

# Structural and functional characterization of *RI*-homologous genes

Inaugural-Dissertation  
zur Erlangung des Doktorgrades  
der Mathematisch-Naturwissen-  
schaftlichen Fakultät der Univer-  
sität Köln

vorgelegt von  
Gábor Miklos Gyetvai  
aus Gronau

Köln 2010

Die vorliegende Arbeit wurde am Max-Planck-Institut für Pflanzenzüchtungsforschung in Köln angefertigt. Die Arbeiten wurden in der Arbeitsgruppe von Dr. PD Christiane Gebhardt innerhalb der Abteilung Pflanzenzüchtung und Genetik von Prof. Maarten Koornneef durchgeführt.

Teile der Arbeit entstanden in Kooperation mit der Universität Aalborg und dem Max-Planck-Institut für Molekulare Pflanzenphysiologie.

Berichterstatter:                    Dr. PD Christiane Gebhardt  
   Prof. Dr. Martin Hülskamp

Tag der mündlichen Prüfung: 09.07.2010

<i>Index of tables and figures</i>	7
<i>Abbreviations and Acronyms</i>	11
<i>Introduction</i>	14
<i>The pathosystem „Solanum tuberosum/Phytophthora infestans” in an agronomical context</i>	14
<i>Molecular interaction between Phytophthora infestans and Solanum tuberosum</i>	18
<i>Genetic basis of resistance of Solanum tuberosum against late blight: Qualitative and quantitative resistance</i>	18
<i>The RXLR-Effectors of Phytophthora infestans</i>	20
<i>Types of resistance mediating genes</i>	21
<i>R1 is a member of the CC-NBS-LRR gene family</i>	23
<i>R1 and the resistance hotspot on Chromosome V</i>	23
<i>Organization of the contigs R1 and r1</i>	24
<i>Solanum nigrum and its role as source for resistance</i>	26
<i>Methods for expression analysis</i>	29
<i>Single target driven methods</i>	29
<i>Methods for transcriptome analysis</i>	30
<i>Massive parallel sequencing</i>	31
<i>Small introduction to the relevant statistical tests for this study</i>	32
<i>Motivation for this study</i>	34
<i>Material and Methods</i>	36
<i>Standard conditions for plant and oomycete cultivation</i>	36
<i>Chemicals</i>	36

<i>Infection assays and plant cultivation</i>	37
<i>Confocal microscopy</i>	37
<i>Molecular biological methods</i>	38
<i>Polymerase chain reaction (PCR)</i>	38
<i>Nucleic acid extraction</i>	41
<i>First strand cDNA synthesis</i>	42
<i>Quantitative polymerase chain reaction (qRT-PCR)</i>	42
<i>Transient expression of effector proteins in Solanum nigrum</i>	43
<i>DeepSAGE</i>	45
<i>Biological material and generation of infected tissue</i>	45
<i>Creation of 3'-tag-libraries</i>	47
<i>Bioinformatical methods for the analysis of transcriptome data</i>	47
Creation of the Standard data set	47
General procedures and sample composition for statistics	48
G-test	50
Fisher's exact test (sage.test)	50
edgeR	51
BaySeq	51
Permu	51
Calculation of the Df and logfoldchange values	52
Cluster analysis	53
Go-term analysis	54
Summary of used Softwares	54
<i>Results</i>	56
<i>Expression analysis of R1-homologous genes of Solanum tuberosum</i>	56

<i>Expression analysis of R1-homologous genes of Solanum nigrum</i>	58
<i>Transient expression of effectors from P. infestans in Solanum nigrum</i>	59
<i>Comparative transcriptome analysis of Solanum tuberosum plants during the early phase of Phytophthora infestans infection</i>	62
<i>Validation of infected Material</i>	62
<i>Technical results</i>	63
<i>Expression levels of indicator-genes</i>	69
<i>Validation of significance statistics</i>	72
<i>Results of global statistical analysis: Cluster and Principal component analysis</i>	81
<i>Results of significance of observation after Fisher's Exact Test</i>	87
<i>Comparison of DeepSAGE-data with semi-quantitative RT-PCR</i>	98
<i>Discussion</i>	103
<i>Evidence for the expression of predicted R1-homologous genes in Solanum tuberosum</i>	103
<i>Evidence for the expression of genomic predicted ORFs in Solanum nigrum</i>	104
<i>Hypersensitive reaction of Solanum nigrum in response to transient expression of effectors of Phytophthora infestans</i>	104
<i>Comparative DeepSAGE-analysis of R1- and ORF45-transgenic Solanum tuberosum plants with wildtype</i>	106
<i>DeepSAGE as method for transcriptome analysis without a reference genome – performance features and uncertainty factors</i>	106
<i>Evaluation of statistical methods – the normal distributed world</i>	108
<i>Validation and characterization of the early stages of infection as phenotype</i>	110
<i>Using Cluster analysis as tool to find coregulated units</i>	111

<i>Identification of candidate genes for a function in the compatible and incompatible interaction with Phytophthora infestans</i>	115
<i>Comparability of DeepSAGE-data with qRT-PCR</i>	118
<i>Concluding remarks - the extraordinary role of the plant hormone synthesis pathway (KEGG ec01070)</i>	119
<i>References</i>	122
<i>Summary (english)</i>	134
<i>Zusammenfassung (deutsch)</i>	135
<i>Protocol for the creation of Ditag-libraries</i>	137
<i>Instructions for the supplementary data</i>	145
<i>Locations of raw data files:</i>	148
<i>Eidesstattliche Erklärung</i>	150
<i>Danksagungen</i>	151
<i>Lebenslauf</i>	154

## Index of tables and figures

### Figures:

<b>Fig. 1.1.1:</b> World map of potato cultivation areas	12
<b>Fig. 1.1.2:</b> The painting: "The King everywhere"	13
<b>Fig. 1.1.3:</b> Alignment of genomic areas of three <i>Phytophthora</i> species	13
<b>Fig. 1.1.4:</b> Scheme of different life stages of <i>Phytophthora infestans</i>	14
<b>Fig. 1.1.5:</b> The life cycle of <i>P. infestans</i>	14
<b>Fig. 1.1.6:</b> Infecting zoospore cysts of <i>Phytophthora cinnamomii</i>	15
<b>Fig. 1.1.7:</b> Symptoms of late blight infections	15
<b>Fig. 1.2.1.1:</b> Race and R-gene nomenclature of <i>Solanum tuberosum</i> and <i>Phytophthora infestans</i>	16
<b>Fig. 1.2.1.2:</b> Scheme of qualitative and quantitative resistance	17
<b>Fig. 1.2.2.1:</b> Scheme of the interaction zone between pathogen and plants	18
<b>Fig. 1.2.3.1:</b> Illustration of the known classes of R-genes	19
<b>Fig. 1.3.1.1:</b> Picture of the <i>R1</i> -specific phenotype	21
<b>Fig. 1.3.2.1:</b> Map of the contigs <i>R1</i> and <i>r1</i>	22
<b>Fig. 1.3.2.2:</b> Phenetic tree of the <i>R1</i> -family	23
<b>Fig. 1.3.2.3:</b> Function map of <i>Solanum tuberosum</i> chromosome V	23
<b>Fig. 1.4.1:</b> Taxonomic view of some members of the <i>Solanaceae</i> family	24
<b>Fig. 1.4.2:</b> Picture of <i>Solanum nigrum</i>	24
<b>Fig. 1.4.3:</b> Illustration of a syntenic region between <i>Solanum tuberosum</i> and <i>Solanum nigrum</i>	25
<b>Fig. 1.4.4:</b> Identification of cosmid clones with <i>R1</i> -homologous sequences in <i>Solanum nigrum</i>	26
<b>Fig. 1.4.5:</b> Southern blot with a <i>R1</i> -specific probe and several cultivars	26
<b>Fig. 1.5.1.1:</b> Scheme of the bDNA-assay	27
<b>Fig. 1.5.2.1:</b> Scheme of the generation of SAGE-libraries	29
<b>Fig. 1.8.1:</b> Model of the mechanism of <i>R1</i>	32
<b>Fig. 2.6.1.1:</b> Overview of infections for DeepSAGE	44
<b>Fig. 2.6.3.1.1:</b> Density plot of mean expression levels	45
<b>Fig. 2.6.3.8.1:</b> Explaining example of the theory behind the $D_r$ -values	50
<b>Fig. 2.6.3.9.1:</b> Example of the centroid expression view of cluster analysis	51

<b>Fig. 3.1.1:</b> Results of the BAC-assay used for the development of primers for the <i>R1</i> -family	54
<b>Fig. 3.1.2:</b> Illustration of the used BACs and the position of the <i>R1</i> -family members	54
<b>Fig. 3.1.3:</b> Verification if transcripts of members of the <i>R1</i> -family	55
<b>Fig. 3.2.1:</b> Verification of expression of <i>R1</i> -homologous genes in <i>Solanum nigrum</i>	56
<b>Fig. 3.3.1:</b> Summary of results after expression of put. effectors from <i>Phytophthora infestans</i> in <i>Solanum nigrum</i>	57
<b>Fig. 3.3.2:</b> Alignment of positive effectors	57
<b>Fig. 3.3.3:</b> Phenetic tree of all tested put. effectors	59
<b>Fig. 3.4.1.1:</b> Microscopy picture of an <i>Phytophthora infestans</i> infected leaf three days after infection	60
<b>Fig. 3.4.2.1:</b> Plot of the number of sequenced tags and the detected unitags per sample after „Next-Generation-Sequencing“	65
<b>Fig. 3.4.2.2:</b> Pie chart of the proportion of possible target sites	66
<b>Fig. 3.4.2.3:</b> Histogram of the quantity of expression levels	67
<b>Fig. 3.4.3.1:</b> Heat map of tags with a go-annotation „defense-response“	68
<b>Fig. 3.4.3.2:</b> Histogram of the expression levels of <i>ef-1a</i>	69
<b>Fig. 3.4.4.1:</b> Histogram of the FDR-values of the G-test	70
<b>Fig. 3.4.4.2:</b> Plot of the FDR-values and the logfoldchanges after G-test	71
<b>Fig. 3.4.4.3:</b> Plot of the FDR-values and the logfoldchanges after Fisher’s exact test (sage.test)	71
<b>Fig. 3.4.4.4:</b> Plot of the FDR-values and the logfoldchanges after Student’s t-test	72
<b>Fig. 3.4.4.5:</b> Plot of the FDR-values and the logfoldchanges after edgeR	73
<b>Fig. 3.4.4.6:</b> Plot of the FDR-values and the logfoldchanges after Permu	74
<b>Fig. 3.4.4.7:</b> Plot of the logP-values and the logfoldchanges after BaySeq	75
<b>Fig. 3.4.4.8:</b> Box-plots of the average logfoldchange values of the significant	

tags of each test	76
<b>Fig. 3.4.4.9:</b> Histogram of the average $D_f$ -values of the significant tags of each test	77
<b>Fig. 3.4.4.10:</b> Venn diagram of the top100 tags by each test	78
<b>Fig. 3.4.5.1:</b> Overview of extracted cluster after k-mean cluster analysis	80
<b>Fig. 3.4.5.2:</b> Apertures of of a heat map after hierarchical cluster analysis	82
<b>Fig. 3.4.5.3:</b> Plot of PC1 and PC2 of a principal component analysis	84
<b>Fig. 3.4.6.1:</b> Overrepresented GO-terms in <i>R1</i> lines 3dpi	91
<b>Fig. 3.4.6.2:</b> Overrepresented GO-terms in wild type lines 1dpi	92
<b>Fig. 3.4.6.3:</b> Overrepresented GO-terms in wild type lines 3 dpi	93
<b>Fig. 3.4.6.4:</b> Overrepresented GO-terms in the ORF45 line 1dpi	94
<b>Fig. 3.4.6.5:</b> Overrepresented GO-terms in the ORF45 line 3dpi	95
<b>Fig. 3.4.7.1:</b> Comparing matrix of DeepSAGE and qRT-PCR	99
<b>Fig. 3.4.7.4:</b> Boxplot of overlapping and not overlapping values	100
<b>Fig. 4.4.3.1:</b> Expression levels of Prb-1b in wildtype	109
<b>Fig. 4.4.4.1:</b> The putative reaction of TC16055	111
<b>Fig. 4.4.5.1:</b> The reaction of ToTAL2	114
<b>Fig. 4.4.5.2:</b> Expression levels of ToTAL2 (StET009648)	114
<b>Fig. 4.4.7.1:</b> Overview of the plant hormone synthesis pathway	119

**Tables:**

<b>Tab. 1.6.2:</b> Performance features of „Next Generation Sequencing“ platforms	29
<b>Tab. 2.5.1.1:</b> List of used oligonucleotides	37
<b>Tab. 2.6.3.2.1:</b> Definition of replica groups and independent experiments	47
<b>Tab. 2.6.3.11.1:</b> Overview of used softwares	53
<b>Tab. 3.3.1:</b> Similarity scores between positive scored put. effectors and <i>avrblb2</i>	58
<b>Tab. 3.4.2.1:</b> Overview of the results of „Next Generation Sequencing“	62
<b>Tab. 3.4.4.1:</b> Summary of the main characteristics of the statistical tests	74
<b>Tab. 3.4.4.2:</b> Correlation matrix of the Df-values if the significant tags by each test	76
<b>Tab. 3.4.5.1:</b> Summary of relevant clusters after k-mean cluster analysis	79
<b>Tab. 3.4.5.2:</b> Summary of identified tags after hierarchical clustering	83
<b>Tab. 3.4.6.1</b> Major significant tags of <i>R1</i> lines	85
<b>Tab. 3.4.6.2</b> Major significant tags of wild type lines	86
<b>Tab 3.4.6.2:</b> Major significant tags of the ORF45 line	89
<b>Tab. 3.4.7.1:</b> Summary of tags used for qRT-PCR experiments	96
<b>Tab. 3.4.7.2:</b> Expression levels of qRT-PCR and according DeepSAGE data	97

## Abbreviations and Acronyms

Abbreviation	Full expression
°C	degree Celsius
A. lfc.	Average logfoldchange
ATP	adenosine 5-triphosphate
Avr	avirulence
BAC	Bacterial artificial chromosome
bp	base pair
C.I.	Compatible interaction
cDNA	complementary DNA
chr.	chromosome
Ct	cycle threshold
cv.	cultivar
D <sub>f</sub>	Differentiation factor
DFCI	Dana-Farber Cancer Institute
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxynucleosidetriphosphate
dpi	days post infection
FDR	False discovery rate; also: false discovery rate corrected p-value
FE	Fisher's exact test
FE	Fisher's exact test
fig.	figure
gDNA	genomic DNA
GFP	green fluorescence protein

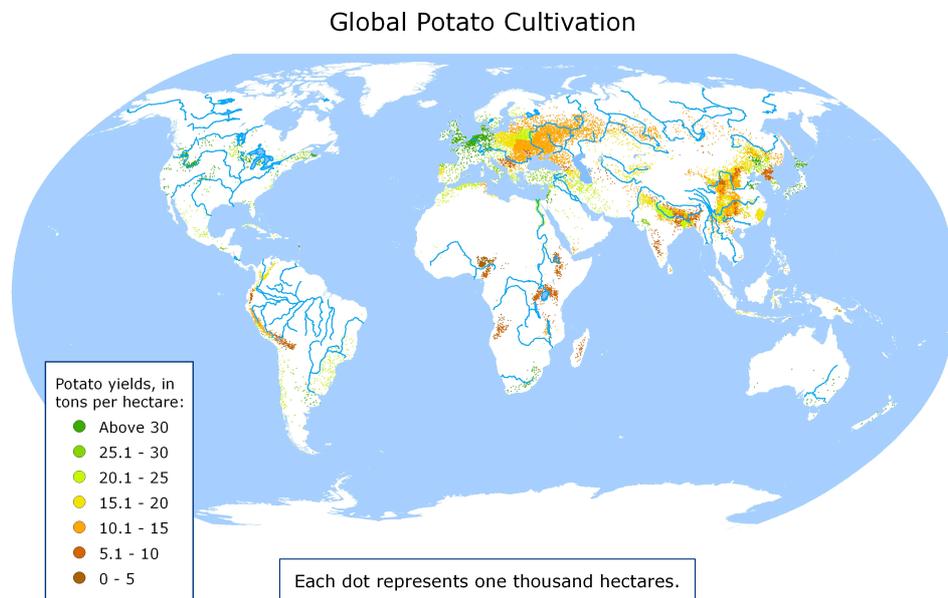
Abbreviation	Full expression
GO	Gene ontology
HR	hypersensitive response
In. I.	Incompatible interaction
JVCI	John Craig Venter Institute
kb	kilobase pairs
kDA	kilodalton
KEGG	Kyoto Encyclopedia of Genes and Genomes
LB	lysogeny broth
log	logarithm
LRR	leucine rich repeat
m	mili
m	mili
M	Molar
ma.lfc	maximal logfoldchange
max	maximum
MES	2-(N-morpholino)ethanesulfonic acid
MgCl <sub>2</sub>	Magnesium chloride
mi.lfc	minimal logfoldchange
min	minimum
NA	No annotation
NBS	nucleotide binding site
NCBI	National Center for Biotechnology Information
OD	optical density
ORF	open reading frame
P.	<i>Phytophthora</i>
PAMP	pathogen-associated molecular pattern

Abbreviation	Full expression
PC	principal component
PCA	principal component analysis
PCR	polymerase chain reaction
PR	pathogenesis related
PVX	Potato virus X
qRT-PCR	quantitative reverse transcriptase polymerase chain reaction
QTL	quantitative trait locus
rel. expr.	relative expression
RT-PCR	reverse transcriptase polymerase chain reaction
S.	<i>Solanum</i>
SAGE	Serial Analysis of Gene Expression
SCRI	Scottish Crop Research Institute
ssp.	subspecies
T <sub>A</sub>	Annealing temperature
TIGR	The Institute for Genomic Research
TIR	Drosophila Toll and human unterleukin-1 receptor
vs	versus
wt	wild type
YEP	Yeast extract peptone
Δ	delta
μ	micro

## 1.Introduction

### 1.1.The pathosystem „*Solanum tuberosum*/*Phytophthora infestans*” in an agronomical context

*Solanum tuberosum* is the most cultivated member of the *Solanaceae* family. With a total yearly production of approximately 323 million tons (source: FAOSTAT). It is after maize, wheat and rice the fourth most important crop plant worldwide. The estimated size of the genome is 840 Mbp which are distributed over 12 linkage groups. The natural occurring



**Fig. 1.1.1:** World map of the areas of potato cultivation worldwide (source: International potato Center).

ploidy level is tetraploid, leading to a total chromosome number of 48 (source: SGN [1]). Differing from most polyploid crop species, the segregation of these four chromosome sets occurs independently. The major producing areas are in Asia and Europe (fig. 1.1.1). Originating from the Andes of South America, where it has been cultivated since at least 7000 B.C., the potato has been introduced into Europe in the course of the discovery of America in the sixteenth century. Originally it has been spread as rare ornamental plant [2]. As cultivated plant for nutrition it was initially used at the end of the seventeenth

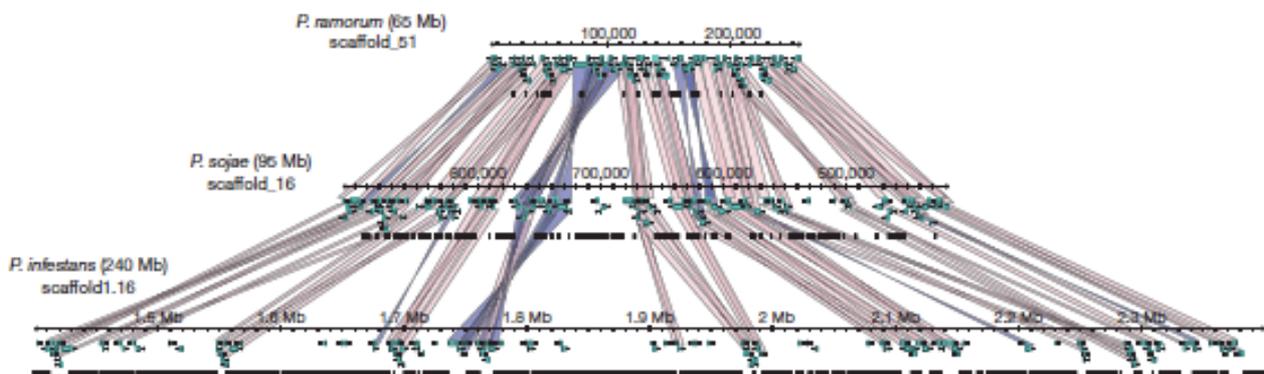
century in Ireland. An additional step forward on its way as agronomical relevant crop plant in Middle Europe was the introduction in Prussia by law in 1756 through King Frederik II (illustrated in fig. 1.1.2). The basic prerequisite was the adaption of the agronomically used potatoes to the Middle European long day conditions. The originally introduced cultivars (mainly from the subspecies *andigena*) were adapted to the South American short day conditions [2, 3].



**Fig. 1.1.2:** Painting with the title: „The King everywhere" by Robert Warthmüller (1886). It presents King Frederik of Prussia inspecting the Potato cultivation. The original is located in the German historic museum in Berlin.

The broad introduction as a major source of nutrition lead to the establishment of large monocultures. These gave the breeding ground for pathogens and associated diseases. One of the most destructive pathogens of *Solanum tuberosum* and related species is the late blight causing oomycete *Phytophthora infestans* [3].

At the moment, the taxon *Phytophthora* contains 110 classified species and additional 232, which are not classified yet (source: NCBI [4]). Among the sequenced genomes of this Phylum, *Phytophthora infestans* has, with 240 Mbp, the largest known genome with large expanding areas compared to the related species *Phytophthora ramorum* and

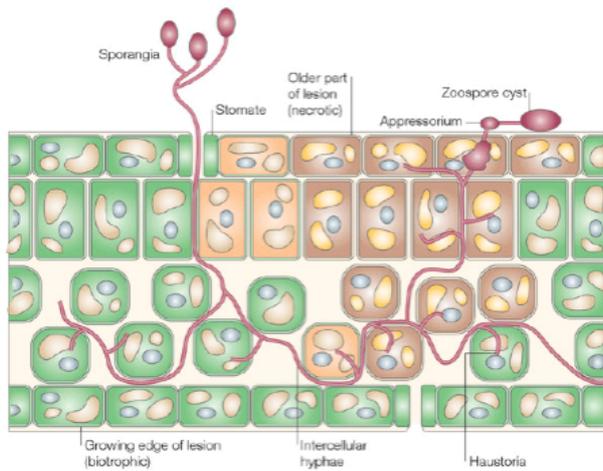


**Fig. 1.1.3:** Schematic alignment of genomic areas from the three *Phytophthora* species *Phytophthora ramorum* (top), *sojae* (middle) and *infestans* (bottom) [5].

*Phytophthora sojae* (fig. 1.1.3). These large intergenic spaces could provide a genetical explanation for the high potential of this species to adapt to changing environments [5].

The life cycle of *Phytophthora infestans* spans three major stages (fig 1.1.4 1.1.5). The oomycete is able to persist in form of sporangia in old infected tissue parts like tubers from

the previous year or in form of thick-walled oospores in the soil [6]. Mainly the oospores are responsible for the infections of the next year [7, 8]. However, during an ongoing



Nature Reviews | Microbiology

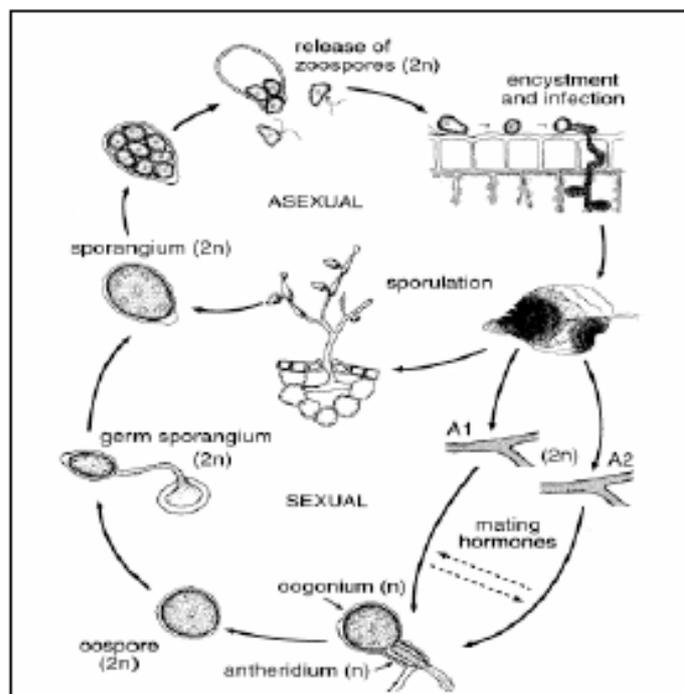
**Fig. 1.1.4:** Scheme of the different life stages of *Phytophthora infestans* [147].

epidemic, the Sporangia are the more active part. The major reason for this is their ability to release mobile zoospores. The released zoospores by themselves are the major infective and virulent element. The biflagellar zoospores, start after recognition of host tissue, to build hyphae so called zoospore cysts [9]. One of the first described stimuli is a perception of a  $\text{Ca}^{2+}$ -concentration gradient [10, 11]. From this cyst an initial hyphae is released and an appressorium is formed under contact with the host. This appressorium is the first interaction zone

between the pathogen and its host from which the infection of the plant tissue starts. Inside

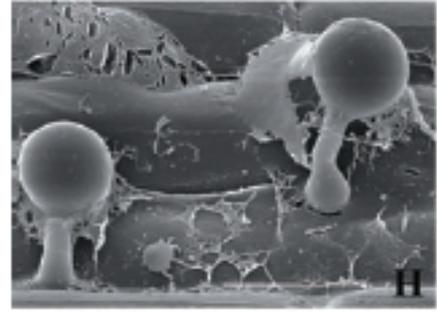
the leaf tissue a mycelium with haustoria is formed. After 5-8 days the mycelium grows out of the plant again and forms new sporangia, which then infect new parts of the plant, new plants or remain in the soil. The reproduction of this organism is sexual and asexual [12]. Comparable to *Saccharomyces cerevisiae*, this pathogen has two mating types – the  $A_1$  and  $A_2$  type. If both types meet, oogonia are formed and inside recombination is performed. From these oogonia new oospores are produced. One of the first and most prominent late blight epidemic has

caused the Irish famine in the 1840s [13]. After this, smaller epidemics occurred in Belgium



**Fig. 1.1.5:** Schematic picture of the life cycle of *P. infestans* [146].

and the Netherlands, leading to an increased focus on this problem [3]. Resistance breeding and agronomic measures reduced this problem in Europe to a minimum until the 1980s. At this time, fields under late blight infection has been observed again. A hypothesis with the aim to explain this phenomenon postulates that to this date, only the A<sub>1</sub> mating type was introduced to Europe and the resistance, mediated by genes introgressed from the wild relative *Solanum demissum*, was durable. With new trading imports from South America, the A<sub>2</sub> mating type was introduced, leading to an increased genomic variability and a fast breaking of the existing resistance genes as well as a tolerance to fungicides like metalaxyl [14].



**Fig. 1.1.6:** Infecting zoospore cysts from the related species *Phytophthora cinnamoni* [9].



**Fig. 1.1.7:** Pictures of a leaf, showing symptoms of a late blight infection (left) and of a field under disease (right) (Source: SCRI).

With its ability to infect large areas in a short time under favorable conditions (high humidity and temperatures between 15 and 20 °C [3]) and its high adaptiveness to environmental conditions, this pathogen is of high agronomic interest. Current estimates calculate worldwide an economical loss of at least 5 billion US\$. These include the costs for the chemical control and the direct losses in yield [15].

## 1.2.Molecular interaction between *Phytophthora infestans* and *Solanum tuberosum*

### 1.2.1.Genetic basis of resistance of *Solanum tuberosum* against late blight: Qualitative and quantitative resistance

The classic view of the gene-for-gene model assumes major single locus resistance genes interacting with corresponding avirulence genes [16, 17]. Eleven avirulence (*avr*) genes in *Phytophthora infestans* and accordingly eleven resistance (*R*) genes in *Solanum tuberosum* are described (fig. 1.2.1.1). Five of these genes, *R1*, *R2*, *R3*, *R6* and *R7*, have been located on the genetic map [18, 19]. Following this categorization system, the race-nomenclature of *Phytophthora infestans* strains is defined. This nomenclature follows the

	Race1 ( <i>vir1</i> )	Race2 ( <i>vir2</i> )	Race3 ( <i>vir3</i> )	Race4 ( <i>vir4</i> )	Race5 ( <i>vir5</i> )	Race6 ( <i>vir6</i> )	Race7 ( <i>vir7</i> )	Race8 ( <i>vir8</i> )	Race9 ( <i>vir9</i> )	Race10 ( <i>vir10</i> )	Race11 ( <i>vir11</i> )
<b>R1</b>	C. I.	In. I.	In. I.	In. I.	In. I.	In. I.	In. I.	In. I.	In. I.	In. I.	In. I.
<b>R2</b>	In. I.	C. I.	In. I.	In. I.	In. I.	In. I.	In. I.	In. I.	In. I.	In. I.	In. I.
<b>R3</b>	In. I.	In. I.	C. I.	In. I.	In. I.	In. I.	In. I.	In. I.	In. I.	In. I.	In. I.
<b>R4</b>	In. I.	In. I.	In. I.	C. I.	In. I.	In. I.	In. I.	In. I.	In. I.	In. I.	In. I.
<b>R5</b>	In. I.	In. I.	In. I.	In. I.	C. I.	In. I.	In. I.	In. I.	In. I.	In. I.	In. I.
<b>R6</b>	In. I.	C. I.	In. I.	In. I.	In. I.	In. I.	In. I.				
<b>R7</b>	In. I.	C. I.	In. I.	In. I.	In. I.	In. I.					
<b>R8</b>	In. I.	C. I.	In. I.	In. I.	In. I.						
<b>R9</b>	In. I.	C. I.	In. I.	In. I.							
<b>R10</b>	In. I.	C. I.	In. I.								
<b>R11</b>	In. I.	C. I.									

**Fig. 1.2.1.1:** Schematic matrix of the classical potato resistance model and the race specification of *Phytophthora infestans*. *R1-R11* stand for the different resistance genes introgressed from *Solanum demissum*. Race1-11 specifies the nomenclature of *Phytophthora* strains, classified after the possible interaction with an *R*-gene. C.I means compatible interaction (leading to an infection of the host). In. I means incompatible interaction (leading to an hypersensitive resistance response of the host).

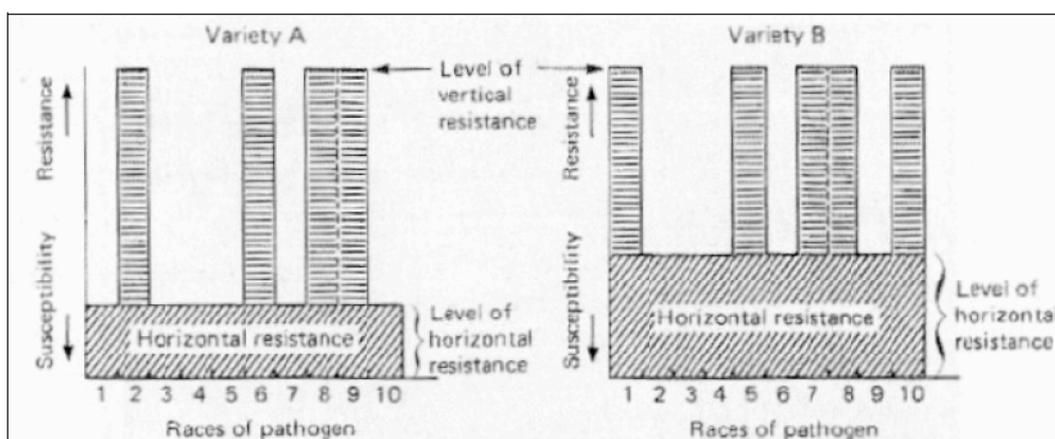
virulence (or the ability to overcome the resistance gene) of the oomycete and consequently introduces a virulence (*vir*) gene nomenclature. The qualitative resistance genes are able to sense the corresponding avirulence gene. This perception of the

pathogen induces a resistance response in form of a hypersensitive cell death reaction. This kind of defense response can be seen in analogy to the inflammation in animal systems [20]. Formal pathology calls this type of host/pathogen interaction as “incompatible interaction“ between the host and the pathogen. The opposite a successful infection of the host, what is termed as “compatible interaction“.

The genetic model of resistance is based on the resistance genes, which have been introgressed before the A<sub>2</sub> mating type was introduced to Europe and the complexity of pathogen strains potentiated. The contemporary active populations in Middle Europe are classified after this scheme and usually exist as complex or semi-complex races, containing a mix of the known virulences or even all. At the moment the virulence identification is carried out in an infection assay using a standard set of potato cultivars. Current resistance breeding aims to introgress new resistance genes from wild species other than *Solanum demissum*, such as the *Rpi-blb* genes [21] from *Solanum bulbocastanum*. This will make a new classification system necessary in the future.

Due to a non-mendelian segregation of *R10* and the detection that the *R3* locus consists out of the two closely linked genes *R3a* and *R3b*, the molecular basis of the *R*-genes stays in discussion [22, 23].

In addition to the qualitative resistance, also the quantitative or field resistance is known [24, 25]. The underlying physiology is rather a strengthening of the plant. This phenotype is additionally known as horizontal resistance. The genetic basis can be genes like *StAOS2* [26], having a role in jasmonate synthesis. Jasmonic acid is involved in unspecific stress signalling [19, 26-28].

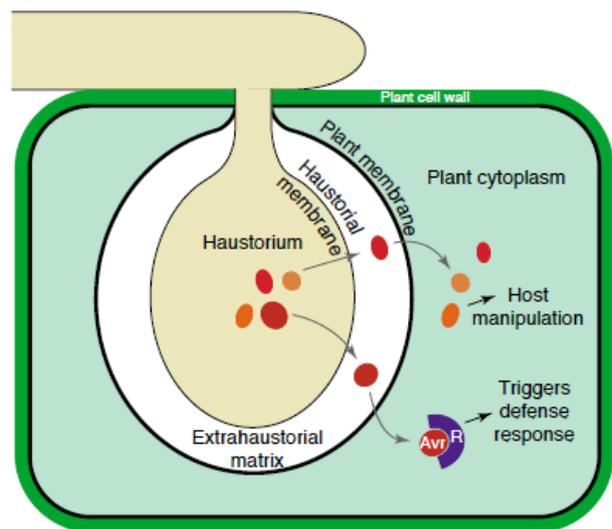


**Fig. 1.2.1.2:** Comparative scheme of the relationship between quantitative and qualitative resistance within the phenotype resistance as whole [146].

The distribution of these phenotypes follows the rules of quantitative traits. This differs from *R*-genes, which follow the distribution of qualitative traits. In plants, both traits are acting together. On one side the horizontal resistance which mediates an unspecific protection against a broad range of pathogens and on the other side the specific resistance genes, which recognize specific stimuli (fig. 1.2.1.2).

### 1.2.2. The RXLR-Effectors of *Phytophthora infestans*

Any phase of the infection requires the establishment of an interaction zone between the host and the pathogen. For this, a special cellular structure, named as „haustorium“ is established (fig. 1.2.2.1). Outgoing from this the further interconnection is directed. During the initial stage, *Phytophthora* secretes various proteins and molecules, which prepare for the further colonization. This progress, known as cellular reprogramming, is essential to protect the pathogen against unspecific defense responses from the host. To this belongs the secretion of antimicrobial proteins, like chitinases, bactinecin, serine- and cysteine protease inhibitors and other direct attacking molecules [29, 30]. These secreted proteins, are currently, together with traditional elicitors, summarized as effectors. These effectors are released in the apoplastic space as well as in the cytoplasm [5, 29]. Recently, two classes of effector protein could be classified. Based on their characteristic amino acid motif they are called RXLR (harboring the sequence RxLR) and crinkler (CRN motif). These motifs seem to be highly conserved among secreted proteins and play a role for the introduction into the host-cell [15, 30-33]. It is assumed that avirulence and virulence genes are present among this genes, playing the major part of the decision between incompatible and compatible interaction [34, 35].

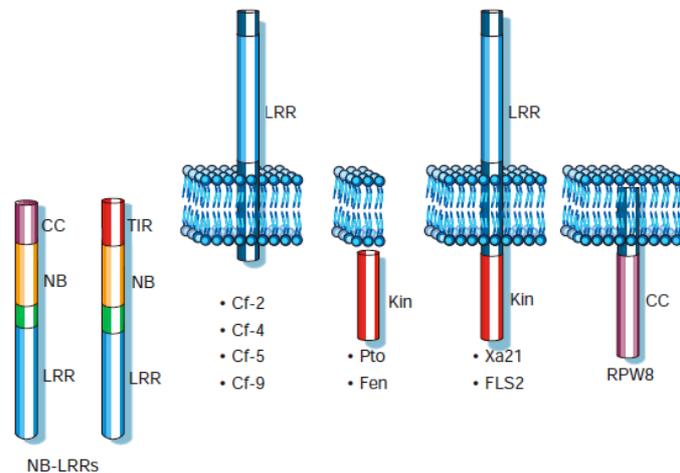


**Fig. 1.2.2.1:** General scheme of the interaction compartment between of pathogens and the plant [148].

### 1.2.3.Types of resistance mediating genes

Currently, 5 different classes of gene products are known in plants, which mediate resistance against pathogens.

These classes are built from five different domains. The major class, the NBS-LRR class, combines a nucleotide binding site (NBS-domain) and leucine rich repeat (LRR domain). Additionally, the N-terminal region contains a coiled-coil domain (CC) or a Toll-interleukin receptor like (TIR) domain. The TIR- and CC-NBS-LRR type genes are summarized as NB-ARC type genes, which



**Fig. 1.2.3.1:** Schematic illustration of the five known classes of resistance genes and their localization in relation to the cellular membrane. Additionally known examples of plant resistance genes are given below [39].

share structural similarities with the NACHT-LRR or CATERPILLER family in animal systems [20, 36, 37]. This class of genes is widely spread and early estimations prognosed more than 200 genes of this type in *Arabidopsis thaliana* [38]. In addition, a kinase domain can be found among resistance genes [34, 39]. The simplest resistance gene, detected so far, is the *pto* gene from *Solanum lycopersicum*, which only contains a kinase domain and is classified as serine/threonine kinase. This gene is not able to mediate resistance on its own and needs *prf* as a second gene, which belongs to the NBS-LRR type class [40-42]. One of the major differences within the resistance genes affects the intracellular localization so far known. On the one hand genes exist, which are associated to the cellular membrane. To this group belong the *cf*-genes or *fls2*. In the case of *fls2*, it has been shown, that the gene is responsible for the extracellular recognition of a certain kind of Flagellin (*flg22*) and belongs to the class of PAMP-receptors [43]. On the other hand, the NBS-LRR genes have not been found to be anchored in the membrane Although an indirect association might be possible [39]. Despite of this fact, it should be remembered, that a clear differentiation of PAMP-triggered immunity and R-gene triggered immunity should not be made as both parts belong to the higher order defense system and

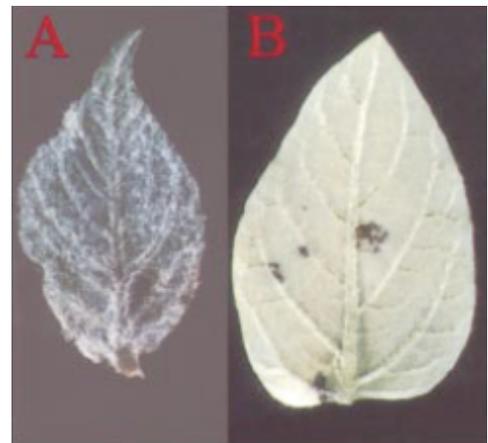
probably both ways stand in a clear cross-talk [44]. As outcome of both mechanisms, the hypersensitive response remains the same [45]. The precise molecular mechanism of the avirulence factor recognition through the NBS-LRR type genes is still unknown. Up to date it is not clear whether it is a direct molecular interaction or a sensing mechanism via an indirect „guarding“ molecule as it is proposed by the guard hypothesis [46].

### 1.3.*R1* is a member of the CC-NBS-LRR gene family

#### 1.3.1.*R1* and the resistance hotspot on Chromosome V

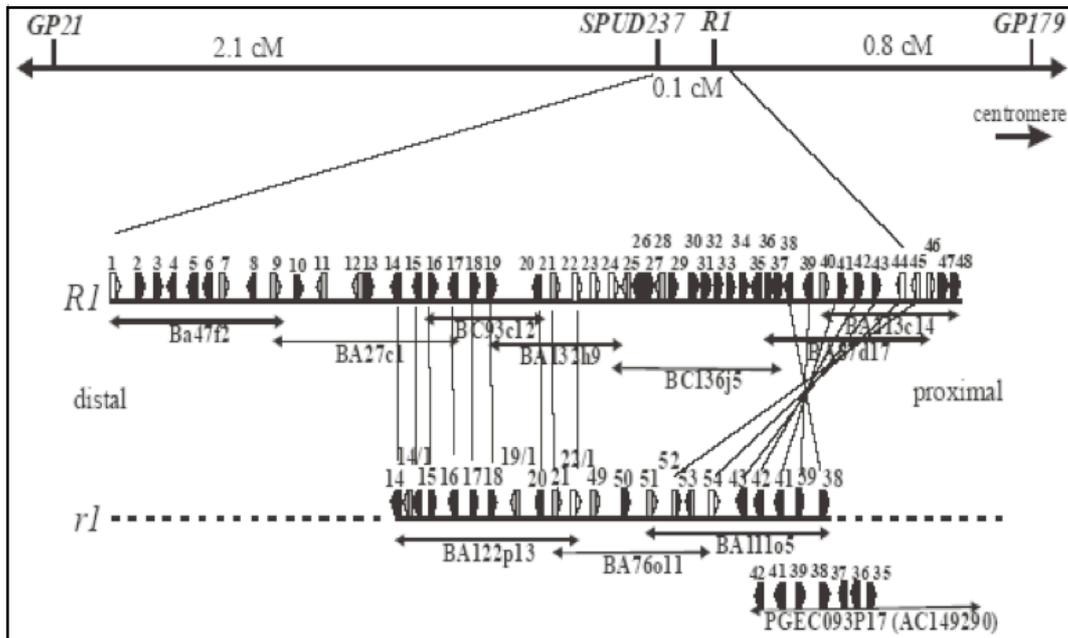
Resistance genes are often organized in clusters. In *Solanum tuberosum*, several of these hot spots for resistance are known. These often co-localize with QTL regions for resistance as well. One of the most prominent hot spots for resistance is localized on chromosome V between the anchoring markers *GP21* and *GP179* [47] (for a comparison see Fig. 1.3.2.3). In addition to the known late blight resistance gene *R1*, this region harbors several genes with function in resistance to nematodes and viruses as well as QTL-regions being involved in resistance to late blight and nematodes. Additionally, a part of a QTL-region for resistance to bacterial diseases [48, 49] is located in this region. Beside many additional relevant regions, other remarkably hot spots for resistance are located on chromosome III tagged by the markers *TG134* and *Pt2* [50-52], on chromosome IV tagged by *TG62* and *GP180-a* [48, 49], on chromosome V close to *GP78* and *GP22* [48, 49, 53-56] and on chromosome XI near *STM2005* and *STM0025* [50, 57].

*R1* by itself has been shown to mediate qualitative resistance [54] and the predicted gene product encodes a protein of 1313 amino acids. Based on the Prosite database, it contains a leucine zipper domain and a nucleotide binding site. Additionally four glycosylation sites, a cAMP-phosphorylations site, various CKcasein kinase II- and phosphokinase c-phosphorylation sites, as well as 3 sites for myristilation and one amidation site can be found (Source: Prosite database [58]). Summing up the predicted motif information, this gene is formally classified as CC-NBS-LLR class gene and belong to the group of classical single locus resistance genes.



**Fig. 1.3.1.1:** Picture of the *R1*-specific phenotype. Seen is a leaf of the potato cultivar Désirée (A) and from a *R1*-transgenic Désirée plant 9 days after infection with a *P. infestans* race 4 isolate [54].

### 1.3.2. Organization of the contigs *R1* and *r1*

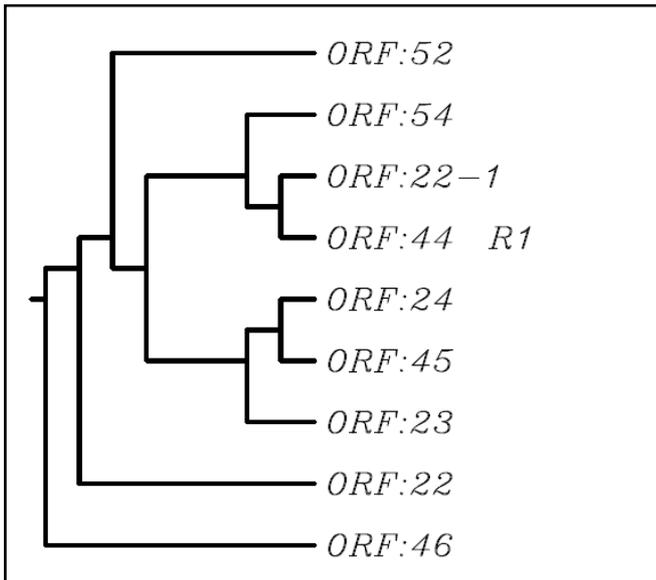


**Fig. 1.3.2.1:** Map of the sequenced contig of *Solanum tuberosum* cv. P6/210. Two differing orthologues regions named as *R1* and *r1* were identified. The *R1* contig harbors the *R1* gene. On both contigs, in total 8 open reading frames have been identified, whose putative products share a high similarity with *R1* [55].

Sequencing and annotation of 10 BAC clones located in the region of the resistance hot spot on chromosome V of *Solanum tuberosum* was able to unravel various open reading frames (ORFs) (see Fig. 1.3.2.1). Among these were eight ORFs, whose predicted products share a high degree of sequence homology to the known resistance gene *R1*. These open reading frames are located on the two contigs *R1* and *r1*, which are likely of different origin. Differing to the *r1*-contig, the *R1*-contig originated from an introgressed region from *Solanum demissum*. The six open reading frames 22, 23, 24, 44 (*R1*), 45, and 46 are located on the introgressed region. In the *r1*-contig, with number 22-1, 52 and 54, three additional ORFs have been detected. Together, these nine open reading frames form the *R1*-family [55].

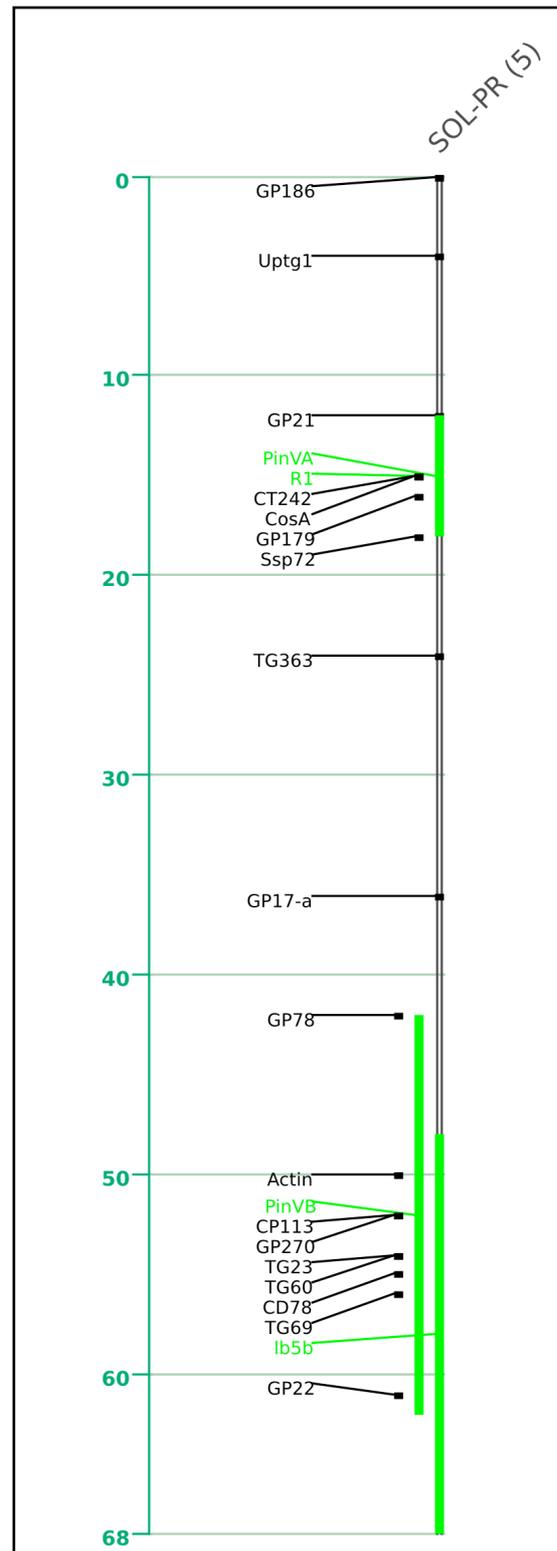
The relationship between the members of the *R1* gene family seems to form two homology groups with three members each and a third group with distantly related members. The

first group contains the *R1* gene, ORF54 and ORF 22-1. Remarkably, both open reading frames being most similar to *R1*, are located on the *r1*-contig. A different observation is made in the second group containing ORF24, ORF45 and ORF23. All of these open reading frames are localized on the *R1*-contig. The greatest similarity



**Fig. 1.3.2.2:** Phenetic tree of nine open reading frames, sharing a high sequence homology with the *R1* gene. The tree is based on the predicted amino acid sequences.

occurs between ORF24 and ORF45. The genomic sequence of these loci share a homology of almost 100% several kilobases up- and downstream of the putative coding region. It is most likely that the members of this family evolved from duplication events and these two open reading frames reflect the youngest. The most diverse member of this group of predicted genes is ORF 46. (see Fig. 1.3.2.2)



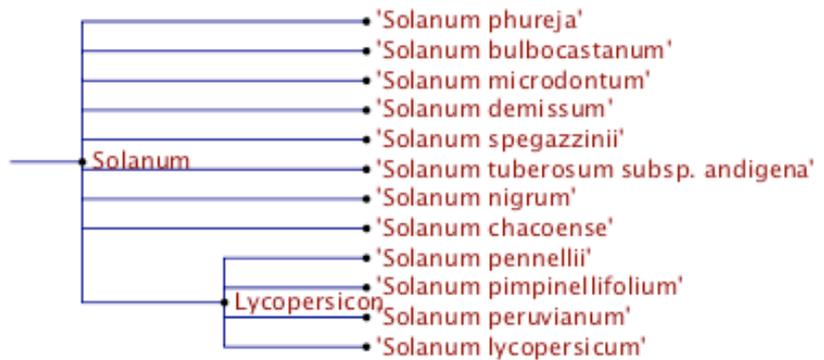
**Fig. 1.3.2.3:** Function map of chromosome V of *S. tuberosum*. The green bars indicate QTL regions for resistance to fungal pathogens and oomycetes (source:PoMaMo (modified) [18]).

## 1.4. *Solanum nigrum* and its role as source for resistance

*Solanum nigrum*, black nightshade with common name, belongs to the *Solanaceae* family (see Fig. 1.4.1). It has its natural habitat all around the world among moderate climate and humidity conditions. As an endemic Eurasian species [59] it grows under the same conditions as cultivated potato. In middle Europe it often can be seen as weed near the potato cultivation areas. In recent times it has been the object of health related studies [60-62]. It is examined for an anti carcinogenic effect of polyphenols, which are present in this plant. In contrast to traditional prejudice, most parts of this plant like the berries, seem not to be poisonous [59].



**Fig. 1.4.2:** Picture of a flowering *Solanum nigrum* plant (Picture taken by Dr. Heibges, this laboratory).



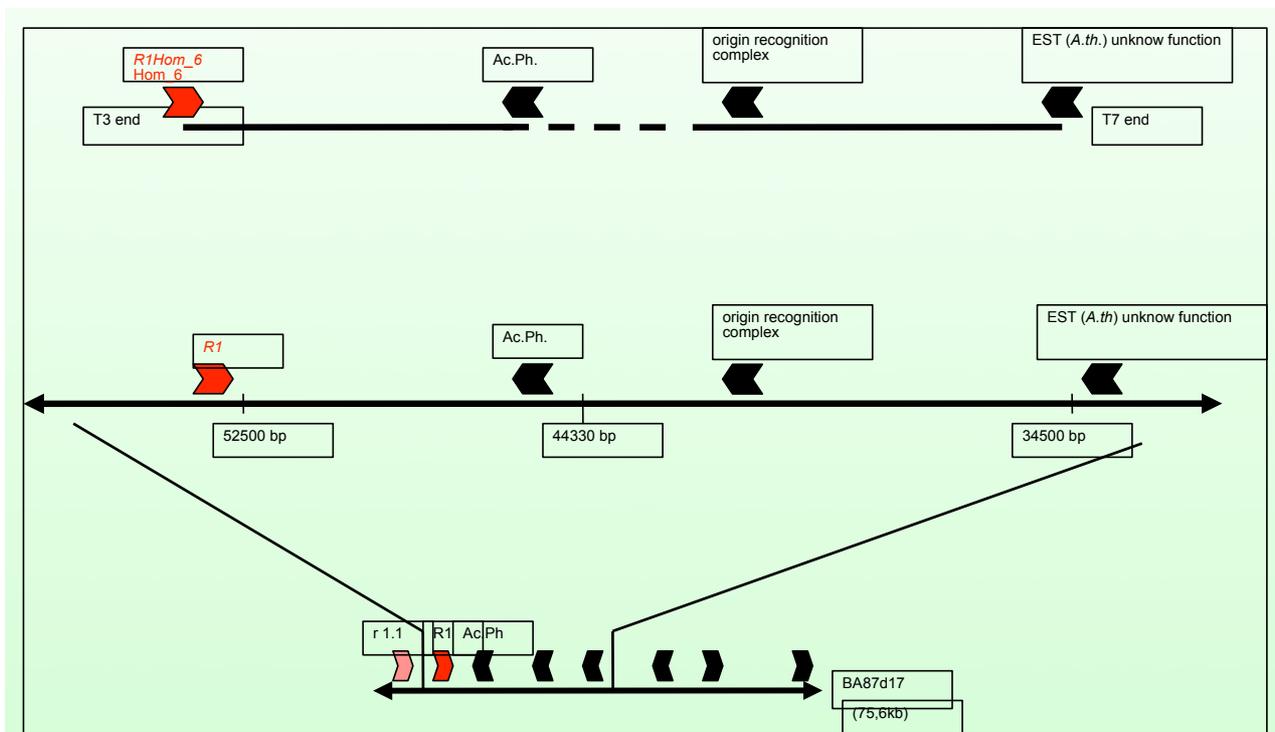
**Fig. 1.4.1:** Taxonomic overview of some members of the Solanaceae family. The taxon *Solanum tuberosum* subsp. *andigena* includes both subspecies *tuberosa* and *andigena* (source: NCBI).

Additionally, mainly due to the work of Ian Baldwin and coworkers (at the MPI for Chemical Ecology), since a few years, this plant is used as a model organism to study plant-herbivore interaction in molecular ecological studies [63, 64]. Although, it has been reported that *Solanum nigrum* can be infected by certain strains of *Phytophthora infestans* [65], the general situation in the field is an incompatible interaction with the pathogen. For potato resistance breeding, *Solanum nigrum* is with its ability to resist

late blight infections to the level of immunity, a very promising plant as genetic source for resistance [66]. Attempts to combine both organisms by somatic hybridization have been

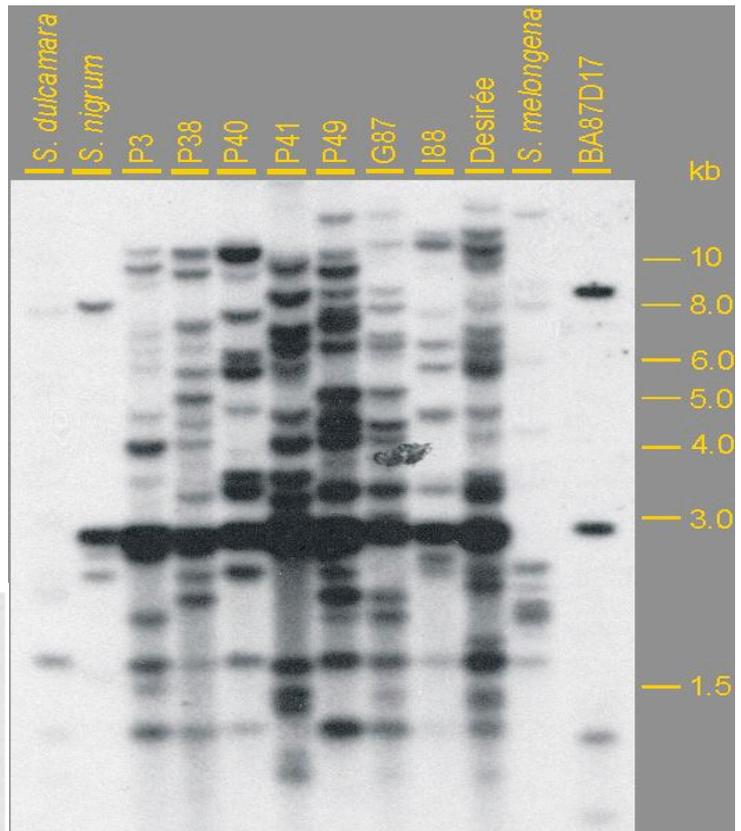
made and showed that it is possible to transfer the resistance capacities of this plants. Nevertheless, this plant also transfers unfavorable characteristics, like the lack of tuber formation. Additionally, somatic hybrids showed a high level of sterility and therefore were useless as pre-breeding material [67, 68].

One important question concerns the underlying mechanism of resistance and at the and the question wether it is a host with a powerful resistance-gene-like recognition system, or wether a PAMP-triggered immune response is the reason for the avoidance of infection. This is not clear yet although genomic sequences with high homology to the *R1* resistance gene of *Solanum tuberosum* have been identified in *S. nigrum* (fig. 1.4.3-1.4.5).

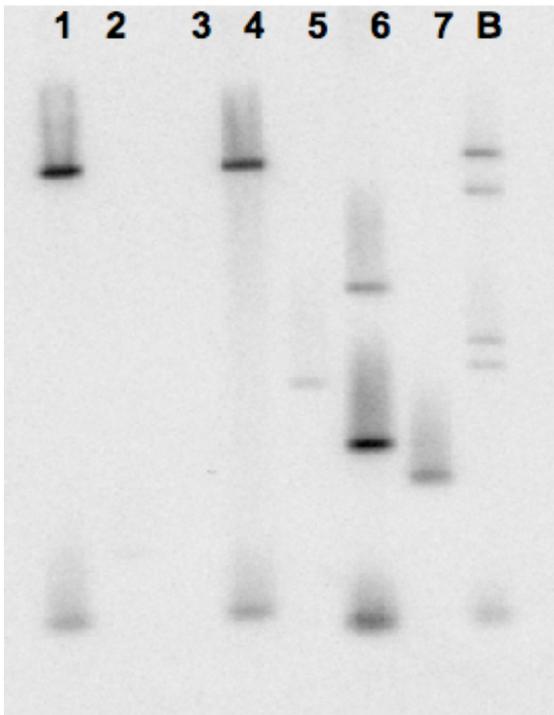


**Fig. 1.4.3:** Identified region of synteny between *Solanum tuberosum* and *Solanum nigrum*. Shown is the organization of a genomic fragment from *Solanum nigrum* (upper) and a corresponding region on BAC BA87d17 from *Solanum tuberosum* (middle and lower) (data from Tatjana von Frey-Jost, this laboratory)

**Fig. 1.4.5:** Results of a Southern Blot after hybridization with a *R1* specific probe. Several cultivars of *Solanum tuberosum* (P3-Desirée) have been used as well as the related species *Solanum dulcamara*, *Solanum nigrum* and *Solanum melongana*. The BAC BA87d17 has been used as control (data from Dr. Heibges, this laboratory).



**Fig. 1.4.4:** Results of a Southern Blot after hybridization with a *R1* specific probe. Seven cosmid clones from a genomic library of *Solanum nigrum* have been used (1-7) and the BAC BA87d17 from *Solanum tuberosum* as control (B) (data from Tatjana von Frey-Jost, this laboratory).



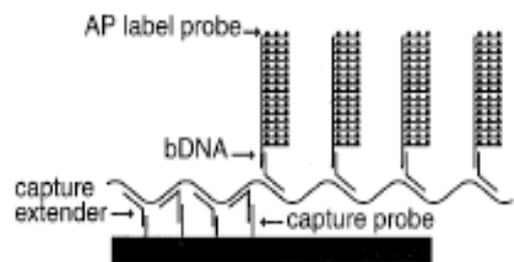
## 1.5.Methods for expression analysis

### 1.5.1.Single target driven methods

The first described method for the analysis of ribonucleic acids has been published in 1977. In this approach, the 1975 published technique of the Southern Blot for the detection of DNA fragments in a complex mixture has been modified to blot RNA onto a nitrocellulose membrane instead of genomic DNA [69, 70]. This new method, known as Northern Blot, was the beginning of expression analysis [70]. The next important step was the ability to synthesize complementary DNA (cDNA) of the RNA [71] in combination with the polymerase chain reaction (PCR) [72].

At present three different methods for the analysis of specific transcripts are known. The method with the broadest use is the reverse transcriptase (RT-)PCR. cDNA is generated and a specific fragment is amplified using a PCR reaction. The signal detection and further analysis is performed using agarose gel electrophoresis. This method has been improved by the quantitative reverse transcriptase (qRT-)PCR, where the amplified fragment detected by a fluorometric measurement during the PCR reaction.

Another method, with broader use in medical applications for detection of retroviruses is the branched DNA or bDNA-assay. This method is mainly used to detect very low amounts of transcripts in blood or tissues but generally can be applied for any kind of nucleic acids. The specific nucleic acid is bound by an immobilized capture probe. Afterwards the signal of the bound molecules is amplified using of bDNA label probes. And since a few years the Northern Blot has found a new application area. This was initiated by the discovery of small RNAs for which the production of cDNA is laborious and far away from being a standard technique. For the analysis of this small molecule class the Northern blot has become the standard technique.



**Fig. 1.5.1.1:** Schematic picture of the principle of a bDNA-assay. AP stand for alkaline phosphatase which is used for the production of a chemoluminescence signal [149].

## 1.5.2.Methods for transcriptome analysis

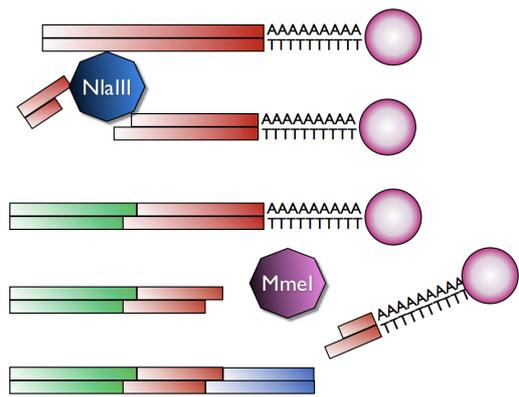
Besides expression analysis for specific target genes, methods have been developed to generate whole transcriptome data. One of the first approaches in this direction the generation of cDNA-libraries and random sequencing of clones to generate EST (expressed sequence tags) information. A first approach in a large scale has been performed in the within the Human Genome Project [73]. In addition, using differential display, a way has been found to compare two libraries [74]. Although being far away from generating quantitative data, this EST-information was the basis for the first whole transcriptome analysis method called cDNA microarray. In this method, the principle of the Northern Blot has been turned around by immobilizing multiple probes on an array and hybridizing this array with first a radioactive and later fluorophoric labeled cDNAs. The first generation was able to detect the presence or absence of a given transcript. The second generation found a way to quantify the fluorometric signal [75-77]. Nevertheless, these array based technologies were limited by the availability of sequence information.

This problem wanted to be solved by the application of an expectation independent method, the first attempt in this direction was to raise the quantitative efficiency of EST-sequencing. This has been done, by the generation of small 3'-sequence tags, which were ligated into concatamers. These concatamers were subsequently cloned and sequenced. This method called serial analysis of gene expression (SAGE) [78], was still limited by the cloning step and the costs of traditional Sanger sequencing [79] and was not able to generate deep transcriptome information in a cost effective manner.

The introduction of „Next Generation Sequencing“ methods (summarized as massive parallel sequencing) is able to multiply the number of sequenced tags per reaction and has opened new possibilities de-novo transcriptome analysis. The major applications of these high throughput sequencing technologies are the sequencing of whole transcriptomes combined with an assembly and an annotation to a given reference genome. The quantification of the amount of a certain transcript is defined by the coverage rate. This method is known as RNAseq or transcriptome sequencing. On the other hand SAGE can be applied in an improved manner. In this method the quantification is done by direct counting of the number of sequenced tags , which are localized on a specific region of the

transcript. These methods are known as Digital gene expression, DeepSAGE, LongSAGE or SuperSAGE [80-86].

DeepSAGE, in particular, works by creating small ditag-libraries of cDNA fragments, which are located close to the 3' end of a transcript. To achieve this, two restriction enzymes are required, which recognize the same restriction site, but cleave different positions. One of the most commonly used enzyme system consists of *NlaIII* (or *DpnII*) and *MmeI*, which both identify a CATG motif. Outgoing from this site, a 18-20 bases large fragment is extracted. Using this principle, it is possible to extract a specific region of a transcript. The advantage of this method compared to whole transcriptome sequencing is that less sequenced molecules are needed to achieve a comparable quantitative information to RNAseq based experiments. The disadvantage is the short length of sequence information which is associated with ambiguities.



**Fig. 1.5.2.1:** Scheme of the generation of SAGE-libraries. Immobilized double strand cDNA is digested with two restriction enzymes using the same CATG restriction site. The final cDNA fragment (red) is flanked by two linkers (blue and green).

## 1.6. Massive parallel sequencing

**Table 1.6.2:** Summary of the performance features of presently available „Next Generation Sequencing“ platforms.\*

System	Provider	Read length	No of sequences per run	Sequencing principle	Template amplification
454 flx	Roche	app. 400 bp	up to $1 \times 10^6$	Pyrosequencing	Emulsion PCR
Genome Analyzer	Illumina	up to 100 bp	up to $200 \times 10^6$	Sanger	Cluster generation
Heliscope	Helicos	25 – 55 bp	$8 - 400 \times 10^6$	Sanger	Not necessary
Solid	Applied Biosystems	50 bp	up to $2 \times 10^9$	Probe ligation	Emulsion PCR

\*The data is based on the specifications of the providers [88-90, 150].

At the moment four different platforms for massive parallel sequencing are available. The most prominent ones are the „Genome Analyzer“ from Illumina, San Diego (USA) (former Solexa) and the 454 FLX system from Roche, Penzberg (Germany) [87]. Quite new is the Solid system from Applied Biosystems, Foster City (USA) and the most recent invented platform is the „Heliscope Sequencer“ from Helios, Cambridge (USA). Interestingly, three of the four different systems follow a different sequencing method. The 454 sequencing technology delivers the longest reads and uses pyrosequencing to generate sequences [88]. The Helicos and Illumina systems use a modification of Sanger sequencing [89]. The Solid system uses a competitive ligation system of fluorescent labeled dibaseprobes as sequencing system. These dibaseprobes are small single strand nucleic acid fragments which consist of two bases at the 3'-end and additional placeholder bases. They are labeled with a fluorescent dye which is specific for the first two bases. The speciality of this system is a two-fold sequence detection of each base, achieving an enhanced sequencing accuracy compared to other methods [90]. The main characteristics are summarized in Table 1.6.2.

### **1.7.Small introduction to the relevant statistical tests for this study**

Whenever large amounts of data have to be analysed like it is the case in transcriptomic experiments, it is necessary to find a way to distinguish between interesting and uninteresting observation. Statistical methods can help to characterize the data mathematically and try to identify the most interesting values in a large amount of data. Especial tests for significance of observation can be a useful tool for the identification of interesting values and a help for the interpretation of results.

One of the most crucial prerequisites for statistical tests is the assessment of the proper probability distribution. One of the most special features of data, generated by Massive parallel sequencing methods is the discrete distribution. Unlike, most empiric data, it does not follow the normal distribution.

During this thesis, five statistical tests, which do not require a normal distribution of the data, have been performed and validated. Additionally some characteristics of Student's t-test are shown.

The G-test statistics [91] uses in principle a series of  $\chi^2$ -distribution tests. In this analysis method are different rules applied. Summarized these rules test the hypotheses if changes

are within a group or between two groups. By the application of separate tests for each hypothesis they have to be rejected or accepted individually to pass the overall test. Usually it is desired that differences between two groups occur and that no differences within a group are detected.

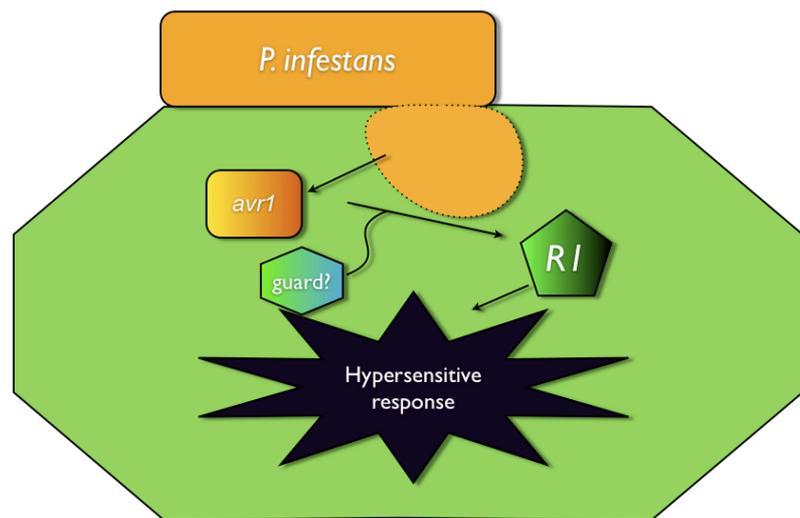
Another implemented test is included in the R-package `Sagenhaft` and is called `sage.test` [92]. This test is a variation of Fisher's Exact test and compares the proportion of each sequence tag in a library to the proportion of the same tag in another library (contingency table statistic). In this test, the original Fisher's exact test has been modified in the way that a negative binomial approach has been used. This was necessary for an application in large data-matrices.

In contrast to the two established tests mentioned above, which were already used for the analysis of traditional SAGE-libraries, `edgeR`, `BaySeq` [93, 94] and a permutation test procedure have been used as well. The first two tests were developed specifically for the analysis of transcriptome data resulting from massive parallel sequencing methods. At the moment, these tests are still in an experimental stage and the permutation test had to be developed specifically.

## 1.8.Motivation for this study

The beginning of the project was the *R1* resistance gene and the genomic region surrounding it that were classified by the ORF-prediction and annotation algorithms as Disease resistance genes. These findings implied the existence of a gene family and showed that the *R1* gene is not unique in *Solanum tuberosum*. This hypothesis was supported by the identification of putative homologous genes in *Solanum nigrum*.

The idea arose that some of these genes have similar functions in pathogen recognition and that among them may be a member, which is involved in the quantitative effect, observed in from the hot spot region on Chromosome V of *Solanum tuberosum*. In *Solanum*



**Fig. 1.8.1:** Model of the mechanism of the function of *R1* at the beginning of this study.

*nigrum*, one of the homologous genes could be involved in the broad resistance of this plant. Additionally, it was necessary to face the truth, that besides the molecular structure, which is originating from the presence of the *R1*-gene, little was known, and just a rough idea about its acting mechanism existed (see fig. 1.8.1).

The prerequisite for further studies on the function of the *R1* family was a further description of the consequences, which result from the presence of this gene. To achieve this, a comparative transcriptome analysis of *R1*-transgenic plants was performed. In this approach, plants were included, which do not harbor the *R1*-gene. This selection has been expanded by including plants which are transformed with the ORF45. The underlying sequence is highly similar but the presence of this open reading frame does not mediate resistance as *R1*.

Another question arose from the predicted open reading frames. Experimental evidence

for expression was lacking. To test the correctness of the in silico data RT-PCR experiments have been performed.

In *Solanum nigrum* an additional question was urgent: Is the presence of a R-gene mediated resistance mechanism the reason for the sensing of *Phytophthora infestans* or is it due to a PAMP-mechanism? To get a hint on this, transient expression of putative effectors of *Phytophthora infestans* has been performed. The aim was the localization of the pathogen recognition event by expressing the effectors directly in the cytosol, using a PVX-based expression system. The identification of necrosis-inducing effectors would be a good argument for the presence of R-genes of the NBS-LRR type class in *Solanum nigrum*, which detects *Phytophthora infestans*. The major aim of this study, was a description of the function of *R1* and the related sequences. A broad data basis was to be generated, which can be used to generate new working hypothesis on the function of *R1* and other members of the *R1* family.

## 2. Material and Methods

### 2.1. Standard conditions for plant and oomycete cultivation

Plants, originating from in vitro culture were grown under longday conditions (16h light (80 mmol photons m<sup>-2</sup>s<sup>-1</sup>), 8h dark) and 21 °C temperature in climate chambers from Brown-Boveri Cie (now: York International), Mannheim (Germany). For race specifications of *Phytophthora infestans* using the standard differential set of *Solanum tuberosum* cultivars from the Scottish Crop Research Institute, the plant material was grown under comparable conditions in the green house.

*Phytophthora infestans* was cultivated on rye agar plates [95], containing small leaves from in vitro culture plants or was propagated on leaflets from susceptible potato cultivars. The climate conditions were 16h light and 8h of darkness and 18°C for short term cultivation (within ongoing infection experiments) or 12°C for long term storage.

### 2.2. Chemicals

Enzymes for RNA-extraction and modification were supplied by Ambion, Austin (USA). Reagents for qPCR-techniques were supplied by Applied Biosystems, Foster City (USA). Reagents for cDNA synthesis were supplied primarily by Fermentas, St. Leon-Roth (Germany) and Invitrogen, Carlsbad (USA). Standard chemicals were supplied by Roth, Karlsruhe (Germany), Sigma Aldrich, St. Louis (USA) and BioBudget, Krefeld (Germany). General enzymes were primarily supplied by Fermentas, St. Leon-Rot (Germany), Invitrogen, Carlsbad (USA), New England Biolabs, Ipswich (USA) and Roche, Penzberg (Germany). All enzymes were used with buffers of the supplier according to the supplier's instructions (if not described differently).

### 2.3. Infection assays and plant cultivation

For the analysis of infected plant material, plants were grown for 6-8 weeks in a climate chamber under standard conditions. At least one week before the beginning of infection experiments, the plants have been transferred to a special chamber. For infections, the *Phytophthora infestans* strain R208m<sup>2</sup> (originating from the laboratory of Felix Mauch, University of Fribourg (Switzerland) and described in Si-Ammour 2003 [96]) was used. During the studies the race specificity of the strain was monitored using a detached leaflet assay and the differential set of potato lines originating from the Scottish crop research institute (SCRI).

The detached leaflet assay was made in 12well-plates, using leaf discs of 2 cm in diameter. The discs were laid on water-soaked Whatman paper-discs (using 300  $\mu$ l of water). Inoculum was produced by washing the sporangia from infected potato leaves (between 8-11 days after infection) using approximately 1-2 ml deionized water for 2-4 infected leaflets (depending on the degree of infection of the used source). As leaf source 6-8 week old plants and the third to fifth leaf were used. The sporangia concentration was quantified using a Neubauer chamber and sporulation of the zoospores was induced by incubating the sporangia suspension at 4-8 °C for 3-6 hours. The climate conditions during infections were 16h day with a temperature of 18 °C and 8 h night with 14 °C in a climate chamber from Ehret, Emmendingen (Germany). The successful sporulation was confirmed optically using a light microscope. Directly before infection experiments, the concentration was adjusted to 30000-60000 sporangia/ml. Infections were done by placing a droplet of 20  $\mu$ l inoculum on the leaf disc. The infection was observed optically with a binocular microscope from the fifth to eighth day post infection. The occurrence of sporangia harboring mycelia on the leaf was the criterion for a positive infection.

For the harvest of material used in chapter 3.1 3.2, 25 plants were grown under conditions described in chapter 2.1 to an age of 10 weeks in case of *Solanum tuberosum* P6/210 and 6-10 in case of *Solanum nigrum* P4. Material was harvested and pooled from 5-15 plants. Harvested material was frozen immediately in liquid nitrogen and stored at -80 °C.

### 2.4. Confocal microscopy

For confocal microscopy, leaves from *Solanum tuberosum* v. Desirée which have been grown for 6-8 weeks under standard conditions as described in chapter 2.1 were used. The infections were made similar to the detached leaflet assay described in chapter 2.3 with following exceptions. Instead of leaf discs, whole leaves were used, which were placed on moist blotting paper on a metal grid within a clear, with parafilm closed plastic box. Infections were made by placing 25-35 droplets of inoculum on each leaflet.

For microscopy small pieces of the infected leaves at three days after infection were stained with a solution of 0,02% diethanol for approximately 10 seconds. Pictures were taken using a confocal laser scanning microscopy (Leica SP2 AOBS, Leica Microsystems, Wetzlar (Germany)). For excitation of the fluorescence laser light with 405 nm wavelength was used. The detection of the fluorescence emission was made using a filter for wavelength of 420-520 nm.

Confocal microscopy pictures were taken three days after infection under aid of Dr. Elmon Schmelzer and the Cemic group at the Max-Planck-Institute for Plant Breeding Research.

## **2.5.Molecular biological methods**

### **2.5.1.Polymerase chain reaction (PCR)**

The standard reaction mixture included following components:

1  $\mu$ l Template-DNA (50 ng genomic DNA, 1 ng BAC-DNA)  
PCR-buffer Amplikon III  
2,5 mM  $MgCl_2$   
200  $\mu$ M dNTP-Mix  
0,25  $\mu$ M Forward-Primer  
0,25  $\mu$ M Reverse-Primer  
1 u recombinant Taq DNA-Polymerase  
+H<sub>2</sub>O to final volume of 20 $\mu$ l

Standard PCR-program included following steps:

2 minutes 94 °C  
 30 seconds 93 °C  
 30 seconds T<sub>A</sub> x35  
 30 second 72 °C  
 5 minutes 72°C

### Linear-PCR for the amplification of ORF23

The polymerase chain reaction in case of ORF23 was done under standard conditions with the exception that the reverse primer was added to the reaction mixture after 15 cycles of amplification.

**Table 2.5.1.1:** Overview of used oligonucleotides for polymerase chain reaction

Name	Sequence (5'-3')	Annealing temperature [°C]	Product size (bp)	Application
TC173455_f	TGCTGGTGAACCCACAAAGCCC	55	306	qRT-PCR
TC173455_r	TCCGCCTTGGTCAACCCCTGC			
TC180208_f	TGCGCAAGGGTTTGGCCTGT	55	337	qRT-PCR
TC180208_r	ACCACCGCCTCCTCCGGTTF			
TC173049_f	GCTCATGGCGGGGAAGGAGG	55	140	qRT-PCR
TC173049_r	CGCGCCGATGGCGAGTAAGT			
TC163043_f	TGTGGAGGCGAGCTCTGGTGT	55	160	qRT-PCR
TC163043_r	GACCTCCAGTTCCGCCGCTG			
TC173953_f	GGCATTGGGGTTACATGGATGGTCC	55	200	qRT-PCR
TC173953_r	TGCCATCTCACTTGTGGATTGCGCC			
TC165331_f	TGGCTGCAGCAGTACGGAACA	55	240	qRT-PCR
TC165331_r	GCTCCACCGATGCAGGACCC			
TC172861_f	TGATTGGCGTGCCAACCCCT	55	144	qRT-PCR
TC172861_r	CCCCCACCTGCAGACCGAGT			
TC168403_f	TGGGTGGCGAGGAAAAGCGG	55	173	qRT-PCR
TC168403_r	AGCAGGGTAAGAGAGTGGGGGT			
TC182394_f	TGGTCTCCGTCTCCGTGGTG	55	378	qRT-PCR
TC182394_r	CCTCTGCTGGTCCGGTGGGA			
TC170569_f	ACCGCAGCAGGTCCATGCAA	55	166	qRT-PCR
TC170569_r	CCCTTTTCGTCTTTCCGCTCACCG			
TC190958_f	CAGCTGAGCAGGGACGGCAG	55	116	qRT-PCR

TC190958_r	GGAGCAGCATCAACGGGTGC	55	140	
TC174692_f	GGAGCGACAGCGATTTGTACGTG	55	200	qRT-PCR
TC174692_r	GCTTGAACCTTCCCCTCGCGT			
TC173012_f	GTACGGCCTTCTGCACCGCT	55	249	qRT-PCR
TC173012_r	GTGACCTTTGCGCCCGCTCT			
TC176096_f	CTGTAGTGGCTGCCCGTGCC	55	169	qRT-PCR
TC176096_r	ACGCGTGCTGCTTTTCCCAT			
TC190958_f	CAGCTGAGCAGGGACGGCAG	55	146	qRT-PCR
TC190958_r	GGAGCAGCATCAACGGGTGC			
TC184597_f	GCTGGACCGCTCAAATGCTGC	55	212	qRT-PCR
TC184597_r	CGGCAAAGAGGGCGAGAAGC			
TC183138_f	AGTGTGCGATGCTGAACTGGTGGGA	55	191	qRT-PCR
TC183138_r	TGCCTTCCCGCTGTCAAATCCT			
st22-1_f	TAGGGATCAGATCAGTACC	52	162	RT-PCR
st22-1_r	GAAATTGAAAAAGCAGGAGAG			
st23_f	GTTCTAAAACCATAAAATGGTACG	55	348	RT-PCR
st23_r	GCTACATTTGTTTGAAACAAAGC			
st24/45_f	CTCAAGAATCAACTTCAAGTTG	55	314	RT-PCR
st24/45_r	CATCTTCAAACCCAACAATTTTC			
st46_f	CGAATGGGGTAGTTACATGC	52	119	RT-PCR
st46_r	GCATGTTTGGTTCATCAGTG			
st52_f	CCAAGCATGGCCCCCTATG	55	278	RT-PCR
st52_r	CCCTGTTTCTTGGAACAAAAAG			
st54_f	CTGTATCTTGAAAAAGTTTGGG	55	586	RT-PCR
st54_r	ATGCTTCTTTTGAGCTTG			
snR1.4_f	ACTGCATCGATGTCAAGATCT	57	320	RT-PCR (developed by Tatjana von Frey- Jost, this laboratory)
snR1.4_r	CCCTTGAGCGTTAAAGATAAC			
snR1.5_f	AGACCTAGATTCTCTACTGAAGC	55	410	RT-PCR (developed by Tatjana von Frey- Jost, this laboratory)
SnR1.5_r	CTTTTTTCTCCGTCTTTGCT			

SnR1.6_f	TTATCCTCACATGCTTTTGC	57	400	RT-PCR (developed by Tatjana von Frey- Jost, this laboratory)
SnR1.6_r	CTCTAAGCACTCGATTTC			
SnR1.7_f	TCTTCAATTGATCAAGCTTCC	55	390	RT-PCR (developed by Tatjana von Frey- Jost, this laboratory)
SnR1.7_r	CTCCGAGTACCCATAAACATG			
0-83 (Pinf-F)	GAAAGGCATAGAAGGTAGA	50	258	Phytophthora infestans detection [145]
0-84 (Pinf-R)	TAACCGACCAAGTAGTAAA			
Tub_for	ACGTATCAATGTTTATTTCAATGA	52	525 (cDNA)	Amplification of Tubulin.
Tub_rev	ATATCATATAGAGCTTCGTTGTCA			
ef-1 $\alpha$ _for	ATGTTTCAGGCGCAAGGTT	60	101	Amplification of <i>ef-1<math>\alpha</math></i> . RT-PCR, qRT-PCR [144]
ef-1 $\alpha$ _rev	TCTGCAACCGGGTCATTCAT			

### 2.5.2. Nucleic acid extraction

RNA was extracted with the TotallyRNA-kit from Ambion, Austin (USA) according to the suppliers instructions.

Genomic DNA has been extracted by adding the same volume of 10 mM TRIS-buffer pH 8.0 to the acidic phenol phase from the RNA-extraction, followed by vortexing and 10 minutes incubation at room temperature. Afterwards, the phenolic and the aqueous phase were separated by 3 minutes centrifugation at maximum velocity. The aqueous phase was then transferred to a fresh tube and residuals of RNA were digested with 1  $\mu$ l RnaseA with an incubation time of 15 minutes at room temperature. Subsequent precipitation was performed by adding two volumes of isopropanol and incubating for at least 30 minutes at -20 °C.

Alternatively genomic DNA was extracted by using the BioSprint 96 platform from Qiagen, Hilden (Germany) or other suitable commercial DNA-extraction solutions.

Quality assessment and quantifications were made by using spectrophotometric measurements on the NanoDrop system of Thermo Scientific, Waltham (USA) and by

separating and visualizing the samples using Agarose gel electrophoresis on a 1,5 % agarose gel for RNA samples and a 1 % agarose gel for DNA samples. Nucleic acids were detected by using ethidium bromide as staining reagent. For the use of qRT-PCR and prior to the creation of ditag-libraries for DeepSAGE, the quantification was made additionally by fluorometric measurements, using the Qubit-system of Invitrogen, Carlsbad (USA). For subsequent PCR-applications traces of DNA were removed by using the TurboDNFree-Kit from Ambion, Austin (USA) according to the suppliers instructions with the exception that for subsequent application of qRT-PCR the amount of enzyme was doubled.

### **2.5.3. First strand cDNA synthesis**

For first strand cDNA synthesis the RevertAid H Minus First strand cDNA synthesis kit from Fermentas, St. Leon-Roth (Germany) was used by following the suppliers instructions with following exceptions:

The synthesis temperature was raised to 45 °C instead of 42 °C in order to increase the synthesis rate.

For subsequent amplification of *R1* or *R1*-homologous genes from the cDNA RnaseH digestion has been performed by the addition of 5 u RnaseH to the reaction mixture after the cDNA synthesis and incubating 20 minutes at 37 °C. The reaction was terminated by 10 minutes incubation at 65 °C.

For qRT-PCR two separate reactions were made, one using a poly(dT)-primer and one using Random Hexamer primers. Afterwards, these samples were pooled and diluted twofold. From this diluted pool, 1  $\mu$ l was used for amplification reactions. For the amplification of *R1* or *R1*-homologues genes, only the poly(dT)-primer was used and the reaction mixture was not diluted.

The quality of the cDNA was tested by amplifying tubulin (table 2.5.1.1) in a PCR-reaction. The used primers span an intron leading to a larger amplicon after amplification of genomic DNA. For this characteristic, the PCR was used as control for the absence of contaminating genomic DNA in the sample as well.

### **2.5.4. Quantitative polymerase chain reaction (qRT-PCR)**

qRT-PCR-experiments were performed on a Mastercycler ep realplex platform from Eppendorf, Hamburg (Germany). Expression quantification and calculation of Ct-values was performed, using the integrated Realplex analysis software. An assay included three biological replicates and two technical replicates, Differences of more than 10% in the technical replicates were sorted out as failed values. For the generation of the standard curve, a pool of all used cDNAs was used in dilutions of 1:1, 1:10 and 1:100.

Assays, with Primer efficiencies below 80% were sorted out as improper.

For the reaction mixture, Power SYBR Green PCR Master Mix from Applied Biosystems , Foster City (USA) was used.

The assessment of the specificity of the PCR-reactions was done by a subsequent melting curve analysis.

Samples with Ct-values above 35 were discarded and not used for further analysis.

A reaction consisted out of following components:

1 $\mu$ l	template DNA
2,5 $\mu$ l	Forward Primer (10 $\mu$ M)
2,5 $\mu$ l	Reverse Primer (10 $\mu$ M)
12,5 $\mu$ l	2x Power SYBR GreenPCR Master Mix
6,5 $\mu$ l	High pure H <sub>2</sub> O
$V_{\text{final}}=25 \mu\text{l}$	

Standard PCR Program was like following:

2 minutes	95 °C	
30 seconds	95 °C	
30 seconds	T <sub>A</sub>	x40
30 second	68 °C	
15 seconds	95°C	
15 seconds	60°C	
20 minute ramp	60°C-95°C (melting curve analysis)	

### **2.5.1. Transient expression of effector proteins in *Solanum nigrum***

For the expression of effectors of *Phytophthora infestans* in *Solanum nigrum* a set of cloned effectors in a PVX-based expression system was used. This set originated from the

laboratory of Dr. Sophien Kamoun, Sainsbury Laboratory (UK) and uses the vector pGR106, which is transformed to *Agrobacterium tumefaciens* LB4404. This set has been described in Oh et al 2009 [97]. For the transient expression, *Solanum nigrum* plants were grown to an age between 4-5 weeks in a climate chamber under conditions described in chapter 2.1. For the local transformation of plant leaves two methods were used. The first was a toothpick inoculation method. For this, the *Agrobacteria* containing YEB-Medium plates were incubated over night at 28 °C. At the following day small wounds were made to the *Solanum nigrum* leaves and a small amount of bacterial material was placed on the wound. 8-11 days after the inoculation, the area around the wound was examined for necrotic symptoms, which exceeded those from the negative control. For this negative control a  $\Delta gfp$ -construct in the same vector was used.

Plants, which were scored positive, were verified by the infiltration of *Agrobacterium* suspension.

For this, *Agrobacteria* were grown for 1-2 days in liquid LB or YEP-medium at 28 °C. Before infiltration the cells were harvested by centrifuging 10-15 minutes at 4000 rpm. The supernatant was decanted and the bacterial pellet was resuspended in „Infiltration solution“ (10 mM MgCl<sub>2</sub>, 1mM MES, 100  $\mu$ M Acetosyringone) to an optical density of 1 (OD<sub>600</sub>=1,0). A small amount of this bacterial suspension was infiltrated to the plants leaves with a 1 ml syringe without tip. This was assessed by a wet area inside the leaf. Afterwards the plants were transferred back to the climate chamber. Eight to eleven days after the infiltration the plants were examined for the occurrence of necrotic symptoms at the infiltrated area.

## 2.6. DeepSAGE

### 2.6.1. Biological material and generation of infected tissue

For the DeepSAGE transcriptome analysis, five different lines originating from the cultivar Desirée were used:

<i>Solanum tuberosum</i> cv. Desirée 10-23/2	containing <i>R1</i> (genomic) [54]
<i>Solanum tuberosum</i> cv. Desirée 10-2/4	containing <i>R1</i> (genomic) [54]
<i>Solanum tuberosum</i> cv. Desirée P4H5K3S2	containing the ORF45 (genomic)
<i>Solanum tuberosum</i> cv. Desirée	wild type
<i>Solanum tuberosum</i> cv. Desirée LV41	empty vector control

The transgenic lines 10-23/2 and 10-2/4 have been shown to be single locus transformants by crossing experiments [98]. In the case of the ORF45 transgenic plant P4H5K3S2 the number of loci is not known. In addition, untransformed plants of the cultivar Desirée were used and the line Desirée LV41, which is transformed with an empty vector. This line was produced during the complementation analysis of *R1* and is transformed with the empty vector pclD04541 [99].

For late blight infection, the *Phytophthora infestans* strain R208m<sup>2</sup> was used [96]. This strain is a race 4 and originally has been transformed with GFP, although this ability has been lost throughout the cultivation.

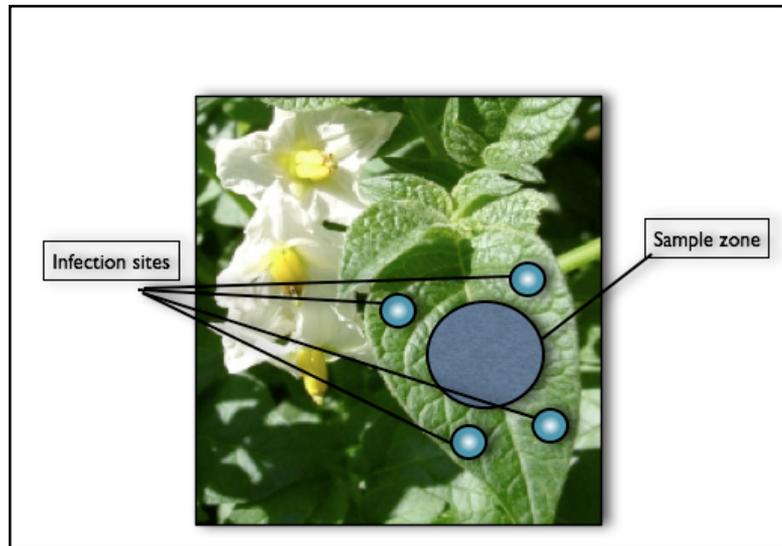
The characteristic of harboring only *vir4* was able to induce the *R1* specific hypersensitive response. It was possible to harvest tissue during the incompatible interaction from lines containing *R1* and material during the compatible interaction from the other lines.

Plants of these lines were propagated in tissue culture and were grown to an age of 5-7 weeks as described in chapter 2.1. At least one week before starting an infection experiment, the plants were transferred to a climate chamber from Vötsch Industrietechnik, Balingen (Germany) with similar light and the same day length conditions. Differing from standard conditions, the temperature was adjusted to 18°C during daytime and to 14°C during night. This was necessary to create favorable infection conditions.

For infections, the inoculum was produced like described in chapter 2.3 with the exception that additionally infected rye agar plates (containing leaves from in vitro plants) were used

as source. Directly before infection experiments, the concentration was adjusted to 30000 sporangia/ml.

Infections were made by infecting 4 leaves with four drops of inoculum of 12,5  $\mu$ l at the corners of each leaf (fig. 2.6.1.1). Infection- and harvesting times were at late afternoon to early evening (approximately between 16:30 - 20:00). Infections started from the third leaf from the top downwards. In total the third to the sixth leaf was infected (in exceptions, depending on the growth shape of the plant, the second to fifth leaf was infected).



**Fig. 2.6.1.1:** Overview over the infected positions and the area of sample harvesting for the transcriptome analysis. The blue dots represent the place of inoculum. The dark area indicates the harvested area.

During the infection the plants were covered with a clear plastic bag to standardize the humidity conditions. As material for further transcriptome analysis, the third to fifth leaves were harvested at zero, one and three days post infection. For this from the central area (between the four infection points) a leaf disc from 2 cm diameter was punched out. These three discs were pooled to one sample and frozen in liquid nitrogen immediately. The exclusive use of the middle area was chosen to avoid *Phytophthora infestans* content inside the sample material. Three samples in one timeline, but from the same batch of plants and with the same inoculum were declared as one independent experiment. For the generation of 0 days post infection samples - the leaves were infected and harvested directly after - with an estimated contact time of one to three minutes. The fourth infected leaf remained at the plant and was harvested after eight days as a control for a successful infection. The content of *Phytophthora infestans* in the fourth leaf was verified by performing a semiquantitative PCR on the genomic DNA of the sample using the Primers 0-83 and 0-84 for the detection of *Phytophthora infestans* and the primers ef-1 $\alpha$ \_f and ef-1 $\alpha$ \_r for the amplification of elongation factor 1 $\alpha$  as standard (table 2.5.1.1).

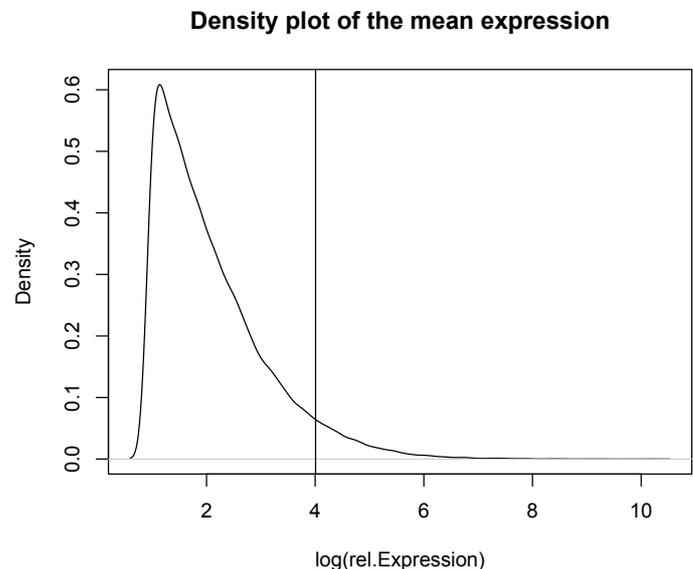
## 2.6.2. Creation of 3'-tag-libraries

The creation of 3'-tag libraries and the sequencing has been made by the group of Kåre Lehmann Nielsen at the department of Biotechnology, Chemistry and Environmental Engineering at Aalborg University (Denmark). A detailed protocol for generation of 3'-tag libraries is given in the Appendix.

## 2.6.3. Bioinformatical methods for the analysis of transcriptome data

### 2.6.3.1. Creation of the Standard data set

For the generation of a standard data set, a number of scripts written in the programming language Perl were applied. In a first step, singleton sequences were removed in order to exclude possible sequencing and pcr based errors. This was achieved by applying the script "CutOffLibsV3.pl" and defining a cutoff value to 2. Afterwards the data of all libraries was combined to one matrix using the script "compareSageV9.pl". Then the tag-counts were transformed to the relative unit counts/million by using the script "NormaliseTagTable.pl". With the application of this script, the final cutoff was defined, which was a minimum detection in three samples based on the presence of three biological replicates or a summed up minimum expression level of 110 counts/million which is derived from the average expression density distribution. The basis is the assumption of a least detection in two samples with a relative expression of



**Fig. 2.6.3.1.1:** Density Plot of the logarithm of the average expression levels of all samples. The additional line indicates the border of the cutoff value of 55 count/million which is the lowest accepted expression value if a tag was observed less than three times.

55 counts/million. This level exceeds the majority (Density 10 %) of the detected expression levels and was not ignored - even if not frequently detected (fig. 2.6.3.1.1). In parallel a combined multiple fastafile was created by using the Potato gene index available from the Dana Faber Institute [100-102] and the available genomic sequence of *Phytophthora infestans* from the Broad institute [5]. This combined multiple fasta file was used for the tag identification under implementation of the script "GlobalSageMap-V14.pl". Considering the high DNA polymorphism of *Solanum tuberosum*, one nucleotide mismatch was allowed for a successful identification. The perl-scripts for the generation for the data matrix were developed by Mads Sønderkær from the group of Prof. K. Lehmann Nielsen at the University of Aalborg (Denmark).

### 2.6.3.2. General procedures and sample composition for statistics

For the subsequent calculations all values, which belong to the same phenotypic class, whether containing the *R1* gene (lines 10-2/4 and 10-23/2), containing the ORF45 (line P4H5K3S2) or having no introduced gene (Desirée wild type and LV41) at a certain timepoint (0 dpi, 1 dpi or 3 dpi) were declared as one replica group. Additionally for the calculation of Fisher's exact test and the logfoldchange values, it was necessary to work with single values. In these cases, time series as they belong to the time series of independent experiments were used. In this experiment, all harvested timepoints in a testing row originated of the same batch of plants and inoculum and for this were declared as independent experiment. This is summarized in table 2.6.3.2.1.

In test statistics, where the original tag counts were necessary, the relative tag count values were calculated back by reconstructing the factor which was used to transform the data to the relative unit „count/million“ from the internal Stat-files and applying it backwards. This was necessary due to the large size of the data matrix which is available before the transformation into the relative unit and applying the final cutoff criteria. Calculations with this matrix exceed the capacities of most standard computer systems. The original tag counts were necessary for the calculation of the G-test, sage.test and BaySeq.

The Permutation test, edgeR and Student's t-test were done by using relative tag counts. For all testings and calculations, values of 0 were replaced by 1 except for the calculation of G-test, which uses an internal replacement factor [91].

After test calculations all p-values were corrected for the False discovery rate (FDR) after the method of Benjamini Hochberg [103]. This was done by implementing the command `p.adjust` of the Software R (on the raw p-values with at least 20 decimal places).

**Table 2.6.3.2.1:** Definition of the samples, the replica groups and the independent experiments\*

Sample	0 dpi	1 dpi	3 dpi	
DesiR1(10-2/4)_Rep1				-> independend experiment
DesiR1(10-2/4)_Rep2				-> independend experiment
DesiR1(10-2/4)_Rep3				-> independend experiment
DesiR1(10-23/2)_Rep1				-> independend experiment
DesiR1(10-23/2)_Rep2				-> independend experiment
DesiR1(10-23/2)_Rep3				-> independend experiment
Desir1(wt)_Rep1				-> independend experiment
Desir1(wt)_Rep2				-> independend experiment
Desir1(wt)_Rep3				-> independend experiment
Desir1(LV41)_Rep1				-> independend experiment
Desir1(LV41)_Rep2				-> independend experiment
Desir1(LV41)_Rep3				-> independend experiment
DesiORF45(P4H5K3S2)_Rep1				-> independend experiment
DesiORF45(P4H5K3S2)_Rep2				-> independend experiment
DesiORF45(P4H5K3S2)_Rep3				-> independend experiment

\*Each colored box represents one sample. The term „Sample“ indicates the sample nomenclature used in this study (expanded by the timepoint of infection (0dpi, 1dpi, 3dpi); the timepoint of infection is indicated by the column 0 dpi, 1 dpi, 3 dpi. Each row represents the samples of an independent experiment. Each group of the same color indicates a replica group. The green samples contain the *R1*-gene, the red samples are untransformed, The yellow samples contain ORF45. The black box represents the sample DesiORF45(P4H5K3S2)\_3dpi\_Rep3, for which the sequencing reaction failed.

For higher order plots, like the generation of Venn-diagrams, box-plots or plots and histograms, which underly a great amount of data the software R was used. For low level plots

and the calculations of Student's t-test (two-sided paired) the softwares NeoOffice Calc or lworks Numbers were used.

For the principal component analysis the R command princomp under default settings was used and the results from principal component 1 to six were extracted.

The calculation of the correlation coefficients in chapter 3.4.4 was made on the basis of  $D_r$  values of significant tags as they are described in the chapters 2.6.3.3-7. For the implementation the Software R in combination with the package „Rcommander“ was used. For the calculation the pearson correlation coefficient was applied.

### 2.6.3.3.G-test

The application of G-test was done using the Software G-test [91] and Microsoft Excel. This test evaluates significance within replica groups and between replica groups separately. Only tags showing no significance within a replica group ( $FDR \geq 0,05$ ) and significance between both tested replica groups ( $FDR \leq 0,05$ ) were accepted as significantly different.

### 2.6.3.4.Fisher's exact test (sage.test)

For the calculations of p-values after a modified version of Fisher's exact test through the command „sage.test“ the software R in combination with the package „Sagenhaft“ [92] was used. The test was done for each experiment as they belong to independent experiments (table 2.6.3.2.1) separately and the FDR-corrected p-value was connected with the criterion of a least twofold expression change (corresponds to a logfoldchange of 0,3). The number of comparisons in a replica group which were assessed as truly significant ( $FDR \leq 0,05$ ;  $0,03 \logfoldchange -0,3$ ) were counted. Those tags, which were evaluated as true significant in at least four of six cases in the wildtype and *R1*-transgenic lines and in at least two of three or two of two cases (at 3dpi) in the ORF45 transgenic line were declared as significantly regulated.

### 2.6.3.5.edgeR

For implementation of edgeR the Software R and the corresponding package „edgeR“ was used. The calculations were done by Dr. Ulrike Göbel from the Bioinformatics group of Dr. Heiko Schoof at the Max-Planck-Institute for Plant Breeding Research. Tags with  $FDR \leq 0,05$  were accepted as significant.

### 2.6.3.6.BaySeq

For implementation of BaySeq the Software R and the corresponding package „BaySeq“ was used. The calculations were done by Dr. Ulrike Göbel from the Bioinformatics group of Dr. Heiko Schoof at the Max-Planck-Institute for Plant Breeding Research. As significance output, BaySeq tests directly for the presence of two different groups and generates logP values as degree of significance. A  $\log P \geq -0,02$  was accepted as significant (corresponding to the criterion  $p \leq 0,05$ ).

### 2.6.3.7.Permu

The computation of the permutation test Permu was done using the Software R. For this test no package was available and it was developed by Dr. Ulrike Göbel from the Bioinformatics group of Dr. Heiko Schoof at the Max-Planck-Institute for Plant Breeding Research. This test begins with a ranking of the tags in most homogeneous groups and a comparison to the comparing condition. Afterwards a p-value is allocated by the test depending on the probability distribution of the imputed data. In the case of the this test, it turned out, that the expression of significance of this test is just weakly comparable to tests. In total just 462 different p-values were generated, leading to an overcorrection for the False discovery rate. This was adjusted, by dividing the p-value through the number of samples. This leads to the acceptance of a FDR-corrected p-value of smaller than 0,3 corresponding to raw-p (without correction)  $\leq 0,0065$ .

### 2.6.3.8. Calculation of the $D_f$ and logfoldchange values

The logfoldchange values were calculated in comparison to the 0dpi value after the equation  $lfc = \log(t_1/t_0)$ ; in this equation lfc stands for logfoldchange value,  $t_1$  for a tested time-point (1dpi or 3 dpi) and  $t_0$  for the 0 dpi value. Primary single independent experiments were calculated like defined in chapter 2.6.3.2. The calculation of the Average logfold-change values was done by calculating the arithmetic mean from the logfoldchange values of the independent experiments.

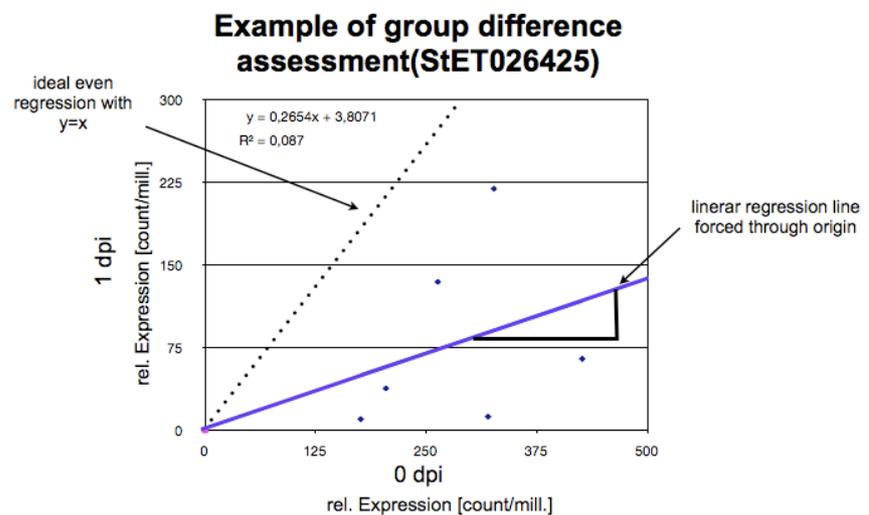
Alternatively, a second assessment of the group differences was made by calculating the linear regressions of the coordinates of both groups according to their behavior in independent experiments.

A correction for probable lower amounts of differences in higher values and for negative differences was made via the forced regression through the origin. In order to achieve the comparability of all calculated values, the data was converted to absolute logarithmic values (fig. 2.6.3.8.1 for illustration).

If the slope of the origin

forced linear regression is defined as  $r_{f0}$  the differentiation factor is defined as:

$$D_f = |\log(r_{f0})|$$



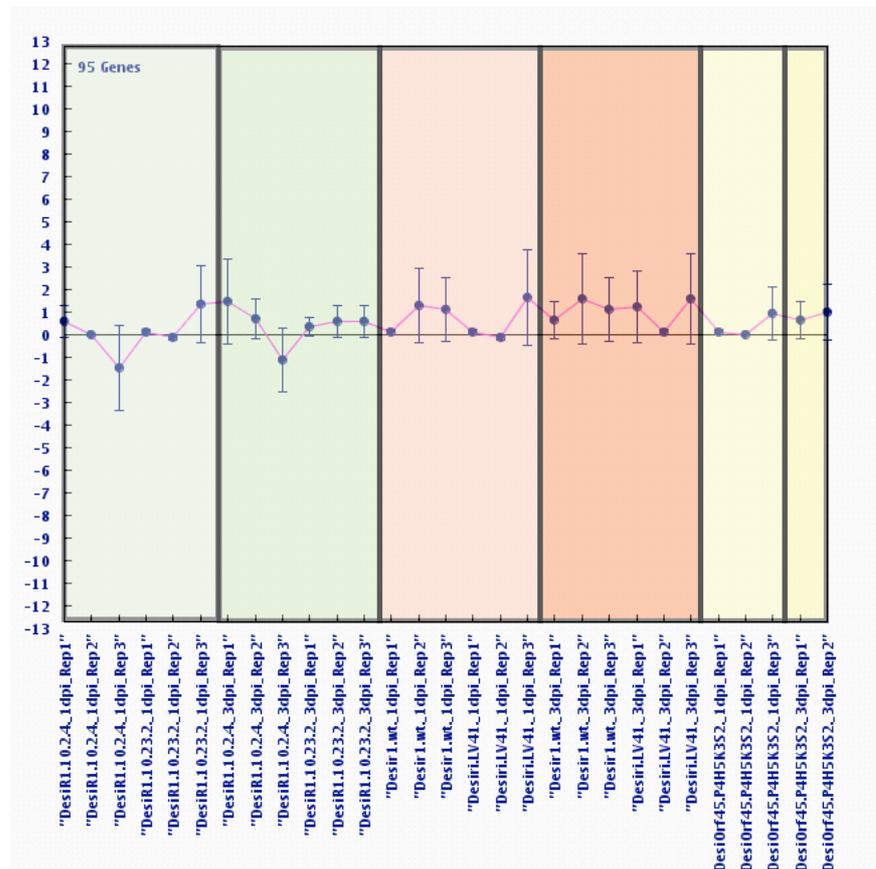
**Fig. 2.6.3.8.1:** Example of the theory behind the alternative evaluation of differences in replica groups. The values are expressed as coordinate corresponding to their relative expression levels. The slope of the linear regression expresses the differences between both groups.

### 2.6.3.9. Cluster analysis

Hierarchical clustering and k-mean clustering was performed based on the logfoldchange values using the web-based services CarmaWeb and GenesisWeb (available under <https://carmaweb.genome.tugraz.at/carma>). For the k-mean clustering a subset of the standard data set was used. This subset contained only those tags which were annotated with grade A or grade B and contained 10323 tags. For the analysis  $6^3$  clusters have been calculated. For clustering methods the pearson correlation coefficient was used.

The extraction of relevant clusters in chapter 3.4.5 was made in two rounds. In a first attempt, the selection was made by eye and all Clusters with a diffuse pattern and no tendency were sorted out.

After this, the clusters were examined by mean expression values. These values were integrated in the analysis and reflect the mean characteristics of all examined tags in a certain sample (fig. 2.6.3.9.1). From this values, those clusters were assessed as interesting, which showed at least in 4 of six values at one replica group (a certain time point in a certain genotype) the same tendency, meaning a positive logfoldchange value or a negative. In cases of ORF45 with three



**Fig. 2.6.3.9.1:** Example of the centroid expression view of the cluster analysis. Each replicate group is indicated by a different color. In this example, the groups of wt and ORF45 at three days after infection would have been assessed as upregulated.

or two values in a replica group, two of three had to show the same tendency at day after infection and for the three days after infection, both values had to show the same tendency

### 2.6.3.10.Go-term analysis

The GO-Term analysis was made using the Software Cytoscape v2.6.3 and the extension BinGO v2.3. As annotation file, the annotation information, available from the Dana faber institute (DFCI) was transformed to a readable format (this was done by Rico Basekow and Maren Imhoff from the group of Dr. Kersten at the MPIMP, Golm). For the assessment of significance of overrepresentation, the included hypergeometric test has been performed and the valid significance level was set to  $p \leq 0,05$  after FDR-correction using the Benjamini Hochberg method. As input for the GO-analysis, tags have been used, which showed a significant change after analysis with the R-package „Sagenhaft“ like defined in chapter 2.6.3.4.

#### 2.6.3.10.1.Summary of used Softwares

A comprehension of the used softwares during this project is given in table 2.6.3.11.1.

**Table 2.6.3.11.1:** Overview of the used softwares during this project

Name	Application	Kind of Software	Cut-off values
CutOffLibsV3.pl	Singleton removal	Perl-script	2
compareSageV9.pl	Matrix-creation	Perl-script	-
NormaliseTagTable.pl	raw counts to relative counts transformation	Perl-script	110 minimum total expression 3 number of detections
GlobalSageMap-V14.pl	Annotation of tags	Perl-script	-
G-test	G-test statistic	Excel communicating program	FDR $\leq$ 0,05*
Microsoft Excel	Basic statistics	Stand-alone Software	-
Sagenhaft	sage.test statistic	R package	FDR $\leq$ 0,05; **- 0,31fco,3
R base	FDR-correction of p-values (p.adjust)	R-package	-
R stats	Principal component analysis (princomp)	R-package	-
edgeR	significance statistic	R-package	FDR $\leq$ 0,05
BaySeq	significance statistic	R-package	logP $\geq$ -0,02
gplots	Venn-diagrams	R-package	-
Rcommander	Calculation of correlation matrix	R-package	-
CarmaWeb	Cluster analysis	Web-source	-
GenesisWeb	Cluster analysis	Web-source	-
Permu	significance statistic	R computed	FDR $\leq$ 0,05***

This table shows the name of the software, the application are, type and Cutoffs.

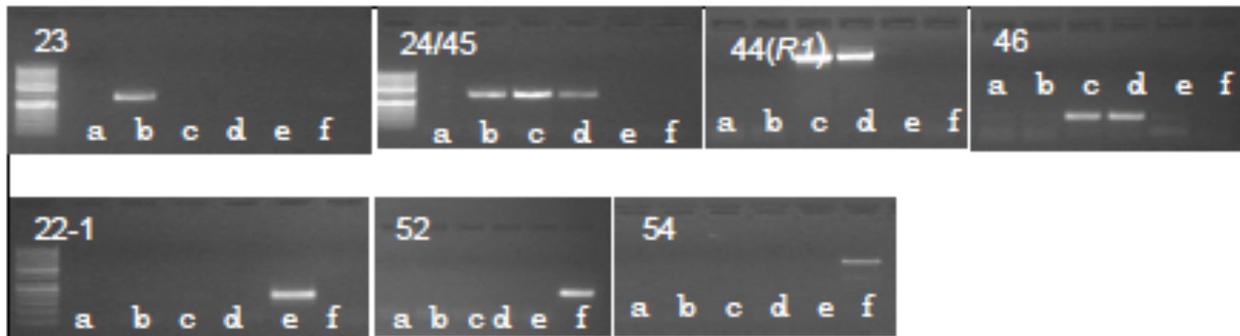
\* in case of G-test, homogeneity and differences were tested separately, for the assessment of homogeneity FDR $\geq$ 0,05 was applied

\*\* For Fisher's exact test each independent experiments was tested separately and true significance was connected with multiple significant observation (chapter2.6.3.4)

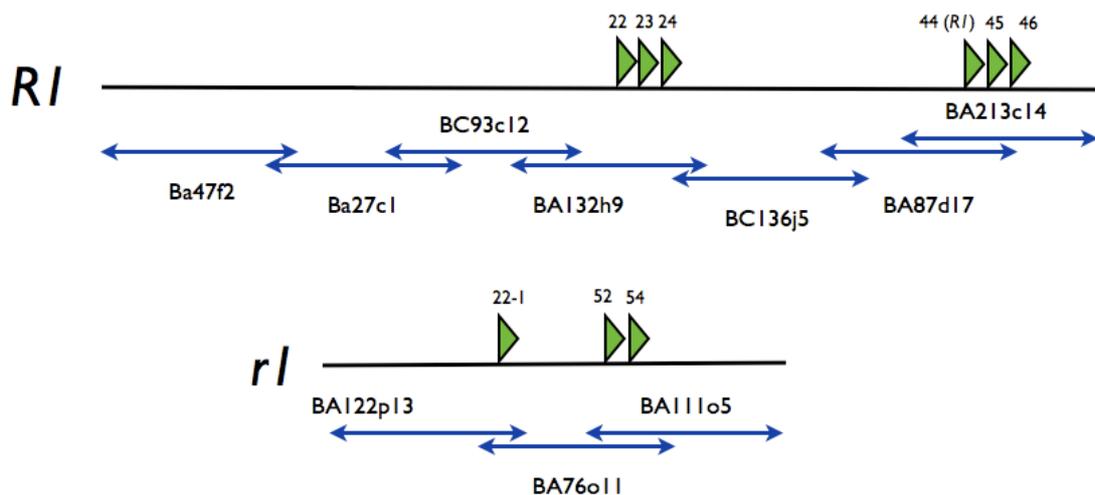
\*\*\* FDR was corrected by a group size factor as described in chapter 2.6.3.7

### 3.Results

#### 3.1.Expression analysis of *R1*-homologous genes of *Solanum tuberosum*

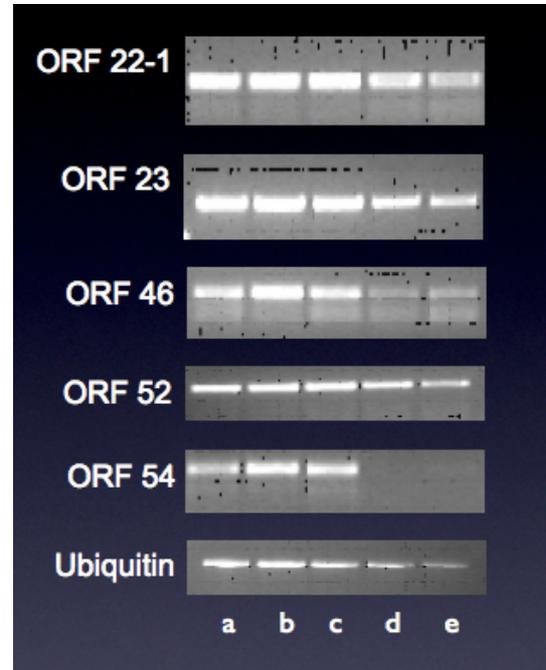


**Fig. 3.1.1:** Overview of the results of the verification of the specificity of the primers used for the amplification of *R1*-homologous genes present on the contig *R1* and *r1*. Each lane (a-f) represents the product after PCR-reaction with a different BAC, using a primerpair designed, to be specific for a certain ORF (table 2.5.1.1) . (a: BC93c12, b: BA132h9, c: BA87d17, d: BA213c14, e: BA76o11, f: BA111o5)



**Fig. 3.1.2:** Schematic view of the BAC-composition of the contigs *R1* and *r1*. The blue arrows show the approximate position of the BAC and the green triangles indicate the position of the open reading frames which form the *R1* family.

The development of gene-specific primers (table 2.5.1.1) for the amplification and detection of the *R1*-homologous genes in *Solanum tuberosum* was a first prerequisite for further analysis. Due to the very high degree of sequence similarity among NBS-LRR genes and of the members of the *R1* and *r1* locus in particular, an assay of different BACs, containing specific members of the *R1* gene family was used to validate the specificity of the primers. This assay showed a specific amplification for ORF 23, ORF 46, ORF 22-1, ORF 52, ORF 54 (fig. 3.1.1 and 3.1.2). In the case of ORF 24 and 45 it was only possible to generate primers, which are specific for both ORFs (due to almost 100% sequence homology). Using cDNA from different tissues of the diploid clone *Solanum tuberosum* P6/210. The expression of the ORFs 22-1, 23, 46, 52 and 54 could be demonstrated in leaf tissues of different ages. Additionally, expression in flowers could be shown for ORFs 22-1, 23, 46 and 52. For ORF 54 no expression in flowers was detected (fig. 3.1.3).



**Fig. 3.1.3:** Expression of five *R1*-homologous genes in different tissues. a) leaf pool b) first leaf c) third leaf d) fourth leaf e) flower. The PCR has been turned out using 35 cycles.

### 3.2.Expression analysis of *R1*-homologous genes of *Solanum nigrum*

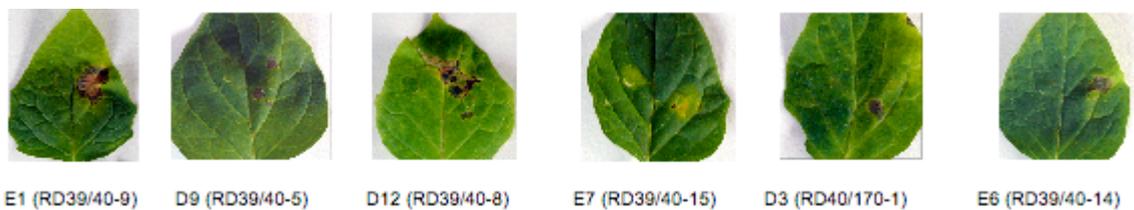
For the detection of *R1*-homologues in *Solanum nigrum*, primers have been used which are specific for four different putative homologous genes (table 2.5.1.1). As template two different tissue pools from *Solanum nigrum* P4 were used. Detection was possible in the cases of *snR1.6* and *snR1.7* (the names indicate the cosmid clone number from which the sequence was obtained, corresponding to fig. 1.4.4). Transcripts of both clones were detected in leaf tissue as well as in flowers. For *snR1.5* only a faint band was detected in floral tissues. The detection of a transcript of *snR1.4* was not possible (fig. 3.2.1).



**Fig. 3.2.1:** Picture of a RT-PCR using cDNA from leaf and floral tissues as well as genomic DNA from *Solanum nigrum*.

### 3.3. Transient expression of effectors from *P. infestans* in *Solanum nigrum*

A transient expression screen using a PVX-system was performed with 68 clones, which contained sequences of putative secreted *Phytophthora infestans* proteins in *Solanum nigrum* P4. Six necrosis inducing clones were identified (see Fig 3.3.1). A comparison of the sequences of the tested clones using nucleic acid sequence alignment and a neighbour-joining method based tree [104] showed that all hypersensitive response



**Fig. 3.3.1:** Picture of *Solanum nigrum* P4 leaves 7 days after infiltration with different *A. tumefaciens* strains containing secreted proteins of *Phytophthora infestans*. The right half was infiltrated with a put. effector clone. The left half with a  $\Delta$ gfp-construct as control.

inducing clones share a high degree of similarity (the sequence information was kindly provided by Dr. Kamoun) (see fig. 3.3.2 and fig. 3.3.3). A comparison of the translated

```

RD39/40-8      -----MSRPLSKTSPDTVATRSLRVEAQEVIQSGRGDGYGGFWKNIIPSTNKI IKKPD I
RD39/40-9      -----MSRPLSKTSPDTVATRSLRVEAQEVIQSGRGDGYGGFWKNIIPSTNKI IKKPD I
RD39/40-5      -----MSRPLSKTSPDTVAPRSLRVEAQEVIQSGRGDGYGGFWKNIIPSTNKI IKKPD I
RD40 170-1     MFPIPDESRLSKTSPDTGATRSLRVEAQEVIQSGRGDGYGGFWKNVFPSTSKI IKKPD I
RD39/40-14     -----MSRPLSKTSPDTGATRSLRVEAQEVIQSGRGDGYGGFWKVAQSTNKIVKRPDI
RD39/40-15     -----MSRPLSKTSPDTVAPRSLRVEAQEVIQSGRGDGYGGFWKVAQSTNKIVKRPDI
                ***** * . ***** ; ***** ; ***** . ; * * . * * ; * * *

RD39/40-8      KISKLIEAAKKAKKK----
RD39/40-9      KISKLIEAAKEAKKK----
RD39/40-5      KISKLIEAAKKAKKK----
RD40 170-1     KISKLIAAAKKAKAKMTKS
RD39/40-14     KIGKLIEAAKKAKAK----
RD39/40-15     KIGKLIEAAKKAKAK----
                ** . *** *** ; ** *

```

**Fig. 3.3.2:** Alignment of the amino acid sequences of six effector clones, which induced a hypersensitive response in *Solanum nigrum*.

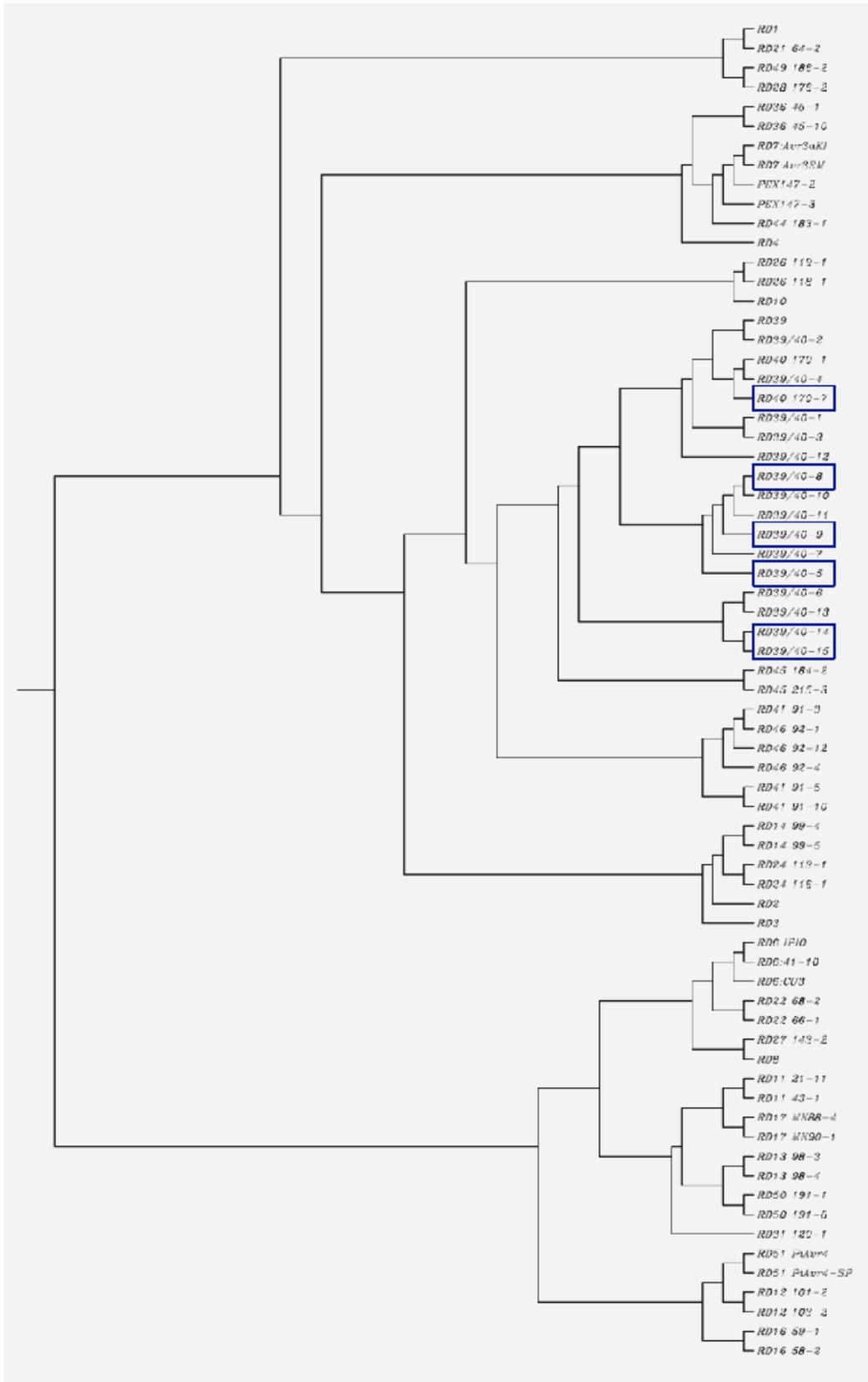
amino acid sequences of the necrosis phenotype inducing clones with the NCBI-database by using the blastp-algorithm, showed, that all clones share high sequence similarity to the known *avrblb2* family effectors ( Acc. EEY54134.1) (see Table 3.3.1) which is highest in the cases of RD 40 170-1, RD39/40-8, RD 39/40-9 and R39/40-15.

**Table 3.3.1:** Comparison of the similarity values of the amino acid sequences compared to *avrblb2*

Clone	Coverage*	E-value**
RD 40 170-1	97%	7 e <sup>-40</sup>
RD39 /40-5	86%	5 e <sup>-22</sup>
RD39 /40-8	98%	1 e <sup>-33</sup>
RD39 /40-9	98%	2 e <sup>-33</sup>
RD39 /40-14	86%	2 e <sup>-28</sup>
RD39 /40-15	98%	2 e <sup>-33</sup>

\*defines the overlapping sequence in of the blast alignment

\*\*indicates the error probability



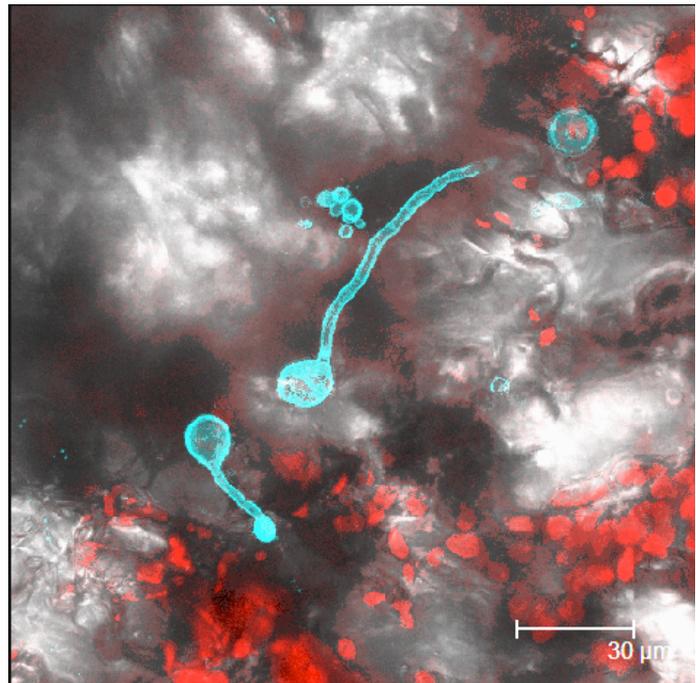
**Fig. 3.3.3:** Phenetic neighbor-joining tree based on the sequences of tested effector clones. The clones with a positive result are marked with a blue frame (note that one of the frames spans two effectors).

### 3.4. Comparative transcriptome analysis of *Solanum tuberosum* plants during the early phase of *Phytophthora infestans* infection

#### 3.4.1. Validation of infected Material

Prior to the application of DeepSAGE, several quality assessments were made on the material, which finally was used for the generation of 3'-tag-libraries. One approach was to achieve an overview of the progress of infection at three days post inoculation. For this, leaves of the susceptible line *Solanum tuberosum* cv. Desirée LV41 were infected with the *P. infestans* strain R208m<sup>2</sup> in a detached leaflet assay. Subsequent confocal laser microscopy pictures showed the initial phase of infection and the formation of appressoria. In figure 3.4.1.1 a layer from the confocal microscopy is displayed. The large blue structures are germinated zoospore cysts of *Phytophthora infestans*. Visible is the original cyst, an initial infection hyphae and in the lower germinated cyst the formation of an appressorium (visible as dense blue round structure). The red dots are plastids, which are under this wavelength (405 nm) autofluorescing. The occurrence of autofluorescing phenolic compounds was almost not visible.

Visually, without microscopical help, an infection phenotype was just detectable in exceptional cases during the experiments. To assess the growth of *P. infestans*, the content of genomic DNA of *Phytophthora infestans* was amplified in the infected samples. This was done using primers specific for the pathogen (table 2.5.1.1) from a control leaf which was



**Fig. 3.4.1.1:** Confocal laser microscopy picture of a leaf from the line *Solanum tuberosum* cv. Desirée LV41 three days post infection with *P. infestans* strain R208m<sup>2</sup>. The cell walls have been stained using Fuchsin. The scale is given by the bar in the lower right corner.

harvested eight days post infection.

This semiquantitative PCR detected in the majority of the tested leaves a PCR-product of the correct size of app. 300 bp (not shown). Additionally, in most of the tested samples originating from the compatible interaction (Desirée LV41 and P4H5K3S2) an increased band intensity was detected. In some samples originating from the incompatible interaction (lines 10-23/2 and 10-2/4), a faint band was detected, too. From this tested samples it was possible to extract complete testing series as defined in chapter 2.6.3.2 for the further application of DeepSAGE. The corresponding PCR-bands are shown in table 3.4.2.1 in chapter 3.4.2.

### **3.4.2. Technical results**

The application of 3'-tag-library sequencing using the Illumina Genome Analyzer platform generated 82314474 sequences. After removal of tags lacking the appropriate CATG-identification key, 51162679 tags remained as high quality reads (for comparison of the sequence quality, please see the file /Raw-data/STATS-Gabor\_Lee\_AHP/All.htm on the supplemental data disc). Of these 47189809 sequences belonged to the experiment described in this study (the difference originates from a positive control which was sequenced within the same sequencing run and a few samples of another experiment). After Singleton-sequences have been removed, 37115611 sequences were used for further analysis.

The data was transformed to the relative unit counts/million and a data-matrix with the values of all 44 samples was constructed. Furthermore, sequences were excluded, which showed an expression of less than 110 counts/million after summing up all samples or which were detected less than 3 times and which contained only adenosine. From the remaining sequences a data set was composed consisting of 44 samples and a total number of 30858 different sequences. The number of unitags per sample varied from 7312 to 24316 (see Fig. 3.4.2.1). The exact number of sequenced tags per sample and the group compositions are summarized in table 3.4.2.1.

**Table 3.4.2.1:** Overview of the results of the „Next Generation Sequencing“

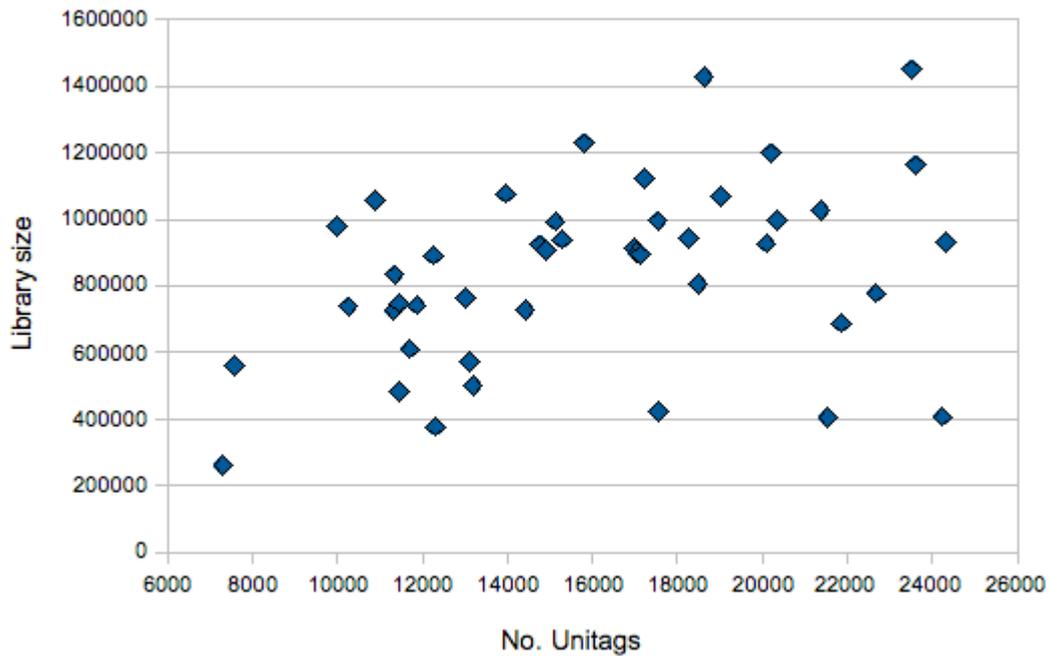
<i>P. infestans</i> control	Ribosomal RNA bands	Sample	No sequenced tags	Phenotype group	No. samples in group
		DesiR1(10-2/4)_0dpi_Rep1	590356	<i>RI</i> -transgenic 0 days post infection (incompatible interaction)	6
		DesiR1(10-2/4)_0dpi_Rep2	1251400		
		DesiR1(10-2/4)_0dpi_Rep3	1404362		
		DesiR1(10-23/2)_0dpi_Rep1	1034900		
		DesiR1(10-23/2)_0dpi_Rep2	532597		
		DesiR1(10-23/2)_0dpi_Rep3	1082820		
		DesiR1(10-2/4)_1dpi_Rep1	1240453	<i>RI</i> -transgenic 1 day post infection (incompatible interaction)	6
		DesiR1(10-2/4)_1dpi_Rep2	795622		
		DesiR1(10-2/4)_1dpi_Rep3	1113138		
		DesiR1(10-23/2)_1dpi_Rep1	1201320		
		DesiR1(10-23/2)_1dpi_Rep2	1640130		
		DesiR1(10-23/2)_1dpi_Rep3	1503866		
		DesiR1(10-2/4)_3dpi_Rep1	844095	<i>RI</i> -transgenic 3 days post infection (incompatible interaction)	6
		DesiR1(10-2/4)_3dpi_Rep2	474425		
		DesiR1(10-2/4)_3dpi_Rep3	1089643		
		DesiR1(10-23/2)_3dpi_Rep1	1084268		
		DesiR1(10-23/2)_3dpi_Rep2	1376621		
		DesiR1(10-23/2)_3dpi_Rep3	1098444		

<i>P. infestans</i> control	Ribosomal RNA bands	Sample	No sequenced tags	Phenotype group	No. samples in group
		Desir1(wt)_0dpi_Rep1	646687	r1 0 days post infection (compatible interaction)	6
		Desir1(wt)_0dpi_Rep2	862680		
		Desir1(wt)_0dpi_Rep3	860094		
		Desir1(LV41)_0dpi_Rep1	1159899		
		Desir1(LV41)_0dpi_Rep2	1192837		
		Desir1(LV41)_0dpi_Rep3	584721		
		Desir1(wt)_1dpi_Rep1	1695177	r1 1 day post infection (compatible interaction)	6
		Desir1(wt)_1dpi_Rep2	1188842		
		Desir1(wt)_1dpi_Rep3	474688		
		Desir1(LV41)_1dpi_Rep1	1186772		
		Desir1(LV41)_1dpi_Rep2	1003347		
		Desir1(LV41)_1dpi_Rep3	659091		
		Desir1(wt)_3dpi_Rep1	1022565	r1 3 days post infection (compatible interaction)	6
		Desir1(wt)_3dpi_Rep2	1133194		
		Desir1(wt)_3dpi_Rep3	1090592		
		Desir1(LV41)_3dpi_Rep1	919803		
		Desir1(LV41)_3dpi_Rep2	1334477		
		Desir1(LV41)_3dpi_Rep3	1191982		

<i>P. infestans</i> control	Ribosomal RNA bands	Sample	No sequenced tags	Phenotype group	No. samples in group
		DesiORF45(P4H5K3S2)_0d-pi_Rep1	1319580	ORF45-transgenic 0 days post infection (compatible interaction)	3
	N.T.*	DesiORF45(P4H5K3S2)_0d-pi_Rep2	1803797		
		DesiORF45(P4H5K3S2)_0d-pi_Rep3	750830		
		DesiORF45(P4H5K3S2)_1d-pi_Rep1	1256793	ORF45-transgenic 1 day post infection (compatible interaction)	3
		DesiORF45(P4H5K3S2)_1d-pi_Rep2	2065877		
		DesiORF45(P4H5K3S2)_1d-pi_Rep3	897277		
		DesiORF45(P4H5K3S2)_3d-pi_Rep1	385086	ORF45-transgenic 3 days post infection (compatible interaction)	2
		DesiORF45(P4H5K3S2)_3d-pi_Rep2	1144661		

\* N.T. = no picture available

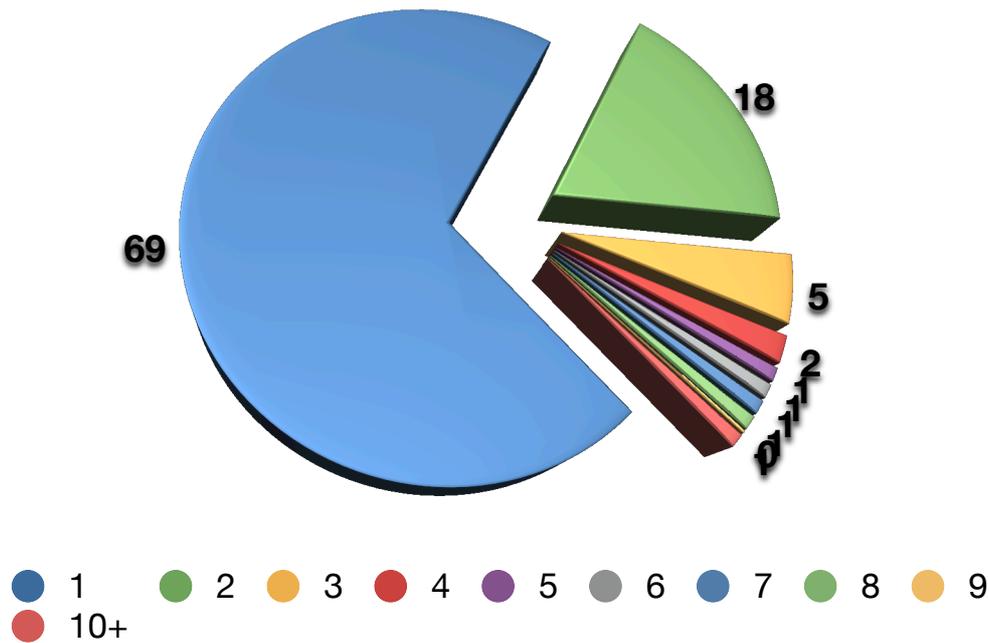
Library size vs number of detected Unitags



**Fig. 3.4.2.1:** Scatterplot of the library sizes (the number of total sequenced tags without normalization) of all sequenced samples and the corresponding number of detected unitags.

The annotation of these tags using a combined reference sequence file consisting of the potato gene index [100-102] and the genomic sequence of *Phytophthora infestans* [5] was successful for 21329 unique tags (8844 with perfect match (grade A) at the 3'-end, 1479 on the 5'-end of the reference sequence but with 100% identity (grade B) and 11006 inside the sequence or with one mismatch (grade C)). This data corresponds to 12185 different transcripts and represents 19,85% of the potato gene index. Another 32 tags have been annotated to the genome of *P. infestans* and for additional 14 tags the origin could be either *Solanum tuberosum* or the pathogen.

### Proportions of possible annotations in the transcriptome data set

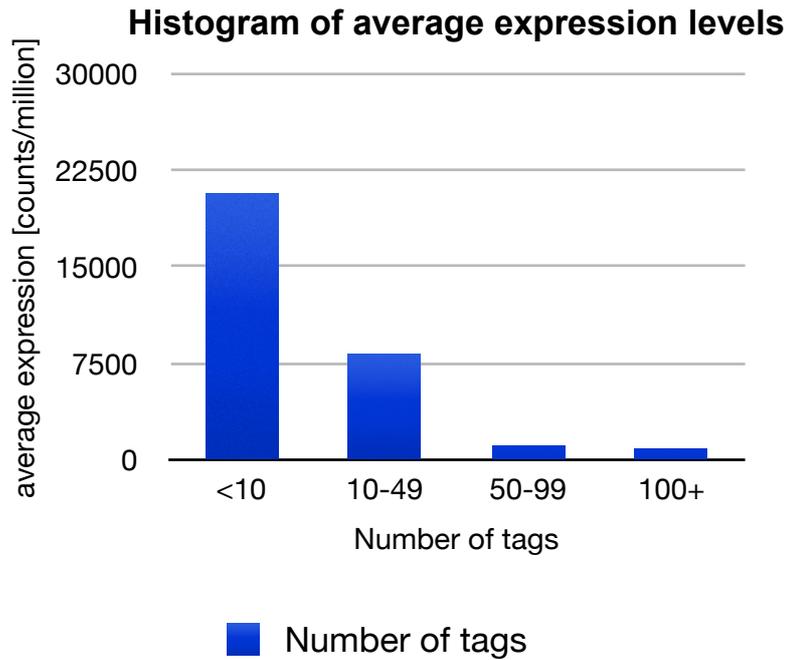


**Fig. 3.4.2.2:** Pie chart of the proportion of possible annotations for the annotated tags. The legend indicates the number of possible annotations while the chart shows their proportion within all annotated tags. The number is the percentage value.

The majority (69%) of the tags were unique in annotation, 18 % could be matched to two transcripts, the remaining 12 % have three and more potential target sequences (see Fig. 3.4.2.2).

Vice versa to the possibility that one tag is allocated to multiple targets, more than one tag is matched to the same transcript as well. Tag TC183365 is an extreme example. For this tentative consensus sequence, 585 different tags matched to this gene, which is annotated as 26S protease subunit 6. The majority of the detected tags shows an average expression of less than 10 counts/million (see Fig. 3.4.2.3).

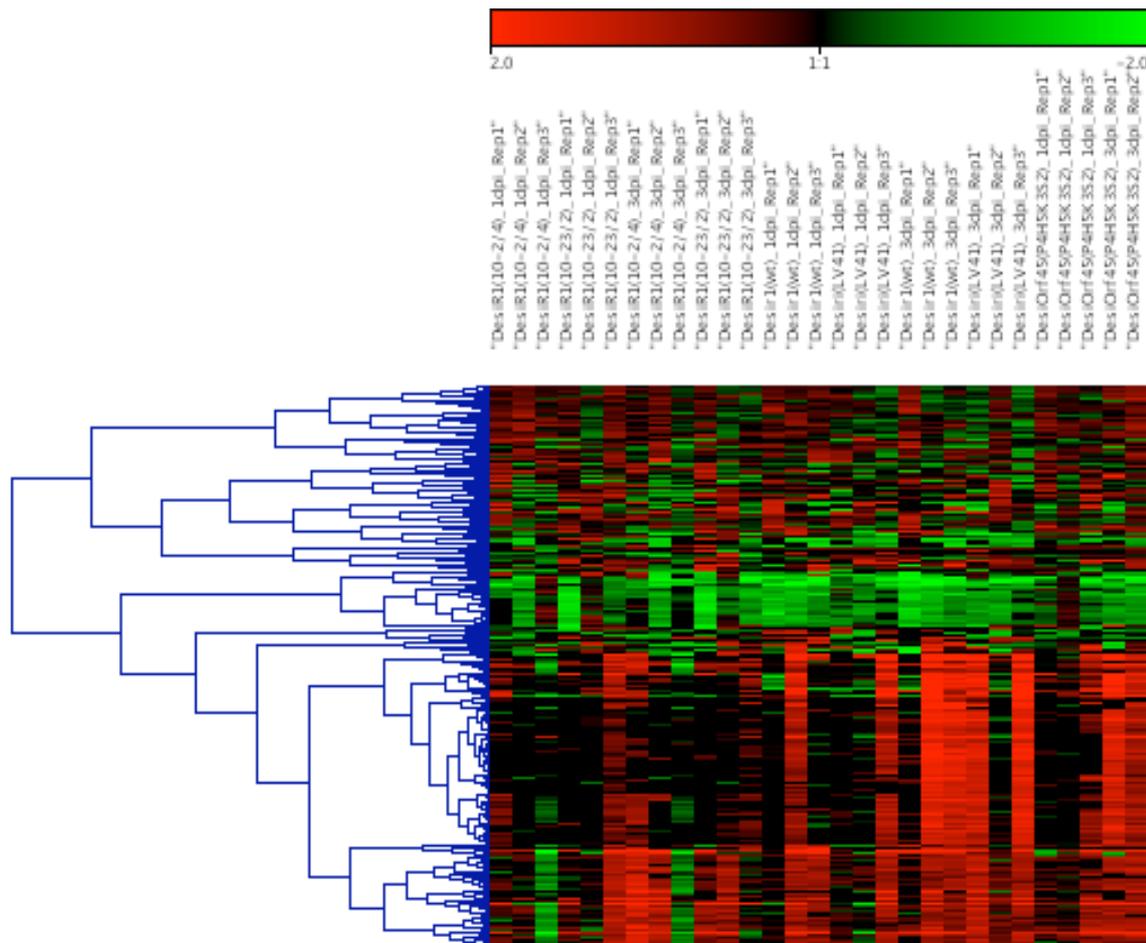
The three highest expressed tags (referring to mean expression levels) with expression levels of 27183,66 counts/million, 10739,95 counts/million and 8221,83 counts/million are StET008016, StET20889 and StET015186. These tags are annotated as a Cell wall protein (TC177374), Photosystem II 10 kDa Polypeptide (TC178472) and Ribulose biphosphate carboxylase small chain 1 (TC167578).



**Fig. 3.4.2.3:** Histogram showing four different categories of expression levels and the corresponding quantity of tags in this category based on the average expression values of all samples.

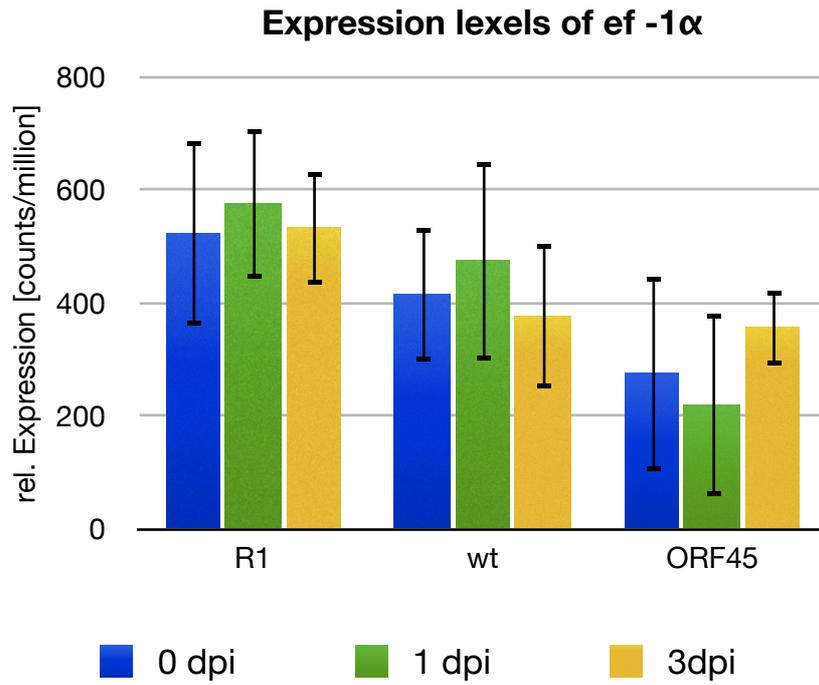
### 3.4.3.Expression levels of indicator-genes

In order to verify the quality of the experiment and to get an idea about the infection state of the samples (see chapter 3.4.2 and fig. 3.4.2.1 for reasons), genes with a putative function in defense response were analyzed separately. To achieve this, all tentative consensus sequences and expressed sequence tags with a gene ontology annotation of defense response (GO:0006592) were extracted from the data set, resulting in the identification of 232 different tags matching to 45 different transcripts. For the majority of the observed transcript assemblies it was possible to identify at least one tag which was under regulatory control. This means that at least four of the six values of a replica group showed a logfoldchange value above 0,3 or below -0,3 (equivalent to approximately two fold increase or reduction). For 14 reference transcripts no obvious up- or downregulation has been found. An overview of the changes in expression levels is given in fig. 3.4.3.1 (a more detailed version is given on the supplementary data disc within the data browser).



**Fig. 3.4.3.1:** Overview of the results of a hierarchical clustering of all tags present in the data set matching to transcripts classified by DFCI as involved in defense response (GO: 0006592)

Beside the validation of a successful response to infection, it is necessary to assess the degree of transcriptional background. For this, the housekeeping gene *ef-1a* was examined in more detail. For this gene one grade A annotated tag was identified (StET018993). The expression dynamic of this tag during the time course does not exceed a twofold change. But the comparison of the expression levels of the genotypes showed that the expression level at one day post infection in *R1*-transgenic plants was more than two times higher than in plants transformed with ORF45 (fig. 3.4.3.2).

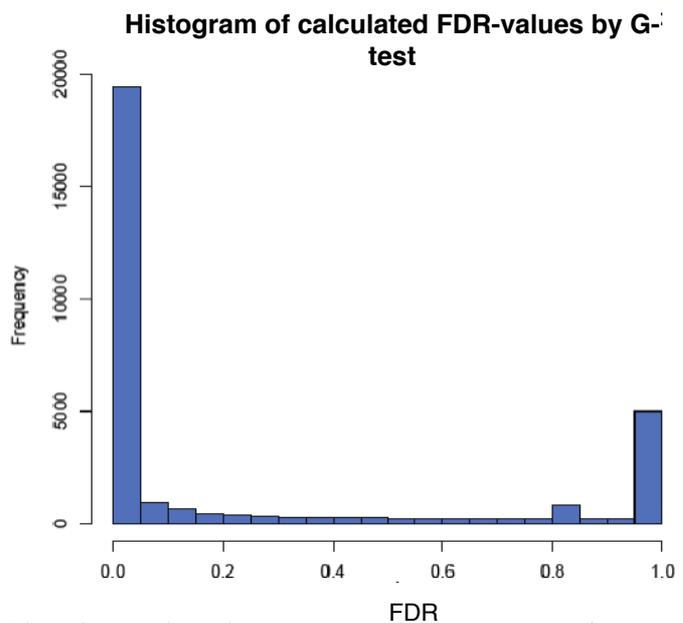


**Fig. 3.4.3.2:** Histogram of the mean expression levels of a tag annotated as elongation factor 1a. Given is the average relative expression and the standard deviation as error bars.

### 3.4.4. Validation of significance statistics

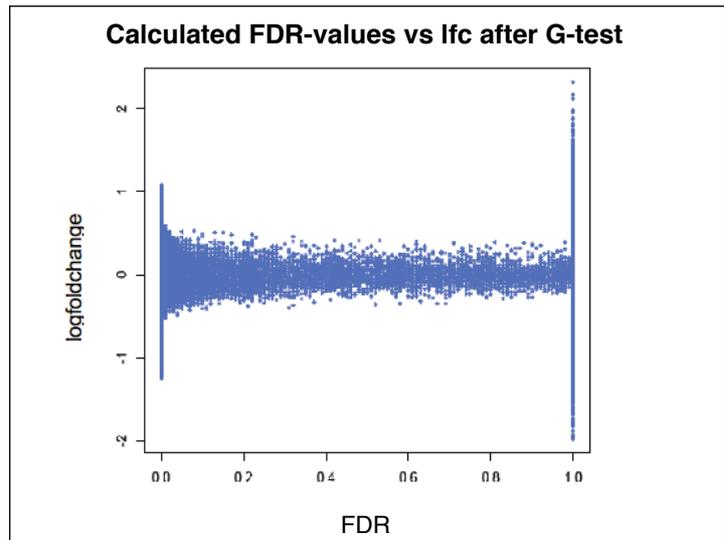
During data analysis, five different statistical tests which do not assume a normal distribution were examined. As orientation, some criteria of Student's t-test were included, although the results are not valid for analysis. The valid tests were the G-test, Fisher's exact test, Bayseq, EdgeR and a permutation based test called Permu. The tests were applied exemplarily to the comparison of wild type plants one day after infection to wild type plants at zero days post infection. The first test implemented was the G-test, which is a more criteria exclusion test (see chapter 1.7 and 2.6.3.3 for details).

The application of this test resulted in a very high number of tags with small p-value (see Fig. 3.4.4.1) in the FDR-value calculations between groups. Nevertheless, after it has been filtered for all criteria which are recommended by the developer [91], 898 tags were left. These tags passed all criteria and

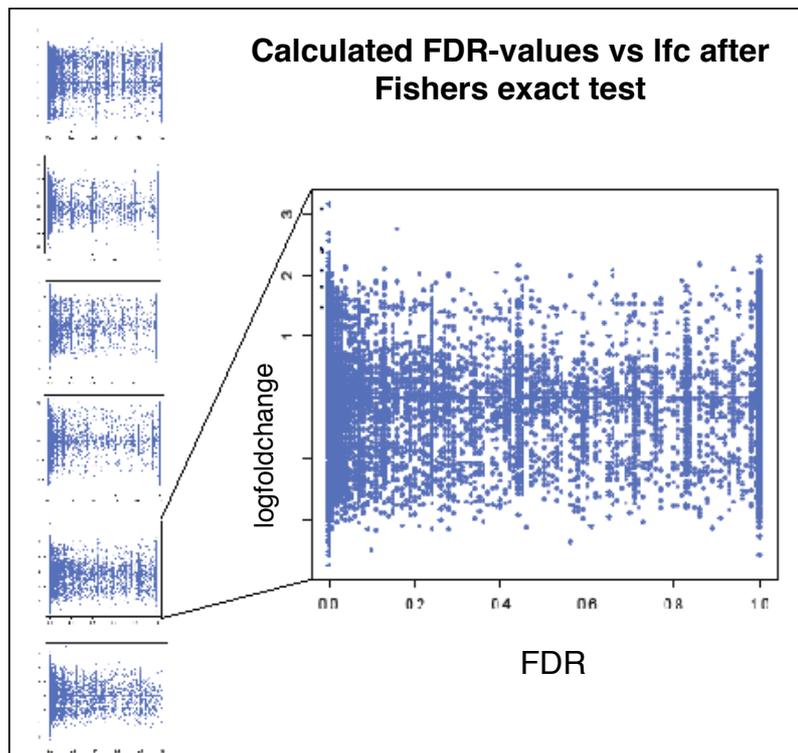


**Fig. 3.4.4.1:** Histogram of the number of raw p-values for the criteria of differences between compared groups. The first bar reflects all FDR-values even or lower than 0,05.

were considered as truly significant. After comparison of the calculated FDR-values and the corresponding expression change values (based on the average logfoldchange), it was obvious that a high proportion of those tags with high expression changes were rejected by the test (see Fig. 3.4.4.2). Moreover, the tags which showed the highest differences were evaluated as not significant. Additionally a Fisher's Exact based test, called „sage.test“, which is included in the SAGE-data analysis package „Sagenhaft“ was performed. Due to the lack of the ability of this test to perform group comparisons, single experiments were compared as they belong to independent experiments (described in chapter 2.6.3.2). Subsequently to the calculation of the FDR-values, these values were combined with the logical parameters described in chapter 2.6.3.5. In all replicates compared, the main characteristics, like p-value distribution over the logfoldchanges were comparable (see fig 3.4.4.3). In the example of Desir1(LV41)Rep2, it is shown that this test, similar to the G-test, rejects a large number of tags, showing a high



**Fig. 3.4.4.2:** Distribution of calculated FDR-values in relation to the average logfoldchange.



**Fig. 3.4.4.3:** Distribution of calculated FDR-values in relation to logfoldchanges. The results of Desir1(LV41)Rep2 are shown in more detail as example

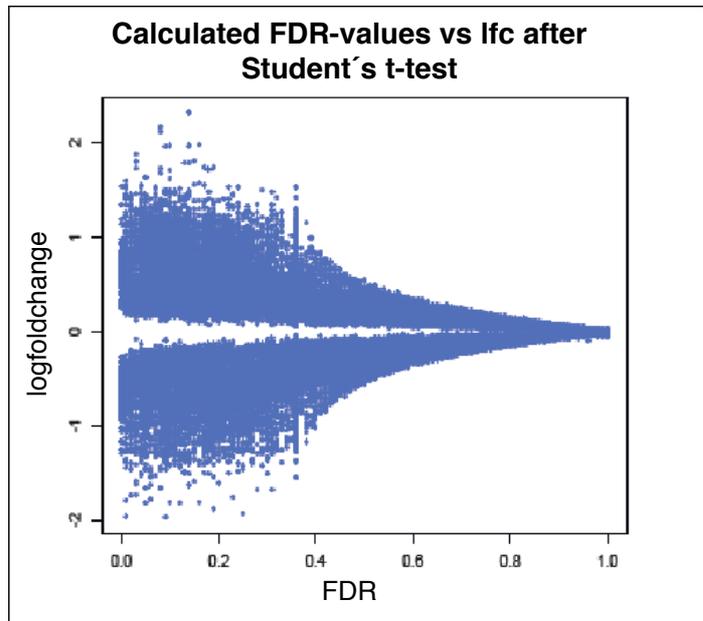
proportion of those tags with high expression changes were rejected by the test (see Fig. 3.4.4.2). Moreover, the tags which showed the highest differences were evaluated as not significant. Additionally a Fisher's Exact based test, called „sage.test“, which is included in the SAGE-data analysis package „Sagenhaft“ was performed. Due to the lack of the ability of this test to perform group comparisons, single experiments were compared as they belong to independent experiments (described in chapter 2.6.3.2). Subsequently to the calculation of the FDR-values, these values were combined with the logical parameters described in chapter 2.6.3.5. In all replicates compared, the main characteristics, like p-value distribution over the logfoldchanges were comparable (see fig 3.4.4.3). In the example of Desir1(LV41)Rep2, it is shown that this test, similar to the G-test, rejects a large number of tags, showing a high

span of average logfoldchange values. A further similarity to the G-test was observed in a high number of low raw FDR-values. In total, almost 14000 tags had a  $FDR \leq 0,05$ . These tendencies were not as strong as in the results of the G-test, and the variation span in those tags with a high FDR-value is not higher than in those with low FDR-values. This leads to the observation that at least genes showing a high differential expression are not excluded per se as it was observed in the G-test.

After combining the FDR-value estimation with the criterion of an at least two fold expression change the number of valid observations was counted (as described in 2.6.3.2). After this testing procedure 34 tags were classified as significant upregulated in 6 of 6 replicates and additional 252 tags showed an upregulation in 5 of 6 replicates.

Another 63 tags were classified as significant downregulated in 6 of 6 replicates and 205 tags showed a significant downregulation in 5 of 6 cases.

The G-test and Fisher's exact test procedures showed no correlation between the FDR-value estimation and the average logfoldchange values. In addition, Student's t-test was performed in order to compare the results with a test, which compares the results to a t-



**Fig. 3.4.4.4:** Plot of the p-values and the corresponding logfoldchange values after Students T-Test.

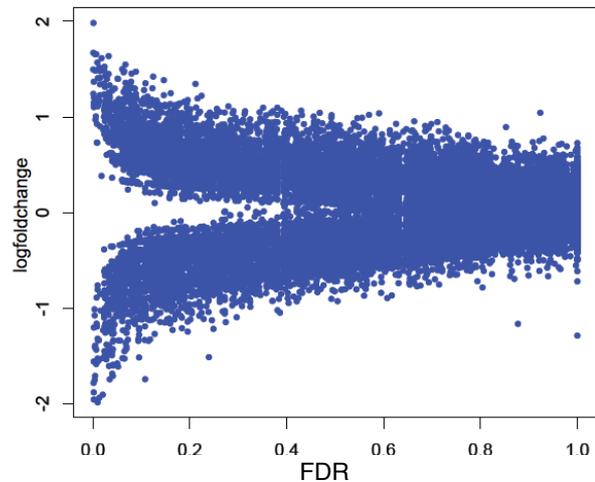
probability distribution. After correlating the calculated FDR-values to the logfoldchange values in this test a clear correlation of increasing FDR-values with decreasing logfoldchange values was observed (fig. 3.4.4.4).

The application of the three more complex tests, edgeR, BaySeq and the in-house developed permutation test „Permu“ showed a similar correlation of the average logfoldchange values to the p-values comparable to the results of Student's t-test. With increasing FDR-

values, a shift of the logfoldchange values towards zero becomes visible and values around zero were not connected with low FDR-value estimations. This tendency was present in all three tests being strongest in the results of BaySeq (fig. 3.4.4.5-7). Please note for the interpretation of the results, that Bayseq generates as significance assessment criterium a logP-value which tests the hypothesis of the presence of two groups. The interpretation of the results is therefore exact the opposite compared to the classical p-value. A result corresponding to a p-value  $\leq 0,05$  or lower is a logP-value  $\geq -0,02$  (maximum = 0). However, in all cases remains the variance in the logfoldchange values at a certain FDR-value high.

One feature of the Permu-test is visible after these plotting procedures. In the area of low FDR-values exists a kind of stuttering. The FDR-values make small jumps toward more rough classes of significance. With higher FDR-values the gap becomes smaller and finally disappears (fig. 3.4.4.6). The main characteristics of each applied test and the number of tags assessed as significant is summarized in table 3.4.4.1.

**Calculated FDR-values vs lfc after edgeR**

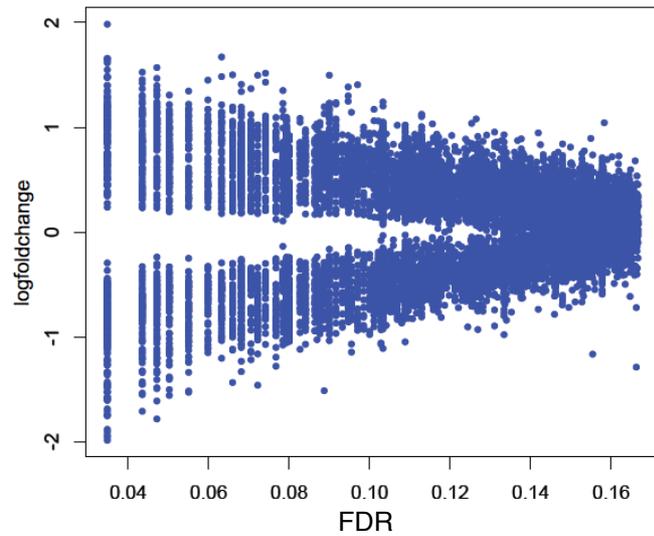


**Fig. 3.4.4.5:** Plot of the FDR-values and the corresponding logfoldchange values after edgeR.

A comparison of the 100 most significant tags from the five tests, which do not use the normal distribution as presumption, resulted in a low overlap of results. The highest congruency existed between BaySeq and edgeR. These both test evaluate 49 tags as high significant in common (see fig. 3.4.4.10).

The implemented test with the most unique tag-list was the G-test. The results showed that 85 of the tags assessed as significant were not among the top 100 in the results of any other test.

**Calculated FDR-values vs lfc after Permu**



**Fig. 3.4.4.6:**Plot of the FDR-values and the corresponding logfoldchange values after Permu.

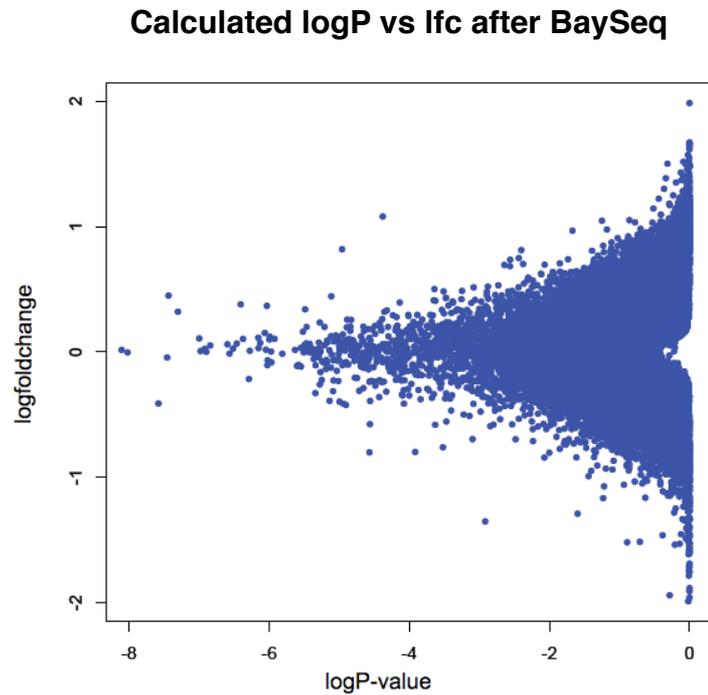
**Table 3.4.4.1:** Summary of the main characteristics of the applied statistical tests

Name	assumed probability distribution	Number of significant
G-test	G-distribution (X <sup>2</sup> -based)	898
sage.test (Fisher's exact test)	negative binomial	2355
edgeR	negative binomial	361
BaySeq	negative binomial	1779
Permu	none (permutation procedure)	710

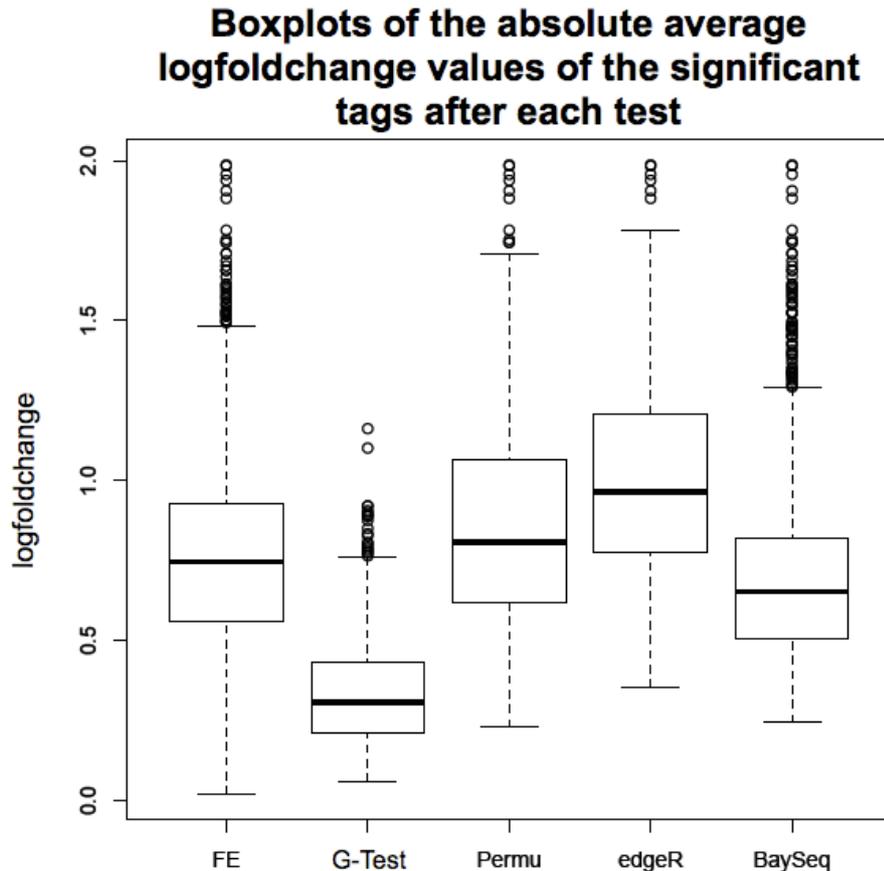
Due to the diverse groups of tags among those with the highest significance, boxplots were created. This opened the possibility to visualize the characteristics of tags, which have been evaluated as significant by each test.

In case of Fisher's Exact test, those tags are presented, which were at least four times significantly different (see chapter 2.4.3.2 for details). For the other tests the p-value had to be smaller or even to 0,05. In cases of Bayseq, the logP-value had to be at least -0,02 (the equivalent to  $p \leq 0,05$ ) (see fig. 3.4.4.5-7). For the creation of the box-plots. The absolute values of the average logfoldchange values were used.

These box-plots showed that the values of the G-test differ remarkably from the other tests. This group shows the highest tendency to smaller logfoldchange values from all applied



**Fig. 3.4.4.7:** Plot of the logP-values and the corresponding logfoldchange values after BaySeq. Given are the logP-values of -8 and more. Above this are a few outliers present, which are not shown for reasons of illustration (please see the supplemental data disc for details).



**Fig. 3.4.4.8:** Boxplots of the average logfoldchange values of those tags, which have been assessed as significant by each test. FE stands for Fisher's exact test.

tests. The highest values are present in the results of edgeR. The results of BaySeq, Fisher's Exact test and Permu are comparable concerning to the logfoldchange behavior (see

**Table. 3.4.4.2:** Matrix table of the correlation coefficients of the as significant assessed tags by each test\*

	BaySeq	edgeR	FE	G-test	Permu
BaySeq	1	0,16998041	-0,03564558	0,027161840	-0,042954454
edgeR	0,16998041	1	-0,01965870	0,016193407	-0,106884208
FE	-0,03564558	-0,01965870	1	-0,039782073	0,075920921
G-test	0,027161840	0,016193407	-0,039782073	1	0,005975535
Permu	-0,042954454	-0,106884208	0,075920921	0,005975535	1

\*this table indicates the correlation coefficient values calculated as described in chapter 2.6.3.2

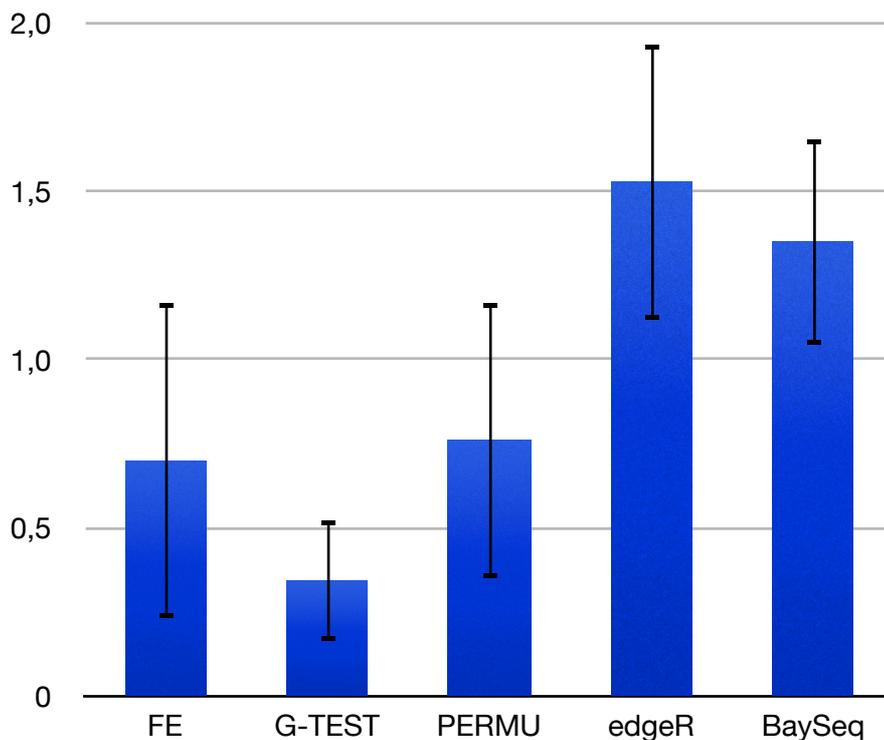
fig. 3.4.4.8).

In order to find describing character for group differences other than the average logfoldchange, the differentiation factor  $D_f$  was introduced and calculated as described in chapter 2.6.3.8.

The value of the differentiation factor increases with increasing difference.

It turned out that the group of significant tags predicted by edgeR and Bayseq showed the greatest internal differences. In general, the analysis of these alternative values corresponds to the comparisons with the logfoldchange values. The only exception is noticed in the results of BaySeq. These corresponded more to this newly introduced factor than on the average logfoldchange values (for comparison see fig. 3.4.4.8 3.4.4.9).

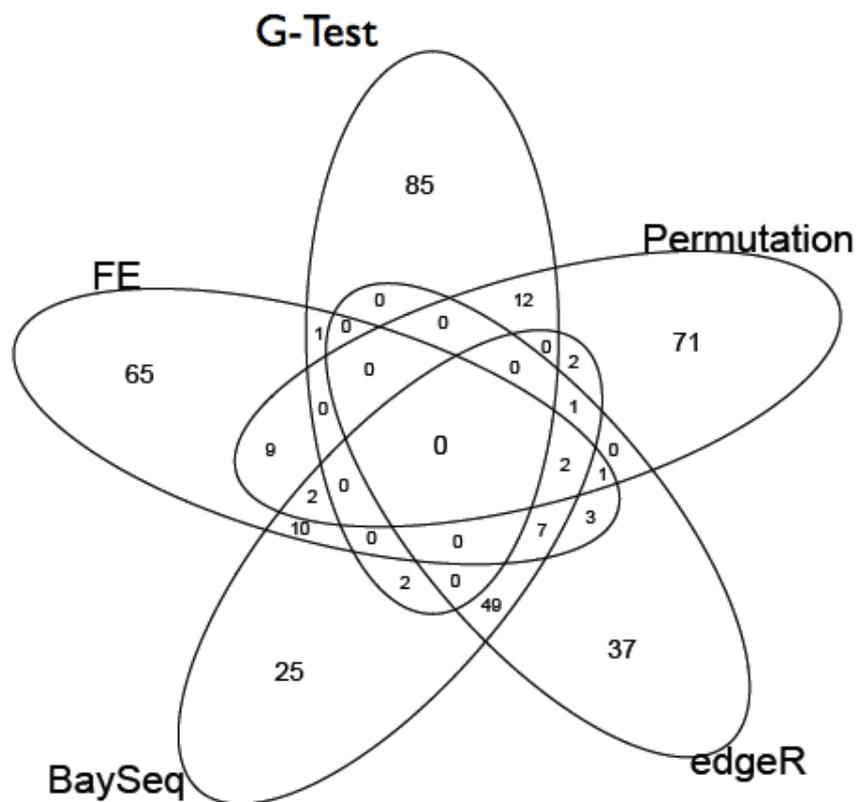
**Average  $D_f$  of the as significant assessed groups by each test**



**Fig. 3.4.4.9:** Histograms of the average  $D_f$ -values of the significant tags by each test. FE stands for Fisher's exact test.

The calculation of a correlation matrix between the test based on this factor in the significant groups resulted in almost no correlation. The highest consensus was found for the values of edgeR and BaySeq. A small congruency can also be seen between the groups resulting from Fisher's exact test and the permutation test. In addition, the resulting lists of edgeR and Permu showed a weak negative correlation.

**Venn diagram of a comparison of the 100 most significant tags of each test**



**Fig. 3.4.4.10:** Venn diagram of a comparison of the 100 most significant tags calculated by each valid test. FE stands for Fisher's exact test. The number inside the overlapping fields show the consensus between the groups.

### 3.4.5. Results of global statistical analysis: Cluster and Principal component analysis

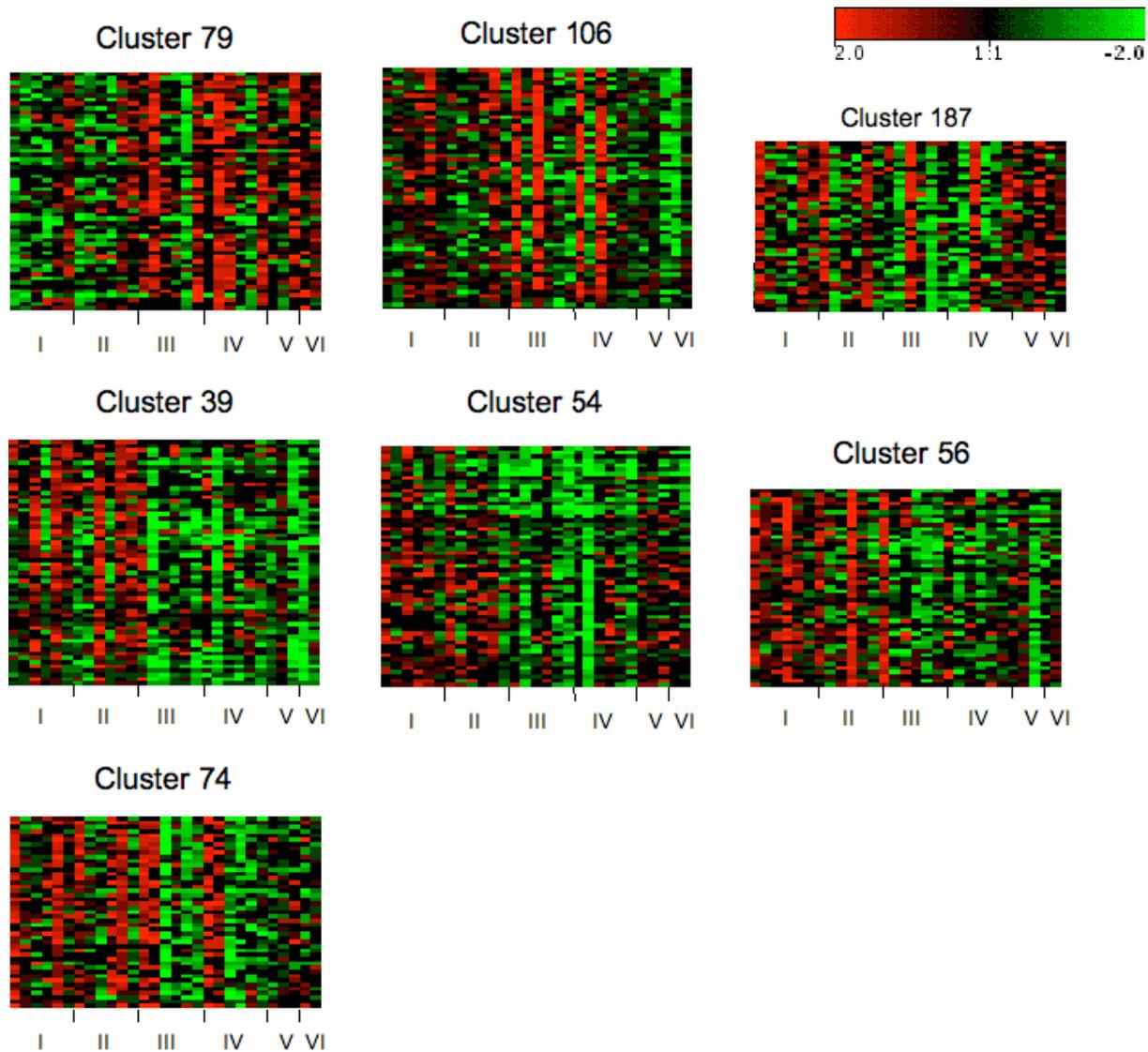
As method to arrange data in similar groups, cluster analysis methods are the standard tools for the identification of genes, whose response to a certain condition is similar. This study included a k-mean clustering as well as hierarchical clustering.

**Table 3.4.5.1:** This table summarizes the relevant characteristics of identified clusters after a k-mean clustering\*

Cluster name	Number of genes	Tendency
Cluster 9	28 (11; 1)	R1(up), wt(1 dpi down)
Cluster 21	35 (4; 1)	R1(up), wt (1 dpi up)
Cluster 39	58 (36; 0)	wt(down), ORF45(down)
Cluster 54	57 (23; 4)	wt(down)
Cluster 56	47 (19; 0)	R1(1 dpi up), wt(down), ORF45 (down)
Cluster 62	46 (11; 1)	wt(up), ORF45(down)
Cluster 67	57 (18; 0)	wt(up), ORF45(up)
Cluster 74	45 (11; 2)	wt(down)
Cluster 79	56 (29; 0)	R1(down)
Cluster 99	39 (15; 0)	R1(down), wt(up), ORF45 (down)
Cluster 103	54 (21; 0)	R1(down), wt(up), ORF45 (down)
Cluster 106	57 (29; 1)	wt(3 dpi down), ORF45(down)
Cluster 114	61 (27; 1)	R1(up), wt(3 dpi up), ORF45 (down)
Cluster 135	87 (57; 8)	wt(3 dpi up)
Cluster 145	68 (30; 0)	wt(up)
Cluster 147	95 (68; 16)	R1(3 dpi up), wt(up), ORF45 (up)
Cluster 180	34 (13; 1)	R1(down), wt(up), ORF45 (down)
Cluster 187	40 (8; 1)	wt(down), ORF45 (up)

\*Given is the genotype group (R1 indicates *R1*-transgenic lines; wt the wild type and the mock-transformed line; ORF45 the line transformed with ORF45) and the time point. If no time point is given, both time points are affected. Additionally the direction of the expression level change is specified by up (meaning an upregulation) or down (meaning downregulation). The middle column gives the number of tags in the cluster. In parentheses is given first the number of meaningful annotated tags and second the number of tags classified as responding to stress, wounding or defense.

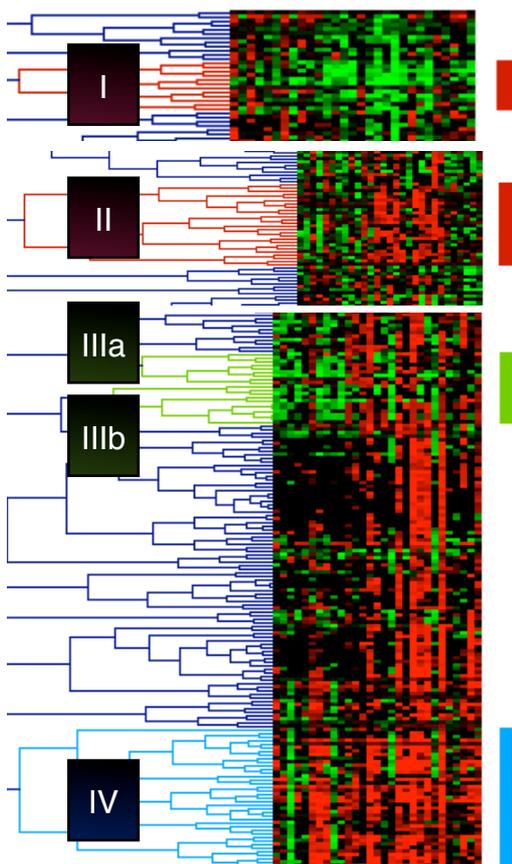




**Fig. 3.4.5.1:** Overview of the extracted clusters after k-mean cluster analysis. The roman numbers indicate the different replica groups: I – R1 1 dpi; II – R1 3 dpi; III wt 1 dpi; IV – wt 3 dpi; V – ORF45 1 dpi; VI – ORF45 3 dpi.

In the k-mean cluster analysis, 216 clusters were calculated. The number was chosen according to the given data. The criteria finally consisted of (6 possible phenotypic groups)<sup>3</sup> possibilities for regulation. As phenotypic groups were defined *R1* at 1dpi, *R1* at 3 dpi, wt at 1 dpi, wt at 3 dpi, ORF45 at 1 dpi and ORF45 at 3 dpi. The values are the result of a comparison to the corresponding 0 dpi value. The three possibilities for regulation were upregulated, downregulated and not regulated. According to this, in the non-hierarchical k-mean clustering 18 clusters were identified which show, based on the mean-cluster characteristics, a tendency towards a common expression response under a specific test condition. (For an overview of extracted clusters see Tab. 3.4.5.1 and Fig. 3.4.5.1 and

chapter 2.6.3.9 for details of the cluster validation, a full list of the 216 clusters is given on the supplemental data disc within the data browser). Within these 18 clusters, the *R1*-transgenic lines showed an upregulation in five clusters and a downregulation in four. For the wild type lines, tags were detected as upregulated in ten clusters and downregulated in six. The ORF45 containing line showed a downregulation in eight clusters and an upregulation in three. The two largest clusters are „Cluster 135“ (87 unitags), which just shows upregulated genes in wild type plants three days after infection and „Cluster



**Fig. 3.4.5.2:** Apertures of the tree image of the hierarchical clustering of the k-mean clustering extracted genes. The as highly relevant observed regions are marked by a colored bar. Green indicates a *R1*-line specific region, red a region responding specifically in wild type plants and blue marks a region, which is upregulated in all infected lines. The arrangement of the samples is comparable to fig. 3.4.6.1. (A larger view of the whole clustering is given on the supplemental data disc.)

147“ (95 unitags), in which all tested lines show an upregulation in at least one of the measured time points. The smallest cluster is Cluster 9 (28 unitags). The genes present in this cluster show an upregulation in *R1*-transgenic plants and are downregulated in wild type plants one day after infection. The proportions of meaningful annotations, what means a match to an enzyme or gene vary from 11% in Cluster 21 to 72 % in Cluster 147. The proportions of genes with a gene ontology term „response to stress, wounding or defense“ vary from 0 % in Cluster 39, 56, 67, 79, 99, 103 and 145 to 17 % in Cluster 147.

These 18 clusters contained together 935 tags. In order to avoid the methodological bias, which occurs by the predefinition of the cluster-number in k-mean clustering methods, and to extract the most interesting groups, the genes of these clusters were combined into one data set and a hierarchical cluster analysis was performed. This enabled the identification of more clearly responding gene groups. By hierarchical cluster analysis, 4 extraordinary tag groups could be identified. One group of genes, which seems to be upregulated in all genotypes (blue; IV), one in which the wild type

**Table 3.4.5.2:** Comprehension of tags identified by hierarchical clustering\*

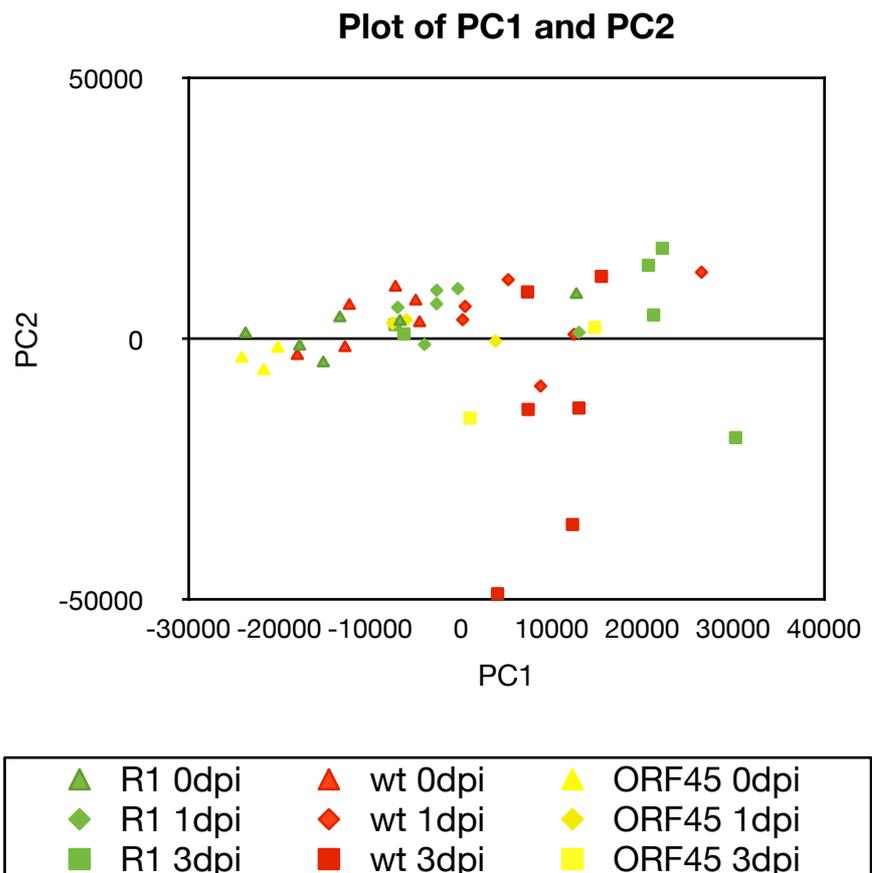
Specific region	number of tags	Annotations (without NA and references to chromosomal scaffolds)
wt, upregulated	26	<ul style="list-style-type: none"> <li>-At4g33100</li> <li>-EXS, C-terminal</li> <li>-Polyubiquitin</li> <li>-Ethylene signaling protein</li> <li>-Proline-rich protein family-like</li> <li>-Protein flbE</li> <li>-91A protein</li> <li>-28 kDa small subunit ribosomal protein;</li> </ul>
wt, downregulated	12	<ul style="list-style-type: none"> <li>- Carbonic anhydrase</li> <li>- Proteinase inhibitor 1</li> <li>- Pectate lyase</li> <li>- Proteinase inhibitor 1</li> <li>- Non-symbiotic hemoglobin 2</li> <li>- Pherophorin-C1 protein precursor</li> <li>- Predicted protein</li> </ul>
<i>RI</i> , downregulated	18	<ul style="list-style-type: none"> <li>- Patatin-like protein</li> <li>- Sec61beta family protein</li> <li>- Avr9/Cf-9 rapidly elicited protein 140</li> <li>- Xyloglucan endotransglucosylase-hydrolase XTH3</li> <li>- Nicotianamine synthase</li> <li>- EDA16 (embryo sac development arrest 16)</li> <li>- Avr9/Cf-9 rapidly elicited protein 65</li> <li>- Phenylalanine ammonia-lyase 1</li> <li>- Phenolic oxidative coupling protein Hyp-1</li> <li>- Ubiquitin carboxyl-terminal hydrolase</li> <li>- Mitochondrial import receptor subunit TOM20</li> </ul>
all, upregulated	39	<ul style="list-style-type: none"> <li>- Calmodulin-binding protein</li> <li>- Citrate binding protein</li> <li>- Delta-tonoplast intrinsic protein;</li> <li>- Glucan endo-1.3-beta-D-glucosidase precursor</li> <li>- WRKY-type DNA binding protein</li> <li>- Polyphenol oxidase</li> <li>- Non-specific lipid-transfer protein</li> <li>- Pathogenesis-related protein 10</li> <li>- Uncharacterized Cys-rich domain</li> <li>- ParA family protein</li> <li>- Non-specific lipid-transfer protein</li> <li>- TATA-box-binding protein</li> <li>- Lipid desaturase-like protein</li> <li>- pathogenesis related protein 10</li> <li>- Alpha-DOX1</li> <li>- Wound-induced protein WIN1 precursor</li> <li>- Ethylene-responsive proteinase inhibitor 1</li> <li>- PR-1</li> <li>- TMV-induced protein I</li> <li>- Osmotin-like protein (2x)</li> <li>- Wound-induced protein WIN2 precursor</li> <li>- Kunitz-type protease inhibitor</li> <li>- Hypoxia-responsive family protein</li> <li>- Peroxidase</li> <li>- Sesquiterpene synthase</li> <li>- Probable glutathione-S-transferase</li> <li>- Extensin-like protein</li> <li>- Formate dehydrogenase. mitochondrial precursor</li> <li>- Osmotin-like protein OSML15 precursor</li> <li>- TMV-induced protein I</li> </ul>

\*Given is the expression behavior, the number of coregulated tags and a shorter version of meaningful annotated tags inside the cluster arm.

plants specifically are downregulated (red; I), one in which the wild type plants specifically are upregulated (red; II), and one region, which contains genes, downregulated in *R1*-transgenic lines (green; IIIaIIIb). For ORF45 no clear group which indicates specific behavior was identified.

The main results of the hierarchical clustering are summarized in Table 3.4.5.2 (the full information is given on the supplemental data disc within the data browser). This table gives the information of the number of detected tags and the annotations given inside these cluster arms. The information of tags lacking an annotation or which are referred as „genomic scaffold of *Vitis vinifera*“ were filtered out. The largest group consisted of 39 tags and showed an upregulation in all genotypic lines. In this group, 30 of 39 tags were annotated meaningful, corresponding to 76,9 %. This was the most successful annotation rate. The other groups showed smaller annotation rates, with 30% in the upregulated wild type plants being the smallest value.

In addition to the cluster analysis, which is able to identify groups of similar expression values, principal component analysis was performed. The method was used to comprise the whole transcriptome data composition from each sample. This analysis resulted in a very rough grouping of the samples by the infection time course in the principal component 1. With the exception of outliers, the majority of the samples grouped over the infection time from left (negative values) to right



**Fig. 3.4.5.3** Biplot of the principal component 1 (PC1) and the principal component 2 (PC2) of a principal component analysis based on the transcriptome expression data of all measured samples.

(positive values). A separation between the tested lines does not occur clearly, although a weak tendency can be seen between the wild type lines and the *R1*- transgenic line in principal component 2 after three days post infection towards negative values in the wild type samples. An independent grouping of ORF45-transgenic plants was not detected (fig 3.4.5.3).

### 3.4.6. Results of significance of observation after Fisher's Exact Test

**Table 3.4.6.1:** Table of summarized results of tags which show independent significant expression changes in *R1* transgenic plants\*

Tag-ID	prim. Annot. ID	min p**	max p**	A.lfc.***	mi.lfc***	ma.lfc***	time point	prim. Annot. name
StET029144	TC163636	0	0,01	-0,92	-1,35	-0,5	1 dpi	Tuber specific and sucrose-responsive element binding factor
StET027738	TC172763	0	0,04	-1,18	-0,62	-0,39	3 dpi	Cluster: EIN3-binding F-box protein
StET027737	unknown	0	0,03	-1,98	-0,71	-0,32	3 dpi	NA
StET018266	CV475279	0	0,02	-1,94	-0,78	-0,36	3 dpi	Chromosome chr7 scaffold_20
StET004373	TC165739	0	0,04	-1,2	-0,86	-0,51	3 dpi	Chromosome chr3 scaffold_8
StET010841	TC176453	0	0,02	-1,56	-1,15	-0,65	3 dpi	Fibrillin 8

\* Given are those tags, which showed a significant expression level change in 6 of 6 replicates and in maximum one replicate under wild type condition.

\*\* The columns min p or max p give the lowest or the highest calculated significance value (after multiple testing correction).

\*\*\*A.lfc gives the average logfoldchange; mi.lfc gives the minimum logfoldchange and ma.lfc indicates the highest logfoldchange (please note, that minimum and maximum is given mathematically, so in downregulated tags, the mi.lfc value indicates the strongest response).

As significance test, which deals most flexible with the biological variance, Fisher's exact test was performed to estimate the significance of observation for all comparison of the infection progress. A comparison between the genotypes was not performed due to the observed uncertainties in the expression of *ef-1a* (see chapter 3.4.3). This test was

performed individually for each independent experiment over time as described in chapter 2.6.3.2 2.6.3.4.

In table 3.4.6.1-3.4.6.3 the results of these tests are summarized. Shown are those tags, which showed a significant change in 6 of 6 experiments and at maximum one significant change in the corresponding control. This control consisted of the expression characteristics of the wildtype in cases of both transgenic lines and of the expression characteristics of *R1*-transgenic lines in cases of the wildtype plants. (For example six significant observations in *R1* transgenic plants and one in wildtype plants.) For reasons of illustration for ORF45 transgenic plants, only those are given which were significant in all cases and showed a logfoldchange of one or above. The supplemental data disc contains the full lists of significant tags up to four of six valid observation (two of three and two of two in case of ORF45). By listing these top significant and unique tags, it turns out that for the *R1*-transgenic lines, six tags fulfilled the mentioned criteria. Five of the significant changes were observed three days after infection and were downregulated. This represents the smallest group of genes.

In wildtype plants, in total 31 tags were identified. 16 were significant at one day post infection and 15 at three days post infection.

**Table 3.4.6.2:** Table of summarized results of tags which only showed significant expression value changes in wild type plants\*

Tag-ID	prim. Annot. ID	min p**	max p**	A.lfc.***	mi.lfc.***	ma.lfc.**	time point	prim. Annot. name
StET017962	TC176177	0	0,01	1,28	1,04	1,58	1dpi	NA
StET030331	TC173012	0	0	1,49	1,04	1,71	1dpi	CBL-interacting protein kinase 25
StET013338	TC165752	0	0,01	1,18	0,39	1,5	1dpi	Chromosome chr17 scaffold_12
StET002687	TC164231	0	0	1,24	0,55	1,48	1dpi	Cap-binding protein CBP20
StET020154	CV504927	0	0,01	1,33	0,42	1,47	1dpi	T4O12.27
StET014099	TC188521	0	0	1,46	0,42	1,7	1dpi	NA

Tag-ID	prim. Annot. ID	min p**	max p**	A.lfc.***	mi.lfc.***	ma.lfc.**	time point	prim. Annot. name
StET019871	TC167304	0	0	1,4	0,42	1,88	1dpi	Chromosome chr5 scaffold_2
StET018100	TC190958	0	0,02	1,19	1	1,86	1dpi	Proton pump interactor
StET025630	CK853475	0	0	1,18	0,39	1,95	1dpi	Serine/threonine protein kinase
StET009293	TC191591	0	0	-0,54	-1,23	-0,37	1dpi	Chloroplast 50S ribosomal protein L2
StET025786	TC173026	0	0	-0,58	-3,24	-0,32	1dpi	Ribulose biphosphate carboxylase small chain 3
StET027128	TC179956	0	0	-0,49	-0,61	-0,31	1dpi	Chromosome chr18 scaffold_61
StET009803	TC191566	0	0	-0,66	-1,18	-0,44	1dpi	AGAP011901-PA
StET029662	TC163337	0	0	-0,51	-0,83	-0,33	1dpi	Oxygen-evolving enhancer protein 2
StET023638	unknown	0	0,02	-0,94	-1,76	-0,5	1dpi	NA
StET005170	TC177397	0	0,01	-1	-1,46	-0,64	1dpi	Tonoplast intrinsic protein
StET027779	TC175508	0	0	-1,33	-1,79	-0,98	3dpi	Phospholipase A2 precursor
StET006445	unknown	0	0,02	-1,31	-2,01	-1,01	3dpi	NA
StET020200	TC189194	0	0,02	-1,27	-1,76	-1,01	3dpi	Pherophorin-C1 protein precursor
StET015408	BQ115199	0	0	-0,77	-2,14	-0,51	3dpi	50S ribosomal protein L27
StET001709	DN906660	0	0,01	-0,84	-2,26	-0,33	3dpi	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit
StET027958	TC174667	0	0	1,97	0,53	2,45	3dpi	Salicylic acid-induced protein 19

Tag-ID	prim. Annot. ID	min p**	max p**	A.lfc.***	mi.lfc.***	ma.lfc.**	time point	prim. Annot. name
StET025045	TC176398	0	0,04	2,2	0,87	2,41	3dpi	Glucan endo-1,3-beta-D-glucosidase precursor
StET005812	TC168480	0	0	2,26	1,15	2,63	3dpi	Chromosome chr18 scaffold_1, whole genome shotgun sequence
StET029512	TC167849	0	0,01	2,24	0,93	2,69	3dpi	Pathogenesis related protein PR-1
StET026244	TC190950	0	0	0,78	0,3	1,41	3dpi	14-3-3 protein
StET005539	unknown	0	0,01	1,82	0,59	2,3	3dpi	NA
StET006462	TC173953	0	0,02	1,27	0,32	1,69	3dpi	Chromosome chr18 scaffold_1
StET006606	TC163449	0	0	1,46	0,64	1,79	3dpi	Cytosolic ascorbate peroxidase 1
StET017601	DR034844	0	0,04	1,34	0,38	1,54	3dpi	Gamma-aminobutyrate transaminase subunit isozyme 3
StET016400	TC190641	0	0,02	1,4	1,14	1,63	3dpi	Chromosome chr1 scaffold_84

\* Given are those tags, which showed a significant expression level change in 6 of 6 replicates and in maximum one replicate under *RI*-transgenic conditions.

\*\*The columns min p or max p give the lowest or the highest calculated significance value.

\*\*\*A.lfc gives the average logfoldchange; mi.lfc gives the minimum logfoldchange and ma.lfc indicates the highest logfoldchange (please note, that minimum and maximum is given mathematically, so in downregulated tags, the mi.lfc value indicates the strongest response).

In the case of the line transformed with the ORF45, for reasons of illustration, only those tags are presented in this chapter which showed a logfoldchange of at least 1. In total after one day of infection 86 tags were downregulated and 196 at three days after infection. Additional 117 tags are significantly upregulated after one day and 334 after three days post infection (table 3.4.6.3).

For the subsequent GO-Term analysis, all tags included in lists on the data disc were

**Table 3.4.6.3:** Table of summarized results of tags which showed significant expression value changes only in ORF45 transgenic plants\*

Tag-ID	prim. Annot. ID	min p**	max p**	A.lfc.***	mi.lfc.***	ma.lfc.***	time point	prim. Annot. name
StET009055	unknown	0	0	-1,32	-1,46	-1,17	1dpi	NA
StET008298	unknown	0	0	-1,16	-1,3	-1,09	1dpi	NA
StET017921	TC174507	0	0	-1,34	-1,54	-1,24	1dpi	Chromosome undetermined scaffold_235
StET002612	unknown	0	0	1,31	1,05	1,57	1dpi	NA
StET004795	unknown	0	0	-1,48	-1,74	-1,31	3dpi	NA
StET020395	unknown	0	0	-1,4	-1,66	-1,24	3dpi	NA
StET001931	unknown	0	0	-1,34	-1,74	-1,14	3dpi	NA
StET006171	unknown	0	0,01	-1,29	-1,41	-1,19	3dpi	NA
StET006494	unknown	0	0,03	-1,29	-1,34	-1,25	3dpi	NA
StET000419	unknown	0	0,01	-1,36	-1,42	-1,3	3dpi	NA
StET000382	unknown	0	0	-1,28	-1,56	-1,11	3dpi	NA
StET014937	BQ114945	0	0,03	-1,2	-1,34	-1,09	3dpi	RuBisCO large subunit-binding protein subunit beta
StET014228	unknown	0	0	-1,3	-1,48	-1,18	3dpi	NA
StET001869	TC165118	0	0	-1,5	-1,69	-1,37	3dpi	Isocitrate lyase
StET018528	TC188425	0	0,01	-1,27	-1,42	-1,16	3dpi	Chromosome chr13 scaffold_17
StET014906	unknown	0	0,01	-1,21	-1,41	-1,07	3dpi	NA
StET005005	unknown	0	0	-1,22	-1,6	-1,02	3dpi	NA
StET029630	TC172868	0	0,03	-1,21	-1,34	-1,11	3dpi	Probable transposase
StET024994	unknown	0	0	1,29	1,02	1,46	3dpi	NA
StET009052	unknown	0	0	1,76	1,23	1,99	3dpi	NA
StET025031	CV477637	0	0	1,51	1,09	1,72	3dpi	Cytochrome d1 heme region precursor
StET016441	TC187954	0	0	1,59	1,12	1,81	3dpi	Predicted protein
StET019804	TC166407	0	0	1,08	1,08	1,09	3dpi	Induced stolen tip protein TUB8
StET018535	unknown	0	0	1,61	1,06	1,85	3dpi	NA
StET022115	unknown	0	0	1,46	1,15	1,64	3dpi	NA
StET016928	TC180376	0	0	1,61	1,02	1,85	3dpi	Chromosome chr4 scaffold_162
StET002242	TC171485	0	0	1,48	1,02	1,69	3dpi	Chromosome chr13 scaffold_210
StET024123	DR037288	0	0	1,14	1,09	1,19	3dpi	NA

Tag-ID	prim. Annot. ID	min p**	max p**	A.lfc.***	mi.lfc.***	ma.lfc.***	time point	prim. Annot. name
StET029144	TC163636	0	0	1,61	1,04	1,85	3dpi	Tuber-specific and sucrose-responsive element binding factor
StET013479	unknown	0	0	1,54	1,09	1,76	3dpi	NA
StET008045	TC177374	0	0,04	1,05	1,02	1,09	3dpi	Cell wall protein
StET004533	TC182782	0	0	1,68	1,24	1,89	3dpi	Chromosome undetermined scaffold_310
StET002188	unknown	0	0	1,8	1,22	2,04	3dpi	NA
StET006157	TC177490	0	0,04	1,04	1,02	1,06	3dpi	NA
StET010511	unknown	0	0	1,51	1,09	1,72	3dpi	NA
StET013632	unknown	0	0	1,6	1,09	1,83	3dpi	NA
StET026236	TC174826	0	0	1,47	1,09	1,67	3dpi	NA

\*Given are those tags, which showed a significant expression level change in all replicates and no replicate under wildtype conditions.

\*The columns min p or max p give the lowest or the highest calculated significance value.

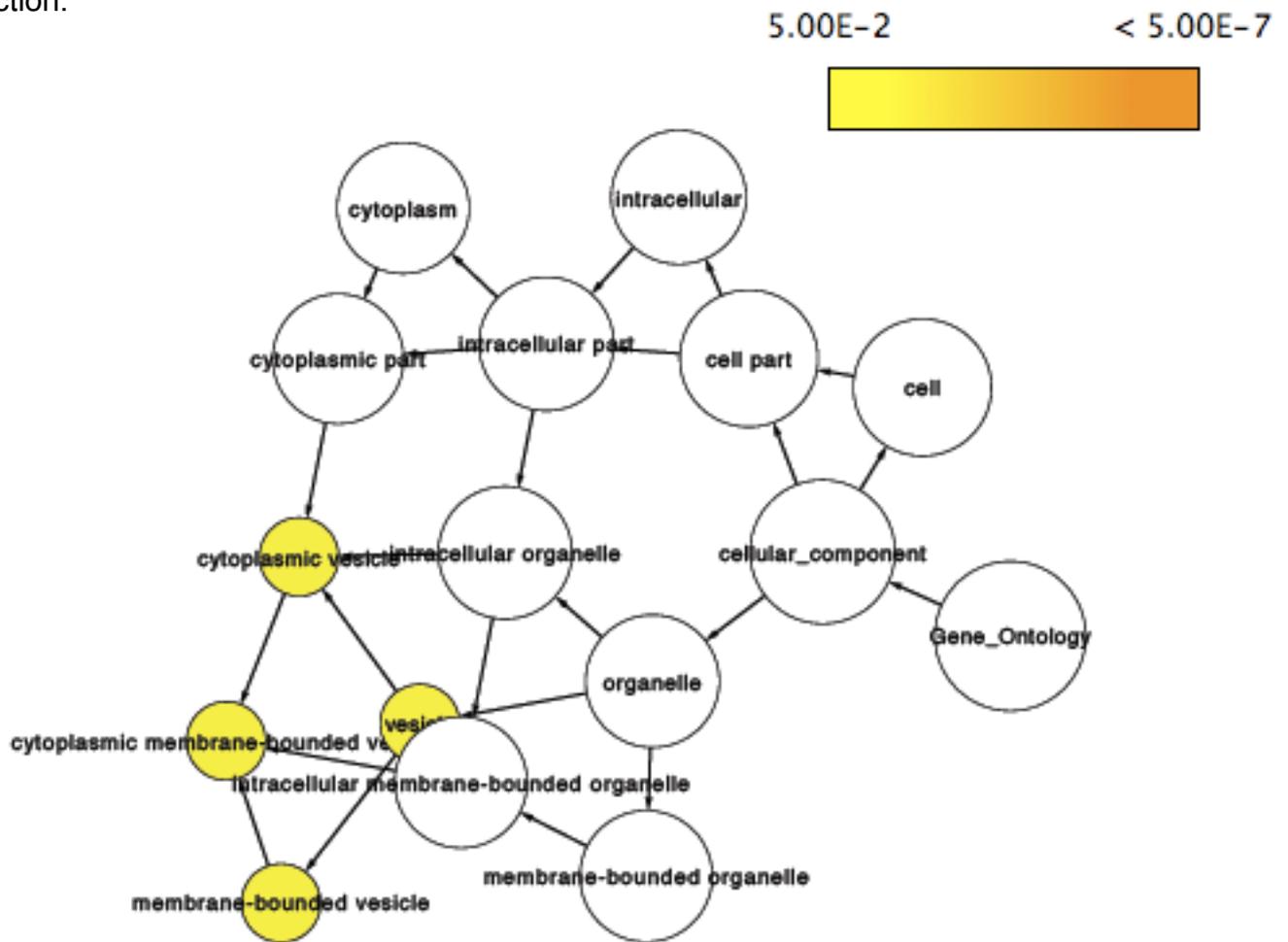
\*\*\*A.lfc gives the average logfoldchange; mi.lfc gives the minimum logfoldchange and ma.lfc indicates the highest logfoldchange (please note, that minimum and maximum is given mathematically, so in downregulated tags, the mi.lfc value indicates the strongest response). To decrease the number of relevant tags this table just shows those tags, which at least showed a tenfold increase or decrease.

considered (the exact criteria are given in chapter 2.4.3.2). An identification of over represented GO-categories was possible for all conditions except one day after infection during the incompatible interaction. The results of the GO-term analysis are illustrated in fig. 3.4.6.1-5. And in more detail on the supplementary data disc. This analysis unraveled following major processes. At three days post infection during the incompatible interaction, it can be seen, that among the significant genes, those are over represented, which are considered to have a function in vesicle formation and vesicle trafficking and in this case membrane bound vesicles are affected as well as cytoplasmatic vesicles. One day post infection during the incompatible interaction no significant overrepresentation of ontology terms has been found.

During the incompatible interaction in wild type much more active processes were observed. In summary, these processes are mainly involved in plastidic activity and carbon fixation one day after infection. Three days after infection the overrepresentation of these gene ontologies is still present and expanded by more defined processes like stroma activity. In addition some new processes were detected for instance response to stress,

systemic acquired resistance and like in the *R1*-transgenic plants, vesicle trafficking processes are active.

The ORF45 containing plant reacted comparable to the wild type plants with two major exceptions. The analysis of these plants, shows in addition two remarkably ontologies. In these plants, an enhanced response to hormones (in detail to gibberelic acid) was detected as well as the ontology term „response to fungus“ and „incompatible interaction“. Common, after the analysis of all lines in the GO-term analysis is that at three days post infection the number of overrepresented categories increased compared to one day post infection.

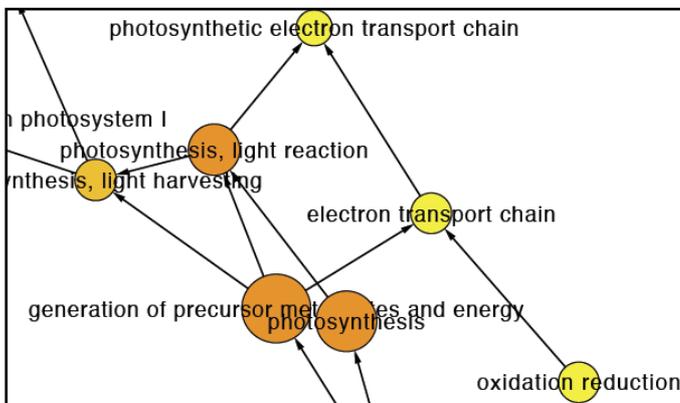
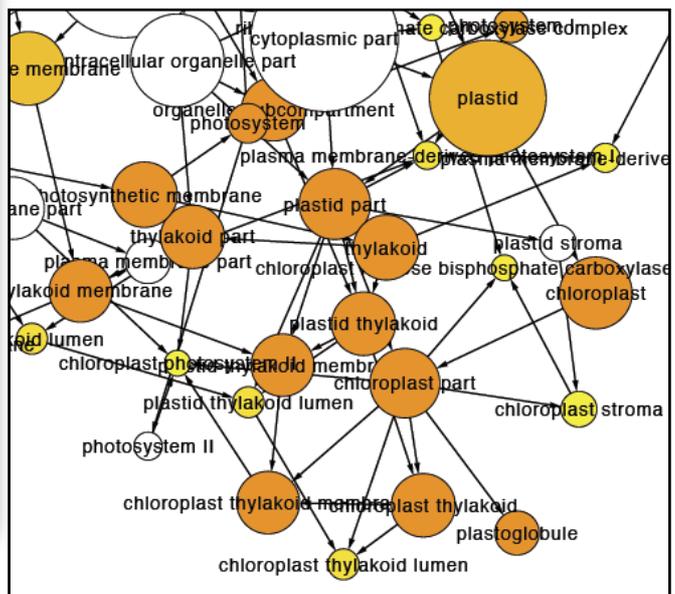
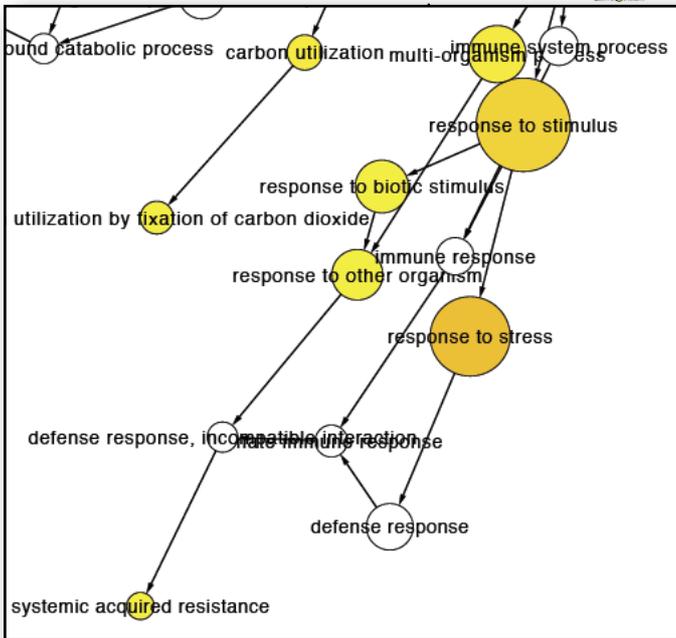
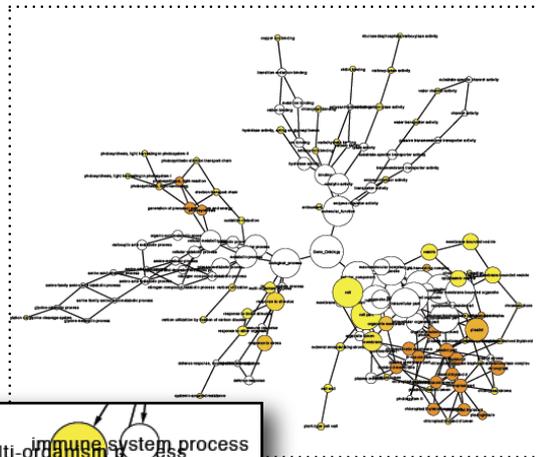


**Fig. 3.4.6.1:** Result of the GO-Term analysis of the significantly up and downregulated tags in *R1* containing plants three days after infection. The size of the circles reflect the number of tags present and the color represents the significance of overrepresentation.



5.00E-2

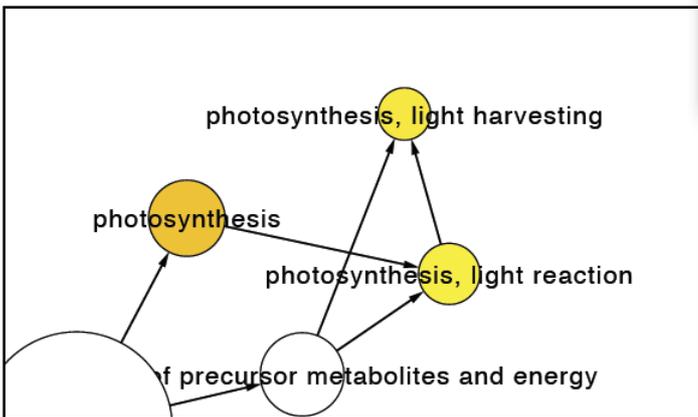
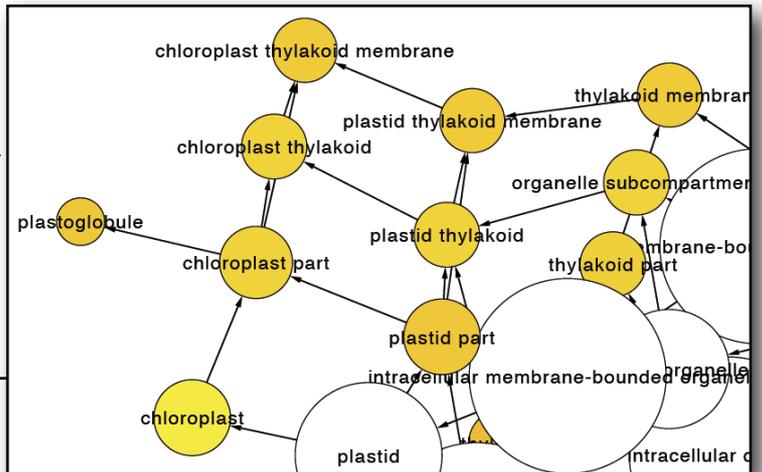
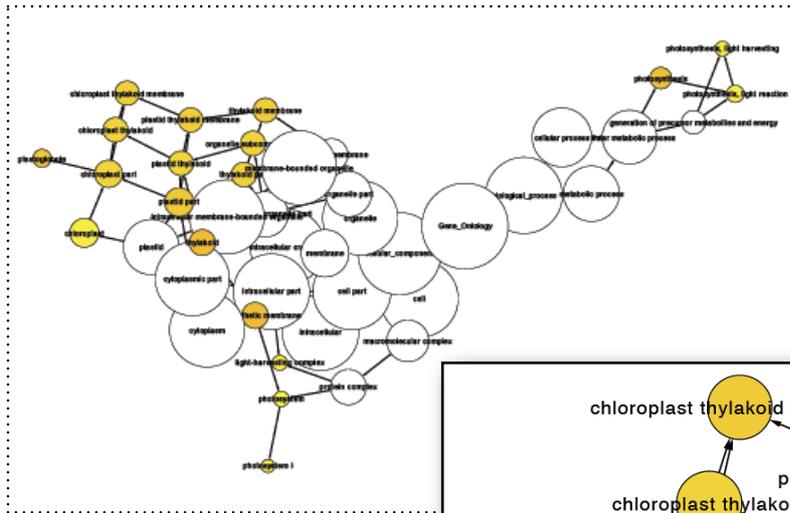
< 5.00E-7



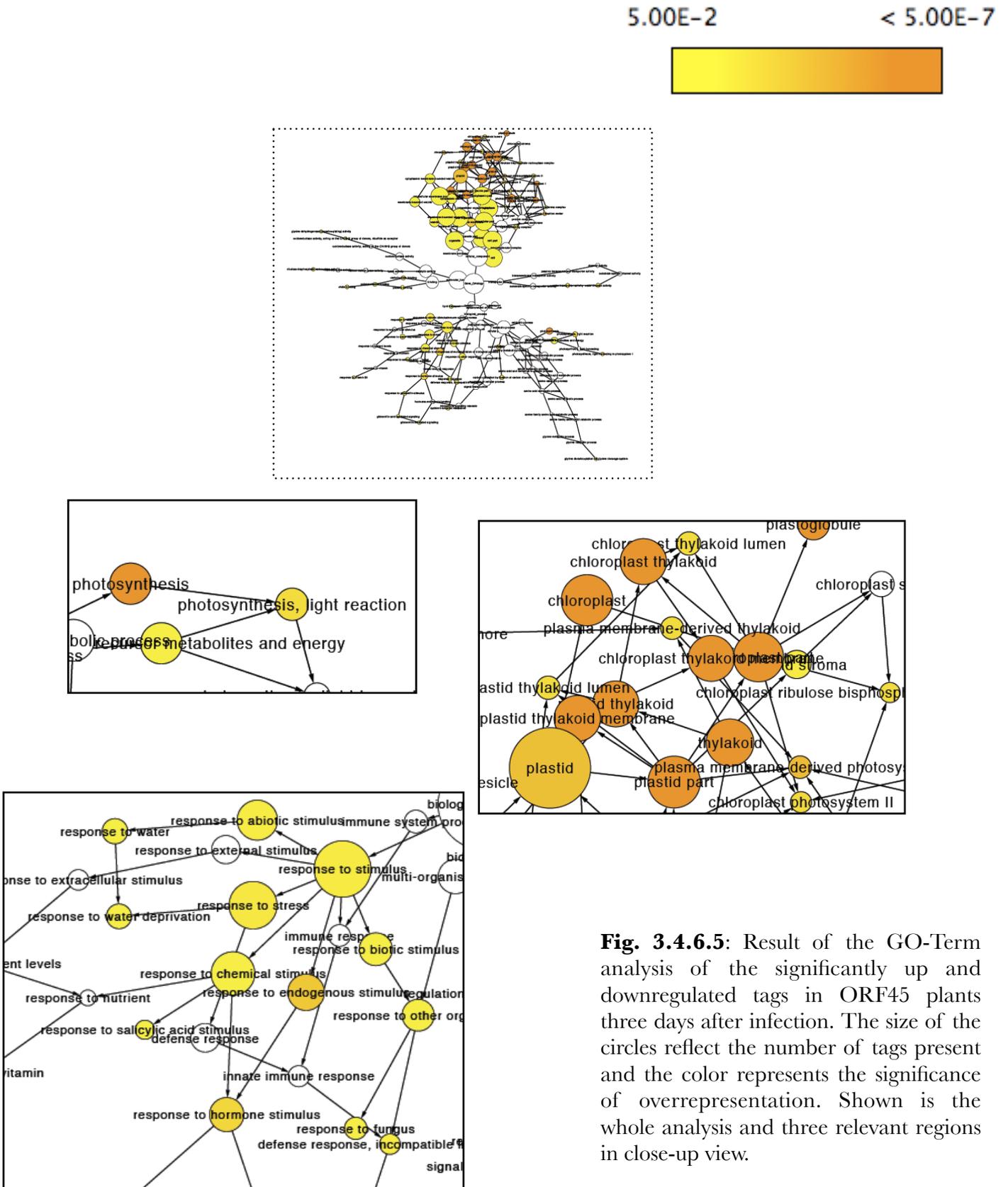
**Fig. 3.4.6.3:** Result of the GO-Term analysis of the significantly up and downregulated tags in wildtype plants three days after infection. The size of the circles reflect the number of tags present and the color represents the significance of overrepresentation. Shown is the whole analysis as overview and three relevant regions in close-up.

5.00E-2

< 5.00E-7



**Fig. 3.4.6.4:** Result of the GO-Term analysis of the significantly up and downregulated tags in ORF45 plants one day after infection. The size of the circles reflect the number of tags present and the color represents the significance of overrepresentation. Shown is the whole analysis as overview and two relevant regions in close-up.



**Fig. 3.4.6.5:** Result of the GO-Term analysis of the significantly up and downregulated tags in ORF45 plants three days after infection. The size of the circles reflect the number of tags present and the color represents the significance of overrepresentation. Shown is the whole analysis and three relevant regions in close-up view.

### 3.4.7. Comparison of DeepSAGE-data with semi-quantitative RT-PCR

A subset of 17 genes was chosen from the transcriptomics data set and the expression of this genes was measured in an independent experiment using normalized qRT-PCR. A list of this genes is given in table 3.4.7.1 and the used primers in table 2.5.1.1.

**Table 3.4.7.1:** Summary of the tags which have been examined using qRT-PCR

Tag-ID	R1 1dpi*	R1 3dpi*	wt 1dpi*	wt 3dpi*	ORF45 1dpi*	ORF45 3dpi*	prim. Annot. Name (partially)
StET013932	0,45	0,5	-0,32	0,61	0,99	0,03	Sulfoquinovosyldiacylglycerol synthase type 2
StET030529	0,22	-0,04	0,2	0,09	0,4	0,66	Splicing factor-like;
StET014686	0,81	1,01	0,45	0,35	0,35	-0,53	Auxin-repressed protein;
StET030029	0,6	0,81	-0,62	-0,33	-0,35	-0,59	Outer envelope protein;
StET014759	-0,26	-0,15	0,64	-0,47	-0,35	-0,05	SGT1-2
StET018100	-0,83	-0,62	1,28	0,5	0,29	-0,18	Proton pump interactor;
StET030331	-0,33	0,26	1,48	1,34	0,24	1,02	CBL-interacting protein kinase 25
StET011957	-0,46	-0,38	0,18	0,68	-0,13	0,3	Chromosome chr14 scaffold_9
StET007124	1,12	0,39	-0,04	-0,4	-0,58	-0,83	Cluster: T8F5.5 protein
StET025959	0,88	0,87	-0,08	0,11	-0,03	-0,61	Glutathione reductase
StET009643	0,95	0,52	0,04	-0,14	0,33	-0,2	Transaldolase ToTAL2
StET030411	0,94	0,88	0,22	-0,22	0,77	1,26	Chromosome undetermined scaffold_383,
StET021458	0,63	0,73	0,34	0,25	-0,02	0,28	GmCK3p
StET001659	0,1	0,18	0,15	0,34	-1,02	-0,75	Chromosome chr18 scaffold_1
StET001804	0,45	-0,04	-0,16	-0,6	1	-0,14	CYP72A56
StET006462	0,13	0,1	0,33	0,91	-0,26	-0,42	Cluster: Chromosome chr18 scaffold
StET029853	0,06	0,13	0,77	0,95	-0,38	0,32	Polyubiquitin;

\* describes the replica group as defined in chapter 2.6.3.2.2

The relative expression value after normalization to the housekeeping gene *ef -1a* can be seen in Table 3.4.7.2. In this table the average values of the biological and in case of the qRT-PCR the technical replicates and the according standard deviations are listed. The basis of these values are the results from the DeepSage experiments on the one hand and from the qRT-PCR experiment on the other hand. To compare the results of both methods

the average logfoldchanges were calculated and the data of both experiments was categorized in three classes. Values below -0,3 were classified as down- and values above 0,3 were classified as upregulated. Logfoldchange values between -0,3 and 0,3 were classified as “no regulatory control”. These three classes were compared and an agreement or disagreement between these classes was determined as class 1 (no agreement) or class 3 (agreement). Due to the fact of possible shifts in the infection progress and the connected shifts in expression response, the intermediate class 2 (partial agreement) was introduced. The results of this comparison are given in figure 3.4.7.1. These comparisons of methods showed just a weak overlap of the results of both experiments. The greatest consensus can be seen in the results for SGT1-2 (TC180208), a glutathione reductase (TC173455) and an auxin-repressed protein (TC173049). In these genes only one disagreement has been noticed. The rest of the tested genes, showed a less amount of overlapping expression behavior, ranging to no agreement at all in T8F5.5 protein (TC183138). In total 102 measurements were compared. Of these, 49 did not show any overlapping tendency, 14 times class 2 has been observed and 39 times congruency has been figured out. The results of this comparison are summarized as matrix in table 3.4.8.3.

**Table 3.4.7.2:** Summary of the *ef-1 $\alpha$*  normalized expression values of the qRT-PCR and the DeepSAGE experiment\*

ID	Method	wt		wt		wt		R1		R1		R1		ORF45		ORF45		ORF45	
		0dpi**		1dpi**		3dpi***		0dpi**		1dpi**		3dpi**		0dpi**		1dpi**		3dpi**	
		rel.	±	rel.	±	rel.	±	rel.	±	rel.	±	rel.	±	rel.	±	rel.	±	rel.	±
***	expr.	expr.	expr.	expr.	expr.	expr.	expr.	expr.	expr.	expr.	expr.	expr.	expr.	expr.	expr.	expr.	expr.	expr.	expr.
StET030529 TC163943	SAGE	0,12	0	0,17	0	0,15	0	0,07	0	0,32	0,2	0,26	0,1	0,11	0,1	0,13	0,1	0,09	0
	RT-PCR	1,4	1	0,96	1	1,3	0	1,48	0,3	0,85	0,7	0,97	0,8	1,35	0,8	1,4	0,1	1,76	0,5
StET014686 TC173049	SAGE	0,19	0	0,22	0	0,2	0	0,2	0,2	0,45	0,1	0,06	0,1	0,03	0	0,12	0,1	0,25	0,2
	RT-PCR	0,99	1	0,99	1	1,49	0	1,32	0,3	1,32	1	0,85	1	1,09	1,1	1,41	0,1	2,91	0,6
StET030029 TC168403	SAGE	0,08	0	0,03	0	0,05	0	0,04	0	0,04	0	0,01	0,1	0,02	0	0,04	0	0,08	0
	RT-PCR	1,05	1	1,26	1	1,52	1	1,38	0,7	0,89	0,6	1,14	0,6	1,61	0,3	1,66	0,2	1,95	0,2
StET014759 TC180208	SAGE	0,02	0	0,05	0	0	0	0,01	0	0,01	0,1	0,01	0	0,01	0	0	0	0	0
	RT-PCR	0,37	1	0,81	1	1,61	1	1,35	0,7	0,62	1	0,93	1,1	0,78	1,1	0,6	1,4	1,33	0,9
StET018100 TC190958	SAGE	0	0	0,06	0	0,03	0	0,01	0	0,19	0	0	0,1	0,03	0	0,01	0	0,01	0
	RT-PCR	1,07	1	1,82	1	1,41	0	1,11	0,8	2,25	0,8	1,4	0,9	1,27	0,4	1,57	0,3	2,82	0,6

StET030331	SAGE	0	0	0,08	0	0,08	0	0,01	0	0,02	0	0,09	0,1	0,04	0,1	0,01	0	0,04	0
TC173012	RT-PCR	1,73	1	1	1	1,44	1	1,03	0,2	1,35	0,3	1,09	0,4	1,43	0,3	1,71	0,3	2,29	0,2
StET011957	SAGE	0	0	0,01	0	0,03	0	0,01	0	0,01	0	0,04	0	0,04	0	0,01	0	0,02	0
TC172861	RT-PCR	0,93	1	0,83	1	0,96	0	0,63	0,3	0,86	0,3	0,7	0,1	0,93	0,1	1,03	0,1	1,57	0,2
StET007124	SAGE	0,09	0	0,03	0	0,04	0	0,06	0	0,01	0	0,02	0,1	0	0	0,05	0	0,02	0
TC183138	RT-PCR	1,42	1	0,86	1	1,21	0	1,51	0,4	0,85	0,6	1,62	0,7	1,5	0,7	1,56	0,3	2,3	0,5
StET025959	SAGE	0,07	0	0,07	0	0,1	0	0,11	0	0,11	0,1	0,03	0,2	0,02	0	0,06	0	0,06	0
TC173455	RT-PCR	0,64	1	1,21	1	2,5	1	1,52	1,4	0,3	1,3	0,21	2,9	0,7	3,3	2,36	4,6	2,09	0,4
StET009643	SAGE	0,02	0	0,02	0	0,02	0	0,02	0	0,07	0	0,01	0	0,01	0	0,05	0	0,03	0
TC165331	RT-PCR	1,62	2	1,13	1	1,51	1	1,47	0,4	1,57	0,2	1,37	0,2	1,6	0,2	2	0,3	1,96	0,2
StET030411	SAGE	0,04	0	0,04	0	0,06	0	0,06	0,1	0,55	0	0,21	0	0,01	0	0,08	0	0,09	0,1
TC184597	RT-PCR	1,64	1	0,75	1	1,34	0	1,47	0,2	0,84	0,5	1,16	0,5	1,37	0,3	1,32	0	1,79	0,4
StET021458	SAGE	0,42	0	0,34	0	0,42	0	0,46	0,2	0,56	0,2	0,5	0,3	0,07	0,1	0,18	0,1	0,26	0,3
TC174692	RT-PCR	1,53	1	0,63	1	2,27	0	1,76	0,3	0,83	1,1	1,41	1,1	1,5	0,5	1,24	0,6	2,49	0,6
StET001659	SAGE	0	0	0	0	0,01	0	0,08	0	0,01	0	0,02	0	0,01	0	0,01	0	0,01	0
TC170569	RT-PCR	2,12	2	1,26	1	1,28	1	1,95	0,5	1,09	0,4	1,07	0,4	1,68	0,4	2,03	0,2	1,95	2,6
StET001804	SAGE	0	0	0,03	0	0,004	0	0,03	0	0,03	0	0	0	0,01	0	0,1	0,1	0	0
TC190995	RT-PCR	1,81	2	1,08	2	2,11	1	2,55	0,6	1,08	1,1	1,38	1,2	1,96	1,1	1,49	0	1,95	1,1
StET006462	SAGE	0,04	0	0,09	0	0,25	0	0,04	0	0,19	0,1	0,05	0,4	0,01	0	0,01	0	0,01	0
TC173953	RT-PCR	1,19	1	1,04	1	0,89	1	1,16	0,6	1,68	0,4	1,22	0,2	1,76	0,3	2,15	0,5	2,93	0,3
StET029853	SAGE	0,04	0	0,12	0	0,15	0	0,09	0	0,08	0,1	0,17	0	0,06	0	0,07	0	0,09	0
TC182394	RT-PCR	2,68	2	1,22	2	1,91	2	1,08	0,3	4,62	0,4	1,62	0,5	1,02	0,5	1,72	0,3	1,6	0,5

\* given are the average expression values (rel. expr.) and the standard deviation ( $\pm$ ) of the different genotype classes and timepoints.

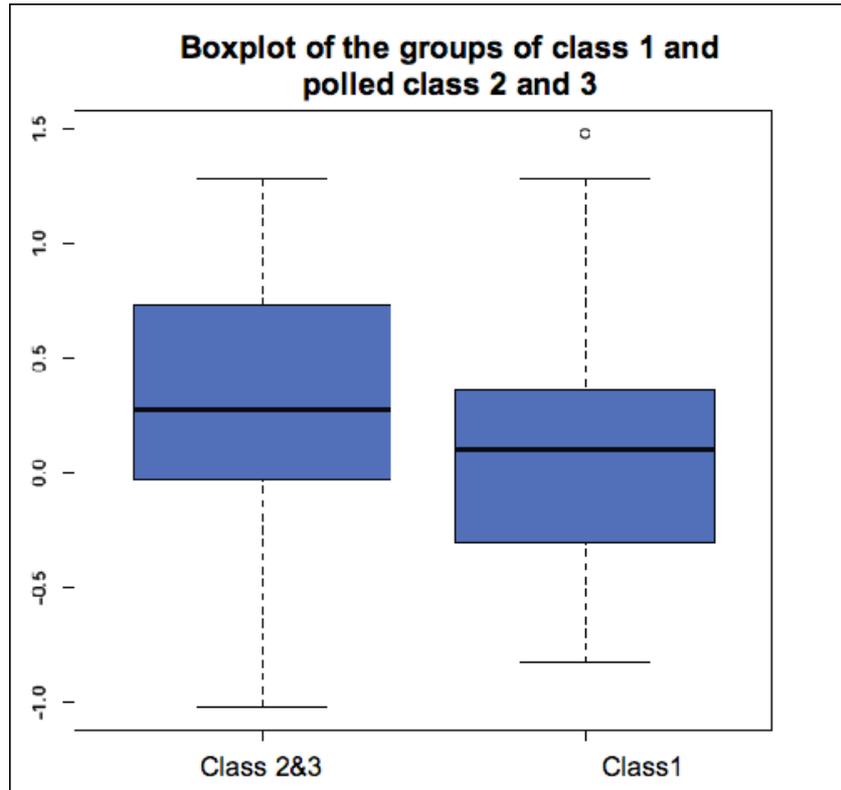
\*\*describes the replica group as defined in chapter 2.6.3.2.2

\*\*\*data originating from the DeepSAGE experiment is given in rows named as SAGE; adatas from the qRT-PCR experiment in given the row named as RT-PCR

A comparison of values from the DeepSage experiment of the class 1 and those in class 2 and 3 showed that the logfoldchange values of both groups differ significantly ( $p=0,0498$ ). As can be seen in the boxplots of both groups (fig. 3.4.7.2), the median and both quantile-values show a shift towards smaller values in the group, which leads to an assessment of class 1.

ID	TC-ID	R1 1dpi	R1 3dpi	wt 1dpi	wt 3dpi	ORF45 1dpi	ORF45 3dpi	Annotation
StET013932	TC176096	1	1	1	1	2	3	Sulfoquinovosyl diacylglycerol synthase type 2
StET030529	TC163943	3	3	3	3	1	3	Splicing factor-like
StET014686	TC173049	2	3	3	3	3	1	Auxin-repressed protein
StET030029	TC168403	3	1	1	3	1	1	Outer envelope protein
StET014759	TC180208	1	3	3	1	3	3	SGT1-2
StET018100	TC190958	1	1	1	1	3	3	Proton pump interactor
StET030331	TC173012	1	3	1	1	1	1	CBL-interacting protein kinase 25
StET011957	TC172861	1	3	3	1	1	1	Chromosome chr14 scaffold_9
StET007124	TC183138	1	1	1	1	1	1	T8F5.5 protein
StET025959	TC173455	3	3	3	2	1	3	Glutathione reductase
StET009643	TC165331	1	1	3	3	1	1	Transaldolase ToTAL2
StET030411	TC184597	1	1	2	3	1	1	Chromosome undetermined scaffold_383
StET021458	TC174692	1	1	2	3	1	3	GmCK3p
StET001659	TC170569	3	1	3	1	1	1	Chromosome chr18 scaffold_1
StET001804	TC190995	1	3	3	3	2	2	CYP72A56
StET006462	TC173953	3	3	1	1	1	3	Chromosome chr18 scaffold_1
StET029853	TC182394	3	3	1	1	2	2	Polyubiquitin

**Fig. 3.4.7.1:** Overview of the different classes of reproduction under different experimental conditions. The columns reflect the identifiers of the different tags of the tested genes in RT-PCR by StET-number, TC-number and partially the annotation is given, as well as the degree of overlap in results is given by the colored numbers. The red number one represents no consensus, the yellow two indicates a conformity in one of the measured timepoints in a given genotype and the green number three reflects agreement in the measurements. Each row reflects the results of one tested gene.



**Fig. 3.4.7.2:** Boxplots of the logfoldchange values from the DeepSage experiment, which lead to a classification 2 or 3 and of those values, which lead to a classification 1.

## 4. Discussion

### 4.1. Evidence for the expression of predicted *R1*-homologous genes in *Solanum tuberosum*

The first step to analyze the *R1*-homologues on the contigs *R1* and *r1* was to generate locus specific primer pairs. This was successful for ORF23, ORF46, ORF 22-1, ORF52 and ORF54 (chapter 3.1). Additionally it was possible to develop one primer combination, which amplifies ORF24 as well as ORF45 (table 2.5.1.1). Together with the existing *R1*-specific primers (table 2.5.1.1), it is now possible to examine the expression of most of the putative *R1*-homologous genes. This was demonstrated by the occurrence of the correct PCR-bands on BAC's, which only harbor the target sequence. In the cases analyzed, no cross-amplification using the PCR-assay was observed. The reason to develop those primers, was given by the high degree of homology of this genes. This high similarity among this genes forced a BAC-assay driven verification. Even with sequence information it was not possible to distinguish completely between all of the members of this family. Moreover in the cases of ORF 22-1 and 46 it was necessary to amplify just small fragments for the specific detection of these genes. This additionally smallered the methodological power of Sanger sequencing. Only in cases of ORF22, the development of those primers failed. Using cDNA of the diploid breeding clone *Solanum tuberosum* P6/210 it was possible to detect transcripts of some *R1*-homologues. This was possible in the cases of ORF22-1, 23, 46, 52 in leaf tissue in different developmental stages and in floral tissues. In the case of ORF54 only in leaf tissue a transcript was detected (chapter 3.1). Although the experiments have been designed a in semiquantitative manner, the interpretation of the results should remain qualitative with respect to the amplification efficiencies (75%) of the used primers (data not shown). Nevertheless in the given cases it was possible to detect transcripts, indicating that the ORF prediction was correct.

## **4.2.Evidence for the expression of genomic predicted ORFs in *Solanum nigrum***

The verification of transcripts in *Solanum nigrum* was successful in the case of three putative *R1*-homologous genes. The results in chapter 3.2 showed the possibility to detect transcripts in cDNA samples for the homologues *snR1.5*, *snR1.6* and *snR1.7* (chapter 3.2). Especially *snR1.6* which shares the highest sequence similarity to the *R1* gene is a good candidate as late blight resistance gene in *Solanum nigrum*.

## **4.3.Hypersensitive reaction of *Solanum nigrum* in response to transient expression of effectors of *Phytophthora infestans***

The transient expression of cloned effectors of *Phytophthora infestans* in *Solanum nigrum* showed a clear hypersensitive response after infiltration with the effectors, which have been presented in chapter 3.3.

This is a decisive hint to answer the question, whether the recognition of *Phytophthora infestans* in *Solanum nigrum* is R-gene driven. This information is derived from the localization of the expressed effector protein. In PAMP-triggered mechanisms, the pathogen is recognized before effectors are secreted in the cytosol and as far as known, the PAMP-receptors are localized to the plasma membrane [105, 106]. The phenotype induced by these receptors is not differentiable from the response of NBS-LRR type R-genes. In both cases, a hypersensitive response is observed. In this transient expression system, the PVX-based vector components are transferred to the plant cell and with them the effector sequence. The expression of the effector occurs directly inside the cell. This is crucial for an assessment of this process and allows a differentiation of both processes.

Although these results do not exclude a PAMP-triggered immunity system in *Solanum nigrum*, they give a strong argument for the existence of an internal R-gene driven mechanism, which is able to recognize effectors from *Phytophthora infestans*. A remarkable feature of the positive effectors is their high degree of similarity to each other on the one hand and the high degree of similarity to the *avrblb2* family effectors. So it is possible, that *Solanum nigrum* as well as *Solanum bulbocastanum* are able to target a class of molecules, whose structure is important for the pathogen and mutations are

therefore under higher selection than in other avr factors.

Another possibility, which is always thinkable in cases of possible resistance genes from wild species, arises from the ecology. In natural populations there is a complex mixture of a resistant and non-resistant environment. A mutation, which is physiologically not beneficial for the pathogen, but leads to recognition of a new hosts, might not be necessary, due to the fact, that nearby a recognizable host is present - this decreases the selective pressure of the environment. This ecology obviously drastically changes after the introgression of a certain resistance gene into a crop monoculture. In such an environment the adaptive pressure on the pathogen is much higher what following leads to a forced adaption to overcome the resistance even under physiological non-beneficial conditions due to the lack of alternatives for survival of the local pathogen population.

During the comparative analysis of the effector sequences it has been observed that clones were included in this set, which did not induce a necrosis but had the same translated amino acid sequence to clones which induced the hypersensitive response. This indicated that these experiments do not allow interpretations in form of exclusions. For a clear identification of sequence motives, which are responsible for a recognition of the effector this distinctive criterion is a prerequisite. It has to be concluded that this effector-set contains clones, which do not work properly and therefore does not allow to interpret negative results. The observation of positive necrotic phenotypes supports the conclusion that in *Solanum nigrum* a R-gene driven resistance mechanism exists, which is able to perceive avirulence factors of *Phytophthora infestans*. This mechanism should be at least in part highly similar to the one of *Solanum bulbocastanum*. To which extend this mechanism correlates to the identified *R1*-homologues has to be clarified by further studies.

## **4.4.Comparative DeepSAGE-analysis of *R1*- and ORF45-transgenic *Solanum tuberosum* plants with wildtype**

### **4.4.1.DeepSAGE as method for transcriptome analysis without a reference genome – performance features and uncertainty factors**

The application of DeepSAGE as massive parallel sequencing supported method for transcriptome analysis was successful. With an annotation of almost 20 % of the potato gene index, a large number of transcripts have been detected and identified. It was possible to annotate 69 % of all sequence tags (chapter 3.4.2). There are still many sequence tags which currently lack any information but can be a valuable source in the future, when such information is available. The precision of annotation, in general was satisfying. 69 % of all annotated tags, had just one possible target inside the used data set of reference sequences. Due to the heterozygosity in *Solanum tuberosum*, it was necessary to allow one mismatch. The number of tags with 100% match, with 10323, was half of the total number of annotated tags (chapter 3.4.2). It should be mentioned that a prerequisite for reaching this annotation-rate was the precise definition of the reference sequences and the tag identification via a tag-extraction-and-matching algorithm (chapter 2.6.3.1). Attempts, to annotate against genomic sequences or using the BLAST algorithm often resulted in various numbers of possibilities and random hits (data not shown).

The currently available TIGR (JVICI)-cDNA microarray chip of potato contains 32448 spots but only 16058 have an annotation [107]. After removing the duplicates 5344 are left. In this context a quality assessment of the DeepSage experiment can be made. With an identification of more than 12000 genes, the performance in terms of information is more than doubled. Additionally 30 % of the sequenced tags, have not been identified, yet. A new annotation of these tags with new sequencing information can easily be done in the future.

One special observation was made during this experiment, which arose from the new technical possibilities. This was the matching of multiple tags on one target. In this quality this was not anticipated from microarray experiments before and will make a discussion of comparability to the hybridization driven data necessary. Although hard to examine

systematically in the given data set, it was noticed that several tags have been annotated more than once, with the extreme example of 26S protease subunit 6 given in chapter 3.4.2. The reason for this multi-tag-families are on the one hand allelic variants of one gene, which might have been detected and on the other hand, gene families where several genes encode homologous transcripts and which are very similar in sequence. In comparison to hybridization and PCR-based methods, it is hard to imagine that these methods are able to reach this resolution at a single base. The quantitative sequencing have the potential to detect such differences. But with this new insights and the new possibility to analyze these findings it is necessary to think about possibilities how this problem can be treated. In our analysis, we decided to use each single tag and its quantitative behavior for searching significant differences in expression. Afterwards, the annotation information has been taken into account - leading for example to a number of upregulated tags, which coded for the same gene. On the contrary, it was observed that a gene was detected by significantly and as well by not significantly regulated tags.

An alternative approach for the analysis could be the pooling of the expression level measurements of annotated EST's and subsequent analysis of levels for significance of pooled observation. This new insights can also be seen as uncertainty factor in the sense that the complete secure allocation of these tags remains open as long as no reference genome is available. Even then, this genome has to be from the same cultivar to avoid ambiguities caused by the allelic states.

In addition no broad experience concerning the sequencing accuracy exists. In studies, where alleles have been cloned and sequenced by traditional Sanger sequencing, only those were accepted, which were detected at least two times [108]. The same criterion is used in the acquisition of massive parallel sequencing and in this approach this criteria has been expanded to a threefold detection. It is not clear yet, whether this is really sufficient. Not much is published about the general error rate of this sequencing technique, and only first approaches to this question exist [109]. Further evidence has to be obtained by a broader application of this methods and the corresponding sequencing platforms.

Nevertheless, comparing our general results to other studies, an overlap can be noticed. One of the published experiments, which fits best to our approach has been done by Restrepo et al [110] and achieved comparable results in terms of the identification of protease inhibitors or the Glutathione reductase [110]. Even in silico modeling approaches showed an indication in this direction [111]. But beyond the rather general results which

are presented in these studies, we find a much more complex situation in our experiment. For example, through the identification of ethylene responsive proteins we have good hints for the involvement of this hormone in the defense response (this will be discussed in the following chapters in more detail). This is in accordance with studies observing *Oryza sativa* upon infection with the bacterial pathogen *Xanthomonas oryzae* [112]. It has been observed, that different genes, which respond to ethylene are showing expression changes whether in the wild type plants or in *R1* transgenics.

#### 4.4.2. Evaluation of statistical methods – the normal distributed world

The comparison of different statistical tests for significance gave surprising results. On the one hand, it turned out that tests, which have been originally designed for the analysis of traditional SAGE-libraries did generate a high number of significant tags in the FDR-value estimation and further criteria were needed for correction (details are given on the supplemental data disc). In case of the G-test this is achieved by the multiple criteria, which are applied like the in-group significance for example and in the case of Fisher's exact test, we introduced modifying criteria manually (see chapter 2.6.3.4).

Another interesting observation was made when comparing the top 100 significant tags of the different tests with each other. In contrast to expectations, there was almost no overlap between the results of the tests.

This is additionally confirmed by the calculation of the correlation coefficients based on the expression differences of the significant tags. As it seems, each test evaluates the behavior of the data values completely different. There was almost no correlation, just a weak tendency of overlap for the results of BaySeq and edgeR, and between Permu and Fisher's exact analysis. Nevertheless, for an interpretation of this result it has to be considered that these tendencies just describe the nearest similarities in a group of unequal ones.

Another interesting feature of the G-test and Fisher's exact test was the correlation to the logfoldchange values. There was no clear decrease in the expression change with increasing FDR-value. This, in contrast, was observed when applying Student's t-test. These results followed the intuitively expected correlation of lower expression changes with increasing FDR-values. The hypergeometric tests do not follow this logic. The G-test decided in extremes. As it seems, the results fall mainly in the category „yes“ with

extremely low FDR-values or the significance category „no“ with a FDR-value of 1. Even more remarkably is the fact that the tags with the highest expression changes fell were assessed as not significant. These characteristics were similar in the results of Fisher's exact test although the scattering of logfoldchange values between 0-1 was larger and the number of raw FDR-values below 0,05 was reduced. This correlation was different in the result of Permu, edgeR and BaySeq. These tests showed a similar tendency in correlation like the t-test.

Concerning the differences in the significant genes by each test, the results of the G-test were not coherent. In both assessment criteria, the characterization via the logfoldchanges and the differentiation factor (via the linear regression), it turned out that those tags which have been evaluated as significant by the G-test showed the least differences compared to those calculated by all other tests. In addition, among the significant tags, it was observed that some expression differences were not conclusive (data not shown, please see the supplemental data disc). This leads to the conclusion that the FDR-value estimations made the G-test is not applicable on such large-scale data.

In contrast, EdgeR and BaySeq showed the greatest pairwise differences within groups of significant tags, and it is likely that these tests are more suitable for the significance analysis of massive parallel sequencing data. At the moment, a disadvantage of these test is their exclusive availability as R-package. The software R demands certain computational skills. Another disadvantage is that these tests (this includes the Permu test as well) are computationally intensive and need corresponding computer capacities. The permutation test, although the results are good especially in terms of homogeneity, has a special problem with the calculated p-values. It is not really clear to which extend the estimated p-values are comparable to those from other tests. One problem affecting this, is the dependency on the group size. As test, identifying primary homogeneous groups against each other, it is more difficult to achieve a similar result the more replicates are present within a group. Finally, it was necessary to introduce a group-size correction factor to achieve a comparability. This factor was introduced intuitively and it is not clear to which degree this affects the total assessment. One origin of this problem is the low number of diverse p-values resulting from this test. In total, only 462 different values have been calculated. This influences the subsequent FDR-correction after Benjamini & Hochberg drastically because this method has a ranking factor included as well which compares the ranking position of the p-value to the number of all values [103]. This is driven out in the

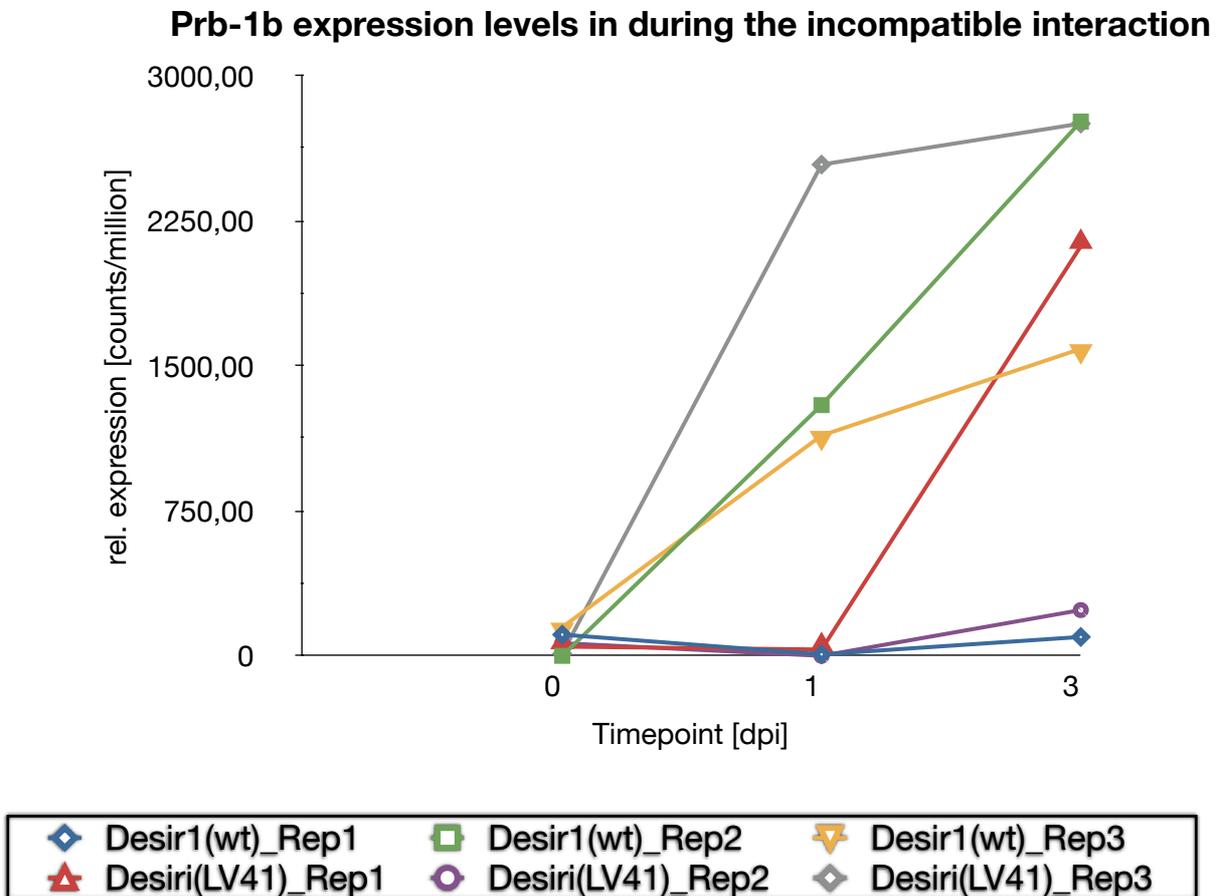
manner that the lowest p-values are less corrected than the highest. If now blocks of equal p-values exist, this influences the ranking and therefore the results of the correction method. Until this problem is solved, it is recommended to interpret the results in the way of ranking lists.

According to this and with similar characteristics in expression differences within the group of significant comparisons as observed in the permutation test, we decided to use the Fisher's exact test statistics for our main analysis. However, this was only possible with the arrangement of replicates in this study. The test procedure was performed in the way of independent experiments. Independent time courses were compared. Whether this kind of significance statistic is suitable for other experiments, in which independent groups are compared, is doubtful.

#### **4.4.3. Validation and characterization of the early stages of infection as phenotype**

Confocal microscopy showed, that 3 days post inoculation the symptoms of infection, are not very distinctive in the given combination of pathogen and host. After three days it became visible that a true infection had occurred although the internal state can not be figured out (see fig. 3.4.1.1). These results correspond to the observation of an almost invisible phenotype during the infection experiments at three days after infection (data not show). On the expression data, this can be correlated to the results of the principal component analysis. Although a gradient in the principal component 1 over the time of infection was detectable, the splitting of the samples and their grouping was not very strict. (see fig. 3.4.5.3) This indicates, that the observed time of the initial phase of the infection reflects a highly variable phenotype, in which environmental factors, which cannot be excluded completely under the artificial conditions of climate chambers, play an important role. Eight days after infection, we were able to detect an increased amount of *Phytophthora infestans* in the samples, especially compared to the incompatible interacting genotypes (see table 3.4.2.1). This PCR-test provides evidence that the plant, which has been used as samples has been infected successfully. However, the individual progress in the early stages was not clear and not measurable without influencing the transcriptome response of the sample. One factor, which can not be influenced, is the local physiological strength of each leaf. This observation is supported by having a closer look

at the expression levels of the gene *Pr-1b*, which was chosen as indicator due to the shown upregulation of PR-1 genes during the compatible interaction [113, 114]. The data of this gene shows a strong expression change in four of six replica, a weak in one and no



**Fig. 4.4.3.1:** Expression levels over time of the tag StET002234. This tag is grade A-annotated for the gene *Prb-1b* (TC169959).

clear visible in another one. This is a good indicator for the variability of infection processes during the initial phase which had to be taken into account for the interpretation of the generated data.

#### 4.4.4. Using Cluster analysis as tool to find coregulated units

The Cluster analysis was performed to find groups of coregulated genes. It was desirable to find groups of genes, which show specific tendencies under the different conditions. This was possible in two approaches. In the non-hierarchical k-mean cluster analysis, groups were identified, which show a trend towards a common response. For both *R1-*

transgenic lines the number of specific clusters was quite low. This observation can be seen as an alternative expression of the results of the PCA, which already indicated that the inter-line differences are rather low compared to the general transcriptomal activities during the infection events. In this chapter four clusters from the k-mean clustering which show specific tendencies will be presented in more detail. With Cluster 79 and 9, two small clusters were identified, which show a different expression pattern in the *R1*-transgenic lines. Cluster 9 shows a specific upregulation and Cluster 79 a downregulation at one and three dpi. Cluster 145 shows the strongest specific upregulation in the wild type plants and Cluster 114 contains genes, which show a specific downregulation in the line harboring the ORF45.

After having a closer view on the tags present in the clusters, it can be seen, that in Cluster 9, for 8 tags GO-information was available (data not shown, please see the data browser for details). Among these were terms like „vesicle“, „nucleic acid binding“ or „Golgi-apparatus“ (GO: 31988, 3676 and 5794).

Besides this, two genes involved in ethylene signalling were within this cluster. One is presented by the tag StET007682 and is matched to TC175466 as Ethylene-responsive transcription factor 7. The other (StET002032) is matched to TC184207 as EIL3 (ETHYLENE-INSENSITIVE3-LIKE3). This gene first has been described in *A. thaliana* in the context of the Sulfur metabolism and in mutants reduced glutathione levels have been detected [115].

The cluster 79 with a specific behavior for the *R1*-transgenic lines did show an involvement of multiple GO terms, among them genes involved in hormone signalling, especially with StET021234, a gene, which is annotated as TC16055 (1-aminocyclopropane-1-carboxylase oxidase). The product of this gene is listed as catalyzing the reaction mechanism EC1.14.17.4. This reaction is present in two reference pathways in the KEGG database [116]. One is the „cysteine and methionine metabolism“ (ec00270) and the other one is „Biosynthesis of plant hormones“ (ec01070). The reaction is directly involved in the synthesis of ethylene in plants. A reduction of this enzyme would lead to a decrease of the ethylene levels.

This would support the hypothesis of a crosstalk between Absidic acid (ABA) and Ethylene during infection events [117] although genes which are known to response specifically to Absidic acid were not identified.

The Cluster 147, which contains tags which show a specific upregulation in wild type

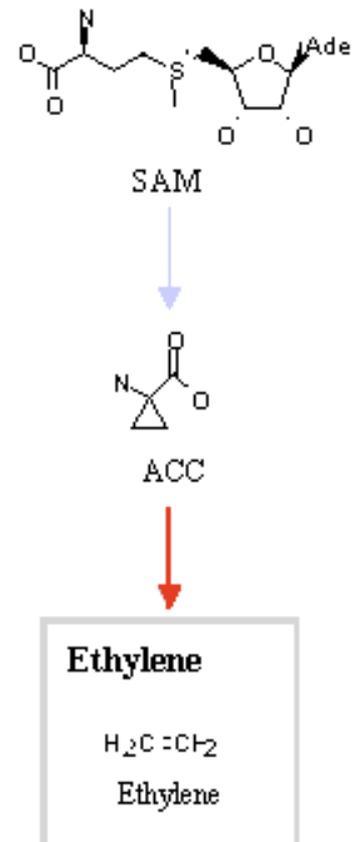
plants, is one of the largest clusters with a large number of identified gene ontologies. Among others, hormone related genes were identified. With reference to CV504466, an Ethylene-responsive small GTP-binding protein was observed (StET017849) as well as TC174667, a gene (annotated as „Salicylic acid-induced protein 19“). Both genes are referred to regulating genes upon influence of the hormones ethylene or Salicylic acid (source: Uniprot [118]). The first gene contains a Ras-domain and might be involved in phosphorylation events. The other gene contains a DNA-binding domain and is listed as regulator of transcription (source: Interpro).

Another gene (StET008480) has been identified as TC164814, annotated as Centrin, a member of the Ef-hand Superfamily (source: Interpro).

The Cluster 114, in which the wild type is upregulated and ORF45 is downregulated contains a broad number of gene ontologies (data not show, please see the supplemental data disc). Additionally to more general ontologies, in this cluster „chromatin binding“ is present. This is due to the presence of StET019390, which led to the identification of TC16680. The annotation for this tentative consensus sequence is F6A14.10

protein. This protein is a member of the nucleosome assembly protein family. This indicates chromatin remodeling events occurring during the infection in this transgenic line. After pooling the genes from 18 clusters and implementing a hierarchical clustering, it was possible to detect smaller, but more clearly identifiable groups. One has been detected to be upregulated in in all genotypes upon infection.

This group contains mainly genes, which are traditionally described as being upregulated upon infection events [119], for instance, genes encoding WIN1 and WIN2 (StET029943 and StET016648). The Uniprot database predicts a chitinase activity for the proteins encoded by the transcripts. It is likely that these genes belong to the group of chitinases, which are likely involved in the general antimicrobial response. In addition, with PR-10 and PR-1 pathogenesis related genes were present. One gene, which responded most clearly, was annotated as lipid desaturase and has been shown to be upregulated after virus



**Fig. 4.4.4.1:** A selected-step of the pathway ec01070 from the KEGG database. The red arrow shows the putative reaction catalyzed by the product of TC16055.

infection in tomato [120].

In the incompatible interaction some genes were found to be jointly downregulated. Among them is a tag annotated as *Hym-1*. This gene originally has been characterized in *Hypericum perforatum* (St. John's wort) [121]. As has been shown biochemically, the product of this gene is able to catalyze a conversion step in the biosynthesis of the secondary metabolite hypericin. An occurrence of this substance in potato is not known.

Another gene belongs to the family of PAL-genes (phenylalanine ammonia-lyase) which have been widely characterized [122] in potato and are upregulated upon wounding and similar stress conditions in potato tubers [123, 124]. Interestingly the transcriptomics experiment shows a clear downregulation during the incompatible interaction. Similarly surprising is the detection of two members of the *avr/cf-9* gene family among the downregulated genes. These genes have been shown to be upregulated during infection of tobacco with *Cladosporium fulvum* [125]. However for a comparison with our results, it has to be kept in mind, that in this study different timepoints have been observed. The used material in the referred study consisted of elicitor induced cell-cultures. So it is doubtful whether these findings reflect the natural situation similar to the observations in the natural interaction system like it has been studied in these experiments.

A third gene is annotated as „Patatin-like“. In addition to the known function as storage protein, this class has been initially found in *Nicotiana tabacum* and connected with phospholipase activity. An accumulation upon infection with Tobacco Mosaic Virus has been detected [126].

About the other members identified in these cluster arms, no clear hints on studies with pathologic background were given. Taking these findings together, we have to notice that during the incompatible interaction experiments, we identified four genes as downregulated, which have been published in similar experiments as upregulated. The only difference is the interaction - in the case of the PAL genes, which is the most comparable example, we find an accumulation during the compatible interaction.

Interestingly, upon the response in *R1*-transgenic plants it is not only the case, that a member of this family is not responding. It has been found that the reaction is the opposite, indicating that the role of the PAL gene family in pathogen interaction is not fully explored, yet.

Among the downregulated genes in wildtype plants, a carbonic anhydrase was detected, supporting the results of Restrepo et al [110] who had comparable results and even

demonstrated a role of this enzyme in the compatible interaction by transient silencing experiments [110]. Additionally two proteinase inhibitors were identified. This class of proteins is involved in the interaction process and targeted during the reprogramming process by the pathogen [35, 127-131]. A gene which is annotated as pectate-lyase is also present in the clusters. These genes have a role in cell wall modifications. Studies on Soybean roots upon inoculation with nematodes demonstrated expression changes in response to infection [132].

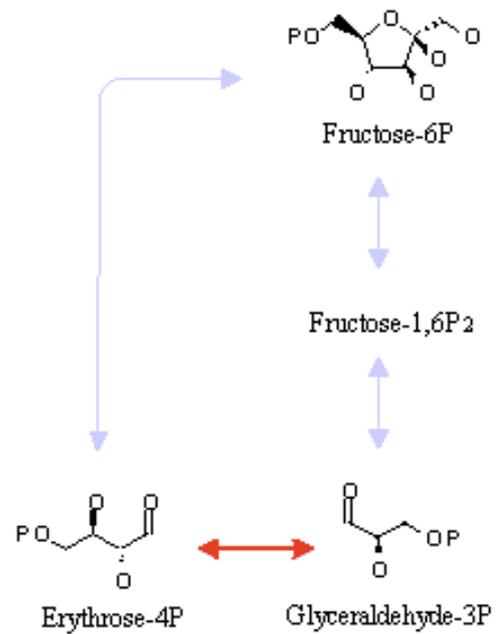
#### **4.4.5. Identification of candidate genes for a function in the compatible and incompatible interaction with *Phytophthora infestans***

After the statistical analysis for significance was implemented in comparisons over time after infection, it turned out that the large transcriptional changes measured during the incompatible interaction were smaller than in compatible interaction. The number of specifically regulated tags in the *R1*-transgenic lines showing a clear significant response was rather low. Only six tags were identified as highly significant and showing no response in the wildtype plants. Remarkably, all these tags were downregulated. At one day post infection, only StET029144 was detected. This tag is annotated as „Tuber-specific and sucrose-responsive element binding factor“ and matches to a gene called *tsf* (EMBL AAG05959.1). The putative gene product is a 364 amino acid protein which is predicted to be localized in the cell nucleus and carries various transcription factor specific domains (source: Interpro database). The domain architecture indicates strongly its classification as myb-type transcription factor. Following domains are identified in this protein: two homeodomains (IPR009057, IPR012287) and four different types of Myb-domains (IPR017930, IPR014778, IPR015495, IPR001005). In the databases, *tsf* has been detected in *Solanum tuberosum* leaves, callus, roots and tubers (Source: DFCI). A role in disease related processes has not been shown yet.

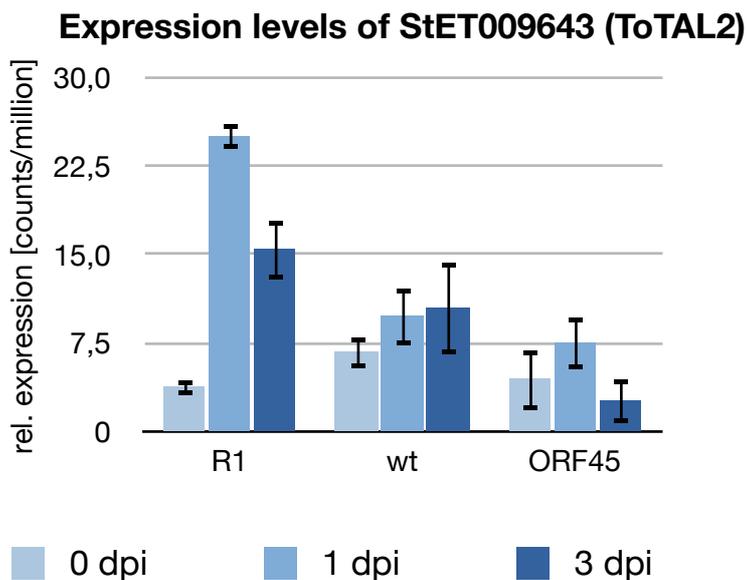
Another gene, which was identified after 3 days of infection through the tag StET027738 is annotated as EIN3-binding-F-box protein. In *Solanum tuberosum*, not very much is known

about this gene. In *Arabidopsis thaliana*, the related genes *ebf-1* and *ebf-2* have been shown to play a role in ethylene signalling and EIN3 itself is thought to be in crosstalk with the 26S proteasome [133].

The third identifiable gene (StET010841) is annotated as Fibrillin-8. This gene was identified in *Solanum tuberosum* in roots, tubers, flowers and leaves (source: DFCI). Almost nothing is known, which could give a hint for a putative function. Only its role in cytoskeleton organization, together with the observations of nucleus movement and cytoplasmatic traffic in the early nineties [134], gives a possible explanation for the significant



**Fig. 4.4.5.1:** Scheme of the reaction, catalyzed by ToTAL2 (EC2.2.1.2). (Source: KEGG database)



**Fig. 4.4.5.2:** Expression levels of StET009643. The bars show the average relative expression levels. Shown are the values for each timepoint in the order 0dpi, 1 dpi, 3dpi. The standard error is given for each bar.

changes in the expression changes of this gene. The other three identified tags have wether no matching target (StET027737) or are referred to chromosomal scaffolds of *Vitis vinifera* (StET018266 and StET004373) and lack any functional information.

Another gene, which is not listed in the high significance list shall be discussed here additionally (please see the data

browser or the additional file /Raw-data/FE-Results/Sig\_upgenes\_FE\_cleared.xls). Among the tags showing significant changes in the incompatible interaction at 1 day post infection, one (StET009643) rises remarkably, beings significantly upregulated in six of six experiments. This tag is annotated as Transaldolase ToTAL2. The product of this gene s

likely to have an enzymatic activity (EC.2.2.1.2). In the KEGG database [116, 135, 136] it is connected to five enzymatic pathways (ec0030, ec01061, ec01063, ec01070, ec01100). It catalyzes the reversible conversion from Erythrose-4P to Glyceraldehyde-3P (see Fig. 4.4.5.1). In plants, no involvement of this gene in disease related processes has been so far shown. But in humans, an expression increase of the TAL gene has been observed in response to oxidative stress. These processes have been connected to an enhanced sensitivity to apoptotic pathways [137-139]. Interestingly, this tag, only is significantly upregulated at 1 dpi, what is true for the results of ORF45 as well (although not as strong and just significant in 2 of 3 replicates). This gene is probably no key player in innate immunity responses but likely a part of correlated processes and a good candidate as quantitative resistance gene.

In a more generalized view the GO-term analysis has been implemented in order to characterize the processes with the most significant changes. It was observed that in the incompatible interaction only 3 days after infection over represented GO-categories were identified. In addition, the number of GO-categories was reduced when compared to the compatible interaction. This phenomenon was already implicated by the PCA analysis and the characterization of the phenotype in chapter 4.4.3 and an additional support for the hypothesis of a soft phenotype. For an interpretation of the results of GO-term analysis, it is important to know that this method is depending on database knowledge. When in a list of genes, the information about the function of these genes is not available, the analysis has no possibility to allocate these genes to the function terms. Here it has to be noticed that the GO-Term „defense response to fungus, incompatible interaction“ (GO:0009817) contains one tentative consensus sequence whereas the higher order gene ontology term defense response to fungus (GO:0050832) contains 131 tentative consensus sequences and includes the compatible interaction. This example reflects the current database situation. Although a couple of genes have been characterized as resistance genes genetically, on the level of molecular functions, these process have often been left behind. Nevertheless, at three days post infection, it was possible to identify four gene ontologies as overrepresented. These can be assigned to the formation of vesicles. The formation of membrane-bound vesicles as well as cytoplasmatic vesicles is present. These processes have been first recognized microscopically [134] and were later on studied on the molecular level in *Arabidopsis thaliana*, leading to the identification of genes like the SNARE-superfamily or PEN1 [140-142].

#### 4.4.6. Comparability of DeepSAGE-data with qRT-PCR

The comparison of the expression data from the DeepSAGE experiment with qRT-PCR as second method resulted in only a small overlap between the results. One problem concerns the different methodological resolutions. After comparing the results it became obvious, that a difference in resolution of both methods exist. Surprising was the fact, that there was no correlation between the logfoldchange-value originating from the DeepSAGE data, and the reproducibility in qRT-PCR (see the supplemental data disc for details). It might be, that the genes chosen for a test in a second experiment did not show sufficient expression changes. The chosen criteria of a minimum two-fold expression change is an ad-hoc criteria and has up to now, due to the lack of a broad application in complex experiments no empiric basis.

One problem might be given by the data resolution generated by both methods. A strong argument for this idea results when the raw values of the housekeeping gene *ef-1a* were compared. The measured expression rates in both methods varied clearly, leading to a decreases normalized expression value in the DeepSAGE data of 10-fold and more. This fact is based on the observation, that the expression levels of this gene (represented by the tag StET018993) with an average expression of 445,09 counts/million is remarkably higher than the expression levels of the other measured genes. This span reflects the resolution of the DeepSAGE method. And as it has been shown in the qRT-PCR-experiments that the expression differences between the housekeeping gene and the other genes were smaller than in the quantitative sequencing data. This is expressed by the values relative to the genes (see chapter 3.4.8.). This might be a possible factor explaining the different results.

Another factor, which scaled down the original measured values was the necessity to reconstruct the logfoldchange values. To get comparable results, two factors are made - on the one hand, the data is expressed in the unit „x-fold *ef-1a*“, afterwards. To achieve a value comparable to DeepSAGE, the data was expressed in a form  $\log\left(\frac{t_1}{(n \times ef-1a)_{t_2}}\right)$ , where  $(n \times ef-1a)$  describes the normalized expression value in a n-times *ef-1a* and  $t_1$  and  $t_2$  indicate the two compared timepoints. The application of this formula influences the initial relationship of values, which have been used for the estimation of significance. However, an expression of the results in form of expression differences is necessary to

compare results when dealing with different units.

Besides technical difficulties uncertainty factors exist. The qRT-PCR experiment was designed in a very challenging manner. Unlike other transcriptomics experiments, a second independent test series with different biological material was used. It is possible that the poor reproduction of results is due to biological variance. With a tag-length of 17 bases, there is also a degree of uncertainty in annotation. With StET029853, annotated as polyubiquitin, one tag was included, which shows this uncertainties as it has two possible target sites. It is possible, that the selected tag does not match the amplified gene.

Nevertheless, given the broad involvement of the genes referred to defense as indicator genes (given in chapter 3.4.4) and the involvement of processes, which have been shown in studies, there is no doubt that a response of the sample tissue to the infection occurred. It has to be admitted, that the molecular phenotype and the transcriptional activities during this event have a high biological variance. It is possible that many of the measured expression changes are of a short lifetime and are very distinct in their occurrence.

#### **4.4.7. Concluding remarks - the extraordinary role of the plant hormone synthesis pathway (KEGG ec01070)**

Throughout multiple analysis methods, many processes like mRNA-processing and many transcription factors (please see chapter 3.4.6 and 3.4.7 and the supplemental data disc for details) have been detected as responding to the infection. Also expression changes in proteinase inhibitors, chitinases and glucanases have been observed. This indicates a vast transcriptomal rearrangement under stress conditions particularly in lines showing a compatible interaction. Whether these huge changes are due to the reprogramming of the pathogen is speculative, but at least in parts probable.

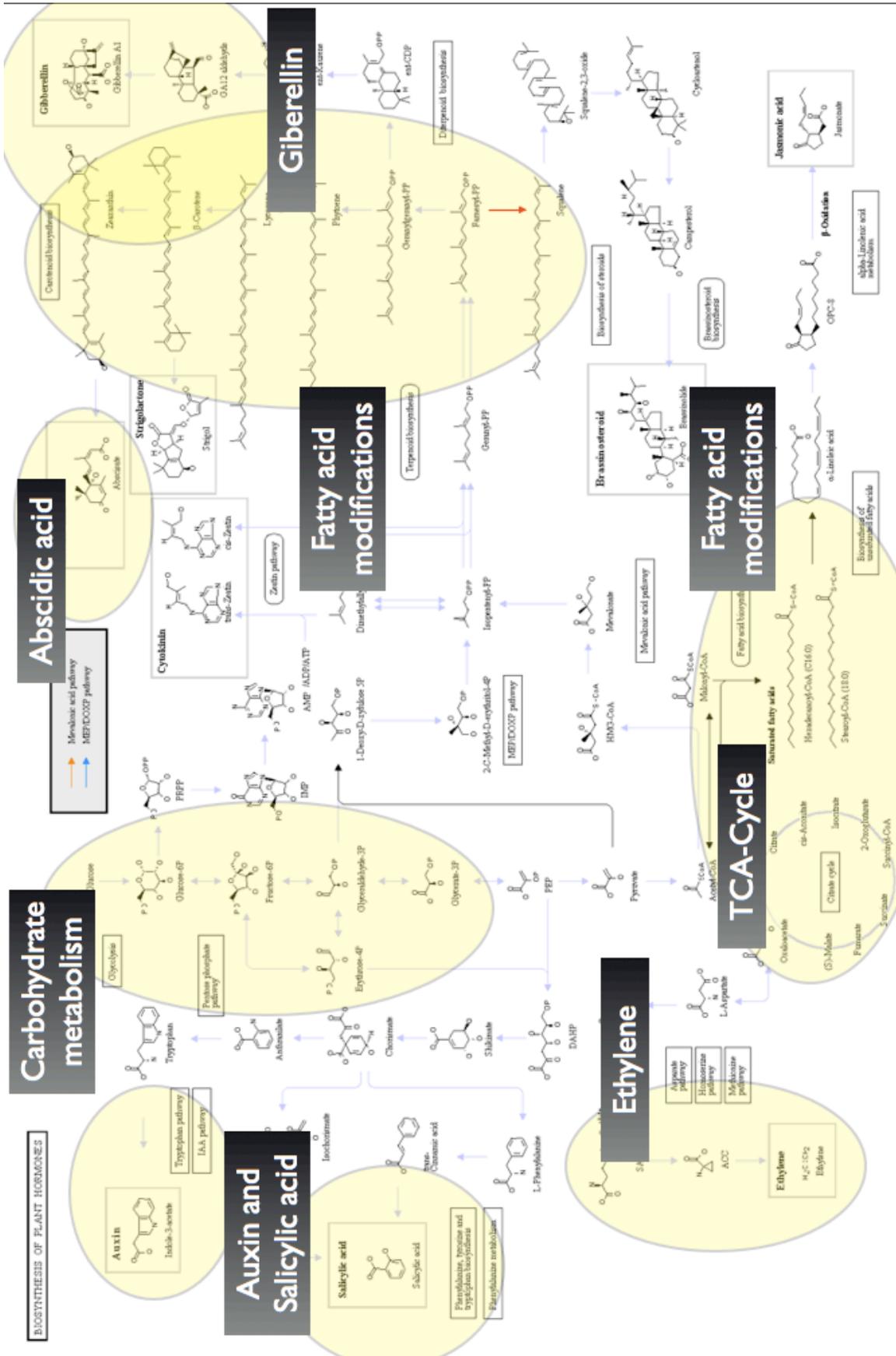
Among the transcriptomal changes, some responding processes have been observed in multiple analysis whether direct or indirect. These processes seem to be correlated to the influence of plant hormones.

In plants, hormone signalling under stress conditions is known and well described, especially the involvement of jasmonic acid and salicylic acid [114, 143]. With our data, we provide further evidence for an intensive involvement of plant hormone signalling as response to host pathogen interaction. For example, the role of auxin as a side-arm of the salicylic acid synthesis is strongly indicated. Additionally, multiple observations support the

hypothesis of ethylene as responding stress hormone, whether it is passive via the proposed cross-talk with absidic acid or under active modulation can not be concluded by expression data. Additionally a wide involvement of the TCA-cycle was detected in both interactions.

In this context the regulatory changes in the RubisCo subunits and photosynthetic active enzymes can be explained as well (see fig. 4.4.7.1). The initial interpretation of the findings thought of a phenotypic artifact, but in the context of the synthesis of hormones it becomes clear that energy recruitment and carbon fixation are the basis for the synthesis of secondary metabolites. This can explain the observed significances of the plastidial processes in the GO-analysis.

A summarizing sentence which expresses the view of a cell under infection can be: „Fight and warn the others“.



**Fig.4.4.7.1:** Overview of the reference pathway ec01070 for plant hormone biosyntheses. Steps where genes have been found to be regulated wether in incompatible or compatible interaction are marked by a yellow area. The involvement in this steps can be wether direct as enzyme or indirect, responding to the synthesized metabolite. (A larger view is given on the supplemental data disc.)

## References

1. Mueller, L.A., et al., *The SOL Genomics Network. A Comparative Resource for Solanaceae Biology and Beyond*. Plant Physiol., 2005. **138**(3): p. 1310-1317.
2. Hawkes, J.G., *The Potato*. 1990, London: Belhaven Press.
3. Schöber-Butin, B., *Late Blight of the Potato and its Causal Agent Phytophthora infestans (Mont.) de Bary*. Mitteilungen aus der Biologischen Bundesanstalt für Land- und Forstwirtschaft. Vol. 384. 2001, Berlin: Parey.
4. Sayers, E.W., et al., *Database resources of the National Center for Biotechnology Information*. Nucl. Acids Res., 2010. **38**(suppl\_1): p. D5-16.
5. Haas, B.J., et al., *Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans*. Nature, 2009. **461**(7262): p. 393-398.
6. Duncan, J., *Phytophthora - an abiding threat to our crops*. Microbiology Today, 1999. **114**(26).
7. Grünwald, N.J. and W.G. Flier, *The biology of Phytophthora infestans at its center of origin*. Annual Reviews in Phytopathology, 2005. **43**: p. 171-190.
8. Zwankhuisen, M.J., F. Govers, and J.C. Zadoks, *Development of Potato Late Blight Epidemics: Disease Foci, Disease Gradients, and Infection Sources*. Phytopathology, 1998. **88**(8): p. 754-763.
9. Hardham, A.R., *Cell biology of plant-oomycete interactions*. Cell Microbiol, 2007. **9**(1): p. 31-9.
10. Judelson, H.S. and F.A. Blanco, *The spores of Phytophthora: weapons of the plant destroyer*. Nature Reviews. Microbiology, 2005. **3**(1): p. 47-58.
11. Tyler, B.M., *Molecular basis of recognition between phytophthora pathogens and their hosts*. Annu Rev Phytopathol, 2002. **40**: p. 137-67.
12. Judelson, H.S., *The genetics and biology of Phytophthora infestans: modern approaches to a historical challenge*. Fungal Genet Biol, 1997. **22**(2): p. 65-76.
13. Fry, W., *Phytophthora infestans: the plant (and R gene) destroyer*. Molecular Plant Pathology, 2008. **9**(3): p. 385-402.
14. Shattock, R.C., *Phytophthora infestans: populations, pathogenicity and phenylamides*. Pest Manag Sci, 2002. **58**(9): p. 944-50.
15. Haldar, K., et al., *Common infection strategies of pathogenic eukaryotes*. Nat Rev Microbiol, 2006. **4**(12): p. 922-31.

16. Belkhadir, Y., R. Subramaniam, and J.L. Dangl, *Plant disease resistance protein signaling: NBS-LRR proteins and their partners*. Current Opinion in Plant Biology, 2004. **7**(4): p. 391-399.
17. Brown, J.K.M., *Evolution. Little else but parasites*. Science (New York, N.Y.), 2003. **299**(5613): p. 1680-1681.
18. Meyer, S., A. Nagel, and C. Gebhardt, *PoMaMo--a comprehensive database for potato genome data*. Nucleic Acids Research, 2005. **33**(Database issue): p. D666-670-D666-670.
19. Gebhardt, C. and J.P. Valkonen, *Organization of genes controlling disease resistance in the potato genome*. Annual Review of Phytopathology, 2001. **39**: p. 79-102.
20. Ting, J.P., S.B. Willingham, and D.T. Bergstralh, *NLRs at the intersection of cell death and immunity*. Nat Rev Immunol, 2008. **8**(5): p. 372-9.
21. van der Vossen, E., et al., *An ancient R gene from the wild potato species Solanum bulbocastanum confers broad-spectrum resistance to Phytophthora infestans in cultivated potato and tomato*. Plant J, 2003. **36**(6): p. 867-82.
22. Bradshaw, J.E., et al., *Mapping the R10 and R11 genes for resistance to late blight (Phytophthora infestans) present in the potato (Solanum tuberosum) R-gene differentials of black*. Theor Appl Genet, 2006. **112**(4): p. 744-51.
23. Huang, S., et al., *The R3 resistance to Phytophthora infestans in potato is conferred by two closely linked R genes with distinct specificities*. Mol Plant Microbe Interact, 2004. **17**(4): p. 428-35.
24. Sliwka, J., *Genetic factors encoding resistance to late blight caused by Phytophthora infestans (Mont.) de Bary on the potato genetic map*. Cell Mol Biol Lett, 2004. **9**(4B): p. 855-67.
25. Solomon-Blackburn, R.M., H.E. Stewart, and J.E. Bradshaw, *Distinguishing major-gene from field resistance to late blight (Phytophthora infestans) of potato (Solanum tuberosum) and selecting for high levels of field resistance*. Theor Appl Genet, 2007. **115**(1): p. 141-9.
26. Pajerowska-Mukhtar, K.M., et al., *Natural variation of potato allene oxide synthase 2 causes differential levels of jasmonates and pathogen resistance in Arabidopsis*. Planta, 2008. **228**(2): p. 293-306.

27. Pajerowska-Mukhtar, K., et al., *Single nucleotide polymorphisms in the allene oxide synthase 2 gene are associated with field resistance to late blight in populations of tetraploid potato cultivars*. Genetics, 2009. **181**(3): p. 1115-1127.
28. Park, T.H., et al., *The late blight resistance locus Rpi-bib3 from Solanum bulbocastanum belongs to a major late blight R gene cluster on chromosome 4 of potato*. Mol Plant Microbe Interact, 2005. **18**(7): p. 722-9.
29. Kamoun, S., *A Catalogue of the effector secretome of plant pathogenic oomycetes*. Annual Reviews of Phytopathology, 2006. **44**: p. 41-60.
30. Kamoun, S., *Groovy times: filamentous pathogen effectors revealed*. Curr Opin Plant Biol, 2007. **10**(4): p. 358-65.
31. Bhattacharjee, S., et al., *The malarial host-targeting signal is conserved in the Irish potato famine pathogen*. PLoS Pathog, 2006. **2**(5): p. e50.
32. Whisson, S.C., et al., *A translocation signal for delivery of oomycete effector proteins into host plant cells*. Nature, 2007. **450**(7166): p. 115-118.
33. Win, J., et al., *Adaptive evolution has targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes*. Plant Cell, 2007. **19**(8): p. 2349-69.
34. Martin, G.B., A.J. Bogdanove, and G. Sessa, *Understanding the functions of plant disease resistance proteins*. Annu Rev Plant Biol, 2003. **54**: p. 23-61.
35. Schornack, S., et al., *Ten things to know about oomycete effectors*. Molecular Plant Pathology, 2009. **10**(6): p. 795-803.
36. Rairdan, G. and P. Moffett, *Brothers in arms? Common and contrasting themes in pathogen perception by plant NB-LRR and animal NACHT-LRR proteins*. Microbes Infect, 2007. **9**(5): p. 677-86.
37. Shen, Q.H. and P. Schulze-Lefert, *Rumble in the nuclear jungle: compartmentalization, trafficking, and nuclear action of plant immune receptors*. Embo J, 2007. **26**(20): p. 4293-301.
38. Ellis, J., P. Dodds, and T. Pryor, *Structure, function and evolution of plant disease resistance genes*. Curr Opin Plant Biol, 2000. **3**(4): p. 278-84.
39. Dangl, J.L. and J.D. Jones, *Plant pathogens and integrated defence responses to infection*. Nature, 2001. **411**(6839): p. 826-33.
40. Mucyn, T.S., et al., *The tomato NBARC-LRR protein Prf interacts with Pto kinase in vivo to regulate specific plant immunity*. Plant Cell, 2006. **18**(10): p. 2792-806.

41. Gutierrez, J.R., et al., *Prf immune complexes of tomato are oligomeric and contain multiple Pto-like kinases that diversify effector recognition*. Plant J, 2009.
42. Mucyn, T.S., et al., *Regulation of tomato Prf by Pto-like protein kinases*. Mol Plant Microbe Interact, 2009. **22**(4): p. 391-401.
43. Chinchilla, D., et al., *The Arabidopsis receptor kinase FLS2 binds flg22 and determines the specificity of flagellin perception*. Plant Cell, 2006. **18**(2): p. 465-76.
44. Jones, J.D. and J.L. Dangl, *The plant immune system*. Nature, 2006. **444**(7117): p. 323-9.
45. Vleeshouwers, V.G., et al., *The hypersensitive response is associated with host and nonhost resistance to Phytophthora infestans*. Planta, 2000. **210**(6): p. 853-64.
46. DeYoung, B.J. and R.W. Innes, *Plant NBS-LRR proteins in pathogen sensing and host defense*. Nat Immunol, 2006. **7**(12): p. 1243-9.
47. Leonards-Schippers, C., et al., *The R1 gene conferring race-specific resistance to Phytophthora infestans in potato is located on potato chromosome V*. Mol Gen Genet, 1992. **233**(1-2): p. 278-83.
48. Leonards-Schippers, C., et al., *Quantitative resistance to Phytophthora infestans in potato: a case study for QTL mapping in an allogamous plant species*. Genetics, 1994. **137**(1): p. 67-77.
49. Bormann, C.A., et al., *Tagging quantitative trait loci for maturity-corrected late blight resistance in tetraploid potato with PCR-based candidate gene markers*. Molecular Plant-Microbe Interactions: MPMI, 2004. **17**(10): p. 1126-1138.
50. Zimnoch-Guzowska, E., et al., *QTL analysis of new sources of resistance to Erwinia carotovora ssp atroseptica in potato done by AFLP, RFLP and resistance-gene-like markers*. Crop Science, 2000. **40**: p. 1156-1167.
51. Ewing, E.E., et al., *Genetic mapping from field tests of qualitative and quantitative resistance to Phytophthora infestans in a population derived from Solanum tuberosum and Solanum berthaultii*. . Molecuar Breeding, 2000. **6**: p. 25-36.
52. Leister, D., et al., *A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants*. Nature Genetics, 1996. **14**(4): p. 421-429.
53. Rickert, A.M., et al., *First-generation SNP/InDel markers tagging loci for pathogen resistance in the potato genome*. Plant Biotechnology Journal, 2003. **1**(6): p. 399-410.

54. Ballvora, A., et al., *The R1 gene for potato resistance to late blight (Phytophthora infestans) belongs to the leucine zipper/NBS/LRR class of plant resistance genes*. The Plant Journal: For Cell and Molecular Biology, 2002. **30**(3): p. 361-371.
55. Ballvora, A., et al., *Comparative sequence analysis of Solanum and Arabidopsis in a hot spot for pathogen resistance on potato chromosome V reveals a patchwork of conserved and rapidly evolving genome segments*. BMC Genomics, 2007. **8**: p. 112.
56. Kuang, H., et al., *The R1 resistance gene cluster contains three groups of independently evolving, type I R1 homologues and shows substantial structural variation among haplotypes of Solanum demissum*. Plant J, 2005. **44**(1): p. 37-51.
57. Oberhagemann, P., et al., *A genetic analysis of quantitative resistance to late blight in potato are strongly correlated with earlyness and vigour*. Molecular Breeding, 1999. **5**: p. 387-398.
58. Hulo, N., et al., *The 20 years of PROSITE*. Nucleic Acids Res, 2008. **36**(Database issue): p. D245-9.
59. Defelice, M.S., *The Black Nightshades, Solanum nigrum L. et al.-Poison, Poultry, and Pie*. Weed Technology, 2003. **17**: p. 421-427.
60. Hsu, J.-D., et al., *Solanum nigrum L. extract inhibits 2-acetylaminofluorene-induced hepatocarcinogenesis through overexpression of glutathione S-transferase and antioxidant enzymes*. Journal of Agricultural and Food Chemistry, 2009. **57**(18): p. 8628-8634.
61. Jeong, J.B., B.O. De Lumen, and H.J. Jeong, *Lunasin peptide purified from Solanum nigrum L. protects DNA from oxidative damage by suppressing the generation of hydroxyl radical via blocking fenton reaction*. Cancer Letters. **In Press, Corrected Proof**.
62. Jimoh, F.O., A.A. Adedapo, and A.J. Afolayan, *Comparison of the nutritional value and biological activities of the acetone, methanol and water extracts of the leaves of Solanum nigrum and Leonotis leonorus*. Food and Chemical Toxicology: An International Journal Published for the British Industrial Biological Research Association, 2010. **48**(3): p. 964-971.
63. Schmidt, D.D., et al., *Solanum nigrum: a model ecological expression system and its tools*. Mol Ecol, 2004. **13**(5): p. 981-95.

64. Schmidt, S. and I.T. Baldwin, *Systemin in Solanum nigrum. The tomato-homologous polypeptide does not mediate direct defense responses*. Plant Physiology, 2006. **142**(4): p. 1751-1758.
65. Flier, W.G., G.B.M. van den Bosch, and L.J. Turkensteen, *Epidemiological importance of Solanum sisymbriifolium, S. nigrum and S. dulcamara as alternative hosts for Phytophthora infestans*. Plant Pathology, 2003. **52**: p. 595-603.
66. Colon, I.T., et al., *Resistance to potato late blight (Phytophthora infestans (Mont.) de Bary) in Solanum nigrum, S. villosum and their sexual hybrids with S. tuberosum and S. demissum*. Euphytica, 1992. **66**(1): p. 55-64.
67. Szczerbakowa, A., et al., *Somatic hybrids Solanum nigrum (+) S. tuberosum: morphological assessment and verification of hybridity*. Plant Cell Reports, 2003. **21**(6): p. 577-584.
68. Zimnoch-Guzowska, E., et al., *Resistance to Phytophthora infestans in somatic hybrids of Solanum nigrum L. and diploid potato*. TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik, 2003. **107**(1): p. 43-48.
69. Southern, E.M., *Detection of specific sequences among DNA fragments separated by gel electrophoresis*. J Mol Biol, 1975. **98**(3): p. 503-17.
70. Alwine, J.C., D.J. Kemp, and G.R. Stark, *Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes*. Proceedings of the National Academy of Sciences of the United States of America, 1977. **74**(12): p. 5350-5354.
71. Myers, J.C., C. Dobkin, and S. Spiegelman, *RNA primer used in synthesis of anti-complementary DNA by reverse transcriptase of avian myeloblastosis virus*. Proc Natl Acad Sci U S A, 1980. **77**(3): p. 1316-20.
72. Saiki, R.K., et al., *Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia*. Science, 1985. **230**(4732): p. 1350-4.
73. Adams, M.D., et al., *Complementary DNA sequencing: expressed sequence tags and human genome project*. Science, 1991. **252**(5013): p. 1651-6.
74. Liang, P., et al., *Differential display and cloning of messenger RNAs from human breast cancer versus mammary epithelial cells*. Cancer Res, 1992. **52**(24): p. 6966-8.

75. Schena, M., et al., *Quantitative monitoring of gene expression patterns with a complementary DNA microarray*. Science, 1995. **270**(5235): p. 467-70.
76. Lockhart, D.J., et al., *Expression monitoring by hybridization to high-density oligonucleotide arrays*. Nat Biotechnol, 1996. **14**(13): p. 1675-80.
77. Lennon, G.G. and H. Lehrach, *Hybridization analyses of arrayed cDNA libraries*. Trends Genet, 1991. **7**(10): p. 314-7.
78. Velculescu, V.E., et al., *Serial analysis of gene expression*. Science, 1995. **270**(5235): p. 484-7.
79. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors*. Proc Natl Acad Sci U S A, 1977. **74**(12): p. 5463-7.
80. Nielsen, K.L., A.L. Høgh, and J. Emmersen, *DeepSAGE--digital transcriptomics with high sensitivity, simple experimental protocol and multiplexing of samples*. Nucleic Acids Res, 2006. **34**(19): p. e133.
81. Nielsen, K.L., *DeepSAGE: higher sensitivity and multiplexing of samples using a simpler experimental protocol*. Methods Mol Biol, 2008. **387**: p. 81-94.
82. Matsumura, H., et al., *Gene expression analysis of plant host-pathogen interactions by SuperSAGE*. Proc Natl Acad Sci U S A, 2003. **100**(26): p. 15718-23.
83. Wei, C.L., et al., *5' Long serial analysis of gene expression (LongSAGE) and 3' LongSAGE for transcriptome characterization and genome annotation*. Proc Natl Acad Sci U S A, 2004. **101**(32): p. 11701-6.
84. Gowda, M., et al., *Robust-LongSAGE (RL-SAGE): a substantially improved LongSAGE method for gene discovery and transcriptome analysis*. Plant Physiol, 2004. **134**(3): p. 890-7.
85. Gowda, M., et al., *Use of robust-long serial analysis of gene expression to identify novel fungal and plant genes involved in host-pathogen interactions*. Methods Mol Biol, 2007. **354**: p. 131-44.
86. Gowda, M. and G.L. Wang, *Robust-LongSAGE (RL-SAGE): an improved LongSAGE method for high-throughput transcriptome analysis*. Methods Mol Biol, 2008. **387**: p. 25-38.
87. Margulies, M., et al., *Genome sequencing in microfabricated high-density picolitre reactors*. Nature, 2005. **437**(7057): p. 376-80.
88. Roche. 2010; Available from: [www.454.com](http://www.454.com).
89. Helicos. 2010; Available from: [www.helicosbio.com](http://www.helicosbio.com).

90. Biosystems, A. 2010; Available from: [www2.appliedbiosystems.com](http://www2.appliedbiosystems.com).
91. Schaaf, G.J., et al., *Statistical comparison of two or more SAGE libraries: one tag at a time*. *Methods Mol Biol*, 2008. **387**: p. 151-68.
92. Beissbarth, T., et al., *Statistical modeling of sequencing errors in SAGE libraries*. *Bioinformatics*, 2004. **20 Suppl 1**: p. i31-9.
93. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data*. *Bioinformatics*, 2010. **26**(1): p. 139-40.
94. Hardcastle, T.J., *baySeq - Empirical Bayesian analysis of patterns of differential expression in count data*. 2009.
95. Caten, C.E. and J.L. Junks, *Spontaneous variability of single isolates of *Phytophthora infestans* I. Clutural variation*. *Canadian Journal of Botany*, 1968. **46**(4): p. 329-348.
96. Si-Ammour, A., B. Mauch-Mani, and F. Mauch, *Quantification of induced resistance against *Phytophthora* species expressing GFP as vital marker:  $\beta$ -aminobutric acid but not BTH protects potato and *Arabidopsis* from infection*. *Molecular Plant athology*, 2003. **4**(4): p. 237-248.
97. Oh, S.-K., et al., *In planta expression screens of *Phytophthora infestans* RXLR effectors reveal diverse phenotypes, including activation of the *Solanum bulbocastanum* disease resistance protein *Rpi-blb2**. *The Plant Cell*, 2009. **21**(9): p. 2928-2947.
98. Bormann, C.A., *Genetic and molecular analysis of quantitative late blight resistance in tetraploid potato*, in *Max-Planck-Institute for Plant Breeding Research*. 2003, University of Hohenheim: Cologne. p. 115.
99. Jones, J.D., et al., *Effective vectors for transformation, expression of heterologous genes, and assaying transposon excision in transgenic plants*. *Transgenic Res*, 1992. **1**(6): p. 285-97.
100. Quackenbush, J., et al., *The TIGR gene indices: reconstruction and representation of expressed gene sequences*. *Nucleic Acids Res*, 2000. **28**(1): p. 141-5.
101. Quackenbush, J., et al., *The TIGR Gene Indices: analysis of gene transcript sequences in highly sampled eukaryotic species*. *Nucleic Acids Res*, 2001. **29**(1): p. 159-64.

102. Lee, Y., et al., *The TIGR Gene Indices: clustering and assembling EST and known genes and integration with eukaryotic genomes*. Nucleic Acids Res, 2005. **33**(Database issue): p. D71-4.
103. Benjamini, Y. and Y. Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing*. Journal of the Royal Statistical Society, 1995. **57**(1): p. 289-300.
104. Saitou, N. and M. Nei, *The neighbor-joining method: a new method for reconstructing phylogenetic trees*. Mol Biol Evol, 1987. **4**(4): p. 406-25.
105. Chakravarthy, S., A.C. Velasquez, and G.B. Martin, *Assay for pathogen-associated molecular pattern (PAMP)-triggered immunity (PTI) in plants*. J Vis Exp, 2009(31).
106. Nicaise, V., M. Roux, and C. Zipfel, *Recent advances in PAMP-triggered immunity against bacteria: pattern recognition receptors watch over and raise the alarm*. Plant Physiol, 2009. **150**(4): p. 1638-47.
107. JVC1. *TIGR Potato cDNA Microarray Description*. 2010; Available from: [http://jcv1.org/potato/sol\\_ma\\_microarrays.shtml](http://jcv1.org/potato/sol_ma_microarrays.shtml).
108. Neigenfind, J., et al., *Haplotype inference from unphased SNP data in heterozygous polyploids based on SAT*. BMC Genomics, 2008. **9**: p. 356-356.
109. Salmela, L., *Correction of sequencing errors in a mixed set of reads*. Bioinformatics, 2010.
110. Restrepo, S., et al., *Gene profiling of a compatible interaction between *Phytophthora infestans* and *Solanum tuberosum* suggests a role for carbonic anhydrase*. Mol Plant Microbe Interact, 2005. **18**(9): p. 913-22.
111. Pinzon, A., et al., *Computational models in plant-pathogen interactions: the case of *Phytophthora infestans**. Theor Biol Med Model, 2009. **6**: p. 24.
112. Kottapalli, K.R., et al., *Transcriptional profiling of indica rice cultivar IET8585 (Ajaya) infected with bacterial leaf blight pathogen *Xanthomonas oryzae* pv *oryzae**. Plant Physiol Biochem, 2007. **45**(10-11): p. 834-50.
113. Wang, X., et al., *Local and distal gene expression of pr-1 and pr-5 in potato leaves inoculated with isolates from the old (US-1) and the new (US-8) genotypes of *Phytophthora infestans* (Mont.) de Bary*. Environmental and Experimental Botany, 2006. **57**(1-2): p. 70-79.

114. Halim, V.A., et al., *Salicylic acid is important for basal defense of Solanum tuberosum against Phytophthora infestans*. Molecular Plant-Microbe Interactions: MPMI, 2007. **20**(11): p. 1346-1352.
115. Maruyama-Nakashita, A., et al., *Arabidopsis SLIM1 is a central transcriptional regulator of plant sulfur response and metabolism*. Plant Cell, 2006. **18**(11): p. 3235-51.
116. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes*. Nucleic Acids Res, 2000. **28**(1): p. 27-30.
117. De Vleeschauwer, D., et al., *Abscisic acid-induced resistance against the brown spot pathogen Cochliobolus miyabeanus in rice involves MAP kinase-mediated repression of ethylene signaling*. Plant Physiol, 2010. **152**(4): p. 2036-52.
118. Jain, E., et al., *Infrastructure for the life sciences: design and implementation of the UniProt website*. BMC Bioinformatics, 2009. **10**: p. 136.
119. Stanford, A., M. Bevan, and D. Northcote, *Differential expression within a family of novel wound-induced genes in potato*. Mol Gen Genet, 1989. **215**(2): p. 200-8.
120. Gadea, J., et al., *Characterization of defense-related genes ectopically expressed in viroid-infected tomato plants*. Mol Plant Microbe Interact, 1996. **9**(5): p. 409-15.
121. Bais, H.P.V., Ramaro, et al., *Molecular and biochemical characterization of an enzyme responsible for the formation of Hypericin in St. John's Wort (Hypericum perforatum L.)*. The Journal of Biological Chemistry, 2003. **278**: p. 32413-32422.
122. Joos, H.J. and K. Hahlbrock, *Phenylalanine ammonia-lyase in potato (Solanum tuberosum L.). Genomic complexity, structural comparison of two selected genes and modes of expression*. European Journal of Biochemistry, 1992. **204**(2): p. 621-629.
123. Rumeau, D., et al., *Extensin and Phenylalanine Ammonia-Lyase gene expression altered in Potato tubers in response to wounding, hypoxia, and Erwinia carotovora infection*. Plant Physiology, 1990. **93**: p. 1134-1139.
124. Rickey, T.M. and W.R. Belknap, *Comparison of the expression of several stress-responsive genes in potato tubers*. Plant Molecular biology, 1991. **16**(6): p. 1009-1018.
125. Durrant, W.E., et al., *cDNA-AFLP reveals a striking overlap in race-specific resistance and wound response gene expression profiles*. Plant Cell, 2000. **12**(6): p. 963-77.

126. Dhondt, S., et al., *Soluble phospholipase A2 activity is induced before oxylipin accumulation in tobacco mosaic virus-infected tobacco leaves and is contributed by patatin-like enzymes*. Plant J, 2000. **23**(4): p. 431-40.
127. Tian, M., B. Benedetti, and S. Kamoun, *A Second Kazal-like protease inhibitor from Phytophthora infestans inhibits and interacts with the apoplastic pathogenesis-related protease P69B of tomato*. Plant Physiol, 2005. **138**(3): p. 1785-93.
128. Tian, M., et al., *A Kazal-like extracellular serine protease inhibitor from Phytophthora infestans targets the tomato pathogenesis-related protease P69B*. J Biol Chem, 2004. **279**(25): p. 26370-7.
129. Tian, M. and S. Kamoun, *A two disulfide bridge Kazal domain from Phytophthora exhibits stable inhibitory activity against serine proteases of the subtilisin family*. BMC Biochem, 2005. **6**: p. 15.
130. Tian, M., et al., *A Phytophthora infestans cystatin-like protein targets a novel tomato papain-like apoplastic protease*. Plant Physiol, 2007. **143**(1): p. 364-77.
131. Valens, M., et al., *A zinc metalloprotease inhibitor, Inh, from the insect pathogen Photorhabdus luminescens*. Microbiology, 2002. **148**(Pt 8): p. 2427-37.
132. Tucker, M.L., et al., *Gene expression profiling for cell-wall modifying proteins associated with soybean cyst nematode infection, petiole abscission, root tips, flowers, apical buds, and leaves*. Journal of experimental Botany, 2007. **58**(12): p. 3395-3406.
133. Binder, B.M., et al., *The Arabidopsis EIN3 binding F-Box proteins EBF1 and EBF2 have distinct but overlapping roles in ethylene signaling*. Plant Cell, 2007. **19**(2): p. 509-23.
134. Freytag, S., et al., *Reversible cytoplasmic rearrangements precede wall apposition, hypersensitive cell death and defense-related gene activation in potato/Phytophthora infestans interactions*. Planta, 1993. **194**: p. 123-135.
135. Kanehisa, M., et al., *From genomics to chemical genomics: new developments in KEGG*. Nucleic Acids Res, 2006. **34**(Database issue): p. D354-7.
136. Kanehisa, M., et al., *KEGG for representation and analysis of molecular networks involving diseases and drugs*. Nucleic Acids Res, 2010. **38**(Database issue): p. D355-60.
137. Banki, K., et al., *Glutathione levels and sensitivity to apoptosis are regulated by changes in transaldolase expression*. J Biol Chem, 1996. **271**(51): p. 32994-3001.

138. Banki, K., et al., *Molecular ordering in HIV-induced apoptosis. Oxidative stress, activation of caspases, and cell survival are regulated by transaldolase.* J Biol Chem, 1998. **273**(19): p. 11944-53.
139. Banki, K., et al., *Elevation of mitochondrial transmembrane potential and reactive oxygen intermediate levels are early events and occur independently from activation of caspases in Fas signaling.* J Immunol, 1999. **162**(3): p. 1466-79.
140. Bhat, R.A., et al., *Recruitment and interaction dynamics of plant penetration resistance components in a plasma membrane microdomain.* Proc Natl Acad Sci U S A, 2005. **102**(8): p. 3135-40.
141. Lipka, V., C. Kwon, and R. Panstruga, *SNARE-ware: the role of SNARE-domain proteins in plant biology.* Annu Rev Cell Dev Biol, 2007. **23**: p. 147-74.
142. Robatzek, S., *Vesicle trafficking in plant immune responses.* Cell Microbiol, 2007. **9**(1): p. 1-8.
143. Halim, V.A., et al., *PAMP-induced defense responses in potato require both salicylic acid and jasmonic acid.* The Plant Journal: For Cell and Molecular Biology, 2009. **57**(2): p. 230-242.
144. Nicot, N., et al., *Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress.* J Exp Bot, 2005. **56**(421): p. 2907-14.
145. Judelson, H.S. and P.W. Tooley, *Enhanced Polymerase Chain Reaction Methods for Detecting and Quantifying Phytophthora infestans in Plants.* Phytopathology, 2000. **90**(10): p. 1112-9.
146. Agrios, G.N., *Plant Pathology.* 1978, San Diego: Academic press.
147. Judelson, H.S. and F.A. Blanco, *The spores of Phytophthora: weapons of the plant destroyer.* Nat Rev Microbiol, 2005. **3**(1): p. 47-58.
148. Ellis, J., A.M. Catanzariti, and P. Dodds, *The problem of how fungal and oomycete avirulence proteins enter plant cells.* Trends Plant Sci, 2006. **11**(2): p. 61-3.
149. Wilber, J.C., *Branched DNA for quantification of viral load.* Immunological Investigations, 1997. **26**(1-2): p. 9-13.
150. Illumina, 2010.
151. Trognitz, F. and B.R. Trognitz, *Survey of resistance gene analogs in Solanum caripense, a relative of potato and tomato, and update on R gene genealogy.* Mol Genet Genomics, 2005. **274**(6): p. 595-605.

## Summary (english)

The major disease for *Solanum tuberosum* is the foliage and tuber blight caused by the oomycete *Phytophthora infestans*. Defeat of this pathogen is often connected with the application of pesticides, which are toxic and increase production costs. However plants dispose over a class of genes which are able to mediate resistance to existing pathogens known as *R*-genes.

These genes play a key-role in effector-induced resistance responses which finally result in a form of programmed cell death (PCD), the hypersensitive response (HR).

*R1* belongs to a major class of resistance genes (NBS-LRR class genes) and is known to mediate race-specific resistance to *P. infestans*. The chromosomal region around *R1* has been sequenced and revealed eight open reading frames (ORFs) with high homology to *R1*. These genes are part of a cluster of resistance genes on *S. tuberosum* chromosome V. The nine ORFs share higher sequence similarity with each other than with any other resistance gene in *S. tuberosum* known so far and are therefore considered as the *R1*-family. Within this project the molecular characterization of the *R1*-family and the response genes was performed.

The major goal of this study was the molecular characterization of the *R1*-mediated resistance phenotype. To achieve this, a complex comparative transcriptome analysis of transgenic plants was performed in the initial stages of late blight infection. Different Desirée lines either with or without the *R1*-gene were used in this study. In addition, transformants with the *R1*-homologous gene ORF45, which shares high homology to *R1* but does not confer resistance was analyzed.

To generate the highest possible output rate, one of the new Next Generation Sequencing methods was used in combination with the DeepSAGE-technology. Performance parameters were validated. Various data analysis techniques were evaluated and at the end a pipeline was composed, which was able to process the data and leads to a state-of-the-art statistical data assessment.

The data analysis was able to provide further evidence for an intensive role of plant hormone signalling during the early stages of infection. Moreover in these key-processes different expression behaviors were observed in the incompatible versus the compatible interaction.

In addition it was possible to give experimental evidence for the expression of five of the members of the *R1*-family and cytosolic resistance mechanism in *Solanum nigrum* which is able to sense the presence of effector molecules from *Phytophthora infestans*.

#### Zusammenfassung (deutsch)

Bei dem für *Solanum tuberosum* primär relevanten Krankheitserreger handelt es sich um den Kraut- und Knollenfäule verursachenden Oomyzeten *Phytophthora infestans*. Die Bekämpfung dieses pilzähnlichen Schädlings ist primär mit dem Einsatz von toxischen Pflanzenschutzmitteln verbunden, die die Produktionskosten erhöhen. Allerdings ist die Pflanze diesem Erreger nicht schutzlos ausgeliefert. Sie verfügt über eine Klasse von Genen, die in der Lage ist, Resistenz gegenüber dem Pathogen zu vermitteln. Diese Gene werden als R-Gene bezeichnet und spielen eine Schlüsselrolle in dem durch Effektoren induzierten programmierten Zelltod, der so genannten Hypersensitiven Antwort.

*R1* gehört zu der Hauptklasse dieser Resistenzgene, den NBS-LRR Genen. Es ist bekannt, dass dieses Gen in der Lage ist, eine rassenspezifische Resistenz zu vermitteln. Durch Sequenzierung des chromosomalen Bereichs um *R1* wurden acht offene Leseraster entdeckt, die eine große Sequenzähnlichkeit zu dem *R1* Gen aufweisen und Teil des Resistenz-Clusters auf Chromosom V von *Solanum tuberosum* sind. Die mit *R1* insgesamt neun Gene zeigen untereinander eine größere Sequenzähnlichkeit als zu jedem anderen Resistenzgen in Kartoffel auf und werden als die *R1*-Familie bezeichnet. Innerhalb dieses Projekts wurde eine molekulare Charakterisierung dieser Familie und der durch *R1* induzierten Abwehrgene durchgeführt.

Das Hauptziel bestand darin, den *R1*-vermittelten Phänotyp molekular zu beschreiben. Um dieses Ziel zu erreichen, wurde eine vergleichende Transkriptomanalyse der Sorte Desirée in der Anfangsphase der Infektion mit Krautfäule durchgeführt. In diesem Zusammenhang wurden verschiedene transgene Linien getestet die entweder das *R1*-Gen enthielten oder nicht. Zudem wurde eine transgene Linie miteinbezogen, die das Gen ORF45 erhielt. ORF45 teilt eine hohe Sequenzähnlichkeit mit dem *R1*-Gen, vermittelt aber nicht den bekannten Resistenzphänotyp.

Um den höchstmöglichen Datendurchsatz zu generieren, wurden eine neue Sequenziermethode der nächsten Generation eingesetzt in Kombination mit der DeepSAGE-Methode. Die Leistungsparameter dieses technischen Ansatzes wurden validiert. Im Verlauf die-

ses Projekts wurden verschiedenste Verfahren zur Datenanalyse getestet, letztendlich wurde eine Auswerte-Pipeline zusammengestellt, die in der Lage ist, die gewonnenen Daten nach aktuellem Erkenntnisstand aufzubereiten und auszuwerten.

Die Analysen zeigten, dass in dieser ersten Phase der Infektion hormonelle Signale in der Pflanze eine große Rolle spielen. Zudem konnten innerhalb dieser Schlüsselprozesse Unterschiede zwischen der kompatiblen (anfälligen) und der inkompatiblen Interaktion (resistenten) beobachtet werden.

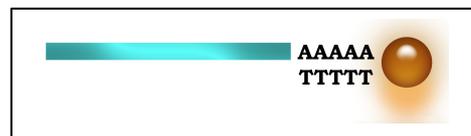
Darüber hinaus gelang es innerhalb der durchgeführten Arbeit, den experimentellen Beweis für die Expression von fünf der offenen Leseraster der *R1*-Familie zu erbringen und einen cytosolischen Resistenzmechanismus in *Solanum nigrum* zu beweisen, der in der Lage ist, Effektormoleküle von *Phytophthora infestans* zu erkennen.

## Protocol for the creation of Ditag-libraries

(the protocol has been kindly provided by Prof. K. Lehmann Nielsen)

### Binding mRNA to Magnetic Beads

1. Thoroughly resuspend the oligo(dT) beads and transfer 10  $\mu$ l to well of a 96-well microtiter plate.
2. Place the plate on plate-magnet until beads are located along the sides and carefully remove the sup. and discard it.
3. Wash beads by resuspending them in 100  $\mu$ l of Lysis Buffer.
4. Place plate on plate-magnet beads are located along the sides.
5. Prepare RNA sample for binding to the beads.  
Use 2  $\mu$ g total RNA and adjust the volume to 150  $\mu$ l with Lysis Buffer
6. Carefully remove the sup and immediately add your RNA sample to the beads



### mRNA Binding

1. After loading the entire 150  $\mu$ l of RNA sample to oligo(dT) beads equilibrated with Lysis Buffer. Seal plate.
2. Mix beads and RNA sample by slowly vibrating plate on a vibrating platform for 30 min at RT and 650 rpm.
3. Place plate on plate-magnet until beads are located along the sides and carefully remove sup.
4. Wash three times with 100  $\mu$ l Wash Buffer B.
5. Wash beads 2 times with 50  $\mu$ l 1X First Strand Buffer by placing the plate on plate-magnet for 1 min and removing the sup between washes. On the second wash, **DO NOT** remove supernatant.
 

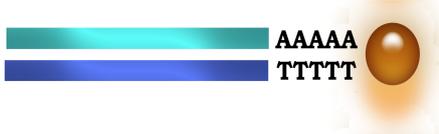
10 ml	1X First Strand Buffer:
2 ml	5X First Strand Buffer
8 ml	DNA H <sub>2</sub> O
6. Before removing sup prepare the first strand cDNA reaction mix.  
Mix the following reagent on ice per sample:

96 samples (x110):

2.4	μl	5X First Strand Buffer	264	μl
0.13	μl	RNase inhibitor (40 U/μl)	14.3	μl
7.6	μl	DEPC H <sub>2</sub> O	836	μl
1.2	μl	0.1 M DTT	132	μl
0.24	μl	dNTP Mix (25 mM)	26.4	μl

7. Incubate the mix at 37°C for 2 min to equilibrate the reagents.
8. Add 47.3 μl SuperScript™ III Reverse Transcriptase (200 U/μl) to the mix.
9. Remove the sup after the second wash and proceed to First Strand Synthesis.

### First Strand cDNA Synthesis



In the 1<sup>st</sup> strand synthesis of cDNA, reverse transcriptase transcribe mRNAs (gene transcript, light blue) into single stranded DNA (Dark blue).

1. Resuspend beads containing mRNA sample in 12 μl first strand cDNA reaction mix.
2. Mix gently and incubate at 37°C for 1 h. Mix gently at every 15 min by moving plate over plate-magnet.
3. Chill the first strand reaction on ice for 2 min and proceed to Second Strand cDNA Synthesis.

### Second Strand cDNA Synthesis



RNase H cleaves the 3'-O-P bond in the RNA-DNA duplex. DNA polymerase replicates the single stranded DNA into double stranded cDNA. DNA ligase link together the DNA strands.

1. Mix the following second strand reagents **in order** and add 80 μl of mix to each well containing 12 μl of the first strand reaction:

Mix for 96 samples (x 100)

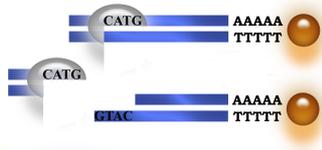
59	μl	DNA H <sub>2</sub> O	5900	μl
17	μl	5X Second Strand Buffer	1700	μl
0.75	μl	dNTP (25 mM)	75	μl
0.65	μl	<i>E. coli</i> DNA ligase (10 U/μl)	65	μl
2.5	μl	<i>E. coli</i> DNA polymerase (10 U/μl)	250	μl
0.25	μl	<i>E. coli</i> RNase H (5 U/μl)	25	μl

2. Incubate reaction mixture at 16°C for 2 h, mixing gently every 30 min by moving plate over plate-magnet. During incubation, preheat Wash Buffer C to 75°C.
3. Place plate on ice and add 30 µl 0.1 M EDTA to stop the reaction.
4. Place plate on plate-magnet for 1 min and carefully remove sup. Add 50 µl warm Wash Buffer C to inactivate the *E. coli* DNA polymerase.
5. Mix well and heat the sample to 75°C for 10 min with intermittent mixing to completely inactivate the polymerase.
6. Place plate on plate-magnet for 1 min. and remove sup.
7. Wash again with 75 µl warm Wash Buffer C. Perform the wash quickly to prevent precipitation of SDS which may trap the beads.
8. Wash sample four times with 75 µl Wash Buffer D.
9. Place plate on plate-magnet until beads are located along the sides and carefully remove sup.
10. Add 25 µl 1X Buffer 4 to each sample well and gently resuspend beads. Transfer the contents to a new well to avoid any traces of exonuclease activity from *E. coli* DNA polymerase. Wash the wells of the old plate with 25 µl 1X Buffer 4 and transfer the contents to the new well containing the reaction mix.

**10 ml      1X Buffer 4:**  
 1 ml      10X Buffer 4  
 9 ml      DNA H<sub>2</sub>O

11. Place plate on plate-magnet until beads are located along the sides and remove sup.
12. Wash plate once with 50 µl 1X Buffer 4.
13. Remove sup. and proceed to Nla III Digestion.

### **Nla III Digestion**

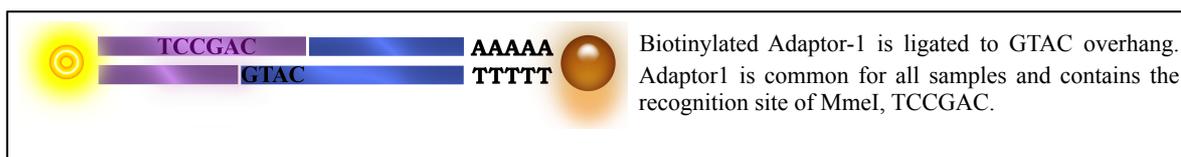


NlaIII recognizes the 4 bp sequence, CATG, which occurs approximately for every 256 bp in a transcript. NlaIII digests to the 3' most recognition site. Transcripts, which do not contain the recognition site, are omitted from the analysis. NlaIII and cDNA released from Oligo(dT) beads are removed during washing of beads.

- Resuspend mag. beads in 25  $\mu$ l of the following mix:
 

21.5 $\mu$ l	LoTE	2236	$\mu$ l
0.25 $\mu$ l	100X BSA	26	$\mu$ l
2.5 $\mu$ l	10X Buffer 4	260	$\mu$ l
0.75 $\mu$ l	Nla III (10 U/ $\mu$ l)	78	$\mu$ l
- Incubate for 1 h at 37°C. Mix occasionally.
- Meanwhile heat Wash Buffer C to 37°C to prevent SDS precipitation.
- After the reaction is complete, place plate on plate-magnet until beads are located along the sides and carefully discard sup.
- Inactivate Nla III by washing each sample twice with 75  $\mu$ l warm Wash Buffer C.
- Wash sample three times with 75  $\mu$ l Wash Buffer D.
- Proceed to Ligating Adaptors to the cDNA or store plate at 4°C ON.

### Ligating Adaptor A to the cDNA



- Place plate on plate-magnet for 1 min and carefully remove sup.
- Wash beads twice with 50  $\mu$ l of 1X T4 DNA Ligase Buffer.
 

<b>10 ml</b>	<b>1X T4 DNA Ligase Buffer:</b>
1 ml	10X T4 DNA Ligase Buffer
9 ml	DNA H <sub>2</sub> O
- Place plate on plate-magnet until beads are located along the sides and carefully remove sup.
- Transfer the plate to ice and add the following reagent to the beads on ice:
 

Reagent	1 Sample	Mix for 96 samples (x 104)
Adaptor A (90 ng/ $\mu$ l)	1 $\mu$ l	104 $\mu$ l
LoTE	12.5 $\mu$ l	1300 $\mu$ l
10X T4 DNA Ligase Buffer	1.5 $\mu$ l	156 $\mu$ l
- Add 15  $\mu$ l of mix to each well.
- Resuspend the beads by moving plate over plate-magnet. Heat the plate for at least 2 min. at 50°C.
- Cool plate for 15 min at RT and then chill on ice.

8. Mix the following:

Reagent	1 Sample	Mix for 96 samples (x104)
LoTE	11.5 $\mu$ l	1196 $\mu$ l
10X T4 DNA Ligase Buffer	1.5 $\mu$ l	156 $\mu$ l
T4 DNA Ligase (5 U/ $\mu$ l)	2 $\mu$ l	208 $\mu$ l

9. Add 15  $\mu$ l of T4 DNA ligase mix to each well and mix well.

10. Incubate ON at 16°C.

11. The following day wash each sample three times with 50  $\mu$ l of Wash Buffer D.

### Cleaving with Tagging Enzyme

Prepare 10X SAM:

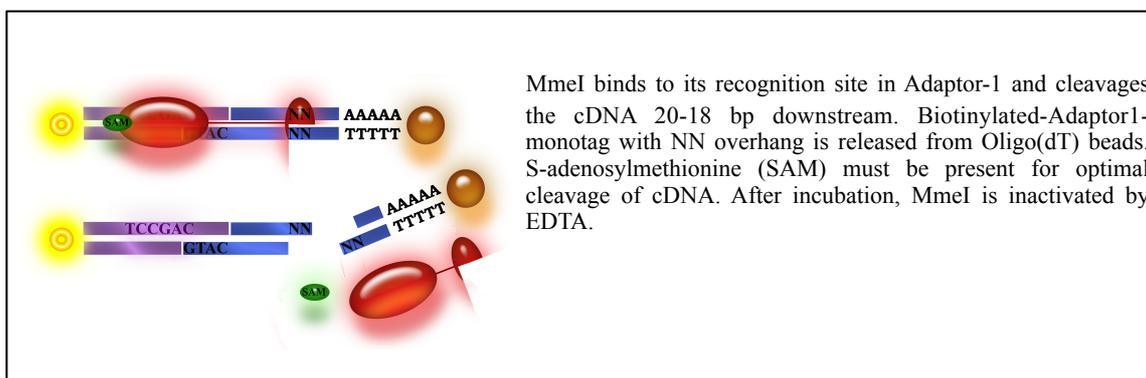
2 $\mu$ l	32-mM SAM (New England Biolabs)
158 $\mu$ l	DNA H <sub>2</sub> O

$$V_{\text{tot}} = 160 \mu\text{l}$$

Prepare 10 ml 1X Buffer 4/1X SAM:

1 ml	10X Buffer 4
9 ml	DNA H <sub>2</sub> O
12.5 $\mu$ l	32-mM SAM (New England Biolabs)

### MmeI Digestion

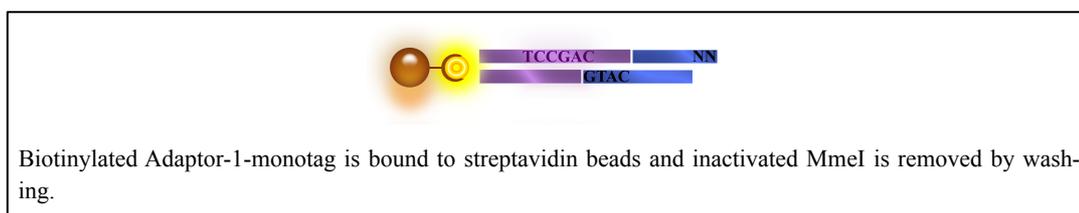


1. Place plate on plate-magnet until beads are located along the sides and remove sup.
2. Wash each well twice with 50  $\mu$ l 1X Buffer 4/1X SAM. Carefully remove and discard the sup and place plate on ice.
3. Add 15  $\mu$ l of the following to each sample: Mix for 96 samples (x104 )

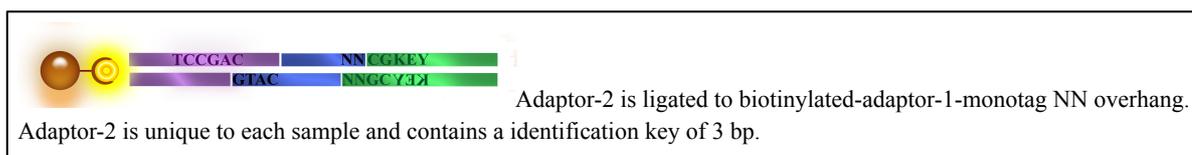
10.5 $\mu$ l	LoTE	1092 $\mu$ l
1.5 $\mu$ l	10X Buffer 4	156 $\mu$ l
1.5 $\mu$ l	10X SAM (400 $\mu$ M)	156 $\mu$ l
1.5 $\mu$ l	MmeI (2 U/ $\mu$ l)	156 $\mu$ l

4. Incubate plate at 37°C for 2.5 h with occasional gentle mixing.
5. 20  $\mu$ l 50 mM EDTA.
6. Place plate on plate-magnet until beads are located along the sides. **Do not discard the supernatant!**
7. Prepare Streptavidin beads.

### Binding linker-monotag to Streptavidin Beads



1. Thoroughly resuspend the Streptavidin beads and transfer 5  $\mu$ l to each well in a new plate.
2. Place plate, on plate-magnet until beads are located along the sides and carefully remove the sup.
3. Wash beads by resuspending them in 50  $\mu$ l 2X Binding and Washing Buffer.
4. Place plate on plate-magnet for 1 min and remove sup.
5. Add 35  $\mu$ l 2X Binding and Washing Buffer.
6. Add 35  $\mu$ l linker-monotag sample to well with 35  $\mu$ l 2X Binding and Washing Buffer.
7. Add 50  $\mu$ l 1X Binding and Washing Buffer.
8. Mix beads and linker-monotag sample by vibrating the plate on a vibrating platform for 15 min at RT and 600 rpm.
9. Place plate on plate-magnet until beads are located along the sides and carefully remove sup.
10. Wash twice with 50  $\mu$ l 1X Binding and Washing Buffer

**Ligating LS Adaptor B LinkerA-Monotag**

12. Place plate on plate-magnet for 1 min and carefully remove sup.

13. Wash beads twice with 50  $\mu$ l of 1X T4 DNA Ligase Buffer.

10 ml	<b>1X T4 DNA Ligase Buffer</b>
1 ml	10X T4 DNA Ligase Buffer
9 ml	DNA H <sub>2</sub> O

14. Place plate on plate-magnet until beads are located along the sides and carefully remove sup.

15. Transfer plate to ice and add the following reagent to the beads on ice:

Reagent	Sample	Mix for 8 samples (x10)
Adaptor B (90 ng/ $\mu$ l)	1 $\mu$ l	10 $\mu$ l
LoTE	12.5 $\mu$ l	125 $\mu$ l
10X T4 DNA Ligase Buffer	1.5 $\mu$ l	15 $\mu$ l

16. Add 15  $\mu$ l of mix to each well. **Remember to change Adaptor B** (max 8 samples per adaptor B per plate)!!!

17. Resuspend the beads by moving plate over plate-magnet. Heat plate for at least 2 min. at 50°C.

18. Cool plate for 15 min at RT and then chill on ice.

19. Mix the following:

Reagent	1 Sample	Mix for 96 samples (x104)
LoTE	11.5 $\mu$ l	1196 $\mu$ l
10X T4 DNA Ligase Buffer	1.5 $\mu$ l	156 $\mu$ l
T4 DNA Ligase (5 U/ $\mu$ l)	2 $\mu$ l	208 $\mu$ l

20. Add 15  $\mu$ l of T4 DNA ligase mix to each well and mix well.

21. Incubate ON at 16°C.

22. The following day ( ) wash each sample three times with 50  $\mu$ l of Wash Buffer D.

23. Wash twice with 50  $\mu$ l LoTE.

24. Resuspend beads in 20  $\mu$ l LoTE.

**PCR Amplification of linker-monotag-linker**

1. Before PCR amplification make the following dilution of the sample: 1/100.
2. Use the following PCR mix for each reaction (RXN):
3. Add 10  $\mu$ l cold DNA H<sub>2</sub>O to each well
4. Add 1  $\mu$ l of template (1/100)
5. Make the following master mix and add 40  $\mu$ l to each sample:

			<b>Mix for 96 samples (x104)</b>		
27.5	$\mu$ l	cold DNA H <sub>2</sub> O	2860	$\mu$ l	cold DNA H <sub>2</sub> O
5	$\mu$ l	DreamTaq Buffer	520	$\mu$ l	10X PCR Buffer
1	$\mu$ l	Primer mix (SOL 1+2)	104	$\mu$ l	Primer mix (SOL 1+2)
4	$\mu$ l	MgCl <sub>2</sub> (25 mM)	416	$\mu$ l	MgCl <sub>2</sub> (25 mM)
1	$\mu$ l	dNTP mix (25 mM)	104	$\mu$ l	dNTP mix (25 mM)
0.5	$\mu$ l	DreamTaq polymerase (5 U/ $\mu$ l)	52	$\mu$ l	DreamTaq polymerase
<b>(Dreamtaq buffer contains 2 mM MgCl<sub>2</sub>)</b>					

6. Use the following PCR program (DIT2):

Temperature	Time	Cycles
94°C	1 min	1
94°C	30 sec	28
53.5°C	1 min	
70°C	1 min	
4°C	~	~

**Analyze the PCR Product**

1. Use a 15% polyacrylamide/bis gel.
2. Load 2  $\mu$ l 25 bp DNA ladder.
3. Load 5  $\mu$ l PCR sample + 1  $\mu$ l loading buffer.
4. Load 2  $\mu$ l 25 bp DNA ladder.
5. Gel electrophoresis was performed in 1X TAE (Running buffer):
6. Stain the gel for 5 min in 25 ml 1X TAE + 5  $\mu$ l EtBr.

## Instructions for the supplementary data

### Go-Term analysis:

Bingo-output architecture:

	A	B	C	D	E	F	G	H	I	J	
	1 File created with BINGO (c) on 07.05.2010 at 11:14:56										
	2										
	3 ontology: all										
	4 curator: GO										
	5										
Internal calculation parameters	6 Selected ontology file : jar:file:/Users/gaborgyetvai/cytoscape/2.8/plugins/BINGO-2.31/BINGO.jar!/GO_Full										
	7 Selected annotation file : /Programme/Cytoscape_v2.8.3/bingo_annotation_file										
	8 Discarded evidence codes :										
	9 Overrepresentation										
	10 Selected statistical test : Hypergeometric test										
	11 Selected correction : Benjamini & Hochberg False Discovery Rate (FDR) correction										
	12 Selected significance level : 0.05										
	13 Testing option : Test cluster versus whole annotation										
	14 The selected cluster:										
Tentative consensus, which were in the tested list	15	TC191323	TC166958	TC187473	TC171965	TC167014	TC163864	CV499228	TC187366	TC192472	TC163524
	16	TC190297	TC180384	TC166257	TC165368	CV502319	CK863857	TC194408	TC184224	TC169662	TC175872
	17	TC170314	TC188800	TC167304	TC165183	TC189501	TC180326	DN923272	TC172077	TC169388	CV503631
	18	TC177428	DN686761	TC163335	TC167994	TC164234	TC174947	TC174073	TC187670	TC182950	TC186710
	19	TC180505	CV497998	TC166561	CV435238	CV505724	TC190730	TC175457	TC186830	TC167490	TC169760
	20	TC181826	TC190549	EG012828	TC168460	TC174895	TC164021	TC167895	TC170661	TC172462	GB X93564
	21	TC172874	TC189088	TC174577	TC174281	TC168882	TC182358				
	22										
Not identified sequences	23 No annotations were retrieved for the following entities:										
	24	TC187473	TC171965	CV499228	TC187366	TC192472	TC188139	TC176637	TC173307	TC193093	TC168117
	25	CK268082	TC175542	TC168403	TC179458	TC185461	TC188274	TC178118	TC170393	CV492775	TC190946
	26	TC167730	TC168758	CV504343	DN938054	CV469939	EG012490	TC169536	DR035737	BQ118938	TC171210
	27	TC166916	TC182945	TC172017	TC188062	TC164697	TC172266	DV622757	TC181895	DN688905	TC182988
	28										
	GO-ID	p-value	corr p-value	x	nX	N	Description	Genes in test set			
29	44435	2,03E-015	5,20E-012	90	266	681	4527 plastid part	TC172110 TC187144 TC1			
30	44434	3,62E-015	5,20E-012	84	242	681	4527 chloroplast par	TC172110 TC187144 TC1			
31	9507	5,98E-015	5,72E-012	96	297	681	4527 chloroplast	TC172110 TC187144 TC1			
32	31984	2,40E-014	1,72E-011	67	177	681	4527 oranelle subo	TC187144 TC172110 TC1			
33	31978	4,69E-014	2,24E-011	66	175	681	4527 plastid thylakoi	TC187144 TC172110 TC1			
34	9534	4,69E-014	2,24E-011	66	175	681	4527 chloroplast thyl	TC187144 TC172110 TC1			
35	9535	9,49E-014	3,03E-011	63	165	681	4527 chloroplast thyl	TC187144 TC172110 TC1			
36	55035	9,49E-014	3,03E-011	63	165	681	4527 plastid thylakoi	TC187144 TC172110 TC1			
37	42851	9,49E-014	3,03E-011	63	165	681	4527 thylakoid mem	TC187144 TC172110 TC1			
38	9579	1,06E-013	3,05E-011	69	190	681	4527 thylakoid	TC187144 TC172110 TC1			
39	44438	1,72E-013	4,49E-011	65	175	681	4527 thylakoid part	TC187144 TC172110 TC1			
40											

Information about the GO-term identifier, the p-value before and after multiple testing correction, the number of TC's identified for a term, the size of the GO-term, the number of tested TC's, the number of TC's in the GO-annotation file, the Go-term description and the TC's in the category

### Tables of Cluster analysis:

This table contains following informations:

Rowname	Description
UniqueID	StET-identifier
Name	Annotation
Sample names (29x)	contains the logfoldchange values (to the 0 dpi value of the sample)

**Expression-data of the DeepSAGEexperiment:**

This table contains following informations:

Rowname	Description
Tag-ID	StET-identifier
Sequence	The sequence inforamtion
Sample name (44x)	measured rel. expression (in counts/million)
Mean_All	Average expression value (over all samples)
prim. Annot. ID	primary identified TC or EST number
prim. Annot. name	primary annotation
pos.	position within the identified sequence
No tags	Number of identified genes for this sequence
All. Annot	Annotation of all possible genes
No Mismatch	Number of mismatches in the identified tag
grade	grade of identification (A= perfect match 3' end; B= perfect match 5' end; C=internal position or 1 not perfect match)
strand	information about the identified strand of the identified sequence)

**Fisher's exact test results:**

This table contains the same information like the expression data table and is expandet by following:

Rowname	Description
min p	lowest calculated p-value in a replica group
max p	highest calculated p-value in a replica group
Average fold change	The mean value of the fold changes
min fold change	the lowest observed fold change value
max fold change	the highest observed fold change value
No TRUE (tested line)	The number of significant comparisons as described in the thesis
No TRUE (tested line2)	For comparison the result of an opposite indicator line, in case of transgenics, the wildtype results are given and in case of wildtype the results of <i>R1</i> -transgenic line are given

**Comprehensive table of all calculated p-values for the comparison of statistical tests:**

This table contains the same information like the expression data table for wildtype lines at 0 dpi and 1 dpi and is expanded by following information:

Rowname	Description
BaySeq; logP	logP values after calculation of BaySeq
edgeR; FDR	FDR-corrected p-values after implementation of edgeR
Permu; rawP and FDR	calculated p-values of the permutation test before and after multiple testing correction
Fisher's exact test; R1-R6	the calculated (corrected) p-values after analysis with sagenhaft for each replica
G-test; Intern, Within 0dpi, 1 dpi and 0 dpi, Within 1dpi	The calculated p-values of G-test for each rule - meaning overall significance, within the group of values for 0dpi, between the group of 1 dpi and 0 dpi and within the group of 1 dpi
T-Test; p	the (corrected) p-values after application of Student's t-test

Rowname	Description
Logfoldchange values; lfc1-lfc6, A.lfc	The logfoldchange values for each replica and the average logfoldchange between 0 dpi and 1 dpi

### Locations of raw data files:

Folder	Filename	Description
/Raw-data/adds	Bingo_outputs.xls	Output files of the GO-term analysis in form of an excel sheet
	GO0006952_Genes.xls	List of all tags identified for the GO-term "defense response"
	Hierarchical Cluster.png	Picture of the hierarchical cluster heatmap, which has been calculated for all tags with grade A and B annotation
	PCA_Principal components1-6.xls	Results of the principal component analysis for the Principal components 1-6
	Potato_Phytophthora.fasta	Multiple fasta file, which contains the used Potato gene index sequences and the published genomic sequence of Phytophthora infestans (basis file for the tag annotation)
	ReferenzdatensatzPoMaMoedit.xls	Table file of all relative expression values gained from the DeepSAGE experiment (in counts/million)
	SUPERPEXRDv3.1sequences.xls	Sequences of the putative Effectors, which have been used in the transient expression screen in Solanum nigrum
/Raw-data/FE-Results	Sig_downgenes_FE_cleared.xls	Table of all significant downregulated tags after Fisher's exact test
	Sig_upgenes_FE_cleared.xls	Table of all significant upregulated tags after Fisher's exact test
/Raw-data/Hierarchical clustering_Clusterextract	Selected Clustertreearms.xls	Table of all marked cluster areas after the hierarchical clustering of the contents of selected k-mean clusters
/Raw-data/k-mean-tables	Cl....xls	Collection of excel sheet which the contents of the clusters calculated during k-mean cluster analysis
/Raw-data/Stat eval	All significantwithDf and alfc.xls	excel sheet containing the data of the wildtype lines at 0 dpi and 1 dpi of all as significant calculated tags – additionally the Df-values and the average logfold-change values are given

	Referenzdatensatz_wildtyp_1dpi_with all p-values.xls	excel sheet containing the expression values and all calculated p-values for the comparison of 0 dpi and 1 dpi in wildtype lines. The table is expanded with the log-foldchange values.
/Raw-data/STATS-Gabor_Lee_AHP	All.htm	please open just this file – leads to a collection of overall pherogram scores of each section of the Illumina sequencing (note that the irritation in the graphs at the end originate from the CATG-Key). The number behind the “s” indicates the lane (up to 8). The following number gives the measured section,

**Eidesstattliche Erklärung**

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen - die anderen Werken im Wortlaut oder Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität vorgelegen hat; dass sie noch nicht veröffentlicht worden ist, sowie dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Dr. PD Christiane Gebhardt betreut worden.

Köln, der \_\_\_\_\_

\_\_\_\_\_  
(Gabor Gyetvai)

## Danksagungen

Ein erster Dank geht an meine Betreuerin Christiane Gebhardt, die es geschafft hat, auch diesen Jungen zu promovieren. Ich danke dir Christiane, dass du auch einfach mal unterschiedliche Meinungen stehen lassen konntest und wirklich nur korrigierend eingegriffen hast, wenn es nicht mehr anders ging. Zu keinem Zeitpunkt, fühlte ich mich nicht frei in meiner Forschung und ich denke das ist ein guter Weg, Wissenschaft zu lernen und auf jeden Fall nicht selbstverständlich.

Ebenso möchte ich mich bei Agim bedanken, der als mein „Schattensupervisor“ immer wieder wichtige Ratschläge, Diskussionen und Rat allen Belangen für mich parat hatte, wenn es um eine Klonierung ging, darüber, dass der Pilz mal wieder nicht will oder einfach nur, wo denn nun die Schlacht bei den Thermopylen genau war.

Zudem möchte ich der kompletten Arbeitsgruppe danken, die es immer wieder verstanden hat, ein großartiges Arbeitsklima zu schaffen und einen Ort, an dem man gerne morgens aufgetaucht ist, um wieder und wieder RNA zu extrahieren und wieder und wieder Proben in den Müll zu werfen, die dann mal einfach für die Tonne waren.... bis man es dann nach drei Jahren endlich konnte. In diesem Zug möchte ich besonders Birgit hervorheben, ohne deren Engagement (trotz fester Stelle), diese Gruppe nicht so gut funktionieren würde, wie sie es tut. Außerdem möchte ich meinen beiden Lieblingsdoktorandinnen Lena und Camilla dafür danken, dass der Status des „Senior-PhD“ eingeführt wurde - auf irgendeine Art und ich weiß nicht genau wie, wirkte es unglaublich motivierend - vielleicht, weil es einem ins Gedächtnis gerufen hat, dass man ja eigentlich gar nicht mehr soweit von seinem Ziel entfernt ist - ich freue mich schon darauf, diesen Titel an Camilla abzugeben.

Außerdem möchte ich meinen vielen Helfern danken. Im besonderen, dir Christine, dafür, dass du immer wieder den Pilz im Leben gehalten hast und in den heißen Phasen, kleine Armeen und Bataillone von kleinen Kartoffeln aus der Zellkultur in die Klimakammern ge-

geben hast und dann natürlich Caren, die diese Armeen dann auch entgegen genommen hat und dafür gesorgt hat, dass aus ihnen richtige Pflanzen werden. Danke dir Caren, dass du immer wieder mitgemacht hast - dein Einsatz und vor Allem die Kooperationsbereitschaft hier ging über das normale Maß hinaus. Auch dir, Sandra möchte ich danken, dass du mir gerade am Anfang mit diesem komischen Pilz (, der ja nichtmal ein richtiger Pilz ist - doofes Teil) geholfen hast und das in einer Zeit, wo er irgendwie nur noch rätselhaft war - mal wächst er, mal nicht. Du warst genau richtig an dem Ort - immerhin hast du nicht gleich das Harakiri-Schwert rausgenommen, wenn zum x-ten mal keine Zoosporen weit und breit zu sehen waren.

Ein weiterer Dank gilt meinen Kooperationspartnern, Kåre und Mads in Aalborg, die über den gesamten Zeitraum immer wieder schnell meine Anfragen bearbeitet haben und was noch wichtiger ist auch weiter Interesse an der Arbeit hatten -auch das kann anders laufen - saubere Sache. Auch möchte ich den Bioinformatikern danken, klar Ulrike, dir vorne weg. Du hast dir immer wieder die Probleme angehört und dann reale Hilfe zur Selbsthilfe gegeben - ohne dich hätte ich wohl nie R gelernt und ich weiß nicht warum, aber ich denke mal, dass werde ich in Zukunft wirklich noch gebrauchen können. Und auch dem Team in Golm um Birgit - Rico, Jost und Maren - ihr habt dann, wenn´s drauf ankam immer gut mitgezogen... und das oftmals auch einfach so - nein! Über meine Kooperationspartner kann ich mich wirklich nicht beschweren.

Und nun noch die besonderen Dankeschöns - Claude, dir Dank, dass du mich immer wieder mit neuem und viel zu teurem Wein versorgt hast - du weißt ja, dass wir bald zum Portugiesen müssen und dafür, dass du mir auch immer wieder erklärt hast, wie der Laden so läuft und tolle Geschichten vom „Lederhosenmann“ und dem Spinner aus Kanada im Schlepptau hattest. Und natürlich wäre ich insgesamt nicht hier, wenn meine Mutter damals nicht gefragt hätte, ob ich auf´s Gymnasium möchte oder auf die Realschule, obwohl

ich nur die Realschulempfehlung hatte. Das stand ja gar nicht zur Diskussion - „Gymnasium!“ kam aus dem 12jährigen Mund - und recht gehabt hat er... wie er immer noch immer wieder Recht hat - und siehe da - jetzt kommt da ein Doktor raus... ansonsten hätte ich wohl was ordentliches gelernt. Nein, im Ernst - ohne deine Hilfe und dein Vertrauen, wär's wohl nicht gegangen.

Und dann noch du Sabrina, immerhin bist du noch da - wer hätte das gedacht - all die Stationen... all die Zeit - wo ist sie hin - als wir uns kennen lernten, war ich AutoklavenHiwi und mein Po-Dekoltée schaute dauernd raus - und du warst in Stufe 2... und jetzt bin ich bald dann doch mal fertig mit studieren und du in Stufe 4 - das ist schon viel Zeit, die du da miterlebt hast - gerade jetzt am Ende war ich froh, dass du hier warst und alles ausgehalten hast... meinen körperlichen und seelischen Verfall - dich darum kümmerst, dass ich dusche und mir die Zähne putze - und immerhin dafür sorgst, dass ich zumindest die wichtigsten Dinge nicht vergesse... das komische an so einer Endphase ist, dass das Leben und die Zeit nicht aufhört zu schlagen - man selbst aber in einer Wolke steckt - einfach nur danke, dass du da bist.

## Lebenslauf

Name: Gyetvai, Gábor Miklos

Anschrift: Regentenstr. 22  
D-51063 Cologne

Tel.: 0049(0)221-1706219

Geburtsdatum und -ort: 03.03.1979 in Gronau

Familienstand: ledig

Nationalität: deutsch

### Bildung:

April 2007 – heute Anfertigen der Dissertation am Max Planck Institut für Züchtungsforschung in Köln

Juli 2006 – März 2007 Anfertigen der Diplomarbeit am Max Planck Institut für Züchtungsforschung und Erwerb des Diploms in Biologie an der Universität Osnabrück (Dipl.-Biol.)

September 2001 – März 2007 Studium der Biologie (Dipl.) an der Universität Osnabrück

Oktober 1999 – September 2001 Studium der Biologie (Dipl.) an der Eberhard-Karls Universität Tübingen

Juni 1999 Beendigung der Schulbildung mit Erwerb der allgemeinen Hochschulreife am Ubbo-Emmius-Gymnasium in Leer (Ostfr.)

Köln, der \_\_\_\_\_

\_\_\_\_\_  
(Gabor Gyetvai)



