into pCAT via pGEM-T Easy vectors

3.1.2.1 cDNA verification and amplification

Purchased plasmids with the cDNAs of interest were supplied either as bacterial stocks or as already isolated plasmids. The purchased plasmids were verified to contain the correct inserts by running a PCR with gene-specific primers and analyzing the PCR products by agarose gel electrophoresis (Fig. 3.4). The approximate sizes of the cDNAs were NUDT7 (800 bp), ATP-BP full (3000 bp), ATP-BP c-terminal (1200 bp), NUDT15 (850 bp) and lastly CHAT homolog (1400 bp). Purchased plasmids were verified to contain correct cDNAs of interest, which were then subcloned into pCAT via pGEM-T Easy vectors.

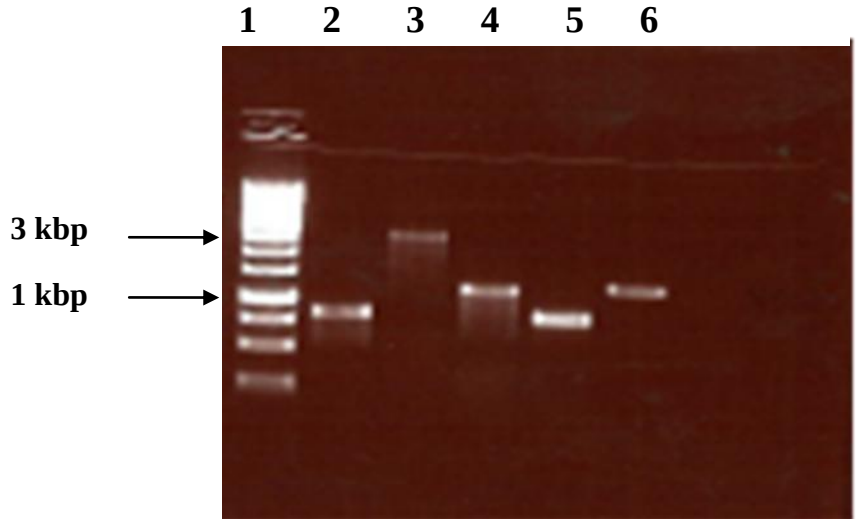


Figure 3.4: cDNA insert verification. PCR was done to check if the purchased plasmids contained the correct insert of the PTS1 protein candidates. 5 µl of each PCR product were used for agarose gel electrophoresis. Loaded in lane 1 was 1Kb DNA ladder; Lane 2 was NUDT7 (800 bp) Lane 3 ATP-BP full (3000 bp); Lane 4 is ATP-BP c-terminal (1200 bp); Lane 5 is NUDT15 (850 bp); Lane 6 is CHAT Homolog (1400 bp).

After verifying the inserts in purchased plasmid, second PCR was carried out in order to amplify the cDNA that was later on cloned into the open pGEM-T Easy vector. . Respective gene-specific primers of cDNA were again used just like in the preceding PCR. To check if PCR products were formed by their respective gene-specific primers, analytical agarose gel electrophoresis was run before preparative agarose gel electrophoresis. Preparative agarose gel electrophoresis was performed so that the cDNAs of the PTS1 candidates could be extracted and be cloned into pGEM-T Easy (see Fig. 3.5). Indeed, all the four PTS1

candidates (cDNAs) were specifically amplified by their respective primer pairs. NUDT15 had the thickest and brightest gel band formed, hence the highest concentration of all. ATP-BP full-length had the thinnest gel band of around 3 kbp that was compared to the DNA ladder. The gel bands for CHAT homolog, NUDT7 and ATP-BP 420 amino acid c-terminal were similar in size and brightness and formed at their respective marks corresponding to their cDNA sizes. The DNA bands were extracted and purified and successfully cloned pGEM-T Easy vector.

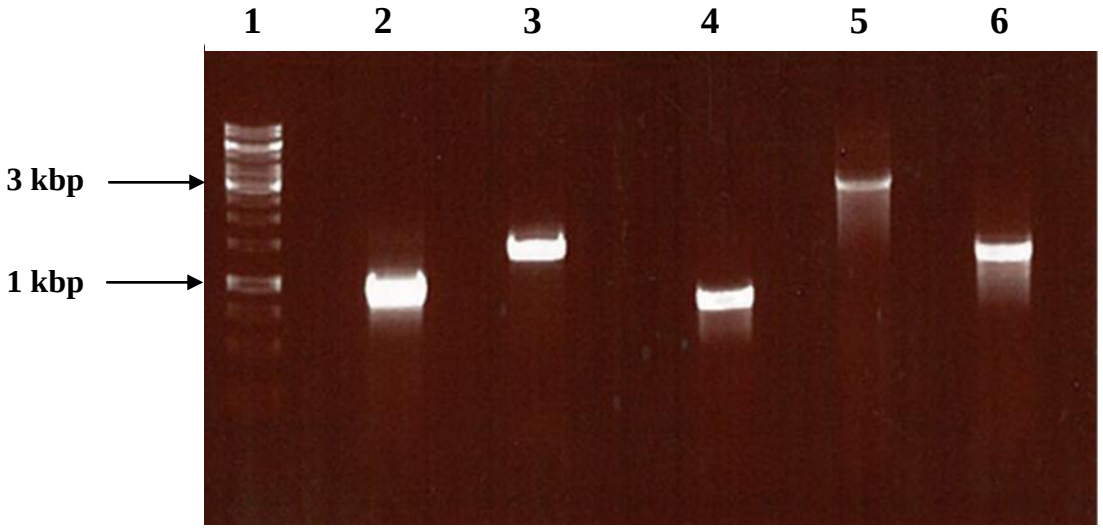


Figure 3.5: PCR-amplified cDNA isolation. 15 µl of PCR amplified cDNA was loaded on an agarose gel. The 1 kb DNA ladder was loaded in lane 1, and lane two to six was NUDT15, CHAT homolog, NUDT7, ATP-BP full length and ATP-BP 420aa c-terminal.

3.1.2.2 PCR fragment isolation

PCR fragment isolation involves agarose gel electrophoresis and PCR fragment extraction from the gel slices. Agarose gel electrophoresis was run with 15 µl PCR product. The aim was to isolate cDNA for NUDT7, NUDT 15, ATP-BP (full length and 420aa C-terminal) and CHAT homolog on the agarose gel. The insert bands were separated on the agarose gel. Therefore, they were then extracted using GE Healthcare GFX PCR DNA and Gel band purification kit (see section 2.1.5).

1 µl of purified PCR product was used to check if the DNA band had been successfully extracted from the gel (data not shown). The agarose gel electrophoresis was also used to determine the DNA band extracted concentration. It was important to precisely determine the

cDNA concentration because an insert to vector molar ratio of 3:1 was required in the DNA ligation step.

Concentration was determined as:

(1) NUDT 7	=	20 ng/µl
(2) ATP-BP full length	=	15 ng/µl
(3) ATP-BP 420aa C-terminal	=	20 ng/µl
(4) NUDT 15	=	20 ng/µl
(5) CHAT homolog	=	20 ng/µl

After the DNA concentration was determined by agarose gel electrophoresis, the isolated PCR fragments were ligated into pGEMT-Easy overnight via TA-cloning. The masses of respective cDNAs subcloned into pGEM-T Easy were 14.4 ng of NUDT 7, 53.3 ng of ATP–BP full-length, 21.4 ng of ATP–BP 420aa, 14.6 ng of NUDT 15 and 23.6 ng of CHAT homolog. 17 ng of pGEM-T Easy was used in all TA-cloning experiments. pGEM-T Easy vector has a Thymine (T) base over hang and the cDNAs when digested with restriction enzymes NotI and XbaI/SacII have an Adenine (A). Therefore, when cDNA with such an over-hang is ligated into the open pGEM-T easy vector, the process is called TA-cloning.

3.1.2.3 Directional cloning into pCAT via pGEM-T Easy vector

E. coli cells were transformed with pGEM-T Easy vector containing the respective cDNA inserts of the four PTS1 protein candidates. Blue-white screening and ampicillin resistance to were used to select E.coli cells that were transformed with pGEM-T Easy containing the cDNA of interest. The blue-white screening technique allows for the quick and easy detection of successful DNA ligation, without the need to individually test each colony. The competent cells were grown in the presence of X-gal (5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside). X-Gal yields colourless and blue products once hydrolyzed by β-galactosidase. A bacterial colony was white if the DNA ligation was successful and blue if ligation was not successful. Colony PCR was done on selected white bacterial colonies to verify the presence of the insert, and then the plasmids were isolated from the successful colony after an overnight incubation in LB medium. To ensure that pGEM-T Easy contained the correct insert, in addition to colony PCR, double restriction enzyme digestion was also done (see next section 3.1.2.5). Indeed, bands for the inserts and backbone vector formed at the correct position. That was also used in double-checking that cDNA of interest had been cloned into

pGEM-T Easy. After verifications that the insert were correct, the samples were sent out for sequencing in Germany to check if there is any mutations. T7 primer was used for sequencing (see section 2.1.6). None of the inserts had any mutations at amino acid level (see section 7.1). cDNA sequences were then ready to be subcloned into the final destination expression vector pCAT-EYFP.

The successfully cloned inserts of PTS1 protein candidates into pGEM-T Easy were digested out from the plasmid after plasmid amplification using *E.coli* cells. Two restriction enzymes were used to cut out the insert from pGEM-T Easy. Restriction enzymes NotI and XbaI were used to digest pGEM-T Easy vector containing NUDT7, NUDT15 and CHAT homolog. NotI and SacII were used to digest the pGEM-T Easy vector containing the ATP-BP insert. The inserts were isolated on an agarose gel. Indeed, all the inserts were isolated from the backbone pGEM-T Easy vector except for ATP-BP full length (Fig. 3.6). ATP-BP full length cDNA and pGEM-T Easy both have the same size (3 kbp), therefore only one band was formed in Lane 3 of the figure shown below.

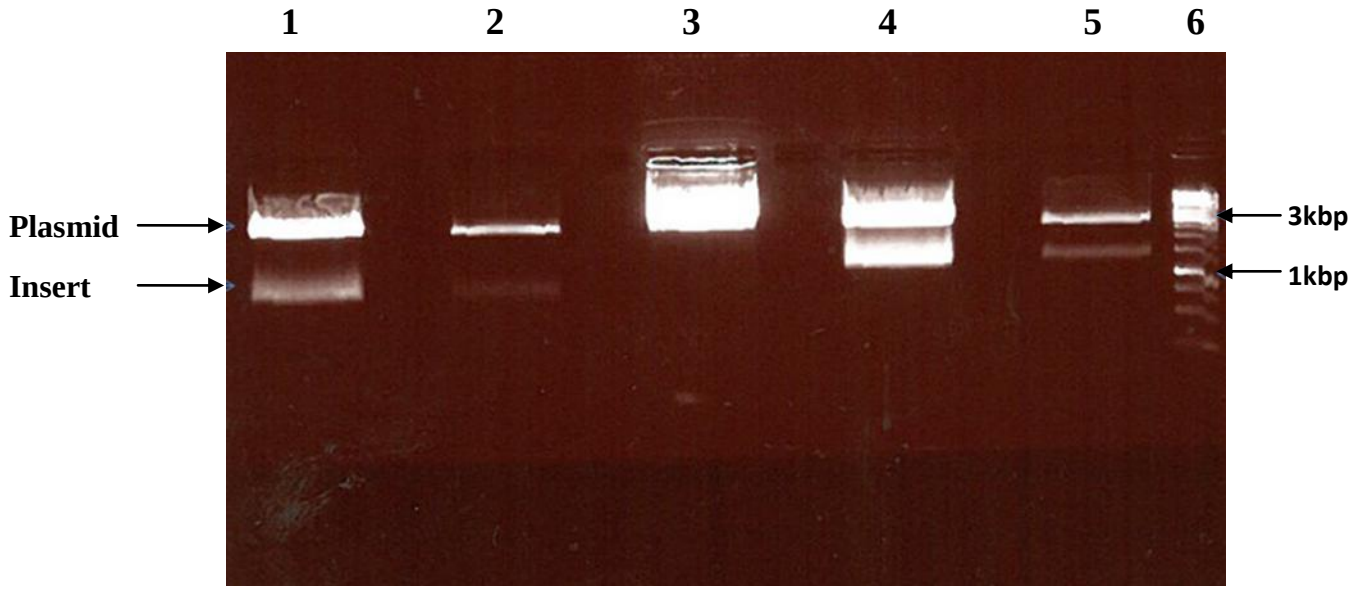


Figure 3.6: cDNA isolation from the pGEM-T Easy vector. The cDNAs were cut out of pGEM-T Easy by using the restriction enzymes NotI/XbaI and NotI/SacII. 20 µl of double digest was separated by agarose gel electrophoresis. Lane 1 to 6 were loaded with NUDT7, NUDT15, ATP-BP full length cDNA, ATP-BP 420aa, CHAT homolog and 1kb DNA ladder (marker) respectively. All lanes had a clear-cut distinction of bands formed between the plasmid backbone and the insert except for lane 3 where ATP-BP (3.1 kbp) and plasmid (3 kbp) formed a single band.

pCAT-EYFP-Hin1R is the vector construct that was used as final destination expression vector that was used subcellular targeting analysis of this MSc thesis (see Fig. 2.2). EYFP is the reporter gene that was N-terminally located to the PTS1 protein candidates. Hin1R was the insert that had to be digested out from the vector so that the cDNA of interest could be expressed as a fusion protein with the report protein (EYFP). The volume of the expression vector needed to be amplified by transformation of *E. coli* cells and then cut out the Hin1R insert so that cDNA of interest could be subcloned into it.

Restriction enzyme double digestion of pCAT with NotI, XbaI and SacII was carried out. NotI and XbaI was the pair of restriction enzymes used to double digest cDNAs of NUDT7, NUDT15 and CHAT homolog. SacII was used instead of XbaI in restriction double digestion of ATP-BP cDNAs. Separate proper bands of pCAT vector backbone and the cut-out Hin1R fragment could be visualized on the gel (data not shown).

The isolated cDNA inserts of PTS 1 proteins were then subcloned into the pCAT-EYFP vector using a directional cloning strategy. NUDT 7, NUDT15 and the CHAT homolog were ligated into pCAT via NotI and XbaI. ATP-BP 420aa was cloned into pCAT-EYFP via Not I and SacII. Ligation of inserts into the pCAT vector was carried out at a stoichiometric ratio of vector to insert of 1:3 like in TA-cloning. *E.coli* cells were then transformed with pCAT-EYFP-PTS1_cDNA. No blue-white screening was done, only colony PCR was used to screen for positive transformants because pCAT plasmid lacks the Lac operon. pCAT plasmids from positive transformants were then isolated and sent for sequencing (see section 7.1.). Only the ligation region between EYFP and inserts was sequenced. Sequencing results showed that all four inserts were in correct orientation and that a fusion protein would be expressed. In addition, no mutations were introduced into the insert cDNAs. The pCAT expression vectors were then ready to transiently express the fusion protein (i.e. EYFP and cDNAs of interest) in onion epidermal layer cells (see section 3.1.3).

3.1.3 *In vivo* validation of predicted peroxisome targeting

Onion epidermal cells were transformed biolistically with EYFP fusion constructs that were C-terminally extended by one of the four PTS1 protein candidates. Subcellular targeting was analyzed by fluorescence microscopy after about18 h expression time at room temperature. An additional 3 to 6 days expression time at 4°C was added for some constructs that did not target any subcellular structures under standard conditions. The EYFP fusion proteins for the CHAT homolog and NUDT15 targeted punctuate subcellular structures which were confirmed to be peroxisomes in double labelling experiment by using a DsRed-SKL as a peroxisomal marker (Matre et al., 2009). EYFP-NUDT7 and EYFP-ATP-BP_420aa did not target any subcellular structures. They remained in the cytosol of the onion epidermal cells. Cytosolic and peroxisomal constructs were reproducibly confirmed either cytosolic or peroxisomal proteins after both short- and long-term expression. EYFP a cytosolic fluorescent protein was used as a negative control to check if there was bacterial contamination. EYFP remained in the cytosol and did not target any subcellular structures whatsoever.

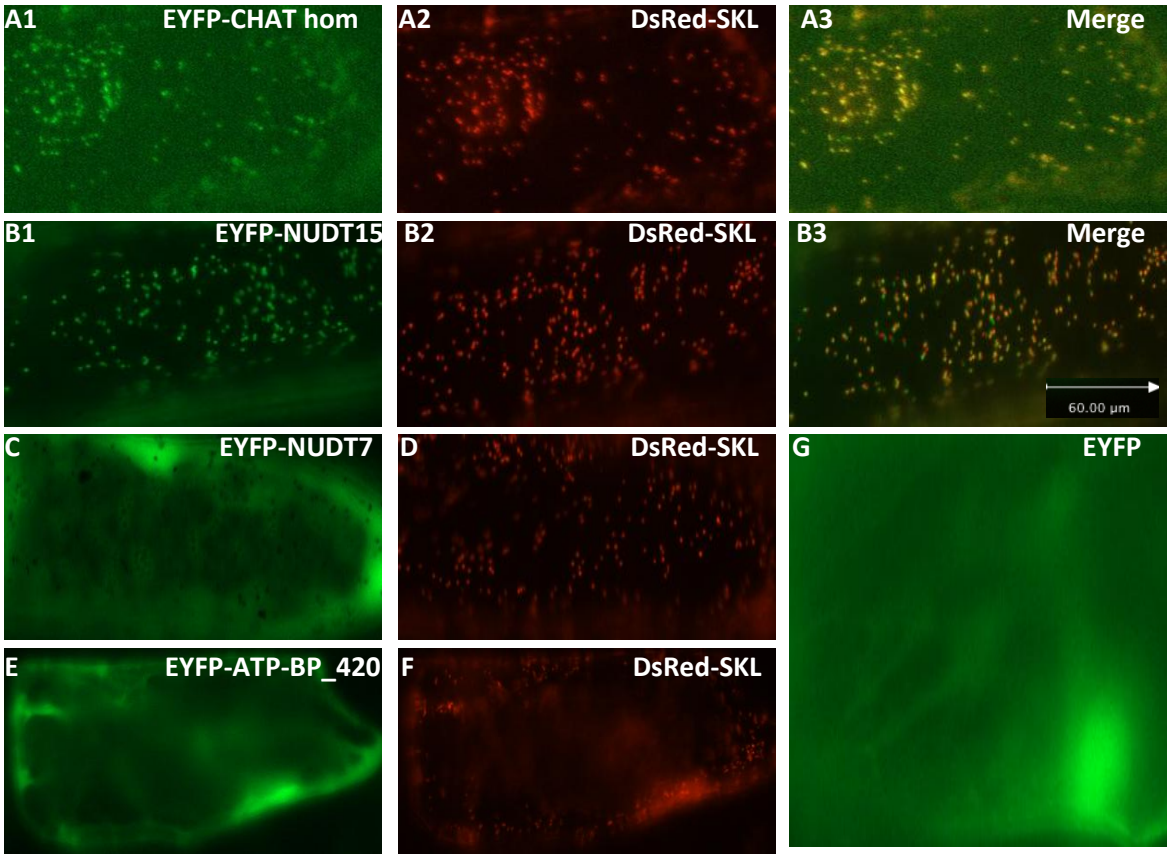


Figure 3.7: *In vivo* subcellular targeting analysis of the four PTS1 protein candidates. Onion epidermal cells were transformed biolistically with EYFP fusion constructs that were C-terminally extended by four PTS1 protein candidates. Subcellular targeting was analyzed by fluorescence microscopy after about18 h expression at room temperature only for ([A1-3], and [B1-3]) or after an additional 3 to 6 days at 4°C for (C, and E). EYFP-CHAT homolog [A1] and EYFP-NUDT15 [B1] targeted punctuate subcellular structures (green dots) that were confirmed to be peroxisomes (yellow dots) by using a peroxisomal marker, DsRed-SKL ([A2] and [B2]) in merge ([A3] and [A3]). EYFP-NUDT7 [C] and YFP-ATP-BP_420aa [E] did not target any subcellular structures, whilst DsRed-SKL (D and F) targeted punctuate structures. Cytosolic constructs [C], and [E] were reproducibly confirmed as cytosolic proteins after both short- and long-term expression. EYFP alone, a cytosolic fluorescent protein [G] was used as a negative control.

3.2 Gene expression analysis of selected NHL and IAN proteins

3.2.1 Bioinformatics analyses of NHL proteins

NHL4, NHL6 and NHL25 (NHL4/6/25) are homologs of NDR1, one of the major pathogen defence protein (Knepper et al., 2011). Bioinformatics analysis on NHL4/6/25 was done so that the function and stimuli for these NDR1 homologs could be know. Conserved domain analysis and various stimuli for NHL4/6/25 were investigated using publicly available databases.

Conserved domains analysis for NHL4/6/25 using from publicly available database, National Centre for Biotechnology Information (**NCBI; http://www.ncbi.nlm.nih.gov)** was carried out. These selected NHL proteins have similar conserved domains called the LEA (Late Embryogenesis Abundant) domains (Fig. 3.8 and appendix section 7.2). The LEA domain has been implicated to have a role during various stress conditions such as dehydration.

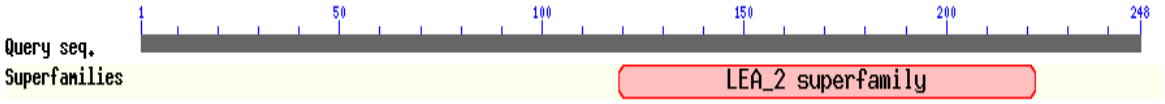


Figure 3.8: NHL25 Conserved domains. NHL25 has one conserved domain called the LEA_2 superfamily (Late Embryogenesis Abundant) located at the C-terminus.

Since, the NHL proteins have the LEA superfamily domain that was reported to be involved in dehydration stress responses (Knepper et al., 2011), gene expression of these proteins is might be stress related. Genevestigator (www.genevestigator.com) was used to search for other stress conditions that would cause the NHL genes of interest to be expressed. Indeed, several conditions were found to either up-regulate or down-regulate the gene expression levels of NHL 4 and NHL6, for NHL25 no microarray data were available. The conditions included both biotic and abiotic stress such as bacterial infection and plant defence hormones. SA treatment was analysed further in detail since it`s the signaling molecule in pathogen defence pathway where NDR1is found interacting with R proteins and RIN4 (Panstruga et al., 2009; Knepper et al., 2011)(Fig. 3.9). Only gene expression of one gene has been presented, nonetheless all our genes of interest were analysed in the same way, provided that their

microarray data was present at Genevestigator. NHL6 gene expression was up-regulated or down-regulated when Arabidopsis plants of different ecotypes were treated with SA. The observed NHL6 gene expression regulation was dependant on, (1) how SA treatment was done. (2) SA concentration, (3) SA treatment duration, (4) Developmental stage of the plant used in the assay and (5) the part (anatomy) of plant used. Other NHL and IAN genes were also regulation by SA treatment (data not shown). In conclusion, in order to validated the bioinformatics analyses carried out, gene expression studies by qPCR method on SA treated plants had to be carried out.

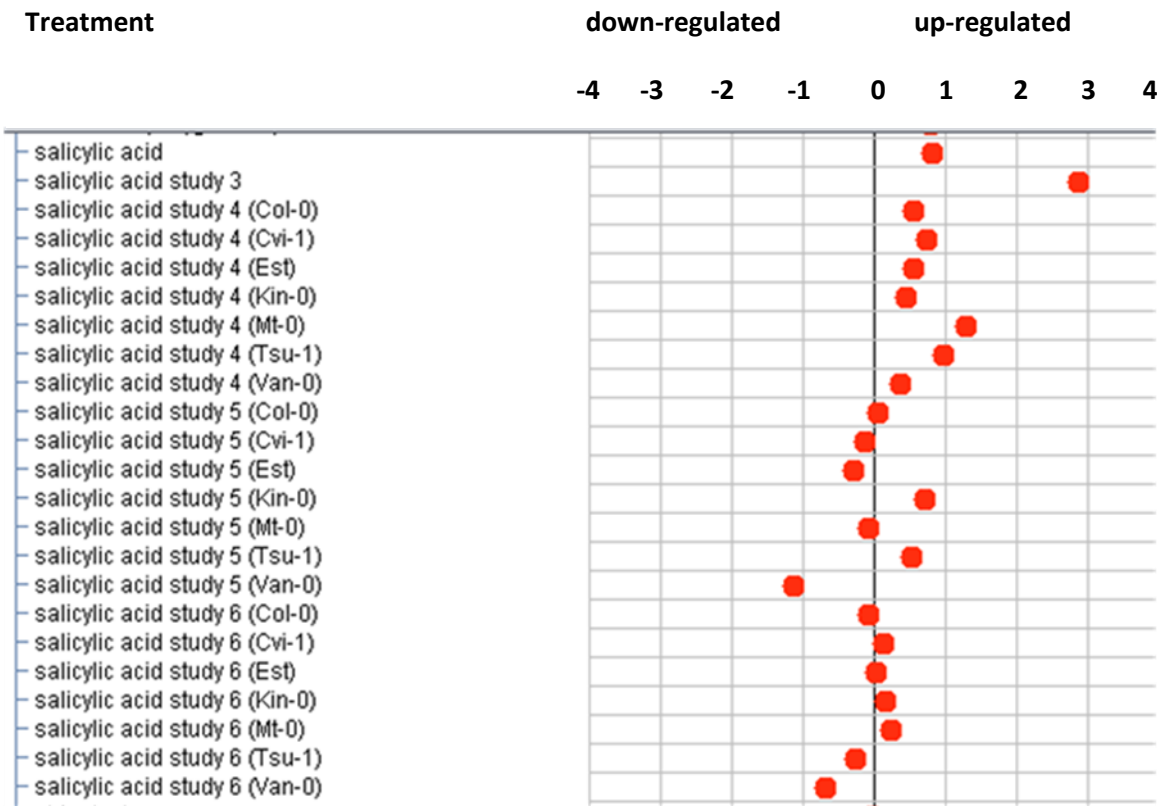


Figure 3.9: NHL6 gene expression analysis. SA treatment gene expression conditions were analysed using Genevestigator. NHL6 gene expression was determined by the way SA treatment was carried out on plants.

3.2.2 RNA isolation and cDNA synthesis

RNA that was isolated needed to have quality and quantity assessment before it could be used in cDNA synthesis. RNA quality was categorized into two i.e RNA integrity and RNA purity. RNA quantity was based on the RNA yield obtained as a result of RNA extraction process. RNA yield was in the range between 12 and 30 µg. RNA integrity was assessed to ensure that samples being compared were of similar integrity. Agarose gel electrophoresis was used for isolated RNA integrity assessment (data not shown). All the RNA used in this study was of good integrity. RNA sample with good integrity had discrete, though thick 28S:18S ribosomal RNA (rRNA) gel bands and in an approximate mass ratio of 2:1.

RNA samples may get contaminated with proteins, phenols and other molecules that interfere with downstream applications. RNA absorbance is measured at a wavelength of 260 nm and proteins absorb at 280nm. All RNA samples used had an A_{260}:A_{280} ratio of >2.0. Pure RNA samples should have an A_{260}:A_{280} ratio of ~2.0 hence the RNA samples used in assays had of high purity levels.

The two-step reverse transcription method was used to synthesize cDNA (refer to section 2.2.3 in materials and methods). In two step method first cDNA was synthesized from RNA and the qPCR later. RNA input concentration was always 100 ng/µl resulting in 50 ng/µl of cDNA synthesized each time reverse transcription was done. Random primers were primers of choice, in order to reverse transcribe all mRNAs (that is targets and endogenous control gene). The High-Capacity cDNA Reverse Transcription kits were used in cDNA synthesis. Synthesized cDNA was diluted to 10ng/µl in most cases and then used in qPCR assays.

3.2.3 Validation experiments

SYBR® Green I chemistry was used in qPCR. SYBR® Green I is a fluorescent dye that binds to the minor grooves of double-stranded DNA (dsDNA). It fluoresces when bound to dsDNA and is excited by a light source. Because SYBR® Green I dye also binds and fluoresces when bound to primer dimers and non-specific amplification products, additional optimization steps were needed to ensure robust amplification and accurate quantitation.

All primer sets that used in qPCR had to first be optimized. Primer optimization was done so that the best primer pair concentrations could be determined and later on be used in the downstream qPCR experiments. Ideally, good primer pairs give relatively low threshold cycle (C_T) value and high delta normalized fluorescence (ΔRn). C_T is the PCR cycle number where the SYBR Green dye fluorescence crosses the baseline. ΔRn is defined as the fluorescence from SYBR Green dye minus the baseline. Primer pair concentrations of 300 and 600nM gave low C_T and high ΔRn (Fig. 3.10). 150 nM primer pair concentrations gave high Ct and low delta normalized fluorescence. Therefore, in all follow-up qPCR experiments shall have minimum primer concentration of 300 nM for both forward and reverse primers. This shall ensure that optimum PCR amplification efficiency of approximately 100%. However, primer optimization shall be coupled with other conditions such as good master mixes, accurate pipeting and Applied Biosystems-approved reagents.

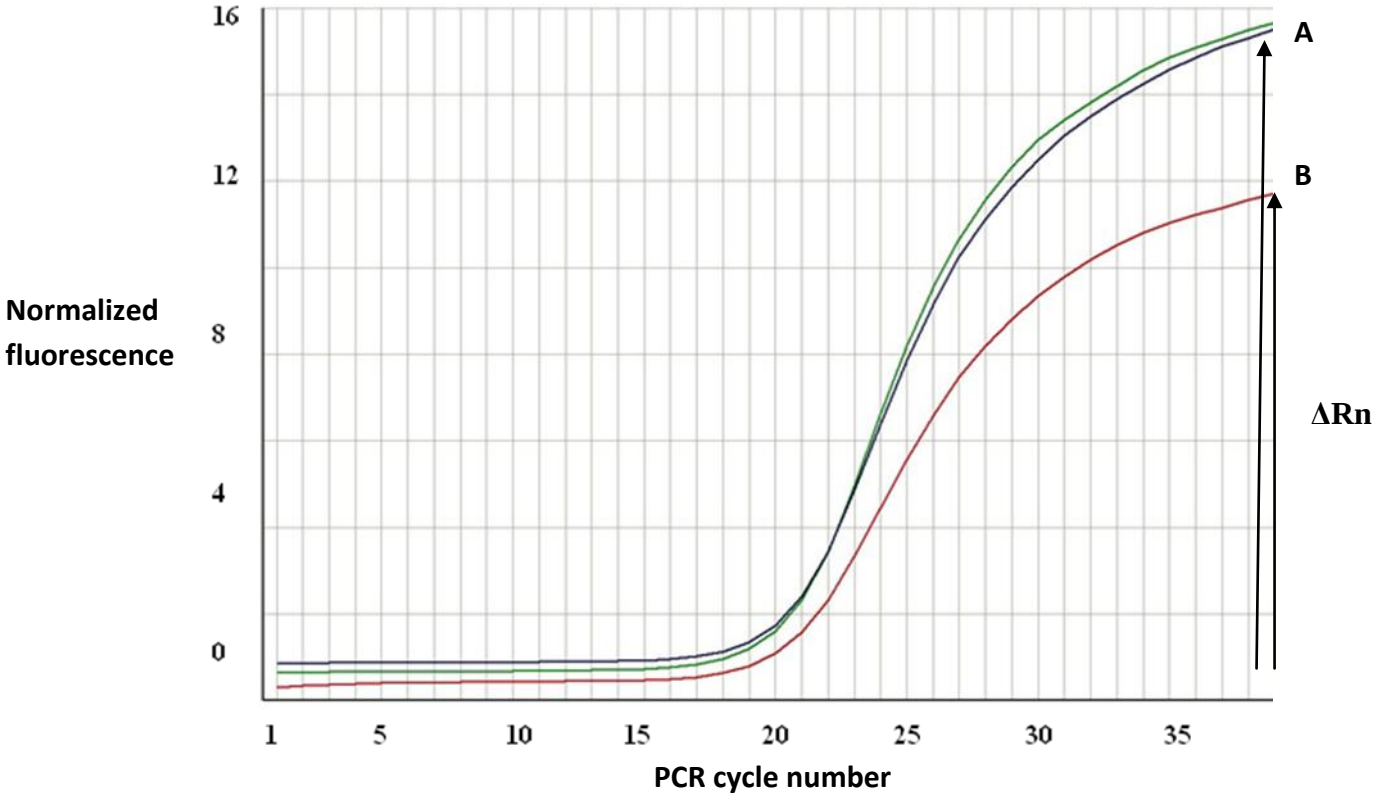


Figure 3.10: qPCR primer optimization. The green and blue curves represent primer pair concentrations of 600 nM and 300 nM respectively. The red curve represents a primer pair concentration of 150 nM. Curves for 600 and 300 nM marked with the letter A, had the same delta normalized fluorescence (ΔRn). ΔRn is the normalized fluorescence minus baseline. primer pairs of 150 nM concentration marked by letter B had lower ΔRn. Curve A gave lower threshold cycle (C_T) values than B. A good primer set should give low C_T values and high ΔRn, therefore, it was concluded that the optimal primer pair concentrations to use in the follow-up qCPR experiments were 300 nM and/or 600 nM.

The selection of an appropriate endogenous control is one of the most important steps in relative quantitation experimental design. Normalization to an endogenous control ensures that differences in amounts of input nucleic acid template do not affect the quantitation of genes. Any gene shown to be expressed at the same level in all study samples can potentially be used as an endogenous control. Nonetheless, the selection was between two of already established endogenous control genes (Livak and Schmittgen, 2001). The endogenous control genes were actin (ACT2) and Ubiquitin (Ubi10) refer to Table 2.12. ACT2 (AT3g18780) formed a single peak along its melting (dissociation) curve and no amplification in No Template Control (NTC) represented by a green line (Fig. 3.11A). This meant that the ACT2 primer pair were specifically designed to amplify only the ACT2 gene and formed no primer dimers. Ubi10 primer pair was nonspecific but formed no primer dimers.

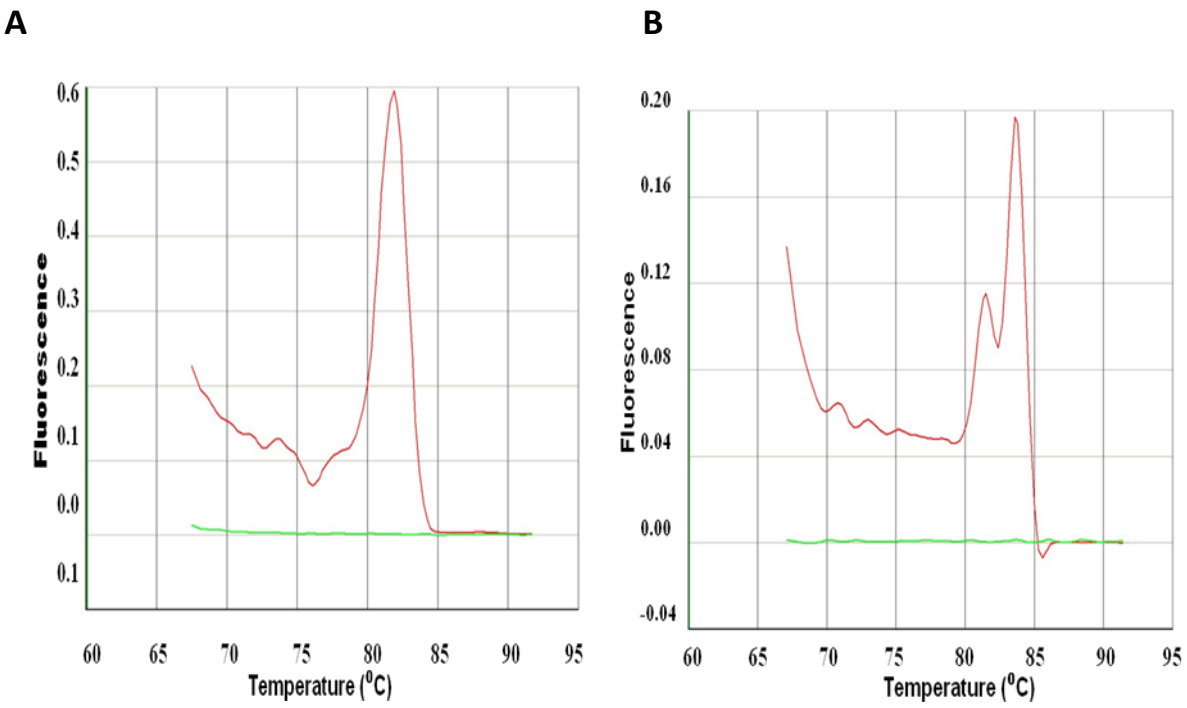


Figure 3.11: Endogenous control gene selection. Specific amplification and primer dimer formation where used as a criterion for a good primer pair for the two endogenous control genes i.e actin (ACT2) and ubiqutin (Ubi10), also see Table 2.12 for more primer information). Green and red curves were for specificity and primer dimer formation tests, respectively. The primer pair for ACT2 (AT3g18780) showed no nonspecific amplification as well as no primer dimer formation (Fig. 3.11A). Ubi10 primer pair showed nonspecific amplification but had no primer dimer formed (Fig. 3.11B).

From primer testing experiment carried out on ACT2 and Ubi10, it was deduced that ACT2 primer pair was better than those for Ubi10, therefore, more validation tests on endogenous control were focused only on ACT2. ACT2 was further tested for uniformity in gene expression in SA treated and untreated samples (Fig. 3.12). ACT2 was found to be uniformly expressed in all samples whether treated or not treated. Taken together, ACT2 gene passed all the endogenous control selection tests and was used, in downstream validation experiments such as comparison of PCR amplification efficiencies of target genes to endogenous control and also in relative quantitation of genes by comparative C_T method.

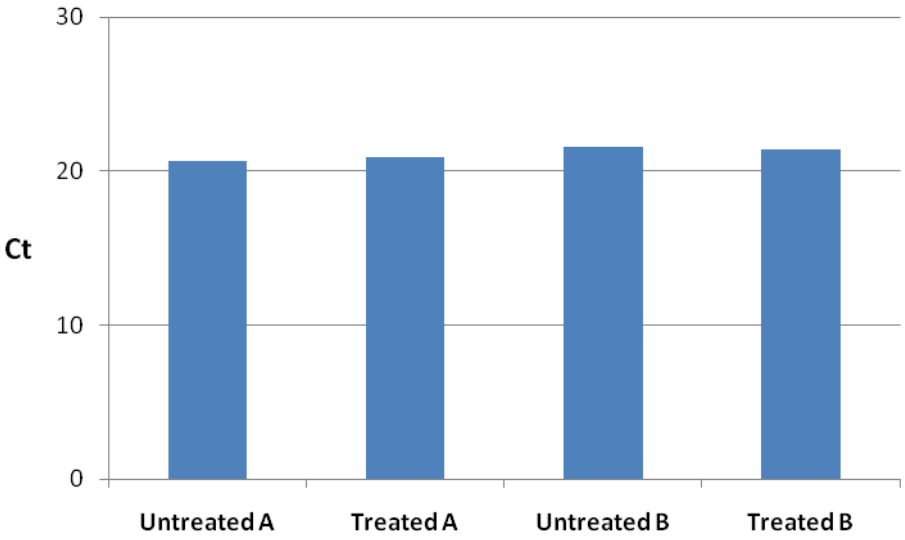


Figure 3.12: Consistency of ACT2 expression level. ACT2 had an average C_T 21±0.3 in all the tested samples. SA treated and untreated samples were tested. SA (100 µM) was exogenously applied to plants and leaves were harvested after 24 hours (see section 2.2.6).

Relative quantitation of genes can be carried out by either Standard relative curve or comparative C_T method. In this present study, comparative C_T method was the method of choice that was used in gene expression analsyis (see section 3.2.5). When performing the comparative C_T method, the target(s) and endogenous control have to have similar or relatively equal PCR efficiencies. The slope of a standard curve was used to estimate the PCR amplification efficiencies of individual genes (Fig. 3.13A and B) or a comparison between ACT2 and target genes i.e. NHL4/6/25 and IAN8/11/12(Fig. 3.13C and D). A real-time PCR standard curve was graphically represented in two ways either as a semi-log regression line plot of; C_T value vs. log of input nucleic acid (Fig. 3.13A and B) or ΔC_T value vs. log of input

nucleic acid (Fig. 3.13C and D). Refer to discussion section 4.2.4 for details on how calculations that were carried out in order to determine the respective PCR amplification efficiencies. A standard curve slope of –3.32 for C_T value vs. log of input nucleic acid line plot indicates a PCR reaction with 100% efficiency (Fig. 3.13B). Additionally, standard curve with a slope of <0.1 for ΔC_T value vs. log of input nucleic acid line plot indicated the two genes (target and endogenous control gene) had an approximately equal efficiency (Fig. 3.13D).

Figures 3.13A and C represents amplification efficiencies of NHL4 alone and a comparison of amplification efficiencies NHL4 and ACT2 respectively. NHL4 had a PCR efficiency of less than 100% because the slope of the graph should be -3.32 and NHL4 had a slope of -0.66 (Fig. 3.13A). NHL4 and ACT2 amplification efficiencies were not approximately in the qPCR because the slope was 3.19 instead of <0.1(Fig. 3.13C). PR5 and ACT2 Real-time PCR amplification efficiencies are approximately equal due to the fact that the slope of the semi-regression curve was 0.031 which is < 0.1 (Fig 3.13D). PR2 Real-time PCR amplification efficiency was 100% since the slope of a semi-log regression line plot of Ct value Vs log of was -3.32 (Fig. 3.13B).

In order for the two genes (target and endogenous control) to be used in gene expression relative quantitation, their respective amplification efficincies had be 100±10% and approximately equal. NHL6, NHL25, PR2 and PR5 had PCR amplification efficiency approximately equal to that of ACT2, therefore, these four genes have been fully optimized and ready to be used in their expression analyses. Whilst NHL4, IAN8, IAN11 and IAN12 amplification efficiencies are not approximately equal to ACT2 and need new primer pairs or using an alternative relative quantitation method such as the relative standard curve method.

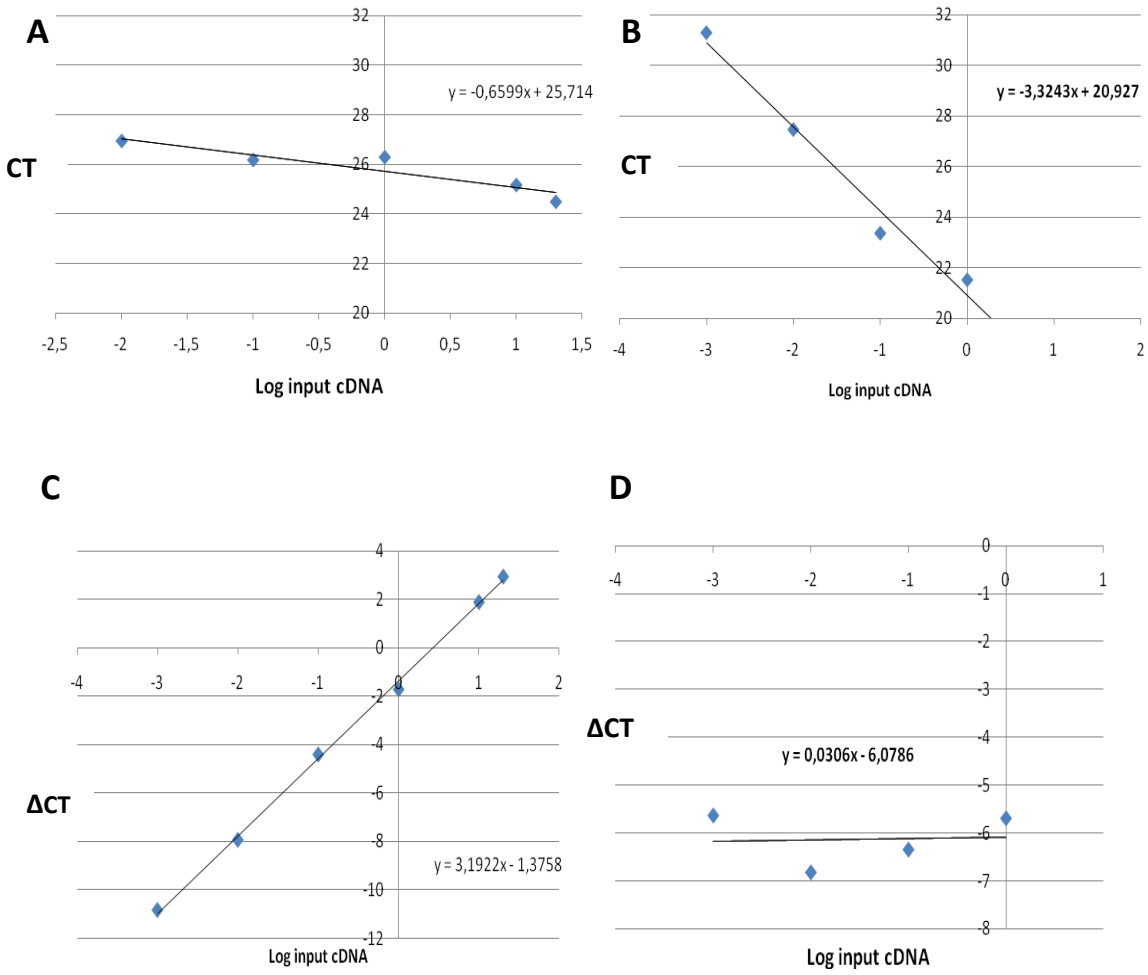


Figure 3.13: qPCR amplification efficiency comparisons.

A real-time PCR standard curve was graphically represented in two ways either as a semi-log regression line plot of; C_T value vs. log of input nucleic acid (Fig. 3.13A and B) or ΔC_T value vs. log of input nucleic acid (Fig. 3.13C and D). Refer to discussion section 4.2.4 for details on how calculations that were carried out in order to determine the respective PCR amplification efficiencies. A standard curve slope of –3.32 for C_T value vs. log of input nucleic acid line plot indicates a PCR reaction with 100% efficiency (Fig. 3.13B). Additionally, standard curve with a slope of <0.1 for ΔC_T value vs. log of input nucleic acid line plot indicated the two genes (target and endogenous control gene) had an approximately equal efficiency (Fig. 3.13D).

3.2.4 Gene expression analysis with qPCR

3.2.4.1 SA treatment of plants

Seven weeks old wild type Arabidopsis plants were treated with 100 µM of SA. The defence hormone treatment was by spraying healthy leaves of three biological replicates. Salicylic acid positive markers were used to confirm the success of the treatment. The positive treatment indicators were pathogenesis related genes (PR2 and PR5). PR2 and PR5 were both induced after treatment with highest recorded fold-change at 24hours post-treatment. PR2 was highly induced than PR5 (Fig. 3.14).

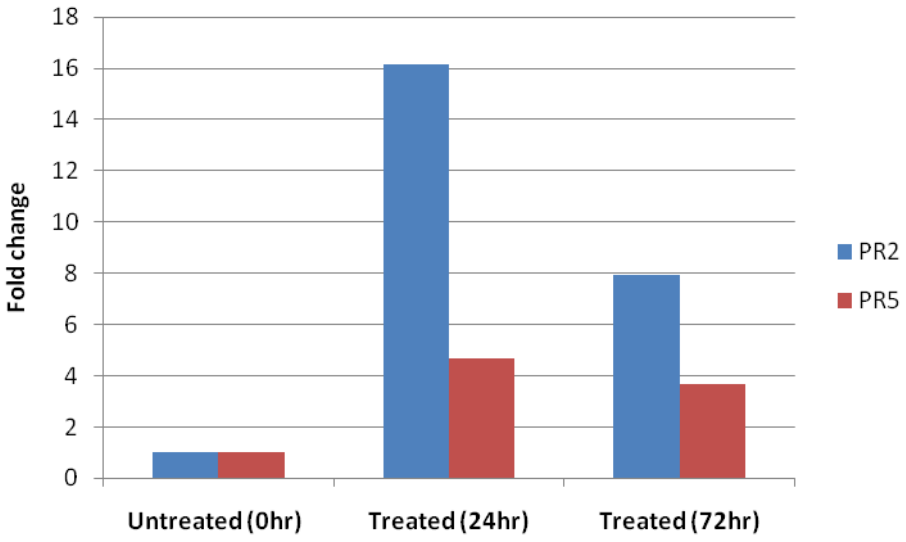


Figure 3.14: PR2 and PR5 (PR2/5) gene expression in three different samples.

PR2 is represented by blue bars whereas PR5 is represented by brown bars. PR2/5 gene expression induction was due to salicylic acid (100µM) treatment of Arabidopsis plants. PR2/5 expression was higher in treated samples than in the untreated sample, with the highest fold change at 24 hours post-treatment. PR2 was highly induced than PR5 in all treated samples with approximately 12 and 4 fold change in 24 and 72 hour treated samples.

3.2.4.2 Relative quantitation of peroxisomal pathogen defence genes

It was evident from the gene expression the salicylic acid (SA) treatment of wild type Arabidopsis plants was successful as SA positive marker genes (PR2 and PR5) were induced (Fig. 3.14). Peroxisomal-targeted pathogen defence genes of interest were also induced (Fig. 3.15). Gene expression levels at 24 hours post treatment was more than before treatment (0 hour) for all the genes tested. However, gene expression levels reduced at 72 hours post treatment for all the genes. IAN12 expression levels had reduced about 15% at 72 hours relative to that at 24 hours. IAN11, NHL4, NHL 6 and NHL25 gene expression levels reduced the most. Only about 30% of these genes relative to their respective expressions at 24 hours were expressed at 72 hours. PR2 gene expression had reduced to about 50% gene expression relative to that at 24 hours. Lastly, PR5 had only 75% gene expression (25% reduction) at 72 hours post treatment (see Fig. 3.14).

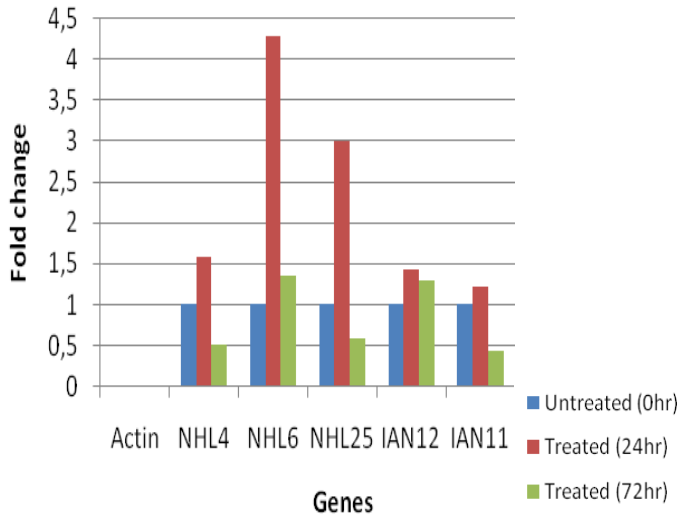


Figure 3.15: Relative quantification of NHL and IAN genes.
10ng/µl of cDNA was used in the relative quantitation of peroxisomal-targeted defence genes. Untreated sample was used as the calibrator in the gene expression qPCR assay of NHL and IAN genes. SA treated samples had NHL and IAN genes expression more at 24 hour posttreatment than at 72hour. Actin was the endogenous control gene hence it was not relatively quantified like the other genes, it was instead used in normalizing the imput cDNA concentration differences among the samples.

3.3 Co-expression analysis

3.3.1 Dataset generation for co-expression analysis

28 Arabidopsis defence proteins were used as queries in the dataset generation for co-expression analysis. Among the 28 queries used was NHL6, NHL4, IAN8, IAN11 and IAN12 (our pathogen defence proteins of interest). Co-expression studies were done the bioinformatics; the Expression Angler (http://bar.utoronto.ca/ntools/cgi-bin/ntools_expression_angler.cgi) and AtGenExpress Pathogen Set. The raw dataset generated was by setting the threshold for correlation score at 0.6 per query (see table 7.3 in appendix). The raw dataset generated then further analysed in order to be able to make proper inferences as to which signal transduction pathways our proteins of interest belong to.

3.3.2 Co-expression groups

28 Arabidopsis pathogen defence proteins were used as queries in co-expression analysis. Correlation score of 0.6 was used as the lowest threshold for co-expression partners. Based on the lowest correlation score used per query, the 28 Arabidopsis defence proteins were grouped into two groups; those highly co-expressed with the higher co-expression score and those not highly co-expressed (Fig. 3.15). 17 pathogen defence proteins were highly expressed (Fig. 3.15A). NHL6 was second highest co-expressed defence protein after RPM1. 8 pathogen defence proteins not highly co-expressed (Fig. 3.15B) and IAN8 was among them. NHL4, IAN11 and IAN12 not co-expressed at all (Fig. 3.15B).

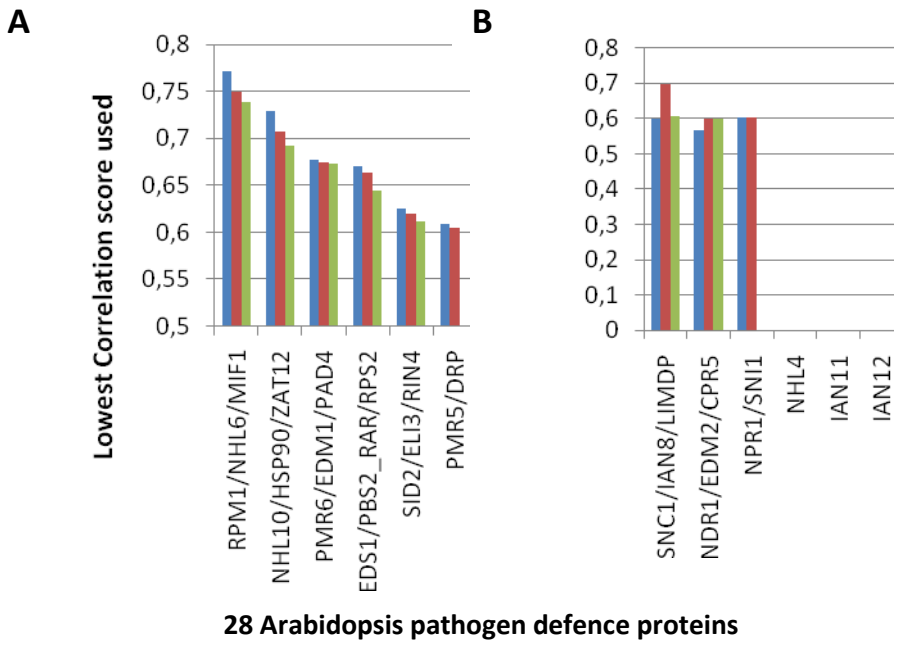


Figure 3.16: Co-expression analysis. 28 Arabidopsis pathogen defence proteins were used as queries in co-expression analysis. Correlation score of 0.6 was used as the lowest threshold for co-expression partners and based on the lowest correlation score used per query the 28 Arabidopsis defence proteins were grouped into two groups; highly co-expression proteins (Fig. 3.16A) and those not highly co-expressed (Fig. 3.16B).

3.3.3 Co-expression patterns of the 28 Arabidopsis pathogen defence proteins

5 defense proteins with a common co-expression protein

RIN4/EDS1/PAD4/EDM1/RPS2

PAD4

4 defense proteins with two common co-expression proteins

RIN4/EDS1/EDM1/RPS2

3 defense proteins (in Red) sharing a common co-expression protein

1.RIN4/EDS1/EDM1 2. RIN4/EDM1/RPS2 3. EDS1/PAD4/RPS2 4. EDS1/PAD4/EDM1

5. RIN4/PBS2/EDM1/RPS2 6. DRP/RIN4/EDM1/RPS2

2 common defense proteins (in Red) sharing a common co-expression protein. In Green is one of our proteins of interest with its co-expression partners. In Blue is NDR1 (NHL4/6/25 homolog) with its co-expression partners

1. IAN8/SID2/EDS1/PAD4 2. LIMDP/EDS1/PAD4 3. NPR1/EDS1/PAD4 4. SID2/EDS1/PAD4

5. PAD4/EDM1/PBS2 6. PAD4/EDM1/PBS2-RAR1 7. DRP/EDM1/RPS2 8. DRP/RIN4/EDM1

9. DRP/RIN4/RPS2 10. EDS1/EDM1/NDR1 11. EDS1/PAD4/NDR1

12. EDM1/RPS2/NDR1 13. EDS1/RPS2/NDR1

Co-expression of our proteins of interest and their close homologs

NDR1	NHL10/NtHin1	NHL6	IAN8
EDS1-NDR1	NHL10/NtHin1-MIF1	NHL6-MIF1	IAN8-SID2
RPS2-NDR1	NHL6-NHL10/NtHin1	NHL6-NHL10/NtHin1	IAN8-ZAT12
EDS1/EDM1/NDR1	NHL10/NtHin1-NHL6	NHL6-SID2	IAN8-NHL6
EDS1/PAD4/NDR1	NHL10/NtHin1-ZAT12	NHL6-HSP90	IAN8-PAD4
EDM1/RPS2/NDR1	NHL10-SID2	NHL6-IAN8	IAN8-CPR5
EDS1/RPS2/NDR1	NHL10-PBS2		

4 Discussion

Peroxisomes carry out metabolic functions important to plants. Much research has been done on the involvement of peroxisomes in lipid metabolism, photorespiration, and hormone biosynthesis. By contrast, very little is known about the role of peroxisome in plant innate immunity (Reumann, 2011). To investigate the recently reported peroxisome function in pathogen defence, gene expression of proteins established to be peroxisomal and homologous to defence proteins was analysed in the present study. Additionally, in order to expand propound more on the peroxisomal pathogen defence role and because of most of peroxisomal matrix proteins that have been identified are PTS1 proteins (Reumann, 2004; Lingner et al., 2011). Four PTS1 carrying proteins previously identified were experimentally validated through *in vivo* subcellular localization analysis.

4.1 Subcellular localization studies

Proteins target the peroxisomes by either PTS1/2 or piggy backing pathways (Reumann, 2004, 2011). Peroxisomal targeting by the PTS1 pathway is dependent in part on the type of PTS1 tripeptide (canonical or noncanonical) and the PTS1 domain (last 15aa residues) (Reumann, 2004; Lingner et al., 2011). The four proteins of interest all carried predicted PTS1 non-canonical (minor) PTS1s (see table 4.1). Often, the targeting efficiency of minor PTS1s depends on enhancing targeting patterns found in the domain immediate upstream of the PTS1. The characteristic features of targeting enhancing targeting patterns include a relative basic PTS1 domain, proline residues and a lack of inhibitory residues upstream of the PTS1(Reumann, 2004; Lingner et al., 2011). Despite carrying predicted non-canonical PTS1 the four proteins were differentially localized in onion epidermal layer cells after being transiently expressed. CHAT homolog and NUDT15 targeted the peroxisomes, whereas NUDT7 and ATP-BP 420aa C-terminal remained cytosolic. The subcellular localization data obtained in this study, for the four PTS1 protein candidates are summarized in Table 4.1 given below.

Table 4.1: Subcellular localization results

Gene locus	Acronym	Annotation	Predicted PTS1	Localization data
At1g72840	ATP-BP 420aa C-terminus	ATP binding protein (Last Exon)	SSL>	Cytosol
At5g17540	CHAT-Homolog	Acetyl CoA: (Z)-3-hexen-1-ol acetyltransferase	SSL>	Peroxisome
At4g12720	NUDT 7	Nudix hydrolase 7	ASL>	Cytosol
At1g28960.2/.4	NUDT15	Nudix hydrolase Homolog 15	PKM>	Peroxisome

4.1.1 NUDT7 is a cytosolic pathogen defence protein in Arabidopsis

NUDT7 has four gene models, which apparently have the same C-terminal domain (last 15 aa residues). They all have the same C-terminal tripeptide (ASL>) that might function as a non-canonical PTS1. NUDT7 has a basic PTS1 domain with an overall charge of +3. Nonetheless, the predicted PTS1 domain lacks residues such as prolines which could help in the enhancing peroxisomal targeting of a protein with a minor predicted PTS1 like ASL> (Reumann, 2004; Lingner et al., 2011).

NUDT7 is a known important pathogen defence protein. It functions as a negative regulator to pathogen defence proteins in Arabidopsis plants (Ge et al., 2007).Latest PTS1 prediction models used such as PWM model, predicted all the four gene models for NUDT7 to be nonperoxisomal (Lingner et al., 2011). Previously, NUDT7 had been reported to be a nonperoxisomal but a cytosolic Arabidopsis protein (Ogawa et al., 2008). Nonetheless, NUDT7 was reported to be a peroxisomal protein in mouse (Gasmi and McLennan, 2001). All this put together, in the quest to identify new peroxisomal targeted defence proteins.

NUDT7 was selected as our PTS1 protein candidate. NUDT7 was successfully cloned in the back of EYFP in pCAT via pGEM-T Easy vector and transiently expressed in onion epidermal cells. NUDT7 subcellular targeting was subsequently analysed by fluorescence microscope after, first 18 hours incubation at room temperature (short-term expression) and then after three to six days cold incubation a 4°C (long-term expression). NUDT7 did not target any subcellular structures after all these varied transient expression conditions. Future subcellular studies on NUDT7 could include creation of Arabidopsis stable lines since only the transient expression was carried out in this current study. Transient expression methods such as tobacco protoplasts should also be carried so that results from both transient expression methods (onion epidermal layer cells and tobacco protoplasts) could be compared. Additionally, C-terminal domain constructs (last 10aa) of NUDT7 such as EYFP-7aa-ASL> should be concomitantly used in subcellular localization analysis with full-length cDNA.

NUDT7 has a basic PTS1 domain and a known functional plant PTS1 but it still failed to target the peroxisomes. The domain lacked known and noticeable enhancing patterns like Proline residue(s) in the upstream region of the PTS1. This shows that basicity of the PTS1 domain alone was not enough to help the minor functional PTS1 to target the protein to the peroxisomes. Nonetheless, NUDT7 has been reported to have a pathogen defence role (Ge et al., 2007). In summary, from the current study, it could not confirm the hypothesis that NUDT7 might be a peroxisomal pathogen defence protein.

4.1.2 NUDT15 is localized in peroxisomes

NUDT15 like NUDT7 had several gene models. It had five gene models that were categorized into two groups; those with a predicted PTS1 and the other group lacking a predicted PTS1. Two gene models have apparently the same predicted PTS1 domain (last 15 aa residues) and the other three gene models also have the same domain. Only two gene models carried a C-terminal tripeptide that might function as a noncanonical PTS1 (PKM>) (see Fig. 3.2). The two gene models with a noncanonical PTS1 have a relatively basic PTS1 domain with an overall charge of +2 (refer to section 1.4). The PTS1 domain also contains residues such as proline and glutamic acid upstream of the PTS1. Proline residues are predicted to enhance peroxisome targeting (Reumann, 2004; Lingner et al., 2011). Residues such as glutamic acid were previously predicted and reported to have an inhibitory effect on peroxisomal targeting (Ma and Reumann, 2008). Nevertheless, PTS1 domain was recently predicted and reported to be very relaxed, implying that it takes a combination of a number of targeting enhancing or inhibitory patterns for a minor PTS1 to localize a protein (Lingner et al., 2011). The relaxation of the PTS1 domain was evident in NUDT15, despite of the presence of acidic residues in the PTS1 domain the noncanonical PTS1 (PKM>) to target the peroxisomes.

Subcellular localization predictions of NUDT15 gene models were made using the PWM model (Lingner et al., 2011). Gene models AT1G28960.2 and AT1G28960.4 were predicted with higher scores to be likely peroxisomal than the other gene models; AT1G28960.1, AT1G28960.3 and AT1G28960.5. Nonetheless, all the five gene models scores were below threshold for predicted peroxisomal proteins by PWM model but within grey zone (Lingner et al., 2011). Based on the prediction scores and the presence of a recognizable functional plant PTS1, gene models AT1G28960.2 and AT1G28960.4 were selected to be used in the current study of subcellular localization studies for NUDT15.

NUDT15 was previously characterized to be a mitochondrial protein (Ogawa et al., 2008). In this study using the cDNAs from AT1G28960.2 and/or AT1G28960.4, NUDT15 was found to be localized in the peroxisomes (Fig. 3.7). It is however possible for a protein like NUDT15 to have dual targeting signals (one to the mitochondria and the other to the peroxisomes). Additionally, depending on which splice variant yields a mature protein, NUDT15 could either be localized in the peroxisomes or mitochondria. Alternatively, the

other three gene models (AT1G28960.1, AT1G28960.3 and AT1G28960.5) were used in the previously localization studies hence NUDT15 distribution in the mitochondria. Therefore, in future further localization analysis should be done on the other three gene models to ascertain which of them could possibly localize NUDT15 to the mitochondria as previously reported.

NUDT15 was transiently expressed in biolistically transformed onion epidermal layer cells. Visible peroxisome targeting took three to six days cold incubation at 4^{o}C, in addition to the standard expression conditions (Lingner et al., 2011). Cold incubation was done because plasmid and report protein degradation at refrigerator temperatures are slower. Therefore, this treatment could have given a low abundance protein such as NUDT15 ample time to be expressed above threshold required for peroxisomal targeting. Secondly, PTS1 domain of NUDT15 had predicted inhibitory residues such as glutamic acid, possibly requiring more time to target peroxisomes. All in all, from this study NUDT15 targeted small punctuate structures which were confirmed to be peroxisomes by using a peroxisomal marker DS-Red-SKL. Future research on NUDT15 would include, investigating its role in pathogen defence by carrying out gene expression analysis assays, and also carrying out assays which would shade light on its interacting partners. Protein to protein interaction assays such as Yeast two Hybrid screening and/ or co-immunoprecipitation.

4.1.3 The CHAT homolog is a peroxisomal protein

The CHAT homolog had only one gene model from the bioinformatics analysis done. The putative PTS1 domain of CHAT homolog had a functional non-canonical PTS1 (SSL>) (Lingner et al., 2011). The PTS1 domain had an overall charge of +3. Other characterized upstream enhancing residues such as proline were lacking in the CHAT homolog PTS1 domain. Nonetheless, inhibitory residues such as the glutamic and aspartic acid were equally absent in the domain.

Using PWM prediction model, the CHAT homolog was highly predicted to be a peroxisomal protein with more than 85% probability (Lingner et al., 2011). Based on annotation and PTS1 prediction high score and microarray data indicating a role in pathogen defence, CHAT homolog was selected and cloned into pCAT so that its subcellular localization could be validated experimentally using biolistically transformed onion epidermal layer cells.

The CHAT homolog was expressed within 18 hours of incubation at room temperature and its subcellular localization was subsequently analysed by fluorescence microscopy. CHAT homolog was targeted to punctuate subcellular structures which were afterwards confirmed to be peroxisomes by using a known peroxisomal protein DsRed-SKL (Fig. 3.7). CHAT homolog did not need any additional incubation at 4°C, in order for it to target the peroxisomes. This implied that the enhancing patterns in the PTS1 domain were very efficient since CHAT homolog carries a noncanonical PTS1 at its extreme C-terminus, which by itself could not have targeted the protein to the peroxisomes as efficiently as observed. In conclusion, CHAT homolog was validated through in vivo subcellular localization studies to be a peroxisomal protein in agreement with PWM model prediction. CHAT homolog functional analysis assays would shade light on its involvement in pathogen defence. Gene expression analysis on SA treated and/or pathogen infected plants should be among the future research studies on CHAT homolog.

4.1.4 ATP-BP is a cytosolic R protein

ATP-BP has a very long gene coding for proteins that could be as big as 100 kDa. It had two gene models AT1G72840.1 and AT1G72840.2 (Fig. 3.3). AT1G72840.2 had a predicted non-canonical PTS1 (SSL>), and its PTS1 domain had an overall charge of -2. This was due to that fact that glutamic acid residue was the most occurrent amino acid in the domain.

The two gene models were predicted to be nonperoxisomal by PWM prediction model. AT1G72840.1 scored very low, way below threshold getting a negative score. AT1G72840.2 was slightly below threshold but in grey zone where several true positive peroxisomal proteins with non-canonical PTS1s are located. Based on ATP-BP sequence homolog to R proteins and possession of a functional PTS1 (SSL>) gene model (AT1G72840.2) was cloned into pCAT vector via pGEM-T Easy vector. Due to that fact that ATP-BP is a big protein, a cloning strategy was employed. Both ATP-BP full length cDNA and the last exon cDNA (420aa C-terminal) were used in the subcellular analysis of ATP-BP.

pGEM-T Easy vector of 3 kbp was used in TA-cloning step. Both the full length and the last exon were successfully cloned into pGEM-T Easy and which was later on double digested with two restriction enzymes. ATP-BP full length cDNA of about 3 kbp could not be isolated from the vector backbone as they were both of the same size (Fig. 3.6). Nevertheless, since there was already a backup plan and that peroxisomal targeting is mediated by the PTS1 domain, ATP-BP subcellular localization analyses proceeded with the ATP-BP 420aa c-terminal. Alternatively, the full-length cDNA should be cloned directly into pCAT vector so that TA-cloning is by-passed.

The biolostically transformed onion epidermal layers cells with pCAT vectors containing ATP-BP 420aa C-terminal cDNA were analysed by fluorescence microscopy. ATP-BP C-terminal 420aa protein remained in the cytosol under both standard and alternated expression conditions. Non peroxisomal targeting by ATP-BP could have been as a result of a negatively charged PTS1 domain. Additionally, enhancing residues such as proline and the basic residues were lacking upstream of the functional plant PTS1 (SSL>). Therefore, ATP BP based on subcellular localization of its last exon, is putatively a cytosolic R protein.

ATP-BP full length separation from the pGEM-T Easy backbone in future studies could be achieved by; (1) using two restriction enzymes that would cut the backbone vector at sites further away from those that were used in this study. This would make the ATP BP full length insert larger than the vector backbone, hence separation by preparative agarose gel electrophoresis could be possible. Then, the ATP-BP insert could be double digested like done in this study. (2) Using a vector that is either greater or less than the size of ATP-BP full length cDNA.

4.2 Gene expression analysis

By measuring the amount of cellular RNA, it was possible to determine to what extent the NHL and IAN genes of interest had been inducible due to SA treatment. qPCR was used in quantifying gene expression levels. Quantitative gene expression using qPCR almost always relied on PCR replicates (Schmittgen and Livak, 2008).

4.2.1 Plant growth and SA treatment

Several conditions could either up-regulate or down-regulate gene expression levels our genes of interest. The conditions included both biotic and abiotic ones. To ensure gene expression of the NHLs and IANs under study was SA-induced, plants were grown in stress free growth chambers. However, had plants been stressed in any way prior to the intended gene induction treatment with exogenously applied SA, gene quantitation as a result of the treatment would have been futile.

SA treatment was among the conditions that caused gene expression of NHL and IAN proteins according to publicly available microarray data. However, these are large-scale data that need to be validated thoroughly before considered reliable.

Healthy seven weeks old Arabidopsis thaliana of ecotype Columbia 0 were used as SA treatment was dependent on plant ecotype, age, anatomy and developmental stage. Additionally, how SA treatment was carried out was also crucial to accurate quantification of gene expression levels.100 µM of SA was exogenously sprayed to plants. Spraying of SA was preferred to other ways of application because no wounding or stress was imposed on plants. Leaves were used for RNA extraction because our genes were relatively expressed in such parts of the plant.

4.2.2 RNA extraction

A major critical step in performing qPCR is the isolation of high quality, intact RNA. Proper treatment and handling of samples prior to RNA isolation and storage of the isolated RNA are crucial components of obtaining high quality RNA. The quality of isolated RNA is crucial for the success of downstream applications such as cDNA synthesis and qPCR.

Quick and thorough sample disruption is another essential component of RNA isolation and affects RNA yield and quality. Samples had to be disrupted quickly as slow disruption would have resulted in RNA degradation by endogenous RNases. Additionally, thorough disruption of samples is recommended because incomplete disruption results in decreased yield, as a result of some RNA remaining trapped in intact cells and unavailable for purification.

RNA yield (concentration) was assessed by Spectrophotometry which measures the absorbance of RNA at a wavelength of 260 nm. This is a very easy and fast assay but there were some disadvantages such as readings may be affected by other molecules that absorb at 260 nm. Accurate RNA concentration determination was very critical as cDNA synthesis was dependant on correctly determined RNA concentrations.

RNA integrity was assessed by agarose gel electrophoresis to ensure that samples being compared were of similar integrity. For total RNA samples, discrete, though thick 28S:18S ribosomal RNA (rRNA) gel bands and in an approximate mass ratio of 2:1 are indications of high integrity. RNA samples may get contaminated with proteins, phenols and other molecules that later on interfere with downstream applications. Pure RNA has an A_{260}:A_{280} ratio of ~2.0. $A_{260/280}$ values < 2.0 may indicate high protein levels in an RNA preparation. An $A_{260/280}$ value < 2.0 does not predict that protein levels are high enough to cause PCR inhibition, but, the more the $A_{260/280}$ deviates from 2.0, the greater the chance of obtaining spurious results. Moreover, genomic DNA is often co-extracted with RNA (especially if lightly fragmented) and can serve as a template in downstream processes such as qPCR. However, the primers were designed such that they span an exon-exon junction. By this strategy genomic DNA was then excluded from serving as a template that would have been detected as second PCR products in the real-time PCR reactions. In addition, genomic DNA contamination could lead to inaccurate RNA quantitation.

In summary, in order to obtain high quality RNA, certain steps were followed carefully; (1) Samples were carefully stored at -80^{o}C prior to RNA isolation, (2) samples were rapidly and completely disrupted mechanically to prevent RNA degradation and to increase RNA yield,

(3) A proper RNA isolation kit (Qiagen kit) was used, (4) RNA was stored in RNase-free solutions and lastly (5) RNA was accurately assessed to confirm that indeed it was of high integrity and purity.

4.2.3 cDNA synthesis and selection of endogenous control gene

Reverse transcription is the process by which RNA was used as a template to synthesize cDNA. This cDNA then served as the template in the real-time reaction. The two-step RT method was used to perform the reverse transcription that cDNA could be used later for qPCR. Random primers were used in order to reverse transcribe all mRNAs (that is targets and endogenous control). 50ng/µl of cDNA was synthesized each time as 100ng/µl of input RNA was used. The High-Capacity cDNA Reverse Transcription kits were used in cDNA synthesis.

A major step in experimental design of relative quantitation is the selection of an appropriate endogenous control. Normalization to an endogenous control (often referred to as a housekeeping gene) allows a correction of results that can be skewed by differing amounts of input nucleic acid template (Livak and Schmittgen, 2001). Any gene shown to be expressed at the same level in all study samples can potentially be used as an endogenous control. Selection of either ACT2 or ubiquitin as an endogenous control was based on the uniformity of endogenous control expression levels and specific amplification (see Fig. 3.11 and .12). Ubiquitin was not selected because there was some nonspecific amplification of cDNA by the primer pair used. ACT2 was instead selected as the endogenous control gene in all qPCR assays in this present study.

The endogenous control was used to normalize differences in the amount of cDNA that was loaded into PCR reaction wells, therefore, endogenous control expression levels were supposed to be the same in all samples. So, it was critical to determine if the SA treatment affected the expression level of Act2 (endogenous control gene). Act2 was constantly expressed in all types of samples whether treated or untreated. Act2 had an average Ct value of 21.12±0.3 in SA treated and untreated samples (see Fig. 3.12). Therefore, ACT2 was a good endogenous control gene for SA treatment gene expression analysis like previously reported (Livak and Schmittgen, 2001).

4.2.4 Validation of gene expression analysis

The two most commonly used methods to analyze data from real-time, quantitative PCR experiments are absolute quantification and relative quantification. Absolute quantification determines the input copy number of the transcript of interest, usually by relating the PCR signal to a standard curve. Relative gene expression presents the data of the gene of interest relative to some calibrator or internal control gene (Livak and Schmittgen, 2001; Schmittgen and Livak, 2008).

In this present study, it was unnecessary to determine the absolute transcript numbers, therefore relative change in gene expression was carried out. Relative quantification was easier to perform than the absolute method because the use of standard curves was not required. Comparative C_T method was selected as the method of choice presenting relative gene expression. Advantages of the comparative C_T method include ease of use and the ability to present data as 'fold change' in expression. Disadvantages of the comparative C_T method include the PCR amplification efficiency of endogenous control and target genes must be approximately equal or the PCR must be further optimized (Livak and Schmittgen, 2001).

Prior to using the comparative C_T method (also known as the $\Delta\Delta C_T$ method), it was important to ensure that the target(s) and endogenous control had similar or relatively equivalent PCR efficiencies. PCR amplification efficiency is the rate at which a PCR amplicon is generated, and expressed as a percentage of the maximum possible value. If a particular PCR amplicon doubles in quantity during the geometric phase of its PCR amplification then the PCR assay has 100% efficiency.

The slope of a standard curve was used to estimate the PCR amplification efficiency of a real-time PCR reaction. A real-time PCR standard curve was graphically represented in two ways, as a semi-log regression line plot of either C_T or ΔC_T value vs. log of input nucleic acid (see Fig. 3.13). A standard curve slope of –3.32 or < 0.1 indicates a PCR reaction with 100% efficiency (Livak and Schmittgen, 2001).

For the comparative C_T method of relative quantitation to be valid, the efficiency of the target amplification and the efficiency of the active reference (endogenous control) amplification

had to be approximately equal. Assessing the relative efficiencies of the target amplification and the endogenous control amplification was achieved by running standard curves for each amplicon utilizing the same sample. The sample in the validation experiment had to express both the target and reference genes. The C_T values generated from equivalent standard curve mass points (target vs. endogenous control) were used in the ΔC_T calculation

($C_{T\ target} - C_{T\ endogenous}$).

ΔC_T values were plotted against log input amount to create a semi-log regression line. The slope of the resulting semi-log regression line was then be used as a general criterion for passing a validation experiment (Livak and Schmittgen, 2001). In a validation experiment that passes, the absolute value of the slope of ΔC_T vs. log input is < 0.1 (approximately zero). Ideally, the absolute values of the slope of the validation experiment were supposed to be <0.1. Nonetheless, not all experimental data gave slopes that were either -3.32 or <0.1, for example slopes in figure 3.13A and C, respectively. In those cases data was re-assessed in an effort to improve the slope. When examining the validation experiment results, it was important to assure that rigorous analysis was done on the data. The outer points of the standard curves (high input and low input) were used in re-assessment of the data; as if inhibitors affect the higher concentration points. Additionally, low levels of target were present at the lower concentration points hence needed to be removed as the lower limits of detection might have been reached. It was also important to look at the precision among the replicates of each dilution set. Any outliers were removed and data re-analysed.

In order for the two genes (target and endogenous control) to be used in gene expression relative quantitation, their respective amplification efficiencies had be 100±10% and approximately equal. Therefore, NHL6, NHL25, PR2 and PR5 had PCR amplification efficiencies approximately equal to that of the endogenous control, ACT2. By contrast, NHL4, IAN8, IAN11 and IAN12 amplification efficiencies were not approximately equal to ACT2. Therefore, the relative quantification could not be accurately done using the comparative Ct method, unless with the relative standard curve method where there is no need to normalize target genes to an endogenous control gene. Nevertheless, all the genes under

expression analysis were studied with the comparative Ct method while bearing in mind that NHL4, IAN8, IAN11 and IAN12 amplification efficiencies were not approximately equal to to the endogenous control gene, ACT2. Relative quantitation of these genes with comparative Ct method was done in order to known whether they could be induced by SA treatment despite not being accurately determined due to their PCR amplification efficiency differences with the endogenous control.

4.2.5 Relative quantification of NHL and IAN genes by the comparative C_T method

Gene relative quantification by the comparative C_T method (ΔΔCt method) was advantageous as no standard curves had to be run on each plate. This could have resulted in reduced usage of very expensive qPCR reagents. However, since the primers were custom-made, an initial validation relative standard curve was necessary in order to validate the PCR efficiencies of the target and endogenous control, and also due to the hypothesis that our genes of interest were of low expression level under standard plant growth conditions like most pathogen defence genes.

The amount of target (fold increase), normalized to an endogenous control and relative to a calibrator (untreated), is given by the formula:

$$2^{-\Delta\Delta C_T}$$

When the ΔΔCt method is used, data is presented as the fold change in gene expression normalized to an endogenous reference gene and relative to the untreated control. For the untreated control sample, ΔΔ*C*T equals zero and 2^0 equals one, so that the fold change in gene expression relative to the untreated control equals one, by definition (Livak and Schmittgen, 2001).

The peroxisomal-targeted defence proteins under study were expressed in all the three samples except for IAN8 that was below qPCR detectable amount in untreated sample. Expression of the NHL and IAN proteins in untreated sample could have been due to some unforeseen stress plants could have been subjected to. However, plants were grown and subjected to less or no stress at all as it was clear from the bioinformatics analysis done on the proteins that they were stress inducible. The genes were highly expressed at 24 hour post-treatment than at 72 hours. This could have been due to the fact that SA treatment was

exogenous and only applied once, the defence hormone concentration could have go too low to keep on inducing the genes by 72 hours post-treatment. In addition, innate immunity of plants has negative regulators such as NUDT7, therefore the expression of the NHL and IAN genes had been negatively regulated. Negative regulation of pathogen defence proteins when there is no more need for the plant to mount defensive mechanisms is important as fatal defensive actions such as hypersensitive response programmed cell death are stopped. NHL4, NHL25 and IAN11 expression levels below their respective expression levels in the untreated sample at zero hour. That could imply that these proteins were more negatively regulated than NHl6 and IAN12. However, NHL6 and IAN12 were also reduced at 72 hours post-treatment. IAN8 was below qPCR detectable amounts in untreated sample therefore, the sequence detecting software (SDS) of qPCR machine could not calculate the fold changes in treated samples. Nonetheless, IAN8 minimum fold change was calculated manually and it had 509 and 26 fold change in 24 hour and 72 hour post treatment samples respectively. Since the least possible Ct value was 40, the undetected Ct value was then assumed to be 40 so that at least the fold change for IAN8 could manually be known.

4.3 Signal transduction pathway analysis for two NHL and three IAN proteins

Co-expression analyses were done on two proteins (NHL4 and NHL6) from the NHL family and also on three proteins (IAN8, IAN11 and IAN12) from the IAN family, so that the signal transduction cascade for these proteins could be known. Co-expression analyses of our five putative defence protein of interest were done concomitantly with the other Arabidopsis established defence proteins. In total, 28 Arabidopsis defence proteins were used in co-expression studies, in an effort to allocate our proteins of interest to known defence pathways. The known and characterized defence pathways in plants are shown in the figure below. Therefore, one of the goals of the current MSc study was to place the proteins of interest to these pathways.

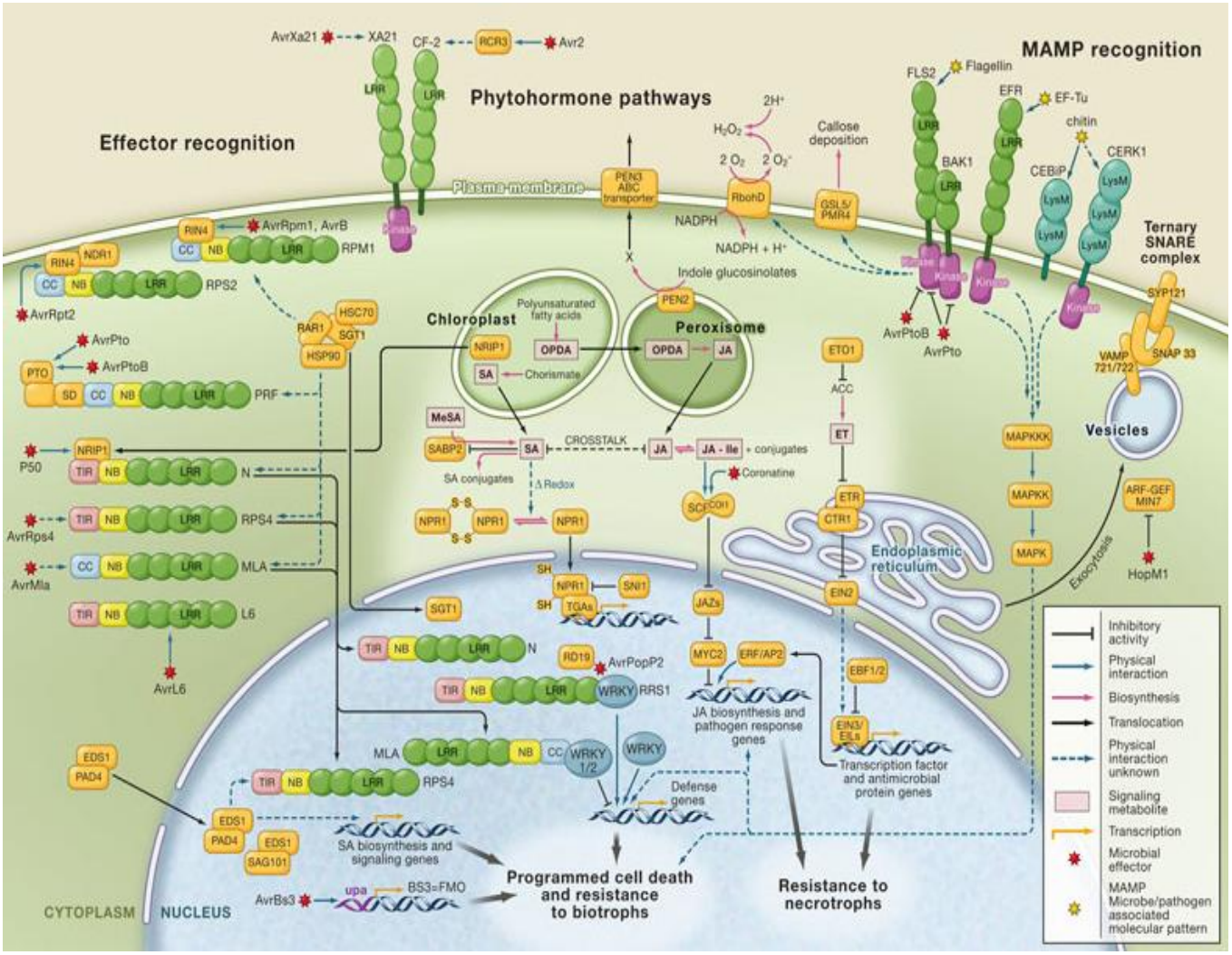


Figure 4.1: Plant immune response pathways (Panstruga et al., 2009). Plant innate immunity comprises of different pathways namely, MAMP recognition pathway, effector recognition and phytohormone pathways.

Multiple and often parallel defence signaling pathways are activated in plants following pathogen infection. Plant innate immunity comprises of different pathways namely, MAMP recognition pathway, effector recognition and lastly phytohormone biosynthesis. The activation of these pathways, still remains to be fully understood as more and more pathogen related proteins and molecules are being discovered.

R-proteins play a pivotal role in the activation of resistance after plants perceive pathogens or effectors (Jones and Dangl, 2006). Nevertheless, R-proteins alone are not sufficient for the initiation of resistance against pathogens. Therefore, numerous auxillary proteins and chaperones are found serving as co-activators of resistance. For example, EDS1 has been

shown to mediate defence signaling through the activation of the TIR domain-containing R-proteins, while NDR1 has been shown to be required for R-proteins that contain the CC domain (Wiermer et al., 2005; Panstruga et al., 2009; Knepper et al., 2011).

EDS1 serves as a central regulatory protein involved in both biotic and oxidative stress signaling. In this capacity, EDS1, through its interaction with PAD4 and SAG101, is required for elicitation of the HR during bacterial infection (Panstruga et al., 2009). RIN4 was co-expressed with RPS2, EDS1, PAD4 and EDM1 (refer to section 3.3.3). RIN4 regulates and activates resistance mediated by the two R proteins namely; RPM1 and RPS2, both belonging to the class CC-NB-LRR of R proteins (see Fig. 4.1). EDS1 together with its interacting partner PAD4 were found to be co-expressed with IAN8 and SID2. SID2 belongs to SA biosynthesis pathway (Abreu and Bosch, 2008). IAN and SID2 were also found to be co-expressed as a pair. IAN8 was also co-expressed as a pair with PAD4. Taken together, IAN8 could be expressed after effector recognition R proteins and signalling by SA. Interestingly, IAN8 was co-expressed with NHL6 another protein of interest in this study. NHL6 is a homolog of NDR1, and NDR1 activates R proteins by its association with RIN4. Apparently, neither IAN8 nor NHL6 were co-expressed with NDR1. NHL6 like IAN8 was co-expressed with SID2.

NHL6 was second highest co-expressed of the 28 Arabidopsis defence proteins used in the co-expression analysis. The highest co-expressed defence protein was an R protein RPM1. RPM1 is negatively regulated by RIN4. RIN4 as NDR1 has its interacting partner (Knepper et al., 2011). NDR1 is an Arabidopsis homolog of NHL6. Taking all this together, NHL6 putatively somewhat positioned in along the effector recognition pathway.

NHL4, IAN11 and IAN12 were not co-expressed with any protein(s). This could have been due to the model limitation of the bioinformatics tools used, as only certain signal transduction settings were possible. Therefore, zero co-expression data of NHL4, IAN11 and IAN12 shows that they do not belong to the major pathogen signal transduction pathways like NHL6 and IAN8. They could however, belong to atypical pathways, which are activated by abiotic stress such as dehydration, salinity, cold, heat and ROS. Future studies on co-expression analysis of NHL4, IAN11 and IAN12 should be done by changing the stimuli from pathogen to any of the aforementioned abiotic stresses. This would help in elucidating the signal transduction pathway they belong to.

5 Conclusion

In this present study, peroxisome-targeted pathogen defence proteins have been identified and validated through subcellular localization studies. By contrast, two pathogen defence proteins possibly carrying PTS1 also could not yet be confirmed peroxisomal proteins and might be cytosolic. Previously predicted PTS1 carrying proteins, NUDT15 and CHAT homolog were validated to peroxisomal proteins through in vivo subcellular targeting analysis. NUDT7 and ATP-BP were the two proteins that were also previously predicted to possibly be carrying PTS1s, nonetheless, their subcellular localization still requires validation by alternative methods.

Three proteins (NHL4, NHL6 and NHL25) from NHL family and three proteins (IAN8, IAN11 and IAN12) from IAN family had their gene expression levels increased after SA treatment. Therefore, NHL4, NHL6, NHL25, IAN8, IAN11 and IAN12 are SA inducible pathogen defence proteins. Additionally, co-expression analysis showed that NHL6 and IAN8 belong to major pathogen signal transduction pathways in plant innate immunity, whereas NHL4, IAN11 and IAN12 do not belong to the typical pathogen pathway.

6 References

Abreu, M.E., and Bosch, S.M. (2008). **Salicylic acid deficiency in NahG transgenic lines and sid2 mutants increases seed yield in the annual plant Arabidopsis thaliana**

Attaran, E., Zeier, T.E., Griebel, T., and Zeier, J. (2009). Methyl salicylate production and jasmonate signaling are not essential for systemic acquired resistance in Arabidopsis. Plant Cell **21,** 954-971.

Bartsch, M., Gobbato, E., Bednarek, P., Debey, S., Schultze, J.L., Bautor, J., and Parker, J.E. (2006). Salicylic acid-independent ENHANCED DISEASE SUSCEPTIBILITY1 signaling in Arabidopsis immunity and cell death is regulated by the monooxygenase FMO1 and the Nudix hydrolase NUDT7. Plant Cell **18,** 1038-1051.

Bonsegna, S., Slocombe, S.P., De Bellis, L., and Baker, A. (2005). AtLACS7 interacts with the TPR domains of the PTS1 receptor PEX5. Arch Biochem Biophys **443,** 74-81.

Brown, L.A., and Baker, A. (2008). Shuttles and cycles: transport of proteins into the peroxisome matrix (review). Mol Membr Biol **25,** 363-375.

Bruinsma, M., Posthumus, M.A., Mumm, R., Mueller, M.J., van Loon, J.J., and Dicke, M. (2009). Jasmonic acid-induced volatiles of Brassica oleracea attract parasitoids: effects of time and dose, and comparison with induction by herbivores. J Exp Bot **60,** 2575-2587.

Dammai, V., and Subramani, S. (2001). The human peroxisomal targeting signal receptor, Pex5p, is translocated into the peroxisomal matrix and recycled to the cytosol. Cell **105,** 187-196.

Gabaldon, T., Snel, B., van Zimmeren, F., Hemrika, W., Tabak, H., and Huynen, M.A. (2006). Origin and evolution of the peroxisomal proteome. Biol Direct **1,** 8.

Gasmi, L., and McLennan, A.G. (2001). The mouse Nudt7 gene encodes a peroxisomal nudix hydrolase specific for coenzyme A and its derivatives. Biochem J **357,** 33-38.

Ge, X., Li, G.J., Wang, S.B., Zhu, H., Zhu, T., Wang, X., and Xia, Y. (2007). AtNUDT7, a negative regulator of basal immunity in Arabidopsis, modulates two distinct defense response pathways and is involved in maintaining redox homeostasis. Plant Physiol **145,** 204-215.

Gould, S.G., Keller, G.A., and Subramani, S. (1987). Identification of a peroxisomal targeting signal at the carboxy terminus of firefly luciferase. J Cell Biol **105,** 2923-2931.

Gould, S.J., Keller, G.A., Hosken, N., Wilkinson, J., and Subramani, S. (1989). A conserved tripeptide sorts proteins to peroxisomes. J Cell Biol **108,** 1657-1664.

Hayashi, M., Yagi, M., Nito, K., Kamada, T., and Nishimura, M. (2005). Differential contribution of two peroxisomal protein receptors to the maintenance of peroxisomal functions in Arabidopsis. J Biol Chem **280,** 14829-14835.

Hofius, D., Schultz-Larsen, T., Joensen, J., Tsitsigiannis, D.I., Petersen, N.H., Mattsson, O., Jorgensen, L.B., Jones, J.D., Mundy, J., and Petersen, M. (2009). Autophagic components contribute to hypersensitive cell death in Arabidopsis. Cell **137,** 773-783.

Jones, J.D., and Dangl, J.L. (2006). The plant immune system. Nature **444,** 323-329.

Kagawa, T., and Beevers, H. (1975). The development of microbodies (glyoxysomes and leaf peroxisomes) in cotyledons of germinating watermelon seedlings. Plant Physiol **55,** 258-264.

Kamigaki, A., Mano, S., Terauchi, K., Nishi, Y., Tachibe-Kinoshita, Y., Nito, K., Kondo, M., Hayashi, M., Nishimura, M., and Esaka, M. (2003). Identification of peroxisomal targeting signal of pumpkin catalase and the binding analysis with PTS1 receptor. Plant J **33,** 161-175.

Knepper, C., Savory, E.A., and Day, B. (2011). Arabidopsis NDR1 Is an Integrin-Like Protein with a Role in Fluid Loss and Plasma Membrane-Cell Wall Adhesion. Plant Physiol **156,** 286-300.

Lingner, T., Kataya, A.R., Antonicelli, G.E., Benichou, A., Nilssen, K., Chen, X.Y., Siemsen, T., Morgenstern, B., Meinicke, P., and Reumann, S. (2011). Identification of Novel Plant Peroxisomal Targeting Signals by a Combination of Machine Learning Methods and in Vivo Subcellular Targeting Analyses. Plant Cell.

Liu, Y., Schiff, M., Czymmek, K., Talloczy, Z., Levine, B., and Dinesh-Kumar, S.P. (2005). Autophagy regulates programmed cell death during the plant innate immune response. Cell **121,** 567-577.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods **25,** 402-408.

Lopez-Huertas, E., Charlton, W.L., Johnson, B., Graham, I.A., and Baker, A. (2000). Stress induces peroxisome biogenesis genes. Embo J **19,** 6770-6777.

Ma, C., and Reumann, S. (2008). Improved prediction of peroxisomal PTS1 proteins from genome sequences based on experimental subcellular targeting analyses as exemplified for protein kinases from Arabidopsis. J Exp Bot **59,** 3767-3779.

Ma, C., Haslbeck, M., Babujee, L., Jahn, O., and Reumann, S. (2006). Identification and characterization of a stress-inducible and a constitutive small heat-shock protein targeted to the matrix of plant peroxisomes. Plant Physiol **141,** 47-60.

Mano, S., Hayashi, M., Kondo, M., and Nishimura, M. (1996). cDNA cloning and expression of a gene for isocitrate lyase in pumpkin cotyledons. Plant Cell Physiol **37,** 941-948.

Matre, P., Meyer, C., and Lillo, C. (2009). Diversity in subcellular targeting of the PP2A B'eta subfamily members. Planta **230,** 935-945.

Nishimura, M., Yamaguchi, J., Mori, H., Akazawa, T., and Yokota, S. (1986). Immunocytochemical Analysis Shows that Glyoxysomes Are Directly Transformed to Leaf Peroxisomes during Greening of Pumpkin Cotyledons. Plant Physiol **81,** 313-316.

Nishimura, M., Hayashi, M., Kato, A., Yamaguchi, K., and Mano, S. (1996). Functional transformation of microbodies in higher plant cells. Cell Struct Funct **21,** 387-393.

Nyathi, Y., and Baker, A. (2006). Plant peroxisomes as a source of signalling molecules. Biochim Biophys Acta **1763,** 1478-1495.

Ogawa, T., Yoshimura, K., Miyake, H., Ishikawa, K., Ito, D., Tanabe, N., and Shigeoka, S. (2008). Molecular characterization of organelle-type Nudix hydrolases in Arabidopsis. Plant Physiol **148,** 1412-1424.

Panstruga, R., Parker, J.E., and Schulze-Lefert, P. (2009). SnapShot: Plant immune response pathways. Cell **136,** 978 e971-973.

Poirier, Y., Antonenkov, V.D., Glumoff, T., and Hiltunen, J.K. (2006). Peroxisomal beta-oxidation--a metabolic pathway with multiple functions. Biochim Biophys Acta **1763,** 1413-1426.

Queval, G., Issakidis-Bourguet, E., Hoeberichts, F.A., Vandorpe, M., Gakiere, B., Vanacker, H., Miginiac-Maslow, M., Van Breusegem, F., and Noctor, G. (2007). Conditional oxidative stress responses in the Arabidopsis photorespiratory mutant cat2 demonstrate that redox state is a key modulator of daylength-dependent gene expression, and define photoperiod as a crucial factor in the regulation of H(2)O(2)-induced cell death. Plant J **52,** 640-657.

Rask, L., Andreasson, E., Ekbom, B., Eriksson, S., Pontoppidan, B., and Meijer, J. (2000). Myrosinase: gene family evolution and herbivore defense in Brassicaceae. Plant Mol Biol **42,** 93-113.

Reumann, S. (2004). Specification of the peroxisome targeting signals type 1 and type 2 of plant peroxisomes by bioinformatics analyses. Plant Physiol **135,** 783-800.

Reumann, S. (2011). Toward a definition of the complete proteome of plant peroxisomes: Where experimental proteomics must be complemented by bioinformatics. Proteomics **11,** 1764-1779.

Reumann, S., Babujee, L., Ma, C., Wienkoop, S., Siemsen, T., Antonicelli, G.E., Rasche, N., Luder, F., Weckwerth, W., and Jahn, O. (2007). Proteome analysis of Arabidopsis leaf peroxisomes reveals novel targeting peptides, metabolic pathways, and defense mechanisms. Plant Cell **19,** 3170-3193.

Reumann, S., Quan, S., Aung, K., Yang, P., Manandhar-Shrestha, K., Holbrook, D., Linka, N., Switzenberg, R., Wilkerson, C.G., Weber, A.P., Olsen, L.J., and Hu, J. (2009). In-depth proteome analysis of Arabidopsis leaf peroxisomes combined with in vivo subcellular targeting verification indicates novel metabolic and regulatory functions of peroxisomes. Plant Physiol **150,** 125-143.

Sakai, Y., Oku, M., van der Klei, I.J., and Kiel, J.A. (2006). Pexophagy: autophagic degradation of peroxisomes. Biochim Biophys Acta **1763,** 1767-1775.

Scherz-Shouval, R., Shvets, E., Fass, E., Shorer, H., Gil, L., and Elazar, Z. (2007). Reactive oxygen species are essential for autophagy and specifically regulate the activity of Atg4. Embo J **26,** 1749-1760.

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

Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative C(T) method. Nat Protoc **3,** 1101-1108.

Wang, Z., and Li, X. (2009). IAN/GIMAPs are conserved and novel regulators in vertebrates and angiosperm plants. Plant Signal Behav **4,** 165-167.

Wiermer, M., Feys, B.J., and Parker, J.E. (2005). Plant immunity: the EDS1 regulatory node. Curr Opin Plant Biol **8,** 383-389.

Yasuda, M., Ishikawa, A., Jikumaru, Y., Seki, M., Umezawa, T., Asami, T., Maruyama-Nakashita, A., Kudo, T., Shinozaki, K., Yoshida, S., and Nakashita, H. (2008). Antagonistic interaction between systemic acquired resistance and the abscisic acid-mediated abiotic stress response in Arabidopsis. Plant Cell **20,** 1678-1692.

7 Appendix

7.1 cDNA sequence analysis in pGEM-T Easy and pCAT vectors

7.1.1 NUDT7

7.1.1.1 NUDT7 (AT4G12720, ASL>) sequence in pGEM-T Easy

Subcloning method: The full-length cDNA amplified was PCR amplified from the ABRC clone NUDT7 and subcloned into pGEMT-Easy. The insert was sequenced from forward end using T7 primer (see section 2.1.10 for more sequencing information).

Nucleotide sequence from Seqlab:

CTGCTCCGGCCGCCATGGCGGCCGCGGGAATTCGATTCAAGTCTAGAGTCAGAGAGAAGCAGAGGCTTGGTCACG
CGATACTTTAAGGCGCTTGGCATGATCCGCATTGCAGTAGATAAAGCTCTCTTTACCAGATGATGTGGTAGTTGG
CACAATGGCGAATCCCAAGTATTCTTCCTCACACTTCTTTTGGCAAATGTTAGCCATGAACTTGAACATCTCGTT
CTTCTTGTTCCATGGTTGGTCTACATACTCTTGGATCGGCATCCACTTAGCTTGCAAGATCTCAGATTTTTGTTC
AGTAATATCGTAAGAGCGCGGACTTAAGACACACAGGAAAAACATATCTGTTTTCTTTTTTAAGATGGCTTTGTG
GCTTTGCCTGAAAGCCAGTACTTCGACAAAATCTGCAATAATTCCAGTTTCTTCTTCCACTTCCCTAGCTACTCC
AGTCCATATATCCTCGCCCTCGTTGATAACACCAGTAGGCAGCTTCCACACATTTTTATCTTTGAAAAACCCACT
CCTCTCCTGGACAACGAGGACCTCTTTAGTATTTTTGTTGATGACCAAAGCACCAGCACCTACAACATGAGAAGC
ATTGGCTGGGATTGTATCAGGAGTTTCAGAGATCCAAGATACAAGCATCAAGTACTCAGGCTCCGCGTGGTGATA
TCTAAATCCTTCACTAACTGCAGCCTCCACAAGATTAGCCAATCCAAGAGGCAGCTTTATCCAAATTCCCTTCTT
CCCCTCTTCTCTCCAATGCGAAAGAGAAGCCCTAAGACTTTCAGTAAAAACCTCAGAATCCATAGGTTCCACCAT
GGTTACAGTAACACCATCGTAATTATCAGTCTCACCTTCAAGTAAAGGAATCTGCTGAGCTCTAGTACCCATAGC
GGCCGCAGTCTTAATCACTAGTGAATTCGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTCCCAACGCGTTGG
ATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAA
ATTGTTATCCGCTCACATTCCNCACAACANCGAGCCGAGCATAAGTGTAAAGCTGGGNTGCTATGANNGAGCTAA
CTCACATTAATTGCNNNTGCGCTNNNNNNCCGNTNCANNCGGANNNNNNNGNCAGCTGCATATGATCGGNNATCG
CTNGGNAAGAAN

This Seqlab sequence is 100% identical at the amino acid level to the cDNA of NUDT7 in the database and the NUDT7 was manually translated at ExPASy for verification.

7.1.1.2 Sequence analysis of NUDT7 in pCAT

Subcloning method: The full-length cDNA amplified was PCR amplified from the ABRC clone NUDT7 and subcloned into pCAT via pGEMT-Easy. The insert was sequenced from forward end using SR321F primer (see section 2.1.10 for sequencing procedure)

Nucleotide sequence from seqlab

CCACGAGAAGCGCGATCCATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTG
TACAAGGCGGCCGCTATGGGTACTAGAGCTCAGCAGATTCCTTTACTTGAAGGTGAGACTGATAATTACGATGGT
GTTACTGTAACCATGGTGGAACCTATGGATTCTGAGGTTTTTACTGAAAGTCTTAGGGCTTCTCTTTCGCATTGG
AGAGAAGAGGGGAAGAAGGGAATTTGGATAAAGCTGCCTCTTGGATTGGCTAATCTTGTGGAGGCTGCAGTTAGT
GAAGGATTTAGATATCACCACGCGGAGCCTGAGTACTTGATGCTTGTATCTTGGATCTCTGAAACTCCTGATACA
ATCCCAGCCAATGCTTCTCATGTTGTAGGTGCTGGTGCTTTGGTCATCAACAAAAATACTAAAGAGGTCCTCGTT
GTCCAGGAGAGGAGTGGGTTTTTCAAAGATAAAAATGTGTGGAAGCTGCCTACTGGTGTTATCAACGAGGGCGAG
GATATATGGACTGGAGTAGCTAGGGAAGTGGAAGAAGAAACTGGAATTATTGCAGATTTTGTCGAAGTACTGGCT
TTCAGGCAAAGCCACAAAGCCATCTTAAAAAAGAAAACAGATATGTTTTTCCTGTGTGTCTTAAGTCCGCGCTCT
TACGATATTACTGAACAAAAATCTGAGATCTTGCAAGCTAAGTGGATGCCGATCCAAGAGTATGTAGACCAACCA
TGGAACAAGAAGAACGAGATGTTCAAGTTCATGGCTAACATTTGCCAAAAGAAGTGTGAGGAAGAATACTTGGGA
TTCGCCATTGTGCCAACTACCACATCATCTGGTAAAGAGAGCTTTATCTACTGCAATGCGGATCATGCCAAGCGC
CTTAAAGTATCGCGTGACCAAGCCTCTGCTTCTCTCTGACTCTAGAGTCCGCAAAAATCACCAGTCTCTCTCTAC
AAATCTATCTCTCTCTATTTTTCTCCAGAATAATGTGTGAGTAGTTCCCAGATAAGGGAATTAGGGTTCTTATGG
NTTCGCTCATGTGTGAGCATATAANAAACCTTAGTATGTNTTGTATTTGTAAAATACTNNATCATAAATTTCTAT
CNAACAANTCANNACTGCAGGCATGCAGCTNNNNCNTCNNTNNACGTCTGACTGGAACCTGNCGTTACCAACTAA
TN

Translation of Seqlab sequence 5` end 99 base pairs at ExPASY gave the amino sequence shown below and the highlighted part is the N-terminal ligation region of the NUDT7 with EYFP via NotI:

5'3' Frame 1

P R E A R S **Met** V L L E F V T A A G I T L G **Met** D E L Y K A A A **Met** G T

>YFP_withNotI

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLV
TTFGYGLQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRI
ELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQN
TPIGDGPVLLPDNHYLSYQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKAAAM

The C-ternimus of YFP_NotI together with the NUDT7 N-terminus were analysed. The NUDT7 was in a proper orientation, therefore, the anticipated fusion protein would be expressed.

7.1.2 NUDT15 (AT1G28960.2/4)

The PTS1 prediction algorithms indicated that 2 splice variants of AtNUDT15 (.2 and .4), which terminate with PKM>, were peroxisomal. However, nucleotide and protein sequences of the five gene models of NUDT15 are presented here together with their respective sequence alignment figures:

Nucleotide sequences of the five NUDT15 gene models

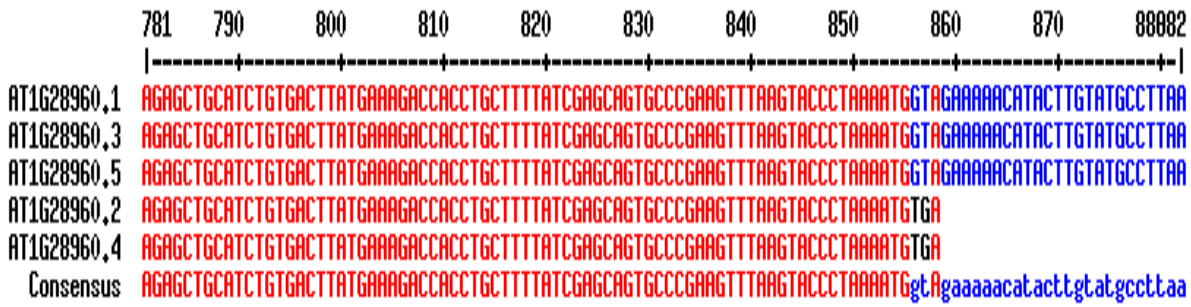


Figure 7.1: NUDT15 gene models sequence alignment at nucleotide level (position 781-882)

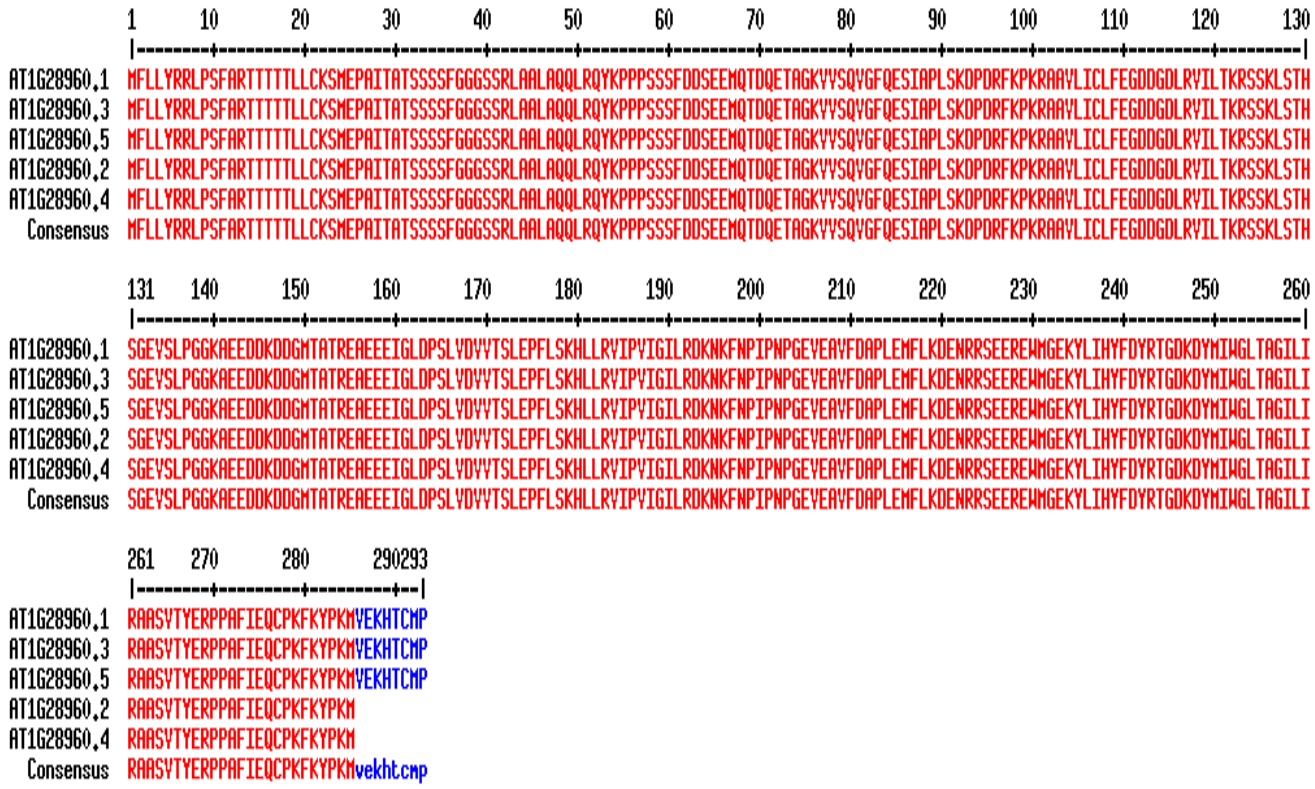


Figure 7.2: NUDT15 gene models sequence alignment at amino acid level.

7.1.2.1 NUDT15 (AT1G28960.2) sequence in pGEM-T Easy

The full-length cDNA amplified was PCR amplified from the RIKEN clone and subcloned into pGEMT-Easy. The insert was sequenced from forward end using T7 primer

Nucleotide sequence from seqlab:

TGCTCCGGCCGCCATGGCGGCCGCGGGAATTCGATTAAGACTGCGGCCGCTATGTTTTTGCTTTATCGTAGGCTT
CCTTCATTTGCACGAACAACAACAACAACCCTTCTCTGCAAATCTATGGAGCCTGCGATAACAGCGACTTCTTCT
TCTTCTTTCGGTGGTGGCTCTTCTCGTCTCGCCGCTTTAGCCCAGCAACTTCGCCAATACAAGCCCCCACCTTCT
TCATCGTTCGATGACTCCGAGGAGATGCAGACAGATCAGGAGACCGCTGGGAAAGTCGTTTCTCAGGTTGGGTTT
CAGGAATCTATTGCTCCTCTTTCGAAAGACCCTGATAGGTTTAAACCCAAGAGAGCCGCTGTGTTGATCTGTCTC
TTTGAAGGAGATGATGGTGATTTGCGTGTTATTCTTACTAAGAGATCTTCCAAATTGTCTACTCACTCTGGAGAA
GTTTCATTGCCAGGTGGTAAAGCGGAGGAGGATGATAAAGATGATGGGATGACTGCTACCAGAGAGGCTGAGGAA
GAGATTGGATTGGACCCTTCTCTTGTTGATGTTGTTACTTCTCTTGAACCATTTCTGTCTAAGCATCTTCTTAGA
GTAATTCCTGTGATAGGCATCTTGAGGGACAAAAATAAATTCAATCCGATACCAAATCCTGGGGAAGTGGAAGCT
GTGTTTGATGCACCCTTGGAAATGTTCCTTAAGGATGAGAATCGAAGATCTGAAGAGAGAGAGTGGATGGGTGAA
AAGTATTTGATCCATTACTTTGACTACAGAACAGGAGATAAGGATTATATGATATGGGGTTTAACTGCTGGGATT
TTGATCAGAGCTGCATCTGTGACTTATGAAAGACCACCTGCTTTTTATCGAGCAGTGCCCGAAGTTTAAGTACTC
TAAATGTGACTTCTAGACTTGAATCACNTAGTGAATTCGCGGCCGCCTGCAGGTCGACCATWATGGGAGAGCTCC
CAACGCGTTGGATGCATAGCTGGAGTATTCTAATAGTGTCACCTTAANNAGCTTNGACGTAATCATGGGTCATAG
CTTGTTTCCTGNGTGAAATTGTTATCCGCTCACATTCNCNNCANCANNNANNNNCGGAAGCNNTAAAGNNTAANC
CTGGGGGGCTAATGATGGAGCTACTCANATTANNGGCNNTGNNCTNNTNGNCCGGCTATTTCANNNGNNANACNN
GGTCTNGTGCNN

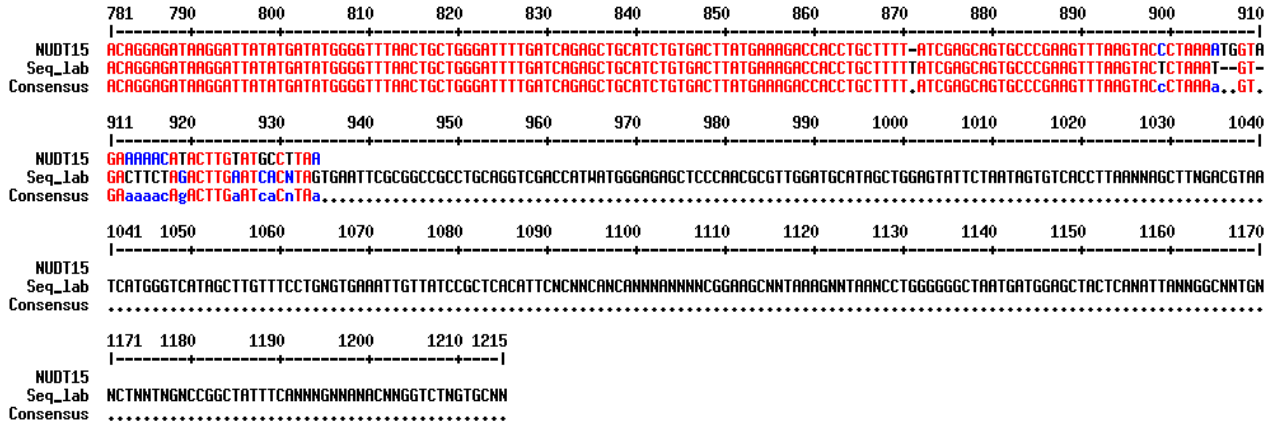


Figure 7.3: NUDT15 sequence alignment. NUDT15 sequence in pGEM-T Easy and that from the database were aligned.

The Seqlab NUDT15 sequence was 99% identical at the amino acid level to the cDNA of NUDT15 in the database and after the SeqLab sequence was translated at ExPASy for verification. The mismatches are only found toward the 3`end of the sequence (after 980 onwards) as shown in figure 7.3. This could be have been due to low confidence level in the sequence efficiency after 800 bp. Therefore, since the sequences are 100% identical upto 870 bp it was assumed that no mutations had been introduced in the subcloning.

7.1.2.2 NUDT15 cDNA in pCAT plasmids

The full-length cDNA amplified was PCR amplified from the RIKEN clone and subcloned into pCAT via pGEMT-Easy. The insert was sequenced from forward end using SR321F primer (see section 2.1.10).

Nucleotide sequence from Seqlab:

AACCACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAG
CTGTACAAGGCGGCCGCTATGTTTTTGCTTTATCGTAGGCTTCCTTCATTTGCACGAACAACAACAACAACCCTT
CTCTGCAAATCTATGGAGCCTGCGATAACAGCGACTTCTTCTTCTTCTTTCGGTGGTGGCTCTTCTCGTCTCGCC
GCTTTAGCCCAGCAACTTCGCCAATACAAGCCCCCACCTTCTTCATCGTTCGATGACTCCGAGGAGATGCAGACA
GATCAGGAGACCGCTGGGAAAGTCGTTTCTCAGGTTGGGTTTCAGGAATCTATTGCTCCTCTTTCGAAAGACCCT
GATAGGTTTAAACCCAAGAGAGCCGCTGTGTTGATCTGTCTCTTTGAAGGAGATGATGGTGATTTGCGTGTTATT
CTTACTAAGAGATCTTCCAAATTGTCTACTCACTCTGGAGAAGTTTCATTGCCAGGTGGTAAAGCGGAGGAGGAT
GATAAAGATGATGGGATGACTGCTACCAGAGAGGCTGAGGAAGAGATTGGATTGGACCCTTCTCTTGTTGATGTT
GTTACTTCTCTTGAACCATTTCTGTCTAAGCATCTTCTTAGAGTAATTCCTGTGATAGGCATCTTGAGGGACAAA
AATAAATTCAATCCGATACCAAATCCTGGGGAAGTGGAAGCTGTGTTTGATGCACCCTTGGAAATGTTCCTTAAG

GATGAGAATCGAAGATCTGAAGAGAGAGAGTGGATGGGTGAAAAGTATTTGATCCATTACTTTGACTACAGAACA
GGAGATAAGGATTATATGATATGGGGTTTAACTGCTGGGATTTTGATCAGAGCTGCATCTGTGACTTATGAAAGA
CCACCTGCTTTTATCGAGCAGTGCCCNAAGTTAANNNCCCTAAAATGTGACTCTNGAGTCCGCAAAATCACCNGT
CTCTCTNTACAATNTATCTCNNNCTATTTTTCTCNNAAATATGTGTGAGTANNTCCNNNAANAGGNAATAAGGNT
CANATAGGNTTCGCTCATGTGTGAGCANNTAGAAACCCTAGNNGNANTNNNTTGTANANACTCNTATCANTAANT
TNTNANTCTGAANCGAATTCTAGNACTGAAGNATNNNGACTGNNNTGNNCGTNNTTNNNACGTCAGGTACNTGAN
AAC

Translation of Seqlab sequence 5` end 99 base pairs at ExPASY gave the amino sequence shown below and the highlighted part is the N-terminal ligation region of the NUDT7 with EYFP via NotI:

5'3' Frame 1

E K R D H **Met** V L L E F V T A A G I T L G **Met** D E L Y K A A A **Met** F N H

>YFP_withNotI

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFGYGLQCFARY
PDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHN
VYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSYQSALSKDPNEKRDHMVLLEFV
TAAGITLGMDELYKAAAM

The Seqlab sequence is 100% identical at the amino acid level to the cDNA of NUDT15 in the database. The C-ternimus of YFP_NotI together with the NUDT15 N-terminus would be expressed into the fusion protien as anticipated.

7.1.3 TIR-NBS-LRR1 (ATP-BP, At1g72840, SSL>)

Cloning strategy of a very long protein such as ATP-BP, possibly containing in the middle two transmembrane domains was by cloning separately of two constructs: a) full-length, b) C-terminal 420-aa domain. **NOTE:** C-terminal 420aa-domain that was used in the cloning strategy was from 624-1042 amino acid residues.

7.1.3.1 Sequence analysis of ATP-BP full-length in pGEM-T Easy

Subcloning method: The full-length cDNA amplified was PCR amplified from the RIKEN clone pda19420 and subcloned into pGEMT-Easy. The insert was sequenced from forward end using T7 primer.

Nucleotide sequence from seqlab:

CATGCTCCGGCCGCCATGGCGGCCGCGGGAATTCGATTAAGACTGCGGCCGCTATGGCTTCCTCGTCATCATCTT
CTGCAACTCGTCTCAGGCACTACGATGTCTTCCTCAGTTTTCGAGGGGTAGATACCCGCCAAACCATCGTCAGCC
ATTTGTATGTGGCTCTACGTAATAATGGAGTTCTTACTTTTAAAGATGATCGGAAGCTCGAGATTGGCGACACCA
TTGCCGATGGTCTAGTCAAAGCTATACAAACTTCGTGGTTTGCGGTGGTTATTCTCTCTGAAAACTACGCTACTT
CGACGTGGTGCTTGGAGGAGCTCCGGTTGATAATGCAGCTTCACAGTGAGGAGCAGATCAAAGTGCTTCCTATCT
TCTACGGCGTAAAACCCTCTGACGTGAGATACCAGGAAGGAAGCTTCGCGACTGCCTTTCAAAGGTACGAAGCAG
ATCCGGAGATGGAGGAGAAGGTTTCTAAATGGAGAAGAGCTCTCACCCAAGTCGCTAATCTATCAGGCAAGCATT
CCAGAAATTGCGTGGATGAGGCAGATATGATAGCCGAGGTAGTTGGAGGCATCTCAAGTCGACTGCCAAGGATGA
AGTCGACAGATTTGATTAATTTAGTTGGAATGGAAGCTCATATGATGAAGATGACTCTCCTTCTGAATATTGGTT
GTGAAGACGAGGTTCATATGATAGGGATCTGGGGAATGGGAGGCATAGGCAAATCCACCATTGCCAAGTGTCTCT
ATGATCGATTTTCACGTCAATTTCCAGCTCACTGTTTTTTGGAAAACGTGTCTAAAGGCTATGATATTAAGCATC
TACAAAAGGAATTGCTTTCCCATATCCTCTATGATGAAGATGTCGAGTTATGGAGCATGGAAGCTGGATCCCAAG
AGATAAAGGAGAGACTCGGGCATCAAAAAGTTTTTGTCGTGCTTGATAATGTCGATAAAGTGGAGCAGTTACATG
GGCCTGACAAAGGACCCAAGCTGGGTTCGTTCCAGGGAGCCGTATCATCATAACCACACGAAGAACAAAGTTTGC
TCAATTCCTGCGGAGTAAACAATTANTATGAGTAAGTGCTGNCGATAGATGCGCTCAGTTTNNAGTTAGCTTTGG
GGAANAAACTNNTCGATGTTNGACNNNNTTAATCANAGCTNTCGGCCNGGCTCACAGNN

This Seqlab sequence is 94% identical at the amino acid level to the cDNA of ATP-BP in the database and the Seq Lab sequence has been translated at ExPASy for verification. The mismatches are only found toward the the end of the sequence (after 980 onwards). This could be have resulted due to low sequence efficiency after 800. Therefore, since the sequences are 100% identical upto 980bp the sample can be assumed not to contain any mutations or alternatively sequence it from the reverse end.

7.1.3.2 ATP-BP 420aa C-terminal sequence in pGEM-T Easy

Subcloning method: The full-length cDNA amplified was PCR amplified from the RIKEN clone and subcloned into pGEMT-Easy. The insert was sequenced from forward end using T7 primer.

Nucleotide sequence from seqlab:

GCTCCGGCCGCCATGGCGGCCGCGGGAATTCGATTAAGACTGCGGCCGCTCTGCTTCCAAACCTACGGATACTAG
ATGTAACAGGATCGAGGAATCTCAGAGAACTTCCAGAACTTTCGACCGCAGTAAATCTTGAAGAGTTGATATTGG
AAAGCTGTACGAGCCTGGTGCAAATCCCAGAGTCTATTAATAGATTATATCTGAGGAAACTAAATATGATGTACT
GTGATGGTCTTGAGGGAGTGATACTCGTCAATGACCTTCAAGAAGCCAGCCTCAGCCGCTGGGGCCTCAAACGGA
TTATACTGAACCTTCCTCATTCAGGGGCGACACTGAGTTCTCTGACAGATCTAGCTATCCAGGGGAAAATATTCA
TTAAGTTGTCGGGTCTCTCGGGTACGGGAGACCATCTGTCTTTTAGTTCTGTGCAGAAGACCGCTCATCAATCAG
TAACACATCTACTTAACTCTGGTTTCTTTGGTTTGAAATCACTCGACATCAAGTGGTTCAGTTACAGGTTGGATC
CTGTTAATTTCAGCTGTCTTAGCTTTGCAGACTTTCCATGTCTGACCGAGCTAAAGCTGATAAACTTAAACATTG
AAGACATCCCTGAAGACATATGTCAGTTGCAGCTCCTAGAGACACTGGACCTCGGTGGAAATGATTTCGTGTATC
TACCCACATCCATGGGACAACTTGCCATGTTAAAGTACCTCAGCCTCAGTAACTGTCGCAGACTTAAGGCACTGC
CACAACTTTCTCAGGTGGAGAGACTCGTACTTTCTGGCTGTGTGAAGCTCGGATCATTGATGGGAATTCTTGGTG
CACGCAGATACAATTTGCTTGATTTTTGCGTTGAAAAATGCAAGAGTCTTGGATCATTGATGGGGAATTCTTAGT
GTGGAAAAATCAGCTCCAGGCAGAAACGAGTTGCTTGAGCTTAGCCCTTGAAAACTGTAAGAGTCTTGTGTCATT
ATCAGAGGAGCTTAGTCATTCNCCAAGTTAACATATCTAGATCTCAGCAGCCTNGAGTTTAGGAGATCCAACAAG

CATCANANAGTTATCCTTTATGAGACTCTCTACCTCAACANTNGCACAAATCTTTTCACTGANNGATCCTNNNNN
AGCNNAAGTATNTCTATGGCATGNNCNAATCNTGGACANGTTAACNTCCTTCGGANTCNATTCCNTTNCANN

7.1.3.3 ATP-BP 420aa C-terminal sequence in pCAT vector

The full-length cDNA amplified was PCR amplified from the RIKEN clone ATP-BP last exon and subcloned into pCAT vector via pGEM-T Easy. The insert was sequenced from forward end using SR321f primer:

Nucleotide sequence from Seqlab:

GTNCCACGAGAAGCGCGATCCATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAG
CTGTACAAGGCGGCCGCTCTGCTTCCAAACCTACGGATACTAGATGTAACAGGATCGAGGAATCTCAGAGAACTT
CCAGAACTTTCGACCGCAGTAAATCTTGAAGAGTTGATATTGGAAAGCTGTACGAGCCTGGTGCAAATCCCAGAG
TCTATTAATAGATTATATCTGAGGAAACTAAATATGATGTACTGTGATGGTCTTGAGGGAGTGATACTCGTCAAT
GACCTTCAAGAAGCCAGCCTCAGCCGCTGGGGCCTCAAACGGATTATACTGAACCTTCCTCATTCAGGGGCGACA
CTGAGTTCTCTGACAGATCTAGCTATCCAGGGGAAAATATTCATTAAGTTGTCGGGTCTCTCGGGTACGGGAGAC
CATCTGTCTTTTAGTTCTGTGCAGAAGACCGCTCATCAATCAGTAACACATCTACTTAACTCTGGTTTCTTTGGT
TTGAAATCACTCGACATCAAGTGGTTCAGTTACAGGTTGGATCCTGTTAATTTCAGCTGTCTTAGCTTTGCAGAC
TTTCCATGTCTGACCGAGCTAAAGCTGATAAACTTAAACATTGAAGACATCCCTGAAGACATATGTCAGTTGCAG
CTCCTAGAGACACTGGACCTCGGTGGAATGATTTCGTGTATCTACCCACATCCATGGGACAACTTGCCATGTTAA
AGTACCTCAGCCTCAGTAACTGTCGCAGACTTAAGGCACTGCCACAACTTTCTCACGTGGAGAGACTCGTACTTT
CTGGCTGTGTGAAGCTCGGATCATTGATGGGAATTCTTGGTGCAGGCAGATACAATTTGCTTGATTTTTGCGTTG
AAAAATGCAAGAGTCTTGGATCATTGATGGGAATTCTTAGTGTGGAAAAATCAGCTCCAGGCAGAAACGAGTTGC
TTGAGCTTAGCCTTGAAAACTGTAAGAGTCTTGTGTCATTATCAGAGGAGCTTAGTCATTTCNCCAAGTTANCAT
ATCTAGATCTCAGCAGCNCGAGTTTAGGANATCCANNGCATCAGAGAGTTATNCNTATGAGANTCTCTACTCACA
CTGCACNAANCTTTNNNGACGAATCTCNANGCCTAGATCNNNTGCATGNNCCNNNTGGACCATGTTTANNNN

Translation of Seqlab sequence 5` end 99 base pairs at ExPASY gave the amino sequence shown below and the highlighted part is the N-terminal ligation region of the NUDT7 with EYFP via NotI:

5'3' Frame 1

X P R E A R S **Met** V L L E F V T A A G I T L G **Met** D E L Y K A A A L L P N L R I L

>YFP_withNotI

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLV
TTFGYGLQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRI
ELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQN
TPIGDGPVLLPDNHYLSYQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKAAA

This Seqlab sequence is 96% identical at the amino acid level to the cDNA of ATP-BP 420aa C-Terminus in the database and the ATP-BP 420aa C-Terminus has been independently translated at ExPASy for verification. The C-ternimus of YFP_NotI together with the ATP-BP 420aa were analysed. The ATP-BP 420aa C-Terminus is in a proper orientation therefore, YFP, NotI and ATP-BP 420aa C-Terminus will be expressed correctly and the anticipated fusion protein will result.

7.1.4 The CHAT homolog (formerly „TF1", At5g17540)

7.1.4.1 The CHAT homolog sequence in pGEM-T Easy

The full-length cDNA amplified was PCR amplified from the French INRA clone, BX830423 and subcloned into pGEMT-Easy. The insert was sequenced from forward end using T7 primer.

Nucleotide sequence from seqlab:

GCTCCGGCCGCCATGGCGGCCGCGGGAATTCGATTAAGACTGCGGCCGCTATGTCCGGGTCACTCACGTTTAAGA
TTTACCGGCAGAAGCCGGAGTTAGTTTCTCCGGCGAAGCCAACGCCAAGAGAGCTCAAACCCCTCTCAGATATTG
ACGACCAAGAAGGACTAAGATTTCACATTCCCACTATCTTTTTCTATAGACACAACCCTACTACTAACTCTGATC
CTGTCGCAGTCATTCGGAGAGCTCTCGCAGAGACGCTGGTTTACTACTATCCGTTCGCCGGTAGGCTTCGGGAAG
GACCGAACCGGAAACTGGCTGTGGATTGTACCGGTGAAGGCGTTTTGTTTATTGAGGCTGATGCTGACGTGACAC
TTGTTGAGTTTGAAGAGAAGGATGCTCTTAAGCCTCCTTTCCCTTGCTTTGAAGAGCTTCTGTTTAACGTTGAAG
GTTCTTGTGAAATGCTCAACACTCCTTTGATGCTCATGCAGGTCACGCGCTTGAAATGCGGCGGTTTTATCTTCG
CCGTCCGTATCAACCACGCAATGTCCGATGCCGGCGGTCTCACGCTGTTCCTCAAAACGATGTGCGAGTTCGTGC
GTGGTTATCATGCACCTACGGTTGCTCCGGTGTGGGAACGTCACCTGCTGAGCGCCAGAGTCCTGCTGCGTGTGA
CACACGCACACCGAGAGTACGACGAAATGCCGGCAATAGGTACAGAACTCGGCAGTAGAAGAGACAATCTGGTAG
GCCGGTCACTCTTCTTCGGTCCCTGCGAGATGTCCGCAATACGCAGGCTCCTCCCACCAAATCTTGTCAACAGCA
GCACCAATATGGAAATGCTAACGTCTTTCTTATGGCGTTATCGCACCATCGCTCTACGACCAGACCAGGACAAGG
AGATGCGGCTCATATTAATTTGTCAACGCACGTTCTAGCTTAAAAATCCACCACTACCTCGAGGATACTACGGAA
ATGCCATTTGCNTCCAGTCNCCATNGCAACAGCTAATGAACTAACTAAGAAACCGTAGANTTTGCNNTGAGACTT
ATNAATGANGNNGAAANCNNAGCGTGANCGGAGGAGTACAGGNNATNACTNGCGGATNNTGATGGTGAAAGGNAG
AACCAAGCTTCTCGTCGGACGGACCTACTTGNNNNNNNNTAGANTTTCNNTATGAATTTCGGNATTTTGGGGGNAA
CATCCTNG

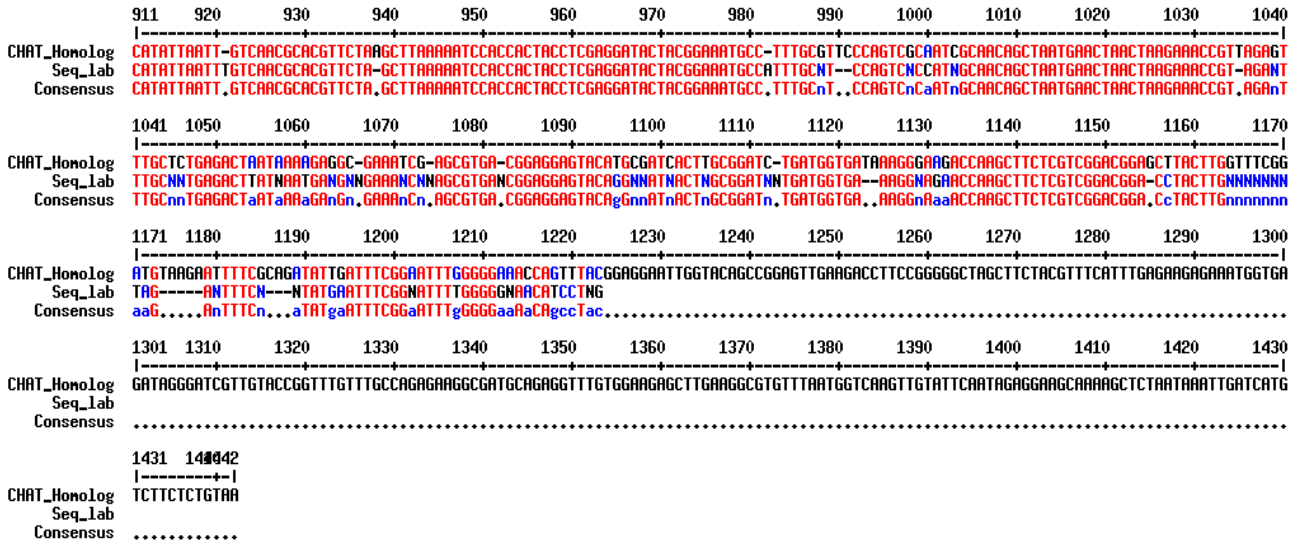


Figure 7.4: The CHAT homolog sequence alignment.

The Seqlab CHAT homolog sequence is 99% identical at the amino acid level to the cDNA of the CHAT homolog in the database and that from Seqlab were aligned (Fig. 7.4). The mismatches were only found toward the end of the sequence (after 920 onwards). This could be have resulted due to low confidence level in the sequence efficiency after 800bp. Therefore, since the sequences are 100% identical upto 920bp the sample can assumed not to contain any mutations.

7.1.4.2 The CHAT homolog sequence in pCAT vector

The full-length cDNA amplified was PCR amplified from the French INRA clone, BX830423, and subcloned into pCAT via pGEMT-Easy. The insert was sequenced from forward end using SR321f primer (see section 2.1.10).

Nucleotide sequence from Seqlab:

GACTACGAGAAGCGCGATCCATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGC
TGTACAAGGCGGCCGCTATGTCCGGGTCACTCACGTTTAAGATTTACCGGCAGAAGCCGGAGTTAGTTTCTCCGG
CGAAGCCAACGCCAAGAGAGCTCAAACCCCTCTCAGATATTGACGACCAAGAAGGACTAAGATTTCACATTCCCA
CTATCTTTTTCTATAGACACAACCCTACTACTAACTCTGATCCTGTCGCAGTCATTCGGAGAGCTCTCGCAGAGA
CGCTGGTTTACTACTATCCGTTCGCCGGTAGGCTTCGGGAAGGACCGAACCGGAAACTGGCTGTGGATTGTACCG
GTGAAGGCGTTTTGTTTATTGAGGCTGATGCTGACGTGACACTTGTTGAGTTTGAAGAGAAGGATGCTCTTAAGC
CTCCTTTCCCTTGCTTTGAAGAGCTTCTGTTTAACGTTGAAGGTTCTTGTGAAATGCTCAACACTCCTTTGATGC
TCATGCAGGTCACGCGCTTGAAATGCGGCGGTTTTATCTTCGCCGTCCGTATCAACCACGCAATGTCCGATGCCG
GCGGTCTCACGCTGTTCCTCAAAACGATGTGCGAGTTCGTGCGTGGTTATCATGCACCTACGGTTGCTCCGGTGT
GGGAACGTCACCTGCTGAGCGCCAGAGTCCTGCTGCGTGTGACACACGCACACCGAGAGTACGACGAAATGCCGG
CAATAGGTACAGAACTCGGCAGTAGAAGAGACAATCTGGTAGGCCGGTCACTCTTCTTCGGTCCCTGCGAGATGT
CCGCAATACGCAGGCTCCTCCCACCAAATCTTGTCAACAGCAGCACCAATATGGAAATGCTAACGTCTTTCTTAT
GGCGTTATCGCACCATCGCTCTACGACCAGACCAGGACAAGGAGATGCGGCTCATATTAATTGTCAACGCACGTT
CTAAGCTTAAAAATCCACCACTACCTCGAGGATACTACGGAAATGCCTTTGCGTTCCCAGTCGCAATCGCAACAG
CTAATGAACTACTAGAACCGTTAGAGTTTGCTCTGAGACTAATAAAAGAGCGAAATCGAGNGNGACGAGNAGTNN
TGCGATCACTGNNGATNTGATGGTGATAAAGNAAGAANNGCTCNGTCGACGACTACTGNNNNANTAGANTTNNGA
ATTNNATTCGGANNTTGGGGGG

Translation of 5` end 99 base pairs of the Seqlab sequence:

5'3' Frame 3

L R E A R S **M** V L L E F V T A A G I T L G **M** D E L Y K A A A **M** S

>YFP_withNotI

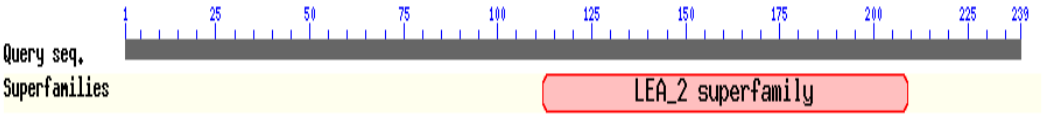
MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLV
TTFGYGLQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRI
ELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQN
TPIGDGPVLLPDNHYLSYQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKAAA

The Seqlab CHAT homolog sequence is 99% identical at the amino acid level to the cDNA of CHAT homolog in the database and the SeqLab sequence. The mismatches are only found toward the end of the sequence (after 920 onwards). This could be have resulted due to low confidence level in the sequence efficiency after 800bp. Therefore, since the sequences are 100% identical upto 920bp the sample can assumed not to contain any mutations. The C-ternimus of YFP_NotI together with the CHAT homolog have been analysed. The CHAT homolog is in a proper orientation therefore, YFP, NotI and CHAT homolog will be expressed correctly.

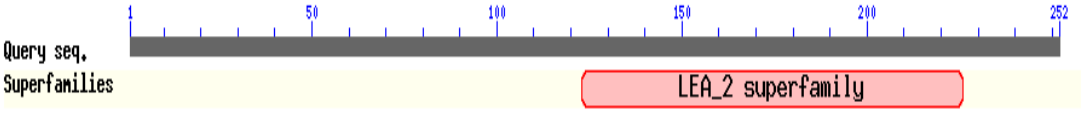
7.2 Gene expression analysis supplementary data

7.2.1 Microarray data of NHL proteins from Genevestigator and NCBI databases

NHL4(At1g54540)



NHL6(At1g65690)



NHL25(At5g36970)

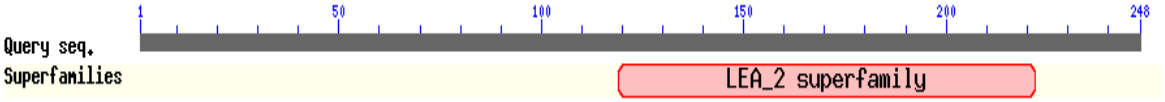


Figure 7.5: Conserved Domains for NHL proteins from publicly available database, National Centre for Biotechnology Information **(NCBI)** http://www.ncbi.nlm.nih.gov/ Different types of LEA (Late Embryogenesis Abudant) proteins are expressed at different stages of late embryogenesis in higher plant seed embryos and under conditions of dehydration stress.

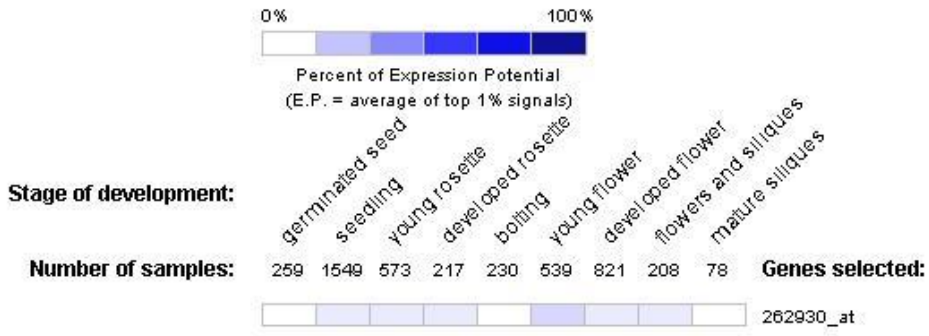


Figure 7.6: NHL6 expression in Arabidopsis plants at different developmental stages

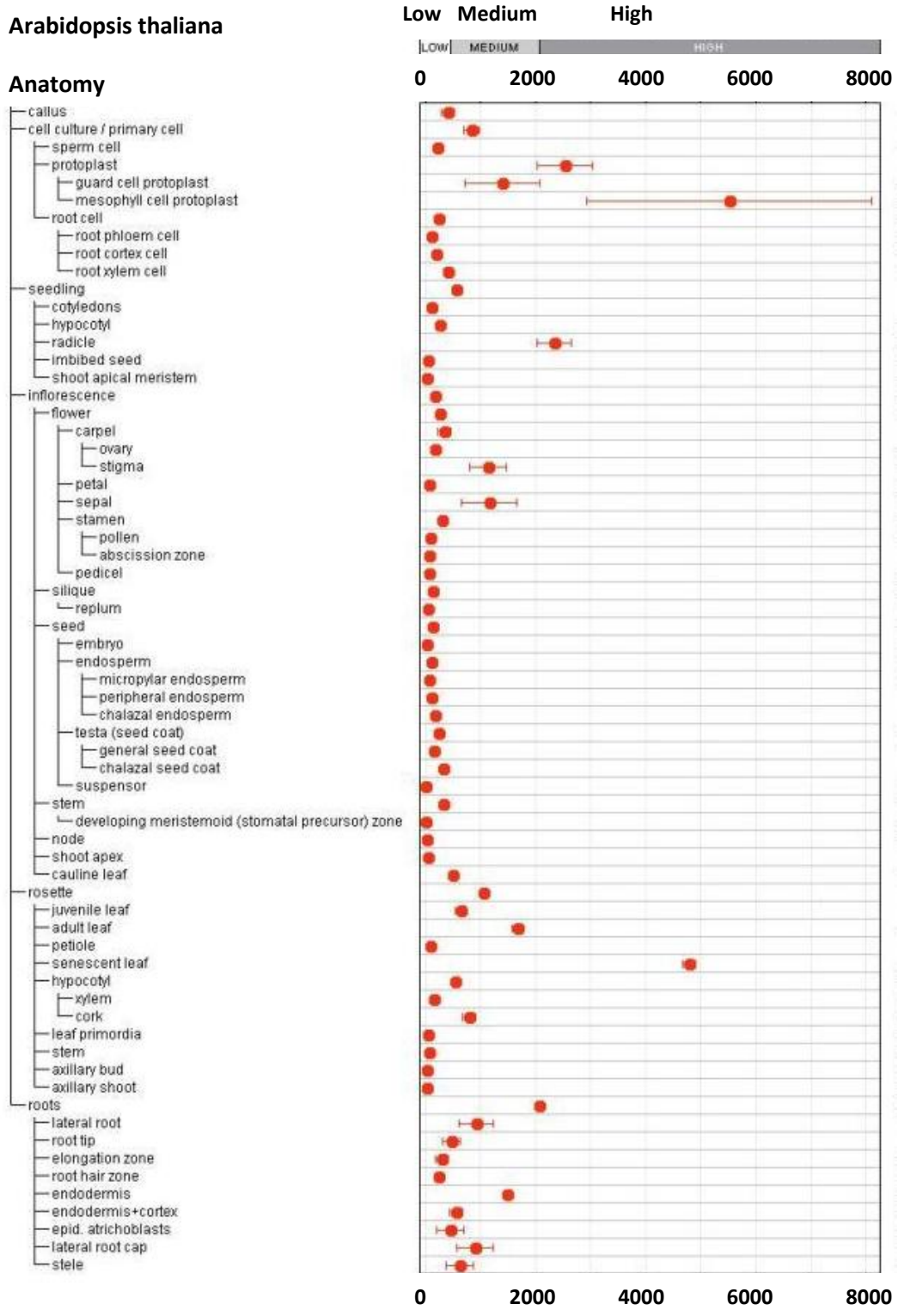


Figure 7.7: NHL6 expression in different parts of the Arabidopsis plant

7.2.2 Relative gene quantification results

Table 7.1: Gene expression of NHLs, IANs and Salicylic acid positive markers (PR2/5)

Gene	**ACT2**				**NHL4**				**IAN8**		
Sample	0 Hour	24 Hours	72 Hours		0 Hour	24 Hours	72 Hours		0 Hour	24 Hours	72 Hours
Mean Ct	20.671	21.489	20.366		27.081	27.243	27.724		>40	31.827	34.996
Fold Change	N/A	N/A	N/A		1	1.576	0.519		N/A	N/A	N/A
Gene	**PR2**				**NHL6**				**IAN11**		
Sample	0 Hour	24 Hours	72 Hours		0 Hour	24 Hours	72 Hours		0 Hour	24 Hours	72 Hours
Mean Ct	25.313	22.119	22.020		28.639	27.359	27.891		29.962	30.494	30.843
Fold Change	1	16.137	7.936		1	4.279	1.359		1	1.219	0.440
Gene	**PR5**				**NHL25**				**IAN12**		
Sample	0 Hour	24 Hours	72 Hours		0 Hour	24 Hours	72 Hours		0 Hour	24 Hours	72 Hours
Mean Ct	24.378	22.972	22.198		34.830	34.061	35.276		29.953	30.263	29.271
Fold Change	1	4.671	3.669		1	3.002	0.594		1	1.422	1.299

7.3 Co-expression analysis supplementary data

Table 7.2: Raw dataset generated from the bioinformatics tool

AGI	Query	Number of co-expressed genes	Lowest correlation score	# of genes used
At3g07040	RPM1	**very many, >100 with score > 0.6**	**0.772**	100
At1g65690	NHL6	**very many, >100 with score > 0.6**	**0.749**	100
At3g51660	MIF1	**very many, >100 with score > 0.6**	**0.738**	100
At2g35980	NHL10/NtHin1	**very many, >100 with score > 0.6**	**0.729**	100
At5g52640	HSP90	**very many, >100 with score > 0.6**	**0.707**	100
At5g59820	ZAT12	**many, >100 with score > 0.6**	0.692	100
At3g54920	PMR6	**many, >100 with score > 0.6**	0.677	100
At4g11260	EDM1	**many, >100 with score > 0.6**	0.674	100
At3g52430	PAD4	**many, >100 with score > 0.6**	0.673	100
At3g48090	EDS1	**many, >100 with score > 0.6**	0.670	100
At5g51700	PBS2/RAR1	**many, >100 with score > 0.6**	0.664	100
At4g26090	RPS2	**many, >100 with score > 0.6**	0.644	100
At1g74710	SID2	**many, >100 with score > 0.6**	0.625	100
At4g37980	ELI3	**many, >100 with score > 0.6**	0.620	100
At3g25070	RIN4	**many, >100 with score > 0.6**	0.612	100
At5g58600	PMR5	**many, >100 with score > 0.6**	0.609	100
At1g58807	DRP	**many, >100 with score > 0.6**	0.604	102
At4g16890	SNC1	some	0.600	46
At1g33960	IAN8	some	0.698	34
At5g17890	LIMDP	some	0.606	28
At3g20600	NDR1	some	0.568	15
At5g55390	EDM2	some	0.600	13
At5g64930	CPR5	some	0.601	11
At1g64280	NPR1	few	0.602	7
At4g18470	SNI1	few	0.604	2
At4g09940	IAN12	none	0	0
At1g54540	NHL4	none	0	0
At4g09930	IAN11	none	0	0

Table 7.3: Annotations of the 28 Arabidopsis defence proteins

Acronym	Annotation
RPM1	RESISTANCE TO PSEUDOMONAS SYRINGAE PV MACULICOLA 1
NHL6	NDR1/HIN1-like 6
MIF1	Macrophage migration Inhibitory Factor 1
NHL10	NDR1/HIN1-like 10
HSP90	HEAT SHOCK PROTEIN 90
ZAT12	zinc finger protein involved in high light and cold acclimation
PMR6	POWDERY MILDEW RESISTANT 6
EDM1	ENHANCED DOWNY MILDEW 1,
PAD4	PHYTOALEXIN DEFICIENT 4
EDS1	ENHANCED DISEASE SUSCEPTIBILITY 1
PBS2/RAR1	PPHB SUSCEPTIBLE 2, REQUIRED FOR MLA12 RESISTANCE 1
RPS2	RESISTANT TO P. SYRINGAE 2
SID2	SALICYLIC ACID INDUCTION DEFICIENT 2
ELI3	elicitor-activated gene 3
RIN4	RPM1 INTERACTING PROTEIN 4
PMR5	POWDERY MILDEW RESISTANT 5
DRP	Disease resistance protein (CC-NBS-LRR class) family
SNC1	SUPPRESSOR OF NPR1-1, CONSTITUTIVE 1
IAN8	Immune-Associated Nucleotide-binding 8
LIMDP	DA1-related protein
NDR1	non race-specific disease resistance
EDM2	ENHANCED DOWNY MILDEW 2
CPR5	CONSTITUTIVE EXPRESSION OF PR GENES 5
NPR1	NONEXPRESSER OF PR GENES 1
SNI1	SUPPRESSOR OF NPR1-1, INDUCIBLE 1
IAN12	Immune-Associated Nucleotide-binding12
NHL4	NDR1/HIN1-like 4
IAN11	Immune-Associated Nucleotide-binding11

7.4 Abbreviations

aa	amino acid
Amp	ampicillin
bp	base pair
cDNA	complementary DNA
CT	Threshold cycle
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphates
E. coli	*Escherichia coli*
ER	endoplasmic reticulum
EYFP	enhanced yellow fluorescent protein
H_2O_2	hydrogen peroxide
IPTG	isopropyl-β-D-thiogalactopyranoside
kbp	kilobase pair
LB medium	Luria-Bertani medium
PCR	polymerase chain reaction
PTS	protein targeting signal
qPCR	Real-time quantitative polymerase chain reaction
ROS	reactive oxygen species
Rn	Normalized fluorescence
SA	Salicylic acid
T_a	annealing temperature
Taq	*Thermus aquaticus*
T_m	melting temperature
UV	ultraviolet
w/v	weight to volume
X-gal	5-bromo-4-chloro-3-indonyl-β-D-galactopyranoside