Characterization of Activation Tagged Potato (*Solanum tuberosum* L.) Mutants

By

Sukhwinder Singh Aulakh

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Horticulture

Richard E. Veilleux
Barry Flinn
Jerzy Nowak
Chuansheng Mei
Allan W. Dickerman

September 14, 2012
Blacksburg, VA

Keywords: Solanaceae, BTB/POZ domain, RNA-seq, GCN5, Sugar transporter, revertant, methylation

Characterization of Activation Tagged Potato (*Solanum tuberosum* L.) Mutants

Sukhwinder Singh Aulakh

ABSTRACT

Generation and characterization of activation tagged potato mutants could aid in functional genomic studies. Morphological and molecular studies were conducted to compare potato cv. Bintje, its two mutants, *underperformer* (*up*), and *nikku* generated using the activation tagging vector pSKI074, and *nikku* revertant plants. Mutant *up* exhibited a dwarf phenotype (plant height 42 cm vs. 73 cm in cv. Bintje), abundant axillary shoot growth (3.1 shoots/plant compared to 0.7 shoots/plant in cv. Bintje; *in vitro* plants), greater tuber yield, altered tuber traits and early senescence compared to wild-type Bintje under *in vitro* conditions. Under *in vivo* conditions, the dwarf and early senescence phenotypes of the mutant were consistent, but the tuber yield of *up* was less (250 g/plant compared to 610 g/plant in wild-type Bintje) and had fewer axillary shoots compared to wild-type (1.9 shoots/plant in *up* vs. 4.7 shoots/plant in Bintje). Mutant *nikku* plants exhibited an extremely dwarf phenotype (plant height 2 cm in *nikku* vs. 6 cm in Bintje), had small hyponastic leaves, were rootless, and infrequently produced small tubers when compared to cv. Bintje. The overall *nikku* phenotype was suggestive of a constitutive stress response, which was further supported by the higher expression levels of several stress-responsive genes in *nikku*. The *nikku* revertant plants exhibited near normal stem elongation, larger leaves and consistent rooting, and it was a case of partial reversion. Southern blot analyses indicated the presence of single T-DNA insertions on chromosome 10 in the *up* and on chromosome 12 in the *nikku* mutant. The reversion in the *nikku* plants was not associated with the loss of enhancer copies from the original *nikku* mutant. Reverse transcriptase PCR analyses indicated transcriptional activation/repression of several genes in the *up* and *nikku* mutants, suggesting pleiotropic effects. In revertant, the expression levels of several genes which were differentially regulated in the *nikku* mutant were similar to Bintje. The gene immediately flanking the right border of the T-DNA insertion, which encoded a novel BTB/POZ (Broad complex, Tramtrac, Bric a brac; also known as Pox virus and Zinc finger) domain-containing protein, was highly up-regulated in the *up* mutant. This protein domain

plays an important role in several important developmental, transcriptional and regulatory pathways. The mRNA-seq analyses resulted in 1,632 genes that were differentially expressed between mutant *up* and Bintje and the total number of up-regulated genes (661) were less than the number of genes down-regulated (971 genes) in the *up* mutant. Further analyses indicated that a variety of biological processes including decreased cell division, cell cycle activity, and abiotic stress responses were modified in the *up* mutant. In the *nikku* mutant, two potato genes, encoding an Acyl-CoA N-acyltransferases (NAT) superfamily protein, and a predicted major facilitator superfamily protein (MFS) were identified and overexpression lines Bintje/35S::NAT1 and Bintje/35S::PMT1 were created for recapitulation of the *nikku* mutant phenotype. Methylated DNA-PCR between the *nikku* and the revertant indicated a change in methylation status of the 35S enhancers, suggesting that the *nikku* revertant phenotype may be associated with some epigenetic modification.

DEDICATION

I dedicate this dissertation to my parents

Late Anokh Singh, Gurmej Kaur and Uncle Bhagwan Singh

My wonderful wife Kavita

And

Aulakh and Burman families

ACKNOWLEDGMENTS

I want to thank the Almighty for bestowing me with the ability to initiate and complete this task. I wish to thank all those who have helped me in this investigation and in writing the dissertation. To them I would like to convey my heartfelt gratitude and sincere appreciation. I would like to acknowledge the financial support provided by the Department of Horticulture, Virginia Tech and the Institute for Advanced Learning and Research.

It gives me great pleasure to acknowledge my supervisor Dr. Richard Veilleux for his guidance, suggestions, constructive criticism and compliments throughout my period of study at Virginia Tech. He is one of the most humble human beings I have ever come across and is the best mentor a graduate student can have.

I want to thank my Co-supervisor Dr. Barry Flinn for giving me the opportunity to work on this great project. The completion of this research was not possible without his guidance. I am grateful for his encouragement, wisdom, incredible patience, for instilling in me incredible work ethics and help throughout my research at the IALR.

I would like to thank my dissertation committee members, Dr. Jerzy Nowak, Dr. Chuansheng Mei and Dr. Allan Dickerman for help in data analysis, continuous support, valuable inputs and comments. Thanks to former members of my committee, Dr. A. Ozzie Abaye and Dr. M. Javed Iqbal for their help and contribution to my project and my training as a researcher.

Grateful acknowledgment is also expressed to staffs of the Department of Horticulture, Virginia Tech and the IALR especially Joyce, Maura, Pris, Jeff (all at VT), Nancy, Jeff, IT, administrative and security personals (all at IALR) for help and troubleshooting the administrative and research related problems. My special thanks go to the members of Veilleux lab (especially Jared and Kendall) for taking care of plants and ISRR lab members (especially Guozhu) for valuable discussions and excellent technical support. Also, I express my thanks to my friends both at Virginia Tech and ISRR, Alejandra, Daljit, Juan, Linda, Rachel, Rajesh, Rubina, Sarah, Scott, Seonhwa, Supriya, Yeun Kyung and fellow graduate students for helping me during my stay at Virginia Tech and IALR.

I express my deep sense of reverence and gratitude to my parents, family members and all my mentors. Last, but certainly not least, I would like to thank my wife, Kavita, for her love, friendship and contribution in this dissertation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTARY FIGURES

LIST OF SUPPLEMENTARY TABLES

Chapter 1

Characterization of *underperformer*, an activation tagged potato mutant

Sukhwinder S. Aulakh[1, 2], Richard E. Veilleux[1*], Guozhu Tang[2] and Barry S. Flinn[1, 2*]

Abstract

The tetraploid (2*n*=4*x*=48) nature of the potato genome, as well as the polyploid nature of many crops, make activation tagged mutants an invaluable resource for functional genomic studies. Morphological and molecular studies were conducted to compare potato cv. Bintje and a Bintje mutant, *underperformer* (*up*), generated using the activation tagging vector, pSKI074. Detailed morphological studies revealed that *up* exhibited a dwarf phenotype (mutant plant height was 42 cm compared to 73 cm in wild-type Bintje; *in vivo* plants), abundant axillary shoot growth (3.1 shoots/plant compared to 0.7 shoots/plant in wild-type Bintje; *in vitro* plants), greater tuber yield, altered tuber traits and early senescence compared to wild type under *in vitro* conditions. Under *in vivo* conditions, the dwarf and early senescence phenotypes of the mutant were consistent, but the tuber yield of *up* was less (250 g/plant compared to 610 g/plant in wild-type Bintje) and had fewer axillary shoots compared to wild-type (1.9 shoots/plant in *up* vs. 4.7 shoots/plant in Bintje). Southern blot analyses indicated the presence of a single T-DNA insertion on chromosome 10 in the mutant. Initial gene expression studies using reverse transcriptase PCR to compare wild-type Bintje and mutant *up* indicated transcriptional activation/repression of several genes in the mutant, suggesting that activation tagging had pleiotropic effects on various genes flanking the insertion. The gene immediately flanking the right border of the T-DNA insertion, which encoded a novel BTB/POZ (Broad complex, Tramtrac, Bric a brac; also known as Pox virus and Zinc finger) domain-containing protein, was highly up-regulated in the mutant. This protein domain plays an important role in mediating protein interactions, affecting global repression of transcription factors, plant defense, protein degradation and cell death, which may explain the molecular basis of the multiple phenotypic changes, observed in the *up* mutant.

Keywords: *Solanum tuberosum*, Solanaceae, BTB/POZ domain, senescence, pleiotropic

[1] Department of Horticulture, Virginia Tech, Blacksburg, Virginia, USA, 24060

[2] The Institute for Sustainable and Renewable Resources (ISRR) at The Institute for Advanced Learning and Research (IALR), Danville, Virginia, USA, 24540

* Both contributed to project development and supervision.

Introduction

Advances in genomic technology have allowed the sequencing and annotation of numerous plant genomes. Potato (*Solanum tuberosum* L.) is the 30th plant species and first asterid for which a genome sequence has been produced providing significant new information on potato gene models. The genome size of the doubled monoploid potato (*S. tuberosum* Group Phureja DM 1-3 566 R44) selected for sequencing is 844 megabases (Mb), and the assembly and annotation of 86% of this 844 Mb have predicted 39,031 protein coding genes and 62.2% repetitive content (The Potato Genome Sequencing Consortium, 2011). Similarly, in the closely related tomato genome, 900 Mb of sequence were generated from inbred cultivar Heinz 1706; the assembly and annotation of 760 Mb (84%) of sequence predicted 34,727 protein coding genes (The Tomato Genome Consortium, 2012). Several other valuable genetic resources are also available for potato, including the transcriptome of the reference potato genome (Massa et al., 2011), 249,761 potato Expressed Sequence Tags (ESTs) from NCBI (dbEST, release 030112) and microarray expression data (Kondrak et al., 2011). Of 23,208 protein coding gene groups of potato, tomato, grape, rice and Arabidopsis, 8,615 genes are common in all genomes (≥2 members of group) (The Tomato Genome Consortium, 2012). The potato genome reveals 2,642 genes specific to this large angiosperm clade, and given the important role of potato in world food production and security, it provides a platform for its genetic improvement. Many traits of interest in potato are quantitative in nature and access to the genome sequence will simplify their characterization and deployment in cultivars (The Potato Genome Sequencing Consortium, 2011), leading to improvement in the potato. However, as for all plant genomes, scientists are faced with the enormous task of functional annotation of the various genes. Mutant populations are a valuable resource for gene discovery, but few such resources are available for potato (Fischer et al., 2008) slowing the process of gene discovery and annotation. Therefore, the challenge for the post-sequencing era is to create valuable mutant resources in potato and to identify the biological functions of the predicted genes. Various approaches can be used to discover gene function, and the most direct is to disrupt the gene(s) and analyze the consequences. In this approach, also known as forward genetics, several methods are available for mutagenesis: chemical mutations such as ethyl methanesulfonate (EMS), fast-neutron treatment, and insertion of a transposable element or transfer DNA (T-DNA) into the genome.

Insertional mutagenesis using transposons or T-DNA has become an extremely valuable research tool for model plant systems. However, all gene disruption approaches also have some limitations. For example, it is difficult to identify the function of redundant genes, or of genes required in early embryogenesis or gametophyte development using the insertional (loss of gene function) mutants.

Activation tagging is a useful method for generating dominant or gain-of-function mutations, complementing conventional insertional mutagenesis and offering many advantages. In this method, a T-DNA or a transposable element containing multimerized cauliflower mosaic virus (CaMV) 35S enhancers (Hayashi et al., 1992; Suzuki et al., 2001) is inserted in the genome. These enhancers can function in either orientation and at a considerable distance from the coding regions, and can cause transcriptional activation of nearby genes, resulting in dominant gain-of-function mutations. Such gene activations may produce novel phenotypes, and can aid in determining important gene functions, including the function of redundant members of a gene family or genes essential for development and survival of an organism. However, T-DNA may also be integrated in the gene itself, causing disruption of gene function and creating a loss-of-function mutation. In Arabidopsis, the number of insertion sites on each Arabidopsis chromosome has been similar, ranging from 8.7-10/Mb, with 317 cases of intragenic (within ORF, leading to disruption of gene function) and 855 cases of intergenic (between ORFs, leading to activation of flanking genes) insertions in a population of 1,172 activation tagged mutants, indicating clear bias for T-DNA integration outside the transcribed region (Ichikawa et al., 2003). Most of plant genome consists of large intergenic regions, which further reduce the probability of a T-DNA landing in an intragenic region. Many genes are present in multiple copies in the genome with partially redundant functions (Hiratsu et al., 2003), and this aspect is even more pronounced in polyploid species like potato, making functional gene annotation difficult. However, the activation tagging approach could prove useful, given that it activates the transcription of flanking endogenous gene(s), and produces an obvious phenotype in most cases. This tagging system has been used for cloning several genes in Arabidopsis (Borevitz et al., 2000; Huang et al., 2001; Kakimoto, 1996; Kardailsky et al., 1999), petunia (Zubko et al., 2002), tomato (Mathews et al., 2003), barley (Ayliffe et al., 2007) and poplar (Busov et al., 2003;

Harrison et al., 2007). The only prerequisite for using activation tagging in any species is the need for at least one genotype to be efficiently transformed. Hence, activation tagging has been growing in popularity in diverse non-model plant species (Busov et al., 2011) but its utility has yet to be demonstrated in potato, a potential model crop for tuber-bearing species.

Potato is the world's third most important crop after wheat and rice, and the most widely grown vegetable crop, with more than half of its global harvest produced in developing countries (http://faostat.fao.org/). The cultivated potato (*Solanum tuberosum* L.) belongs to a large genus that includes 160 tuber bearing species, eight of which are cultivated. Most commercial potato cultivars are autotetraploid (2*n*=4*x*=48), having originated from diploid ancestors in the Andean Mountains of South America. As the cultivated potato is tetraploid, and most diploid potatoes are either self-incompatible or marginally fertile, the development of loss-of-function mutants is difficult. Improvement in potato is hampered by its large and complex genome, high heterozygosity, acute inbreeding depression, susceptibility to many devastating pests and abiotic stress. The development of resources and methods to elucidate the function of many potato genes remains an important goal for the realization of many of the potential benefits of the potato genome sequence (The Potato Genome Sequencing Consortium, 2011). The generation and characterization of activation tagged potato lines provide one such resource.

To assist the gene discovery process in potato, the Canadian Potato Genome Project (CPGP), created a population of activation tagged mutants by using the activation tagging vector pSKI074, that has been extensively used in Arabidopsis (Kardailsky et al., 1999; Weigel et al., 2000) and other species (Zubko et al., 2002). This collection of approximately 8,000 activation tagged lines was derived from the potato cv. Bintje, an important European cultivar that is a sterile hybrid from a cross between 'Munstersen' and 'Fransen'. A subset of these mutant lines has been screened for phenotypic traits but no detailed studies have been conducted. We selected two mutants (indexed as AT601 and AT615 in CPGP) from this collection based on obvious visible phenotypes under *in vitro* conditions. AT601 plants were extremely dwarf and rootless, whereas AT615 plants were dwarf, had more axillary branches but normal roots. In this paper, we report the results of in-depth characterization of the AT615 mutant, which we

termed *underperformer* (*up*), due to its overall poor performance, reduced organ size and early senescence when compared to wild type Bintje plants.

Results

Generation and screening of potato activation tagged mutants

The potato cv. Bintje was selected for the generation of activation tagged mutants due to its better response to shoot regeneration and transformation in a preliminary study (under CPGP) involving two cultivars and several different media. The activation tagging approach was successfully used by the Canadian Potato Genome Project (Regan et al., 2006) to create approximately 8,000 potato mutants from cv. Bintje using the activation tagging vector pSKI074 (Weigel et al., 2000). The mutant line AT615, which we designated *underperformer* (*up)* was selected in an *in vitro* screen. Following one month of growth, several new axillary stem structures were formed in the *up* plants, leading to branching, whereas wild-type Bintje exhibited a single stem (**Figure 1.1**). Transgenic plantlets were initially screened by polymerase chain reaction (PCR) to confirm the presence of the T-DNA plasmid sequences in the plant (data not shown).

Determination of T-DNA copy number by Southern blot analysis

To determine the T-DNA copy number in the *up* mutant, we performed Southern blot analysis with genomic DNA, using two restriction enzymes (*Pci*I and *Psi*I) that cut within the T-DNA, in the region just left of the tetramerized enhancer sequences. Using a probe that corresponded primarily to the enhancer region of pSKI074, we observed a single hybridizing band for both restriction enzymes, and no hybridization with wild-type Bintje DNA, indicating the presence of a single T-DNA insertion in the mutant (**Figure 1.2**).

In vitro plant growth and tuberization

In order to assess growth characteristics, we multiplied Bintje and *up* plants from single node explants taken from 4-5 week old plantlets and recorded data after 4 weeks of *in vitro* growth. Mutant *up* plants exhibited mean stem length (3.6 cm), and mean number of axillary branches (3.1), which were significantly different compared to the Bintje control, where the mean stem length was longer (7 cm) with fewer mean number of axillary branches (0.7). However, no

significant differences were found in number of roots per plant (**Figure 1.3**). *In vitro tuberization experiments were conducted to assess tuberization ability of the up mutant.* All recorded parameters (number of tubers/explant, number of stolons/explant and tuber weight/explant) were significantly greater for *up* (**Figure 1.4 and Figure 1.5A-B**) compared to wild-type Bintje. We also observed differences in tuber shape and size between Bintje and *up* plants. Bintje tubers were round and more uniform compared to the longer tubers observed with *up* (**Figure 1.5C-D**). As cytokinin is known to impact tuberization, kinetin effects on tuberization were assessed with both Bintje and *up*, to see if any difference in responsiveness existed in the mutant. *In vitro* tuberization using tuberization medium supplemented with 10 µM kinetin improved the number and weight of tubers for both Bintje and *up*, but the increase was somewhat greater for Bintje (**Figure 1.6**). The number of stolons was also increased for Bintje but not *up*.

Phenotyping under growth chamber and green house conditions

The *up* plants exhibited multiple phenotypes when grown *ex vitro* in soil, providing more information about mutant morphology and life cycle. New phenotypes were observed for *up* plants, in addition to the traits observed under *in vitro* conditions (**Figure 1.7**). Mutant plants were dwarf (mutant plant height was 42 cm compared to 73 cm in Bintje; **Table1.1**), had smaller sized leaves (**Figure 1.7A-B & 1.8**), produced long and fingerlike tubers (**Figure 1.7C-D & 1.8**), had small floral buds which did not open and dropped prematurely (**Figure 1.7E-G**), and exhibited early senescence of the leaves (**Figure 1.7H-I**). When freshly harvested tubers of the *up* plants were replanted in soil without any treatment to break tuber dormancy, 60% (15 out of 25) of tubers sprouted within 40 days **(Figure 1.7K)**, in contrast to the normally dormant tubers of Bintje which failed to sprout during the same time period (**Figure 1.7 J**). The detailed plant and tuber data (**Figure 1.8 & 1.9 and Table 1.1**) indicated that the *up* mutant exhibited significantly reduced plant size, smaller leaves and tubers and less number of tubers. Biomass and yield traits like aboveground fresh weight, total fresh weight, total dry biomass weight, total tuber weight per plant, number of size one (1) tubers, and marketable yield were also significantly lower in the mutant compared to wild-type Bintje **(Figure 1.9)**. Other tuber qualitative traits like tuber shape, tuber color, tuber eye depth etc. were also different between

the two genotypes **(Table 1.1)**. In addition, the mutant had more tubers that exhibited external defects (like green, misshapen, growth cracked and rotted) and lower specific gravity (**Table 1.1**).

Identification of potato genomic region flanking the T-DNA

We used the GenomeWalker (Clontech) technique to identify and amplify the genomic region flanking the T-DNA insertion. As the pSKI074 vector has few unique sequences between the T-DNA right border and the tetramerized enhancer region, we selected the left border region of pSKI074 for genome walking. Gene specific primers and nested gene specific primers were designed against the mannopine synthase gene sequences and used with arbitrary primer 1 (AP1) and arbitrary primer 2 (AP2) to amplify the potato genomic region flanking the left border of the T-DNA (**Table S1.1**). Primary and secondary PCR products from GenomeWalker were generated (**Figure S1.1**), the secondary products cloned and sequenced. The resultant genomic DNA sequence flanking the left border T-DNA region (**Figure S1.2**) was used to query the potato genomePGSC_DM_v3_superscaffolds database (http://solanaceae.plantbiology.msu.edu/pgsc_download.shtml). The location of the T-DNA insertion was found on superscaffold PGSC0003DMB0149: 763110 (**Figure 1.10**), more specifically on chromosome 10 position. Further PCR and sequence analysis of the left border T-DNA/plant DNA junction site revealed that a vector backbone of 250 bp (7802 to 8053 sequence of pSKI074) outside of the T-DNA was also found in the *up* mutant (data not shown). In a similar analysis of the right border T-DNA/plant DNA junction site in the *up* mutant, we found that 22 bp of the T-DNA right border (4273 to 4294 sequence of pSKI074) and the same number of base pairs from potato genome (PGSC0003DMB0149: 763111 to 763132) were lost during the T-DNA integration event (data not shown). Using the potato genome browser (http://solanaceae.plantbiology.msu.edu/cgi-bin/gbrowse/pgsc-potato-dm/), we identified several gene model predictions on both sides of the T-DNA insertion. The first potato gene open reading frame flanking the right and left borders of T-DNA were approx. 5.2 kb and 3 kb from right and left borders of T-DNA, respectively (**Figure 1.10**).

Transcript analysis of potato genes flanking the T-DNA

Transcript analysis of potato genes flanking the T-DNA was done by reverse transcriptase PCR, using primers designed against the reading frame of each gene model (**Table S1.1**) and with cDNA made from leaf RNA. We analyzed approximately 400 kb regions both upstream and downstream of the insertion site, and observed up- and down-regulation of potato genes flanking both right and left borders of the T-DNA (**Figures 1.11 and 1.12**). The most significant change in gene expression observed in the mutant was strong up-regulation of the first two predicted genes (conserved genes of unknown function), which started 5.2 kb and 11.5 kb from the right border (**Figure 1.11**). The other gene whose expression was up-regulated in the mutant was a SAM-dependent methyltranserase, which was located 41 kb downstream of the right border. Several genes within the 400 kb region off of the right border were down-regulated (F-box family protein, GA_2-oxidase 1, P21-rho-binding domain-containing protein, adenosine deaminase, Dicer homolog, DNA-3 methyladenine glycosylase) in the mutant. Only one gene on the left border of the T-DNA, a conserved gene of unknown function, was observed to be highly up-regulated in the *up* mutant, whereas a Protein phosphatase 2C and another conserved gene of unknown function were down-regulated in the mutant (**Figure 1.12**).

BTB/POZ domain-containing protein

The transcript analysis of the flanking genes on both side of the T-DNA revealed that insertion of the T-DNA had resulted in modification of the expression for many genes in the *up* mutant, not only the first (or few) genes immediately adjacent to the right border and associated tetramerized enhancers. In order to characterize transcriptome changes on a greater scale, we used RNA-sequencing (RNA-seq) of leaves to obtain an accurate global gene expression profile of the mutant. This analysis resulted in a unique dataset of 1,632 genes that were differentially expressed between *up* and Bintje, which will be described in a subsequent chapter. Here we will focus on the first two potato genes flanking the right border, which we observed to be highly up-regulated in the *up* mutant. In the potato genome browser, these genes (PGSC0003DMG401011239 and PGSC0003DMG402011239) are annotated as two separate genes (**Figure 1.10, Figure 1.11 - R1 and R2 genes, Figure 1.13B**), as conserved genes of

unknown function. Our previous RT-PCR results indicated that both genes were up-regulated to a similar degree (**Figure 1.11**), and RNA-seq reads covering these genes were much more abundant in the *up* mutant compared with Bintje, providing secondary verification of our PCR results (Figure 13A). In addition, RNA-seq transcript reads matched the predicted exon reads for these two genes, with the exception of the region around 771,200 (**Figure 1.13**).

When assessing the potato gene models, we noted that the two genes showed little physical separation, and that the tomato gene (Solyc10g005600.2.1) model prediction actually indicated these exons comprised a single gene (**Figure 1.13C**). Based on these observations, and the similar level of expression noted for these two genes in the *up* mutant, we hypothesized that the potato gene model may be wrong, and the tomato gene prediction correct. To test this, we used primers designed against the start and stop codons based on the tomato gene model, and carried out RT-PCR using leaf cDNA from the mutant. If the potato gene model was correct, no amplification products would be obtained. If the tomato gene model was correct, an amplification product of approximately 3 kb would result. Our PCR results generated a 3 kb fragment (**Figure 1.14**), indicating that the two predicted genes of the potato model actually comprised a single gene. To obtain the complete transcript sequence, 5' and 3'RACE were carried out, generating cDNA of approximately 3.7 kb, an open reading frame of 3 kb, encoding a predicted protein of 1000 amino acids. The closest plant ortholog for this gene was At1g04390, a BTB/POZ domain containing protein coding gene in Arabidopsis.

Discussion

To the best of our knowledge, this study is the first comprehensive description of an activation-tagged potato mutant, describing the isolation and characterization of a novel potato mutant, *up*, which was identified by its dwarf and axillary branching phenotype under *in vitro* conditions. Furthermore, *in vitro* and *in vivo* studies revealed additional *up* mutant phenotypes including loss of flowering, small floral buds, early bud drop, early senescence, reduced tuber dormancy and changes in tuber size, tuber shape, tuber number and number of eyes on the tubers. There are few reports that describe pleiotropic phenotypes in plants. Under-expression of a *Solanum chacoense* NOTCHLESS gene (*ScNLE*) encoding a WD-protein caused reduction in

aerial organ size, a reduction in some organ numbers, delayed flowering, and an increase in stomatal index (Chantha and Matton, 2007) and *sticky peel* (*pe*) mutant of tomato that revealed several phenotypes indicative of a defect in epidermal cell function and role of CUTIN DEFICIENT2 (CD2) and ANTHOCYANINLESS (ANL2) genes in regulating diverse metabolic pathways are a few examples. Our *in vivo* and *in vitro* growth data supported each other, with the exception of the number and weight of tubers and number of axillary branches. *In vitro up* plants produced more (3.1 shoots/plants in *up* vs. 0.7 shoots/plants in Bintje) axillary branches; however we observed less (1.9 shoots/plants in *up* vs. 4.7 shoots/plants in Bintje) axillary branches in the mutant *up* plants during *in vivo* plant growth compared to wild-type Bintje. This trait may explain, to some extent, the larger number of tubers in *up*, and the subsequent greater tuber weight per mutant. Under greenhouse and growth chamber experiments, *up* mutant plants exhibited earlier senescence, resulting in less time for tuber formation and bulking, resulting in lower tuber number and overall reduction in yield. Some observed differences in phenotypes between *in vitro* and *ex vitro* growth conditions could be due to environmental factors which are less controlled in soil grown plants as compared to *in vitro* plants.

In our study, reverse transcriptase PCR results showed both up- and down-regulation of genes on both sides of the T-DNA. Also the effect of our inserted tag was not limited to the immediate flanking gene on the right border, but expression of genes several hundred kb away was also modified, which is contrary to the expectation for activation tagged mutants. Although we included several flanking predicted gene models covering approximately 800 kb total genomic region (400 kb each on RB and LB flanking region), that included 23 gene models flanking the right border and 16 gene models flanking the left border, we left some genes out of RT-PCR analysis. Predicted gene models which were too small and/or genes with little or poor annotation; for example L6 to L10; were not included in the RT-PCR analysis. In most activation tagged mutant studies, overexpression of the gene immediately flanking the T-DNA accounts for the mutant phenotype. However, microarray analysis of an Arabidopsis activation tagged mutant, K3571, revealed RNA abundance of 1,273 genes greater than two-fold. Among 244 genes in a 984 kb region near right border, 201 genes were up-regulated and six were down-

regulated (Ahn et al., 2007). There are several explanations for the activation of multiple genes, including: 1) the scanning action of an activator bound to the 35S enhancer, which scans the genome and activates the genes in its path until it meets with an insulator (Bell and Felsenfeld, 1999), 2) chromatin changes and re-arrangements that may occur due to T-DNA insertion and 3) the effect of a differentially expressed nearby gene on downstream genes. For example, if the nearby gene modified by the activation tag were a transcription factor, its increased protein product would interact with other gene promoters and/or other proteins to cause pleiotropic effects. This could also explain the variation in gene expression observed between Bintje and the *up* mutant. Insertion of the T-DNA tag may have had an impact on the conformation of the DNA, affecting other DNA-binding protein interactions with adjacent DNA domains, some of which may have been tissue-specific.

The first two genes flanking the right border of the T-DNA (Figure 11, R1 and R1.1) were annotated as conserved gene of unknown function and gene models for these were less than 100 bp apart in the potato genome browser. Furthermore, our RT-PCR and RNA-seq data indicated that the increases in expression levels for these two genes in the *up* mutant were similar, suggesting that these two gene models could be part of a single gene. RNA-seq transcript data and exon regions of R1 and R1.1 genes in the potato browser gene model match well except in the 771,200 region in which no exons are shown in the potato gene model. However, we detected transcripts during RNA-seq analyses, which we confirmed by RT-PCR. Furthermore, we amplified transcripts flanking both gene models (results not shown) which indicate that R1 and R1.1 are part of same gene. Later a single full length transcript of 3 kb was amplified from potato, which confirmed this as a single gene. The RNA-seq transcript pattern matches the tomato gene model, which is a single gene of about 3 kb. The closest orthologs of this gene in Arabidopsis (At1g04390) and tomato (Solyc10g005600.2.1) encode a BTB/POZ domain-containing protein. BTB/POZ (Broad complex, Tramtrac, Bric a brac; also known as Pox virus and Zinc finger) domain-containing proteins play important roles in various biological processes by regulating the transcriptional activities of downstream genes (Liu et al., 2011). This protein domain mediates protein interactions, affecting global repression of transcription factors, plant defense, protein degradation and cell death which may explain the molecular

basis of the multiple phenotypic changes observed in the *up* mutant. A study consisting of chromatin immunoprecipitation sequencing (Chip-seq) analysis of BACH1 (BTB and CNC homology 1) target genes in Human Embryonic Kidney 293 (HEK 293) cells with knockdown of BACH1, followed by transcriptome profiling analysis, revealed the impact of *BACH1* repression on transcription and identified BACH1 target genes affecting cell cycle, apoptosis, subcellular transport and redox regulation (Warnatz et al., 2011). A novel member of the human BTB/POZ protein family, ZBTB1, with an N-terminal BTB/POZ domain and eight C-terminal zinc-finger motifs acts as a transcription repressor in CREB activation and the cAMP-mediated signal transduction pathway to mediate cellular functions (Liu et al., 2011). In *Helicoverpa armigera*, a BTB domain-containing protein is up-regulated by immune challenge indicating its role in insect innate immunity (Wang et al., 2011). Similarly, in Arabidopsis, two BTB proteins (POB1 and POB2), function redundantly as negative regulators of defense against fungal pathogen *Botrytis cinerea* and oomycete pathogen *Hyaloperonospora arabidopsis* (Qu et al., 2010). A small clade (5 members) of Arabidopsis BTB domain-containing proteins termed as BT1- BT5 (BTB and TAZ domain proteins) are multifunctional scaffolds that act in or interconnect multiple cellular pathways and perform important functions in Arabidopsis development (Robert et al., 2009). However, these BT proteins are short-lived and are themselves targets of ubiquitination which is in contrast to BTB domain-containing proteins that label other proteins for ubiquitination-dependent protein degradation. A member of this clade, BT2 mediates multiple responses to nutrients, stress and hormones in Arabidopsis. Constitutive expression of BT2 imparted resistance to both sugars and ABA at germination, and positively regulated certain auxins, suggesting that BT2 suppresses sugar and ABA responses which further explains the early sprouting in mutant tubers (Mandadi et al., 2009). Furthermore, MATH (Meprin and TRAF homology)/BTB domain protein acts as substrate adaptor for CUL3 (CULLIN family protein)-based ubiquitin E3 ligase and targets the homeodomain Leucine Zipper *ATHB6* gene to modulate ABA signaling (Lechner et al., 2011). We amplified the full length cDNA clone of the BTB/POZ domain containing protein which we hope to use to generate an over-expression construct to recapitulate the *up* phenotype by over-expressing in Bintje.

One interesting component of this research is that it has allowed us to start describing uncharacterized gene model predictions from the recently completed potato genome sequencing effort. In the present study we tagged a conserved gene of unknown function in the potato genome that was later found to be a close homolog (expect (E) value, 8×10^{-93}) of an Arabidopsis BTB/POZ domain-containing protein. The two predicted gene models in potato genome were in fact a single large gene that we validated with RNA-seq and RT-PCR data and cloned full length cDNA of that gene. These data are an example of how we can contribute to gene prediction validation. Although many questions are yet to be answered, our findings shed new light on the utility of activation tagging in potato gene function studies and suggest that forward genetic approaches may be useful in plant species with a wide variety of genome size as long as at least one highly transformable genotype can be identified. Mutants and gene tagging no longer need to be restricted to model plant species; activation tagging allows effective forward genetics in virtually any species of interest. Due to advantages of activation tagging over conventional insertional mutagenesis this may become a method of choice for potato gene discovery.

Methods

Generation and screening of potato activation tagged mutants

The wild-type Bintje and activation tagged mutant AT 615 (*up*) mutant plantlets were received from BioAtlantech (http://bioatlantech.nb.ca/site/solanumgenomics), having been generated previously through the Canadian Potato Genome Project (Regan et al., 2006). The mutant lines were generated as follows: Potato (*Solanum tuberosum* L. cv. Bintje) plants were grown on plantlet growth medium [PGM; MS (Murashige and Skoog, 1962) minimal organic medium, sucrose 3%, pH 5.7 and solidified with 0.7% agar, sterilized] from 1 cm nodal stem sections and kept under 16 h photoperiod, 20˚C day/18˚C night and 70-100 µE $m^{-2}s^{-1}$ light intensity. For transformation, internodal stem segments from 4-week-old plants were excised and placed on infiltration medium (IM; MS minimal organic medium, sucrose 3%, MES 0.5g/L, mannitol 20 g/L, pH 5.5, filter-sterilized) to prevent the explants from drying out.

The activation tagging vector pSKI074 (GenBank accession no. AF218466) was transformed into *Agrobacterium tumefaciens* strain GV3101 (pMP90RK) by electroporation. The T-DNA of vector pSKI074 encompasses left and right borders, kanamycin selectable marker, T3 RNA polymerase promoter, T7 RNA polymerase promoter and four CaMV 35S enhancer repeats (Weigel et al., 2000). The transformed colonies were selected by growing them on ampicillin-containing Luria broth (LB) plates. For potato transformation, *Agrobacterium* was grown from freshly streaked LB plates supplemented with 100 mg/L ampicillin, 50 mg/L kanamycin and 100 mg/L rifampicin and plates were incubated for 2 days at 28°C in the dark. A single colony of *Agrobacterium* was picked and cultured overnight in liquid LB medium containing 100 mg/L ampicillin, 50 mg/L kanamycin and 100mg/L rifampicin, at 28°C in the dark. *Agrobacterium* inoculum was grown to an OD_{600} of 0.6-0.8 . Bacterial cells were pelleted by centrifugation at 3,000 rpm for 15 min, re-suspended in IM with 200 µM acetosyringone at a final concentration corresponding to an OD_{600} of 0.3.

Excised internodal explants (placed on IM medium) were transferred to a 50 ml tube and IM medium containing the re-suspended *Agrobacterium* and acetosyringone was added. Tubes were inverted gently for 2 to 5 min. For co-cultivation, explants were put on sterile cheesecloth or filter paper to remove excess *Agrobacterium* before transfer to petri plates containing 25 ml of callus growth medium [CGM; MS minimal organic medium, glucose 2%, adenine sulfate 40 mg/L, MES 0.5g/L, polyvinylpyrrolidone (PVP) 0.5 g/L, glutamine 200 mg/L, pH 5.7 and solidified with 0.22% gelrite, sterilized and supplemented with 0.1 mg/L trans-zeatin and 0.1 mg/L IAA] for 2 days. Petri plates were sealed with micropore tape and incubated at 24°C with 16 h photoperiod under low-light (60-80 µE $m^{-2}sec^{-1}$) in a plant growth chamber.

After 2 days of co-cultivation, explants were washed 2-3 times with sterile distilled water containing 300 µg/ml claforan, blotted on sterile filter paper and placed on callus selection medium (CSM = CGM + kanamycin 100 µg/ml and claforan 300 µg/ml). Petri plates were sealed with micropore tape and incubated at 22-24°C, 16 h photoperiod under low-light in a plant growth chamber. After 10 days, explants were transferred to fresh CSM and higher light intensity (150-200 µE $m^{-2}sec^{-1}$) and incubated for 10 days. A second transfer was done onto

fresh CSM and explants incubated for 10 days for callus induction (total 30 days on CSM) before the explants were transferred to shoot growth medium. Calli were transferred to shoot growth medium (SGM; MS minimal organic medium, glucose 2%, adenine sulfate 40 mg/L, MES 0.5 g/L, polyvinylpyrrolidone(PVP) 0.5 g/L, glutamine 200 mg/L, pH 5.7 and solidified with 0.22% gelrite, sterilized and supplemented with 0.1 mg/L trans-zeatin, 50 mg/L kanamycin and 300 mg/L claforan) and incubated for 6-8 weeks, transferring to fresh SGM every 7-10 days to facilitate shoot elongation. Shoots were transferred to root growth medium (RGM; MS minimal organic medium, sucrose 2%, pH set to 5.7 and solidified with 0.7% agar, sterilized and supplemented with 50 mg/L kanamycin and 300 mg/L claforan) or to PGM when 2 cm long with have several leaves.

Transformed bacterial colonies were screened for the presence of the plasmid before infection of internodal explants and also after 2 days of co-cultivation on medium by colony PCR. Bacteria were transferred to fresh LB plates with antibiotics (ampicillin, kanamycin and rifampicin) and incubated for 2 days at 28°C in dark. Colony PCR with 074 Red and Fred primers (**Table S1.1**) was done to check plasmid sequences. For screening of transformed plantlets, total DNA was extracted from leaves of putative transgenic plants and PCR was done with the same set of primers (074 Red and Fred) to confirm the presence of plasmid.

Identification of potato genomic regions flanking the T-DNA

Genomic DNA was extracted using Plant DNAzol Reagent (Life Technologies; Grand Island, NY, USA) following manufacturer's instructions. Regions flanking the T-DNA were amplified using GenomeWalker universal kit (Mountain View; CA, USA. Cat. 638904) following manufacturer's instructions. Briefly, genomic DNA from *up* mutant plants was completely digested with a set of restriction enzymes *MscI, NaeI, ScaI* and *SspI* that generate blunt ends in separate tubes, and then ligated with GenomeWalker adapters to create "GenomeWalker libraries". The primary PCR amplification was done with outer adapter primer (AP1), gene specific primer (GSP1) and GenomeWalker library DNA as templates which were followed by secondary PCR amplification using nested adapter primer (AP2), nested gene specific primer (GSP2) [**Table S1.1**] and diluted primary PCR product as template. Secondary PCR amplification products were separated by

electrophoresis and purified using a gel extraction kit (Qiagen; Valencia, CA, USA); cloned and sequenced. Sequencing was performed in the ISRR lab (IALR, Danville, VA) using a CEQ 8800 genetic analysis system (Beckman Coulter, Brea, CA, USA). Recovered genomic sequences flanking the left border of the T-DNA insertion sites were identified and used to perform Basic Local Alignment Search Tool (BLAST) searches in the GenBank database (accessed on 4-5-2010) of the National Center for Biotechnology Information (NCBI), and in the potato genome (http://potatogenomics.plantbiology.msu.edu/blast.html) with default search parameters (expected threshold 10, maximum number of alignments 100, max number of descriptions 100, word length 11, no filter and both strands). This allowed the T-DNA insertion site to be located on superscaffold PGSC003DMB000000149 within the potato genome. Approximately 400 kb of potato sequence upstream and downstream of the insertion site were analyzed for the presence of annotated gene models using the potato genome browser (http://solanaceae.plantbiology.msu.edu/cgi-bin/gbrowse/pgsc-potato-dm/). Sequence homology searches and analyses were performed using the NCBI BLAST server. Sequence alignments were carried out by the CLUSTALW2 (Larkin et al., 2007) method and using the EMBL server (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

Determination of T-DNA copy number by Southern blot analysis

Genomic DNA of cv. Bintje and *up* was digested with *Pci*I and *Psi*I restriction enzymes and separated on a 0.8% agarose gel. Digested DNA was transferred to positive charged Zeta probe nylon membrane (Bio-Rad; Hercules, CA, USA) using downward capillary transfer setup and NaOH buffer (Sambrook and Russell, 2001). DNA was transferred to the membrane and rinsed in 2X SSC, then placed on a sheet of Whatman 3 MM filter paper and allowed to air dry for 30 min. The probe (640 bp) used for hybridization was amplified from 4X35S enhancer region of pSKI074 plasmid by PCR using primers (**Table S1.1**) and labeled using Amersham AlkPhos Direct Labeling Reagents (GE Healthcare Life Sciences; Buckinghamshire, U.K. Catalog No. RPN3682) following manufacturer's protocol. Overnight hybridization with the probe was done in Alkphos Direct hybridization buffer (Amersham GE) at 55°C. After hybridization, the blot was washed to remove excess probe and signal detected with CDP-Star (GE Healthcare Life Sciences; Buckinghamshire, U.K.) chemiluminescent detection reagent. For final signal

generation, the blot was exposed to autoradiography film and the film was developed with Kodak GBX developer and fixer solution.

***In vitro* plant growth and tuberization**

For *in vitro* plant growth assays, Bintje and *up* plants were grown from single node explants taken from 4-5 week old plants. Explants were placed on MS basal medium with vitamins (Phytotech, Lenexa, KS, USA) supplemented with 3% sucrose, pH 5.8 and solidified with 0.7% agar. Plants were grown under 16 h photoperiod, 24±1°C and 70-100 µE light intensity in a plant growth chamber and data were recorded after 4 weeks. For *in vitro* tuberization assays, stem sections, 3-5 cm long, with 4-5 nodes per section were prepared from 4-5 weeks old *in vitro* 'Bintje' and *up* plants by removing all leaves. These stem sections were grown in 40 ml liquid propagation medium [MS basal medium + 148 mg l^{-1} NaH_2PO_4 + 0.4 mg l^{-1} thiamine HCl + 100 mg l^{-1} inositol + 3% sucrose, pH 5.9] (Radouani, 2003) at 16 h photoperiod, 24±1°C and 70-100 µE for 3-4 weeks in 250 ml conical flasks capped with ventilated plugs. After 3-4 weeks, propagation medium was replaced with microtuberization medium [MS basal + 8% sucrose, 2X the amounts of NH_4NO_3, KH_2PO_4 and KNO_3 and without NaH_2PO_4, pH 5.8] (Radouani, 2003). Culture flasks were kept at 18-20°C in the dark to induce tubers and data were recorded after 30 days. All *in vitro* experiments were done in randomized block design with three replicates and each experiment was repeated at least three times. Data were analyzed in JMP, version 9.0.0 (SAS Institute Inc., Cary, NC) and tables and graphs were drawn in Microsoft Excel (2010).

Phenotyping under growth chamber and greenhouse conditions

Phenotyping was done on plants grown in walk-in growth chambers in the Department of Forest Resources and Environmental Conservation, Virginia Tech. These chambers are capable of mimicking the outdoor field growth conditions. Ten equal size (by weight) tubers of Bintje (average tuber weight 24 g) and *up* (average tuber weight 22 g) were planted in 30 cm diam pots filled with commercial Miracle-Gro potting mix (Scotts Miracle-Gro, Marysville, OH). Standard plant growth practices were followed. Plants were grown under 16 h photoperiod; 22°C day/16°C night and 350 µE light intensity and data were recorded 75 days after tuber

planting (45 days after plant emergence). Growing plants were supported by a plastic net and manually watered twice per week or as needed.

The *up* mutant was also compared to 'Bintje' under greenhouse conditions. 'Bintje' and *up* plants were grown in the greenhouse of the Institute for Sustainable and Renewable Resources (ISRR) at the Institute for Advanced Learning and Research (IALR), Danville, Virginia, USA. Twenty tubers each of wild-type and mutant (after equalizing the weight) were planted directly into 30 cm diam. pots filled with commercial Miracle-Gro potting mix. Standard plant growth practices were followed. The greenhouse was set to 24-30°C. Data were recorded 75 days after planting (45 days after emergence) and analyzed in JMP.

Reverse transcriptase PCR analysis

Leaf samples for RNA extraction were harvested from plants in the walk-in-growth chamber at 60 days after planting. Leaves at the same developmental stage (not fully mature) were selected. Ten leaves were harvested to make one composite sample (one leaf per plant) for each genotype and three replicates of composite samples per genotype were collected. Leaves were immediately put in liquid nitrogen and then stored at -80°C until needed. RNA was extracted using either Concert Plant RNA Reagent (Invitrogen, Carlsbad, CA) following manufacturer's instructions or a combination of Trizol Reagent (Invitrogen, Carlsbad, CA, USA) extraction and RNA mini kit (Qiagen, Valencia, CA, USA) purification. For extraction with Trizol Reagent, frozen tissues were ground with a mortar and pestle to a fine powder in liquid nitrogen and 0.2 g - 0.5 g powder per 1 ml of prewarmed (35-40°C) Trizol was used for extraction. Powder was mixed with Trizol immediately by vortexing to protect it from RNases. This mixture was incubated for 5 min at room temperature (RT), frequently mixed by vortex followed by centrifugation at 12,000 g for 10 min at 2-8°C. After removal of supernatant to a new tube, 0.2 ml chloroform was added and mixed by vortex for 15 sec. Tubes were incubated for 1 min at RT, mixed by vortex for 15 sec and incubated at RT for 2-3 min, followed by centrifugation at 12,000 g for 15 min at 2-8°C to separate the phases. The upper layer was removed and divided into two parts of 200 µl each (one part for backup was stored at -20°C). To another 200 µl from the top layer, 700 µl of Qiagen RLT buffer (after addition of 1 µl β-

mercaptoethanol to 1 ml Buffer RLT) and 500 µl of 96-100% ethanol were added and mixed well by vortexing. Samples were applied to Qiagen MinElute spin column and washed with RPE buffer, followed by two washes with 750 µl of 80% ethanol (Sigma) to remove all guanidine salts that may inhibit downstream applications. RNA was eluted with 20 µl of RNAase free water.

DNAse treatment of RNA was done with DNA-free kit (Ambion; Foster City, CA, USA) following the manufacturer's instructions. First strand cDNA was synthesized using SuperScript III (Invitrogen; Carlsbad, CA, USA) using 1 µg of total RNA. After cDNA synthesis the final volume was adjusted to 100 µl of diluted cDNA for each cDNA synthesis reaction. Reverse transcription polymerase chain reaction (RT-PCR) was performed using gene-specific primers (**Table S1.2**). Equal amounts of cDNA were used in each reaction and *Solanum tuberosum* Elongation factor 1-α (StEF1-α) gene expression was used as control. Gel images were acquired by Alpha Innotech (now part of Cell Biosciences) gel doc system (San Leandro, CA, USA).

mRNA-sequencing

Leaves were harvested 60 days after planting tubers (about 30 days after emergence). Single lateral leaves, 5-6 cm long and 2-3 cm broad, which were at the same developmental stage (not fully mature) were used for RNA extraction. Ten leaves, harvested from ten plants were pooled to form a composite sample for each genotype. Three composite samples (3 x 10 = 30 leaves) were harvested in total, frozen in liquid nitrogen and stored at -80°C until needed. RNA was prepared by grinding tissue in liquid nitrogen and extracting the RNA using Trizol Reagent following manufacturer's instructions followed by purification with RNA mini kit. For DNase treatment we used Turbo DNA free protocol following manufacturer's instruction with some modifications. The Illumina TruSeq RNA sample preparation kit (low-throughput protocol) was used for poly-A based mRNA enrichment and thermal mRNA fragmentation followed by cDNA synthesis using 4 µg of total RNA sample as starting material for each enrichment reaction. Sequencing (mRNA-seq) was performed at the Research Technology Support Facility (RTSF) of Michigan State University using the Illumina Genome Analyzer (GAII) and HiSeq 2000 system. Sequence reads were mapped to *S. tuberosum* Group Phureja DM1-3 516-R44 (DM)

superscaffolds using Tophat v2.0.4 (Trapnell et al., 2009), which made use of Bowtie. Reads were counted in regions of genes based on the PGSC_DM_v3_2.1.10_pseudomolecule annotation. Transcript abundance was expressed as raw read counts and also in a normalized unit, reads per kilobase per million reads (RPKM) as implemented in Cufflinks v2.0.2 (Trapnell et al., 2010), which allow comparison both within and among samples.

Acknowledgments

We thank Jared Carter for helping in growth chamber experiments and Allan Dickerman for assistance with RNA-seq analysis. This work was funded through Special Grants (2003-38891-02112, 2008-38891-19353 and 2009-38891-20092) from the United States Department of Agriculture, and operating funds from the Commonwealth of Virginia to the Institute for Advanced Learning and Research.

Figure 1.1 *In vitro* phenotypes of wild type 'Bintje' (left) and the mutant AT615, *underperformer* (right).

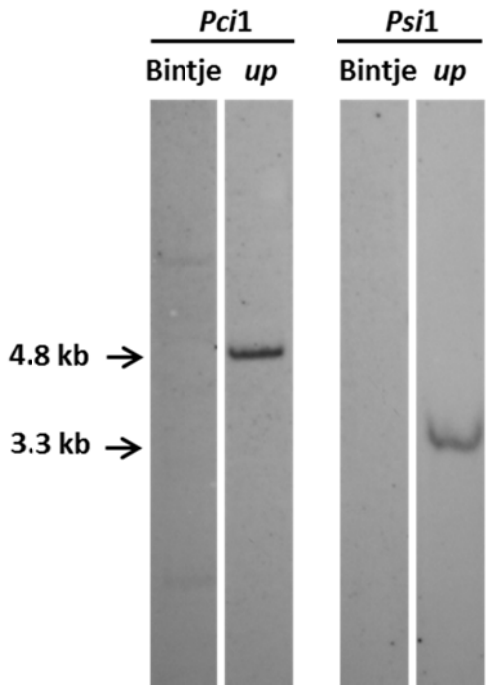


Figure 1.2 Determination of T-DNA copy number by Southern blot analysis. Fifteen µg genomic DNA of 'Bintje' (WT) and mutant *underperformer* were digested with restriction enzymes (*Pci*I or *Psi*I) that cut within the T-DNA, and probed with an AlkPhos direct labeled 640 bp probe, and the signal detected with the CDP-Star detection kit.

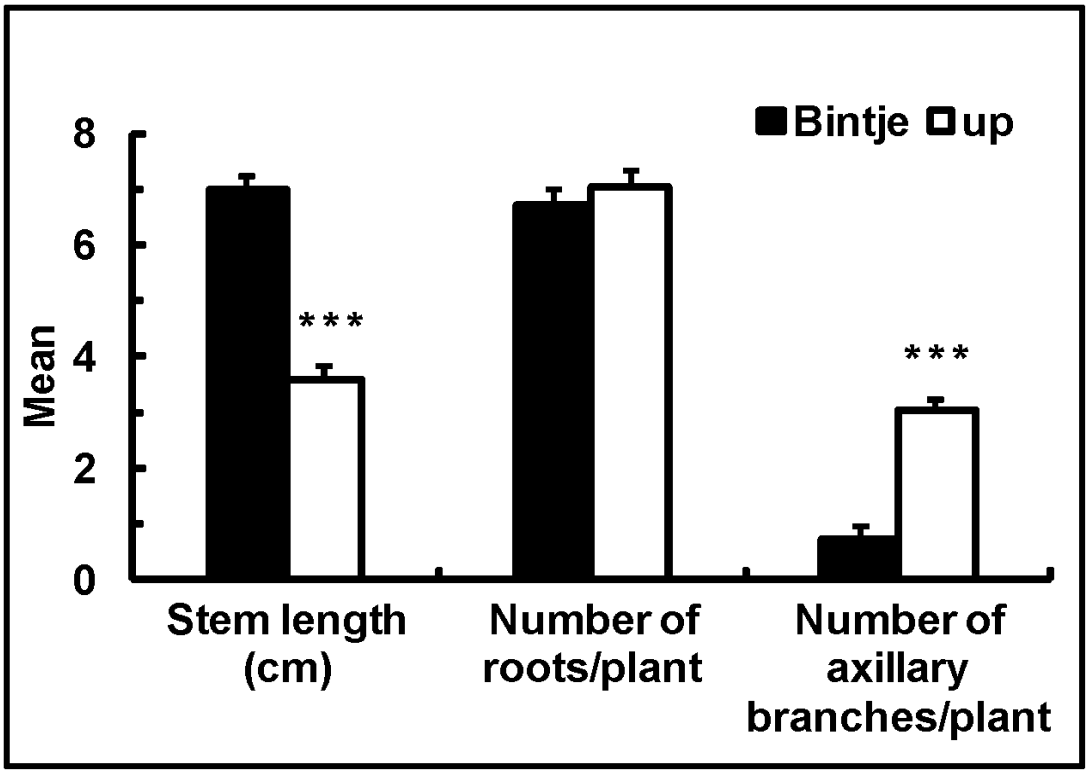


Figure 1.3 *In vitro* plant growth. Single node explants from 4-5 week-old potato plantlets of 'Bintje' and the *up* mutant were placed on MS medium with vitamins and data were recorded after 30 days. The values are means ± SE (sample size, n = 30). The asterisks indicate statistical significance of means in the same parameter estimated using Tukey's HSD test (* = P < 0.05-0.005, ** = P < 0.005- 0.0005, *** = P < 0.0005).

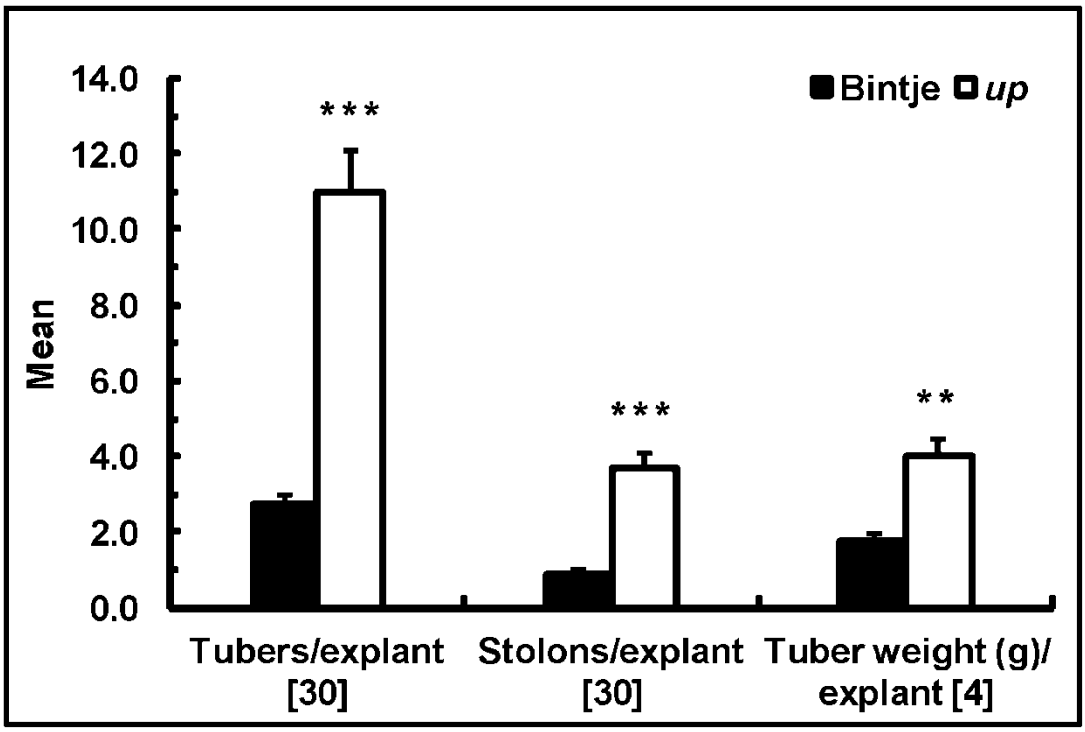


Figure 1.4 *In vitro* tuberization. Explants 3-4 cm long with 4-5 nodes were cultured in liquid propagation medium for 3-4 weeks under 16 h photoperiod and then transferred to high sucrose liquid tuberization medium and plants were transferred to the dark before data were recorded after an additional 30 days. The values are means ± SE (sample size, n = [in the parentheses]). The asterisks indicate statistical significance of means in the same parameter estimated using Tukey's HSD test (* = P < 0.05-0.005, ** = P < 0.005- 0.0005, *** = P < 0.0005).

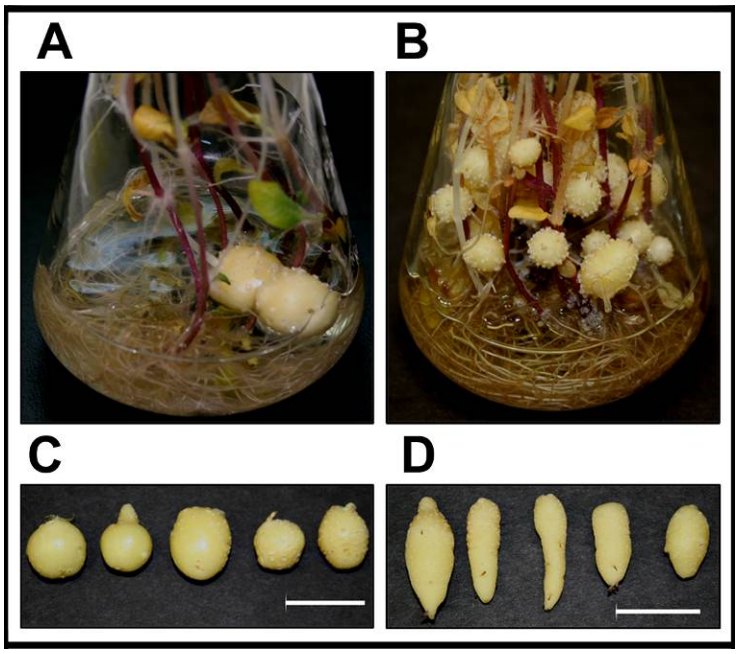


Figure 1.5 *In vitro* plant growth and tuberization. A (Bintje) and B (*up*) tuberized plants in liquid tuberization after 30 days and the harvested tubers of Bintje (C) and *underperformer* (D) showing different shape and size. Stem sections of 3-5 cm in length were grown in 40 ml liquid propagation medium at 24±1°C, a 16h photoperiod and 70-100 µE $m^{-2}sec^{-1}$ light intensity for 3-4 weeks, followed by one month of plant culture in microtuberization medium at 18-20°C in the dark to induce tubers. Data were collected after 30 days of microtuberization. Bar = 2 cm.

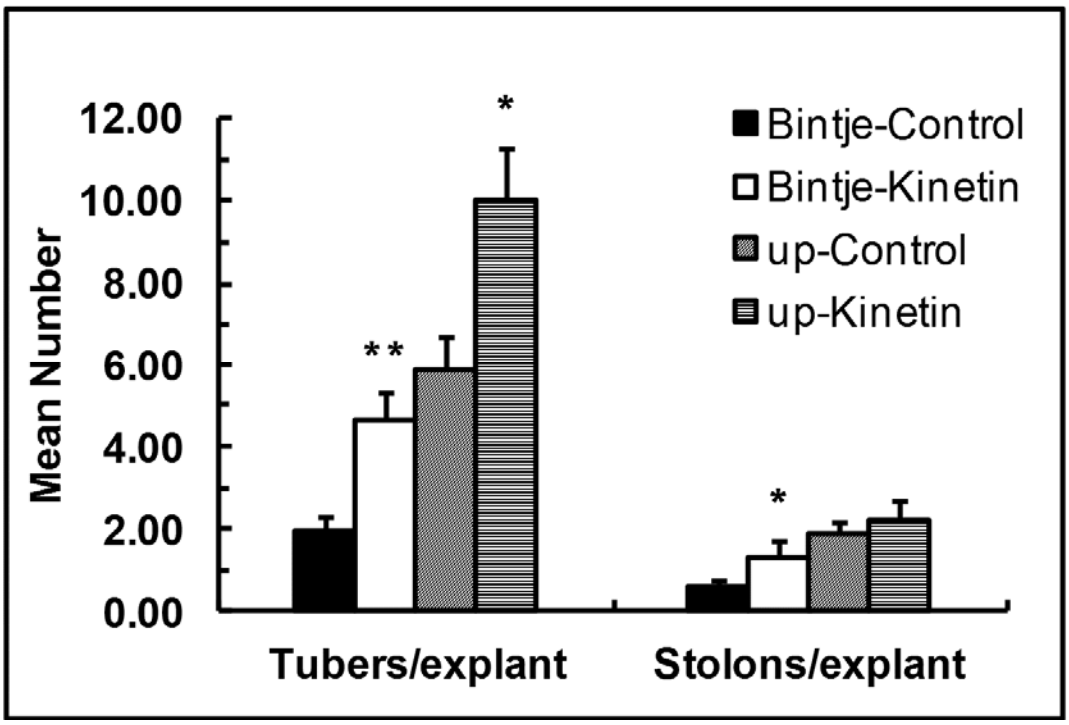


Figure 1.6 *In vitro* tuberization using kinetin. Explants 3-4 cm long with 4-5 nodes were cultured in liquid propagation medium for 3-4 weeks in light and then transferred to high sucrose liquid tuberization medium supplemented with 10 µM kinetin and plants were grown for an additional 30 days. The values are means ± SE (sample size, n = 15). The asterisks indicate statistical significance of means in the same parameter estimated using Tukey's HSD test (* = P < 0.05-0.005, ** = P < 0.005- 0.0005, *** = P < 0.0005).

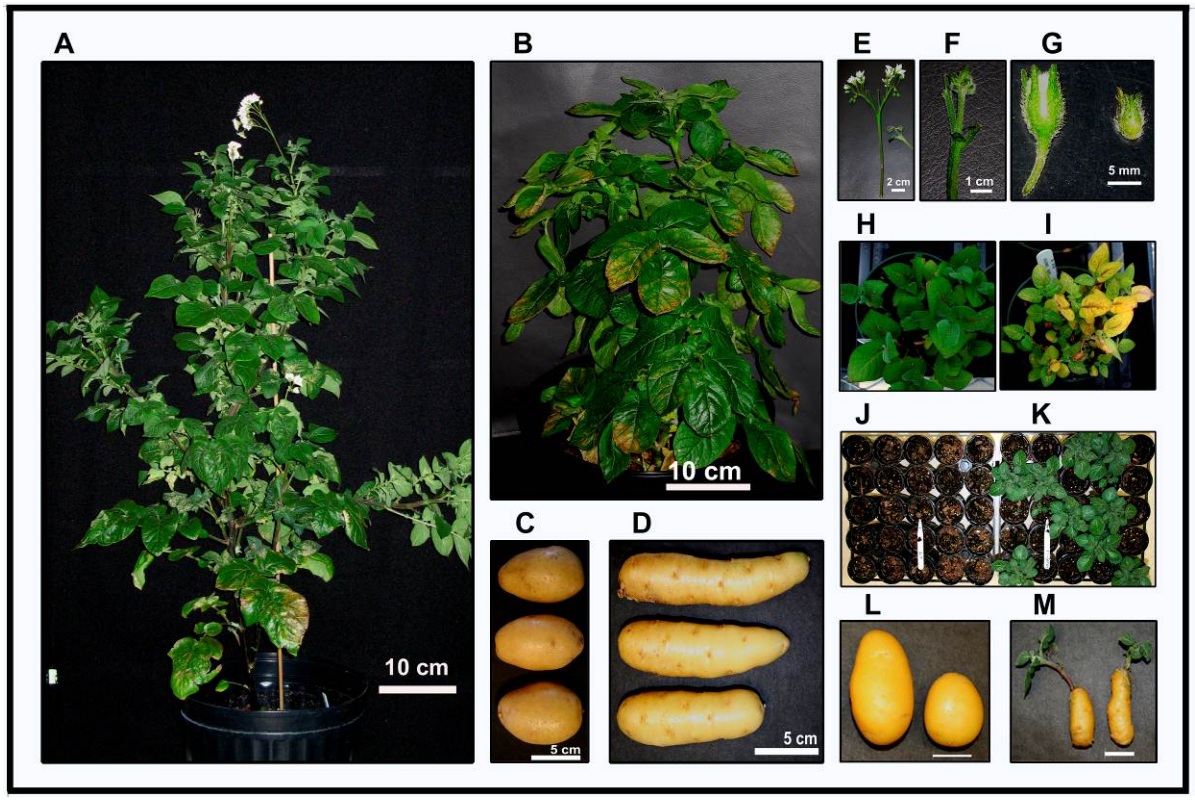


Figure 1.7 Phenotyping of 'Bintje' and *underperformer* (*up*) under growth chamber and greenhouse conditions. A (Bintje) and B (*up*) plants growing in a walk-in-chamber, 80 days after planting. C (Bintje) and D (*up*) tubers after harvesting from soil. E (Bintje) and F (*up*), inflorescence. G, flower buds, left Bintje, right *up*. Senescence in *up* plants after 35 days of planting (I), H (Bintje). Reduced dormancy phenotype, Bintje (J) and *up* (K) plants 30 days after tubers were planted without cold storage. Reduced tuber dormancy at the time of freshly harvested tubers of *up* (M), Binte (L). Bar = 2 cm.

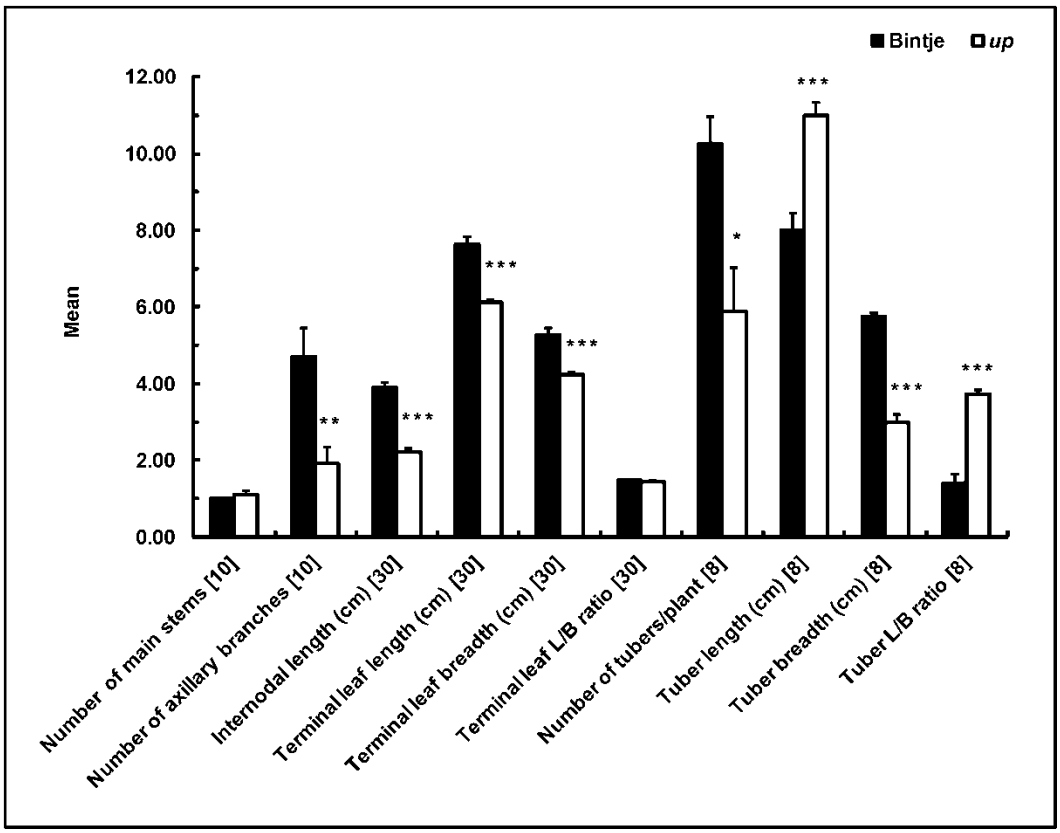


Figure 1.8 Bintje and *up* plant growth characteristics of the walk-in-growth chamber grown plants. Plants were grown from tubers directly planted in soil after breaking the tuber dormancy (cold storage for 45 days). Plant data were recorderd, 75 days after planting (DAP). Tuber data were recorded at final harvesting (130 DAP). The values are means ± SE (sample size, n = [in the parentheses]). The asterisks indicate statistical significance of means in the same parameter estimated using Tukey's HSD test (* = P < 0.05-0.005, ** = P < 0.005- 0.0005, *** = P < 0.0005).

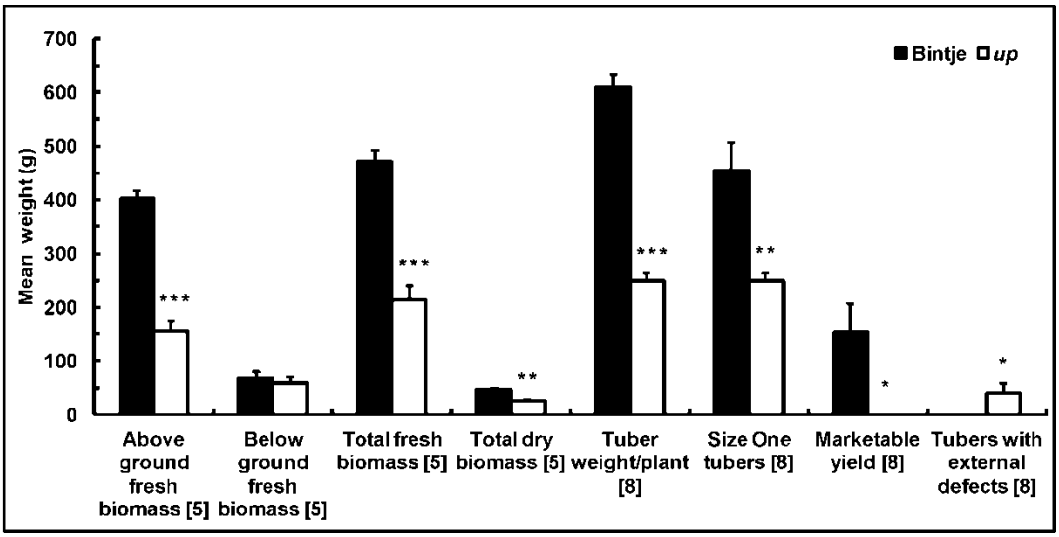


Figure 1.9 Bintje and *up* plant yield traits in the walk-in-growth chamber. Plants were grown from tubers directly planted in soil after breaking the tuber dormancy (cold storage for 45 days). Biomass data were recorderd, 75 days after planting (DAP). Tuber data were recorded at final harvesting (130 DAP). The values are means ± SE (sample size, n = [in the parentheses]). The asterisks indicate statistical significance of means in the same parameter estimated using Tukey's HSD test (* = P < 0.05-0.005, ** = P < 0.005- 0.0005, *** = P < 0.0005). The Standardized NE1014 rating codes for plant and tuber characteristics are given in Appendix 1.1.

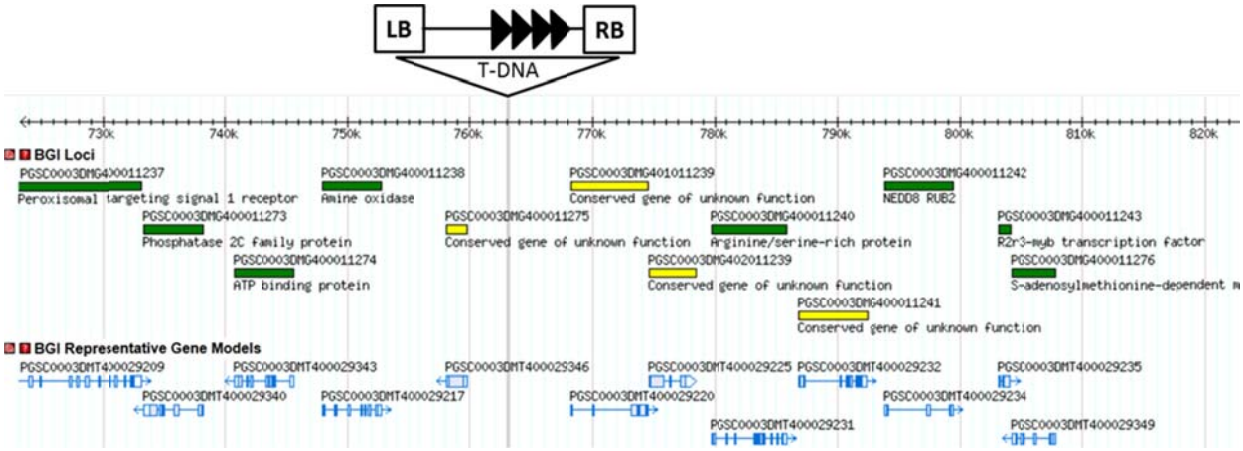


Figure 1.10 Identification of potato genomic regions flanking the T-DNA. A screen shot of a 100 kb region scaffold PGSC0003DMB000000149 of the potato DM1-3 516R44 PGSC Genome Assembly, showing the T-DNA insertion site and flanking gene prediction models.

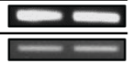
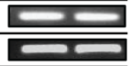
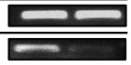
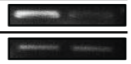
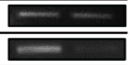
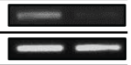
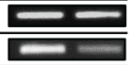
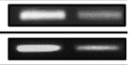
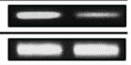
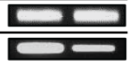
Gene serial number	Gene identifier PGSC003DMG	Gene annotation based on potato genome	Direction of the transcript	Gene model distance from RB (kb)	RT-PCR
					Bintje *up*
	Control	*S.tuberosum* elongation factor 1-α (EF1-α)			
R1	401011239	Conserved gene of unknown function	⇒	5.2	
R1.1	402011239	Conserved gene of unknown function	⇒	11.5	
R2	400011240	Arginine serine-rich protein	⇒	16.6	
R3	400011241	Conserved gene of unknown function	⇒	24	
R4	400011242	NEDD8 RUB2	⇒	31	
R6	400011276	S-adenosyl methionine dependent methyltransferase	⇐	41	
R8	400011245	Dead-box ATP dependent RNA helicase	⇒	70	
R9	400011246	Glyceraldehyde 3 phosphate dehydrogenase	⇒	78	
R10	400011247	Cation diffusion facilitator 9	⇒	85	
R11	400011248	Sno-protein	⇒	89	
R11.1	400011249	F-box family protein	⇒	131	
R12	400011277	Conserved gene of unknown function	⇐	135	
R13	400011250	MYB transcription factor	⇒	143	
R14	400011278	Equlibrative nucleoside transporter	⇐	151	
R18	400011280	Pyruvate orthophosphate dikinase regulatory protein	⇐	261	
R19	400011254	Gibberellin 2-oxidase 1	⇒	271	
R20	400011281	Xyloglucan endotransglucosylase/hydrolase protein A	⇐	279	
R21	400011255	P21-rho-binding domain-containing protein	⇒	281	
R22	400011256	Homeobox-leucine zipper protein	⇒	291	
R24	400011257	Adenosine deaminase	⇒	320	
R28	400011259	Endoribonuclease dicer homolog	⇒	338	
R29	400011260	Conserved gene of unknown function	⇒	380	
R31	400011287	DNA-3 methyladenine glycosylase	⇐	409	

Figure 1.11 Reverse transciptase PCR analysis of various potato genes flanking the right border of the activation tag T-DNA insertion of the *up* mutant. RNA was extracted from leaves of 60 days old plants grown in walk-in-growth chamber. In all gel images mutant transcript is on right side and wild-type Bintje on the left. The direction of the arrow indicates the orientation of the predicted open reading frame. Primer information used in RT-PCR is given in supplementary Table S1.

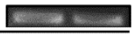
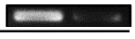
Gene serial number	Gene identifier PGSC003DMG	Gene annotation based on potato genome	Direction of the transcript	Gene model distance from LB (kb)	RT-PCR
					Bintje *up*
StEF1-α	Control	*S.tuberosum* Elongation Factor 1-α			
L1	400011275	Conserved gene of unknown function	⇐	3.0	
L2	400011238	Amine oxidase	⇒	10.3	
L3	400011274	ATP binding protein	⇐	17.5	
L4	400011273	Phosphatase 2C family protein	⇐	25	
L5	400011237	Peroxisomal targeting signal 1 receptor	⇒	30	
L11	400011270	Conserved gene of unknown function	⇐	98	
L16	400008687	Conserved gene of unknown function	⇒	150	No product
L17	400008686	Serine/threonine protein kinase	⇒	154	
L18	400008685	Neighbor of COX4	⇒	162	
L19	400008691	Conserved gene of unknown function	⇐	168	
L20	400008684	Pre-mRNA - splicing factor cwc15	⇒	172	
L23	400014404	Conserved gene of unknown function	⇒	206	
L25	400014370	Heat shock cognate protein 80	⇒	224	
L27	400014372	Uridine kinase	⇐	238	No product
L29	Not found	TFIIH basal transcription factor complex subunit		257	
L41	Not found	EDM2, Transcription factor		400	

Figure 1.12 Reverse transciptase PCR analysis of various potato genes flanking the left border of T-DNA. RNA was extracted from leaves of 60 days old plants grown in walk-in-growth chamber. In all gel images mutant transcript is on right side. The direction of the arrow indicates the orientation of the predicted open reading frame. Primer information used in RT-PCR is given in supplementary Table S1.

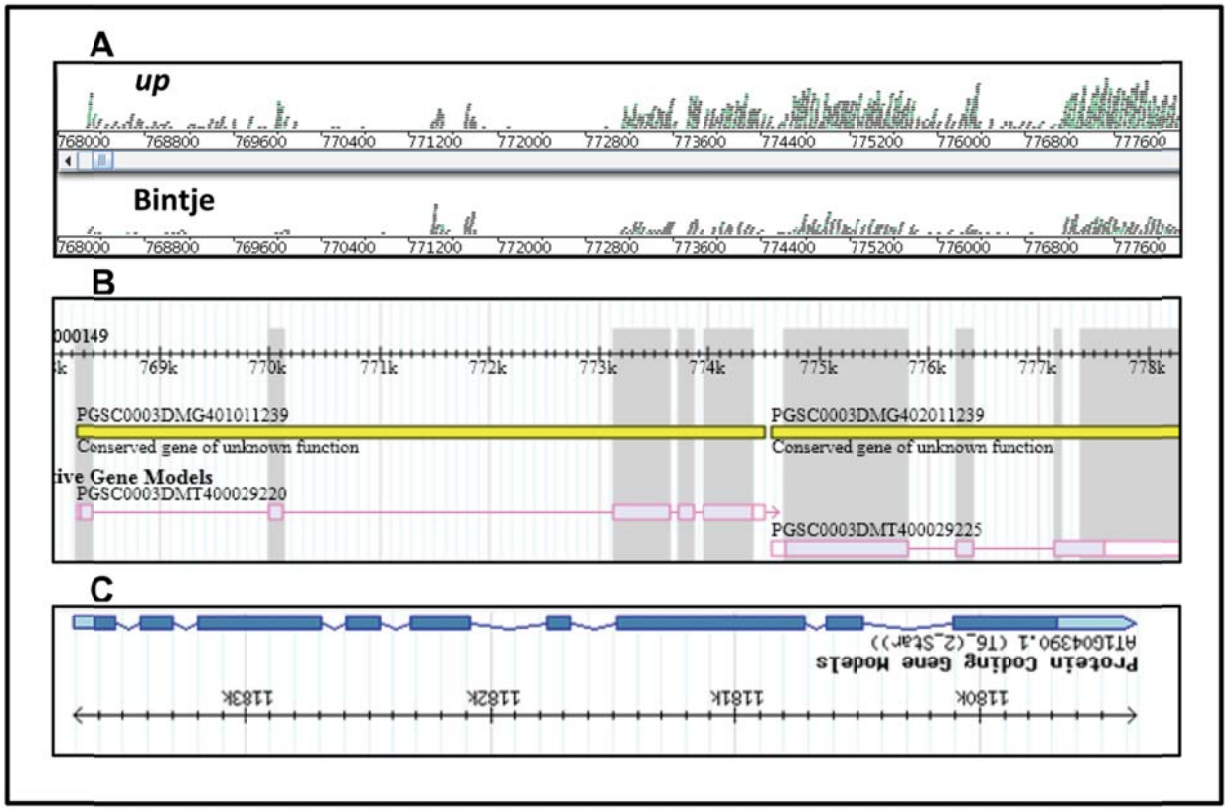


Figure 1.13 Structure of first potato gene flanking the right border of T-DNA. Panel A, Bamview visualization of Illumina transcript reads showing differential expression between mutant (upper reads) and Bintje (lower reads), along PGSC_DM_v3_2.1.10 genome chromosome 10, covering the annotated genes PGSC0003DMG401011239 and PGSC0003DMG402011239. Panel B, Corresponding region from potato genome showing two gene models. Grey highlighted area corresponds to cDNA transcripts sequenced during full-length Race PCR. Panel C, Gene model of Tomato ortholog.

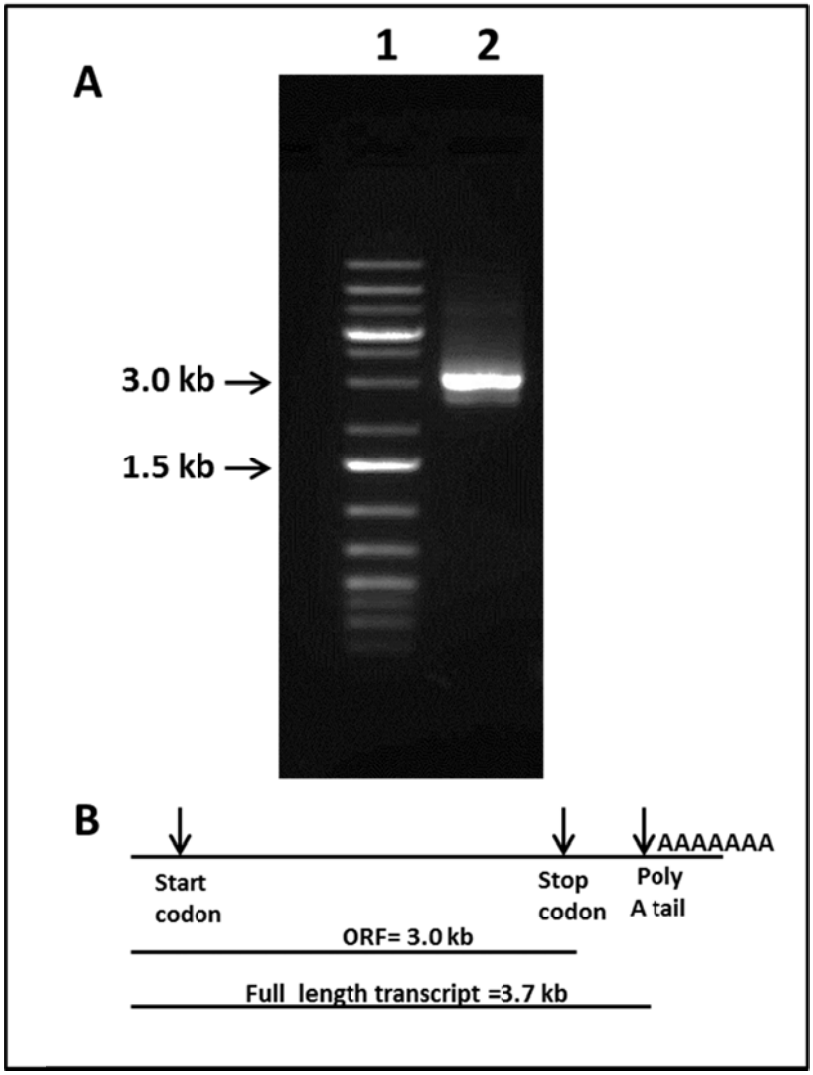


Figure 1.14 Confirmation that the two predicted potato gene models are actually a single gene, consistent with the proposed Tomato ortholog gene model. PCR primers (Table S2) were designed against the predicted start and stop codons of the single gene model, and yielded a single PCR product of the predicted size.

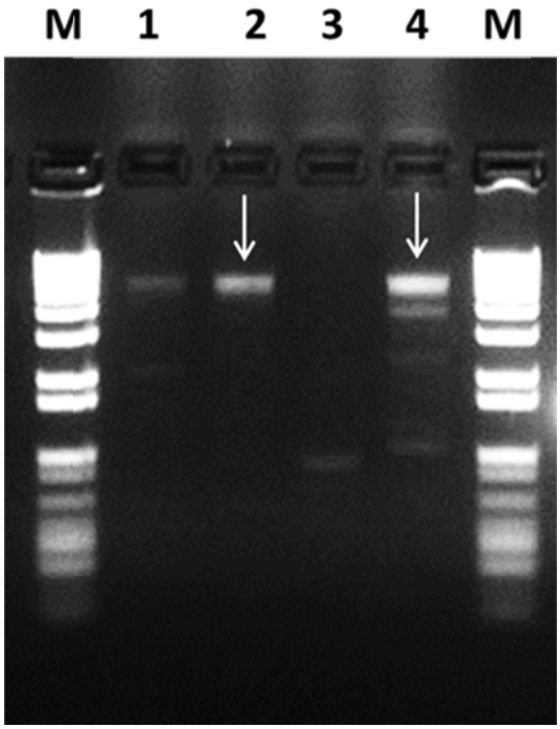


Figure S1.1 GenomeWalker secondary PCR products. Gene specific primers (GSP1) and nested gene specific primers (GSP2) were designed against the mannopine synthase gene sequences and used with arbitrary primer 1 (AP1) and arbitrary primer 2 (AP2) to amplify potato genomic regions flanking the left border of the T-DNA. Secondary PCR products from GenomeWalker are shown here. Lane 1, *Ssp*I digestion; lane 2, *Sca*I digestion; lane 3, *Nae*I digestion; and lane 4, *Msc*I digested mutant *underperformer* genomic DNA GenomeWalker libraries. (M= Invitrogen 1 kb+ DNA mass ladder). Sequences of primers used in genome walker are given in Table S1.1. Arrow = PCR product sequenced.

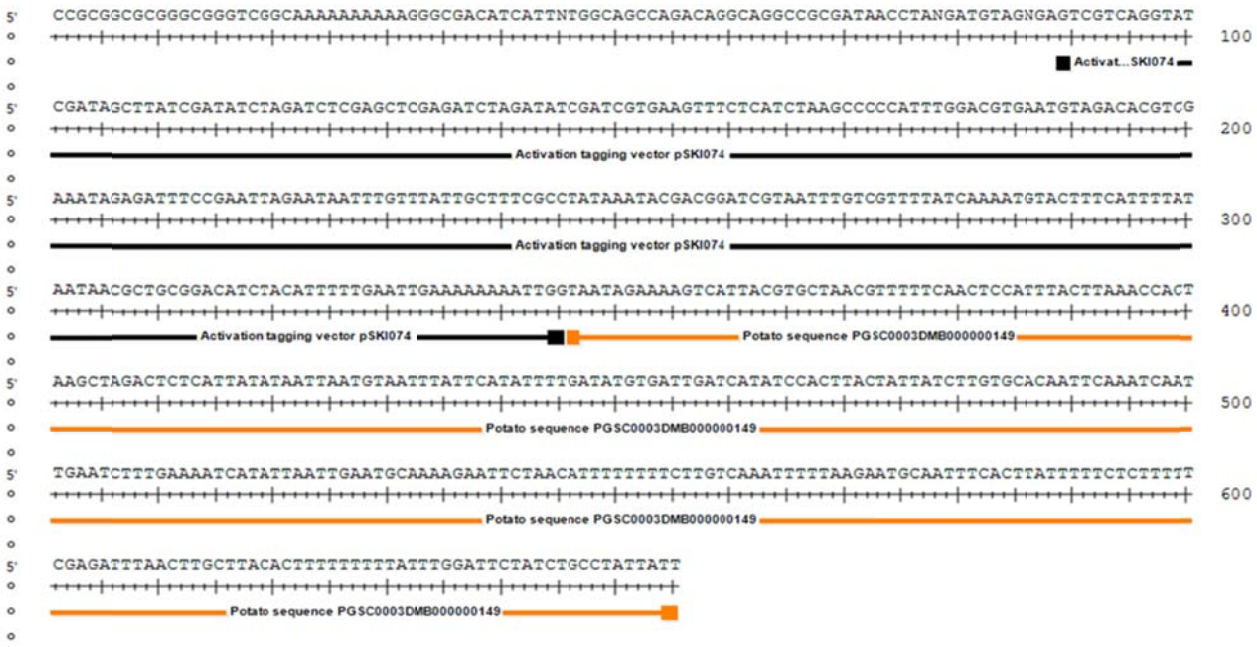


Figure S1.2 Sequence from GenomeWalker AT615-2 SP6 used to query potato scaffold. Activation tagging vector pSKI074 sequence is underlined with the Black line and corresponds to pSKI074: 8057...7801. Potato genome sequence is underlined with the Orange line and corresponds to PGSC0003DMB000000149: 763110...762906 in the PGSC data.

Table 1.1 Phenotyping the *underperformer* mutant in walk-in-growth chamber. Mean ± SE values are given. (N= sample size). The Standardized NE1014 rating codes for plant and tuber characteristics are given in Appendix 1.1.

Description	N	Bintje	*up*
Plant Height (cm)	10	72.9±1.3	41.8±1.4
Chlorophyll Content Index	30	46.1±1.7	40.8±1.5
Tuber shape	8	4.3±0.4	8.6±0.2
Tuber color	8	6.6±0.2	6.8±0.2
No of eyes/tuber	8	13.8±0.9	35±2.2
Tuber eye depth	8	8±0.2	5.1±0.2
Tuber skin texture	8	7±0.3	7.3±0.3
Specific Gravity	3	1.11±0.006	1.08±0.004

Table S1.1 Primers used for PCR amplification of pSKI074 vector sequences, backbone sequences and methylation-enriched DNA.

Primer name	Primer Sequence	Purpose
pSKI074-Fred-5'	GCGTGGCTTTATCTGTCTTTGTATTG	Screening of transgenic plants
pSKI074-Red-3'	GGCCTACTTTAATTGCTTCCAGTGTTA	Screening of transgenic plants
Aulakh35F	TGGCAAGTGTAGCGGTCACG	Southern blot probe
Aulakh35R	CTTGCTTTGAAGACGTGGTTGGAACGT	Southern blot probe
Adapter Primer 1 (AP1)	GTAATACGACTCACTATAGGGC	GenomeWalker
Nested AP1(AP2)	ACTATAGGGCACGCGTGGT	GenomeWalker
SKI074H3(GSP1)	GCTCTCTCGAGGTCGACGG	GenomeWalker
Nested SKI074H3(GSP2)	GAGGTCGACGGTATCGATAAGC	GenomeWalker
AT615 Primer 1	GAATTCTTTTGCATTCAATTAATATGATTTTC	GenomeWalker primers to amplify flanking LB sequence
AT615 Nested to Primer 1	GAATTGTGCACAAGATAATAGTAAGTGGATATGATCA	GenomeWalker primers to amplify flanking LB sequence
Potato AT primer set 4-5'	AAGGGAATAAGGGCGACACGGAAA	Right border T-DNA
pSKI074-35 S promoter-3'	CACTGATAGTTTCGGATCTAGATATC	Right border T-DNA
pSKI074-Fred-5'	GCGTGGCTTTATCTGTCTTTGTATTG	Right border T-DNA
Potato AT primer set 4-3'	CCAACTTAATCGCCTTGCAGCACA	Right border T-DNA
pSKI074-8068-8090-5'	CAACAGAGCCTGGCGTTCCCCTT	Left border T-DNA
pSKI074-Red-3'	GGCCTACTTTAATTGCTTCCAGTGTTA	Left border T-DNA
pSKI-4242..71-5'	GATATCTAGATCCGAAACTATCAGTGTTTG	Right border backbone
pSKI-4266..90-5'	TGTTTGACAGGATATATTGGCGGGT	Right border backbone
pSKI-4359..82-3'	GGCATGCACATACAAATGGACGAA	Right border backbone
pSKI-4573..93-3'	GCGATCGAGGATTTTTCGGCG	Right border backbone
pSKI-4874..97-3'	CGATATCATTACGACAGCAACGGC	Right border backbone
pSKI-5379..5402-3'	AGACGAACGAAGAGCGATTGAGGA	Right border backbone
pSKI-4242..71	GAT ATC TAG ATC CGA AAC TAT CAG TGT TTG	RB T-DNA/plant junction
615-77offRB	GAGGTAAAACGTACACATTATTGTGTC	RB T-DNA/plant junction
615-442off RB	CCC ATA GTT TGA TTT TGA TAT GTG ATT G	RB T-DNA/plant junction
615-1512offRB	CAA TTT TCT CGC CTG TTA TAA AGA TGT C	RB T-DNA/plant junction
615R1-GSP2-5'	GGAGTTTGGATGCTGGAAAGCTTGCTAAGGACC	RACE PCR
615R1-GSP1-3'	TGTGAGGCATCCGAGGATATCCCAAACATATGG	RACE PCR
615R1-NGSP2-5'	TGTCCATACTAGTCGAGGCATGCCGTTTAGCT	RACE PCR
615R1-NGSP1-3'	GACCTCTTCTAACTTGACGATCTGCTCTTGCA	RACE PCR
UPM long	CTAATACGACTCACTATAGGGCAAGCAGTGG TATCAACGCAGAGT	RACE PCR
615R1-ORF-5'	ATG AGG TCA TCG TCT TCT TCA AAG C	BTB/POZ(615R1) amplification
615R1-ORF-3'	TGT CAA TGT GAC CAA GTG TCC ATT TC	BTB/POZ(615R1) amplification
615R1-FLT-3'	TAT GAA ACA AGT AAT ATA ATG TCA AAA GTT TG	BTB/POZ(615R1) amplification

Table S1.2 Primers used for semi-quantitative PCR. R1 to R31 are primers for genes flanking the right border and L1-L41 are primers for genes flanking the left border (LB= Left border, RB= Right border).

Primer name	Forward primer sequence	Reverse primer sequence
AT615R1	TGTAGCTGTTAAGATCGTCCGC	GTCTCATTTCCATGCATCTTGG
AT615R1.1	AACCAGCATCTAGAGCAGTCCTT	CCACTCAAACGTAGAGTTGCTGATA
AT615R2	TTACCCTGCCTGAACGAAAG	GTCTTAGTGGAGGTGATGGTGA
AT615R3	CAACTGGGACTCCCAAAATAA	CTGGCAATGGGCGTGACTTA
AT615R4	GGCTTATATACGCTGGAAAACA	CACCAAAAAAGGGAAGAGAAGA
AT615R6	CCTCCCACCAGGATGGC	TGTCAGGTTAGTTGCTTTCAACT
AT615R8	TCTCCGGACAGGATGTGTTA	CGGGAGTTGCTATCACCAC
AT615R9	GTGGTGCAGGACAAAACATCAT	GCTGCTTTCACATCATCATAAGA
AT615R10	GCGAGGACAGTAGCAGAAAATG	AAGTCTATATGGACGAAAGCACG
AT615R11	TTGTAGTGCATGCTTTGAACA	TATGACGAGACCATCAAATTCTGG
AT615R11.1	AGGATAGTGTGTGGAGTTGAACAT	CCTCTGTGTCCATTTGCATAAA
AT615R12	AATGCTAATTATGTTGTTGAACCA	AATCACATACTTTCCCAGCATTTA
AT615R13	GGCCAGATATTAAAAGAGGAAAA	TTGTAGAGTGAAGGGTTGAAAAT
AT615R14	TGCTTCAGTTTGGATACCTGC	CAGTGTTGTTTCTGGCTATCTCC
AT615R18	CGGTGTCTTCCCACTTCGTG	GGCTCAATTTGCCAAACAAG
AT615R19	TTCATATGATGTTCTCTGACACCA	AAAAACCCCATTCTCGCGAT
AT615R20	CTCAATATCTTGGCAATAATCATAA	CCTAACGAGAAGCCAATGTACT
AT615R21	TTGGTTAGGCTATGGAATGATTTAT	CAACCAATATGTGCCACATGTTTAA
AT615R22	CTATGTTATCTACGCTCCAGTTGAT	AGTTTTGCTGTTGGGACAGAATC
AT615R24	AATGTCTCCAAAGGGATGAACA	GTTTCCATTGCAGCTTCTGTG
AT615R28	AGATCTTGGAAGCATTGACCG	GGAGAAAATCGATCTGCTTGG
AT615R29	AGCTGCCATTCCCACTGC	ATCCGGGTCATCGTCATCTT
AT615R31	GCACATTGGACATGTCTTGGAC	AGTACCTATGGGGATTTGTGAGC
AT615L1	ACTATCATCAACAGAGAGTCCCCG	TATCTCTTTCATCGTCTTCTTCTTCA
AT615L2	AGAGAAATGACAGAAGATGGAAGG	CAATATCAAGAAATTGCTTAGAATCA
AT615L3	ATCTTGAGATCCATGAAATCGG	TTATGAAAGAAAGGAGTTGGTATTG
AT615L4	CCCATCAGTTGCTAGGAGGAC	ATAAATGTAAAGGAAGGGTTTTTG
AT615L5	CAGGTTAATGCTCTATTGTCTTCA	GCTCAAATTCAGATGCCCAG
AT615L11	CCATATCTGCTGGTAAAGGAGGAC	GTTCATATCTTCCCGTTGATTACTG
AT615L16	GGAAGAATTCCATGTTGTGAAAAG	ACCTATTGCCAAGAACAGAGTGA
AT615L17	ATGACTAGGGTGTACATGAGGTGA	TTGGTTGTTGCTTCTGCCAT
AT615L18	GATGCATCTAGGAGCTGGAAACTA	GCAACTTCAATTGAAGAGCTCTG
AT615L19	CGACCTTCAAACACTAACTGTAATG	CAATACCGCACTCTGCAAGGAC
AT615L20	TGTCAGGTGGGATGATGATGTG	CAAAACAAAACAGTTCAATTCGC
AT615L23	GAAACTGATGTTAATGGGGAAAT	ATAGTACTGGTTACTCTCTTCGTGC
AT615L25	TTACTGGTGAGAGCAAGAAGGC	TCCACAACACGGTCTGAAACG
AT615L27	GGGTTGCTTAACACGTTCATCTAT	TGACTATTTTATGTTGGTGGCACA
AT615L29	TGTACAATTTCCTCAGGCAGTTAC	GCCCATTACGACGCAGATATA
AT615L41	AGCAGACTGAGGATATTGGTGAAT	GCACACAATATCGTCTTCTAAGGC

References

Ahn, J.H., et al. (2007). Isolation of 151 mutants that have developmental defects from T-DNA tagging. Plant and Cell Physiology 48:169-178.

Ayliffe, M.A., Pallotta, M., Langridge, P., and Pryor, A.J. (2007). A barley activation tagging system. Plant Mol. Biol. 64:329-347.

Bell, A.C., and Felsenfeld, G. (1999). Stopped at the border: boundaries and insulators. Curr. Opin. Genet. Dev. 9:191-198.

Borevitz, J.O., Xia, Y.J., Blount, J., Dixon, R.A., and Lamb, C. (2000). Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell 12:2383-2393.

Busov, V., et al. (2011). Activation tagging is an effective gene tagging system in *Populus*. Tree Genet. Genom. 7:91-101.

Busov, V.B., Meilan, R., Pearce, D.W., Ma, C.P., Rood, S.B., and Strauss, S.H. (2003). Activation tagging of a dominant gibberellin catabolism gene (*GA 2-oxidase*) from poplar that regulates tree stature. Plant Physiol. 132:1283-1291.

Chantha, S.C., and Matton, D.P. (2007). Underexpression of the plant *NOTCHLESS* gene, encoding a WD-repeat protein, causes pleitropic phenotype during plant development. Planta 225:1107-1120.

Fischer, L., Lipavska, H., Hausman, J.F., and Opatrny, Z. (2008). Morphological and molecular characterization of a spontaneously tuberizing potato mutant: an insight into the regulatory mechanisms of tuber induction. BMC Plant Biol. 8:13.

Harrison, E.J., et al. (2007). Diverse developmental mutants revealed in an activation-tagged population of poplar. Canadian Journal of Botany 85:1071-1081.

Hayashi, H., Czaja, I., Lubenow, H., Schell, J., and Walden, R. (1992). Activation of a plant gene by T-DNA tagging: auxin-independent growth *in vitro*. Science 258:1350-1353.

Hiratsu, K., Matsui, K., Koyama, T., and Ohme-Takagi, M. (2003). Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in *Arabidopsis*. Plant J. 34:733-739.

Huang, S.S., Cerny, R.E., Bhat, D.S., and Brown, S.M. (2001). Cloning of an *Arabidopsis* patatin-like gene, STURDY, by activation T-DNA tagging. Plant Physiol. 125:573-584.

Ichikawa, T., et al. (2003). Sequence database of 1172 T-DNA insertion sites in *Arabidopsis* activation-tagging lines that showed phenotypes in T1 generation. Plant J. 36:421-429.

Kakimoto, T. (1996). *CKI1*, a histidine kinase homolog implicated in cytokinin signal transduction. Science 274:982-985.

Kardailsky, I., et al. (1999). Activation tagging of the floral inducer *FT*. Science 286:1962-1965.

Kondrak, M., Marincs, F., Kalapos, B., Juhasz, Z., and Banfalvi, Z. (2011). Transcriptome analysis of potato leaves expressing the *Trehalose-6-Phosphate Synthase 1* gene of yeast. Plos One 6.

Larkin, M.A., et al. (2007). Clustal W and clustal X version 2.0. Bioinformatics 23:2947-2948.

Lechner, E., et al. (2011). MATH/BTB CRL3 receptors target the homeodomain-leucine zipper ATHB6 to modulate abscisic acid signaling. Dev. Cell 21:1116-1128.

Liu, Q., et al. (2011). Novel human BTB/POZ domain-containing zinc finger protein ZBTB1 inhibits transcriptional activities of CRE. Mol. Cell. Biochem. 357:405-414.

Mandadi, K.K., Misra, A., Ren, S., and McKnight, T.D. (2009). BT2, a BTB protein, mediates multiple responses to nutrients, stresses, and hormones in *Arabidopsis*. Plant Physiol. 150:1930-1939.

Massa, A.N., Childs, K.L., Lin, H.N., Bryan, G.J., Giuliano, G., and Buell, C.R. (2011). The transcriptome of the reference potato genome *Solanum tuberosum* Group Phureja clone DM1-3 516R44. Plos One 6.

Mathews, H., et al. (2003). Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. Plant Cell 15:1689-1703.

Qu, N., Gan, W., Bi, D., Xia, S., Li, X., and Zhang, Y. (2010). Two BTB proteins function redundantly as negative regulators of defense against pathogens in *Arabidopsis*. Botany-Botanique 88:953-960.

Radouani, A. (2003). Effect of plant growth regulators, plantlet attributes and nutrient medium on *in vivo* and *in vitro* potato tuberization. In: Science Agronomiques Rabat, Morocco: Institut Agronomique & Vétérinaire Hassan II.

Regan, S., et al. (2006). Finding the perfect potato: using functional genomics to improve disease resistance and tuber quality traits. Can. J. Plant Path. 28:S247-S255.

Robert, H.S., Quint, A., Brand, D., Vivian-Smith, A., and Offringa, R. (2009). BTB and TAZ domain scaffold proteins perform a crucial function in *Arabidopsis* development. Plant J. 58:109-121.

Sambrook, J., and Russell, D.W. (2001). Molecular cloning : A laboratory manual. Cold Spring Harbour: Cold Spring Harbour Laboratory Press.

Suzuki, Y., et al. (2001). A novel transposon tagging element for obtaining gain-of-function mutants based on a self-stabilizing *Ac* derivative. Plant Mol. Biol. 45:123-131.

The Potato Genome Sequencing Consortium. (2011). Genome sequence and analysis of the tuber crop potato. Nature 475:189-U194.

The Tomato Genome Consortium. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635-641.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105-1111.

Trapnell, C., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28:511-U174.

Wang, G., Liu, P.-C., Wang, J.-X., and Zhao, X.-F. (2011). A BTB domain-containing gene is upregulated by immune challenge. Arch. Insect Biochem. Physiol. 77:58-71.

Warnatz, H.-J., et al. (2011). The BTB and CNC Homology 1 (BACH1) target genes are involved in the oxidative stress response and in control of the cell cycle. J. Biol. Chem. 286:23521-23532.

Weigel, D., et al. (2000). Activation tagging in *Arabidopsis*. Plant Physiol. 122:1003-1013.

Zubko, E., Adams, C.J., Machaekova, I., Malbeck, J., Scollan, C., and Meyer, P. (2002). Activation tagging identifies a gene from *Petunia hybrida* responsible for the production of active cytokinins in plants. Plant J. 29:797-808.

Chapter 2

Global gene expression profiling of leaves of an activation tagged potato mutant, *underperformer* by mRNA-seq analysis

Sukhwinder S. Aulakh[1, 2], Allan Dickerman[3], Richard E. Veilleux[1*] and Barry S. Flinn[1, 2*]

Abstract

The transcriptome represents the complete set of transcripts in a cell, and their quantity, for a specific developmental stage or physiological condition. RNA-sequencing (RNA-seq) is a relatively new whole transcriptome analysis strategy that uses Next Generation Sequencing (NGS) platforms to carry this out, and offers several advantages over hybridization based methods like microarrays. In this report, we studied gene expression changes between potato cv. Bintje and its activation tagged mutant *underperformer* (*up*) on a global scale, using mRNA-seq on leaf samples harvested from 60 day-old 'Bintje' control and mutant plants. The mRNA-seq analyses resulted in a unique dataset of 1,632 genes that were differentially expressed between mutant *up* and Bintje leaf samples, with the total number of up-regulated genes (661 genes) were less than the number of genes down-regulated (971 genes) in the mutant. Visualization of differentially expressed genes was achieved with the MapMan software, which allowed the grouping of modified genes into categories according to biological processes or pathways. Our results indicated that a variety of biological processes and pathways were modified in the mutant, including increased carbon allocation through carbohydrate metabolism and reduced lipid metabolism, decreased cell division and cell cycle activity, and abiotic stress responses. We relate the modifications in gene expression encountered in the mutant to the overall phenotype of the *up* mutant, and to the tentatively-identified activation tagged gene, a member of the BTB/POZ (Broad complex, Tramtrac, Bric a brac; also known as Pox virus and Zinc finger) domain-containing protein family.

Keywords: *Solanum tuberosum*, transcriptome, Solanaceae, BTB/POZ domain, Next generation sequencing

[1] Department of Horticulture, Virginia Tech, Blacksburg, Virginia, USA, 24060

[2] The Institute for Sustainable and Renewable Resources (ISRR) at The Institute for Advanced Learning and Research (IALR), Danville, Virginia, USA, 24540

[3] Virginia Bioinformatics Institute (VBI), Blacksburg, Virginia, USA, 24060

* Both contributed to project development and supervision.

Introduction

The transcriptome represents the complete set of transcripts in a cell, and their quantity, for a specific developmental stage or physiological condition (Wang et al., 2009). The main goal of whole transcriptome analysis is to identify, characterize and catalogue all the transcripts expressed within a specific cell/tissue, at a particular stage, with the potential to determine the correct splicing patterns and structure of genes, and to quantify the differential expression of transcripts under both physiological and pathological conditions (Costa et al., 2010). The traditional workhorse of transcriptome analysis, microarray, was the method of choice for gene expression studies and facilitated important advances in a wide range of biological processes. However, microarrays have limitations, such as high background, limited dynamic range, and the need to carry out complex and costly experimental replications. Prior sequence information of probes is also a requirement. However, with the arrival of Next Generation Sequencing (NGS) platforms [Genome Sequence FLX (Roche), Genome Analyzer IIx and HiSeq 2000 (Illumina), SOLiD 4 and 5500 Series Genetic Analyzer (Applied Biosystems)] over the last few years, sequencing capabilities and associated gene expression profiling capabilities have achieved a new level. These NGS platforms are capable of producing hundreds of millions of short reads (normally 25-400 bp) in parallel sequencing. For example, GS FLX titanium produces greater than 1 million reads in excess of 400 bp, while the Helicos Bioscience Heliscope produces 400 million reads of 25-35 bp. While each platform uses different technologies, the basic concept is similar. DNA is randomly broken into millions of smaller-sized pieces, these pieces or templates are attached to a solid support and then millions of individual miniature sequencing reactions take place in parallel with single DNA molecules in each reaction (Costa et al., 2010). In addition to generating large amounts of sequencing data, the cloning step used in traditional sequencing is eliminated, reducing costs. The development of these NGS sequencing platforms has resulted in several new sequencing approaches: RNA-sequencing (RNA-seq) for transcriptome analysis, chromatin immunoprecipitation sequencing (Chip-seq) for DNA-protein interaction analysis, DNase-seq for identification of the most active regulatory regions, CNV-seq for copy number variation analysis, and methyl-seq for genome wide profiling of epigenetic marks. Amongst these, RNA-seq is the most powerful and perhaps the most complex NGS technology.

RNA-seq offers several advantages over microarray hybridization based methods. It can identify new transcripts and isoforms, determine exon-intron boundaries to a single nucleotide level and thereby correct gene annotation, is useful for organisms with no sequence information, is highly sensitive with low background, require less RNA sample and there is no upper limit for quantification of transcripts. The detailed RNA-seq method, its advantages and limitations have been recently reviewed (Cloonan and Grimmond, 2008; Costa et al., 2010; Jain, 2011; Wang et al., 2009). Many research groups are now using NGS platforms for transcriptome analysis and a brief summary of recent publications on RNA-seq use with plants is provided in **Table 2.1**. The sensitivity and large amount of data produced by NGS platforms requires the use of supercomputers (or computer clusters) and new bioinformatics and statistical tools to draw valuable conclusions from the data. Typical steps in an RNA-seq experiment include sample preparation, mRNA enrichment, conversion of mRNA into cDNA and preparation of cDNA libraries and sequencing. After completion of RNA-sequencing, the downstream steps of data storage, quality control, alignment and assembly of reads, statistical analysis and visualization and interpretation of transcriptome data represent specialized requirements, and several open source tools and third party software have been developed for this purpose. Some of these tools are: 1) TopHat (Trapnell et al., 2009), an open-source software designed to align reads from a RNA-seq experiment to a reference genome without relying on known splice sites; 2) Cufflinks (Trapnell et al., 2010), which assembles transcripts, estimates their abundance, and tests for differential expression and regulation in RNA-seq samples; 3) Multi-Experimental Viewer (MeV) (Howe et al., 2011) and 4) S-MART (Zytnicki and Quesneville, 2011), which allows advanced analysis of gene expression data through an intuitive graphical interface; 5) RNA-Seq by Expectation Maximization (RSEM), which quantifies gene and isoform abundance from single-end and paired-end RNA-seq data without any reference genome (Li and Dewey, 2011); 6) DESeq, a method based on negative binomial distribution and implemented in the R/Bio-conductor package, used to infer differential expression data with good statistical power (Anders and Huber, 2010); as well as several others. Recently, RNA-seq analysis tools designed to run in parallel on Amazon cloud computing infrastructure have also been developed (Hong et al., 2012). Visualization of pathway data can be achieved with MapMan (Thimm et al., 2004),

which displays RNA-seq data sets onto diagrams of metabolic pathways and other biological processes, or other similar tools previously developed for microarrays.

We recently characterized the activation-tagged potato mutant *underperformer* (**Chapter 1**), a member of a potato activation-tagged mutant population created by the Canadian Potato Genome Project from the 'Bintje' variety using the pSKI074 activation tag vector (Weigel et al. 2000). The *underperformer* (*up*) plants exhibited multiple phenotypes when grown *in vitro* and in soil. These plants were dwarf, had smaller sized leaves, produced long and fingerlike tubers with reduced tuber dormancy, had small floral buds which did not open, and dropped prematurely, and exhibited early leaf senescence (**see Chapter 1 – Figure 1.7**). Additionally, *up* mutant plants exhibited a significant reduction in several traits (above ground fresh weight, chlorophyll content, total fresh weight, total dry biomass, total tuber weight per plant, number of size one tubers, and marketable yield), compared to wild-type Bintje (**Chapter 1 – Figures 1.8, 1.9 and Table 1.1**). The location of the T-DNA insertion was found on superscaffold PGSC0003DMB0149: 763110, more specifically on chromosome 10 position. We analyzed approx. 400 kb regions both upstream and downstream of the insertion site, and observed up- and down-regulation of genes off both right and left borders of the T-DNA. The detailed morphological and molecular analysis of the *underperformer* mutant is described in detail in **Chapter 1**. In this chapter we report the results of mRNA-seq analysis to characterize gene expression changes on a global level in the leaves of the mutant compared with wild-type 'Bintje' plants. This analysis resulted in a unique dataset of 1,632 genes that were differentially expressed between *up* and Bintje. We attempted to present the broader picture of gene expression changes, derive the biological meaning from this data and relate this to the role of activation tagging in the mutant phenotype.

Results

Tissue samples and mRNA-sequencing

We analyzed the gene expression profiles of potato cv. Bintje and its activation tagged mutant *underperformer* (*up*). Leaf samples harvested at 60 days after planting were used for mRNA-seq. Sequencing experiments were performed with two independent biological replicates; each

biological replicate consisted of two technical replicates. This study generated >207 million mRNA-seq reads in total, out of which 57 million, 36 bp long reads were generated with Illumina's GAII sequencer, and 150 million, 50 bp reads were generated with Illumina's HiSeq 2000 NGS platform (**Table 2.2**). Both sets of reads were sequenced from a single end. The average base quality score for the GAII reads was 38.2, and that of the HiSeq reads was 38.4, and the resultant gene expression data in general was agreeable for both platforms. The GC content (%GC) of GAII reads (49%) and HiSeq reads (44%) were both higher than the 34.8% GC content of the potato reference sequences (The Potato Genome Sequencing Consortium, 2011). Out of the total 207 million mRNA-seq reads, 183 million reads (88%) were mapped to potato genome assembly PGSC_DM_v3_2.1.10_pseudomolecules.zip using TopHat v2.0.4. However, when reads were mapped separately for each dataset with the potato genome assembly, 84% and 89% reads were mapped for GAII and HiSeq 2000 datasets, respectively (**Table 2.2**).

Differential gene expression analyzes

The GAII dataset flagged 371 differentially expressed genes as significant compared to 209 such genes with the HiSeq dataset between 'Bintje' and the *underperformer* mutant. However, our combined dataset flagged 1632 genes as showing significantly different expression levels (**Table 2.2**) between two genotypes. Out of the 1632 differentially expressed genes between Bintje and *up*, 661 genes (40.5%) were found to be up-regulated in the mutant, whereas 971 genes (59.5%) were down-regulated in the mutant. While the T-DNA insertion in the mutant was previously identified as occurring on chromosome 10, the distribution of up- and down-regulated genes followed a random pattern across all 12 potato chromosomes (**Figure 2.1**). The most up-regulated genes (110) were found on chromosome 9, and the most down-regulated genes (133) were found on chromosome 2 (**Figure 2.1**).

Visualization of gene expression data

We used MapMan to assign the 1632 differentially-expressed genes to various functional categories and subcategories based on MapMan GO annotation, and the genes were shown as different elements for major plant pathways. MapMan displays large datasets onto diagrams of

metabolic pathways or other processes as pictorial diagrams that symbolically depict areas of biological function. The classification was based on hierarchical functional categories (BINs, subBINs, individual enzymes). In the next several sections, we will describe the major biological pathways/processes, those most relevant to our RNA-seq dataset and into which most of our differential expressed genes were clustered. We use the terms up- and down-regulation in gene expression levels of mutant genes relative to wild-type 'Bintje' gene expression.

Metabolism overview

This pathway provides an overview of the major metabolic functions of an organism. A total of 196 genes grouped in 58 categories (BINs) was represented on the metabolism overview diagram (**Figure 2.2**). When considering the 196 differentially expressed genes between 'Bintje' and *up*, 94 genes were up-regulated (blue squares), whereas the remaining 102 were down-regulated (red squares) in mutant leaves. Amongst the genes up-regulated in the mutant, the tetrapyrrole synthesis gene glutamyl-tRNA reductase [PGSC0003DMP400054292; Bin: 19.2] exhibited the highest log2-fold change in expression level (3.9) whereas for down-regulated genes, a transferase family protein involved in secondary metabolism [PGSC0003DMP400047121; Bin: 16.2] exhibited the highest log2-fold change in expression level (-4.8) in *up* compared to Bintje. Most of the genes involved in photosynthesis pathways, including light reactions and Calvin cycle [Bin: 1.1; 1.3], as well as major carbohydrate metabolism [Bin 2.1; 2.2], minor carbohydrate metabolism [Bin: 3.1 to 3.6], glycolysis [Bin: 4], fermentation [Bin 5], glyoxylate cycle [Bin 6], TCA [Bin: 8.1; 8.3], two cell wall synthesis genes GAE6 (UDP-D-Glucuronate 4-Epimerase 6) [Bin: 10.1] and tetrapyrolle [Bin: 19] were up-regulated in the *up* mutant (**Figure 2.2**).

Genes regulating cell wall synthesis and degradation displayed mixed expression levels, with genes in some of the subcategories showing up-regulation, and others down-regulation. Most of the pectin esterase genes were down-regulated in the *up* mutant except AGP [Bin: 10.5.1], as was cellulose and hemicellulose synthesis [Bin: 10.2], while most xyloglucan endotransglycosylase [XETs; Bin: 10.7] genes were up-regulated in the mutant. In the expansin category [Bin: 10.8] both up- and down-regulation of genes were observed. Genes associated with lipid metabolism that include fatty acid (FA) synthesis and FA elongation [Bin: 11.1] and

lipid degradation lipases [Bin: 11.9.2] were down-regulated, whereas genes belonging to FA desaturation [Bin: 11.2] and lipid degradation lysophospholipases [Bin: 11.9.3] were up-regulated in the mutant. Secondary metabolism genes [Bin: 16] that include terpenes, flavonoids, phenylpropanoids and phenolic compounds; amino acid metabolism genes [Bins: 13.1; 13.2] and genes involved in nucleotide metabolism [Bin: 23] were also down-regulated in the mutant. The detailed Bin ids, Bin description, peptide ids, gene annotation and log2-fold change values are given in supplementary **Table S2.1** under the tab metabolism overview.

Cell function overview

There were 1441 genes involved in cell function processes which were significantly differentially expressed in the present dataset. These were further divided into two groups; 560 genes were up-regulated and 881 genes were down-regulated in the mutant (**Figure 2.3**). However, approximately one third of the genes (465 out of 1441) were not assigned to any functional category, and so were grouped together [Bin: 35.2]. Amongst the up-regulated genes, the highest fold increase in expression levels (7.3 and 6.8 log2fold) were observed for a gene involved in regulation of transcription [PGSC0003DMP400055149] and in a homeobox transcription factor family gene encoding peptide PGSC0003DMP400024694 [Bins: 27.3.9 and 27.3.22 respectively]. Overall, however, most genes involved in regulation of transcription [Bin: 27.3] were down-regulated. Amongst the down-regulated genes in the mutant, the gene which showed the greatest level of down-regulation (-5.3) was found to be a glutaredoxin gene PGSC0003DMP400004213 [Bin: 21.4]. We observed up-regulation of expression for a majority of the genes belonging to hormone metabolism [Bin: 17], abiotic stress [Bin: 20.0], glutaredoxins [Bin: 21.4], protein degradation [Bin: 29.5] and transport categories [Bin: 34] in the mutant. Genes involved in DNA and RNA synthesis [Bins: 28.1; 27.2], DNA repair [Bin: 28.2], cell organization [Bin: 31.1], cell division [Bin: 31.2], cell cycle [Bin: 31.3], cell vesicle transport [Bin: 31.4] were down-regulated in the mutant. However, there were also a third group consisting of biotic stress [Bin: 20.1], regulation of transcription [Bin: 27.3], signaling [Bin: 30], development [Bin: 33], posttranslational protein modification [Bin: 29.4], various enzyme families [Bin: 26] which displayed a similar mix of up- and down-regulated genes, although the majority of the genes in these categories/subcategories were down-regulated in the mutant.

Regulation overview

In the regulation pathway (**Figure 2.4**) there were 452 genes displayed which were grouped into 22 categories/subcategories (BINs/subBINs). However, the majority of these genes (312) clustered into three groups [Bins: 27.3; 29.4; 29.5] and the results have already been described in the cell function overview section (**Figure 2.3**). The differentially expressed genes in the remaining three groups; hormones [Bin: 17.1 to 17.8], regulation [Bins: 30.1 to 30.5; 30.11] and redox state [Bins: 21.1 to 21.6] were unique to this process. Amongst hormone metabolism, a gene encoding the abscisic acid (ABA) biosynthetic enzyme, 9-*cis*-epoxycarotenoid dioxygenase 3 (NCED3), which is induced during drought [Bin: 17.1], several auxin responsive genes; PIN5, DFL1, SAURs [Bin: 17.2], SMT2; a gene encoding sterol-C24-methyltransferases (SMTs) involved in sterol biosynthesis [Bin: 17.3] and cytokinin oxidase/dehydrogenase1 gene [Bin: 17.4] were strongly up-regulated in the mutant. Out of 20 genes in the ethylene metabolism category, 12 were up-regulated and 8 were down-regulated in the mutant [Bin: 17.5]. The majority of genes associated with salicylic acid [Bin: 17.8] and gibberellic acid [Bin: 17.6] metabolism were down-regulated in the mutant. In the regulation group, genes in the receptor kinase category, LRR and DUF [Bin: 30.2]; calcium regulation, calcium ion binding and calmodulin binding proteins [Bin: 30.3] and G-proteins [Bin: 30.5] were down-regulated in the mutant. In the redox group, genes belonging to most of the subgroups, like thioredoxin [Bin: 21.1], heme [Bin: 21.3], glutaredoxin [Bin: 21.4] and dismutase/catalase [Bin: 21.6] were up-regulated, whereas glutathione dehydrogenase (ascorbate) gene [Bin: 21.2] was down-regulated in the mutant.

Cellular response and stress related genes

In the cellular response pathway, 66 genes were up-regulated and 97 were down-regulated in the mutant from a total set of 163 genes representing 14 categories or BINs (**Figure 2.5**). Results for gene categorized in the redox cluster [Bins: 21.1 to 21.6] were reported in the regulation overview section, and results of cell division [Bin: 31.2], cell cycle [Bin: 31.3] and development [Bin: 33] genes were presented in the cell function overview section, and genes involved in biotic stress [Bin: 20.1] group will be reported in detail in the following paragraph. Therefore, here we will report on the expression level of genes involved in abiotic stress response [Bin: 20.2.1 to 20.2.4; 20.2.99]. Genes involved in heat and cold stress and wound

responses, like heat shock proteins (HSPs), cold regulated 413-plasma membrane 2 (COR413-PM2) and senescence associated gene 20 (SAG20) were up-regulated in the mutant. However there was a single early-responsive to dehydration protein-related gene (peptide id: DMP40010930) which was 2-fold (log2 value) down-regulated in the mutant.

In the biotic stress pathway, total of 322 genes were represented, with 25 clusters or major BINs (**Figure 2.6**). Results have already been reported for categories including hormones, cell wall, proteolysis or protein degradation, abiotic stress, redox state signaling and heat shock proteins. Gene clusters which are important for biotic stress response include Salicylic Acid (SA) and Jasmonic Acid (JA), cell wall genes, proteolysis, transcription factors and secondary metabolites. *Two peptides (Bin: 17.8; PGSC*0003DMP400038439 and DMP400022754) encoding indole-3-acetic acid (IAA) carboxyl methyltransferase-1 (IAMT1), involved in converting IAA to methyl-IAA ester (MeIAA) were down-regulated in the mutant. In addition, a gene encoding an S-adenosyl-L-methionine:carboxyl methyltransferase family protein (Bin: 17.8; DMP400033205) was highly up-regulated (3.8 log2fold) in the mutant. Another gene [Bin: 17.7; DMP400018673], which was moderately similar to allene oxide synthase (AOS) was more than fourfold (4.13) up-regulated in the mutant, potentially increasing Jasmonic Acid (JA) levels. However, a Jasmonate-ZIM-domain protein 1 (JAZ1) gene [Bin: 17.7; DMP400005281] was down-regulated (-1.1 log2fold) in the mutant. The plant β-1, 3-glucanases are a highly complex gene family involved in pathogen defense, as well as a wide range of normal developmental processes. In the *up* mutant, most of the genes of the β-glucanase cluster [Bin: 26.4] were down-regulated. Pathogenesis-related (PR) proteins showed both up- and down-regulation of genes in the mutant [Bin: 20.1.7]. Most of genes in peroxidase gene family [Bin: 26.12] including the RCI3 (Rare Cold Inducible Gene 3) were down-regulated in the mutant. Several gene members of the Tau class Glutathione-S-Transferases [Bin: 26.9], involved in the oxidative stress response were up-regulated in the mutant. However, majority of the transcription factors involved in abiotic stress like ERF, bZIP, WRKY, MYB, DOF [Bins: 27.3 and sub-Bins within] were down-regulated in the mutant.

Secondary metabolism

In the secondary metabolism pathway, 44 genes were differentially expressed between 'Bintje' and the mutant, representing 11 Bins (**Figure 2.7)**. Most of the genes in this pathway were down-regulated in the mutant. However, some genes were up-regulated, including a non-mevalonate pathway gene, geranylgeranyl pyrophosphate synthase [Bin: 16.1.1; DMP400012558], a gene similar to terpene synthase 21 [Bin: 16.1.5; DMP400020890; 1.5 log2fold], O-methyltransferase family 2 protein [Bin: 16.2; DMP400001619; 2.1 log2fold], 4-coumarate:COA ligase 2 (4CL2) [Bin: 16.2; DMP40002548; 1.6 log2fold], irregular xylem 12 (IRX12)/laccase 4 (LAC4) [Bin: 16.10; DMP400035339; 3.7 log2fold], as well as other genes [Bins: 16.1.2; 16.2.1; 16.4.1 and 16.8.3].

Large enzyme families and RNA-protein synthesis

There were 12 large gene families for which significant differential gene expression was noted in our dataset, with 110 genes represented (**Figure 2.8**). In most of the gene families, we observed more down-regulated than up-regulated genes in the mutant. For example, in the cytochrome P450 gene family, 11 genes were up-regulated and 14 were down-regulated [Bin: 26.10]. Similarly, 10 genes were up-regulated and 18 were down-regulated in UDP glycosyltransferases gene family [Bin: 26.2], 3 genes were up-regulated and 16 down-regulated for the GDSL-lipases, and one gene was up-regulated and 7 down-regulated for the nitrilase gene family.

In the RNA-protein synthesis pathway there were 141 genes displayed (**Figure 2.9**). However, most of gene clusters presented here, have already reported in other sections (see above). Genes involved in RNA synthesis [Bin: 27.2], RNA processing [Bin: 27.1], protein synthesis and amino acid activation [Bins: 29.2.1; 29.2.3 and 29.2.4], vesicle transport and protein targeting [Bin: 29.3], protein modifications [Bin: 29.4], protein degradation [Bin: 29.5] were presented in the cell function overview section, and the majority of these genes were down-regulated in the mutant. Similarly in the transcription factor gene cluster [Bin: 29.3], four genes were down-regulated and 2 were up-regulated in the mutant.

BTB/POZ domain containing protein

A variety of genes across all 12 chromosomes, involved in numerous pathways and processes, were found to be up- or down-regulated in the mutant compared to wild-type Bintje. The mutant, *underperformer*, is an activation-tagged potato mutant, with the first gene adjacent to the right border of the T-DNA insertion representing a conserved gene of unknown function, whose closest orthologs in Arabidopsis (At1g04390) and tomato (Solyc10g005600.2.1) encode a BTB/POZ (Broad complex, Tramtrac, Bric a brac; also known as Pox virus and Zinc finger) domain-containing protein. The RNA-seq reads corresponding to this gene region covered two annotated potato gene models PGSC0003DMG402011239 and PGSC0003DMG401011239 (PGSC0003DMP400019889). These raw reads from the Illumina GAII platform for both Bintje and *up* mutant (**Figure S2.3**) indicated higher raw read counts for the mutant *underperformer* (1236) compared to cv. Bintje (284), illustrating the up-regulation of the BTB/POZ domain-containing protein transcript level in the mutant. The corresponding gene expression value for this region was found to be 2.3-fold (log2 value) higher in the mutant. Furthermore, this particular BTB/POZ domain-containing protein clustered in the protein degradation pathway (**Figure 2.10**), suggesting that this pathway may be important for the overall mutant phenotype.

Discussion

This study has demonstrated the efficacy of high-throughput NGS sequencing technology to assess global gene expression levels in wild-type and mutant potato samples. A unique aspect of this study was that our mRNA-seq data were produced using two Illumina platforms, which generated different outputs. The GAII platform generated 36 bp reads, compared to the longer 50 bp reads of the HiSeq 2000 platform. In addition, the number of reads from the HiSeq platform was approx. 3-fold greater than from the GAII platform. Both GAII and HiSeq use sequencing-by-synthesis (SBS) chemistry, but HiSeq has a 2- to 5-fold increased data acquisition rate. Although we did not report on a comparison of platform-specific bias related to quality of reads for mRNA-seq experiments in this study, we found that both platforms produced high quality mRNA-seq reads. The average base quality scores for GAII and HiSeq reads were comparable, but there were some differences in the GC content of reads generated with both platforms, which were higher than observed for the reference genome. Similar results were

found in a study in which GAII and HiSeq data were compared (Minoche et al., 2011). Since we used the same developmental stage (60 DAP leaf) for RNA sampling in two biological replicates for both platforms, this could be a useful dataset to compare GAII and HiSeq platforms in the future. Our analyses suggest that the Illumina sequencing data were highly reproducible. We sequenced each sample in two lanes as technical replicates, which was sufficient for our experiment, since each additional lane provided only modest gains in the number of genes detected, as found in an assessment of technical reproducibility of RNA-seq data (Marioni et al., 2008). Another key feature of our study was to analyze RNA-seq data using different analysis methods. RNA-seq is a relatively new technique for studying gene expression levels and standardized methods to analyze RNA-seq data are still being developed (Salzman et al., 2011). RNA-seq reads follow different distributions, with about 81% of the reads following a general linear model (similar to microarray), 14% a poisson distribution and 4% a negative binomial distribution. That means, the majority of genes still follow the same linear distribution pattern, and raw read counts should provide a rough idea of gene expression. Based on read counts only, we identified ~4500 differentially expressed genes in the GAII RNA-seq data (**Figure S2.1**). When we analyzed the same dataset (GAII) based on negative binomial distribution (Anders and Huber, 2010), we found 206 genes which were differentially expressed (**Figure S2.2**). We adopted the widely used RPKM values as a measure of gene expression, which are normalized units and allow comparison both within and between the samples, and are based on Cufflinks (Trapnell et al., 2010). Genes were flagged as significant based on the q value of multiple comparisons. We found 371 and 209 significant differentially expressed genes in GAII and HiSeq datasets, respectively. Our approach of combining both datasets worked well, and indicated 1632 genes which exhibited significant differential gene expression levels between Bintje and *up* mutant leaf samples. Hence, our multiple comparisons provided stronger and more sensitive statistical validation.

Visualization of gene expression data and answering the biological question

Differential gene expression data derived from mRNA-seq of potato cv. Bintje and the activation-tagged mutant *underperformer* were displayed using MapMan (Thimm et al., 2004). This allowed the data to be viewed at different levels of resolution, from the global level, the

level of discrete processes, or down to the single gene level. The overview of expression levels indicated that the total number of genes down-regulated was greater than the total number of genes up-regulated in the activation tagged mutant *underperformer*. As noted previously (**Chapter 1**), the *underperformer* mutant exhibited multiple phenotypes, including dwarfness, smaller leaves, finger-like tubers, early senescence, loss of flowering and reduced dormancy. In addition, the large number of genes modified in their expression patterns in leaves of the *up* mutant indicated that the insertion of the activation tag in the mutant resulted in numerous pleiotropic effects. Hence, the various trait phenotypes are most likely due to the involvement of many genes, some of which (e.g. transcription factors) may activate or repress other genes.

Our mRNA-seq data indicated 1632 differentially-expressed genes distributed across numerous metabolic, developmental, regulation, cell function and stress pathways. Most genes in the photosynthesis pathway (light reaction, Calvin cycle, and photorespiration), hormone metabolism (except SA and JA), abiotic and biotic stress pathways, starch and sucrose metabolism, minor carbohydrate metabolism, cell wall degradation proteins, proteolysis pathway, and redox reaction pathway and transport were up-regulated in the mutant. Furthermore, there appeared to be a shift in carbon metabolism, with increased gene expression towards carbohydrate metabolism, and reduced gene expression toward lipid metabolism in the mutant. In contrast, most of the genes involved in DNA synthesis, cell division and cell cycle, cell organization, development, transcription factors, signaling (receptor kinase, G-proteins and calcium regulation), protein modification and secondary metabolism were down-regulated in the mutant.

Given the significant role of hormones in controlling plant development, it was not surprising that various hormone signaling pathways appeared to be modified in the *up* mutant. For example, auxins are major plant hormones closely connected with light signal transduction, affect numerous plant growth and development processes (eg. lateral root differentiation and apical dominance), and they are also known to induce the expression of several genes like DFL1 and SAURs (Nakazawa et al., 2001). In addition, the overexpression of the IAMT1 gene, which encodes an indole-3-acetic acid (IAA) carboxyl methyltransferase that converts IAA to methyl-

IAA ester (MeIAA), affects auxin-regulated processes (Qin et al., 2005). In the present study DFL1 and SAUR genes were up-regulated, whereas IAMT1 was down-regulated in the *up* mutant, suggesting modifications in auxin response potential.

Apart from auxin, other hormone pathways were also modified in the mutant. Cytokinins are growth control hormones, which promote cell division, nutrient mobilization and leaf longevity, modulate defense responses via accumulation of salicyclic acid (Choi et al., 2011). SMT2 is an enzyme of the sterol pathway involved in the synthesis of 24-ethyl sterols (e.g. sitosterol), and over-expression of SMT2 (leading to high sitosterol content) in plants lead to overall reduced stature, small siliques and other phenotypic differences (Schaeffer et al., 2001), due to reduced cell division and cell size. Another enzyme, cytokinin oxidase/dehydrogenase (CKX1), is responsible for most cytokinin catabolism, inactivating the hormone in single enzymatic step and thereby control cytokinin-dependent processes by controlling the local cytokinin levels (Schmulling et al., 2003). We observed strong up-regulation of SMT2 and CKX1 genes in the *up* mutant and the resemblance of the *up* phenotypes to others described above indicates the direct or indirect contribution of these hormone pathways to the overall phenotype. In addition, the observed down-regulation of cell division and cell cycle genes would further contribute to the reduced growth phenotype of the mutant.

Several genes involved in biotic and abiotic defense response were up-regulated in the mutant. For example, allene oxide synthase (AOS), has known effects of increasing endogenous jasmonic Acid (JA) levels, pathogen-related (PR) protein expression and host resistance to fungal infections (Mei et al., 2006). Another gene, encoding a putative S-adenosyl-L methionine: salicylic acid carboxyl methyltransferase1 (SAMT1), is induced by treatment with chemical inducers of disease resistance like benzothiadiazole (BTH) and salicylic acid (SA) in rice (Xu et al., 2006) suggesting a role in disease response. Both of these genes were up-regulated in the mutant. However, there were other genes which were down-regulated in the *up* mutant, but are still attributed to a role in stress response. For example, jasmonate-ZIM domain proteins (JAZs) which mediate various jasmonate-regulated processes, including fertility, root growth, anthocyanin accumulation, senescence, and defense as negative regulators by

repressing JA responses were down-regulated in the mutant. JAZ proteins are identified as substrate of the SCFcol1 complex and interact with MYB21 and MYB 24 transcription factors (Song et al., 2011). Similarly, plant β-1,3-glucanases, involved in pathogen defense as well as a wide range of normal developmental processes like cell division and cell wall remodeling (Doxey et al., 2007) were also down-regulated in the mutant *underperformer*.

Interestingly, oxidative stress response genes were down-regulated in the mutant. Tau class GSTs participate in a broad network of catalytic and regulatory functions involved in the oxidative stress response. High GST levels were linked with increased resistance to adverse conditions and high yields in wheat and related species, and over-expression of a homologous BI-GST Tau class enzyme from tobacco was shown to enhance resistance to chilling and salt stress in seedlings (Kilili et al., 2004). All these genes just described were down-regulated to various levels in the *up* mutant. The gene RCI3 (Rare Cold Inducible gene 3) encodes an active cationic peroxidase, regulation occurs at the transcriptional level during plant development and is involved in the tolerance to water and salt stress in Arabidopsis (Llorente et al., 2002). The potato ortholog of RCI3 was also down-regulated in the mutant.

Contribution of enhancer elements in global gene activation/repression and mutant phenotype

As reported in Chapter 1, the first activated gene adjacent to the right border was a conserved gene of unknown function, whose closest orthologs in Arabidopsis (At1g04390) and tomato (Solyc10g005600.2.1) encode a BTB/POZ domain-containing protein. The BTB/POZ domain-containing proteins play important roles in various biological processes by regulating the transcriptional activities of downstream genes. In the present study, our RNA-seq data confirmed previous RT-PCR results (**Chapter 1**), indicating that this conserved gene of unknown function (peptide id: PGSC0003DMP400019889) was up-regulated in the mutant. The BTB/POZ protein encoded by this gene is expected to play an important role in mediating protein interactions, possibly affecting global repression of transcription factors (Hur et al., 2001; Liu et al., 2011), plant defense (Boyle et al., 2009; Qu et al., 2010), protein degradation (Yuasa et al., 2008) and cell death (Sadanandom et al., 2008). Our RNA-seq data indicated that our tentative

activation-tagged BTB/POZ domain-containing protein localized within the ubiquitin-dependent protein degradation pathway, and we also observed a significant up-regulation of many genes involved in the protein degradation pathway. As many BTB/POZ proteins are interaction partners with the Cullin component of the SCF-like E3 ubiquitin ligase complex (Krek, 2003; Pintard et al., 2004), these observations suggest that proteolysis is a key contributor to the overall phenotype of *underperformer*. A variety of plant proteins is targeted for degradation via this complex, including many associated with plant hormone signaling (Santner and Estelle, 2010). Hence, this may explain the molecular basis of the multiple phenotypic changes observed, as well as the apparent broad gene expression and metabolic pathway perturbations, in the *up* mutant.

Conclusions

The present study focused on gene expression comparisons, obtained through RNA-seq, between wild-type Bintje potato and the activation-tagged mutant, *underperformer*. The RNA-seq data not only allowed a study of differential gene expression, but through the use of MapMan software, we were able to identify numerous pathways that were modified in the mutant. Our analyses of the leaf transcriptome revealed that various metabolic pathways and biological processes were modified in the mutant, indicating that the single T-DNA activation tag insertion resulted in numerous pleiotropic effects. Several of the pathways and processes that were modified, such as the significant down-regulation of genes associated with cell division, cell cycle, DNA synthesis, transcription factors and secondary metabolism, made sense in light of the slow growing, dwarf and early senescence characteristics of the mutant. While we assessed gene expression in 60 day-old leaves, a more complete understanding of the true implications of the T-DNA insertion on the overall mutant phenotype will require RNA-seq comparisons in the future with wild-type Bintje, for other various tissues and developmental stages (e.g. roots, tubers, flower buds). In summary, RNA-seq appears to be an extremely promising technology for measuring mRNA expression and identifying differentially expressed genes. It also allowed us to start describing uncharacterized gene model predictions from the recently completed potato genome sequencing effort.

Methods

Plant growth and tissue sampling

Phenotyping was done on plants grown in walk-in growth chambers at the Department of Forest Resources and Environmental Conservation, Virginia Tech. These chambers were capable of mimicking the outdoor field growth conditions. Ten equal size (by weight) tubers of Bintje (average tuber weight 24 g) and *up* (average tuber weight 22 g) were planted in 30 cm diameter pots filled with commercial Miracle-Gro potting mix (Scotts Miracle-Gro, Marysville, OH). Standard plant growth practices were followed. Plants were grown under 16 h photoperiod; 22°C day/16°C night temperature and 350 µE $m^{-2}s^{-1}$ light intensity. Growing plants were supported by a plastic net and watered manually twice a week or as needed. Leaves samples were harvested 60 days after planting tubers (about 30 days after emergence). Single lateral leaves, 5-6 cm long and 2-3 cm broad, which were at the same developmental stage (not fully mature) were harvested per plant. Ten leaves, harvested from ten plants were pooled to form a composite sample for each genotype. Three composite samples (3 x 10 = 30 leaves) were harvested in total and frozen in liquid nitrogen and were stored at -80°C until needed.

Preparation of RNA

RNA was prepared by grinding tissue in liquid nitrogen and extracting the RNA using Trizol Reagent (Invitrogen; Carlsbad, CA, USA) following manufacturer's instructions followed by purification with RNA mini kit (Qiagen; Valencia, CA, USA). For extraction with Trizol Reagent, 1 g powder was added to 1 ml of pre-warmed (35-40°C) Trizol and mixed immediately by vortex to protect it from RNases. This mixture was incubated for 5 min at room temperature (RT), frequently mixed by vortex followed by centrifugation at 12,000 g for 10 min at 2-8°C. After removal of supernatant to a new tube, 0.2 ml chloroform was added and mixed by vortex for 15 sec. Tubes were incubated for 1 min at RT, mixed by vortex for 15 sec and incubated at RT for 2-3 min, followed by centrifugation at 12,000 g for 15 min at 2-8°C to separate the phases. The upper layer was removed and divided into two parts of 200 µl each (one part for backup was stored at -20°C). To another 200 µl from the top layer, 700 µl of Qiagen RLT buffer (after addition of 1 µl β-mercaptoethanol to 1 ml Buffer RLT) and 500 µl of 96-100% ethanol were added and mixed well by vortex. Samples were applied to Qiagen MinElute spin column and

washed with RPE buffer, followed by two washes with 750 µl of 80% ethanol (Sigma) to remove all guanidine salts that may inhibit downstream applications. RNA was eluted with 20 µl of RNAase free water. RNA concentration was measured using 1 µl of each RNA sample on the NanoDrop ND-1000 Spectrophotometer. For DNase treatment we used Turbo DNA free protocol following manufacturer's instruction with some modifications. In brief, each DNase treatment reaction was done in 50 µl volume using 12 µg of total RNA to which 5 µl buffer and 1µl DNase were added and mixture were incubated at 37°C for 30 min. After incubation 5 µl of inactivation buffer were added and mixed by vortex for 2 min. Samples were centrifuged at 13,000 rpm for 1 min and supernatant was transferred to RNAase free tubes. 50x elution of this sample were run on an Agilent 2100 Bioanalyzer (Agilent technology) to check RNA quality. Samples with an RNA Integrity Number (RIN) value > 8 were subjected to cDNA synthesis.

cDNA synthesis and Illumina library preparation

The Illumina TruSeq RNA sample preparation kit (low-throughput protocol) was used for poly-A based mRNA enrichment and thermal mRNA fragmentation followed by cDNA synthesis using 4 µg of total RNA sample as starting material for each enrichment reaction. The fragmented mRNA samples were used for cDNA synthesis using reverse transcriptase (Super-Script II) and random primers. The cDNA was further converted into double stranded DNA using the reagent supplied in the kit and resulting dsDNA was used to prepare a library of template molecules suitable for high-throughput DNA sequencing using the TruSeq Illumina kit.

mRNA-sequencing

Sequencing (mRNA-seq) was performed at the Research Technology Support Facility (RTSF) of Michigan State University using the Illumina Genome Analyzer (GAII) and HiSeq 2000 system, which uses a massively parallel sequencing-by-synthesis four-dye approach to generate billions of bases of high-quality DNA sequence per run. Briefly, randomly fragmented cDNA molecules were attached to an optically transparent surface flow cell and these attached fragments were extended and amplified to create an ultra-high density flow cell generating up to 1.5 billion clusters. Each of these clusters contained ~1,000 copies of the same template and was sequenced using a proprietary reversible terminator-based chemistry. A total of two biological

replicates each consisting of two technical replicates was sequenced. For one set of biological replicates we used the GAII platform generating single end 36 bp reads and for the second biological set we used the HiSeq 2000 machine generating single end 50 bp reads.

Bioinformatics and statistical analyzes

Sequence reads were mapped to *S. tuberosum* Group Phureja DM1-3 516-R44 (DM) superscaffolds using Tophat v2.0.4 (Trapnell et al., 2009), which made use of Bowtie. Reads were counted in regions of genes based on the PGSC_DM_v3_2.1.10_pseudomolecule annotation. Transcript abundance was expressed as raw read counts and also in a normalized unit, reads per kilobase per million reads (RPKM) as implemented in Cufflinks v2.0.2 (Trapnell et al., 2010), which allow comparison both within and among samples. Transcripts with RPKM value >0.001 and 95% confidence interval were considered for further analysis. Differentially expressed genes were flagged significant based on multiple comparisons values, q. Only genes with q values <0.05 were considered significant. Individual datasets (GAII or HiSeq) as well combined dataset (GA and Hiseq) were used for cuffdiff analyzes and the resulted output files were scanned for significant differentially expressed genes. A more strict statistical analysis was performed using differential expression analysis tool that is based on negative binomial distribution (Anders and Huber, 2010) and genes which had significantly different expression level at $\alpha < 0.05$ level were considered. The Hierarchical clustering of genes filtered using Cuffdiff analysis was done using MapMan v3.5.1 software (Thimm et al., 2004). Protein domain architecture analysis was done with Simple Modular Architecture Research Tool (SMART).

Visualization of gene expression data

We used MapMan (version 3.5.1), which displays large data sets onto pictorial diagrams (Maps) that symbolically depict areas of biological function. Maps are diagrams of pathways or processes in which experimental data are displayed. MapMan organizes data into functional categories (BINs) which are themselves split into subcategories (subBINs). Each Bin was given a corresponding numerical code (e.g., 'photosynthesis'=1) which can be extended in a hierarchical manner (e.g., the subBINs 'light reactions'= 1.1; 'photorespiration' = 1.2) and so on (Thimm et al., 2004). The change of expression is displayed via false color code: blue for

increased and red for decreased expression. A setting was selected in which it saturates at a value of 2 (2^2 = a fourfold change). Potato mapping file Stub_PGSC_DM_v3.4 was downloaded from the MapMan website http://mapman.gabipd.org/web/guest/mapmanstore. The mapping file represents the file in which the association of genes, metabolites or proteins to MapMan BINs is stored. This was a linking file that links the experimental data with the MapMan ontology. An experimental data file with potato peptide IDs (PGSC0003DMP4000xxxxx) and relative expression levels (log2fold values) was created and uploaded in the MapMan dataset folder. The experimental data file contained the measurements from the RNA-seq experiment in the log2 fold changes between a treatment (mutant *up*) and control (Bintje). MapMan not only displays a large dataset but also incorporates statistical analysis. For example deviation of response of items in a particular BIN from the response of all other items is calculated by Wilcoxon test, which is similar to Student's t-test. MapMan also has a built-in multiple comparison test (Benjamini Hochberg) which calculates corrected P values (Usadel et al., 2009).

Acknowledgments

We want to thank Robin Buell and Brieanne Vaillancourt, department of Plant Biology, Michigan State University for assistance with mRNA-sequencing. This work was funded through Special Grants (2003-38891-02112, 2008-38891-19353 and 2009-38891-20092) from the United States Department of Agriculture, and operating funds from the Commonwealth of Virginia to the Institute for Advanced Learning and Research.

Table 2.1 Selected papers recently published on plant RNA-sequencing.

Plant species	Cell/tissue type	NGS platform	Reference
Malus sp. (Apple)	Shoots	Illumina HiSeq 2000	(Zhang et al., 2012)
Brassica napus (Oilseed rape)	Leaves	Illumina GAIIx	(Bancroft et al., 2011)
Miscanthus sinensis (Maiden grass)		Illumina GAII	(Swaminathan et al., 2012)
Arabidopsis thaliana Col-0 ecotype (Arabidopsis)	Meiocytes, anthers and seedlings	Illumina GA	(Chen et al., 2010)
Beta vulgaris sp. vulgaris (Sugarbeet)	Shoot apex	Illumina HiSeq 2000	(Mutasa-Goettgens et al., 2012)
Triticum x *Secale* (Triticale)	Root, leaf, stem, flower	Roche-454	(Xu et al., 2011)
Oryza sativa ssp. japonica (Rice)	Callus, leaf, root, panicle before flowering, panicle after flowering, seed, and shoot	Illumina GAIIx	(Sakai et al., 2011)
Solanum lycopersicum and S. *habrochaites* (Tomato)	Trichomes	Roche-454 GS20	(Bleeker et al., 2011)
Cucumis sativus var. sativus line 9930 (Cucumber)	Root, stem, leaf, male and flower, fertilized and unfertilized ovary basal part of tendril and tendril	Illumina GAII	(Li et al., 2011)
Panicum hallii (Swithgrass)	Leaf, inflorescence (developing fruits, florets, spikelets and panicle material), stem, node, crown, root and seeds.	Roche-454 FLX Titanium SOLiD	(Meyer et al., 2012)
Solanum tuberosum Group Phureja clone DM1-3 516R44 (Potato)	Leaves, petioles, stolons, tubers, flower, fruit (berries), shoots and roots (in vitro-grown plants), biotic stress, abiotic (heat , salt, osmotic) and hormonal stress conditions	Illumina GAII	(Massa et al., 2011)

Table 2.2 Summary of mRNA-seq reads and differently expressed genes in potato leaves harvested at 60 days after planting (DAP) in two biological replicates.

NGS platform [Read length]	Genotype [Biological replicate]	Number of mRNA-Seq reads[#]	Number of mapped reads[§] [% age reads mapped]	Number of genes with significantly different expression[*]
Illumina GAII [36 bp SE]	Bintje [1]	28,882,858	24,284,175 [84.1%]	371
	up [1]	28,372,335	23,883,511 [84.2%]	
Illumina HiSeq 2000 [50 bp SE]	Bintje [2]	79,641,059	71,430,878 [89.7%]	209
	up [2]	70,376,313	63,126,457 [89.7%]	
After combining data from both (GAII + HiSeq 2000) platforms				1632

SE - single end; bp - base pair.

up - *underperformer* mutant.

[#]read numbers are total of two technical replicates.

[§]RNA-Seq reads mapped to the potato genome assembly, PGSC_DM_v3_2.1.10_pseudomolecules.zip

[*]based on q values < 0.05.

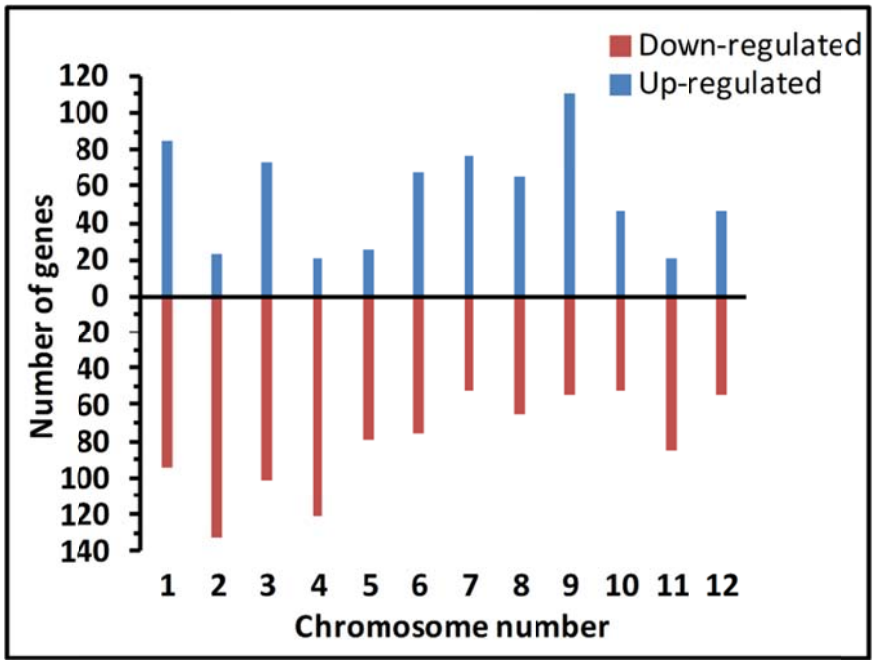


Figure 2.1 Chromosomal view of differential gene expression pattern. 1632 significant differentially expressed genes from combined RNA-seq datasets (GAII + HiSeq) were grouped according to their location on the potato chromosomes. Leaf RNA from 60 day old plants was used for RNA-sequencing. Genes up-and down-regulated in mutant *underperformer* were presented relative to wild-type Bintje. Significant cutoff was based on $q < 0.05$ values.

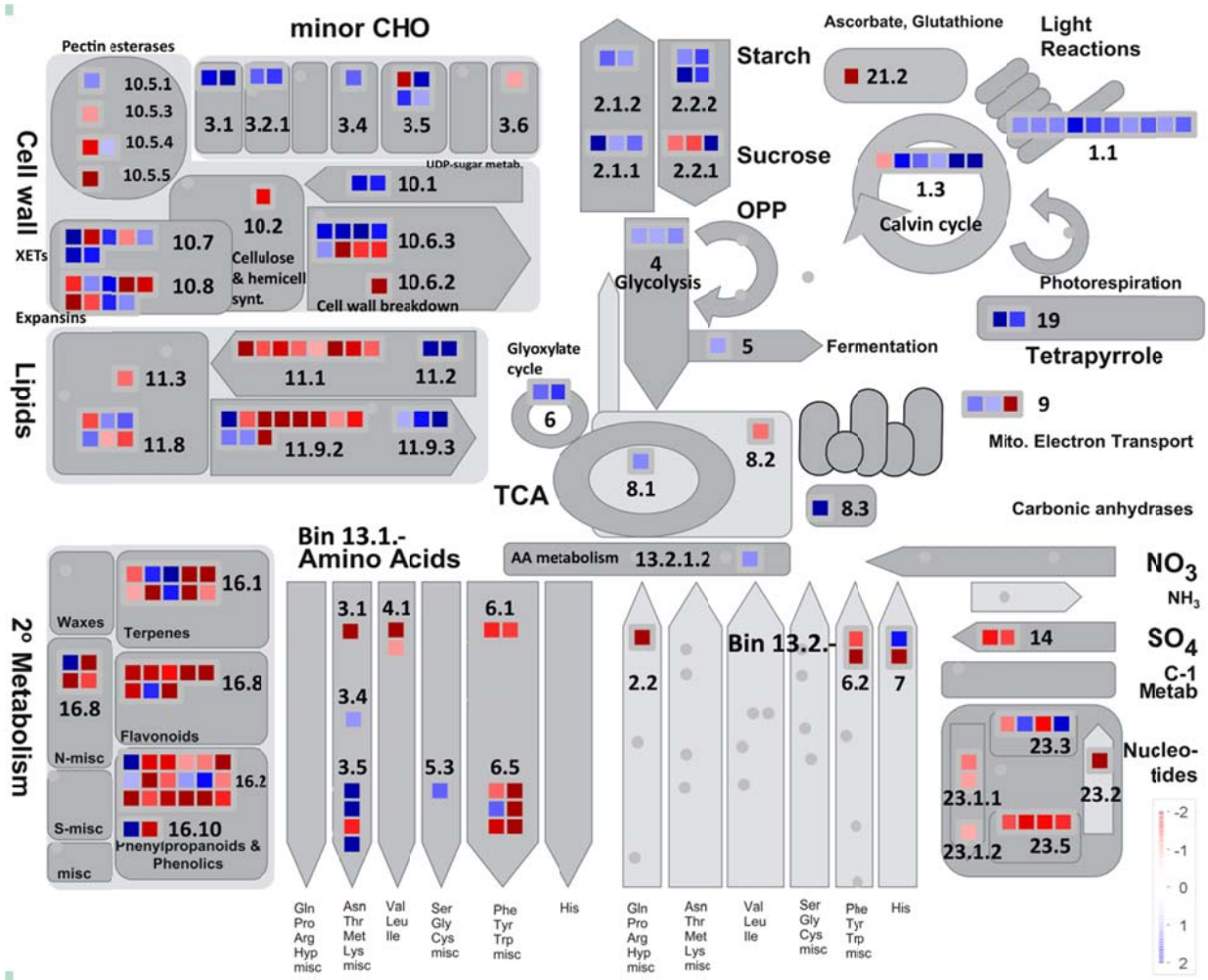


Figure 2.2 Overview display of differentially-expressed genes assigned to metabolism. Changes in transcript levels were measured by RNA-sequencing of leaves harvested from activation tagged mutant *underperformer* (*up*) and wild-type Bintje plants at 60 days after planting (DAP). Transcript abundance was expressed in Reads per kilobase of exon model per million mapped reads (RPKM) and converted to relative (*up*/Bintje) log2 fold values. Relative log2 fold values were visualized using MapMan software. The change in expression levels were displayed via false color code: blue for increased and red for decreased expression in the mutant. Each square represent one gene (peptide). A color setting was selected in which it saturates at a value of 2 (2^2 = a fourfold change). Genes with significantly ($q < 0.05$) differential expression levels were displayed. Number near each cluster corresponds to MapMan BIN ids. Grey dots were genes not present in our dataset. Detailed information for each data point including Bin ids, Bin description, peptide ids, gene annotation and log2 fold values are given in supplementary **Table S2.1**.

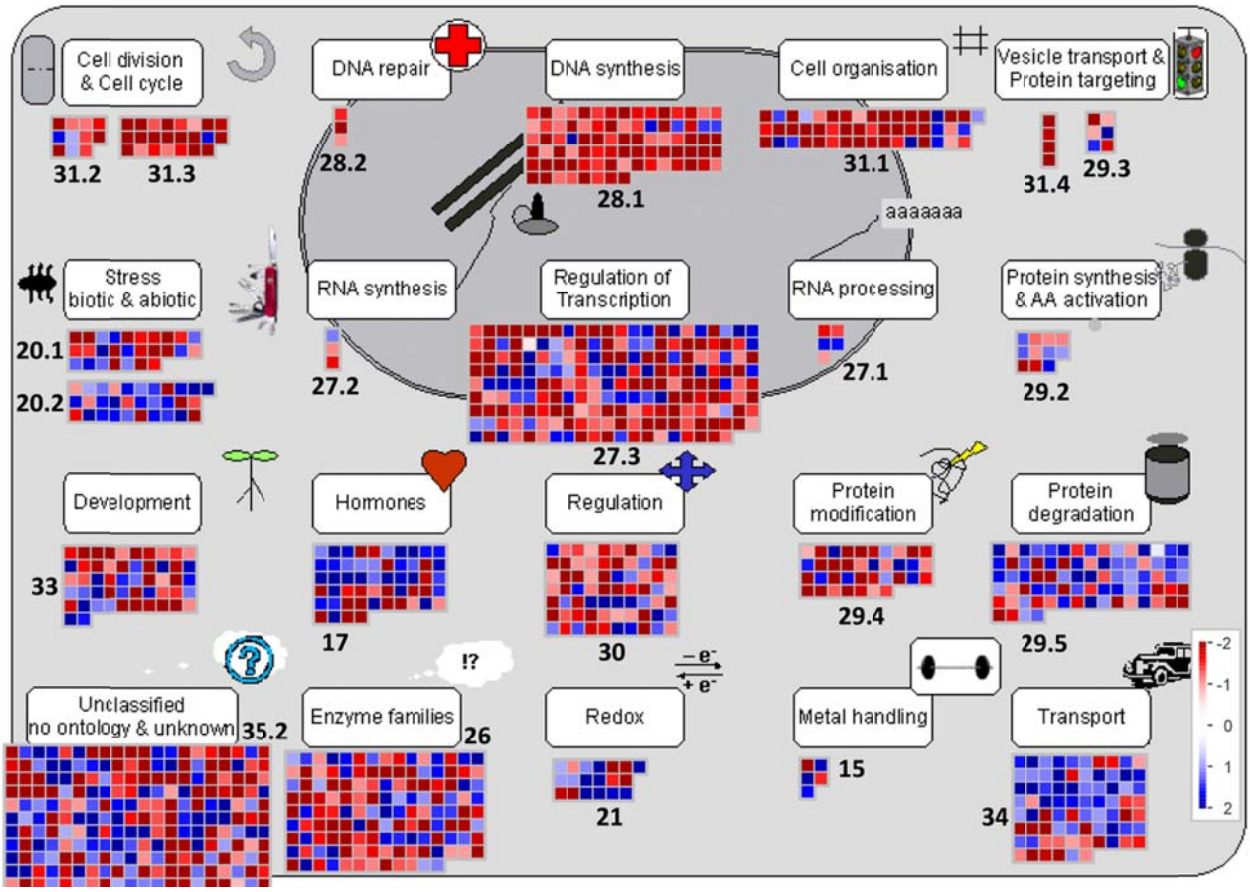


Figure 2.3 Overview display of differentially-expressed genes assigned to cell functions. Changes in transcript levels were measured by RNA-sequencing of leaves harvested from activation tagged mutant *underperformer* (*up*) and wild-type Bintje plants at 60 days after planting (DAP). Transcript abundance was expressed in Reads per kilobase of exon model per million mapped reads (RPKM) and converted to relative (*up*/Bintje) log2 fold values. Relative log2 fold values were visualized using MapMan software. The change in expression levels were displayed via false color code: blue for increased and red for decreased expression in the mutant. Each square represent one gene (peptide). A color setting was selected in which it saturates at a value of 2 (2^2 = a fourfold change). Genes with significantly ($q < 0.05$) differential expression levels were displayed. Number near each cluster corresponds to MapMan BIN ids. Grey dots were genes not present in our dataset. Detailed information for each data point including Bin ids, Bin description, peptide ids, gene annotation and log2 fold values are given in supplementary **Table S2.1**.

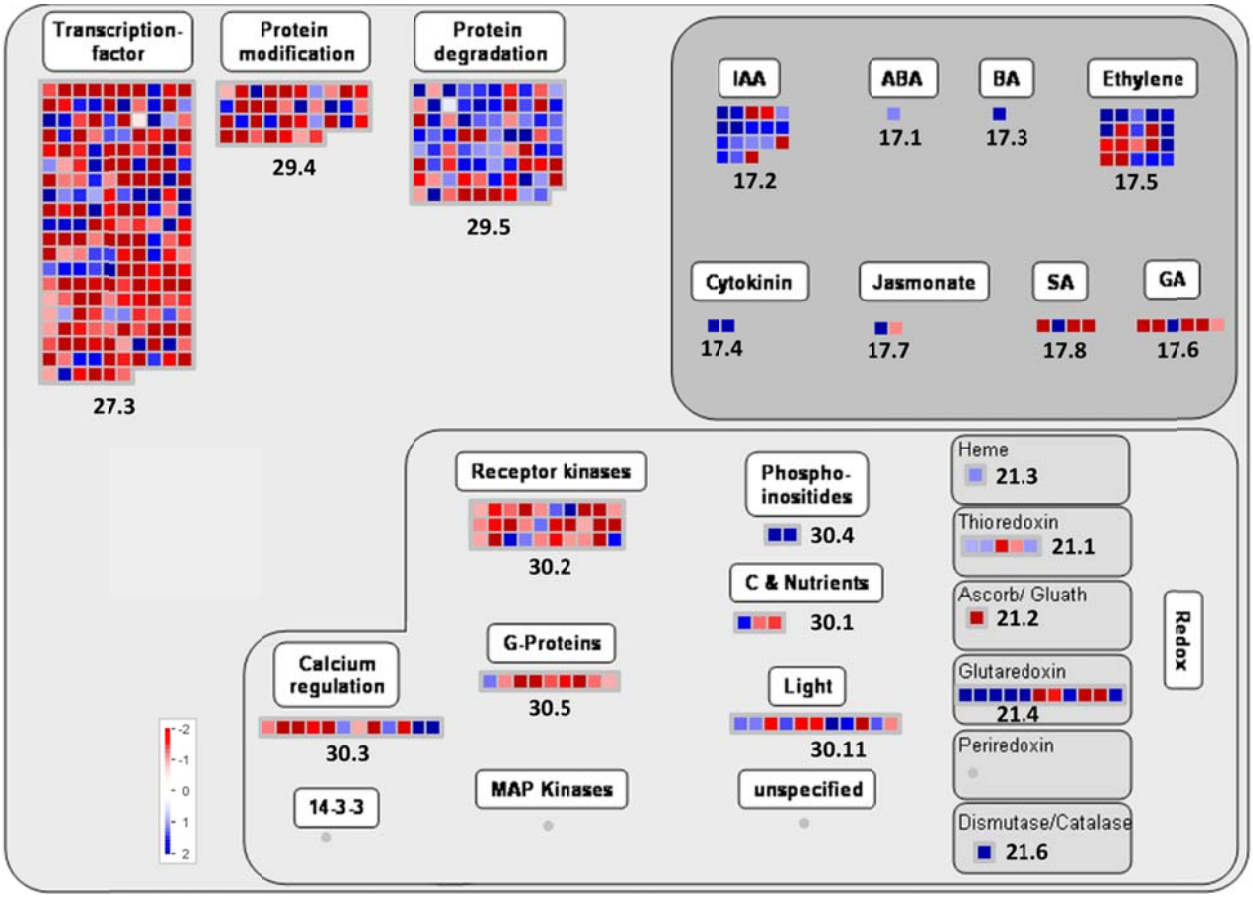


Figure 2.4 Overview display of differentially-expressed genes assigned to regulation pathway. Changes in transcript levels were measured by RNA-sequencing of leaves harvested from activation tagged mutant *underperformer* (*up*) and wild-type Bintje plants at 60 days after planting (DAP). Transcript abundance was expressed in Reads per kilobase of exon model per million mapped reads (RPKM) and converted to relative (*up*/Bintje) log2 fold values. Relative log2 fold values were visualized using MapMan software. The change in expression levels were displayed via false color code: blue for increased and red for decreased expression in the mutant. Each square represent one gene (peptide). A color setting was selected in which it saturates at a value of 2 (2^2 = a fourfold change). Genes with significantly ($q < 0.05$) differential expression levels were displayed. Number near each cluster corresponds to MapMan BIN ids. Grey dots were genes not present in our dataset. Detailed information for each data point including Bin ids, Bin description, peptide ids, gene annotation and log2 fold values are given in supplementary **Table S2.1**.

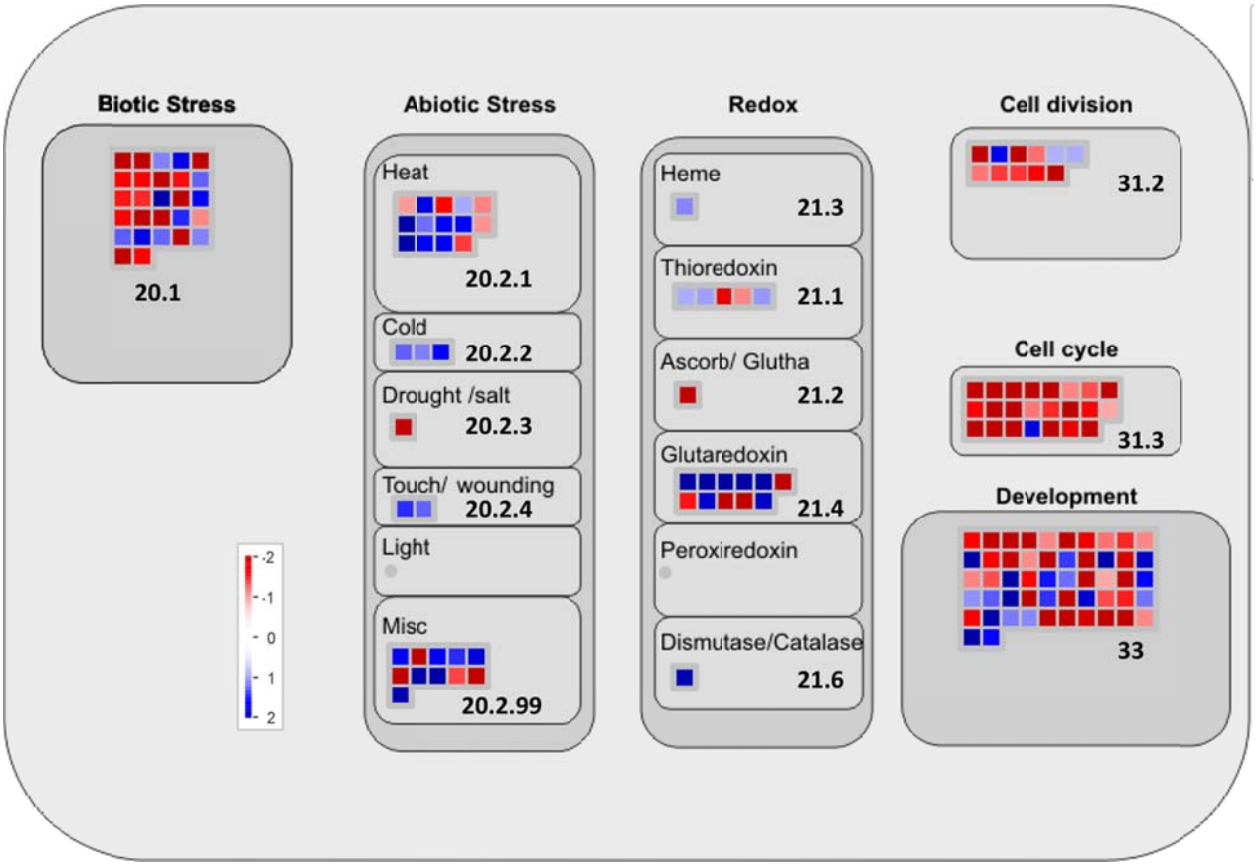


Figure 2.5 Overview display of differentially-expressed genes assigned to cellular responses. Changes in transcript levels were measured by RNA-sequencing of leaves harvested from activation tagged mutant *underperformer* (*up*) and wild-type Bintje plants at 60 days after planting (DAP). Transcript abundance was expressed in Reads per kilobase of exon model per million mapped reads (RPKM) and converted to relative (*up*/Bintje) log2 fold values. Relative log2 fold values were visualized using MapMan software. The change in expression levels were displayed via false color code: blue for increased and red for decreased expression in the mutant. Each square represent one gene (peptide). A color setting was selected in which it saturates at a value of 2 (2^2 = a fourfold change). Genes with significantly ($q < 0.05$) differential expression levels were displayed. Number near each cluster corresponds to MapMan BIN ids. Grey dots were genes not present in our dataset. Detailed information for each data point including Bin ids, Bin description, peptide ids, gene annotation and log2 fold values are given in supplementary **Table S2.1**.

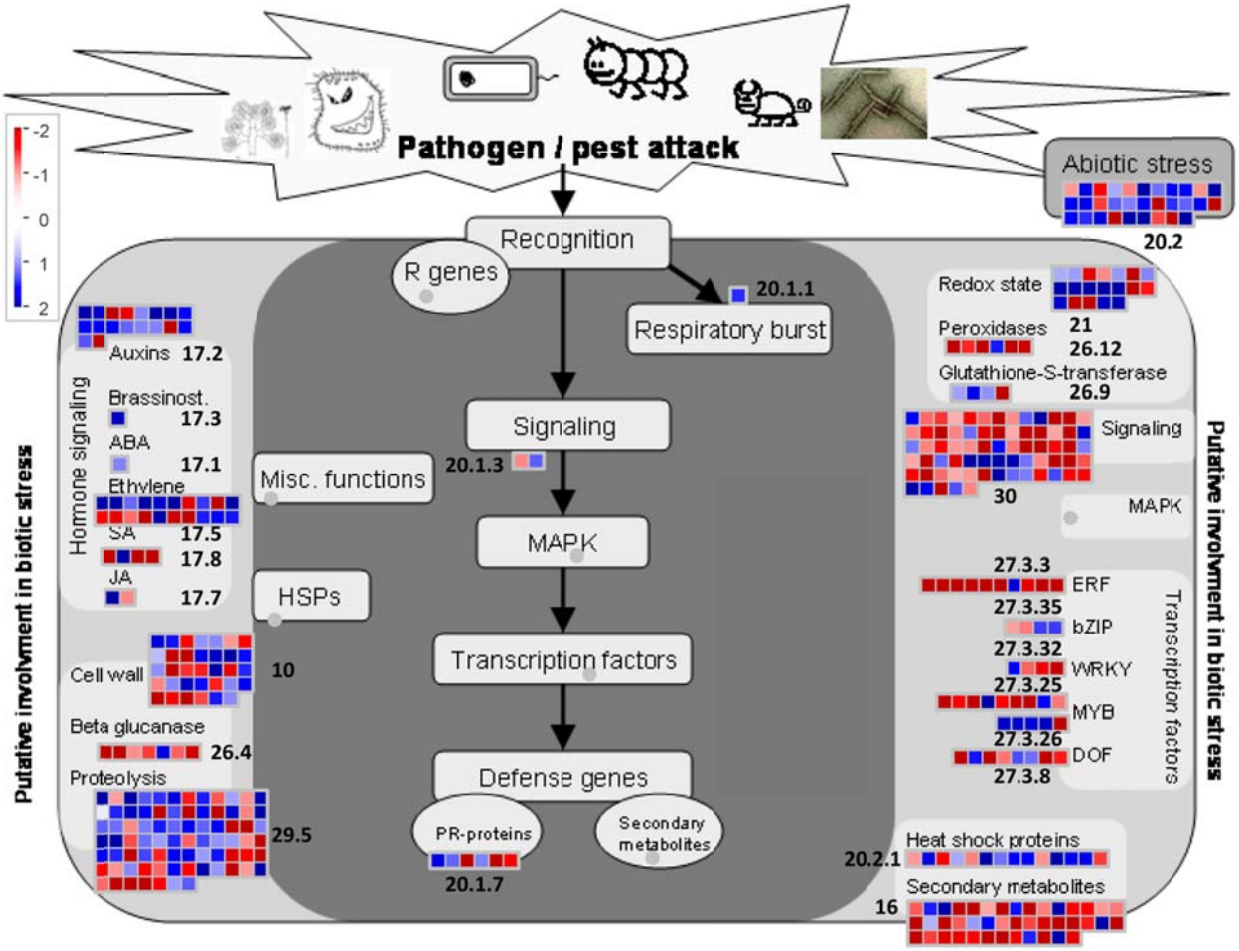


Figure 2.6 Overview display of differentially-expressed genes assigned to stress response pathways. Changes in transcript levels were measured by RNA-sequencing of leaves harvested from activation tagged mutant *underperformer* (*up*) and wild-type Bintje plants at 60 days after planting (DAP). Transcript abundance was expressed in Reads per kilobase of exon model per million mapped reads (RPKM) and converted to relative (*up*/Bintje) log2 fold values. Relative log2 fold values were visualized using MapMan software. The change in expression levels were displayed via false color code: blue for increased and red for decreased expression in the mutant. Each square represent one gene (peptide). A color setting was selected in which it saturates at a value of 2 (2^2 = a fourfold change). Genes with significantly ($q < 0.05$) differential expression levels were displayed. Number near each cluster corresponds to MapMan BIN ids. Grey dots were genes not present in our dataset. Detailed information for each data point including Bin ids, Bin description, peptide ids, gene annotation and log2 fold values are given in supplementary **Table S2.1**.

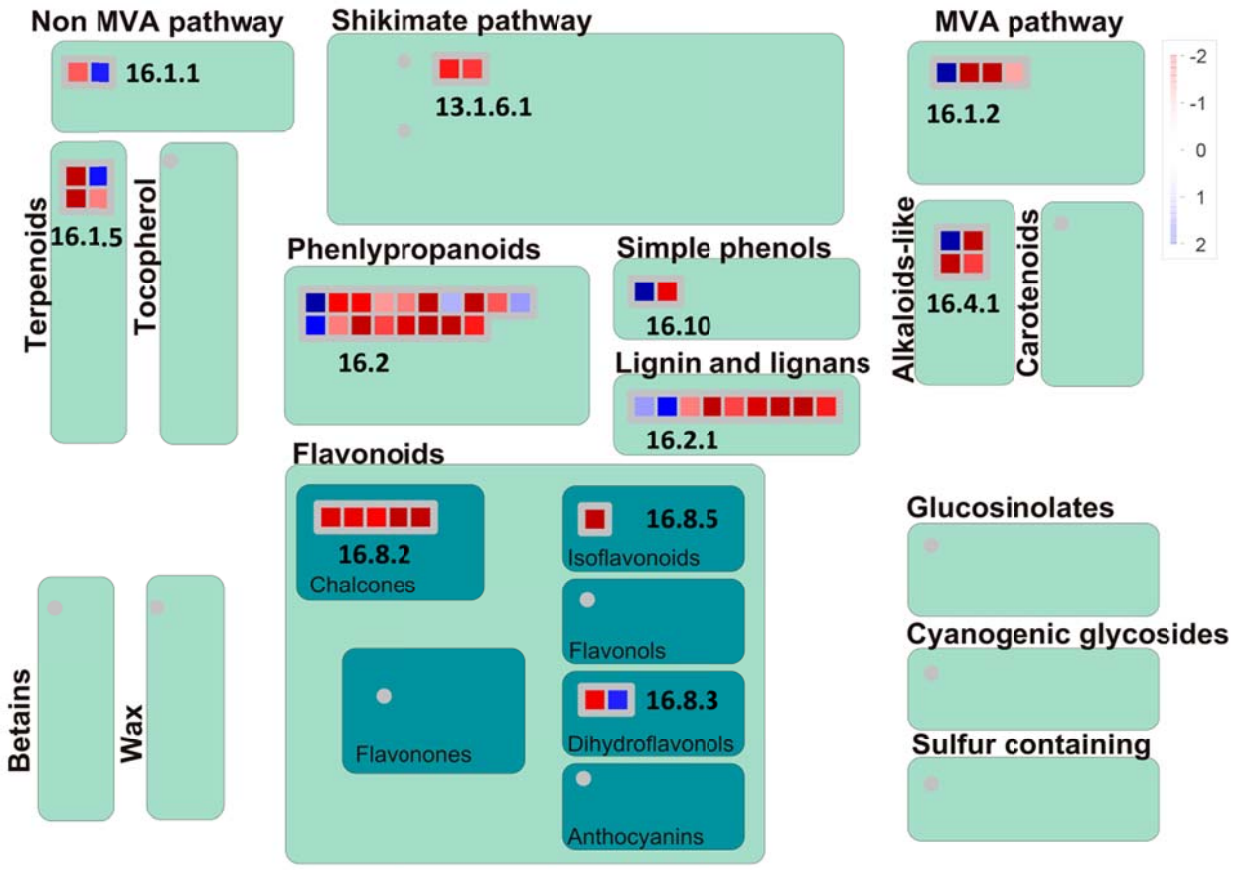


Figure 2.7 Overview display of differentially-expressed genes assigned to secondary metabolism. Changes in transcript levels were measured by RNA-sequencing of leaves harvested from activation tagged mutant *underperformer* (*up*) and wild-type Bintje plants at 60 days after planting (DAP). Transcript abundance was expressed in Reads per kilobase of exon model per million mapped reads (RPKM) and converted to relative (*up*/Bintje) log2 fold values. Relative log2 fold values were visualized using MapMan software. The change in expression levels were displayed via false color code: blue for increased and red for decreased expression in the mutant. Each square represent one gene (peptide). A color setting was selected in which it saturates at a value of 2 (2^2 = a fourfold change). Genes with significantly ($q < 0.05$) differential expression levels were displayed. Number near each cluster corresponds to MapMan BIN ids. Grey dots were genes not present in our dataset. Detailed information for each data point including Bin ids, Bin description, peptide ids, gene annotation and log2 fold values are given in supplementary **Table S2.1**.

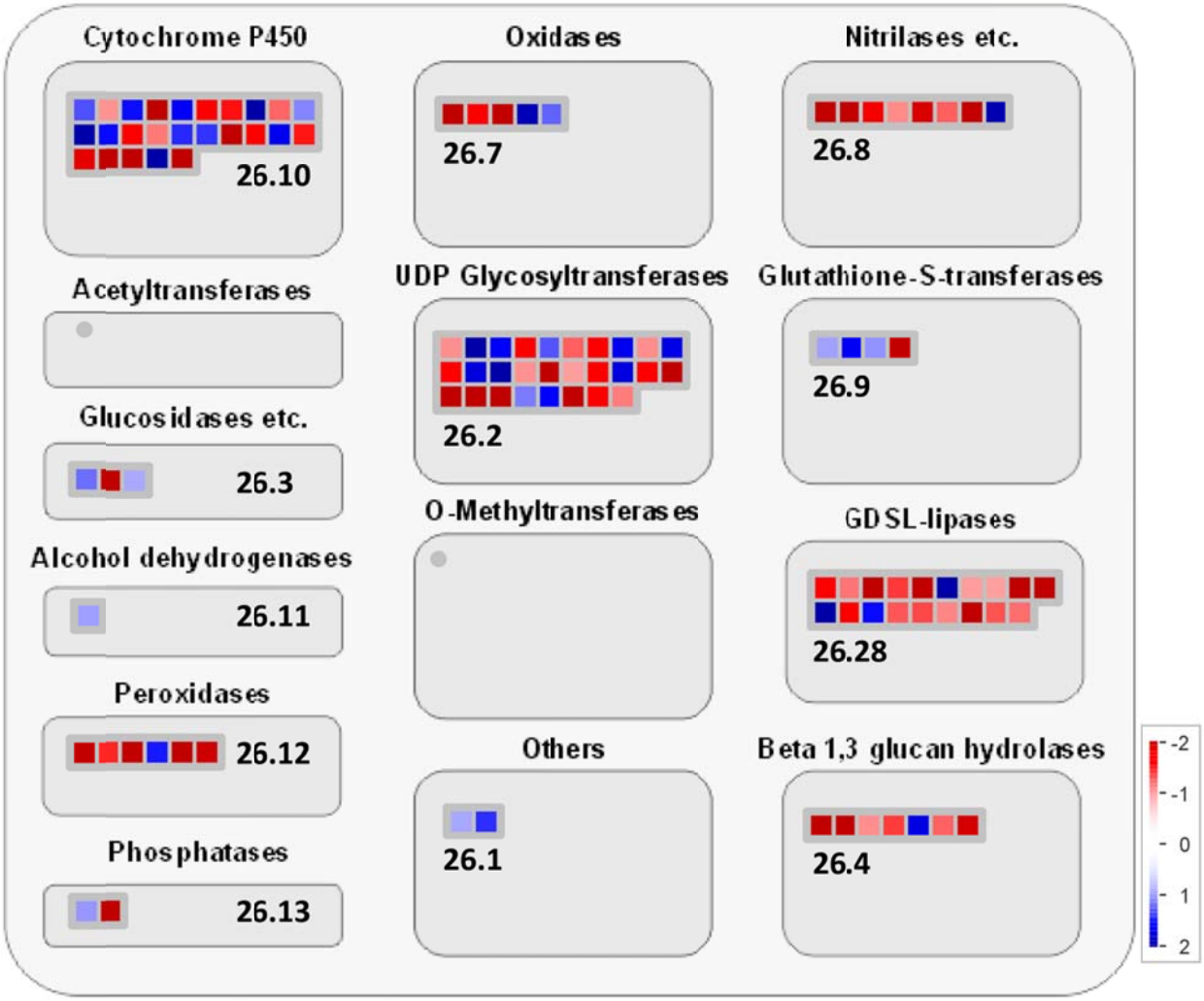


Figure 2.8 Overview display of differentially-expressed genes assigned to large gene families. Changes in transcript levels were measured by RNA-sequencing of leaves harvested from activation tagged mutant *underperformer* (*up*) and wild-type Bintje plants at 60 days after planting (DAP). Transcript abundance was expressed in Reads per kilobase of exon model per million mapped reads (RPKM) and converted to relative (*up*/Bintje) log2 fold values. Relative log2 fold values were visualized using MapMan software. The change in expression levels were displayed via false color code: blue for increased and red for decreased expression in the mutant. Each square represent one gene (peptide). A color setting was selected in which it saturates at a value of 2 (2^2 = a fourfold change). Genes with significantly ($q < 0.05$) differential expression levels were displayed. Number near each cluster corresponds to MapMan BIN ids. Grey dots were genes not present in our dataset. Detailed information for each data point including Bin ids, Bin description, peptide ids, gene annotation and log2 fold values are given in supplementary **Table S2.1**.

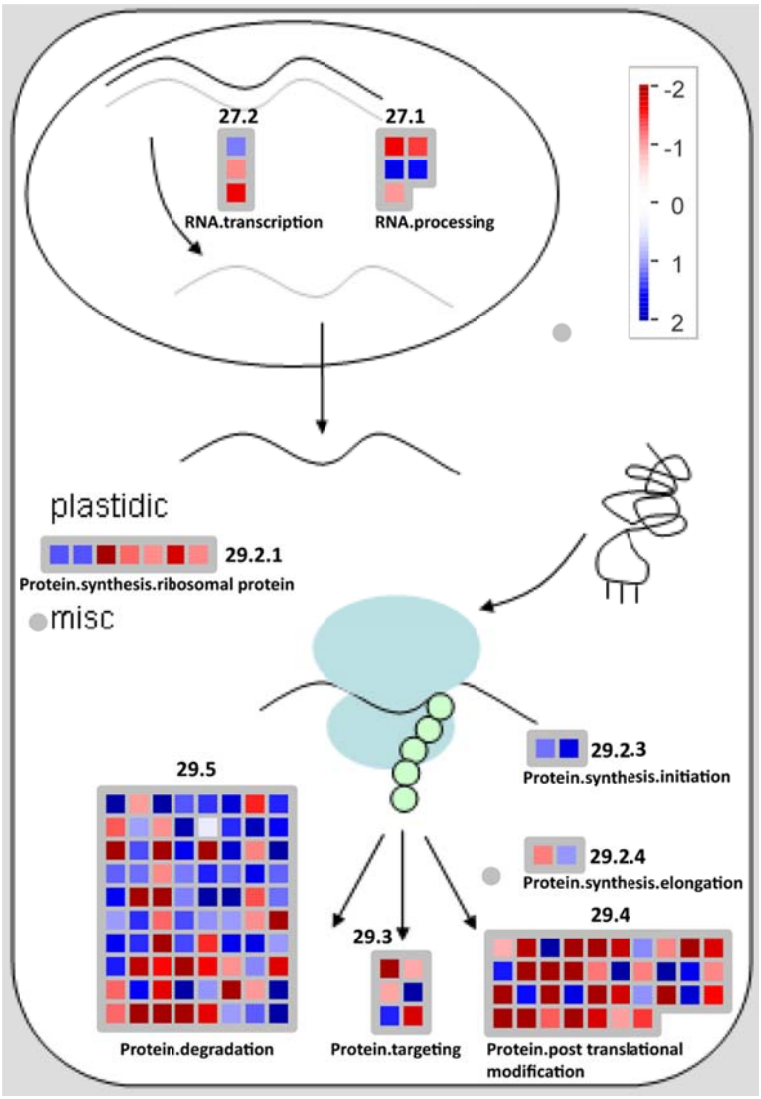


Figure 2.9 Overview display of differentially-expressed genes assigned to RNA-protein synthesis pathway. Changes in transcript levels were measured by RNA-sequencing of leaves harvested from activation tagged mutant *underperformer* (*up*) and wild-type Bintje plants at 60 days after planting (DAP). Transcript abundance was expressed in Reads per kilobase of exon model per million mapped reads (RPKM) and converted to relative (*up*/Bintje) log2 fold values. Relative log2 fold values were visualized using MapMan software. The change in expression levels were displayed via false color code: blue for increased and red for decreased expression in the mutant. Each square represent one gene (peptide). A color setting was selected in which it saturates at a value of 2 (2^2 = a fourfold change). Genes with significantly ($q < 0.05$) differential expression levels were displayed. Number near each cluster corresponds to MapMan BIN ids. Grey dots were genes not present in our dataset. Detailed information for each data point including Bin ids, Bin description, peptide ids, gene annotation and log2 fold values are given in supplementary **Table S2.1**.

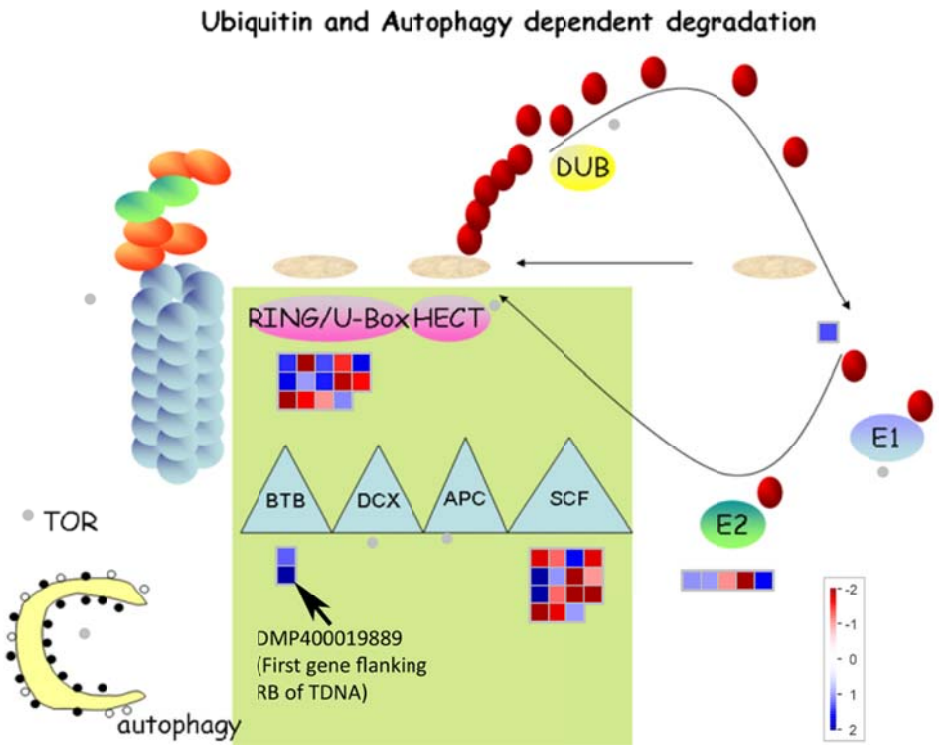


Figure 2.10 Overview display of differentially-expressed genes assigned to Ubiquitin and autophagy dependent degradation. Changes in transcript levels were measured by RNA-sequencing of leaves harvested from activation tagged mutant *underperformer* (*up*) and wild-type Bintje plants at 60 days after planting (DAP). Transcript abundance was expressed in Reads per kilobase of exon model per million mapped reads (RPKM) and converted to relative (*up*/Bintje) log2 fold values. Relative log2 fold values were visualized using MapMan software. The change in expression levels were displayed via false color code: blue for increased and red for decreased expression in the mutant. Each square represent one gene (peptide). A color setting was selected in which it saturates at a value of 2 (2^2 = a fourfold change). Genes with significantly ($q < 0.05$) differential expression levels were displayed. Number near each cluster corresponds to MapMan BIN ids. Grey dots were genes not present in our dataset. Detailed information for each data point including Bin ids, Bin description, peptide ids, gene annotation and log2 fold values are given in supplementary **Table S2.1**.

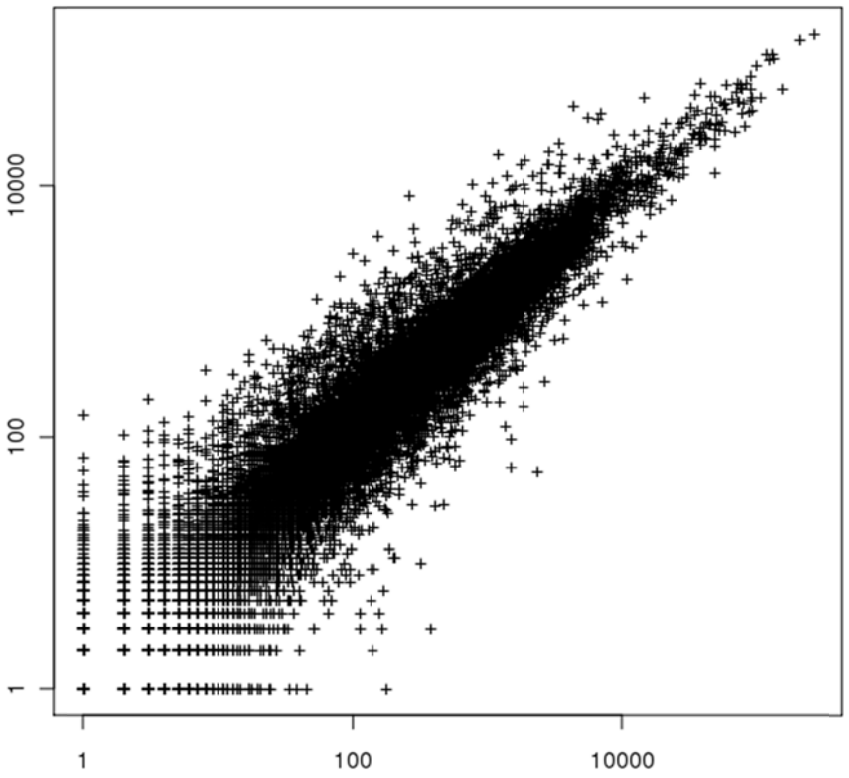


Figure S2.1 Log-log plot of Illumina GAII reads per gene of Bintje and *underperformer* (*up*) mutant. Reads on X-axis are of *up* mutant, and those on Y-axis are of Bintje.

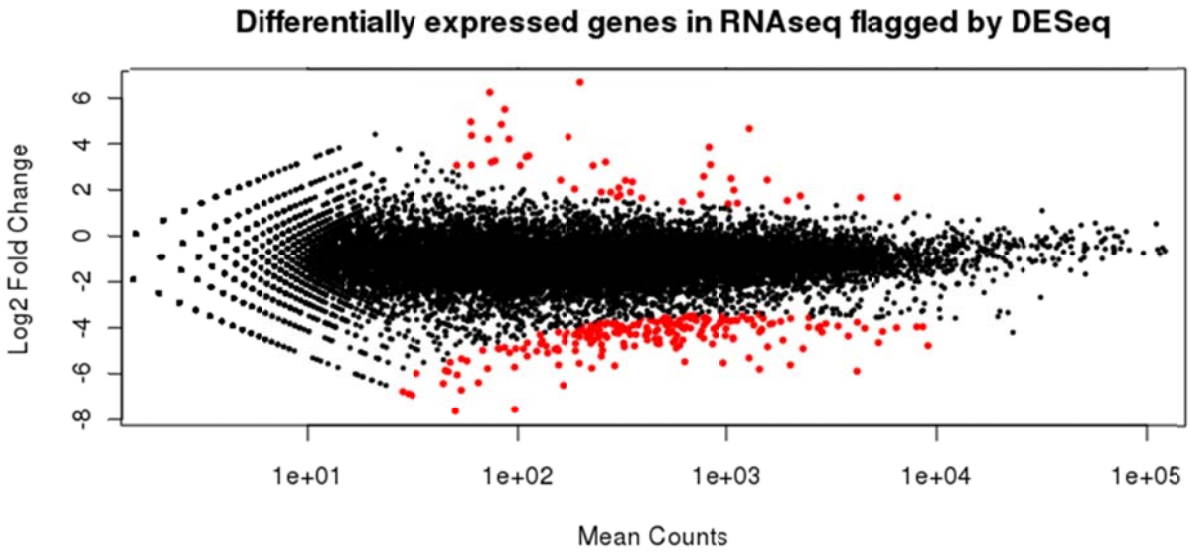


Figure S2.2 Differentially expressed genes from Illumina GAII reads comparing activation tagged mutant *underperformer* versus Bintje potato using the DESeq program of the Bioconductor package. Genes with read counts differing at the 0.05 alpha levels were marked in red. Total of 50 genes (upper red dots) were up-regulated and 217 (lower red dots) down-regulated in the mutant.

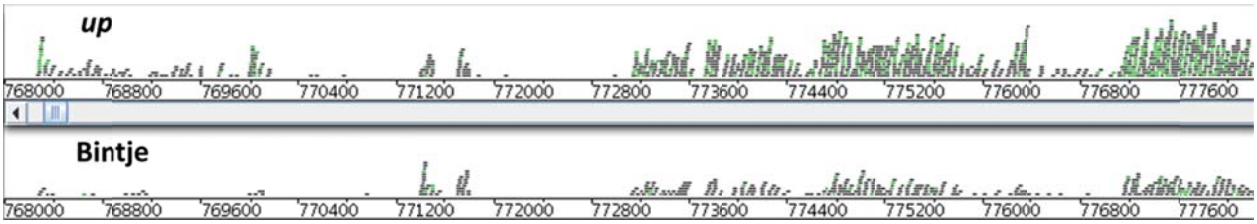


Figure S2.3 BamView visualization of GAII reads along PGSC_DM_v3_2.1.10 genome chromosome 10 covering the regions of annotated genes PGSC0003DMG401011239 and PGSC0003DMG402011239. These two gene models correspond to a BTB/POZ domain-containing protein (single ORF), which was also the first activated gene in the mutant flanking the right border of T-DNA. Upper panel, mutant *underperformer* (*up*) reads and lower panel Bintje reads.

References

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biology 11:R106.

Boyle, P., et al. (2009). The BTB/POZ domain of the *Arabidopsis* disease resistance protein NPR1 interacts with the repression domain of TGA2 to negate its function. Plant Cell 21:3700-3713.

Choi, J., Choi, D., Lee, S., Ryu, C.M., and Hwang, I. (2011). Cytokinins and plant immunity: old foes or new friends? Trends Plant Sci. 16:388-394.

Cloonan, N., and Grimmond, S.M. (2008). Transcriptome content and dynamics at single-nucleotide resolution. Genome Biology 9.

Costa, V., Angelini, C., De Feis, I., and Ciccodicola, A. (2010). Uncovering the complexity of transcriptomes with RNA-Seq. J. Biomed. Biotechnol.

Doxey, A.C., Yaish, M.W.F., Moffatt, B.A., Griffith, M., and McConkey, B.J. (2007). Functional divergence in the *Arabidopsis* beta-1,3-glucanase gene family inferred by phylogenetic reconstruction of expression states. Mol. Biol. Evol. 24:1045-1055.

Hong, D., et al. (2012). FX: an RNA-Seq analysis tool on the cloud. Bioinformatics 28:721-723.

Howe, E.A., Sinha, R., Schlauch, D., and Quackenbush, J. (2011). RNA-Seq analysis in MeV. Bioinformatics 27:3209-3210.

Hur, M.W., Lee, D.K., and Kwon, H.S. (2001). Molecular mechanism of transcription repression by HDAC and BTB/POZ-domain transcription factors. FASEB J. 15:A517-A517.

Jain, M. (2011). A next-generation approach to the characterization of a non-model plant transcriptome. Curr. Sci. 101:1435-1439.

Kilili, K.G., et al. (2004). Differential roles of Tau class glutathione S-transferases in oxidative stress. J. Biol. Chem. 279:24540-24551.

Krek, W. (2003). BTB proteins as henchmen of Cul3-based ubiquitin ligases. Nat. Cell Biol. 5:950-951.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12.

Liu, Q., et al. (2011). Novel human BTB/POZ domain-containing zinc finger protein ZBTB1 inhibits transcriptional activities of CRE. Mol. Cell. Biochem. 357:405-414.

Llorente, F., Lopez-Cobollo, R.M., Catala, R., Martinez-Zapater, J.M., and Salinas, J. (2002). A novel cold-inducible gene from *Arabidopsis*, *RCI3*, encodes a peroxidase that constitutes a component for stress tolerance. Plant Journal, 32 (1) pp. 13-24, 2002.

Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. (2008). RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 18:1509-1517.

Mei, C., Qi, M., Sheng, G., and Yang, Y. (2006). Inducible overexpression of a rice allene oxide synthase gene increases the endogenous jasmonic acid level, *PR* gene expression, and host resistance to fungal infection. Mol. Plant-Microbe Interact. 19:1127-1137.

Minoche, A.E., Dohm, J.C., and Himmelbauer, H. (2011). Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. Genome Biology 12.

Nakazawa, M., et al. (2001). *DFL1*, an auxin-responsive *GH3* gene homologue, negatively regulates shoot cell elongation and lateral root formation, and positively regulates the light response of hypocotyl length. Plant J. 25:213-221.

Pintard, L., Willems, A., and Peter, M. (2004). Cullin-based ubiquitin ligases: Cul3-BTB complexes join the family. EMBO J. 23:1681-1687.

Qin, G.J., et al. (2005). An Indole-3-Acetic Acid Carboxyl Methyltransferase regulates *Arabidopsis* leaf development. Plant Cell 17:2693-2704.

Qu, N., Gan, W., Bi, D., Xia, S., Li, X., and Zhang, Y. (2010). Two BTB proteins function redundantly as negative regulators of defense against pathogens in *Arabidopsis*. Botany-Botanique 88:953-960.

Sadanandom, A., Mesmar, J., Yang, C., Ewan, R., Carr, C., and O' Donnell, E. (2008). The cell death regulator *At*PUB17 directly interacts with the BTB/POZ domain transcriptional repressor, *At*BTB1 to control disease resistance in plants. Comp. Biochem. Physiol., A: Mol. Integr. Physiol. 150:S178-S179.

Salzman, J., Jiang, H., and Wong, W.H. (2011). Statistical modeling of RNA-Seq data. Statistical Science 26:62-83.

Santner, A., and Estelle, M. (2010). The ubiquitin-proteasome system regulates plant hormone signaling. Plant J. 61:1029-1040.

Schaeffer, A., Bronner, R., Benveniste, P., and Schaller, H. (2001). The ratio of campesterol to sitosterol that modulates growth in *Arabidopsis* is controlled by *STEROL METHYLTRANSFERASE 2;1*. Plant J. 25:605-615.

Schmulling, T., Werner, T., Riefler, M., Krupkova, E., and Manns, I.B.Y. (2003). Structure and function of cytokinin oxidase/dehydrogenase genes of maize, rice, *Arabidopsis* and other species. J. Plant Res. 116:241-252.

Song, S., et al. (2011). The Jasmonate-ZIM domain proteins interact with the *R2R3-MYB* transcription factors *MYB21* and *MYB24* to affect jasmonate-regulated stamen development in *Arabidopsis*. Plant Cell 23:1000-1013.

The Potato Genome Sequencing Consortium. (2011). Genome sequence and analysis of the tuber crop potato. Nature 475:189-U194.

Thimm, O., et al. (2004). MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J. 37:914-939.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105-1111.

Trapnell, C., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28:511-U174.

Usadel, B., Poree, F., Nagel, A., Lohse, M., Czedik-Eysenberg, A., and Stitt, M. (2009). A guide to using MapMan to visualize and compare omics data in plants: a case study in the crop species, maize. Plant Cell and Environment 32:1211-1229.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10:57-63.

Xu, R., Song, F., and Zheng, Z. (2006). *OsBISAMT1*, a gene encoding S-adenosyl-L-methionine : salicylic acid carboxyl methyltransferase, is differentially expressed in rice defense responses. Mol. Biol. Rep. 33:223-231.

Yuasa, S., et al. (2008). The Novel BTB/POZ protein, Rhobtb3, regulates mitochondrial function via the ubiquitin proteasome system. Circul. Res. 103:E53-E54.
Zytnicki, M., and Quesneville, H. (2011). S-MART, a software toolbox to aid RNA-seq data analysis. Plos One 6.

Chapter 3

Characterization of *nikku*, an activation tagged potato mutant

Sukhwinder S. Aulakh[1, 2], Richard E. Veilleux[1*], Guozhu Tang[2] and Barry S. Flinn[1, 2*]

Abstract

Potato, a member of the Solanaceae, is the world's third most important staple food crop after wheat and rice, and the most widely grown vegetable crop. Generation and characterization of activation tagged mutants of potato could aid in functional genomic studies. Morphological and molecular studies were conducted to compare potato cv. Bintje and a Bintje mutant, *nikku*, generated using the activation tagging vector, pSKI074. Morphological studies revealed that *nikku* plants exhibited an extremely dwarf phenotype (mutant plant height was 2 cm compared to 6 cm in Bintje), had small hyponastic leaves, were rootless, and infrequently produced small tubers when compared to wild type Bintje. The overall phenotype was suggestive of a constitutive stress response, which was further supported by the higher expression levels of several stress-responsive genes in *nikku*. Southern blot analyses indicated the presence of a single T-DNA insertion on chromosome 12 in the mutant. Initial gene expression studies comparing Bintje and *nikku* plants on genes flanking both sides of the insertion indicated transcriptional activation/repression of several genes in the mutant, suggesting that activation tagging had pleiotropic effects on a variety of genes flanking the insertion. Two predicted potato genes, encoding an Acyl-CoA N-acyltransferases (NAT) superfamily protein, and a predicted Major Facilitator Superfamily protein (MFS) were situated at a distance of 117 kb and 172 kb, respectively, from the right T-DNA border. These transcripts were significantly up-regulated in *nikku*. Overexpression lines Bintje/35S::NAT1 and Bintje/35S::PMT1 were created for recapitulation of mutant phenotype.

Keywords: *Solanum tuberosum*, Solanaceae, Acetyltransferase, Sugar transporter, GCN5

[1]Department of Horticulture, Virginia Tech, Blacksburg, Virginia, USA, 24060

[2]Institute for Sustainable and Renewable Resources (ISRR) at Institute for Advanced Learning and Research (IALR), Danville, Virginia, USA, 24540

* Both contributed to project development and supervision.

Introduction

Potato (*Solanum tuberosum* L.) is a member of the Solanaceae, an economically important family of plants that includes tomato, pepper, aubergine (eggplant), petunia and tobacco. Potato is also the world's fifth most important commercial crop, third most important staple food crop after wheat and rice, and the most widely grown vegetable crop. Its worldwide importance, especially within the developing world, is growing rapidly, with the total production of 324 million tonnes, valued at 44.5 billion USD in 2010 (http://www.fao.org). Potato tubers are a globally important dietary source of starch, protein, antioxidants and vitamins. Most commercial potato cultivars are autotetraploid (2*n*=4*x*=48). Following the completion of the potato genome (The Potato Genome Sequencing Consortium, 2011), one of the goals has been the elucidation of the functions of all the genes in the potato genome, and the acquisition of a greater understanding of the mechanisms underlying the development of this plant. Mutagenesis and selection remain key strategies in the investigation of biological processes. Genetic mutants deficient in a biological process enable mutant and wild-type genes involved in the process to be identified and isolated. In agricultural species, mutagenesis has the added advantage of potentially creating agronomically useful phenotypes. A variety of mutagenesis methods is available to facilitate the generation of phenotypically altered organisms. The most widely used method is insertional mutagenesis, in which DNA elements such as transposons and T-DNA sequences are inserted into the genome, where they act as a molecular tag that can be useful for cloning genes (Ayliffe and Pryor, 2007). Insertional mutagenesis using transposons or T-DNA has become a valuable research tool for model plant systems. However, all gene disruption approaches also have some limitations. Insertional mutagenesis commonly, but not always, generates mutant phenotypes as a consequence of loss-of-function of a gene. The effect is usually recessive, requiring homozygous mutant stocks, and may be seldom identified in genes that are redundant or compensated by alternative biochemical processes (Ayliffe et al., 2007) rendering these methods unsuitable in these cases. This problem becomes more salient when either two or more redundant genes perform certain functions, when a mutation induces early embryonic lethality (Ahn et al., 2007) or when absence of a visible phenotype can obscure the identification of such mutants.

An alternative strategy to circumvent the problems inherent in the classical loss-of-function approach has been activation tagging, which can generate dominant gain-of-function phenotypes (Weigel et al., 2000). These mutant phenotypes are a consequence of altered gene function due to native gene overexpression rather than gene inactivation. The different mutation spectrum offered by this process has the potential for generating agronomically advantageous phenotypes previously unobtainable by conventional mutagenesis (Ayliffe et al., 2007). Enhancer or promoter sequences can potentially lead to different types of transcriptional perturbations and may produce novel phenotypes, which can aid in determining important gene functions. The dominant character of the mutation, caused by transcriptional activation of flanking plant genes, has proven suitable for the isolation of genes regulating complex developmental and metabolic pathways as we reported in characterization of activation tagged mutant *underperformer* in Chapter 1 and Chapter 2. The detailed methodology, potential advantages and limitations of activation tagging are also discussed in Chapter 1.

In Chapter 1, we described the morphological and molecular characterization of an activation tag line, AT615, named *underperformer* (*up*). In this paper, we present the results of the characterization of a second activation tagged potato mutant, AT601 generated by the Canadian Potato Genome Project (CPGP). This mutant was extremely dwarfed, deficient in stem and leaf growth, rooting and tuberization. Based on its phenotype, we named the AT601 mutant line '*nikku'* (Hindi- *little one*). This mutation was caused by an insertion on chromosome 12, and an assessment of activated genes flanking the T-DNA insertion indicated that several predicted genes flanking the right border, known to be associated with stress response, were up-regulated. We discuss their possible contributions to the overall *nikku* phenotype. In addition to gene discovery in potato, these mutants can also serve as material for post-genome analysis including transcriptomics, proteomics, and metabolomics.

Results

Generation and screening of potato activation tagged mutants

The potato cv. Bintje was used as the genetic background to generate the activation tagged mutant lines, due to its more reliable shoot regeneration and transformation in a preliminary study involving two cultivars and several different media. The *nikku* mutant line AT601 was selected for further study due to its distinct phenotype observed *in vitro*. In contrast to wild type 'Bintje' plants grown for 1 month, mutant plants displayed slow shoot growth, short internodes, small leaves which exhibited hyponasty, and formed no roots (**Figure 3.1**). Mutant plantlets were initially screened by polymerase chain reaction (PCR) to confirm the presence of the T-DNA plasmid sequences in the plant (data not shown).

Determination of T-DNA copy number by southern blot analysis

To determine the T-DNA copy number in the *nikku* mutant, we performed Southern blot analysis with genomic DNA, using two restriction enzymes (*Pci*I and *Psi*I) that cut within the T-DNA, in the region, just left of the tetramerized enhancer sequences. Using a probe that corresponded to the enhancer region of pSKI074, we observed a single hybridizing band for both restriction enzymes, which were close in size in the *nikku* mutant DNA. No bands were observed with Bintje, indicating the presence of a single T-DNA insertion in the mutant (**Figure 3.2**).

In vitro plant growth and tuberization

In order to quantify growth characteristics, we multiplied Bintje and *nikku* plants from 1 cm long nodal explants derived from 4-5 week old *in vitro* plantlets, and recorded data after 4 weeks of *in vitro* growth (**Figure 3.3**). Mutant *nikku* plants exhibited a mean stem length of 2 cm which was significantly less compared to the Bintje control, where the mean stem length was 3 times longer (6 cm). As initially observed, the *nikku* mutant plants rarely developed roots, (0.16 roots/plant), while the mean number of roots observed in Bintje plants was 5.8 roots/plant (**Figure 3.3**). *In vitro* tuberization experiments were conducted to assess the tuberization ability of the *nikku* mutant. Tuberization of the mutant was poor (**Figure 3.4**), with the mean number of tubers/explant (0.39) significantly lower than observed for Bintje (2.73) (**Figure 3.5**). However, the mutant exhibited more stolons/explant (2.27) than Bintje (0.88) (**Figure 3.5**).

Although most of the *nikku* plantlets did not form tubers, the few tubers that did develop were different in size and shape from those of Bintje. Under *in vitro* tuberization conditions, Bintje explants produced both roots and tubers that were round and uniform, whereas *nikku* explants were rootless and developed smaller, irregular, tuber-like structures (**Figure 3.4**). As cytokinin is known to impact tuberization, we assessed the effects of kinetin on *in vitro* tuberization with both Bintje and *nikku*. *In vitro* tuberization using tuberization medium supplemented with 10 µM kinetin improved the number and weight of tubers for Bintje, but there was no effect on *nikku* tuberization (data not reported).

The *nikku* phenotype is associated with a constitutive stress response

The dwarf, slow-growing mutant phenotype resembled that of a constitutive stress response. In order to assess this, we used RT-PCR analysis to investigate the expression of several known stress-induced genes in *in vitro*-grown Bintje and *nikku* shoots (**Figure 3.6**). The transcript levels of four genes correlated with stress response were elevated in the *nikku* mutant. These genes include the CBF1/CBF2 (PGSC0003DMB000000154:1180500...1181620) genes encoding cold-induced transcription factors, an ABA-induced LEA gene (PGSC0003DMB000000032: 1950973...1952454), P5CS1 (PGSC0003DMB000000088:1830300...1820500), encoding the proline biosynthetic enzyme *Δ*1-pyrroline-5-carboxylate synthase and the ascorbate peroxidase-encoding gene APX2 (PGSC0003DMB000000614:83000...86560). Of these genes, the strongest induction in the mutant occurred with P5CS1.

***In vitro* assays to recapitulate the mutant phenotype**

Several experiments were conducted with Bintje explants in an attempt to recapitulate the mutant phenotype. As stress responses are often mediated through ABA, and several of the genes listed above are induced by ABA, we subjected Bintje plantlets to growth *in vitro* in the presence of various concentrations of ABA (**Figures 3.7 and 3.8**). We observed that increasing concentrations of ABA decreased stem length, leaf size and rooting (**Figure 3.7**). A concentration of 5 µM ABA reduced stem length 3-fold (2.8 cm) compared to the control treatment (8.5 cm), with increasing concentrations displaying further reduction (**Figure 3.8**). Root length was also significantly reduced (0.42 cm) with 25 µM ABA compared to control

treatment (4 cm) (**Figure 3.8**). At the highest levels of ABA tested (500-1000 µM), the explants appeared similar to the *nikku* mutant. We also exposed Bintje explants to a salt stress assay. At 25 mM NaCl, little effect was observed on stem length, root length and root number compared to control medium (**Figures 3.9 and 3.10**). However, greater levels of salt stress (>50 mM) significantly reduced all the observed traits, plants were significantly dwarfed and rootless at 200 mM NaCl, and died at 500 mM NaCl (**Figures 3.9 and 3.10**). Of the ABA and NaCl treatments, the ABA-treated Bintje explants at the highest concentrations tested appeared the most similar to the *nikku* phenotype.

Attempts at *ex vitro* growth characterization of *nikku*

Due to the lack of rooting with *nikku* shoots, all of the growth data obtained was from *in vitro* experiments. In an attempt to generate consistently rooted plantlets for subsequent *ex vitro* growth experiments, we tested various concentrations of auxin (IBA) with *nikku* shoots (**Figure 3.11**). Our experimental treatments represented 0, 20, 50, 100 and 200 µM IBA, which were tested with both Bintje and *nikku* shoots. In the case of wild type Bintje shoots, which rooted in the absence of any supplied auxin, increasing IBA concentrations greater than 50 µM significantly reduced root growth, and induced some basal callus formation in the root region (**Figure 3.11A-E**). There was a stimulation of root number at 20 µM IBA. The *nikku* shoots also displayed increased basal callusing and additional dwarfing. In no cases were any roots induced by the IBA treatment (**Figure 3.11F-J**). Additional attempts at root induction with *nikku* shoots involved treatment with the bacterial endophyte *Burkholderia phytofirmans* strain PsJN, and high temperature (28°C), but were not successful (data not shown). Attempts were also made to *in vitro* graft *nikku* scions on Bintje rootstocks, but the grafts were not able to take. However, we were able to grow a limited number of mutant plants (2) from the *in vitro* tubers produced. Mutant shoots grew from the tubers, maintaining the slow growth phenotype in the soil, and no roots were observed to form. Little increase in stem length was observed over a period of 8 months, such that they still remained significantly smaller than Bintje plants grown from tubers (**Figure 3.12**). In contrast to the typical compound leaves of potato (**Figure 3.12A**), we noticed that the small-sized leaves of the mutant plants were simple (**Figure 3.12B**). After 8

months, plants were discarded and the few small-sized tubers that could be harvested were much smaller and knobbier than the tubers from 3 month-old Bintje (**Figure 3.12C-D**).

Identification of potato genomic regions flanking the T-DNA

In order to identify the location of the integrated T-DNA, we used GenomeWalker (Clontech) technology to isolate the genomic region flanking the T-DNA insertion. A gene-specific primer (GSP1) and a nested gene-specific primer (GSP2) were designed against the mannopine synthase gene sequences to amplify the genomic DNA flanking the left border region of pSKI074. These primers (**Table S3.1**) were used with GenomeWalker arbitrary primer 1 (AP1) and arbitrary primer 2 (AP2) to amplify potato genomic regions flanking the left border of the T-DNA. Primary and secondary PCR products were generated (**Figure S3.1**), and the secondary products were cloned and sequenced. The resultant genomic DNA sequences flanking the left border of the T-DNA region were used to query the potato genome PGSC_DM_v3_superscaffolds database (http://potatogenomics.plantbiology.msu.edu/blast.html). The location of the T-DNA insertion was found on superscaffold PGSC0003DMB000000116: 916193 (**Figure 3.13**), more specifically on chromosome 12. Further PCR and sequence analysis of the left border T-DNA/plant DNA junction site revealed that a vector backbone of 278 bp (7777 to 8053 sequence of pSKI074) outside of the T-DNA was also found in the *nikku* mutant (data not shown). A similar analysis of the right border T-DNA/plant DNA junction site in the *nikku* mutant revealed that there was no backbone T-DNA sequence integration at the right border. Using the potato genome browser (http://solanaceae.plantbiology.msu.edu/cgibin/gbrowse/pgsc-potato-dm/), we identified several gene model predictions on both sides of the T-DNA insertion. The first predicted open reading frames flanking the T-DNA was approx. 3.4 kb and 9 kb from the right and left borders, respectively, of the T-DNA (**Figure 3.13**).

Transcript analysis of potato genes flanking the T-DNA

Transcript analysis of the predicted potato genes flanking the T-DNA was performed by reverse transcriptase PCR, using primers designed against the reading frame of gene models (**Table S3.2**). We analyzed an approx. 700 kb potato genomic region off of the right border and an

approx. 400 kb region off of the left border of the insertion site, and observed up- and down-regulation of potato genes flanking both right and left borders of the T-DNA (**Figures 3.14 and 3.15**). For genes flanking the right border (**Figure 3.14**), the gene closest to the insertion showed little change in expression between Bintje and *nikku*. However, the next gene, ethylene inducing xylanase, showed some down-regulation in the mutant. Several of the genes downstream from here showed increases in expression, including a conserved gene of unknown function, an E3 ubiquitin ligase RMA1H1, a ribosomal protein alanine (referred to as NAT1 due to its homology with Acyl-CoA N-acyltransferase) and a predicted sugar transporter (referred as PMT1, due to its homology with polyol/monosaccharide transporters), a receptor-like kinase SRF3 and a DNA-binding protein 6b-interacting protein 1.

We also assessed the expression patterns in the leaves for several genes situated within 400 kb genomic region flanking the left border (**Figure 3.15**). One gene, encoding a senescence-associated protein, was observed to be down-regulated in the mutant. Several other genes displayed some up-regulation in the mutant, such as a WAX2 protein, a serine-threonine protein phosphatase, a conserved gene of unknown function, and a DNA-binding protein phosphatase 2C.

Analysis of contributions of the candidate genes in the mutant phenotype

The transcript analysis of the flanking genes on both sides of the T-DNA revealed that insertion of this DNA had resulted in changes in expression of many genes in the *nikku* mutant, not just the first (or few) gene(s) immediately adjacent to the right border and associated tetramerized enhancers. To determine which of the modified genes may have primarily contributed to the overall *nikku* phenotype, we selected two genes, NAT1 and PMT1, located 117 kb and 172 kb, respectively from right border both of which were up-regulated in the mutant (**Figure 3.14**). Cloning of NAT1 and PMT1 revealed open reading frames of 468 bp for NAT1 and 1,560 bp for PMT1, encoding predicted proteins of 155 amino acids (NAT1) and 519 amino acids (PMT1) in size. Using these cloned genes, independently derived transgenic Bintje lines constitutively expressing the native potato genes NAT1 or PMT1 under the control of the CaMV 35S promoter

were regenerated, to see if they would recapitulate the original *nikku* mutant phenotype. Initial observations indicated Bintje/35S::NAT1 plants were slow growing and showed some similarities with original *nikku* mutant phenotype, whereas Bintje/35S::PMT1 plants were more similar to the 'Bintje' plants.

Discussion

This report describes the isolation and characterization of a novel, activation-tagged potato mutant, *nikku*, which was identified by its extreme dwarf and lack of rooting phenotype under *in vitro* conditions. Our studies revealed additional mutant characteristics like small, simple hyponastic leaves and little-to-no tuberization, with the few tubers that formed being small with a knobby, irregular appearance. The overall growth characteristics suggested that the mutant was exhibiting a constitutive stress response, as is often reported for plants showing a severe dwarf phenotype (Jungkunz et al., 2011; Williams et al., 2005). Our application of stress treatments to wild-type Bintje nodal sections, through high levels of ABA and NaCl, generated similar dwarf phenotypes. In addition, it has been reported that tuber production during stress situations is greatly reduced, and the tubers were smaller in size (Silva et al., 2001), similar to our observations during *nikku* tuberization experiments. Furthermore, several genes known to be up-regulated under stress conditions (LEA, CBF 1/2 and P5CS1) were expressed at higher levels in *nikku* compared to wild-type Bintje. All of these observations taken together further support that the *nikku* phenotype is representative of a strong, constitutive stress response.

The stress-response genes analyzed earlier and shown to be expressed at higher levels in *nikku* are known to be up-regulated by the plant stress-response growth regulator ABA. Furthermore, our application of various levels of ABA to wild-type Bintje nodal sections reconstituted a phenotype similar to that of the *nikku* mutant, namely lack of shoot, root and leaf growth. These results, together with the fact that ABA is known to reduce microtuber size (Gopal et al., 2004), suggest the involvement of components of the ABA signaling pathway as part of this constitutive stress response. While we did not analyze endogenous ABA content in *nikku* tissues, this is one area to be targeted in the future.

Our analysis of the T-DNA integration in *nikku* indicated a single insertion on chromosome 12. Subsequent analysis of the chromosome sequences flanking the insertion site revealed that several genes flanking both ends of the T-DNA were modified in gene expression patterns compared to wild-type Bintje. We observed a similar characteristic during the analysis of another activation-tagged potato mutant, *underperformer*, as discussed in Chapter 1. Similar observations of gene expression modifications flanking either the left border or right border of the T-DNA insertion have been reported with activation-tagged mutants for other plants (Ahn et al., 2007; Hsing et al., 2007; Imaizumi et al., 2005). The concept of activation tagging has revolved around the random insertion of multimerized CaMV 35S enhancers in genomic DNA, and the ability of these enhancers to cause transcriptional activation of nearby genes (Hayashi et al., 1992). Initial studies and interpretations suggested that activation-tagging was a relatively simple system, with the overexpressed genes occurring immediately adjacent to the enhancers, flanking the right border of the T-DNA (Weigel et al., 2000). However, several studies have suggested that activation tagging may not be as simple. In addition to the fact that modified gene expression patterns can be observed adjacent to and further downstream from the left border (e.g., not adjacent to the enhancers), the activation tag also appears to enhance gene expression if located downstream (e.g., off of the 3'end) of the tagged gene (Hsing et al., 2007; Imaizumi et al., 2005; Wan et al., 2009). Furthermore, the activated genes can occur at significant distances from the T-DNA insertion site, and are not always immediately adjacent to the T-DNA/enhancer sequences. While Ahn et al. (2007) noted that in many cases, the gene adjacent to the enhancers was activated, an analysis of seven randomly selected genes within 984 kb downstream of the right border/enhancers indicated that six were up-regulated. Two of these were located 720 kb and 890 kb from the T-DNA/enhancers. These results suggested that broad regions of the chromosome could be impacted by the 35S enhancers, and that any gene within that broad region, and not simply the gene immediately adjacent to the T-DNA, could be responsible for the dominant plant phenotype (Ahn et al., 2007). This long distance modification of gene expression may not be surprising, as the chromatin structure surrounding the enhancer insertion site can strongly facilitate enhancer action over large DNA distances (Rubtsov et al., 2006).

While our previous analysis of *underperformer* in Chapter 1 revealed a strong activation of the gene immediately adjacent to the enhancers, our results with *nikku* revealed no enhancement in expression of the genes adjacent to the right border and enhancers. In fact, the second gene downstream of the right border (ethylene-inducing xylanase) exhibited down-regulation. However, several genes located from 36 kb to 705 kb downstream of the enhancer sequences showed up-regulation in *nikku*. Off of the T-DNA left border, our analyses indicated various genes located from 22 kb to 106 kb exhibited up-regulation in the mutant, while one gene 32 kb from the left border appeared to be down-regulated in the mutant. These results agree with the comments by Ahn et al. (2007), indicating gene expression along a broad range of the chromosome being affected by the T-DNA insertion.

Regarding the activation-tagging process, most studies have suggested that the genes downstream of the enhancer sequences are associated with the mutant phenotype, although determining the major contributor to the mutant phenotype may be more difficult to assess, as suggested by Ahn et al. (2007). Many genes on chromosome 12 were modified in their expression patterns in the *nikku* mutant, with several genes downstream of the enhancers associated with stress responses. One gene, an E3 ubiquitin ligase, was located 36 kb downstream from the right border. These ubiquitin ligases are associated with the regulation of several stress responses, modulating the degradation of different proteins via the 26S proteasome, including many stress-responsive transcription factors (Lyzenga and Stone, 2012). Furthermore, other reports have indicated that E3 ligases play a role in ABA-dependent or ABA–independent stress responses (Lee and Kim, 2011). Our potato gene showed homology with the RING E3 ligase, RING membrane-anchor 1 homologue 1 (Rma1H1), which is a dehydration-induced gene (Park et al., 2003) capable of enhancing tolerance to drought stress (Lee et al., 2009). This is mediated via the degradation of the plasma membrane water transporter, aquaporin, which is associated with symplastic water transport. Hence, the increased expression of the RING E3 ligase Rma1H1 in *nikku* may result in reduced aquaporin levels and less water transport, reducing the growth potential of *nikku* and contributing to the extreme dwarf phenotype. We would also expect that other unknown proteins are targeted for

degradation via the action of the E3 ligase, which could also contribute to overall plant phenotype. However, E3 ligase regulation usually occurs as a downstream component of stress signaling, further removed from the stress, the stress sensor, and the transducers (calcium, kinases, phosphatases) associated with the signal transduction pathway (Lyzenga and Stone, 2012).

Two other genes of interest downstream of the enhancers which were up-regulated on chromosome 12 were a ribosomal-protein-alanine, encoding an Acyl-CoA N-acyltransferase (NAT) superfamily protein, and a member of the Major Facilitator Superfamily, encoding a sugar transporter of the polyol/monosaccharide transporter (PMT) family. Both of these genes were of potential relevance to the overall stress phenotype observed with the *nikku* mutant. The potato NAT1 (located 117 kb downstream of the right border) up-regulated in *nikku* displayed similarity to the histone acetyltransferase GCN5. Histone acetylation involves the transfer of an acetyl group from acetyl-CoA to the N-terminal lysine residues of histones H3, H4, H2A and H2B by histone acetyltransferases (HATs). These histone modifications are important in chromatin remodeling and gene regulation (Horn and Peterson, 2002), with active genes preferentially associated with highly acetylated histones, while inactive genes are associated with hypo-acetylated histones. HATs can be classified into two groups, HAT A (nuclear) and HAT B (cytoplasmic), according to intracellular location and substrate specificity (Brownell and Allis, 1996). A-type HATs are involved in the post-synthetic acetylation of all four core nucleosomal histones, impacting on chromatin assembly and gene transcription, and can be classified into five families, one of them the GNATs (for GCN5-related actyltransferase).

In Arabidopsis, AtGCN5-dependent histone acetylation is closely associated with many developmental pathways, by controlling the expression of key developmental regulatory genes involved in the cell cycle, flowering time, response to environmental stresses, hormone signals and epigenetic processes (Chen and Tian, 2007). More specifically, H3K9 acetylation is found to be associated with actively transcribed genes and shown to influence numerous developmental and biological processes in plants (Benhamed et al., 2006). Chromatin

immunoprecipitation experiments in Arabidopsis indicated that approximately 40% of promoters associated with AtGCN5 (Benhamed et al., 2008), suggesting the potential for GCN5-mediated histone acetylation to affect many genes, and the potential for numerous pleiotropic effects should any alteration in GCN5-mediated histone acetylation occur.

Histone acetyltransferases (HATs) interact with transcription factors and are involved in activating stress-responsive genes, such as those for drought-induced gene expression (Kim et al., 2008) and cold stress (Stockinger et al., 2001). GCN5 serves as the catalytic subunit of the Spt-Ada-Gcn5 acetyltransferase (SAGA), forming complexes with these and other proteins, with GCN5 coordinating with the other complex subunits to stimulate transcriptional activation (Servet et al., 2010). Studies in Arabidopsis (Stockinger et al., 2001) revealed that CBF (CRT/DRE binding factor), transcriptional activators involved in cold regulated gene expression are dependent upon activity of GCN5 in association with ADA2a and ADA2b proteins. AtGCN5, AtADA2b and the AGA complex may interact with CBF transcription factors to promote cold stress regulated COR gene expression. Mutant *nikku* exhibited increased levels of CBF1/2 transcripts. Hence, given the role of GCN5 in gene activation, including those associated with stress response, the up-regulation of GCN5 in *nikku* may contribute to the increased stress responsive gene expression and overall phenotype of the mutant.

The second candidate gene, located just downstream of the potato GCN5 (172 kb from the right T-DNA border), was a member of the major facilitator superfamily (MFS) of proteins. Currently, there are 74 known MFS families, each transporting a different set of related compounds, such as monosaccharides, oligosaccharides, amino acids, peptides, vitamins, enzymes cofactors, drugs, nucleotides, metal chelates, organic and inorganic anions and cations (Reddy et al., 2012). The largest MFS family represents sugar porters and permeases (Pao et al., 1998). Our gene of interest, which was up-regulated in *nikku*, was a member of the polyol/monosaccharide transporters (PMTs), which we designated as potato PMT1. Various members of the PMT subfamily are plasma membrane-localized, and can transport various hexoses and sugar alcohols (Klepek et al., 2010).

Sugar transporters play key roles in plant growth and development, as they are required to carry out the efficient movement and partitioning of carbohydrates between distinct organs and tissues (Buettner, 2010), and the PMTs play a specific role in the apoplastic loading of sugar alcohols into the phloem for transport (Slewinski, 2011). In addition to development, numerous sugar transporters are up-regulated during drought, salt and cold stress (Aluri and Buettner, 2007; Kiyosue et al., 1998; Wormit et al., 2006; Yamada et al., 2010), as well as exogenous ABA (Oliver et al., 2007), serving to enhance sugar accumulation in plant tissues. This is believed to be protective in function via membrane stabilization and osmotic adjustment, as well as play a role in redox and sugar signaling (Slewinski, 2011). So, the up-regulation of PMT1 observed in *nikku* fits into the suggestion of an overall constitutive stress response observed for this mutant.

Many mutants altered in sugar transport exhibit modifications to overall plant growth and development. Trehalose is a minor sugar involved in plant responses to stress. There are several examples in which the positive and negative effects of trehalose in plants are demonstrated or suspected (Fernandez et al., 2010). One example, the '*sweetie*' mutant, exhibited strongly modified carbohydrate metabolism leading to elevated levels of trehalose, trehalose-6-phosphate and starch, and significantly reduced growth and development, including severe dwarfism (Veyres et al., 2008). These authors also noted significant changes in gene expression in the mutant, including increased expression of sugar metabolism genes (including sugar transporters), as well as numerous genes associated with the response to abiotic or biotic stress. The severe dwarfism, and increased sugar transporter and stress response gene transcript levels of *sweetie* were somewhat similar to our observations of *nikku*. However, while *nikku* exhibited some similarities with *sweetie*, our *nikku* mutant did not exhibit leaf expansion or any significant rooting, indicating that *nikku* has a more severe phenotype. It is possible that increased sugar transporter activity may have resulted in severe osmotic modifications in *nikku*, significantly inhibiting growth and expansion. However, we have not yet determined if *nikku* plants contain higher levels of sugars/starch than wild-type Bintje.

Another gene on chromosome 12 that showed some level of up-regulation in *nikku* was a member of the *STRUBBELIG-RECEPTOR FAMILY* (*SRF*), SRF3, located approximately 227 kb downstream of the enhancer sequences. The SRF family encodes leucine rich repeat – receptor-like kinases (LRR-RLKs) believed to control cellular morphogenesis in many plant organs (Chevalier et al., 2005). However, ectopic expression of *35S::SRF3 in Arabidopsis resulted in male sterility, but no other obvious phenotypic variation (Eyueboglu et al., 2007), suggesting little involvement of SRF3 up-regulation in the nikku phenotype.*

In addition, a gene encoding a 6b-interacting protein 1-like protein (**Figure 3.14**, R18)also displayed increased expression levels in the mutant. This protein is a member of the trihelix group of transcription factors, a group associated with responses to salt and pathogen stress, the development of stomata, trichomes, perianth organs, the seed abscission layer, and the regulation of late embryogenesis (Kaplan-Levy et al., 2012). There are five clades in this family of transcription factors, with the 6b-interacting protein 1-like protein falling into the SIP1 clade (Kaplan-Levy et al., 2012). However, little is known about the members of this clade of trihelix transcription factors. The first plant member of 6b-interacting protein 1 was described from tobacco, and was a protein that bound to the *Agrobacterium tumefaciens* oncogenic protein 6b, but its function remains unknown (Kitakura et al., 2002). Two closely-related genes in Arabidopsis, ASIL1 and ASIL2, are associated with the repression of numerous late embryo developmental genes during embryogenesis and in seedlings (Gao et al., 2009; Willmann et al., 2011). However, one member of the SIP1 group results in a dwarf seedling with pale green and aberrant leaves when mutated (Kuromori et al., 2006). This sounds somewhat similar to our *nikku* mutant, although the expression of the potato 6b-interacting protein 1 is up-regulated in *nikku*. Kaplan-Levy et al. (2012) suggested an emerging theme for these transcription factors in suppression of growth. As this is the case for *nikku*, perhaps the over-expression of the 6b-interacting protein 1-like gene contributes to the mutant phenotype. In *nikku* mutant, we have observed the up-regulation of DNA-binding protein phosphatase 2C gene (**Figure 3.15**, L13) which is flanking the left border of T-DNA. In Arabidopsis, a protein phosphatase 2C interacts with AtGCN5 and dephosphorylates AtGCN5 *in vitro* and mutation of PP2C gene increases

H3K14 acetylation (Servet et al., 2008), suggesting that HAT activity of AtGCN5 may be regulated by phosphorylation /dephosphorylation.

Although any single gene in the above discussion could underlie the *nikku* phenotype, with so many differences in gene expression between *nikku* and wild type Bintje, it may be speculated that altered expression of a single transcription factor resulted in a cascade of differential gene expression giving rise to the multifaceted phenotype. The premise of activation tagging, i.e., over-experssion of an adjacent gene, may often be too simplistic in cases where patterns of differential gene expression are complex. Activation tagging may be useful for the identification of developmental mutants, but it is insufficient for elucidating the mechanism of all kinds of genes especially those which induce lethality or developmental arrest or dominant-negative genes with no obvious phenotype.

Methods

Generation and screening of potato activation tagged mutants

The wild-type Bintje and activation tagged mutant AT 601 (*nikku*) mutant plantlets were received from BioAtlantech (http://bioatlantech.nb.ca/site/solanumgenomics), having been generated by the Canadian Potato Genome Project (Regan et al. 2006). The mutant lines were generated as follows: Potato (*Solanum tuberosum* L. cv. Bintje) plants were grown on plantlet growth medium [PGM; MS (Murashige and Skoog, 1962) minimal organic medium, sucrose 3%, pH 5.7 and solidified with 0.7% agar, sterilized] from 1 cm nodal stem sections and kept under 16 h photoperiod, 20˚C day/18˚C night and 70-100 µE $m^{-2}s^{-1}$ light intensity. For transformation, internodal stem segments from 4-week-old plants were excised and placed on infiltration medium (IM; MS minimal organic medium, sucrose 3%, MES 0.5g/L, mannitol 20 g/L, pH 5.5, filter-sterilized) to prevent the explants from drying out.

The activation tagging vector pSKI074 (GenBank accession no. AF218466) was transformed into *Agrobacterium tumefaciens* strain GV3101 (pMP90RK) by electroporation. The T-DNA of vector pSKI074 encompasses left and right borders, kanamycin selectable marker, T3 RNA polymerase promoter, T7 RNA polymerase promoter and four CaMV 35S enhancer repeats (Weigel et al., 2000). The transformed colonies were selected by growing them on ampicillin-containing LB

plates. For potato transformation, *Agrobacterium* was grown from freshly streaked LB plates supplemented with 100 mg/L ampicillin, 50 mg/L kanamycin and 100 mg/L rifampicin and plates were incubated for 2 days at 28°C in the dark. A single colony of *Agrobacterium* was picked and cultured overnight in liquid LB medium containing 100 mg/L ampicillin, 50 mg/L kanamycin and 100mg/L rifampicin, at 28°C in the dark. *Agrobacterium* inoculum was grown to an OD_{600} of 0.6-0.8. Bacterial cells were pelleted by centrifugation at 3,000 rpm for 15 min, re-suspended in IM with 200 µM acetosyringone at a final concentration corresponding to an OD_{600} of 0.3.

Excised internodal explants (placed on IM medium) were transferred to a 50 ml tube and IM medium containing the re-suspended *Agrobacterium* and acetosyringone was added. Tubes were inverted gently for 2 to 5 min. For co-cultivation, explants were put on sterile cheesecloth or filter paper to remove excess *Agrobacterium* before transfer to petri plates containing 25 ml of callus growth medium [CGM; MS minimal organic medium, glucose 2%, adenine sulfate 40 mg/L, MES 0.5g/L, polyvinylpyrrolidone (PVP) 0.5 g/L, glutamine 200 mg/L, pH 5.7 and solidified with 0.22% gelrite, sterilized and supplemented with 0.1 mg/L trans-zeatin and 0.1 mg/L IAA] for 2 days. Petri plates were sealed with micropore tape and incubated at 24°C with 16 h photoperiod under low-light (60-80 µE $m^{-2}sec^{-1}$) in a plant growth chamber.

After 2 days of co-cultivation, explants were washed 2-3 times with sterile distilled water containing 300 µg/ml claforan, blotted on sterile filter paper and placed on callus selection medium (CSM = CGM + kanamycin 100 µg/ml and claforan 300 µg/ml). Petri plates were sealed with micropore tape and incubated at 22-24°C, 16 h photoperiod under low-light in a plant growth chamber. After 10 days, explants were transferred to fresh CSM and higher light intensity (150-200 µE $m^{-2}sec^{-1}$) and incubated for 10 days. A second transfer was done onto fresh CSM and explants incubated for 10 days for callus induction (total 30 days on CSM) before the explants were transferred to shoot growth medium. Calluses were transferred to shoot growth medium (SGM; MS minimal organic medium, glucose 2%, adenine sulfate 40 mg/L, MES 0.5 g/L, polyvinylpyrrolidone(PVP) 0.5 g/L, glutamine 200 mg/L, pH 5.7 and solidified with 0.22% gelrite, sterilized and supplemented with 0.1 mg/L trans-zeatin, 50 mg/L kanamycin and

300 mg/L claforan) and incubated for 6-8 weeks, transferring to fresh SGM every 7-10 days to facilitate shoot elongation. Shoots were transferred to root growth medium (RGM; MS minimal organic medium, sucrose 2%, pH set to 5.7 and solidified with 0.7% agar, sterilized and supplemented with 50 mg/L kanamycin and 300 mg/L claforan) or to PGM when 2 cm long and had several leaves.

Transformed bacterial colonies were screened for presence of the plasmid before infection of internodal explants and also after 2 days of co-cultivation on medium by colony PCR. Bacteria were transferred to fresh LB plates supplemented with antibiotics (ampicillin, kanamycin and rifampicin) and incubated for 2 days at 28°C in the dark. Colony PCR with 074 Red and Fred primers was done to check plasmid sequences. For screening of transformed plantlets, total DNA was extracted from leaves of transgenic plants and PCR was done with same set of primers (074 Red and Fred) as described earlier to confirm the presence of plasmid (**Table S3.1**)

Determination of T-DNA copy number by Southern blot analysis

Genomic DNA of Bintje and *nikku* was digested with *Pci*I and *Psi*I restriction enzymes and separated on 0.8% agarose gel. Digested DNA was transferred to positive charged Zeta probe nylon membrane (Bio-Rad; Hercules, CA, USA) using downward capillary transfer setup and NaOH buffer (Sambrook and Russell, 2001). DNA was transferred to the membrane and rinsed in 2X SSC, then placed on a sheet of Whatman 3 MM filter paper and allowed to air dry for at least 20 min. The probe (640 bp) used for hybridization was amplified from 4X35S enhancer region of pSKI074 plasmid by PCR using primers (35S-F, 35S-R, **Table S3.1**) and labeled using Amersham AlkPhos Direct Labeling Reagents (GE Healthcare Life Sciences; Buckinghamshire, U.K. Catalog No. RPN3682) following manufacturer's protocol. Overnight hybridization with the probe was done in Alkphos Direct hybridization buffer (Amersham GE) at 55°C. After hybridization, the blot was washed to remove excess probe and signal detected with CDP-Star (GE Healthcare Life Sciences; Buckinghamshire, U.K.) chemiluminescent detection reagent. For final signal generation, the blot was exposed to autoradiography film and the film was developed with Kodak GBX developer and fixer solution.

***In vitro* plant growth and tuberization**

For *in vitro* plant growth assays, Bintje plants were grown from single node explants taken from 4-5 week old plants (for *nikku* mutant, it was difficult to obtain single node explants due to small size of internodes, so ~1 cm long explants equivalent in size to Bintje were used instead). Explants were placed on MS basal medium with vitamins (Phytotech, Lenexa, KS, USA) supplemented with 3% sucrose, pH 5.8 and solidified with 0.7% agar. Plants were grown under 16 h photoperiod, 24±1°C and 70-100 µE light intensity in a plant growth chamber and data were recorded after 4 weeks. Each experiment was conducted in a randomized block design with three replicates.

For *in vitro* tuberization assays, stem sections, 3-5 cm long, with 4-5 nodes per section were prepared from 4-5 week old *in vitro* Bintje and *nikku* plants by removing all leaves. These stem sections were grown in 40 ml liquid propagation medium [MS basal medium + 148 mg l^{-1} NaH_2PO_4 + 0.4 mg l^{-1} thiamine HCl + 100 mg l^{-1} inositol + 3% sucrose, pH 5.9] (Radouani, 2003) at 16 h photoperiod, 24±1°C and 70-100 µE for 3-4 weeks in 250 ml conical flasks capped with ventilated plugs. After 3-4 weeks, propagation medium was replaced with microtuberization medium [MS basal + 8% sucrose, 2X the amounts of NH_4NO_3, KH_2PO_4 and KNO_3 and without NaH_2PO_4, pH 5.8] (Radouani, 2003). Culture flasks were kept at 18-20°C in the dark to induce tubers and data were recorded after 30 days. The average of three independent experiments was presented. Each experiment consisted of three replicates in a randomized block design. Data were analyzed in JMP, version 9.0.0 (SAS Institute Inc., Cary, NC) and tables and graphs were drawn in Microsoft Excel (2010).

Identification of potato genomic regions flanking the T-DNA

Genomic DNA was extracted using Plant DNAzol Reagent (Life Technologies; Grand Island, NY, USA) following manufacturer's instructions. Regions flanking the T-DNA were amplified using the GenomeWalker universal kit (Mountain View; CA, USA. Cat. 638904) following manufacturer's instructions. Briefly, genomic DNA from *nikku* mutant plants was completely digested with a set of restriction enzymes *MscI, NaeI, ScaI* and *SspI* that generate blunt ends in separate tubes, and then ligated with GenomeWalker adapters to create "GenomeWalker

libraries." The primary PCR amplification was done with outer adapter primer (AP1), a gene specific primer (GSP1) and the GenomeWalker library DNA as templates which were followed by secondary PCR amplification using nested adapter primer (AP2), a nested gene specific primer (GSP2) and diluted primary PCR product as template. Secondary PCR amplification products were separated by electrophoresis and purified using a gel extraction kit (Qiagen; Valencia, CA, USA); cloned and sequenced. Sequencing was performed in the ISRR lab (IALR, Danville, VA) using a CEQ 8800 genetic analysis system (Beckman Coulter, Brea, CA, USA). Recovered genomic sequences flanking the left border of the T-DNA insertion sites were identified and used to perform Basic Local Alignment Search Tool (BLAST) searches in the GenBank database (accessed 1-28-2010) of the National Center for Biotechnology Information (NCBI) and in the potato genome (http://potatogenomics.plantbiology.msu.edu/blast.html) with default search parameters (expected threshold 10, maximum number of alignments 100, max number of descriptions 100, word length 11, no filter and both strands). This allowed the T-DNA insertion site to be located on superscaffold PGSC0003DMB000000116 within the potato genome. Approximately 700 kb of potato sequence from right border and 400 kb from left border of T-DNA insertion site were analyzed for the presence of annotated gene models using the potato genome browser (http://solanaceae.plantbiology.msu.edu/cgi-bin/gbrowse/pgsc-potato-dm/). Sequence homology searches and analyses were performed using the NCBI BLAST server. Sequence alignments were carried out by the CLUSTALW2 (Larkin et al., 2007) method and using the EMBL server (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

RNA extraction and PCR analysis

RNA was extracted from 4-5 week old *in vitro* plants. All tissues above the culture medium level were collected (leaves and stems but no roots) and immediately put in liquid nitrogen and stored at -80°C until needed. RNA was extracted using either Concert Plant RNA Reagent (Invitrogen, Carlsbad, CA) following manufacturer's instructions or a combination of Trizol Reagent (Invitrogen, Carlsbad, CA, USA) extraction and RNA mini kit (Qiagen, Valencia, CA, USA) purification. For extraction with Trizol Reagent, frozen tissues were ground with a mortar and pestle to a fine powder in liquid nitrogen and 1 g powder per 1 ml of prewarmed (35-40°C) Trizol was used for extraction. Powder was mixed with Trizol immediately by vortexing to

protect it from RNases. This mixture was incubated for 5 min at room temperature (RT), frequently mixed by vortex followed by centrifugation at 12,000 g for 10 min at 2-8°C. After removal of supernatant to a new tube, 0.2 ml chloroform was added and mixed by vortex for 15 sec. Tubes were incubated for 1 min at RT, mixed by vortex for 15 sec and incubated at RT for 2-3 min, followed by centrifugation at 12,000 g for 15 min at 2-8°C to separate the phases. The upper layer was removed and divided into two parts of 200 µl each (one part for backup was stored at-20°C). To another 200 µl from the top layer, 700 µl of Qiagen RLT buffer (after addition of 1 µl β-mercaptoethanol to 1 ml Buffer RLT) and 500 µl of 96-100% ethanol were added and mixed by vortexing. Samples were applied to Qiagen MinElute spin column and washed with RPE buffer, followed by two washes with 750 µl of 80% ethanol (Sigma) to remove all guanidine salts that may inhibit downstream applications. RNA was eluted with 20 µl of RNAase free water. DNAse treatment of RNA was done with DNA-free kit (Ambion; Foster City, CA, USA) following the manufacturer's instructions. First strand cDNA was synthesized using SuperScript III (Invitrogen; Carlsbad, CA, USA) using 1 µg of total RNA. After cDNA synthesis the final volume was adjusted to 100 µl of diluted cDNA for each cDNA synthesis reaction.

Reverse transcription polymerase chain reaction (RT-PCR) was performed using gene-specific primers (**Table S3.2**). Equal amounts of cDNA were used in each reaction and *Solanum tuberosum* elongation factor 1-α (StEF1-α) gene expression was used as control. Taq Polymerase (Takara) was used in all amplification reactions. Gel images were acquired by Alpha Innotech (now part of Cell Biosciences) gel doc system (San Leandro, CA, USA).

ABA, NaCl and IBA assays

For *in vitro* plant stress assays, Bintje single node explants were taken from 4-5 week old plants. For abscisic acid (ABA) and salt (NaCl) stress experiments, we used MS basal medium with vitamins (Phytotech, Lenexa, KS, USA) supplemented with 3% sucrose, pH 5.8 and solidified with 0.7% agar as control medium with the following treatments: ABA (5, 25, 100, 500 or 1000 µM); NaCl (25, 50, 100,200 or 500 mM); agar (6, 8, 10, 12, 14, 16 g/L; 6 g/L). For Indole-3-butyric acid (IBA) experiments we used Bintje and *nikku* explants taken from 4-5 week old *in vitro* plantets. MS medium with vitamins was used as control and treatments consisted of different levels of IBA (20, 50, 100 or 200 µM). Explants on these media were grown under 16 h

photoperiod, 24±1°C and 70-100 µE light intensity in a plant growth chamber and data were recorded after 4 weeks. All experiments consisted of three replications applied in a randomized block design.

StNAT1 and StPMT1 over expression constructs

Full length Open Reading Frames (ORFs) of StNAT1 (PGSC003DMG400007868) and StPMT1 (PGSC0003DMG400007865) were amplified from *nikku* cDNA transcribed from RNA extracted from 1-month-old *in vitro* plants. Amplifications were carried out in a thermal cycler (MyCycler, Biorad) with gene specific primers [**Table S3.1**; StNAT1 (primers 5'StGCN5 ORF *Xho*I and 3'StGCN5 ORF *Xba*I)] and [StPMT1 (primers 5'StGCN5 ORF *Xho*I and 3'StGCN5 ORF *Xba*I)]. The PCR products were cloned into pGEMT-Easy vector and sequenced. Restriction sites for *Xho*1 and *Xba*I added during PCR at 5' and 3' ends of ORFs, respectively, were used for directional cloning of ORFs into a pHannibal vector. Double digestion of the pHannibal vector with *Xho*I and *Xba*I released the 741 bp Pdk intron from the vector and ORFs of StNAT1 or StPMT1 were inserted downstream of 35S CaMV promoters in two separate pHannibal vectors. The gene cassettes, 2.7 kb for StNAT1 and 3.8 kb for StPMT1, consisted of CaMV 35S promoter, ORF (StNAT1 or StPMT1) and OCS terminator were then released with *Not*I digestion and inserted in the binary vector pART27. These two overexpression constructs were used to transform *Agrobacterium tumefaciens* strain GV3101 (pMP90RK). Potato transformation was done using the same method as described earlier.

Phenotyping under growth chamber and greenhouse conditions

Bintje and *nikku* plants were grown in the greenhouse of the Institute for Sustainable and Renewable Resources (ISRR) at the Institute for Advanced Learning and Research (IALR), Danville, Virginia, USA. Planting was done during end of September 2010 and tubers were harvested in early January 2011. Microtubers of Bintje and *nikku*, produced during *in vitro* experiments (this study) were planted in 30 cm diameter pots filled with commercial Miracle-Gro potting mix (Scotts Miracle-Gro, Marysville, OH). Standard plant growth practices were followed. The greenhouse settings were 24-30°C and 16 h photoperiod was maintained by supplementary lighting with 1000 watt, 5.7 Kelvin daylight spectrum, metal halide bulbs.

Acknowledgments

We thank BioAtlantch for the gift of the activation tagged mutant line. This work was funded through Special Grants (2003-38891-02112, 2008-38891-19353 and 2009-38891-20092) from the United States Department of Agriculture, and operating funds from the Commonwealth of Virginia to the Institute for Advanced Learning and Research.

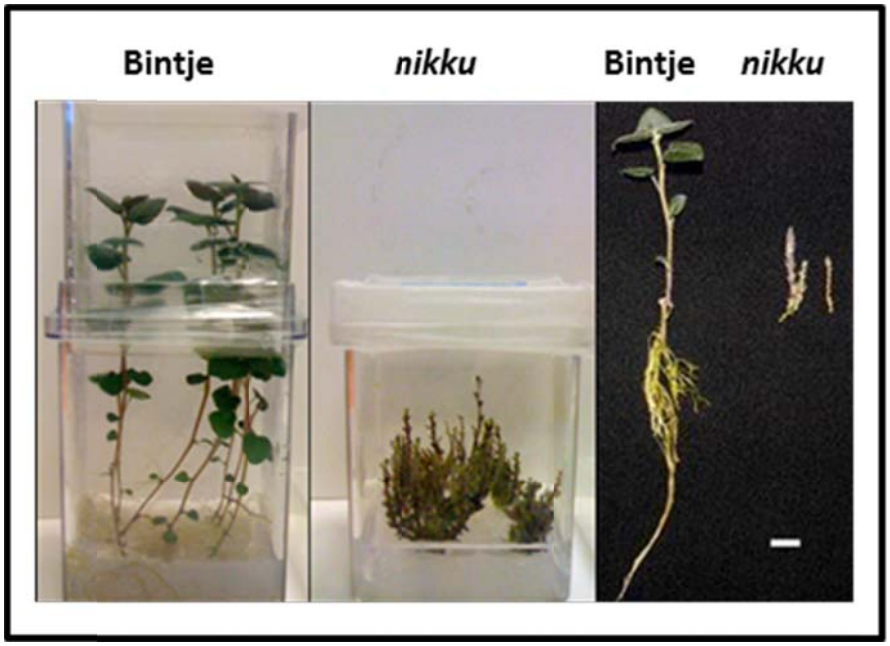


Figure 3.1 Dwarf and rootless phenotype of the *nikku* mutant compared with wild-type Bintje. One month old Bintje and activation tagged mutant *nikku* plants were grown on Murashige and Skoog (MS) medium with vitamins supplemented with 3% sucrose. One cm nodal explants were cultured and grown at 24±1°C temperature with a 16 h photoperiod at 70-100 µE light intensity in a plant growth chamber for one month. (Scale bar=1 cm).

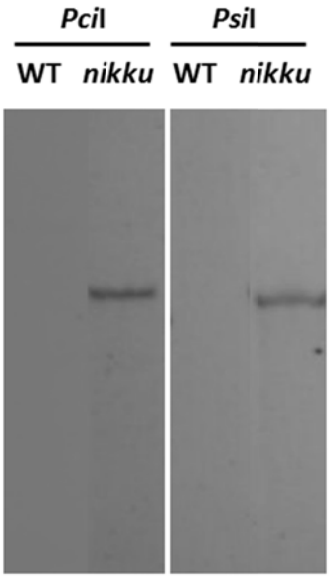


Figure 3.2 Determination of T-DNA copy number by Southern blot analysis. Fifteen µg genomic DNA of Bintje (WT) and mutant *nikku* were digested with restriction enzymes (*Pci*I or *Psi*I) that cut within the T-DNA, and probed with an AlkPhos direct labeled 600 bp probe, and the signal detected with the CDP-Star detection kit.

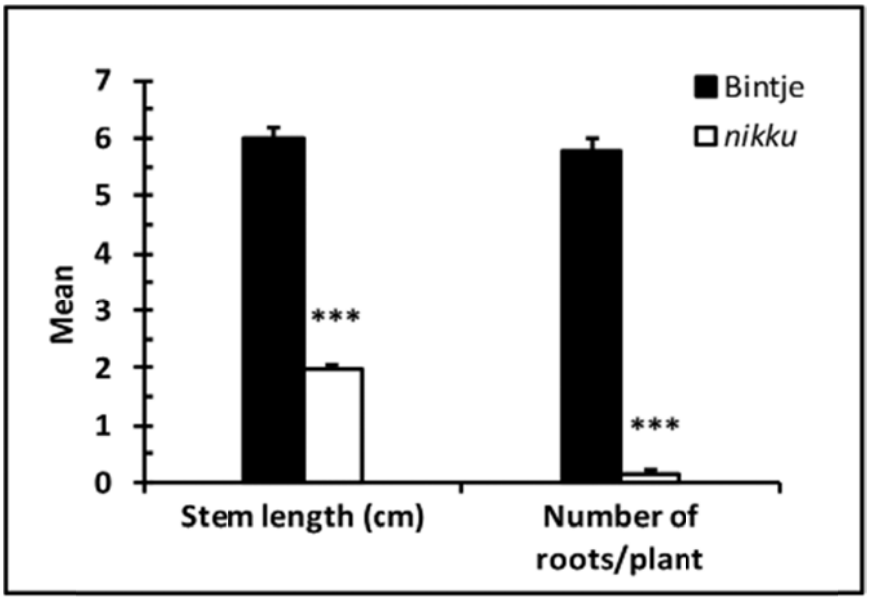


Figure 3.3 *In vitro* plant growth recorded after 4 weeks of growth. The values are means ± SE (sample size, n=50). The asterisks indicate statistical significance of means in the same parameter estimated using Tukey's HSD test (* = P < 0.05-0.005, ** = P < 0.005- 0.0005, *** = P < 0.0005).

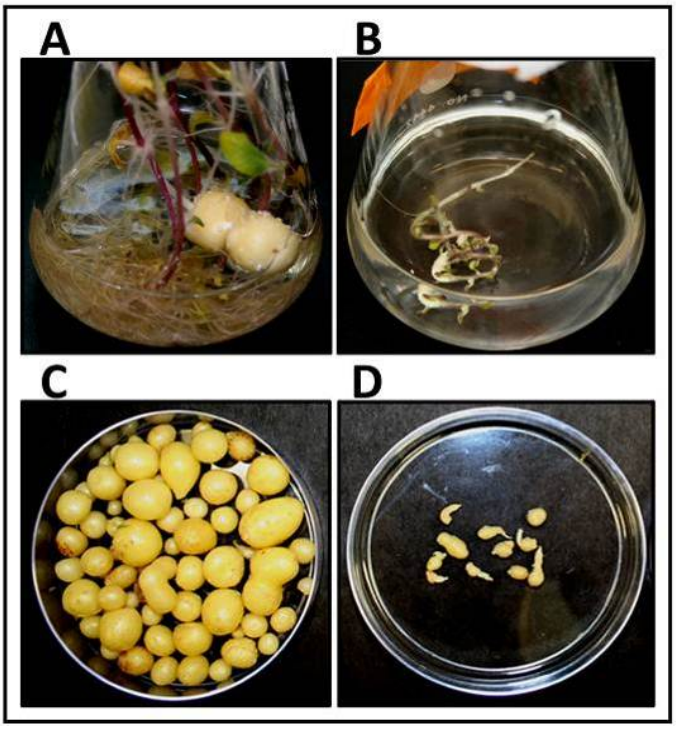


Figure 3.4 *In vitro* tuberization. Bintje (A) and *nikku* (B) explants during tuberization in the flasks, and the harvested tubers of Bintje (C) and *nikku* (D). Stem sections of 3-5 cm in length were grown in 40 ml liquid propagation medium at 24±1°C, a 16h photoperiod and 70-100 µE $m^{-2}sec^{-1}$ light intensity for 3-4 weeks, followed by one month of plant culture in microtuberization medium at 18-20°C in the dark to induce tubers. Explants were analyzed after 30 days of microtuberization.

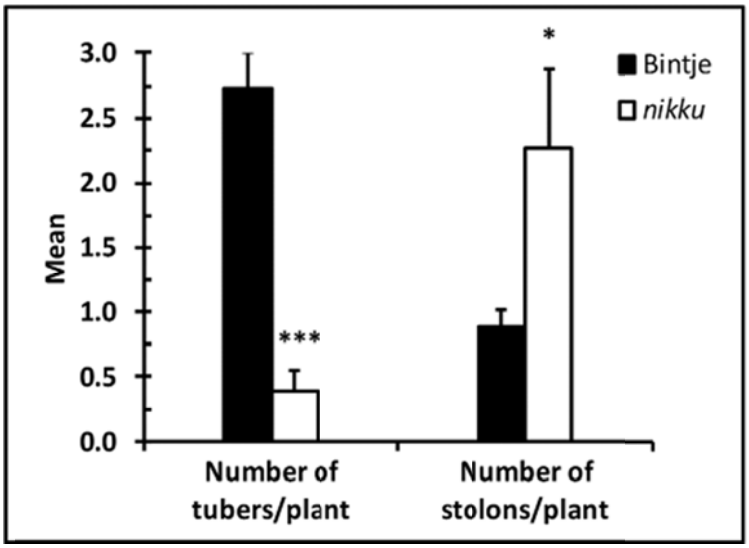


Figure 3.5 *In vitro* tuberization results after 1 month after culture on tuberization medium. The values are means ± SE of sample (sample size, n=33) . The asterisks indicate statistical significance of means in the same parameter estimated using Tukey's HSD test (* = P < 0.05-0.005, ** = P < 0.005- 0.0005, *** = P < 0.0005).

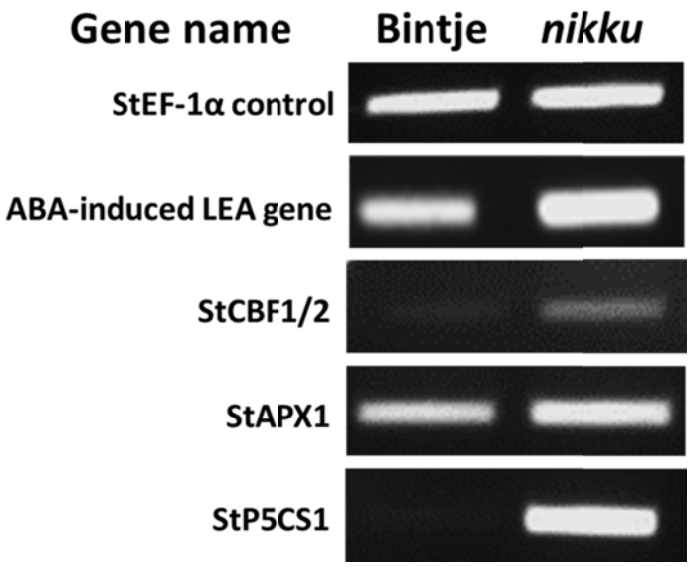


Figure 3.6 The *nikku* phenotype is associated with the constitutive stress response. Reverse transciptase PCR analysis of various stress-induced genes in *in vitro*-grown 'Bintje' and *nikku* shoots were up-regulated *nikku* (AT601) mutant. RNA was extracted from *in vitro* plantlets (leaves+stem). In all gel images the transcript from the mutant is on the right side and the transcript from wild-type Bintje is on the left. StEF1α transcript expression was used as control.

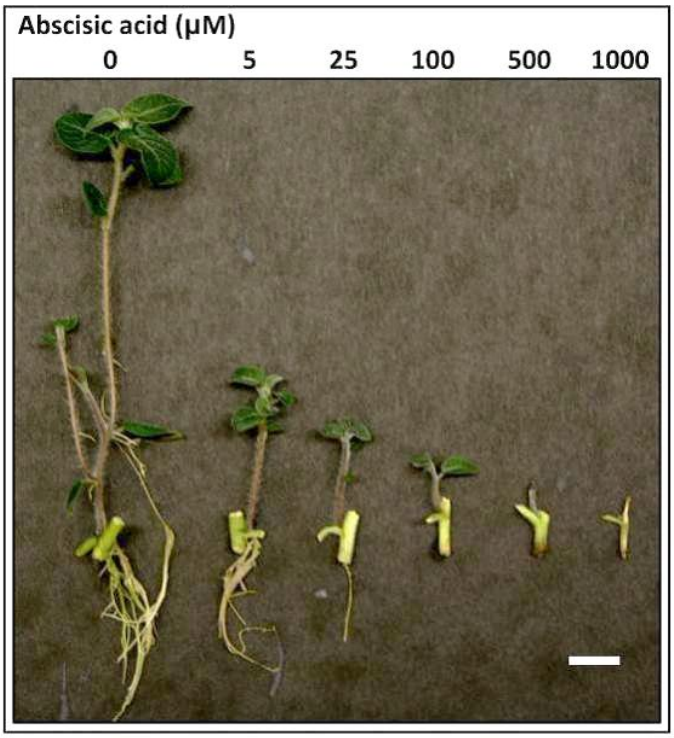


Figure 3.7 Bintje explants cultured with different amounts of Abscisic acid (ABA) in the culture medium. Data were recorded after 1 month of growth on MS medium with vitamins or MS medium with vitamins supplemented different levels of ABA (Scale bar= 1cm).

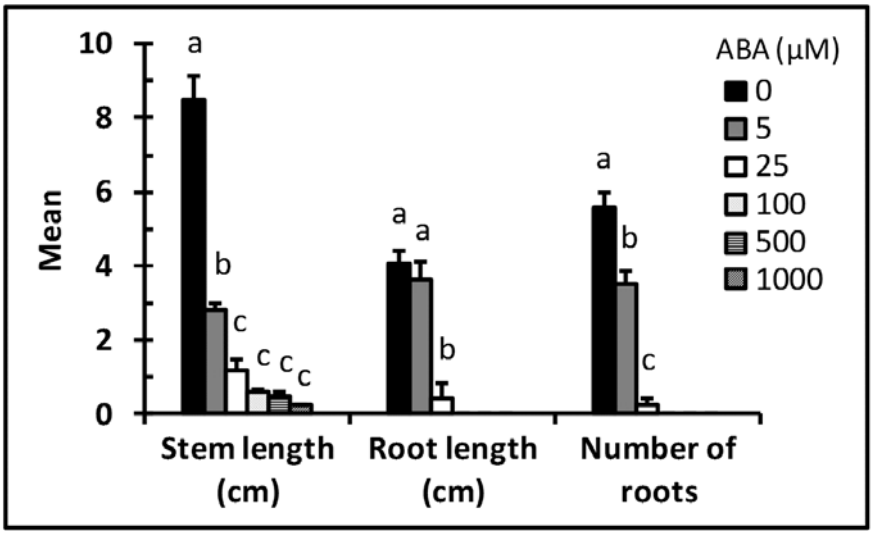


Figure 3.8 Growth data for Bintje explants cultured with different amounts of Abscisic acid (ABA) in the culture medium. Data were recorded after 1 month of growth on MS medium with vitamins or MS medium with vitamins supplemented different levels of ABA. The values are means ± SE (sample size, n=12) of n observations per treatment. The asterisks indicate statistical significance of means in the same parameter estimated using Tukey's HSD test (* = P < 0.05-0.005, ** = P < 0.005- 0.0005, *** = P < 0.0005).

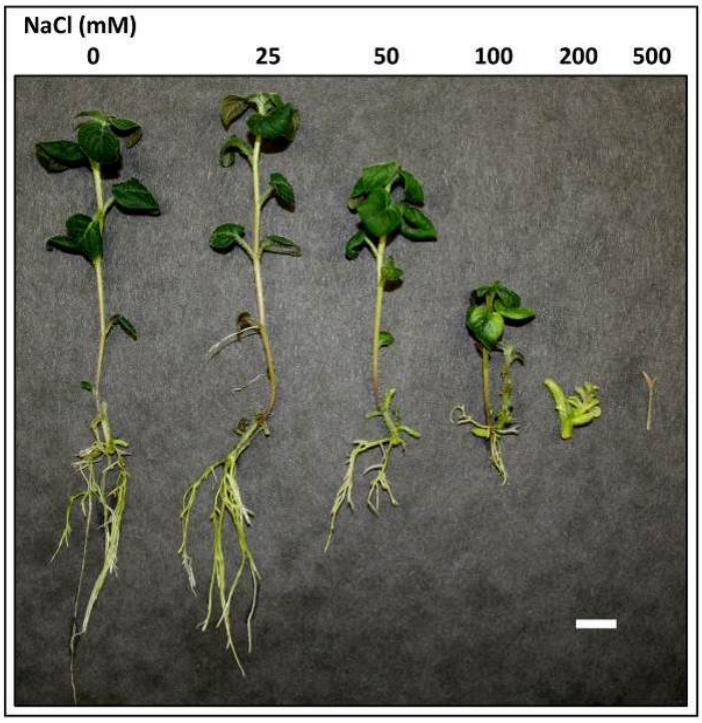


Figure 3.9 Bintje explants cultured with different amounts of NaCl in the culture medium. Data were recorded after 1 month of growth on MS medium with vitamins or MS medium with vitamins supplemented different levels of NaCl (Scale bar= 1cm).

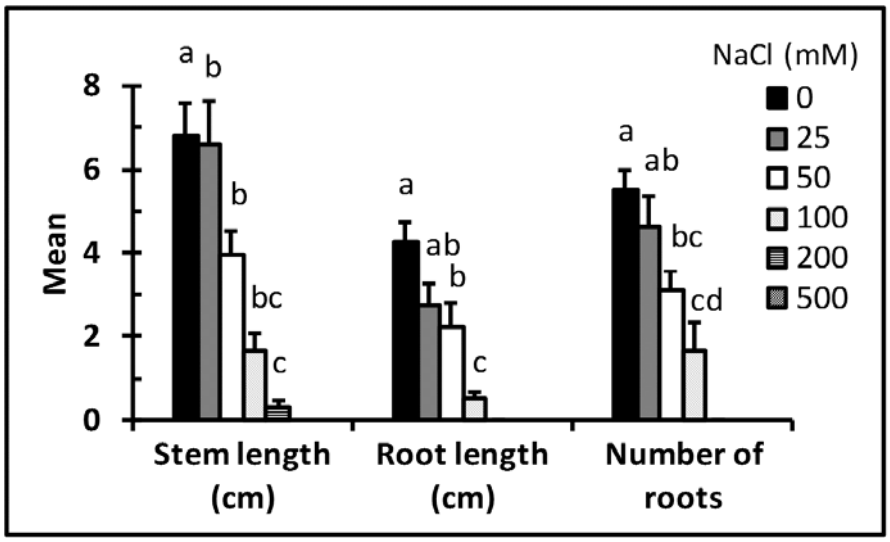


Figure 3.10 Growth data for Bintje explants cultured with different amounts of NaCl in the culture medium. Data were recorded after one month of growth on MS medium with vitamins or MS medium with vitamins supplemented different levels of NaCl. The values are means ± SE (sample size, n=12). The asterisks indicate statistical significance of means in the same parameter estimated using Tukey's HSD test (* = P < 0.05-0.005, ** = P < 0.005- 0.0005, *** = P < 0.0005).

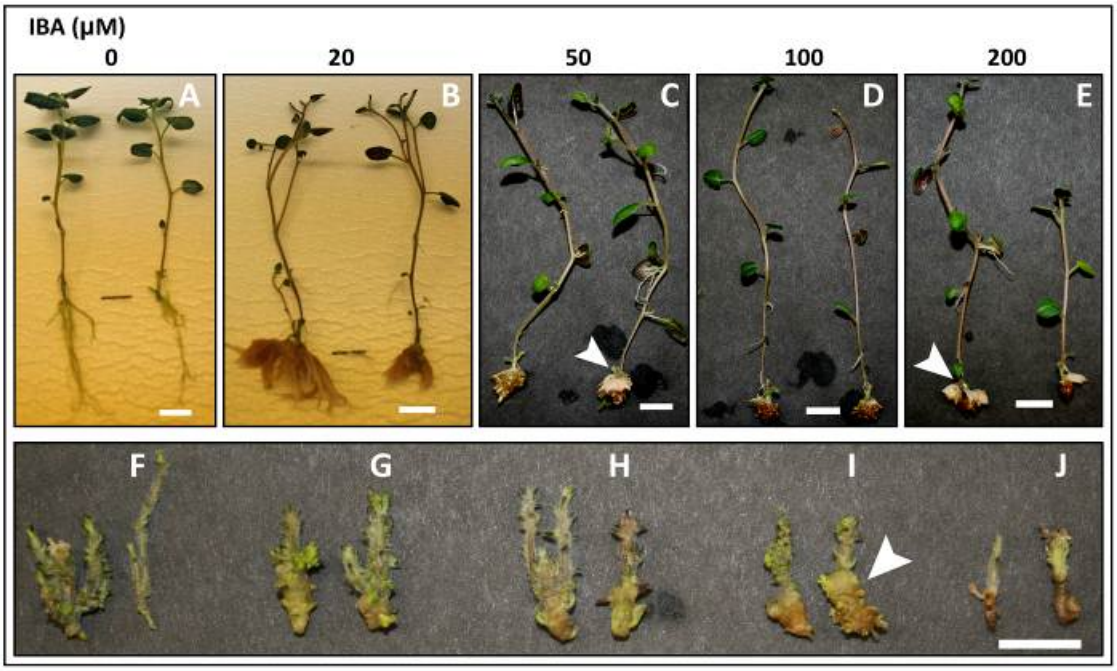


Figure 3.11 Potato explants cultured with different amounts of Indole-3-butyric acid (IBA) in the culture medium. Bintje (upper panel) and *nikku* (lower panel) plants after one month of culture on MS medium with vitamins (A, F) and MS medium with vitamins supplemented with 20 µM (B, G), 50 µM (C, H), 100 µM (D, I) 200 µM (E, J) IBA. Arrowheads- callus like structures. (Scale bar= 1 cm).

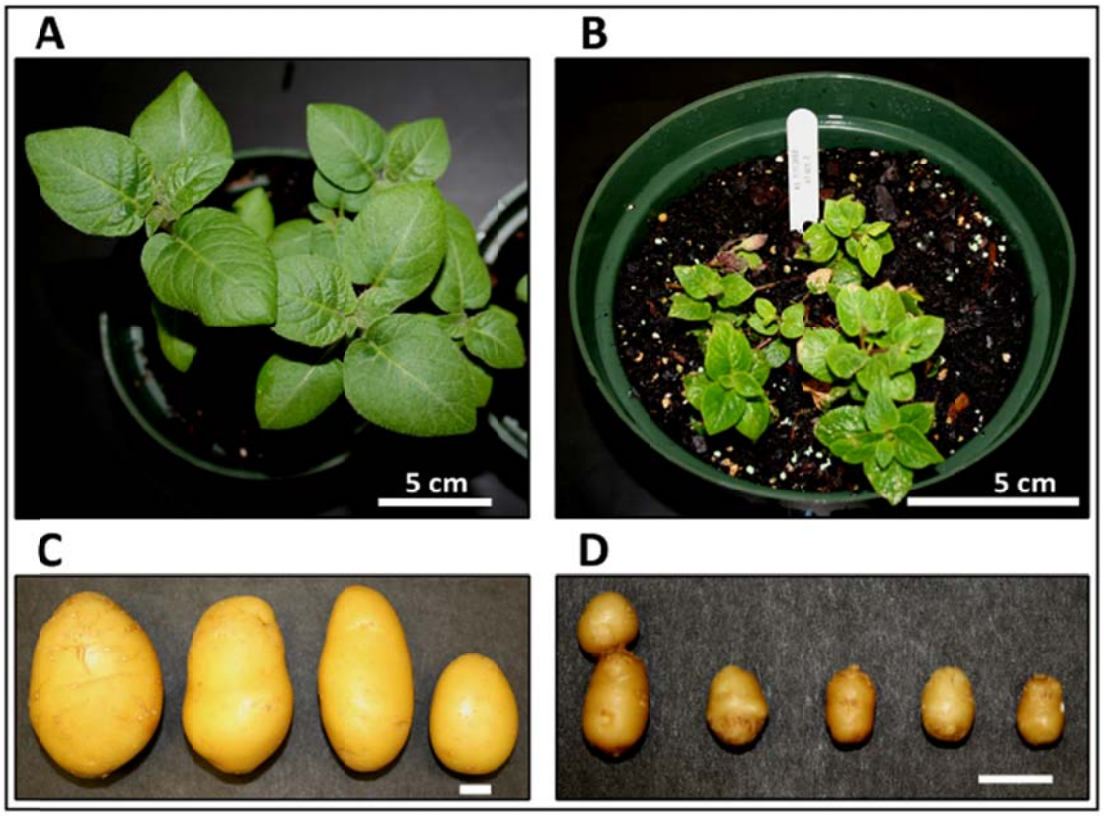


Figure 3.12 Potato phenotyping under greenhouse conditions. (A) 15 day-old Bintje plant and (B) 8-month old *nikku* plant grown from microtubers. (C) Bintje tubers harvested after 3 months of growth and (D) *nikku* tubers harvested from 8 month old plants. (Scale bar in C and D = 1 cm).

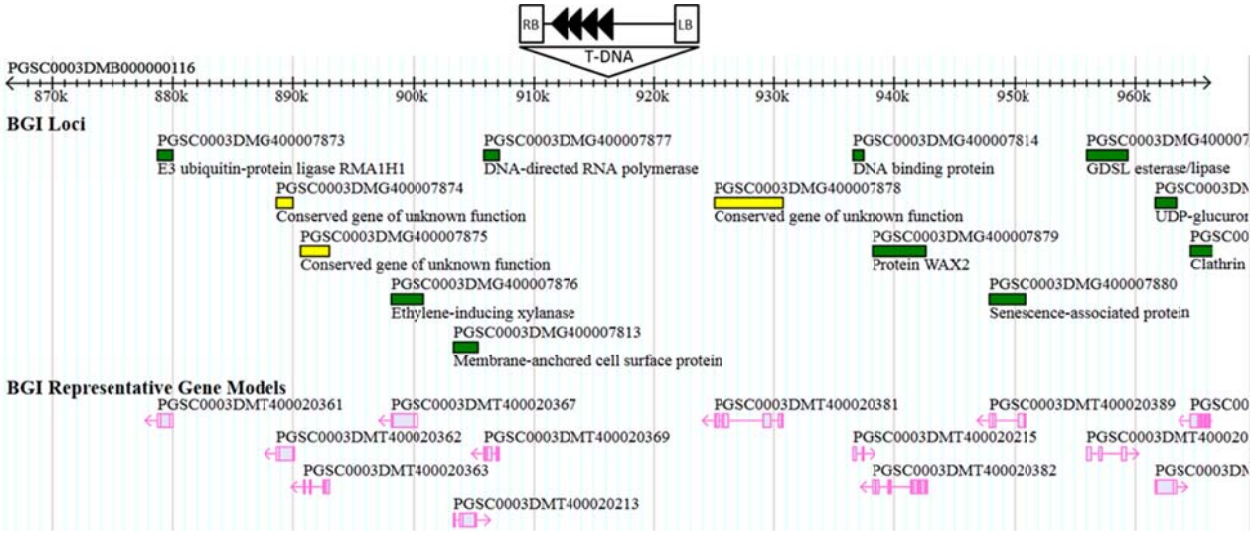


Figure 3.13 Identification of potato genomic regions flanking the T-DNA insertion in the *nikku* mutant. A screen shot of a 100 kb region of scaffold PGSC0003DMB000000116 of the potato DM1-3 516R44 PGSC Genome Assembly, showing the T-DNA insertion site and the flanking gene prediction models.

Gene serial number	Gene identifier PGSC0003DMG	Gene annotation based on potato genome	Direction of the transcript	Gene model distance from RB (kb)	RT-PCR	
					Bintje	*nikku*
	Control	*S.tuberosum* elongation factor 1-α (EF1-α)				
R1	400007877	DNA –directed RNA polymerase	⇐	3.4		
R2	400007876	Ethylene-inducing xylanase	⇐	15.3		
R3	400007845	Conserved gene of unknown function	⇐	23		
R4	400007874	Conserved gene of unknown function	⇐	26		
R5	400007873	E3 ubiquitin-protein ligase RMA1H1	⇐	36		
R6	400007872	NBS-LRR protein	⇐	68		
R7	400007868	Ribosomal-protein-alanine acetyltransferase	⇐	117		
R7.1	400007866	Ribosomal-protein-alanine acetyltransferase	⇐	159	No product	
R8	400007865	Sugar transporter	⇐	172		
R9	400007812	Conserved gene of unknown function	⇒	176		
R10	400007811	Conserved gene of unknown function	⇒	202		
R11	400007862	SRF3	⇐	227		
R16	400007807	14-3-3 protein	⇒	315		
R18	400007806	6b-interacting protein 1	⇒	325		
R19	400007856	F-box family protein	⇐	337		
R26	400007792	Vacuolar cation/proton exchanger	⇒	574		
R27	400046891	Glucose-1-phosphate adenylyltransferase	⇒	596		
R30	400007788	WRKY transcription factor	⇒	663		
R31	400007786	2,4-D inducible glutathione S-transferase	⇒	700		
R32	400007843	Ethylene receptor 1	⇐	705		

Figure 3.14 Reverse transciptase PCR analysis of various potato genes flanking the right border of the activation tag T-DNA insertion of the *nikku* mutant. RNA was extracted from *in vitro* plantlets (leaves+stem). In all gel images the transcript from the mutant is on the right side and the transcript from wild-type Bintje is on the left. The direction of the arrow indicates the orientation of the predicted open reading frame.

Gene serial number	Gene identifier PGSC0003DMG	Gene annotation based on potato genome	Direction of the transcript	Gene model distance from LB (kb)	RT-PCR	
					Bintje	*nikku*
	Control	*S.tuberosum* elongation factor 1-α (EF1-α)				
L1	400007878	Conserved gene of unknown function	⇐	9		
L2	400007879	Protein WAX2	⇐	22		
L3	400007880	Senescence-associated protein	⇐	32		
L4	400007815	GDSL esterase/lipase	⇒	40		
L5	400007816	UDP-glucuronate 5-epimerase	⇒	46		
L6	400007881	Clathrin assembly protein	⇐	49	No product	
L7	400007882	Eukaryotic peptide chain release factor subunit 1-1	⇐	55		
L8	400007817	Oxysterol-binding protein	⇒	60	No product	
L11	400007884	Serine/threonine-protein phosphatase	⇐	79		
L12	400007818	Conserved gene of unknown function	⇒	86		
L13	400007819	DNA-binding protein phosphatase 2C	⇒	106		
L23	400007825	Translation initiation factor	⇒	339		
L26	400007899	CBL-interacting protein kinase 7	⇐	403		

Figure 3.15 Reverse transciptase PCR analysis of various potato genes flanking the left border of the activation tag T-DNA insertion of the *nikku* mutant. RNA was extracted from *in vitro* plantlets (leaves+stem). In all gel images the transcript from the mutant is on the right side and the transcript from wild-type Bintje is on the left. The direction of the arrow indicates the orientation of the predicted open reading frame.

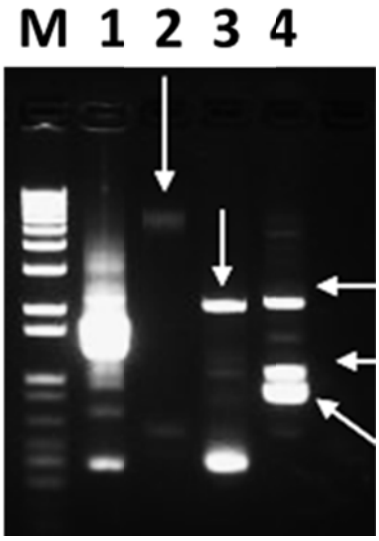


Figure S3.1 GenomeWalker secondary PCR products. Gene specific primers (GSP1) and nested gene specific primers (GSP2) were designed against the mannopine synthase gene sequences and used with arbitrary primer 1 (AP1) and arbitrary primer 2 (AP2) to amplify potato genomic regions flanking the left border of the T-DNA. Secondary PCR products from GenomeWalker are shown here. Lane 1, *Ssp*I digestion; lane 2, *Sca*I digestion; lane 3, *Nae*I digestion; and lane 4, *Msc*I digested mutant *nikku* genomic DNA GenomeWalker libraries. (M= Invitrogen 1 kb+ DNA mass ladder).

Table S3.1 Primers used for PCR amplification of pSKI074 vector sequences, backbone sequences, GenomeWalker and cloning of NAT1 and PMT1 genes.

Primer name	Primer sequence	Purpose
pSKI074-Fred-5'	GCGTGGCTTTATCTGTCTTTGTATTG	Screening of transgenic plants
pSKI074-Red-3'	GGCCTACTTTAATTGCTTCCAGTGTTA	Screening of transgenic plants
Aulakh35F	TGGCAAGTGTAGCGGTCACG	Southern blot probe
Aulakh35R	CTTGCTTTGAAGACGTGGTTGGAACGT	Southern blot probe
Adapter Primer 1 (AP1)	GTAATACGACTCACTATAGGGC	GenomeWalker
Nested AP1 (AP2)	ACTATAGGGCACGCGTGGT	GenomeWalker
SKI074H3 (GSP1)	GCTCTCTCGAGGTCGACGG	GenomeWalker
Nested SKI074H3 (GSP2)	GAGGTCGACGGTATCGATAAGC	GenomeWalker
AT601 Primer 1	GTGTCATTAAATAATTGTCATTTTGCAAC	GenomeWalker primers to amplify flanking LB sequence
AT601 Nested to Primer 1	GAGGGAATAGTACTTTCATGTTAGTGGATGGT	GenomeWalker primers to amplify flanking LB sequence
Potato AT primer set 4-5'	AAGGGAATAAGGGCGACACGGAAA	Right border T-DNA
pSKI074-35 S promoter-3'	CACTGATAGTTTCGGATCTAGATATC	Right border T-DNA
pSKI074-Fred-5'	GCGTGGCTTTATCTGTCTTTGTATTG	Right border T-DNA
Potato AT primer set 4-3'	CCAACTTAATCGCCTTGCAGCACA	Right border T-DNA
pSKI074-8068-8090-5'	CAACAGAGCCTGGCGTTCCCCTT	Left border T-DNA
pSKI074-Red-3'	GGCCTACTTTAATTGCTTCCAGTGTTA	Left border T-DNA
pSKI-4242..71-5'	GATATCTAGATCCGAAACTATCAGTGTTTG	Right border backbone
pSKI-4266..90-5'	TGTTTGACAGGATATATTGGCGGGT	Right border backbone
pSKI-4359..82-3'	GGCATGCACATACAAATGGACGAA	Right border backbone
pSKI-4573..93-3'	GCGATCGAGGATTTTTCGGCG	Right border backbone
pSKI-4874..97-3'	CGATATCATTACGACAGCAACGGC	Right border backbone
pSKI-5379..5402-3'	AGACGAACGAAGAGCGATTGAGGA	Right border backbone
AT601 offLB-5'	GTGTCATTAAATAATTGTCATTTTGCAAC	Left border backbone
pSKI074-Red-3'	GGCCTACTTTAATTGCTTCCAGTGTTA	Left border backbone
5'StGCN5 ORF *Xho*I	AAAACTCGAG ATGGCAGAGATAGTTGAATTGCAGAG	Cloning of NAT1 (GCN5) gene
3'StGCN5 ORF *Xba*I	GGGGTCTAGA TTACTCCATGTCAAAATCCAAATACATTCTATAGG	Cloning of NAT1 (GCN5) gene
5'StSTP ORF *Xho*I	AAAACTCGAG ATGGGTATGGCAGGTGTCCAAGA	Cloning of PMT1 (STP) gene
3'StSTP ORF *Xba*I	GGGGTCTAGA TCACTTCTGCACTAGATGTTCAACATCTCC	Cloning of PMT1 (STP) gene

Table S3.2 Primers used for semi-quantitative PCR (RT-PCR). R1 to 32 are primers for genes flanking the right border and L1-L26 are primers for genes flanking the left border (LB= Left border, RB= Right border).

Primer name	Forward primer sequence	Reverse primer sequence
StEF 1-α	TGATTGAGAGGTCTACCAACCTCGA	GTTCCTTACCTGAACGCCTGTCA
AT601R1	CCGAATTTCGATATCAGCAGTT	TGATGCAAGTAGAAGACGAAGG
AT601R2	TATATCTTATTGCCCATGTTTTCC	GGTTCACTGTAGTAACAATAATCAGG
AT601R3	TAAGTCGCAAATAAAAGTCAGCA	AGGACCAGTTCCTATTCAAGAAG
AT601R4	TCGTTGAATCTTTAGCAATCACAT	AAGCATCATTAGCATCTACCAAGG
AT601R5	TGACTATACCAAAATTCGGGGC	TGGATGTTGTACATGATCCTGTTG
AT601R6	TTATATCTGCAAAATCTTGAGGGAT	CTTCTTGTCATGGAAGGATTTAAGC
AT601R7	GATACAAGCAGAGAGTGAAGAAGG	AGTGAGAATAGAGAAGAAGATTTTCC
AT601R8	AACAACAACTCTATCTGCTCCAAC	AAATCTTTCCCTTAAGGCTTCGT
AT601R9	TTGGAGAGTTGTTCGTTTACATTC	CTTTCTTGAGAGCCTTGAACATAG
AT601R10	TGCAGGGTAATGTTTGGATTTC	CAAGAGAGGGCAACTTCAGTATT
AT601R11	AGCCATCCTCATGCGAGTATTC	CAAAAGGATGATGAGTTTCTTGAC
AT601R16	GATGGTCGATGCAATGAAGAATG	GGAATCAGATGCTCATCAATCACA
AT601R18	AGACACCTCCGTTGCTTCCTCCT	CTGCTTAGCATCTTCAACTCGTTCG
AT601R19	TGTGGTGGCCAGTAATCAAGC	GTGTCGGTGATCTTTTAACTCCTT
AT601R26	ACTTTACTAGTGCTTGGATGTGC	TGGTTGGTCAGCTGAAAAAAT
AT601R27	GATGCGGTCAGACAGTATTTATGG	GCATGTACCTGCATTTCTTTCAAC
AT601R30	GAACCAAGATTTGCCTTCAT	TTAATATTTTACCTGGGGAAAG
AT601R31	CAAGAATTGCATTAGATGAAAAAGG	GTCAATAAAATCAGCCCAAAACC
AT601R32	CCTGATCAGCTACCACTTCAAC	AATTCTCCTGTCGCAAAACTTC
AT601L1	AGGGAAGTGAAGGATTATCAGC	GGAACAATTCCCACAAAGTAGA
AT601L2	ACTGCTGCTGTTAATGTATTTCC	CCTTTTGCAACAAATGGTATTAA
AT601L3	CATCCAACAGAGTAAACAAGAATG	GCAAACCATCAAATGACTGC
AT601L4	TTAAGGCTTTGGGCTTTCTTG	CATTACCTGAGGACGTTGAACC
AT601L5	AACGGCGTCGTATTTGAGTT	CACCGTCTCCACGTTTTTTC
AT601L6	CTATCTTAACGTACTTCTGCCCT	GGACTATTCTACGTGGATTCG
AT601L7	TAGTATCAGGCCGGAGACATTG	CGCGAGGTCCTTCATAAGTTTA
AT601L8	CCATCTGAAGATGGTCTTAGTTCT	CTCCCCCATTCATATGCTTGAT
AT601L11	GGGGATTTTGGTCAGACATACG	TCAGGAACAAAATCAAGTTCATG
AT601L12	TCCTTGCTATGGGTTTCTCAG	CCTTGACGCCAGTCCATAATA
AT601L13	GCAGATCATTACTCGTGGCAAAT	GGAACTCGTCCTCCTTAGTCAGC
AT601L23	CAGGTACATCCTCATGCATCTATTGA	TCCTGATCTCATTAACACTGGGTTG
AT601L26	CCACAAGTTGTATGCAATAAACCATC	CGAATTGTTCGAAAAAGTAGCTAAAG
AT601-LEA	ATGGCTGATCAGTACGAACAGAAC	CGTCACTCGAGCTACTAGAGCTGC
St CBF1/2	GCCAGCTGGAAGGAAGAAGTTTC	GCCGCCTTTTGAATATCTTTAGAG
St APX	CAAGGCTGTTGACAAATGTAAAAGG	GTGGAAATCAGCATGGGAGAGG
St P5CS1	TGGAATCTTTTGGGATAATGACAG	CTAGTTATCACAACGGGAATGCC

References

Ahn, J.H., et al. (2007). Isolation of 151 mutants that have developmental defects from T-DNA tagging. Plant and Cell Physiology 48:169-178.

Aluri, S., and Buettner, M. (2007). Identification and functional expression of the *Arabidopsis thaliana* vacuolar glucose transporter 1 and its role in seed germination and flowering. Proceedings of the National Academy of Sciences of the United States of America 104:2537-2542.

Ayliffe, M.A., Pallotta, M., Langridge, P., and Pryor, A.J. (2007). A barley activation tagging system. Plant Mol. Biol. 64:329-347.

Ayliffe, M.A., and Pryor, A.J. (2007). Activation tagging in plants - generation of novel, gain-of-function mutations. Aust. J. Agric. Res. 58:490-497.

Benhamed, M., Bertrand, C., Servet, C., and Zhou, D.-X. (2006). *Arabidopsis* GCN5, HD1, and TAF1/HAF2 interact to regulate histone acetylation required for light-responsive gene expression. Plant Cell 18:2893-2903.

Benhamed, M., et al. (2008). Genome-scale *Arabidopsis* promoter array identifies targets of the histone acetyltransferase GCN5. Plant J. 56:493-504.

Brownell, J.E., and Allis, C.D. (1996). Special HATs for special occasions: Linking histone acetylation to chromatin assembly and gene activation. Curr. Opin. Genet. Dev. 6:176-184.

Buettner, M. (2010). The *Arabidopsis* sugar transporter (AtSTP) family: an update. Plant Biol. 12:35-41.

Chen, Z.J., and Tian, L. (2007). Roles of dynamic and reversible histone acetylation in plant development and polyploidy. Biochimica Et Biophysica Acta-Gene Structure and Expression 1769:295-307.

Chevalier, D., et al. (2005). STRUBBELIG defines a receptor kinase-mediated signaling pathway regulating organ development in *Arabidopsis*. Proceedings of the National Academy of Sciences of the United States of America 102:9074-9079.

Eyueboglu, B., et al. (2007). Molecular characterisation of the STRUBBELIG-RECEPTOR FAMILY of genes encoding putative leucine-rich repeat receptor-like kinases in *Arabidopsis thaliana*. BMC Plant Biol. 7.

Fernandez, O., Bethencourt, L., Quero, A., Sangwan, R.S., and Clement, C. (2010). Trehalose and plant stress responses: friend or foe? Trends Plant Sci. 15:409-417.

Gao, M.-J., et al. (2009). Repression of Seed Maturation Genes by a Trihelix Transcriptional Repressor in *Arabidopsis* Seedlings. Plant Cell 21:54-71.

Gopal, J., Chamail, A., and Sarkar, D. (2004). In vitro production of microtubers for conservation of potato germplasm: Effect of genotype, abscisic acid, and sucrose. In Vitro Cellular & Developmental Biology-Plant 40:485-490.

Hayashi, H., Czaja, I., Lubenow, H., Schell, J., and Walden, R. (1992). Activation of a plant gene by T-DNA tagging: auxin-independent growth *in vitro*. Science 258:1350-1353.

Horn, P.J., and Peterson, C.L. (2002). Molecular biology: Chromatin higher order folding: Wrapping up transcription. Science 297:1824-1827.

Hsing, Y.I., et al. (2007). A rice gene activation/knockout mutant resource for high throughput functional genomics. Plant Mol. Biol. 63:351-364.

Imaizumi, R., et al. (2005). Activation tagging approach in a model legume, Lotus japonicus. J. Plant Res. 118:391-399.
Jungkunz, I., Link, K., Vogel, F., Voll, L.M., Sonnewald, S., and Sonnewald, U. (2011). AtHsp70-15-deficient *Arabidopsis* plants are characterized by reduced growth, a constitutive cytosolic protein response and enhanced resistance to TuMV. Plant J. 66:983-995.
Kaplan-Levy, R.N., Brewer, P.B., Quon, T., and Smyth, D.R. (2012). The trihelix family of transcription factors - light, stress and development. Trends Plant Sci. 17:163-171.
Kim, J.-M., et al. (2008). Alterations of Lysine Modifications on the Histone H3 N-Tail under Drought Stress Conditions in *Arabidopsis thaliana*. Plant and Cell Physiology 49:1580-1588.
Kitakura, S., Fujita, T., Ueno, Y., Terakura, S., Wabiko, H., and Machida, Y. (2002). The protein encoded by oncogene 6b from Agrobacterium tumefaciens interacts with a nuclear protein of tobacco. Plant Cell 14:451-463.
Kiyosue, T., Abe, H., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1998). ERD6, a cDNA clone for an early dehydration-induced gene of *Arabidopsis*, encodes a putative sugar transporter. Biochimica Et Biophysica Acta-Biomembranes 1370:187-191.
Klepek, Y.-S., et al. (2010). *Arabidopsis thaliana* POLYOL/MONOSACCHARIDE TRANSPORTERS 1 and 2: fructose and xylitol/H+ symporters in pollen and young xylem cells. J. Exp. Bot. 61:537-550.
Kuromori, T., et al. (2006). A trial of phenome analysis using 4000 Ds-insertional mutants in gene-coding regions of *Arabidopsis*. Plant J. 47:640-651.
Larkin, M.A., et al. (2007). Clustal W and clustal X version 2.0. Bioinformatics 23:2947-2948.
Lee, H.K., Cho, S.K., Son, O., Xu, Z., Hwang, I., and Kim, W.T. (2009). Drought Stress-Induced Rma1H1, a RING Membrane-Anchor E3 Ubiquitin Ligase Homolog, Regulates Aquaporin Levels via Ubiquitination in Transgenic *Arabidopsis* Plants. Plant Cell 21:622-641.
Lee, J.H., and Kim, W.T. (2011). Regulation of abiotic stress signal transduction by E3 ubiquitin ligases in *Arabidopsis*. Molecules and Cells 31:201-208.
Lyzenga, W.J., and Stone, S.L. (2012). Abiotic stress tolerance mediated by protein ubiquitination. J. Exp. Bot. 63:599-616.
Oliver, S.N., Dennis, E.S., and Dolferus, R. (2007). ABA regulates apoplastic sugar transport and is a potential signal for cold-induced pollen sterility in rice. Plant and Cell Physiology 48:1319-1330.
Pao, S.S., Paulsen, I.T., and Saier, M.H. (1998). Major facilitator superfamily. Microbiol. Mol. Biol. Rev. 62:1-+.
Park, J.A., et al. (2003). Isolation of cDNAs differentially expressed in response to drought stress and characterization of the Ca-LEAL1 gene encoding a new family of atypical LEA-like protein homologue in hot pepper (Capsicum annuum L. cv. Pukang). Plant Sci. 165:471-481.
Radouani, A. (2003). Effect of plant growth regulators, plantlet attributes and nutrient medium on *in vivo* and *in vitro* potato tuberization. In: Science Agronomiques Rabat, Morocco: Institut Agronomique & Vétérinaire Hassan II.
Reddy, V.S., Shlykov, M.A., Castillo, R., Sun, E.I., and Saier, M.H. (2012). The major facilitator superfamily (MFS) revisited. FEBS J. 279:2022-2035.

Rubtsov, M.A., Polikanov, Y.S., Bondarenko, V.A., Wang, Y.H., and Studitsky, V.M. (2006). Chromatin structure can strongly facilitate enhancer action over a distance. Proceedings of the National Academy of Sciences of the United States of America 103:17690-17695.

Sambrook, J., and Russell, D.W. (2001). Molecular cloning : A laboratory manual. Cold Spring Harbour: Cold Spring Harbour Laboratory Press.

Servet, C., Benhamed, M., Latrasse, D., Kim, W., Delarue, M., and Zhou, D.-X. (2008). Characterization of a phosphatase 2C protein as an interacting partner of the histone acetyltransferase GCN5 in *Arabidopsis*. Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms 1779:376-382.

Servet, C., Conde e Silva, N., and Zhou, D.-X. (2010). Histone acetyltransferase AtGCN5/HAG1 is a versatile regulator of developmental and inducible gene expression in *Arabidopsis*. Molecular Plant 3:670-677.

Silva, J.A.B., Otoni, W.C., Martinez, C.A., Dias, L.M., and Silva, M.A.P. (2001). Microtuberization of Andean potato species (Solanum spp.) as affected by salinity. Scientia Horticulturae 89:91-101.

Slewinski, T.L. (2011). Diverse Functional Roles of Monosaccharide Transporters and their Homologs in Vascular Plants: A Physiological Perspective. Molecular Plant 4:641-662.

Stockinger, E.J., Mao, Y.P., Regier, M.K., Triezenberg, S.J., and Thomashow, M.F. (2001). Transcriptional adaptor and histone acetyltransferase proteins in *Arabidopsis* and their interactions with CBF1, a transcriptional activator involved in cold-regulated gene expression. Nucleic Acids Res. 29:1524-1533.

The Potato Genome Sequencing Consortium. (2011). Genome sequence and analysis of the tuber crop potato. Nature 475:189-U194.

Veyres, N., et al. (2008). The *Arabidopsis sweetie* mutant is affected in carbohydrate metabolism and defective in the control of growth, development and senescence. Plant J. 55:665-686.

Wan, S.Y., et al. (2009). Activation tagging, an efficient tool for functional analysis of the rice genome. Plant Mol. Biol. 69:69-80.

Weigel, D., et al. (2000). Activation tagging in *Arabidopsis*. Plant Physiol. 122:1003-1013.

Williams, M.E., et al. (2005). Mutations in the *Arabidopsis* phosphoinositide phosphatase gene *SAC9* lead to overaccumulation of PtdIns(4,5)P-2 and constitutive expression of the stress-response pathway. Plant Physiol. 138:686-700.

Willmann, M.R., Mehalick, A.J., Packer, R.L., and Jenik, P.D. (2011). MicroRNAs regulate the timing of embryo maturation in *Arabidopsis*. Plant Physiol. 155:1871-1884.

Wormit, A., et al. (2006). Molecular identification and physiological characterization of a novel monosaccharide transporter from *Arabidopsis* involved in vacuolar sugar transport. Plant Cell 18:3476-3490.

Yamada, K., et al. (2010). Functional analysis of an *Arabidopsis thaliana* abiotic stress-inducible facilitated diffusion transporter for monosaccharides. J. Biol. Chem. 285:1138-1146.

Chapter 4

Characterization of a partial revertant of the potato activation-tagged mutant, *nikku*

Sukhwinder S. Aulakh[1, 2], Richard E. Veilleux[1*] and Barry S. Flinn[1, 2*]

Abstract

The silencing of resident plant genes triggered by transgenes is a widely known phenomenon, although the actual mechanisms are still being clarified. In this study, we performed morphological and molecular analysis of revertant plants of a potato activation tagged mutant, *nikku*. Unlike the *nikku* mutant, which exhibited severe dwarfism, small hyponastic leaves, and little to no root development, the revertant exhibited near normal shoot elongation, larger leaves and consistent rooting. However, the revertant did not completely resemble wild-type Bintje. Reverse transcriptase PCR analysis of selected genes in revertant plants showed that the expression levels of several, but not all, genes which were differentially regulated in the *nikku* mutant were similar to wild-type Bintje gene expression levels. The reversion appeared partial, and was not the result of a loss of enhancer copies from the original *nikku* mutant, as there was no change in transgene structure and location in the revertant. Our PCR analysis using methylated DNA as template indicated a change in methylation status of the 35S enhancers between *nikku* and revertant plants, suggesting that the *nikku* mutant reversion phenotype may be associated with some epigenetic modification.

Keywords: *Solanum tuberosum*, methylation, 35S, gene silencing, epigenetics

[1]Department of Horticulture, Virginia Tech, Blacksburg, Virginia, USA, 24060

[2]Institute for Sustainable and Renewable Resources (ISRR) at Institute for Advanced Learning and Research (IALR), Danville, Virginia, USA, 24540

* Both contributed to project development and supervision.

Introduction

Insertional mutagenesis is a useful approach for functional gene discovery, as evident from the concentrated development of both loss-of-function and gain-of-function mutant populations for several plant species. The most widely used constitutive promoter in many T-DNA cassettes for plant transformation was derived from the 35S RNA of cauliflower mosaic virus (35S CaMV). Transcriptional enhancers derived from the 35S CaMV promoter were first used in plants in the T-DNA of a plasmid for *Agrobacterium*-based transformation of Arabidopsis to generate gain-of-function mutants (Hayashi et al., 1992); this strategy has become known as activation tagging. Several activation tagging vectors were constructed or modified, which included transposons or other complex constructs, and these have been used successfully to discover gene functions in various plant species (Ayliffe et al., 2007; Borevitz et al., 2000; Busov and Strauss, 2007; Busov et al., 2011; Busov et al., 2003; Jeong et al., 2002; Kardailsky et al., 1999; Marsch-Martinez et al., 2002; Mathews et al., 2003; Suzuki et al., 2001; van der Graaff et al., 2000; van der Graaff et al., 2003; Walden et al., 1994; Weigel et al., 2000). A typical activation tagging vector has strong enhancer elements near one end of the T-DNA, theoretically to increase the expression of genes within a few kilobase pairs of chromosomal DNA, thus producing a gain-of-function phenotype. In activation tagged lines with no visible phenotype, there is a possibility that the 35S promoters and/or enhancers were not transcribed or were silenced. Silencing of the endogenous resident genes is another possibility, although the exact mode of action is unclear. Homology-dependent epigenetic gene silencing and methylation of transgenes are possible mechanisms of gene silencing. Higher plants are predominantly methylated at the dinucleotide CpG and trinucleotide CpNpG (where N is any nucleotide) in the form of 5-methylcytosine (Harding, 1994). DNA methyltranferase recognizes hemi-methylated CpG sites and methylates the cytosine in the unmethylated strand. Transgenes might be silenced by transcriptional gene silencing (TGS) or post-transcriptional gene silencing (PTGS); in either case hyper-methylation of the transgene generally occurs in the transgene promoter region (Mishiba et al., 2010). Additionally, the copy number of a transgene is also thought to be a contributing factor to transgene silencing (Yamasaki et al., 2011b). Alternative activation tagging vector systems like pAYDH, pAYDB and pAYDB8, which are based on the promoter from the pararetrovirus (virus that contain a gene for reverse transcriptase, but cannot insert

itself in a host genome), cassava vein mosaic virus (CsVMV), have been developed to overcome the 35S promoter-induced gene silencing (Dong and von Arnim, 2003). However, vectors consisting of enhancers derived from the CaMV promoters are still widely used.

The activation tagging vector pSKI074 comprises tetramerized 35S enhancers, T3 and T7 RNA polymerase promoters, kanamycin selection marker [neomycin phosphotransferase (*npt*)], left and right borders and the mannopine synthase terminator. The CaMV 35S enhancers in the pSKI074 vector correspond to nucleotides -417 to -86 relative to the transcription start site. This vector was first used to generate gain-of-function mutants in Arabidopsis (Weigel et al., 2000). During this study, reversion of the mutant phenotype was reported for several mutants where the dominant phenotype was gradually lost in subsequent generations. Recently we characterized two activation tagged potato mutants, *underperformer* (Chapter 1) and *nikku* (Chapter 3), generated from potato cv. Bintje using the activation tagging vector pSKI074. Mutant *nikku* plants were dwarf and rootless, slow growing, had hyponasty and rudimentary leaves, and produced either no or occasional small tubers. A variety of genes on either side of the T-DNA insertion was modified in their expression, suggesting a constitutive stress response phenotype. During routine subculture of *nikku* shoots growing *in vitro*, we observed that in some cases the mutant phenotype was unstable and revertant plants were observed. These presumed revertant plants formed functional roots, developed more prominent leaves and were larger than *nikku* plants. The revertant plants maintained their phenotype through subsequent subcultures. Morphological and molecular analyses of these revertant plants were performed and the revertants were compared to the original *nikku* mutant and wild-type Bintje plants. Changes in expression of genes flanking the T-DNA were studied using semi-quantitative PCR and compared with 'Bintje' and *nikku* mutant. We also examined the methylation status of 35S enhancer sequences and present evidence suggesting that the differential methylation of the inserted 35S CaMV enhancers may contribute to the differences between the *nikku* and *revertant* phenotypes.

Results

Identification of the *nikku* revertant phenotype

The revertant plants exhibited a phenotype which was intermediate between the original *nikku* mutant and wild-type Bintje phenotype, but certainly more similar to Bintje. The original *nikku* plantlets were characterized *in vitro* by slow growth, lack of roots, hyponastic and rudimentary leaves and little to no tuber formation. A detailed description of *nikku* was provided in Chapter 3. After several generations (>6) of subculture on plantlet growth medium (PGM), some *nikku* plants formed roots, their leaves became less hyponastic and larger than the original *nikku* mutant and the stems grew longer in length (**Figure 4.1**). We termed these phenotypically different plants as revertants. However, the revertant phenotype also differed from wild-type Bintje plants and appeared to represent a case of partial reversion. We confirmed the presence of a single T-DNA insertion in *nikku* plants by Southern blot experiments and the T-DNA was located on superscaffold PGSC0003DMB000000116:910110 on chromosome 12 when sequences obtained using GenomeWalker were queried against the potato genome PGSC_DM_v3_superscaffolds.fasta and PGSC_DM_v3_2.1.10_pm.fasta (accessed on 02-10-2011) databases, respectively (Chapter 3). Since revertant plants were vegetatively propagated from *nikku* mutants, it was expected that the T-DNA copy number and location were similar to the original *nikku* mutant; this assumption of similar location was later confirmed by sequencing PCR products amplified from genomic DNA of revertant plants spanning the plant/T-DNA junction (data not shown).

In vitro plant growth and tuberization

After our original observation of differences between the growth habit of *nikku* and revertant plants, we carried out *in vitro* growth and tuberization experiments to quantify these differences and further characterize the revertant plants. We found the mean stem length of revertants (4.1 cm) after 4 weeks of growth was significantly longer than observed for *nikku* (0.9 cm), but not significantly different from Bintje (4.7 cm). The mean number of roots in the revertant (7.5) was significantly greater than for Bintje (4.7) and the rootless *nikku* plants (**Figure 4.2**). The revertant plants also produced 3x more axillary branches than both Bintje and *nikku* plants. *In vitro* tuberization experiments revealed some variation in phenotype between

Bintje, *nikku* and the revertant *(***Figure 4.1** and **Figure 4.4**). The mean number of tubers/explant (0.8) and weight of tubers/explant (0.2 g) for the revertant were not significantly different from *nikku* (number of tubers = 0.2 and weight of tubers = 0.04 g) but were significantly less than Bintje (number of tubers = 2.4 and weight of tubers = 2.3 g). However, the mean number of stolons/explant was significantly greater for the revertant (3.2) than *nikku* (0.7) and Bintje (0.2) explants (**Figure 4.4**)**.** We found that most revertant plants were morphologically different, in that they were bushier than *nikku* and Bintje plants (**Figure 4.3A-C**). However, there was little difference in tuber shape and size between revertant and *nikku* plants, and both formed tuberoid like structures. However, both *nikku* and the revertant had much smaller tubers than those observed with Bintje (**Figure 4.3D-F**), which were also rounder and smoother. Additionally, during the tuberization experiment various revertant plants appeared to express different degrees of reversion (**Figure S4.1**).

The revertant plant phenotype of soil-grown plants

Some revertant plants were grown under greenhouse conditions from microtubers and compared with similarly propagated Bintje (**Figure 4.5**). However, there were no soil-grown *nikku* mutants as our efforts to grow them in soil were unsuccessful. The revertant plants had larger terminal leaves without lateral leaflets, resulting in a simple leaf compared to the typical compound leaf of wild-type Bintje (**Figure 4.6**). The petioles of the revertant were shorter and the rachis was absent. Additionally, leaf trichomes appeared more prominent in revertant plants than Bintje (see inset - **Figure 4.6**). We also observed differences in tuber set and tuber yield between Bintje and the revertant (data not reported).

Determining the presence of 35S enhancers in the revertant

The T-DNA of activation tagging vector pSKI074 consisted of four copies of 35S enhancers, which had been integrated into the *nikku* genome during transformation. One possible underlying cause of the reversion phenotype was the loss of 35S enhancer sequences from the *nikku* mutant during subculture. In order to check this, we amplified different T-DNA components of revertant genomic DNA by PCR. We found that the revertant genomic DNA had both right and left T-DNA border regions, and the correctly-sized tetramerized enhancer

fragment. The T-DNA fragments amplified near the right border of pSKI074, using revertant genomic DNA and primers designed from T-DNA sequences (**Table S4.1**) were identical in size and sequences to those of *nikku* and these fragments were amplified across the tetramer enhancers (**Figure 4.7A and 4.7B**). Similarly, PCR with primers designed near the left border of the T-DNA also produced identical size gel bands with both revertant and *nikku* genomic DNA (**Figure 4.7C** and **Table S4.1**). Therefore, it appears that revertant plants still harbor the complete T-DNA without loss of enhancer copies. Further PCR analysis at the T-DNA/plant DNA junctions in revertant plants, with primers designed from *nikku* genomic DNA sequences revealed the presence of 277 bp (7776 to 8053 sequence of pSKI074) of left border backbone and no right border backbone (data not shown). These results agreed with the integration of T-DNA in *nikku* reported in Chapter 3. Collectively, these results suggest that the T-DNA structure, size, sequences and position within the genomes of *nikku* and revertant were similar.

Comparison of gene transcripts flanking the T-DNA in Bintje, *nikku* and revertant plants

Transcript analysis of several potato genes flanking the T-DNA insertion was performed using reverse transcriptase PCR (RT-PCR), with primers designed against the reading frame for the various gene model predictions. The genes flanking the right and left border of the T-DNA were given consecutive numbers based on increasing distance from the insert with prefix R-(for right border) and L-(for left border), respectively. Details of primer sequences used in RT-PCR analysis are given in supplementary **Table S4.2**. The expression levels of many genes, which were up-regulated in *nikku* (R3, R4, R5, R7, R8, R11, R27, R30, L4, L11 and L23), were lower in the revertant, similar to levels found in Bintje. Similarly, expression levels of two genes [Ethylene inducing xylanase (R2) and Senescence associated protein (L3)] which were down-regulated in *nikku* plants were more highly expressed in the revertant, again similar to levels found in Bintje (**Figure 4.8**). However, there was no change in expression levels of three genes (R26, L13 and L26) between Bintje, *nikku* and *revertant*. Hence, the revertant was more wild-type-like based on the expression of several genes, although still not completely, further suggesting a partial reversion phenotype.

Differential methylation of 35S enhancer sequences in *nikku* and revertant

Having previously shown that the 35S tetramerized enhancer complement was intact; we pursued another possibility underlying the phenotypic variation between *nikku* and the revertant, i.e., differences in DNA methylation. To address this possibility, we used a methylated DNA enrichment method, followed by PCR amplification of the 35S enhancers from the enriched fraction, to assess methylation levels of this DNA sequence in *nikku* and the revertant. The restriction enzyme *Sau*3AI was used to digest genomic DNA for the different samples after which methylated DNA was selected on magnetic beads using an attached methylated DNA-binding protein. The methylated DNA fraction bound to the beads was eluted, and used for PCR of methylation-enriched DNA with 35X *Sau*3AI primers (**Table S4.1**) designed to amplify individual 321 bp enhancer sequences. We consistently observed fewer PCR products from the revertant than for *nikku*, even though equal amounts of methylation-enriched DNA were used as template (**Figure 4.9**). The PCR results indicated hypo-methylation of enhancer sequences in the revertant compared to the *nikku* mutant. However, based on the methodology of this assay, we do not know what site-specific variation in cytosine methylation may exist between *nikku* and the revertant, and what contribution this may play in the modification of gene expression between the two.

Discussion

Reports of reversion of plant mutant phenotypes in transgenic plants occasionally appear in the literature and in most cases, reversion have been attributed to transgene silencing, resulting from multiple copies of T-DNA integrated into the plant genome (Matzke and Matzke, 1998). Transgene silencing in single copy transgenic plants may also take place in a sequence-dependent manner (Mishiba et al., 2010; Yamasaki et al., 2011b). The underlying reasons for this silencing or under-expression of a transgene are still unclear. Possible reasons for gene silencing are positional effects of an inserted transgene, methylation of the transgene, and homology-dependent silencing of endogenous genes. In activation-tagged mutants, the methylation of enhancer sequences may reduce their ability to over-express adjacent native genes, thereby eliminating the visible phenotype. The low frequency of visible phenotypes (1 in 1000 in Arabidopsis) of dominant gain-of-function mutants generated using activation tagging

vector pSKI074 (Weigel et al., 2000) has been attributed to the methylation of enhancer sequences (Chalfun et al., 2003) in multiple insertions. Endogenous and environmental factors further influence 35S promoter methylation, as revealed in a study involving petunia plants carrying a single copy of the maize *A1* gene, encoding a *dihydroflavonol reductase*, which confers a salmon red flower color (Meyer et al., 1992). In this study, all 13 tested variegated plants showed hyper-methylation of the 35S promoter directing *A1* gene expression and seasonal flower color variation was correlated with methylation of the 35S promoter. After 1 to 2 weeks at 4°C, *Agrobacterium tumefaciens* has been found to lose one CaMV enhancer copy on average due to recombination proficiency of *Agrobacterium* (Weigel et al., 2000). However, this does not appear to be a problem *in planta*, as we did not find evidence of loss of an enhancer sequence after integration into the plant genome. Weigel et al. (2000) reported that several Arabidopsis mutant lines generated using the pSKI074 vector became attenuated in subsequent generations, with a reversion to wild-type phenotype. This was attributed to the selective inactivation of the CaMV 35S enhancers, but no corresponding analysis of the enhancer sequences was provided. Our results indicated that the T-DNA structure and integration site were intact in the revertant plants and the reversion of the *nikku* phenotype was not due to enhancer loss. Although it is hard to explain with certainty the mechanism behind reversion of the *nikku* mutant, our data based on methylated DNA enrichment and PCR analysis of the enriched methylated DNA fraction for *nikku* and the revertant suggest the involvement of differential methylation. The original *nikku* mutant had a single T-DNA insertion; therefore, multiple T-DNA insertions were not responsible for the reversion phenotype. Single copy transgenic plants can exhibit transgene silencing, due to hyper-methylation in the 35S promoter leading to silencing of the transgene (Mishiba et al., 2010). In this study, the authors also observed methylation of the NOS promoter, but the frequency of methylation in the 35S promoter region was much greater and methylation was independent of copy number and integration site. Specifically, hyper-methylation at CpG and CpNpG sites and moderate methylation at CpHpH (where H=A, C or T) were observed in all single copy transgenic plants. The reversion of the *nikku* mutant observed here would suggest some silencing of the promotive effect of the tetramerized 35S enhancer sequences. Our observations of hypo-

methylation of the 35S tetramer enhancer regions of the revertant in comparison to *nikku* are in contrast to the observations of Mishiba et al. (2010), associating 35S enhancer hyper-methylation with gene silencing. These authors identified a consensus target region for *de novo* methylation of the 35S promoter, with distinct footprints associated with the 35S promoter and the enhancer region crucial for methylation. However, tetramerized 35S enhancers were not analyzed in that study. Hence, it is not known how the pattern of cytosine methylation might change in this tetramerized enhancer context during silencing. Unfortunately, the nature of our methodology allowed for the detection of overall methylation levels in single 35S enhancers amplified by PCR, but did not allow us to determine site-specific methylation differences between *nikku* and the revertant. In addition, our methodology did not allow us to determine the pattern of cytosine methylation in the ordered series of 35S enhancers comprising the tetramer for *nikku* and revertant. Therefore, it is possible that the reduced effect of 35S enhancers in the revertant was a result of variation in methylation of specific cytosine residues, rather than the overall level of methylation. Recent progress in high-throughput technologies has greatly contributed to the development of DNA methylation profiling, and further characterization of the 35S enhancer methylation status for both *nikku* and the revertant will require the use of such techniques, such as bisulfite sequencing.

In general, gene regulation is controlled by chromatin modifications and changes in methylation of the promoter/transcribed regions, and these two processes are linked with each other. There are common links that connect these two well-known regulators of gene expression. Methyl-CpG binding domain (MBD) proteins couple DNA methylation to transcriptional repression of genes (Lopez-Serra et al., 2006) and MBD proteins are associated with histone deacetylase, mechanistically linking this process to histone deacetylation-mediated transcriptional repression (Wade et al., 1999). Histone acetylation-deacetylation itself is a reversible process coupled with differential methylation and their involvement in many common development processes; growth, gene regulation, stress response and common pathways during gene regulation may provide some explanation for the reversion of the *nikku* phenotype. Several studies related to methylation and histone modification have appeared recently but few analyze potential common links between these two phenomena (Yamasaki et al., 2011a).

Hence, in addition to 35S enhancer methylation, future studies with *nikku* and the revertant should also explore variation in histone acetylation.

Methods

Screening of *nikku* revertant phenotype

In vitro plantlets of Bintje and its activation tagged mutant *nikku* were received from BioAtlantech (http://bioatlantech.nb.ca/site/solanumgenomics). The *nikku* mutant harbors T-DNA from the activation tagging vector pSKI074, which contains tetramerized 35S CaMV enhancer sequences near the right border. The generation and screening of activation tagged mutant *nikku*, was described in Chapter 3. Bintje and *nikku* plantlets were maintained in *vitro* on plantlet growth medium [PGM; MS (Murashige and Skoog, 1962) minimal organic medium, sucrose 3%, pH 5.7 and solidified with 0.7% agar, sterilized] in magenta boxes. Plantlets were propagated from 1 cm nodal stem sections and kept under 16 h photoperiod, 20˚C day/18˚C night and 70-100 µE $m^{-2}s^{-1}$ light intensity. The revertant plants were first noticed after several generations (>6) of subculture and multiplication of *nikku* plants on PGM. The revertant plants were separated from the original mutant *nikku* plants and maintained separately on PGM for all the experiments.

In vitro plant growth and tuberization

For *in vitro* plant growth assays, wild-type Bintje, the original *nikku* mutant and revertant plants were grown from 1 cm long explants (with at least one node) taken from 4-5 week old *in vitro* plantlets. Explants were placed on MS basal medium with vitamins (Phytotech, Lenexa, KS, USA) supplemented with 3% sucrose, pH 5.8 and solidified with 0.7% agar. Plants were grown under a 16 h photoperiod, 24±1°C and 70-100 µE $m^{-2}s^{-1}$ light intensity in a plant growth chamber. Experiment was conducted in a randomized block design with three replicates. Growth data were recorded after 4 weeks on MS medium. For *in vitro* tuberization assays, stem sections, 3-5 cm long, with 4-5 nodes per section were prepared from 4-5 week old *in vitro* Bintje, *nikku* and revertant plants by removing all leaves. These stem sections were grown in 40 ml liquid propagation medium [MS basal medium + 148 mg l^{-1} NaH_2PO_4 + 0.4 mg l^{-1} thiamine HCl + 100 mg l^{-1} inositol + 3% sucrose, pH 5.9] (Radouani, 2003) at 16 h photoperiod, 24±1°C and 70-100

µE $m^{-2}s^{-1}$ light intensity for 3-4 weeks in 250 ml conical flasks capped with ventilated plugs. After 4 weeks, propagation medium was replaced with microtuberization medium [MS basal + 8% sucrose, 2X the amounts of NH_4NO_3, KH_2PO_4 and KNO_3 and without NaH_2PO_4, pH 5.8] (Radouani, 2003). Culture flasks were kept at 18-20°C in the dark to induce tubers and data were recorded after 30 days of growth on tuberization medium. Each tuberization experiment consisted of three replications in a randomized block design and total of two experiments were conducted. Data were analyzed in JMP, version 9.0.0 (SAS Institute Inc., Cary, NC) and tables and graphs were drawn in Microsoft Excel (2010).

Determination of 35S enhancer copies in the revertant

Genomic DNA was extracted using Plant DNAzol Reagent (Life Technologies; Grand Island, NY, USA) following manufacturer's instructions. PCR analysis was performed using several sets of primers designed to amplify fragments of the pSKI074 T-DNA part of pSKI074 vector from genomic DNA. The primer sequences used in these PCR reactions are given in supplementary **Table S4.1**. Each PCR reaction was performed in total reaction volume of 20 µl, with 5-10 units of ExTaq polymerase (Takara Bio Inc., Japan) and 50-100 ng of DNA template in a MyCycler (Bio-Rad) thermal cycler.

Extraction of RNA and Reverse transciptase PCR

RNA was extracted from 4 week old *in vitro* plantlets. All tissues above the culture medium level were collected (leaves and stems, without roots) and immediately frozen in liquid nitrogen and stored at -80°C until needed. RNA was extracted using either the Concert Plant RNA Reagent (Invitrogen; Carlsbad, CA, USA) following manufacturer's instructions, or a combination of Trizol Reagent (Invitrogen; Carlsbad, CA, USA) extraction and RNA mini kit (Qiagen; Valencia, CA, USA) purification and was previously described in Chapter 3. DNAse treatment of RNA was done using the DNA-free kit (Ambion; Foster City, CA, USA) following the manufacturer's instructions. First strand cDNA was synthesized using SuperScript III (Invitrogen; Carlsbad, CA, USA) using 1 µg of total RNA and following the supplier's protocol. After cDNA synthesis the final volume was adjusted to 100 µl of diluted cDNA for each cDNA synthesis reaction. Reverse transcription polymerase chain reaction (RT-PCR) was performed using gene-specific primers, as mentioned

in supplementary material (**Table S4.2**). Equal amounts of cDNA were used in each reaction and *Solanum tuberosum* elongation factor 1-α (StEF1-α) gene expression was used as control. ExTaq Polymerase from Takara (Takara Bio Inc., Japan) was used in all amplification reactions. Gel images were acquired using an Alpha Innotech (Cell Biosciences, San Leandro, CA, USA) gel doc system.

Methylation DNA enrichment

Methylated DNA was enriched using the EpiXplore Methylated DNA enrichment Kit (Clontech, Mountain View, CA, USA) following the manufacturers protocol. In brief, for each enrichment reaction, first 100 µl of the TALON magnetic beads were prewashed twice with 100 µl of 1x binding/wash buffer in a 1.5 ml nuclease free microcentrifuge tubes and re-suspended in 100 µl of 1x binding/wash buffer. Diluted 100 µl methyl-CpG binding domain (MBD2) protein 2 [10 µl of MBD2 and 90 µl 1x binding/wash buffer] were transferred to each tube containing 100 µl of TALON magnetic beads. The mixture was incubated on a platform shaker for 1 h at room temperature to allow coupling of MBD2 protein with TALON magnetic beads. After binding, the MBD2/TALON magnetic bead mixture was washed three times with 100 µl of 1x binding/wash buffer. Bintje, *nikku* and revertant genomic DNA was digested with restriction enzyme *Sau*3AI, with four *Sau*3AI cleavage sites located within the tetramerized enhancer sequences of T-DNA (**Figure S4.2**). For each methylated DNA enrichment reaction, 25 µl of 4x binding/wash buffer and 1 µg of digested genomic DNA were added to a microcentrifuge tube and volume were adjusted to 100 µl with nuclease free water. These diluted DNA samples were transferred to tubes containing the washed MBD2 protein/TALON magnetic bead mixture and tubes were incubated on a platform shaker for 1 h at room temperature. Tubes containing the mixture were centrifuged briefly and placed on a magnetic separator to allow the magnetic separator to pull the beads to the wall of the tubes. The supernatant was removed and transferred to a fresh nuclease free tube and labeled as Fraction I (consisting of non-methylated and hypo-methylated DNA). MBD2 protein/TALON magnetic beads were washed twice with 200 µl of 1x binding/wash buffer and the supernatant separated, pooled and labeled as Fraction II (wash). The enriched methylated DNA from MBD2 protein /TALON magnetic beads complex was eluted twice with 200 µl of elution buffer (high), pooled and labeled as Fraction III (methylated DNA).

DNA fraction III was used for downstream PCR reactions. To perform PCR with methylated enriched DNA of *nikku* and revertant plants, we designed primers within the *Sau*3AI digested products (**Figure S4.2** and **TableS4.1**) to give a PCR product of 321 bp.

Plant growth in soil

Microtubers of Bintje and revertant plants produced during *in vitro* tuberization experiments (this study) were planted in 15 cm diam pots filled with commercial Miracle-Gro potting mix (Scotts Miracle-Gro, Marysville, OH). These were grown in spring 2012 (2-7-2012 to 5-20-2012) in the greenhouse of the Institute for Sustainable and Renewable Resources (ISRR) at the Institute for Advanced Learning and Research (IALR), Danville, Virginia, USA. Standard plant growth practices were followed. The greenhouse settings were 24-30°C temperature and day-length was extended to 16h with 1000 watt, 5.7 Kelvin daylight spectrum metal halide bulbs.

Acknowledgments

We wish to thank BioAtlantech for supplying the activation tagged mutant *nikku* and Bintje germplasm. We also want to thank Guozhu Tang for her help with GenomeWalker in the original *nikku* mutant. This work was funded through Special Grants (2003-38891-02112, 2008-38891-19353 and 2009-38891-20092) from the United States Department of Agriculture, and operating funds from the Commonwealth of Virginia to the Institute for Advanced Learning and Research.

Figure 4.1 Comparison of *in vitro* plants of Bintje, *nikku* mutant and *nikku* revertant. One cm nodal stem explants cultured on Murashige and Skoog (MS) basal medium with vitamins supplemented with 3% sucrose. Plants were grown a using 16 h photoperiod, 24±1°C and 70-100 µE $m^{-2}s^{-1}$ light intensity in a plant growth chamber. Pictures were taken after 4 weeks of growth. (Scale bar = 1 cm).

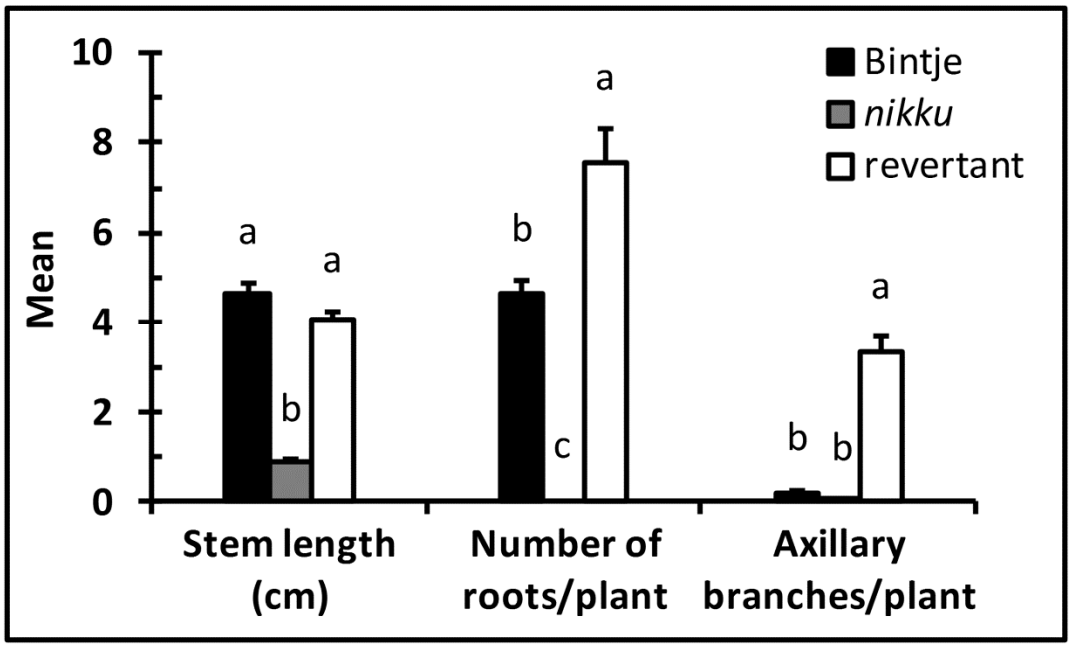


Figure 4.2 *In vitro* plant growth data for Bintje, *nikku* mutant and *nikku* revertant. Data were recorded following 4 weeks of culture. The values are means ± SE (n=13). Statistical differences were calculated by one-way ANOVA. Different letters indicate means that were statistically different by Tukey's HSD multiple testing method ($P < 0.05$) for genotypes within a given growth parameter.

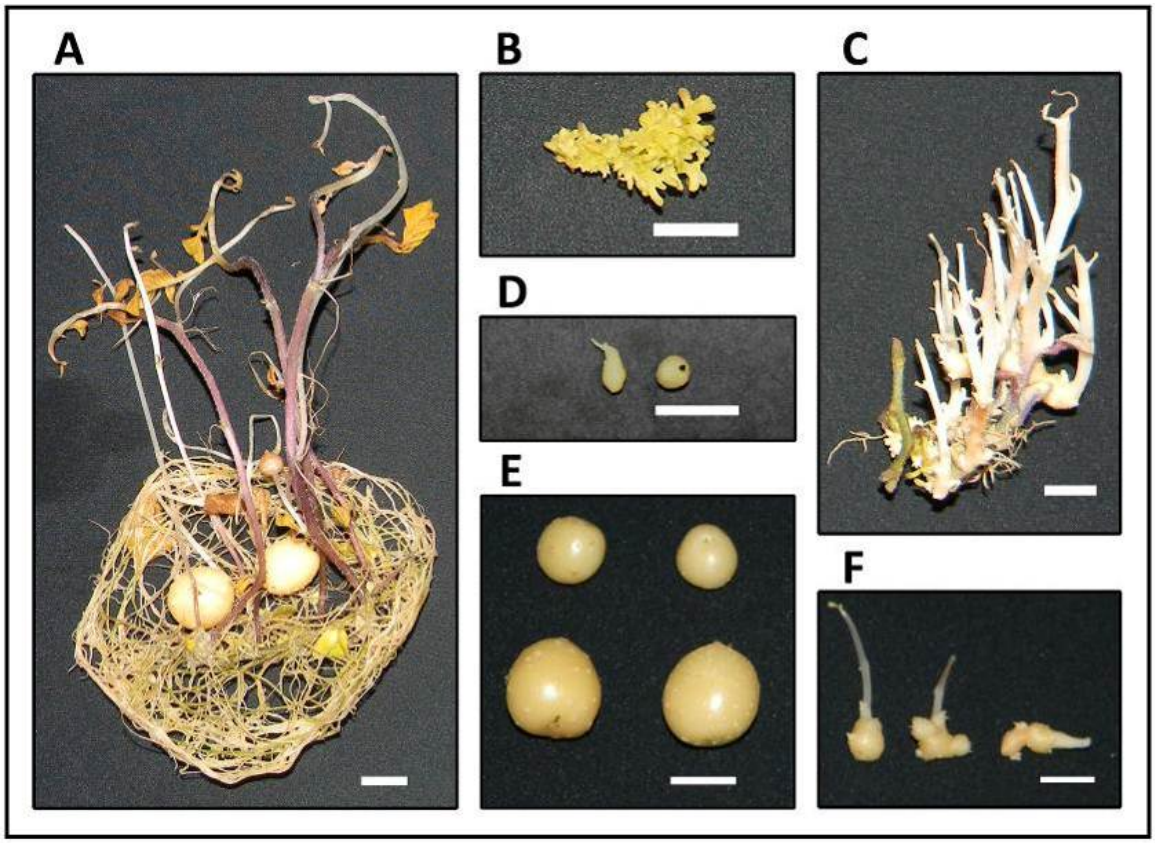


Figure 4.3 *In vitro* tuberization. Bintje (A), *nikku* (B) and *nikku* revertant (C) explants after 30 days of growth in tuberization medium. Harvested tubers of Bintje (E), *nikku* (D) and *nikku* revertant (F) at the end of the tuberization experiment. Stem sections of 3-5 cm in length were grown in 40 ml liquid propagation medium at 24±1°C, 16h photoperiod and 70-100 µE $m^{-2}sec^{-1}$ light intensity for 4 weeks in followed by 30 days of plant culture in microtuberization medium at 18-20°C in the dark. (Scale bar = 1 cm).

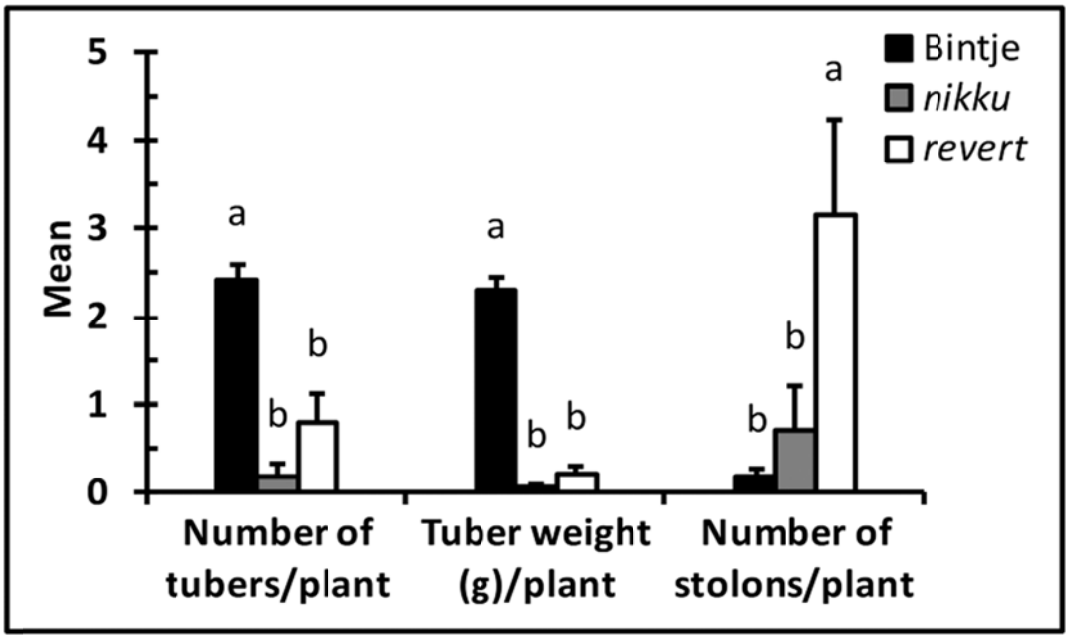


Figure 4.4 Growth responses during *in vitro* tuberization for Bintje, *nikku* and *nikku* revertant. Data were recorded 30 days after culture on tuberization medium. The values are means ± SE (n=24). Statistical differences were calculated by one-way ANOVA. Different letters indicate means that were statistically different by Tukey's HSD multiple testing method ($P < 0.05$) for genotypes within a given growth parameter. (Revert = revertant).

Figure 4.5 Potato phenotyping under greenhouse conditions. Bintje (left) and revertant (right) plants grown from microtubers. These microtubers were produced during *in vitro* tuberization experiments and were planted in soil after breaking dormancy.

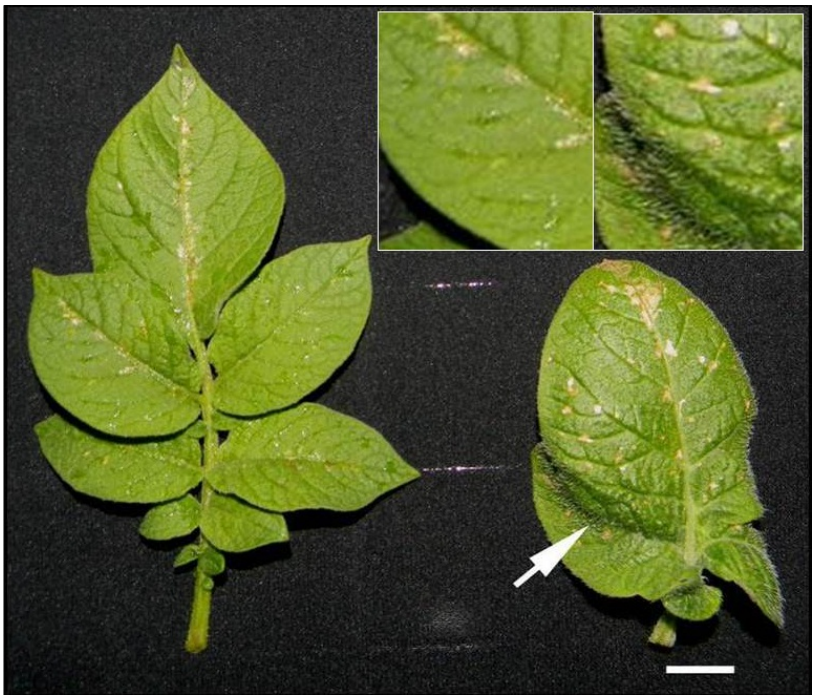


Figure 4.6 Potato leaf phenotypes under greenhouse conditions. Typical compound leaf of potato cv. Bintje with oval lateral leaflets (left) and simple leaf of revertant without lateral leaflets (right). Leaves were taken from corresponding Bintje (left) and revertant (right) plants shown in **Figure 4.5**. Arrow pointed at the trichomes and at the point of fusion between terminal leaf and lateral leaflets in the revertant leaf. Inset shows a close-up of both leaves to illustrate the prominent trichomes in the revertant (Scale bar = 1cm).

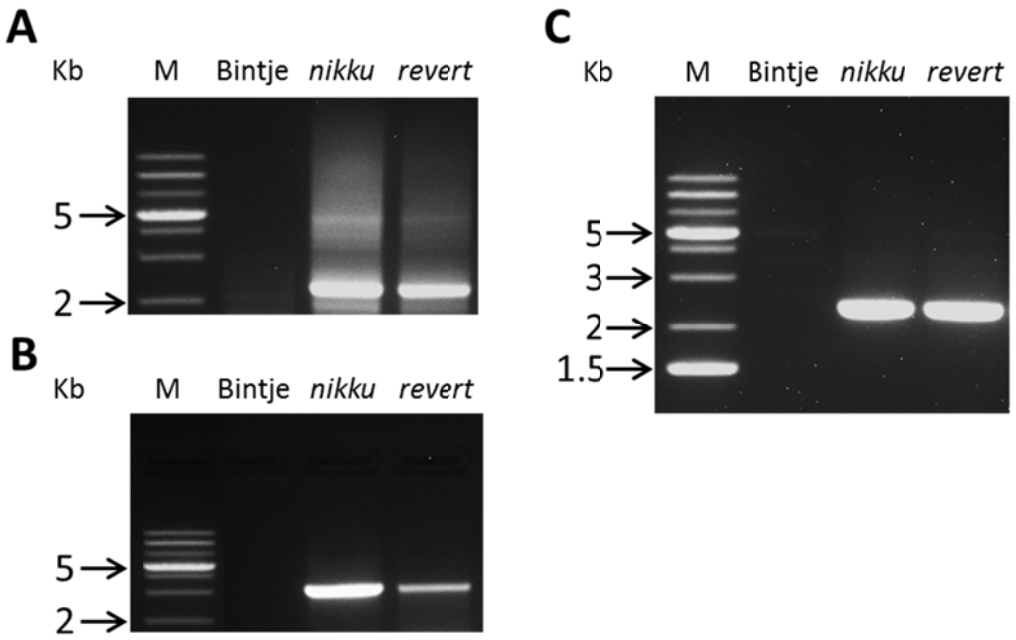


Figure 4.7 Determining the presence of T-DNA in the revertant (revert). PCR products were amplified from genomic DNA of the revertant to amplify right border (A and B) and left border (C) T-DNA fragments. Bintje and *nikku* genomic DNA was used as templates for negative and positive control reactions, respectively. Primer sequences used in A [AT set4-5' and pSKI074-35S promoter-3' (2232 bp)], B [pSKI074Fred-5' and ATset4-3' (3294 bp)] and C [pSKI074-8068...8090-5' and pSKI074Red-3' (2318 bp)] are given in **Table S4.1**. Numbers in small parentheses are the expected product sizes. (M= Gene ruler 1 kb+ DNA mass ladder).

Gene Sr. No.	Gene identifier PGSC0003DMG	Gene annotation based on potato genome	Direction & distance from T-DNA insert (kb)	Reverse transcriptase PCR (RNA from *in vitro* plants)		
				Bintje	*nikku*	*revertant*
RB	Control	*S.tuberosum* elongation factor 1-α (EF1-α)				
R2	400007876	Ethylene-inducing xylanase	⇐ 15			
R3	400007845	Conserved gene of unknown function	⇐ 23			
R4	400007874	Conserved gene of unknown function	⇐ 26			
R5	400007873	E3 ubiquitin-protein ligase RMA1H1	⇐ 36			
R6	400007872	NBS-LRR protein	⇐ 68			
R7	400007868	Ribosomal-protein-alanine	⇐ 117			
R8	400007865	Sugar transporter	⇐ 172			
R9	400007812	Conserved gene of unknown function	⇒ 176			
R10	400007811	Conserved gene of unknown function	⇒ 202			
R11	400007862	SRF3	⇐ 227			
R26	400007792	Vacuolar cation/proton exchanger	⇒ 574			
R27	400046891	Glucose-1-phosphate adenylyltransferase	⇒ 596			
R30	400007788	WRKY transcription factor	⇒ 663			
R31	400007786	2,4-D inducible glutathione S-transferase	⇒ 700			
R32	400007843	Ethylene receptor 1	⇐ 705			
LB	Control	*S.tuberosum* elongation factor 1-α (EF1-α)				
L3	400007880	Senescence-associated protein	⇐ 32			
L4	400007815	GDSL esterase/lipase	⇒ 40			
L11	400007884	Serine/threonine-protein phosphatase	⇐ 79			
L12	400007818	Conserved gene of unknown function	⇒ 86			
L13	400007819	DNA-binding protein phosphatase 2C	⇒ 106			
L23	400007825	Translation initiation factor	⇒ 339			
L26	400007899	CBL-interacting protein kinase 7	⇐ 403			

Figure 4.8 Comparison of various gene transcripts flanking the right and left border of the T-DNA insert in Bintje, *nikku* and *revertant*. Reverse transcriptase PCR analysis was done using gene specific primers (**Table S4.2**). RNA was extracted from *in vitro* plantlets (leaves+stem). The direction of the arrow indicates the orientation of the predicted open reading frame. (LB = left border, RB = right border).

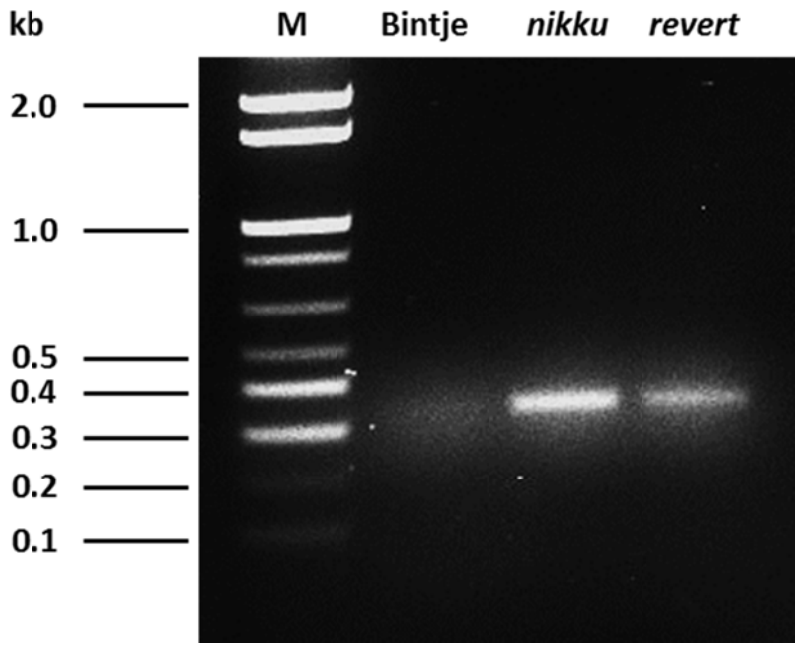


Figure 4.9 Differential methylation status of 35S enhancers in *nikku* and revertant. Genomic DNA of *nikku* and revertant plants- digested with restriction enzyme Sau3AI was enriched for cytosine methylated fragments (see methods section) and equal amounts of enriched DNA were used as template for PCR. Sequences of primers (35X4S *Sau*3AI- 5' and 35X4S *Sau*3AI- 3') are given in **Table S4.1**. (M= Invitrogen 1 kb+ DNA mass ladder).

Figure S4.1 Mutant *nikku* revertant plant phenotypes appeared during *in vitro* tuberization experiment. From left to right, plants with increasing degree of phenotype reversion, as denoted by increased rooting and leaf/stem development. Stem sections of 3-5 cm in length were grown in 40 ml liquid propagation medium at 24±1°C, 16h photoperiod and 70-100 µE $m^{-2}sec^{-1}$ light intensity for 4 weeks followed by 30 days of plant culture in microtuberization medium at 18-20°C in the dark. (Scale bar = 1 cm).

Methylator analysis of the pSKI074 tetramerized enhancers

GATC = *Sau*3AI and *Dpn*I site pSKI074:2894...4180 = 1287 bp

GATCCCCAACATGGTGGAGCACGACACTCTCGTCTACTCCAAGAATATCAAAGATACAGTCTCAGAAGACCAGAGGG
CTATTGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTCATC
GAAAGGACAGTAGAAAAGGAAGATGGCTTCTACAAATGCCATCATTGCGATAAAGGAAAGGCTATCGTTCAAGATGC
CTCTACCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAACATCGTGGAAAAAGAAGACGTTCCAACCACGT
CTTCAAAGCAAGTGGATTGATGTGATATCTAGATCCCCAACATGGTGGAGCACGACACTCTCGTCTACTCCAAGAAT
ATCAAAGATACAGTCTCAGAAGACCAGAGGGCTATTGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGG
ATTCCATTGCCCAGCTATCTGTCACTTCATCGAAAGGACAGTAGAAAAGGAAGATGGCTTCTACAAATGCCATCATT
GCGATAAAGGAAAGGCTATCGTTCAAGATGCCTCTACCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAAC
ATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATATCTAGATCCCCAACATGGT
GGAGCACGACACTCTCGTCTACTCCAAGAATATCAAAGATACAGTCTCAGAAGACCAGAGGGCTATTGAGACTTTTC
AACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTCATCGAAAGGACAGTAGAA
AAGGAAGATGGCTTCTACAAATGCCATCATTGCGATAAAGGAAAGGCTATCGTTCAAGATGCCTCTACCGACAGTGG
TCCCAAAGATGGACCCCCACCCACGAGGAACATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGG
ATTGATGTGATATCTAGATCCCCAACATGGTGGAGCACGACACTCTCGTCTACTCCAAGAATATCAAAGATACAGTC
TCAGAAGACCAGAGGGCTATTGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCATTGCCCAGC
TATCTGTCACTTCATCGAAAGGACAGTAGAAAAGGAAGATGGCTTCTACAAATGCCATCATTGCGATAAAGGAAAGG
CTATCGTTCAAGATGCCTCTACCGACAGTGGTCCCAAAGATGGACCCCCACCCAC

Prediction Results

Methylated cytosines of CpG Dinucleotides are shown as larger red colored letters.

1	GATCCCCAAC	ATGGTGGAGC	ACGACACTCT	CGTCTACTCC	AAGAATATCA	AAGATACAGT	CTCAGAAGAC	CAGAGGGCTA	TTGAGACTTT	TCAACAAAGG	**100**
101	CTAATATCGG	GAAACCTCCT	CGGATTCCAT	TGCCCAGCTA	TCTGTCACTT	CATCGAAAGG	ACAGTAGAAA	AGGAAGATGG	CTTCTACAAA	TGCCATCATT	**200**
201	GCGATAAAGG	AAAGGCTATC	GTTCAAGATG	CCTCTACCGA	CAGTGGTCCC	AAAGATGGAC	CCCCACCCAC	GAGGAACATC	GTGGAAAAAG	AAGACGTTCC	**300**
301	AACCACGTCT	TCAAAGCAAG	TGGATTGATG	TGATATCTAG	ATCCCCAACA	TGGTGGAGCA	CGACACTCTC	GTCTACTCCA	AGAATATCAA	AGATACAGTC	**400**
401	TCAGAAGACC	AGAGGGCTAT	TGAGACTTTT	CAACAAAGGG	TAATATCGGG	AAACCTCCTC	GGATTCCATT	GCCCAGCTAT	CTGTCACTTC	ATCGAAAGGA	**500**
501	CAGTAGAAAA	GGAAGATGGC	TTCTACAAAT	GCCATCATTG	CGATAAAGGA	AAGGCTATCG	TTCAAGATGC	CTCTACCGAC	AGTGGTCCCA	AAGATGGACC	**600**
601	CCCACCCACG	AGGAACATCG	TGGAAAAAGA	AGACGTTCCA	ACCACGTCTT	CAAAGCAAGT	GGATTGATGT	GATATCTAGA	TCCCCAACAT	GGTGGAGCAC	**700**
701	GACACTCTCG	TCTACTCCAA	GAATATCAAA	GATACAGTCT	CAGAAGACCA	GAGGGCTATT	GAGACTTTTC	AACAAAGGGT	AATATCGGGA	AACCTCCTCG	**800**
801	GATTCCATTG	CCCAGCTATC	TGTCACTTCA	TCGAAAGGAC	AGTAGAAAAG	GAAGATGGCT	TCTACAAATG	CCATCATTGC	GATAAAGGAA	AGGCTATCGT	**900**
901	TCAAGATGCC	TCTACCGACA	GTGGTCCCAA	AGATGGACCC	CCACCCACGA	GGAACATCGT	GGAAAAAGAA	GACGTTCCAA	CCACGTCTTC	AAAGCAAGTG	**1000**
1001	GATTGATGTG	ATATCTAGAT	CCCCAACATG	GTGGAGCACG	ACACTCTCGT	CTACTCCAAG	AATATCAAAG	ATACAGTCTC	AGAAGACCAG	AGGGCTATTG	**1100**
1101	AGACTTTTCA	ACAAAGGGTA	ATATCGGGAA	ACCTCCTCGG	ATTCCATTGC	CCAGCTATCT	GTCACTTCAT	CGAAAGGACA	GTAGAAAAGG	AAGATGGCTT	**1200**
1201	CTACAAATGC	CATCATTGCG	ATAAAGGAAA	GGCTATCGTT	CAAGATGCCT	CTACCGACAG	TGGTCCCAAA	GATGGACCCC	CACCCAC		**1287**

Figure S4.2 Tetramerized 35S enhancer sequences of activation tagged vector pSKI074. Upper Panel: Sau3AI restriction sites (GATC) are shown in red font. The sequence of primers used for PCR analysis are underlined and also given in **Table S4.1**. Lower Panel: Output result of prediction of methylated cytosines of CpG dinucleotides, as determined using Methylator prediction software (**http://bio.dfci.harvard.edu/methylator/**). The predicted methylated cytosine residues are shown in bold and red color compared to the rest of the cytosine residues.

Table S4.1 Primers used for PCR amplification of pSKI074 vector sequences, backbone sequences and methylation-enriched DNA.

Primer name	Primer Sequence	Purpose
Potato AT primer set 4-5'	AAGGGAATAAGGGCGACACGGAAA	Right border T-DNA
pSKI074-35 S promoter-3'	CACTGATAGTTTCGGATCTAGATATC	Right border T-DNA
pSKI074-Fred-5'	GCGTGGCTTTATCTGTCTTTGTATTG	Right border T-DNA
Potato AT primer set 4-3'	CCAACTTAATCGCCTTGCAGCACA	Right border T-DNA
pSKI074-8068-8090-5'	CAACAGAGCCTGGCGTTCCCCTT	Left border T-DNA
pSKI074-Red-3'	GGCCTACTTTAATTGCTTCCAGTGTTA	Left border T-DNA
pSKI-4242..71-5'	GATATCTAGATCCGAAACTATCAGTGTTTG	Right border backbone
pSKI-4266..90-5'	TGTTTGACAGGATATATTGGCGGGT	Right border backbone
pSKI-4359..82-3'	GGCATGCACATACAAATGGACGAA	Right border backbone
pSKI-4573..93-3'	GCGATCGAGGATTTTTCGGCG	Right border backbone
pSKI-4874..97-3'	CGATATCATTACGACAGCAACGGC	Right border backbone
pSKI-5379..5402-3'	AGACGAACGAAGAGCGATTGAGGA	Right border backbone
AT601 offLB-5'	GTGTCATTAAATAATTGTCATTTTGCAAC	Left border backbone
pSKI074-Red-3'	GGCCTACTTTAATTGCTTCCAGTGTTA	Left border backbone
4x 35S Sau3AI-5'	CAACATGGTGGAGCACGACACTCTC	Methylation enrichment PCR
4x 35S Sau3AI-3'	CAATCCACTTGCTTTGAAGACGTGGTTGG	Methylation enrichment PCR

Table S4.2 Primers used for semi-quantitative PCR. R2 to 32 are primers for genes flanking the right border and L3-L26 are primers for genes flanking the left border (LB= Left border, RB= Right border).

Primer name	Forward primer sequence	Reverse primer sequence
StEF 1-α	TGATTGAGAGGTCTACCAACCTCGA	GTTCCTTACCTGAACGCCTGTCA
AT601R2	TATATCTTATTGCCCATGTTTTCC	GGTTCACTGTAGTAACAATAATCAGG
AT601R3	TAAGTCGCAAATAAAAGTCAGCA	AGGACCAGTTCCTATTCAAGAAG
AT601R4	TCGTTGAATCTTTAGCAATCACAT	AAGCATCATTAGCATCTACCAAGG
AT601R5	TGACTATACCAAAATTCGGGGC	TGGATGTTGTACATGATCCTGTTG
AT601R6	TTATATCTGCAAAATCTTGAGGGAT	CTTCTTGTCATGGAAGGATTTAAGC
AT601R7	GATACAAGCAGAGAGTGAAGAAGG	AGTGAGAATAGAGAAGAAGATTTTCC
AT601R8	AACAACAACTCTATCTGCTCCAAC	AAATCTTTCCCTTAAGGCTTCGT
AT601R9	TTGGAGAGTTGTTCGTTTACATTC	CTTTCTTGAGAGCCTTGAACATAG
AT601R10	TGCAGGGTAATGTTTGGATTTC	CAAGAGAGGGCAACTTCAGTATT
AT601R11	AGCCATCCTCATGCGAGTATTC	CAAAAGGATGATGAGTTTCTTGAC
AT601R26	ACTTTACTAGTGCTTGGATGTGC	TGGTTGGTCAGCTGAAAAAAT
AT601R27	GATGCGGTCAGACAGTATTTATGG	GCATGTACCTGCATTTCTTTCAAC
AT601R30	GAACCAAGATTTGCCTTCAT	TTAATATTTTACCTGGGGAAAG
AT601R31	CAAGAATTGCATTAGATGAAAAAGG	GTCAATAAAATCAGCCCAAAACC
AT601R32	CCTGATCAGCTACCACTTCAAC	AATTCTCCTGTCGCAAAACTTC
AT601L3	CATCCAACAGAGTAAACAAGAATG	GCAAACCATCAAATGACTGC
AT601L4	TTAAGGCTTTGGGCTTTCTTG	CATTACCTGAGGACGTTGAACC
AT601L11	GGGGATTTTGGTCAGACATACG	TCAGGAACAAAATCAAGTTCATG
AT601L12	TCCTTGCTATGGGTTTCTCAG	CCTTGACGCCAGTCCATAATA
AT601L13	GCAGATCATTACTCGTGGCAAAT	GGAACTCGTCCTCCTTAGTCAGC
AT601L23	CAGGTACATCCTCATGCATCTATTGA	TCCTGATCTCATTAACACTGGGTTG
AT601L26	CCACAAGTTGTATGCAATAAACCATC	CGAATTGTTCGAAAAAGTAGCTAAAG

References

Ayliffe, M.A., Pallotta, M., Langridge, P., and Pryor, A.J. (2007). A barley activation tagging system. Plant Mol. Biol. 64:329-347.

Borevitz, J.O., Xia, Y.J., Blount, J., Dixon, R.A., and Lamb, C. (2000). Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell 12:2383-2393.

Busov, V., and Strauss, S.H. (2007). Gene discovery in *Populus* using activation tagging. In Vitro Cellular & Developmental Biology-Animal 43:S22-S22.

Busov, V., et al. (2011). Activation tagging is an effective gene tagging system in *Populus*. Tree Genet. Genom. 7:91-101.

Busov, V.B., Meilan, R., Pearce, D.W., Ma, C.P., Rood, S.B., and Strauss, S.H. (2003). Activation tagging of a dominant gibberellin catabolism gene (*GA 2-oxidase*) from poplar that regulates tree stature. Plant Physiol. 132:1283-1291.

Chalfun, A., Mes, J.J., Mlynarova, L., Aarts, M.G.M., and Angenent, G.C. (2003). Low frequency of T-DNA based activation tagging in *Arabidopsis* is correlated with methylation of CaMV 35S enhancer sequences. FEBS Lett. 555:459-463.

Dong, Y.Z., and von Arnim, A.G. (2003). Novel plant activation-tagging vectors designed to minimize 35S enhancer-mediated gene silencing. Plant Molecular Biology Reporter 21:349-358.

Harding, K. (1994). The methylation status of DNA derived from potato plants recovered from slow growth. Plant Cell Tissue and Organ Culture 37:31-38.

Hayashi, H., Czaja, I., Lubenow, H., Schell, J., and Walden, R. (1992). Activation of a plant gene by T-DNA tagging: auxin-independent growth *in vitro*. Science 258:1350-1353.

Jeong, D.H., et al. (2002). T-DNA insertional mutagenesis for activation tagging in rice. Plant Physiol. 130:1636-1644.

Kardailsky, I., et al. (1999). Activation tagging of the floral inducer *FT*. Science 286:1962-1965.

Lopez-Serra, L., Ballestar, E., Fraga, M.F., Alaminos, M., Setien, F., and Esteller, M. (2006). A profile of methyl-CpG binding domain protein occupancy of hypermethylated promoter CpG islands of tumor suppressor genes in human cancer. Cancer Res. 66:8342-8346.

Marsch-Martinez, N., Greco, R., Van Arkel, G., Herrera-Estrella, L., and Pereira, A. (2002). Activation tagging using the *En-I* maize transposon system in *Arabidopsis*. Plant Physiol. 129:1544-1556.

Mathews, H., et al. (2003). Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. Plant Cell 15:1689-1703.

Matzke, A.J.M., and Matzke, M.A. (1998). Position effects and epigenetic silencing of plant transgenes. Curr. Opin. Plant Biol. 1:142-148.

Meyer, P., Linn, F., Heidmann, I., Meyer, H., Niedenhof, I., and Saedler, H. (1992). Endogenous and environmental factors influence 35S promoter methylation of a maize *A1* gene construct in transgenic petunia and its colour phenotype. Molecular & General Genetics 231:345-352.

Mishiba, K.-i., et al. (2010). Strict *de novo* methylation of the 35S enhancer sequence in gentian. Plos One 5.

Radouani, A. (2003). Effect of plant growth regulators, plantlet attributes and nutrient medium on *in vivo* and *in vitro* potato tuberization. In: Science Agronomiques Rabat, Morocco: Institut Agronomique & Vétérinaire Hassan II.

Suzuki, Y., et al. (2001). A novel transposon tagging element for obtaining gain-of-function mutants based on a self-stabilizing *Ac* derivative. Plant Mol. Biol. 45:123-131.

van der Graaff, E., Den Dulk-Ras, A., Hooykaas, P.J.J., and Keller, B. (2000). Activation tagging of the *LEAFY PETIOLE* gene affects leaf petiole development in *Arabidopsis thaliana*. Development 127:4971-4980.

van der Graaff, E., Nussbaumer, C., and Keller, B. (2003). The *Arabidopsis thaliana rlp* mutations revert the ectopic leaf blade formation conferred by activation tagging of the *LEP* gene. Mol. Genet. Genomics 270:243-252.

Wade, P.A., Gegonne, A., Jones, P.L., Ballestar, E., Aubry, F., and Wolffe, A.P. (1999). Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. Nat. Genet. 23:62-66.

Walden, R., Fritze, K., Hayashi, H., Miklashevichs, E., Harling, H., and Schell, J. (1994). Activation tagging - A means of isolating genes implicated as playing a role in plant -growth and development Plant Mol. Biol. 26:1521-1528.

Weigel, D., et al. (2000). Activation tagging in *Arabidopsis*. Plant Physiol. 122:1003-1013.

Yamasaki, S., et al. (2011a). Epigenetic modifications of the 35S promoter in cultured gentian cells. Plant Sci. 180:612-619.

Yamasaki, S., et al. (2011b). *De novo* DNA methylation of the 35S enhancer revealed by high-resolution methylation analysis of an entire T-DNA segment in transgenic gentian. Plant Biotechnol. 28:223-230.

Appendix 1.A Standardized NE1014 rating codes for plant and tuber characteristics

Tuber color	**Tuber shape**	**Tuber skin texture**
1. Purple 2. Red 3. Pink 4. Dark brown 5. Tan/light brown 6. Buff 7. Cream	1. Very round 2. Mostly round 3. Round to oblong 4. Mostly oblong 5. Oblong 6. Oblong to long 7. Mostly long 8. Long 9. Cylindrical	1. Partial russet 2. Heavy russet 3. Moderate russet 4. Light russet 5. Netted 6. Slight net 7. Moderately smooth 8. Smooth 9. Very smooth
Tuber eye depth	**Size classification for round tubers**	**Size classification for long tubers**
1. – 2. Deep 3. + 4. – 5. Medium 6. + 7. – 8. Shallow 9. +	Size1= 1 to 1 $^{7/8}$ Size 2= 1 $^{7/8}$ to 2 ½ Size 3= 2 ½ to 3 ¼ Size 4= 3 ¼ to 4 Size 5= > 4” diameter	Size1= 0 to 4 oz Size 2= 4 to 8 oz Size 3= 8 to 12 oz Size 4= 12 to 16 oz Size 5= > 16 oz
Total external defects Green Misshapen –Knobs Growth cracks Rot		

Adapted from Appendix 2: North Carolina potato variety trial and breeding report 2005, page 44