Functional analysis of agrobacterium virulence genes

Niu, X.

Chapter 4

The influence of the *Agrobacterium tumefaciens* virulence proteins VirE3 and VirF on the *Arabidopsis thaliana* transcriptome

Xiaolei Niu, Christiaan Henkel, G Paul H van Heusden and Paul JJ Hooykaas

Department of Molecular and Developmental Genetics, Institute of Biology, Leiden University, Sylviusweg 72, 2333BE Leiden, the Netherlands

Abstract

During *Agrobacterium*-mediated transformation of plant cells a part of the tumor inducing plasmid, the T-DNA, is integrated into the host genome. In addition, a number of virulence proteins are translocated from the bacterium into the host cell. It has been shown that the virulence protein VirE3 binds to the *Arabidopsis thaliana* pBrp protein, a plant-specific general transcription factor of the TFIIB family, suggesting that VirE3 may influence transcription in the host cell (García-Rodríguez *et al*., 2006). The virulence protein VirF may be involved in degradation of the *Arabidopsis* VIP1 transcription factor. To study a possible role of these two virulence proteins in transcriptional regulation, we stably expressed *virE3* and *virF* in the plant *A. thaliana* under control of a tamoxifen-inducible promoter. By RNA sequencing the effect of expression of *virE3* and *virF* on the genome-wide transcription profile was determined. Our results showed that after correction for the effect of tamoxifen on the control plant lines, the RNA levels of 1472 genes were increased more than two-fold and those of 1310 genes decreased more than two-fold upon *virE3* expression. Expression of *virF* had much less pronounced effects. Our data support a role of *virE3* in influencing host cell's transcription machinery.

Introduction

The plant pathogen *Agrobacterium tumefaciens* can cause crown gall disease by transferring a part of its tumor inducing (Ti) plasmid, called T-DNA, to plant cells, where it integrates into the genome; for review see: (Citovsky *et al*, 2007; Hooykaas, P.J.J and Beijersbergen, 1994; Gelvin, 2000; Pitzschke & Hirt, 2010). Due to the expression of oncogenes located on the T-DNA the transformed plant cells are proliferating uncontrollably and form crown gall tumors. The genes responsible for T-DNA production and transfer are present in the virulence region (*vir*) of the Ti plasmid. The *vir* region encodes virulence proteins which have different functions in T-DNA processing and delivery. Some of these virulence proteins (VirE2, VirE3, VirD5 and VirF) are directly translocated into the recipient cells independent of the T-DNA. The virulence protein VirD2 generates nicks at the border sequences of the T-DNA region and covalently binds to the 5' end of the T-strand, a single stranded copy of the T-DNA which is produced inside *Agrobacterium* (Vogel & Das, 1992). The VirD2-T-strand complex is transported into the plant cells via a type IV secretion system (Vergunst *et al*, 2000). VirE2 is a single strand DNA-binding protein which binds to the T-strand in the host cell cytoplasm to prevent its degradation by host nucleases. Both VirD2 and VirE2 have nuclear localization signals (NLSs) which facilitate targeting of the T-strand

to the host cell nucleus.

VirE3 is highly conserved between different *Agrobacterium* species (García-Rodríguez *et al*, 2006), suggesting that VirE3 plays an important role during *Agrobacterium*-mediated transformation (AMT). VirE3 has two potential NLS sequences and it has been shown that VirE3 binds to importins-α (García-Rodríguez *et al*, 2006). In line with these observations VirE3 expressed in plant cells is localized in the nucleus (Lacroix *et al*, 2005; García-Rodríguez *et al*, 2006). Also VirE3 was reported to mimic the function of the transcription factor VIP1 and therefore, it may promote the entry of the T-complex into the host nucleus (Lacroix *et al*, 2005). Furthermore, VirE3 interacts with pBrp, a plant specific general transcription factor of the TFIIB family, suggesting that VirE3 may influence transcription in the host cell (García-Rodríguez *et al*, 2006). However, until now direct evidence for a role of VirE3 in transcriptional regulation is still lacking.

VirF contains a putative F-box motif through which it interacts with plant homologues of the yeast Skp1 protein (ASK1 and ASK2) to form a Skp-Cullin-F-box protein (SCF) complex. This interaction was abolished by mutation of the F-box motif by alanine substitution of two conserved leucine and proline residues (Schrammeijer *et al*, 2001). It has been reported that SCF^{virF} targets the plant transcription factor VIP1 and its associated *Agrobacterium* VirE2 protein for degradation by the 26S proteasome both in yeast and *in planta* (Tzfira *et al*, 2004). As VIP1 is a plant transcription factor, VirF may affect transcription in the host cell.

As indicated above, the function of both VirE3 and VirF during AMT may be disturbance of the host cell's transcription machinery. In Chapter 3, we have shown that expression of *virF* in yeast has only a very minor effect on the genome-wide transcription profile. Also, the effect of expression *of virE3* in yeast is very limited (data not shown). As plants are considered to be the natural hosts of *Agrobacterium* and the transformation of plants by *Agrobacterium* in general is more efficient than that of yeast, in this chapter the effect of expression of *virE3* and *virF* on the genome-wide transcription profile of the plant *Arabidopsis thaliana* was investigated. To this end, expression cassettes for both genes with a tamoxifen-inducible promoter were integrated into the *A. thaliana* genome and the effect on the genome-wide transcription profile was investigated using RNA sequencing. The results show that expression of VirE3 affects the expression of many host genes, suggesting that VirE3 is indeed able to affect the host cell's transcription machinery.

Material and Methods

Binary constructs and plant transformation

The *Agrobacterium virE3* and *virF* coding sequences were amplified by PCR from plasmid 34VN-VirE3 (kindly provided by Philippe Sakalis) and pTOPO[virF] (see chapter 3), respectively, using *Phusion*™ *High-Fidelity* DNA Polymerase. For *virE3*, the oligonucleotides 5'- aaagcggccgcaaatggtgagcactacgaagaaa-3' and 5'-aaagcggccgcttagaaacctctggaggtgga-3' were used. For *virF*, the oligonucleotides 5'-aaagcggccgcaaatgagaaattcgagtttgcgtg-3' and 5'-aaagcggccgctcatagaccgcgcgttgatcga-3' were used. PCR fragments were inserted into the CloneJET™ PCR Cloning vector (Fermentas) as recommended by the manufacturer, yielding pJET1.2 [virE3] and pJET1.2 [virF]. DNA fragments containing *virE3* or *virF* were obtained by digestion of pJET1.2 [virE3] or pJET1.2 [virF], respectively, with *NotI* restriction enzyme and were cloned into the binary vector pGPINTAM-Not (Friml *et al*, 2004), digested with *Not*I, producing pGPINTAM-Not-VirE3 (pSDM3480) and pGPINTAM-Not-VirF (pSDM3478). All PCR fragments were verified by sequencing before using them for plasmid constructions. Correct ligation was checked by restriction analysis and by sequencing the junctions between the plasmids and DNA fragments.

The binary vectors pGPINTAM-Not-VirE3, pGPINTAM-Not-VirF and pGPINTAM-Not (as a negative control) were introduced into *A. tumefaciens* strain *AGL1* by triparental mating (Ditta *et al*, 1980). *Arabidopsis* plants (Columbia, Col-0) were transformed using the floral dip method (Clough & Bent, 1998). Dipped plants were grown at 21°C in a growth chamber (16h light/ 8h dark, 2500 lux) for approximately 2 months until seeds could be harvested. Seeds were surface-sterilized by incubation for 1 minute in 70% ethanol and for 20 minutes in 1% hypochlorite, followed by six rinses with sterile water. Then, the seeds were resuspended in 0.1% agarose and stored at 4 °C for 4 days. To select for transgenic plants, seeds (1 g) were plated on ten plates of ½ MS medium (Murashige and Skoog, 1962) containing 100 mg/L timentin, 50 mg/L Kanamycine and 100 mg/L nystatin and incubated at 21°C in a growth chamber for 14 days. The seedlings were transferred to soil and grown until new seeds could be harvested. Then, seeds were taken from 3 *virE3* plants, 5 *virF* plants and 6 control plants and stored to be used for further experiments. An aliquot of the seeds was germinated as described above and after 10 days DNA was extracted from leaf samples as described by Sylvia de Pater (de Pater *et al*, 2009). T-DNA integration was checked by PCR using *Phusion*™ *High-Fidelity* DNA Polymerase. For transgenic *virE3* plants, the primers 5'-gtacccggggatctgtcgacctcgatcgagat-3' and 5'- aaagcggccgcttagaaacctctggaggtgga -3' were used. For transgenic *virF* plants, the primers 5'-gtacccggggatctgtcgacctcgatcgagat-3' and 5'-aaagcggccgctcatagaccgcgcgttgatcga-3' were used. For transgenic control plants, the primers 5'-gtacccggggatctgtcgacctcgatcgagat-3' and

5’-agtcgactcatagaccgcgcgttgatcgaggt-3’ were used.

RNA extraction and RT-PCR analysis
Transgenic seeds were germinated as described above and after 10 days, 15 to 20 seedlings were transferred to a 50 ml polypropylene tube containing 10 ml liquid ½ MS medium (Murashige and Skoog, 1962) without antibiotics. Tubes were incubated at 21 °C at 120 rpm for four additional days. Treatments with tamoxifen were performed by adding 10 µl of 10 µM (+/-)-tamoxifen (Sigma-Aldrich, T5648-1G) dissolved in DMSO. In control incubations 10 µl of DMSO was added. Four hours after the addition of tamoxifen or of DMSO the plantlets were rapidly frozen in liquid nitrogen.

For RNA isolation, the frozen plantlets were ground under liquid nitrogen in a TissueLyser II (Qiagen^{R}). RNA was isolated with the RNeasy Mini Kit (Qiagen), using the optional on-column DNAse treatment. RNA samples were stored at -80°C. For RT-PCR analysis, 1 µg RNA of each sample was first treated again with DNase I (Ambion^{R}), in the presence of 0.5 µl RNasin (Promega^{R}) in a total volume of 10 µl. Subsequently, 0.5 µg treated RNA of each sample was used for Oligo (dT)-primed cDNA synthesis, using M-MLV Reverse Transcriptase (Promega^{R}) according to the protocol recommended by the supplier. In a control reaction the reverse transcriptase was omitted. The reaction mix was diluted 10-fold in water and 2.5 µl was used for PCR analysis in a total volume of 25 µl using *Phusion™ High-Fidelity* DNA Polymerase. For analysis of *virE3* expression, the primers 5’-aaaactagtatggtgagcactacgaagaa-3’ and 5’-aaagaattcgaaacctctggaggtggaacg-3’ were used; for analysis of *virF* expression the primers 5’-aaggatccatgagaaattcgagtttgcgtgatg-3’ and 5’-aagtcgactcatagaccgcgcgttgatcgaggt-3’ were used; for analysis of ROC expression the primers 5’-gaacggaacaggcggtgagtc-3’ and 5’-ccacaggcttcgtcggctttc-3’ were used.

Illumina mRNA-seq library preparation and sequencing
Illumina mRNA-seq libraries were generated and sequenced by BaseClear BV (Leiden, The Netherlands). mRNA was used as input for the Illumina TruSeq RNA sample preparation kit v2 with adaptations. The mRNA fraction was purified from total RNA by polyA capture, fragmented and converted to double-stranded cDNA using random hexamers. DNA adapters including sample-specific barcodes were ligated to both ends of the DNA fragments and subjected to PCR amplification. The resultant libraries were checked on a Bioanalyzer (Agilent) and quantified. The libraries were multiplexed, clustered, and sequenced on an Illumina HiSeq 2000 (TruSeq v3 chemistry) with a single-read 50 cycles sequencing protocol and indexing. The sequencing run was analyzed with the Illumina CASAVA pipeline (v1.8.2), with demultiplexing based on sample-specific barcodes. Sequence reads of low quality (only “passing filter” reads were selected) and reads

containing adaptor sequences or PhiX control sequences (with an in-house filtering protocol) were removed.

Data processing
Raw sequencing data were processed using CLC Genomic Workbench version 4.6. Processed reads were aligned to the *A. thaliana* annotated coding sequences (TAIR v10; www.arabidopsis.org). Absolute reads were normalized to reads per kilo base per million to compensate for the difference in total reads obtained per sample. Expression data were processed using Microsoft Excel.

q-RT-PCR analysis
To verify the differential gene expression detected by RNA-sequencing, quantitative RT-PCR was performed using a DNA Engine Thermal Cycler (MJ Research) equipped with a Chromo4 real-time PCR detection system (Bio-Rad) with SYBR Green I (SYBR® Green I Nucleic Acid Gel Stain, 10,000X concentrate in DMSO, S-7563, Invitrogen) as the fluorescent detection dye. First-strand cDNA was synthesized from 0.5 µg of DNase I –treated RNA, as described above. A dilution series (1:6 to 1: 1296) of the cDNA was made and 5 µl of the diluted cDNA was used for qPCR analysis in a total volume of 25 µl using 0.1 ul *Phusion*™ *High-Fidelity* DNA Polymerase plus 1.25 ul 20X SYBR Green I stock solution. Specific primers were designed for reference and target-genes using the Primer 3 online software (http://biotools.umassmed.edu/bioapps/primer3_www.cgi) to generate specific fragments of around 150 bp. The primers used are listed in Table 1. For all qPCR reactions the conditions were: 98°C for 3min, followed by 40 cycles of 98 °C for 10 s, 60 °C for 20 s, and 72 °C for 20 s. The fold-change in transcript levels following induction by tamoxifen was calculated using the following equation: (primer efficiency) $^{\Delta Ct}$, where ΔCt is the number of PCR amplification cycles to reach the threshold in non-induced sample minus the number of PCR amplification cycles to reach the threshold in induced samples. The efficiency was determined using a serial dilution of the cDNA from one of the induced samples.

Results

Construction of transgenic plants expressing *virE3* or *virF*
In order to investigate whether the *Agrobacterium* virulence proteins VirE3 and VirF influence the host cell's transcriptional machinery, T-DNA constructs were made containing the coding sequences of *virE3* or *virF* under control of a tamoxifen-inducible promoter. These constructs were integrated into the *A. thaliana* genome by *Agrobacterium*-mediated floral dip transformation (see Materials and Methods). In addition, control lines lacking sequences coding for virulence proteins were generated. Three independent lines of the plants with

virE3 and five independent lines of plants with *virF* and six of the control plants were selected for further analysis. To this end, seedlings from each line were grown on solid medium for 10 days, transferred to liquid medium, grown for an additional four days and subsequently tamoxifen was added to half of the seedlings. Induction of the virulence genes did not result in an obvious effect on the morphology in any of the plant lines (data not shown). Four hours after the addition of tamoxifen, RNA was isolated from the plantlets and analyzed for the expression of *virF* and *virE3* by semi-quantitative RT-PCR. As shown in Figure 1, in all three lines containing *virE3* coding sequences *virE3* RNA was detected, only after addition of tamoxifen. In the control lines and in the lines with *virF* coding sequences *virE3* RNA could not be detected, whereas the *A. thaliana* ROC RNA could easily be found. Similarly, *virF* RNA was detected in the lines containing *virF* coding sequences after addition of tamoxifen and not in the other lines (Figure 1).

Table 1. Primers for q-PCR analysis

Primer	Gene description	Sequence
At4g38740-F	Cytosolic cyclophilin ROC1	5'AGAGAAAGGTGTTGGCGGTA3'
At4g38740-R	Cytosolic cyclophilin ROC1	5'TCGAACTTGCTCCCGTAGAT3'
At3g18780-F	Actin2	5'GCAGAGCGGGAAATTGTAAG3'
At3g18780-R	Actin2	5'CAGCACCAATCGTGATGACT3'
At5g12250-F	BETA-6 TUBULIN	5'TGGCAAGATGAGCACAAAAG3'
At5g12250-R	BETA-6 TUBULIN	5'TGTCGATGCCATTGAGAGAC3'
At5g60390-F	GTP binding Elongation factor Tu family protein (EF1)	5' TGGTGACGCTGGTATGGTTA3'
At5g60390-R	GTP binding Elongation factor Tu family protein (EF1)	5' TCCTTCTTGTCCACGCTCTT3'
At5g54720-F	Ankyrin repeat family protein	5'CGACGGAAACACTCCTCTTC3'
At5g54720-R	Ankyrin repeat family protein	5'ACAGCGACGTGATTGTTGAG3'
AT5G13320-F	Auxin-responsive GH3 family protein	5'TGAGTCAAGCGAAGCTCGTA3'
AT5G13320-R	Auxin-responsive GH3 family protein	5'CTCAAACTGTGCCGTCTCCT3'
AT4G19925-F	Toll-Interleukin-Resistance (TIR) domain family protein	5'AGCCTAGCTTCACGTCAGGA3'
AT4G19925-R	Toll-Interleukin-Resistance (TIR domain family protein	5'CAACCGAAGCAGAAAGAAGC3'
AT1G19250-F	Flavin-dependent monooxygenase 1	5'CTAAGCCGGCTAGTGGATGA3'
AT1G19250-R	Flavin-dependent monooxygenase 1	5'CAGCATGTTGGATGCTGAAC3'

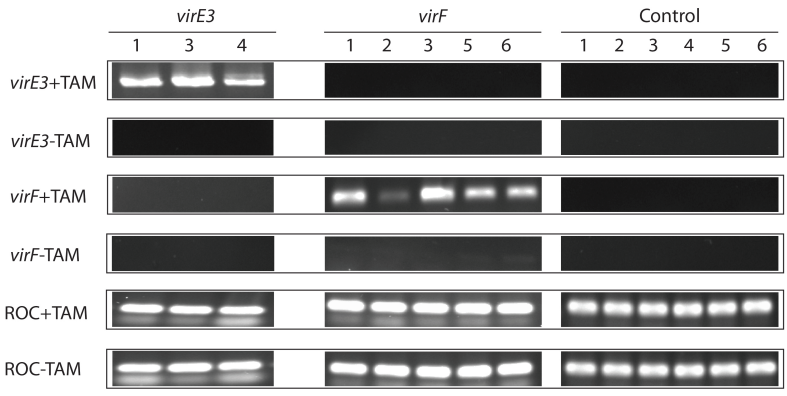


Figure 1. Analysis of tamoxifen-dependent expression of *virE3*, *virF* and ROC in transgenic *A. thaliana* lines by semi-quantitative RT-PCR. TAM: tamoxifen. Numbers identify the different plant lines analyzed.

Transcriptome analysis by RNA-seq

To investigate the effect of VirE3 and VirF on the genome-wide transcription profile, one line expressing *virE3* (line 1), one line expressing *virF* (line 5) and one control line (line 3) were selected and the RNA from tamoxifen-induced and non-induced plantlets was analyzed by Illumina RNA sequencing. An overview of the results is shown in Table 2. The obtained sequences were aligned to the *A. thaliana* annotated coding sequences (TAIR v10 www.arabidopsis.org) to obtain information on the genome-wide transcription profile. In this way, alignment was observed to 25782 *A. thaliana* annotated genes. The number of reads aligning to each of the *A. thaliana* genes is shown in the supplementary Table S1.

From our RNA sequencing data it is clear that the expression levels of the various *A. thaliana* genes differ considerably. The five genes with the highest RNA levels in the non-induced control plants are: CAB1 (AT1G29930; chlorophyll A/B binding protein 1), RBCS1A (AT1G67090, ribulose bisphosphate carboxylase small chain 1A), MT2B (AT5G02380, metallothionein 2B), AT2G43150 (Proline-rich extensin-like family protein) and CWLP (AT3G22120, cell wall-plasma membrane linker protein), with expression levels of 7381, 4318, 2872, 2711 and 2673 reads per million per kilobase, respectively. 19,202 genes had an expression level of more than 1 read per million per kilobase.

Table 2. Overview of Illumina sequencing of RNA isolated from tamoxifen-induced and non-induced *A. thaliana* seedlings expressing *virE3* or *virF*.

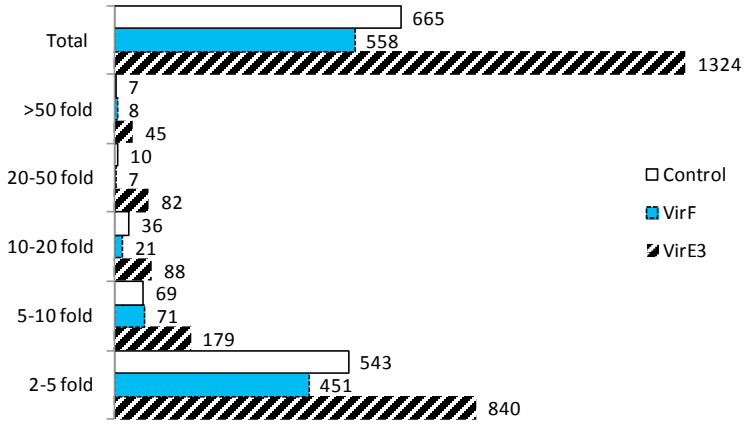
Sample	Plant line	Tamoxifen induction	Number of Reads	Sequenced bases (in MB)	Average Quality Scores (Phred)
1	Control (3)	-	28,665,911	1,461	37.12
2	Control (3)	+	38,869,998	1,982	37.09
3	virE3 (1)	-	30,089,370	1,534	37.15
4	VirE3 (1)	+	27,040,112	1,379	37.04
5	VirF (5)	-	33,742,133	1,720	37.10
6	VirF (5)	+	24,323,529	1,240	37.18

Transcriptional changes after 4 hours of *virE3* or *virF* gene expression

Comparison of the expression profile of tamoxifen-induced and non-induced plantlets showed that in the control plantlets tamoxifen-treatment resulted in a more than 2-fold higher RNA level of 665 genes and a more than 2-fold lower RNA level of 465 genes. Similarly, induction of *virF* resulted in a more than 2-fold higher RNA level of 558 genes and a more than 2-fold lower RNA level of 331 genes. The effect of *virE3* induction is much larger: for 1324 genes RNA levels increased more than two-fold and for 725 genes RNA levels decreased more than two-fold. A more detailed overview of the number of genes affected in the *virE3*, *virF* and control lines is shown in Figure 2.

As tamoxifen treatment of the control plantlets already had a considerable effect on the transcription profile, the effect of the expression of *virE3* and *virF* on the transcription profile may be masked by the effect of tamoxifen. Therefore, we corrected the effects of *virE3* and *virF* expression for the effects of tamoxifen in the control plantlets. After this correction, 1472 genes were found to be up-regulated more than two-fold and 1310 genes were down-regulated more than two-fold upon induction of *virE3*. After correction, Induction of *virF* expression resulted in up-regulation of 136 genes and down-regulation of 172 genes. The 50 most up-regulated genes and the 50 most down-regulated genes after induction of *virF* are listed in Tables 3 and 4, respectively. The 50 most up-regulated genes and the 50 most down-regulated genes after induction of *virE3* are listed in Tables 5 and 6, respectively.

A Numbers of up-regulated genes in transgenic *virE3*, *virF* and control plant lines



B Numbers of down-regulated genes in transgenic *virE3*, *virF* and control plant lines

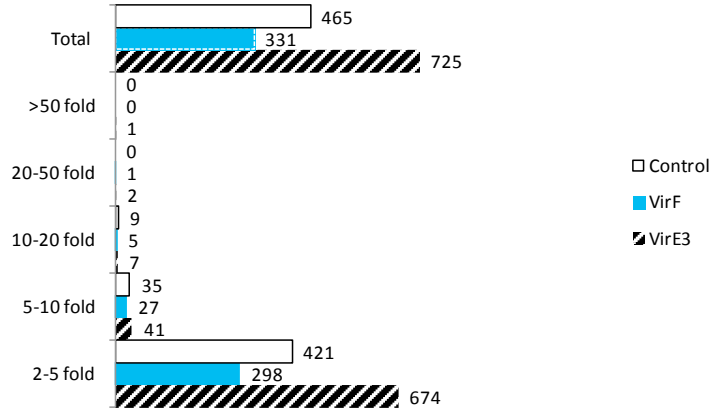


Figure 2. Graphical representation of the numbers of genes that have a more than two-fold higher or lower RNA level after induction by tamoxifen compared to the non-induced controls in transgenic *virE3*, *virF* and control plant lines. A, numbers of up-regulated genes; B, numbers of down-regulated genes**.** The number of induced genes was calculated by TE/NE, TF/NF and TC/NC, see Table S1.

To obtain information on the genes affected by expression of *virE3* or *virF* we analyzed the 50 most up-regulated and 50 most down-regulated genes for the biological process in which they are involved, their molecular function and the cellular component in which their gene products are predicted to localize, using Superviewer (Provart & Zhu, 2003) and the data available at the TAIR database (www.arabidopsis.org). In figure 3 the number of genes found in each class corrected for the total number of *A. thaliana* genes in that class is given. This analysis shows that genes up-regulated by *virE3* expression are highly enriched for genes involved in signal transduction, stress response and response to

biotic and abiotic stimuli. On the other hand, genes involved in DNA and RNA metabolism are underrepresented. Interestingly, genes down-regulated by *virE3* expression are highly enriched for genes encoding ribosomal proteins. The genes affected after expression of *virF* are not clearly enriched for a specific class.

Validation of RNA-seq data by q-RT-PCR analysis

To verify the differential gene expression detected by RNA-sequencing, four genes found to be highly induced upon *virE3* expression, were selected for analysis by quantitative RT-PCR. The effect of tamoxifen-induced *virE3* expression on the expression of these four selected genes was determined not only in the plant line (line 1) that was used for RNA-seq analysis, but also in two independent transgenic lines (lines 3 and 4) containing *virE3* under control of a tamoxifen-inducible promoter. As shown in Table 7, qRT-PCR analysis confirmed that these four selected genes were strongly up-regulated upon *virE3* expression. This up-regulation was also found in the other two independent plant lines, indicating that the effect of *virE3* on the expression of these four genes is not an effect restricted to one plant line.

Table 7. Analysis by qRT-PCR of the expression levels of four genes (At5g54720, AT5G13320, AT4G19925 and AT1G19250) found by RNA-seq to be up-regulated upon *virE3* expression and of four control genes (Roc1, Actin2, Beta6-Tubulin and EF1).

Genes	**Fold-change found by RNA-Seq**		**Fold-change found by qRT-PCR analysis**		
	Fold-change Line 1 (TE/NE)[a]	Fold-change relative to control line (TE/NE)/(TC/NC)[a]	Line 1	Line 3	Line 4
Roc1	0.6	0.5	2.2	3.0	2.3
Actin2	1.1	1.4	4.2	4.7	3.4
Beta-6 Tubulin	1.1	1.2	3.8	3.1	2.5
EF1	1.0	0.9	1.1	1.5	1.2
At5g54720	381	657	49	15	49
AT5G13320	243	72	104	164	38
AT4G19925	353	68	464	261	582
AT1G19250	92	52	74	203	90

a: TE, expression level (reads per kilobase per million) in induced *virE3* plants; NE, expression level (reads per kilobase per million) in non-induced *virE3* plants. TC, expression level (reads per kilobase per million) in induced control plants. NC, expression level (reads per kilobase per million) in non-induced control plants.

Discussion

During AMT the virulence proteins VirE3 and VirF are translocated from the bacterium into the host cell. Here, they may disturb transcription to facilitate the transformation process and/or disrupt the host cells defense system. To investigate this possibility, we inducible expressed *virE3* and *virF* in *A. thaliana* and determined the effect on the genome-wide transcription profile using RNA sequencing. As described in Chapter 3, we performed similar experiments in yeast. From these experiments we were unable to obtain evidence for a role of VirF in AMT of yeast. We showed that the effect of expression of *virF* on the genome-wide transcription profile was very minor. RNA levels of none of the yeast genes were more than 1.6-fold changed. Also the effect of expression *of virE3* in yeast is very limited (data not shown). As plants are considered to be the natural hosts of *Agrobacterium* and the transformation of plants by *Agrobacterium* in general is more efficient than that of yeast, in this Chapter the effect of expression of *virE3* and *virF* on the genome-wide transcription profile of the plant *A. thaliana* was investigated. To this end expression cassettes for both genes with a tamoxifen-inducible promoter were integrated into the *A. thaliana* genome and the effect on the genome-wide transcription profile was investigated using RNA sequencing. These studies showed that expression of *virE3* had a considerable effect on the genome-wide transcription profile. RNA levels of 1324 genes were more than two-fold increased and that of 725 genes were more than two-fold decreased after induction of *virE3* expression. After correction of the effect of tamoxifen on the control lines, 1472 and 1310 genes were found to be up- or down-regulated, respectively. Previously, we showed that VirE3 interacts with pBrp, a plant specific general transcription factor of the TFIIB family (García-Rodríguez *et al*, 2006). This may explain the effects of expression of *virE3* on the *A. thaliana* transcriptome. However, more research is needed to find out whether the effects of *virE3* expression are caused by the interaction with pBrp. The observation that many *A. thaliana* genes are affected by expression of *virE3* may imply that VirE3 plays a role in disturbing transcription in the host cell. However, it has to be kept in mind that we only analyzed one plant line and performed RNA sequencing on one RNA preparation. On the other hand, for four genes we showed that the effect is not restricted to one plant line (Table 7), strengthening the conclusion that VirE3 affects transcription. However, further research is needed to draw conclusions on the exact role of VirE3 in transcriptional regulation.

The effect of expression of *virF* was considerably lower. Induction of *virF* expression resulted in a more than two-fold increase in the RNA levels of 558 genes and a more than two-fold decrease of the RNA levels of 331 genes. After correction of the effect of tamoxifen on the control lines, 136 and 172

genes were found to be up- or down-regulated, respectively. It has been reported that VirF may target VirE2 and VIP1 for proteosomal degradation thus uncoating the VirE2—VIP1—T-DNA complex enabling T-DNA integration (Tzfira *et al*, 2004). VIP1 is an *A. thaliana* transcription factor and expression of *virF* may result in an altered expression of genes regulated by VIP1. It has been shown that over-expression of VIP1 resulted in an increased expression of four genes: AT4G19230 (CYP707A1), AT5G45340 (CYP707A3), AT1G69880 (TH8) and AT5G67300 (MYB44) (Pitzschke *et al*, 2009; Tsugama *et al*, 2012). However, RNA levels of none of these genes were affected by *virF* expression. This may indicate that either VIP1 is not a target of VirF or that VIP1 is only degraded in a VirF-dependent way when complexed to T-DNA and VirE2. Furthermore, an effect of VirF on VIP1-dependent transcription may only be seen after activation of VIP1 by MPK3-dependent phosphorylation induced by *Agrobacterium* infection and subsequent translocation to the nucleus (Djamei *et al*, 2007).

Our analyses showed that genes up-regulated by *virE3* expression are highly enriched for genes involved in signal transduction, stress response and response to biotic and abiotic stimuli (Figure 2). On the other hand, genes involved in DNA and RNA metabolism are underrepresented. Interestingly, genes down-regulated by *virE3* expression are highly enriched for genes encoding ribosomal proteins. The biological significance of these findings remains unclear.

Acknowledgements

This study was supported by the Chinese Scholarship Council (CSC) and by the division Chemical Sciences (CW) of the Netherlands Organisation of Research (NWO), TOP grant number 700.56.303.

References

Citovsky V, Kozlovsky S., Lacroix B, Zaltsman A, Dafny-Yelin M, Vyas S, Tovkach A & Tzfira T (2007) Biological systems of the host cell involved in Agrobacterium infection. *Cell Microbiol* **9:** 9-20

Clough SJ & Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J* **16:** 735-43

Ditta G, Stanfield S, Corbin D & Helinski DR (1980) Broad host range DNA cloning system for gram-negative bacteria: construction of a gene bank of Rhizobium meliloti. *Proc Natl Acad Sci U S A* **77:** 7347-7351

Djamei A, Pitzschke A, Nakagami H, Rajh I & Hirt H (2007) Trojan horse strategy in Agrobacterium transformation: abusing MAPK defense signaling. *Science* **318:** 453-456

Friml J, Yang X, Michniewicz M, Weijers D, Quint A, Tietz O, Benjamins R, Ouwerkerk PBF, Ljung K, Sandberg G, Hooykaas PJ. ., Palme K & Offringa R (2004) A PINOID-dependent binary switch in apical-basal PIN polar targeting directs auxin efflux. *Science* **306:** 862-865

García-Rodríguez FM, Schrammeijer B & Hooykaas PJJ (2006) The Agrobacterium VirE3 effector protein: a potential plant transcriptional activator. *Nucleic Acids Res* **34:** 6496-504

Gelvin SB (2000) Agrobacterium and plant genes involved in T-DNA transfer and integration. *Annu Rev Plant Physiol Plant Mol Biol* **51:** 223-256

Hooykaas, P.J.J and Beijersbergen AGM (1994) The virulence system of Agrobacterium tumefaciens. *Annu Rev Phytopathol* **32:** 157-179

Lacroix B, Vaidya M, Tzfira T & Citovsky V (2005) The VirE3 protein of Agrobacterium mimics a host cell function required for plant genetic transformation. *EMBO J* **24:** 428-437

Murashige T & Skoog F (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol Plant* **15:** 473-497

de Pater S, Neuteboom L., Pinas JE, Hooykaas PJ. & van der Zaal B. (2009) ZFN-induced mutagenesis and gene-targeting in Arabidopsis through Agrobacterium-mediated floral dip transformation. *Plant Biotechnol J* **7:** 821-835

Pitzschke A, Djamei A, Teige M & Hirt H (2009) VIP1 response elements mediate mitogen-activated protein kinase 3-induced stress gene expression. *Proc Natl Acad Sci U S A* **106:** 18414-18419

Pitzschke A & Hirt H (2010) New insights into an old story: Agrobacterium-induced tumour formation in plants by plant transformation. *EMBO J* **29:** 1021-1032

Provart N & Zhu T (2003) A Browser-based Functional Classification SuperViewer for Arabidopsis Genomics. *Currents in Computational Molecular Biology***:** 271-272

Schrammeijer B, Risseeuw E, Pansegrau W, Regensburg-Tuink, T. J., Crosby,W. L. & Hooykaas PJJ (2001) Interaction of the virulence protein VirF of Agrobacterium tumefaciens with plant homologs of the yeast Skp1 protein. *Curr Biol* **11:** 258-262

Tsugama D, Liu SK & Takano T (2012) A bZIP protein, VIP1, is a regulator of osmosensory signaling in Arabidopsis. *Plant Physiol* **159:** 144-155

Tzfira T, Vaidya M & Citovsky V (2004) Involvement of targeted proteolysis in plant genetic transformation by Agrobacterium. *Nature* **431:** 87-92

Vergunst A., Schrammeijer B, den Dulk-Ras A, de Vlaam CM., Regensburg-Tuink TJ. & Hooykaas PJJ (2000) VirB/D4-dependent protein translocation from Agrobacterium into plant cells. *Science* **290:** 979-982

Vogel AM & Das A (1992) Mutational analysis of Agrobacterium tumefaciens virD2: tyrosine 29 is essential for endonuclease activity. *J Bacteriol* **174:** 303-308

Table 3. The 50 most up-regulated genes after induction of virF. The fold up-regulation is calculated by dividing the ratio between the expression level in induced virF plants (TF) to the non-induced virF plants (NF) by the ratio between the expression level in induced control plants (TC) to the non-induced control plants (NC). TE, expression level in induced virE3 plants; NE, expression level in non-induced virE3 plants. RPKM, reads per kilobase per million. The data are filtered for genes of which the sum of the expression levels in the six samples is more than 15 RPKM.

Rank	Locus	VirF Up-reg. (TF/NF)/(TC/NC)	TF RPKM	NF RPKM	TC RPKM	NC RPKM	TE RPKM	NE RPKM	Name/Description
1	AT5G41612	7.1	6.4	1.7	1.5	2.7	7.0	3.9	Other RNA
2	AT5G04630	6.4	23.8	0.0	26.5	0.1	4.2	0.2	Cytochrome P450, family 77, subfamily A, polypeptide 9
3	AT5G50630	5.1	8.7	5.4	2.9	9.2	9.7	6.1	Major facilitator superfamily protein
4	AT4G04900	4.5	1.1	3.0	0.5	5.7	0.9	6.2	ROP-interactive CRIB motif-containing protein 10
5	AT3G05950	3.9	48.5	88.2	24.8	176.3	64.7	159.2	RmlC-like cupins superfamily protein
6	AT2G35945	3.9	3.7	1.9	1.4	2.8	3.5	5.4	Other RNA
7	AT2G05250	3.8	6.1	10.0	2.0	12.3	11.1	9.3	DNAJ heat shock N-terminal domain-containing protein
8	AT1G70440	3.5	2.0	2.1	1.5	5.5	2.8	4.9	Similar to RCD one 3
9	AT2G27228	3.4	4.9	3.3	3.8	8.7	4.3	6.1	Conserved peptide upstream open reading frame 6
10	AT4G04972	3.3	4.5	2.7	2.9	5.7	3.2	4.5	Unknown protein.
11	AT3G50825	3.3	5.4	5.8	2.7	9.6	4.4	9.2	SnoRNA
12	AT1G67100	3.3	4.4	7.8	2.3	13.4	3.3	3.8	LOB domain-containing protein 40
13	AT1G56250	3.3	4.6	0.1	1.7	0.1	30.5	0.1	Phloem protein 2-B14
14	AT3G22235	3.2	4.1	2.5	1.4	2.7	6.6	6.2	Unknown protein
15	AT4G17098	3.2	3.9	1.6	2.8	3.8	3.7	2.5	Other RNA
16	AT4G05380	3.2	3.0	4.5	2.6	12.3	4.5	8.4	P-loop containing nucleoside triphosphate hydrolases superfamily protein
17	ATCG00560	3.2	2.6	2.7	1.7	5.4	3.4	0.6	Photosystem II reaction center protein L
18	AT1G50390	3.2	3.9	3.3	2.9	8.1	2.9	4.0	PfkB-like carbohydrate kinase family protein
19	AT5G60180	3.2	15.8	0.3	15.4	0.8	2.2	0.1	Pumilio 19
20	AT5G45428	3.1	6.2	4.8	4.4	10.6	10.1	7.0	Conserved peptide upstream open reading frame 24
21	AT1G74080	3.1	3.3	0.9	1.8	1.6	25.6	0.8	Myb domain protein 122
22	AT3G48640	3.1	0.3	0.1	0.1	0.2	34.8	0.2	Unknown protein
23	AT1G07901	3.0	5.7	2.9	4.0	6.0	3.1	2.8	Unknown protein

88 **Table 3.** Continued

Rank	Locus	VirF Up-reg. (TF/NF)/(TC/NC)	TF RPKM	NF RPKM	TC RPKM	NC RPKM	TE RPKM	NE RPKM	Name/Description
24	AT5G38540	3.0	5.6	7.1	2.9	11.0	8.9	9.2	Mannose-binding lectin superfamily protein
25	AT5G46295	2.9	3.2	0.5	2.0	1.0	20.6	0.9	Unknown protein
26	AT1G05680	2.9	5.6	2.6	7.5	10.2	11.0	5.1	Uridine diphosphate glycosyltransferase 74E2
27	AT5G47980	2.9	8.9	5.3	6.6	11.4	4.4	7.1	HXXXD-type acyl-transferase family protein
28	AT5G24105	2.9	2.4	3.0	1.6	6.0	0.8	4.4	Arabinogalactan protein 41
29	AT3G03670	2.9	2.1	6.4	0.8	6.9	11.9	11.2	Peroxidase superfamily protein
30	AT5G18080	2.9	5.3	3.6	1.6	3.1	2.0	1.2	SAUR-like auxin-responsive protein family
31	AT5G03890	2.9	2.2	2.2	1.4	4.0	3.0	4.0	Unknown protein;
32	AT3G52770	2.8	4.0	3.0	2.1	4.4	1.6	1.5	Protein binding
33	AT2G06255	2.8	2.9	2.4	2.0	4.7	2.2	3.7	ELF4-like 3
34	AT2G01023	2.8	120.8	9.2	105.9	22.5	15.4	13.3	Unknown protein;
35	AT2G46400	2.8	10.4	3.1	9.5	7.9	131.6	6.7	WRKY DNA-binding protein 46
36	ATMG01080	2.8	5.0	3.5	4.7	9.4	6.1	3.6	Mitochondrial F0-ATPase subunit 9
37	AT5G59310	2.8	29.2	35.1	11.9	39.9	26.7	13.2	Lipid transfer protein 4
38	AT5G44306	2.7	2.3	1.8	2.5	5.5	5.0	6.1	Unknown protein;
39	AT3G11840	2.7	3.6	1.9	1.8	2.6	17.0	2.2	Plant U-box 24
40	AT3G09600	2.7	5.4	3.1	2.2	3.4	3.5	4.0	Homeodomain-like superfamily protein
41	AT1G57560	2.7	5.3	4.9	4.0	10.2	13.4	6.4	Myb domain protein 50
42	AT5G44460	2.7	5.5	5.7	2.6	7.1	11.9	8.0	Calmodulin like 43
43	AT4G37390	2.7	6.2	3.0	3.3	4.3	9.6	3.8	Auxin-responsive GH3 family protein
44	AT1G35830	2.7	4.6	3.1	2.7	4.8	3.7	2.8	VQ motif-containing protein
45	AT3G51240	2.6	4.8	6.0	2.8	9.3	4.7	6.4	Flavanone 3-hydroxylase
46	AT1G77145	2.6	4.7	4.3	3.8	9.0	5.5	5.0	Protein of unknown function (DUF506)
47	AT4G15258	2.6	2.4	1.2	2.6	3.5	2.1	3.6	SNOR37-2; snoRNA
48	AT1G73640	2.6	3.3	2.1	1.8	2.9	1.5	3.6	RAB GTPase homolog A6A
49	AT1G58410	2.6	1.9	2.3	1.5	4.6	3.1	2.8	Disease resistance protein (CC-NBS-LRR class) family
50	AT3G59880	2.6	3.3	1.9	2.6	3.9	5.6	3.2	Unknown protein

Table 4. The 50 most down-regulated genes after induction of *virF*. The fold up-regulation is calculated by dividing the ratio between the expression level in non-induced *virF* plants (NF) to the induced *virF* plants (TF) by the ratio between the expression level in non-induced control plants (NC) to the induced control plants (TC). TE, expression level in induced *virE3* plants; NE, expression level in non-induced *virE3* plants. RPKM, reads per kilobase per million. The data are filtered for genes of which the sum of the expression levels in the six samples is more than 15 RPKM.

Rank	Locus	VirF Down-reg. (NF/TF)/ (NC/TC)	TF RPKM	NF RPKM	TC RPKM	NC RPKM	TE RPKM	NE RPKM	Name/Description
1	AT2G04495	9.9	0.2	0.4	0.5	0.1	16.1	0.8	unknown protein
2	AT4G30662	9.0	2.7	3.8	7.4	1.2	2.2	6.5	unknown protein
3	AT1G28480	7.9	0.5	0.2	1.2	0.1	14.7	0.4	Thioredoxin superfamily protein
4	AT1G33760	7.8	5.8	0.3	7.6	0.1	4.4	0.0	Integrase-type DNA-binding superfamily protein
5	ATCG00080	7.7	5.5	3.7	7.7	0.7	2.1	10.2	photosystem II reaction center protein I
6	AT5G39130	7.1	1.8	2.4	2.7	0.5	5.0	4.4	RmlC-like cupins superfamily protein
7	AT3G09280	7.1	3.9	1.0	14.7	0.5	13.4	0.3	unknown protein
8	ATCG00630	6.1	13.6	9.8	16.4	1.9	9.6	13.4	PSAJ
9	AT3G60961	5.6	3.6	7.0	7.4	2.6	4.8	1.2	P-loop containing nucleoside triphosphate hydrolases superfamily protein
10	AT1G69050	5.2	2.6	2.6	3.2	0.6	3.5	2.7	unknown protein
11	AT2G24850	5.2	5.9	3.0	3.0	0.3	23.9	3.7	tyrosine aminotransferase 3
12	AT2G44460	4.8	6.3	2.7	4.6	0.4	4.5	0.4	beta glucosidase 28
13	AT5G52471	4.8	3.0	5.9	6.0	2.5	2.1	1.4	U60.1F; snoRNA
14	AT1G19550	4.7	4.4	3.7	12.7	2.3	4.0	7.9	Glutathione S-transferase family protein
15	AT2G39030	4.6	6.5	7.7	4.6	1.2	3.0	6.1	Acyl-CoA N-acyltransferases (NAT) superfamily protein
16	AT2G11810	4.5	5.4	2.1	5.9	0.5	2.9	0.9	monogalactosyl diacylglycerol synthase 3
17	AT3G20450	4.3	2.8	4.7	5.0	1.9	2.6	4.7	B-cell receptor-associated protein 31-like
18	AT3G25190	4.2	21.6	4.3	30.0	1.4	14.6	3.6	Vacuolar iron transporter (VIT) family protein Print.pdf
19	AT3G47340	4.2	6.2	1.4	5.9	0.3	6.6	0.4	glutamine-dependent asparagine synthase 1
20	AT1G03420	4.0	11.2	12.0	9.6	2.5	3.0	15.1	transposable element gene
21	AT5G01881	4.0	2.1	5.4	2.4	1.6	1.6	5.2	Unknown protein
22	AT4G38825	3.9	3.1	8.4	5.6	3.9	4.5	9.7	SAUR-like auxin-responsive protein family
23	AT5G46900	3.9	0.6	10.6	2.4	11.5	0.8	11.2	Bifunctional inhibitor/lipid-transfer protein/ seed storage 2S albumin superfamily protein
24	AT4G38560	3.8	0.5	0.8	1.0	0.4	20.8	0.7	Arabidopsis phospholipase-like protein (PEARLI 4) family

Table 4. Continued

Rank	Locus	VirF Down-reg. (NF/TF)/ (NC/TC)	TF RPKM	NF RPKM	TC RPKM	NC RPKM	TE RPKM	NE RPKM	Name/Description
25	AT5G26220	3.8	21.6	7.8	16.2	1.5	7.2	2.2	ChaC-like family protein
26	AT1G53620	3.7	2.2	0.6	1.9	0.1	38.7	0.8	Unknown protein
27	AT4G04293	3.7	3.3	10.3	4.0	3.4	3.3	12.7	transposable element gene
28	AT1G05880	3.7	0.1	0.2	0.4	0.1	22.5	0.2	RING/U-box superfamily protein
29	AT1G52890	3.6	1.7	3.2	2.9	1.5	5.9	2.3	NAC domain containing protein 19
30	AT3G48940	3.6	0.6	2.9	3.3	4.7	1.1	4.6	Remorin family protein
31	AT2G43000	3.5	1.8	0.7	2.4	0.3	16.7	0.4	NAC domain containing protein 42
32	AT5G24580	3.5	2.7	4.7	4.7	2.4	3.0	2.9	Heavy metal transport/detoxification superfamily protein
33	AT2G26010	3.4	31.1	86.4	58.1	47.4	54.1	38.7	Plant defensin 1.3
34	AT2G43390	3.3	3.6	1.6	4.3	0.6	10.0	2.1	Unknown protein
35	AT5G25260	3.3	0.4	0.4	0.4	0.1	71.3	0.3	SPFH/Band 7/PHB domain-containing membrane-associated protein family
36	AT5G17350	3.3	1.7	1.1	1.9	0.4	12.2	0.3	Unknown protein
37	AT3G49570	3.3	12.3	5.9	12.4	1.8	10.3	3.4	Response to low sulfur 3
38	AT2G23690	3.2	3.4	4.0	3.9	1.4	2.0	3.0	Unknown protein
39	AT1G31750	3.2	6.7	1.2	10.7	0.6	0.9	1.4	Proline-rich family protein
40	AT1G53541	3.2	5.1	2.0	2.7	0.3	2.5	5.4	Unknown protein
41	AT3G05730	3.2	77.0	115.5	67.7	32.1	57.6	55.5	Encodes a defensin-like (DEFL) family protein.
42	AT2G26020	3.2	4.3	8.6	8.4	5.3	6.4	6.1	Pant defensin 1.2b
43	AT1G53542	3.1	12.2	11.5	12.0	3.6	7.8	17.9	Unknown protein
44	AT1G56240	3.1	6.5	0.5	2.7	0.1	52.1	0.0	Phloem protein 2-B13
45	AT3G16360	3.1	7.5	1.5	6.8	0.5	4.5	1.7	HPT phosphotransmitter 4
46	AT1G74453	3.0	10.0	14.0	14.6	6.7	12.1	22.2	SnoRNA
47	AT4G35720	3.0	3.5	2.9	3.2	0.9	4.7	2.2	Unknown function
48	AT1G07400	3.0	9.3	2.3	10.0	0.8	7.2	3.2	HSP20-like chaperones superfamily protein
49	AT3G62760	3.0	1.1	5.1	2.3	3.5	1.2	5.1	Glutathione S-transferase family protein
50	AT4G17810	3.0	4.8	6.8	6.0	2.9	3.5	6.9	C2H2 and C2HC zinc fingers superfamily protein

Table 5. The 50 most up-regulated genes after induction of *virE3*. The fold up-regulation is calculated by dividing the ratio between the expression level in induced *virE3* plants (TE) to the non-induced *virE*3 plants (NE) by the ratio between the expression level in induced control plants (TC) to the non-induced control plants (NC). TF, expression level in induced *virF* plants; NF, expression level in non-induced *virF* plants. RPKM, reads per kilobase per million. The data are filtered for genes of which the sum of the expression levels in the six samples is more than 15 RPKM.

Rank	Locus	VirE3 Up-reg. (TE/NE)/ (TC/NC)	TE RPKM	NE RPKM	TC RPKM	NC RPKM	TF RPKM	NF RPKM	Name/Description
1	AT5G54720	657.0	35.7	0.1	0.5	0.9	0.8	1.1	Ankyrin repeat family protein
2	AT3G48640	296.5	34.8	0.2	0.1	0.2	0.3	0.1	Unknown protein
3	AT4G04540	212.0	42.7	0.2	0.5	0.4	0.5	0.6	Cysteine-rich RLK (RECEPTOR-like protein kinase) 39
4	AT1G24147	142.0	22.9	1.3	0.0	0.4	0.1	0.4	Unknown protein
5	AT3G46090	102.3	48.9	0.3	3.0	2.1	3.3	1.3	C2H2 and C2HC zinc fingers superfamily protein
6	AT3G60120	75.5	31.9	0.0	0.9	0.0	1.2	0.1	Beta glucosidase 27
7	AT5G13320	71.9	135.0	0.6	1.6	0.5	1.9	0.8	Auxin-responsive GH3 family protein
8	AT5G25260	68.0	71.3	0.3	0.4	0.1	0.4	0.4	SPFH/Band 7/PHB domain-containing membrane-associated protein family
9	AT4G19925	67.6	16.6	0.0	0.3	0.0	1.1	0.1	Toll-Interleukin-Resistance (TIR) domain family protein
10	AT1G19250	52.0	63.3	0.7	2.3	1.3	3.4	1.1	Flavin-dependent monooxygenase 1
11	AT1G21120	51.5	142.0	0.5	8.4	1.5	7.7	0.8	O-methyltransferase family protein
12	AT1G01560	50.1	37.0	0.3	0.7	0.3	1.0	0.4	MAP kinase 11
13	AT1G16420	47.9	27.8	0.9	0.3	0.4	0.8	1.1	Metacaspase 8
14	AT1G26380	42.1	91.9	0.6	3.1	0.9	3.4	0.9	FAD-binding Berberine family protein
15	AT1G30370	41.7	28.2	1.9	2.1	5.8	3.0	4.6	Alpha/beta-Hydrolases superfamily protein
16	AT1G56240	40.7	52.1	0.0	2.7	0.1	6.5	0.5	Phloem protein 2-B13
17	AT4G23210	40.1	23.6	0.4	0.7	0.5	0.5	0.5	Cysteine-rich RLK (RECEPTOR-like protein kinase) 13
18	AT3G23230	39.6	12.2	0.3	1.1	1.1	0.9	0.7	Integrase-type DNA-binding superfamily protein
19	AT1G66090	38.1	89.2	1.2	2.1	1.0	3.2	1.0	Disease resistance protein (TIR-NBS class)
20	AT2G45760	37.2	109.6	0.3	0.4	0.0	2.0	0.3	BON association protein 2
21	AT1G05880	36.6	22.5	0.2	0.4	0.1	0.1	0.2	RING/U-box superfamily protein
22	AT4G19520	33.4	27.7	0.9	1.5	1.7	1.3	1.4	Disease resistance protein (TIR-NBS-LRR class) family
23	AT3G22121	32.5	4.9	0.3	1.7	3.4	1.6	5.0	Other RNA
24	AT5G66640	32.3	15.0	0.6	0.9	1.1	1.6	0.9	DA1-related protein 3

Table 5. Continued

Rank	Locus	VirE3 Up-reg. (TE/NE)/ (TC/NC)	TE RPKM	NE RPKM	TC RPKM	NC RPKM	TF RPKM	NF RPKM	Name/Description
25	AT1G01680	31.8	139.5	0.5	4.2	0.5	4.2	0.3	Plant U-box 54
25	AT1G01680	31.8	139.5	0.5	4.2	0.5	4.2	0.3	Plant U-box 54
26	AT5G64870	31.2	21.2	0.3	1.3	0.7	2.1	0.9	SPFH/Band 7/PHB domain-containing membrane-associated protein family
27	AT5G41750	29.3	59.6	1.0	2.1	1.0	2.8	1.0	Disease resistance protein (TIR-NBS-LRR class) family
28	AT1G56250	28.8	30.5	0.1	1.7	0.1	4.6	0.1	Phloem protein 2-B14
29	AT1G74080	28.1	25.6	0.8	1.8	1.6	3.3	0.9	Myb domain protein 122
30	AT1G56060	27.9	109.2	0.3	6.1	0.5	11.1	0.4	Unknown protein
31	AT5G58120	27.7	12.3	0.6	0.6	0.8	0.8	0.8	Disease resistance protein (TIR-NBS-LRR class) family
32	AT5G22520	27.2	73.4	0.8	2.2	0.7	4.2	0.6	Unknown protein
33	AT1G24140	26.1	18.2	0.3	0.8	0.3	1.4	0.5	Matrixin family protein
34	AT4G19720	26.0	16.4	0.4	1.9	1.1	2.2	1.0	Glycosyl hydrolase family protein with chitinase insertion domain
35	AT5G52760	25.7	144.5	2.5	5.2	2.3	7.1	1.6	Copper transport protein family
36	AT4G11170	25.1	37.4	0.6	0.6	0.2	1.0	0.3	Disease resistance protein (TIR-NBS-LRR class) family
37	AT2G21900	24.8	33.5	1.2	0.6	0.5	1.1	1.1	WRKY DNA-binding protein 59
38	AT5G57220	23.7	157.6	2.2	7.6	2.5	9.4	2.8	Cytochrome P450, family 81, subfamily F, polypeptide 2
39	AT1G57630	23.7	96.2	1.1	2.4	0.6	2.6	1.1	Toll-Interleukin-Resistance (TIR) domain family protein
40	AT1G10340	23.2	23.6	1.1	0.6	0.6	0.7	0.8	Ankyrin repeat family protein
41	AT3G02800	22.7	27.6	0.8	3.0	2.1	6.1	2.4	Tyrosine phosphatase family protein
42	AT5G22380	22.4	18.2	0.8	0.3	0.3	0.3	0.2	NAC domain containing protein 90
43	AT4G19920	21.0	44.3	3.5	2.7	4.6	4.4	3.7	Toll-Interleukin-Resistance (TIR) domain family protein
44	AT5G06090	20.6	18.4	0.4	1.2	0.6	0.7	0.6	Glycerol-3-phosphate acyltransferase 7
45	AT2G04070	20.6	10.4	5.6	0.8	9.0	0.4	6.1	MATE efflux family protein
46	AT5G26920	20.3	60.7	1.4	3.0	1.4	4.0	1.4	Cam-binding protein 60-like G
47	AT4G04490	20.0	23.5	0.7	1.0	0.6	0.4	0.7	Cysteine-rich RLK (RECEPTOR-like protein kinase) 36
48	AT5G27420	19.7	67.9	2.0	7.3	4.1	7.9	3.2	Carbon/nitrogen insensitive 1
49	AT5G10410	19.1	15.6	2.0	1.2	2.8	2.3	2.4	ENTH/ANTH/VHS superfamily protein
50	AT1G14540	18.9	41.5	0.2	3.9	0.4	7.0	0.3	Peroxidase superfamily protein

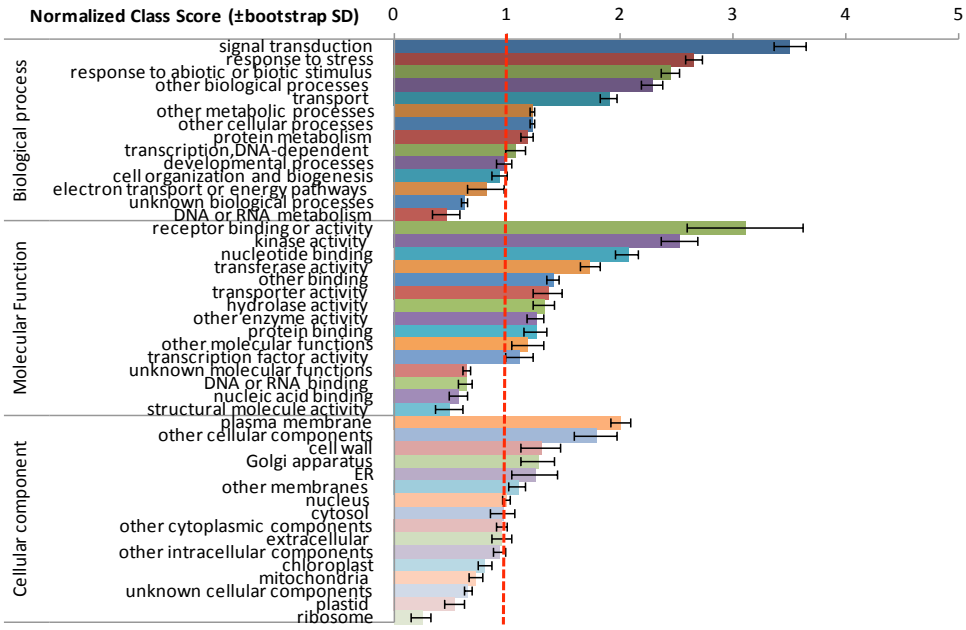
Table 6. The 50 most down-regulated genes after induction of *virE3*. The fold up-regulation is calculated by dividing the ratio between the expression level in non-induced *virE3* plants (NE) to the induced *virE3* plants (TE) by the ratio between the expression level in non-induced control plants (NC) to the induced control plants (TC). TF, expression level in induced *virF* plants; NF, expression level in non-induced *virF* plants. RPKM, reads per kilobase per million. The data are filtered for genes of which the sum of the expression levels in the six samples is more than 15 RPKM.

Rank	Locus	VirE3 Down-reg. (NE/TE)/ (NC/TC)	TE RPKM	NE RPKM	TC RPKM	NC RPKM	TF RPKM	NF RPKM	Name/Description
1	ATCG00080	55.6	2.1	10.2	7.7	0.7	5.5	3.7	Photosystem II reaction center protein I
2	AT1G31750	26.0	0.9	1.4	10.7	0.6	6.7	1.2	Proline-rich family protein
3	AT3G14452	19.9	1.3	10.6	6.4	2.5	5.6	3.5	Unknown protein
4	AT1G03420	19.2	3.0	15.1	9.6	2.5	11.2	12.0	Transposable element gene
5	AT4G30662	18.2	2.2	6.5	7.4	1.2	2.7	3.8	Unknown protein
6	AT1G53541	18.0	2.5	5.4	2.7	0.3	5.1	2.0	Unknown protein
7	AT5G08210	15.6	3.0	8.9	3.9	0.7	6.7	2.9	MIR834a; miRNA
8	AT2G18800	15.3	0.0	5.5	0.5	4.1	0.6	5.6	Xyloglucan endotransglucosylase/hydrolase 21
9	AT5G12020	14.6	9.4	1.2	17.9	0.2	23.1	0.3	17.6 kDa class II heat shock protein
10	AT4G08250	13.9	1.7	0.9	11.2	0.5	9.3	0.7	GRAS family transcription factor
11	AT2G01008	13.6	19.5	19.3	133.6	9.7	115.7	11.3	Unknown protein
12	AT3G27865	13.4	0.4	3.8	2.4	1.6	5.2	2.7	SnoRNA
13	ATCG00630	11.9	9.6	13.4	16.4	1.9	13.6	9.8	PSAJ
14	AT1G19550	10.8	4.0	7.9	12.7	2.3	4.4	3.7	Glutathione S-transferase family protein
15	AT5G09443	9.8	11.4	9.1	91.9	7.5	66.6	8.3	Other RNA
16	AT1G68700	9.0	1.1	5.8	4.1	2.3	1.6	1.3	Unknown protein
17	AT3G29633	9.0	2.0	3.7	5.8	1.2	6.8	2.9	Unknown protein
18	AT1G04425	8.8	2.0	7.9	8.4	3.8	5.4	7.0	Other RNA
19	AT2G47115	8.7	35.9	149.3	188.4	90.2	106.5	42.4	Unknown protein
20	AT3G06690	8.7	2.1	8.6	6.7	3.2	5.5	2.2	Acyl-CoA oxidases; oxidoreductases, acting on the CH-CH group of donors; FAD binding; oxidoreductases; acyl-CoA oxidases
21	ATCG01160	8.6	2.3	7.9	2.3	0.9	2.2	1.8	Chloroplast-encoded 5S ribosomal RNA, which is a component of the 50S large subunit of the plastidic ribosome.
22	AT3G42658	8.5	0.0	0.4	3.2	3.7	9.6	5.2	Transposable element gene
23	AT5G21910	8.2	3.4	17.5	13.4	8.5	13.1	6.8	Unknown protein
24	AT2G05765	8.2	7.5	27.7	16.4	7.5	13.9	13.5	SnoRNA
25	AT3G06900	8.1	1.3	5.3	4.7	2.4	2.3	2.5	U4.2; snRNA

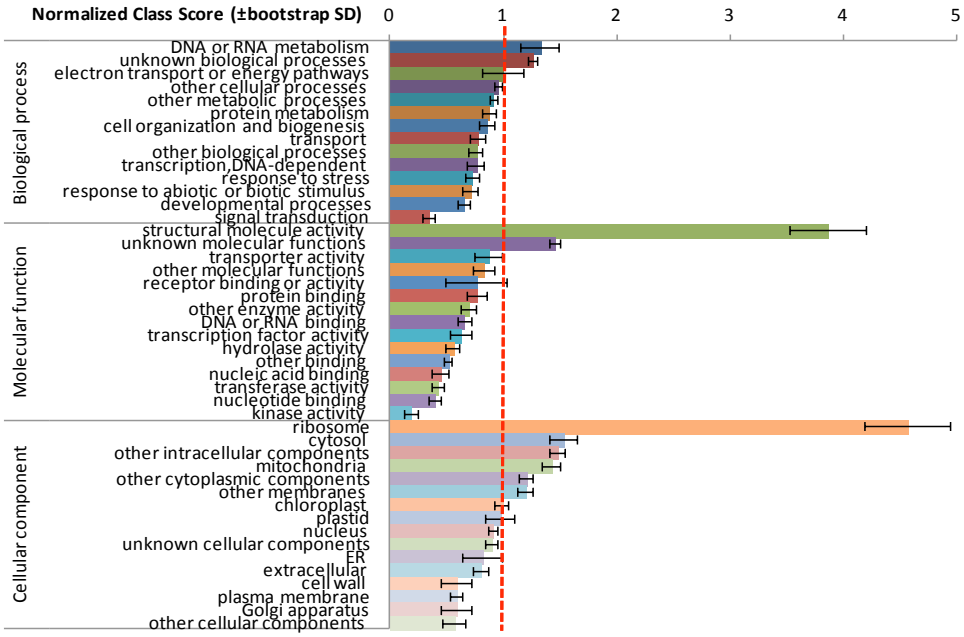
Table 6. Continued

Rank	Locus	VirE3 Down-reg. (NE/TE)/ (NC/TC)	TE RPKM	NE RPKM	TC RPKM	NC RPKM	TF RPKM	NF RPKM	Name/Description
26	AT3G56705	8.0	9.9	27.6	31.3	10.9	18.0	9.7	U2.6; snRNA
27	AT2G39030	7.9	3.0	6.1	4.6	1.2	6.5	7.7	Acyl-CoA N-acyltransferases (NAT) superfamily protein
28	AT4G01575	7.9	1.5	3.2	5.1	1.4	2.8	2.1	Serine protease inhibitor, Kazal-type family protein
29	AT4G39361	7.8	3.3	8.5	7.4	2.5	7.7	2.9	SnoRNA
30	AT1G53542	7.7	7.8	17.9	12.0	3.6	12.2	11.5	Unknown protein
31	AT3G47710	7.6	0.7	10.6	3.1	6.1	3.9	7.9	BANQUO 3
32	AT1G80320	7.4	2.0	3.5	3.4	0.8	4.4	1.2	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
33	AT1G28590	7.4	2.7	0.1	14.0	0.1	12.7	0.1	GDSL-like Lipase/Acylhydrolase superfamily protein
34	AT5G04630	7.1	4.2	0.2	26.5	0.1	23.8	0.0	Cytochrome P450, family 77, subfamily A, polypeptide 9
35	AT4G31398	7.0	1.1	3.9	3.3	1.7	4.1	1.5	Other RNA
36	AT5G16160	6.9	7.9	24.0	24.3	10.8	17.0	11.0	Unknown protein
37	ATMG01370	6.7	1.8	5.4	6.5	3.0	3.8	2.5	Hypothetical protein
38	AT3G41762	6.4	12.9	7.3	102.0	9.0	61.7	5.4	Unknown protein.
39	AT4G38330	6.2	3.1	11.0	11.7	6.7	8.0	5.3	Integral membrane protein hemolysin-III homolog
40	AT3G46230	6.1	7.3	1.1	18.3	0.5	15.9	0.6	Heat shock protein 17.4
41	AT3G13710	6.1	3.0	9.1	7.5	3.7	4.5	2.6	Prenylated RAB acceptor 1.F4
42	AT2G32235	6.1	3.6	14.3	7.6	4.9	6.3	3.6	Unknown protein
43	AT2G46192	6.0	1.9	4.1	4.6	1.7	3.7	3.7	Other RNA
44	AT2G20480	5.8	2.5	9.3	8.9	5.8	5.3	5.0	Unknown protein
45	AT4G20420	5.7	1.7	5.2	3.9	2.1	4.7	3.9	Tapetum specific protein TAP35/TAP44
46	AT4G18510	5.7	0.8	15.3	3.6	12.5	2.1	10.3	CLAVATA3/ESR-related 2
47	AT2G20722	5.7	4.7	19.3	15.8	11.3	14.6	11.7	SnoRNA
48	AT2G43745	5.7	24.5	60.8	53.1	23.0	49.7	28.6	Unknown protein
49	AT3G16360	5.6	4.5	1.7	6.8	0.5	7.5	1.5	HPT phosphotransmitter 4
50	AT1G49620	5.5	8.3	5.5	37.0	4.5	31.0	4.2	Cyclin-dependent kinase inhibitor family protein

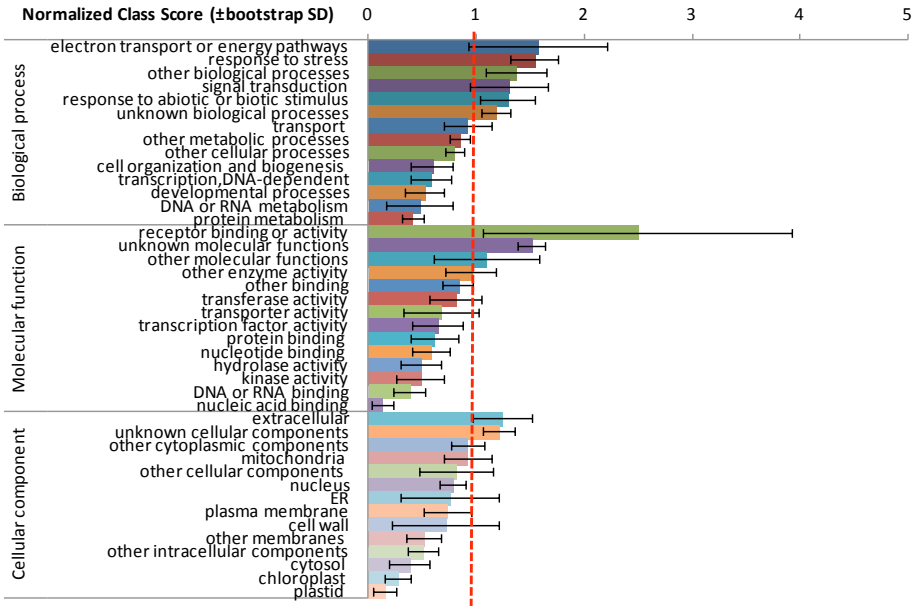
A. *VirE3 Up-regulated genes*



B. *VirE3 down-regulated genes*



C. *VirF Up-regulated genes*



D. *VirF down-regulated genes*

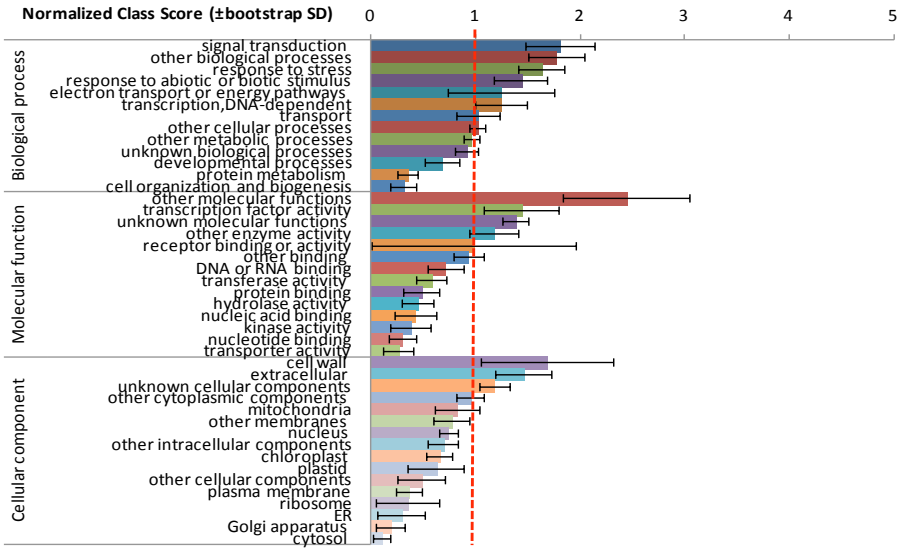


Figure 3. Characterization of the genes up-regulated more than two-fold by expression of *virE3* (A), down-regulated more than two-fold by the expression of *virE3* (B), up-regulated more than two-fold by the expression of *virF* (C) and down-regulated more than two-fold by the expression of *virF* (D). The data are normalized for the total number of *A. thaliana* genes present in that class. Analyses were performed using Superviewer (Provart & Zhu, 2003) and the data available at the TAIR database (www.arabidopsis.org).