Genomics and Transcriptomics of the sebacinoid fungi *Piriformospora indica* and *Sebacina vermifera*

Dissertation

zur

Erlangung des Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

Dem Fachbereich Biologie

der Philipps-Universität Marburg

vorgelegt von

Urs Lahrmann
geboren am 27.04.1982 in Ankum

Marburg an der Lahn im November 2013

Die Untersuchungen der vorliegenden Arbeit wurden von April 2010 bis September 2013 unter Betreuung von Frau Dr. Alga Zuccaro in Marburg am Max-Planck-Institut für terrestrische Mikrobiologie in der Abteilung Organismische Interaktionen durchgeführt.

Vom Fachbereich

der Philipps-Universität Marburg als Dissertation

angenommen am: 07.02.2014

Erstgutachter: Frau Dr. Alga Zuccaro

Zweitgutachter: Herr Prof. Dr. Hans-Ulrich Mösch

Drittgutachter: Herr Prof. Dr. Michael Bölker

Tag der mündlichen Prüfung: 27.02.2014

Eigenständigkeitserklärung

Hiermit bestätige ich, dass ich die vorliegende Dissertation mit dem Titel „Genomics and Transcriptomics of the sebacinoid fungi *Piriformospora indica* and *Sebacina vermifera*“ selbstständig angefertigt habe und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Dissertation, welche dem Wortlaut oder dem Sinn nach anderen Werken entnommen sind, wurden unter Angabe der Quelle kenntlich gemacht.
Diese Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den 05.11.2013

Teile dieser Arbeit wurden in folgenden Artikeln veröffentlicht:

Lahrmann, U., Ding, Y., Banhara, A., Rath, M., Hajirezaei, M.R., Döhlemann, S., Wirén, N. von, Parniske, M., and Zuccaro, A. (2013). Host-related metabolic cues affect colonization strategies of a root endophyte. PNAS *110*, 13965–13970.

Lahrmann, U., and Zuccaro, A. (2012). *Opprimo ergo sum* - Evasion and Suppression in the Root Endophytic Fungus *Piriformospora indica*. Molecular Plant-Microbe Interactions *25*, 727–737.

Zuccaro, A., Lahrmann, U., Güldener, U., Langen, G., Pfiffi, S., Biedenkopf, D., Wong, P., Samans, B., Grimm, C., Basiewicz, M., et al. (2011). Endophytic Life Strategies Decoded by Genome and Transcriptome Analyses of the Mutualistic Root Symbiont *Piriformospora indica*. PLoS Pathog *7*, e1002290.

Zusammenfassung

Der Wurzelendophyt *Piriformospora indica* und der Orchideen-Mykorrhiza Pilz *Sebacina vermifera* (Sebacinales, Basidiomycota) sind in der Lage eine mutualistische Symbiose mit Pflanzen zu etablieren. Beide Pilze kolonisieren dabei die Wurzelrinde einer Vielzahl von Gefäßpflanzen, einschließlich der monokotyledonen Gerste (*Hordeum vulgare*) und des dikotyledonen Ackerschmalwands (*Arabidopsis thaliana*). Die Besiedelung der Wirtspflanze durch die Pilze führt zu einer Wachstumsförderung, sowie einer erhöhten Resistenz gegen abiotische und biotische Stressfaktoren. Während der Entwicklung der Pilze innerhalb der Wurzel sind diese sowohl in toten Zellen der Wurzelrinde, als auch in einer biotrophen Interaktion mit lebenden Pflanzenzellen zu finden. Diese Eigenschaften zusammen mit der Möglichkeit, die Pilze auf synthetischen Medien zu kultivieren, offenbaren eine umfangreiche phänotypische Flexibilität, welche sich auch in ihrem Erbgut widerspiegeln.

In dieser Arbeit wurden die Genome von *Piriformospora indica* und *Sebacina vermifera* charakterisiert. Es konnte gezeigt werden, dass bestimmte Gene und funktionelle Proteindomänen in beiden Spezies expandiert sind. Zu diesen gehören Proteine, für die eine Funktion im intra- und extrazellulären Transport (Transporter), in der Proteolyse (Peptidasen), im Abbau von Kohlenhydraten (Hydrolasen) und in der Bindung von Kohlenhydraten (Lektine) vorhergesagt wurde. Zusätzlich wurde eine neuartige Familie von kleinen sekretierten Proteinen in *P. indica* identifiziert, welche sich durch regelmäßig verteilte Histidine und Alanine, sowie ein konserviertes, sieben Aminosäuren-Motiv ("RSIDELD") am C-terminus der Proteine auszeichnet. Auf der anderen Seite konnte gezeigt werden, dass die Anzahl an Proteinen welche eine Funktion in der Produktion von Sekundärmetaboliten besitzen, insbesondere Polyketide und nicht-ribosomale Peptidsynthetasen, in beiden Pilzen stark reduziert ist. Dies ist bezeichnend für den nicht pathogenen Charakter von *P. indica* und *S. vermifera*.

Des weiteren konnte unter Verwendung von "Microarrays" und "RNA-Sequenzierung" gezeigt werden, dass die Expression von Genen in *P. indica* während der Besiedelung von Gersten- oder Arabidopsis-Wurzeln zeit- und wirtsabhängig reguliert ist. Eine erste vergleichende Untersuchung der Genexpression in *S. vermifera* und *P. indica* während der Besiedelung von Arabidopsis legt nahe, dass definierte Unterschiede während der Besiedelung dieses Wirtes durch die beiden untersuchten Pilze existieren.

Summary

The root endophyte *Piriformospora indica* and the orchid mycorrhiza *Sebacina vermifera* (Sebacinales, Basidiomycota) are able to establish a mutualistic symbiosis with plants. Both fungi colonize the root cortex of a wide range of vascular plants, including the monocot barley (*Hordeum vulgare*) and the dicot *Arabidopsis thaliana*. Colonization by the fungi results in growth promotion and induced resistance against abiotic and biotic stresses. Fungal development in roots combines biotrophic growth in living plant cells and cell-death associated colonization of dead cortex cells. These features together with the possibility to cultivate the fungi on synthetic media reveal substantial phenotypic plasticity which is reflected in their genomic traits.

In this study, the genomes of *Piriformospora indica* and *Sebacina vermifera* were characterized. It could be shown that certain gene and functional protein domain expansions occurred in both species. These included proteins predicted to be involved in intra- and extracellular transport (Transporters), proteolysis (Peptidases), degradation of carbohydrates (Hydrolases) and non-destructive carbohydrate binding (Lectins). Additionally, a novel family of small secreted proteins was identified in *P. indica* which is characterized by regular distributed histidine and alanine residues and a conserved seven amino acid motif ("RSIDELD") at the C-terminus. On the other side, the number proteins involved in secondary metabolism, in particular polyketide and nonribosomal peptide synthetases, were shown to be strongly reduced in both fungi which is indicative of the non pathogenic character of *P. indica* and *S. vermifera*.

By using microarrays and RNA-sequencing, a time- and host-specific expression of genes could be shown in *P. indica* during colonization of barley- or Arabidopsis roots. A first comparative analyses of genes expressed in *S. vermifera* and *P. indica* during colonization of Arabidopsis suggests that defined differences exist during the colonization of this host by both analysed fungi.

Abbreviations and technical terms

aa	Amino acids	JA	Jasmonic acid
AM	Arbuscular mycorrhiza	kb / kbp	kilo basepairs
AP	Animal pathogen	LRR	Leucine-rich repeat
Blast	Basic local alignment search tool	MAMP	Microbe associated molecular pattern
bp	Base pairs	Mb / Mbp	Megabases
BR	Brown rot	MTI	MAMP-triggered immunity
C	Carbon	min	minutes
cf.	compare	N	Nitrogen
cDNA	complementary DNA	N-terminal / 5'	aminoterminal
CM	Complete medium	NGS	Next generation sequencing
C-terminal / 3'	carboxyterminal	NH_4^+	Ammonium
CEG	Conserved eukaryotic genes	nr	non-redundant
CWDE	Cell wall degrading enzyme	NO_3^-	Nitrate
DNA	Deoxyribonucleic acid	ORF	Open reading frame
DMSO	Dimethylsulfoxid	PAGE	Polyacrylamide gel electrophoresis
d	days	P / P_i	Phosphorus
dpi	days post inoculation	PCR	Polymerase chain reaction
dpt	days post treatment	Plant pathogen	
e.g.	for example ("exempli gratia")	PRR	Pattern recognition receptors
e-Val	e-value threshold	qRT-PCR	quantitative real-time PCR
ECM	Ectomycorrhiza	RNA	Ribonucleic acid
EP	Endophyte	RNA-seq	RNA sequencing
ER	Endoplasmic reticulum	rpm	rounds per minute
EST	Expressed sequence tag	s	seconds
ETI	Effector-triggered immunity	SA	Salicylic acid
Fig	Figure	SD	Soil decomposer
GPD / GAPDH	Glycerinaldehyd-3-Phosphate-Dehydrogenase	SNP	Single nucleotide polymorphism
GFP	Green Fluorescence Protein	SSR	Simple sequence repeat
h	hours	Tab	Table
ddH_2O	double distilled water	Tris	Trishydroxymethylaminomethane
HMM	Hidden markov model	WGA	Wheat germ agglutinin
HR	Hypersensetive response	WR	White rot
Hyg	Hygromycin B	wt	Wildtype

Index of contents

1. Introduction 1

1.1. Interactions between plant and fungi 1

1.2. Fungal symbionts and endophytes 3

1.2.1. Sebacinales 4

1.2.1.1. *Piriformospora indica* 5

1.2.1.2. *Sebacina vermifera* 7

1.3. Genome research 7

1.3.1. Sequencing and sequence assembly 8

1.3.2. Fungal genome programs 10

1.4. Aims of this thesis 11

2. Results 13

2.1. Comparison and characterization of the *P. indica* and *S. vermifera* genomes 13

2.1.1. General genomic features within the Sebacinales 13

2.1.2. Comparative genomics 14

2.1.3. Annotation and analyses of protein coding genes 20

2.1.4. Key genes in the plant fungus interaction 24

2.1.4.1. Transmembrane transporter 25

2.1.4.2. Hydrolases and peptidases 27

2.1.4.3. Carbohydrate binding proteins 29

2.1.4.4. Secondary metabolite production 37

2.1.4.5. Effector-like proteins 38

2.2. Transcriptional regulations in *P. indica* and *S. vermifera* during host colonization 42

2.2.1. Transcriptional profiling of *P. indica* during barley and Arabidopsis colonization 43

2.2.1.1. Barley microarray design and evaluation 43

2.2.1.2. Arabidopsis microarray design and evaluation 47

2.2.1.3. Transcriptional upregulation of selected functional groups *in planta* 52

2.2.2. Conserved regulation patterns within the Sebacinales 57

2.2.3. Arabidopsis transcriptional responses to *P. indica* colonization 61

2.2.3.1. Biotic stress responses in Arabidopsis due to *P. indica* colonization 62

3. Discussion 67

3.1. The Sebacinales share genomic traits common in biotrophic and saprotrophic fungi 67

3.2. Gene/domain expansions in the Sebacinales and their regulation during plant colonization 67

3.2.1. Potential to avoid recognition by masking of elicitors 69

3.2.2. Hydrolytic degradation of plant cell wall components by Sebacinoid fungi 72

3.2.3. Proteolytic cleavage in the fungus plant interaction 75

3.2.4. Nutrient transport and detoxification mechanisms in *P. indica* and *S. vermifera* 76

3.2.5. Small secreted proteins and putative effector candidates 78

3.3. *P. indica* colonizes barley and Arabidopsis roots in a host-dependent way 79

3.4. Differences in the transcriptional response to colonization of Arabidopsis by *S. vermifera* and *P. indica* 82

3.5. Defence pathways transcriptionally induced in Arabidopsis during *P. indica* colonization 83

4. Material and Methods 86

4.1. Materials and sources of supply 86

4.1.1. Chemicals and "kits" 86

4.1.2. Buffers, solutions and media 86

4.1.3. Enzymes 86

4.1.4. Oligonucleotides 86

4.2. Strains and vectors 89

4.2.1. *E. coli* strains 89

4.2.2. *P. indica* and *S. vermifera* strains 90

4.2.3. *A. thaliana* and *H. vulgare* strains 90

4.2.4. Vectors 91

4.2.4.1. Vectors for TA cloning of PCR products 91

4.2.4.2. *P. indica* transformation vectors 91

4.3. Bacterial, fungal and plant cultivation 92

4.3.1. *E. coli* cultivation 92

4.3.2. *P. indica* and *S. vermifera* cultivation 92

4.3.3. *A. thaliana* and *H. vulgare* cultivation 95

4.4. Microbiological and biochemical methods 96

4.4.1. Heat shock transformation of *E. coli* 96

4.4.2. Collection of *P. indica* chlamydospores 97

4.4.3. Preparation of *S. vermifera* inoculation culture 98

4.4.4. *P. indica* - barley interaction studies 98

4.4.5. *P. indica* - Arabidopsis interaction studies 98

4.4.6. *S. vermifera* - Arabidopsis interaction studies 99

4.4.7. *P. indica* protoplast preparation 99

4.4.8. PEG-mediated transformation of *P. indica* protoplasts 100

4.5. Molecular methods 101

4.5.1. DNA isolation 101

4.5.2. RNA isolation 103

4.5.3. Polymerase chain reaction (PCR) 103

4.5.4. Quantitative real-time PCR 105

4.5.5. Agarose gel electrophoresis 106

4.5.6. TA cloning of PCR products 107
4.5.7. Southern blot 108
4.5.8. DIG labelling 109
4.5.9. Microarray experiments 110
4.5.9.1. Experimental design 110
4.5.9.2. Custom design of Agilent microarray chips 111
4.5.9.3. Labelling and hybridization of custom designed arrays 111
4.5.9.4. Microarray analyses and verification 112
4.5.10. DNA sequencing 112
4.5.11. RNA sequencing 113
4.6. Microscopy 113
4.7. Bioinformatic methods 113
4.7.1. JAVA programs 113
4.7.1.1. Augustus2fasta 114
4.7.1.2. BlastAnalyzer / BlastEntry 114
4.7.1.3. Blast2go2goeast 114
4.7.1.4. CodonUsage 115
4.7.1.5. FastaLoader / FastaEntry / DNAtranslator 115
4.7.1.6. IntergenicSpacePlot 116
4.7.1.7. MicroarrayDataSelection 116
4.7.1.8. MotifSeeker 116
4.7.1.9. MultipleFastaSelection 117
4.7.1.10. MultipleSequenceAnalyzer 117
4.7.1.11. Pfam2dog / Smart2dog 117
4.7.1.12. PfamScanAnalyzer / PfamScanEntry / LectinFinder 117

4.7.1.13. ProbeInfo / ProbeEvaluator 118
4.7.1.14. SequenceTableParser 119
4.7.1.15. SmartBatchParser 119
4.7.1.16. TargetPParser 119
4.7.1.17. TmhmmParser 120
4.7.1.18. WolfpsortParser 120
4.7.2. R scripts 120
4.7.2.1. MicroarrayAnalyses 120
4.7.2.2. PCA 121
4.7.2.3. MAPlot 121
4.7.2.4. DensityPlot 121
4.7.2.5. DomExpTTest 121
4.7.2.6. Heatmap 121
4.7.2.7. IntergenicSpacePlot_2D 122
4.7.3. Shell scripts 122
4.7.3.1. FastaStandardization 122
4.7.3.2. NucLocPrediction 122
4.7.3.3. PfamScanMultiExe 122
4.7.3.4. PfamScanSelection 122
4.7.3.5. SecretionPrediction 123
4.7.4. Comparative genomics 123
4.7.5. Identification of single nucleotide polymorphisms 126
4.7.6. Repetitive elements 127
4.7.7. Transfer RNAs / codon usage 127
4.7.8. Gene modelling 127

4.7.9. Annotation of predicted open reading frames 128

4.7.9.1. Blast2GO 128

4.7.9.2. Sub cellular localization of proteins 129

4.7.9.3. Automated secretome prediction 129

4.7.9.4. Functional protein domains 129

4.7.9.5. Specialized protein database 129

4.7.10. Phylogenetic analyses 130

4.7.11. Cluster analyses 131

4.7.12. Evolutionary analysis of protein families 131

4.7.13. Enrichment analysis 132

4.7.14. Assembly and analysis of the mitochondrion 132

4.7.15. Orthologues and Synteny prediction 132

5. References 134

6. Appendix 163

1. Introduction

The kingdom fungi encompasses an estimated number of 5 million species which are distributed all over the world living partially under extreme environmental conditions like heat, cold or high acidity (Blackwell, 2011; Selbmann et al., 2012; Sterflinger et al., 2012). These achievements are accompanied by the evolution of different lifestyles and different forms of interaction with other organisms in the same habitat. In this context, three major lifestyles can be differentiated in which the fungus growth as a saprotroph on dead or decaying organic matter, as symbiont with beneficial effects for its interaction partner or as parasite at the expense of the fitness of its interaction partner (Selbmann et al., 2012). Boundaries between these lifestyles however, are not fixed but rather fluent, meaning that a fungus is not restricted to one certain lifestyle and may switch it depending on different environmental stimuli (O'Connell et al., 2012; Vaario et al., 2012).

1.1. Interactions between plant and fungi

Fungi are heterotrophic organisms, lacking chlorophyll and vascular tissue, and the demand for nutritional resource therefore depends on the feeding on organic matter. In the vast majority, these organic compounds required for growth are plant derived and may be acquired using different strategies. In this context, the above stated fungal lifestyles are further refined using the terms saprotrophy, necrotrophy and facultative/obligate biotrophy (Lewis, 1973). While necrotrophs actively kill plant cells via the secretion of toxins and hydrolytic enzymes or via induction of the hypersensitive response (HR) in order to feed saprotrophically on the dead plant material, biotrophs depend on the interaction with living plant cells to satisfy their nutritional demands. These lifestyles can be transient depending on the host and its metabolic status (Kämper et al., 2006; López-Berges et al., 2010; Stuttmann et al., 2011; Tyler et al., 2006; Lahrmann et al., 2013). In any case, fungal colonization of a potential plant host is accompanied by the interaction with the innate plant immune system. This will recognize foreign organisms initially based on conserved microbe-associated molecular patterns (MAMP) which are characteristic and indispensable molecules often exposed at the microbes surface, like chitin in fungi or flagellin in bacteria (Boller and He, 2009; Lahrmann and Zuccaro, 2012; Miya et al., 2007; Pel and Pieterse, 2012). MAMPs are perceived by the plant via pattern recognition receptors (PRR) which may induce MAMP-triggered immunity (MTI) (He et al., 2007; Jones and Dangl, 2006). This first defence against the potentially harmful invader can include different intracellular responses like localized cell wall reinforcements,

rapid changes in gene expression and production of reactive-oxygen species (Luna et al., 2011; Zipfel, 2008). The establishment of a successful colonization (compatible interaction) depends therefore on the ability of the fungus to either suppress or evade recognition via PRRs. In order to identify essential determinants involved in the establishment of a compatible interaction, research in the last years has focused on the analyses of so called effector proteins (Jones and Dangl, 2006; de Jonge et al., 2010; Lahrmann and Zuccaro, 2012; Pel and Pieterse, 2012). These are small proteins that are secreted by the fungus and play an important role during the interaction with its plant host. Effector functions that were described so far are diverse and a fungus is probably not using one single strategy to overcome plant defence reactions. Such a multi level colonization strategy has for example been well described for *Cladosporium fulvum*, a biotrophic leaf pathogen of tomato. It could be shown that the fungus prevents chitin-elicited PRR responses via ECP6, a lectin-like protein, that specifically sequesters chitin fragments released from the fungal hyphae during penetration of a plant cell thereby hiding them from plant PRRs (de Jonge et al., 2010). As this is probably not an absolute protection, AVR4, another chitin-binding lectin-like protein, is secreted by the fungus to bind to exposed chitin at the outer surface of the fungal hyphae probably to shield the cell wall against plant derived chitinases (van Esse et al., 2007). Besides these passive strategies, *C. fulvum* also secretes AVR2, a general cysteine protease inhibitor, that prevents proteolytic degradation of fungal proteins in the plant apoplast (van Esse et al., 2008). While these were examples of effectors that have a function in the host–microbe interface, effectors might also be translocated into the host cell where they interfere with cytoplasmic processes or regulate target gene expression in the nucleus (Djamei et al., 2011; Plett et al., 2011). Plant hosts on the other hand may have evolved mechanisms to recognize fungal effectors leading eventually to HR via effector-triggered immunity (ETI) (Boller and He, 2009; Jones and Dangl, 2006). Alternatively, microbes can induces the secretion of secondary metabolites like terpenes, phytoalexins or glucosinolates (Ahuja et al., 2012; Bednarek et al., 2009; Gershenzon and Dudareva, 2007; Kliebenstein, 2012) as well as the secretion of certain anti-microbial proteins and enzymes like defensins, protease inhibitors or hydrolytic enzymes (De Coninck et al., 2013; Koiwa et al., 1997; Spoel and Dong, 2012; Stassen and Van den Ackerveken, 2011). Fungal lifestyle transitions and the establishment of compatible plant fungi interactions via effector secretion was long thought to be a characteristic of fungal pathogens, but recent studies revealed similar mechanisms to be also involved in mutualistic plant fungus relations.

1.2. Fungal symbionts and endophytes

It is estimated that as many as 90% of all plants depend on a symbiotic relation with fungi (Smith and Read, 2008). The most common and best described interaction of symbiotic fungi with terrestrial plants is the so called mycorrhiza. In this symbiosis, the fungus provides the plant with minerals from the soil, mainly phosphor and nitrogen, that are taken up from the soil by specialised transporters in the fungal membrane predominately in the form of inorganic phosphate, ammonium (NH_4^+) and nitrate (NO_3^-). This leads to improved growth and increased photosynthetic rates of the host plant. Simultaneously, there may well be a cost to the autotrophic partner of AM symbioses, as the fungus partner could possibly act as a carbon sink consuming as 4 - 20% of recently fixed photosynthetic carbon in the form of simple sugars to maintain its activity (Black et al., 2000; Bonfante and Genre, 2010; Kaschuk et al., 2009; Perotto et al., 2012; Wright et al., 1998) (Figure 1.1).

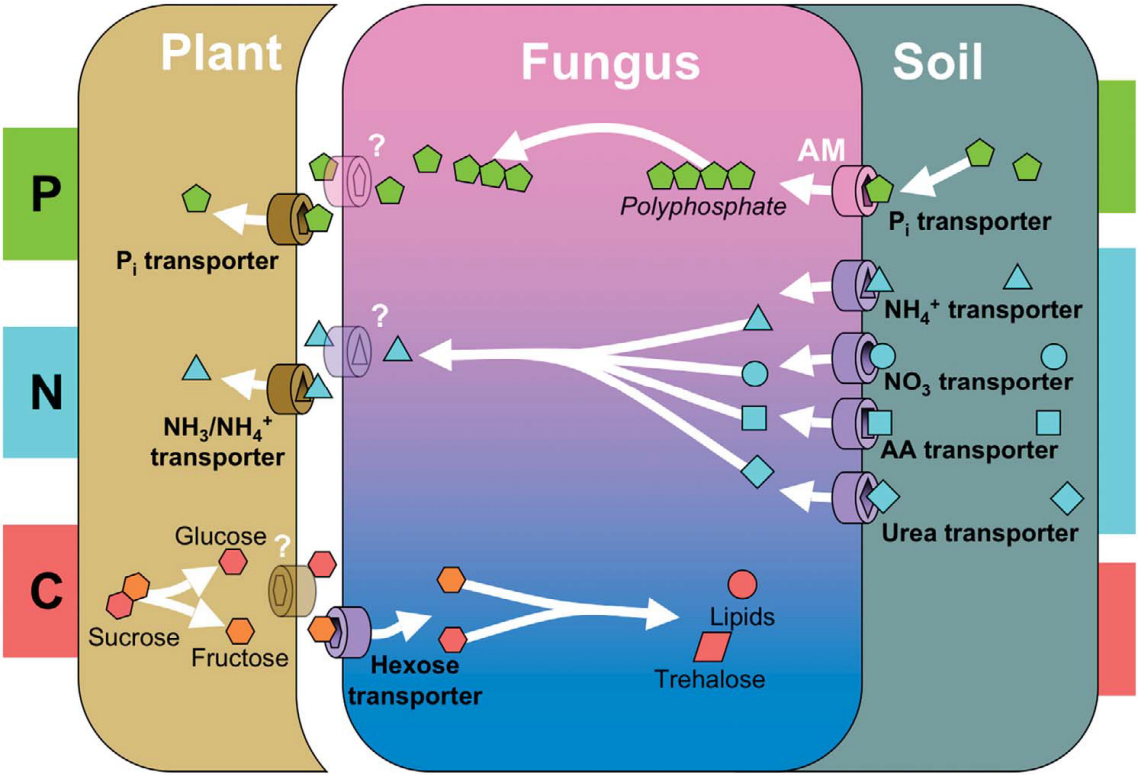


Figure 1.1: Summary of nutrient flow from soil to mycorrhizal fungus to plant host (Bonfante and Genre, 2010). In the mycorrhizal fungus-plant interaction, the fungus provides the plant with phosphorus (P) and nitrogen (N) and gets carbon (C) in return.

Additionally, the symbiosis may result in increased plant resistance against different biotic and abiotic stresses (Pozo et al., 2010; Rodriguez et al., 2004; Schützendübel and Polle, 2002). This increase in resistance is not only associated with the improved nutritional status

of the plant, but includes also changes in the phytohormone homeostasis (Jung et al., 2009; Lopez-Raez et al., 2010; Pozo et al., 2010). The mycorrhizal association is classified based on structural characteristics of fungal hyphae during the colonization of the plant host. According to (Brundrett, 2004), these associations are divided in four main classes: endomycorrhizal, ectomycorrhizal, orchid and ericoid mycorrhizal fungi. The arbuscular mycorrhiza is a form of endomycorrhiza formed by fungi of the order Glomeromycota. The fungus grows inside the plant cell and forms specialized, tree-like structures, the so called arbuscules. In the second class which includes ectomycorrhizal fungi root colonization is associated with a hyphal mantle enclosing the root and the so called Hartig net, a well developed net of intercellular growing hyphae. The third and fourth classes, respectively, encompass orchid and ericoid mycorrhizal fungi which are grouped due to their predominant hosts of the plant family Orchidaceae and Ericaceae. Here hyphal coils (pelotons) are formed which are discussed to be the predominate site of interactions between the fungus and its host plant in this type of symbiosis. In the orchid mycorrhizal association, the pelotons are continuously digested and re-established in living plant cells possibly in order to supply the orchid with carbon not only during seedling development but also at later stages (Beyrle and Smith, 1993; Rasmussen and Rasmussen, 2009; Smith and Read, 2008).

In contrast to the term mycorrhiza which always implies a mutualistic relation between a plant and a fungus, the term endophyte, when applied for fungi, is used more general in literature for every organism that lives, at least transiently, symptomless within a plant host (Bacon and White, 2000). Fungal endophytes usually have a broad host range and do not form specialized mycorrhizal-like structures and prominent fruiting bodies, although beneficial effects, when present, are similar to those described for mycorrhizal associations. While mycorrhizal fungi are often obligate biotrophs, endophytes are normally also able to satisfy their nutritional demands via saprotrophic growth. Because of the inconspicuous nature of the endophytes this group of fungi is often overlooked and literature of this group of fungi is therefore still limited. Some of them have only recently been recognised to be wide spread within the order Sebacinales.

1.2.1. Sebacinales

The order Sebacinales occupies a basal position within the Agaricomycetes (Basidiomycota) and was first described in 2004 based on molecular and ultrastructural studies (Weiss et al., 2004). Together with other basal orders of the Basidiomycota, the Sebacinales share longitudinally septate basidia and imperforate parenthesomes and they also lack cystidia and

clamp connections (Varma et al., 2013; Weiß et al., 2011). Ecological studies and molecular characterization showed the presence of the Sebacinales in field specimens of bryophytes (moss), pteridophytes (fern) and all families of herbaceous angiosperms (flowering plants) from temperate, subtropical and tropical regions. These natural host plants include, among others, liverworts, wheat, maize and the non-mycorrhizal genetic model plant *Arabidopsis thaliana* (Selosse et al., 2002a; Weiß et al., 2011). The associations of Sebacinoid species with the host plant can be partially separated according to the phylogenetically distinct clades A and B within the Sebacinales. Sequences from species belonging to clade A were obtained from fruiting bodies, achlorophyllous and photosynthetic orchids as well as different trees mainly within an ecto- or orchid mycorrhizal association (Julou et al., 2005; McKendrick et al., 2002; Moyersoen, 2006; Selosse et al., 2002b, 2004; Taylor et al., 2003; Urban et al., 2003; Weiß and Oberwinkler, 2001). In contrast, clade B Sebacinales were found in autotrophic orchids, liverworts and Ericaceae mainly within an ericoid- or orchid mycorrhizal association (Allen et al., 2003; Berch et al., 2002; Bougoure and Cairney, 2005; Bougoure et al., 2005; Kottke et al., 2003; Selosse et al., 2007; Setaro et al., 2006; Warcup, 1988). For clade B Sebacinales no production of fruiting bodies could be observed so far. Besides the ability to interact with different plants, an intimate association between Sebacinales and different endobacteria from the genera Paenibacillus, Acinetobacter and Rhodococcus could be identified (Sharma et al., 2008). The current data available on the Sebacinales suggested that the multitude of mycorrhizal interactions may have arisen from an ancestral endophytic habitat by specialisation. Considering their proven beneficial influence on plant growth and their ubiquity, endophytic Sebacinales may be a previously unrecognised universal hidden force in the shaping of natural and cultured plant ecosystems (Deshmukh et al., 2006; Varma et al., 2013; Weiß et al., 2011; Weiss et al., 2004). This establishes fungi belonging to this order as interesting candidates to study general mechanisms of plant fungi interactions and requirements for adaptation to host diversification.

1.2.1.1. *Piriformospora indica*

Piriformospora indica is the most studied fungus of the order Sebacinales. It is a filamentous fungus that was first isolated in 1997 from the rhizosphere of *Prosopis juliflora* and *Zizyphus nummularia* where it was found in close association with spores of the arbuscular mycorrhizal fungus *Glomus mossae* (Verma et al., 1998). *P. indica* is classified as a root endophyte because it is not causing any visible disease symptoms during colonization of the plant host. Instead, the interaction is rather mutualistic with benefits for both, the plant and the fungus.

This includes growth promotion, increased nutrient uptake, enhanced seed production and increased resistance against different biotic and abiotic stresses could be observed in the host when colonized by *P. indica* (Achatz et al., 2010; Baltruschat et al., 2008; Peškan-Berghöfer et al., 2004; Sherameti et al., 2008; Sun et al., 2010; Vadassery et al., 2008; Waller et al., 2005, 2008; Yadav et al., 2010). *P. indica* was shown to be capable of colonizing a variety of different plants (Qiang et al., 2012a; Varma et al., 2013), but research on the molecular mechanisms underlying symbiosis has so far focused on the agronomical important monocotyledonous crop plant barley (*Hordeum vulgare*) and the dicotyledonous genetic model plant *Arabidopsis thaliana*. It could be shown that the plant host induces the expression of defence-related genes upon colonization by *P. indica* which is in turn actively suppressed by the fungus (Camehl et al., 2011; Jacobs et al., 2011; Schäfer et al., 2009). Furthermore, increased disease resistance could be linked to a modulation of the basal phytohormone levels within the host, in particular jasmonate, ethylene and gibberellin (Camehl et al., 2010; Khatabi et al., 2012; Schäfer et al., 2009; Stein et al., 2008). Cytological studies showed for barley that *P. indica* establishes first a biotrophic interaction in this host which is characterized by an intact, invaginated plant plasma membrane surrounding penetrating fungal hyphae (Schäfer et al., 2009; Zuccaro et al., 2011). This interaction lasts in barley five to seven days and is accompanied by the expression of small secreted proteins (Zuccaro et al., 2011). After this initial biotrophic interaction, *P. indica* switches to a cell death associated growth which is characterized by a massive proliferation and eventually spore formation within dead rhizodermal and cortical cells (Deshmukh et al., 2006; Lahrmann et al., 2013; Zuccaro et al., 2011). It was further shown that gene expression of the negative cell death regulator BAX inhibitor-1 (HvBI-1) is suppressed in *P. indica* colonized barley roots from five days onwards (Deshmukh et al., 2006). This correlates with an increased activity of the cell death marker VPE (Vacuolar Processing Enzyme) within inoculated root material from five days onwards (Lahrmann et al., 2013). However, for cereal crop plants, in particular barley and wheat, programmed root cortical cell death can be observed in uninfected, two day old root segments. This type of cell death has been characterized mainly by DNA fragmentation and nucleosomal fragmentation, one of the most important markers for apoptosis in animal cells (Hector, 1938; Henry and Deacon, 1981; Liljeroth and Bryngelsson, 2001). Nevertheless, the colonization of barley roots by *P. indica* is also at later stages of the interaction beneficial for the plant as indicated by a strong growth promotion effect (Lahrmann et al., 2013). Similar to the interaction with barley, colonization of Arabidopsis roots by *P. indica* were also shown to occur in two distinct phases. However, while the initial

biotrophic colonization is comparable in both hosts, late colonization of Arabidopsis cells was shown to be accompanied by an ER stress–induced caspase-dependent cell death (Qiang et al., 2012b). Besides the ability of *P. indica* to colonize different plant hosts, this fungus grows well on different synthetic media and many saprotrophic traits could be identified in its genome (Zuccaro et al., 2011). The beneficial effects for the plant host together with the saprotrophic traits are further discussed to be important for the broad distribution of the fungus and eventually for shaping natural and managed ecosystems by promoting plant adaptation to sub-optimal conditions (Basiewicz et al., 2012; Lahrmann and Zuccaro, 2012; Weiß et al., 2011). Due to the interesting lifestyle of *P. indica* and the plant promoting effects, the genome and transcriptome of *P. indica* were analyzed in order to identify key determinants involved in the endophytic mutualism (Zuccaro et al., 2011).

1.2.1.2. *Sebacina vermifera*

The orchid mycorrhizal fungus *Sebacina vermifera* (MAFF 305830) was first isolated from the Australian orchid *Cyrtostylis reniformis* (Warcup, 1988). Although less research has been performed with this fungus, available results indicate similar colonization mechanisms as described for *P. indica* (Deshmukh et al., 2006). Similar to *P. indica*, enhanced seed germination and biomass production (Ghimire and Craven, 2011; Ghimire et al., 2009) as well as increased resistance against biotic and abiotic stresses (Ghimire and Craven, 2011; Waller et al., 2008) could be observed in plants colonized by *S. vermifera*. Inhibition of ethylene signalling in *Nicotiana attenuata* was reported to be important for the growth promoting effects by *S. vermifera* in this host (Barazani et al., 2007). This observation seems to be in contrast to *P. indica* because Camehl and colleagues reported that restriction of fungal growth by ethylene signalling components is required for the beneficial interaction between *P. indica* and *A. thaliana* (Camehl et al., 2010). On the other has ethylene been shown to have diverse and opposite functions in supporting or repressing plant defence responses to microbes (Broekaert et al., 2006). The different requirements for ethylene signalling in the growth promoting effects of both Sebacinoid fungi might therefore just be a host-dependent mechanism.

1.3. Genome research

The genome of an organism is the entirety of its hereditary information and includes protein coding genes as well as their regulatory elements and other non-coding sequences. The term refers normally to the nuclear genome, whereas the genome of organelles (such as

mitochondrion and chloroplast) are explicitly named when meant. In order to determine the entire DNA sequence and analyse the function and structure of genomes, the discipline of genome research or simple genomics, combines bioinformatic tools with recombinant DNA and DNA sequencing methods (Klug et al., 2011; Seyffert, 2003). Due to the advent of new sequencing technologies and decreasing costs together with improved algorithms optimized for analyzing the huge amount of data produced, genome research became an important field of study in biology.

1.3.1. Sequencing and sequence assembly

The term sequencing is mainly used for the method of determining the nucleotide composition of a DNA or RNA strand. With the invention of the sequencing method of Sanger (Sanger et al., 1977) which uses dideoxynucleotides as chain terminators, sequencing became an accurate, fast and affordable technique to analyse protein coding genes and genome organisation on DNA level. Within the last years, several so called next generation sequencing (NGS) methods were developed which lowered the costs further while the sequencing speed increased drastically. Besides the sequencing of whole genomes, these technologies aid nowadays also in different applications ranging from chromatin immunoprecipitation, mutation mapping and polymorphism discovery to non-coding RNA discovery (Mardis, 2008). While there are different NGS technologies available on the market, two are predominately used in fungal genomics. That is the Illumina® Genome Analyzer utilizing the Solexa sequencing technology which uses PCR on a surface and the Genome Sequencer FLX System from Roche Applied Science which utilizes the pyrosequencing technology developed by 454 Life Sciences (Shendure and Ji, 2008). Both technologies work on a "sequencing-by-synthesis" principle where an incorporated nucleotide is identified already during its assembly into the growing DNA strand (Metzker, 2010). In the Illumina system, this is achieved by using nucleotides to which chain terminators and different fluorescent dyes are reversibly attached. During the PCR step, all four nucleotides compete for the incorporation. Afterwards, unbound nucleotides are washed away and a four colour image is acquired. Fluorescent dyes and terminating groups are then removed and the next cycle starts (Bentley et al., 2008; Metzker, 2010). Decoupling the enzymatic removal of the fluorescent label and the image acquisition in the Illumina system, allows a high parallelization which results in a very high yield of sequence reads. In contrast to that, pyrosequencing uses the enzymes sulfurylase and luciferase to emit light from pyrophosphate released after successful incorporation of a nucleotide (Margulies et al., 2005). This method is

faster than the Illumina approach and produces longer reads. However, it is more expensive and may produce errors in homopolymeric regions which are stretches of the same nucleotide whose length is determined from the duration and intensity of the emitted light flash. Additionally can the lack of competition between nucleotides be a source of error in the pyrosequencing method. Using these sequencing methods, millions to billions of sequence reads can be produced within a few days, from which the original chromosomes and genome, respectively, need to be reconstructed by assembly (Figure 1.2).

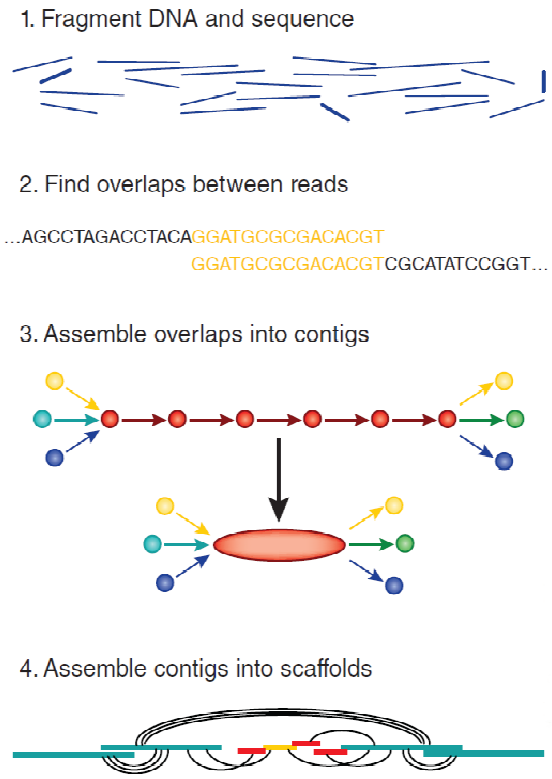


Figure 1.2: Steps in the assembly of genomes from sequence reads (Baker, 2012). Genomes are sequenced in reads (1) which are assembled into contigs by assemblers (2, 3). Contigs may thereby be elongated as long as sufficient read support is given and need to break up if sequencing depth drops and no unique matching read overlap can be found (3). Using paired end sequencing, the contigs may further be assembled into scaffolds (4).

Specialized computer programs (Assemblers) are therefore used to find overlapping regions between reads. Using these overlaps, the reads are assembled into so called contigs where the average number of reads used to support the contig sequence is referred to as the coverage of the contig. The size of those contigs can vary greatly from some hundreds base pairs to several megabases. This is mainly influenced by the coverage of the contig and the occurrence of long stretches of repetitive DNA within the genome. While the coverage is rather a technical problem, genome fragmentation after assembly is mainly reasoned by repetitive sequences. Due to the relative small read length of less than 1 kb, the assembler often fails to

identify the correct order of the repetitive sequences and genes in their proximity, resulting in a sequence rearrangement and/or inversion (Phillippy et al., 2008). Alternatively, reads covering multiple repeats might also be collapsed to a single one or vice versa (Phillippy et al., 2008). These problems are partially circumvented by using paired-end tags which are short sequences that can be, due to their preparation method, linked to each other giving the precise distance between two mating tags (Fullwood et al., 2009). With this additional information, contigs may be further assembled into so called scaffolds which would represent chromosomes in an optimal case (Figure 1.2). After the assembly has been performed, it is furthermore difficult to provide information about the quality of it. To address this question for eukaryotes, Parra and colleagues have identified and published a set of 248 highly conserved genes deduced from 28 analyzed organisms which are supposed to be exclusively present in all eukaryotes (Parra et al., 2009). As a quantitative measurement of assembly quality, the N50 value is often used, which indicates the size of the last scaffold needed to cover 50% of the overall genome sequence.

1.3.2. Fungal genome programs

The field of fungal genomics and with that also eukaryotic genomics, started with the report on the complete sequence of the yeast *S. cerevisiae* in 1996 (Goffeau et al., 1996). As sequencing progress was initially slow, a consortium of mycologists and scientists from Broad Institute (previously: Whitehead institute / MIT center for genome research) started the Fungal Genome Initiative (FGI) in 2000 (http://www.broadinstitute.org/scientific-community/science/projects/fungal-genome-initiative). Until now over 50 genomes have been sequenced and annotated within the consortium, among them the genome of the model organism and first sequenced filamentous fungus, *Neurospora crassa* (Galagan et al., 2003), and the genome of the first sequenced plant pathogen, *Magnaporthe grisea* (Dean et al., 2005). The aim of the FGI is the study of evolution, eukaryotic biology, and medical studies via comparative genomics of relevant fungal genomes. In November 2011 the Department of Energy (DoE) in the USA awarded funding for the 1000 fungal genomes project (1KFG) in which 13 principle investigators coordinate the sequencing two species from 500 fungal families within five years ((Grigoriev et al., 2011); http://1000.fungalgenomes.org/home/). In contrast to the aims of the FGI, main focus lies on alternative fuel, carbon cycling and bioremediation, while projects in the context of human health or disease are not allowed. Sequencing, genome assembly and gene prediction are performed at the Joint Genome Institute (JGI) which additionally tries to collect and combine genome data from other

institutes or private investigators. The database of the JGI contains currently 260 fungal genomes, mainly of the two major fungal groups, i.e. Ascomycetes and Basidiomycetes (Figure 1.3).

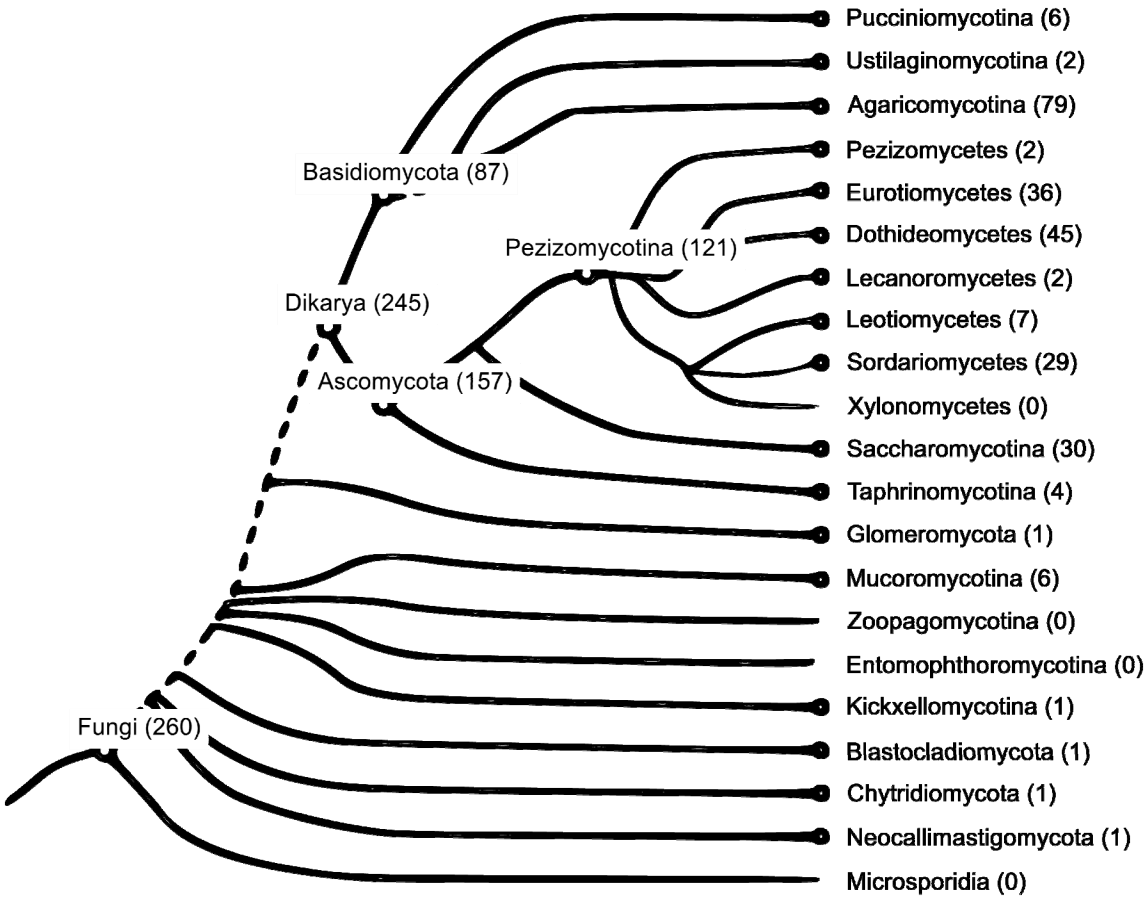


Figure 1.3: Sequencing progress in the fungal tree of life. Simplified phylogeny of fungal groups with numbers of currently available genomes within them (Adapted from (Grigoriev et al., 2012))

1.4. Aims of this thesis

The scientific work conducted during this thesis focuses primarily on the genomic and transcriptomic analyses of *Piriformospora indica* and *Sebacina vermifera* and is schematically depicted in Figure 1.4. The aim was to discover saprotrophic and biotrophic features in the genome and transcriptome of these two symbionts and to identify and characterize conserved and specific sebacinoid genes as key determinant in the Sebacinales symbioses. For the genome analyses, this was achieved by a detailed analysis of predicted protein coding genes in both genomes and a comparison of these with genomes of other Basidiomycetes and Ascomycetes. The ability of *P. indica* to grow as a saprotroph or as a biotroph during colonization of barley roots was addressed by microarray analyses of

transcriptional regulations in *P. indica* during colonization of either living or autoclaved barley roots. Information gained from the colonization of barley roots was then compared to the colonization of Arabidopsis roots. Finally, the transcriptional data obtained from *P. indica* were compared to that of *S. vermifera* during colonization of Arabidopsis roots. This will help to answer the question how the broad host range and worldwide distribution of these two fungi are a result of the described mechanisms. Data and knowledge provided in this thesis support towards understanding of the symbiotic plant fungus interaction and serve as a basis for a detailed, functional characterization of key factors predominately involved in this.

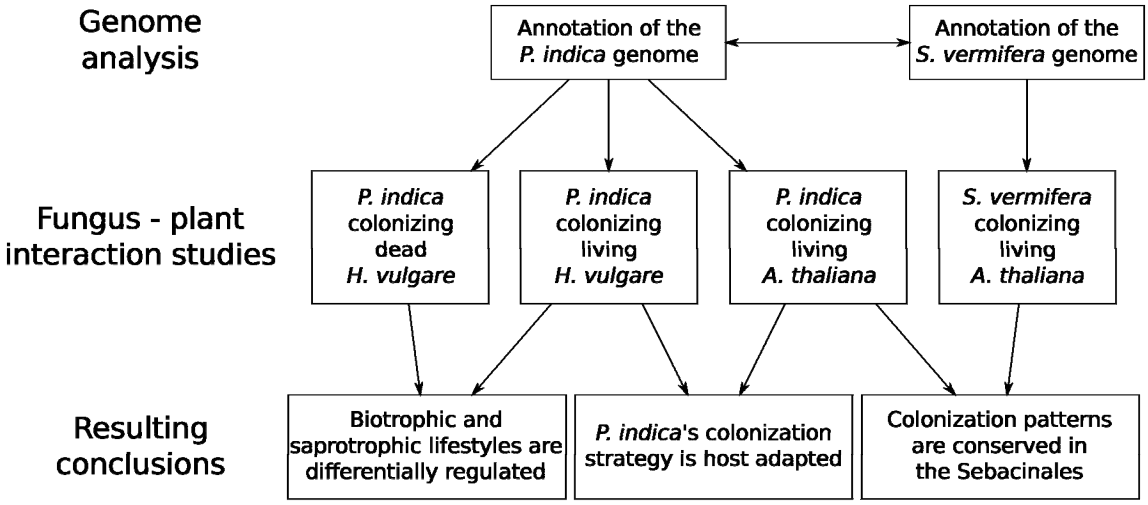


Figure 1.4: Main chapters of this thesis and their interconnection. The scheme is showing the four main chapters and the comparisons made between the main topics.

2. Results

2.1. Comparison and characterization of the *P. indica* and *S. vermifera* genomes

Both analyzed fungi are able to grow saprotrophically in the soil and colonize different plant hosts. In order to exploit the genomic background that is required to live and proliferate in those unequal environments, the genomes of both fungi were sequenced and analyzed using a variety of bioinformatic tools. By comparing the identified gene sets to each other and selected other fungi with different lifestyles, specific and shared genomic traits of the Sebacinales could be identified. These traits provide hints towards evolution and lifestyle classification of fungi belonging to this order.

2.1.1. General genomic features within the Sebacinales

The analysis of the *P. indica* genome has been published in detail (Zuccaro et al., 2011) and is therefore just briefly described here and updated where newer and/or additional results are available. Pyrosequencing of the *P. indica* genome was performed in parallel to RNA-Seq of cDNA pooled from different fungal developing stages. The genome was assembled into 1884 scaffolds (N50: 51.83 kb) containing 2359 contigs with an average read coverage of ~22 and a genome size of 24.97 Mb. Computational and manual analysis of the *P. indica* genome assembly revealed the presence of single nucleotide polymorphisms (SNP) with two variants (Chapter 4.7.5). A total number of 60493 equally distributed SNPs could be identified in scaffolds covering 23.15 Mb (92%) of the whole genome sequence (2.6 SNPs per kb). 11769 gene models were manually selected based on gene models predicted by the programs AUGUSTUS (Stanke and Waack, 2003), GeneMark (Lukashin and Borodovsky, 1998) and FGenesH (Salamov and Solovyev, 2000). Predicted open reading frames were afterwards validated by mapping unique expressed sequence tags (EST) to the scaffolds. Through experimental validation, the gene number has now been slightly corrected to 11767. To assess the gene space covered by all predicted gene models, *P. indica* genes were compared against 248 highly conserved eukaryotic core genes (CEG) defined by Parra and colleagues (Parra et al., 2009) and against 246 highly conserved single copy genes defined in the FunyBASE database (Marthey et al., 2008) using Psi-Blast (e-Val < 10^{-5}). The Blast-output was afterwards manually evaluated and 243 (98%) CEG orthologs and 245 (99.6%) orthologs to FunyBASE entries were identified in *P. indica* with high confidence. Using Blastx (e-Val <

10^{-3}) to compare all *P. indica* genes against the non-redundant (nr) database from the NCBI identified the ectomycorrhizal fungus *Laccaria bicolor* and the saprotrophic fungus *Coprinus cinerea* as the closest related organisms at the time the genome was published (This statement was re-evaluated and verified last in June 2012).

The analysis of the *S. vermifera* genome followed that of *P. indica*. The genome and transcriptome were sequenced using Illumina and pyrosequencing in parallel. Genome assembly resulted in 546 scaffolds (N50: 319.3 kb) containing 2457 contigs with an average read coverage of 117 and a genome size of 38.09 Mb. Like for *P. indica*, SNPs with two variants could be predicted in the *S. vermifera* genome assembly. With a total number of 191952 equally distributed SNPs in scaffolds covering 97.3% of the whole genome sequence (5.2 SNPs per kb), SNPs are about two times more frequent in *S. vermifera* than in *P. indica*. Gene prediction was performed with support from 83352 ESTs and resulted in 15312 automatically generated gene models. Orthologs to 246 from 248 (99%) CEGs and to all 246 (100%) FunyBASE entries could be identified in the *S. vermifera* genome, indicating an equally well covered gene space as in *P. indica*. Blastx (e-Val < 10^{-3}) comparisons against the nr-database identified exclusively *P. indica* as close relative (10289 top blast matches, 67.2%), followed by *L. bicolor* (180, 1.2%) and *C. cinerea* at 10th position (93, 0.6%). Besides these Blast hits, 2969 (19.4%) *S. vermifera* genes showed no significant similarity to any sequence in the database and were therefore termed as *S. vermifera* specific genes. Initial genome comparisons focused therefore on the genomes of *P. indica*, *S. vermifera*, *L. bicolor* and *C. cinerea*. Additionally, the genome of the biotrophic leaf pathogen *Ustilago maydis* was used as outgroup.

2.1.2. Comparative genomics

It is known that the genome size does not correlate with the complexity of an organism. This is rather determined by the genes encoded on the genome and the machinery regulating their expression. Main features related to these determinants, like the number of protein coding genes or repetitive sequences in the intergenic space, were collected in 52 genomes of different Basidiomycetes and compared to those of *P. indica* and *S. vermifera* (Table 5.17). The results from this analysis show that the genome of *P. indica* strongly differs from the average in some of the defined main genomic features and that the same features in *S. vermifera* are closer to the average. Both genomes are smaller than the average, but encode more genes in relation to this size (Table 2.1, gene density). Because the average length of protein coding genes is comparable in all analyzed fungi and because the genetic region

covered by coding regions is bigger in *P. indica* and *S. vermifera*, a higher gene density implies a reduced intergenic space in these fungi (Table 2.1). This is also reflected by a lower abundance of simple sequence repeats (SSR), but not by less repetitive elements in general (Table 2.1, repetitiveness and simple sequence repeats).

Table 2.1: Main genome feature of Basidiomycetes fungi

	P. indica	***S. vermifera***	***L. bicolor***	***C. cinerea***	***U. maydis***	**Basidios[1]**
Genome size (Mb)	25	38	61	36	20	46
GC content (%)	50.7	48.9	47	51.6	54	50.1
Repetitiveness (%)	4.7	9.3	27.9	5.4	2	(9.9)[2]
Simple Sequence Repeats	602	1490	6104	2050	2462	3266
Protein coding genes	11767	15312	23130	13342	6787	14315
Gene density (Genes/Mb)	471	402	381	368	345	337
Gene coding regions (%)	61.8	49.5	40.6	51	61	41.9
Average intergenic space (bp)	530	813	1524	980	1059	1446
Secreted proteins (SP)	867	1024	1295	1118	518	939
Average length of SPs	511	506	519	530	576	530
SSPs (SP < 300 aa)	386	460	747	549	239	441
Average length of SSPs	172	170	158	188	198	170
SNPs[3]	60534	191952	-	-	-	-

[1] Average of all used 52 Basidiomycetes genomes listed in Table 5.17 including also the listed five.
[2] Repetitiveness was only predicted on the five shown fungi and the average is therefore not representing the whole dataset
[3] SNPs were only predicted in *P. indica* and *S. vermifera.*

The overall similarity of the five above stated genomes and the genes encoded on them was evaluated using synteny and ortholog predictions. Syntenic regions were predicted with the program SyMAP (Soderlund et al., 2006) (cf. also chapter 4.7.15). Due to set number and size restrictions of the program, it was not possible to use the high number of genomic scaffolds as presently available. Therefore, the scaffolds of each of the five genomes were concatenated each into 8-10 more or less equally sized super-scaffolds and synteny predicted on these. An overall number of 4992 syntenic regions could be identified which could be grouped into 210 adjacent blocks (Table 2.2, Figure 2.1).

Table 2.2: Syntenic hits and adjacent syntenic blocks between *P. indica*, *S. vermifera*, *L. bicolor*, *C. cinerea* and *U. maydis*.

	P. indica	*S. vermifera*	*L. bicolor*	*C. cinerea*	*U. maydis*
P. indica	0 (0)	-	-	-	-
S. vermifera	1486 (115)	0 (0)	-	-	-
L. bicolor	75 (8)	169 (15)	36 (2)	-	-
C. cinerea	53 (4)	129 (13)	3020 (50)	0 (0)	-
U. maydis	0 (0)	0 (0)	17 (2)	7 (1)	0 (0)

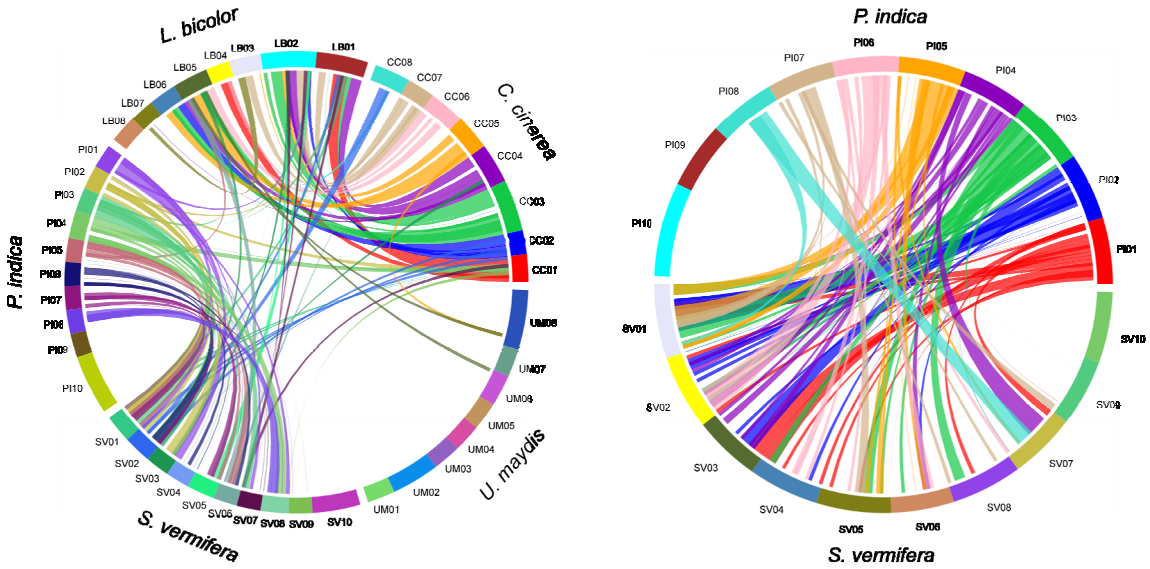


Figure 2.1: Syntenic blocks between *P. indica*, *S. vermifera*, *L. bicolor*, *C. cinerea* and *U. maydis*. Syntenic blocks were calculated on concatenated genome sequences and from these graphs produced using the program SyMAP with standard settings (Soderlund et al., 2006). Left graph: Syntenic blocks between all five tested species. Right graph: Syntenic blocks between *P. indica* and *S. vermifera* only.

Most of the hits within this set were due to similarities between *L. bicolor* and *C. cinerea* with a genome sequence coverage of syntenic regions of 61% (*L. bicolor*) and 75% (*C. cinerea*), respectively (3020 syntenic hits (60.5% of all hits), 50 blocks of adjacent syntenic hits (23.8% of all blocks)). Compared to that resulted the comparison between *P. indica* and *S. vermifera* in a lower genome sequence coverage with 52% (*P. indica*) and 51% (*S. vermifera*), respectively. Syntenic regions are also on average smaller and were therefore combined in more syntenic blocks (1486 hits (29.8%), 115 blocks (54.8%)). Furthermore, both groups show only minor synteny between each other (426 hits (8.5%), 40 blocks (19.1%)) and almost none with *U. maydis* (Table 2.1, Figure 2.1). It need to be taken into account that these results might be biased due to the differences in the genome sequence

fragmentation (number of scaffolds) of the presently available data. These data therefore only account for the minimal possible synteny and that higher values and longer overlapping regions might be possible in individual groups. This is especially true for the comparison between *P. indica* and *S. vermifera* which genomes are much stronger fragmented than those of the other fungi and becomes most evident by looking on the last two super-scaffolds in *P. indica* and *S. vermifera* (Figure 2.1, right graph). These super-scaffolds contain several small scaffolds which, if correctly assembled, might also contain syntenic regions that cannot be detected yet. Nevertheless showed the analysis that many regions of high sequence similarity exist in both genomes. In order to support this observation and to get an estimation of the similarity between the genomes that is independent of the described fragmentation, groups of orthologous genes were calculated to estimate the similarity in the gene repertoire between the five species. Hence, all genes from *P. indica* (11767), *S. vermifera* (15312), *L. bicolor* (23130), *C. cinerea* (13342) and *U. maydis* (6787) were used as input for the program proteinortho (Lechner et al., 2011) and the final result manually summarized (Figure 2.2).

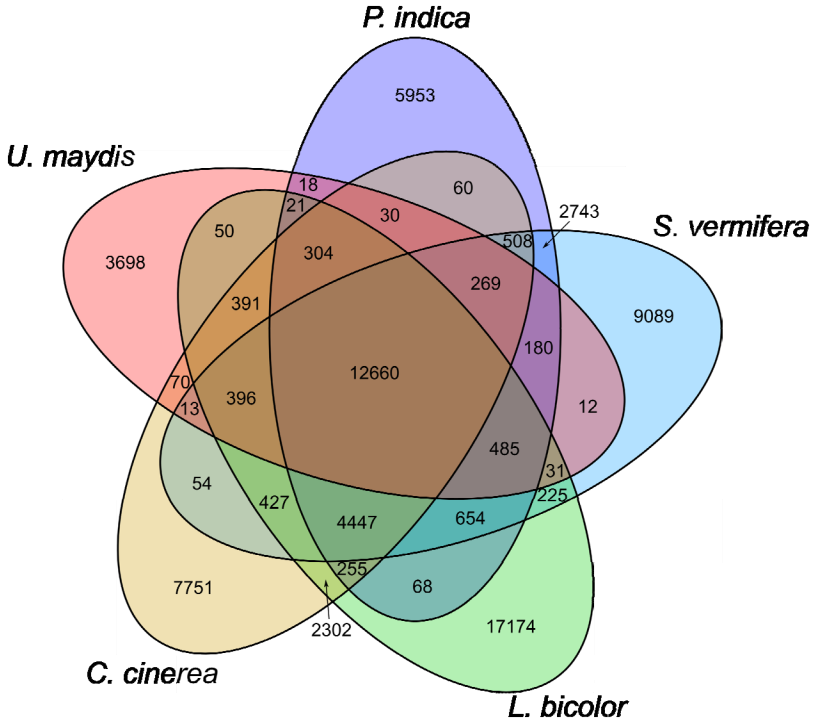


Figure 2.2: Orthologous groups between *P. indica*, *S. vermifera*, *L. bicolor*, *C. cinerea* and *U. maydis*. Orthologues genes within the whole set of 70338 genes were calculated using the program proteinortho with standard settings (Lechner et al., 2011) and the output manually summarized in a venn diagram.

Most of the 70338 genes do not have an orthologous gene in one of the other organisms (43665, 62.1%) or are orthologous in all organisms (12660, 18%). Correlating with the observations from the synteny prediction, the next highest numbers of orthologous genes were

found in the group containing all organisms except *U. maydis* (4447, 6.3%), in the *P. indica* and *S. vermifera* group (2743, 3.9%) and in the *L. bicolor* and *C. cinerea* group (2302, 3.3%) (Figure 2.2). Within these combination, 6830 genes of *P. indica* (58%) were predicted to have an orthologue in *S. vermifera*. The other way around, 7421 genes of *S. vermifera* (49%) were predicted to have an orthologue in *P. indica*.

Following this analysis, the gene organisation on *P. indica* and *S. vermifera* was compared to the genomes of 52 Basidiomycetes listed in Table 5.17. In comparison to these fungi, the genomes of *P. indica* and *S. vermifera* show a significantly (one sample t-test, $p < 0.001$) higher gene density with 471 and 402 open reading frames (ORF) per Mb (Average in Basidiomycetes: 334 (w/o *P. indica*) and 336 (w/o *S. vermifera*) ORFs/Mb) (Table 2.1). This observation is accompanied by a significantly (one sample t-test, $p < 0.001$) higher space occupied by coding regions (exons) in the genomes of *P. indica* and *S. vermifera* than in other Basidiomycetes with 61.83% and 49.48% (Average in Basidiomycetes: 41.5% (w/o *P. indica*) and 41.7% (w/o *S. vermifera*)) (Table 2.1). This indicates that rather the space between genes than the gene length is reduced in both fungi. To support this statement, the 5' and 3' distances of all genes to their respective next neighbouring genes were analyzed in the genomes of the *P. indica*, *S. vermifera*, *L. bicolor*, *C. cinerea* and *U.* maydis (Figure 2.3). For all genomes, the majority of genes have a 5' and 3' space of at most 1000 basepairs (1 kb), but outliers above this threshold occur much less frequent in *P. indica* than in the other genomes. Especially in *L. bicolor*, intergenic distances of more than 1 kb can be observed for many genes, which is coincides with an increased proportion of repetitive elements that has been described for this fungus (Martin et al., 2008). The comparable low number of genes with minor 5' distances in *U. maydis* are probably either due to more defined promoter regions in this fungus or just due to a better/different gene prediction which did not allow smaller regions. Interesting but is in that context that those genes with a larger 5' distance do not necessarily have a larger 3' distance, indicating that those genes are orientated in opposite directions to their respective neighbouring genes. This could have evolved as a mechanism to reduce intergenic space and thereby increase gene density. However, with 345 ORFs/Mb *U. maydis* has an average gene density and a manual revision of genes in *P. indica* and *S. vermifera* with big 5' and small 3' distances could also not support this hypothesis.

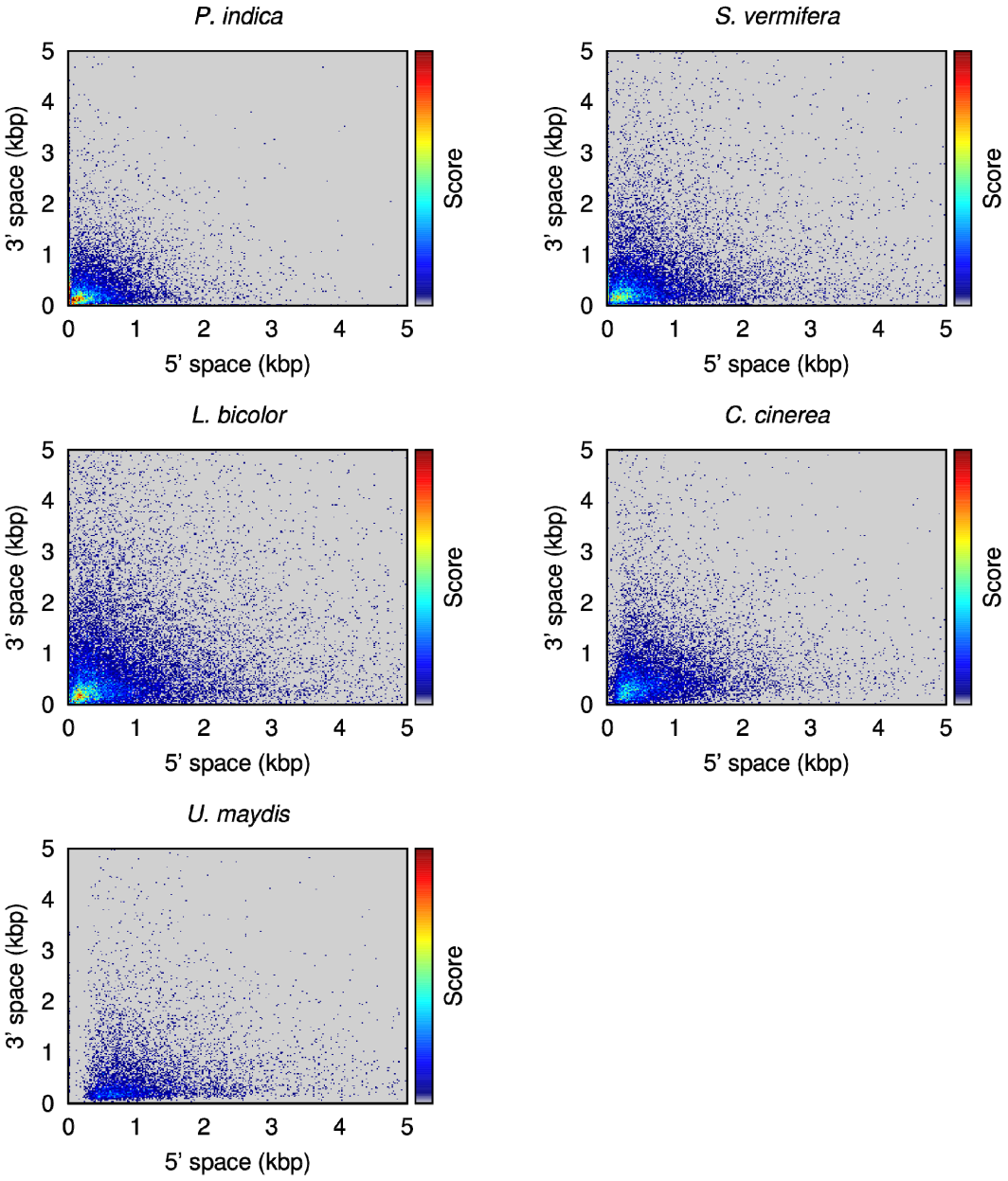


Figure 2.3: Intergenic distances of the *P. indica*, *S. vermifera*, *L. bicolor*, *C. cinerea* and *U. maydis* genes. Each dot resembles a group of genes with an intergenic space within a 20 bp range. Colour code indicates the number of genes within each group from 0 (grey) over green (13) to red (25). Plots were created using gnuplot (Williams and Kelley, 2012) and are limited to 5 kbp regions for visualization purposes.

In order to analyze whether the higher gene density in *P. indica* and *S. vermifera* occurred by retro-transposition, the type and abundance of repetitive elements were predicted in both genomes (Chapter 4.7.6). Using the program RepeatModeler (Smit and Hubley, 2008), 4.7% of the *P. indica* genome and 9.3% of the *S. vermifera* genome were identified as repetitive regions (Table 2.1). Only a few of these repeats could be classified and showed that in particular LTR/Gypsy and LTR/Copia are much less frequent in *P. indica* (Gypsy: 1; Copia: 2) and *S. vermifera* (Gypsy: 3; Copia: 3) than in other fungal genomes (Muszewska et al.,

2011). The presence of repetitive sequences is often linked to the presence of the reverse transcriptase domain 1 or 2 (RVT_1 or RVT_2) in the genome. This could but not be confirmed here, as 41 of these domains were predicted in the genome *P. indica*, while the *S. vermifera* genome contains only 4. Furthermore, the highly repetitive genomes of *L. bicolor* and *T. melanosporum* (Martin et al., 2008, 2010) contain also only 12 and 1 of these domains, respectively (cf. also chapters 2.1.4 and 3.1). Like transposable elements, also the relative abundance of simple sequence repeats (SSR) is significantly (one sample t-test, $p < 0.005$) reduced in the *P. indica* and *S. vermifera* genome with 24 SSRs/Mb and 39 SSR/Mb (Average in Basidiomycetes: 79 (w/o *P. indica*) and 78 (w/o *S. vermifera*) SSRs/Mb) (Table 2.1).

2.1.3. Annotation and analyses of protein coding genes

A higher gene density with reduced space for regulatory elements of genes, like transcription promoters and terminators, might result in or even require modifications in the tRNA content and codon usage of protein coding genes. Therefore, transfer RNAs (tRNA) and the codon usage preferences were predicted on both genomes (Chapter 4.7.7). The prediction resulted in 58 tRNAs in *P. indica* and 70 tRNAs in *S. vermifera* covering all standard 20 proteinogenic amino acids, but these numbers are smaller than in the compared genomes of *L. bicolor* (330), *C. cinerea* (289) and *U. maydis* (111) and no correlation between the amount of tRNAs, gene density or the number of putative protein coding genes could be identified. Additionally, a similar nucleotide preference at the first and second position of all codon triplets could be identified in all analyzed genomes which is indicative of a comparable codon usage (Figure 2.4). In all analyzed genomes, guanine is preferred in the first position, while thymidine is unfavoured. In the second position guanine is unfavoured and either adenine (62%) or cytosine (38%) are preferred. In the third position, either cytosine (77%) or thymine (21%) are preferred and either adenine (87%) or guanine (13%) are unfavoured. Using these data as reference, nucleotide preferences within codon triplets in *P. indica* and *S. vermifera* perfectly resemble the average (Figure 2.4).

The annotation of protein coding genes was performed using partially available and partially newly developed bioinformatic tools in order to allow a high quality estimation of putative functions. The analyses described in the following chapter thereby focus on similarity searches for known and partially functionally verified sequences (Blast) and probabilistic predictions of functional protein domains (hidden markov models, HMM). Protein domains are conserved parts of a protein that have a distinct function and structure. These domains may occur alone, in multiple repeats of the same domain or as multi-domain structures with

several different domains (Björklund et al., 2005, 2006). The Pfam database version 24 (Finn et al., 2010) was used to predict functional protein domains in a set of 60 fungal genomes (52 Basidiomycetes and 8 Ascomycetes) of different lifestyles (Chapter 4.7.9.4 and Table 5.17). From the 11912 different classes of protein domains listed in this release of the database, about 2700 were predicted on average within the analyzed genomes.

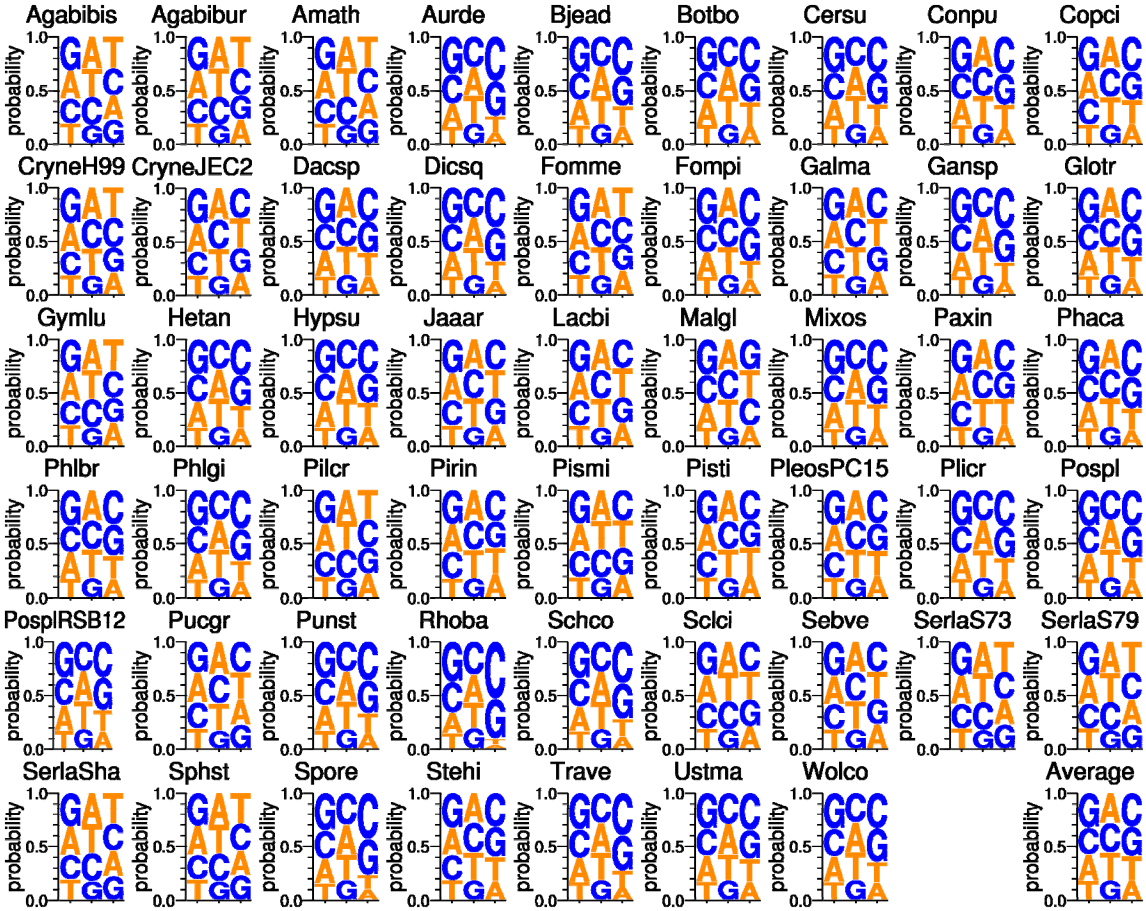


Figure 2.4: Nucleotide preferences in codon triplets. Codon triplets off all protein coding genes of each organism were used to draw frequency-plots with the program WebLogo 3 (Crooks et al., 2004), whereas bigger letters and a higher position in the plot refer to a more frequent occurrence of the nucleotide at the respective position. For organism abbreviations see Table 5.17.

In order to gain an overview of the kind of domains that are over- or underrepresented in *P. indica* and *S. vermifera* and if these are linked to a certain lifestyle, three comparisons were performed. The first two analyses focused on domains expanded/constrained in *P. indica* or *S. vermifera*, respectively, in relation to all other genomes. In the third analysis, the 8 genomes from Basidiomycetes living in a symbiosis-related lifestyle were compared against the 29 genomes from saprotrophic white- and brown-rot Basidiomycetes. A significantly (one sample t-test, $p < 0.05$) increased number of protein domains was found for 119 in *P. indica* and 114 in *S. vermifera* in comparison to other fungi (Figure 2.5). 7 of these expanded

domains overlap with those of the symbionts group and include domains involved in protein/carbohydrate binding (Pfam accessions: PF00400 (WD40), PF05729 (NACHT), PF01822 (WSC)), signalling and regulation of cellular responses to stress and nutrient availability (PF00931 (NB-ARC), PF02145 (Rap_GAP)) and two domains of unknown function (PF07173 (DUF1399), PF10914 (DUF2781)).

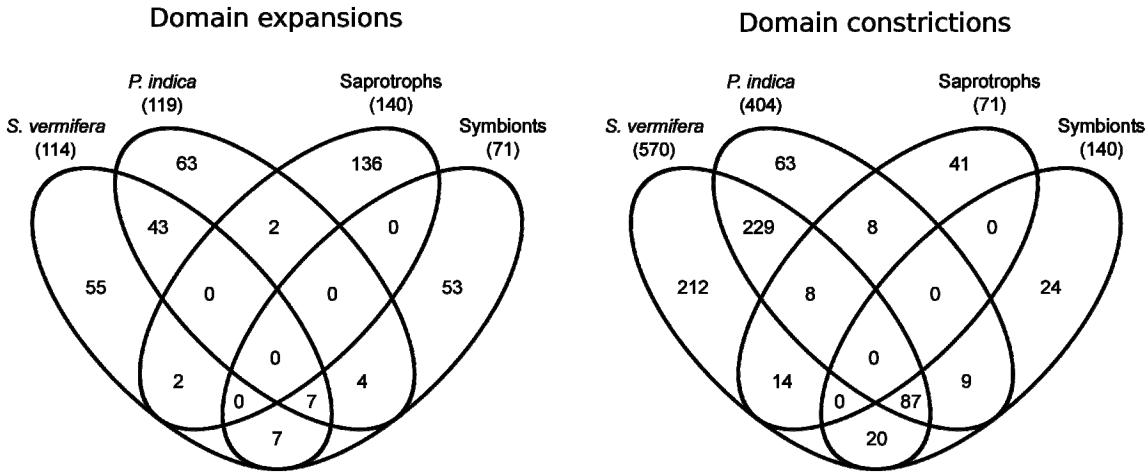


Figure 2.5: Functional domain expansions/constriction in *P. indica* and *S. vermifera*. Functional protein domains were predicted using the Pfam database (Finn et al., 2010). Domain expansions and constrictions were identified by comparing *P. indica* or *S. vermifera* against all other genomes or by comparing fungi living in a plant-fungus-symbiosis with fungi living saprotrophically on dead plant material.

Domains specifically, but not necessarily exclusively, expanded in *P. indica* and *S. vermifera* include domains predicted to be involved in hydrolytic degradation of plant cell wall carbohydrates (Glycosyl hydrolase families GH10, GH11 and GH61), carbohydrate binding (LysM, WSC and CBM_1), protein binding (WD40, NACHT and TPR_4) together with proteins most probably involved in signalling and regulation of cellular responses to stress and nutrient availability (NB-ARC, G-alpha protein, F-box, RAS and RHO families) (Table 2.3). Domains that are strongly overrepresented in *P. indica* but not in *S. vermifera* are found in proteins involved in proteolysis (Metallopeptidases families M36 and M43) and reverse transcription (RVT1 and RVT2). In contrast to this, domains overrepresented in *S. vermifera* but not in *P. indica* are found in proteins involved in membrane degradation (Patatin), interaction (Adaptin) or transport (HEAT) (Table 2.3).

Table 2.3: Protein domain expansions in *P. indica* and *S. vermifera*.

Domain	Pfam Accession	*P. indica*[1,2] *S. vermifera*	Population mean[1]	Significance[1]
Degradation of plant derived carbohydrates				
Glycosyl hydrolase family 10	PF00331	22 11	3.6 3.8	3.32E-044 2.21E-019
Glycosyl hydrolase family 11	PF00457	16 7	1.1 1.3	1.10E-052 1.91E-023
Glycosyl hydrolase family 61	PF03443	27 36	11.9 11.8	2.51E-015 5.80E-025
Proteolysis				
Peptidase_M36	PF02128	9 2	3.3 3.4	1.63E-014 0.02
Peptidase_M43	PF05572	12 3	2.1 2.3	2.57E-024 0.2
Carbohydrate binding				
Chitin / Peptidoglycan binding (LysM)	PF01476	58 27	8.9 9.4	2.04E-047 2.35E-019
Beta-glucan binding (WSC)	PF01822	109 99	13.6 13.8	3.32E-045 2.80E-041
Cellulose binding (CBM_1)	PF00734	69 55	13.7 14	1.13E-033 4.94E-026
Protein-protein interaction				
NACHT domain	PF05729	168 222	49.9 49	9.05E-021 1.23E-029
WD domain, G-beta repeat	PF00400	1458 2431	741.3 724.8	2.11E-014 1.47E-033
Tetratricopeptide repeat 4	PF07721	2 8	0.4 0.3	1.73E-014 4.77E-063
NB-ARC domain	PF00931	48 111	8.1 7	6.50E-025 3.59E-056
Reverse transcriptase domains expanded in *P. indica*				
Reverse transcriptase (RNA-dependent DNA polymerase) rvt1	PF00078	17 2	9.1 9.4	2.80E-005 7.75E-005
Reverse transcriptase (RNA-dependent DNA polymerase) rvt2	PF07727	24 2	4 4.3	2.53E-031 0.01

Integrase core domain	PF00665	20 3	4.5 4.8	2.76E-019 0.1
Membrane associated domains expanded in *S. vermifera*				
Patatin-like phospholipase	PF01734	10 184	9 6.1	0.8 4.37E-079
Adaptin N terminal region	PF01602	9 48	10.2 9.5	0.1 7.50E-070
HEAT repeat	PF02985	1 24	3.5 3.1	1.06E-006 2.66E-059

[1]if values differed between both fungi, the upper value refers to *P. indica*, the lower to *S. vermifera*

[2]domain numbers in red are significantly ($p < 0.05$) overrepresented. Numbers in blue refer to significantly underrepresented domains. Upper rows refer to values for *P. indica*, lower ones for *S. vermifera*.

2.1.4. Key genes in the plant fungus interaction

The identification and characterization of genes involved in retrieving nutrients in the background of the interaction with a living plant host was in the focus of the performed analyses. Genomic capabilities of *P. indica* and *S. vermifera* were therefore focused on five main categories of proteins predicted to be involved in intra- and extracellular transport (Transporters), proteolysis (Peptidases), degradation of carbohydrates (Hydrolases), non-destructive carbohydrate binding (Lectins) and small, secreted proteins which are important for successful plant colonization ("Effectors"). Characteristics of both fungi within these groups were identified by a comparison against the 58 genomes of Basidiomycetes and Ascomycetes fungi listed in Table 5.17 which were additionally grouped into 4 categories: Ectomycorrhizal and lichen forming symbiosis (ECM/LF), endophytic symbiosis (EP), Saprotrophic fungi including soil decomposer, white rots and brown rots (SD/WR/BR) and animal and plant pathogens (AP/PP). Marked differences could be identified within these groups in both Sebacinales fungi in comparison to the other Basidiomycetes (Figure 2.6) which will be explained in more detail in the following four subchapters.

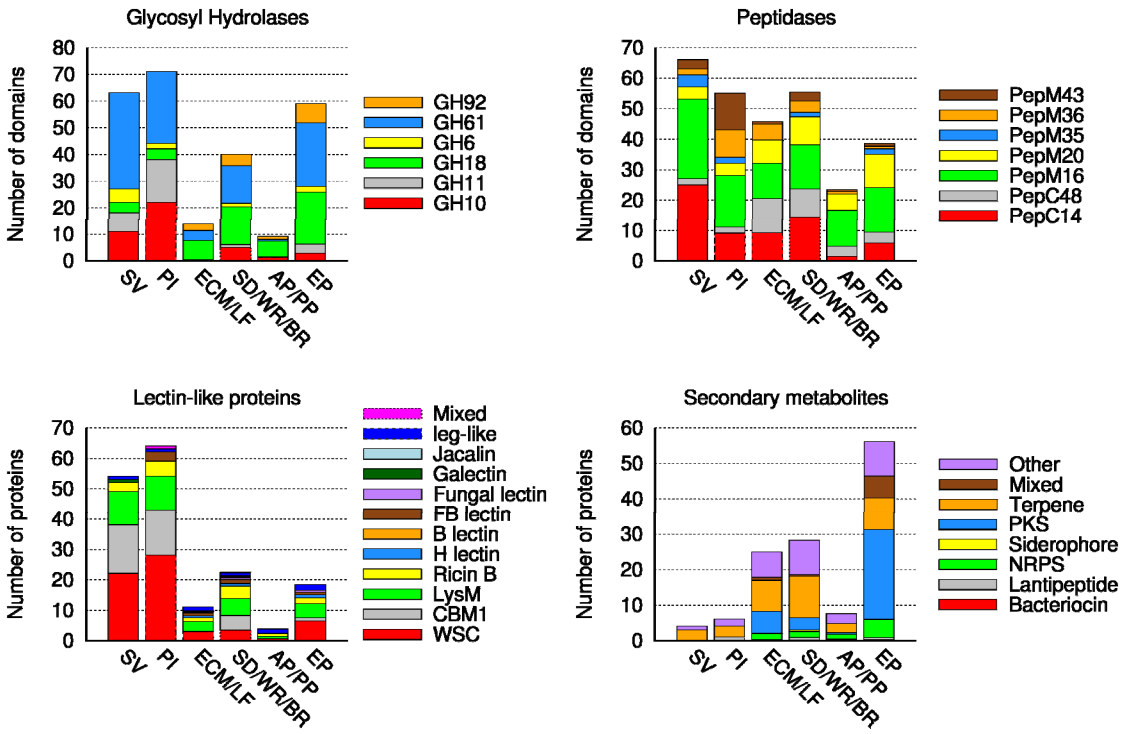


Figure 2.6: Protein and domain expansions in *P. indica* and *S. vermifera*. Functional protein domains were predicted using the Pfam database (Finn et al., 2010). Prediction of peptidases is supported by a comparison against the MEROPS database (Rawlings et al., 2011). Lectin-like proteins are proteins which only contain one or a combination of the shown non-enzymatic domains and were identified with the LectinFinder program (Chapter 4.7.1.12). Restrictions in genes involved in secondary metabolite production were verified using the antiSMASH program (Medema et al., 2011). Bar charts were created using gnuplot (Williams and Kelley, 2012).

2.1.4.1. Transmembrane transporter

Using a combination of the TransportDB prediction server (Li et al., 2008) and the Transporter Classification DataBase (TCDB) (Saier et al., 2009) 357 proteins putatively involved in the transmembrane transport of different substances were identified in *P. indica*. According to the TCDB classification system, these transporters could be grouped into 56 families of which 11 (25 proteins) were channel-type facilitators or membrane spanning pores, 36 (249 proteins) were secondary carrier-type facilitators whose transport is driven by an electrochemical potential and 9 (83 proteins) were primary active transporters whose transport is ATP driven. Similar numbers were found for *S. vermifera* with a total of 54 transporter families (391 proteins) divided into 9 (30 proteins) channel-type facilitators, 36 (271 proteins) secondary carrier-type facilitators and 9 (90 proteins) primary active transporters. Besides similar numbers, also the types of identified transporters in both fungi are predominately very similar including nutrient transporters involved in the translocation of carbohydrates, nitrogen, phosphate and metal ions as well as a large number of ATP-binding

cassette (ABC) transporters which are involved in the detoxification of different secondary metabolites (Coleman and Mylonakis, 2009; Coleman et al., 2011; Stefanato et al., 2009; Wang et al., 2013a; Weber et al., 2012). The major difference between both fungi is the presumable lack of a classical nitrate transporter in *P. indica*. To verify that this transporter was not missed due to wrong gene prediction, the identified nitrate transporter of *S. vermifera* was used to search *P. indica* proteins and scaffolds using Psi-Blast and tBlastn (e-Val < 10^{-3}). Both Blast comparisons did not result in any significant hit supporting the lack of this transporter in *P. indica*. This could be experimentally verified by growing both fungi on media containing different nitrogen sources. On medium with nitrate as sole nitrogen source, *P. indica* showed a significantly (one-way ANOVA: $p < 0.01$) reduced growth which was comparable to growth on medium without any nitrogen source (Figure 2.7, left chart and Figure 2.8).

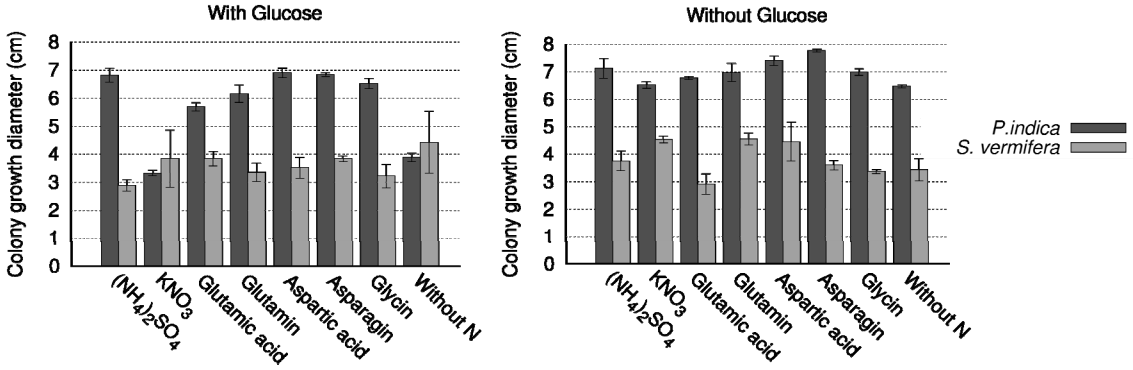


Figure 2.7: Growth assay on different nitrogen sources. Minimal medium plates were inoculated as described in chapter 4.3.2 and *P. indica* and *S. vermifera* colonies grown for 3 weeks at 25°C in the dark. Values are shown in relation to the control where no nitrogen source was added to the medium. Error bars are transformed standard errors of three independent biological repetitions. Bar charts were created using gnuplot (Williams and Kelley, 2012).

The significant difference in growth disappeared when no carbon source was added to the medium (Figure 2.7, right chart) which might be related to the production of thinner and faster growing hyphae on these media (Figure 2.8, *P. indica* +/- Glc). For *S. vermifera* a ring like growth could be observed with a complete removal of older hyphae during growth on medium without carbon source (Figure 2.8, *S. vermifera* "w/o Nitrogen - Glc" and "Nitrate - Glc"). Because this phenotype could be partially complemented by using amino acids as nitrogen source and completely complemented by adding glucose, the removal of hyphae by *S. vermifera* is probably due to carbon limiting conditions.

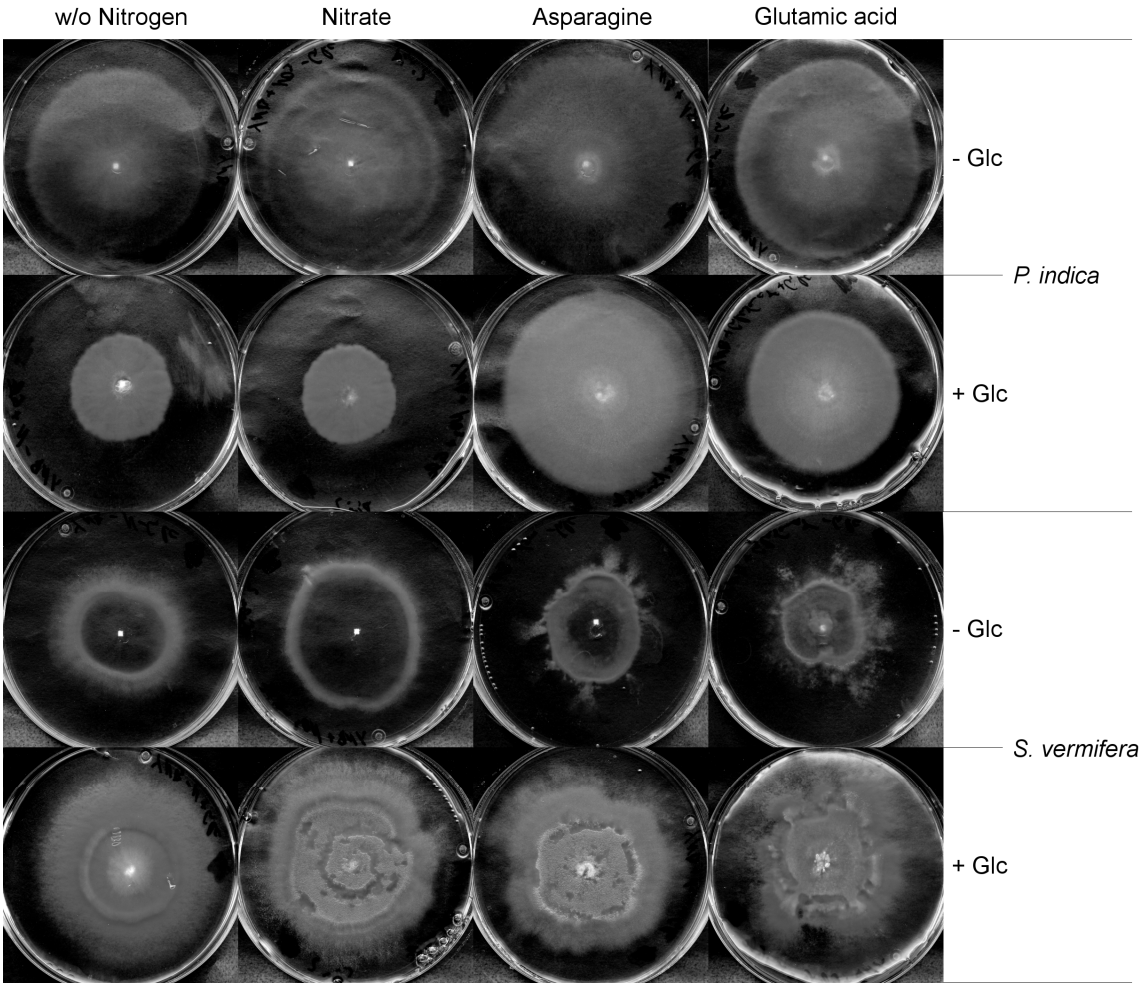


Figure 2.8: Growth effect of different nitrogen sources in the presence and absence of glucose. Minimal medium plates were inoculated as described in chapter 4.3.2 and *P. indica* and *S. vermifera* colonies grown for 3 weeks at 25°C in the dark. White dots in the middle of the culture plates indicate the point of inoculation.

2.1.4.2. Hydrolases and peptidases

The cell wall of plants consists mainly of cellulose, lignin and hemicellulose and the breakdown of these polysaccharides by fungi is accomplished by a variety of different enzymes, many of them described under the collective term glycosyl hydrolases (GH) (Murphy et al., 2011). Domain prediction could identify members of 45 different GH families in the analysed dataset of which 25 were present in 45 (75%) of the 60 analyzed genomes. Members of 33 (153 proteins) different GH families were represented in *P. indica* and 31 (145 proteins) in *S. vermifera*, respectively. By grouping the analyzed fungi according to their different lifestyles, a reduced number of GH domains could be observed for fungi living in an ectomycorrhizal or lichen-forming symbiosis as well as fungi which live as pathogens (Figure 2.6, Glycosyl Hydrolases). The latter observation is here slightly biased as the genomes of

three animal pathogens were included which naturally lack hydrolytic enzymes acting on plant substrates, but the data is also confirmed when focusing on plant pathogenic fungi alone. The differences in GH domains between the Sebacinales and other fungi are mainly due to six families. Domains belonging to GH 10, 11 (xylanases) and 61 (monooxygenases) are strongly overrepresented in both fungi (Table 2.3 and Figure 2.6) and GH6 (cellulases, glucanases) domains only in *S. vermifera*. On the contrary, no GH92 (mannosidases) and a strongly reduced number of GH18 (chitinases, acetylglucosaminidases) domains could be identified in both Sebacinoid fungi although these domains are well represented in the other analyzed fungal genomes (Figure 2.6, Glycosyl Hydrolases).

The amount and different types of predicted peptidase domains were overall comparable between both fungi with some defined differences (Figure 2.6, Peptidases). To support the domain prediction for peptidases and peptidase inhibitors, a batch Blast (e-Val < 10^{-4}) of all *P. indica* and *S. vermifera* protein coding genes against the MEROPS database (Rawlings et al., 2011) was performed. Combining domain prediction and Blast results, a total number of 355 peptidases and peptidase inhibitors could be identified in *P. indica* which further group based on their catalytic type into 46 aspartic-, 62 cysteine-, 121 metallo-, 100 serine- and 20 threonine-peptidases. Furthermore, one protein with an unknown catalytic type was identified as well as 5 peptidase inhibitors. The prediction for *S. vermifera* resulted in overall similar numbers with 361 peptidases grouped into 24 aspartic-, 78 cysteine-, 119 metallo-, 113 serine-, 20 threonine-peptidases. Also here, one protein with an unknown catalytic type was found and 6 peptidase inhibitors. However, distinctive differences could be observed within some of these families. The genome of *P. indica* for example encodes 31 proteins belonging to a family of transposon peptidases (A11) which were shown to be involved in the retrotransposition process in *S. cerevisiae* (Garfinkel et al., 1991) while the *S. vermifera* genome contains only 6 of those. On the contrary, the genome of *S. vermifera* encodes 26 putative caspases/metacaspases (C14), which are primary activators of apoptotic cell death (Earnshaw et al., 1999; Thornberry and Lazebnik, 1998), while there are only 9 of these proteins in *P. indica*. Further differences in the expansions of peptidases in *P. indica* and *S. vermifera* are found in the metallo-peptidase families M35 (deuterolysin; 5 in *P. indica* and 18 in *S. vermifera*), M36 (fungalysin; 9 in *P. indica* and 2 in *S. vermifera*) and M43 (cytophagalysin; 13 in *P. indica* and 3 in *S. vermifera*) (Boldt et al., 2001; Brouta et al., 2001; Doi et al., 2003).

2.1.4.3. Carbohydrate binding proteins

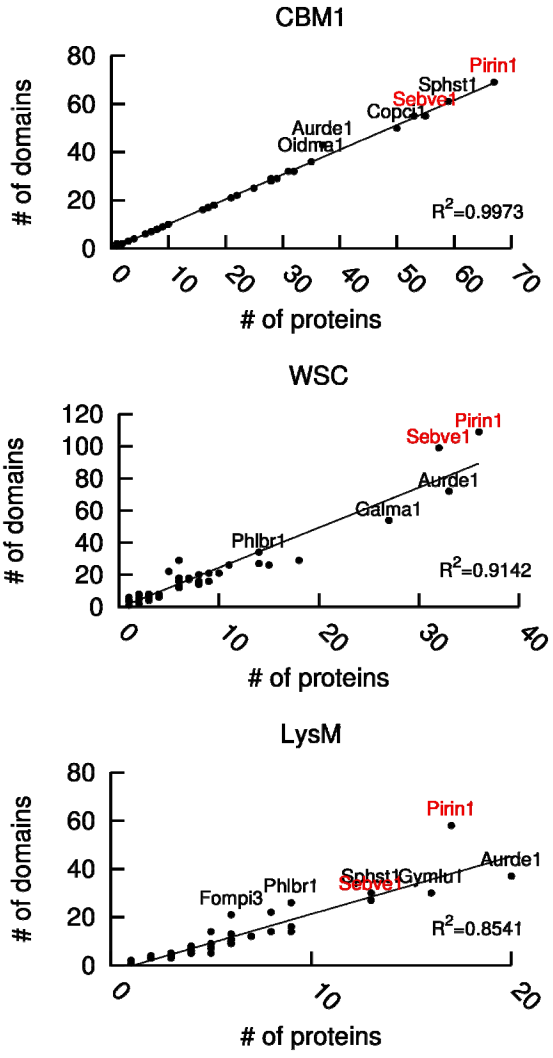


Figure 2.9: Expansion of carbohydrate binding domains (CBD) and CBD-containing proteins in *P. indica* and *S. vermifera*. Names of genomes with low abundance of CBDs are not shown for visualization purposes. Plots were created using gnuplot (Williams and Kelley, 2012).

In total, 120 proteins in *P. indica* and 100 in *S. vermifera* were predicted to contain a carbohydrate binding domain that is putatively involved in binding to chitin (LysM), glucan (WSC) or cellulose (CBM1). 65 (54%) of these proteins in *P. indica* (50 (50%) in *S. vermifera*) occur in combination with catalytic enzyme domains like deacetylase or hydrolase domains. The remaining proteins only contain one or more of the carbohydrate binding domains thereby resembling a lectin-like structure. Looking on both types in all 60 analyzed genomes, a strong correlation ($R^2 > 0.85$) could be identified between the numbers of proteins containing a carbohydrate binding domain and the overall number of the domain (Figure 2.9). This indicates that, although these proteins are probably involved in different biological processes, the carbohydrate binding property requires a relatively fixed number of domains per protein. It is interesting in this context that the number of carbohydrate-binding domains, while similar in other fungi, was found to be increased for WSC domain containing proteins in *S. vermifera* and *P. indica* and for LysM domain containing proteins only in *P. indica*. The evolution of the WSC and LysM domains was therefore further analyzed. For this, a special focus was put on lectin-like LysM and WSC proteins, respectively. The identification of these proteins in all analyzed genomes was performed using the LectinFinder program (Chapter 4.7.1.12). Based on the output of the Pfam prediction, proteins were searched which contain exclusively one or a combination of the following domains: WSC (Pfam accession: PF01822), CBM_1 (PF00734), LysM (PF01476), Lectin_C (PF00059), Ricin_B_lectin (PF00652), Gal_Lectin (PF02140), H_lectin

(PF09458), Lectin_legB (PF00139), Fim-adh_lectin (PF09222), Lectin_N (PF03954), B_lectin (PF01453), FB_lectin (PF07367), Fungal_lectin (PF07938), Gal-bind_lectin (PF00337), Intimin_C (PF07979), Jacalin (PF01419), Lectin_leg-like (PF03388). A total number of 1147 lectin-like proteins containing 1761 domains (1.54 domains per protein) could be identified in all 60 analyzed genomes. While the majority of the proteins contained only one type of lectin domain, a combination of different types was found only in 10 proteins. These include WSC/CBM_1 combinations in *A. delicata* (3 proteins), Fungal_lectin/Jacalin combinations in *B. adusta* (1 protein), *P. carnosa* (2 proteins) and *P. gigantea* (1 protein), WSC/LysM combinations in *P. indica* (1 protein) as well as CBM_1/Fungal_lectin combinations in *S. lacrymans* (2 proteins).

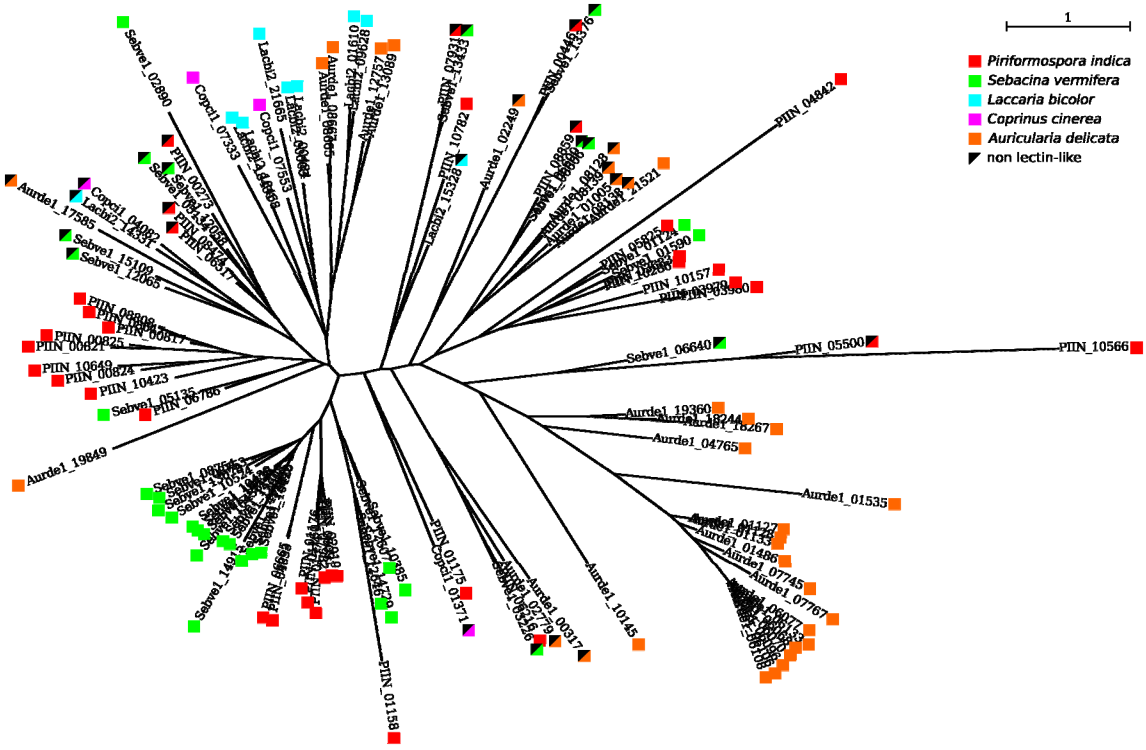


Figure 2.10: Expansions of WSC domain containing proteins in selected Basidiomycetes. The phylogram was created using seaview (Gouy et al., 2010). An initial alignment of concatenated WSC domains was created using MUSCLE (Edgar, 2004) and afterwards manually refined. Phylogenetic relations were calculated based on this tree using PhyML (Guindon et al., 2010) with a LG model and BioNJ as starting tree. The optimal tree was then searched by applying both available methods, Nearest Neighbor Interchanges (NNI) and Subtree Pruning and Regrafting (SPR) and the best performing method used for the tree. Different colours added to the genes indicate the respective origin as shown in the figure legend while additional black triangles indicate that the WSC domains are combined with non-carbohydrate binding domains in the gene.

Based on the previous observations and the identification of lectin-like proteins, WSC and LysM domain expansions were characterised phylogenetically in *P. indica* (18 LysM, 36 WSC domain containing proteins), *S. vermifera* (13 LysM, 32 WSC), *L. bicolor* (2 LysM, 9

WSC), *C. cinerea* (4 LysM, 4 WSC), *U. maydis* (1 LysM, 0 WSC) and *A. delicata* (20 LysM, 33 WSC) (Figure 2.10 and Figure 2.12). As shown before, major expansions of proteins containing WSC domains were found in *P. indica*, *S. vermifera* and *A. delicata*. From the phylogenetic analyses different groups could be identified. Lectin-like WSC proteins from the five analysed fungi clustered thereby in individual groups while non lectin-like proteins from different fungi grouped together (Figure 2.10). Together with the strong expansion of lectin-like proteins and the likely assumption that lectins and non-lectins are involved in different biological processes, the WSC domains of both groups are probably under different evolutionary pressures which results in species-specific gene duplications.

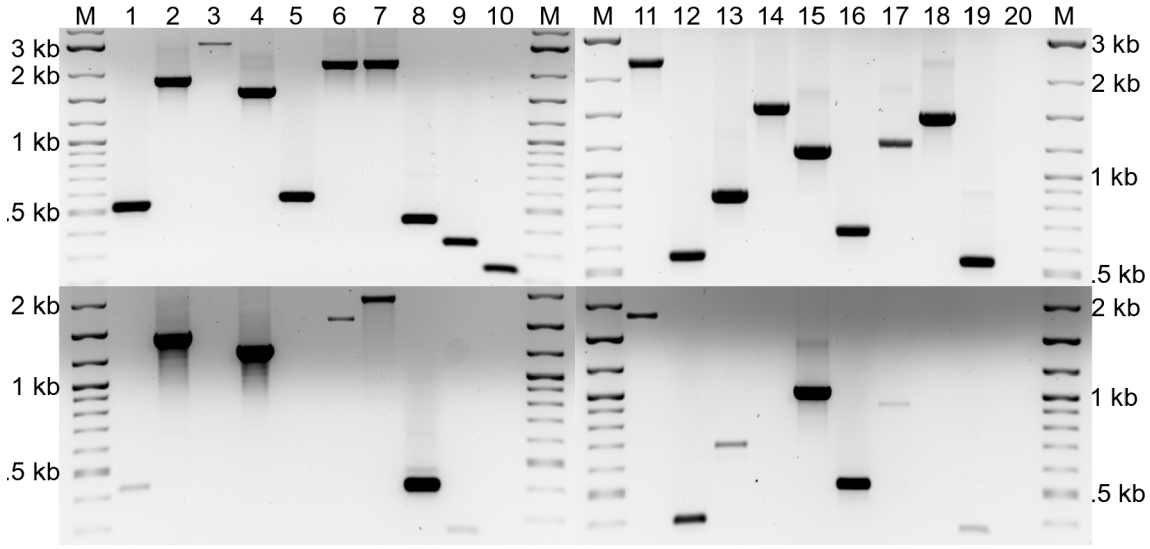


Figure 2.11: Verification of gene prediction and expression of LysM proteins. PCR on genomic (upper panels) and cDNA (lower panels) with specific primers for each LysM containing protein within the *P. indica* genome. (1) *PIIN_00867*, (2) *PIIN_02169*, (3) *PIIN_02170*, (4) *PIIN_02172*, (5) *PIIN_02781*, (6) *PIIN_03654*, (7) *PIIN_03655*, (8) *PIIN_05278*, (9) *PIIN_05722*, (10) *PIIN_05723*, (11) *PIIN_06786*, (12) *PIIN_08720*_left_part, (13) *PIIN_08720*_right_part, (14) *PIIN_08720*_left2right_part, (15) *PIIN_08721*, (16) *PIIN_08723*, (17) *PIIN_08724*, (18) *PIIN_08725*, (19) *PIIN_11109*, (20) neg. control.

Similar observation could also be made for LysM domain containing proteins. 18 of these proteins were initially identified in the genome of *P. indica*. As the high number as well as the domain structure were quite unusual in comparison to LysM proteins found in other fungi, including *S. vermifera*, these proteins were used to verify the domain combinations and with that the gene prediction experimentally. Therefore, primers were designed spanning the gene region from within the left most domain to the right most. Using a mix of cDNA templates from *P. indica* grown on CM and different timepoints of colonized barley roots, the expression and domain structure of almost all LysM proteins could be verified (Figure 2.11). Furthermore, three modifications to the actual gene set were made based on the obtained

results. An alternative gene prediction was used for gene *PIIN_02171* which has a reduced length that better fits its putative paralog *PIIN_02170* and the gene was then renamed to *PIIN_20000*. Gene *PIIN_08720* was shown to represent two separate genes and two gene prediction fitting this model were chosen and renamed to *PIIN_20001* and *PIIN_20002* (Figure 2.11, lane 12-14). The genes *PIIN_05722* and *PIIN_11109* which are predicted as partial genes at the end of their respective contigs 140 and 1310 were shown to be one single gene. By DNA sequencing of the gap, the full gene was constructed and renamed to *PIIN_20003* and contig 1310 was fused to the start of contig 140.

Analysis of this final set of 18 LysM containing proteins showed that 16/18 were predicted to be secreted. Most of the proteins show a lectin-like structure and their respective genes occur physically clustered in the genome (*PIIN_20000*, *PIIN_20001*, *PIIN_08723*, *PIIN_08724*, *PIIN_08725*), yet forms which combine LysM domains with a chitin deacetylase domain (*PIIN_02169*, *PIIN_02172* and *PIIN_03655*) a peptidases domain (*PIIN_02170* and *PIIN_20000*), a hydrolase domain (*PIIN_03654*) or a transmembrane domain (*PIIN_02781*) exist. Similarly, 11/13 LysM domain containing proteins in *S. vermifera* have a lectin-like structure and their respective genes also cluster physically (*Sebve1_01227*, *Sebve1_01228*, *Sebve1_01229*, *Sebve1_01230*). The phylogenetic analysis of the LysM proteins resulted in a similar clustering pattern as for the WSC proteins but with a more distinct separation of the individual groups (Figure 2.12). Two groups could be identified containing a single member from each fungus (Figure 2.12, lower right branches). Except for proteins from *A. delicata*, proteins in the left of these two groups contain no signal peptides, single LysM domains and a transmembrane domain. Similar, all proteins in the right group, except those from *A. delicata*, contain a signal peptide and 2 LysM domains. Based on the present data, also the LysM domain containing proteins seem to have evolved by several gene duplication events. However, no transposable elements could be identified in the proximity of any of the LysM or WSC proteins in *P. indica* or *S. vermifera* and it is therefore more likely that the observed protein expansions are a result of unequal recombination events that occurred either in the ancestral or independently in each fungus. It remains unclear from this analysis, why more LysM domains occur per protein in *P. indica* in comparison to other fungi (Figure 2.9, LysM). Phylogenetic analyses were therefore performed on these domains alone (Figure 2.13).

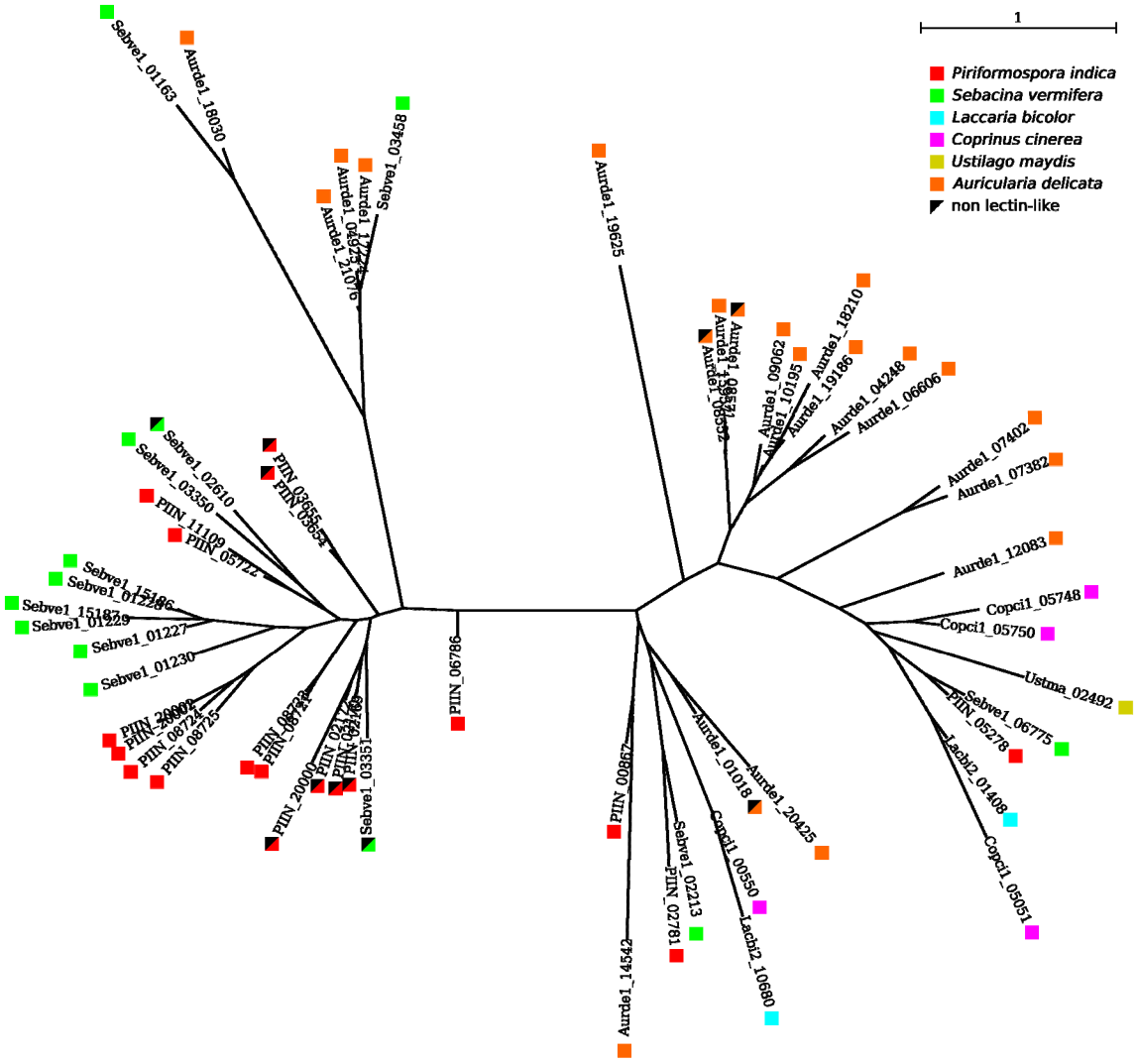


Figure 2.12: Expansions of LysM domain containing proteins in selected Basidiomycetes. The phylogram was created using seaview (Gouy et al., 2010). An initial alignment of concatenated LysM domains was created using MUSCLE (Edgar, 2004) and afterwards manually refined. Phylogenetic relations were calculated based on this tree using PhyML (Guindon et al., 2010) with a LG model and BioNJ as starting tree. The optimal tree was then searched by applying both available methods, Nearest Neighbor Interchanges (NNI) and Subtree Pruning and Regrafting (SPR) and the best performing method used for the tree. Different colours added to the genes indicate the respective origin as shown in the figure legend while additional black triangles indicate that the LysM domains are combined with non-carbohydrate binding domains in the gene.

The results suggest that LysM domain containing proteins initially duplicated in *P. indica* and that single domains afterwards duplicated within these proteins. Such duplication events could not be identified in LysM proteins of *S. vermifera* or *A. delicata* which is consistent with the average domain per protein ratio in these fungus (Figure 2.9, LysM).

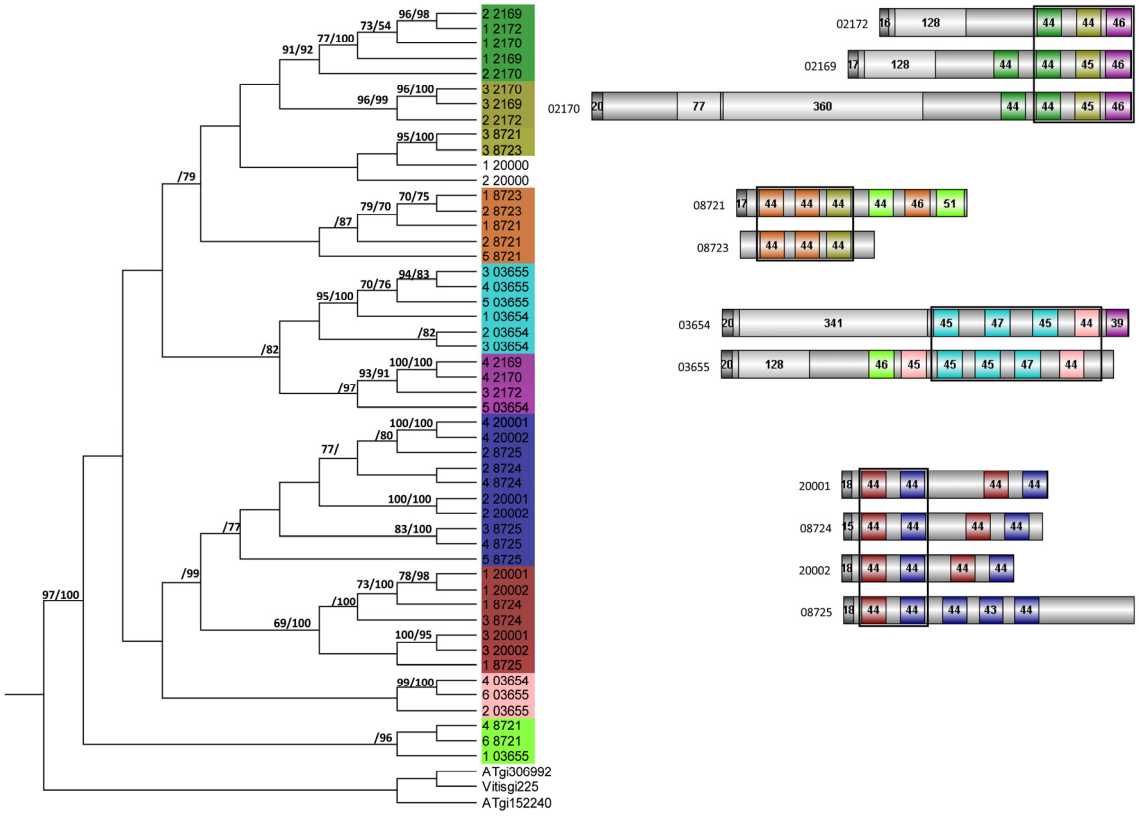


Figure 2.13: Structure of LysM domains within the *P. indica* genome. From (Zuccaro et al., 2011).

It has been shown that the tomato pathogen *C. fulvum* secrets lectin-like LysM effectors during plant colonization that are binding either to the fungal hyphae or chitin fragments released during penetration of plant cells. *P. indica* has a broad host range and lectin-like domains are expanded within its genome. It was therefore tested whether *P. indica* uses similar strategies to evade plant recognition by fusing two candidate proteins to a green fluorescent protein (GFP) and localization analyzed using confocal microscopy. The first protein (PiChiDe1, encoded by *PIIN_02172*) is predicted to contain a signal peptide (SP), a polysaccharide deacetylase 1 domain and three LysM domains and the second protein (PiLysM1, encoded by *PIIN_08721*) contained a SP and six LysM domains (Figure 2.14).

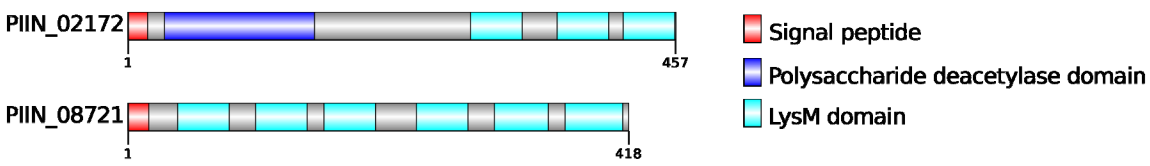


Figure 2.14: Domain structure of the PiChiDe1 (PIIN_02172) and PiLysM1 (PIIN_08721) proteins from *P. indica*. Proteins and predicted domains are drawn in size relations based on Pfam predictions. Graphs were created using DOG2.0.1 (Ren et al., 2009).

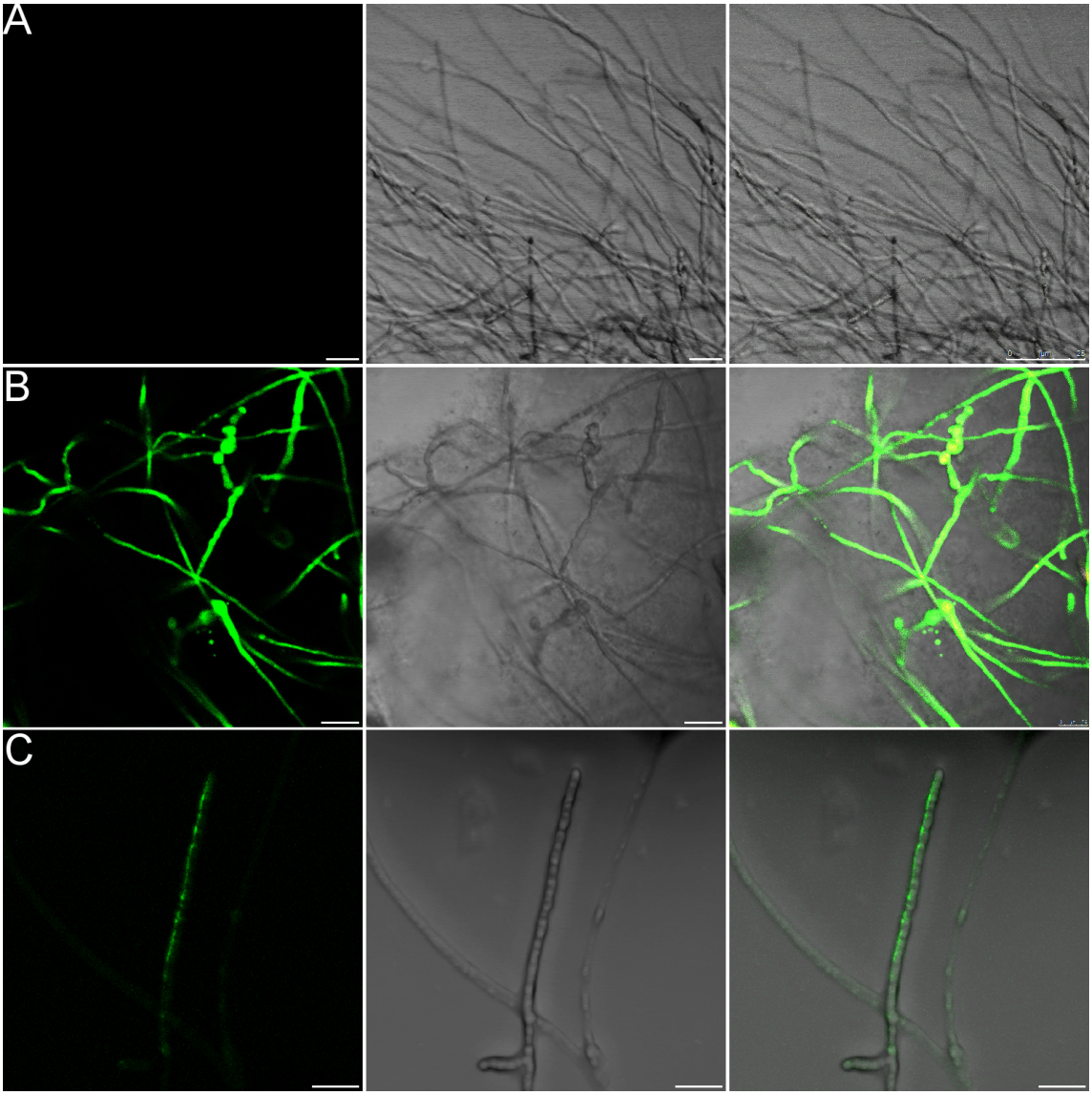


Figure 2.15: Localization of constitutively produced PiLysM1 (PIIN_08721) in axenic culture. Auto-fluorescence of the wildtype (negative control) could not be observed with the used settings (A). Fluorescence of intracellular GFP (positive control) could be observed all over the cytoplasm (B). GFP C- terminally fused to PiLysM1 could be observed at the hyphal cell wall (C). Expression of the construct is under control of the constitutively active GPD promoter. Scale bars represent 10 µm.

The deacetylase domain is found, among others, within chitin deacetylases, enzymes which are involved in the hydrolytic conversion of chitin into chitosan. It was therefore hypothesized that PiChiDe1 binds chitin in the hyphal cell wall via its LysM domains and deacetylates it afterwards to chitosan. For PiLysM1 a function similar to one of the *C. fulvum* effectors was assumed. Both genes were initially cloned under the control of their native promoters, but no fluorescent signals could be observed at several different tested timepoints. Because a control expressing intracellular GFP under the control of the constitutively active GPD promoter resulted in a strong GFP signal distributed all over the hyphae, it could be excluded that this is due to the GFP-tag in general (Figure 2.15, B). Both genes were then cloned in front of GFP

and under to control of the GPD promoter in order to test whether the lack of fluorescence is an expression related problem. Six to eight different transformants of each construct were tested in axenic culture. A distinct fluorescence signal could be observed for both genes at the hyphal wall (Figure 2.15 C and Figure 2.16 A-C). For PiChiDe1, the distribution was patchy and concentrated at hyphal tips (Figure 2.16 A) and hyphal septa (Figure 2.16 B). Because these are known sides of secretion, the observed fluorescence might not indicate the real localization target (Hayakawa et al., 2011; Wösten et al., 1991). The fluorescence signals shown here were very rare and could almost only be observed close to the tips of young hyphae. Furthermore, for none of the constructs could a clear signal be observed *in planta*.

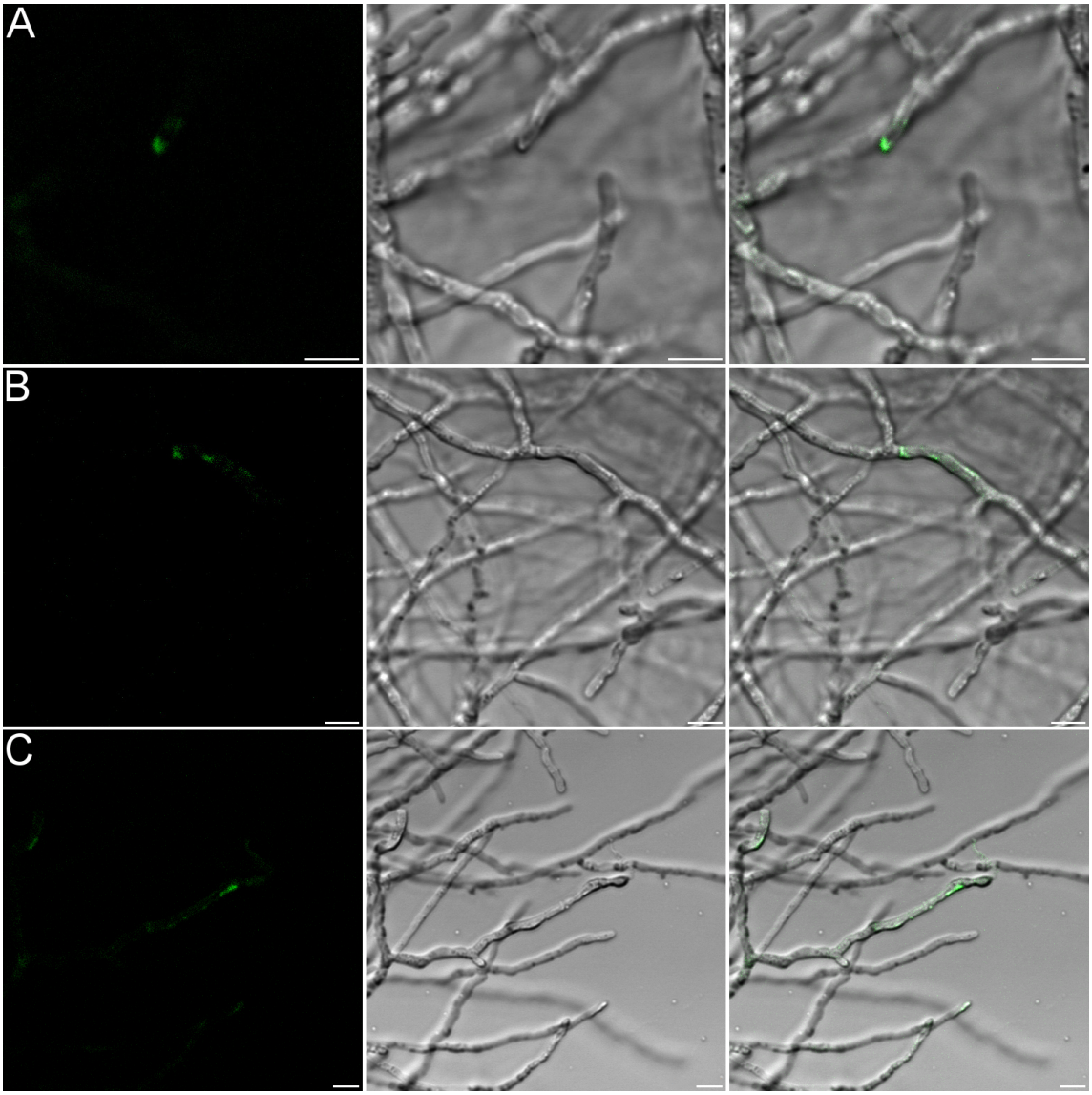


Figure 2.16: Localization of constitutively produced PiChiDe1 (PIIN_02172) in axenic culture. Fluorescence of GFP C-terminally fused to PiChiDe1 could be observed at hyphal tips (apical secretion; A), at cell junctions (septal secretion; B) and at the hyphal cell wall (B, C). Expression of the construct is under control of the constitutively active GPD promoter. Scale bars represent 5 µm.

2.1.4.4. Secondary metabolite production

The identification of domains and proteins involved in secondary metabolism can provide important information towards the lifestyle and interaction behaviour with other organisms. Therefore, proteins that may function in the production of secondary metabolites like polyketides or non-ribosomal peptides were identified on the basis of characteristic domains within them (Brakhage, 2013; Keller et al., 2005). All of the protein domains were found to be underrepresented in both, *P. indica* and *S. vermifera* (Table 2.4).

Table 2.4: Secondary metabolite domains in *P. indica* and *S. vermifera*.

Domain	Pfam Accession	*P. indica*[1] *S. vermifera*	Population mean[1]	Significance[1]
Acyl transferase	PF00698	1	8.9	5.95E-007
β-ketoacyl synthase (N- and C-terminal)[2]	PF00109	1	8.9	5.31E-007
Retrotransposon gag protein	PF03732	3 1	21 21.1	2.49E-004 5.60E-005
Phosphopantetheine attachment site	PF00550	3 2	18 18.1	6.63E-009 1.09E-009
Transposase family tnp2	PF02992	3 10	12.5 12.4	3.51E-006 0.2
Heterokaryon incompatibility protein (HET)	PF06985	0	41	5.22E-005
O-methyltransferase	PF00891	0	11.5	9.13E-014
Condensation	PF00668	0	6.1	6.73E-008
CDR ABC transporter	PF06422	0	5.1	7.52E-020
β-ketoacyl reductase	PF08659	0	4.9	7.48E-006

[1]if values differed between both fungi, the upper value refers to *P. indica*, the lower to *S. vermifera*

[2]only values for the N-terminal domain are shown but they are almost identical for the C-terminal domain

The results from the domain analysis were verified using the antiSMASH database (Medema et al., 2011). Consistent with the domain prediction, a screening of the *P. indica* and *S. vermifera* genomes for proteins involved in the production of antibiotics and secondary metabolites resulted in the classification of two proteins in *P. indica* and one in *S. vermifera* that are putatively involved in the synthesis of nonribosomal peptides (NRP). Blastp analysis of these proteins against the nr-database (e-Val < 10^{-3}) showed high similarity to an alpha-aminoadipate reductase, an enzyme involved in the biosynthesis of lysine. All three proteins have a similar domain structure, while one of the *P. indica* proteins seems to be a truncated

version of the other. Domain predictions revealed the presence of a putative condensation domain, an essential domain in non-reducing polyketide synthases (NRPKSs), in the untruncated proteins in *P. indica* and *S. vermifera*, but the predictions were discarded due to a low significance value. Additionally, 3 genes involved in terpene production were found in both, *P. indica* and *S. vermifera*. Furthermore, two genes in *P. indica* were predicted to be involved in the production of lantipeptides, ribosomally synthesized and posttranslationally modified peptides with a potential antibiotic activity (Schnell et al., 1988; Wang and van der Donk, 2012). In comparison to other fungi, the Sebacinoid isolates have a strongly reduced sets of proteins putatively involved in the production of secondary metabolites (Figure 2.6, Secondary metabolites).

2.1.4.5. Effector-like proteins

The term “effector” is applied to proteins which are secreted and have a crucial role during host colonization by interfering with the plant defence system or metabolism (Djamei et al., 2011; de Jonge et al., 2011; Kloppholz et al., 2011; De Wit et al., 2009). Furthermore, as effectors are expected to be effectively produced and translocated to their respective place of action, they are often smaller than 300 aa in size. Because putative effector candidates might be important determinants in the establishment of a compatible plant fungus interaction, the following annotation steps were performed on the Sebacinales proteins additionally to Blast2GO analyses and domain predictions. 1) Sequence similarity to experimentally verified effectors: Databases like the PHI (Pathogen-Host-Interaction) database contain curated molecular and biological information on genes proven to affect the outcome of pathogen-host interactions (Winnenburg et al., 2008). 2) Structure prediction: The tertiary structure prediction of proteins based only on their amino acid sequence focuses on known folding mechanisms and structure homology to crystallized proteins (Kelley and Sternberg, 2009). The prediction of such a structure might aid in estimating a function. 3) Motifs: Amino acid motifs are small, conserved stretches in the amino acid sequence. These motifs, although rarely giving indications of putative functions, can help in classifying effector-candidates into functional groups. Amino acid motifs searched for included an translocation signal from the oomycete *Phytophthora infestans* RxLR-EER (Whisson et al., 2007), an N-terminal motif of haustoria-producing biotrophic and hemibiotrophic fungi [Y/F/W]xC (Godfrey et al., 2010) and yet uncharacterized motif of necrotrophic oomycetes YxSL[R/K] (Lévesque et al., 2010). 4) Transposable elements (TE): Effectors are suspected to be crucial for host colonization and are thus strongly subjected to an evolutionary "arms race" with its host (Boller and He,

2009; Jones and Dangl, 2006; Kamoun, 2006). A higher density of TEs and repeats and an association of these with effector candidates may indicate a higher rate of genome evolution, possibly due to selection pressure (Laurie et al., 2012).

Using these criteria, putative effector candidates were predicted in *P. indica* and *S. vermifera*. In total 867 proteins in *P. indica* and 1024 in *S. vermifera* were predicted to be secreted (RC < 3; TM <=1; Chapter 4.7.9.3) with an average length of 511 and 506 aa, respectively. Of these secreted proteins, 386 in *P. indica* and 460 in *S. vermifera* had a length of less than 301 aa with an average length of 172 and 170, respectively. These values are similar to those identified for the other 58 analyzed Basidiomycete genomes with an average of 939 secreted proteins (average length of 531 aa) of which 442 are smaller than 301 aa (average length 170 aa). Functional domains could be predicted in 126 of the *P. indica* and 120 of the *S. vermifera* small secreted proteins (SSP). 16 proteins in *P. indica* and 20 in *S. vermifera* contain domains of unknown function (DUF) or domains with a conserved structure but yet unknown function (e.g. CFEM). Together with Blast comparisons against the nr-database (e-Val < 10^{-3}) and against the PHI database (e-Val < 10^{-5}), a final set of 261 effector candidates in *P. indica* and 337 in *S. vermifera* could be identified. Manual revision of the amino acid sequences of all *P. indica* effectors resulted in the identification of a family of 25 highly similar proteins which share three different but highly conserved regions within their amino acid sequence (Figure 2.17).

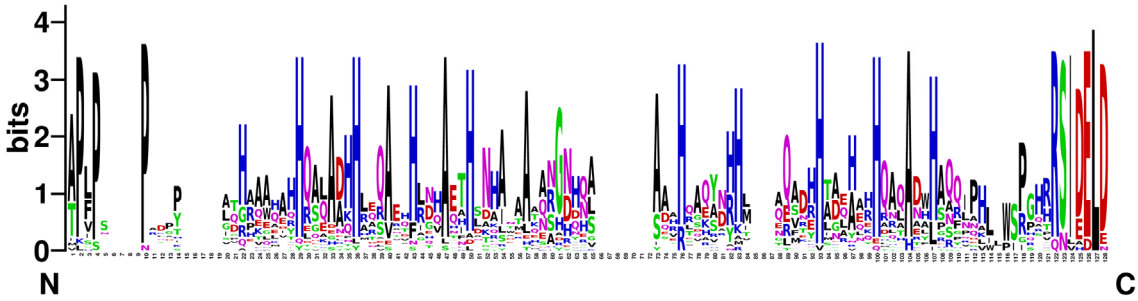


Figure 2.17: Conserved amino acid residues of the *P. indica* DELD proteins. Regions with regularly distributed histidine (blue) and alanine (black) residues are forming two anti parallel α-helices which are separated by a loop around the central glycine. The sequence logo was created using WebLogo (Crooks et al., 2004) based on a multiple sequence alignment of the 25 DELD proteins without the predicted signal peptides using MUSCLE (Edgar, 2004).

The central region contains several regular distributed histidine and alanine residues (Figure 2.17) which are separated into two parts by a central glycine in most of the proteins. Using the program Phyre[2] (Kelley and Sternberg, 2009) to predict the tertiary structure of these proteins showed that these histidine and alanine rich parts form two anti parallel α-helices which are

separated by a loop around the central glycine. Furthermore, the program as well as Blast results showed similarities to HRPII, a histidine rich protein in *Plasmodium falciparum*, the causative agent of malaria disease in humans (Parra et al., 1991). In the effectome of *S. vermifera* only two proteins with high similarity to the *P. indica* DELD proteins could be found. Besides searching for regular distributed histidines, also lysines and arginines were included as they may be functional redundant due to their shared basic character, which resulted in few additional matches. The search was then further expanded towards all secreted proteins including those larger than 300 aa and this revealed a group of 26 secreted proteins which have a degenerated RSIDELD motif at the C-terminus and a central region of regular distributed lysines, but are on average three times longer than DELD proteins from *P. indica* (452 aa).

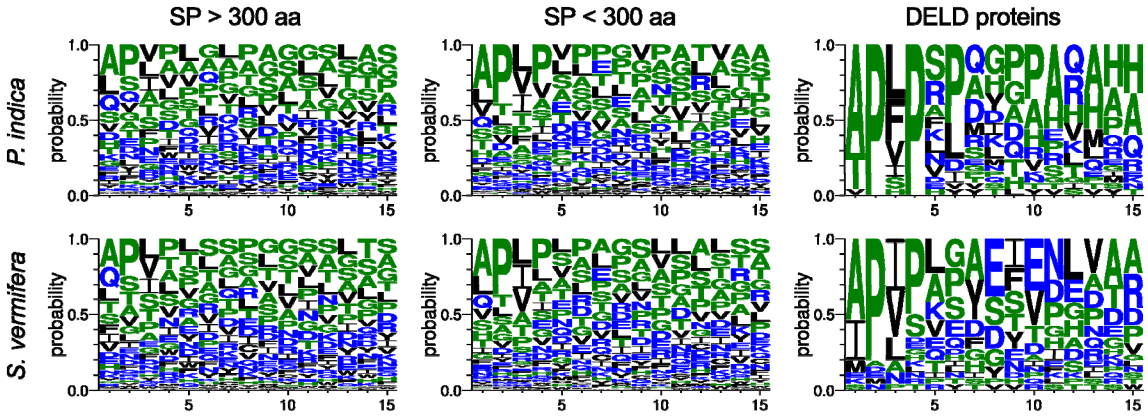


Figure 2.18: Amino acid abundance after the predicted signal peptide cleavage site in *P. indica* and *S. vermifera*. Secretion was predicted using TargetP with cleavage site predictions from SignalP (Emanuelsson et al., 2000; Nielsen et al., 1997). For *P. indica* 481 secreted proteins were longer than 300 aa and 386 smaller. For *S. vermifera* 564 secreted proteins were longer than 300 aa and 460 smaller. Sequence from 25 DELD proteins were used for *P. indica* (all smaller than 300 aa) and 26 for *S. vermifera* (5 smaller than 300 aa). The sequence logo was created using WebLogo (Crooks et al., 2004), whereas bigger letters and a higher position in the plot refer to a more frequent occurrence of the amino acid at the respective position.

At the N-terminal sequence and after the predicted signal peptide cleavage site two to four prolines residues occur of which the prolines at the +2 and +4 position are most highly conserved. At least a proline at the +2 position has been shown to be required for endoplasmic reticulum export in mammalian cells (Tsukumo et al., 2009). Looking at all secreted proteins in *P. indica* and *S. vermifera*, a proline occurs at this position in about 20% of all proteins that are secreted, but longer than 300 aa and in about 40% of all proteins that are secreted but smaller than 300 aa. Furthermore, all of the *P. indica* DELD proteins and the majority of the *S. vermifera* DELD proteins have a proline at the +2 position (Figure 2.18).

In contrast to that is the C-terminal consensus sequence "RSIDELD" (Arg-Ser-Ile-Asp-Glu-Leu-Asp) which is strongly reminiscent of the well characterised ER retention signals "KDEL" and "HDEL" (Gomord et al., 1997; Pelham, 1990). Proteins containing one of these ER retention signals were searched in *P. indica* and *S. vermifera* and their predicted signal peptide cleavage sites compared. In total five proteins in *P. indica* contain such a signal at the C-terminus of their sequence, but none contained a proline at the +2 position after the signal peptide cleavage site (SP_CVS): PIIN_03127 (hypothetical protein, HDEL), SP_CVS: AFA-LR; PIIN_03849 (putative S9 peptidase, KDEL), SP_CVS: no predicted signal peptide; PIIN_05777 (putative GH31 glucosidase, HDEL), SP_CVS: GFA-TQ; PIIN_06182 (putative hsp70, HDEL), SP_CVS: GHA-EE; PIIN_08996 (putative DNA binding protein, KDEL), SP_CVS: VLG-GR. Similarly do the six proteins in *S. vermifera* with an ER retention signal also not possess a proline at the +2 position after the SP_CVS: Sebve1_02564 (hypothetical protein, HDEL), SP_CVS: no predicted signal peptide; Sebve1_03945 (putative GH31 glucosidase, HDEL), SP_CVS: VRA-AS; Sebve1_07494 (putative hsp70, HDEL), SP_CVS: VRA-EE; Sebve1_10364 (putative glycosyltransferase, HDEL), SP_CVS: VTA-SV; Sebve1_10658 (hypothetical protein, HDEL), SP_CVS: SLA-LR; Sebve1_11628 (putative α-glucosidase, KDEL), SP_CVS: TTS-AA. Randomly selected proteins for which ER retention was experimentally verified do also not contain a proline at the +2 positions after the SP_CVS. This includes β-glucosidases like PYK10 in *A. thaliana* (KDEL, SP_CVS: ANA-DG) (Matsushima et al., 2003) and protein disulfide isomerase like one in *Triticum durum* (KDEL, SP_CVS: ARA-EE) (Ciaffi et al., 2001). It is therefore likely that the RSIDELD motif, although similar to classical ER retention signals, is not involved in such a process and the proteins are therefore targeted to the extracellular environment. By using the MotifSeeker program (Chapter 4.7.1.8) on small secreted proteins, the RSIDELD motif could be identified in 25 *P. indica* proteins, but only occasionally in the effectomes of other analyzed fungi (Figure 2.19). Two leucine residues have been found to be enriched at the C-terminus of the putative effectors. These have been implicated in the regulation of secretion of a mannose-6-phosphate receptor and the lutropin hormone in mammals (Jablonka-Shariff and Boime, 2011; Tikkanen et al., 2000).

Effectors found in filamentous fungi and in particular in the oomycete *Phytophthora infestans* were shown to be located in gene-poor regions and in association with transposable elements (Raffaele et al., 2010; Sacristán et al., 2009). To test this observation for the DELD family, 5' and 3' intergenic spaces of all DELD proteins were correlated to the identified transposable elements in the genome of *P. indica* (Figure 2.3 and chapter 2.1.2). Consistent with the

literature, all DELD proteins in *P. indica* had a 5' and 3' intergenic of more than 1 kb and were associated with transposable elements in a proximity of less than 2 kb.

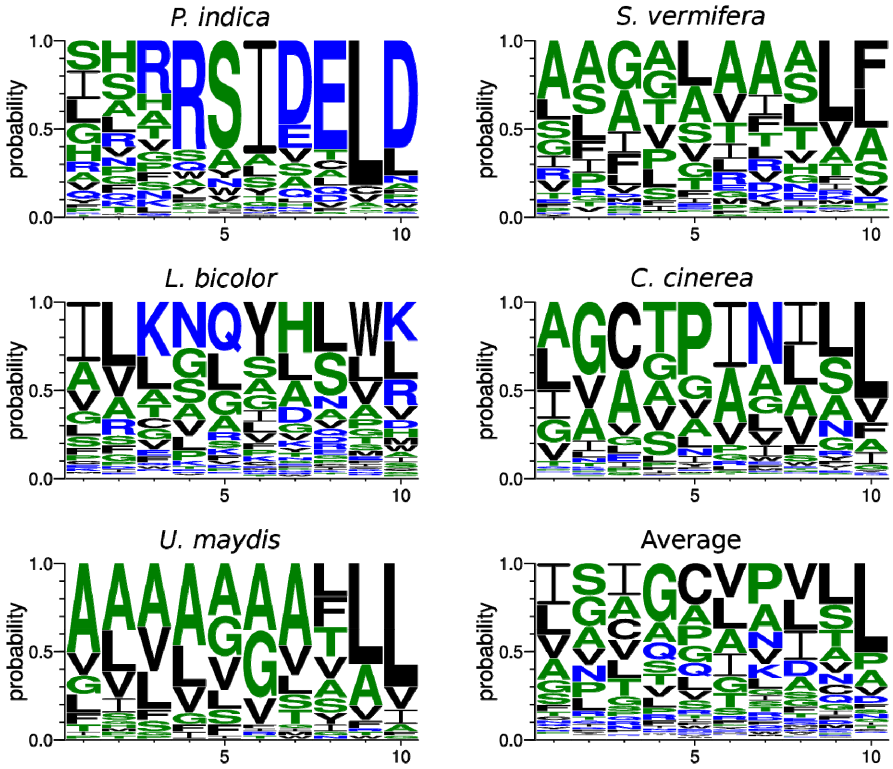


Figure 2.19: Motif enrichment in the effectomes of analyzed fungi. The program MotifSeeker was used to identify motifs within the last 10 aa in the effectomes of the shown fungi. The average includes all remaining 55 effectomes of fungi listed in Table 5.17. Sequence logos from the program output were generated with the program WebLogo 3 (Crooks et al., 2004), whereas bigger letters and a higher position in the plot refer to a more frequent occurrence of the amino acid at the respective position.

2.2. Transcriptional regulations in *P. indica* and *S. vermifera* during host colonization

For the identification of genes and metabolic pathways required during plant colonization of *P. indica* a global transcriptional profiling was performed by using microarray experiments. Because most of the functional experiments and cytological studies that have been performed so far with *P. indica* focus on barley and *Arabidopsis thaliana*, these plant hosts were also used in this study in order to allow a comparison, verification and expansion of the present knowledge of the colonization strategies and mechanisms of *P. indica*. The obtained data are complemented by an analysis of transcriptional regulations of *A. thaliana* genes due to *P. indica* colonization, in particular genes involved in defence reactions and stress resistance.

RNAseq data from *S. vermifera* colonized *A. thaliana* roots were additionally used to find differing and conserved colonization mechanisms in the Sebacinales.

2.2.1. Transcriptional profiling of *P. indica* during barley and Arabidopsis colonization

2.2.1.1. Barley microarray design and evaluation

For gene expression analysis of *P. indica* colonizing living and autoclaved barley roots, 2x105k Agilent custom microarrays were designed with up to 6 probes for each *P. indica* gene (Chapter 4.5.9). This strategy was thought to be useful for minimizing false positive and false negative signals due to cross hybridization or incorrect gene predictions. Using different JAVA programs (Chapters 4.7.1.7 and 4.7.1.13) to evaluate the probe performance, the probe position was identified as the major factor influencing signal intensity (Figure 2.20).

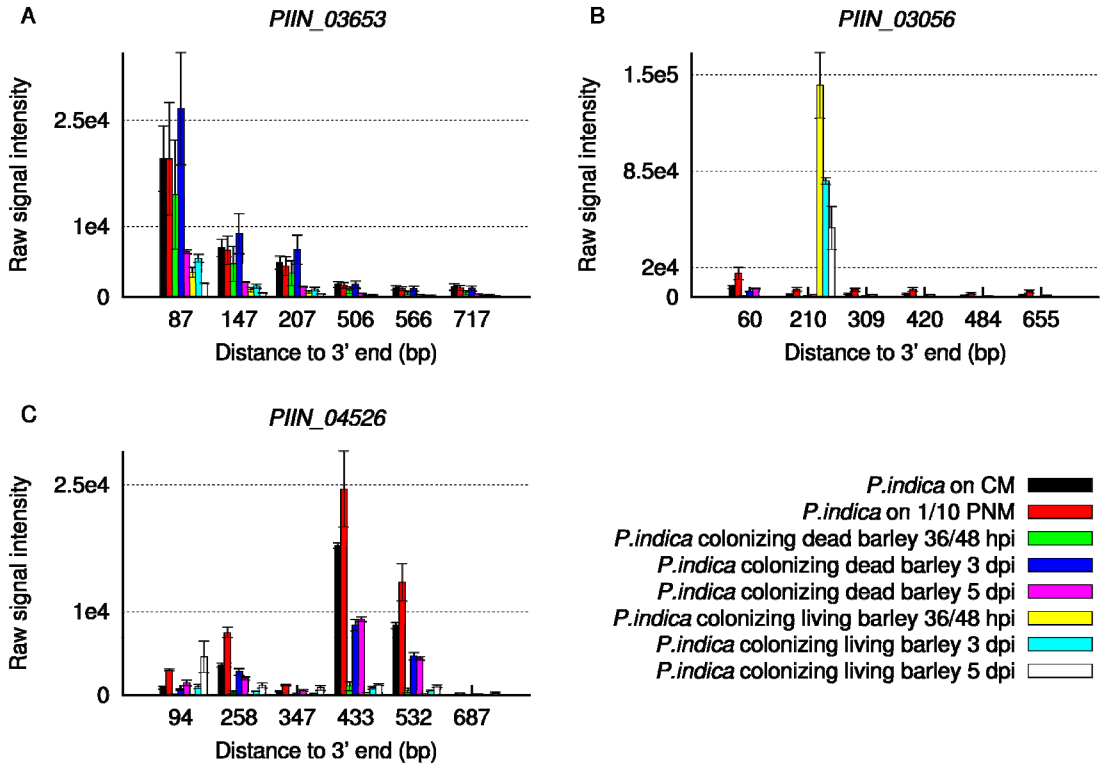


Figure 2.20: Influences of probe position and cross-hybridization on the signal intensities of microarrays. Raw, unnormalized signal intensity values of probe signals were assigned to their respective position within the gene. Averaged signal intensities and standard errors are calculated from two independent biological replicates. Three genes were selected as representatives for three cases: A) A decrease in signal intensities from the 3' to the 5' end (*PIIN_03653*), B) a strong signal at one specific probe (which is not the closest to the 3' end) while signal intensities from other probes are strongly reduced (*PIIN_03056*) and C) a signal as a result of putative cross-hybridization (*PIIN_04526*). Bar charts were created using gnuplot (Williams and Kelley, 2012).

The majority of all genes had an expression profile as shown for *PIIN_03653* (Figure 2.20 A) where the signal intensity decreased in a hyperbolic model with increasing distance from the 3' end of the gene. For *PIIN_03056* (Figure 2.20 B) only one probe showed a signal, while others do not. No putative cross-hybridization target was predicted for this probe. It is therefore possible, that an intron at the end of the gene was not correctly predicted leading to an earlier stop codon. The expression profile shown for *PIIN_04526* (Figure 2.20 C) showed signals only for probes for which putative cross-hybridization targets were predicted. The respective target was in this case strongly expressed at the same conditions and the expression of *PIIN_04526* therefore marked as false positive and ignored in downstream analyses. All genes with a predicted cross-hybridization that were included in the described analysis were manually verified in order to minimize the effect of artefacts. Although this included less than 3% of all genes, these results underline the importance of a second verification of microarray results through qRT-PCR. Furthermore, it indicates that an initial chip design using multiple probes per gene can aid in finding a probe that works best for each gene.

To study the differences of global gene expression in *P. indica* during its saprotrophic growth on dead plant material and its biotrophic interaction with living plant roots, cDNA microarray experiments were performed with *P. indica* colonizing dead (autoclaved) or living barley roots under sterile conditions. Barley seedlings were germinated for 3 days on filter paper, afterwards transferred on sugar-free plant minimal and then inoculated with $2.5x10^7$ *P. indica* chlamydospores (Chapters 4.4.2 and 4.4.4). Fungal mycelium and inoculated roots were harvested 1.5, 2, 3 and 5 dpi and RNA extracted (Chapter 4.5.2). RNA from the 36 hpi and 48 hpi samples were mixed in equal amounts and are further referred to as pre-penetration timepoint, in contrast to 3 dpi (early colonization) and 5 dpi (late colonization). As a control, *P. indica* was grown on complete medium (CM) and RNA from 1.5, 2, 3 and 5 dpi was also mixed in equal amounts. Two independent biological replications of the 2 x 3 colonization samples and the control were hybridized to 2x105k custom-designed Agilent microarrays and differential regulations calculated using R (Chapters 4.5.9.3 and 4.5.9.4). 13 fungal genes were chosen to verify the obtained microarray results via qRT-PCR (Table 2.5). Fold changes expression of fungal genes were calculated using the 2^{-ddCt} method (Livak and Schmittgen, 2001) with the *P. indica* transcription elongation factor alpha (*PiTEF*, *PIIN_03008*) as reference gene and cDNA from *P. indica* grown on CM as control condition. With the exception of two genes (*PIIN_02172* and *PIIN_08724*), all tested genes with a significant log2 expression change (adj. $p < 0.05$; log2EC >= 1) in the performed microarray experiments showed an induction comparable to values obtained from qRT-PCR experiments.

Table 2.5: Log2 expression changes of selected genes in performed *P. indica* - barley microarrays and qRT-PCR verification.

Gene ID	Gene Description	*P. indica – H. vulgare* 36/48 hpi		*P. indica – H. vulgare* 3 dpi		*P. indica – H. vulgare* 5 dpi	
		Micro array[1]	qRT-PCR[1]	Micro array[1]	qRT-PCR[1]	Micro array[1]	qRT-PCR[1]
PIIN_00867	Lectin-like LysM protein	3.1 -	2.5 -	- -	- -	- -	- -
PIIN_02169	Chitin deacetylase	-2.3 -3	-2 -3.8	-2.5 -2.2	-3.9 -5.3	-3.1 -1.6	-6.1 -4.4
PIIN_02172	Chitin deacetylase	-1.5 -2.1	1.9 1.7	-1.3 -	1 -	-1.5 -	-0.9 2.3
PIIN_02952	Metallo-peptidase	5.2 4	8.5 6	5.6 6.4	11.1 10.8	5.5 6.1	10.1 11.2
PIIN_03737	Lectin-like CBM1 protein	1.1 1.8	1.2 -1.3	- 1.2	- 1.5	- -	- -
PIIN_05825	Lectin-like WSC protein	1.3 -	2.8 -	1.7 1.9	1.4 2.1	1.6 -	1.3 -
PIIN_06067	Lectin-like CBM1 protein	2.7 -	1.8 -	- -	- -	- -	- -
PIIN_06665	Lectin-like WSC protein	6.1 -	1.5 -	4.5 -	1.6 -	3.8 -	1.7 1.4
PIIN_08721	Lectin-like LysM protein	- -2.6	- -2.1	- -1.4	- -2.3	- -	- -
PIIN_08724	Lectin-like LysM protein	- -	-3.5 -3.8	- -	-1.2 -3	- -	- -
PIIN_09643	RSIDELD protein	- -	1.1 -	2.4 2.6	4.9 3.3	2.9 3.9	6.4 7.4
PIIN_09687	RSIDELD protein	5.4 -	2.3 -	4 -	1.7 -	3.5 -	1.9 1.4
PIIN_11595	RSIDELD protein	4.2 -	1.2 1	3 -	3.6 2.5	2.5 -	2.5 2.9

[1]values in the upper row refer to data obtained from *P. indica* colonizing living barley roots while values in the lower row are from *P. indica* colonizing autoclaved barley roots.

A direct comparison of genes differentially expressed during saprotrophic growth of *P. indica* resulted in a total number of 1617 (13.7% of all genes) genes, 1176 stronger expressed in living roots and 441 stronger expressed in dead roots (Figure 2.21 A and B). These difference were expected as it is likely that more proteins are needed to interact with the living host than

to feed on decaying substances, especially under sterile conditions where no other nutritional competitors are present. The expression of the symbiotic gene set strongly differs at the 3 timepoints with little overlap (Figure 2.21 A).

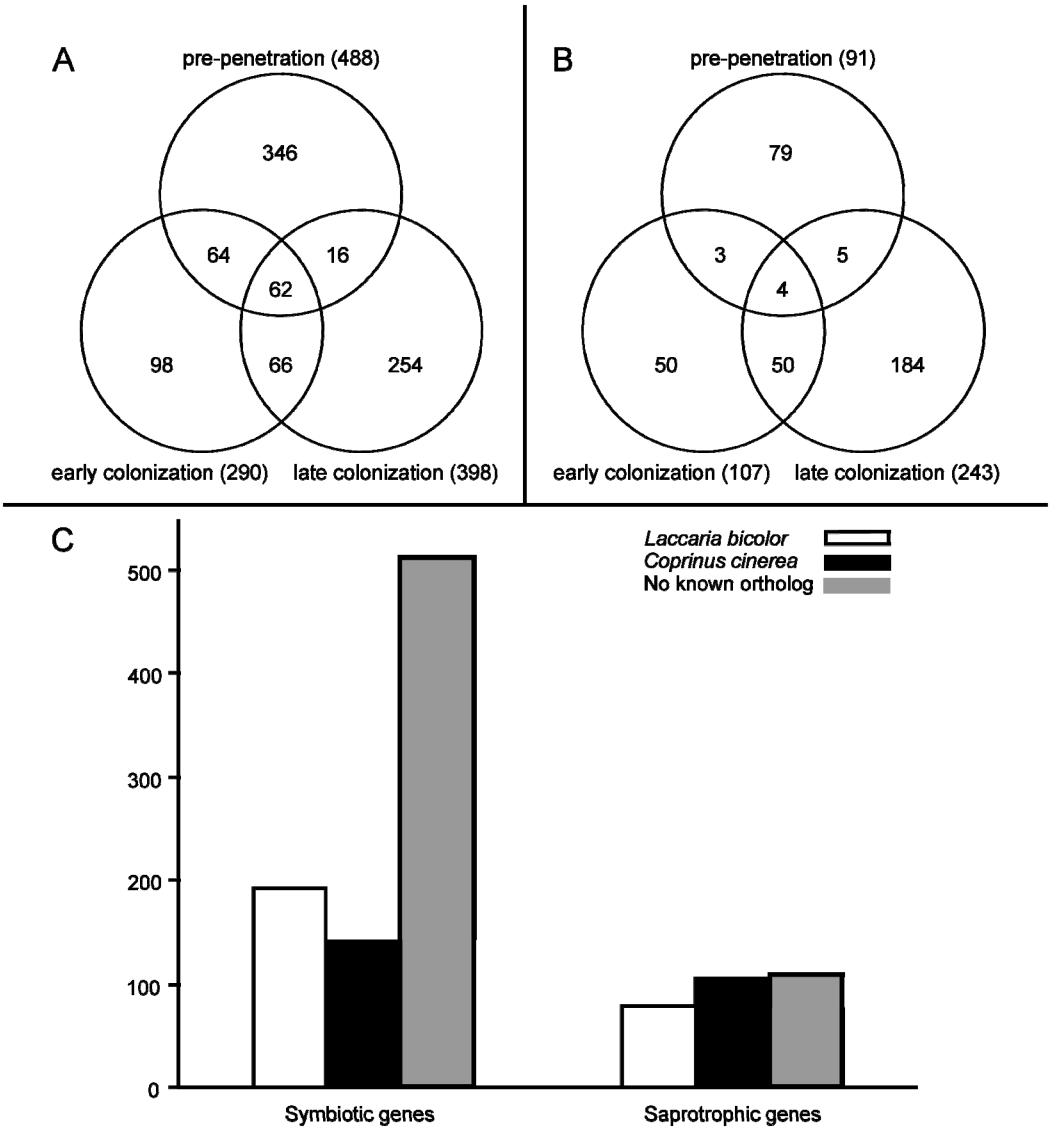


Figure 2.21: Symbiotic and saprotrophic *P. indica* genes induced upon barley root colonization. Venn diagrams showing *P. indica* genes significantly induced during colonization of living barley roots compared to colonization of autoclaved roots (A) and vice versa (B). Numbers in brackets indicate the number of all genes induced at the respective timepoint, numbers in overlapping circles indicate the number of genes induced at all involved timepoints. The bar chart (C) is showing the number of symbiotic and saprotrophic genes with top blast hits for genes from *L. bicolor* or *C. cinerea*. No known ortholog is referring to genes with no Blast hit against the NCBI nr-database (adapted according to (Zuccaro et al., 2011)).

This is in agreement with cytological data showing that at the penetration phase *P. indica* is attempting to penetrate barley cells. At this stage, the fungus needs to overcome physical plant barriers and interact with the innate plant immune system. At 3 dpi (early colonization), a biotrophic interaction is established, which then switches already at 5 dpi (late colonization)

to a saprotrophic nutrition. During saprotrophic feeding on autoclaved roots, *P. indica* does not need to overcome the immune system of the host and this may explain the lower number of expressed genes (Figure 2.21 B). Blastx comparisons against the NCBI nr-database (e-Val $< 10^{-3}$) showed further that symbiotic genes show a higher similarity to the ectomycorrhizal symbiont *L. bicolor* than to the soil decomposer *C. cinerea* (Figure 2.21 C). Furthermore, a larger number of genes does not show any significant similarity to a sequence in the nr-database, indicating that the interaction of *P. indica* with its host is defined mainly by so far uncharacterized proteins. Consistently, saprotrophic genes show a higher similarity to *C. cinerea* genes with little *P. indica* specific genes, emphasizing the existence of transcriptionally defined gene sets for biotrophic and saprotrophic lifestyles.

2.2.1.2. Arabidopsis microarray design and evaluation

Based on a the previously discussed microarray design, an optimized version was created for experiments with *P. indica* colonizing Arabidopsis as described in the chapter 4.5.9. With the described workflow, one single best probes (BP) was found for ~93% of all *P. indica* genes and used for comparative expression analyses between barley and Arabidopsis. Additional to probes specific to *P. indica* genes, probes for Arabidopsis genes were included in this design. Plant and fungal probes were the separated prior to normalization. To exclude biases in downstream analyses, a putative cross-hybridization of fungal derived cRNA with plant probes and vice versa was analyzed. Raw signal intensities were evaluated for the *P. indica* probes on arrays hybridized with *A. thaliana* derived cRNA (Figure 2.22, left graph) and for *A. thaliana* probes on arrays hybridized with *P. indica* derived cRNA (Figure 2.22, right graph). Both probe sets showed no cross-hybridization with few exceptions. *P. indica* genes for which a strong signal could be observed include two ubiquitin-related proteins with 77% (PIIN_04265) and 73% (PIIN_01523) DNA sequence similarity with their homologues in *A. thaliana*. Furthermore an ADP/ATP carrier protein with 75% similarity and two histon-related proteins (PIIN_05985 and PIIN_04354) with 79% to 85% similarity gave a weak signal. *A. thaliana* genes with a strong signal due to cross-hybridization from *P. indica* cRNA include conserved proteins, like polyubiquitin3 (AT5G03240), ubiquitin 13 (AT1G65350) and a tubulin family protein (AT4G14960).

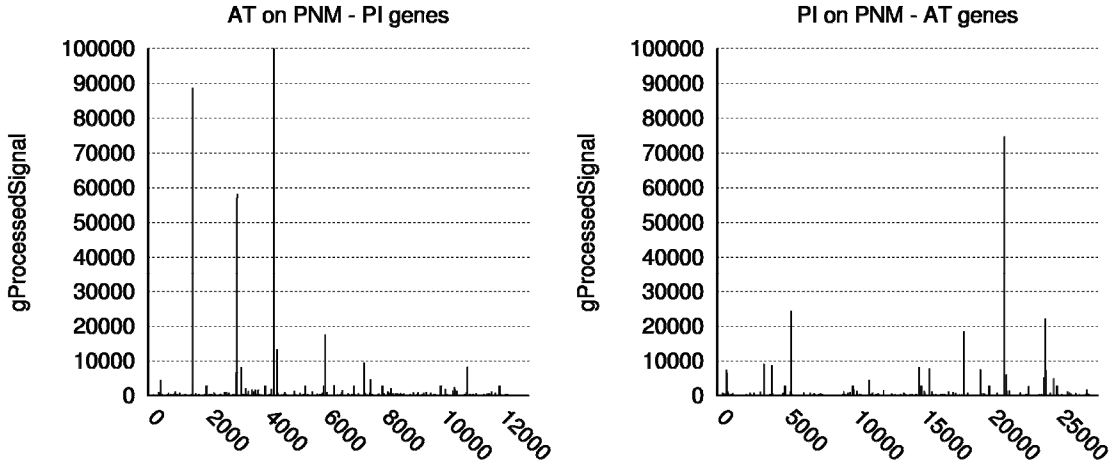


Figure 2.22: Raw signal intensities of negative control arrays. Left chart: Raw, unnormalized signals from *P. indica* specific probes on chips hybridized with *A. thaliana* cRNA only. Right chart: Raw, unnormalized signals from *A. thaliana* specific probes on chips hybridized with *P. indica* cRNA only. Charts were generated using gnuplot (Williams and Kelley, 2012).

After verification that plant and fungal probes are only weakly influenced by cross-hybridization between both species, suitability of the performed normalization method as well as variation between the performed experiments and replications was assessed by principle component analysis (PCA) (Hotelling, 1933) (Figure 2.23 and Figure 2.24).

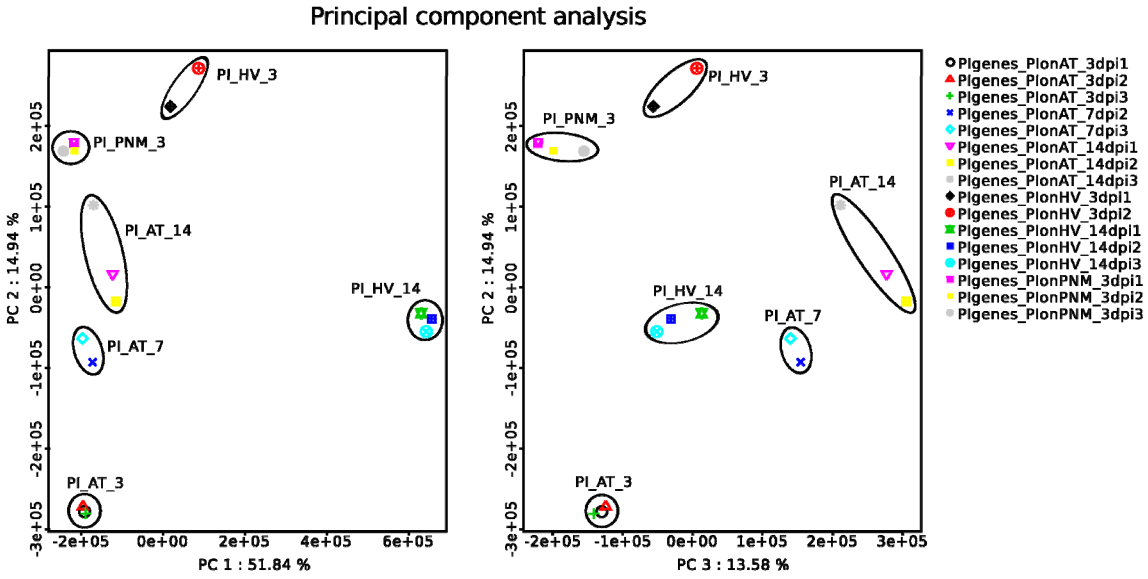


Figure 2.23: Principle component analysis of *P. indica* genes. Overall variability between individual microarray hybridizations were calculated by principle components analysis and graphs of the first three principle components were created using R (R Development Core Team, 2011). The analysis included *P. indica* genes (PIgenes) from the following experiments: *P. indica* colonizing *A. thaliana* roots 3 (PIonAT_3), 7 (PIonAT_7) and 14 dpi (PIonAT_14), *P. indica* colonizing *H. vulgare* roots 3 (PIonHV_3) and 14 dpi (PIonHV_14) and *P. indica* growing on minimal medium 3 dpt (PIonPNM_3).

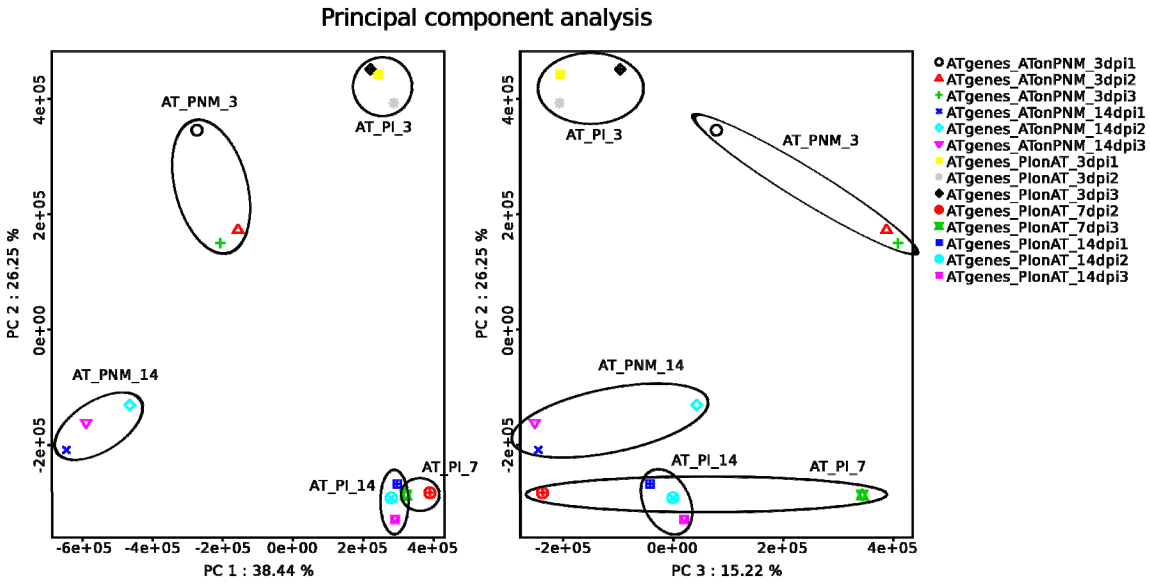


Figure 2.24: Principle component analysis of *A. thaliana* genes. Overall variability between individual microarray hybridizations were calculated by principle components analysis and graphs of the first three principle components were created using R (R Development Core Team, 2011). The analysis included *A. thaliana* genes (ATgenes) from the following experiments: Mock treated *A. thaliana* roots growing on minimal medium 3 (ATonPNM_3) and 14 dpt (ATonPNM_14) and *P. indica* colonizing *A. thaliana* roots 3 (PIonAT_3), 7 (PIonAT_7) and 14 dpi (PIonAT_14).

PCA is a method to calculate the overall variability within large datasets. The original dataset is thereby expressed in so called principle components which have no covariations and eigenvalues in decreasing order. For *P. indica* genes, ~80% of the overall variability between genes and experiments could be explained by the first three principle components (PC). The analysis shows that signal intensities from biological replications are highly similar (Figure 2.23). Colonization of barley roots by *P. indica* 14 dpi (PI_HV_14) shows thereby to be more different to the other experiments than these are among each other.

Also for *A. thaliana* genes, an overall variability of ~80% could be explained by the first three PC's. Opposite to *P. indica* genes, a high variability between biological replications could be observed within the third PC (Figure 2.24, right graph). Similar variations could not be observed for *P. indica*. It can therefore be excluded that *A. thaliana* roots were unequally colonized by *P. indica* in the biological replications as well as that other technical problems occurred during the hybridization. Furthermore, such differences could not be observed in any other PC nor in other performed tests and gene expression results from qRT-PCR experiments strongly support the correctness of the applied statistical methods.

Nine fungal and three plant genes were chosen to verify the obtained microarray results via qRT-PCR (Table 2.6). Fold changes expression of fungal genes were calculated using the 2^{-ddCt} method (Livak and Schmittgen, 2001) with the *P. indica* transcription elongation factor

alpha (*PiTEF*, PIIN_03008) as reference gene and cDNA from *P. indica* grown for 3 days on 1/10 PNM as control condition. For plant genes, *A. thaliana* ubiquitin 5 (*AtUBI*, AT3G62250) was used as reference. The control conditions in qRT-PCR experiments were always cDNA from respective uncolonized root material. For microarray experiments, the 3 and 14 dpi timepoints were also compared to data from respective uncolonized material while 7 dpi were compared against both control conditions individually (Table 2.6). All tested genes with a significant log2 expression change (adj. $p < 0.05$; log2EC >= 2) predicted in the performed microarray experiments showed an overall high similarity with values obtained from qRT-PCR experiments.

Table 2.6: Log2 expression changes of selected genes in performed *P. indica* - Arabidopsis microarrays and qRT-PCR verification

		P. indica – A. thaliana 3 dpi		*P. indica – A. thaliana* 7 dpi		*P. indica – A. thaliana* 14 dpi	
Gene ID[1]	Gene Description	Micro array	qRT-PCR	Micro array	qRT-PCR	Micro array	qRT-PCR
PIIN_02036	Ammonium transporter 1	2.11	1.63	3.17	2.26	1.61	1.43
PIIN_04582	Putative signalling protein	1.62	1	6.27	6.32	8.16	8.22
PIIN_05015	Hypothetical protein	-6.02	-6.14	-4.95	-5.48	-5.44	-5.56
PIIN_05825	Lectin-like WSC protein	-	-	2.19	1.98	1.45	1.87
PIIN_05872	RSIDELD protein	-	-	-	-	-	-1.14
PIIN_06069	Lectin-like CBM1 protein	5.23	6.01	3.65	4.23	2.44	2.5
PIIN_07534	Tryptophan transaminase 1	-	-	-	-	-	-
PIIN_08721	Lectin-like LysM protein	-	-	1.46	1.77	3.01	2.93
PIIN_09643	RSIDELD protein	-	-	4.94	6.07	4.39	5.23
AT2G14610	Pathogenesis-related gene 1	-	-	4.43[2]	6.59	1	-
AT5G01900	WRKY transcription factor 62	-	-	2.56[2]	3.39	3.69	4.29
AT5G22570	WRKY transcription factor 38	-	-	2.24[2]	2.49	3.16	3.56

[1]Gene IDs starting with PIIN are from *P. indica*, starting with AT from *A. thaliana*

[2]Values are averages of the comparison 7 dpi vs. 3 days control and 7 dpi vs. 14 days control

In total were 2069 genes transcriptionally induced in *P. indica* during colonization of Arabidopsis roots in comparison to growth on minimal medium and in at least one of the analyzed timepoints (Figure 2.25, A). Like for barley shown before (Figure 2.21 A), the amount of genes stronger expressed during colonization is highest in the beginning of the interaction and is decreasing over the time. Additionally, half of genes induced at the 3 dpi timepoint were not induced at one of the other analyzed timepoints while only ~1/10 of the genes at 7 dpi and ~1/3 of the genes at 14 dpi were induced in such a timepoint-specific manner. About 15% (310) of the genes are induced at all timepoints and similar numbers of genes are commonly induced at 3 and 7 dpi and at 7 and 14 dpi (Figure 2.25, A). The timepoints 3 and 14 dpi were then chosen to analyze similarities and differences in the number and type of genes induced during colonization of either Arabidopsis or barley roots (Figure 2.25, B).

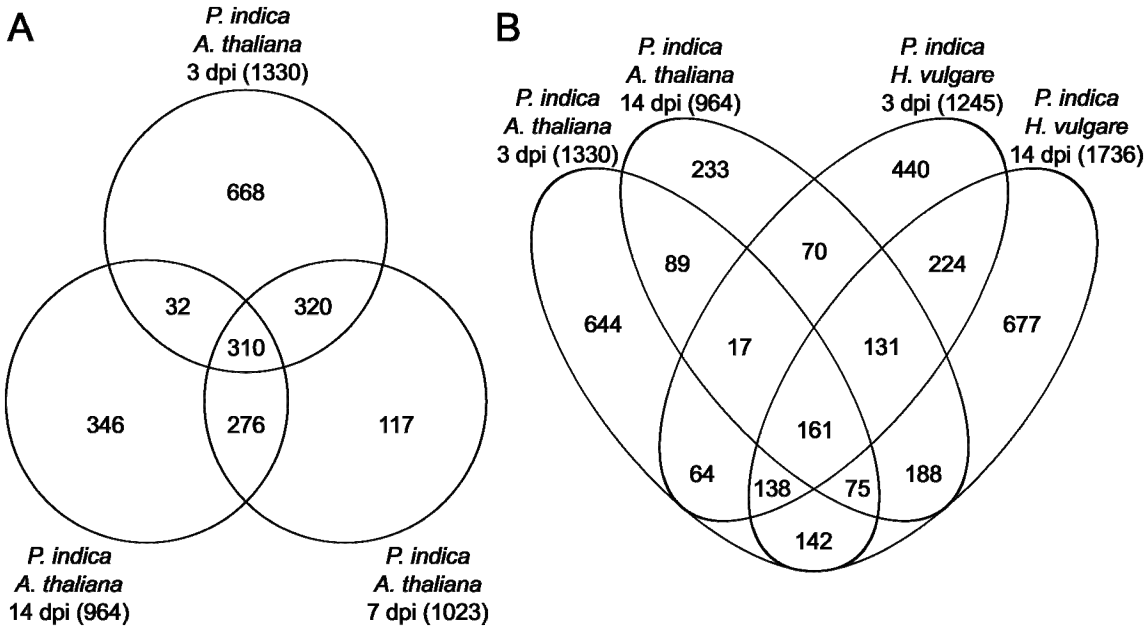
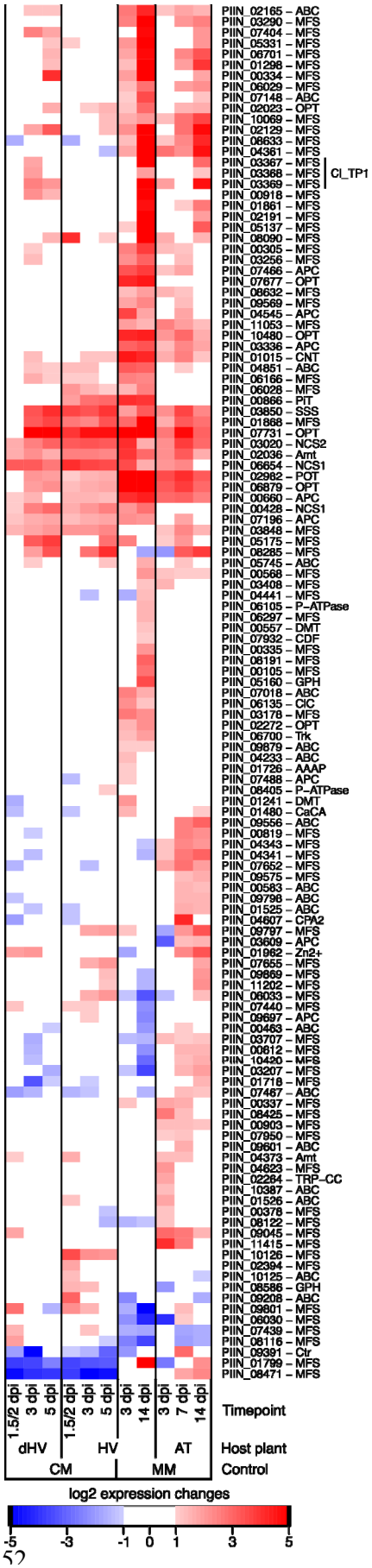


Figure 2.25: Transcriptionally induced genes in *P. indica* at different timepoints during colonization of Arabidopsis and barley roots. A) Venn diagram showing similarities and differences of *P. indica* genes significantly induced during colonization of Arabidopsis roots 3, 7 and 14 dpi compared to fungal growth on a minimal medium control. B) Venn diagram showing similarities and differences of *P. indica* genes significantly induced during colonization of either Arabidopsis roots 3 and 14 dpi or barley roots 3 and 14 dpi. Numbers in brackets indicate the number of all genes induced at the respective timepoint, numbers in overlapping shapes indicate the number of genes induced at all involved timepoints.

While the total numbers of genes induced during colonization of Arabidopsis (1952) and barley (2327) are similar, strong differences can be found for the late colonization timepoints (14 dpi) with almost twice as many genes induced in barley than in Arabidopsis (Figure 2.25, B). The number of genes induced only during early Arabidopsis colonization remained high



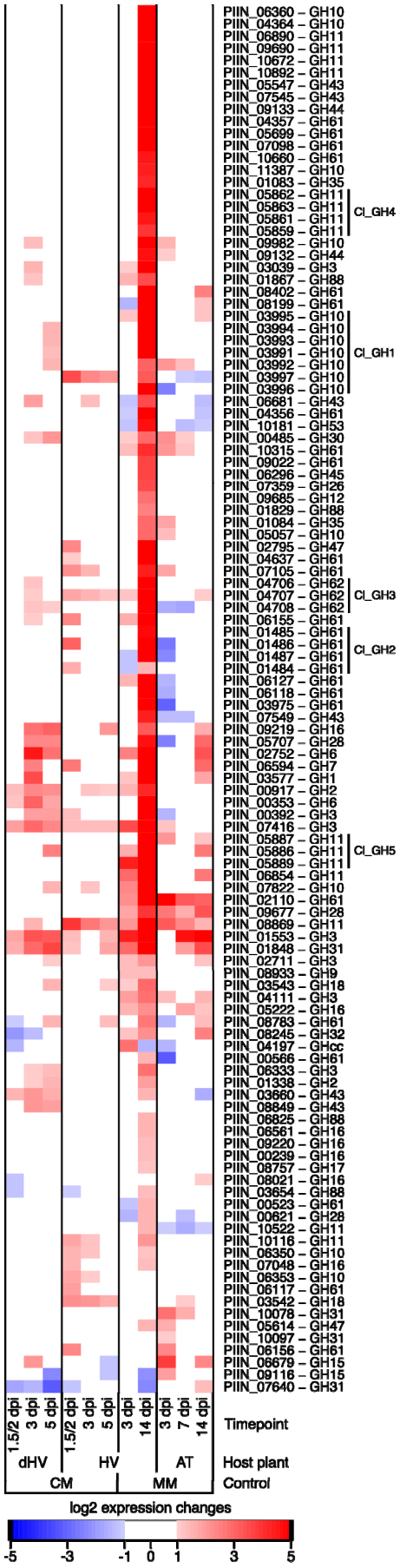
in both comparison (Figure 2.25, 668 vs. 644) which reveals not only a time- but probably also host-specific requirement for some of these genes. A more detailed analyses of the individual regulation patterns of the previously identified expanded gene families was therefore conducted.

2.2.1.3. Transcriptional upregulation of selected functional groups *in planta*

Transporter

Within the performed microarray experiments 127 of the 357 proteins (~36%) predicted to be involved in transmembrane transport were identified as transcriptionally upregulated *in planta* compared to the control conditions (Figure 2.26). Following Blast comparisons against the Transporter Classification Database (TCDB) and the non-redundant database of the NCBI, expression patterns could be assigned to different transporter types. Putative nitrogen and amino acids transporters of the APC, AAAP, Amt, MFS and OPT families were predominately induced in both host plants. These include different permeases (urea-, uracil-, purine-, amino acid-), three peptide transporters and a high-affinity ammonium transporter (PIIN_02036) which was shown to be induced under nitrogen starvation (Lahrmann et al., 2013). Most putative carbohydrate transporters belonging to the GPH and MFS families were predominately induced at 14 dpi of

Figure 2.26: Upregulated transmembrane transporters in *P. indica* during colonization of barley (HV), dead barley (dHV) and Arabidopsis (AT) roots. Shown are log2 expression changes calculated against gene expression during growth on CM or PNM. Expression changes are capped at a maximum induction of 2^5 for visualization purposes. Gene expressions are hierarchically clustered using Ward's method (Everitt et al., 2001). Modified accoring to (Lahrmann et al., 2013; Zuccaro et al., 2011).

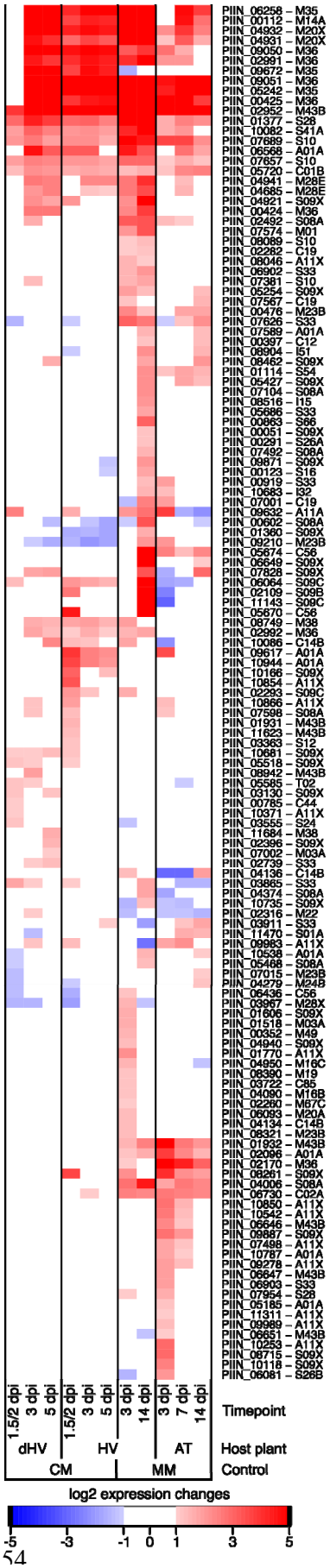


barley roots and during growth on autoclaved barley roots. These transporters were annotated as general sugar transporters, in particular of glucose, hexose and lactose. Three of the putative lactose transporters cluster physically on the genome of *P. indica* (Figure 2.26, Cl_TP1). Opposite to the carbohydrate transporters, multi-drug resistance transporters and toxin-efflux carrier proteins of the ABC, DMT and MFS families were more often induced at early timepoints in barley and at all timepoints in Arabidopsis. Transporters with a less specific expression pattern were classified as cation transporters, channels and antiporters of the CaCA, CDF, CIC, Ctr, Trk, MFS, P-ATPase, TRP and ZIP families.

Glycosyl hydrolases

153 proteins in *P. indica* contain at least one domain of the GH family and were therefore considered to be involved in the hydrolysis of glycosidic bonds. Because members of the GH61 family were shown to be involved in lignocellulose decomposition and an substantial enhancement of cellulose hydrolysis, they are also included here although they have been re-classified as copper-dependent monooxygenases (Harris et al., 2010; Phillips et al., 2011; Žifčáková and Baldrian, 2012). 114 of the 153 proteins (~75%) were identified as transcriptionally upregulated *in*

Figure 2.27: Upregulated glycosyl hydrolases in *P. indica* during colonization of barley (HV), dead barley (dHV) and Arabidopsis (AT) roots. Shown are log2 expression changes calculated against gene expression during growth on CM or PNM. Expression changes are capped at a maximum induction of 2^5 for visualization purposes. Gene expressions are hierarchically clustered using Ward's method (Everitt et al., 2001). Modified accoring to (Lahrmann et al., 2013; Zuccaro et al., 2011).



planta compared to the control conditions (Figure 2.27). The majority of GH's were strongly induced at 14 dpi in barley and belong predominately to the GH10, GH11 (xylanases, (Enkerli et al., 1999)) and GH61 families. Five clusters of glycosyl hydrolases could be identified in the genome of *P. indica* containing at least three adjacent glycosyl hydrolases. Members of these clusters were shown to transcriptionally regulated in a similar way (Figure 2.27, Cl_GH1 - Cl_GH5).

Peptidases

137 of the 354 (~39%) predicted peptidases were identified as transcriptionally upregulated *in planta* compared to the control conditions (Figure 2.28). Genes encoding putative peptidase could be separated into three distinct groups. The first group contains metallopeptidases, in particular M35 (deuterolysin) and M36 (fungalysin), that were strongly upregulated at all timepoints except early saprotrophic growth. These peptidases closely match the M36 peptidase family in *C. cinerea* (Lilly et al., 2008). Metalloproteases belonging to other families were induced in a more host-specific manner and more strongly at earlier timepoints. The second group contains serine proteases of the S09 family (prolyl oligopeptidases) which were primarily induced at 14 dpi in barley and during growth on autoclaved barley roots. The third group, which is predominately induced at early timepoints in Arabidopsis and barley, contains aspartic peptidases of the A01 (pepsin) and A11 (Copia transposon peptidase) family. Metalloproteases and serine proteases were, like glycosyl hydrolases, found in clusters of three to seven genes in the genome of *P. indica*, however, genes

Figure 2.28: Upregulated peptidases in *P. indica* during colonization of barley (HV), dead barley (dHV) and Arabidopsis (AT) roots. Shown are log2 expression changes calculated against gene expression during growth on CM or PNM. Expression changes are capped at a maximum induction of 2^5 for visualization purposes. Gene expressions are hierarchically clustered using Ward's method (Everitt et al., 2001).

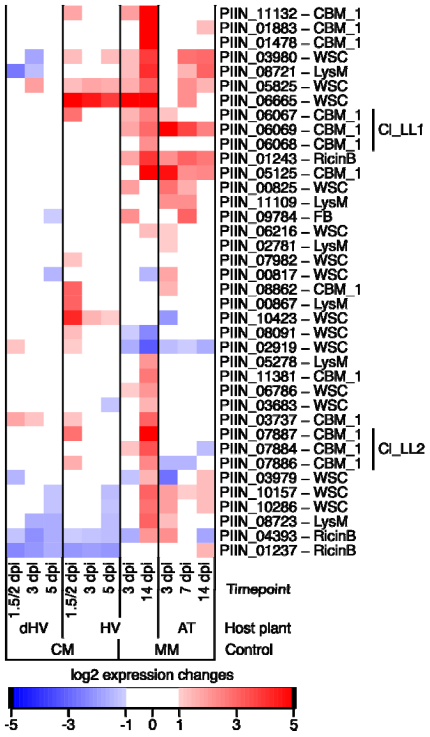


Figure 2.29: Upregulated lectin-like proteins in *P. indica* during colonization of barley (HV), dead barley (dHV) and Arabidopsis (AT) roots. Shown are log2 expression changes calculated against gene expression during growth on CM or PNM. Expression changes are capped at a maximum induction of 2^5 for visualization purposes. Gene expressions are hierarchically clustered using Ward's method (Everitt et al., 2001).

belonging to these clusters did not show a similar expression pattern. Four pairs of adjacent metalloproteases could be identified with a highly similar expression (PIIN_00424, PIIN_00425 - M36; PIIN_02991, PIIN_02992 - M36; PIIN_04931, PIIN_04932 - M20; PIIN_09050, PIIN_09051 - M36).

Lectin-like proteins

38 of the 63 (~60%) lectin-like proteins identified in *P. indica* were transcriptionally induced during colonization of living roots. These included predominately proteins involved in the putative binding of chitin (LysM), glucan (WSC) and cellulose (CBM1). Proteins from each type were induced at different timepoints during barley or Arabidopsis root colonization, but most of the CBM1 proteins were stronger induced at 14 dpi in barley roots. This correlates with the observed induction of glycosyl hydrolases and CBM1 domains are often combined with hydrolytic enzymes domains in *P. indica* and in other fungi (Cantarel et al., 2009). Because some of the CBM1 lectins are clustered in the genome of *P. indica* (Figure 2.29, Cl_LL1 and Cl_LL2) and because no putative catalytic enzyme domain could be predicted within a 1kb range around the genes, these genes are probably true lectins which are not involved in hydrolytic digestion of plant material at this timepoint. The induction of WSC and LysM containing lectins on the other hand was for most of the proteins not restricted to one host but rather induced in both hosts at similar timepoints.

Putative effector candidates

Plant-responsive putative effectors were separated into *P. indica* specific effectors (Figure 2.30, right heatmap) and effectors which have a significant (e-Val < 10^{-3}) similarity to putative effectors in the nr-database (Figure 2.30, left heatmap).

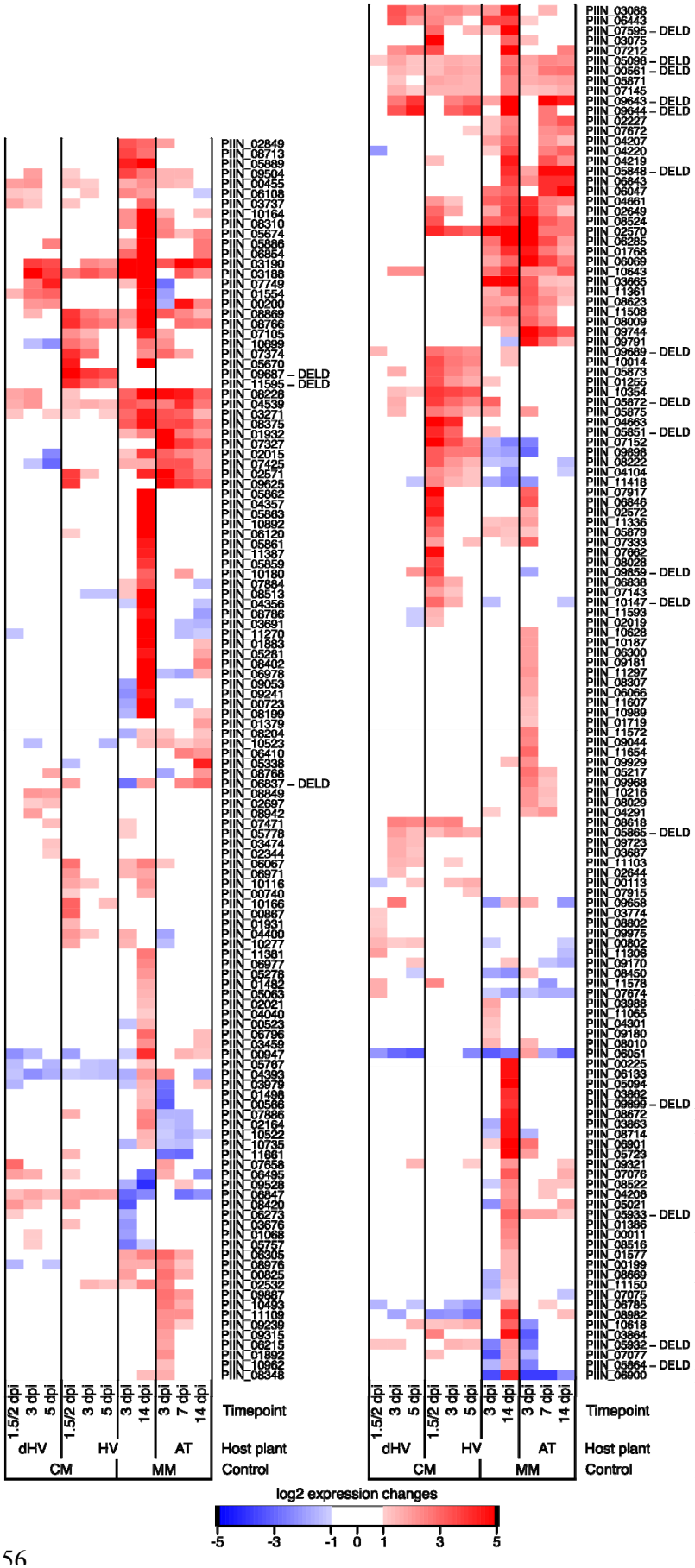


Figure 2.30: Upregulated effector candidates in *P. indica* during colonization of barley (HV), dead barley (dHV) and Arabidopsis (AT) roots. Putative effectors were separated in those which have significant similarities to other putative effectors in the nr-database (left heatmap) and those which are *P. indica* specific effectors (right heatmap). Shown are log2 expression changes calculated against gene expression during growth on CM or PNM. Expression changes are capped at a maximum induction of 2^5 for visualization purposes. Gene expressions are hierarchically clustered using Ward's method (Everitt et al., 2001).

From all 261 plant-responsive putative effectors in *P. indica*, 120 showed an induced expression in barley roots compared to the complete medium control, while 173 were induced in barley compared against the minimal medium control. Induced in Arabidopsis compared against the minimal medium control were 134 putative effectors (Figure 2.30). Most of the putative effectors were not commonly induced under all analyzed conditions but showed a time- and host specific expression. An induced expression could be shown for 19 of the 29 members of the putative novel effectors family of the DELD proteins. 12 of the 19 were thereby only induced during colonization of barley roots while 7 were additionally to barley also induced during Arabidopsis colonization (Figure 2.30, "DELD"). Effectors with putative orthologues in other fungi were predominately identified as cell wall degrading enzymes (glucanases, xylanases, lyases, monooxygenases or esterases) while no function could be predicted for *P. indica* specific effectors.

2.2.2. Conserved regulation patterns within the Sebacinales

As part of the mycorrhizal genome initiative (MGI) framework (http://mycor.nancy.inra.fr/IMGC/MycoGenomes/), the colonization of Arabidopsis by *S. vermifera* was characterised by RNA sequencing of fungal material grown in axenic culture and a mix of RNA from colonized roots 3, 7 and 14 dpi. The obtained data was compared to microarray data from *P. indica* colonizing Arabidopsis and barley in order to identify conserved genes and colonization strategies within the Sebacinales. But because expression values and deduced gene inductions/repressions from RNAseq and microarray analyses are not comparable and because a mix of different colonization timepoints was used for RNAseq of *S. vermifera* colonizing *A. thaliana*, only tendencies were compared. Therefore, a similar induction was noted if an upregulated *S. vermifera* gene has an orthologue in *P. indica* that is induced during at least one of the analysed colonization timepoints. From the 15312 *S. vermifera* genes, 2016 genes showed a significantly (adj. p-value < 0.05) stronger expression in the colonized roots than in the control, while 694 genes were significantly weaker expressed (Figure 2.31). From all *P. indica* genes transcriptionally induced upon plant colonization according the performed microarrays, 2108 were orthologues to *S. vermifera* genes and of these 1148 were induced at least at one timepoint during Arabidopsis colonization and 1601 at least at one timepoint during barley colonization (Figure 2.31, left diagram). Strong overlaps could be observed between the number of genes induced in *P. indica* during barley and Arabidopsis colonization, but only minor overlaps between *S. vermifera* and *P. indica* colonization. Surprisingly, more genes were found induced in *S.*

vermifera which have an orthologue in *P. indica* that is only induced during barley colonization than in Arabidopsis. Similar relations were also found for genes lower expressed during plant colonization than under control conditions (Figure 2.31, right diagram).

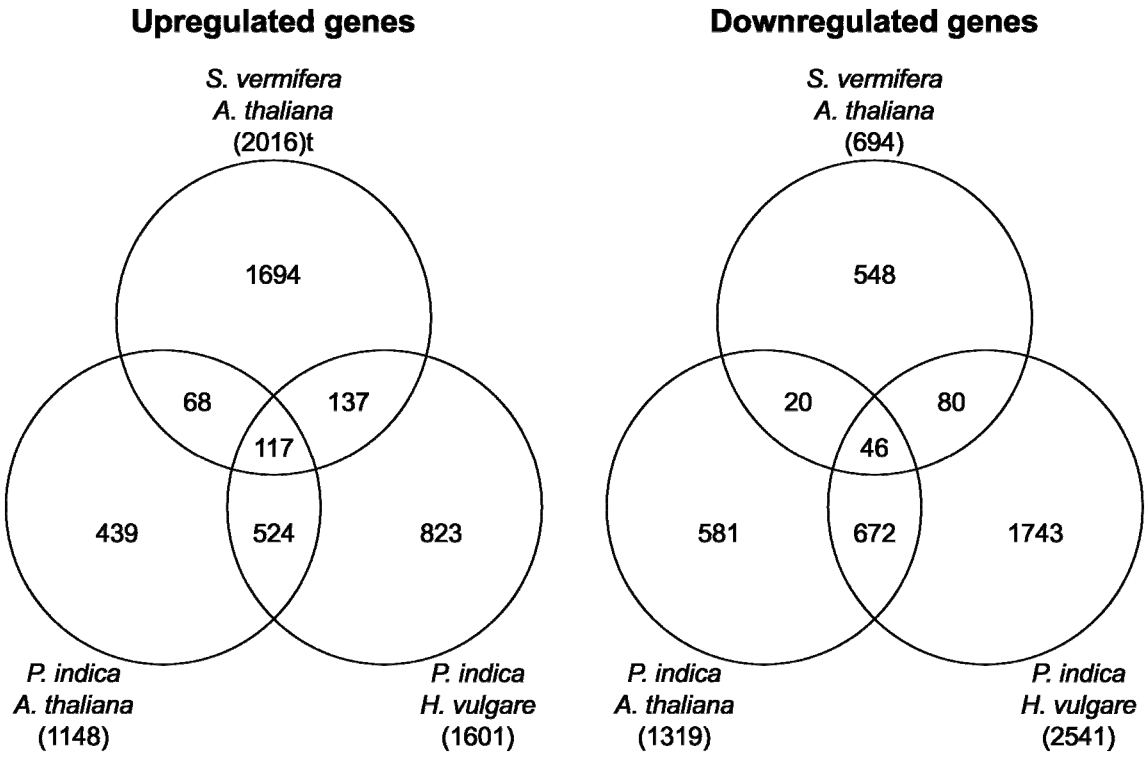


Figure 2.31: Transcriptional regulation of *S. vermifera* genes and their respective orthologues during plant colonization. Venn diagram showing similarities and differences of the *in planta* regulation of *S. vermifera* genes compared to their respective orthologue in *P. indica*. Orthologues were predicted as described in chapter 4.7.15. Numbers in brackets indicate the number of all genes induced at the respective timepoint, numbers in overlapping shapes indicate the number of genes induced at all involved timepoints.

A further comparison focused, like for *P. indica*, on genes involved in nutrient acquisition and transport (Transporters, Peptidases and Hydrolases) and on genes thought to be key determinants in the interaction with the plant host (SSPs and Lectins). For this, log2 expression changes of *S. vermifera* genes deduced from the performed RNAseq experiments were compared to the induction/repression of the respective *P. indica* orthologue during plant colonization (Figure 2.32). In the shown heatmaps, only whether a *P. indica* orthologue is induced (red) or repressed (blue) *in planta* is indicated, while blank indicate either that the respective *S. vermifera* does not have an orthologue in *P. indica* or that the orthologue is not transcriptionally regulated *in planta*. This relatively small selection of proteins include 46 of the 117 (~40%) consistently induced genes in both fungi and both hosts, underlying the importance of these genes during plant colonization.

391 transporters were identified in the genome of *S. vermifera* of which 106 were differentially regulated according to the RNAseq data. Of these 106, transcripts for 79 were at least twice as abundant in colonized roots than in the controls (adj. $p < 0.05$; FC >= 2), with the strongest induction found for transporters involved in nitrogen transport and toxin efflux (Figure 2.32, Transporter). Most of the induced genes which have on orthologue in *P. indica*, that is also transcriptionally regulated *in planta*, were induced during colonization of both hosts. A similar observation could be made for small secreted proteins (SSPs) and lectin-like proteins, but with lower numbers of orthologues identified in *P. indica* (Figure 2.32, SSPs / lectins). Slightly different from these is the induction of putative peptidases and glycosyl hydrolases which orthologues in *P. indica* were more often induced during colonization of barley roots, in particular during the late, saprotrophic colonization phase, than during colonization of Arabidopsis roots. 361 peptidases were identified in the genome of *S. vermifera* of which 84 were differentially regulated according to the RNAseq data. Of these 84, transcripts for 72 were at least twice as abundant in colonized roots than in the controls (adj. $p < 0.05$; FC >= 2) (Figure 2.32, Peptidases). About 2/3 of these proteins either did not have an ortholog in *P. indica* or this is not induced *in planta*. However, the six most strongly induced peptidases as well as two others had orthologs in *P. indica* which were shown to belong to a cluster of strongly and host independently induced metallo- and serine-peptidases (Figure 2.32, Peptidases). Noticeable is also a group of lower, but still significantly, induced metallo- and cysteine-peptidases which orthologous genes in *P. indica* are downregulated during colonization of barley roots. For the 145 proteins containing a hydrolase domain in the genome of *S. vermifera* a significantly (adj. $p < 0.05$; FC >= 2) induced expression could be observed for 27 genes in comparison to control conditions (Figure 2.32, Glycosyl hydrolases). Although the number is relatively small in comparison to *P. indica*, about 1/3 of these genes have an orthologue that is only induced during colonization of barley roots, in particular during the late, saprotrophic colonization phase (Figure 2.32, Glycosyl hydrolases).

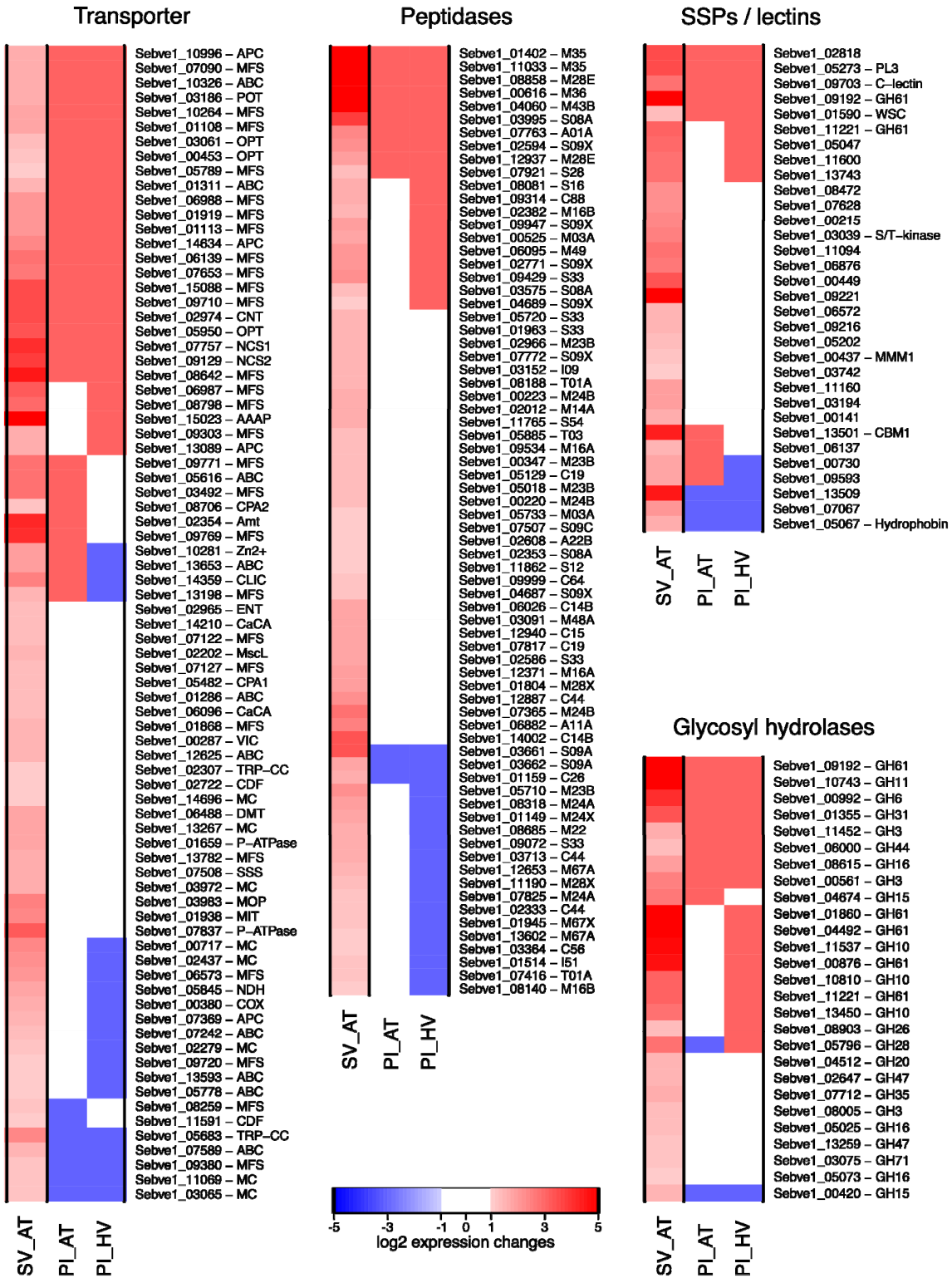


Figure 2.32: Transcriptional regulation of *S. vermifera* genes and their respective orthologues during plant colonization. Orthologues were predicted as described in chapter 4.7.15. Log2 expression changes in *S. vermifera* during colonization of Arabidopsis were calculated against growth in axenic culture (SV_AT). For orthologues in P. indica expression changes were set to 3 if the gene was induced during colonization of Arabidopsis (PI_AT) or barley (PI_HV), to -3 if it was repressed and to 0 otherwise. Expression changes are capped at a maximum induction of 2^5 for visualization purposes. Gene expressions are hierarchically clustered using Ward's method (Everitt et al., 2001).

2.2.3. Arabidopsis transcriptional responses to *P. indica* colonization

Transcriptional responses of Arabidopsis genes to the colonization by *P. indica* were calculated by comparing microarray data from *P. indica* colonized samples to data from mock treated roots. The comparison resulted for *P. indica*- vs. mock-treated samples 3 dpi in 1481 differentially (829 up- and 652 down-) regulated genes (adj. $p < 0.05$) and for 14 dpi in 3194 differentially (1689 up- and 1505 down-) regulated genes (adj. $p < 0.05$) (Figure 2.33).

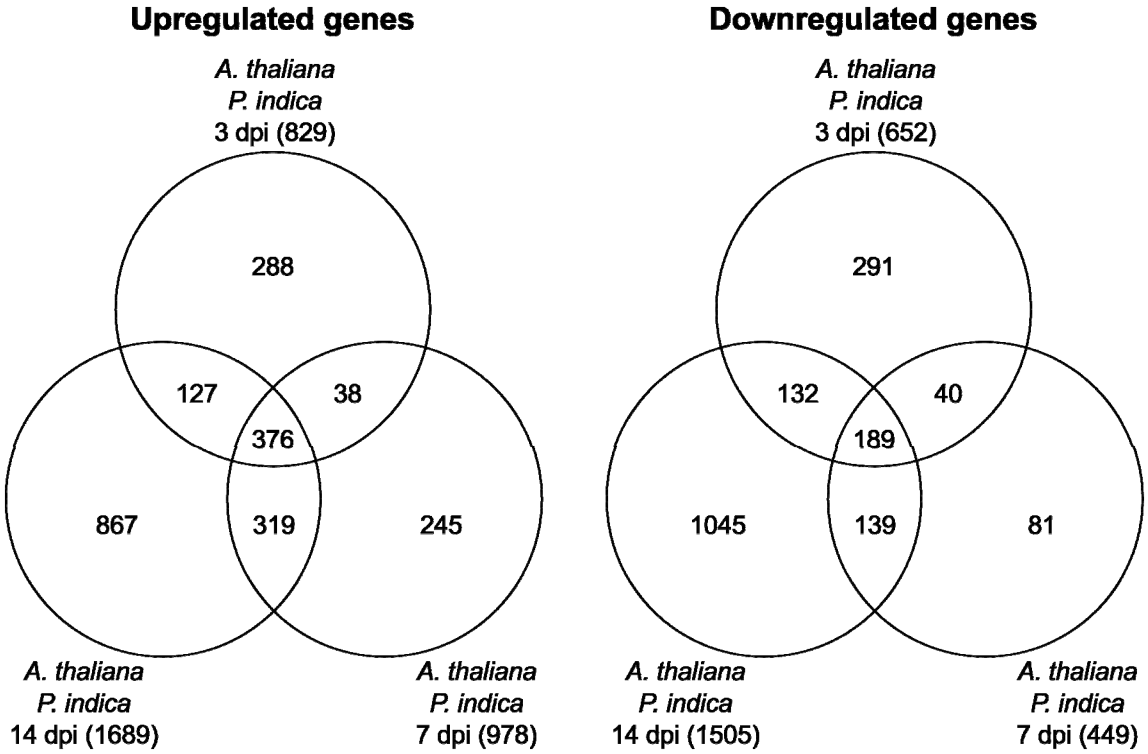


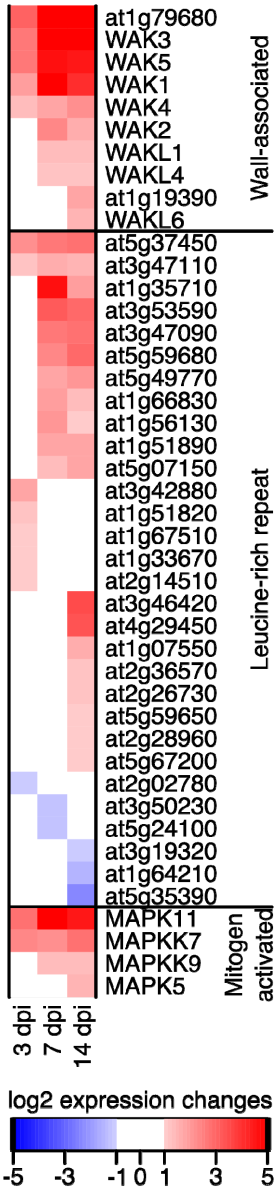
Figure 2.33: Numbers of differentially regulated Arabidopsis genes in response to *P. indica* colonization. Numbers indicate genes that are significantly up- (left graph) or down- (right graph) regulated in Arabidopsis roots 3, 7 and 14 days after *P. indica* colonization. Numbers in brackets indicate the number of all genes induced at the respective timepoint, numbers in overlapping shapes indicate the number of genes induced at all involved timepoints.

Due to the lack of the appropriate control, the data from *P. indica* inoculated roots 7 dpi were compared to both mock-treated controls, 3 dpt and 14 dpt. To distinguish gene regulation at 7 dpi as a result from the fungal colonization from developmental responses, only significantly regulated genes were used in further analyses that were up- or down-regulated in comparison to both controls. Log2-expression changes of these values were then averaged and resulted in 1427 differentially (978 up- and 449 down-) regulated genes (adj. $p < 0.05$) (Figure 2.33, 7 dpi).

2.2.3.1. Biotic stress responses in Arabidopsis due to *P. indica* colonization

1/3 of all Arabidopsis genes transcriptionally regulated during colonization by *P. indica* (3 dpi: 458/1481 (31%), 7 dpi: 489/1427 (34%), 14 dpi: 996/3194 (31%)) were predicted to be involved in biotic and abiotic stress responses according to the MapMan software (Usadel et al., 2005). These genes are involved in intracellular signalling mechanisms, in particular WRKY transcription factors, pathogenesis-related (PR) gene regulation, secondary metabolite biosynthesis, hormone homeostasis, and abiotic stress responses.

Signalling



Transcriptionally induced genes were predominately receptor kinases, in particular leucine-rich repeat receptor kinases (LRR-RK), wall associated kinases (WAK) and mitogen-activated protein kinases (MAPK). These kinases are involved in the general recognition of different stimuli, including fungal, bacterial or plant derived molecules (Kohorn and Kohorn, 2012; Pitzschke et al., 2009; Roux et al., 2011). MAPK11 and the MAP kinase kinase (MAPKK) 7 were induced at all three timepoints while MAPKK9 was induced at 7 dpi and 14 dpi and MAPK5 only at 14 dpi. MAPK11 has been shown to be induced upon treatment with different MAMPs suggesting a general role in plant defence responses (Eschen-Lippold et al., 2012). Different reports however have been published on the function of MAPKK7 and MAPKK9. While both kinases are discussed to have a redundant function due to their ~89% sequence similarity on amino acid level, MAPKK7 is involved in auxin transport and induction of basal as well as systemic acquired resistance (SAR) in Arabidopsis (Dai et al., 2006; Zhang et al., 2007). MAPKK9 was shown to phosphorylate MAPK3 and MAPK6 which are important components in plant defence as well as abiotic stress responses (Galletti et al., 2011; Xu et al., 2008; Yoo et al.,

Figure 2.34: Selected upregulated *A. thaliana* proteins involved in intracellular signalling. Shown are log2 expression changes calculated against gene expression during growth on PNM. Expression changes are capped at a maximum induction of 2^5 for visualization purposes.

2008). Much less is known for MAPK5 except its involvement in induction of several transcription factors of the WRKY family (Popescu et al., 2009).

WRKY transcription factors

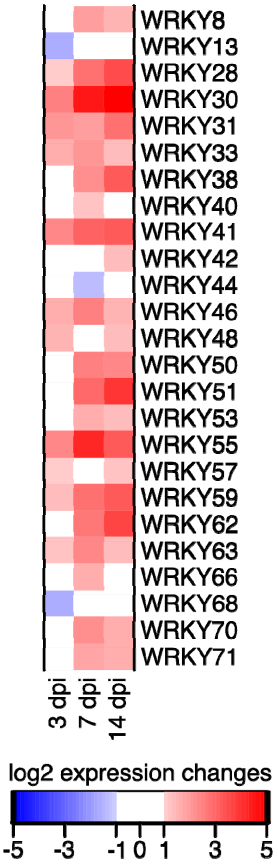


Figure 2.35: Selected upregulated *A. thaliana* transcription factors of the WRKY family. Shown are log2 expression changes calculated against gene expression during growth on PNM. Expression changes are capped at a maximum induction of 2^5 for visualization purposes.

Members of the WRKY family have been shown to be involved in the regulation of different transcriptional responses to biotic and abiotic stresses (Dong et al., 2003; Eulgem et al., 2000). During colonization by *P. indica*, the expression of WRKY transcription factors was either highest at 7 dpi or increased over the three analyzed timepoints (Figure 2.35). Implicated functions in defence of these WRKY transcription factors have been shown by studies on the Arabidopsis resistance or susceptibility towards the necrotrophic pathogens *Botrytis cinerea*, *Sclerotinia sclerotiorum* and *Erwinia carotovora* and the biotrophic pathogens *Pseudomonas syringae* and *Golovinomyces orontii*. Five of the characterized transcription factors (WRKY8, 28, 33, 40, 48) where thereby shown to have a function in defence induction against necrotrophic pathogens (Birkenbihl et al., 2012; Chen et al., 2010, 2013; Mao et al., 2011; Pandey et al., 2010; Schön et al., 2013; Xing et al., 2008) and eight (WRKY38, 41, 46, 50, 51, 53, 62, 70) were shown to have a function in defence induction against biotrophic pathogens (Gao et al., 2011; Higashi et al., 2008; Hu et al., 2012; Kim et al., 2008; Miao et al., 2013; Shim et al., 2013). Other WRKYs induced upon colonization by *P. indica* have either a function in abiotic stress responses (WRKY8, 30, 57, 63, 70) (Hu et al., 2013; Jiang et al., 2012; Li et al., 2013; Ren et al., 2010; Scarpeci et al., 2013) or have not been characterized yet (WRKY13, 31, 42, 44, 55, 59, 68, 71). Most of the studies have further shown that the WRKY transcription factor family mainly regulates defence responses by alteration of the hormone homeostasis, in particular salicylic- and jasmonic acid, and the production of the phytoalexin camalexin.

PR-proteins and secondary metabolites

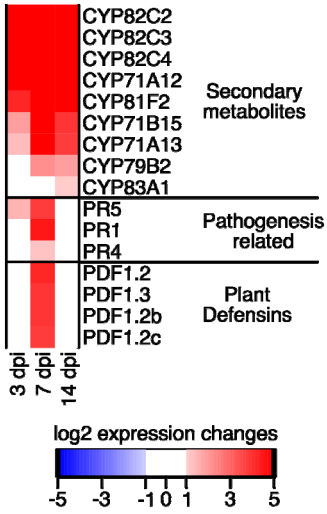


Figure 2.36: Selected upregulated *A. thaliana* defence related genes. Shown are log2 expression changes calculated against gene expression during growth on PNM. Expression changes are capped at a maximum induction of 2^5 for visualization purposes.

The transcriptional induction of pathogenesis related (PR) genes is associated with a regulation of different defence responses in plants (Kesarwani et al., 2007; Kombrink and Somssich, 1997; Van Loon and Van Strien, 1999; Stintzi et al., 1993; Wang et al., 2013b). Expression of PR1, PR2 and PR5 in Arabidopsis is related to SA-dependent defence induction, while expression of PR3 and PR4 is related to JA-dependent defence induction (Hamamouch et al., 2011). The plant defensins (PDF) induced during colonization by *P. indica* were shown to be involved in JA- and ET-dependent defence responses (Brown et al., 2003; Leon-Reyes et al., 2010; Manners et al., 1998). In the performed microarrays, genes encoding PR proteins and PDF were strongly induced at 7 dpi. These included in particular PR1, 4 and 5 and PDF1.2a, b, c and PDF1.3 (Figure 2.36). Different reports have so far shown that secondary metabolites like glucosinolates and phytoalexins are molecules with a broad-spectrum antimicrobial activity (Bednarek et al., 2009; Chen et al., 2012; Clay et al., 2009; Freitas et al., 2013; Kliebenstein, 2004; Saavedra et al., 2010, 2012; Tierens et al., 2001). In Arabidopsis, in particular members of the P450 monooxygenase superfamily have been shown to be essential for the pathogen-induced accumulation of glucosinolates and the phytoalexin camalexin. During colonization by *P. indica* genes encoding members of the CYP82C subfamily (CYP82C2, CYP82C3 and CYP82C4) as well as CYP81F2 are strongly upregulated at all analysed timepoints. These monooxygenases are involved in the biosynthesis of glucosinolates (Bednarek, 2012; Liu et al., 2010). Additionally are genes strongly induced which are involved in the camalexin biosynthesis pathway, like CYP71A12, CYP71A13 and CYP71B15 (PAD3) (Ahuja et al., 2012; Hiruma et al., 2013; Møldrup et al., 2013).

Hormones

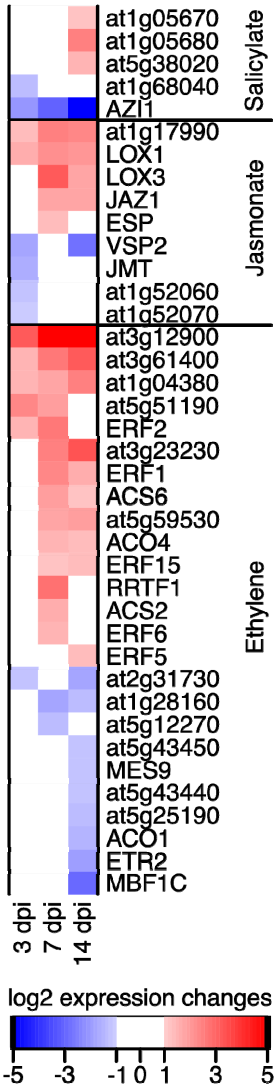
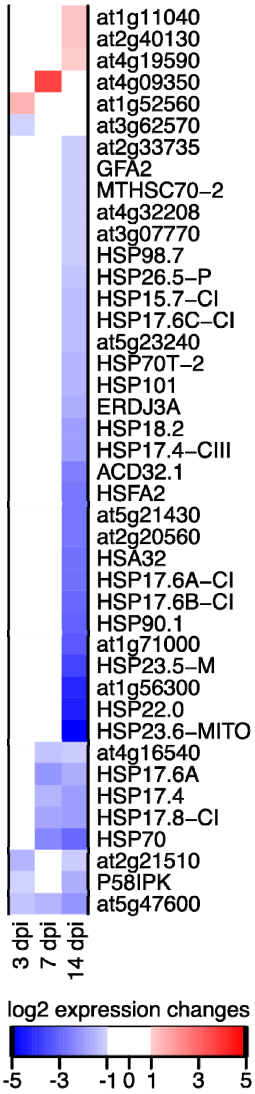


Figure 2.37: Selected upregulated *A. thaliana* genes involved in hormone biosynthesis and signalling. Shown are log2 expression changes calculated against gene expression during growth on PNM. Expression changes are capped at a maximum induction of 2^5 for visualization purposes.

The phytohormones jasmonic acid, salicylic acid and ethylene are important signals in involved in abiotic stress responses as well as the induction of basal defence responses against biotrophic (SA) or necrotrophic (JA/ET) pathogens in Arabidopsis (Van der Ent et al., 2009; Nawrath and Métraux, 1999; Wees et al., 2003) and plants in general. Only a minor number of genes that have been characterized to be involved in hormone-mediated signalling were identified in the performed microarrays (Figure 2.37). For salicylate, this included the *AZI1* (Azelaic acid induced 1) gene which was shown to be involved in the accumulation of SA (Jung et al., 2009) and is induced under cold stress (Xu et al., 2011). For jasmonate, *LOX1* and *LOX3* (Lipoxygenase 1 and 3) were induced for which functions in defence responses, lateral root development and leaf senescence have been proposed (He et al., 2002; Vellosillo et al., 2007). An additional induction was observed for *JAZ1* (JA ZIM-domain 1), a key repressor of JA signalling (Hou et al., 2010) and ESP (Epithiopecifier protein), a protein promoting the hydrolysis of glucosinolates to nitriles (Lambrix et al., 2001). The Ethylene Response Factor (ERF) family contains 122 members in Arabidopsis of which five (ERF1, 2, 5, 6 and 15) were induced during *P. indica* colonization. These were shown to be involved in mediating responses to ethylene (Büttner and Singh, 1997; Solano et al., 1998), abiotic stresses (Fujimoto et al., 2000; Zhu et al., 2010) and pathogen infection (Maleck et al., 2000; Oñate-Sánchez and Singh, 2002).

Abiotic stress responses



The majority of the proteins involved in responses against abiotic stress in Arabidopsis that were regulated during *P. indica* colonization were heat shock proteins (Hsps). These proteins are an essential part of the heat stress response and are assumed to play a central role in acquired thermotolerance in plants (Evrard et al., 2013). Furthermore have Hsps been shown to be induced under high light conditions and are involved the detoxification process in the light induced increase of reactive oxygen species (ROS) (Panchuk et al., 2002; Rossel et al., 2002). A downregulation of several heat shock proteins has been linked to an improved nutritional status of Arabidopsis plants feeding on non-pathogenic *E. coli* (Paungfoo-Lonhienne et al., 2010a, 2010b).

Figure 2.38: Gene encoding regulated *A. thaliana* heat shock proteins. Shown are log2 expression changes calculated against gene expression during growth on PNM. Expression changes are capped at a maximum induction of 2^5 for visualization purposes.

3. Discussion

3.1. The Sebacinales share genomic traits common in biotrophic and saprotrophic fungi

The Sebacinales are a basal order within the Agaricomycotina and includes obligate as well as facultative biotrophs like *P. indica* and *S. vermifera*. Within the Sebacinales, obligate biotrophy is discussed to be derived from saprotrophy (Basiewicz et al., 2012; Hibbett et al., 2000; Weiß et al., 2011; Zuccaro et al., 2011) and fungi like *P. indica* and *S. vermifera* which share traits of both lifestyles are therefore ideal candidates to analyse the transition state from saprotrophy to biotrophy. Comparative analyses showed that the genome of *S. vermifera* is 50% larger and contains 30% more predicted genes than the genome of *P. indica*. The number of repetitive elements, in particular simple sequence repeats (SSR), is strongly reduced in *P. indica* in comparison to *S. vermifera* or other plant pathogens and symbionts (Martin et al., 2010; Muszewska et al., 2011; Thon et al., 2006). Although the number of protein coding genes encoded on both genomes is significantly different, strong similarities could be identified in functionally related protein groups represented within both genomes. These similarities include in particular proteins that are potentially involved in nutrient acquisition (transporters, cell wall degrading enzymes (CWDE), peptidases) or are required for interaction with organisms in the environment (Effectors, Lectins). CWDE, including enzymes acting on cellulose, xylan, glucan, pectin, lignin, arabinose and glucose, are typically associated with saprotrophism and probably contribute to *P. indica*'s and *S. vermifera*'s ability to grow readily on different synthetic media and colonize a wide range of mono- and dicotyledonous plants. This further indicates that both genomes did not undergo host driven specialization as observed in typical obligate biotrophs (Spanu et al., 2010).

3.2. Gene/domain expansions in the Sebacinales and their regulation during plant colonization

The evolution of novel gene functions has mainly evolved by modification of existing genes rather than by creation of new ones *ab initio* (Cohen-Gihon et al., 2011). Gene duplication has in this context the advantage that one gene copy can be modified while the original one maintains its function. Mutations in the gene copy can thereby lead to the following three scenarios: (1) The mutations do not affect the function of the protein and both genes maintain their functional properties; (2) The mutated gene may acquire a new or additional function in

comparison to the original one (Hickman and Rusche, 2010); (3) The mutated genes loses its function partially or completely and becomes a pseudogene (van der Burgt et al., 2013). In both Sebacinoid genomes were members of the expanded gene families found in physical clusters of two to seven adjacent genes. Additionally, indications for each of the three described scenarios have been found in the genome of *P. indica* and partially also in the genome of *S. vermifera*. The duplication of genes which maintained their function can be assumed for clusters of different xylanases (GH10 and GH11 members) and single gene duplication of metallopeptidases. Although amino acid sequences of genes within these clusters can differ strongly, the length and exon/intron structure is highly similar. Genes in some of these clusters were further shown to be transcriptionally co-regulated during plant colonization supporting the hypothesis that they are involved in similar processes. Acquisition of a new function during gene duplication may further be hypothesized for genes of the expanded carbohydrate binding protein families in *P. indica* and *S. vermifera*. In particular for LysM and WSC domain containing proteins different domain structures and domain combinations could be identified which are unique in comparison to the other analyzed fungal genomes. Gene duplication and domain rearrangements have also been suggested to occur simultaneously during evolution of new functions rather than each process acting independently (Cohen-Gihon et al., 2011). Some of the carbohydrate binding proteins display an induced expression *in planta* while others are not regulated. This further indicates that they function in other processes as suggested earlier allowing the Sebacinoid fungi to adapt faster to different environmental stimuli. Pseudogenization on the other hand might have occurred in some of the effectors of the *P. indica* DELD family. The size, amino acid sequence and exon/intron structure of this protein family is well conserved over all 29 members, but for four proteins no signal peptide could be predicted. No sign for a wrong gene prediction could be identified bioinformatically and because these four genes are not induced during plant colonization in contrast to most of the other DELDs, it is more likely that these are pseudogenes rather than DELD proteins that have an intercellular function.

A strongly reduced set of repetitive sequences could be found in *P. indica* in comparison to other analyzed fungi which indicates that processes like retrotransposition occur rarely in this fungus. The transposition machinery requires four components. The reverse transcriptase produces cDNA from a RNA template; the RNAse H degrades the RNA template; the integrase introduced the cDNA into the genome; aspartic proteases process larger transposon fragments (Muszewska et al., 2011). The transcriptase and invertase may thereby be identified by their conserved domains (Representative Pfam domains RVT_1, RVT_2 and rve, gag_pre-

integrs), while RNAses (RNase_H) and the aspartic proteases (RVP_2, Asp) involved in this process show a higher sequence variability and are therefore more difficult to identify (Muszewska et al., 2011). The transcriptase and invertase domains were found to be significantly expanded in the genome of *P. indica*, but not in *S. vermifera*, and several genes putatively involved in the retrotransposition process were identified in the proximity of them. Because repetitive sequences are rare in the *P. indica* genome it might be concluded that most repetitive sequences could not be assembled in the *P. indica* draft genome and were therefore discarded. Though there is no evidence that sequence information is missing in the draft genome after manual revision of the reads. Estimations of the genome size by pulse field gel electrophoresis and qRT-PCR rather support a genome size that is even smaller than the actual 25 Mb (Basiewicz et al., 2012). Homologues could be identified to almost all conserved eukaryotic genes (CEG) further supporting that only a minor part of the genome may have been discarded due to assembly problems. It might therefore be speculated that retrotransposition is tightly controlled in *P. indica* by mechanisms as transcriptional gene silencing (TGS) through histone modification or posttranscriptional gene silencing (PTGS) via RNA interference (RNAi) or both (Nakayashiki, 2011). The functionality of RNAi in *P. indica* has been shown experimentally via targeted downregulation of the tryptophan transaminase 1 (*PiTam1*) and the high-affinity ammonium transporter 1 (*PiAmt1*) genes (Hilbert et al., 2012; Lahrmann et al., 2013). On the other hand, studies of the genome of *Neurospora crassa* revealed that mobile genetic elements can be silenced through mechanisms such as repeat-induced gene silencing (RIP) (Galagan and Selker, 2004; Selker, 2002). RIP emerged after gene duplication in *N. crassa* resulting in the inactivation of mobile elements (Galagan and Selker, 2004; Galagan et al., 2003). Hence, it is possible that transposable elements could only rarely been identified in *P. indica* and *S. vermifera* because mechanisms like RIP exist in the Sebacinales.

3.2.1. Potential to avoid recognition by masking of elicitors

Invading microbes can be recognize by plants through their non-self detection system. This system includes pattern recognition receptors (PRR) and nucleotide-binding and leucine-rich repeat (NB-LRR) proteins to identify any potential threat to the plant (Eitas and Dangl, 2010; Jones and Dangl, 2006; Schulze-Lefert and Panstruga, 2011). Defence responses which are induced upon perception of such a threat differ between plant species and may include the secretion of broad spectrum anti-microbial secondary metabolites as well as compounds specifically evolved against distinct microbial species. The host range of a microbe is

therefore related to its ability to overcome this multi-layered detection and defence system (Schulze-Lefert and Panstruga, 2011). Hence, fungi with a broad host range, like *P. indica* and *S. vermifera*, must evolve strategies to target and suppress conserved defence mechanisms or hiding for the innate immune system by preventing recognition or, more likely, both. This might explain the high numbers of lectin-like proteins in *P. indica* and *S. vermifera* in comparison to the other analyzed fungi. Especially lectin-like LysM proteins have been shown to be important virulence factors in *Cladosporium fulvum* and *Mycosphaerella graminicola* (van Esse et al., 2007; de Jonge et al., 2010; Marshall et al., 2011). The LysM proteins from these fungi were shown to prevent chitin-triggered host immune responses by sequestering chitin fragments in the apoplast thereby masking them. Additionally, a protective function against plant derived chitinases could be observed for a LysM protein. It could be shown by constitutive overexpression and fusion to a GFP tag, that at least one of the lectin-like LysM proteins in *P. indica* binds to the hyphal cell wall. As the transcription of the protein is upregulated during colonization of *A. thaliana*, it is possible that is has a similar function as Avr4 from *C. fulvum* in protection of surface exposed chitin in the fungal cell wall. It remains to be clarified if this is the only function of the protein or if it is also able to sequester free chitin molecules within the apoplastic space. Interestingly, comparative genomics revealed a significant expansion of proteins containing the LysM domain, the cell wall integrity and stress response component (WSC) or the carbohydrate-binding module family 1 (CBM_1) in the genome of *P. indica* and *S. vermifera* which are partially plant responsive. This significant expansion of carbohydrate-binding proteins is not observed in any other fungal genome that has been analyzed. It is therefore tempting to speculate that WSC containing lectin-like proteins are important elements in root generalists and may also function in protecting and masking the fungal hyphae. There is little information available about the function of WSC domains in fungi. The *Saccharomyces cerevisiae* protein Wsc1, which binds β-glucan via its WSC domain, has been shown to be localized in the cell membrane, where it acts as sensor of cell wall integrity and alkaline pH stress (Heinisch et al., 2010; Serrano et al., 2006). Hence, it is possible that some of the WSC proteins in *P. indica* and *S. vermifera* have a similar function in cell wall stress signalling. Due to the strong expansion of these proteins however, it seems unlikely that all WSC proteins are involved in such a mechanism. A similar function in signalling has been reported for WSC proteins from *Aspergillus nidulans* (wscA and wscB) and *Neurospora crassa* (WSC-1) (Futagami et al., 2011; Maddi et al., 2012). These proteins seem to be sensors of osmotic and acidic pH stresses, thereby regulating the synthesis of α-1,3-glucan based on the different environmental

stresses. The synthesis of α-glucans has further been implicated as a method to avoid host recognition. In the rice pathogen *Magnaporthe oryzae* as well as in the human pathogen *Histoplasma capsulatum*, surface exposed β-glucans are covered with α-1,3-glucans upon host colonization in order to prevent recognition through glucan-binding receptors (Esteban et al., 2011; Fujikawa et al., 2012; Rappleye et al., 2007). Based on these data, WSC proteins in the Sebacinales might have a similar function in the synthesis of α-1,3-glucan in order to avoid host recognition. On the contrary, α-1,3-glucan could not be detected on the outside of haustoria of the powdery mildew fungus *Golovinomyces orontii* during Arabidopsis colonization although the fungus is able to produce it (Micali et al., 2011). Some of the WSC proteins in *P. indica* and *S. vermifera* contain a signal peptide and their respective genes are transcriptionally induced upon plant colonization. Taking into account what is known about the LysM proteins, it is possible that these proteins have a function in protection β-glucans in the fungal cell wall from being hydrolyzed by plant derived glucanases or by sequestering glucan fragments that are released during the penetration of the host cell. Morphological changes could be observed in the hyphae of *P. indica* upon plant colonization. It is therefore possible that protection of β-glucans in the fungal cell wall, either indirect by WSC-mediated synthesis of α-1,3-glucan or direct by lectin-like WSC proteins, depends on the colonization status. Additionally, WSC lectins might bind specifically to β-(1,3)-(1,4) glucans released from the plant cell wall upon penetration which could be perceived by the plant as danger-associated molecular patterns (DAMP) (Aziz et al., 2007). Besides the identified lectin-like proteins that bind to fungal cell wall components, proteins containing the cellulose-binding domain CBM_1 may act only on plant cell wall derived cellulose. Swollenin from *Trichoderma reesei* and cellulose-binding elicitor lectin (CBEL) from *Phytophthora infestans* and *Phytophthora parasitica* which resemble a similar structure to the proteins found in *P. indica* and *S. vermifera*, have been reported to possess an expansin-like function on cellulose-containing cell walls (Gaulin et al., 2002, 2006; Jones and Ospina-Giraldo, 2011; Saloheimo et al., 2002). Genes encoding CBM_1 containing lectins which are predicted to be secreted by *P. indica* are induced during early (3 dpi) plant colonization stages. If these proteins do have an expansin-like function, they might facilitate colonization of the epidermis by increasing the permeability of the cells at the site of penetration. Cellodextrin, a water soluble cellulose derivates, was shown to act as a DAMP and induce an oxidative burst in grapevine (*Vitis vinifera* L.) (Aziz et al., 2007). The CBM_1 lectins might therefore also have a function in masking of cellulose hence preventing DAMP-triggered immune responses. The transcriptional induction of CBM_1 lectins at the late stages (14 dpi) correlates with the

induction of different hydrolytic enzymes. This increased hydrolytic activity might also increase the release of, among others, cellulose from the cell wall. This would support the requirement for CBM_1 lectins in DAMP binding at this timepoint. In contrast to *P. indica*, none of the putatively secreted CBM_1 containing lectins in *S. vermifera* showed a transcriptional induction during plant colonization. This could be a result of the combined sequencing of mixed RNA from 3, 7 and 14 dpi which could mask a specific upregulation during early colonization. The strong expansion of lectin like proteins in *P. indica* and *S. vermifera* might have evolved as a mechanism to prevent initial immunity responses in the plant host by masking fungal as well as plant derived carbohydrates.

3.2.2. Hydrolytic degradation of plant cell wall components by Sebacinoid fungi

The cell wall of plants is a network of the different polysaccharides, where cellulose, hemicellulose and pectin are the major components (Mohnen et al., 2008; Stumpf and Conn, 1988). In order to retrieve carbon from cell wall bound polysaccharides, microbes can hydrolyse these complex compounds into monosaccharides and oligosaccharides by a variety of different cell wall degrading enzymes (CWDE) (Gilbert, 2010). The secretion of CWDE by fungi is considered to be predominantly a feature of saprotrophic and necrotrophic fungi while obligate biotrophs and symbiotic fungi lost most of these enzymes during their evolution (Eastwood et al., 2011; Kikot et al., 2009; Nagendran et al., 2009; Di Pietro et al., 2009; Spanu, 2012). Regarding this aspect it is interesting that some families of the glycosyl hydrolases, in particular GH10, 11 (xylanases), GH61 (monooxygenases) and GH5/6 (cellulases/cellobiohydrolases/glucanases), are expanded in *P. indica* and *S. vermifera*. Although reports not only in barley, but also Arabidopsis have shown that *P. indica* requires cell-death for successful colonization (Deshmukh et al., 2006; Qiang et al., 2012b), the reasons behind these observations remain speculative as fungal colonization leads to growth promotion and increased disease resistance in both plants (Camehl et al., 2010; Hilbert et al., 2012; Lahrmann et al., 2013; Stein et al., 2008; Waller et al., 2005). It needs to be additionally mentioned that a natural (non-pathogenic) root cortex death (RCD) has been described in cereals which results in about 40% dead cortex cells in barley after 3 weeks of growth (Henry and Deacon, 1981). This RCD was measured in terms of cells lacking a nucleus and was most evident at the differentiation zone of the root and partially occurred in the elongation zone. Interestingly, intracellular colonization of barley roots by *P. indica* is pronounced in this zone and occurs only rarely at the meristematic zone or the root tip (Deshmukh et al., 2006).

Deshmukh and colleagues further showed that colonization of *P. indica* results in downregulation of BAX inhibitor 1 (BI-1), a negative regulator of apoptotic cell death and congruently, that overexpression of BI-1 reduces *P. indica* colonization. Although Deshmukh and colleagues concluded from this observation that *P. indica* requires cell death for successful host colonization, microarray analyses of the *P. indica* – barley interaction provide hints that *P. indica* rather proliferates faster on the already dead plant material than be restricted by living cells. In this context, downregulation of BI-1 could also be interpreted as a normal immune response to the colonization of a biotrophic fungus.

The transcriptional induction of genes encoding CWDE in *P. indica* was observed predominately during the late, cell-death associated, colonization of barley roots and only partially during the early biotrophic phase (Lahrmann et al., 2013; Zuccaro et al., 2011). Genes of these families are among the most highly upregulated genes at all analyzed timepoints during the interaction, but interestingly most of them are not continuously induced but rather expressed at specific timepoints which is more during the early interaction or during late interaction. The control of cell wall degrading enzymes (CWDE) in plant colonizing fungi general is influenced by the availability of carbon by a process called catabolite or glucose repression (Nadal et al., 2010; Ruijter and Visser, 1997; Tonukari et al., 2000). The observed pattern of induction and non-induction of CWDE could in this context mean that the fungus is not acquiring sufficient amounts of carbon in the early and late phase of the barley colonization but adequate amounts in between both phases. On the other hand, it is possible that CWDE induced in the pre-penetration and early interaction phase are differentially regulated due to different functional properties. For copper-dependent monooxygenases (GH61) different binding specificities have been suggested depending on the carbohydrate binding module (CBM) with which they are combined in the gene (Lo Leggio et al., 2012; Phillips et al., 2011; Žifčáková and Baldrian, 2012). In *P. indica* 10 GH61 were found combined with CBM1, 15 with no other functional domain. Furthermore have xylanases of the families GH10 and 11 been shown to be crucial for plant infection in *Magnaporthe oryzae* (Nguyen et al., 2011) and *Fusarium graminearum* (Sella et al., 2013), possibly by being locally expressed at the site of penetration. In both contexts could the induction of CWDE explained in the following way. During the early interaction phase, hydrolases and monooxygenases are secreted locally to allow or facilitate penetration of cells. During late interaction, these enzymes are together with other glycosyl hydrolases used to decompose the cell walls of dead barley cortex cells for carbon acquisition. The latter case would additionally fit with the induction of several genes encoding proteins putatively

involved in protein degradation and nitrogen transport which showed an increased induction over time and are particularly strong expressed during fungal proliferation on autoclaved barley roots. The increasing RCD in barley could account for this similarity in the expression profile between living and autoclaved roots. The expression levels of various key genes normally affected by starvation, such as those involved in autophagy or coding for metacaspases, acetyl-CoA synthetase and enoyl-CoA hydratase (Keon et al., 2007; Todd et al., 1998), were transcriptionally unaffected or even downregulated during symbiosis. This observation is consistent with nutrient availability during biotrophic colonization in barley. By using a RNAi mediated gene silencing approach it could be shown that nitrogen availability is a signal for the expression of CWDE in *P. indica* (Lahrmann et al., 2013). The silencing of the high-affinity nitrogen transporter PiAmt1, which is a nitrogen importer with an additional sensor function, results in a prolonged biotrophic colonization of barley roots characterized by a lower expression of CWDE.

The strongly reduced expression of xylanases in Arabidopsis compared to barley might be explained by differences in the cell wall composition of these two plants, because xylan is more abundant in the cell wall of monocots than in the cell wall of dicots (Burke et al., 1974). On the contrary, also monooxygenases, which act synergistically with cellulolytic enzymes in cellulose degradation (Žifčáková and Baldrian, 2012) and cellulases are strongly induced during late barley colonization while only partially induced during Arabidopsis colonization. As cellulose is more abundant in dicots than in monocots (Burke et al., 1974), the observed induction pattern shows that *P. indica* massively degrades plant cell walls in barley at later timepoints while it does not in Arabidopsis. The induction pattern could further indicate, that a state of glucose depletion is reached in barley as expression of xylanases and also other CWDE was shown to be repressed in the presence of glucose (Nadal et al., 2010; De Vries et al., 1999). In contrast to those xylanases involved in the degradation of cell walls another group might be distinguished which are important for penetration of living cells and are transcriptionally induced during different timepoints of plant colonization by *P. indica* and during Arabidopsis colonization by *S. vermifera*. It was shown in *Magnaporthe oryzae* that RNAi mediated down-regulation of a group of xylanases resulted in an increased hypersensitive response in penetrated host cells (Nguyen et al., 2011). The authors favoured explanation for this is that the delayed colonization by the knock-down mutants provides more time for the plant cell to prepare an effective defence against the pathogen. They furthermore show that xylanases from the same group, which are not targeted by the RNAi construct, show an induction in gene expression over wildtype levels. This indicates that

xylanases in phytopathogenic fungi and maybe also in the Sebacinales were expanded to increase functional redundancy and penetration speed in order to allow a rapid penetration followed by downregulation of the plant defence system.

3.2.3. Proteolytic cleavage in the fungus plant interaction

Fungal proteases are involved in a variety of different processes ranging from nutrient liberation during germination via substrate hydrolysis for nutrition to cell death initiation. Proteases are grouped into eight catalytic types based on the functional group at the active site, namely asparagine, aspartic, cysteine, glutamic, metallo, serine, threonine and unknown. Secreted aspartic, serine and metalloproteases have so far been studied most in fungi as they are considered to be important virulence factors in many pathogenic species (Behnsen et al., 2010; Bhadauria et al., 2010; Fan et al., 2010; Hung et al., 2005; Yike, 2011). Studies on the opportunistic human pathogens *Aspergillus fumigatus* and *Aspergillus flavus* have further shown that expression and secretion of proteases is substrate specific and that a certain functional redundancy among proteases from different families exists (Farnell et al., 2012; Ramesh and Kolattukudy, 1996; Sriranganadane et al., 2011).

In both, *P. indica* as well as *S. vermifera*, genes encoding putative metalloproteases of the families M35 (Deuterolysin), M36 (Fungalysin) and M43B (Pappalysin-1) are among the most strongly induced genes upon colonization of plant root tissue. A deuterolysin in the human pathogen *Coccidioides posadasii* was shown to digest a cell surface antigen during endospore differentiation thereby preventing host recognition (Hung et al., 2005; Li et al., 2012). It is possible that comparable mechanisms exist in the plant fungus interaction and that *P. indica* and *S. vermifera* utilize deuterolysins in a similar way to prevent recognition by plant receptors. A further implication of metalloproteases in virulence was found for a chitinase-modifying protein from the plant pathogen *Fusarium verticillioides* (Fv-cmp). This protein was shown to be a fungalysin that is able to truncate, thereby inactivate class IV chitinases during plant infection (Naumann, 2011; Naumann and Price, 2012). Metalloproteases could, however, be identified in all of the fungi analyzed within this thesis which indicates that they can not only be involved in pathogenicity. This is supported by a report on an expanded family of fungalysins in *C. cinerea* (Lilly et al., 2008). The authors showed by phylogenetic analyses that only one of the eight fungalysins from *C. cinerea* has putative orthologues in pathogenic fungi while the remaining seven grouped with three fungalysins from *L. bicolor*. This supports the hypothesis that functionally different groups exist within each of the described families. It might further explain why most of the

metalloproteases induced in *P. indica* are also expressed during growth on autoclaved barley roots. These proteases are probably more generally involved in the liberation of nitrogen from dead or decaying plant source as it has been suggested for metalloproteases of the M36 family in *L. bicolor* (Plett and Martin, 2011). In contrast to the expansion of metalloproteases in *P. indica*, certain cysteine protease families were shown to be expanded in *S. vermifera* but not in *P. indica*. These included mainly caspases and metacaspases of the family C14 which were, however, not transcriptionally induced during colonization of Arabidopsis roots. The metacaspase YCA1 of *Saccharomyces cerevisiae* is a member of the C14 family which has been described to be involved in apoptosis (Khan et al., 2005). Further studies of the protein revealed another function in protein quality control which leads to an improved fitness and adaptability of growing yeast (Lee et al., 2010). Concerning apoptosis, studies on *Aspergillus niger* and *Aspergillus nidulans* described carbon starvation as a possible cause of the occurrence of this cell death mechanism (Emri et al., 2005; Nitsche et al., 2012). It was observed that *S. vermifera* degrades its own mycelium under carbon-limiting conditions and it might be possible that this degradation occurs due to apoptosis mediated or accompanied by proteases of the C14 family. If this hypothesis holds true, *S. vermifera* is probably not under carbon limitation during colonization of Arabidopsis roots. Whether the ability to degrade its own mycelium is stimulated in a mycoheterotrophic orchid host which is dependent on fungal derived carbon remains speculative as data stressing this hypothesis are missing.

3.2.4. Nutrient transport and detoxification mechanisms in *P. indica* and *S. vermifera*

In the classical arbuscular plant fungus symbiosis, phosphorus and nitrogen are transported by the fungus into the host plant in return for photosynthetically fixed carbon (Bonfante and Genre, 2010; Smith et al., 1994). Most research within this field has focussed so far on ecto- and arbuscular mycorrhizal associations and information of transporter mechanisms in endophytic- and orchid mycorrhizal symbiosis is still rare. Two transporters that might be important in the plant fungus interaction have so far been characterized experimentally in *P. indica*. The phosphate transporter PiPT (encoded by *PIIN_08122*) was shown to be involved in the transport of phosphate from the fungus into the host plant (Yadav et al., 2010). It was further shown that expression of *PiPT* is induced under phosphate limiting conditions and that phosphate transport mediated by PiPT contributes to the growth promoting effects of *P. indica* (Kumar et al., 2011; Yadav et al., 2010). It is interesting in this context that the gene encoding PiPT was upregulated at 3 dpi in Arabidopsis but downregulated at 3 and 14 dpi in

barley compared to the minimal medium control. Experiments with the AM fungus *Glomus intraradices* colonizing tomato plants have shown that mycorrhizal phosphate transport was repressed, when the phosphate status of the tomato plant was high (Nagy et al., 2009). On the contrary, studies with *Rhizophagus irregularis* on Medicago showed an induction of phosphate transporters when the phosphate availability is high (Fiorilli et al., 2013). In comparison to the described reports, phosphate availability is high within the minimal medium used in this study (cf. Table 5.5). If mechanisms similar to *G. intraradices* and *R. irregularis* control phosphate transport in the symbiotic interaction of *P. indica* with its host, it can be concluded that the phosphate status of barley is high throughout the interaction and that the initial demand for phosphate in Arabidopsis is met during late interaction with *P. indica*. The other characterized transporter in *P. indica* which plays an important role during the plant fungus interaction is the ammonium transporter 1 (PiAmt1 encoded by *PIIN_02036*). The *PiAMT1* gene was shown to be induced under nitrogen limiting conditions and to be involved in ammonium uptake (Lahrmann et al., 2013; Zuccaro et al., 2011). Silencing of *PiAMT1* resulted in enhanced colonization of barley roots by *P. indica*, whereas it had no effect on the colonization of Arabidopsis. A possible function as nitrogen sensor could be shown for PiAmt1 which is linked to the expression of CWDE in *P. indica* ((Lahrmann et al., 2013); Ding, Lahrmann, Zuccaro, unpublished). These results showed that nitrogen availability is low in late interaction with barley while it remains high in Arabidopsis. This nitrogen availability seems to be a result of amino acid re-allocation in Arabidopsis which were shown to be usable alternative nitrogen sources for *P. indica*. The respective orthologous genes of *PiPT* and *PiAMT1* in *S. vermifera* were not regulated during the interaction with *A. thaliana*. This could indicate that the respective transporters in *S. vermifera* have functions similar to those proposed in *P. indica*. In contrast to *P. indica* could a nitrate transporter be identified in *S. vermifera* which probably enables this fungus to use nitrate as nitrogen source. This suggests that *S. vermifera* is not under nitrogen limiting conditions while growing on plant minimal medium (PNM). It would be interesting to analyse the expression of hydrolytic enzymes during colonization of barley roots to verify the effect of the medium composition on barley colonization.

Besides nutrient transporters, different toxin efflux and (multi) drug resistance proteins were found in high abundance in the genomes of *P. indica* and *S. vermifera*. Because a strongly reduced set of genes involved in the production of secondary metabolites could be identified in the genomes, it is unlikely that the toxin efflux transporters are involved in the translocation of toxic substances into host cells like it has been described for pathogenic fungi

(Coleman and Mylonakis, 2009; Howlett, 2006; Keller et al., 2005). Instead, examples also from pathogenic fungi show that toxin efflux as well as drug resistance transporters are able to confer resistance against different antibiotics and plant derived secondary metabolites, in particular glucosinolates and phytoalexins (Coleman and Mylonakis, 2009; Kretschmer et al., 2009; de Waard et al., 2006). An ABC transporter from *Botrytis cinerea* (BcatrB) has for example been shown to be a virulence factor due to its ability to export Arabidopsis derived camalexin from the fungus (Stefanato et al., 2009). This transporter shows high similarity (Blastp, e-Val < 10^{-10}) to several ABC efflux proteins and a multidrug resistance protein in *P. indica* and *S. vermifera*. The respective genes encoding these proteins are up-regulated at 7 and 14 dpi in *P. indica* and in *S. vermifera* during *A. thaliana* colonization suggesting a similar function in camalexin resistance as for the *B. cinerea* transporter. Besides conferring camalexin resistance, further studies of BcatrB have shown that this transporter is also involved in the export of other plant derived secondary metabolites with fungal toxicity, including 2,4-DAPG (Schouten et al., 2008) or Eugenol (Schoonbeek et al., 2003). As several transporters are encoded in the genome of *P. indica* and *S. vermifera* with similarity to BcatrB and as many of them are transcriptionally induced during the colonization of Arabidopsis roots, it seems likely that the production of plant derived secondary metabolites is not actively suppressed by both fungi during colonization. Instead, a passive defence via detoxification of those compounds is induced. This explanation is in line with the broad host range of *P. indica* and putatively also *S. vermifera*. It is tempting to assume that a highly effective, general working detoxification machinery could be a determinating factor in tolerating different toxin-related defence strategies without the need for specific adaptations. This mechanism could therefore be important for the broad host-range and world-wide distribution of the Sebacinales.

3.2.5. Small secreted proteins and putative effector candidates

It was shown that *P. indica* actively suppresses basal plant defence responses like MAMP-triggered immunity (MTI) during its colonization of Arabidopsis roots (Jacobs et al., 2011). As this suppression did not only include chitin- but also flagellin-induced HR, it is unlikely that this is only a result of masking of potential elicitors, but rather involves proteins secreted into the plant cell to manipulate defence responses. Such effector proteins have been shown to be key determinants in the host-fungus interaction independent whether the relationship is symbiotic or pathogenic (de Jonge et al., 2011; Kloppholz et al., 2011; Plett et al., 2011; De Wit et al., 2009). Additionally, effectors in oomycetes and fungi were shown to contain

conserved sequence motifs like the "RxLR" motif in *P. infestans* which is required but not sufficient for translocation into the host cell (Tyler et al., 2013). About 1/3 of the genes encoding putative small secreted proteins (SSPs; size < 300 aa) are induced in *P. indica* upon root colonization. The majority of these proteins (~65%) are induced either during Arabidopsis or during barley root colonization which indicates that they may play an important role in determining the success of the endophytic interaction. SSPs putatively involved in penetration, suppression of plant immunity and growth within living cells show a rather host-specific expression pattern which might be due to strong differences in root structure and defence responses between Arabidopsis and barley. More than half of the SSPs show no significant similarity to other known proteins in the nr-database (e-Val < 10^{-3}) and do not contain any known functional protein domain. These include putative effectors of the DELD protein family in *P. indica*. The majority of these proteins in *P. indica* are secreted, smaller than 200 aa in size and transcriptionally induced upon either barley or Arabidopsis colonization. Based on the expression pattern, *P. indica* DELD proteins may be divided into two groups, one solely induced at early timepoints and one induced at later timepoints of the colonization. As such a pattern has been observed for hydrolytic enzymes, it might be that also the different DELD groups are involved in two different processes during colonization of and proliferation within the root. All members of this protein family occur in close proximity of transposable elements strongly suggesting a significant co-expansion between DELD paralogs and transposon sequences that benefited *P. indica* in some way during adaptation to different hosts. Among other SSPs, those containing glycosyl hydrolase domains are most strongly induced, followed by lectin-like proteins.

3.3. *P. indica* colonizes barley and Arabidopsis roots in a host-dependent way

The colonization of barley roots has been shown to occur in two distinct phases, an initial biotrophic and a later cell-death associated phase (Lahrmann et al., 2013; Zuccaro et al., 2011). Although reports not only in barley, but also Arabidopsis have shown that *P. indica* requires cell-death for successful colonization (Deshmukh et al., 2006; Qiang et al., 2012b), the reasons behind these observations remain speculative as fungal colonization leads to growth promotion and increased disease resistance in both plants. It needs to be additionally mentioned that a natural (non-pathogenic) root cortex death (RCD) has been described in cereals which results in about 40% dead cortex cells in barley after 3 weeks of growth (Henry and Deacon, 1981). This RCD was measured in terms of cells lacking a nucleus and was most

evident at the differentiation zone of the root and partially occurred in the elongation zone. Interestingly, intracellular colonization of barley roots by *P. indica* is pronounced in this zone and occurs only rarely at the meristematic zone or the root tip (Deshmukh et al., 2006). Deshmukh and colleagues further showed that colonization of *P. indica* results in downregulation of BAX inhibitor 1 (BI-1), a negative regulator of apoptotic cell death and congruently, that overexpression of BI-1 reduces *P. indica* colonization. Although Deshmukh and colleagues concluded from this observation that *P. indica* requires cell death for successful host colonization, microarray analyses of the *P. indica* – barley interaction provide hints that *P. indica* rather proliferates faster on the already dead plant material than be restricted by living cells. In this context, downregulation of BI-1 could also be interpreted as a normal immune response to colonization of a biotrophic fungus. This becomes most evident by looking on genes predicted to be involved in the degradation of plant cell walls, in particular glycosyl hydrolases involved in the degradation of cellulose (GH61) and xylan (GH10, 11). Genes of these families are among the most highly upregulated genes at all analyzed timepoints during the interaction, but interestingly most of them are not continuously induced but rather expressed at specific timepoints which is more during the early interaction or during late interaction. The control of cell wall degrading enzymes (CWDE) in plant colonizing fungi general is influenced by the availability of carbon by a process called catabolite or glucose repression (Nadal et al., 2010; Ruijter and Visser, 1997; Tonukari et al., 2000). The observed pattern of induction and non-induction of CWDE could it this context mean that the fungus is not acquiring sufficient amounts of carbon in the early and late phase of the barley colonization but it is getting “enough” in between both phases. On the other hand it is possible that CWDE induced in the pre-penetration and early interaction phase are differentially regulated due to different functional properties. For copper-dependent monooxygenases (GH61) different binding specificities have been suggested depending on the carbohydrate binding module (CBM) with which they are combined in the gene (Lo Leggio et al., 2012; Phillips et al., 2011; Žifčáková and Baldrian, 2012). In *P. indica* 10 GH61 combined with CBM1, 15 with nothing else. Furthermore have xylanases of the families GH10 and 11 been shown to be crucial for plant infection in *Magnaporthe oryzae* (Nguyen et al., 2011) and *Fusarium graminearum* (Sella et al., 2013), possibly by being locally expressed at the site of penetration. In both contexts could the induction of CWDE explained in the following way. During the early interaction phase, hydrolases and monooxygenases are secreted locally to allow or facilitate penetration of cells. During late interaction, these enzymes are together with other glycosyl hydrolases used to

decompose the cell walls of dead barley cortex cells in order to acquire sufficient amounts of carbon. The latter case would additionally fit with the induction of several genes encoding proteins putatively involved in protein degradation and nitrogen transport which showed an increased induction over time and are particularly strong expressed during fungal proliferation on autoclaved barley roots. The increasing RCD in barley could account for this similarity in the expression profile between living and autoclaved roots. The expression levels of various key genes normally affected by starvation, such as those involved in autophagy or coding for metacaspases, acetyl-CoA synthetase and enoyl-CoA hydratase (Keon et al., 2007; Todd et al., 1998), were transcriptionally unaffected or even downregulated during symbiosis. This observation is consistent with nutrient availability during biotrophic colonization in barley.

While *P. indica* undergoes major transcriptional rearrangements during the switch from biotrophic to saprotrophic nutrition in barley, it maintains a predominantly biotrophic nutrition in Arabidopsis. Early interactions are biotrophic in both hosts with comparably low numbers of hydrolytic enzymes and carbon transporters induced, but with increased expression of genes encoding putative toxin efflux proteins. This interaction is cytologically characterized by thicker bulbous invasive hyphae in epidermal cells. These hyphae are enveloped by a host derived plasma membrane which can be stained by the endocytosis marker FM4-64 but prevents staining of hyphal chitin by WGA (Lahrmann and Zuccaro, 2012). These invasive hyphae are present throughout the Arabidopsis colonization process, leading to a non-destructive progression within the epidermis and cortex layers. Consistent with this observation is the reduced expression of *P. indica* genes involved in host cell wall and lipid degradation at 3 and 14 dpi in this host compared to the situation in barley roots were several hydrolytic enzyme are strongly upregulated 14 dpi and invasive hyphae are found more rarely. Instead of these, secondary thinner hyphae are the dominant hyphal structures found at this timepoint and which even completely fill the root cortex cell 30 dpi (Lahrmann et al., 2013). At this stage colonization by *P. indica* does not result in any macroscopically visible disease symptoms, like necrotic lesions, but on the contrary, to growth promotion and overall healthy looking plants (Hilbert et al., 2012; Lahrmann et al., 2013; Waller et al., 2005). The occurrence of a long-term biotrophic nutrition in Arabidopsis and of a switch to a saprotrophic nutrition in barley is further supported by comparative expression analyses of fungal genes involved in primary metabolism and nutrient transport. By using an RNAi mediated gene silencing approach it could be shown that nitrogen availability is a signal for the expression of CWDE in *P. indica* (Lahrmann et al., 2013). The silencing of the high-affinity nitrogen transporter PiAmt1, which is a nitrogen importer with

an additional sensor function, results in a prolonged biotrophic colonization of barley roots characterized by a lower expression of CWDE.

3.4. Differences in the transcriptional response to colonization of Arabidopsis by *S. vermifera* and *P. indica*

The major difference between *S. vermifera* and *P. indica* at the genome level is the strong expansion of patatin domains in *S. vermifera*. The patatin domain is found alone or linked to NACHT/WD40 or NB-ARC/TPR10/TPR12 domains which were shown to be involved, among others, in protein-protein interaction processes (D'Andrea and Regan, 2003; Damiano et al., 2004; Stirnimann et al., 2010). Furthermore, ten of the patatin-like proteins are induced during colonization of *A. thaliana* and nine of these have a combination with the described domains. First characterized in potato, patatins are storage proteins with an additional phospholipase activity involved in the cleavage of fatty acids from membranes (Mignery et al., 1988). Patatin-like proteins are further discussed to be involved in anti-fungal defence mechanisms through a putative glucanase activity (Camera et al., 2009; Kim et al., 2009; Liu et al., 2003). The sugar-dependent 1 protein (SDP1) of Arabidopsis is a patatin-like lipase that is involved in storage oil breakdown during early seed germination thereby being an important factor for providing carbon during postgerminative growth (Eastmond, 2006; Quettier and Eastmond, 2009; Rudolph et al., 2011). The expansion of patatin-like proteins in *S. vermifera* could be explained with the ability of orchid mycorrhiza to stimulate germination of the orchid. This however cannot be a general mechanism because only ten patatin domains could be identified in the green orchid mycorrhizal fungus *Tulasnella calospora.* Alternatively, if patatins are carbon storage proteins, it might be possible that the protein expansion in *S. vermifera* evolved as an additional carbon storage which is used by the orchid host. This could be of special importance in the interaction with myco-heterotrophic orchids which are parasitizing mycorrhizal fungi in order to acquire carbon (Ogura-Tsujita et al., 2009; Rasmussen and Rasmussen, 2009).

In comparison to *in planta* induced genes in *P. indica*, genes induced in *S. vermifera* upon colonization of Arabidopsis roots provide hints towards a slightly different colonization strategy of *S. vermifera*. While expression patterns of peptidases are overall highly similar with a small number of conserved and highly expressed metallopeptidases, a relative high number of CWDE is induced in *S. vermifera* and orthologues to these in *P. indica* are also more often induced during barley colonization. This could indicate that *S. vermifera* colonizes

Arabidopsis more aggressively than *P. indica* or that more cell death occurs in *S. vermifera* colonized Arabidopsis roots which induces a saprotrophic program. This may also explain Arabidopsis plants which show putative stress symptoms like purple coloured leafs (putatively increased anthocyanin accumulation, (Shirley et al., 1995)) and early flowering which occurs rarely in *S. vermifera* colonized plants but was on the other hand never observed during *P. indica* colonization. On the contrary, the observed stress symptoms might be a result from the experimental setup. Plants were inoculated with crushed mycelium instead of spores, because *S. vermifera* does not produce spores in culture. Although the mycelium was regenerated and washed prior to inoculation, it can be expected that a large amount of fungal biomass is dead and therefore an easily accessible nutritional resource. Hence, the production of hydrolytic enzymes may be induced in hyphae outside the plant root in order to liberate carbon from dead hyphae for further growth. Because *A. thaliana* roots are extremely fragile and washing of the roots is therefore difficult, non-colonized hyphae partially remained on the surface and the observed increased induction of hydrolases might be attributed to these hyphae. Putative transporters induced during Arabidopsis colonization by *S. vermifera* include, like in *P. indica*, some strongly induced carbohydrate transporters and predominately toxin efflux and nitrogen transporters. This indicates that nutrient transport is likewise regulated in both fungi during colonization of Arabidopsis. Interestingly, only a low number of genes encoding SSPs and lectins were induced in *S. vermifera*. As these proteins are rather time- and host-specifically induced in *P. indica* upon root colonization, the mixing of inoculation timepoints might have resulted in a dilution of the time-specific responses. Further studies focussing on specific interaction timepoints and probably a different method of plant inoculation will therefore reveal if the colonization of *S. vermifera* differs significantly in certain pathways from that of *P. indica*. Based on the present data, in particular patatin-like proteins and the nitrate transporter seem to be interesting first targets for this further investigation.

3.5. Defence pathways transcriptionally induced in Arabidopsis during *P. indica* colonization

Members of the Brassicaceae are, with a few exceptions, non-hosts for mycorrhizal fungi which is mainly attributed to the production of secondary metabolites and the absence of key genes in the symbiosis (SYM) pathway (Bressan et al., 2009; Koch et al., 2011; Oldroyd and Downie, 2006; Regvar et al., 2003; Venkateshwaran et al., 2013; Vierheilig et al., 2000). Tryptophan derived secondary metabolites like glucosinolates or Camalexin (phytoalexin) are

a well studied virulence factors in *A. thaliana* that confers disease resistance against different pathogens and parasites like *Phytophthora brassicae* (Schlaeppi et al., 2010), *Sclerotinia sclerotiorum* (Stotz et al., 2011) or *Myzus persicae* (Kettles et al., 2013). Members of the cytochrome P450 superfamily have been characterized to be crucial for the production of these secondary metabolites. Synthesis of camalexin requires for example the induction of *CYP71A12*, *CYP71A13* and *CYP71B15* (PAD3) (Ahuja et al., 2012; Hiruma et al., 2013; Møldrup et al., 2013; Nafisi et al., 2007; Stefanato et al., 2009), while *CYP82C2*, *CYP82C3*, *CYP82C4*, *CYP79B2*, *CYP81F2*, *CYP83A1* and *CYP83B1* are involved in broad spectrum disease resistance via glucosinolate production (Bak et al., 2001; Bednarek, 2012; Bednarek et al., 2009; Clay et al., 2009; Liu et al., 2010; Mikkelsen et al., 2000; Naur et al., 2003). Because these secondary metabolites are the strongest induced genes upon colonization by *P. indica* and have been shown to be involved in the defence against phytopathogenic fungi, it could be assumed that they also have a defence function against *P. indica* in Arabidopsis. An induction of drug resistance transporters by *P. indica* would have evolved in order to overcome this defence mechanism. Besides secondary metabolite production, a strong but transient induction of *PR1* and *PDF1.2* at 7 dpi could be observed in the performed microarray analyses. These two genes were shown to be transcriptionally induced in Arabidopsis against *Fusarium sporotrichioides* (Asano et al., 2012) and *B. cinerea* (Buxdorf et al., 2013). The induction of these genes further correlates with the reported active cell death induction in this host (Qiang et al., 2012b) and might indicate that defence responses comparable to those triggered against necrotrophic fungi are induced at this timepoint. Because the induced expression of *PR1* and *PDF1.2* vanishes at 14 dpi this defence response would then be actively suppressed by *P. indica*. Such a suppression has been shown by Jacobs and colleagues who showed a manipulation of JA and gibberellin (GA) responses in Arabidopsis by *P. indica* (Jacobs et al., 2011). Jacobs and colleagues for example showed enhanced colonization of Arabidopsis quintuple della mutants but repressed colonization of jin1-1 mutants by *P. indica.* This is in agreement with an observed induced expression of the gene encoding JA ZIM-domain 1 (JAZ1, aka JIN1) from 7 dpi onwards. JAZ1 is a key repressor of JA signalling which is inhibited by the DELLA proteins (Hou et al., 2010). In this context, *P. indica* might inhibit JA signalling, at least partially, by interfering with DELLA mediated regulation of the *JAZ1* gene. Besides *JAZ1* , also the JA marker VSP2 (Vegetative Storage Protein 2) (Jacobs et al., 2011) was found to be downregulated by *P. indica* as well as the gene encoding AZI1 (Azelaic Acid Induced 1). AZI1 is involved in the accumulation of SA was shown to have a systemic defence priming effect in plants towards

resistance against bacterial pathogens (Jung et al., 2009). Furthermore was the key SA-regulator NPR1 (Wu et al., 2012) not found to be regulated at one of the analyzed timepoints. Although neither JA/SA biosynthesis nor signalling was observed to be induced in Arabidopsis several members of the WRKY transcription factor family were transcriptionally induced. Interestingly were five of these transcription factors shown to act in the defence against necrotrophic pathogens while eight were involved in defence responses against biotrophic pathogens. Furthermore have five WRKYs an implicated function in abiotic stress responses.

Taken together, different defence responses were shown to be upregulated in Arabidopsis during colonization by *P. indica*. These responses most often increased over the time and included pathways that have been shown before to be involved in the defence against biotrophic as well as necrotrophic pathogens. The defence program against *P. indica* includes an unbalance in the JA/SA biosynthesis and signalling pathways which might be a result of cell death induction. This cell death seems however to be transient as genes putatively involved in this pathway are downregulated during the late interaction and because Arabidopsis plants benefit from the colonization by *P. indica* in terms of growth promotion and increased seed production (Camehl et al., 2011; Lahrmann et al., 2013; Shahollari et al., 2007). Other defence responses induced by *P. indica* are independent of these pathways and include predominately an increased expression of genes involved in secondary metabolite biosynthesis.

4. Material and Methods

4.1. Materials and sources of supply

4.1.1. Chemicals and "kits"

All chemicals used during this thesis were of analytical grade and predominately purchased from the companies Difco (Augsburg), Sigma-Aldrich Buchs (Buchs/Switzerland), Merck (Darmstadt), Roth (Karlsruhe) and Sigma-Aldrich (Deisenhofen). For the extraction and purification of DNA fragments from agarose gels, the PROMEGA Wizard® SV Gel and PCR Clean-Up System (Promega, Mannheim) was used. For the extraction of plasmids from *E. coli*, the DNA QIAprep® Spin Miniprep Kit and the QIAGEN® Plasmid Midi Kit (QIAGEN, Hilden) were used. The TOPO TA cloning kit (Invitrogen, Karlsruhe) was used for cloning of PCR products. Other chemicals and kits are described together with the methods they were used in.

4.1.2. Buffers, solutions and media

If not indicated differently, all buffers, solutions and media were autoclaved for 5 min at 121°C. In case of heat instable ingredients, solutions were filter sterilized using Rotilabo® -syringe filters (pore size 0.45 µm or 0.22 µm; Roth, Karlsruhe).

4.1.3. Enzymes

Restriction enzymes were purchased from New England Biolabs (NEB, Frankfurt). The DNA polymerases Taq (Fermentas), Pfu (Promega) and Easy-A (Agilent) were used for standard PCR reactions.

4.1.4. Oligonucleotides

Oligonucleotides used as primers in PCR reactions were purchased from Eurofins MWG Operon (Ebersberg) and Sigma-Aldrich and are listed in table 5.1. All primers, except those used for cloning, were designed to have an optimal annealing temperature of 59°C. Primers used in quantitative real-time PCR reactions were additionally designed with a 3' bias.

Table 5.1: Oligonucleotides designed and used during this thesis

#	Name	Sequence (5'→3')	Usage
01	00867_FULL_FOR	GTCCTCTCGTTGGTCGTGAT	Gene verification and expression analysis of the putative chitin binding (LysM
02	00867_qRT_FOR	CATGGTGTGACACTCCGAGA	

03	00867_REV	ACGTCGAGCTAGAGCCAGAA	domain) lectin-like protein PIIN_00867
04	02169_FULL_FOR	CTCGCACAGGTCATAAGCAA	Gene verification and expression analysis of the putative chitin deacetylase PIIN_02169
05	02169_qRT_FOR	ACAGGACTTGTGCAGCTGTG	
06	02169_REV	GTGTTGGGCCACACATCAT	
07	02170_FULL_FOR	CTGCAACAGCGTTTATCGAA	Gene verification and expression analysis of the putative chitin binding protein PIIN_02170
08	02170_REV	TTGTGCTTGTGCTTGTAGGC	
09	20000_FULL_FOR	AGAACGTTCGGAAAGGGAAC	Gene verification and expression analysis of the putative chitin binding protein PIIN_20000
10	20000_FULL2_FOR	CTACGAGTCGCATGACAGGA	
11	20000_REV	TTCTCGTACAGCCCATCCAC	
12	02172_FULL_FOR	ATGGACCAGGTCATCTCGAA	Gene verification and expression analysis of the putative chitin deacetylase PIIN_02172
13	02172_qRT_FOR	GGGTGCACGCTCAAGATTT	
14	02172_REV	GTCCACACGTCGTTGCAGT	
15	02781_FULL_FOR	GTCTCGTCCAGATCGTTGGT	Gene verification and expression analysis of the putative chitin binding (LysM domain) lectin-like protein PIIN_02781
16	02781_REV	CAAGCCCGGGTTATACTTGA	
17	03654_FULL_FOR	GCGGCAATTGATGATAGACA	Gene verification and expression analysis of the putative chitin binding protein PIIN_03654
18	03654_REV	TTCCAGTAGCGAATGTTGGA	
19	03655_FULL_FOR	GCAACCACGACTTTGCTTTT	Gene verification and expression analysis of the putative chitin binding protein PIIN_03655
20	03655_REV	GATCGAGTTGCAGTCCACAA	
21	05278_FULL_FOR	TCGAAACTACACGGTTGCTG	Gene verification and expression analysis of the putative chitin binding protein PIIN_05278
22	05278_REV	CCTCCTCCGTAATCTGGTCA	
23	05722_FULL_FOR	ACACTTGCGATTCGATTGCT	Gene verification and expression analysis of the putative chitin binding protein PIIN_05722
24	05722_REV	AACCCCCGATACTGGTCATT	
25	05723_FULL_FOR	CGTTGACTCTGCTCGCTGT	Gene verification and expression analysis of the putative chitin binding protein PIIN_05723
26	05723_REV	GCCAGGATGAAAGAGTTGGA	
27	06786_FULL_FOR	CAAAAGAGCACCCAATCTCC	Gene verification and expression analysis of the putative chitin (LysM domain) and glucan (WSC domain) lectin-like protein PIIN_06786
28	06786_REV	TATGACCAAACGAGGGGAAC	
29	08720_LeftPart_FOR	TCGTCTATGTCAACGCTGCT	Gene verification of the putative chitin binding (LysM domain) lectin-like protein PIIN_08720
30	08720_LeftPart_REV	GATAGGAGTGCCATCCCAAA	
31	08720_RightPart_FOR	ATACCCAACCTTGCTGGACA	
32	08720_RightPart_REV	CGCCTGGTTTGGAGATGTAT	
33	08720_Left2Right_FOR	TATCGCTCGCAGCTCTATGA	
34	08720_ Left2Right_REV	GTCCAGCAAGGTTGGGTATC	
35	08721_FULL_FOR	GACTTGGACGCCTGTGTGTA	Gene verification and expression analysis of the putative chitin binding (LysM
36	08721_qRT_FOR	GCCCATACAAGATTGGAGGA	

37	08721_REV	GTTGACCAAGGGATGCAGAT	domain) lectin-like protein PIIN_08721
38	08723_FULL_FOR	TCCTGTTTGCGAGGAGTTCT	Gene verification and expression analysis of the putative chitin binding (LysM domain) lectin-like protein PIIN_08723
39	08723_REV	ATGTGGCAACTGGAGGGATA	
40	08724_FULL_FOR	TCCCTGGTGATACGTGTGAA	Gene verification and expression analysis of the putative chitin binding (LysM domain) lectin-like protein PIIN_08724
41	08724_qRT_FOR	GCATCTATGGTCAATCGCAAG	
42	08724_REV	ACCAAAGCGCGTAGCAATAG	
43	08725_FULL_FOR	CGGGAGTTGTTAATGCAACC	Gene verification and expression analysis of the putative chitin binding (LysM domain) lectin-like protein PIIN_08725
44	08725_REV	CCTCGGCAGAAGACGAATAA	
45	11109_FULL_FOR	TCGTTGCAGCTTCTCTTGTC	Gene verification and expression analysis of the putative chitin binding (LysM domain) lectin-like protein PIIN_11109
46	11109_REV	ACCAATACGTGAGGCAACCT	
47	02036_qRT_FOR	CTGGGTCATGCACTTTATCC	Gene expression analysis of the ammonium transporter 1of *P. indica* (*PiAmt1*)
48	02036_REV	CGGCATCAATGCCAACATAG	
49	02952_qRT_FOR	TGCCCCAGTTCACAAGACTC	Gene expression analysis of the putative metallo-peptidase PIIN_02952
50	02952_REV	AGGTGGCACGGGTGATCT	
51	03008_qRT_FOR	GCAAGTTCTCCGAGCTCATC	Reference for fungal gene expression analysis; targeting PIIN_03008 (*PiTEF*)
52	03008_qRT_REV	CCAAGTGGTGGGTACTCGTT	
53	03737_qRT_FOR	AATTATTTGGGCAGGCAGAG	Gene expression analysis of the putative cellulose binding (CBM1 domain) lectin-like protein PIIN_03737
54	03737_REV	TCGAGATCGTTGAGCTGTGT	
55	04582_ qRT _FOR	TCACGTGGTGAAGAATTGGA	Gene expression analysis of the putative transcription factor of the PAS superfamily PIIN_04582
56	04582_ qRT _REV	AGAGGGAATCAAGACGAGCA	
57	05015_ qRT _FOR	CCTCGCAACATACCTCCCTA	Gene expression analysis of the putative protein PIIN_05015
58	05015_ qRT _REV	ACGCTATAACGGCACAAACC	
59	05825_ qRT _FOR	CGGACAGCTACGAAAAGAGG	Gene expression analysis of the putative glucan binding (WSC domain) lectin-like protein PIIN_05825
60	05825_ qRT _REV	GCTTGCCTAACTCCAATCCA	
61	05872_qRT_FOR	GGGCTTAGACCCAAATGGAC	Gene expression analysis of the effector candidate (RSIDELD) PIIN_05872
62	05872_ qRT _REV	TGCCTCTCTGTTGTGCGTAG	
63	06067_ qRT _FOR	ATTCTAGCCATCCCCTCTCC	Gene expression analysis of the putative cellulose binding (CBM1 domain) lectin-like protein PIIN_06067
64	06067_ qRT _REV	GCATAGACTCGTTTGGGTTGT	
65	06069_ qRT _FOR	TAAAGCCGACGACCGTGTAT	Gene expression analysis of the putative cellulose binding (CBM1 domain) lectin-like protein PIIN_06069
66	06069_ qRT _REV	ATTGTCCCCACAAGGAAGTG	
67	06665_ qRT _FOR	ACCCATGACGGTTCAAAACT	Gene expression analysis of the putative glucan binding (WSC domain) lectin-like protein PIIN_06665
68	06665_ qRT _REV	CCCTCGCATTTCATGATACA	
69	07534_ qRT _FOR	AAAGAAAAGGCGCTCGAAG	Gene expression analysis of the tryptophan transaminase 1 of *P. indica* (*PiTam1*)
70	07534_ qRT _REV	CTCAGCCCCTCGTCTACATC	

71	09643_qRT_FOR	CACCTTCATCATGCAGAGACA	Gene expression analysis of the effector candidate (RSIDELD) PIIN_09643
72	09643_ qRT _REV	AAGTTGGTCGGCCTTAGCTT	
73	09687_ qRT _FOR	CACCTGGATCATGCACAAAC	Gene expression analysis of the effector candidate (RSIDELD) PIIN_09687
74	09687_ qRT _REV	GTGAATGTCGGCCTGTGTT	
75	11595_qRT_FOR	CAAACCAGGCGATATTCCAC	Gene expression analysis of the effector candidate (RSIDELD) PIIN_11595
76	11595_qRT_REV	CGCTTGTTGCATATGGTGAC	
77	AtUBI_F	CCAAGCCGAAGAAGATCAAG	Reference for plant gene expression analysis in *A. thaliana*; targeting *AT3G62250* (Ubiquitin 5)
78	AtUBI_R	ACTCCTTCCTCAAACGCTGA	
79	at5g01900_F	GATCCCACATGATCCGATTC	Gene expression analysis of the transcription factor WRKY62 of *A. thaliana*
80	at5g01900_R	TGGGATGGAGACCAATCTTC	
81	at5g22570_F	TCAGGCTCTTATCCTCCTTCA	Gene expression analysis of the transcription factor WRKY38 of *A. thaliana*
82	at5g22570_R	CTGATCATAACGATCCCACG	
83	at2g14610_F	CGTGAACATGTGGGTTAGC	Gene expression analysis of the pathogenesis-related gene 1 (PR1) of *A. thaliana*
84	at2g14610_R	CACCTCACTTTGGCACATC	
85	02172_cPro_GFP_F	GCCATCGATATGTTCGCGTCACTCCTCCT	Cloning of PIIN_02172 into the GoGFP vector using ClaI
86	02172_cPro_GFP_R	GCCATCGATGTGTCTTACGCAAAGCGGCG	
87	08721nLocNatP_F	GTTGGGCCCACAAGACCTAGCTCCATGTT	Cloning of PIIN_08721 into the GoGFP vector using ClaI (constitutive promoter) or ClaI and ApaI (native promoter)
88	08721nLocConP_F	GCATCATCGATATGACTGCATTCGTCTCTCT	
89	08721nLoc_R	GCATCATCGATGTAAAAGGTTGACCAAGGGA	
90	GFP_579_F	CCGAGTAGATGCGTCTGAAA	Sequencing primer for cloning verification into the GoGFP vector
91	GFP_1125_R	GAAGTCGTGCTGCTTCATGT	

4.2. Strains and vectors

4.2.1. *E. coli* strains

For all bacterial experiments described in this thesis, the *E. coli* strain Top10 (Invitrogen, Karlsruhe) was used. This strain derived from the *E. coli* K12 strain and has the following genotype: F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(StrR) endA1 λ-.

4.2.2. *P. indica* and *S. vermifera* strains

The fungal strains used in this thesis as well as the strains generated during this thesis are listed in table 5.2.

Table 5.2: Used and generated fungal strains

Organism	Strain	Genotype	Resistance[1]	Reference
P. indica	DSM11827	-	-	Deutsche Sammlung von Mikroorganismen und Zellkulturen (Braunschweig)
P. indica	GoGFP	pGPD::oGFP-tnos, pTEF::HYG-tnos	amp, hyg	(Hilbert et al., 2012)
P. indica	c2172:GoGFP	pGPD::PIIN02172: :oGFP-tnos, pTEF::HYG-tnos	amp, hyg	this thesis
P. indica	c8721:GoGFP	pGPD::PIIN08721: :oGFP-tnos, pTEF::HYG-tnos	amp, hyg	this thesis
P. indica	n8721:GoGFP	p08721::PIIN0872 1::oGFP-tnos, pTEF::HYG-tnos	amp, hyg	this thesis
S. vermifera	MAFF305830	-	-	National Institute of Agrobiological Sciences, Tsukuba, Japan

[1]Ampicillin (amp), Hygromycin (hyg)

4.2.3. *A. thaliana* and *H. vulgare* strains

Plant microbe interaction studies were performed using the plants species listed in table 5.3

Table 5.3: Used plant species

Organism	Strain	Reference
Arabidopsis	*Arabidopsis thaliana* Ecotype Columbia-0	LEHLE SEEDS (Round Rock, USA)
Barley	*Hordeum vulgare* L. cv. Golden promise	Prof. Diter von Wettstein (Washington State University)

4.2.4. Vectors

All vectors described in this thesis contain an ampicillin resistance cassette as selection marker for transformed *E. coli* cells and a hygromycin resistance cassette as selection marker for *P. indica* transformants. Cloning steps for vector construction were performed using restriction enzymes, Antarctic Phosphatase and T4-DNA ligase from New England Biolabs (NEB, Frankfurt) and the *E. coli* strain Top10 from Invitrogen. All PCR reactions were performed using a proofreading polymerase (Pfu, Promega or Easy-A, Agilent). The success of cloning was confirmed by colony PCR (Taq polymerase, Fermentas) and sequencing (Eurofins MWG Operon).

4.2.4.1. Vectors for TA cloning of PCR products

pCRII-TOPO (Invitrogen, Karlsruhe): Vector used for cloning of PCR products. The vector is linearized and contains 3'-thymidine overhangs with covalently bound Topoisomerase I to aid in straight forward insertion of a DNA fragment with 3'-adenine overhangs. Inserted DNA segments can be cut out with EcoRI. Verification of the inserted DNA can be performed by sequencing using the primers M13 or M13Rev (Table 5.1). The vector possesses an ampicillin and a kanamycin resistance cassettes as selection markers for *E. coli* transformation.

4.2.4.2. *P. indica* transformation vectors

pGOGFP: Vector used for localization studies in *P. indica* as described in (Hilbert et al., 2012). Genes can be cloned with ClaI or HindIII to the N-terminus of a green fluorescent protein (GFP) that is codon optimized for an expression by *P. indica*. Expression of the fusion construct is under the control of the constitutively active promoter from the Glycerol-3-phosphate dehydrogenase (GPD) gene from *P. indica* and the nos terminator from *A. tumefaciens*.

pGPD::2172::GOGFP: Vector used for localization studies of the secreted chitin deacetylase *PiChiDe1* (PIIN_02172) in *P. indica*. The gene was amplified from cDNA by PCR and cloned into the pGOGFP vector by using ClaI.

pCRII-TOPO::8721: Vector containing the *PiLYSM1* (PIIN_08721) gene. Cloning was performed by PCR amplification from cDNA, subsequent A-tailing and insertion into the pCRII-TOPO vector.

pGPD::8721::GOGFP: Vector used for localization studies of the secreted chitin binding lectin-like protein *PiLysM1* (PIIN_08721) in *P. indica*. The gene was PCR amplified from the pCRII-TOPO::8721 vector and cloned into the pGOGFP vector by using ClaI.

p08721::8721::GOGFP: Vector used for localization studies of the secreted chitin binding lectin-like protein *PiLysM1* (PIIN_08721) in *P. indica*. The gene including its estimated promoter was PCR amplified from genomic DNA and cloned into the pGOGFP vector by using ApaI and ClaI.

4.3. Bacterial, fungal and plant cultivation

4.3.1. *E. coli* cultivation

Bacterial cell cultures were propagated at 37°C using dYT medium (Sambrook et al., 1989) (Table 5.4). Ampicillin (100 µg/ml) was supplemented as selection marker for transformed strains. Liquid cultures were propagated in baffled flasks at 200 rpm. For long-term storage, an over-night culture was mixed 1:1 with 100% sterile glycerol and stored at -80°C.

Table 5.4: Bacterial growth media

Medium / supplements	Ingredient / Final concentration
dYT medium	1.6% (w/v) tryptone
	1% (w/v) yeast extract
	0.5% (w/v) NaCl
	1.3% (w/v) agar (for solid medium)
	autoclaved in ddH_2O
Selection marker	100 µg/ml Ampicillin

4.3.2. *P. indica* and *S. vermifera* cultivation

Piriformospora indica cultures were propagated at 28°C using modified complete medium (CM, Table 5.5) (Pham et al., 2008). Hygromycin B (80 µg/ml) was supplemented as

selection marker for transformed strains. Liquid cultures were propagated in baffled flasks at 130 rpm. *Sebacina vermifera* cultures were propagated at 25°C using MYP medium (Table 5.5). Liquid cultures were propagated in baffled flasks at 120 rpm. For long-term storage, spores and mycelium pieces were scratched from 3 to 4 weeks old culture plates, washed once with sterile tween water (0.002% Tween20 in ddH_2O), mixed 1:1 with 100% sterile glycerol and stored at -80°C.

For microarray experiments, *P. indica* was cultivated on 1/10 PNM (Basiewicz et al., 2012) (Table 5.5) in a Conviron growth chamber under controlled conditions (day: 16 h, 23°C, 108 µmol m^{-2} s^{-1} / night: 8 h, 18°C) for up to 14 days.

To test preferred usage of different nitrogen sources of *P. indica* and *S. vermifera*, fungal cultures were cultivated at 25°C using minimal medium (MM) supplemented with different nitrogen sources with and without glucose (Table 5.5).

Table 5.5: Fungal growth media

Medium / supplements	**Ingredient / Final concentration**
Complete medium (CM)	2% (w/v) glucose
	0.2% (w/v) peptone
	0.1% (w/v) yeast extract
	0.1% (w/v) casamino acids
	1x salt solution
	1x microelements
	1.5% (w/v) agar (for solid medium)
	autoclaved in ddH_2O
20x salt solution	12% (w/v) $NaNO_3$
	1.04% (w/v) KCl
	1.04% (w/v) $MgSO_4$ x $7H_2O$
	3.04% (w/v) KH_2PO_4
	autoclaved in ddH_2O
1000x microelements	0.6% (w/v) $MnCL_2$ x $4H_2O$

	0.15% (w/v) H_3BO_3
	0.265% (w/v) $ZnSO_4$ x $7H_2O$
	0.075% (w/v) KI
	0.00024% (w/v) Na_2MO_4 x $2H_2O$
	0.013% (w/v) $CuSO_4$ x $5H_2O$
	autoclaved in ddH_2O
Selection marker	80 µg/ml Hygromycin B
MYP medium	0.7% (w/v) malt extract
	0.1% (w/v) peptone
	0.05% (w/v) yeast extract
	1.2% (w/v) agar (for solid medium)
	autoclaved in ddH_2O
1/10 PNM	0.1% (v/v) 500 mM KNO_3
	0.1% (v/v) 367 mM KH_2PO_4
	0.1% (v/v) 144 mM K_2HPO_4
	0.1% (v/v) 2 M $MgSO_4$ x H_2O
	0.1% (v/v) 200 mM $Ca(NO_3)_2$
	0.1% (v/v) Fe-EDTA
	0.1% (v/v) 428 mM NaCl
	0.4% (w/v) Gelrite
	autoclaved in dH_2O
Fe-EDTA	2.5 g $FeSO_4$ x $7H_2O$
	3.6 g Na_2EDTA
	dissolved in 400 ml ddH_2O, briefly boiled and cooled down, afterwards filled up to 450 m1 with ddH_2O
Minimal medium (MM)	60% (v/v) ddH_2O
	2% (w/v) agar (for solid medium)

	autoclaved
	10% (v/v) 10x YNB (Yeast Base w/o amino acids and ammonium sulfate (Difco); filter sterilized)
	10% (v/v) K-P-Buffer (pH 6.0); filter sterilized
	10% (v/v) carbon source solution
	10% (v/v) nitrogen source solution
Supplemented carbon source	20% (w/v) glucose
	ddH_2O (= w/o C-source)
Supplemented nitrogen source	20 mM $NaNO_3$
	20 mM NH_4SO_4
	10 mM Glutamine
	20 mM Glutamic acid
	10 mM Asparagine
	20 mM Aspartic acid
	20 mM Glycine
	ddH_2O (= w/o N-source)

4.3.3. *A. thaliana* and *H. vulgare* cultivation

Seeds from *Arabidopsis thaliana* were surface sterilized by incubating the seeds in 70% ethanol for 5 min and 6% sodium hypochlorite for 15 minutes under continuous shaking on an orbital shaker. Seeds were afterwards washed 6 times for 5 min with sterile ddH_2O. Surface sterilized seeds were stored on sterile wet filter paper at 4°C in the dark for 3 days to break dormancy (stratification) and then germinated on 1/10 PNM (Table 5.5) in a Conviron growth chamber under controlled conditions (day: 16 h, 23°C, 108 µmol m^{-2} s^{-1} / night: 8 h, 18°C) for 14 days. Seedlings of roughly the same size were afterwards transferred to the upper quarter of fresh 1/10 PNM in square petri dishes. For colonization studies plant roots were then inoculated with *P. indica* or *S. vermifera* as described in chapters 5.4.5 and 5.4.6 and further grown as described before for up to 14 days.

Seeds from barley (*Hordeum vulgare*) were surface sterilized by incubating the seeds in 70% ethanol for 1 min and 12% sodium hypochlorite for 1.5 hours under continuous shaking on an orbital shaker. Seeds were afterwards washed 3 times for 1 hour with sterile ddH_2O. Surface sterilized seeds were germinated at room temperature in the dark for 3 days on sterile wet filter paper. For colonization studies the plant roots were inoculated with *P. indica* as describe in chapter 5.4.4 and grown in a Conviron growth chamber under controlled conditions (day: 16 h, 22°C, 108 µmol m^{-2} s^{-1} / night: 8 h, 18°C) for up to 14 days.

4.4. Microbiological and biochemical methods

4.4.1. Heat shock transformation of *E. coli*

Preparation of chemical competent cells and chemical transformation of *E. coli* succeeded according to a modified protocol of (Cohen et al., 1972). An overnight culture of *E. coli* cells was diluted 1:100 to an $OD_{600} \approx 0.5$ with LB medium (Table 5.6) supplemented with 10 mM $MgCl_2$ and 10 mM $MgSO_4$ and again grown overnight at 200 rpm (37°C). Next day, the culture was centrifuged at 3000 rpm (4°C) for 15 min (Sorvall RC 5B Plus). The supernatant was discarded and the cell pellet resuspended in 33 ml of pre-chilled RF1 solution (Table 5.6) followed by incubation at 4°C for 30 to 60 min. Afterwards, the suspension was centrifuged again at 3000 rpm (4°C) for 15 min. The supernatant was discarded and the cell pellet resuspended in 5 ml of pre-chilled RF2 solution (Table 5.6) followed by incubation on ice for 15 min. The suspension of now competent cells was snap frozen in 30 µl aliquots in liquid nitrogen and stored at -80°C.

For the transformation of chemical competent *E. coli* cells, a concentrated aliquot was thawed on ice for 2 min and then diluted 1:10 with RF2 solution (Table 5.6). For each transformation, 50 µl of diluted competent cells were mixed with 2 µl ligation reaction, gently mixed and incubated on ice for 15 to 30 min. *E. coli* cells were then "heat shocked" at 42°C for 45 sec and immediately chilled down on ice for 2 min. For the regeneration of transformed cells, 500 µl dYT medium (Table 5.4) without antibiotic was added and the cells were incubated at 37°C for 30 min. Finally, 50 µl of *E. coli* cell suspension was streaked out on a dYT agar plate supplemented with 100 µg/ml ampicillin. The rest of the suspension was briefly centrifuged, resuspended in 50 µl dYT and streaked out on a second dYT agar plate supplemented with 100 µg/ml ampicillin. All plates were then incubated at 37°C over night.

Table 5.6: Solution and media for heat shock transformation of *E. coli*

Medium / supplements	Ingredient / Final concentration
RF1 solution	100 mM RbCl
	50 mM $MnCl_2$ x $4H_2O$
	30 mM potassium acetate
	10 mM $CaCl_2$ x $2H_2O$
	15% (v/v) glycerol
	pH 5.8 (adjusted with NaOH)
	Filter sterilized
RF2 solution	10 mM MOPS
	10 mM RbCl
	75 mM $CaCl_2$ x $2H_2O$
	15% (v/v) glycerol
	pH 5.8 (adjusted with NaOH)
	Filter sterilized
LB medium	1% (w/v) peptone
	0.5% (w/v) yeast extract
	0.5% (w/v) NaCl
	1.5% (w/v) agar (for solid medium)
	pH 7.0 (adjusted with NaOH)

4.4.2. Collection of *P. indica* chlamydospores

Chlamydospores from *P. indica* were collected from 3 to 4 weeks old culture plates using sterile tween water (0.002% Tween20 in ddH_2O). Spores and mycelium pieces were thoroughly scratched from the plates and then filtered through Calbiochem® Miracloth. While the mycelium stays in the filter, chlamydospores were collected in a sterile falcon tube, centrifuged at 3500 rpm (RT) for 7 min and washed three times with sterile tween water. For

plant interaction studies, the spore concentration was adjusted to 5 x 10^5 chlamydospores per ml using a Neubauer improved counting chamber (Marienfeld-Superior, Lauda Königshofen).

4.4.3. Preparation of *S. vermifera* inoculation culture

Crushed mycelium of *S. vermifera* was used for plant interaction studies. For this, 200 ml MYP medium (Table 5.5) was inoculated with 10 plugs from a 4 weeks old *S. vermifera* MYP agar plate and the culture then cultivated under continuous shaking at 120 rpm (25°C) for 7 days. The mycelium was crushed for 16 sec in a sterile blender (Microtron MB 550, Kinematica AG) and afterwards washed 3 times with 0.9% NaCl through Calbiochem® Miracloth.

4.4.4. *P. indica* - barley interaction studies

For the interaction studies of *P. indica* with barley plants, experiments were performed under sterile conditions using surface sterilized and germinated barley seeds (Chapter 5.3.3). Three days old germlings were dipped for 30 minutes in a chlamydospores suspension (0.75 ml of 5 x 10^5 chlamydospores per ml per germling) and transferred into sterile jars containing 1/10 PNM. Afterwards, 1 ml of chlamydospores suspension was added additionally on the medium in the proximity of the roots. As a control, germlings were treated with sterile tween water. Inoculated barley germlings were cultivated in a Conviron growth chamber under controlled conditions (day: 16 h, 22°C, 108 µmol m^{-2} s^{-1} / night: 8 h, 18°C) for up to 14 days. Root samples were collected after 1, 3, 7 and 14 dpi and carefully washed in sterile ddH_2O. For all experiments described in this thesis, only the stronger colonized differentiation zone (up to 4 cm below the seed) was used and immediately frozen in liquid nitrogen, while the elongation and meristematic zone (the rest of the root including the root tip) was discarded. All experiments were prepared in 3 to 4 technical replicates and 2 to 3 independent biological repetitions.

4.4.5. *P. indica* - Arabidopsis interaction studies

For the interaction studies of *P. indica* with Arabidopsis plants, experiments were performed under sterile conditions using surface sterilized and germinated Arabidopsis seeds (Chapter 5.3.3). 20 to 25 fourteen days old germlings of roughly the same size were transferred to the upper quarter of a square petri dish containing 1/10 PNM. 1 ml of chlamydospores suspension (5 x 10^5 chlamydospores per plate) was directly applied on the roots of the transferred plants. Tween water was used for control plants. The plates were left for drying for 1 to 2 h and the

inoculated plants then cultivated in a Conviron growth chamber under controlled conditions (day: 16 h, 23°C, 108 µmol m^{-2} s^{-1} / night: 8 h, 18°C) for up to 14 days. In order to prevent root growth to the upper quarter of the petri dish, the plates were arranged upwards in a 45° angle. Root samples were collected after 3, 7 and 14 dpi, carefully washed in sterile ddH_2O and immediately frozen in liquid nitrogen. All experiments were prepared in 3 to 4 technical replicates (60 - 100 plants in total) and 3 independent biological repetitions.

4.4.6. *S. vermifera* - Arabidopsis interaction studies

For the interaction studies of *S. vermifera* with Arabidopsis plants, plants were prepared as described before (Chapter 5.4.5). *S. vermifera* inoculation culture (Chapter 5.4.3) was afterwards directly applied on the roots of the transferred plants. The plates were left for drying for 1 to 2 h and the inoculated plants then cultivated in a Conviron growth chamber under controlled conditions (day: 16 h, 23°C, 108 µmol m^{-2} s^{-1} / night: 8 h, 18°C) for up to 14 days. In order to prevent root growth to the upper quarter of the petri dish, the plates were arranged upwards in a 45° angle. Root samples were collected after 3, 7 and 14 dpi, carefully washed in sterile ddH_2O and immediately frozen in liquid nitrogen. All experiments were prepared in 4 technical replicates (60 - 80 plants in total) and 3 independent biological repetitions.

4.4.7. *P. indica* protoplast preparation

Chlamydospores from *P. indica* were collected according to chapter 5.4.2 and adjusted to 2.5 x 10^6 chlamydospores per ml. 200 ml CM medium (Table 5.5) was inoculated with 1 ml of this suspension and the culture then cultivated under continuous shaking at 130 rpm (28°C) for 7 days. Afterwards, the culture was filtered through Calbiochem® Miracloth and the mycelium washed with 0.9% NaCl. The flow through was discarded and the mycelium crushed for 10 seconds in 60 ml fresh CM using a sterile blender (Microtron MB 550, Kinematica AG). 20 ml of the crushed mycelium was mixed with 130 ml fresh CM in a sterile baffled flask and cultivated at 130 rpm (28°C) for 3 days. The young mycelium was filtered through Miracloth, washed once with 0.9% NaCl and added into 10 to 20 ml of pre-chilled enzyme solution (Table 5.7). After protoplastation at RT for 20 to 30 min, protoplast formation was inspected using a light microscope (Axiostar, Zeiss). When amount and size of protoplasts were sufficient, the protoplastation was stopped by filtering through Miracloth followed by adding 10 to 20 ml cold STC buffer (Table 5.7) to the flow through. Subsequently, protoplasts were centrifuged at 4000 rpm (RT) for 10 min and the protoplast

pellet washed 3 times with cold STC. For transformation of *P. indica*, the protoplast concentration was adjusted to 10^7 to 10^9 protoplasts per ml using a Neubauer improved counting chamber (Marienfeld-Superior, Lauda Königshofen).

Table 5.7: Solutions for protoplastation of *P. indica*

Medium / supplements	Ingredient / Final concentration
Enzyme solution	2.5 mg/ml Novozyme (Novo Nordisc, Kopenhagen, Dänemark) in SCS
	filter sterilized (0.2 µm, Roth)
SCS	20 mM sodium citrate, pH 5.8
	1 M sorbitol
STC	10 mM Tris HCl, pH 7.5
	50 mM $CaCl_2$ x $2H_2O$
	1.33M sorbitol

4.4.8. PEG-mediated transformation of *P. indica* protoplasts

Transformation of *P. indica* was performed according to (Hilbert et al., 2012). For this, 8 to 10 µg of the vector to be transformed were linearized with 20 units of an appropriate restriction enzyme for 2 h and gel-purified using the PROMEGA Wizard® SV Gel and PCR Clean-Up System kit. Afterwards, 70 µl of freshly prepared protoplasts (Chapter 5.4.7) were mixed with the linearized and cleaned vector and 1 µl heparin (15 mg/ml). The mixture was incubated on ice for 10 min, 500 µl of STC/PEG solution (Table 5.8) and 20 units of the restriction enzyme used for linearizing the vector added and the mixture then incubated on ice for another 15 min. 5 ml of fresh but pre-chilled top agar (Table 5.8) was added, gently mixed and poured on plates containing 20 ml solidified bottom agar (Table 5.8) and the transformants then cultivated at 28°C. After 10 to 14 days, growing transformants were transferred onto CM plates (Table 5.5) containing 80 µg/ml hygromycin.

Table 5.8: Regeneration agar for *P. indica* transformants

Medium / supplements	Ingredient / Final concentration
STC/PEG solution	40% (v/v) PEG4000 in STC (Table 5.7)

Regeneration agar - Bottom agar	MYP medium (Table 5.5)
	0.3 M sucrose
	1.2% agar
	100 µg/ml Hygromycin B
Regeneration agar - Top agar	MYP medium (Table 5.5)
	0.6% agar

4.5. Molecular methods

4.5.1. DNA isolation

Harvested fungal material from pure cultures was thoroughly grounded in liquid nitrogen with mortar and pestle. About 100 to 200 mg from the resulting powder was mixed with 1 ml of extraction buffer (Table 5.9) in a 2 ml Eppendorf tube and incubated under gentle shaking (RT) for 10 minutes. After the addition of 1 ml CIA (chloroform:isoamylalkohol in a 24:1 relation) and another incubation under gentle shaking (RT) for 5 min, samples were centrifuged at 13000 rpm (RT) for 20 min. The aqueous phase was transferred into a new 2 ml Eppendorf tube and 0.2 reaction volumes of 100% ethanol were carefully added. After incubation under continuous shaking for 5 min, 1 reaction volume of CIA was added and the mixture incubated for another 5 min followed by centrifugation at 13 rpm (RT) for 20 min. The aqueous phase was again transferred into a new 2 ml Eppendorf tube, 500 µl isopropanol was added and the DNA precipitated overnight at 4°C. The next day, samples were centrifuged at 13000 rpm (4°C) for 30 min and DNA pellets washed with 900 µl cold 70% ethanol. After another centrifugation at 13000 rpm (4°C) for 15 min, the ethanol was removed, DNA pellets air-dried for 5 min and then dissolved in 50 µl sterile TE buffer (pH 8.0) (Table 5.9) or ddH_2O. RNA was digested by adding 1 µl of RNase A (10 mg/ml, pH 7.4) and incubation at 37°C for 30 to 60 min. DNA quality and quantity was confirmed on an 1% agarose gel and with a NanoDrop-1000 Spectrophotometer.

For DNA isolation from plants inoculated with either *P. indica* or *S. vermifera*, a modified protocol from (Doyle and Doyle, 1987) was used. For this, harvested material from inoculated plants was thoroughly grounded to a fine powder in liquid nitrogen with mortar and pestle. 100 mg of grounded material was mixed with 700 µl of pre-warmed (65°C) D&D extraction

buffer (Table 5.9) in a 2 ml Eppendorf tube and incubated under gentle shaking (65°C) for 30 min. After the addition of 0.7 ml CIA (chloroform:isoamylalkohol in a 24:1 relation) and an incubation under gentle shaking (RT) for 5 min, samples were centrifuged at 13000 rpm (4°C) for 20 min. The aqueous phase was transferred into a new 2 ml Eppendorf tube and mixed with another 0.6 ml of CIA followed by another incubation under gentle shaking (RT) for 5 min and centrifugation at 13000 rpm (4°C) for 10 min. The aqueous phase was again transferred into a new 2 ml Eppendorf tube, mixed with 50 µl 10 M ammonium acetate and 60 µl 3 M sodium acetate (pH 5.5) an incubated under gentle shaking (RT) for 5 min. Afterwards, 500 µl isopropanol was added and DNA precipitated overnight at 4°C. The next day, samples were centrifuged at 13000 rpm (4°C) for 30 min and DNA pellets washed two times with 500 µl 10 mM ammonium acetate (dissolved in 70% ethanol). After another centrifugation at 13000 rpm (4°C) for 15 min, the ethanol was removed, DNA pellets air-dried for 5 min and then dissolved either in 50 µl sterile TE buffer (pH 8.0) (Table 5.9) or ddH_2O. RNA was digested by adding 1 µl of RNase A (10 mg/ml, pH 7.4) and incubation at 37°C for 30 to 60 min. DNA quality and quantity was confirmed on an 1% agarose gel and with a NanoDrop-1000 Spectrophotometer.

Table 5.9: Buffers and solutions for DNA extraction

Medium / supplements	**Ingredient / Final concentration**
Extraction buffer	100 mM Tris-HCl (pH 7.5)
	50 mM EDTA (pH 8.0)
	1.5 M NaCl
	2% (w/v) CTAB
	0.05% (v/v) ß-mercaptoethanol (added just before use)
D&D extraction buffer	100 mM Tris-HCl (pH 8.0)
	20 mM EDTA (pH 8.0)
	1.4 M NaCl
	2% (w/v) CTAB
	1% (w/v) $Na_2S_2O_5$
	0.2% (v/v) ß-mercaptoethanol (added just before use)

1 M Tris-HCl (pH 8.0)	8.88% (w/v) Tris-HCl
	5.3% (w/v) Tris-Base
	pH adjusted with HCl
0.5 M EDTA (pH 8.0)	18.21% Na_2EDTA x $2H_2O$
	pH adjusted with NaOH
TE buffer (pH 8.0)	10 mM Tris-HCl (pH 8.0)
	1 mM EDTA (pH 8.0)

4.5.2. RNA isolation

Total RNA from fungal and plant material was extracted using TRIzol (Invitrogen, Karlsruhe). For this, 100 mg of grounded frozen material was mixed with 1 ml TRIzol in a 2 ml Eppendorf tube and immediately vortexed on a vibrax shaker at 1500 rpm for 5 min. Afterwards, 200 µl chloroform was added, the samples were briefly vortexed for 20 sec and then centrifuged at 13000 rpm (4°C) for 30 min. 500 µl of the upper/aqueous phase was transferred into a new 2 ml Eppendorf tube and RNA precipitated for 1 to 12 h at 4°C by adding 500 µl isopropanol. The RNA was pelleted by centrifugation at 13000 rpm (4°C) for 30 min, washed with 1 ml of 75% ethanol (diluted in DEPC treated ddH_2O) and again centrifuged at 13000 rpm (4°C) for 15 min. The ethanol was removed, RNA pellets air-dried for 5 min and then dissolved in 50 µl RNase-free water. Amount and quality of extracted RNA was estimated with a NanoDrop-1000 Spectrophotometer and on a 1% agarose gel. For microarray and RNAseq experiments, the RNA was additionally purified using the clean-up protocol from the RNeasy Plus Universal Kit (Qiagen) and the RNA integrity confirmed using a Bioanalyzer 2100 (Agilent, Santa Clara, USA).

4.5.3. Polymerase chain reaction (PCR)

Amplification of DNA fragments was performed for the verification of gene structures, cloning of coding sequences and verification of cloning success in a T-professional Basic Thermocycler (Biometra). A proof reading polymerase (Pfu or Easy-A) was used if the amplified DNA fragment was used for subsequent cloning in a protocol like described in table 5.10.

Table 5.10: Standards for PCR experiments with a proofreading polymerase

PCR reaction mix (50 µl)	x µl genomic DNA or cDNA (10 to 100 ng)			
	5 µl 10x Pfu buffer containing $MgCl_2$ (Promega)			
	5 µl 2.5 mM dNTPs (Fermentas)			
	1 µl forward primer (10 pM)			
	1 µl reverse primer (10 pM)			
	0.5 µl Pfu polymerase (Promega)			
	(37.5 – x) µl ddH_2O			
PCR protocol	initial denaturation	95°C	5 min	
	denaturation	95°C	30 sec	30 cycle
	Annealing	55 – 65°C	45 sec	
	Elongation	72°C	2 – 3 min	
	final elongation	72°C	8 min	

In all other cases, the Taq polymerase was used for the amplification in a protocol like described in table 5.11.

Table 5.11: Standards for PCR experiments with the Taq polymerase

PCR reaction mix (25 µl)	x µl genomic DNA or cDNA (10 to 100 ng)			
	2.5 µl 10x Taq buffer containing $(NH_4)_2SO_4$			
	2.5 µl 2.5 mM dNTPs (Fermentas)			
	2 µl $MgCl_2$ (Fermentas)			
	0.5 µl forward primer (10 pM)			
	0.5 µl reverse primer (10 pM)			
	0.625 µl Taq polymerase (Fermentas)			
	(16.375 – x) µl ddH_2O			
PCR protocol	initial denaturation	95°C	5 min	
	denaturation	95°C	30 sec	35 cycle

	annealing	55 – 65°C	45 sec	
	elongation	72°C	1 – 2 min	
	final elongation	72°C	8 min	

After amplification, PCR products were separated and visualized by electrophoresis in an agarose gel (Chapter 5.5.5).

4.5.4. Quantitative real-time PCR

For a quantitative analysis of gene expression and colonization efficiency, quantitative real-time polymerase chain reactions (qRT-PCR) were performed. In case of gene expression analysis, 1 µg RNA (Chapter 5.5.2) were reverse transcript to cDNA using the First Strand cDNA synthesis kit (Fermentas, St. Leon-Rot). The qRT-PCR was then performed with either 10 ng cDNA or genomic DNA (Chapter 5.5.1) in the SYBR green Supermix kit (Bio-Rad) according to the manufacturers instruction (Table 5.12). 100 nM Fluoresceinisothiocyanat (FITC) was used as a reference dye and the reaction was performed in a Bio-Rad iCycler system using the amplification protocol listed in table 5.12. Relative expression and relative amounts of DNA, respectively, were calculated according to the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

Table 5.12: Standards for quantitative real-time PCR experiments

qRT-PCR reaction mix (20 µl)	4 µl cDNA / DNA (2.5 ng/µl)			
	10 µl SYBR green mix			
	0.2 µl forward primer (10 pM)			
	0.2 µl reverse primer (10 pM)			
	0.2 µl FITC (1µM)			
	3.4 µl ddH_2O			
qRT-PCR protocol	initial denaturation	95°C	5 min	
	denaturation	95°C	30 sec	
	annealing	59°C	30 sec	40 cycle
	elongation	72°C	30 sec	

	final elongation	72°C	1 min
	melt curve analysis (60 to 80°C in 100 steps)		

4.5.5. Agarose gel electrophoresis

The agarose gel electrophoresis was used for the separation of PCR amplified DNA fragments (Chapter 5.5.3) or to assess the quality of isolated total DNA or RNA (Chapter 5.5.1 and 5.5.2). The agarose concentration for isolated total DNA or RNA was always 1% and varied from 1% to 2% for PCR amplified DNA fragments depending on their expected size. For DNA, the appropriate amount of agarose (Biozym) was dissolved in 1x TAE or 0.5x TBE buffer (Table 5.13) by cooking and cooled down to 60°C. For RNA, the agarose was dissolved in water, cooled down to 60°C and then mixed 1:1:8 with 37% formaldehyde and 10x MOPS buffer (Table 5.13).

Table 5.13: Buffers for agarose gels

Medium / supplements	Ingredient / Final concentration
50x TAE buffer	2 M Tris-HCl
	2 M acetic acid
	50 mM EDTA, pH 8.0
5x TBE buffer	440 mM Tris-HCl
	440 mM H_3BO_3
	10 mM EDTA, pH 8.0
10x MOPS buffer	200 mM MOPS
	50 mM sodium acetate
	10 mM EDTA
	prepared with DEPC-treated water

For the visualization of either DNA or RNA, ethidium bromide (0.5 µg/ml; Roth) was added in a 1:10000 dilution and the gel poured into appropriate gel casts. Separation took place in an electrophoresis chamber filled with the same buffer as used for the gel. Prior to the gel run,

DNA and RNA samples were then mixed with an appropriate loading dye (Table 5.14) separated at 120 V for 40 to 60 min and at 130 V for 15 min, respectively.

Table 5.14: Used DNA and RNA loading dyes

Medium / supplements	Ingredient / Final concentration
6x DNA loading dye	40% (w/v) sucrose
	0.25% (w/v) bromphenol blue
	Dissolved in ddH_2O
6x RNA loading dye	50% (w/v) sucrose
	0.25% (w/v) bromphenol blue
	0.25% (w/v) xylencyanol FF
	Dissolved in 1x MOPS buffer (Tab)

Afterwards, bands were visualized under UV light and pictures were taken using an UV solo TS imaging system.

4.5.6. TA cloning of PCR products

Genes that should be cloned into more than one destination vector, were initially cloned in the pCRII-TOPO vector (Invitrogen) according to the manufactures protocol (Table 5.15). The vector is linearized and has a single 3´-terminal thymidine overhang as well as covalently bound topoisomerase I at both ends. This allows a very fast insertion if the fragment to be integrated contains compatible 3' deoxyadenosine overhangs as produced by either Taq or Easy-A polymerase.

Table 5.15: Standards for cloning

Medium / supplements	Ingredient / Final concentration
pCRII-TOPO cloning	4 µl PCR product
	1 µl salt solution
	1 µl TOPO vector

After gentle mixing of the components, ligation was performed at room temperature for 5 to 15 min, subsequently transformed into competent *E. coli* cells as described in chapter 5.4.1. Successful gene insertion was tested with grown colonies by PCR amplification with the M13 forward primer and a gene specific reverse primer (Table 5.1). Correctness of the inserted sequence was verified by sequencing.

4.5.7. Southern blot

The PEG mediated transformation of *P. indica* (Chapter 5.4.8) results in a random integration of the vector into the genome of *P. indica*. Besides the continuous growth of transformants on selection medium, stable genome integration of the transformed vector was verified by southern blot analysis. DNA from a fungal pure culture was isolated (Chapter 5.5.1.) and 30 to 50 µg digested with an appropriate enzyme overnight in a total volume of 100 µl. Digestion was confirmed on an 1% agarose gel and the DNA then concentrated to 3-5 µg/µl using sodium acetate precipitation. For this, the DNA was precipitated with 0.1 reaction volumes of 3 M sodium acetate and 2.5 reaction volumes of isopropanol for 10 min on ice. After centrifugation at 13000 rpm (4°C) for 10 min, the pellet was washed once with 70% ethanol and centrifuged again at 13000 rpm (4°C) for 10 min. The ethanol was removed and the pellet air-dried for 5 min. The DNA pellet was dissolved in ddH_2O, loading dye (Table 5.14) was added and loaded on a 0.8% agarose gel (1x TAE) and run at 80V for 2.5 to 3 h. Afterwards, the DNA was depurinated in 0.25 N HCl and subsequently denatured in 0.4 M NaOH under gentle shaking for 15 min each. Transfer of DNA to a nylon membrane (Amersham Biosciences Hybond-N+, GE Healthcare) was performed overnight. For this, a filter paper bridge (Whatman paper) was placed on a glass slide with its ends dipping in 0.4 M NaOH solution. Carefully preventing the introduction of air bubbles between the layer, the agarose gel, the nylon membrane and another 3 pieces of filter paper were arrange on top of the bridge. To increase capillary flow, a thick layer of absorbent papers were put on top and everything pressed together with a ~3 kg weight. After blotting over night, the DNA was cross-linked to the membrane in an UV Stratalinker 1800 (Stratagene) using an auto-crosslinking program (1200 x 100 µJoule). The membrane was then pre-incubated with 20 to 30 ml southern hybridization buffer (Table 5.16) in a hybridization tube at 65°C for 0.5 to 2 h followed by hybridization for 3 days with DIG labelled probes.

4.5.8. DIG labelling

Probes were prepared by denaturation of 1 to 3 µg DNA template (cleaned PCR product) at 95°C for 10 min in a final volume of 16 µl followed by fast chilling on ice. Afterwards, 4 µl of DIG-High Prime (Roche) were mixed with the denatured DNA and incubated at 37°C for 3 to 20 h. Labelling was stopped by the addition of 2 µl 0.2 M EDTA (pH 8) and incubation at 65°C for 10 min. The probe was then again denatured at 95°C for 10 min, added into 30 to 50 ml of pre-warmed hybridization buffer and used for hybridization at 65°C overnight. After hybridization, the membrane was washed twice with Southern wash buffer at 65°C for 20 min and then equilibrated in 30 ml of DIG-wash buffer under gentle shaking (RT) for 5 min. Blocking was performed in 30 ml of freshly prepared DIG II buffer at RT for 30 min. Afterwards, the membrane was incubated in antibody solution under gentle shaking (RT) for 30 min, followed by washing twice with 30 ml DIG-wash buffer for 45 min. The membrane was equilibrated in 30 ml DIG III buffer at RT for 5 min and then incubated with CDP-Star-solution at RT for 5 min. Tightly wrapped in a plastic bag, the membrane was incubated in the dark at 37°C for 15 min and subsequently used for signal detection. Under red light, a x-ray film (CEA RP New medical X-ray screen film blue sensitive) was placed on top of membrane in a Roentgen-cassette and incubated for 30 sec to 60 min depending on the signal intensity. The film was developed in a Fuji medical Film processor FPM-100A.

Table 5.16: Buffers and solutions for southern blots

Medium / supplements	**Ingredient / Final concentration**
1 M Na-P-Buffer (pH 7)	61% (v/v) 1 M Na_2HPO_4
	39% (v/v) 1 M NaH_2PO_4 x H_2O
Southern hybridization buffer	0.5 M Na-P-Buffer (pH 7)
	7% (w/v) SDS
	stored at 37°C
Southern wash buffer	0.1 M Na-P-Buffer (pH 7)
	1% (w/v) SDS
DIG I buffer (pH 7.5)	0.1 M maleic acid
	0.15 M NaCl
	pH was adjusted with NaOH

DIG wash buffer	0.3% (v/v) Tween20 in DIG I buffer
Blocking solution (stock)	10% (w/v) blocking reagent (Roche) in DIG I buffer (pH 7.5)
	stored at 4°C
DIG II buffer	10% (v/v) Blocking solution (stock)
	90% (v/v) DIG I buffer (pH 7.5)
DIG III buffer (pH 9.5)	0.1 NaCl
	0.05 M $MgCl_2$ x $6H_2O$
	pH was adjusted with Tris-HCl
Antibody solution	1 µl Anti-DIG-AP (Roche) in 10 ml DIG II buffer
CDP-Star solution	100 µl CDP-Star (Roche) in 10 ml DIG III buffer

4.5.9. Microarray experiments

4.5.9.1. Experimental design

Microarray analysis were used for whole transcriptome expression analysis in three connected experiments addressing different questions. Because *P. indica* is able to colonize living plant roots as well as dead plant material, the *P. indica* - barley experiments addressed the fungal gene expression in these two environments. Therefore, living and dead (autoclaved) barley roots were inoculated with *P. indica* as described in chapter 5.4.4. *P. indica* from cultures grown in liquid CM or 1/10 PNM (Chapter 5.3.2) were used as controls. For all four conditions plant/fungal material was harvested after 24 h, 36 h, 3 d, 7 d and 14 d. *P. indica* has a very broad host range and is able to colonize monocotyledonous like barley as well as dicotyledonous like *Arabidopsis thaliana*. In order to analyse and compare the fungal gene expression in both plants, *P. indica* - Arabidopsis experiments were performed as described in chapter 5.4.5. Inoculated root material was harvested after 3 d, 7 d and 14 d. *P. indica* grown on solid 1/10 PNM (Chapter 5.3.2) for 3 d was used as fungal control and roots from tween water treated Arabidopsis plants harvested after 3 d, 7 d and 14 d were used as plant controls. For all conditions plant/fungal material was frozen in liquid nitrogen directly after harvest and stored at -80°C.

4.5.9.2. Custom design of Agilent microarray chips

The microarray design for the *P. indica* - barley experiments was performed using eArray (https://earray.chem.agilent.com/earray/). Up to six 60-mer probes were calculated with the best distribution methodology. Additionally, probes for 265 barley genes (including genes involved in defence and transport), 158 *A. tumefaciens* genes (bacterial control) and 11 *P. indica* housekeeping genes (positive control) were generated. To evaluate the hybridization efficiency within one array, probes from 10 *P. indica* genes were hybridised randomly in 10 replicates. Based on this microarray design and the resulting data, an optimized 6x80k custom-designed Agilent microarray was created for the *P. indica* - Arabidopsis experiments. For this, the JAVA programs ProbeInfo and ProbeEvaluator were used as described in chapter 5.7.1.13. New probes (NP) were calculated for those genes that had no or no suitable probe on the old design by using the eArray program or Picky (Chou, 2010) with adjusted parameters. Additionally, probes for all protein coding genes of *Arabidopsis thaliana* as described in the Tair10 genome release (http://www.arabidopsis.org/index.jsp) were generated using the eArray program. Control probes were loaded as described for the first design.

4.5.9.3. Labelling and hybridization of custom designed arrays

For preparation and hybridization of custom designed microarray chips, total RNA was extracted, purified and quality-controlled as described in chapter 5.5.2. For the *P. indica* - barley experiments, two independent biological replicates for each treatment were performed. RNA from *P. indica* colonized roots 36 and 48 hpi were pooled together and referred to as the pre-penetration sample. Two more time points were selected for the hybridization, 72 hpi (early colonization) and 120 hpi (late colonization). Further RNA from 36, 48, 72 and 120 hpi of *P. indica* grown on CM or 1/10 PNM were pooled together and used as controls. The labelling preparation was performed according to Agilent's One-Color Microarray-Based Gene Expression Analysis (Quick Amp Labeling) with Tecan HS Pro Hybridization protocol (version 6.0). Briefly, 500 ng total RNA was mixed with a "spike-in" mix containing *in vitro* synthesized 55-mer sequences from the human adenovirus type 6 E1A 13S gene as controls for linearity and sensitivity of the hybridization efficiency. cDNA was then synthesized from the total RNA and subsequently transcribed to cRNA which resulted in double stranded RNA with cyanine-3 (Cy3) labelled nucleotides. The cRNA was purified by using the RNeasy Plus Universal Kit (Qiagen) and cRNA quantity and Cy3 activity measured with a NanoDrop-1000 Spectrophotometer. Labelled cRNA was afterwards hybridised to 2×105k custom-designed Agilent microarrays according to Agilent's One-Color Microarray-Based Gene Expression

Analysis (Quick Amp Labeling) protocol (version 5.7). For the *P. indica* - Arabidopsis experiments, total RNA from *P. indica* colonized and tween water treated Arabidopsis roots 3, 7 and 14 dpi was used. Additionally, total RNA extracted from 3 days old *P. indica* hyphae grown on 1/10 PNM-agar were used as control. All samples were labelled and hybridized to 8x60k custom-designed Agilent microarray chips according to Agilent's One-Color Microarray-Based Gene Expression Analysis Low Input Quick Amp Labeling protocol (version 6.5).

4.5.9.4. Microarray analyses and verification

Microarray image files were analyzed using Agilent's Feature Extraction software v. 10.5 which calculates for each spot a background corrected signal intensity value (gProcessedSignal) that was used for further analysis. For statistical analysis of the raw data, the R environment (www.r-project.org; version 2.15.1) including the Bioconductor package 'Limma' was used in the following way. 1) Raw data were standardized by quantile normalization to compare expression values between experiments. 2) Intensity values from replicate probes were averaged and log2-ratios between experiments were calculated. 3) Quality and suitability of the applied statistics was estimated by generating density and MA plots. 4) The degree of variability between the experiments was shown by principle component analysis. 5) For the selection of differentially expressed genes, fold changes were calculated from the log2-ratios and Student's t-statistic applied to test for significance. The selection was based on a fold change of 2 and a false discovery rate-adjusted significance level (adj. p-value) of less than 0.05. The *P. indica* expression data from the *P. indica* - barley and *P. indica* - Arabidopsis experiments are stored in the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/) under the accession numbers GSE31266 and GSE201936, which complies with MIAME (minimal information about a microarray experiment) guidelines. Obtained microarray data were verified by quantitative real-time PCR (qRT-PCR) as described in chapter 5.5.4 from three biologically independent kinetics for each host (time points: 3, 7 and 14 dpi for Arabidopsis; 3, 5, 10 and 14 dpi for barley). The constitutively expressed *P. indica* gene encoding for the "translation elongation factor 1 alpha" (*PiTEF*) was used as reference.

4.5.10. DNA sequencing

Genomic DNA was extracted from fungal material grown either in CM (*P. indica*) or MYP (*S. vermifera*) liquid culture using the CTAB protocol of Doyle and Doyle (Chapter 5.5.1).

Sequencing of the genome of *P. indica* was performed by Eurofins MWG Operon, Ebersberg, Germany, using the 454 GS FLX Titanium platform. Sequencing of the genome of *S. vermifera* was performed by DOE joint genome institute, Walnut Creek (CA), USA using a combination of paired end- pyrosequencing and Illumina sequencing in a 4- and 8 kb library, respectively.

4.5.11. RNA sequencing

Total RNA was extracted from fungal material grown either in CM (*P. indica*) or MYP (*S. vermifera*) liquid culture using the TRIzol reagent (Invitrogen, Darmstadt, Germany) (Chapter 5.5.2). Sequencing of the *P. indica* transcriptome was performed as described before (Zuccaro et al., 2011). Sequencing of the transcriptome of *S. vermifera* was performed by the DOE joint genome institute, Walnut Creek (CA), USA using Illumina sequencing in a 4- and 8kb library, respectively. In order to find similar and differing mechanisms in the colonization strategy of *S. vermifera* and *P. indica*, RNA sequencing was performed. For this, *A. thaliana* roots were inoculated and grown as described in chapter 5.4.6. As a control, *S. vermifera* was grown in liquid MYP (Chapter 5.3.2). Root and fungal material was harvested after 3 d, 7 d and 14 d and extracted RNA pooled with equal volumes. RNA sequencing was then performed by IGA Technology Services Srl Parco Scientifico eTecnologico di Udine (Udine, Italy). Data were analyzed in cooperation with Dr. Annegret Kohler from the French national institute for agricultural research (Nancy, France)

4.6. Microscopy

For chlamydospore and protoplast counting, an optical microscope (Objective: A-plan 20x; Axiostar, Zeiss) was used. GFP fluorescence was analyzed using a confocal laser scanning microscope (TCS-SP5, Leica, Bensheim, Germany). The green fluorescence was excited with an argon laser (30% strength) at 488 nm and detected at 505-555 nm.

4.7. Bioinformatic methods

4.7.1. JAVA programs

For analyzing and parsing of sequences and program outputs, JAVA programs were written using JAVA version 7 (www.sun.com) and the JAVA development environment Eclipse version 4.1 (http://www.eclipse.org/; (Vaughan-Nichols, 2003)). A main principle behind JAVA is the object-oriented programming (OOP). Objects in JAVA are described in classes

and can be interpreted as computational descriptions of real-world objects. For this, an object in JAVA needs to have attributes to describe the object and methods to be able to work with or manipulate the attributes. The following sub-chapters shall briefly describe all self-written programs, their intended usage and a short description of the required in- and produced output. The source code of all programs is supplied on a CD with this thesis due to space and formatting issues.

4.7.1.1. Augustus2fasta

The output file of the gene prediction program AUGUSTUS (Stanke and Waack, 2003) contains the precise exon/intron structure as well as the nucleotide and amino acid sequence of each predicted gene. Given this file as input, the program copies the predicted coding sequence and deduced amino acid sequence and produces two new fasta formatted files.

4.7.1.2. BlastAnalyzer / BlastEntry

BLAST (Altschul et al., 1990, 1997) as a stand-alone version includes different programs with many different parameters. In order to compare Blast outputs and eventually optimize parameter settings based on this comparison, the program BlastAnalyzer summarizes the similarities of the Blast hits. For this, a tab-delimited Blast output file is needed as produced with the "-outfmt 6" parameter. Similarities are then grouped in 10% ranges and the final summary written to plain text file. By comparing these outputs, parameters could be deduced which were more suitable for individual analysis.

The BlastEntry class holds the definitions for a Blast object to be described in JAVA. These definitions include the attributes a Blast object has, like alignment length or bit score, and the methods that can be performed on it, like retrieving or setting attribute values.

4.7.1.3. Blast2go2goeast

Gene ontology (GO) enrichment analyses were performed using GOEAST (http://omicslab.genetics.ac.cn/GOEAST/index.php) which needs two files, one containing all genes with all assigned GOs and one list containing a subset genes for which the enrichment shall be calculated. The first file, must have a special format and the program Blast2go2goeast converts the GO description file produced by the Blast2GO suite into such a format.

4.7.1.4. CodonUsage

The program counts all codon triplets that occur in the sequences of a provided fasta formatted file and produces two files based on this. The first one is a codon table showing the absolute number of each codon and the relative percentage of the codon in comparison to all codons. The second file is a list of all codons required for the creation of a frequency plot using the program WebLogo (Crooks et al., 2004).

4.7.1.5. FastaLoader / FastaEntry / DNAtranslator

For easy and consistent processing of large sequence files and because present headers in downloaded genomes caused problems in many programs, a standardized fasta-format was designed. Therefore, the program FastaLoader takes a fasta formatted sequence file and an optionally supplied general feature format (gff) file and converts each fasta entry into a standardized format in four steps. 1) The sequence of each entry is concatenated, thereby reduced to one line per entry. 2) A new header is created consisting of (a) an abbreviation of the Latin name of the organism and (b) a generic number given by its position in the input file. If a gff file is provided, (c) the ID of the scaffold in which the gene was predicted and (d) the position of the gene on its scaffold are additionally added to the header. 3) If the sequences in input file are recognized as DNA sequences, these are translated into amino acid sequences 4) DNA and AA sequences are written into separate new fasta files.

The FastaEntry class holds the definitions for a Fasta object to be described in JAVA. These definitions include the attributes a Fasta object has, like the original identifier or the standardized identifier or the sequence length, and the methods that can be performed on it, like retrieving or setting attribute values.

The DNAtranslator class contains additional methods that can be performed on fasta objects. The sequence can be reversed (reverseSequence), the complementary strand may be predicted (complementarySequence), in can be transformed to contain only upper case letters (low2upSequence), a one line Fasta format can be generated (fasta2onelinefasta), a nucleotide sequence can be translated to amino acid (dna2protein) and sequences of fasta objects can concatenated (concatFasta). Additionally, a method for splitting a multiple fasta file into smaller ones (fastaSplitter) is integrated into this class.

4.7.1.6. IntergenicSpacePlot

The program counts genes with an intergenic region within a defined range. The output is a tab-delimited text file containing x, y and z coordinates that can be used as input for the IntergenicPlot R script (Chapter 5.7.2.7). The x and y values are representing the position in the 2D plot, whereas the z defines the intensity of the colour.

4.7.1.7. MicroarrayDataSelection

When only a subset of entries from a microarray file was required, the program MicroarrayDataSelection was used. Two input files are expected, the first one is a microarray file produced by the Agilent Feature Extraction software. The second one is a plain text file containing a list of IDs referring either to the ID of the probe (ProbeID) or the gene target of the probe (GeneID). Entries in the microarray file containing either the ProbeID or GeneID in the selection file are copied into the text file. If an entry in the selection lists is not found in the microarray file, a warning is printed.

4.7.1.8. MotifSeeker

In order to screen genomes for known and unknown motifs in the amino acid sequence, the program MotifSeeker was written. Given a set of protein sequences and a fixed search space within each sequence, the program either predicts new motifs or searches for a given pattern within the defined region. The latter is based on pattern finding with regular expressions were the exact regular expression pattern to search for has to be provided. The search for new motifs, in contrast, uses a simple similarity scoring for the prediction. For this, every sequence fragment to be analysed was pairwise aligned with all other fragments. A similarity score was increased by 5 for same amino acids at same positions, by 4 if the position differed by +- 1, by 3 if the position differed by +-2 and by 2 for greater differences. An additional block score of 1 was added if amino acids were matching in continuous blocks and a penalty of 2 was subtracted from the score if an amino acid did not occur in the tested sequence. The minimal score of an alignment to be predicted as a motif was set to be 7n / 2, were n is the length of the sequence fragment. The output is a list of sequences that were analyzed and eventually matching motifs in other sequences together with the similarity score and the ID of the sequence were the putative new motif was identified.

4.7.1.9. MultipleFastaSelection

The program expects two input files. The first one is a Fasta formatted sequence file. The second one is a plain text file containing (a) either only a list of IDs referring to the Fasta entries in the first file or (b) a file as described before but with two additional columns specifying the position of a subsequence within the defined fasta entry. If selection file a is provided, only those fasta entries which are listed in the selection file are copied into a new fasta file. If selection file b is provided the program creates two new fasta files. The first one contains only the specified parts from the listed sequences, the second one contains the same sub-sequences, but those belonging to the same original sequence are concatenated. If an entry in the selection lists is not found in the sequence file, a warning is printed.

4.7.1.10. MultipleSequenceAnalyzer

For each nucleotide sequence in a provided fasta formatted sequence file, the program calculates the relative number of every base in relation to the length of the sequence, but excluding unknown (N) bases. Afterwards, the sequence is translated into amino acids and the relative number of every amino acid in relation to the length of the sequence calculated, but with excluding unknown amino acids (X). Furthermore, the relative number of hydrophobic, hydrophilic, basic and acidic amino acids is given as well as the expected molecular weight of the protein. At last, the chance of getting a soluble protein when overexpressing it in *E. coli* is calculated based on the algorithm of (Koschorreck et al., 2005).

4.7.1.11. Pfam2dog / Smart2dog

In principal, the programs are doing the same but with different input files. Pfam2dog requires the output file from the PfamScan Perl script in which the fields are separated by tab stops (Chapter 5.7.3.3). Smart2dog requires the output from the SmartBatchParser program (Chapter 5.7.1.15). Both programs need further a plain text file containing the genes to be used for the prediction. Taking the gene length from the sequence file and domain coordinates from the program outputs, a non-valid XML file is created which can be read by the program DOG (Ren et al., 2009) to visualize multiple proteins with their respective domain structure.

4.7.1.12. PfamScanAnalyzer / PfamScanEntry / LectinFinder

Functional domains were predicted by using the PfamScan Perl script from the Pfam database (Finn et al., 2010). In order to compare predictions from different genomes, the program PfamScanAnalyzer reduces the output of the Perl script in the following way. 1) A list

containing all domain IDs present in the Pfam database is loaded and used as reference. 2) For each domain in the reference list, the output file from the Perl script is searched and counts (a) the overall occurrence of the domain and (b) the average length of the domain together with the minimal and maximal size. 3) Proteins containing the specified domain are counted additionally, independent whether they contain the domain just once or more often. 4) The domain ID, domain description, number of domains, average size of the domain and number of proteins containing the domain are written in tab-delimited text file.

The PfamscanEntry class holds the definitions for a Pfamscan object to be described in JAVA. These definitions include the attributes a Pfamscan object hast, like the gene- or domain identifier, and the methods that can be performed on it, like retrieving or setting attribute values.

With the LectinFinder program, those proteins can be extracted from the Pfamscan output file, which only contain a defined subset of domains. Therefore, all predicted domains of each protein are compared against a provided list of domain identifiers. The protein then only remains in the output if all domains are listed and is discarded otherwise.

4.7.1.13. ProbeInfo / ProbeEvaluator

Performance of the probes from first microarray design (Chapter 5.5.9.1) was assessed in two steps. While the first program collects required information in a new file, the second one evaluates probe efficiency based on these data. The program ProbeInfo expects four input files. 1) The microarray file produced by the Agilent Feature Extraction software. 2) The tab-delimited text file containing the probe IDs and their respective sequence which can be downloaded from Agilent's eArray platform after manufacturing of the respective microarray. 3) The probe summary file that is created after probe design an contains the probe sequence, the target gene ID of the respective probe, the position of the probe in the gene as well as other characteristics of the probe sequence. 4) The fasta file of the genes that was used to predict probes. From these files, the program extracts the following fields and combines them into one single file. The ID of the probe and signal on the array is taken from (1). With the ID, the sequence of the probe is retrieved from (2). The sequence is then necessary to get the probe location, target ID and cross-hybridization potential from (3). In the last step, the target ID is used to get the target sequence and its length from (4).

When more than one probe per gene was used in a microarray experiment, the program ProbeEvalutor can identify the best performing probe (BP) per gene based on the following

criteria: 1) The output file(s) from the ProbeInfo program are loaded. 2) Probes with a cross hybridization potential are discarded if an alternative probe for the gene could be used. If this is not possible, three probes with different cross hybridization targets are chosen and used (XP1-XP3). 3) The signal intensity of each probe is compared to its position in the gene. The probe is defined to be a BP if its signal intensity is high and its position is close to the 3' end of the respective gene. Thus, the BP probe either has the strongest signal of all probes of the respective gene and is located at most 300 bp apart from the 3' end or it is the closest to the 3' end and has at least 90% signal intensity of the strongest probe. 4) The signal intensity of the best probe defined in step 3 was compared against those of its intra- and inter-array replicates in order to exclude non-uniformity outliers. If the signal intensity ranged between +/- 1.42 of the interquartile range, it was accepted as the best probe. Alternatively, step 4 was repeated with the second (third, etc.) best probe.

4.7.1.14. SequenceTableParser

All data produced by the Blast2GO suite (Conesa et al., 2005) can be retrieved in the so called sequencetable file. The program SequenceTableParser uses this file and separates the field containing applied gene ontologies into three according to the main GO categories cellular component (C), biological process (P) and molecular function (F).

4.7.1.15. SmartBatchParser

Domain prediction via SMART (Letunic et al., 2009) may be performed on multiple amino acid sequences via a free Perl script (http://smart.embl-heidelberg.de/help/SMART_batch.pl). Using the script, however, results in single processing of sequences and results are also stored in separate text files. The program SmartBatchParser therefore collects all significant domain predictions and writes them into a new file in a tabular output format.

4.7.1.16. TargetPParser

TargetP (Emanuelsson et al., 2000) including cleavage site predictions by SignalP (Nielsen et al., 1997) was used to predict signal peptides in protein sequences. Based on the output of these programs and a provided threshold value, the program TargetPParser produces two files. One is a list of IDs of proteins that were predicted to have a signal peptide with a reliability below or equal to the given threshold. The other one is a summary of all results without selection.

4.7.1.17. TmhmmParser

TMHMM (Sonnhammer et al., 1998) was used to predict transmembrane domains in protein sequences. Based on the output of this program and a provided threshold value, the program TmhmmParser produces two files. One is a list of IDs of proteins that were predicted to have a number of transmembrane domains below the provided threshold value or none at all. Furthermore, the prediction is summarized in a table. The other one is a summary of all results without selection.

4.7.1.18. WolfpsortParser

Sub-cellular localization of proteins was predicted using WoLF PSORT (Horton et al., 2007). The output file and a threshold value are expected by the program WolfpsortParser in order to interpret localization predictions. Possibly multiple localization targets for a protein are ranked by WoLF PSORT and the provided threshold value defines the minimal difference between the first and second possible target. If the difference is higher than the threshold, the first target is set as a localization target of the respective protein and no target is assigned otherwise.

4.7.2. R scripts

The statistic language R (R Development Core Team, 2011) was used for all purposes where statistical analyses were necessary. Whenever these analyses included processing of many statistical tests or graphical visualization, a script was written to facilitate the analyses. Function and usage of all R scripts used during this thesis shall be described in the following sub-chapters.

4.7.2.1. MicroarrayAnalyses

One colour microarray files produced by Agilent's feature extraction software are read and compared depending on the given settings. First, the gProcessedSignal from the arrays is read as green signal. This signal value is the adjusted green signal value were the background signal is subtracted and a multiplicative detrending algorithm is applied. The algorithm is another correction which plays a role if the hybridization efficiency on the array was not equal. This is determined by replicate probes on the chip and no correction is applied when the signal intensities of the replicates are homogeneous. The gProcessedSignal is then quantile-normalized between all loaded arrays and afterwards log2 transformed. Differential

expression values are then calculated between the experiments and t statistics applied on the calculation.

4.7.2.2. PCA

The script follows the MicroarrayAnalyses script until the quantile-normalization of the signal. Afterwards, principle components (PC) are calculated using the prcomp function from the limma package. The first three PCs are drawn in two 2D plots (PC1 vs. PC2 and PC2 vs. PC3).

4.7.2.3. MAPlot

MA plots were used to assess the suitability of the performed normalization method. Therefore, the malowess R-script from the program Robin (Lohse et al., 2010) was used in an own script were M is the difference of the green signal from the average signal intensity and A half of the signal plus the average signal intensity.

4.7.2.4. DensityPlot

Density plots were used, as MA plots, to assess the suitability of the performed normalization method. Therefore, the plotSigDensities R-script from the program Robin (Lohse et al., 2010) was used in an script in order to plot all raw and all normalized signal distributions in two separate graphs next to each other.

4.7.2.5. DomExpTTest

Significant differences between numbers of domains were calculated by applying the "t.test" of function to supplied data. The input is expected to be a tab-delimited text file were the first row and first column are header and domain descriptions, respectively. A one sample t-test is then performed for each row using the value in the second column as μ. For the two sample t-tests, the μ was replaced with the column range of the data that should be compared.

4.7.2.6. Heatmap

In order to show a graphical representation of a selection of differential expressed genes, heatmaps were produced using the heatmap.2 function. If not stated otherwise, log2-transformed expression values were used and the columns of heatmap clustered using the hclust function while the rows were manually sorted.

4.7.2.7. IntergenicSpacePlot_2D

Based on a given data matrix containing x, y and z values, the script draws a coloured density plot where x and y are the coordinates within the plot and z refers to a colour value ranging from blue via yellow to red.

4.7.3. Shell scripts

In those cases, were multiple genomes were analyzed and/or multiple programs were used one after another, shell scripts were written to automate and direct the workflow. These scripts shall be described briefly in the following sub-chapters.

4.7.3.1. FastaStandardization

The JAVA program FastaLoader is applied to all genomes listed in a provided text file. For each genome, the full path to the nucleotide sequences should be given. If a gff file belonging to an organism is located in the same folder as the nucleotide sequences, this is additionally assigned to the JAVA program.

4.7.3.2. NucLocPrediction

Providing a fasta-formatted amino acid sequence file, the script first creates single fasta files in a sub folder. The program predictNLS (Cokol et al., 2000) is then used with every fasta file in the subfolder and the individual output files saved in another sub folder. After completion of all predictions, all output files are combined in one single file and the IDs of all proteins containing a predicted nuclear localization signal were extracted using the Unix shell commands cat and grep.

4.7.3.3. PfamScanMultiExe

Executes the PfamScan Perl-script (Chapter 5.7.1.12) on a list of amino acid sequences. Afterwards, the output is modified in such a way that continuous spaces are replaced by tab-stops using the Unix command sed.

4.7.3.4. PfamScanSelection

Based on a tab-delimited PfamScan output and a text file containing a list of Pfam domain accession numbers, the script copies entries belonging to a listed accession into separate files.

4.7.3.5. SecretionPrediction

The script coordinates the execution of the different JAVA programs and Unix commands in the following steps. 1) The input sequences are converted to a one-line Fasta format and, if nucleotide sequences are provided instead of amino acid, translated by the programs fasta2onelineFasta and DNAtranslator (Chapter 5.7.1.5). 2) Signal peptides are predicted by using the TargetP software package v1.1 (Emanuelsson et al., 2000) (including cleavage site predictions by SignalP (Nielsen et al., 1997) with standard settings for non-plant networks. Because the TargetP software does not allow more than 1000 sequences to be processed at a time, the input file is first split and afterwards merged using the Unix commands split and grep. The merged output is then parsed using the TargetPParser program (Chapter 5.7.1.16) with a RC cutoff value of 3. 3) The resulting selection file and all protein sequences of the respective organism are used as inputs for the MultipleFastaSelection program (Chapter 5.7.1.9) and the filtered sequence file then used for the prediction of transmembrane domains using TMHMM v2.0 (Sonnhammer et al., 1998) with standard settings. The resulting output is parsed using the TMHMMParser program (Chapter 5.7.1.17) with a TM cutoff value of 1. The resulting selection file and all protein sequences of the respective organism were again used as inputs for the MultipleFastaSelection program and the filtered sequence file for all further analyses referred to as the secretome of the analyzed organism.

4.7.4. Comparative genomics

Large scale comparisons of genomic features in the analyzed genomes of *P. indica* and *S. vermifera* were compared against those of 52 selected Basidiomycetes and 8 selected Ascomycetes genomes (Table 5.17). All genomes are included in the fungal genome program of the JGI and genome/gene sequences and gff files were downloaded from the MycoCosm webportal (Grigoriev et al., 2011, 2012). In agreement with the data usage policy of the JGI, only those sequences were used that were either published in a peer-review journal or released latest 12 month before this thesis was finished. The last updated version of all sequences as present at the 1st of February, 2013 was used for all analyses.

Table 5.17: Genome sequences used in this thesis

Name	Abb[1]	Lifestyle[2]	Reference or date of sequence release
Basidiomycetes			

Sebacina vermifera MAFF 305830	Sebve	Symbiotic (OM)	Release: Mar, 2012
Piriformospora indica DSM 11827	Pirin	Symbiotic (EP)	(Zuccaro et al., 2011)
Laccaria bicolor	Lacbi	Symbiotic (ECM)	(Martin et al., 2008)
Paxillus involutus ATCC 200175	Paxin	Symbiotic (ECM)	Release: Mar, 2011
Piloderma croceum F 1598	Pilcr	Symbiotic (ECM)	Release: Feb, 2012
Pisolithus microcarpus 441	Pismi	Symbiotic (ECM)	Release: Feb, 2012
Pisolithus tinctorius Marx 270	Pisti	Symbiotic (ECM)	Release: Mar, 2012
Scleroderma citrinum Foug A	Sclci	Symbiotic (ECM)	Release: Mar, 2012
Auricularia delicata SS-5	Aurde	Saprotrophic (WR)	(Floudas et al., 2012)
Bjerkandera adusta	Bjead	Saprotrophic (WR)	Release: Jun, 2011
Ceriporiopsis (*Gelatoporia*) *subvermispora* B	Cersu	Saprotrophic (WR)	Release: May, 2010
Dichomitus squalens	Dicsq	Saprotrophic (WR)	(Floudas et al., 2012)
Fomitiporia mediterranea	Fomme	Saprotrophic (WR)	(Floudas et al., 2012)
Galerina marginata	Galma	Saprotrophic (WR)	Release: Jun, 2011
Ganoderma sp. 10597 SS1	Gansp	Saprotrophic (WR)	Release: Mar, 2011
Heterobasidion annosum	Hetan	Saprotrophic (WR)	(Floudas et al., 2012)
Phanerochaete carnosa HHB-10118-Sp	Phaca	Saprotrophic (WR)	(Floudas et al., 2012)
Phlebia brevispora HHB-7030 SS6	Phlbr	Saprotrophic (WR)	Release: Mar, 2011
Phlebiopsis gigantea	Phlgi	Saprotrophic (WR)	Release: Jun, 2011
Pleurotus ostreatus PC15	PleosPC 15	Saprotrophic (WR)	Release: Jun, 2010
Plicaturopsis crispa	Plicr	Saprotrophic (WR)	Release: Nov, 2011
Punctularia strigosozonata	Punst	Saprotrophic (WR)	(Floudas et al., 2012)
Schizophyllum commune	Schco	Saprotrophic (WR)	(Floudas et al., 2012)
Stereum hirsutum FP-91666 SS1	Stehi	Saprotrophic (WR)	(Floudas et al., 2012)
Trametes versicolor	Trave	Saprotrophic (WR)	(Floudas et al., 2012)

Agaricus bisporus var bisporus (H97)	AgabiBis	Saprotrophic (SD)	(Morin et al., 2012)
Agaricus bisporus var. burnettii JB137-S8	AgabiBur	Saprotrophic (SD)	Release: May 2010
Amanita thiersii Skay4041	Amath	Saprotrophic (SD)	(Wolfe et al., 2012)
Coprinopsis cinerea	Copci	Saprotrophic (SD)	(Floudas et al., 2012)
Sphaerobolus stellatus	Sphst	Saprotrophic (SD)	Release: Apr, 2012
Coniophora puteana	Conpu	Saprotrophic (BR)	(Floudas et al., 2012)
Dacryopinax sp. DJM 731 SSP1	Dacsp	Saprotrophic (BR)	(Floudas et al., 2012)
Fomitopsis pinicola FP-58527 SS1	Fompi	Saprotrophic (BR)	(Floudas et al., 2012)
Gloeophyllum trabeum	Glotr	Saprotrophic (BR)	(Floudas et al., 2012)
Jaapia argillacea	Jaaar	Saprotrophic (BR)	Release: Aug, 2011
Postia placenta MAD 698-R	Pospl	Saprotrophic (BR)	(Floudas et al., 2012)
Postia placenta MAD-698-R-SB12	PosplRSB12	Saprotrophic (BR)	(Floudas et al., 2012)
Serpula lacrymans S7.3	SerlaS73	Saprotrophic (BR)	(Floudas et al., 2012)
Serpula lacrymans S7.9	SerlaS79	Saprotrophic (BR)	(Floudas et al., 2012)
Serpula lacrymans var shastensis SHA21-2	SerlaSha	Saprotrophic (BR)	(Floudas et al., 2012)
Wolfiporia cocos MD-104 SS10	Wolco	Saprotrophic (BR)	(Floudas et al., 2012)
Botryobasidium botryosum	Botbo	Saprotrophic	Release: Aug, 2011
Gymnopus luxurians	Gymlu	Saprotrophic	(Lackner et al., 2012)
Hypholoma sublateritium	Hypsu	Saprotrophic	Release: Aug, 2011
Mixia osmundae IAM 14324	Mixos	Pathogenic (PP)	(Nishida et al., 2011)
Puccinia graminis	Pucgr	Pathogenic (PP)	(Duplessis et al., 2011)
Rhodotorula graminis WP1	Rhoba	Pathogenic (PP)	Release: Sep, 2010
Sporisorium relianum	Spore	Pathogenic (PP)	(Schirawski et al., 2010)
Ustilago maydis	Ustma	Pathogenic (PP)	(Kämper et al., 2006)

Cryptococcus neoformans var neoformans JEC21	CryneJEC21	Pathogenic (AP)	(Floudas et al., 2012)
Cryptococcus neoformans var. grubii H99	CryneH99	Pathogenic (AP)	(Floudas et al., 2012)
Malassezia globosa	Malgl	Pathogenic (AP)	(Floudas et al., 2012)
Ascomycetes			
Cladonia grayi	Clagr	Symbiotic (LF)	Release: Jul, 2010
Xanthoria parietina	Xanpa	Symbiotic (LF)	Release: Jan, 2011
Daldinia eschscholzii EC12	DalEC12	Symbiotic (EP)	Release: Oct, 2011
Hypoxylon sp. Cl-4A	HypCl4A	Symbiotic (EP)	Release: Oct, 2011
Hypoxylon sp. CO27-5	HypCO275	Symbiotic (EP)	Release: Jan, 2012
Hypoxylon sp. EC38	HypEC38	Symbiotic (EP)	Release: Feb, 2012
Oidiodendron maius	Oidma	Symbiotic (EP)	Release: May, 2011
Tuber melanosporum	Tubme	Symbiotic (ECM)	Martin et al., 2010

[1]Genome abbreviations used in this thesis

[2]OM: Orchid Mycorrhiza, EP: Endophyte, ECM: Ectomycorrhiza, WR: White Rot, SD: Soil Decomposer, BR: Brown Rot, PP: Plant Pathogen, AP: Animal Pathogen, LF: Lichen Forming

4.7.5. Identification of single nucleotide polymorphisms

The presence of two haploid genomes in the sequence information of *P. indica* was bioinformatically verified by searching for single nucleotide polymorphisms (SNPs) using the swap454 program from the Broad Institute (Brockman et al., 2008). According to the protocol (http://www.broadinstitute.org/science/programs/genome-biology/computational-rd/454-help) a new standard flowgram format (SFF) file was created from the raw read sequence fasta and quality files. For the creation of a coverage map the assembled contig sequences were used as reference. The SNP calling parameters were chosen in such a way that at least 10% of the reads had to differ from the reference sequence in order to be counted as a SNP. The predicted number and position of SNPs in the contigs was afterwards manually validated in ~100 randomly chosen contigs using the assembly viewer eagleview (Huang and Marth, 2008).

4.7.6. Repetitive elements

The prediction of transposable elements (TE) was performed using the program RepeatModeler (Smit and Hubley, 2008) on genomic scaffolds. RepeatModeler combines the conveniently used tools RECON (Bao and Eddy, 2002), RepeatScout (Price et al., 2005), TRF (Benson, 1999) and RepeatMasker (Smit et al., 1996) to predict, classify and mask repetitive sequences in genomes. Identified TEs were classified in this pipeline using Blast-based database searches and analysis of structural characteristics. Blast comparisons were performed against Repbase, a reference database of eukaryotic repetitive sequences (Jurka et al., 2005) and the NCBI non-redundant database (Pruitt et al., 2007) to distinguish TEs from cellular genes. Analysis of the TE structure included subfamily identification for DNA transposons, pattern analysis of sequence alignments for DNA transposons, long terminal repeats (LTRs) and long interspersed nuclear elements (LINEs) as well as identification of RNA polymerase III (pol III) promoter elements for short interspersed nuclear elements (SINEs). Additionally to TEs, simple sequence repeats (SSR) were predicted using the MISA tool (http://pgrc.ipk-gatersleben.de/misa/) with standard settings.

4.7.7. Transfer RNAs / codon usage

For the prediction of tRNAs the program tRNAscan-SE (Lowe and Eddy, 1997) and Aragorn (Laslett and Canback, 2004) were used. The prediction was performed on the nucleotide sequences of scaffolds and contigs with the default search mode and eukaryotic gene model. Obtained results were afterwards combined manually by discarding all tRNAs that were only predicted by one of the two programs.

Codon triplets and a corresponding codon table were calculated from nucleotide sequences of predicted genes using the JAVA program CodonUsage (Chapter 5.7.1.4). The list of codons produced by the program was then used to calculate frequency plots using WebLogo (Crooks et al., 2004). The plots show which nucleotide is preferred in each position of the codon triplets.

4.7.8. Gene modelling

Gene modelling for *P. indica* was first performed using the *ab initio* gene prediction program AUGUSTUS (Stanke and Waack, 2003). Gene models derived from *Laccaria bicolor* and *Coprinus cinerea* were used as training sets and *P. indica* ESTs were used as hints. The obtained results were further expanded and manually revised in cooperation with Dr. Ulrich

Güldener from the institute of bioinformatics and systems biology (IBIS) at the Munich Information Center for Protein Sequences (MIPS). Briefly, a collection of possible gene definitions was created by applying Fgenesh (Salamov and Solovyev, 2000), GeneMark-ES (Ter-Hovhannisyan et al., 2008) and AUGUSTUS with different training sets, all gene definitions were manually revised and one "best-fitting" gene model selected. For the evaluation of the putative gene models, *P. indica* ESTs were mapped onto the genome using Blat (Kent, 2002). The gene modelling for *S. vermifera* as well as the verification of the predictions by mapping of ESTs was performed at the DOE Joint Genome Institute (JGI) in Walnut Creek (CA), USA. To estimate the genome completeness on both genomes, reciprocal Blastp comparisons against a database of core eukaryotic genes (CEG) (Parra et al., 2009) was performed with an eValue threshold of 10^{-3} and manual revision of the result.

4.7.9. Annotation of predicted open reading frames

For a first automated prediction of all protein coding genes predicted in the genome of *P. indica* the PEDANT system (Walter et al., 2009) as well as the IMG system (Markowitz et al., 2011) have been used. Results from these analyses for *P. indica* are accessible at http://pedant.gsf.de/pedant3htmlview/pedant3view?Method=analysis&Db= p3_t65672_Pir_indic and http://genome.jgi.doe.gov/Pirin1/Pirin1.home.html, respectively. For the annotation of proteins predicted in the genome of *S. vermifera* only the IMG system was used and results are accessible at http://genome.jgi-psf.org/Sebve1/Sebve1.home.html. In order to expand and refine these annotations, more specialized programs were used that shall be described in the following sub-chapters.

4.7.9.1. Blast2GO

An additional *in silico* approach for batch analyses of new sequences is the program Blast2GO (http://www.blast2go.org/home) that was designed with the main purpose of enabling Gene Ontology (GO) based data mining on sequence data for which no GO annotation is yet available (Conesa et al., 2005). For that purpose the application provides data analyses in 4 major steps. 1) BLAST: Supplied nucleotide sequences are compared via blastx searches against the NCBI nr-database. 2) Mapping: GO terms are mapped on the BLAST results based on a monthly updated GO database. 3) Annotation: Based on user chosen parameters an annotation rule is applied to find the most specific annotations with a certain level of reliability. 4) Visualization: The data can be viewed in several different ways,

including charts and KEGG maps. During this thesis the steps 1 to 3 were used straightforward with standard settings if not stated otherwise.

4.7.9.2. Sub cellular localization of proteins

Putative cellular targets of proteins were predicted by WoLF PSORT (Horton et al., 2007) which estimates possible targets based on the prediction of leading sequences and the general amino acid composition of a sequence. The program outputs all possible targets of protein ranked descending based on the likeliness of the prediction. To reduce the number of false positives, predictions were filtered by allowing only those, were the "first neighbour" was more than 50% higher than the "second neighbour". Nuclear localization signals were additionally verified by using PredictNLS (Cokol et al., 2000). Because the program is restricted to one fasta files containing just a single sequence, the prediction was automated in a shell script (Chapter 5.7.3.2).

4.7.9.3. Automated secretome prediction

The secretome was defined as a subset of predicted genes which are most likely to be secreted extracellularly. To allow an automated prediction of the secretome in one or more genomes, published prediction programs, self-written JAVA programs and basic Unix commands were combined in a shell script (Chapter 5.7.3.5).

4.7.9.4. Functional protein domains

A protein domain is a conserved part of a protein that has a specific and evolutionary conserved function. Such domains were predicted using the PfamScan Perl-script, ftp://ftp.sanger.ac.uk/pub/databases/Pfam/Tools/PfamScan.tar.gz) from the Pfam database (Finn et al., 2010) and the results were validated with the SMART analysis pipeline (Letunic et al., 2009) using standard settings. For the analyses of present domains and comparisons between predictions, the output of the script was used together with the description file of all Pfam domains (version 24 release) as input for the PfamScanAnalyzer program (Chapter 5.7.1.12). Results were used for domain clustering (Chapter 5.7.11) and/or smaller comparisons of single domains.

4.7.9.5. Specialized protein database

In recent years, not only the amount of functional *in silico* predictions but also the quality has improved constantly. This is mainly a result of using biological data to improve available

algorithms. However, such a combination results also in many specialized programs and databases. Because these programs are more accurate in their predictions than more general approaches, the annotations described before were verified and/or modified according to the databases and underlying prediction algorithms described in table 5.18.

Table 5.18: Specialized protein annotation databases

Name	Focus of prediction	Website	Reference
AntiSMASH	Secondary metabolites	http://antismash.secondarymetabolites.org/	(Medema et al., 2011)
FTFD	Fungal transcription factors	http://ftfd.snu.ac.kr/index.php?a=view	(Park et al., 2008)
FunyBASE	Fungal Phylogenomics	http://genome.jouy.inra.fr/funybase/	(Marthey et al., 2008)
PHI-base	Pathogen host interaction	http://www.phi-base.org/release_notes.php	(Winnenburg et al., 2008)
MEROPS	Proteolytic enzymes	http://merops.sanger.ac.uk/	(Rawlings et al., 2011)
mycoCLAP	Ligno cellulytic enzymes	https://mycoclap.fungalgenomics.ca/mycoCLAP/	(Murphy et al., 2011)
TCDB	Transporter	http://www.tcdb.org/	(Saier et al., 2009)
TransportTP	Transporter	http://bioinfo3.noble.org/transporter/	(Li et al., 2008)

4.7.10. Phylogenetic analyses

Phylogenetic relations were analysed with predicted nucleotide sequences or deduced amino acid sequences. In cases were functional domains of proteins were analysed, only those domains were aligned, either the single domains or domains belonging to one protein concatenated. In all cases, the program MultipleFastaSelection (Chapter 5.7.1.9) was used to extract the desired sequences from a larger set, were domain boundaries were taken from Pfam predictions (chapter 5.7.9.4). Alignments and tree building was then performed within SeaView (Gouy et al., 2010). For sequence alignments, MUSCLE (Edgar, 2004) was used with a doubled number of iterations (-maxiter 32) to increase accuracy and manual correction afterwards if necessary. If not stated differently, phylogenetic relations were then computed

using PhyML (Guindon et al., 2010) starting with a bio neighbour joining (BioNJ) tree and the LG model as replacement matrix (Le and Gascuel, 2008). The best tree was then estimated from the tree space by Subtree Pruning and Regrafting (SPR) (Hordijk and Gascuel, 2005). Branch support is given as an approximate Likelihood-Ratio Test (aLRT) (Anisimova and Gascuel, 2006).

4.7.11. Cluster analyses

Protein sequences and functional protein domains were grouped in clusters based on two different methods. In the first approach, the program mcl (Enright et al., 2002) was used according to the online available workflow protocol (http://micans.org/mcl/man/clmprotocols.html#blast). The inflation parameter was defined by clustering with increasing inflation parameters going from 1 to 4 in steps of 0.2. All results were compared with respect to their ability to group proteins containing a LysM and WSC domain separately while clustering only *P. indica* or only *S. vermifera* proteins. Based on these results an optimal inflation parameter of 1.4 could be estimated and was used for all further clustering procedures. To identify *P. indica* or *S. vermifera* specific protein families in the Basidiomycetes group, a blastp (eVal: 10^{-3}) "all vs all" comparison of all protein sequences from Basidiomycetes genomes listed in table 5.20??? was performed and used as input for the mcl workflow. In the second clustering approach, functional domains were predicted as described in chapter 5.7.10.4 in the same sequence set as used for the mcl clustering. A decreased/increased number of proteins domains in comparison to the other genomes was determined by calculating significant differences from either *P. indica* or *S. vermifera* separately against the other genomes (one sample t-test) or both together (two sample t-test). In both cases, a protein domain were considered to be enriched/constraint if its number was significantly ($p<0.01$) higher/lower than the population mean by a factor of two or more.

4.7.12. Evolutionary analysis of protein families

Evolutionary changes in protein family size were analyzed using CAFE (Bie et al., 2006). For the identification of protein expansions/contractions, all protein families from the mcl cluster analysis were used that contained at least 5 proteins. From this set, all protein families that are unique to one of the analyzed genomes were excluded. A phylogenetic tree was constructed based on 10 core eukaryotic genes highly conserved in all tested organisms.

4.7.13. Enrichment analysis

To identify significantly enriched gene ontology (GO) terms from the microarray hybridization experiments, the Gene Ontology Enrichment Analysis Software Toolkit (GOEAST) was used (http://omicslab.genetics.ac.cn/GOEAST/index.php) with settings for customized microarray platforms. For the enrichment analysis the probe annotation file for gene ontology terms produced by Blast2GO was used and the format adjusted to fit the GOEAST requirements. Induced genes during symbiosis or during growth on autoclaved root material were analyzed using the recommended parameter settings.

4.7.14. Assembly and analysis of the mitochondrion

For the assembly of the *P. indica* mitochondrion, all contigs with either a high coverage or a low GC-content were assembled in a single scaffold using the contig assembler seqMan (Swindell and Plasterer, 1997). Circularity was verified by PCR with primers designed at the beginning and at the end of the scaffold. Genes on the mitochondrion were predicted using a program pipeline with different bioinformatical tools. 1) Different *in silico* sheared fragments were analyzed by Blast2GO to identify all genes on the mitochondrion of *P. indica*. The exon/intron structure of these genes was then refined by building consensuses from multiple sequences alignments produced by the program protein2genome of the Exonerate package (Slater and Birney, 2005). For the assembly of the *S. vermifera* mitochondrion, circularity of three scaffolds with low GC-content was verified by PCR with primers designed at the beginning and at the end of the scaffolds. Genes predicted on the *S. vermifera* mitochondrion were then mapped as described before.

4.7.15. Orthologues and Synteny prediction

Orthologues are genes or proteins that evolved separately in different species but have a common ancestor (Fitch, 2000). Such genes may be identified by their sequence similarity under the assumption that orthologues have the highest similarity between each other in the organisms analyzed. A common way of predicting orthologues in large data sets is the detection of reciprocal best hits (RBH) via blastp or blastn analyses (Hirsh and Fraser, 2001). The program Proteinortho (Lechner et al., 2011) was used to predict orthologues and clusters of orthologues, respectively, via RBHs with some adjustments to the standard settings. To allow partially predicted genes to be used reasonable in the analysis, the minimal coverage of the best blast hit was reduced from 0.5 to 0.25 and the minimal similarity for additional hits from 0.95 to 0.5. Because the program just predicts orthologue clusters when used with

standard settings, the parameters pairs, to output single orthologue pairs, and singles, to output proteins without an orthologue, were additionally used.

Syntenic regions in genomes were predicted and visualized using Symap (Soderlund et al., 2006). For improved performance of the program and to allow a graphical display, genome sequences were concatenated into up to 10 super-scaffolds with the concatFasta method (Chapter 5.7.1.5) before use in the Symap program.

5. References

Achatz, B., Rüden, S. von, Andrade, D., Neumann, E., Pons-Kühnemann, J., Kogel, K.-H., Franken, P., and Waller, F. (2010). Root colonization by *Piriformospora indica* enhances grain yield in barley under diverse nutrient regimes by accelerating plant development. Plant Soil *333*, 59–70.

Ahuja, I., Kissen, R., and Bones, A.M. (2012). Phytoalexins in defense against pathogens. Trends in Plant Science *17*, 73–90.

Allen, T.R., Millar, T., Berch, S.M., and Berbee, M.L. (2003). Culturing and direct DNA extraction find different fungi from the same ericoid mycorrhizal roots. New Phytologist *160*, 255–272.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl. Acids Res. *25*, 3389–3402.

Anisimova, M., and Gascuel, O. (2006). Approximate Likelihood-Ratio Test for Branches: A Fast, Accurate, and Powerful Alternative. Syst Biol *55*, 539–552.

Asano, T., Kimura, M., and Nishiuchi, T. (2012). The defense response in *Arabidopsis thaliana* against *Fusarium sporotrichioides*. Proteome Science *10*, 61.

Aziz, A., Gauthier, A., Bézier, A., Poinssot, B., Joubert, J.-M., Pugin, A., Heyraud, A., and Baillieul, F. (2007). Elicitor and resistance-inducing activities of β-1,4 cellodextrins in grapevine, comparison with β-1,3 glucans and α-1,4 oligogalacturonides. J. Exp. Bot. *58*, 1463–1472.

Bacon, C.W., and White, J. (2000). Microbial Endophytes (CRC Press).

Bak, S., Tax, F.E., Feldmann, K.A., Galbraith, D.W., and Feyereisen, R. (2001). CYP83B1, a Cytochrome P450 at the Metabolic Branch Point in Auxin and Indole Glucosinolate Biosynthesis in Arabidopsis. Plant Cell *13*, 101–111.

Baker, M. (2012). De novo genome assembly: what every biologist should know. Nat Meth *9*, 333–337.

Baltruschat, H., Fodor, J., Harrach, B.D., Niemczyk, E., Barna, B., Gullner, G., Janeczko, A., Kogel, K.-H., Schäfer, P., Schwarczinger, I., et al. (2008). Salt tolerance of barley induced by the root endophyte *Piriformospora indica* is associated with a strong increase in antioxidants. New Phytologist *180*, 501–510.

Bao, Z., and Eddy, S.R. (2002). Automated De Novo Identification of Repeat Sequence Families in Sequenced Genomes. Genome Res. *12*, 1269–1276.

Barazani, O., von Dahl, C.C., and Baldwin, I.T. (2007). *Sebacina vermifera* Promotes the Growth and Fitness of *Nicotiana attenuata* by Inhibiting Ethylene Signaling. Plant Physiol *144*, 1223–1232.

Basiewicz, M., Weiß, M., Kogel, K.-H., Langen, G., Zorn, H., and Zuccaro, A. (2012). Molecular and phenotypic characterization of *Sebacina vermifera* strains associated with orchids, and the description of *Piriformospora williamsii* sp. nov. Fungal Biology *116*, 204–213.

Bednarek, P. (2012). Chemical warfare or modulators of defence responses – the function of secondary metabolites in plant immunity. Current Opinion in Plant Biology *15*, 407–414.

Bednarek, P., Pislewska-Bednarek, M., Svatos, A., Schneider, B., Doubsky, J., Mansurova, M., Humphry, M., Consonni, C., Panstruga, R., Sanchez-Vallet, A., et al. (2009). A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum antifungal defense. Science *323*, 101–106.

Behnsen, J., Lessing, F., Schindler, S., Wartenberg, D., Jacobsen, I.D., Thoen, M., Zipfel, P.F., and Brakhage, A.A. (2010). Secreted *Aspergillus fumigatus* Protease Alp1 Degrades Human Complement Proteins C3, C4, and C5. Infect. Immun. *78*, 3585–3594.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucl. Acids Res. *27*, 573–580.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., et al. (2008). Accurate Whole Human Genome Sequencing using Reversible Terminator Chemistry. Nature *456*, 53–59.

Berch, S.M., Allen, T.R., and Berbee, M.L. (2002). Molecular detection, community structure and phylogeny of ericoid mycorrhizal fungi. Plant and Soil *244*, 55–66.

Beyrle, H.F., and Smith, S.E. (1993). The effect of carbohydrate on the development of a Cattleya hybrid in association with its mycorrhizal fungus. Mycorrhiza *3*, 57–62.

Bhadauria, V., Banniza, S., Wang, L.-X., Wei, Y.-D., and Peng, Y.-L. (2010). Proteomic studies of phytopathogenic fungi, oomycetes and their interactions with hosts. Eur J Plant Pathol *126*, 81–95.

Bie, T.D., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFE: a computational tool for the study of gene family evolution. Bioinformatics *22*, 1269–1271.

Birkenbihl, R.P., Diezel, C., and Somssich, I.E. (2012). Arabidopsis WRKY33 Is a Key Transcriptional Regulator of Hormonal and Metabolic Responses toward *Botrytis cinerea* Infection. Plant Physiol. *159*, 266–285.

Björklund, Å.K., Ekman, D., Light, S., Frey-Skött, J., and Elofsson, A. (2005). Domain Rearrangements in Protein Evolution. Journal of Molecular Biology *353*, 911–923.

Björklund, Å.K., Ekman, D., and Elofsson, A. (2006). Expansion of Protein Domain Repeats. PLoS Comput Biol *2*, e114.

Black, K.G., Mitchell, D.T., and Osborne, B.A. (2000). Effect of mycorrhizal-enhanced leaf phosphate status on carbon partitioning, translocation and photosynthesis in cucumber. Plant, Cell & Environment *23*, 797–809.

Blackwell, M. (2011). The Fungi: 1, 2, 3 … 5.1 million species? Am. J. Bot. *98*, 426–438.

Boldt, B., M, O., L, L., K, W., L, S.-J., and C, O. (2001). Mutational analysis of the proteolytic domain of pregnancy-associated plasma protein-A (PAPP-A): classification as a metzincin.

Boller, T., and He, S.Y. (2009). Innate Immunity in Plants: An Arms Race Between Pattern Recognition Receptors in Plants and Effectors in Microbial Pathogens. Science *324*, 742–744.

Bonfante, P., and Genre, A. (2010). Mechanisms underlying beneficial plant–fungus interactions in mycorrhizal symbiosis. Nat Commun *1*, 48.

Bougoure, D.S., and Cairney, J.W.G. (2005). Assemblages of ericoid mycorrhizal and other root-associated fungi from *Epacris pulchella* (Ericaceae) as determined by culturing and direct DNA extraction from roots. Environmental Microbiology *7*, 819–827.

Bougoure, J.J., Bougoure, D.S., Cairney, J.W.G., and Dearnaley, J.D.W. (2005). ITS-RFLP and sequence analysis of endophytes from Acianthus, Caladenia and Pterostylis (Orchidaceae) in southeastern Queensland. Mycological Research *109*, 452–460.

Brakhage, A.A. (2013). Regulation of fungal secondary metabolism. Nat Rev Micro *11*, 21–32.

Bressan, M., Roncato, M.-A., Bellvert, F., Comte, G., Haichar, F. el Z., Achouak, W., and Berge, O. (2009). Exogenous glucosinolate produced by *Arabidopsis thaliana* has an impact on microbes in the rhizosphere and plant roots. ISME J *3*, 1243–1257.

Brockman, W., Alvarez, P., Young, S., Garber, M., Giannoukos, G., Lee, W.L., Russ, C., Lander, E.S., Nusbaum, C., and Jaffe, D.B. (2008). Quality scores and SNP detection in sequencing-by-synthesis systems. Genome Res. *18*, 763–770.

Broekaert, W.F., Delauré, S.L., De Bolle, M.F.C., and Cammue, B.P.A. (2006). The Role of Ethylene in Host-Pathogen Interactions. Annual Review of Phytopathology *44*, 393–416.

Brouta, F., Descamps, F., Fett, T., Losson, B., Gerday, C., and Mignon, B. (2001). Purification and characterization of a 435 kDa keratinolytic metalloprotease from *Microsporum canis*. Medical Mycology *39*, 269–275.

Brown, R.L., Kazan, K., McGrath, K.C., Maclean, D.J., and Manners, J.M. (2003). A Role for the GCC-Box in Jasmonate-Mediated Activation of the PDF1.2 Gene of Arabidopsis. Plant Physiol. *132*, 1020–1032.

Brundrett, M. (2004). Diversity and classification of mycorrhizal associations. Biological Reviews *79*, 473–495.

Van der Burgt, A., Karimi, M., Bahkali, A.H., and de Wit, P.J.G.M. (2013). Pseudogenization in pathogenic fungi with different host plants and lifestyles might reflect their evolutionary past. Molecular Plant Pathology n/a–n/a.

Burke, D., Kaufman, P., McNeil, M., and Albersheim, P. (1974). The Structure of Plant Cell Walls VI. A Survey of the Walls of Suspension-cultured Monocots. Plant Physiol. *54*, 109–115.

Büttner, M., and Singh, K.B. (1997). *Arabidopsis thaliana* ethylene-responsive element binding protein (AtEBP), an ethylene-inducible, GCC box DNA-binding protein interacts with an ocs element binding protein. PNAS *94*, 5961–5966.

Buxdorf, K., Rahat, I., Gafni, A., and Levy, M. (2013). The Epiphytic Fungus *Pseudozyma aphidis* Induces Jasmonic Acid- and Salicylic Acid/Nonexpressor of PR1-Independent Local and Systemic Resistance. Plant Physiol. *161*, 2014–2022.

Camehl, I., Sherameti, I., Venus, Y., Bethke, G., Varma, A., Lee, J., and Oelmüller, R. (2010). Ethylene signalling and ethylene-targeted transcription factors are required to balance beneficial and nonbeneficial traits in the symbiosis between the endophytic fungus *Piriformospora indica* and *Arabidopsis thaliana*. New Phytologist *185*, 1062–1073.

Camehl, I., Drzewiecki, C., Vadassery, J., Shahollari, B., Sherameti, I., Forzani, C., Munnik, T., Hirt, H., and Oelmüller, R. (2011). The OXI1 Kinase Pathway Mediates *Piriformospora indica*-Induced Growth Promotion in Arabidopsis. PLoS Pathog *7*, e1002051.

Camera, S.L., Balagué, C., Göbel, C., Geoffroy, P., Legrand, M., Feussner, I., Roby, D., and Heitz, T. (2009). The *Arabidopsis* Patatin-Like Protein 2 (PLP2) Plays an Essential Role in Cell Death Execution and Differentially Affects Biosynthesis of Oxylipins and Resistance to Pathogens. Molecular Plant-Microbe Interactions *22*, 469–481.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res *37*, D233–D238.

Chen, H., Wang, C., Ye, J., Zhou, H., and Chen, X. (2012). Antimicrobial activities of phenethyl isothiocyanate isolated from horseradish. Natural Product Research *26*, 1016–1021.

Chen, L., Zhang, L., and Yu, D. (2010). Wounding-Induced WRKY8 Is Involved in Basal Defense in *Arabidopsis*. Molecular Plant-Microbe Interactions *23*, 558–565.

Chen, X., Liu, J., Lin, G., Wang, A., Wang, Z., and Lu, G. (2013). Overexpression of AtWRKY28 and AtWRKY75 in Arabidopsis enhances resistance to oxalic acid and *Sclerotinia sclerotiorum*. Plant Cell Rep *32*, 1589–1599.

Chou, H.-H. (2010). Shared probe design and existing microarray reanalysis using PICKY. BMC Bioinformatics *11*, 196.

Ciaffi, M., Paolacci, A.., Dominici, L., Tanzarella, O., and Porceddu, E. (2001). Molecular characterization of gene sequences coding for protein disulfide isomerase (PDI) in durum wheat (*Triticum turgidum* ssp. durum). Gene *265*, 147–156.

Clay, N.K., Adio, A.M., Denoux, C., Jander, G., and Ausubel, F.M. (2009). Glucosinolate Metabolites Required for an Arabidopsis Innate Immune Response. Science *323*, 95–101.

Cohen, S.N., Chang, A.C.Y., and Hsu, L. (1972). Nonchromosomal Antibiotic Resistance in Bacteria: Genetic Transformation of *Escherichia coli* by R-Factor DNA. PNAS *69*, 2110–2114.

Cohen-Gihon, I., Sharan, R., and Nussinov, R. (2011). Processes of fungal proteome evolution and gain of function: gene duplication and domain rearrangement. Phys. Biol. *8*, 035009.

Cokol, M., Nair, R., and Rost, B. (2000). Finding nuclear localization signals. EMBO Reports *1*, 411–415.

Coleman, J.J., and Mylonakis, E. (2009). Efflux in Fungi: La Pièce de Résistance. PLoS Pathog *5*, e1000486.

Coleman, J.J., White, G.J., Rodriguez-Carres, M., and VanEtten, H.D. (2011). An ABC Transporter and a Cytochrome P450 of *Nectria haematococca* MPVI Are Virulence Factors on Pea and Are the Major Tolerance Mechanisms to the Phytoalexin Pisatin. Molecular Plant-Microbe Interactions *24*, 368–376.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics *21*, 3674–3676.

De Coninck, B., Cammue, B.P.A., and Thevissen, K. (2013). Modes of antifungal action and in planta functions of plant defensins and defensin-like peptides. Fungal Biology Reviews *26*, 109–120.

Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E. (2004). WebLogo: A Sequence Logo Generator. Genome Res. *14*, 1188–1190.

D'Andrea, L.D., and Regan, L. (2003). TPR proteins: the versatile helix. Trends in Biochemical Sciences *28*, 655–662.

Dai, Y., Wang, H., Li, B., Huang, J., Liu, X., Zhou, Y., Mou, Z., and Li, J. (2006). Increased Expression of MAP KINASE KINASE7 Causes Deficiency in Polar Auxin Transport and Leads to Plant Architectural Abnormality in Arabidopsis. Plant Cell *18*, 308–320.

Damiano, J.S., Oliveira, V., Welsh, K., and Reed, J.C. (2004). Heterotypic interactions among NACHT domains: implications for regulation of innate immune responses. Biochemical Journal *381*, 213.

Dean, R.A., Talbot, N.J., Ebbole, D.J., Farman, M.L., Mitchell, T.K., Orbach, M.J., Thon, M., Kulkarni, R., Xu, J.-R., Pan, H., et al. (2005). The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature *434*, 980–986.

Deshmukh, S., Hückelhoven, R., Schäfer, P., Imani, J., Sharma, M., Weiss, M., Waller, F., and Kogel, K.-H. (2006). The root endophytic fungus *Piriformospora indica* requires host cell death for proliferation during mutualistic symbiosis with barley. PNAS *103*, 18450–18457.

Djamei, A., Schipper, K., Rabe, F., Ghosh, A., Vincon, V., Kahnt, J., Osorio, S., Tohge, T., Fernie, A.R., Feussner, I., et al. (2011). Metabolic priming by a secreted fungal effector. Nature *478*, 395–398.

Doi, Y., Lee, B.R., Ikeguchi, M., Ohoba, Y., Ikoma, T., Tero-Kubota, S., Yamauchi, S., Takahashi, K., and Ichishima, E. (2003). Substrate Specificities of Deuterolysin from *Aspergillus oryzae* and Electron Paramagnetic Resonance Measurement of Cobalt-substituted Deuterolysin. Bioscience, Biotechnology, and Biochemistry *67*, 264–270.

Dong, J., Chen, C., and Chen, Z. (2003). Expression profiles of the Arabidopsis WRKY gene superfamily during plant defense response. Plant Mol Biol *51*, 21–37.

Doyle, J., and Doyle, J. (1987). Genomic plant DNA preparation from fresh tissue-CTAB method. Phytochemical Bulletin *19*.

Duplessis, S., Cuomo, C.A., Lin, Y.-C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., Joly, D.L., Hacquard, S., Amselem, J., Cantarel, B.L., et al. (2011). Obligate biotrophy features unraveled by the genomic analysis of rust fungi. PNAS.

Earnshaw, W.C., Martins, L.M., and Kaufmann, S.H. (1999). MAMMALIAN CASPASES: Structure, Activation, Substrates, and Functions During Apoptosis. Annual Review of Biochemistry *68*, 383–424.

Eastmond, P.J. (2006). SUGAR-DEPENDENT1 Encodes a Patatin Domain Triacylglycerol Lipase That Initiates Storage Oil Breakdown in Germinating Arabidopsis Seeds. Plant Cell *18*, 665–675.

Eastwood, D.C., Floudas, D., Binder, M., Majcherczyk, A., Schneider, P., Aerts, A., Asiegbu, F.O., Baker, S.E., Barry, K., Bendiksby, M., et al. (2011). The Plant Cell Wall–Decomposing Machinery Underlies the Functional Diversity of Forest Fungi. Science *333*, 762–765.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl. Acids Res. *32*, 1792–1797.

Eitas, T.K., and Dangl, J.L. (2010). NB-LRR proteins: pairs, pieces, perception, partners, and pathways. Current Opinion in Plant Biology *13*, 472–477.

Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000). Predicting Subcellular Localization of Proteins Based on their N-terminal Amino Acid Sequence. Journal of Molecular Biology *300*, 1005–1016.

Emri, T., Molnár, Z., and Pócsi, I. (2005). The appearances of autolytic and apoptotic markers are concomitant but differently regulated in carbon-starving *Aspergillus nidulans* cultures. FEMS Microbiology Letters *251*, 297–303.

Enkerli, J., Felix, G., and Boller, T. (1999). The Enzymatic Activity of Fungal Xylanase Is Not Necessary for Its Elicitor Activity. Plant Physiol. *121*, 391–398.

Enright, A.J., Dongen, S.V., and Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. Nucl. Acids Res. *30*, 1575–1584.

Van der Ent, S., Van Wees, S.C.M., and Pieterse, C.M.J. (2009). Jasmonate signaling in plant interactions with resistance-inducing beneficial microbes. Phytochemistry *70*, 1581–1588.

Eschen-Lippold, L., Bethke, G., Palm-Forster, M.A.T., Pecher, P., Bauer, N., Glazebrook, J., Scheel, D., and Lee, J. (2012). MPK11--a fourth elicitor-responsive mitogen-activated protein kinase in *Arabidopsis thaliana*. Plant Signal Behav *7*, 1203–1205.

Van Esse, H.P., Bolton, M.D., Stergiopoulos, I., de Wit, P.J.G.M., and Thomma, B.P.H.J. (2007). The Chitin-Binding *Cladosporium fulvum* Effector Protein Avr4 Is a Virulence Factor. Molecular Plant-Microbe Interactions *20*, 1092–1101.

Van Esse, H.P., Klooster, J.W. van't, Bolton, M.D., Yadeta, K.A., Baarlen, P. van, Boeren, S., Vervoort, J., Wit, P.J.G.M. de, and Thomma, B.P.H.J. (2008). The *Cladosporium fulvum*

Virulence Protein Avr2 Inhibits Host Proteases Required for Basal Defense. Plant Cell *20*, 1948–1963.

Esteban, A., Popp, M.W., Vyas, V.K., Strijbis, K., Ploegh, H.L., and Fink, G.R. (2011). Fungal recognition is mediated by the association of dectin-1 and galectin-3 in macrophages. PNAS *108*, 14270–14275.

Eulgem, T., Rushton, P.J., Robatzek, S., and Somssich, I.E. (2000). The WRKY superfamily of plant transcription factors. Trends in Plant Science *5*, 199–206.

Evrard, A., Kumar, M., Lecourieux, D., Lucks, J., von Koskull-Doring, P., and Hirt, H. (2013). Regulation of the heat stress response in Arabidopsis by MPK6-targeted phosphorylation of the heat stress factor HsfA2. Peerj *1*.

Fan, Y., Pei, X., Guo, S., Zhang, Y., Luo, Z., Liao, X., and Pei, Y. (2010). Increased virulence using engineered protease-chitin binding domain hybrid expressed in the entomopathogenic fungus *Beauveria bassiana*. Microbial Pathogenesis *49*, 376–380.

Farnell, E., Rousseau, K., Thornton, D.J., Bowyer, P., and Herrick, S.E. (2012). Expression and secretion of *Aspergillus fumigatus* proteases are regulated in response to different protein substrates. Fungal Biology *116*, 1003–1012.

Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K., et al. (2010). The Pfam protein families database. Nucl. Acids Res. *38*, D211–D222.

Fiorilli, V., Lanfranco, L., and Bonfante, P. (2013). The expression of GintPT, the phosphate transporter of *Rhizophagus irregularis*, depends on the symbiotic status and phosphate availability. Planta *237*, 1267–1277.

Fitch, W.M. (2000). Homology: a personal view on some of the problems. Trends in Genetics *16*, 227–231.

Floudas, D., Binder, M., Riley, R., Barry, K., Blanchette, R.A., Henrissat, B., Martínez, A.T., Otillar, R., Spatafora, J.W., Yadav, J.S., et al. (2012). The Paleozoic Origin of Enzymatic Lignin Decomposition Reconstructed from 31 Fungal Genomes. Science *336*, 1715–1719.

Freitas, E., Aires, A., Rosa, E.A. de S., and Saavedra, M.J. (2013). Antibacterial activity and synergistic effect between watercress extracts, 2-phenylethyl isothiocyanate and antibiotics against 11 isolates of *Escherichia coli* from clinical and animal source. Letters in Applied Microbiology *57*, 266–273.

Fujikawa, T., Sakaguchi, A., Nishizawa, Y., Kouzai, Y., Minami, E., Yano, S., Koga, H., Meshi, T., and Nishimura, M. (2012). Surface α-1,3-Glucan Facilitates Fungal Stealth Infection by Interfering with Innate Immunity in Plants. PLoS Pathog *8*, e1002882.

Fujimoto, S.Y., Ohta, M., Usui, A., Shinshi, H., and Ohme-Takagi, M. (2000). Arabidopsis Ethylene-Responsive Element Binding Factors Act as Transcriptional Activators or Repressors of GCC Box–Mediated Gene Expression. Plant Cell *12*, 393–404.

Fullwood, M.J., Wei, C.-L., Liu, E.T., and Ruan, Y. (2009). Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. Genome Res. *19*, 521–532.

Futagami, T., Nakao, S., Kido, Y., Oka, T., Kajiwara, Y., Takashita, H., Omori, T., Furukawa, K., and Goto, M. (2011). Putative Stress Sensors WscA and WscB Are Involved in Hypo-Osmotic and Acidic pH Stress Tolerance in *Aspergillus nidulans*. Eukaryotic Cell *10*, 1504–1515.

Galagan, J.E., and Selker, E.U. (2004). RIP: the evolutionary cost of genome defense. Trends in Genetics *20*, 417–423.

Galagan, J.E., Calvo, S.E., Borkovich, K.A., Selker, E.U., Read, N.D., Jaffe, D., FitzHugh, W., Ma, L.-J., Smirnov, S., Purcell, S., et al. (2003). The genome sequence of the filamentous fungus *Neurospora crassa*. Nature *422*, 859–868.

Galletti, R., Ferrari, S., and Lorenzo, G.D. (2011). Arabidopsis MPK3 and MPK6 Play Different Roles in Basal and Oligogalacturonide- or Flagellin-Induced Resistance against *Botrytis cinerea*. Plant Physiol. *157*, 804–814.

Gao, Q.-M., Venugopal, S., Navarre, D., and Kachroo, A. (2011). Low Oleic Acid-Derived Repression of Jasmonic Acid-Inducible Defense Responses Requires the WRKY50 and WRKY51 Proteins. Plant Physiol. *155*, 464–476.

Garfinkel, D.J., Hedge, A.M., Youngren, S.D., and Copeland, T.D. (1991). Proteolytic processing of pol-TYB proteins from the yeast retrotransposon Ty1. J Virol *65*, 4573–4581.

Gaulin, E., Jauneau, A., Villalba, F., Rickauer, M., Esquerré-Tugayé, M.-T., and Bottin, A. (2002). The CBEL glycoprotein of *Phytophthora parasitica* var-nicotianae is involved in cell wall deposition and adhesion to cellulosic substrates. J Cell Sci *115*, 4565–4575.

Gaulin, E., Dramé, N., Lafitte, C., Torto-Alalibo, T., Martinez, Y., Ameline-Torregrosa, C., Khatib, M., Mazarguil, H., Villalba-Mateos, F., Kamoun, S., et al. (2006). Cellulose Binding Domains of a Phytophthora Cell Wall Protein Are Novel Pathogen-Associated Molecular Patterns. Plant Cell *18*, 1766–1777.

Gershenzon, J., and Dudareva, N. (2007). The function of terpene natural products in the natural world. Nat Chem Biol *3*, 408–414.

Ghimire, S.R., and Craven, K.D. (2011). Enhancement of Switchgrass (*Panicum virgatum* L.) Biomass Production under Drought Conditions by the Ectomycorrhizal Fungus *Sebacina vermifera*. Appl. Environ. Microbiol. *77*, 7063–7067.

Ghimire, S.R., Charlton, N.D., and Craven, K.D. (2009). The Mycorrhizal Fungus, *Sebacina vermifera*, Enhances Seed Germination and Biomass Production in Switchgrass (*Panicum virgatum* L). Bioenerg. Res. *2*, 51–58.

Gilbert, H.J. (2010). The Biochemistry and Structural Biology of Plant Cell Wall Deconstruction. Plant Physiol. *153*, 444–455.

Godfrey, D., Böhlenius, H., Pedersen, C., Zhang, Z., Emmersen, J., and Thordal-Christensen, H. (2010). Powdery mildew fungal effector candidates share N-terminal Y/F/WxC-motif. BMC Genomics *11*, 317.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., et al. (1996). Life with 6000 Genes. Science *274*, 546–567.

Gomord, V., Denmat, L.-A., Fitchette-Lainé, A.-C., Satiat-Jeunemaitre, B., Hawes, C., and Faye, L. (1997). The C-terminal HDEL sequence is sufficient for retention of secretory proteins in the endoplasmic reticulum (ER) but promotes vacuolar targeting of proteins that escape the ER. The Plant Journal *11*, 313–325.

Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. Mol Biol Evol *27*, 221–224.

Grigoriev, I.V., Cullen, D., Goodwin, S.B., Hibbett, D., Jeffries, T.W., Kubicek, C.P., Kuske, C., Magnuson, J.K., Martin, F., Spatafora, J.W., et al. (2011). Fueling the future with fungal genomics. Mycology: An International Journal on Fungal Biology *2*, 192–209.

Grigoriev, I.V., Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., Kuo, A., Minovitsky, S., Nikitin, R., Ohm, R.A., et al. (2012). The Genome Portal of the Department of Energy Joint Genome Institute. Nucleic Acids Res *40*, D26–D32.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst Biol *59*, 307–321.

Hamamouch, N., Li, C., Seo, P.J., Park, C.-M., and Davis, E.L. (2011). Expression of Arabidopsis pathogenesis-related genes during nematode infection. Molecular Plant Pathology *12*, 355–364.

Harris, P.V., Welner, D., McFarland, K.C., Re, E., Navarro Poulsen, J.-C., Brown, K., Salbo, R., Ding, H., Vlasenko, E., Merino, S., et al. (2010). Stimulation of Lignocellulosic Biomass Hydrolysis by Proteins of Glycoside Hydrolase Family 61: Structure and Function of a Large, Enigmatic Family. Biochemistry *49*, 3305–3316.

Hayakawa, Y., Ishikawa, E., Shoji, J., Nakano, H., and Kitamoto, K. (2011). Septum-directed secretion in the filamentous fungus *Aspergillus oryzae*. Molecular Microbiology *81*, 40–55.

He, P., Shan, L., and Sheen, J. (2007). Elicitation and suppression of microbe-associated molecular pattern-triggered immunity in plant–microbe interactions. Cellular Microbiology *9*, 1385–1396.

He, Y., Fukushige, H., Hildebrand, D.F., and Gan, S. (2002). Evidence Supporting a Role of Jasmonic Acid in Arabidopsis Leaf Senescence. Plant Physiol. *128*, 876–884.

Hector, J.M. (1938). Introduction to the botany of field crops. Vol. I Cereals. South African Agric. Ser. Vol. 16 xxxiv + 478 pp.

Heinisch, J.J., Dupres, V., Wilk, S., Jendretzki, A., and Dufrêne, Y.F. (2010). Single-Molecule Atomic Force Microscopy Reveals Clustering of the Yeast Plasma-Membrane Sensor Wsc1. PLoS ONE *5*, e11104.

Henry, C.M., and Deacon, J.W. (1981). Natural (non-pathogenic) death of the cortex of wheat and barley seminal roots, as evidenced by nuclear staining with acridine orange. Plant and Soil *v. 60(2) p. 255-274*.

Hibbett, D.S., Gilbert, L.-B., and Donoghue, M.J. (2000). Evolutionary instability of ectomycorrhizal symbioses in basidiomycetes. Nature *407*, 506–508.

Hickman, M.A., and Rusche, L.N. (2010). Transcriptional silencing functions of the yeast protein Orc1/Sir3 subfunctionalized after gene duplication. PNAS *107*, 19384–19389.

Higashi, K., Ishiga, Y., Inagaki, Y., Toyoda, K., Shiraishi, T., and Ichinose, Y. (2008). Modulation of defense signal transduction by flagellin-induced WRKY41 transcription factor in *Arabidopsis thaliana*. Mol Genet Genomics *279*, 303–312.

Hilbert, M., Voll, L.M., Ding, Y., Hofmann, J., Sharma, M., and Zuccaro, A. (2012). Indole derivative production by the root endophyte *Piriformospora indica* is not required for growth promotion but for biotrophic colonization of barley roots. New Phytologist *196*, 520–534.

Hirsh, A.E., and Fraser, H.B. (2001). Protein dispensability and rate of evolution. Nature *411*, 1046–1049.

Hiruma, K., Fukunaga, S., Bednarek, P., Piślewska-Bednarek, M., Watanabe, S., Narusaka, Y., Shirasu, K., and Takano, Y. (2013). Glutathione and tryptophan metabolism are required for Arabidopsis immunity during the hypersensitive response to hemibiotrophs. PNAS *110*, 9589–9594.

Hordijk, W., and Gascuel, O. (2005). Improving the efficiency of SPR moves in phylogenetic tree search methods based on maximum likelihood. Bioinformatics *21*, 4338–4347.

Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., and Nakai, K. (2007). WoLF PSORT: protein localization predictor. Nucl. Acids Res. *35*, W585–W587.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology *24*, 417–441.

Hou, X., Lee, L.Y.C., Xia, K., Yan, Y., and Yu, H. (2010). DELLAs Modulate Jasmonate Signaling via Competitive Binding to JAZs. Developmental Cell *19*, 884–894.

Howlett, B.J. (2006). Secondary metabolite toxins and nutrition of plant pathogenic fungi. Current Opinion in Plant Biology *9*, 371–375.

Hu, Y., Dong, Q., and Yu, D. (2012). Arabidopsis WRKY46 coordinates with WRKY70 and WRKY53 in basal resistance against pathogen *Pseudomonas syringae*. Plant Science *185–186*, 288–297.

Hu, Y., Chen, L., Wang, H., Zhang, L., Wang, F., and Yu, D. (2013). Arabidopsis transcription factor WRKY8 functions antagonistically with its interacting partner VQ9 to modulate salinity stress tolerance. The Plant Journal *74*, 730–745.

Huang, W., and Marth, G. (2008). EagleView: A genome assembly viewer for next-generation sequencing technologies. Genome Res. *18*, 1538–1543.

Hung, C.-Y., Seshan, K.R., Yu, J.-J., Schaller, R., Xue, J., Basrur, V., Gardner, M.J., and Cole, G.T. (2005). A Metalloproteinase of *Coccidioides posadasii* Contributes to Evasion of Host Detection. Infect. Immun. *73*, 6689–6703.

Jablonka-Shariff, A., and Boime, I. (2011). A dileucine determinant in the carboxyl terminal sequence of the LHβ subunit is implicated in the regulated secretion of lutropin from transfected GH3 cells. Molecular and Cellular Endocrinology *339*, 7–13.

Jacobs, S., Zechmann, B., Molitor, A., Trujillo, M., Petutschnig, E., Lipka, V., Kogel, K.-H., and Schäfer, P. (2011). Broad-Spectrum Suppression of Innate Immunity Is Required for Colonization of Arabidopsis Roots by the Fungus *Piriformospora indica*. Plant Physiol. *156*, 726–740.

Jiang, Y., Liang, G., and Yu, D. (2012). Activated Expression of WRKY57 Confers Drought Tolerance in Arabidopsis. Mol. Plant *5*, 1375–1388.

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

Jones, R.W., and Ospina-Giraldo, M. (2011). Novel Cellulose-Binding-Domain Protein in Phytophthora Is Cell Wall Localized. PLoS ONE *6*, e23555.

De Jonge, R., Peter van Esse, H., Kombrink, A., Shinya, T., Desaki, Y., Bours, R., van der Krol, S., Shibuya, N., Joosten, M.H.A.J., and Thomma, B.P.H.J. (2010). Conserved Fungal LysM Effector Ecp6 Prevents Chitin-Triggered Immunity in Plants. Science *329*, 953–955.

De Jonge, R., Bolton, M.D., and Thomma, B.P. (2011). How filamentous pathogens co-opt plants: the ins and outs of fungal effectors. Current Opinion in Plant Biology *14*, 400–406.

Julou, T., Burghardt, B., Gebauer, G., Berveiller, D., Damesin, C., and Selosse, M.-A. (2005). Mixotrophy in orchids: insights from a comparative study of green individuals and nonphotosynthetic individuals of *Cephalanthera damasonium*. New Phytologist *166*, 639–653.

Jung, H.W., Tschaplinski, T.J., Wang, L., Glazebrook, J., and Greenberg, J.T. (2009). Priming in Systemic Plant Immunity. Science *324*, 89–91.

Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. Cytogenetic and Genome Research *110*, 462–467.

Kamoun, S. (2006). A Catalogue of the Effector Secretome of Plant Pathogenic Oomycetes. Annual Review of Phytopathology *44*, 41–60.

Kämper, J., Kahmann, R., Bölker, M., Ma, L.-J., Brefort, T., Saville, B.J., Banuett, F., Kronstad, J.W., Gold, S.E., Müller, O., et al. (2006). Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. Nature *444*, 97–101.

Kaschuk, G., Kuyper, T.W., Leffelaar, P.A., Hungria, M., and Giller, K.E. (2009). Are the rates of photosynthesis stimulated by the carbon sink strength of rhizobial and arbuscular mycorrhizal symbioses? Soil Biology and Biochemistry *41*, 1233–1244.

Keller, N.P., Turner, G., and Bennett, J.W. (2005). Fungal secondary metabolism - from biochemistry to genomics. Nat. Rev. Microbiol. *3*, 937–947.

Kelley, L.A., and Sternberg, M.J.E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protocols *4*, 363–371.

Kent, W.J. (2002). BLAT—The BLAST-Like Alignment Tool. Genome Res. *12*, 656–664.

Keon, J., Antoniw, J., Carzaniga, R., Deller, S., Ward, J.L., Baker, J.M., Beale, M.H., Hammond-Kosack, K., and Rudd, J.J. (2007). Transcriptional Adaptation of *Mycosphaerella*

graminicola to Programmed Cell Death (PCD) of Its Susceptible Wheat Host. Molecular Plant-Microbe Interactions *20*, 178–193.

Kesarwani, M., Yoo, J., and Dong, X. (2007). Genetic Interactions of TGA Transcription Factors in the Regulation of Pathogenesis-Related Genes and Disease Resistance in Arabidopsis. Plant Physiol. *144*, 336–346.

Kettles, G.J., Drurey, C., Schoonbeek, H., Maule, A.J., and Hogenhout, S.A. (2013). Resistance of *Arabidopsis thaliana* to the green peach aphid, *Myzus persicae*, involves camalexin and is regulated by microRNAs. New Phytologist n/a–n/a.

Khan, M.A.S., Chock, P.B., and Stadtman, E.R. (2005). Knockout of caspase-like gene, YCA1, abrogates apoptosis and elevates oxidized proteins in *Saccharomyces cerevisiae*. PNAS *102*, 17326–17331.

Khatabi, B., Molitor, A., Lindermayr, C., Pfiffi, S., Durner, J., von Wettstein, D., Kogel, K.-H., and Schäfer, P. (2012). Ethylene Supports Colonization of Plant Roots by the Mutualistic Fungus *Piriformospora indica*. PLoS ONE *7*, e35502.

Kikot, G.E., Hours, R.A., and Alconada, T.M. (2009). Contribution of cell wall degrading enzymes to pathogenesis of Fusarium graminearum: a review. Journal of Basic Microbiology *49*, 231–241.

Kim, J.-Y., Park, S.-C., Hwang, I., Cheong, H., Nah, J.-W., Hahm, K.-S., and Park, Y. (2009). Protease Inhibitors from Plants with Antimicrobial Activity. Int J Mol Sci *10*, 2860–2872.

Kim, K.-C., Lai, Z., Fan, B., and Chen, Z. (2008). Arabidopsis WRKY38 and WRKY62 Transcription Factors Interact with Histone Deacetylase 19 in Basal Defense. Plant Cell *20*, 2357–2371.

Kliebenstein, D.J. (2004). Secondary metabolites and plant/environment interactions: a view through *Arabidopsis thaliana* tinged glasses. Plant, Cell & Environment *27*, 675–684.

Kliebenstein, D.J. (2012). Plant Defense Compounds: Systems Approaches to Metabolic Analysis. Annual Review of Phytopathology *50*, 155–173.

Kloppholz, S., Kuhn, H., and Requena, N. (2011). A Secreted Fungal Effector of Glomus intraradices Promotes Symbiotic Biotrophy. Current Biology *21*, 1204–1209.

Klug, W.S., Cummings, M.R., Spencer, C.A., and Palladino, M.A. (2011). Concepts of Genetics, Books a la Carte Plus MasteringGenetics -- Access Card Package (Benjamin Cummings).

Koch, A.M., Antunes, P.M., Barto, E.K., Cipollini, D., Mummey, D.L., and Klironomos, J.N. (2011). The effects of arbuscular mycorrhizal (AM) fungal and garlic mustard introductions on native AM fungal diversity. Biol Invasions *13*, 1627–1639.

Kohorn, B.D., and Kohorn, S.L. (2012). The Cell Wall-Associated Kinases, WAKs, Regulate Cell Expansion and the Stress Response. In Receptor-like Kinases in Plants, F. Tax, and B. Kemmerling, eds. (Springer Berlin Heidelberg), pp. 109–124.

Koiwa, H., Bressan, R.A., and Hasegawa, P.M. (1997). Regulation of protease inhibitors and plant defense. Trends in Plant Science *2*, 379–384.

Kombrink, E., and Somssich, I.E. (1997). Pathogenesis-Related Proteins and Plant Defense. In Plant Relationships, P.D.G.C. Carroll, and P.D.P. Tudzynski, eds. (Springer Berlin Heidelberg), pp. 107–128.

Koschorreck, M., Fischer, M., Barth, S., and Pleiss, J. (2005). How to find soluble proteins: a comprehensive analysis of alpha/beta hydrolases for recombinant expression in *E. coli*. BMC Genomics *6*, 49.

Kottke, I., Beiter, A., Weiss, M., Haug, I., Oberwinkler, F., and Nebel, M. (2003). Heterobasidiomycetes form symbiotic associations with hepatics: Jungermanniales have sebacinoid mycobionts while *Aneura pinguis* (Metzgeriales) is associated with a *Tulasnella* species. Mycological Research *107*, 957–968.

Kretschmer, M., Leroch, M., Mosbach, A., Walker, A.-S., Fillinger, S., Mernke, D., Schoonbeek, H.-J., Pradier, J.-M., Leroux, P., De Waard, M.A., et al. (2009). Fungicide-Driven Evolution and Molecular Basis of Multidrug Resistance in Field Populations of the Grey Mould Fungus *Botrytis cinerea*. PLoS Pathog *5*, e1000696.

Kumar, M., Yadav, V., Kumar, H., Sharma, R., Singh, A., Tuteja, N., and Johri, A.K. (2011). *Piriformospora indica* enhances plant growth by transferring phosphate. Plant Signaling & Behavior *6*, 723–725.

Lackner, G., Misiek, M., Braesel, J., and Hoffmeister, D. (2012). Genome mining reveals the evolutionary origin and biosynthetic potential of basidiomycete polyketide synthases. Fungal Genetics and Biology *49*, 996–1003.

Lahrmann, U., and Zuccaro, A. (2012). *Opprimo ergo sum* - Evasion and Suppression in the Root Endophytic Fungus *Piriformospora indica*. Molecular Plant-Microbe Interactions *25*, 727–737.

Lahrmann, U., Ding, Y., Banhara, A., Rath, M., Hajirezaei, M.R., Döhlemann, S., Wirén, N. von, Parniske, M., and Zuccaro, A. (2013). Host-related metabolic cues affect colonization strategies of a root endophyte. PNAS *110*, 13965–13970.

Lambrix, V., Reichelt, M., Mitchell-Olds, T., Kliebenstein, D.J., and Gershenzon, J. (2001). The Arabidopsis Epithiospecifier Protein Promotes the Hydrolysis of Glucosinolates to Nitriles and Influences *Trichoplusia ni* Herbivory. Plant Cell *13*, 2793–2807.

Laslett, D., and Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucl. Acids Res. *32*, 11–16.

Laurie, J.D., Ali, S., Linning, R., Mannhaupt, G., Wong, P., Güldener, U., Münsterkötter, M., Moore, R., Kahmann, R., Bakkeren, G., et al. (2012). Genome Comparison of Barley and Maize Smut Fungi Reveals Targeted Loss of RNA Silencing Components and Species-Specific Presence of Transposable Elements. Plant Cell *24*, 1733–1745.

Le, S.Q., and Gascuel, O. (2008). An Improved General Amino Acid Replacement Matrix. Mol Biol Evol *25*, 1307–1320.

Lechner, M., Findeiß, S., Steiner, L., Marz, M., Stadler, P.F., and Prohaska, S.J. (2011). Proteinortho: Detection of (Co-)orthologs in large-scale analysis. BMC Bioinformatics *12*, 124.

Lee, R.E.C., Brunette, S., Puente, L.G., and Megeney, L.A. (2010). Metacaspase Yca1 is required for clearance of insoluble protein aggregates. PNAS *107*, 13348–13353.

Lo Leggio, L., Leggio, L.L., Welner, D., and Maria, L.D. (2012). A structural overview of GH61 proteins – fungal cellulose degrading polysaccharide monooxygenases. Comp & Struct Biotech *2*.

Leon-Reyes, A., Du, Y., Koornneef, A., Proietti, S., Körbes, A.P., Memelink, J., Pieterse, C.M.J., and Ritsema, T. (2010). Ethylene Signaling Renders the Jasmonate Response of *Arabidopsis* Insensitive to Future Suppression by Salicylic Acid. Molecular Plant-Microbe Interactions *23*, 187–197.

Letunic, I., Doerks, T., and Bork, P. (2009). SMART 6: recent updates and new developments. Nucl. Acids Res. *37*, D229–D232.

Lévesque, C.A., Brouwer, H., Cano, L., Hamilton, J.P., Holt, C., Huitema, E., Raffaele, S., Robideau, G.P., Thines, M., Win, J., et al. (2010). Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. Genome Biology *11*, R73.

Lewis, D.H. (1973). Concepts in Fungal Nutrition and the Origin of Biotrophy. Biological Reviews *48*, 261–277.

Li, H., Dai, X., and Zhao, X. (2008). A nearest neighbor approach for automated transporter prediction and categorization from protein sequences. Bioinformatics *24*, 1129–1136.

Li, J., Yu, L., Tian, Y., and Zhang, K.-Q. (2012). Molecular Evolution of the Deuterolysin (M35) Family Genes in *Coccidioides*. PLoS ONE *7*, e31536.

Li, J., Besseau, S., Törönen, P., Sipari, N., Kollist, H., Holm, L., and Palva, E.T. (2013). Defense-related transcription factors WRKY70 and WRKY54 modulate osmotic stress tolerance by regulating stomatal aperture in Arabidopsis. New Phytologist *200*, 457–472.

Liljeroth, E., and Bryngelsson, T. (2001). DNA fragmentation in cereal roots indicative of programmed root cortical cell death. Physiologia Plantarum *111*, 365–372.

Lilly, W.W., Stajich, J.E., Pukkila, P.J., Wilke, S.K., Inoguchi, N., and Gathman, A.C. (2008). An expanded family of fungalysin extracellular metallopeptidases of *Coprinopsis cinerea*. Mycological Research *112*, 389–398.

Liu, F., Jiang, H., Ye, S., Chen, W.-P., Liang, W., Xu, Y., Sun, B., Sun, J., Wang, Q., Cohen, J.D., et al. (2010). The Arabidopsis P450 protein CYP82C2 modulates jasmonate-induced root growth inhibition, defense gene expression and indole glucosinolate biosynthesis. Cell Res *20*, 539–552.

Liu, Y.-W., Han, C.-H., Lee, M.-H., Hsu, F.-L., and Hou, W.-C. (2003). Patatin, the Tuber Storage Protein of Potato (*Solanum tuberosum* L.), Exhibits Antioxidant Activity in Vitro. J. Agric. Food Chem. *51*, 4389–4393.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. Methods *25*, 402–408.

Lohse, M., Nunes-Nesi, A., Krüger, P., Nagel, A., Hannemann, J., Giorgi, F.M., Childs, L., Osorio, S., Walther, D., Selbig, J., et al. (2010). Robin: An Intuitive Wizard Application for R-Based Expression Microarray Quality Assessment and Analysis. Plant Physiol. *153*, 642–651.

Van Loon, L., and Van Strien, E. (1999). The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. Physiological and Molecular Plant Pathology *55*, 85–97.

López-Berges, M.S., Rispail, N., Prados-Rosales, R.C., and Pietro, A.D. (2010). A Nitrogen Response Pathway Regulates Virulence Functions in *Fusarium oxysporum* via the Protein Kinase TOR and the bZIP Protein MeaB. Plant Cell *22*, 2459–2475.

Lopez-Raez, J.A., Verhage, A., Fernandez, I., Garcia, J.M., Azcon-Aguilar, C., Flors, V., and Pozo, M.J. (2010). Hormonal and transcriptional profiles highlight common and differential host responses to arbuscular mycorrhizal fungi and the regulation of the oxylipin pathway. J Exp Bot *61*, 2589–2601.

Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. Nucl. Acids Res. *25*, 0955–0964.

Lukashin, A.V., and Borodovsky, M. (1998). GeneMark.hmm: New solutions for gene finding. Nucl. Acids Res. *26*, 1107–1115.

Luna, E., Pastor, V., Robert, J., Flors, V., Mauch-Mani, B., and Ton, J. (2011). Callose Deposition: A Multifaceted Plant Defense Response. Molecular Plant-Microbe Interactions *24*, 183–193.

Maddi, A., Dettman, A., Fu, C., Seiler, S., and Free, S.J. (2012). WSC-1 and HAM-7 Are MAK-1 MAP Kinase Pathway Sensors Required for Cell Wall Integrity and Hyphal Fusion in Neurospora crassa. PLoS ONE *7*, e42374.

Maleck, K., Levine, A., Eulgem, T., Morgan, A., Schmid, J., Lawton, K.A., Dangl, J.L., and Dietrich, R.A. (2000). The transcriptome of *Arabidopsis thaliana* during systemic acquired resistance. Nat Genet *26*, 403–410.

Manners, J.M., Penninckx, I.A.M.A., Vermaere, K., Kazan, K., Brown, R.L., Morgan, A., Maclean, D.J., Curtis, M.D., Cammue, B.P.A., and Broekaert, W.F. (1998). The promoter of the plant defensin gene PDF1.2 from Arabidopsis is systemically activated by fungal pathogens and responds to methyl jasmonate but not to salicylic acid. Plant Mol Biol *38*, 1071–1080.

Mao, G., Meng, X., Liu, Y., Zheng, Z., Chen, Z., and Zhang, S. (2011). Phosphorylation of a WRKY Transcription Factor by Two Pathogen-Responsive MAPKs Drives Phytoalexin Biosynthesis in Arabidopsis. Plant Cell *23*, 1639–1653.

Mardis, E.R. (2008). Next-Generation DNA Sequencing Methods. Annual Review of Genomics and Human Genetics *9*, 387–402.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., et al. (2005). Genome Sequencing in Open Microfabricated High Density Picoliter Reactors. Nature *437*, 376–380.

Markowitz, V.M., Chen, I.-M.A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., Ratner, A., Jacob, B., Huang, J., Williams, P., et al. (2011). IMG: the integrated microbial genomes database and comparative analysis system. Nucleic Acids Research *40*, D115–D122.

Marshall, R., Kombrink, A., Motteram, J., Loza-Reyes, E., Lucas, J., Hammond-Kosack, K.E., Thomma, B.P.H.J., and Rudd, J.J. (2011). Analysis of Two in Planta Expressed LysM Effector Homologs from the Fungus *Mycosphaerella graminicola* Reveals Novel Functional Properties and Varying Contributions to Virulence on Wheat. Plant Physiol. *156*, 756–769.

Marthey, S., Aguileta, G., Rodolphe, F., Gendrault, A., Giraud, T., Fournier, E., Lopez-Villavicencio, M., Gautier, A., Lebrun, M.-H., and Chiapello, H. (2008). FUNYBASE: a FUNgal phYlogenomic dataBASE. BMC Bioinformatics *9*, 456.

Martin, F., Aerts, A., Ahrén, D., Brun, A., Danchin, E.G.J., Duchaussoy, F., Gibon, J., Kohler, A., Lindquist, E., Pereda, V., et al. (2008). The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. Nature *452*, 88–92.

Martin, F., Kohler, A., Murat, C., Balestrini, R., Coutinho, P.M., Jaillon, O., Montanini, B., Morin, E., Noel, B., Percudani, R., et al. (2010). Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. Nature *464*, 1033–1038.

Matsushima, R., Kondo, M., Nishimura, M., and Hara-Nishimura, I. (2003). A novel ER-derived compartment, the ER body, selectively accumulates a β-glucosidase with an ER-retention signal in Arabidopsis. The Plant Journal *33*, 493–502.

McKendrick, S.L., Leake, J.R., Taylor, D.L., and Read, D.J. (2002). Symbiotic germination and development of the myco-heterotrophic orchid *Neottia nidus-avis* in nature and its requirement for locally distributed *Sebacina* spp. New Phytologist *154*, 233–247.

Medema, M.H., Blin, K., Cimermancic, P., Jager, V. de, Zakrzewski, P., Fischbach, M.A., Weber, T., Takano, E., and Breitling, R. (2011). antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucl. Acids Res. *39*, W339–W346.

Metzker, M.L. (2010). Sequencing technologies — the next generation. Nat Rev Genet *11*, 31–46.

Miao, Y., Jiang, J., Ren, Y., and Zhao, Z. (2013). The single-stranded DNA binding protein WHIRLY1 represses WRKY53 expression and delays leaf senescence in a developmental stage-dependent manner in *Arabidopsis thaliana*. Plant Physiol. pp.113.223412.

Micali, C.O., Neumann, U., Grunewald, D., Panstruga, R., and O'Connell, R. (2011). Biogenesis of a specialized plant–fungal interface during host cell internalization of *Golovinomyces orontii* haustoria. Cellular Microbiology *13*, 210–226.

Mignery, G.A., Pikaard, C.S., and Park, W.D. (1988). Molecular characterization of the patatin multigene family of potato. Gene *62*, 27–44.

Mikkelsen, M.D., Hansen, C.H., Wittstock, U., and Halkier, B.A. (2000). Cytochrome P450 CYP79B2 from Arabidopsis Catalyzes the Conversion of Tryptophan to Indole-3-acetaldoxime, a Precursor of Indole Glucosinolates and Indole-3-acetic Acid. J. Biol. Chem. *275*, 33712–33717.

Miya, A., Albert, P., Shinya, T., Desaki, Y., Ichimura, K., Shirasu, K., Narusaka, Y., Kawakami, N., Kaku, H., and Shibuya, N. (2007). CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in Arabidopsis. PNAS *104*, 19613–19618.

Mohnen, D., Bar-Peled, M., and Somerville, C. (2008). Biosynthesis of plant cell walls. Biomass Recalcitrance. Oxford: Blackwell Publishing 94–187.

Møldrup, M.E., Salomonsen, B., Geu-Flores, F., Olsen, C.E., and Halkier, B.A. (2013). De novo genetic engineering of the camalexin biosynthetic pathway. Journal of Biotechnology *167*, 296–301.

Morin, E., Kohler, A., Baker, A.R., Foulongne-Oriol, M., Lombard, V., Nagye, L.G., Ohm, R.A., Patyshakuliyeva, A., Brun, A., Aerts, A.L., et al. (2012). Genome sequence of the button mushroom *Agaricus bisporus* reveals mechanisms governing adaptation to a humic-rich ecological niche. PNAS *109*, 17501–17506.

Moyersoen, B. (2006). *Pakaraimaea dipterocarpacea* is ectomycorrhizal, indicating an ancient Gondwanaland origin for the ectomycorrhizal habit in Dipterocarpaceae. New Phytologist *172*, 753–762.

Murphy, C., Powlowski, J., Wu, M., Butler, G., and Tsang, A. (2011). Curation of characterized glycoside hydrolases of Fungal origin. Database *2011*, bar020–bar020.

Muszewska, A., Hoffman-Sommer, M., and Grynberg, M. (2011). LTR Retrotransposons in Fungi. PLoS ONE *6*, e29425.

Nadal, M., Garcia-Pedrajas, M.D., and Gold, S.E. (2010). The *snf1* Gene of *Ustilago maydis* Acts as a Dual Regulator of Cell Wall Degrading Enzymes. Phytopathology *100*, 1364–1372.

Nafisi, M., Goregaoker, S., Botanga, C.J., Glawischnig, E., Olsen, C.E., Halkier, B.A., and Glazebrook, J. (2007). Arabidopsis Cytochrome P450 Monooxygenase 71A13 Catalyzes the Conversion of Indole-3-Acetaldoxime in Camalexin Synthesis. Plant Cell *19*, 2039–2052.

Nagendran, S., Hallen-Adams, H.E., Paper, J.M., Aslam, N., and Walton, J.D. (2009). Reduced genomic potential for secreted plant cell-wall-degrading enzymes in the ectomycorrhizal fungus *Amanita bisporigera*, based on the secretome of *Trichoderma reesei*. Fungal Genetics and Biology *46*, 427–435.

Nagy, R., Drissner, D., Amrhein, N., Jakobsen, I., and Bucher, M. (2009). Mycorrhizal phosphate uptake pathway in tomato is phosphorus-repressible and transcriptionally regulated. New Phytologist *181*, 950–959.

Nakayashiki, H. (2011). The Trickster in the genome: contribution and control of transposable elements. Genes to Cells *16*, 827–841.

Naumann, T.A. (2011). Modification of recombinant maize ChitA chitinase by fungal chitinase-modifying proteins. Molecular Plant Pathology *12*, 365–372.

Naumann, T.A., and Price, N.P.J. (2012). Truncation of class IV chitinases from Arabidopsis by secreted fungal proteases. Molecular Plant Pathology *13*, 1135–1139.

Naur, P., Petersen, B.L., Mikkelsen, M.D., Bak, S., Rasmussen, H., Olsen, C.E., and Halkier, B.A. (2003). CYP83A1 and CYP83B1, Two Nonredundant Cytochrome P450 Enzymes

Metabolizing Oximes in the Biosynthesis of Glucosinolates in Arabidopsis. Plant Physiol. *133*, 63–72.

Nawrath, C., and Métraux, J.-P. (1999). Salicylic Acid Induction–Deficient Mutants of Arabidopsis Express PR-2 and PR-5 and Accumulate High Levels of Camalexin after Pathogen Inoculation. Plant Cell *11*, 1393–1404.

Nguyen, Q.B., Itoh, K., Van Vu, B., Tosa, Y., and Nakayashiki, H. (2011). Simultaneous silencing of endo-β-1,4 xylanase genes reveals their roles in the virulence of *Magnaporthe oryzae*. Molecular Microbiology *81*, 1008–1019.

Nielsen, H., Engelbrecht, J., Brunak, S., and Heijne, G. von (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. *10*, 1–6.

Nishida, H., Nagatsuka, Y., and Sugiyama, J. (2011). Draft genome sequencing of the enigmatic basidiomycete *Mixia osmundae*. The Journal of General and Applied Microbiology *57*, 63–67.

Nitsche, B.M., Jørgensen, T.R., Akeroyd, M., Meyer, V., and Ram, A.F. (2012). The carbon starvation response of *Aspergillus niger* during submerged cultivation: Insights from the transcriptome and secretome. BMC Genomics *13*, 380.

O'Connell, R.J., Thon, M.R., Hacquard, S., Amyotte, S.G., Kleemann, J., Torres, M.F., Damm, U., Buiate, E.A., Epstein, L., Alkan, N., et al. (2012). Lifestyle transitions in plant pathogenic Colletotrichum fungi deciphered by genome and transcriptome analyses. Nat Genet *44*, 1060–1065.

Ogura-Tsujita, Y., Gebauer, G., Hashimoto, T., Umata, H., and Yukawa, T. (2009). Evidence for novel and specialized mycorrhizal parasitism: the orchid *Gastrodia confusa* gains carbon from saprotrophic Mycena. Proc Biol Sci *276*, 761–767.

Oldroyd, G.E., and Downie, J.A. (2006). Nuclear calcium changes at the core of symbiosis signalling. Current Opinion in Plant Biology *9*, 351–357.

Oñate-Sánchez, L., and Singh, K.B. (2002). Identification of Arabidopsis Ethylene-Responsive Element Binding Factors with Distinct Induction Kinetics after Pathogen Infection. Plant Physiol. *128*, 1313–1322.

Panchuk, I.I., Volkov, R.A., and Schöffl, F. (2002). Heat Stress- and Heat Shock Transcription Factor-Dependent Expression and Activity of Ascorbate Peroxidase in Arabidopsis. Plant Physiol. *129*, 838–853.

Pandey, S.P., Roccaro, M., Schön, M., Logemann, E., and Somssich, I.E. (2010). Transcriptional reprogramming regulated by WRKY18 and WRKY40 facilitates powdery mildew infection of Arabidopsis. The Plant Journal *64*, 912–923.

Park, J., Park, J., Jang, S., Kim, S., Kong, S., Choi, J., Ahn, K., Kim, J., Lee, S., Kim, S., et al. (2008). FTFD: an informatics pipeline supporting phylogenomic analysis of fungal transcription factors. Bioinformatics *24*, 1024–1025.

Parra, G., Bradnam, K., Ning, Z., Keane, T., and Korf, I. (2009). Assessing the gene space in draft genomes. Nucl. Acids Res. *37*, 289–297.

Parra, M.E., Evans, C.B., and Taylor, D.W. (1991). Identification of *Plasmodium falciparum* histidine-rich protein 2 in the plasma of humans with malaria. J Clin Microbiol *29*, 1629–1634.

Paungfoo-Lonhienne, C., Schmidt, S., and Lonhienne, T.G.A. (2010a). Uptake of non-pathogenic E. coli by Arabidopsis induces down-regulation of heat shock proteins. Plant Signaling & Behavior *5*, 1626–1628.

Paungfoo-Lonhienne, C., Rentsch, D., Robatzek, S., Webb, R.I., Sagulenko, E., Näsholm, T., Schmidt, S., and Lonhienne, T.G.A. (2010b). Turning the Table: Plants Consume Microbes as a Source of Nutrients. PLoS ONE *5*, e11915.

Pel, M.J.C., and Pieterse, C.M.J. (2012). Microbial recognition and evasion of host immunity. J. Exp. Bot.

Pelham, H.R.B. (1990). The retention signal for soluble proteins of the endoplasmic reticulum. Trends in Biochemical Sciences *15*, 483–486.

Perotto, S., Angelini, P., Bianciotto, V., Bonfante, P., Girlanda, M., Kull, T., Mello, A., Pecoraro, L., Perini, C., Persiani, A.M., et al. (2012). Interactions of fungi with other organisms. Plant Biosystems - An International Journal Dealing with All Aspects of Plant Biology *0*, 1–11.

Peškan-Berghöfer, T., Shahollari, B., Giong, P.H., Hehl, S., Markert, C., Blanke, V., Kost, G., Varma, A., and Oelmüller, R. (2004). Association of *Piriformospora indica* with *Arabidopsis thaliana* roots represents a novel system to study beneficial plant–microbe interactions and involves early plant protein modifications in the endoplasmic reticulum and at the plasma membrane. Physiologia Plantarum *122*, 465–477.

Pham, G.H., Kumari, R., Singh, A., Malla, R., Prasad, R., Sachdev, M., Kaldorf, M., Buscot, F., Oelmüller, R., Hampp, R., et al. (2008). Axenic Culture of Symbiotic Fungus *Piriformospora indica*. In Plant Surface Microbiology, A. Varma, L. Abbott, D. Werner, and R. Hampp, eds. (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 593–613.

Phillippy, A.M., Schatz, M.C., and Pop, M. (2008). Genome assembly forensics: finding the elusive mis-assembly. Genome Biol *9*, R55.

Phillips, C.M., Beeson, W.T., Cate, J.H., and Marletta, M.A. (2011). Cellobiose Dehydrogenase and a Copper-Dependent Polysaccharide Monooxygenase Potentiate Cellulose Degradation by Neurospora crassa. ACS Chem. Biol. *6*, 1399–1406.

Di Pietro, A., Roncero, M.I.G., and Roldán, M.C.R. (2009). From Tools of Survival to Weapons of Destruction: The Role of Cell Wall-Degrading Enzymes in Plant Infection. In Plant Relationships, P.D.H.B. Deising, ed. (Springer Berlin Heidelberg), pp. 181–200.

Pitzschke, A., Schikora, A., and Hirt, H. (2009). MAPK cascade signalling networks in plant defence. Current Opinion in Plant Biology *12*, 421–426.

Plett, J.M., and Martin, F. (2011). Blurred boundaries: lifestyle lessons from ectomycorrhizal fungal genomes. Trends in Genetics *27*, 14–22.

Plett, J.M., Kemppainen, M., Kale, S.D., Kohler, A., Legué, V., Brun, A., Tyler, B.M., Pardo, A.G., and Martin, F. (2011). A Secreted Effector Protein of *Laccaria bicolor* Is Required for Symbiosis Development. Current Biology *21*, 1197–1203.

Popescu, S.C., Popescu, G.V., Bachan, S., Zhang, Z., Gerstein, M., Snyder, M., and Dinesh-Kumar, S.P. (2009). MAPK target networks in *Arabidopsis thaliana* revealed using functional protein microarrays. Genes Dev. *23*, 80–92.

Pozo, M.J., Jung, S.C., López-Ráez, J.A., and Azcón-Aguilar, C. (2010). Impact of Arbuscular Mycorrhizal Symbiosis on Plant Response to Biotic Stress: The Role of Plant Defence Mechanisms. In Arbuscular Mycorrhizas: Physiology and Function, H. Koltai, and Y. Kapulnik, eds. (Springer Netherlands), pp. 193–207.

Price, A.L., Jones, N.C., and Pevzner, P.A. (2005). De novo identification of repeat families in large genomes. Bioinformatics *21*, i351–i358.

Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res *35*, D61–D65.

Qiang, X., Weiss, M., Kogel, K.-H., and Schäfer, P. (2012a). *Piriformospora indica*—a mutualistic basidiomycete with an exceptionally large plant host range. Molecular Plant Pathology *13*, 508–518.

Qiang, X., Zechmann, B., Reitz, M.U., Kogel, K.-H., and Schäfer, P. (2012b). The Mutualistic Fungus *Piriformospora indica* Colonizes Arabidopsis Roots by Inducing an Endoplasmic Reticulum Stress–Triggered Caspase-Dependent Cell Death. Plant Cell *24*, 794–809.

Quettier, A.-L., and Eastmond, P.J. (2009). Storage oil hydrolysis during early seedling growth. Plant Physiology and Biochemistry *47*, 485–490.

R Development Core Team (2011). R: A Language and Environment for Statistical Computing (Vienna, Austria).

Raffaele, S., Win, J., Cano, L.M., and Kamoun, S. (2010). Analyses of genome architecture and gene expression reveal novel candidate virulence factors in the secretome of *Phytophthora infestans*. BMC Genomics *11*, 637.

Ramesh, M.V., and Kolattukudy, P.E. (1996). Disruption of the serine proteinase gene (sep) in *Aspergillus flavus* leads to a compensatory increase in the expression of a metalloproteinase gene (mep20). J. Bacteriol. *178*, 3899–3907.

Rappleye, C.A., Eissenberg, L.G., and Goldman, W.E. (2007). *Histoplasma capsulatum* α-(1,3)-glucan blocks innate immune recognition by the β-glucan receptor. Proc Natl Acad Sci U S A *104*, 1366–1370.

Rasmussen, H.N., and Rasmussen, F.N. (2009). Orchid mycorrhiza: implications of a mycophagous life style. Oikos *118*, 334–345.

Rawlings, N.D., Barrett, A.J., and Bateman, A. (2011). MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Research *40*, D343–D350.

Regvar, M., Vogel, K., Irgel, N., Wraber, T., Hildebrandt, U., Wilde, P., and Bothe, H. (2003). Colonization of pennycresses (Thlaspi spp.) of the Brassicaceae by arbuscular mycorrhizal fungi. Journal of Plant Physiology *160*, 615–626.

Ren, J., Wen, L., Gao, X., Jin, C., Xue, Y., and Yao, X. (2009). DOG 1.0: illustrator of protein domain structures. Cell Res *19*, 271–273.

Ren, X., Chen, Z., Liu, Y., Zhang, H., Zhang, M., Liu, Q., Hong, X., Zhu, J.-K., and Gong, Z. (2010). ABO3, a WRKY transcription factor, mediates plant responses to abscisic acid and drought tolerance in Arabidopsis. The Plant Journal *63*, 417–429.

Rodriguez, R.J., Redman, R.S., and Henson, J.M. (2004). The Role of Fungal Symbioses in the Adaptation of Plants to High Stress Environments. Mitigation and Adaptation Strategies for Global Change *9*, 261–272.

Rossel, J.B., Wilson, I.W., and Pogson, B.J. (2002). Global Changes in Gene Expression in Response to High Light in Arabidopsis. Plant Physiol. *130*, 1109–1120.

Roux, M., Schwessinger, B., Albrecht, C., Chinchilla, D., Jones, A., Holton, N., Malinovsky, F.G., Tör, M., Vries, S. de, and Zipfel, C. (2011). The Arabidopsis Leucine-Rich Repeat Receptor–Like Kinases BAK1/SERK3 and BKK1/SERK4 Are Required for Innate Immunity to Hemibiotrophic and Biotrophic Pathogens. Plant Cell *23*, 2440–2455.

Rudolph, M., Schlereth, A., Körner, M., Feussner, K., Berndt, E., Melzer, M., Hornung, E., and Feussner, I. (2011). The lipoxygenase-dependent oxygenation of lipid body membranes is promoted by a patatin-type phospholipase in cucumber cotyledons. J. Exp. Bot. *62*, 749–760.

Ruijter, G.J., and Visser, J. (1997). Carbon repression in Aspergilli. FEMS Microbiol. Lett. *151*, 103–114.

Saavedra, M.J., Borges, A., Dias, C., Aires, A., Bennett, R.N., Rosa, E.S., and Simões, M. (2010). Antimicrobial Activity of Phenolics and Glucosinolate Hydrolysis Products and their Synergy with Streptomycin against Pathogenic Bacteria. Medicinal Chemistry *6*, 174–183.

Saavedra, M.J., Dias, C.S.P., Martinez-Murcia, A., Bennett, R.N., Aires, A., and Rosa, E.A.S. (2012). Antibacterial Effects of Glucosinolate-Derived Hydrolysis Products Against *Enterobacteriaceae* and *Enterococci* Isolated from Pig Ileum Segments. Foodborne Pathogens and Disease *9*, 338–345.

Sacristán, S., Vigouroux, M., Pedersen, C., Skamnioti, P., Thordal-Christensen, H., Micali, C., Brown, J.K.M., and Ridout, C.J. (2009). Coevolution between a Family of Parasite Virulence Effectors and a Class of LINE-1 Retrotransposons. PLoS ONE *4*, e7463.

Saier, M.H., Yen, M.R., Noto, K., Tamang, D.G., and Elkan, C. (2009). The Transporter Classification Database: recent advances. Nucl. Acids Res. *37*, D274–D278.

Salamov, A.A., and Solovyev, V.V. (2000). Ab initio Gene Finding in Drosophila Genomic DNA. Genome Res. *10*, 516–522.

Saloheimo, M., Paloheimo, M., Hakola, S., Pere, J., Swanson, B., Nyyssönen, E., Bhatia, A., Ward, M., and Penttilä, M. (2002). Swollenin, a *Trichoderma reesei* protein with sequence similarity to the plant expansins, exhibits disruption activity on cellulosic materials. European Journal of Biochemistry *269*, 4202–4211.

Sambrook, J., Fritsch, E., and Maniatis, T. (1989). Molecular cloning: A laboratory manual+ Cold Spring Harbor (New York: Cold Spring Harbor Laboratory Press).

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. PNAS *74*, 5463–5467.

Scarpeci, T.E., Zanor, M.I., Mueller-Roeber, B., and Valle, E.M. (2013). Overexpression of AtWRKY30 enhances abiotic stress tolerance during early growth stages in *Arabidopsis thaliana*. Plant Mol Biol *83*, 265–277.

Schäfer, P., Pfiffi, S., Voll, L.M., Zajic, D., Chandler, P.M., Waller, F., Scholz, U., Pons-Kühnemann, J., Sonnewald, S., Sonnewald, U., et al. (2009). Manipulation of plant innate immunity and gibberellin as factor of compatibility in the mutualistic association of barley roots with *Piriformospora indica*. The Plant Journal *59*, 461–474.

Schirawski, J., Mannhaupt, G., Münch, K., Brefort, T., Schipper, K., Doehlemann, G., Di Stasio, M., Rössel, N., Mendoza-Mendoza, A., Pester, D., et al. (2010). Pathogenicity determinants in smut fungi revealed by genome comparison. Science *330*, 1546–1548.

Schlaeppi, K., Abou-Mansour, E., Buchala, A., and Mauch, F. (2010). Disease resistance of Arabidopsis to *Phytophthora brassicae* is established by the sequential action of indole glucosinolates and camalexin. The Plant Journal *62*, 840–851.

Schnell, N., Entian, K.-D., Schneider, U., Götz, F., Zähner, H., Kellner, R., and Jung, G. (1988). Prepeptide sequence of epidermin, a ribosomally synthesized antibiotic with four sulphide-rings. Nature *333*, 276–278.

Schön, M., Töller, A., Diezel, C., Roth, C., Westphal, L., Wiermer, M., and Somssich, I.E. (2013). Analyses of *wrky18 wrky40* Plants Reveal Critical Roles of SA/EDS1 Signaling and Indole-Glucosinolate Biosynthesis for *Golovinomyces orontii* Resistance and a Loss-of Resistance Towards *Pseudomonas syringae* pv. *tomato* AvrRPS4. Molecular Plant-Microbe Interactions *26*, 758–767.

Schoonbeek, H., Nistelrooy, J.G.M. van, and Waard, M.A. de (2003). Functional Analysis of ABC Transporter Genes From *Botrytis cinerea* Identifies BcatrB as a Transporter of Eugenol. European Journal of Plant Pathology *109*, 1003–1011.

Schouten, A., Maksimova, O., Cuesta-Arenas, Y., Van Den Berg, G., and Raaijmakers, J.M. (2008). Involvement of the ABC transporter BcAtrB and the laccase BcLCC2 in defence of *Botrytis cinerea* against the broad-spectrum antibiotic 2,4-diacetylphloroglucinol. Environmental Microbiology *10*, 1145–1157.

Schulze-Lefert, P., and Panstruga, R. (2011). A molecular evolutionary concept connecting nonhost resistance, pathogen host range, and pathogen speciation. Trends in Plant Science *16*, 117–125.

Schützendübel, A., and Polle, A. (2002). Plant responses to abiotic stresses: heavy metal-induced oxidative stress and protection by mycorrhization. J. Exp. Bot. *53*, 1351–1365.

Selbmann, L., Egidi, E., Isola, D., Onofri, S., Zucconi, L., De Hoog, G.S., Chinaglia, S., Testa, L., Tosi, S., Balestrazzi, A., et al. (2012). Biodiversity, evolution and adaptation of fungi in extreme environments. Plant Biosystems - An International Journal Dealing with All Aspects of Plant Biology *0*, 1–10.

Selker, E.U. (2002). Repeat-induced gene silencing in fungi. In Advances in Genetics, Jay C. Dunlap and C.-ting Wu, ed. (Academic Press), pp. 439–450.

Sella, L., Gazzetti, K., Faoro, F., Odorizzi, S., D'Ovidio, R., Schäfer, W., and Favaron, F. (2013). A *Fusarium graminearum* xylanase expressed during wheat infection is a necrotizing factor but is not essential for virulence. Plant Physiology and Biochemistry *64*, 1–10.

Selosse, M.-A., Bauer, R., and Moyersoen, B. (2002a). Basal hymenomycetes belonging to the Sebacinaceae are ectomycorrhizal on temperate deciduous trees. New Phytologist *155*, 183–195.

Selosse, M.-A., Weiß, M., Jany, J.-L., and Tillier, A. (2002b). Communities and populations of sebacinoid basidiomycetes associated with the achlorophyllous orchid *Neottia nidus-avis* (L.) L.C.M. Rich. and neighbouring tree ectomycorrhizae. Molecular Ecology *11*, 1831–1844.

Selosse, M.-A., Faccio, A., Scappaticci, G., and Bonfante, P. (2004). Chlorophyllous and Achlorophyllous Specimens of *Epipactis microphylla* (Neottieae, Orchidaceae) Are Associated with Ectomycorrhizal Septomycetes, including Truffles. Microb Ecol *47*, 416–426.

Selosse, M.-A., Setaro, S., Glatard, F., Richard, F., Urcelay, C., and Weiß, M. (2007). Sebacinales are common mycorrhizal associates of Ericaceae. New Phytologist *174*, 864–878.

Serrano, R., Martín, H., Casamayor, A., and Ariño, J. (2006). Signaling Alkaline pH Stress in the Yeast *Saccharomyces cerevisiae* through the Wsc1 Cell Surface Sensor and the Slt2 MAPK Pathway. J. Biol. Chem. *281*, 39785–39795.

Setaro, S., Weiß, M., Oberwinkler, F., and Kottke, I. (2006). Sebacinales form ectendomycorrhizas with *Cavendishia nobilis*, a member of the Andean clade of Ericaceae, in the mountain rain forest of southern Ecuador. New Phytologist *169*, 355–365.

Seyffert, W. (2003). Lehrbuch der Genetik (Spektrum Akademischer Verlag).

Shahollari, B., Vadassery, J., Varma, A., and Oelmüller, R. (2007). A leucine-rich repeat protein is required for growth promotion and enhanced seed production mediated by the endophytic fungus *Piriformospora indica* in *Arabidopsis thaliana*. The Plant Journal *50*, 1–13.

Sharma, M., Schmid, M., Rothballer, M., Hause, G., Zuccaro, A., Imani, J., Kämpfer, P., Domann, E., Schäfer, P., Hartmann, A., et al. (2008). Detection and identification of bacteria intimately associated with fungi of the order Sebacinales. Cellular Microbiology *10*, 2235–2246.

Shendure, J., and Ji, H. (2008). Next-generation DNA sequencing. Nat Biotech *26*, 1135–1145.

Sherameti, I., Tripathi, S., Varma, A., and Oelmüller, R. (2008). The Root-Colonizing Endophyte *Pirifomospora indica* Confers Drought Tolerance in Arabidopsis by Stimulating the Expression of Drought Stress–Related Genes in Leaves. Molecular Plant-Microbe Interactions *21*, 799–807.

Shim, J.S., Jung, C., Lee, S., Min, K., Lee, Y.-W., Choi, Y., Lee, J.S., Song, J.T., Kim, J.-K., and Choi, Y.D. (2013). AtMYB44 regulates WRKY70 expression and modulates antagonistic

interaction between salicylic acid and jasmonic acid signaling. The Plant Journal *73*, 483–495.

Shirley, B.W., Kubasek, W.L., Storz, G., Bruggemann, E., Koornneef, M., Ausubel, F.M., and Goodman, H.M. (1995). Analysis of Arabidopsis mutants deficient in flavonoid biosynthesis. The Plant Journal *8*, 659–671.

Slater, G.S., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics *6*, 31.

Smit, A., and Hubley, R. (2008). RepeatModeler Open-1.0.

Smit, A., Hubley, R., and Green, P. (1996). RepeatMasker Open-3.0.

Smith, S.E., and Read, D.J. (2008). Mycorrhizal Symbiosis (Academic Press).

Smith, S.E., Gianinazzi-Pearson, V., Koide, R., and Cairney, J.W.G. (1994). Nutrient transport in mycorrhizas: structure, physiology and consequences for efficiency of the symbiosis. Plant Soil *159*, 103–113.

Soderlund, C., Nelson, W., Shoemaker, A., and Paterson, A. (2006). SyMAP: A system for discovering and viewing syntenic regions of FPC maps. Genome Res. *16*, 1159–1168.

Solano, R., Stepanova, A., Chao, Q., and Ecker, J.R. (1998). Nuclear events in ethylene signaling: a transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and ETHYLENE-RESPONSE-FACTOR1. Genes Dev. *12*, 3703–3714.

Sonnhammer, E.L., von Heijne, G., and Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol *6*, 175–182.

Spanu, P.D. (2012). The Genomics of Obligate (and Nonobligate) Biotrophs. Annual Review of Phytopathology *50*, 91–109.

Spanu, P.D., Abbott, J.C., Amselem, J., Burgis, T.A., Soanes, D.M., Stüber, K., Themaat, E.V.L. van, Brown, J.K.M., Butcher, S.A., Gurr, S.J., et al. (2010). Genome Expansion and Gene Loss in Powdery Mildew Fungi Reveal Tradeoffs in Extreme Parasitism. Science *330*, 1543–1546.

Spoel, S.H., and Dong, X. (2012). How do plants achieve immunity? Defence without specialized immune cells. Nat Rev Immunol *12*, 89–100.

Sriranganadane, D., Reichard, U., Salamin, K., Fratti, M., Jousson, O., Waridel, P., Quadroni, M., Neuhaus, J.-M., and Monod, M. (2011). Secreted glutamic protease rescues aspartic protease Pep deficiency in *Aspergillus fumigatus* during growth in acidic protein medium. Microbiology *157*, 1541–1550.

Stanke, M., and Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics *19*, ii215–ii225.

Stassen, J.H., and Van den Ackerveken, G. (2011). How do oomycete effectors interfere with plant life? Current Opinion in Plant Biology *14*, 407–414.

Stefanato, F.L., Abou-Mansour, E., Buchala, A., Kretschmer, M., Mosbach, A., Hahn, M., Bochet, C.G., Métraux, J.-P., and Schoonbeek, H. (2009). The ABC transporter BcatrB from *Botrytis cinerea* exports camalexin and is a virulence factor on *Arabidopsis thaliana*. The Plant Journal *58*, 499–510.

Stein, E., Molitor, A., Kogel, K.-H., and Waller, F. (2008). Systemic Resistance in Arabidopsis Conferred by the Mycorrhizal Fungus *Piriformospora indica* Requires Jasmonic Acid Signaling and the Cytoplasmic Function of NPR1. Plant Cell Physiol *49*, 1747–1751.

Sterflinger, K., Tesei, D., and Zakharova, K. (2012). Fungi in hot and cold deserts with particular reference to microcolonial fungi. Fungal Ecology *5*, 453–462.

Stintzi, A., Heitz, T., Prasad, V., Wiedemann-Merdinoglu, S., Kauffmann, S., Geoffroy, P., Legrand, M., and Fritig, B. (1993). Plant "pathogenesis-related" proteins and their role in defense against pathogens. Biochimie *75*, 687–706.

Stirnimann, C.U., Petsalaki, E., Russell, R.B., and Müller, C.W. (2010). WD40 proteins propel cellular networks. Trends in Biochemical Sciences *35*, 565–574.

Stotz, H.U., Sawada, Y., Shimada, Y., Hirai, M.Y., Sasaki, E., Krischke, M., Brown, P.D., Saito, K., and Kamiya, Y. (2011). Role of camalexin, indole glucosinolates, and side chain modification of glucosinolate-derived isothiocyanates in defense of Arabidopsis against *Sclerotinia sclerotiorum*. The Plant Journal *67*, 81–93.

Stumpf, P.K., and Conn, E.E. (1988). The Biochemistry of Plants: Carbohydrates (Academic Press).

Stuttmann, J., Hubberten, H.-M., Rietz, S., Kaur, J., Muskett, P., Guerois, R., Bednarek, P., Hoefgen, R., and Parker, J.E. (2011). Perturbation of Arabidopsis Amino Acid Metabolism Causes Incompatibility with the Adapted Biotrophic Pathogen *Hyaloperonospora arabidopsidis*. Plant Cell *23*, 2788–2803.

Sun, C., Johnson, J.M., Cai, D., Sherameti, I., Oelmüller, R., and Lou, B. (2010). *Piriformospora indica* confers drought tolerance in Chinese cabbage leaves by stimulating antioxidant enzymes, the expression of drought-related genes and the plastid-localized CAS protein. Journal of Plant Physiology *167*, 1009–1017.

Swindell, S.R., and Plasterer, T.N. (1997). SEQMAN. Contig assembly. Methods Mol. Biol. *70*, 75–89.

Taylor, D.L., Bruns, T.D., Szaro, T.M., and Hodges, S.A. (2003). Divergence in mycorrhizal specialization within *Hexalectris spicata* (Orchidaceae), a nonphotosynthetic desert orchid. Am. J. Bot. *90*, 1168–1179.

Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y.O., and Borodovsky, M. (2008). Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res. *18*, 1979–1990.

Thon, M.R., Pan, H., Diener, S., Papalas, J., Taro, A., Mitchell, T.K., and Dean, R.A. (2006). The role of transposable element clusters in genome evolution and loss of synteny in the rice blast fungus *Magnaporthe oryzae*. Genome Biology *7*, R16.

Thornberry, N.A., and Lazebnik, Y. (1998). Caspases: Enemies Within. Science *281*, 1312–1316.

Tierens, K.F.M.-J., Thomma, B.P.H.J., Brouwer, M., Schmidt, J., Kistner, K., Porzel, A., Mauch-Mani, B., Cammue, B.P.A., and Broekaert, W.F. (2001). Study of the Role of Antimicrobial Glucosinolate-Derived Isothiocyanates in Resistance of Arabidopsis to Microbial Pathogens. Plant Physiol. *125*, 1688–1699.

Tikkanen, R., Obermüller, S., Denzer, K., Pungitore, R., Geuze, H.J., Von Figura, K., and Höning, S. (2000). The Dileucine Motif Within the Tail of MPR46 is Required for Sorting of the Receptor in Endosomes. Traffic *1*, 631–640.

Todd, R.B., Andrianopoulos, A., Davis, M.A., and Hynes, M.J. (1998). FacB, the *Aspergillus nidulans* activator of acetate utilization genes, binds dissimilar DNA sequences. EMBO J *17*, 2042–2054.

Tonukari, N.J., Scott-Craig, J.S., and Waltonb, J.D. (2000). The *Cochliobolus carbonum* SNF1 Gene Is Required for Cell Wall–Degrading Enzyme Expression and Virulence on Maize. Plant Cell *12*, 237–247.

Tsukumo, Y., Tsukahara, S., Saito, S., Tsuruo, T., and Tomida, A. (2009). A Novel Endoplasmic Reticulum Export Signal PROLINE AT THE +2-POSITION FROM THE SIGNAL PEPTIDE CLEAVAGE SITE. J. Biol. Chem. *284*, 27500–27510.

Tyler, B.M., Tripathy, S., Zhang, X., Dehal, P., Jiang, R.H.Y., Aerts, A., Arredondo, F.D., Baxter, L., Bensasson, D., Beynon, J.L., et al. (2006). Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science *313*, 1261–1266.

Tyler, B.M., Kale, S.D., Wang, Q., Tao, K., Clark, H.R., Drews, K., Antignani, V., Rumore, A., Hayes, T., Plett, J.M., et al. (2013). Microbe-independent entry of oomycete RxLR effectors and fungal RxLR-like effectors into plant and animal cells is specific and reproducible. Molecular Plant-Microbe Interactions 130403113643006.

Urban, A., Weib, M., and Bauer, R. (2003). Ectomycorrhizas involving sebacinoid mycobionts. Mycological Research *107*, 3–14.

Usadel, B., Nagel, A., Thimm, O., Redestig, H., Blaesing, O.E., Palacios-Rojas, N., Selbig, J., Hannemann, J., Piques, M.C., Steinhauser, D., et al. (2005). Extension of the Visualization Tool MapMan to Allow Statistical Analysis of Arrays, Display of Coresponding Genes, and Comparison with Known Responses. Plant Physiol *138*, 1195–1204.

Vaario, L.-M., Heinonsalo, J., Spetz, P., Pennanen, T., Heinonen, J., Tervahauta, A., and Fritze, H. (2012). The ectomycorrhizal fungus *Tricholoma matsutake* is a facultative saprotroph in vitro. Mycorrhiza *22*, 409–418.

Vadassery, J., Ritter, C., Venus, Y., Camehl, I., Varma, A., Shahollari, B., Novák, O., Strnad, M., Ludwig-Müller, J., and Oelmüller, R. (2008). The Role of Auxins and Cytokinins in the Mutualistic Interaction Between Arabidopsis and *Piriformospora indica*. Molecular Plant-Microbe Interactions *21*, 1371–1383.

Varma, A., Kost, G., and Oelmüller, R. (2013). *Piriformospora indica* (Springer).

Vaughan-Nichols, S.J. (2003). The battle over the universal Java IDE. Computer *36*, 21 – 23.

Vellosillo, T., Martínez, M., López, M.A., Vicente, J., Cascón, T., Dolan, L., Hamberg, M., and Castresana, C. (2007). Oxylipins Produced by the 9-Lipoxygenase Pathway in Arabidopsis Regulate Lateral Root Development and Defense Responses through a Specific Signaling Cascade. Plant Cell *19*, 831–846.

Venkateshwaran, M., Volkening, J.D., Sussman, M.R., and Ané, J.-M. (2013). Symbiosis and the social network of higher plants. Current Opinion in Plant Biology *16*, 118–127.

Verma, S., Varma, A., Rexer, K.-H., Hassel, A., Kost, G., Sarbhoy, A., Bisen, P., Butehorn, B., and Franken, P. (1998). *Piriformospora indica*, gen. et sp. nov., a New Root-Colonizing Fungus. Mycologia *90*, 896.

Vierheilig, H., Bennett, R., Kiddle, G., Kaldorf, M., and Ludwig-Müller, J. (2000). Differences in glucosinolate patterns and arbuscular mycorrhizal status of glucosinolate-containing plant species. New Phytologist *146*, 343–352.

De Vries, R.P., Broeck, H.C. van den, Dekkers, E., Manzanares, P., Graaff, L.H. de, and Visser, J. (1999). Differential Expression of Three α-Galactosidase Genes and a Single β-Galactosidase Gene from *Aspergillus niger*. Appl. Environ. Microbiol. *65*, 2453–2460.

De Waard, M.A., Andrade, A.C., Hayashi, K., Schoonbeek, H., Stergiopoulos, I., and Zwiers, L.-H. (2006). Impact of fungal drug transporters on fungicide sensitivity, multidrug resistance and virulence. Pest Management Science *62*, 195–207.

Waller, F., Achatz, B., Baltruschat, H., Fodor, J., Becker, K., Fischer, M., Heier, T., Hückelhoven, R., Neumann, C., Wettstein, D. von, et al. (2005). The endophytic fungus *Piriformospora indica* reprograms barley to salt-stress tolerance, disease resistance, and higher yield. PNAS *102*, 13386–13391.

Waller, F., Mukherjee, K., Deshmukh, S.D., Achatz, B., Sharma, M., Schäfer, P., and Kogel, K.-H. (2008). Systemic and local modulation of plant responses by *Piriformospora indica* and related Sebacinales species. Journal of Plant Physiology *165*, 60–70.

Walter, M.C., Rattei, T., Arnold, R., Güldener, U., Münsterkötter, M., Nenova, K., Kastenmüller, G., Tischler, P., Wölling, A., Volz, A., et al. (2009). PEDANT covers all complete RefSeq genomes. Nucl. Acids Res. *37*, D408–D411.

Wang, H., and van der Donk, W.A. (2012). Biosynthesis of the Class III Lantipeptide Catenulipeptin. ACS Chem. Biol. *7*, 1529–1535.

Wang, Y., Lim, L., DiGuistini, S., Robertson, G., Bohlmann, J., and Breuil, C. (2013a). A specialized ABC efflux transporter GcABC-G1 confers monoterpene resistance to *Grosmannia clavigera*, a bark beetle-associated fungal pathogen of pine trees. New Phytologist *197*, 886–898.

Wang, Y., Bouwmeester, K., van de Mortel, J.E., Shan, W., and Govers, F. (2013b). Induced expression of defense-related genes in Arabidopsis upon infection with *Phytophthora capsici*. Plant Signaling & Behavior *8*, e24618.

Warcup, J.H. (1988). Mycorrhizal associations of isolates of *Sebacina vermifera*. New Phytologist *110*, 227–231.

Weber, S.S., Kovalchuk, A., Bovenberg, R.A.L., and Driessen, A.J.M. (2012). The ABC transporter ABC40 encodes a phenylacetic acid export system in *Penicillium chrysogenum*. Fungal Genetics and Biology *49*, 915–921.

Wees, S.C.M. van, Chang, H.-S., Zhu, T., and Glazebrook, J. (2003). Characterization of the Early Response of Arabidopsis to *Alternaria brassicicola* Infection Using Expression Profiling. Plant Physiol. *132*, 606–617.

Weiß, M., and Oberwinkler, F. (2001). Phylogenetic relationships in Auriculariales and related groups—hypotheses derived from nuclear ribosomal DNA sequences. Mycological Research *105*, 403–415.

Weiss, M., Selosse, M.-A., Rexer, K.-H., Urban, A., and Oberwinkler, F. (2004). Sebacinales: a hitherto overlooked cosm of heterobasidiomycetes with a broad mycorrhizal potential. Mycological Research *108*, 1003–1010.

Weiß, M., Sýkorová, Z., Garnica, S., Riess, K., Martos, F., Krause, C., Oberwinkler, F., Bauer, R., and Redecker, D. (2011). Sebacinales Everywhere: Previously Overlooked Ubiquitous Fungal Endophytes. PLoS ONE *6*, e16793.

Whisson, S.C., Boevink, P.C., Moleleki, L., Avrova, A.O., Morales, J.G., Gilroy, E.M., Armstrong, M.R., Grouffaud, S., van West, P., Chapman, S., et al. (2007). A translocation signal for delivery of oomycete effector proteins into host plant cells. Nature *450*, 115–118.

Williams, T., and Kelley, C. (2012). gnuplot 4.6.0.

Winnenburg, R., Urban, M., Beacham, A., Baldwin, T.K., Holland, S., Lindeberg, M., Hansen, H., Rawlings, C., Hammond-Kosack, K.E., and Köhler, J. (2008). PHI-base update: additions to the pathogen–host interaction database. Nucl. Acids Res. *36*, D572–D576.

De Wit, P.J.G.M., Mehrabi, R., Van Den Burg, H.A., and Stergiopoulos, I. (2009). Fungal effector proteins: past, present and future. Molecular Plant Pathology *10*, 735–747.

Wolfe, B.E., Kuo, M., and Pringle, A. (2012). *Amanita thiersii* is a saprotrophic fungus expanding its range in the United States. Mycologia *104*, 22–33.

Wösten, H.A.B., Moukha, S.M., Sietsma, J.H., and Wessels, J.G.H. (1991). Localization of growth and secretion of proteins in *Aspergillus niger*. J Gen Microbiol *137*, 2017–2023.

Wright, D.P., Scholes, J.D., and Read, D.J. (1998). Effects of VA mycorrhizal colonization on photosynthesis and biomass production of *Trifolium repens* L. Plant, Cell & Environment *21*, 209–216.

Wu, Y., Zhang, D., Chu, J.Y., Boyle, P., Wang, Y., Brindle, I.D., De Luca, V., and Després, C. (2012). The Arabidopsis NPR1 Protein Is a Receptor for the Plant Defense Hormone Salicylic Acid. Cell Reports *1*, 639–647.

Xing, D.-H., Lai, Z.-B., Zheng, Z.-Y., Vinod, K.M., Fan, B.-F., and Chen, Z.-X. (2008). Stress- and Pathogen-Induced Arabidopsis WRKY48 is a Transcriptional Activator that Represses Plant Basal Defense. Mol. Plant *1*, 459–470.

Xu, J., Li, Y., Wang, Y., Liu, H., Lei, L., Yang, H., Liu, G., and Ren, D. (2008). Activation of MAPK Kinase 9 Induces Ethylene and Camalexin Biosynthesis and Enhances Sensitivity to Salt Stress in Arabidopsis. J. Biol. Chem. *283*, 26996–27006.

Xu, Z.-Y., Zhang, X., Schläppi, M., and Xu, Z.-Q. (2011). Cold-inducible expression of AZI1 and its function in improvement of freezing tolerance of *Arabidopsis thaliana* and *Saccharomyces cerevisiae*. Journal of Plant Physiology *168*, 1576–1587.

Yadav, V., Kumar, M., Deep, D.K., Kumar, H., Sharma, R., Tripathi, T., Tuteja, N., Saxena, A.K., and Johri, A.K. (2010). A Phosphate Transporter from the Root Endophytic Fungus *Piriformospora indica* Plays a Role in Phosphate Transport to the Host Plant. J. Biol. Chem. *285*, 26532–26544.

Yike, I. (2011). Fungal Proteases and Their Pathophysiological Effects. Mycopathologia *171*, 299–323.

Yoo, S.-D., Cho, Y.-H., Tena, G., Xiong, Y., and Sheen, J. (2008). Dual control of nuclear EIN3 by bifurcate MAPK cascades in C2H4 signalling. Nature *451*, 789–795.

Zhang, X., Dai, Y., Xiong, Y., DeFraia, C., Li, J., Dong, X., and Mou, Z. (2007). Overexpression of Arabidopsis MAP kinase kinase 7 leads to activation of plant basal and systemic acquired resistance. The Plant Journal *52*, 1066–1079.

Zhu, Q., Zhang, J., Gao, X., Tong, J., Xiao, L., Li, W., and Zhang, H. (2010). The Arabidopsis AP2/ERF transcription factor RAP2.6 participates in ABA, salt and osmotic stress responses. Gene *457*, 1–12.

Žifčáková, L., and Baldrian, P. (2012). Fungal polysaccharide monooxygenases: new players in the decomposition of cellulose. Fungal Ecology *5*, 481–489.

Zipfel, C. (2008). Pattern-recognition receptors in plant innate immunity. Current Opinion in Immunology *20*, 10–16.

Zuccaro, A., Lahrmann, U., Güldener, U., Langen, G., Pfiffi, S., Biedenkopf, D., Wong, P., Samans, B., Grimm, C., Basiewicz, M., et al. (2011). Endophytic Life Strategies Decoded by Genome and Transcriptome Analyses of the Mutualistic Root Symbiont *Piriformospora indica*. PLoS Pathog *7*, e1002290.

6. Appendix

Due to space limitation and formatting issues are all self-written programs and scripts supplied on a CD to this thesis. This CD contains the following files:

1. The JAVA programs described in the chapters 4.7.1.1. to 4.7.1.18.
2. The R scripts described in the chapters 4.7.2.1. to 4.7.2.7.
3. The shell scripts described in the chapters 4.7.3.1. to 4.7.3.5.

Danksagungen

An erste Stelle möchte ich mich bei Alga Zuccaro bedanken, die mir dich Möglichkeit gegeben hat, diese Arbeit in ihrer Arbeitsgruppe anzufertigen. Im Besonderen danke ich ihr dabei für die vielen hilfreichen Diskussionen und Anregungen, sowie für die Freiheiten auch meinen eigenen wissenschaftlichen Interessen und Vorlieben nachzugehen.

Besonderer Dank gilt weiterhin Professor Regine Kahmann und der International Max-Planck Research School (IMPRS) für die finanzielle Unterstützung, hilfreichen Diskussionen und Weiterbildungsmöglichkeiten.

Den Mitgliedern meines Prüfungskomitees, Professor Hans-Ulrich Mösch, Professor Michael Bölker, Professor Alfred Batschauer und Professor Anke Becker danke ich für die Zeit, die sie sich für mich und die Bewertung meiner Arbeit genommen haben.

Weiter bedanke ich mich für sehr angenehme Kooperationen bei Ulrich Güldener und Philip Wong vom Münchener Informationszentrum für Proteinsequenzen (MIPS) und bei Annegret Kohler und Francis Martin vom Nationalen französischen Institut für Agrarforschung (INRA).

Auch wenn die Doktorarbeit vor allem viel Arbeit bedeutete, freue ich mich im ganzen Institut, Kollegen und Freunde gefunden zu haben, die dafür Sorge getragen haben, dass auch wissenschaftlich weniger gute Zeiten schnell vorbeigingen. Für eine schöne Zeit und eine gute Arbeitsatmosphäre danke ich daher allen, die dazu beigetragen haben. Insbesondere möchte ich mich dabei bei Magdalena Hilbert bedanken, für viel Verständnis, Diskussionsbereitschaft und ein stetes Lächeln auf den Lippen.

Natürlich bedanke ich mich auch bei allen Freunden und Verwandten, die mich auf meinem Weg begleitet und unterstützt haben. Insbesondere gilt dieser Dank meinen Eltern, meinem Bruder, meiner Schwägerin, meiner Patentante, meiner "richtige" Tante, meinem Onkel und alle weiteren Mitglieder meiner Familie und der meiner Frau, derer es leider zu viele sind um sie hier gerechterweise alle aufzulisten. Ich danke euch allen aus tiefstem Herzen!

Meiner größter Dank geht an meine Ehefrau Catharina, die trotz der räumlichen Distanz mit ihrer Liebe immer bei mir war, die immer Zeit für mich hatte und die mir immer Kraft gegeben hat weiterzumachen.