The genome of the fungal plant pathogen *Thecaphora thlaspeos*

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Lamprinos Frantzeskakis

aus Athen/Griechenland

Düsseldorf, Mai 2016

aus dem Institut für Mikrobiologie
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. Michael Feldbrügge
Korreferent: Prof. Dr. Joachim Ernst
Tag der mündlichen Prüfung: _____

Eidesstattliche Erklärung

Ich versichere an Eides Statt, dass die Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist. Die Dissertation wurde in ihrer jetzigen oder ähnlichen Form noch bei keiner anderen Hochschule eingereicht. Ich habe zuvor keine erfolglosen Promotionsversuche unternommen.

_____ _____

Ort, Datum Lamprinos Frantzeskakis

Die Untersuchungen zur vorliegenden Arbeit wurden von September 2012 bis Mai 2016 in Düsseldorf an der Heinrich-Heine-Universität in dem Institut für Mikrobiologie unter der Betreuung von Herrn Prof. Dr. Michael Feldbrügge durchgeführt.

TABLE OF CONTENTS

Summary I

Zusammenfassung II

List of abbreviations III

1 Introduction 4

1.1 Pathogens, endophytes and modern agriculture 4

1.2 The importance of the smut fungi 5

1.3 Mating: a crucial checkpoint in the life of smut fungi 6

1.4 Genomics and genomes of smut fungi 8

1.5 *Thecaphora thlaspeos*, a fungal pathogen of Brassicaceae 9

1.6 Aim of this work 9

2 Results 11

2.1 Part I: Collection and culture of *T. thlaspeos* 12

2.1.1 Sample collection and macroscopic symptoms 12

2.1.2 Teliospore germination of *T. thlaspeos* 15

2.1.3 The biotrophic phase of the *T. thlaspeos* life cycle 16

2.1.4 A brief report on other *Thecaphora* species 19

2.2 Part II: The genome of *T. thlaspeos* 22

2.2.1 *De-novo* assembly of the *T. thlaspeos* genome 22

2.2.2 Gene prediction and functional annotation 24

2.2.3 Synteny 25

2.2.4 Core genome of smut fungi 27

2.2.5 The pheromone/pheromone receptor a locus 29

2.2.6 Interactions in the *T. thlaspeos b* mating type locus 32

2.2.7 Gain of a GH78 family enzyme – An adaptation to the dicot cell wall? 35

2.2.8 Loss of the flocculosin cluster in *T. thlaspeos* 37

2.2.9 Candidate effector proteins in the *T. thlaspeos* secretome 39

2.2.10 The genome of the closely related species, *T. seminis-convolvuli* 45

3 Discussion 47

3.1 Summary 48

3.2 An overview and open questions in the life cycle of *T. thlaspeos* 48

3.3 A first comparative analysis of the *T. thlaspeos* genome reveals novel genes 50
3.4 Conclusion and future prospects 52
4 Materials and Methods 54
4.1 Molecular biology and Mycology methods 54
4.1.1 Chemicals and Enzymes 54
4.1.2 PCR, Plasmid extraction, Cloning 54
4.1.3 Media 54
4.1.4 Plant growth chamber 55
4.1.5 Germination of *T. thlaspeos* teliospores and culture of strains 55
4.1.6 Germination of *Thecaphora spp*. teliospores and culture of strains 56
4.1.7 Genomic DNA and RNA extraction 57
4.1.8 cDNA synthesis 57
4.1.9 Yeast two-hybrid assays 57
4.1.10 Transient expression of TtNLP1 and TtNLP2 in *N. benthamiana* 58
4.1.11 Infected plant tissue staining 59
4.1.12 Nuclear staining 59
4.1.13 Microscopy 59
4.1.14 Oligonucleotides list 60
4.1.15 Plasmid list 64
4.1.16 Strain list 66
4.2 Bioinformatics and computational methods 67
4.2.1 General purpose software 67
4.2.2 Genome sequencing and assembly 67
4.2.3 RNA dataset 67
4.2.4 Structural annotation 68
4.2.5 Functional annotation 68
4.2.6 Orthology with other species 69
4.2.7 Multi-locus phylogeny 70
4.2.8 Synteny 70
4.2.9 Expression of genes during culture 70
5 Appendix 71
5.1 Supplementary Figures and Tables 71

5.2 References 83
5.3 Figure and Table list 103
5.3.1 Figures 103
5.3.2 Tables 103
5.3.3 Supplementary figures 104
5.3.4 Supplementary tables 104
5.3.5 Supplementary data 105
5.4 Acknowledgements 106
5.5 CV 107

SUMMARY

Fungal plant pathogens are still an important limitation to global produce increase, despite the development of chemical and genetic means for their control. In many grain crops, diseases that are caused by members of the *Ustilaginales* lead to yield loss and severe reduction of the grain quality. These fungi have a biotrophic life style which is completed with the production of dark pigmented teliospores usually in the reproductive organs of the host. This characteristic symptom has coined the term "smut fungi" for this group of pathogens. Research for this disease is dominated by the model organism and causal agent of the corn smut, *Ustilago maydis*.

In recent years, many advances have been made in characterizing virulence factors and processes necessary for host colonization, mainly driven by well-established lab protocols and rich genomic information for these species. A challenge that persists however is the description of the host responses and the elucidation of possible defense strategies to this disease. The difficulty is owed to the genetic complexity of maize and other host grasses of the smut fungi. Overcoming this issue would require the migration of the current knowledge on smut fungi to a host-pathogen pair that includes a model plant and a smut pathogen with the same lab-friendly attributes of *U. maydis*.

In order to explore this option, this study presents *Thecaphora thlaspeos*, a fungal plant pathogen of the *Brassicaceae* species *Arabis hirsuta* and *Ar. ciliata*. After describing the main symptoms, culture characteristics and its genome, it is shown that this species presents significant novelties compared to other smut fungi. One remarkable feature of *T. thlaspeos* is that its teliospores germinate filamentously after the perception of a host derived signal and in contrast to many of the grass smuts, filamentous growth persists in haploid culture even in the absence of a compatible *b* heterodimer. In addition, comparative analysis of the *T. thlaspeos* genome shows that although the majority of the genes between the so far sequenced smut fungi are homologous, a group of approximately 350 genes is unique for this species. This group includes small, putatively secreted and cysteine-rich proteins that are likely to be necessary for successful host infection. Selected examples of this group are discussed and examined in detail.

With this work, a foundation is laid to explore *T. thlaspeos* as a potential complimentary organism to *U. maydis*. It is anticipated that the strains derived from this study along with the annotated genome will facilitate further exploration of the ability of smut fungi to infect *Brassicaceae* and most importantly to determine if this species is able to complete its life cycle in the model plant *Arabidopsis thaliana*.

ZUSAMMENFASSUNG

Pilzliche Pflanzenpathogene sind nach wie vor eine wichtige Limitierung für die globale Ertragserhöhung von pflanzlichen Erzeugnissen trotz der Entwicklung chemischer und genetischer Möglichkeiten für ihre Kontrolle. In vielen Getreidearten führen Krankheiten, die durch Mitglieder der Ustilaginales hervorgerufen werden, zum Ertragsverlust und erheblicher Minderung der Kornqualität. Diese Pilze haben einen biotrophen Lebensstil, der durch die Produktion dunkel pigmentierter Teliosporen, üblicherweise in den reproduktiven Organen des Wirtes, vollendet wird. Dieses charakteristische Symptom hat den Begriff „Brandpilz“ für diese Gruppe von Pathogenen geprägt. Die Forschung über diese Krankheiten wird angeführt vom Modellorganismus und kausalen Erreger des Maisbeulenbrandes, *Ustilago maydis*.

In den letzten Jahren wurden viele Fortschritte bezüglich der Charakterisierung von Virulenzfaktoren und Prozessen, die für die Wirtskolonisierung notwendig sind, gemacht. Diese wurden hauptsächlich durch gut etablierte Laborprotokolle und ausgiebige genomische Informationen für diese Art vorangetrieben. Eine Herausforderung, die dennoch fortbesteht, ist die Beschreibung der Wirtsantworten und die Schilderung möglicher Abwehrstrategien für diese Krankheit. Die Schwierigkeit liegt dabei in der genetischen Komplexität von Mais und anderen Wirtsgräsern der Brandpilze. Diesen Aspekt zu überwinden würde voraussetzen das derzeitige Wissen über Brandpilze auf ein Wirt-Pathogen Paar zu übertragen, welches eine Modellpflanze und einen Brandpilz mit denselben laborfreundlichen Attributen wie *U. maydis* involviert.

Um diese Option zu untersuchen, wird in dieser Arbeit *Thecaphora thlaspeos* präsentiert, ein pflanzenpathogener Pilz der Brassicaceae Spezies *Arabis hirsuta* und *Ar. ciliata*. Nachdem die Hauptsymptome, Kulturcharakteristika und sein Genom beschrieben werden, wird gezeigt, dass diese Spezies signifikante Neuerungen gegenüber anderen Brandpilzen aufweist. Eine bemerkenswerte Eigenschaft von *T. thlaspeos* ist, dass die Teliosporen filamentös keimen nachdem ein vom Wirt stammendes Signal wahrgenommen wird. Im Gegensatz zu vielen Gräser befallenden Brandpilzen wird sogar in Abwesenheit eines kompatiblen b-Heterodimers das filamentöse Wachstum aufrechterhalten. Zusätzlich zeigten vergleichende Analysen des *T. thlaspeos* Genoms, dass obwohl die Mehrheit der Gene zwischen den bis jetzt sequenzierten Brandpilzen homolog sind, es eine Gruppe von etwa 350 Genen gibt, die spezifisch für diesen Organismus ist. Diese Gruppe enthält kleine, voraussichtlich sekretierte und Cystein-reiche Proteine, die wahrscheinlich für die erfolgreiche Infektion des Wirtes benötigt werden. Ausgewählte Beispiele dieser Gruppe werden diskutiert und detailliert untersucht.

Mit dieser Arbeit wird ein Fundament gelegt, um zu untersuchen ob *T. thlaspeos* das Potenzial als komplementärer Organismus zu *U. maydis* hat. Es wird antizipiert, dass die aus dieser Arbeit hervorgehenden Stämme zusammen mit dem annotierten Genom weitere Forschung der Fähigkeit von Brandpilzen, Brassicaceae zu infizieren, ermöglichen. Der wichtigste Aspekt dieser Stämme wird jedoch sein, zu bestimmen ob diese Spezies in der Lage ist, seinen Lebenszyklus in der Modellpflanze *Arabidopsis thaliana* zu vollenden.

LIST OF ABBREVIATIONS

µl	Microliter	**min**	Minutes
AD	Activation domain	**mm**	Millimeter
AED	Annotation Edit Distance	**MSN**	Murashige & Skoog medium incl. Nitsch vitamins
bbs	*b* binding sequence	**NADPH**	Nicotinamide adenine dinucleotide phosphate
bE	bEast	**NCBI**	National Center for Biotechnology Information
bp	Basepair	**NEP**	Necrosis and ethylene inducing proteins
bW	bWest	**NIP**	Necrosis-inducing proteins
CA	Carbonic anhydrases	**NLP**	Necrosis inducing-like proteins
CAZymes	Carbohydrate active enzymes	**NLS**	Nuclear localization signal
cAMP	Cyclic adenosine monophosphate	**nr**	non-redundant
CBM	Carbohydrate-binding module	**^{o}C**	Celsius degree
cDNA	Complementary DNA	**ORF**	Open reading frame
CDS	Coding DNA sequence	**PBS**	Phosphate-buffered saline
cm	Centimeter	**PCR**	Polymerase chain reaction
CM	Complete Medium	**PI**	Propidium iodine
CW	Calcofluor White	**PL**	Polysaccharide Lyases
DIC	Differential interference contrast	**rDNA**	ribosomal DNA
DNA	Deoxyribonucleic acid	**RNA**	Ribonucleic acid
GB	Gigabase	**RNAi**	RNA interference
gDNA	Genomic DNA	**ROS**	Reactive oxygen species
GH	Glycoside hydrolases	***spp.***	Species
GMC oxidoreductase	Glucose-methanol-choline oxidoreductase	**tRNA**	transfer RNA
HR	Hyper-sensitive immune response	**UA**	Ustilagic Acid
i.e	*id est*	**UV**	Ultra-violet
Kb	Kilobase	**WGA**	Wheat germ agglutinin
MAPK	Mitogen-activated protein kinases	**Y2H**	Yeast Two-Hybrid
MB	Megabase		
MEL	Mannosylerythritol lipids		
MHz	Megahertz		

1 INTRODUCTION

1.1 Pathogens, endophytes and modern agriculture

It is well established that the main challenge agricultural production faces today, and will be facing in the next decades, is the increase of yield production (Godfray *et al*., 2010), despite the adverse effects of climate change (Long *et al*., 2006; Tubiello *et al*., 2007; Lobell *et al*., 2011). Biotic stress has its toll on yield, although diseases are successfully managed today by the advances that have been made by the chemical industry and the introgression of genetic resistance (Oerke & Dehne, 2004). Nevertheless, management of plant diseases is not always durable due to the ability of the pathogens to overcome host resistance or susceptibility to pesticides (Parlevliet, 2002; Lucas *et al*., 2015).

Fungal pathogens present a significant challenge to agricultural production either directly, as a bottleneck towards high yields (Oerke, 2006), or indirectly, as a cost in production derived by the use of fungicides (as demonstrated for winter wheat in Wiik & Rosenqvist, 2010; Thompson *et al*., 2014). Moreover, it is very interesting that fungal diseases can be very dynamic, both in respect to spatial movement (Bebber *et al*., 2013) but also genetically, with the emergence of new races (Singh *et al*., 2011) or hybrid species (Menardo *et al*., 2015).

On the other hand, fungal organisms have been proven to be useful associates of plants which help alleviate stress coming from the environment (Evelin *et al*., 2009), deter harmful insects (Omacini *et al*., 2001) or induce resistance against pathogens (Harman *et al*., 2004). A prime example is arbuscular mycorrhizal fungi which have been long studied for their beneficial effects on plant growth (Smith & Smith, 2012; Berruti *et al*., 2016). In addition, recent studies show that plants have a very rich community of microbes inhabiting the belowground (Bulgarelli *et al*., 2012; Lundberg *et al*., 2012) and aboveground tissues (Bai *et al*., 2015), with fungi being not as abundant as bacteria but still a significant part of it (García *et al*., 2012; Horton *et al*., 2014). Out of these studies, novel mutualistic plant-microbe interactions have emerged (Hiruma *et al*., 2016), throwing the spotlight on endophytes, a group of fungal organisms that is able to inhabit plant tissue without causing macroscopic symptoms.

Research on endophytes has demonstrated that there is a certain balance between the plant host and the fungus that can turn a commensal or mutualistic association to a pathogenic one. Such an example has been demonstrated between *Colletotrichum magna* and different plant species (Redman *et al*., 2001). It is hypothesized that fungi previously thought to be exclusive saprotrophs and live of dead organic matter, might also have an endophytic phase in their life cycle. For example, recently it was shown that *Neurospora crassa*, a fungus usually found on burned vegetation (Perkins & Turner, 1988), can colonize the Scots pine *Pinus sylvestris* (Kuo *et al*., 2014). Most interestingly, pathogens can turn to endophytes depending on small changes in the genome of the host, as it has been shown with *Sporisorium reilianum f. sp zeae* and maize. In that case, the presence of the wall-associated kinase ZmWAK allows endophytic growth of *S. reilianum* but inhibits the development of the symptoms (Zuo *et al*., 2014; Zhao *et al*., 2015).

Vice versa, the disruption of genetic elements of the pathogen can render a pathogenic relationship to a endophytic one as it has been shown for example with *Ustilago maydis* by the deletion of effectors in the cluster 19 (Brefort *et al*., 2014) and the *path-1* mutation in *C*. *magna* (Freeman & Rodriguez, 1993; Redman *et al*., 1999).

The study of the molecular interactions between pathogenic fungi and their respective hosts not only benefits modern agriculture and the industry, but also enriches our knowledge on basic biological functions of eukaryotic organisms. Still, much information, hidden in the genomes of pathogenic, mutualistic and endophytic fungi is to be recovered (Stajich *et al*., 2009) and utilized for new products and applications.

1.2 The importance of the smut fungi

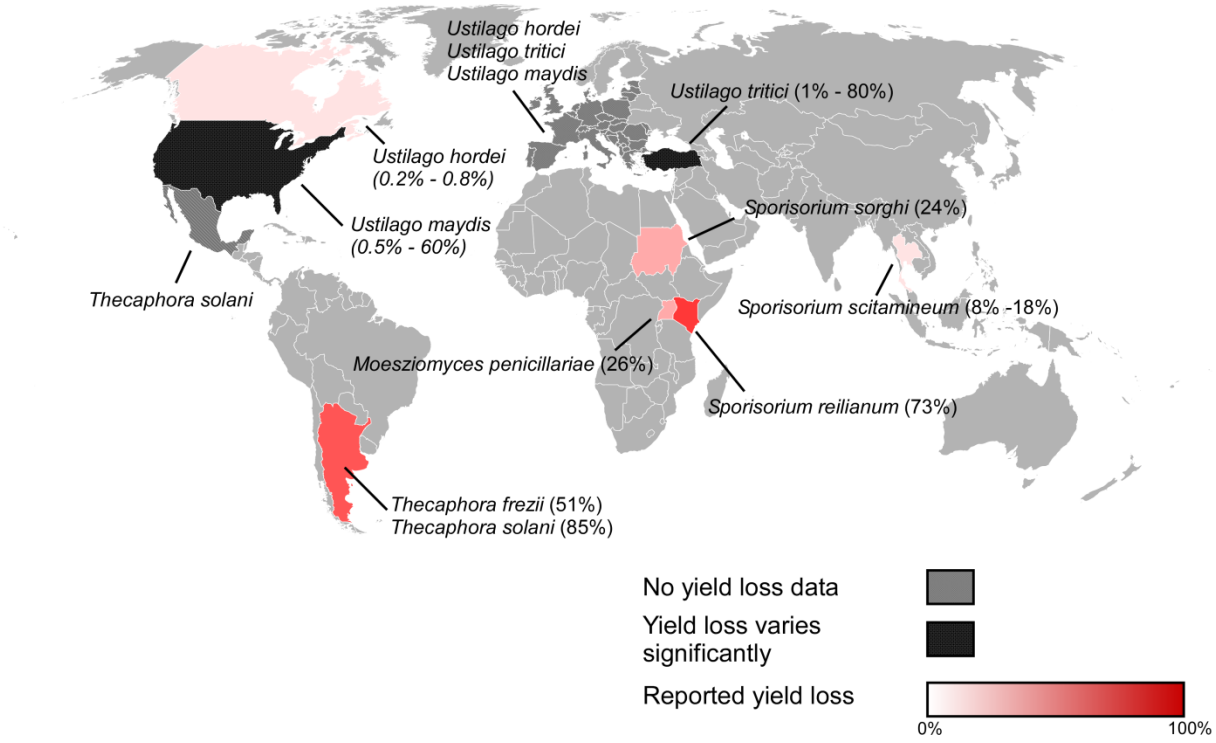


Figure 1. **Occurrence of smut diseases in several parts of the world**. Based on recent literature, the severity of smut diseases varies in different areas of the world depending on agricultural practices. The most striking disease reports come from the sub-Saharan Africa with *S*. *reilianum* on sorghum, but also from Latin America with the recent occurrence of *Thecaphora* diseases on the potato and peanut crops (Christensen, 1963; Parlak, 1981; Sundar, AR; Barnabas, EL; Malathi, P; Viswanathan, 1989; Ngugi *et al*., 2002; Andrade *et al*., 2004; Grewal *et al*., 2008; Smith *et al*., 2009; Little *et al*., 2012; Conforto *et al*., 2013; Lubadde *et al*., 2014).

Smut fungi are biotrophic plant pathogens of many agricultural crops, wild grasses and dicotyledonous weeds, that belong in the order *Ustilaginales* (Vanky, 2012). The occurrence of smut diseases is worldwide, however severity heavily varies depending on the agricultural practices (Figure 1). Most of the species have a narrow host range restricted between one and

five host species coming from the same family (approximately 86% of the records, Begerow *et al*., 2004). The symptoms caused, although they differ from species to species, have in common the deposition of dark or light brown colored teliospores (Begerow *et al*., 2014). In many hosts, the symptoms are restricted to the floral organs (Vánky & Lutz, 2007; Roets *et al*., 2008) which leads the reduction or complete failure of setting healthy seeds (Batts, 1955; Malik & Batts, 1960; Christensen, 1963). However some other species cause tissue rupture of the leaves (*U. striiformis* Alderman *et al*., 2007; *Thecaphora schwarzmaniana* Vasighzadeh *et al*., 2014) or of the belowground organs (*T. solani* Andrade *et al*., 2004; *T. frezii* Conforto *et al*., 2013). Furthermore, a more important symptom for the agricultural industry is also the quality reduction of the grains (Kronstad, 1997).

In addition to their significance as pests, members of the smut fungi have been developed into valuable model organisms for molecular biology and genetics. *U. maydis* in particular, is the most well-studied smut fungus and has been utilized since the 1960s (Holliday, 1961a, 1974) as a model organism for studying basic cell processes of eukaryotic organisms (Holliday, 1961b; Leaper *et al*., 1980; Kmiec & Holloman, 1986; Bauchwitz & Holloman, 1990). Since then, this species has been recruited from a number of labs around the world to study plant-microbe interactions (Skibbe *et al*., 2010; Hemetsberger *et al*., 2012; Lanver *et al*., 2014), fungal biology (Boyce *et al*., 2005) and cell biology (Becht *et al*., 2006). Most recently, *U. maydis* has been used as a heterologous protein expression platform for the biotech industry (Stock *et al*., 2012) giving to this species an application potential that no other fungal plant pathogen has had before.

1.3 Mating: a crucial checkpoint in the life of smut fungi

Mating has a central role in the life cycle of these pathogens, as they can reproduce sexually (teliospores) and asexually (sporidia) (Ingold, 1983). Most of our knowledge on the molecular mechanisms of the mating behavior of smut fungi comes from *U. maydis*. Once its diploid spores land on plant tissue, whether this is a leaf or the inflorescence of maize, they germinate and go through meiosis (Christensen, 1963; O'Donnell & McLaughlin, 1984). This process gives rise to four haploid sporidia. Non-self recognition and compatibility after mating is dependent on the two mating type loci, *a* and *b*. In *U. maydis*, the *a* locus has two alleles (Bölker *et al*., 1992), carrying one pheromone receptor-coding gene called *pra* and a pheromone precursor coding gene called *mfa*. The *b* locus which governs the compatibility of the dikaryon after cell fusion has multiple alleles (Holliday, 1961a; Puhalla, 1968).

In the *b* locus, two divergently transcribed genes can be found, named *bE* and *bW*, encoding proteins with a single homeobox domain each (Schulz *et al*., 1990; Gillissen *et al*., 1992; Kämper *et al*., 1995). For *U. maydis* and the other smut fungi, the *b* locus has been under extensive investigation since it is indispensable for infection (Gillissen *et al*., 1992). The *b* proteins, bE and bW, when originating from different alleles, form a heterodimeric transcription factor (Kämper *et al*., 1995), which targets genes necessary for filamentous growth, pathogenicity and appressoria formation (Wahl *et al*., 2010). Dimer formation is depended on a variable N-terminal domain that exists in both bE and bW (Kämper *et al*., 1995; Yee & Kronstad, 1998) (Figure 2).

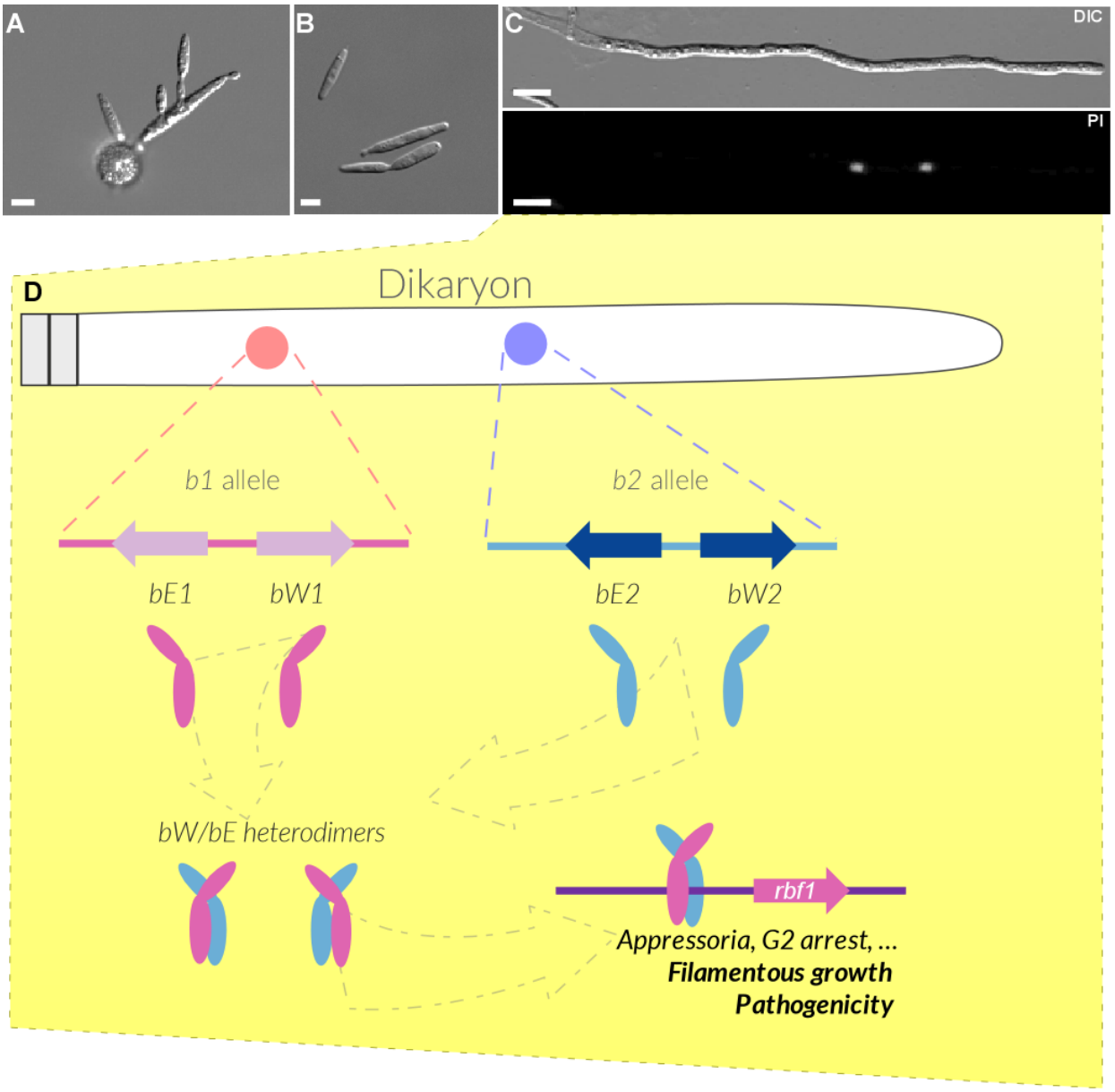


Figure 2. **The *b* locus of *U. maydis*.** Teliospores of *U. maydis* germinate with the growth of a probasidium, which after meiosis gives rise to 4 haploid sporidia (A). These sporidia grow asexually by budding and represent the saprophytic part of the life cycle (B). Once sporidia of opposing mating types come in close proximity, they identify each other by pheromone signaling driven by the genes in the *a* locus. After cell fusion, compatibility of the dikaryon is established by the products of the *b* locus. The products *bE* and *bW* if compatible, dimerize and initiate filamentous growth and also the expression of targets related to appressorium formation and pathogenicity (C, D). Bar = 50µm

In other smut fungi, mating functions under the same principle (Kellner *et al*., 2011) however variations on the linkage of the two loci exist. For example, *U. hordei* has a bipolar mating system with two loci called MAT-1 and MAT-2 (Lee *et al*., 1999) where the pheromone/pheromone receptor locus is linked with the homeobox domain-coding gene locus.

Other basidiomycetes have been found to carry the same genetic modules (Kües, 2015), although compatible interactions have different phenotypes. For example, in *Coprinus cinereus*, b-compatible interactions are responsible for clamp cell formation (Kües *et al*., 1994; Asante-owusua *et al*., 1996; Spit *et al*., 1998), while in *Cryptococcus neoformans*, the b homologues SXI1a and SXI2a are necessary for sexual development but not directly involved in virulence (Hull *et al*., 2005).

1.4 Genomics and genomes of smut fungi

The advances in short and long read sequencing technology as well as the decreasing costs are transforming mycology and plant pathology into genome-enabled sciences (Raffaele & Kamoun, 2012; Hibbett *et al*., 2013). Key genomic features of plant pathogens, such as the different families and abundance of carbohydrate active enzymes (CAZymes) (Zhao *et al*., 2014) or unique sets of small secreted proteins (Mueller *et al*., 2008; Guyon *et al*., 2014; Alfaro *et al*., 2014; Pellegrin *et al*., 2015) have become a vivid topic of study. Previously unknown features of fungal genomes, such as the role of small RNAs, have been brought to light and are now intensively researched (Weiberg *et al*., 2013; Yang, 2015).

The earliest and most important advance in understanding the genomes of smut fungi was the genome sequencing of *U. maydis* (Kämper *et al*., 2006). In this seminal paper, it was shown that the RNAi machinery was absent and also that virulence is dependent on a number of clustered small secreted proteins. Later on, this finding was supported with studies characterizing whole clusters, as a 24 effector-containing locus residing in the chromosome 19 (Brefort *et al*., 2014) and specific effectors within, as the effector *tin2* (Tanaka *et al*., 2014). In addition, certain genetic elements overrepresented in genomes of other fungi, for example CAZymes (Ma *et al*., 2010; de Wit *et al*., 2012) or high number of transposons (Spanu *et al*., 2010), in this genome were in low abundance, correlating with the biotrophic and sexual lifestyle and of this organism.

Subsequently, sequencing the genomes of *S. reilianum* (Schirawski *et al*., 2010) and *U. hordei* (Laurie *et al*., 2012) revealed new features, as for example species-specific and conserved secreted effector proteins but also an influence of transposable and repetitive elements in virulence and genome organization (Ali *et al*., 2014). The recently sequenced genome of *S. scitamineum* (Que *et al*., 2014; Taniguti *et al*., 2015), reinforces the idea that clustering of secreted effectors and a relatively low abundance of CAZymes is a characteristic of the genomes of smut fungi. Further efforts on sequencing additional species yielded interesting results, as in the genome of the *Melanopsichium pennsylvanicum*, where it was revealed that many conserved secreted effector proteins have been lost from its genome, suggesting a connection with host jumps and de-functionalization of otherwise important genes (Sharma *et al*., 2014). In a similar manner, the genome of *Anthracocystis flocculosa* (=*Pseudozyma flocculosa*) showed an analogous loss of secreted effectors which is hypothesized to be linked with its epiphytic lifestyle. Its increase of chitinase-encoding genes, belonging to the CAZyme family GH18, might explain its successful role as a biocontrol agent against foliar pathogens (Lefebvre *et al*., 2013).

The genomes of the anamorphic yeasts *P. brasiliensis* (Oliveira *et al*., 2014), *P. antarctica* (Morita *et al*., 2014), *P. hubeiensis* and *P. aphidis*, although available through the NCBI database, have not been yet fully described. Despite the fact that they have a close relation to plant pathogens, the studies available are focused more towards their biotechnologically relevant

lipid metabolism and plant cell wall degrading enzymes rather than on elements that could reveal a plant associated lifestyle.

Certainly these genomes constitute a small but very important sample out of the approximately 1400 smut fungi (Vanky, 2012) and have brought forward characteristics of a backbone that could exist in all *Ustilaginales* genomes. Additional genomes could help in answering specific questions regarding the evolutionary adaptation of smut fungi to new hosts but also how certain structural phenomena, for example, the clustering of secreted proteins, came to be.

1.5 *Thecaphora thlaspeos*, a fungal pathogen of Brassicaceae

Although there is a detailed understanding of many molecular pathways and elements that are utilized during infection in the genomes of smut fungi, mechanisms of resistance in the respective host plants remain unclear. Grasses, which have been studied as hosts of smut fungi, have notoriously complex and oversized genomes, and therefore responses to fungal infections are underexplored. Despite extensive breeding efforts (Baumgarten *et al*., 2007; Grewal *et al*., 2008; Knox *et al*., 2014) there is no identified resistance locus to smut fungi that has been molecularly characterized. On the other hand, a wealth of information on plant defense responses can be found in the model plant *Arabidopsis thaliana* (Dangl & Jones, 2001; Jones & Dangl, 2006; van der Hoorn & Kamoun, 2008) and consequently, it would be advantageous if research on smut fungi was complemented with a species that is able to infect *A. thaliana* or close relatives in the Brassicaceae.

This direction might also be helpful in understanding the relationship between smut fungi and dicot plants. To date, our information on the adaptation of smut fungi to dicot hosts is restricted to botanical descriptions of the fungal spores, their macroscopic symptoms (Vánky, 1994) and their phylogenetic relationship (Vánky *et al*., 2007; Begerow *et al*., 2014). Even though the first genome of a dicot infecting smut fungus, *M. pennsylvanicum*, has been recently available (Sharma *et al*., 2014), our knowledge on the infection strategies of the dicot smut fungi and their life cycle is quite limited. It has to be noted that *M. pennsylvanicum* is a unique case of a species residing in the *Ustilaginaceae* infecting non-grass hosts (Begerow *et al*., 2006), in contrast with the majority of the dicot smuts which are classified outside the *Ustilaginaceae*. Unfortunately, with *M. pennsylvanicum*, we are also limited in our ability to examine the responses of the host plant *Polygonium sp*., since there are no efficient transformation methods or any knowledge on its defense responses. Therefore, we are still in need of pathosystem where it would be possible to genetically manipulate both the smut pathogen and its dicot host.

1.6 Aim of this work

In the effort to gain more knowledge on the dicot infecting smuts and the responses of their hosts, the *Glomosporiaceae* family presents a significant opportunity. This family contains the single genus *Thecaphora* with the highly significant plant pathogens of potato and peanut, *T. solani* (Andrade *et al*., 2004) and *T. frezii* (Conforto *et al*., 2013).

Most importantly, *T. thlaspeos* is a species of this family that is able to infect several Brassicaceae species (Vánky *et al*., 2007) and has the potential to develop as a supplementary pathosystem to *U. maydis* and maize. This possibility has been recently highlighted by members of the smut fungi community (Wollenberg & Schirawski, 2014).

In order to investigate the biology of these species and their effector repertoire, this thesis aims:

1. To culture and isolate haploid cells of the fungal plant pathogen *T*. *thlaspeos*.
2. To describe the life cycle of *T*. *thlaspeos* by presenting its teliospore germination, macro- and microscopic symptoms on the host.
3. Sequence and annotate the genome of the species in order to analyze and compare its features to the genomes of other sequenced smut fungi.

2 RESULTS

2.1 Part I: Collection and culture of *T. thlaspeos*

2.1.1 Sample collection and macroscopic symptoms

Reports of *T. thlaspeos* date back to the beginning of the 20th century (Frantisek, 1916), under the species name *Tilletia* or *Ustilago thlaspeos* infecting the host plants *Arabis hirsuta*, *Ar. petraea*(= *Arabidopsis lyrata subsp. petraea*) and *Thlaspi alpestre* in central and north Europe. Since then, *T. thlaspeos* has been reported in 15 host species in several locations in Europe as well as in South Korea (Vanky, 1999; Denchev *et al*., 2007). Unfortunately, only the most recent report is accompanied by molecular evidence such as the sequence of the internal transcribed spacer (ITS) or the rDNA large subunit (LSU) (Vánky *et al*., 2007), and thus the precise comparison of the isolates is unfeasible. During the years 2012 and 2013, *T. thlaspeos* was collected in Germany, Slovenia, and Greece in the three different *Brassicaceae* species, *Ar. hirsuta, Ar. ciliata*, and *Ar. sagitatta* (Table 1). Subsequent collections the following years hint that both the pathogen and host populations are established in the area. In all three collections, the LSU sequence matched the reference sequence, with the occurrence of one polymorphism in all isolates and one polymorphism exclusive in the Greek isolate (Suppl. Data 1). It has to be noted that *Ar. ciliata* was not reported before as host for this pathogen, suggesting that *T. thlaspeos* might have a broader range of hosts in the *Brassicaceae* than considered before.

Table 1. Collections of *T. thlaspeos* teliospores.

Hosts were identified based on macroscopic characteristics and their ITS sequence and *T. thlaspeos* based LSU sequence. Genomic DNA from the hosts was obtained by the leaf tissue and for the pathogen from teliospores.

Location	Host	Year	Evidence	Collector or Annotation of sample
Petrota, Greece	*Ar. sagitatta*	2013	LSU	L. Frantzeskakis
Ronheim, Germany	*Ar. hirsuta*	2013	LSU	M. Feldbrügge, J. Kruse
Bad Berneck, Germany	*Ar. hirsuta*	2013	LSU	M. Feldbrügge, J. Kruse
Hohe Leite, Germany	*Ar. hirsuta*	2013	LSU	M. Feldbrügge, J. Kruse
Eselsburg, Germany	*Ar. hirsuta*	2013	LSU	M. Feldbrügge, J. Kruse
Kranjska Gora, Slovenia	*Ar. ciliata*	2012	LSU	T. Langner, E. Vollmeister, R. Kellner

In all cases, infected plants were identified by the presence of teliospores within the silique tissue (Fig. 3B, C). During the collections, no other symptom, such as stunted growth, distorted leaf growth, abnormal root, shoot or flower development, was consistently observed in the infected plants. Occasionally, infected and healthy plants were found in very close proximity (Fig 3D)

suggesting that the infection does not spread as aggressively as other fungal diseases (for example rust fungi). Conidia in the flower anthers, identified as the anamorph *Rhombiella cardamine*s in past reports (Vánky, 1994), were not observed.

Figure 3. **Symptoms of *T. thlaspeos* infection on *Ar. hirsuta* and *Ar. ciliata***. Infected and healthy host plant *Ar. hirsuta* as collected from the field (A, right and left respectively, Bar = 2cm). In siliques from the infected *Ar. hirsuta* plants spores cover the developing seeds (B, up, healthy down, Bar = 1 mm). Similarly, infected *Ar. ciliata* siliques present the same symptoms (C, Bar = 1mm). Infected and healthy *Ar. ciliata* plants were found occasionally in close proximity (D, infected plants are marked with an asterisk).

Collected samples from different plants were examined to further describe the symptoms. Spores develop simultaneously with the seed in the silique. As the seed development progresses, spores turn from hyaline/white to light brown and cover a larger portion of the seed (Fig. 4A-C). While in the literature it is reported that the seeds of the infected plants are completely replaced by spores (Vanky, 1999), here it was observed that there is a range of the severity of the symptoms, extending from the full replacement of the seeds by the teliospores to normal development of the seeds being coated with teliospores (Fig. 4 A-C).

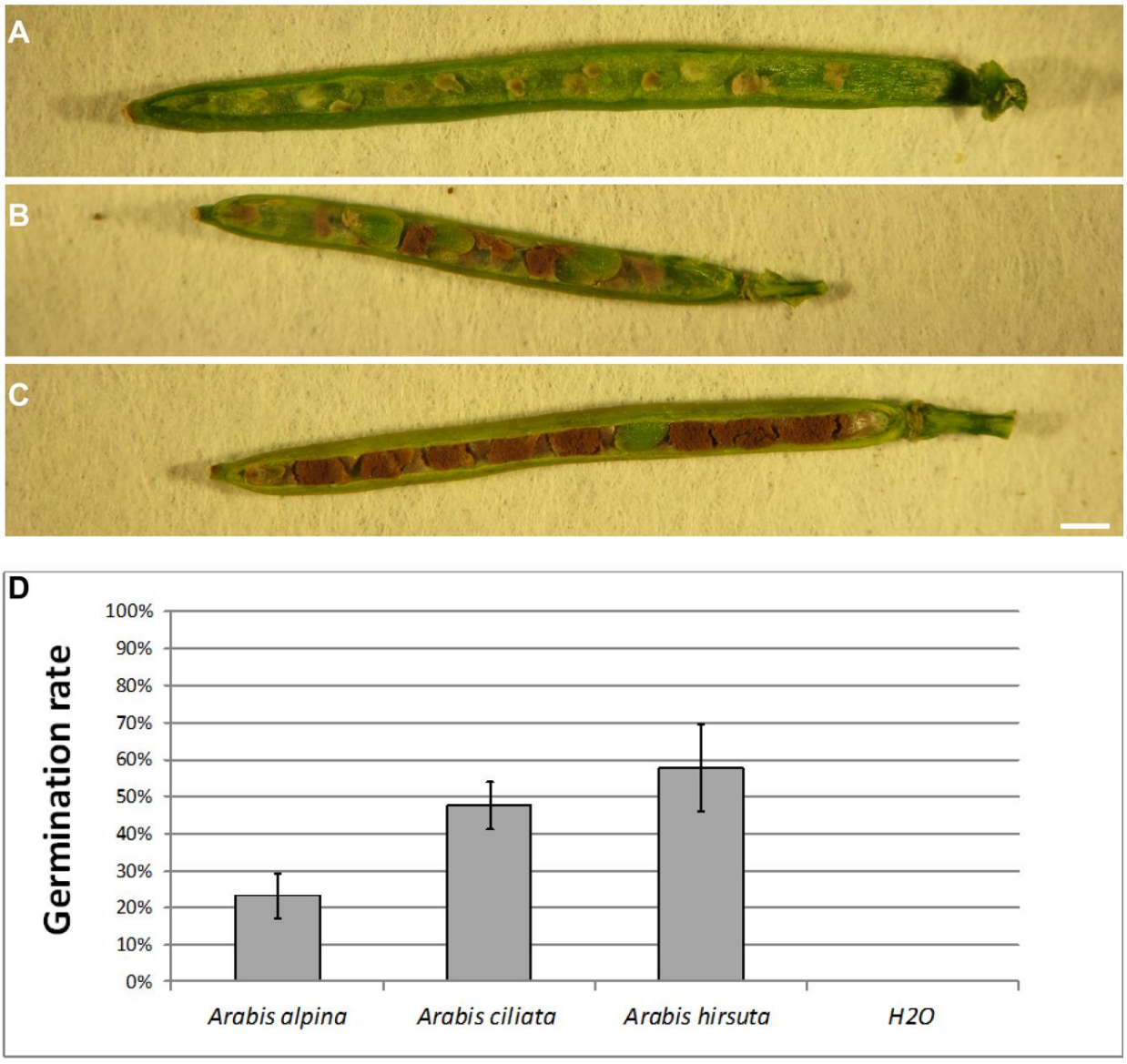


Figure 4. **Maturation of teliospores in *Ar. ciliata* infected siliques**. Infected siliques of *Ar*. *ciliata* plant collected in Slovenia 2013. Spores initially appear in the seed stem having white/hyaline color (A). As the seed develops they turn brown and start covering the whole seed (B, C). Bar = 1mm. Germination of teliospores is initiated by incubation in Brassicaceae leaf extracts (D). Bars show standard deviation between three experiments.

2.1.2 Teliospore germination of *T. thlaspeos*

As it has been shown by C. T. Ingold in his extensive studies of the basidium that teliospore germination in smut fungi is not uniform (Ingold, 1983, 1989, 1997, 1998). In general, teliospores germinate readily in minimal media such as 0.2% malt agar or water agar, without any requirements for specific temperature or humidity (summarized in Suppl. Table S1). *T. thlaspeos* teliospores however, remain dormant in many tested standard, nutrient rich or nutrient poor media (Suppl. Table S2). Interestingly, in media containing leaf extracts from Brassicaceae, filaments emerged from the teliospores and kept proliferating (Figure 4D, 5A). The same response is observed when spores are placed next to germinating seeds (Figure S1, Movie 1). This suggests that the plant signal that is responsible for the initiation of teliospore germination, does not exist exclusively in the aboveground tissues (Figure S1, Movie 1). Other closely related *Thecaphora* species, which were examined in this thesis, did not present the same requirement for germination (Paragraph 2.1.4).

T. thlaspeos does not produce haploid sporidia as *U. maydis* and other grass smuts, and neither are the emerging filaments in cell cycle arrest. Germinated spores form compact filamentous colonies on plates, which are growing at a much slower rate than *U. maydis* sporidial colonies.

Nuclear staining shows that the filaments of the teliospores are initially monokaryotic or multinucleate. Mitotic divisions of the nuclei are likely to be leading in proliferation of the filaments, which branch and insert of empty sections at the basal pole of every cell (Figure 5A, Movie 2). This growth pattern resembles some species from the genus *Thecaphora* (Andrade *et al*., 2004; Vánky *et al*., 2007), but clearly differs from other smuts which develop a probasidium to form haploid sporidia after the meiosis of a diploid nucleus (Ingold, 1988, 1989). Prolonged culture of the *T. thlaspeos* filaments on plate does not result in clamp cells or coiled structures as reported for the potato smut *T. solani* (Andrade *et al*., 2004), nor conjugation tubes between compatible compartments as reported for some *Ustilaginaceae* species (Ingold, 1988, 1989). In vigorously shaking liquid cultures, the filaments separate at the empty sections and growth is more uniform. These filaments are exclusively monokaryotic and their branching pattern is consistent (Figure 5B, Movie 2).

Metagenome sequencing of the spores and mating type specific markers did show that the spores from different infected samples carry at least two mating types, suggesting that they are diploid as in *U. maydis*. Resulting cultures obtained from filamentous colonies after extensive re-streaking and sub-culturing, were shown to be haploid and carry a single mating type based on genome sequencing (Figure 12C, Paragraphs 2.2.1, 2.2.4), suggesting that meiosis is likely occurring at some point after spore germination.

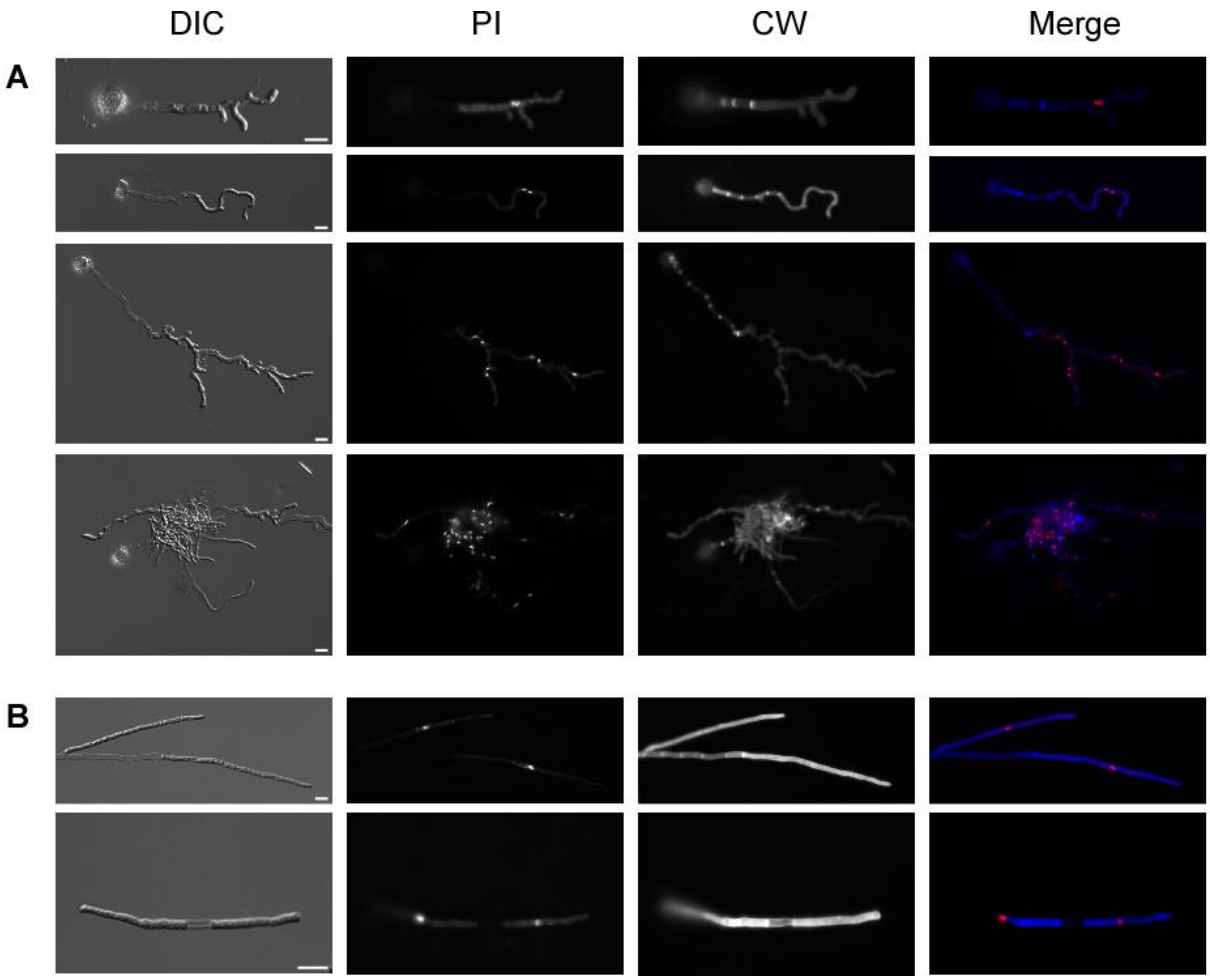


Figure 5. **Germination of *T. thlaspeos* teliospores**. Early stages of *T. thlapseos* spore germination (A) and nuclear content after propidium iodine (PI) staining. Dikaryotic filaments grow by inserting septa at the basal pole (shown by Calcofluor White staining, CW). Prolonged incubation (>14 days) leads to distorted growth, of multinucleate filaments and clumps. Vigorous shaking of such clumps leads to isolation of haploid filaments (B) which divide mitotically by inserting empty sections. Bar = 10μm.

2.1.3 The biotrophic phase of the *T. thlaspeos* life cycle

In order to characterize the infection process of *T. thlaspeos* host plants that were collected during 2012 and 2013 were sampled and stained with wheat germ agglutinin (WGA) and propidium iodine (PI). In the infected plants, fungal hyphae were found in the rosette and shoot leaves, as well as in the infected siliques. This suggests that *T. thlaspeos* grows systemically in the whole plant and possibly, since some hosts are perennials, is able to overwinter even after the inflorescence is removed. The fungal growth is limited in the vascular bundle of the leaf, and it is observed both the main and the secondary veins. Both studied hosts, *Ar. ciliata* and *Ar. hirsuta* showed the same symptom (Figure 6).

Table 2. Infection experiments with *T. thlaspeos* by co-incubation of teliospores and host seeds. Infection was scored by WGA/PI staining of whole plantlets grown for 2 weeks

Experiment date	Host	Spore sample	Infection result
18.7.2013	*A. thaliana Col-0*	Slovenia – 21.5-2	0/10
18.7.2013	*A. thaliana cerk1*	Slovenia – 21.5-2	0/10
19.7.2013	*Ar. saggittata*	Slovenia – 21.5-2	0/10
19.7.2013	*Ar. hirsuta*	Slovenia – 21.5-2	0/10
19.7.2013	*Ar. alpina pep1*	Slovenia – 21.5-2	0/10
19.7.2013	*Ar. ciliata*	Slovenia – 21.5-2	0/10
28.7.2013	*A. thaliana Col-0*	Slovenia – 21.5-2	1/16
28.7.2013	*A. thaliana cerk1*	Slovenia – 21.5-2	1/16
17.9.2013	*A. thaliana Col-0*	Slovenia – SL71	0/10
17.9.2013	*Ar. alpina pep1*	Slovenia – SL71	0/10
24.9.2013	*A. thaliana Col-0*	Slovenia – SL71	0/10
24.9.2013	*Ar. ciliata*	Slovenia – SL71	0/16
3.10.2013	*Ar. hirsuta*	Slovenia – SL75	1/16
3.10.2013	*Ar. hirsuta*	Germany – RH13	0/16
3.10.2013	*Ar. ciliata*	Slovenia – SL75	0/16
3.10.2013	*Ar. ciliata*	Germany – RH13	1/16
17.1.2014*	*A. thaliana Col-0*	Germany - Es	9/24

* This assay was conducted in 24-well plates and not on soil

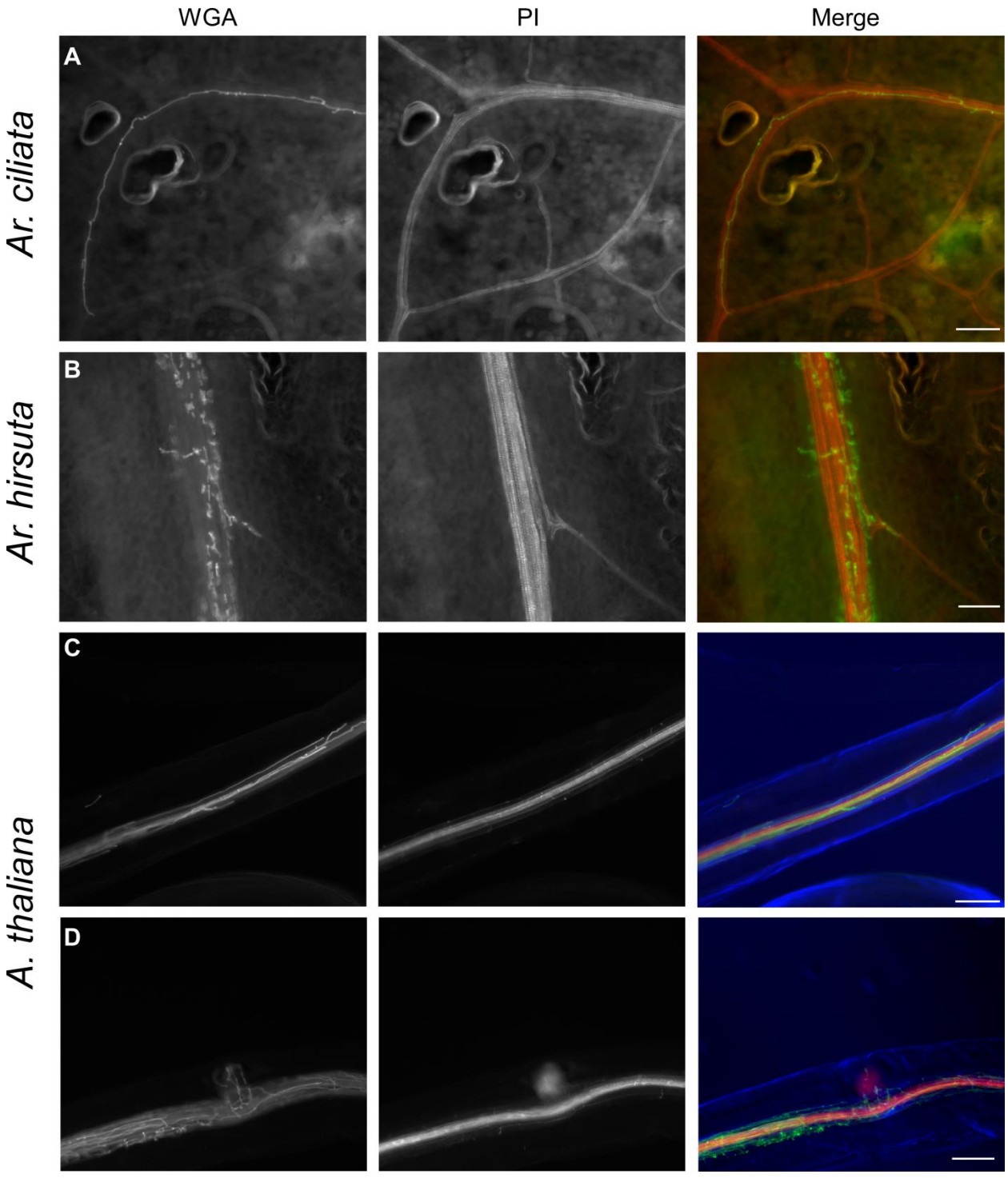


Figure 6. Staining of infected host plants from the field collections or lab infection assays. *Ar*. *ciliata* collected from the field during 2013 in Slovenia was stained with WGA and PI (A). Hyphae were found to be growing along the vascular bundle. The same was observed with *Ar*. *hirsuta* plants collected the same year in Bad Berneck, Germany (B). In the infection assays with *A*. *thaliana* both the WT (C) and the *cerk1* (D) roots were found to be colonized in the same way by hyphae. Bar = 100µm

In a first attempt to develop the *T. thlaspeos - A. thaliana* pathosystem and to replicate the observations from the field collections, artificial infection experiments were designed. When *T. thlaspeos* spores and *A. thaliana* Col-0 were co-incubated in 24-well plates with plant media (1/2 MSN, 1% Sucrose), spores germinated and hyphae were found to be attached to the roots. The same result was obtained in assays in the soil, and with the *A. thaliana cerk1* mutant. The *cerk1* mutant is chitin insensitive (Petutschnig *et al*., 2010) and it was chosen for the infection assays in order to determine if it would be more susceptible to *T. thlaspeos*.

However, the assays had very low infection rates (summarized in Table 2) and due to limitations of the available material (spores), the experiments were not followed up. The reasons for the low infection rate could be both technical (amount of the initial inoculum, proximity of spores to the root, environmental) and related to the material (different spore batches, possible incompatibility with the host). Indeed, infection assays based on this protocol yielded reproducible infection at higher rates when using spore samples from other collections (Courville K., Göhre V., pers. communication).

In the plants colonized by fungal hyphae in these assays, the pattern of growth was identical with the plants collected from the field, independent of the host species. Fungal hyphae, which could be traced back to teliospores, attached and penetrated the root tissue (Figure 6).

2.1.4 A brief report on other *Thecaphora* species

During the course of this project, additional species were investigated in order to understand the germination and growth behavior of *Thecaphora* species in general. During the years 2013-2014 the three additional *Thecaphora* species, *T. oxalidis*, *T. saponariae* and *T. seminis-convolvuli* and one Exobasidiomycete, *Tilletiopsis pallescens*, were collected (Table 3).

Table 3. Additional species collection

Location	Species	Host	Year	Evidence	Collector or Annotation of sample
Germany, Düsseldorf	*T. seminis-convolvuli*	*C. sepium*	2013	LSU	M. Feldbrügge
Germany, Isteiner Klotz	*T. saponariae*	*S. officinalis*	2013	LSU	M. Feldbrügge, J. Kruse
Germany, Röslau	*T. oxalidis*	*O. stricta*	2013	LSU	M. Feldbrügge, J. Kruse
Germany, Röslau	*Ti. pallescens*	*O. stricta*	2013	LSU	L. Frantzeskakis

Oxalis stricta plants infected by *T*. *oxalidis* did not present any loss of vigor, had their anthers covered with conidia and their seed pots filled with teliospores (Figure 7M).

Infected *Saponaria officinalis* plants by *T*. *saponariae* produced green enlarged flower buds instead of flowers, which dried in a period of 2-3 weeks releasing the teliospores (Figure 7H, I). In these infected plants, conidia were not detected as reported previously (Vánky & Lutz, 2007), however since *T*. *saponariae* infects multiple species, conidial formation might be depended on host.

Calystegia sepium plants infected by *T*. *seminis-convolvuli* are easily identified because the anthers of their flowers are shortened compared to healthy ones (Figure 7D, E). These shortened anthers are covered with conidia. Seed development in infected individuals progresses normally, however the endosperm of the seeds is filled with teliospores. Infected seeds are able to germinate normally, resembling the symptomatology of *T*. *thlaspeos* infections (Suppl. Figure S12, Paragraph 2.1.1).

The teliospores of the three *Thecaphora* species germinated between 3-6 weeks in water-agar plates at room temperature (~20 ^{o}C) and did not present a requirement for a host signal. Germination was filamentous as in *T*. *thlaspeos*, however *T*. *saponariae* germinates also by producing aerial hyphae which bear conidia (Figure 7F). These observations are in agreement with previous descriptions of these species (Ingold, 1987; Vánky & Lutz, 2007; Vánky *et al*., 2007).

Filaments of all three species are proliferating as in *T*. *thlaspeos*. Cultures can be obtained by incubating the filaments in nutrient rich media, as the Complete Medium for *U*. *maydis*. Interestingly, *T*. *seminis-convolvuli* filaments produce conidia at their tips when in rich media (Movie 3). This behavior stops and growth becomes solely filamentous after prolonged incubation of cultures in the same medium, suggesting that there might be a link to the availability of nutrients (Suppl. Figure S6B).

The conidia found on the anthers of *T*. *oxalidis* and *T*. *seminis-convolvuli* infected plants. germinate in a period of 6-24 hours and occasionally form conjugation tubes between different compartments (Figure 7C, Movie 4, 5).

Sampling of tissues from infected *Oxalis stricta* plants, suggests that *T*. *oxalidis* also grows close to the vascular bundle (Suppl. Figure S6A), as is *T*. *saponariae* in *S*. *officinalis* (Suppl. Figure S6A).

Overall, the germination behavior of these *Thecaphora* species is quite comparable with *T*. *thlaspeos* with the exception of the host signal requirement. In the future, it would be interesting to explore whether the differences in the germination behavior of *T*. *thlaspeos* compared to the other *Thecaphora* species is also reflected in their genomes.

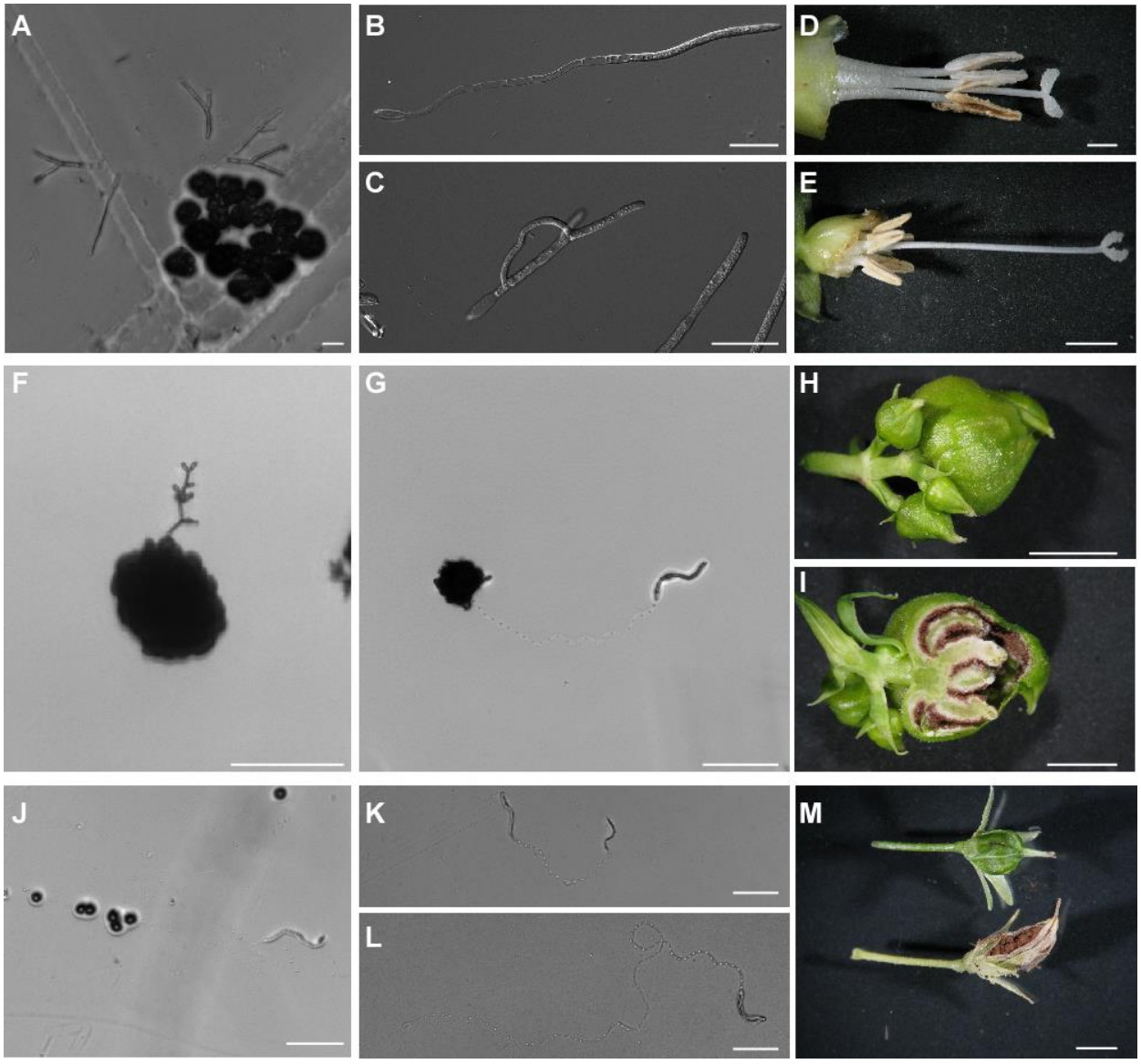


Figure 7. Germination, growth and symptoms of three *Thecaphora* species. Spores of *T*. *seminis-convolvuli* germinate filamentously and so are the conidia (A, B, Bar = 100µm). Occasionally, conjugation tubes between compatible compartments are observed (C) Bar = 100µm. A distinctive symptom of *T*. *seminis-convolvuli* infected plants is the shortening of the anthers (E) compared to healthy plants (D). Bar = 1cm. *T*. *saponariae* germinates either filamentously or by producing conidia (F, G Bar = 100µm). *S*. *officinalis* plants do not produce flowers, but flower buds get swollen and filled with spores (H, I bar =0,5cm). *T*. *oxalidis* spores germinate filamentously with no observed variation (J, Bar = 100µm). Conidia germinate also filamentously, occasionally with more than one filament (K, L, Bar = 100**n**m). *O*. *stricta* plants that are infected do not produce seeds, instead seed pots get swollen and are filled with spores (M, Bar = 0,5cm).

2.2 Part II: The genome of *T. thlaspeos*

2.2.1 *De-novo* assembly of the *T. thlaspeos* genome

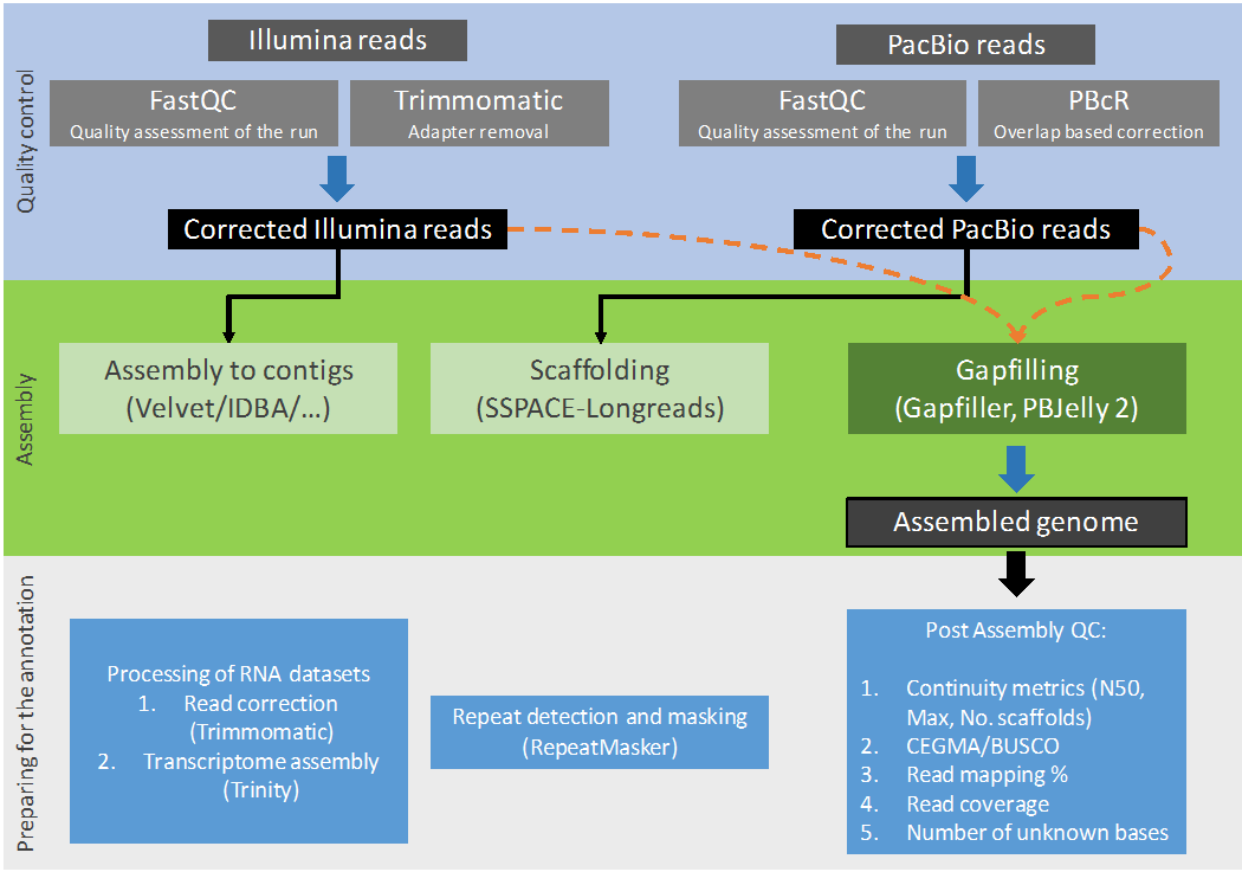


Figure 8. Overview of the *de-novo* genome assembly pipeline used. The pipeline consists of three parts, the initial quality control, the assembly of the reads and a final part of post-assembly quality control and data preparation for the structural annotation.

In order to assemble the *T. thlaspeos* genome the haploid culture *LF1*, of the mating type *a1b1* (Figure 12C, noted as "ES3"), was selected for sequencing. The Mi-Seq Illumina platform was used to perform paired-end sequencing of a library with insert size 500-700bp. The two sequencing rounds resulted in 6.75 GB of data (1st round 2.75 GB, 2nd round 4 GB). This dataset was later supplemented with two additional PacBio P6-C4 runs, which resulted in 49.8 GB of data. In addition, the second haploid culture LF2 (Figure 12C, noted as "RH13b"), of the mating type (a2b2), was sequenced at high coverage rate, following the same pipeline with the Mi-Seq Illumina platform (3 GB of data).

Table 4. Synopsis of the genomic assembly[#]

	T. thlaspeos	***A. flocculosa***	***U. maydis***	***U. hordei***	***S. reilianum***
	Assembly Statistics				
Number of scaffolds*	77	39	27 (23)	713 (23)	45 (23)
Total contig length (MB)	20	23.2	19.7	20.7	18.2
Total scaffold length (MB)	20.4	23.4	19.7	21.2	18.5
Average base coverage**	84x/39x	28x	10x	20x	20x
N50 contig (Kb)	28.2	38.6	127.4	48.7	50.3
N50 scaffold (Kb)	893.6	919.9	885.1	307.7	772.4
GC content (%)	61.2	65.1	54	52	59.7
GC content (Coding)	63.16	66.3	56.3	54.3	62.6
GC content (Noncoding)	58.56	63.7	50.5	49	54.3
	Coding sequences				
Percentage coding (%)	56,8	54.3	61.1	57.3	64.7
Average gene size (bp)***	1783	2097	1836	1782	1858
Average gene density (genes/MB)	319	291.5	344.5	335.4	360.7
Protein coding genes	6509	6877	6787	7111	6673
Ab-initio only predictions	299 (4,5%)	183 (2.7%)	312 (4.6%)	668 (9.4%)	223 (3.3%)
Exons	18209	19318	9778	10995	9777
Average exon size	637	658	1230	1103	1222
Exons/gene	2.79	2.8	1.44	1.55	1.47
tRNA genes	164	176	111	110	96
	Noncoding sequences				
Introns	11700	12427	2991	3884	3104
Introns/gene	2.28	1.8	0.44	0.55	0.47
Intron-free genes	1381 (21,2%)	1565 (22.7%)	4910 (72.3%)	4821 (67.8%)	4732 (70.9%)
Average intron length (bp)	129	141	146	141	146
Average intergenic length (bp)	1120	1273	1061	1098	919
Transposable elements and repeats (%)	3.7	3.4	2	10	0.8

* In parenthesis the number of chromosomes, where it is known

** In T. thlaspeos the coverage corresponds to Illumina and PacBio data respectively

***Excluding introns

Data for *A. flocculosa, U. maydis, U. hordei, S. reilianum* are based on the (Lefebvre et al., 2013)

In order to assemble the genome, the pipeline described in (Faino & Thomma, 2014) was followed. During this process, different assemblers were tested for their performance in the dataset, and the results were scored based on sequence continuity. Briefly, the trimmed Illumina

reads were assembled with IDBA-UD resulting in 1441 contigs. Afterwards they were scaffolded using SSPACE-Longreads and the corrected PacBio reads resulting in 86 scaffolds. Gap-filling using the Illumina and the corrected PacBio reads resulted in the final assembly (version 0.9) of 78 scaffolds with 2326 unknown nucleotides total. The mitochondrial genome was fully assembled in the Scaffold 2, with a resulting size of 126.5 kb. This dataset has sequence continuity comparable to the finished genomes of *U. maydis* and *S. reilianum*, especially considering the fact that 30 of 78 scaffolds have sizes shorter than 3 kb.

The total genome size is 20.5 MB, with a low repeat content of 3.76% mainly composed of simple repeats (Table 4, Suppl. Table S3). In the scaffolds 3, 7, 9, 11, 12 the telomeric repeat sequence TTAGGG can be found in the beginning and the end of the sequence, suggesting that these sequences might represent full chromosomes. In 25 of the scaffolds, the motif can be found either in the beginning or in the end, suggesting that these sequences might represent chromosomes where the sequencing or the assembly of the centromere region was not feasible.

Completeness analysis based on BUSCO (Simao *et al*., 2015) shows that the majority (Table 5, extended results in Suppl. Table S4) of single copy orthologs from the dataset for fungi can be found in this assembly. The results are comparable to the other smut genomes (Suppl. Table S4), indicating that orthologs that cannot be detected in this assembly might be a result of fragmentation or due to the detection limitations of the software. Similar results, showing high completeness, were obtained with the CEGMA analysis (Suppl. Table S5, Parra *et al*., 2007).

A draft assembly of the *T. thlaspeos LF2* genome was provided by Andreas Brachmann (Biocenter of the LMU Munich). The genome was assembled using CLC Genomics workbench and resulted in 1,485 contigs. The resulting assembly, although less complete as reflected in the BUSCO results (Table 5), can still be used for specific purposes as for example identifying the second mating locus.

Table 5. BUSCO analysis results

	T. thlaspeos LF1		***LF2***		***U. maydis***	
	#	%	#	%	#	%
Complete Single-Copy BUSCOs	1327	92	1236	86	1317	91
Complete Duplicated BUSCOs	74	5.1	34	2,4	84	5.8
Fragmented BUSCOs	50	3.4	117	8,2	83	5.7
Missing BUSCOs	61	4.2	80	5,6	38	2.6
Total BUSCO groups searched	1438		1438		1438	

2.2.2 Gene prediction and functional annotation

For the annotation of the genome, the MAKER pipeline was utilized (Holt & Yandell, 2011). As supporting evidence for the gene calls, the proteomes of 13 smut fungi (listed in Materials & Methods), the Uniprot protein dataset and an RNAseq dataset coming from a culture of *T.*

thlaspeos LF1 were used. Compared to other methods, MAKER was considered as the optimal choice since it is fully automated and provides several metrics on the final gene models, such as the Annotation Edit Distance (Eilbeck *et al*., 2009; Holt & Yandell, 2011). In addition this pipeline has successfully been used before to annotate genomes of eukaryotic plant pathogens (Lévesque *et al*., 2010; Cantu *et al*., 2013).

After the initial run of MAKER, 397 gene models were manually curated to generate a dataset for further training of the gene predictors AUGUSTUS and SNAP. The genes for this dataset were picked randomly, with the restriction to have good RNA coverage (>20 reads/bp throughout the predicted gene model). Subsequently, the gene predictors were iteratively trained with the outcome of each MAKER round. Appropriate gene models were picked with the provided MAKER scripts adjusted for SNAP and AUGUSTUS.

The resulting final annotation set contains 6509 gene models of which only 299 are completely *ab-initio* predictions with no supporting evidence (AED = 1.00). The overall metrics on the annotation result, as summarized by the AED metric (Cantarel *et al*., 2008), show that the majority of the gene models can be considered confident, based on the fact that the supplied evidence correlate with the splice sites and overall gene model structure (<0.25 AED for 87% of the gene models, Suppl. Figure S2). Inconsistencies that were observed after manual inspection between the final gene model and the mapping of the RNA reads are most likely coming from potential alternative splicing of the genes. Notably, the majority of the genes in *T*. *thlaspeos* contain at least one intron, a trend that seems to be shared with *A*. *flocculosa*, while other characteristics, as for example gene density or gene size, seem to be similar to the other smut genomes (Table 4).

2.2.3 Synteny

Synteny between the genomes of plant pathogenic fungi has been a very useful approach to identify regions with high rate of evolution. In both ascomycete (O'Connell *et al*., 2012) and basidiomycete (Schirawski *et al*., 2010) fungal pathogens, loci of low synteny harbor genes that confer species specificity, however synteny between species of the two phyla seems to have different attributes (Hane *et al*., 2011). In smut fungi, high levels of synteny have been observed so far between *U*. *maydis*, *S*. *reilianum*, *U*. *hordei* and *S*. *scitamineum* (Schirawski *et al*., 2010; Laurie *et al*., 2012; Taniguti *et al*., 2015), following their close phylogenetic relationship.

In *T*. *thlaspeos*, genomic synteny correlates with the multi-locus phylogeny results (described in paragraph 2.2.4). High levels of synteny can be observed between *T*. *thlaspeos* and *A*. *flocculosa*, while with *U*. *maydis* synteny is more limited (Figure 9). Certain *U*. *maydis* chromosomes though, are completely homologous to scaffold sequences of *T*. *thlaspeos* (e.g. *U*. *maydis* Chromosome 4 and Scaffold 9 in Figure 9). Between *A*. *flocculosa* and *T*. *thlaspeos*, inversions and translocations that disrupt synteny can be observed (e.g. *T*. *thlaspeos* Scaffold 3 and *A*. *flocculosa* Scaffold00002 in Figure 9) although interpretation of these events should be cautious since both genomes are not assembled to chromosomes. In accordance with these observations, areas important for virulence, such as the Cluster 19A of *U*. *maydis* (Brefort *et al*., 2014) are absent from *T*. *thlaspeos* although synteny with the flanking areas is conserved (Suppl. Figure S3A).

Interestingly, chromosomes 3 and 5 of *U*. *maydis* have one region each with complete loss of any synteny to *T*. *thlaspeos* (Suppl. Figure S4). The same locations can be observed in *U*. *maydis* and

A. flocculosa comparisons, but not with species residing in the grass smut clade (Supp. Figure S4). Although it is hypothesized that hyperactive areas like those contribute to pathogenicity, the actual impact these rearrangements have on the genome is still poorly understood (Stukenbrock & Croll, 2014). In the future, it would be interesting to explore in genomes closely related to *T. thlaspeos* and *A. flocculosa* the origin of these re-arrangements.

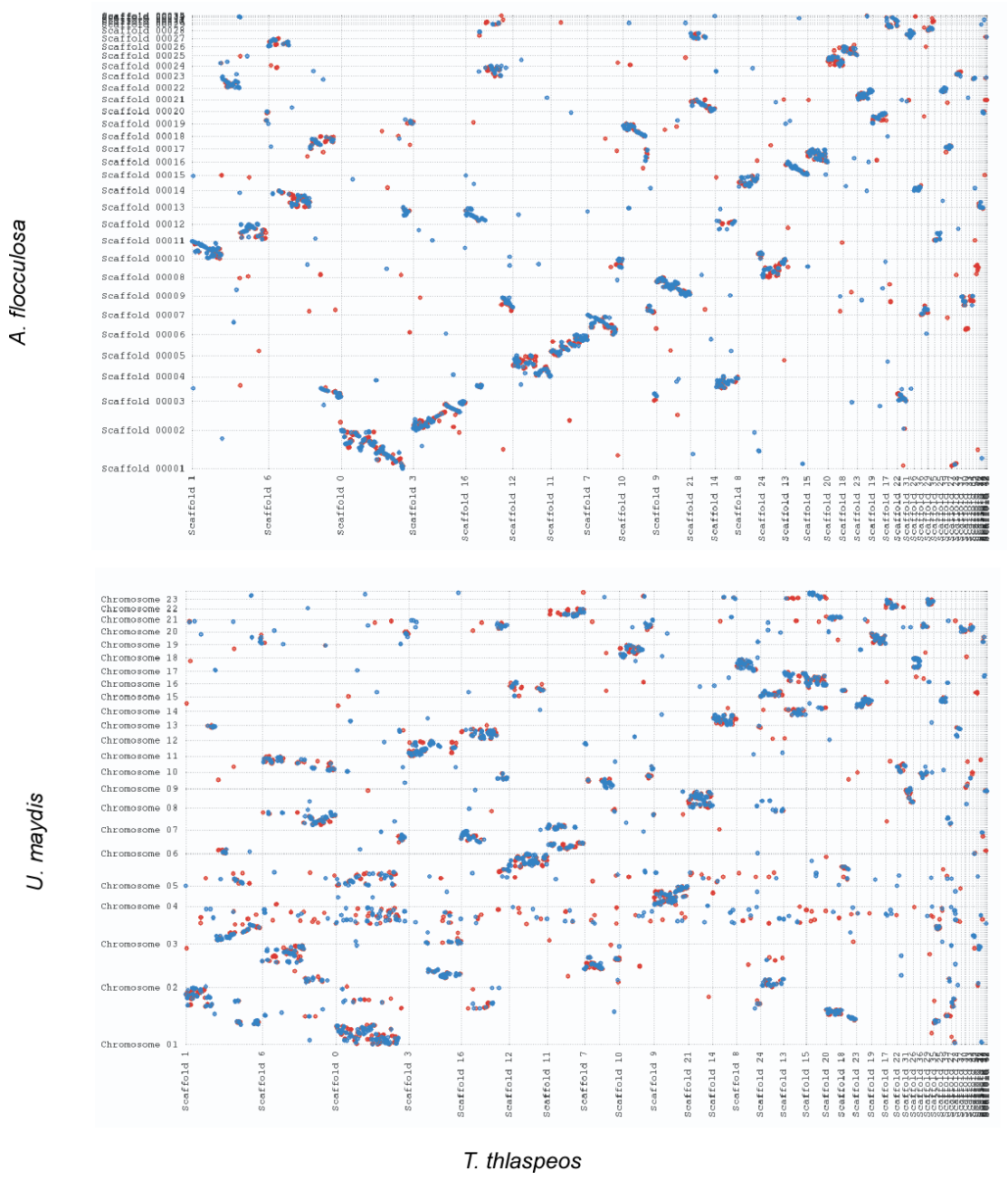


Figure 9. Synteny of the *T. thlaspeos* genome with related smut fungi. Synteny between *T. thlaspeos* and *A. flocculosa* (A), and *T. thlaspeos* and *U. maydis* (B). Blue dots indicate sense to sense alignment blocks, and red dots indicate sense to anti-sense alignment blocks. Scaffolds in *T. thlaspeos*, *U. maydis* and *A. flocculosa* with a size smaller than 10 kb were removed.

2.2.4 Core genome of smut fungi

Genes that are shared between the sequenced smut fungi and *T. thlaspeos* were investigated in order to define the genomic characteristics of this group of pathogens but also to identify unique genes, which might be involved in virulence. For this task OrthoFinder (Emms & Kelly, 2015) was used to detect shared orthologs between the 14 smut genomes (Table 9). The analysis resulted in 12775 orthogroups, including singletons.

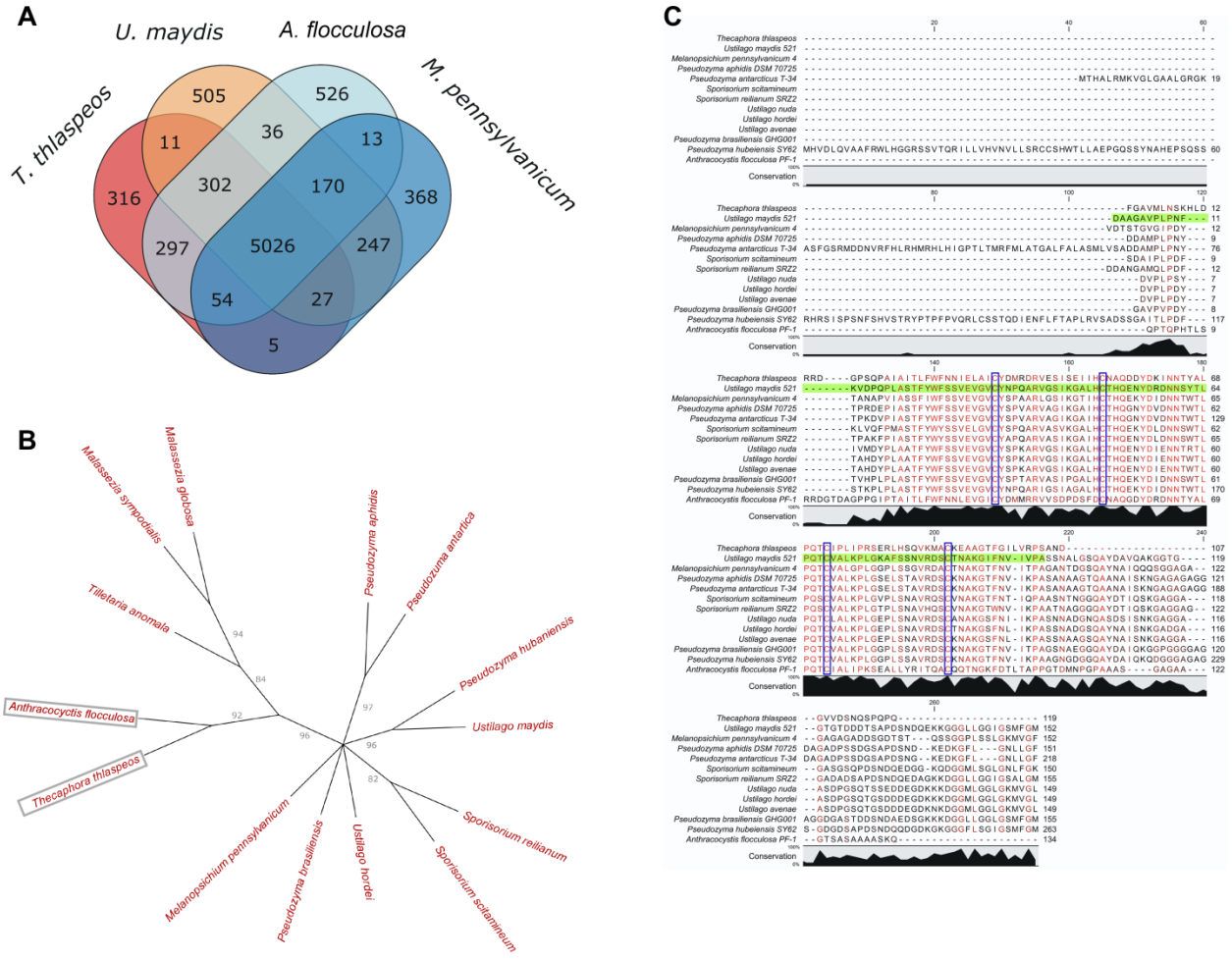


Figure 10. Phylogeny and ortholog clustering with the *T. thlaspeos* genome. Multi-locus phylogeny based on 1773 single copy orthologs shows *A. flocculosa* as the closest sequenced relative to *T. thlaspeos* (A). Ortholog clustering with OrthoFinder shows that *T. thlaspeos* shares exclusively more genes with *A. flocculosa* than the genomes of *U. maydis* or *M. pennsylvanicum* (B). Alignments of TtPep1 homologs after removing the predicted signal peptide (SignalP4, CLC was used). Cysteines and core Pep1 domain are highlighted (C).

Including the two *Malassezia* species in the analysis reduces the total amount of shared orthogroups between all sequenced *Ustilaginales* to 2205, mainly because of those species'

reduced genomes (Xu *et al*., 2007). However, when examining only the six known plant pathogens (*U. maydis, S. reilianum, U. hordei, S. scitamineum, T. thlaspeos, M. pennsylvanicum*) the shared orthogroups increase to 4827 (5092 genes). This trend persists even if a more reduced and diverse group of species is sampled, as shown in Figure 10A. In the universally shared gene families, most genes have a functional annotation (81% of the total), associated with conserved functions (primary metabolism, cell cycle, DNA repair etc.) and with the most frequent domain annotation PF00400 – WD domain, G-beta repeat. A more thorough list of the functional domain annotations shared between the six plant pathogenic smut fungi can be found in the Suppl. Table S15.

Among these shared gene families, elements necessary for the plant-dependent part of the life cycle are also detected. For example *pep1*, a secreted effector that has been characterized in previous studies demonstrating its indispensable function in plant tissue penetration (Doehlemann *et al*., 2009; Hemetsberger *et al*., 2012), has a homolog in *T. thlaspeos*. This agrees with the notion that *pep1* has a conserved function in different smut species (Hemetsberger *et al*., 2015). In addition, with the current analysis it can be supported that on one hand its conservation extends outside the clade of the grass smuts but also that there are more homologs than previously thought in the *Pseudozyma spp*. (Suppl. Figure 5). Although these species are not known to have a biotrophic stage, their homolog versions of *pep1* show conservation of the cysteines and the core Pep1 domain (Figure 10C , Hemetsberger *et al*., 2015). The effectors *pit2* (Mueller *et al*., 2013), *see1* (Redkar *et al*., 2015a) and *cmu1* (Djamei *et al*., 2011) cannot be found in *T. thlaspeos*, indicating that they might be effectors adapted to monocot hosts.

Similarly, clusters of secreted effectors from *U. maydis* (Kämper *et al*., 2006) show different levels of conservation between the species (Suppl. Table S6), with *T. thlaspeos* having homologs in the clusters 1A, 3A and 9A. These clusters however do not seem to have a detrimental role in the infection, since their deletion does not stop colonization (Kämper *et al*., 2006). The fact that some of the homologs of these genes extend to the *Malassezia spp*. (Clusters 3A and 9A, Suppl. Table S6), hint towards the idea that these clusters, although they contain secreted protein coding genes, are not necessarily related with biotrophy but probably rather with more generic processes of the life cycle. Indeed, the genes in the cluster 9A contain GMC oxidoreductase domains and the genes in cluster 3A contain peptidase domains.

In all genomes, genes with no orthologs in the other species could be found (Table 6). In plant pathogens, genes with no orthologs and functional annotation most likely represent host specific effectors (Lo Presti *et al*., 2015) or cases of horizontal gene transfer (Soanes & Richards, 2014). As expected, many of the predicted *T. thlaspeos* unique and secreted protein coding genes are upregulated during infection (pers. communication Courville K.) hinting towards their function during the biotrophic stage of the life cycle.

It has to be noted, that in genomes that have not been through extensive manual curation, such as *T. thlaspeos*, *A. flocculosa* or *M. pennsylvanicum*, genes with no homology to other species might also be false negatives of the OrthoFinder analysis, truncated genes or misannotations. Therefore, to identify the final set of *T. thlaspeos* specific effectors, thorough experimentation will be required.

Interestingly, *M. pennsylvanicum* has a very small number of exclusively shared genes with *T. thlaspeos* although they both infect dicot plants. On the contrary, *M. pennsylvanicum* shares a high number of genes with the grass infecting smut *U. maydis* (Figure 10A). At the same time *T.*

thlaspeos shares exclusively many genetic elements with *A. flocculosa*, an observation that is also verified by the multi-locus phylogeny analysis (Figure 10B).

In addition, this comparative analysis revealed secreted proteins from *U. maydis* that are unique to this species and not in known effector clusters providing targets for further studies (Suppl. Table S7).

Table 6. Percent of non-orthologs with other sequenced *Ustilaginomycotina* per species

	Predicted genes	Genes with no orthologs to other smut fungi	(%)	With PFAM annotation	Secreted
U. hordei	7111	369	5.19%	3	34
M. pennsylvanicum	6279	284	4.52%	10	18
U. maydis	6784	270	3.98%	21	28
S. reilianum	6672	96	1.44%	23	13
A. flocculosa	6877	474	6.89%	37	152
T. thlaspeos	6509	306	4.70%	52	47
Mal. globossa	4286	254	5.93%		
Mal. sympodialis	3318	19	0.57%		
P. antarctica	6640	173	2.61%		
P. aphidis	6011	195	3.24%		
P. brasiliensis	5765	17	0.29%		
P. hubeiensis	7472	1256	16.81%		
S. scitamineum	7711	794	10.30%		
Til. anomala	6808	1347	19.79%		

2.2.5 The pheromone/pheromone receptor a locus

In the genomic assembly of the *T. thlaspeos* strain LF1 (mating type a1b1) both mating loci are located in different scaffolds, reinforcing the hypothesis that *T. thlaspeos* has a tetrapolar mating system. In such systems, there is no linkage between the pheromone/receptor locus and the homeobox transcription factor locus and therefore they are segregating independently leading to a high number of mating types (Kothe, 1996).

Reinforcing the idea that mating is part of the life cycle, the components of the cAMP and MAPK signaling as described in (Feldbrügge *et al*., 2004; Brefort *et al*., 2009; Kruzel & Hull, 2010) can be found in *T. thlaspeos* (Suppl. Table S8), also indicating that most of these genes reside in the core genome of smut fungi (Suppl. Data 2).

The *a* locus of *T. thlaspeos* has a low level of synteny with *U. maydis*, showing a rearrengment in the left flank. In more detail, the *right border gene of the a locus* (*rba)* has conserved its proximity to the pheromone receptor but the *left border gene of the a locus* (*lba)* is placed

approximately 350kb downstream (Figure 11). A similar conformation is observed in *A. flocculosa* (distance between the *rba* and *lba* homologues is approximately 360 kb), suggesting that the conformation of the locus is different compared to the core grass smuts (Kellner *et al*., 2011). The second allele of the locus *a2* shows an expected homology to *a1* of *T. thlaspeos* however due to the fragmentation of the assembly further conclusions on the synteny are not possible.

Surprisingly, while *a2* allele contains the locus specific gene *rga2*, lacks *lga2*, which is shown to be specifically upregulated during dikaryon formation in *U. maydis* (Romeis *et al*., 2000). *Lga2* is not found elsewhere in the *LF2* genome and is absent in other smut species such as *Ustanciosporium gigantosporum* (Kellner *et al*., 2011), indicating that this gene might be grass smut specific. In both alleles, a novel gene is predicted (*THTG_02791*) with no homology to other smut fungi, coding for a secreted nuclease.

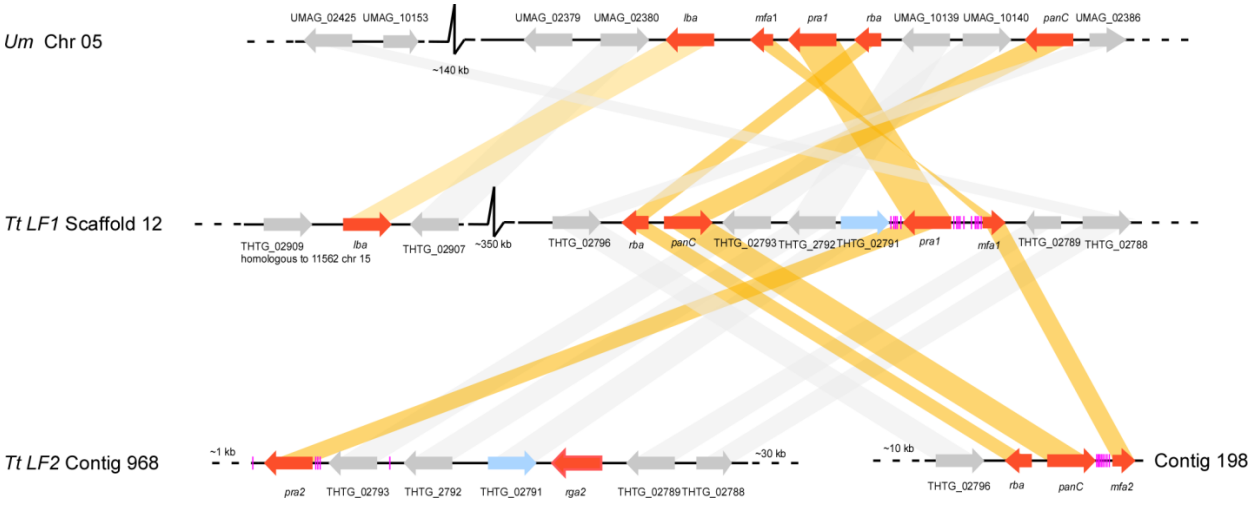


Figure 11. Inter- and intra-species synteny of a locus of *T. thlaspeos*. Arrows in red depict genes associated with the mating loci, as well as the pheromone and pheromone receptors. In blue shade are novel genes in *T. thlaspeos*. Pink lines indicate sequences homologous to the *U. maydis* pheromone response element.

In the promotor area of the pheromone and the pheromone receptor genes, pheromone response elements are found (sequence ACAAAGGGA, Figure 11) (Urban *et al*., 1996). These sequences occur also upstream of *pra1*, suggesting that as in *U. maydis*, there might have been a second pheromone in the *a1* allele that has been pseudogenized.

The overall gene model structure of *Ttpra* genes is comparable to the ones of *U. maydis*. They contain three (*Ttpra1*) and two (*Ttpra2*) introns and have a size of 388 amino acids (Suppl. Figure S7). Maximum likelihood phylogeny placed TtPra1 in the Pra1 family and TtPra2 in the Pra2 family of pheromone receptors (Figure 12A). In the spore samples collected from different locations, the pair *a1* with the *a2* allele (Figure 12C) was repeatedly recovered, indicating for a

successful infection both mating types are that most likely required. From these spore samples, cultures with single mating types can be obtained after extensive subculturing (Figure 12B).

The predicted pheromone precursors *Ttmfa1* and *Ttmfa2* were identified based on the C-terminal CaaX motif (Spellig *et al*., 1994) and their proximity to mating related genes (Figure 11, Figure S8). In these two sequences, a potential cleavage site of the pheromone is located 14 and 11 amino-acids upstream the CaaX motif of TtMfa1 and TtMfa2 respectively, leading to processing into the mature pheromone (Brown & Casselton, 2001; Szabó *et al*., 2002). Both *Ttmfa* and *Ttpra* are expressed during culture (Figure 12B), suggesting that *T*. *thlaspeos* has an active pheromone/pheromone perception module although it grows filamentously, a state which is typically associated with post-mating development in *U*. *maydis* (Feldbrügge *et al*., 2004).

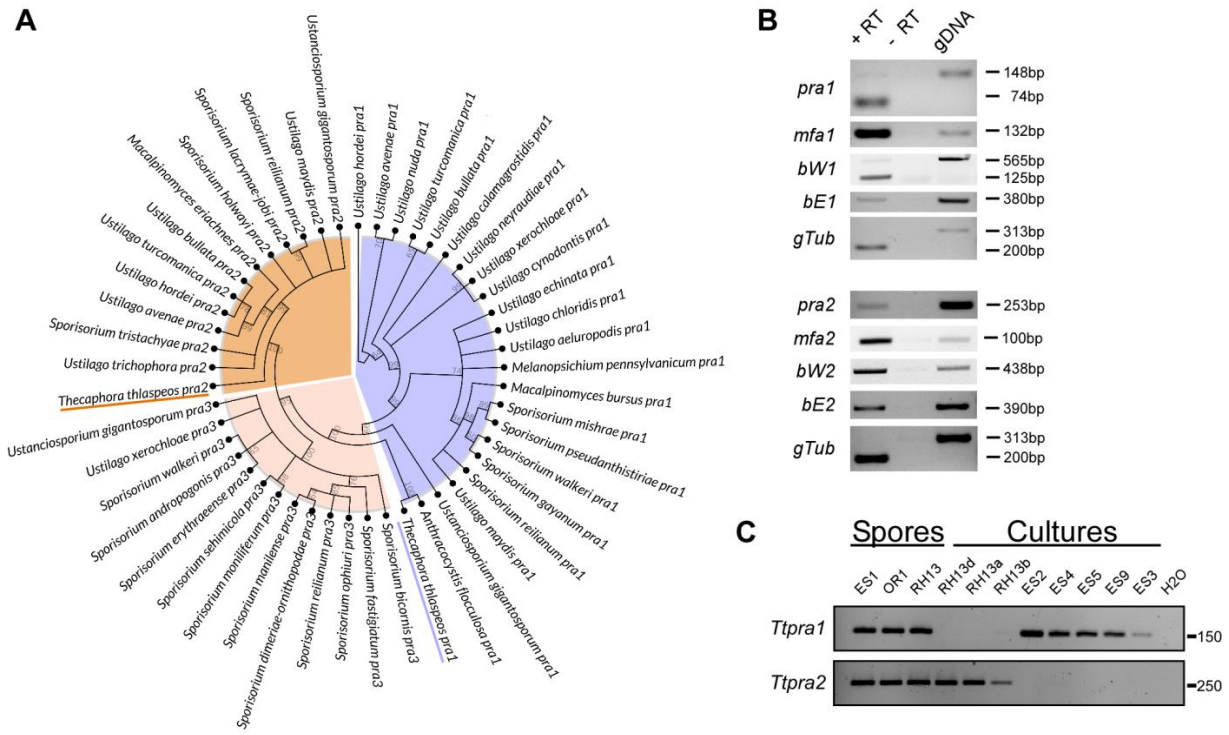


Figure 12. Phylogeny and expression of *T. thlaspeos* mating genes. (A) Maximum likelihood phylogeny of the pheromone receptor genes of *T*. *thlaspeos* Pra1 and Pra2. (B) Mating genes from both loci and both alleles are expressed during culture. (C) PCR with pra1/pra2 specific markers reveals that spores from different strains harbor both mating types. Cultures of isolated filaments derived from these spores are haploid.

2.2.6 Interactions in the *T. thlaspeos b* mating type locus

As mentioned in the introduction, the mating system of smut fungi is composed of two loci with the the *b* locus encoding two homeobox domain proteins bW, bE (Kämper *et al*., 1995). The role of the *b* locus genes in the sexual reproduction of *U. maydis* has been taken advantage experimentally, either by designing infection assays with two partners carrying mutations in background of the opposing mating types *a1b1* and *a2b2* or by using synthetic diploids like the strain SG200 (a1::mfa2 bW2bE1, Bölker *et al*., 1995). Generation of synthetic diploid strains and also of strains where filamentous growth is artificially induced (Brachmann *et al*., 2001), is based on the manipulation of the *b*-mating type genes. In order to investigate why *T. thlaspeos* haploid cells grow filamentously and to develop an infection assay similar to the one used in *U. maydis*, the *b* locus system of *T. thlaspeos* was investigated.

In the *LF1* assembly, the *TtbE (THTG_01472)* and *TtbW (THTG_01473)* were identified based on their conserved homeobox domain and the syntenic organization of the locus in Scaffold 1 to the *U. maydis b* locus. The domain with the highest conservation is the homeobox domain (82% similarity to *U. maydis*) while the rest of the protein is poorly conserved (Figure 13). None of the deposited *b* alleles of *U. maydis* in NCBI shows higher homology to *T. thlaspeos*. Other identified domains as the activation domain (AD) of bW (Kämper *et al*., 1995), or predicted nuclear localization signal (NLS) domains in bE are not conserved although it very likely that they exist in *T. thlaspeos* bW and bE (Suppl. Figure S9). Between the two alleles in *T. thlaspeos*, a variable domain and a conserved domain can be easily distinguished (Figure 12) following what has been shown for a number of other smut fungi (Schulz *et al*., 1990; Schirawski *et al*., 2005; Yan *et al*., 2016).

In contrast to *U. maydis*, *T. thlaspeos bW* and *bE* are expressed in haploid culture (Figure 12B, Quadbeck-Seeger *et al*., 2000). It was hypothesized that a functional homodimer between TtbW and TtbE could exist and result in filamentous growth. In order to investigate this hypothesis a yeast-two-hybrid (Y2H) approach, similar to the one that led to the initial discovery of the variable domains (Kämper *et al*., 1995), was used.

The Y2H results show that as in *U. maydis* and other basidiomycetes (Casselton & Olesnicky, 1998), TtbE and TtbW formed a heterodimer only when coming from different alleles and this interaction is due to their variable domain (Figure 14A, C, D). Surprisingly however, TtbE1 is able to form homodimers with itself and heterodimers with TtbE2 (Figure 14B). In addition, this interaction seems to be due to the conserved domain (Figure 14C).

In order to conclude on the contribution of TtbE in filamentous growth, further experiments are required, favorably with a non-functional mutant of *TtbE* in *T. thlaspeos*. Since a transformation protocol is not yet available, UV-mutagenesis and screening for aberrant growth phenotypes could be an alternative. It would be also interesting to test in the Y2H system *b* alleles from species with intermediate growth phenotypes, such as *T. seminis-convolvuli*, to see if the bE has a similar behavior. During this study the *b* alleles of *T. seminis-convolvuli* were also cloned, but not utilized in this assay.

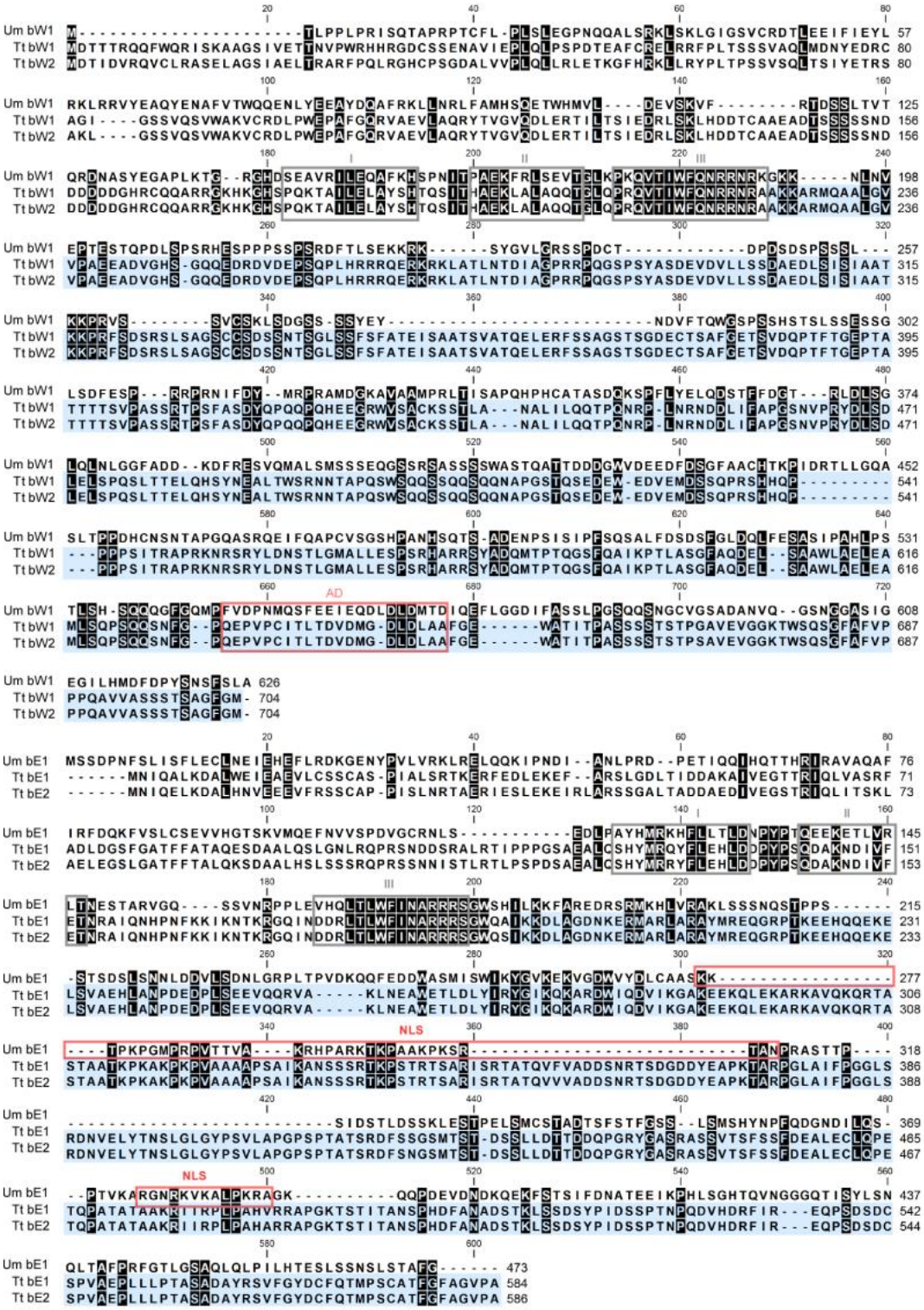


Figure 13. Alignments of the TtbE and TtbW. Amino acid alignments between the b alleles of *T*. *thlaspeos* and the homologs of *U*. *maydis* show high conservation only in the putative helices of the homeodomain (grey boxes). The activation domain of bW (Kämper *et al*., 1995) and the putative NLS domains in bE shown in red boxes are poorly conserved between the two species. Blue boxes show conserved domains of the TtbE and TtbW.

In further search for homologs of elements involved in filamentous growth and activated by the bE/bW homodimer (described in Kruzel & Hull, 2010), it was found that all the known targets are conserved in *T. thlaspeos* (Supplementary Table S8) and only the homologue of the *U. maydis hdp1* is missing. An *hdp1* homologue exists in the closely related *A. flocculosa*, suggesting that this gene might be exclusively absent in the *Thecaphora* genus. Interestingly, Δ*hdp1* strains of *U. maydis* show a non-canonical number of nuclei in their filaments (Pothiratana, 2007), indicating an involvement of *hdp1* in cell cycle arrest.

Another noteworthy finding was that the repressor of *bE* and *bW rum1* in *U. maydis* (Quadbeck-Seeger *et al*., 2000) is homologous to *THTG_03084* and was found to be expressed during filamentous growth in *T. thlaspeos* (Suppl. Table 18). This however, does not inhibit the expression of *TtbW* and *TtbE* (Suppl. Table 18).

Finally, in the promoter area of the putative *rbf1* homolog *THTG_04136*, the bW/bE homodimer target sequence ($TGAN_9TGA$, *bbs*) could be found (Supplementary Figure S11). This suggests that class 1 genes exist in *T. thlaspeos* as in *U. maydis* (Flor-parra & Vranes, 2006; Heimel *et al*., 2010). Although the Tt*bW* and Tt*bE* genes are transcriptionally active during haploid filamentous growth (Figure 12C), this class 1 target is not highly expressed, suggesting that the putative TtbE homodimers might target, if at all, other non-*bbs* sequences.

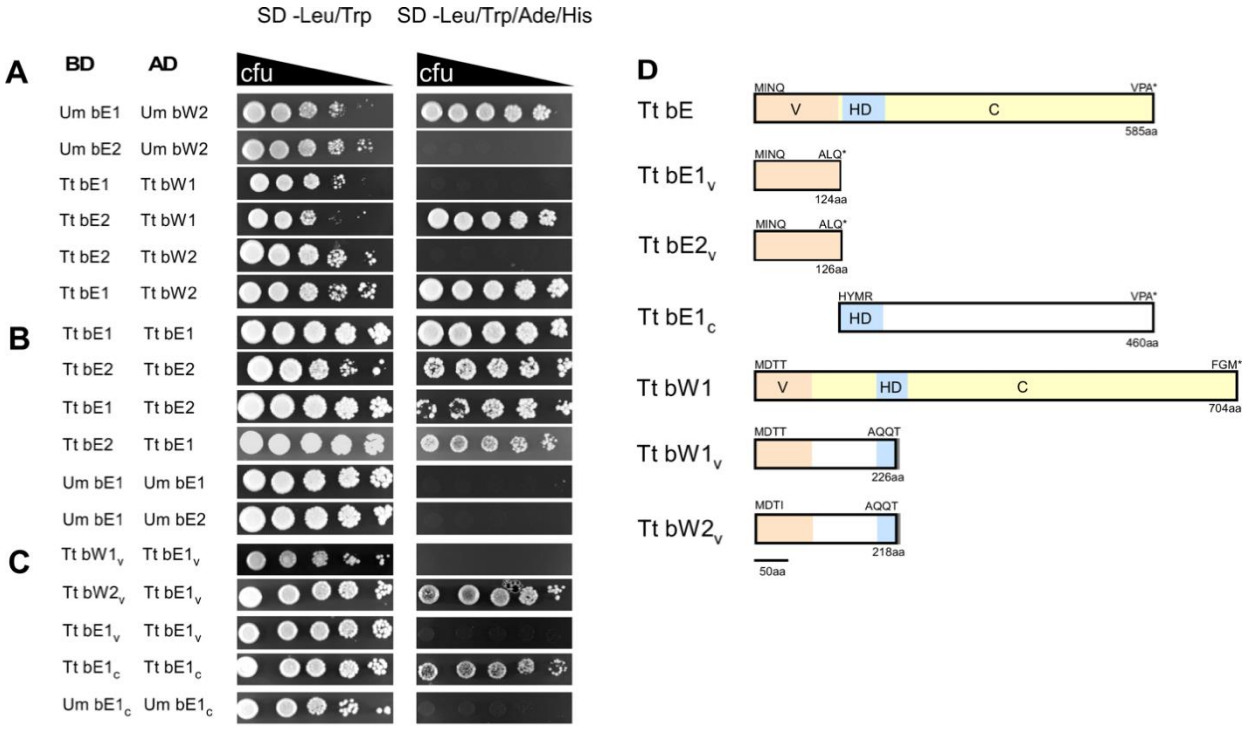


Figure 14. Yeast-two-hybrid analysis of the interactions between the bE and bW proteins. *T. thlaspeos* bE and bW present the same heterodimer formation behavior as in described in *U. maydis* (A), however bE in *T. thlaspeos* is able to form homodimers with itself as with other bE alleles (B). The heterodimer formation in *T. thlaspeos* is governed by the variable domains of bE and bW as in demonstrated in *U. maydis* (C), however for the *T. thlaspeos* bE homodimer formation the conserved part of the protein might have a role. In (D) schematic representation of the protein truncations and constructs used is shown.

2.2.7 Gain of a GH78 family enzyme – An adaptation to the dicot cell wall?

Penetration through the plant cell wall is the first step of infection in many fungal plant pathogens, and although different strategies are utilized, successful invasion is largely dependent on the utilization of carbohydrate activity enzymes that remove different layers of protection from the plant tissue (Walton, 1994). Using the hidden Markov models provided by the dbCAN database (Yin *et al*., 2012), the plant cell wall degradation capabilities of *T. thlaspeos* and 13 related smut species were investigated. Similar approaches have been used so far in larger datasets (Zhao *et al*., 2014) as well as for several fungal genomes (Gan *et al*., 2013; Williams *et al*., 2016).

As expected, the grass smuts (*U. maydis*, *U. hordei*, *S. reilianum*, *S. scitamineum*), the dicot smuts *(M. pennsylvanicum*, *T. thlaspeos*) and the yeast anamorphs (*A. flocculosa*, *P. antartica*, *P. aphidis*, *P. brasiliensis*, *P. hubeiensis*) show a high degree of similarity in the CAZy members they contain (Suppl. Table S16). On the other hand, the animal associated yeasts *Mal. globulosa*, *Mal. sympodialis* and the soil fungus *Tilletaria anomala*, have reduced numbers or lack specific CAZy genes hinting towards their adaptation to a non-plant pathogenic lifestyle.

Differences between the plant pathogens included in this study can be observed when looking at specific CAZy families. The cell wall of dicot plants is richer in pectin than the one of the grasses (Mohnen, 2008), and this is usually reflected in the genomes of pathogens which exclusively infect one of the two groups (Kubicek *et al*., 2014). For example, in contrast to the grass smuts, the *T. thlaspeos* genome contains a single member of the GH78 a-L-Rhamnosidase family. This is likely to be an adaptation to the dicot cell wall which is rich in a-L-Rhamnose (contained in the Rhamnogalacturonan I and II pectins). Additional pectin degrading enzymes, belonging to PL3, PL4 and GH53 families, specifically found in *T. thlaspeos* and *A. flocculosa* might support this function. Surprisingly, not all GH78 homologues from basidiomycetes cluster together in the constructed maximum phylogeny tree. Particularly, *T. thlaspeos* and *A. flocculosa* homologues seem to be closer related to bacteria than other smut fungi (Figure 15A). Alignment of THTG_02733 with the homologs of the closely related *A. flocculosa*, the basidiomycete plant pathogen *Rhizoctonia solani* and the bacterium *Rhizobacter sp*. show that the amino acid sequence is similar but beyond the GH78 domain, there is low conservation (Figure 15B).

Hemicellulase content and abundance in the genomes of plant pathogens, such as in the family GH16, can be attributed to the difference in the composition of the secondary cell wall between monocots and dicots (Suppl. Table S16, Scheller & Ulvskov, 2010). Members of the GH16 family are shown to have xyloglucanase activity, degrading the non-cellulosic part of the plant cell wall. The loss of members of this family in *T. thlaspeos* might be compensated by the activity of GH5 members which use the same substrate. The loss might also be explained by the fact that members of those families, such as the GH12 XEG1 from *Phytophthora sojae*, are able to induce plant immune responses (Ma *et al*., 2015) and therefore gene loss would be beneficial for the pathogen.

Finally, although *A. flocculosa* and *T. thlaspeos* are closely related, *T. thlaspeos* does not have a high number of CBM50 or CBM18 members, which is associated with parasitism on other fungi (Lefebvre *et al*., 2013; Kubicek *et al*., 2014). The higher number of those genes in *A. flocculosa* might be genus or species specific.

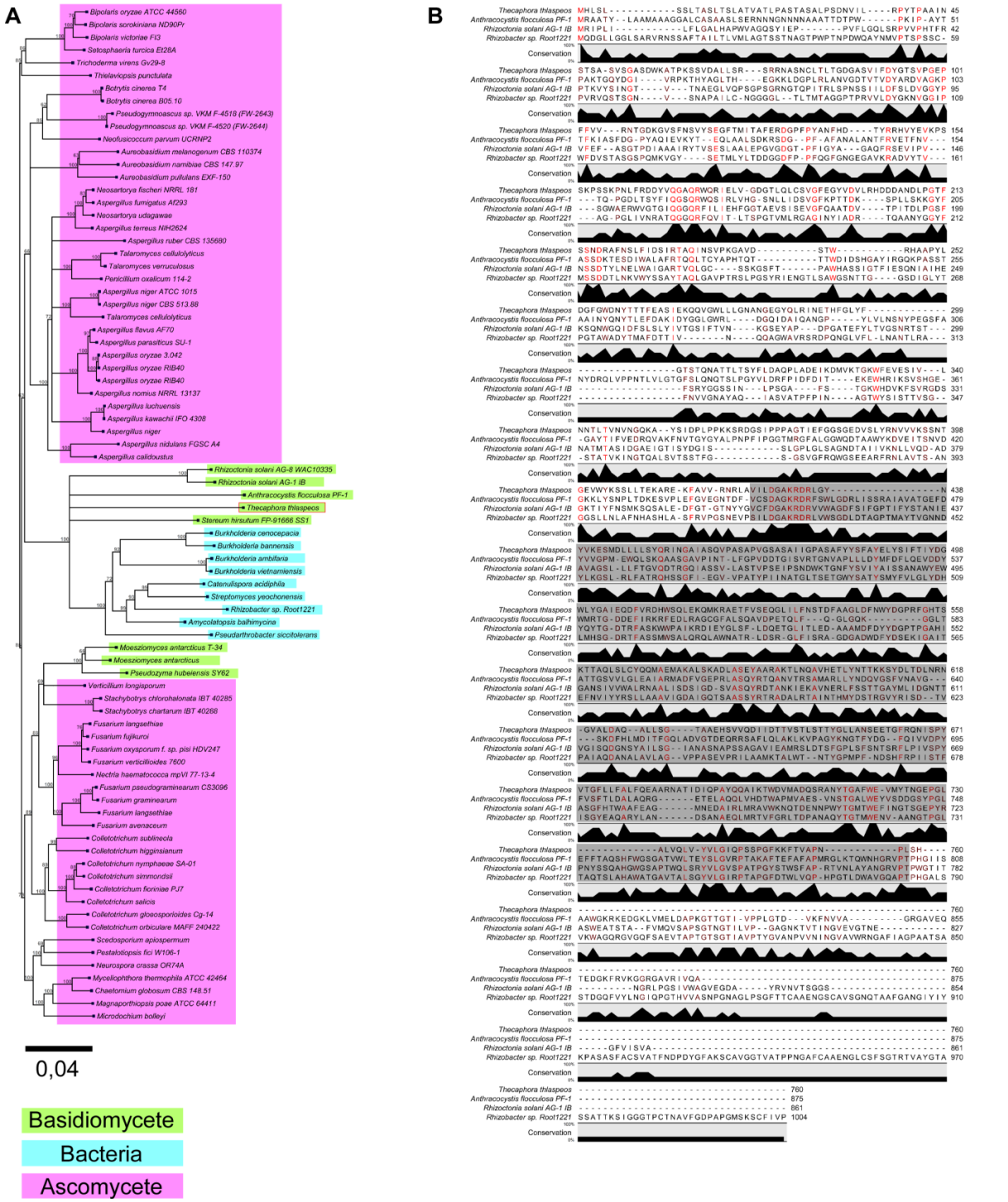


Figure 15. Phylogeny of the *T. thlaspeos* a-L-Rhamnosidase (THTG_02733). Homologues of *T.thlaspeos THTG_02733* identified by blastp to the NCBI nr database and were aligned using MUSCLE. Maximum likelihood phylogeny tree was generated using CLC Bio Workbench (default settings). Bootstrap values represent 1000 iterations and branches with support lower than 60% were collapsed (A). Alignment of THTG_02733 and homologs, including the bacterial a-L-Rhamnosidase from *Rhizobacter sp*. The GH78 domain as predicted by HMMER for THTG_02733 is highlighted in the grey box.

2.2.8 Loss of the flocculosin cluster in *T. thlaspeos*

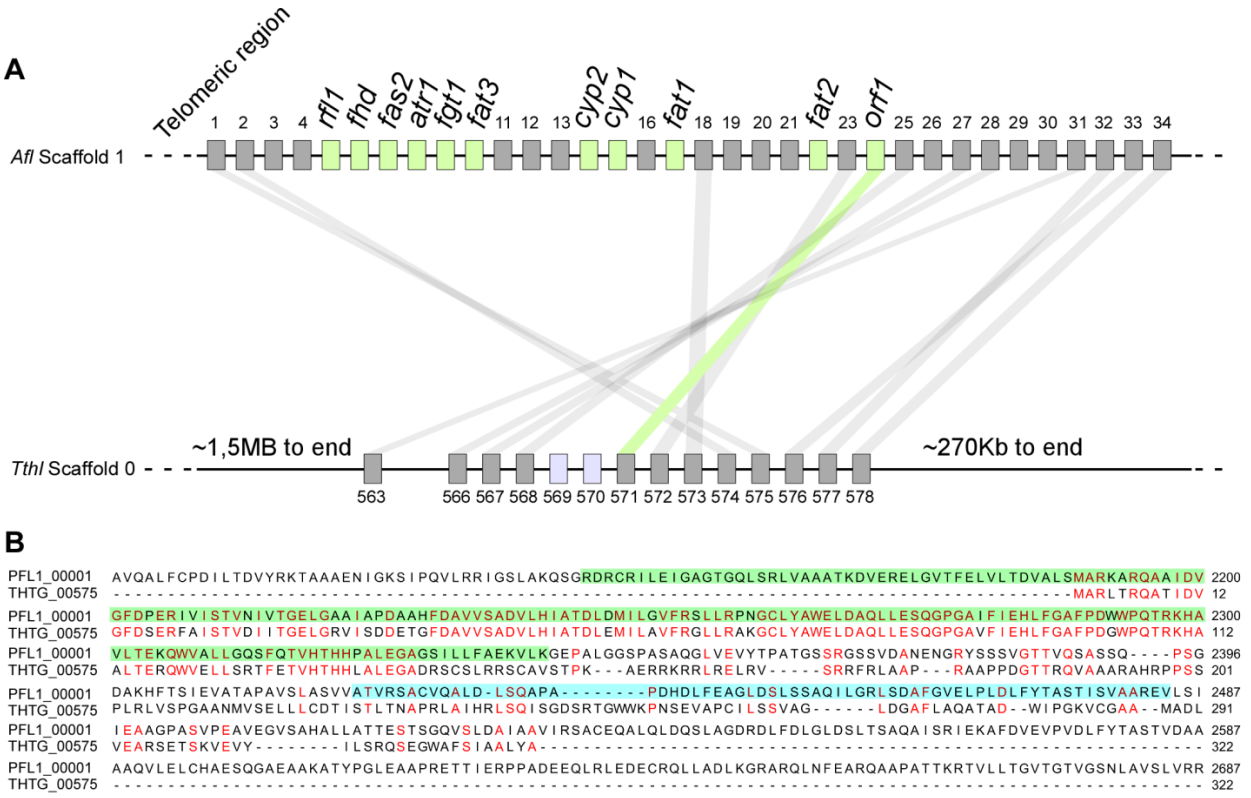


Figure 16. Loss of the flocculosin cluster in *T. thlaspeos*. Schematic representation of the flocculosin cluster in *A. flocculosa* and the homologous genomic region in *T. thlaspeos*. Green boxes represent genetic elements involved in the regulation or the production of flocculosin, grey are uncharacterized ORFs and blue are genes unique to *T. thlaspeos*. Numbering follows the assigned numbers in the annotation (i.e for *A.flocculosa* 1 represents PFL1_00001 and in *T. thlaspeos* 566 represents THTG_00566) (A). Alignment of the PFL1_00001 C-terminal domain to THTG_00575. Green box indicates predicted a SAM-dependent methyltransferase domain and blue box indicates a phosphopantetheine attachment site (B).

In addition to effector proteins, many bacterial and fungal plant pathogens utilize their secondary metabolism to successfully invade their host or, especially in the case of necrotrophs, induce plant cell death. Examples of these toxins originating from pathogenic fungi are the T-toxin from *Cochliobolus heterostrophus* (Kono & Daly, 1979), the AAL-toxin from *Alternaria alternate f. sp lycopersici* (Gilchrist, 1976; Tsuge *et al*., 2013) and Cercosporin from *Cercospora sp*. (Daub & Briggs, 1983). An example of a bacterial toxin is the coronatin from *Pseudomonas syringae* pv. *tomato* DC3000 (Brooks *et al*., 2004). These toxins can be host or non-host specific (Walton, 1996; Stergiopoulos *et al*., 2013).

In smut fungi, secondary metabolites functioning as toxins have not been identified, although modules of metabolite production have been described (Bölker *et al*., 2008), with ustilagic acid (UA) being the best characterized molecule (Lemieux *et al*., 1951; Teichmann *et al*., 2010). UA does not have a direct effect on the infection process (Hewald *et al*., 2005), however it might

have an effect on microbial antagonism on the host surface as it has been demonstrated that it inhibits infection by *Botrytis cinerea* (Teichmann *et al*., 2007). Additionally, the production of UA seems to be a conserved trait as it is also found in closely related species, including *A. flocculosa* (Kulakovskaya *et al*., 2005; Teichmann *et al*., 2011). In *A. flocculosa* the metabolite is known as flocculosin.

Interestingly, the genes necessary for UA/flocculosin production are not found in *T. thlaspeos*. In order to verify this observation, synteny of the flocculosin cluster (Scaffold 1, range 0 – 120kb) to the *T. thlaspeos* genome was examined and it could be shown that the cluster has been deleted together with many associated genes (Figure 16, A). The loss is also observed in the second sequenced strain *LF2*. An exception is the gene *orf1* which has been shown to be upregulated during nitrogen starvation (Teichmann *et al*., 2011). Moreover, downstream of *THTG_00575*, a homologous non-coding sequence to the *regulator of flocculosin 1* (*rfl1*) can be found (sequence APRHTRRGPNGRKGWCSLCTEGEWYWM, 89% identity). This indicates that the deletion of this locus is probably recent, as the sequence has not diverged significantly.

An additional observation is that in contrast to *U. maydis* and *A. flocculosa*, the homologous region of the flocculosin cluster in *T. thlaspeos* is not found in a subtelomeric region. Genomic location of the secondary metabolite clusters is crucial for their regulation (Palmer & Keller, 2010) and it could be hypothesized that loss of this regulation due to change of position might have contributed in the loss of the locus.

In the same genomic location with the flocculosin locus, the gene *pfl1_00001* has a polyketide synthase domain homologous to bacterial polyketide synthases, suggesting that proximal genes to the locus might also be involved in secondary metabolism. Notably, in *T. thlaspeos* only the C-terminal domain of PFL1_00001 has homology to THTG_00575 (Figure 15, B) while the rest of the gene is absent.

Similarly, homologous genes to the secondary metabolite clusters for mannosylerythritol lipids (MEL) and itaconic acid are absent from the *T. thlaspeos* genome. Out of the 6 genes that constitute the MEL cluster (Roelants *et al*., 2014), only the transporter *mmf1* is conserved (homologous to *THTG_03173)*. For the cluster producing the antimicrobial substance itaconic acid (Geiser *et al*., 2016) homologs to some of the genes of the cluster exist in *T. thlaspeos* but there is no synteny and the key enzyme *adi1*, an aconitate-Δ-isomerase, is absent.

To further characterize the secondary metabolite production capabilities of *T. thlaspeos*, the anti-SMASH 3.0.5 software was used. In *T. thlaspeos,* seven putative metabolic cluster locations could be detected (Suppl. Table S9). All of the genes identified are homologous to other smut fungi; however, the resulting metabolite might be different in each species (as exemplified by ustilagic acid and flocculosin). In *U. maydis* and *A. flocculosa* more putative metabolic clusters were identified (13 and 10 respectively based on proteome based predictions, Suppl. Data 3). This might not necessarily represent gene loss in *T. thlaspeos* but might have to do with the sensitivity of the software (Medema *et al*., 2011).

Overall, *T. thlaspeos* was not found to have an extensive catalog of secondary metabolite synthesis genes while it also seems that specific genes necessary for the production of known metabolites in smut fungi are not present in the genome.

2.2.9 Candidate effector proteins in the *T. thlaspeos* secretome

The predicted genes of *T. thlaspeos* were manually inspected to find elements that are potentially responsible for host colonization. These elements are widely referred with the term 'effector proteins'. In some plant pathogens, effectors can be detected by searching for conserved motifs, as the RxLR motif in *Phytophthora spp*. and the Y/F/WxC motif in *Blumeria graminis* f.sp. *hordei* (Tyler *et al*., 2006; Godfrey *et al*., 2010). In pathogens with no known conserved effector domains other criteria can be applied, as for example the detection a of secretion signal peptide, cysteine content, absence of known functional domains or even the size of the predicted protein (Saunders *et al*., 2012; Nemri *et al*., 2014; Alfaro *et al*., 2014). It is known however that such criteria generate false negatives (Sperschneider *et al*., 2015a) and debated if the bioinformatics methods commonly used can be applied to all plant pathogens universally (Sperschneider *et al*., 2015b).

Since in the smut fungi conserved effector domains have not been found and effector proteins are known to be secreted, the first step was to define the putative secretome. In this study a method similar to (De Jonge, 2012) was used, which it combines the results of two different signal peptide prediction programs (SignalP, WOLFPsort) and two transmembrane prediction programs (Phobius, TMHMM). For further consideration, the results from the less stringent SingalP prediction, filtered by the transmembrane prediction program TMHMM, are also provided (Suppl. Table S10, S11). It has to be noted that neither of the methods was able to accurately predict all the putative *U. maydis* effectors (Suppl. Table S12). On the other hand, *U. maydis* is also known to secrete proteins unconventionally (Stock *et al*., 2012), a fact that advices for caution when interpreting the prediction results.

In *T. thlaspeos* 310 genes are predicted to code for secreted proteins, a number comparable to the other smut species, which represents 4,76% of the genome (Table 7). In the future transcriptomic dataset from *T. thlaspeos* infected tissue, might reveal more genes, currently not annotated, that are secreted.

Table 7. Secretome prediction

	T. thlaspeos	***A. flocculosa***	***U. maydis***	***U. hordei***	***S. reilianum***	***M. penns.***
WOLFPsort	730	949	850	747	861	617
SingalP	441	622	625	536	629	419
Final prediction	310	488	471	380	466	287
% of the genome	4.76	7.10	6.94	5.34	6.98	4.57

In the predicted secreted proteins that have a PFAM annotation (57%), there is no overrepresented domain (Suppl. Table S13), but there are proteins with domains not shared with

other smut fungi, especially the grass smuts. Among the most interesting functional annotations is one eukaryotic-type carbonic anhydrase (PFAM00194), three necrosis-inducing proteins (PFAM05630) and a secreted catalase (PFAM00199). The above three cases are presented in the following paragraphs. Furthermore, 31 of the putative secreted proteins have a predicted nuclear localization signal (NLS) which might also pose an interesting subset to study in the future for their localization in the host cell nucleus (Suppl. Table S14) (Kloppholz *et al*., 2011; Plett *et al*., 2011).

Carbonic anhydrases (CA) are enzymes that catalyze carbon dioxide to bicarbonate in a reversible reaction. These enzymes are ubiquitous in eukaryotic organisms and in fungi are found in two classes, α and β. The latter is being overrepresented in basidiomycetes (Elleuche & Pöggeler, 2009a). β-CA have been found to be involved in CO_2 sensing, and therefore virulence, in *Cryptococcus neoformans* and *Candida glabrata* (Bahn *et al*., 2005; Cottier *et al*., 2013) but also in the sexual development of filamentous fungus *Sordaria macrospora* (Elleuche & Pöggeler, 2009b). α-CA are structurally different from β-CA but catalyze the same reaction (Lehneck & Pöggeler, 2014). To date, few things are known about the members of this class, except for the fact that they are more present in ascomycetes and some of them are secreted (Elleuche & Pöggeler, 2009a). Surprisingly, α-CAs exist in oomycete and ascomycete plant pathogens and are upregulated during infection (Raffaele *et al*., 2010; Rudd *et al*., 2015). All the smut fungi studied in this work have a single copy of a β-CA (Suppl. Data 2, Orthogroup OG001598, *THTG_00198*), but *T. thlaspeos* is the only sequenced species in the *Ustilaginales* that has an additional secreted α-CA. The *T. thlaspeos* α-CA *THTG_05674* shows reduced homology to the few α-CA that have been found so far in Basidiomycetes (closest homolog in the ectomycorrhiza *Piloderma croceum,* similarity 48,59%) but nevertheless, the important aminoacids for activity are conserved (Figure 17).

Catalases are enzymes that convert peroxide (H_2O_2) to water (H_2O) and oxygen (O_2). Although fungi maintain catalases which are involved in the detoxification of primary cellular metabolism (Scott & Eaton, 2008; Heller & Tudzynski, 2011), plant pathogens have to utilize extracellular versions of these enzymes in order to cope with basal plant defenses and to inhibit hypersensitive immune responses (HR) (Levine *et al*., 1994). For example, the plant pathogenic fungi *Claviceps purpurea, B. cinerea* and *B. graminis* f.sp. *hordei* secrete a catalase specifically during infection (Garre *et al*., 1998; Schouten *et al*., 2002; Zhang *et al*., 2004). Diverse roles for catalases in the infection process might exist, as shown with *Magnaporthe grisea* which utilizes a secreted catalase to protect its cell wall from stress (Skamnioti *et al*., 2007).

In smut fungi, secreted catalases have not been identified so far. Moreover, *U. maydis* and *S. reilianum* do not have a catalase or catalase-like enzymes homologous to the ones in *T. thlaspeos* (Figure 18B, Suppl. Data 2, 4). A catalase-peroxidase that exists in all *Ustilaginales*, including *U. maydis* (UMAG_11067), is not involved in ROS detoxification during infection (Molina & Kahmann, 2007) but rather in primary metabolism. Surprisingly, the *Z. tritici* homolog Mycgr3P67250 is secreted and upregulated 9 days post infection (Rudd *et al*., 2015).

T. thlaspeos has two catalases, the non-secreted THTG_00409 that is expressed during culture, and the secreted THTG_00174 that is not expressed during culture (Suppl. Table 18). The inferred phylogeny of these to genes with the homologs found in the NCBI non-reduntant (nr) database argues towards a case of gene loss in some of the grass smuts for both genes (Figure 18, B). THTG_00409 has homologs in the smut plant pathogens (*U. hordei, S. scitamineum, M.*

pennsylvanicum) and the anamorphic yeasts *P. hubeiensis* and *P. brasiliensis*. On the contrary, THTG_00174 has no homologs in the *Ustilaginacae* but shows a close relationship to homologs from ascomycete plant pathogens (Figure 18A, B). Not all the homologs of THTG_00174 are secreted, suggesting that this trait is not constant; however, secreted catalases are predominantly coming from plant pathogens (Figure 18, B).

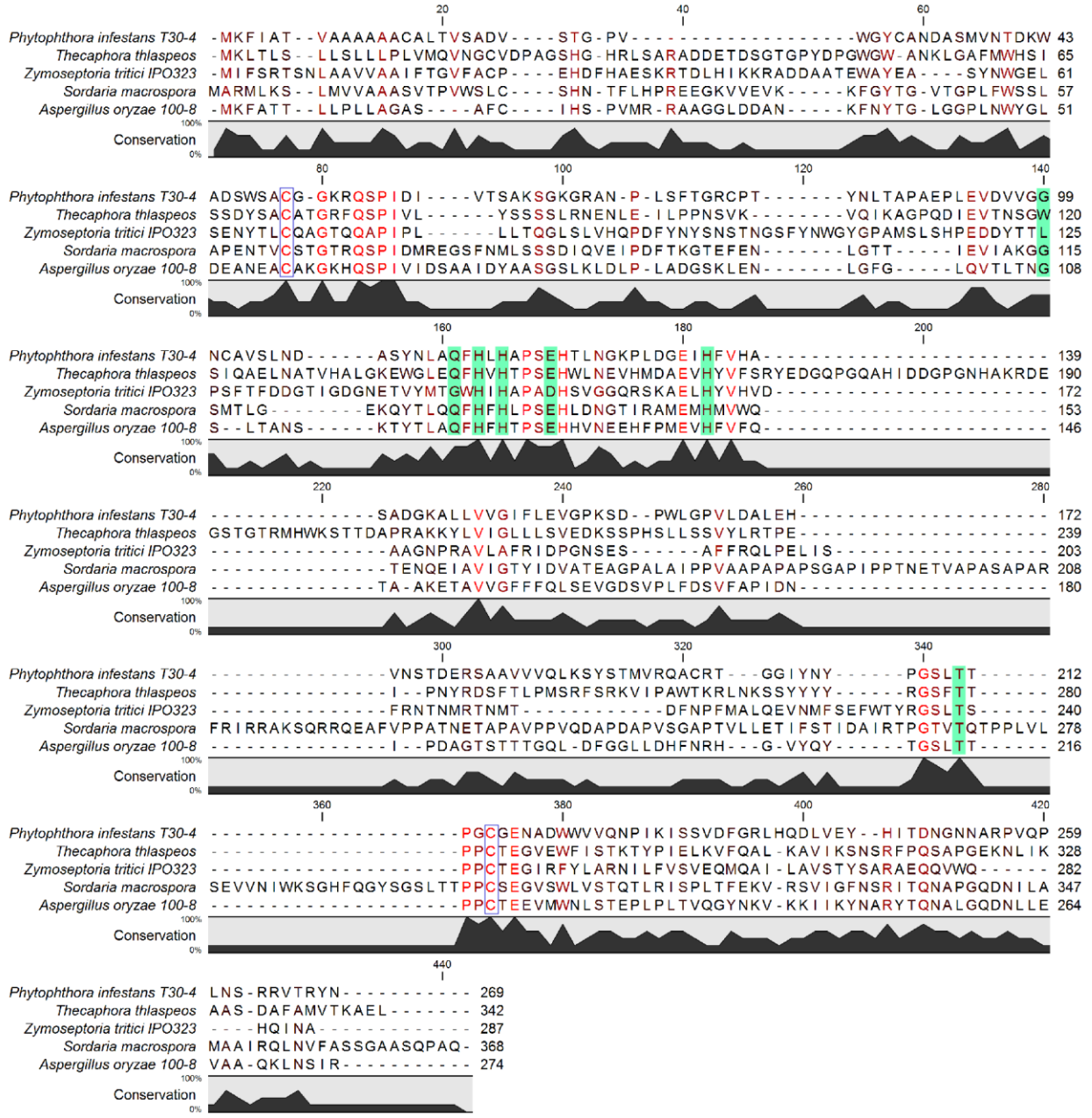


Figure 17. Novel elements in the genome of *T. thlaspeos* – α-CA. Alignment of fungal alpha type carbonic anhydrases with the *T. thlaspeos* THTG_05674. Cysteines that are predicted to form disulphide bridges are highlighted with a blue box while amino acids important for zinc binding and enzymatic activity are highlighted in green boxes.

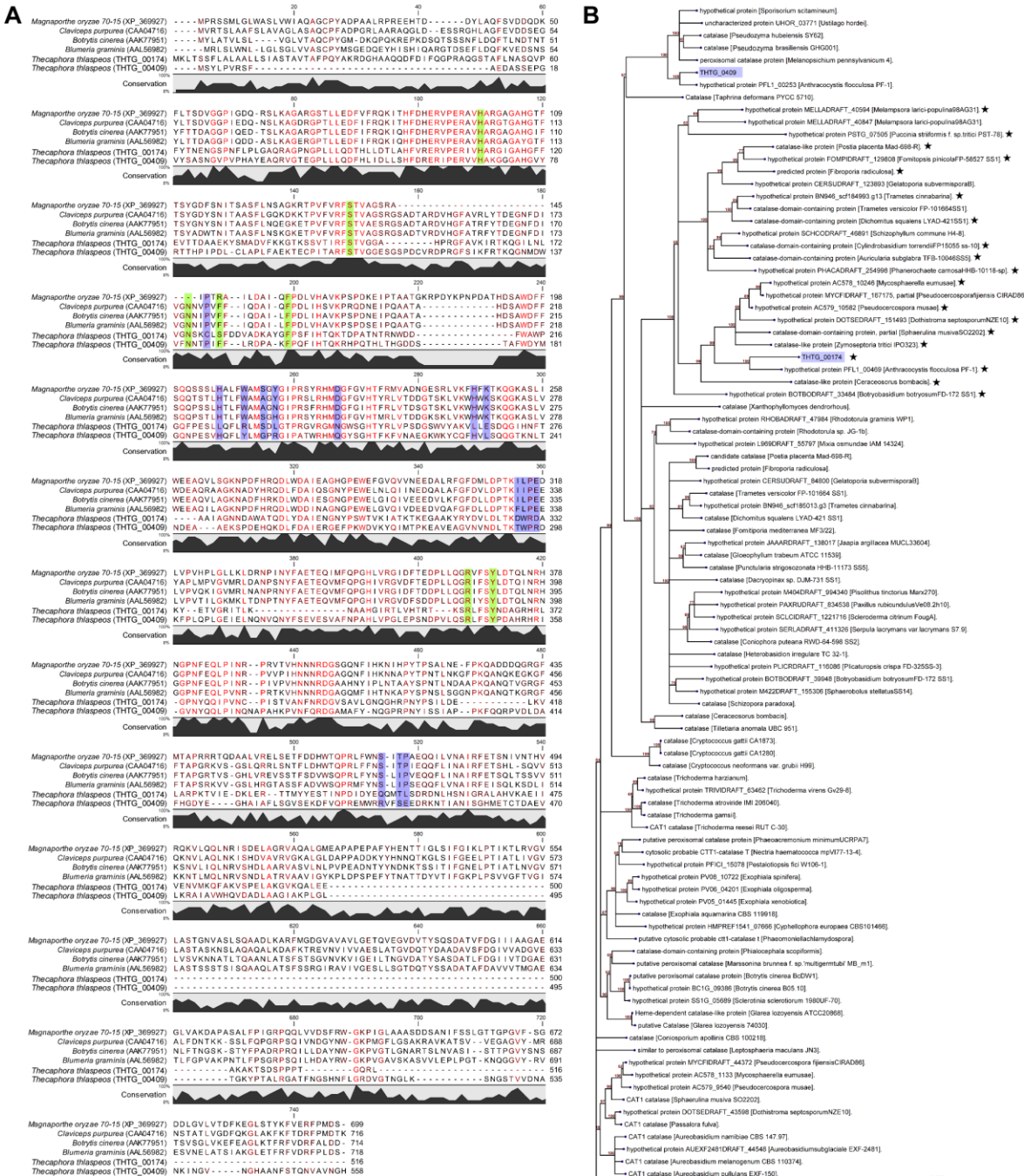


Figure 18. The secreted catalase THTG_00174 of *T. thlaspeos*. Alignment of the two *T. thlaspeos* catalases with characterized secreted homologues of other plant pathogens. In green positions important for heme-binding activity are highlighted. In blue, amino acids important for NADPH binding, as identified by the NCBI Conserved Domain Database. (A). Maximum likelihood phylogeny of THTG_00174, THTG_00409 and their homologs. Bootstrap values calculated after 1000 iterations, branches with a value lower than 60 were collapsed. Catalases with a signal peptide identified by SignalP 4.1 are marked with a black star (B).

Finally, another interesting set of genes in *T. thlaspeos* is related to HR inducing proteins having the PFAM domain 05639, known as necrosis-inducing proteins (NIP), necrosis and ethylene inducing proteins (NEP) or necrosis inducing-like proteins (NLP). Here, they will be referred as NLP proteins.

This type of effector proteins was first discovered in supernatants of *Fusarium oxysporum* cultures (Bailey, 1995; Bailey *et al*., 1997) and since then they have been extensively characterized in a number of plant pathogens, including fungi (Kleemann *et al*., 2012; Santhanam *et al*., 2012), bacteria (Mattinen *et al*., 2004) and oomycetes (Cabral *et al*., 2012; Dong *et al*., 2012). Recently, it was shown that NLP proteins are ubiquitous in microbes' genomes (Oome & Van den Ackerveken, 2014). They exist mainly in the genomes of dicot plant pathogens, but a rare case in a monocot specific pathogen has been reported (Motteram *et al*., 2009). In the wheat pathogen *Z. tritici,* the single NLP member of the genome (Motteram *et al*., 2009) is upregulated during the switch from biotrophy to necrotropy (Mycgr3G88451, Rudd *et al*., 2015) but its role in infection is not known. Interestingly, NLPs which cause necrosis on dicot plants, have no effect on grasses (Bailey, 1995; Motteram *et al*., 2009).

The necrotic effect of the proteins is related to a highly conserved heptapeptide sequence (GHRHDWE, Ottmann *et al*., 2009), but the heptapeptide sequence by itself was unable to cause necrosis in peptide infiltration experiments (Schouten *et al*., 2008). The ethylene inducing effect is due to a 20 aminoacid domain, also known as nlp20 (Böhm *et al*., 2014). Infiltration of nlp20 peptides from different plant pathogens is able to cause immunogenic responses and more recently, the receptor for the nlp20 RLP23 has been identified in *A. thaliana* (Albert *et al*., 2015). Certain NLP proteins exist in biotrophs and do not elicit HR-related cell death. So far their exact role is also not known, except that they are expressed during infection (Cabral *et al*., 2012).

In the sequenced *Ustilaginacae,* NLP genes are absent, however in the genome of *A. flocculosa* two members were found, hypothesizing that they might be utilized against other fungi (Lefebvre *et al*., 2013). In *T. thlaspeos,* three genes with the functional annotation PF05639 – Necrosis inducing protein (NPP1) were found, and all three of them had signal peptides (THTG_00343, THTG_00351, THTG_04815, Suppl. Data 4). The *T. thlaspeos* NLPs are found in two different loci, one in Scaffold0 and one in Scaffold 15. The locus containing *Ttnlp3* is homologous to the Scaffold00016 of *A. flocculosa* and the gene *PFL1_04735*. The locus containing *TtNLP2 (THTG_00343)* and *TtNLP1 (THTG_00351)*, harbors more putatively secreted genes that could be involved in infection (Figure 19A). *TtNLP1* is over-expressed during infection; however, TtNLP2 and TtNLP3 are not (Suppl. Table 18, pers. communication Courville, K.).

Amino acid positions important for cation binding and necrosis (Ottmann *et al*., 2009) are not conserved in *T. thlaspeos* members, as also positions important for ethylene induction (Böhm *et al*., 2014) (Figure 18B). Further transient expression assays in *Nicotiana benthamiana* show that indeed TtNLP1 and TtNLP2 fail to cause any symptoms of HR response, in contrast to necrotic the *Phytophthora sojae* NIP (Figure 18C, Qutob *et al*., 2002).

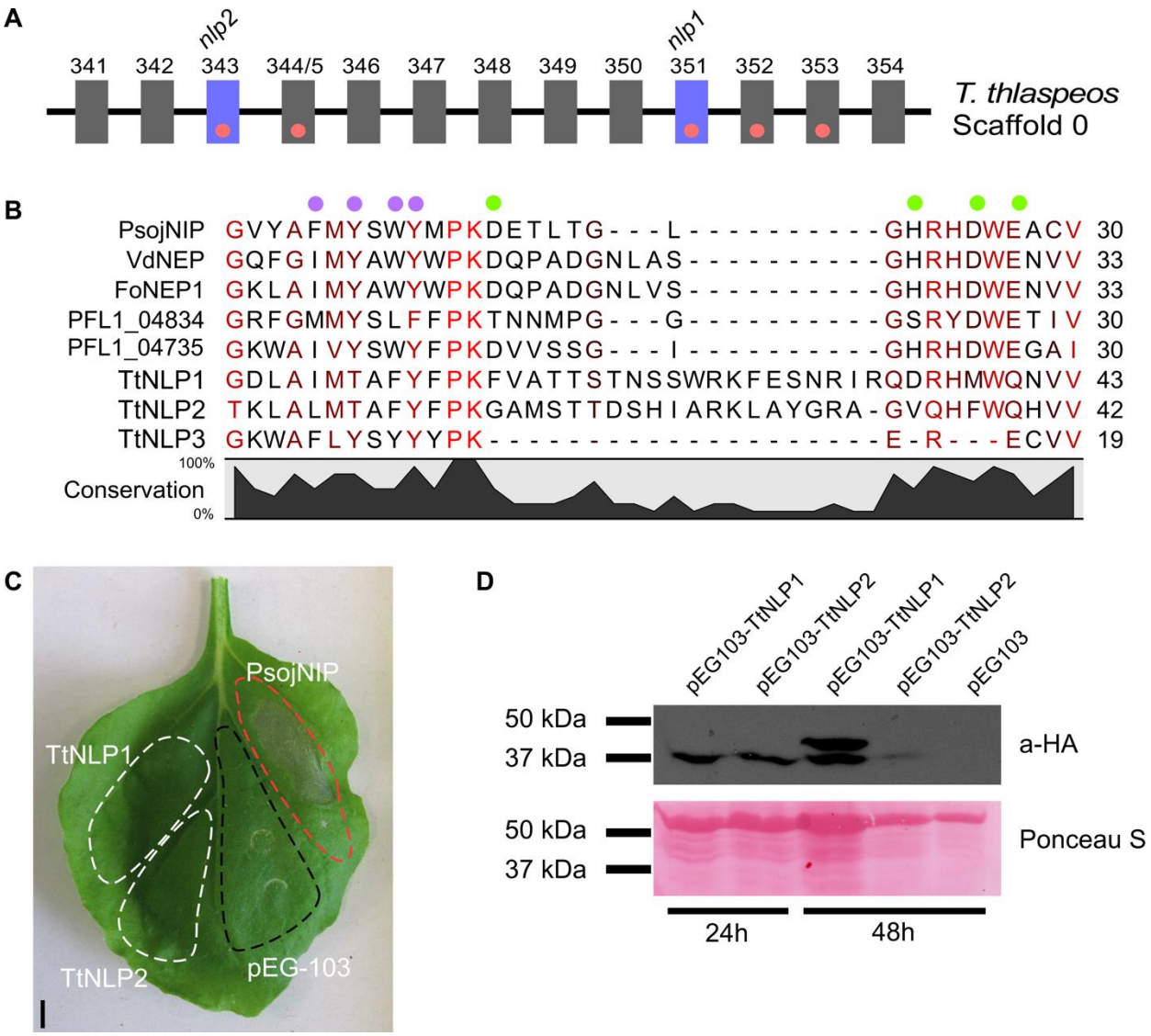


Figure 19. Necrosis inducing-like proteins in *T. thlaspeos*. A. The locus of *nlp1* and *nlp2* in Scaffold 0. Genes encoding proteins with a signal peptide are marked with a red dot (A). Alignment of the region including the ethylene inducing domain and the heptapeptide sequence of *T. thlaspeos* NLP and homologs. Amino acids indispensable for ethylene induction are marked with a purple dot and amino acids indispensable for HR response are marked with a green dot. (B). Transient expression of *Ttnlp1* and *Ttnlp2* in *N. benthamiana*, along with the positive control PsojNIP and the negative control pEG-103 (Bar = 0,5cm) (C). Agro-infiltration of *Ttnlp1* and *Ttnlp2* in the pEG-103 transient expression vector, results in heterologous protein production in *N. benthamiana*.(D)

2.2.10 The genome of the closely related species, *T. seminis-convolvuli*

Using the same tools and pipeline applied to *T. thlaspeos*, the genome of the *T. seminis-convolvuli* strain *4e5-1* was sequenced, assembled and annotated. The first draft assembly (provided by Dr. Yao Pan, Heinrich Heine University Düsseldorf, Group Algorithmic Bioinformatics) contains 754 scaffolds, a total size of 22.5 MB and 5.5% repeat content. This assembly provided genome completeness comparable to the *T. thlaspeos* and *U. maydis* (Table 5, 8). The final version of the structural annotation done with MAKER resulted in 6995 genes, including 1001 *ab-initio* predictions. After manual curation, the resulting gene models were 6741.

The motivation behind sequencing this species genome lays in the fact that *T. thlaspeos* and *T. seminis-convolvuli* (*Tsc*) are very closely related (Vánky *et al*., 2007). Therefore, comparative analysis will reveal on one hand adaptations that are likely to be conserved in the *Thecaphora* genus, emphasizing those elements that are necessary for infection of dicot plants, and on the other separate genes or events that are specific only to *T. thlaspeos*.

Initial analysis of the mating type loci shows that the strain is haploid and carries a single pheromone/pheromone receptor mating allele that is homologous to *T. thlaspeos a1* (Figure 20A). A single allele of the *b* mating locus also exists with TscbE and TscbW having high amino acid sequence identity with TtbE1 and TtbW1 (70% and 69% respectively).

Overall, the genomes of the two *Thecaphora* species are syntenic between each other and with *A. flocculosa* but to the same extend with *U. maydis* (Figure 20B, Suppl. Data 10). This further supports that *A. flocculosa* is a species closely related to *Thecaphora*. In order to get a better insight on the synteny, improvement of the *Tsc* assembly with the PacBio sequencing would be desirable.

Finally, the so-far comparative analysis of the two species has brought forward two important findings. One, that the NLPs are likely to be conserved in the *Thecaphora* clade, since the *T. thlaspeos* NLPs have homologs in *T. seminis-convolvuli* (Figure 20C). And second, that the loss of the UA/flocculosin cluster is probably a specific event for *T. thlaspeos,* since homologous genes could be found in *T. seminis-convolvuli* (Figure 20D).

In the future, it would be interesting to investigate if NLP homologs exist in other *Thecaphora* species such as the *T. saponariae* and *T. oxalidis*, but also to further explore the secondary metabolism of these plant pathogens.

Table 8. Completeness analysis with BUSCO for the *T. seminis-convolvuli* genome

	#	%
Complete Single-Copy BUSCOs	1284	89
Complete Duplicated BUSCOs	68	4.7
Fragmented BUSCOs	77	5.3
Missing BUSCOs	77	5.3
Total BUSCO groups searched	1438	

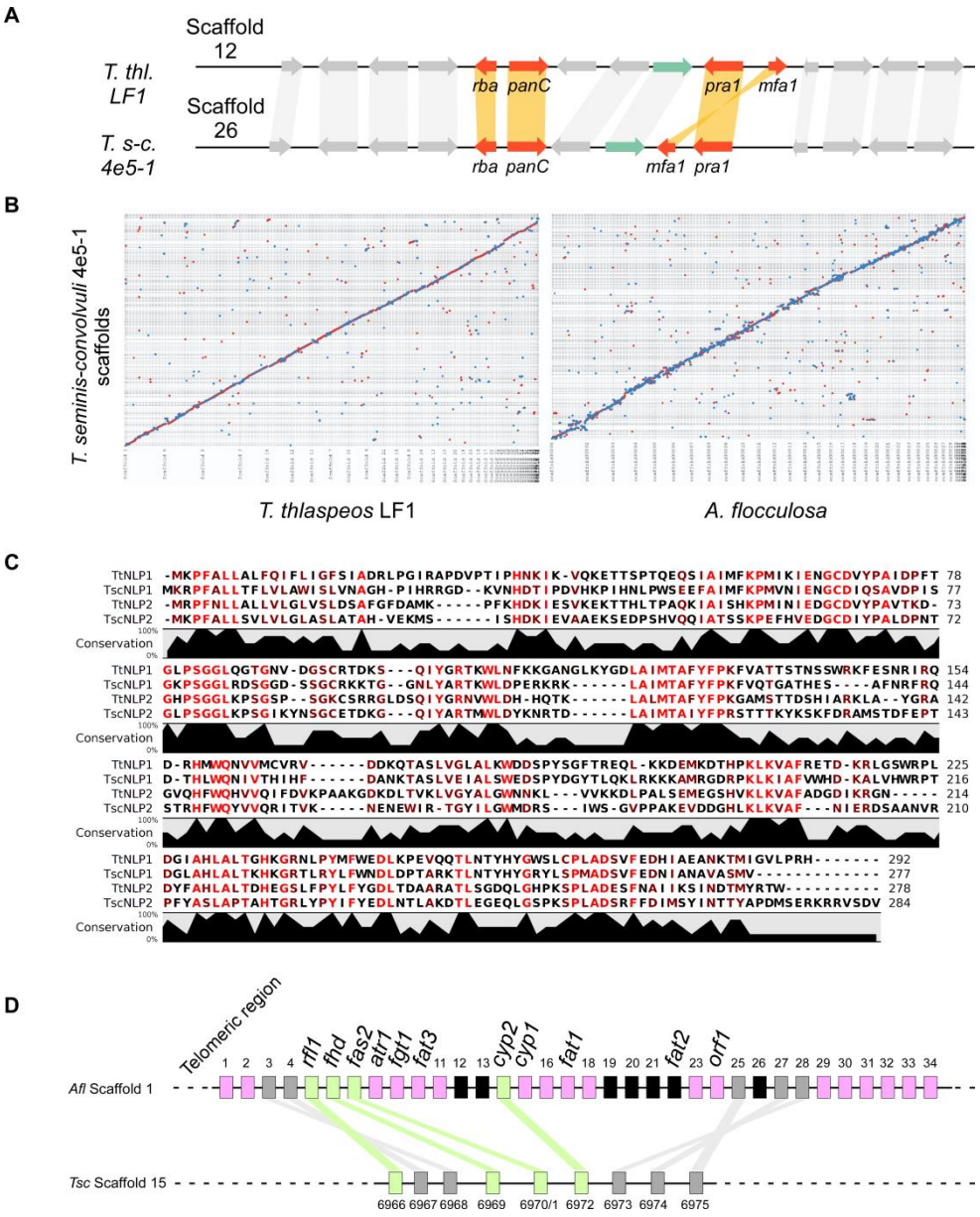


Figure 20. Synopsis of the *T. seminis-convolvuli* genome analysis. The sequenced strain carries a single pheromone/pheromone receptor locus which is syntenic and homologous to the a1 locus of *T. thlaspeos* LF1 (A). The genome is overall very syntenic to the closely related genomes of *T. thlaspeos* and *A. flocculosa* (B). Homologs of the two *T. thlaspeos* NLPs exists also in the *T. seminis-convolvuli* genome (C). A homologous locus to the flocculosin cluster exist in *T. seminis-convolvuli* scaffold 15 (D). In green boxes are genes that are necessary for flocculosin production and have homologs in scaffold 15. Grey boxes are genes residing in the locus but not involved in flocculosin production. Pink boxes are genes with putative homologs elsewhere in the *T. seminis-convolvuli* genome. Black boxes are genes with no homologs found in the genome.

3 DISCUSSION

3.1 Summary

The aim of this dissertation was to describe the life cycle and the genome of the fungal plant pathogen *T*. *thlaspeos*. Although more questions have come forward than the ones that have been addressed here, the fundamental materials that will help research with *T*. *thlaspeos* are now available.

During this work, a protocol to obtain cultures of *T*. *thlaspeos* from spores was developed and the culture of two strains with opposing mating types was possible. Considering the fact that *T*. *thlaspeos* is able to colonize the model plant *A*. *thaliana,* controlled experiments using these two strains can now be designed in order to investigate if and how *T*. *thlaspeos* is able to complete its life cycle in this model host and if not, to answer why.

An additional resource presented here, the genome of *T*. *thlaspeos*, will complement every effort during this process, from generating expression vectors and markers for infection to deciphering the mechanisms of filamentous growth in this species. Moreover, the annotation and the analysis of this genomic data set, which is a continuous process, will be always enhanced by new data and discoveries in other organisms. Most importantly at this current time point, a novel set of *T*. *thlaspeos*-specific genes demands further investigation.

3.2 An overview and open questions in the life cycle of *T*. *thlaspeos*

In all *T*. *thlaspeos* infected plants found in the field, regardless the location or the host, teliospore formation on the siliques was the only observed symptom (Figure 2). Therefore, it would safe to suggest that the main way of dispersal of this pathogen is the wind. In addition, since it was observed that in many cases the host seeds were still in the seed pod and their development was not affected by the teliospores (Figure 3 and S1), it can also be assumed that an additional way of dispersal is by infecting the immediate progeny of the host. This is also observed in small grain smut fungi, where infectious hyphae grows in the developing seed (Kronstad, 1997). This close association between host and pathogen is remarkable, as it is known that plants confronted with pathogens are able to use their "epigenetic memory" to counter biotic stress in the following generations (Molinier *et al*., 2006; Slaughter *et al*., 2012). Thus, the first question in the life cycle of *T*. *thlaspeos* is how and if it is able to infect subsequent generations of the same host.

Another interesting aspect of this smut pathogen is that its teliospores germinate in the presence of a host specific signal. The signal can originate either from a germinating seed or from leaf tissues (Figure 3D and S1). This behavior is unique in the smut fungi, since this association has never been observed before. Moreover, it seems that it is a specific feature of *T*. *thlaspeos* since the teliospores of the close relatives *T*. *seminis-convolvuli, T*. *saponariae* and *T*. *oxalidis* can germinate even in water-agar. So far it is known that signals originating from plants are involved, to some extent, in the life cycle of *U*. *maydis* and of the more distant anther smut *Microbotryum violaceae* (Kokontis & Ruddat, 1986; Mendoza-Mendoza *et al*., 2009), but these do not regulate spore germination. Therefore, it would be interesting to investigate what is the nature of this plant signal and how *T*. *thlaspeos* is able to perceive it. It can be envisioned that in

the future a transcriptome survey of the *T. thlaspeos* teliospore germination and further comparative analysis with the *T. seminis-convolvuli* genome would be very helpful in identifying candidates for study.

Still, the teliospore germination, presents another noteworthy feature. While many of the smut fungi germinate by producing haploid sporidia, *T. thlaspeos* produces haploid filaments. In this study, it was attempted to shed some light on this process. On one hand, bioinformatic methods were used to investigate the genome and describe the mating system and on the other, experimental interaction assays between the two core components of the *b* mating locus bW and bE were used to show if and in what way they interact in *T. thlaspeos*. Surprisingly, in addition to the widespread observation that bE and bW form heterodimers when coming from different alleles (Figure 2 and Figure 13), it was observed that TtbE is able to form homodimers with itself. However, it was not possible to verify the expression of the protein constructs residing in the pGAD vector via Western blot analysis (Suppl. Figure S10), despite following the suggested (MatchMaker manual, Clonetech) and optimized protocols (pers. communication Pohlmann, T.). Nevertheless, the control interactions presented in the Suppl. Figure S9, argue that the TtbE-TtbE homodimer is likely not an artifact. Follow up experiments designed to knockout the bE gene in *T. thlaspeos*, either by homologous recombination or by UV mutagenesis, would support this hypothesis and address whether TtbE drives filamentous growth in haploid cells.

An alternative hypothesis is that environmental factors are responsible for the filamentous growth in *T. thlaspeos*; meaning that under favorable conditions, *T. thlaspeos* is able to produce sporidia. Filamentous growth independent of an active *b* heterodimer has been observed in *U. maydis*. Nitrogen availability in *U. maydis* is a factor that can lead to hyphal growth of sporidial cells (Lovely *et al*., 2011) and even more interestingly, *T. seminis-convolvuli* switches between conidial production and filament elongation during culture, probably based on the availability of nutrients (Movie 3, Suppl. Figure 6). It would be also interesting to monitor the expression of *TscbE* during the switch from conidia to filament elongation, to see if there is a correlation.

A third hypothesis can be that *T. thlaspeos*, although closely related to *U. maydis*, has lost the ability to grow in a sporidial form. An example for this argument is the filamentous *Ashbya gossypii* when compared to the yeast *Saccharomyces cerevisiae*. These two species have genomes with high orthology (95% of the genes) and synteny (90%) (Dietrich *et al*., 2004), yet orthologous proteins have different roles during growth, leading to a different growth morphology (Philippsen *et al*., 2005). Targeted analysis of the genome and further description of the genes involved in filamentous growth in *U. maydis* might point out such differences.

Further in the life cycle, preliminary confrontation assays between the strains *T. thlaspeos* LF1 and LF2 that show filament fusion (Movie 6), and the finding that in the teliospores two mating types are detected (Figure 12, C), suggests that *T. thlaspeos* has an active mating behavior. Considering the role of mating in other smut fungi, it is likely that mating is also necessary for successful infection. Post-penetration of the host tissue, whether this is a leaf or the root, *T. thlaspeos* grows asymptomatically in the proximity of the vascular bundle. Systemic growth seems to be a conserved behavior in the *Thecaphora* species (Figure 6 and Suppl. Figure S6A) but also in the smut fungi in general (Begerow *et al*., 2014), as it is exemplified in *S. reilianum* (Ghareeb *et al*., 2011). This is in contrast to *U. maydis* where infection is local (Snetselaar & Mims, 1994) and leads to an extensive reprogramming of the host tissue (Skibbe *et al*., 2010; Redkar *et al*., 2015a). Transcriptome sequencing of lab-infected plants in the future will reveal

which host pathways have to be suppressed by *T. thlaspeos* in order to have this symptomless, endophytic growth. This is especially interesting when considering that *T. thlaspeos* infects and most likely persists for more than one season in perennial hosts (*Ar. ciliata, Ar. hirsuta*).

Finally, as the last part in the life cycle, teliospore formation is observed only during the development of the host seeds. In the future it would be very interesting to investigate if this is again a response to a host related signal. Giving that it is very difficult to suggest what this signal could be, proanthocyanidins that accumulate in early seed development (Nesi *et al*., 2001) and trehalose which regulates flowering time and embryo maturation (Eastmond *et al*., 2002; Wahl *et al*., 2013) could be initial targets.

3.3 A first comparative analysis of the *T. thlaspeos* genome reveals novel genes

The genome of *T. thlaspeos* is very similar structurally and content wise with the genomes of the *Ustilaginales* (Table 4). Yet, genes with functions not found in other smut fungi and a group of secreted genes with no known functions is most likely, what makes this species able to specialize in the infection of Brassicaceae hosts.

A first important outcome out this study is the evidence that the adaptation of smut fungi in dicot hosts has happened with at least two different ways. Although this is an overall bold claim, the genomes of *T. thlaspeos* and *M. pennsylvanicum* greatly support this argument. Besides the phylogenetic evidence that already exist (Begerow *et al*., 2000; Wang *et al*., 2015) and are supplemented in this study (Figure 10), the lack of exclusively homologous genes bearing typical effector characteristics between these two species shows that the invasion of the dicot host plant is realized in a different way. On one hand, *M. pennsylvanicum* likely utilizes many effectors that are homologous to the ones of the grass smuts and on the other *T. thlaspeos* has elements that are new to smut fungi and have the typical characteristics of effectors (Paragraph 2.2.4, 2.2.9).

For example, the effector *see1*, with a known function in tumor formation during infection of maize by *U. maydis*, is conserved in the *Ustilaginales* clade including *M. pennsylvanicum* (Redkar *et al*., 2015b), but is absent from *T. thlaspeos*. The effector *see1* is targeting the maize homolog of the *S. cereviseae* SGT1 (Kitagawa *et al*., 1999; Redkar *et al*., 2015a) which is involved in the initiation of cell death upon confrontation with plant pathogens (Peart *et al*., 2002; Wang *et al*., 2010). Not surprisingly, SGT1 homologs are also known to be targeted directly (El Oirdi & Bouarab, 2007; Liu *et al*., 2016) or indirectly (Cui *et al*., 2010) by other plant pathogens. A similar interaction has been even shown between the *Salmonella enterica* effector SspH2 and the human SGT1 homolog (Bhavsar *et al*., 2013). Considering that these common effectors between *M. pennsylvanicum* and grass smuts target conserved host defense hubs in dicots and monocots, it can only be hypothesized that in *T. thlaspeos* another set of non-homologous effectors has developed to target the same processes. Therefore, it will be extremely interesting to see in what way *T. thlaspeos* targets the SGT1b of *A. thaliana*.

Genes that are exclusively found in the *T. thlaspeos* genome, when compared to other smut fungi, contribute to this hypothesis. For example, SA-dependent responses are important for plant defense and thus effectors that (down)regulate the SA production have been found in

bacterial and oomycete plant pathogens (DebRoy *et al*., 2004; Caillaud *et al*., 2013). In smut fungi, this is utilized by the conserved effector *cmu1* (Djamei *et al*., 2011, Suppl. Table S6), which is not found in *T. thlaspeos* genome. *Cmu1* codes for a secreted chorismate mutase that interferes with salicylic acid (SA) production upon infection. An enzyme that could complement the absence of a *cmu1* homolog in *T. thlaspeos* might be the secreted α-carbonic anhydrase (THTG_05674), since β and α-carbonic anhydrases have been shown to bind to SA (Slaymaker *et al*., 2002; Bayram *et al*., 2008).

In a similar manner, *U. maydis* and *M. pennsylvanicum* are likely to have different enzymes that respond to basal plant defense than *T. thlaspeos*. This is illustrated with the catalases of *T. thlaspeos* which have reduced or no homology to the *Ustilaginales*. Further sequencing of the *Thecaphora* spp. which were briefly presented here (Paragraph 2.1.4) would give an important insight on the origin of these genes and their conservation.

Another important point is that this study brings forward is the close relation of *T. thlaspeos* with *A. flocculosa*. This is very surprising considering that an extensive phylogenetic study (Wang *et al*., 2015) based on seven different loci clearly clusters this species in a sister clade to the grass pathogen genus *Sporisorium,* while the multi-locus phylogeny here (Figure 9) shows that this species more closely related to the dicot smut fungi of *Thecaphora*. This result is supplemented by the gene orthology analysis (Figure 9), where it is shown that *A. flocculosa* shares exclusively genes, some of which code for putative secreted proteins or known effectors as the NLPs, with *T. thlaspeos*. Knowing that hybridizations between pathogenic fungi (Newcombe *et al*., 2000; Kellner *et al*., 2011; Menardo *et al*., 2015) are possible, it would be interesting to investigate if *A. flocculosa* is a natural hybrid of a grass and a dicot smut. Specifically, since the closest sequenced homolog of TtPra1 is coming from *A. flocculosa*, a confrontation assay between these two species and description of the interaction would reinforce, this hypothesis.

Considering the above, more questions are raised about the genomes and the lifestyle of the species in the "catch-all" genus *Pseudozyma*. The evidence here support the argument that these species might not just be mycoparasites (Neveu *et al*., 2007; Gafni *et al*., 2015) or saprotrophs but anamorphs of plant pathogens (Stoll *et al*., 2005; Wang *et al*., 2015). The most striking finding is that *P. hubeiensis* has 15 homologs to the 25 genes contained in the effector cluster 19 (Suppl. Table S6), nearly as many as the closely related grass pathogen *S. reilianum* (17 out of 25). Another piece of evidence is the repertoire of CAZyme families. Abundance or loss of CAZyme families is an indication of lifestyle as it can be seen in the Suppl. Table S16. It has also been suggested that CAZymes can be considered as factors that determine a host/non-host interaction (Walton, 1994). In that sense, the GH88 and GH39 CAZy families that are missing in *T. thlaspeos* but exist in *Pseudozyma spp*., might serve as an indirect indication that in these *Pseudozyma spp*. there is a specific adaptation to plant cell walls, since these families can be used for degradation of xylans and pectins.

Finally, the genome of *T. thlaspeos* presents an interesting case of species specific loss of a secondary metabolite cluster (Figure 16, 20D). *T. thlaspeos* does not have the biosurfactant producing cluster of flocculosin, which is widely conserved between smut species. A first hypothesis could be that this metabolite might be toxic or induce HR responses in the Brassicaceae and therefore its loss is justified. Alternatively, it could be hypothesized that *T. thlaspeos* does benefit from an intact microbiome in the host tissue, or it uses other ways to manipulate it and outcompete other pathogens. In the latter point, it would be interesting to

entertain the involvement of TtNLPs, since it was recently shown that a peptide sequence originating from an *Ph*. *parasitica* NLP (Ppnlp20, sequence AIMYSWYFPKDSPVTGLGHR) could prime the plant's immune system against pathogen infections (Böhm *et al*., 2014). Peptide infiltrations with a homologous Ttnlp20 peptide and subsequent infections with the bacterial plant pathogen *Ps*. *syringae* could show if this is also true for *T*. *thlaspeos*.

Nevertheless, further exploration of the metabolic capabilities of *T*. *thlaspeos* could be aided by using histone deacetylase or DNA methyltransferase inhibitors, which have been shown to promote the expression of silent biosynthetic pathways in *U*. *maydis* and *Aspergillus niger* (Fisch *et al*., 2009; Yang *et al*., 2013).

All these findings in the *T*. *thlaspeos* genome constitute a small part of the information contained in this dataset. The extended biotrophic phase and the systemic infection of *T*. *thlaspeos* present a great opportunity to explore the time and tissue specific expression of the genes mentioned here.

3.4 Conclusion and future prospects

The effort put in this thesis has resulted in few but significant accomplishments. *T*. *thlaspeos,* a fungal organism for which we previously had very limited knowledge, now has a foundation of methods and resources required to be a useful tool for the exploration of the plant-microbe interactions. It is able to grow as haploid cells in artificial media both in liquid and on plates, different mating types are available and based on microscopic observations, a draft life cycle can be suggested (Figure 20). In addition, sequence information from the genome of two different strains, gives the ability to molecularly investigate the interaction of this fungus with its hosts and the environment. Nevertheless, in order to make the transition to a model organism, additional steps have to be made.

First, methods for reproducible infection and genetic transformation have to be developed. Both goals can be achieved by focusing on the concept of an artificial diploid as it is used in *U*. *maydis* (Bölker *et al*., 1995). Although transformation protocols have not yet been developed for any of the *Thecaphora* species, it is very encouraging that the close relative *A*. *flocculosa* is amendable to transformation by using plasmids with *U*. *maydis* derived sequences (Neveu *et al*., 2007).

Second, further support of the assembly with optical mapping and consistent curation of annotated genes would be very advantageous (Thomma *et al*., 2015; Faino *et al*., 2015). Both of these two points are reflected in the genome of *U*. *maydis*. The collaborative effort of the smut fungi community has achieved a high standard reference genome for *U*. *maydis*, since both the annotations and the available sequences are very reliable. In order to achieve this for *T*. *thlaspeos*, the use of already developed tools and resources such as the databases Ensebl Fungi and PhytopathDB should be considered (Kersey *et al*., 2015; Pedro *et al*., 2015).

Finally, additional sequence information from *T*. *thlaspeos* strains obtained from different hosts and locations would reveal the population structure and diversity of *T*. *thlaspeos*. At the same time, sequence information of other *Thecaphora* species will help in evolutionary studies of effectors and other elements necessary for these species' survival on their hosts. It is more

exciting when considering that the genomes of smut fungi, with their low repeat content (average <5%, Table 4) and small genome size (±20 MB) are ideal for such studies.

I hope that in the future, we will be able to see a large database of smut genomes, harvesting as much as 1002 different species, readily available for analysis.

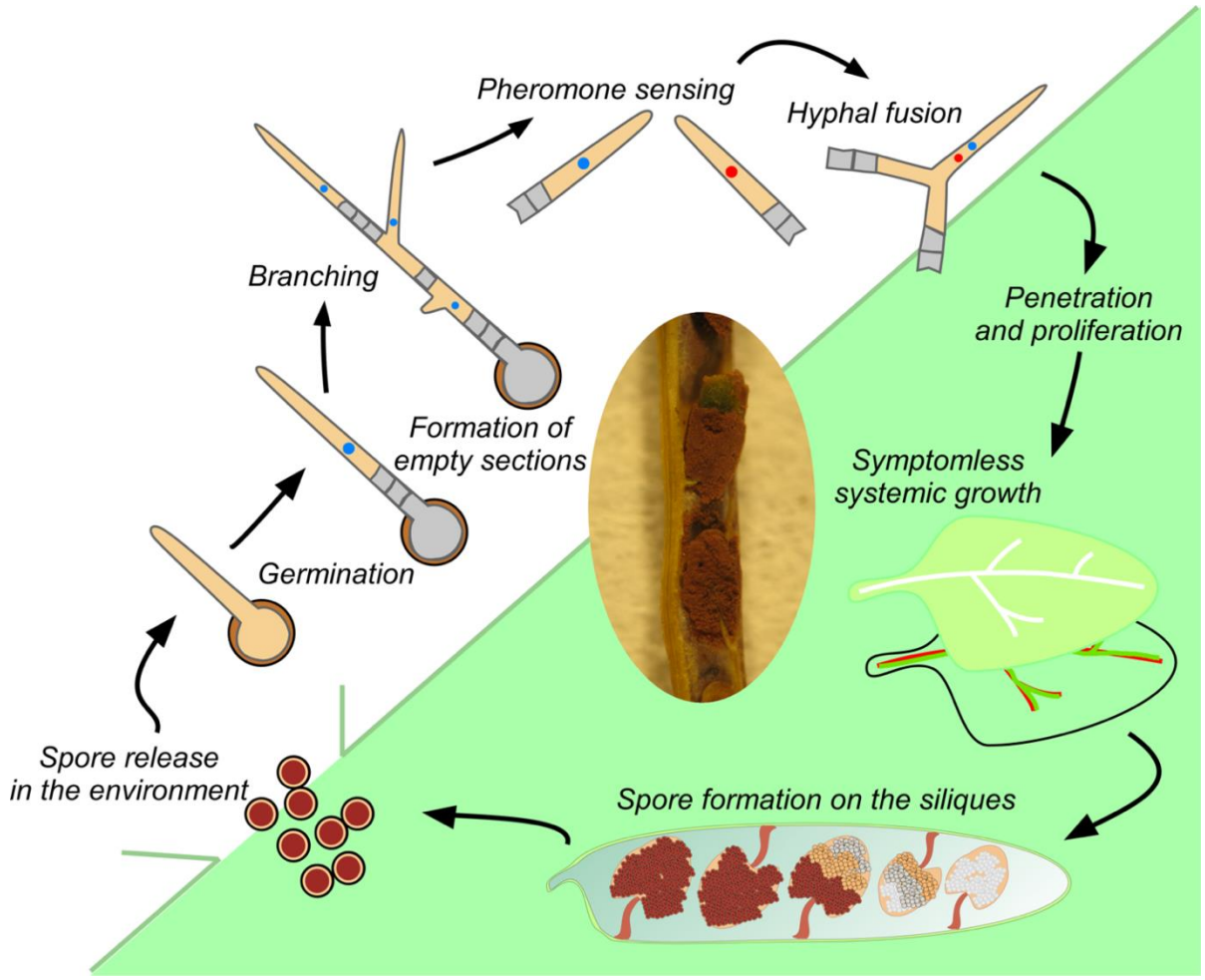


Figure 21. A proposed life cycle for *T. thlaspeos*. Spores germinate filamentously and branch leaving empty sections at every branch point. Resulting filaments are monokaryotic and haploid. Pheromone sensing is likely to contribute in sensing of filaments with different mating types leading to hyphal fusion. As in *U. maydis*, dikaryotic filaments are expected to be infectious. Penetration and proliferation in the host tissue follows. Growth in the host tissue is asymptomatic and restricted in the proximity of the vascular bundle. Later, in parallel to the development of the seeds, teliospores are formed covering the host seed. Teliospore deposition is gradual, starting with hyaline/white spores and finished with brown colored teliospores. Release in the environment occurs through the wind or together with the seed in the soil.

4 MATERIALS AND METHODS

4.1 Molecular biology and Mycology methods

4.1.1 Chemicals and Enzymes

Chemicals used in this work were purchased from the following companies unless otherwise specified:

- Sigma-Aldrich, Merck,
- Thermo Scientific,
- Fluka,
- Roche,
- Difco,
- BioRad,
- GE Healthcare,
- Pharmacia,
- Invitrogen
- Carl Roth.

All restriction enzymes used were purchased from New England Biolabs.

For DNA ligation, Quick-DNA ligase was used, purchased from New England Biolabs.

For plasmid isolation form *Escherichia coli* lysozyme purchased from Merck was used. Subsequent digestion of RNA was done using RNAse A purchased from Roche.

For all PCR reactions Phusion DNA polymerase from New England Biolabs was used.

4.1.2 PCR, Plasmid extraction, Cloning

DNA isolation, amplification and cloning was carried out with the methods described in (Sambrook *et al*., 1989)

4.1.3 Media

Media used for spore germination and culture are described in the Suppl. Table S2. For solid plates, the Phytagel gelling agent (SIGMA, P8169) was used since the Bacto-Agar (BD, 214010) was causing growth inhibition of *T*. *thlaspeos*.

4.1.4 Plant growth chamber

All plants in soil were grown in GroBanks CLF PlantClimatics BrightBoyXXL2 plant chamber. The program used was: 16hrs at 22°C, 8hrs at 18°C, Fluorescent light intensity approximately 125 µEinstein, LED intensity approximately 125 µEinstein.

All plants grown in plates were grown in Panasonic Versatile Environment Test Chamber (MLR-352H Series). The program used was: 12hrs at 21°C, 12hrs at 18°C, Fluorescent light intensity approximately 110 µEinstein.

4.1.5 Germination of *T. thlaspeos* teliospores and culture of strains

Germination assays:

- For germination assays by co-germination, *T. thlaspeos* teliospores were harvested from infected siliques by opening the infected siliques in an 1,5 ml Eppendorf tube. After brief centrifuging of the spores (30s, 13000rpm) distilled H_2O with Ampicillin (100 µg/ml) was added and vortexed. Spores are left to incubate for 24h at root temperature, shaking vigorously (1000 rpm).
  Seeds to be used in the assay are disinfected with EtOH (3x 100% EtOH for 1 min) and let dry under a clean bench.
  Teliospores are mixed with seeds of *A. thaliana Col-0* or *Ar. hirsuta* or *Ar. ciliata* seeds in half-strength MSN liquid medium and placed in a growth chamber at 21 ^{o}C, 8h light. Teliospore germination occurs 3-7 days post seed germination.

- For germination assays with leaf extracts, rosette leaves are harvested from *Ar. hirsuta* or *Ar. alpina* or *Ar. ciliata* grounded with a mortar and a pestle and a 0,075% w/v solution with dH_20 is made.
  The solution can be autoclaved or filter-sterilized and the teliospores are disinfected as mentioned above. Germination is observed after 3-7 days of incubation at room temperature.

Culturing of filaments:

After 4-8 weeks of co-incubation with the host plants, hyphal clumps were formed by spores not attached to host tissue (roots). These clumps were transferred in 20ml barrier flasks with CM liquid medium (200rpm/16 ^{o}C). The medium was refreshed by carefully changing the liquid every 5 days with a pipet. Caution was required in order to avoid removing the hyphal clumps. After 1 month the culture was dense enough, to exchange the medium by pelleting (10.000 rpm/20min). In order to get cultures originating from a single cell, 200µl of culture were plated in CM-phytoagar and colonies were restreaked for at least 5 times. Cultures coming from colonies were screened with the an *a* and *b* locus PCR marker. Cultures that had a single allele for each mating type were considered to be originating from a single cell.

Verification of the mating type locus:

The markers used are for the *a* locus:

- Primers DD209/210 for *pra1*
- Primers DD207/208 for *pra2*

And for the *b* locus:

- Primers DD15/DD175, amplifying the variable parts and the promoter region of bW and bE. Products of this PCR have to be sequenced.

With this protocol the cultures LF1 and LF2 were generated. LF1 originated from *T*. *thlaspeos* teliospores from a single silique of an *Ar*. *hirsuta* infected plant collected in Eselsburg, Germany in 2013 and LF2 originated from *T*. *thlaspeos* teliospores from a single silique of an *Ar*. *hirsuta* infected plant collected in Ronheim, Germany in 2013.

4.1.6 Germination of *Thecaphora spp*. teliospores and culture of strains

Germination:

Teliospores of *T*. *saponariae*, *T*. *seminis-convolvuli* and *T*. *oxalidis* were harvested from infected samples of their respective hosts. Spores were then plated in CM-phytagel and water-phytagel medium and placed on RT (±20 ^{o}C). The media were supplemented with Kanamycin (50 μg/ml) and Ampicilin (100 μg/ml). Spores germinated after 4-8 weeks.

Culturing:

Forming colonies were picked and restreaked at least 5 times in CM-phytagel plates. Cultures started by picking single colonies and placing the cells in CM liquid medium in 20ml barrier flasks (200rpm shaking, 16 ^{o}C).

With this protocol three strains were generated:

1. *T*. *oxalidis* 4-1
2. *T*. *saponariae* P-1
3. *T*. *seminis-convolvuli* 4e5-1

Conidia of *T*. *seminis-convolvuli* and *T*. *oxalidis* were harvested by the anthers of infected host, by vortexing the anthers in a 1,5ml Eppendorf tube filled with H_2O. The suspension was plated in water-phytagel supplemented with Kanamycin (50 μg/ml) and Ampicilin (100 μg/ml).

Germination occurs 3-24 hours later.

4.1.7 Genomic DNA and RNA extraction

Genomic DNA from teliospores for metagenomic sequencing was prepared by pooling the spores of three infected siliques from an *Ar. ciliata* plant, sampled in Slovenia during 2012, in an 1,5ml Eppendorf tube. Spores were crushed by adding 3 steel balls and then shaking in a RetschMill (30 MHz, 5min). Genomic DNA was then extracted by using a standard SDS/Phenol protocol as for *U. maydis*.

Genomic DNA for genome sequencing or PCR was prepared by pelleting 200ml of dense turbid 1-week old culture (10.000 rpm/20min). The pellet was frozen in liquid nitrogen and was ground to fine powder. The powder was stored in 2ml Eppendorf tubes in liquid nitrogen. Genomic DNA was then extracted using a standard SDS/Phenol protocol as described in (Raeder & Broda, 1985). Quality of the DNA was assessed by Nanodrop measurements, and integrity of the extracted DNA fragments was analyzed by gel electophoresis. For PacBio sequencing, samples had a fragment size larger than 20 kb.

Bacterial DNA contamination checks were performed by PCR using the primers 27F and 1391R. Samples which contained contaminating bacterial DNA were discarded.

RNA extraction from *T. thlaspeos*, *T. saponariae*, *T. oxalidis*, *T. seminis-convolvuli* for transcriptome sequencing was carried out using the Qiagen RNAeasy kit (QIAGEN Cat No./ID 74104) or the Trizol RNA isolation solution (Thermofischer Scientific, Product number 15596026).

4.1.8 cDNA synthesis

For all RT-PCR reactions and cloning, cDNA was synthesized using the Protoscript kit (NEB, E6560L) following the manufacturer's instructions.

4.1.9 Yeast two-hybrid assays

For the Y2H assays the Matchmaker Gold Yeast Two-Hybrid system was used (Takara-Clontech, 630489). For all steps required for the assay (cloning, preparation of yeast competent cells, etc.) the recommendations and guidelines of the manufacturer as stated in the "Matchmaker Gold Yeast Two-Hybrid System User Manual" and "Yeastmaker Yeast Transformation System 2 User Manual" were followed, unless stated otherwise in the text.

In order to clone the *TtbW* alleles from the *T. thlaspeos* strains LF1 and LF2, RNA was extracted from cultures incubated in CM for 5 days. RNA extraction and cDNA synthesis were executed as mentioned before. Afterwards *TtbW1* was amplified using the SfiI primers DD232/236 (*TtbW2* with DD239/236) and cloned to the pGAD and pGBK plasmids pUMa1624, pUMa1625 carrying

the GAL activation domain and GAL binding domain respectively. The resulting plasmids are the pUMa2494/2502 carrying *TtbW1* and pUMa2730/2731 carrying *TtbW2*.

The TtbE alleles were amplified from gDNA using primers DD232/233 (*TtbE2* with DD240/233) and cloned to the pCRII- TOPO vector (ThermoFischer, 45-0640). Afterwards TtbE was digested with SfiI and cloned to the pGAD and pGBK plasmids (pUMa1624, pUMa1625). The resulting plasmids are the pUMa2496/2504 carrying *TtbE1* and pUMa 2631/2632 carrying *TtbE2*.

To generate the truncated version of TtbE1var, the variable domain was amplified with the primers DD232/234 from pUMa2496 and cloned in pUMa1624, pUMa1625. Similarly, the TtbE1 conserved domain was amplified with the primers MB35/DD233 from the plasmid pUMa2496 and cloned in pUMa1624, pUMa1625.

The truncated version TtbW1var and TtbW2var, carrying the variable domains of TtbW, was generated after digestion with AatI of pUMa2494 and pUMa 2730, Klenow treatment and relegation. The resulting plasmids are pUMa2495 and 2498.

The *UmbE* and *UmbW* alleles were subcloned from the original Y2H plasmids (pUMa08/09/10, Kämper *et al*., 1995) to the Matchmaker Gold Yeast Two-Hybrid pGAD and pGBK plasmids pUMa1624, pUMa1625. The conserved domain of *UmbE* was amplified with the primers MB36/DD441 from the plasmid pUMa2576 and cloned to the pGAD and pGBK plasmids.

For control interactions the plasmids pUma1636, 1637, 1638 were used as suggested by the manufacturer (Clonetech).

4.1.10 Transient expression of TtNLP1 and TtNLP2 in *N. benthamiana*

In order to transiently express *T. thlaspeos* NLP homologs in *N. benthamiana*, the genes *TtNLP1* and *TtNLP2* were synthesized by GeneArt (ThermoFischer Scientific) and cloned in the plasmids pINVITROGEN_NLP1 and pINVITROGEN_NLP2. Afterwards TtNLP1 and TtNLP2 were amplified with the primers DD856/857/858 and DD859/860/861 respectively and cloned in the plasmid pENTR-D-TOPO. Using the Gateway cloning system, *TtNLP1* and *TtNLP2* were inserted in the plasmid pEarleyGate-103 (Earley *et al*., 2006) resulting in the plasmids pUMa2782-2789. Plasmids with a GFP fusion were not used for this study. Additional information and resources on these plasmids can be found in the web-page http://www.indiana.edu/~pikweb/pEarleyGate%20plasmid%20vectors%20copy/Table%20of%20vectors.html .

The *Ph. sojae PsojNIP* effector was amplified by the original plasmid provided by Kamoun S. (Qutob *et al*., 2002) using the primers DD804/805/806 and cloned in pEarlyGate-103 after digestion with XhoI.

Preparation of competent cells and transformation of *Agrobacterium tumefaciens* GV3101 pMP90-RK was carried out as described in (Weigel & Glazebrook, 2006). Infiltrations of *A. tumefaciens* in *N. bentamiana* were carried out as described in (Leuzinger *et al*., 2013).

4.1.11 Infected plant tissue staining

In order to described the biotrophic growth of *T. thlaspeos* and other Thecaphora *spp*. tissue samples of infected plants we harvested and put in 2ml Eppendorf tubes with 100% EtOH. EtOH was exchanged until the tissues had no more chlorophyll and turned completely white. Afterwards, EtOH is removed, KOH is added and the samples are incubated in 99 ^{o}C for 10min. KOH is removed with caution, so not to damage the sample.

The sample was then washed once with 1x PBS, and incubated with WGA (10 µg/ml in 1x PBS, SIGMA L4895) and PI (1:1000 dilution of 10mg/ml stock solution, diluted in 1x PBS) for 15 min. Sample was finally washed with PBS and kept at RT until microscopy.

To prepare 10x PBS buffer solution the following protocol was used:

	For 1000ml
79 mM Na2HPO4	8,9g
145 mM KH2PO4	19,7g
5 mM MgCl2*6H2O	0,9g
27 mM KCl (0,2% w/v))	2,0g
1,37 M NaCl (8% w/v)	80g

Adjust the pH with HCl or NaOH to 7,2.

4.1.12 Nuclear staining

Filaments of *Thecaphora spp*. or germinated teliospores were kept overnight in 70% EtOH. After the EtOH was removed, TE buffer with RNAse A (10 mg/ml, ROCHE 10109169001) was added, and the solution incubated for 1h at 50 ^{o}C. The samples were kept in PBS. PI (1:1000 dilution of 10mg/ml stock solution, diluted in 1x PBS) and/or CW (1µl of 0,002 mg/ml solution) was added before microscopy.

4.1.13 Microscopy

All macroscopy was carried out in the Axio Imager M1 microscope (ZEISS), and digital images of the cells were taken using a CCD camera Pursuit (SPOT). For the processing of the images MetaMorph (version 7) was used.

4.1.14 Oligonucleotides list

ID	Comment	Sequence
RL1093 LF_Tt_Frw	ITS Marker for T. thlaspeos	CTTTCGGCAACGGATCTCTAGG
RL1094 LF_Tt_Rev	ITS Marker for T. thlaspeos	AAGATTCGATGATTCACTGAC
RL1124_LF_BGI-27F	16SrRNA - Bacteria	AGAGTTTGATCMTGGCTCAG
RL1125_LF_BGI-1391R	16SrRNA - Bacteria	GACGGGCRGTGWGTRCA
RL1126_LF_BGI-342R	16SrRNA - Bacteria	CTGCTGCSYCCCGTAG
RL1127_LF_BGI-1492R	16SrRNA - Bacteria	TACGGYTACCTTGTTACGACTT
RL1358_Tt_LBA_FRW	probable lba in T thlaspeos	GTCGTTGTTCTTCTGCGGCGG
RL1359_Tt_LBA_REV	probable lba in T thlaspeos	CAGCGCAAGAACCCCAACGGC
RL1360_Tt_aro4_FRW	probable aro4 in T thlaspeos	CGGCGTTGAGCCTATCGAGCA
RL1361_Tt_aro4_REV	probable aro4 in T thlaspeos	GACATCAACGGCAGCTTCC
RL1362_Tt_panC_FRW	probable panC in T thlaspeos	CAAGTCGGTCGGCTTCGTCGC
RL1363_Tt_panC_REV	probable panC in T thlaspeos	GAACAGCTTGGTGACGACGGT
RL1364_Tt_tub2_FRW	probable tub2 in T thlaspeos	GGGGTTGTAGAGATTCTTGTA
RL1365_Tt_tub2_REV	probable tub2 in T thlaspeos	GCAGACGACGAGCACTACAT
RL1442_Ac_pep1_FRW	pep1 from Arabis ciliata	GATGTTTTAGAGTGAATTTTT
RL1443_Ac_pep1_REV	pep1 from Arabis ciliata	CCAAATGTTAATATTTCGTGA
RL1444_NL1	LSU primer	GCATATCAATAAGCGGAGGAAAAG
RL1445_NL4	LSU primer	GGTCCGTGTTTCAAGACGG
RL1446_NSIP	Plant specific primer forward	GATTGAATGATCCGGTGAAG
RL1447_NLBP	Plant specific primer reverse	GCTGTCACCCTCTCAGGC
RL1448_Tt_tub2-2_FRW	Thecaphora tubulin forward	CACCTTACAACGATAGCCTAG
RL1449_Tt_tub2-2_REV	Thecaphora tubulin reverse	GGCAGGCGTCCAGAGGGGCCG
RL1450_Tt_tub2-2_FRW2	Thecaphora tubulin forward in CDS	TCGTCCTCGACAATGCCGCGC
RL1451_Tt_tub2-2_REV2	Thecaphora tubulin reverse in CDS	TACGACGTCATGAGGAAGTGG
RL1452_asp1_FRW	Thecaphora anti silencing protein 1	GTCAACTCGTTTGAATTCGA
RL1453_asp1_REV	Thecaphora anti silencing protein 1	CATTTGATGTTAAACCTGGTG
RL1507_rbaFRW	Rba Thecaphora contig 6160	CATCGCAAGCACACCATGTC
RL1508_rbaFRW2	Rba Ustilago maydis	CAAAGCATTTACACACGATGTC
RL1509_rbaREV	Rba Thecaphora contig 6160	GCATGTCACGGCACTCACAG
RL1510_rbaREV2	Rba Thecaphora contig 6160	TTGTCGTGCGCGATGTGG
RL1511_rbaREV3	Rba Ustilago maydis	TCACTTACAAAAGCGACATCC
RL1512_lbafrw2	Lba Thecaphora contig 564	ATCCGTGTGCCATCGAGAG
RL1513_lbafrw3	Lba Thecaphora and Ustilago maydis	CAGCCGTTGGGGTTCTTG
RL1557_tub2F	Thecaphora thlaspeos tub2 forward	CGCACACCACACCTTACAACG
RL1558_tub2R	Thecaphora thlaspeos tub2 reverse	ATGTATGGCGAGGCAGCATGGC
MB189_DD233b	Amplifying TtbE1 from plamid	ATATGGCCGAGGCGGCCCTAGGCTGG
MB_280_Ttmfa2_rev	Ttmfa2 RT PCR	AGCTGCCGGGGAAGATCTTGC

MB_281_Ttmfa2_frw	Ttmfa2 RT PCR	CCCTATCTTTGCTCCCATCAC
MB316_ttCAT2_FRW1	secreted catalase THTG_00174	GTCACCATCCGCTTCTCAAC
MB317_ttCAT2_REV1	secreted catalase THTG_00174	GTCTTGGCCTTTGCCTCTTC
MB318_ttCAT2_FRW2	secreted catalase THTG_00174	GGCTTTGCCGTCAAGATCCG
MB319_ttCAT2_REV2	secreted catalase THTG_00174	TGAGGCGCAGGAACTGCAAC
MB320_ttCAT1_FRW1	catalase THTG_00409	GGCCTTCTGGGACTACATGG
MB321_ttCAT1_REV1	catalase THTG_00409	CCTCGTCGTTGGTGAGGTTC
MB322_ttCA_FRW	carbonic anhydrase THTG_05674	GACGCTTCCAATCGCCCATC
MB323_ttCA_REV	carbonic anhydrase THTG_05674	CATCCGCTGTTGGTCACCTC
MB324_ttSCA_FRW	secreted carbonic anhydr THTG_00198	TGGATCGGCTGTGCAGACTC
MB325_ttSCA_REV	secreted carbonic anhydr THTG_00198	TCGCGGTCTCGACGATGTTC
MB326_ttrbf1_FRW	THTG_04136 bbs1 target	ACGCAATGATCGACCCTTCC
MB327_ttrbf1_REV	THTG_04136 bbs1 target	AGGGCGATGATGACAAGCTC
DD09_Tsc_bWREV1	T. seminis-convolvuli bWest reverse	TTCTTGCTCTCGGTCATTGG
DD10_Tsc_bEREV1	T. seminis-convolvuli bEast reverse	GAGCGAGAAAGGGACGAAAC
DD11_Tsc_bWConsF	T. seminis-convolvuli bWest domain forward	AACGCCACTCACCCAAATCG
DD12_Tsc_bEConsRev2	T. seminis-convolvuli bEast domain reverse	AGGCGGTCGTCGTTGATCTC
DD13_Tsc_bEConsFrw1	T. seminis-convolvuli bEast domain forward	ACTGCCGAAGCCCTACAATC
DD14_Tth_bWFrw1	T. thlaspeos bWest forward	ATCTCGACGTCTTCCCACTC
DD15_Tth_bEREV	T. thlaspeos bEast reverse	CATGTAGTGCGACTGTAGGG
DD16_Tth_bEFRW1	T. thlaspeos bEast domain forward	GGGAAGAGGGCGGATGATAC
DD64_TtbWESTREV1	bWEST Tthlaspeos	GTCAGCGATCACATCGGACAG
DD65_TtbEASTREV1	bEAST Tthlaspeos	AGGCCATAAGAAAGGGCAAGG
DD66_TtbWESTFRW1	bWEST Tthlaspeos	TCCCTTCCTAGACGCTATCAC
DD67_TtbEASTFRW1	bEAST Tthlaspeos	AGCCACCACGATGAACATCC
DD68_Tsc_bEFRW2	bEAST Tseminis convolvuli	GGCTTCCCTCACATCCACATC
DD69_Tsc_bWFRW1	bWEST Tseminis convolvuli	CCCGTCTATCTTCCTCTCAC
DD146_TscbEastrev2	Tsc bEast new reverse primer	CGCGGAGAGTACAGGAAACG
DD168_TtbWest_seqR1	sequencing primer for bW from T. thlaspeos	GTGTTTGGGTCCGGTAGGTG
DD169_TtbW_start	for cDNA amplification	ATGGATACTACCACTCGACAA
DD170_TtbW_stop	for cDNA amplification	TCACATGCCGAATCCAGCCGA
DD171_TtbE_start	for cDNA amplification	ATGAACATCCAGGCACTCAAG
DD172_TtbE_stop	for cDNA amplification	CTAGGCTGGAACGCCAGCGAA
DD175_TtbWest_seqR2	sequencing primer for bpromoter from T. thlaspeos	CGAGAGTCGGTCTTCGATAC
DD207_Ttpra2-F	a locus marker	ACCACCGCCGACGTCTTTAC
DD208_Ttpra2-R	a locus marker	GTCAACGGCACCACGAAAGC
DD209_Ttpra1-F	a locus marker	ATATGGGCGAGATCAGTAGG

DD210_Ttpra1-R	a locus marker	TGGAAGACATGACGGATACG
DD211_TtbW1-intronF	bW intron primer	GCACACTCAGAGCATTAC
DD212_TtbW1-intronR	bW intron primer	GCATCCTGGCTTTCTTAG
DD232_bE1_sfi-start	to be used in Y2H vectors,reordered 24.11.15	GATCGGCCATTACGGCCATGAACAT CCAGGCACTCAAGG
DD233_bE1_sfi-stop-585	to be used in Y2H vectors,reordered 20.1.15	TAATGGCCGAGGCGGCCCTAGGCTG GAACGCCAGCGA
DD234_bE1_sfi-stop-127	to be used in Y2H vectors	GGCCGAGGCGGCCTCAGTAGTGCGA CTGGAGAGCCTC
DD235_bW1_sfi-start	to be used in Y2H vectors	GGCCATTACGGCCATGGATACTAC CACTCGACAA
DD236_bW1_sfi-stop-705	to be used in Y2H vectors,reordered 20.1.15	TATAGGCCGAGGCGGCCTCACATG CCGAATCCAGCCGA
DD237_bW1_sfi-stop-146	to be used in Y2H vectors	GGCCGAGGCGGCCTCACTCGGCCG CGCAGGTGTCGT
DD238_TtbW1-intron-alt	alternative splicing in bW primer	GTGAGCAGCAGTCGTGTCTCT
DD239_TtbW2-sfi-start-RH13D-3	to be used in Y2H vectors,reordered 20.1.15	ATGCGGCCATTACGGCCATGGATA CCATCGATGTACG
DD240_TtbE2-sfi-start-RH13D-3	to be used in Y2H vectors,reordered 20.1.15	GATCGGCCATTACGGCCATGAACA TCCAGGAACTCAAG
DD241_TtbW3-sfi-start-SL75j-2	to be used in Y2H vectors	GGCCATTACGGCCATGGACGCCACC ACTGGACAA
DD242_TtbE3-sfi-start-SL75j-2	to be used in Y2H vectors	GGCCATTACGGCCATGAACATCCAG GCACTCAAG
DD243_Tt-mfa1_a1	contig verification	ATGGACGCCATCTTCAACAC
DD244_Tt-mfa1_a1-um10153-stop	contig verification	AAACGTGGCGAGGATGAGGAC
DD245_Tt-mfa1-stop_a1	contig verification	TTAGGCAATGGTGCAGAAGG
DD434_pUMa010-bW2-FRW-sfi	Y2H Ustilago maydis controls	GATCGGCCATTACGGCCATGCTACCG CCACTGCCAAGA
DD435_pUMa010-bW2-REV-sfi	Y2H Ustilago maydis controls	GATCGGCCGAGGCGGCCCATATGAA GGATACCTTCTC
DD436_pUMa008-bE1-FRW-sfi	Y2H Ustilago maydis controls	GTACGGCCATTACGGCCATGTCCAGC GACCCGAATTTC
DD437_pUMa008-bE1-REV-sfi	Y2H Ustilago maydis controls	GATAGGCCGAGGCGGCCTCAGCCAA ACGCAGTAGAAAG
DD440_pUMa009-bE2-FRW-sfi	Y2H Ustilago maydis controls	GATCGGCCATTACGGCCATGTCTAAC TACCCGAACTTT
DD441_pUMa009-bE2-REV-sfi	Y2H Ustilago maydis controls	GATAGGCCGAGGCGGCCTCAGCCAA ACGCAGTAGAAAG
DD792-TtNLP1-F1	Tt NLPs for RT-PCR	CAACCATCCCTCACAACA
DD793 TtNLP1-R1	Tt NLPs for RT-PCR	CTTTCATCTCGTCCTTCT
DD794 TtNLP1-F2	Tt NLPs for RT-PCR	TTCAAGAAAGGGGCCAAT
DD795 TtNLP1-R2	Tt NLPs for RT-PCR	GAGTTTGTGCTTGTGGTT
DD796 TtNLP2-F1	Tt NLPs for RT-PCR	ACAGCCTTCTACTTCCCC
DD797 TtNLP2-R1	Tt NLPs for RT-PCR	ATTTCCTCTCTTTATGTCGCC
DD798 TtNLP2-R2	Tt NLPs for RT-PCR	TTCCAGCCGAGAGCGTAT
DD799 TtNLP3-F1	Tt NLPs for RT-PCR	gtgggcttttctttactctt
DD800 TtNLP3-R1	Tt NLPs for RT-PCR	tgcgctgtattggatctt

DD801 TtNLP3-F2	Tt NLPs for RT-PCR	ctccatctctccctcattt
DD802 Ttpep1-F1	Ttpep1 for RT-PCR	CTTTTCCTCTTCACCTCC
DD803 Ttpep1-R1	Ttpep1 for RT-PCR	TCTTGTCGTAATCGTCCT
DD804 PsojNIP-F1-xhoI	PsojNIP cloning to pEG103	CTCGAGATGAACCTCCGCCCTGCACTC
DD805 PsojNIP-R1-STOP-xhoI	PsojNIP cloning to pEG103	CTCGAGTTAAGCGTAGTAGGCGTTGCC
DD806 PsojNIP-R1-NOSTOP-xhoI	PsojNIP cloning to pEG103	CTCGAGAGCGTAGTAGGCGTTGCC
DD856_TT_NLP1_F	NLP1 start	CACCATGAAACCGTTCGCCCTTCTC
DD857_TT_NLP1_R_STOP	NLP1 end with stop codon	CTAgtaatctggaacatcgta
DD858_TT_NLP1_R_NOSTOP	NLP1 end with no stop codon	gtaatctggaacatcgtatgg
DD859_TT_NLP2_F	NLP2 start	CACCATGAGACCGTTCAATCTCCTC
DD860_TT_NLP2_R_STOP	NLP2 end with stop codon	CTAgtaatctggaacatcgta
DD861_TT_NLP2_R_NOSTOP	NLP2 end with no stop codon	gtaatctggaacatcgtatgg
DD889_Tt_Ualocus_gap_check_F1	GAP check for the ustilagic acid in Tt LF1	GGCGGCTAAGTAGTATAGAG
DD890_Tt_Ualocus_gap_check_R1	GAP check for the ustilagic acid in Tt LF1	AGTGCGATGAGGCCTATAAG
DD891_Tt_461_gap_check_F1	GAP check in the gene Tt00461_rd7-1g	CGATGTCTTCGCTGACAAAC
DD892_Tt_461_gap_check_R1	GAP check in the gene Tt00461_rd7-1g	TCAAGATGGCCAAGATCACC
DD893_Tt00459rd7_1g_F1	Homologue of ORF2 in UA cluster and Pf	CCTCACCGTATCGATTTACC
DD894_Tt00459rd7_1g_R1	Homologue of ORF2 in UA cluster and Pf	TCATGCCCTTCTGCTTAGCC
DD895_Tt00459rd7_1g_F2	Homologue of ORF2 in UA cluster and Pf	ATGATGCCTTATACTCACTC
DD896_Tt00457rd7_1g_F1	Homologue of Pfl000018, umag01428	CGGCTGCCCTTGTTCCATAG
DD897_Tt00457rd7_1g_R1	Homologue of Pfl000018, umag01428	CCACCCGATGACTTGCACTC
DD898_Tt00455rd7_1g_F1	Homologue of Pfl000001, no umag hit	TTTGCGAGAGGCGGTGAATC
DD899_Tt00455rd7_1g_R1	Homologue of Pfl000001, no umag hit	CTCCGGGAGCGGCAAATATG
DD900_umag6465_F1	ORF2 in UA cluster	CTCGGCAAGGTTTCGGGTAG
DD901_umag6465_R1	ORF2 in UA cluster	GACACAGCGCCGTAGTGTAG
DD902_umag6462_F1	uat1 in UA cluster	CCTGGTGCTTGATTATCTCC
DD903_umag6462_R1	uat1 in UA cluster	AAGCTCGGCAAGCCTTATAC
DD904_ttACT1_F1	act1 homolog in thecaphora thlaspeos for RT-PCR	ACCACCCGTTGCTCATGACC
DD905_ttACT1_R1	act1 homolog in thecaphora thlaspeos for RT-PCR	TCTGCGCCGTAGAGGCTTTG

4.1.15 Plasmid list

ID	*Plasmid*	Contributor	Description
pUMa1636	pGADT7-T	Thomas Pohlmann	Kit Control Plasmid
pUMa1637	pGBKT7-Lam	Thomas Pohlmann	Kit Control Plasmid
pUMa1638	pGBKT7-p53	Thomas Pohlmann	Kit Control Plasmid
pUMa08	GalBD-E1-pBT9	Michael Reichmann	(Kämper *et al*., 1995)
pUMa09	GalBD-E2-pBT9	Michael Reichmann	As above
pUMa10	W2-GalAD-pGADcterm	Michael Reichmann	As above
pUMa1624	pGAD-DS@1	Kai Heimel	Matchmaker Backbone AD vector
pUMa1625	pGBKT7 -Sfi I MCS_1	Kai Heimel	Matchmaker Backbone BD vector
pUMa2435	pTOPO_TSC_bWEST	L. Frantzeskakis / L. Pluecker	T. seminis-convolvuli bW1 genomic
pUMa2436	pTOPO_TSC_bEAST	L. Frantzeskakis / L. Pluecker	T. seminis-convolvuli bE1 genomic
pUMa2437	pTOPO_TTH_bWEST	L. Frantzeskakis / L. Pluecker	T. thlaspeos bW1 genomic
pUMa2438	pTOPO_TTH_bEAST	L. Frantzeskakis / L. Pluecker	T. thlaspeos bE1 genomic
pUMa2449	pTOPO_TSC_bPromoter	L. Frantzeskakis / L. Pluecker	*b* promotor sequence from Tsc
pUMa2450	pTOPO_Tt_bPromoter	L. Frantzeskakis / L. Pluecker	*b* promotor sequence from Tthlaspeos
pUMa2473	Topo-TtbW1-cDNA	L. Frantzeskakis	T. thlaspeos bW1 cDNA
pUMa2474	Topo-TtbE1-cDNA	L. Frantzeskakis	T. thlaspeos bE1 cDNA
pUMa2494	pGAD-DS-bW1-full	L. Frantzeskakis	y2h matchmaker, strain LF1, primers: dd235-236
pUMa2495	pGAD-DS-bW1-partial	L. Frantzeskakis	digestion with aatI, klenow and ligation
pUMa2496	pGAD-DS-bE1-full	L. Frantzeskakis	y2h matchmaker, strain LF1, primers: dd232-233
pUMa2497	pGAD-DS-bE1-partial	L. Frantzeskakis	y2h matchmaker, strain LF1, primers: dd232-234
pUMa2498	pGAD-DS-bW2-partial	L. Frantzeskakis	digestion with aatI, klenow and ligation
pUMa2499	pGAD-DS-bW3-partial	L. Frantzeskakis	y2h matchmaker, strain LF3, primers: dd241-237
pUMa2500	pGAD-DS-bE2-partial	L. Frantzeskakis	y2h matchmaker, strain LF2, primers: dd240-234
pUMa2501	pGAD-DS-bE3-partial	L. Frantzeskakis	y2h matchmaker, strain LF3, primers: dd242-234
pUMa2502	pGBKT7-bW1-full	L. Frantzeskakis	y2h matchmaker, strain LF1, primers: dd235-236
pUMa2503	pGBKT7-bW1-partial	L. Frantzeskakis	digestion with aatI, klenow and ligation
pUMa2504	pGBKT7-bE1-full	L. Frantzeskakis	y2h matchmaker, strain LF1, primers: dd232-233
pUMa2505	pGBKT7-bE1-partial	L. Frantzeskakis	y2h matchmaker, strain LF1, primers: dd232-234

pUMa2506	pGBKT7-bW2-partial	L. Frantzeskakis	digestion with aatI, klenow and ligation
pUMa2507	pGBKT7-bW3-partial	L. Frantzeskakis	y2h matchmaker, strain LF3, primers: dd241-237
pUMa2508	pGBKT7-bE2-partial	L. Frantzeskakis	y2h matchmaker, strain LF2, primers: dd240-234
pUMa2509	pGBKT7-bE3-partial	L. Frantzeskakis	y2h matchmaker, strain LF3, primers: dd242-234
pUMa2539	pTOPO-TtbW1-full with sfi sites	L. Frantzeskakis	T. thlaspeos bW1 cDNA with SfiI sites
pUMa2574	pGAD-DS-UmbW1-full	L. Frantzeskakis / L. Pluecker	Moving from the old plasmids to the Matchmaker plasmids
pUMa2575	pGBKT7-UmbW1-full	L. Frantzeskakis / L. Pluecker	Moving from the old plasmids to the Matchmaker plasmids
pUMa2576	pGAD-DS-UmbE1-full	L. Frantzeskakis / L. Pluecker	Moving from the old plasmids to the Matchmaker plasmids
pUMa2577	pGBKT7-UmbE1-full	L. Frantzeskakis / L. Pluecker	Moving from the old plasmids to the Matchmaker plasmids
pUMa2578	pGAD-DS-UmbE2-full	L. Frantzeskakis / L. Pluecker	Moving from the old plasmids to the Matchmaker plasmids
pUMa2579	pGBKT7-UmbE2-full	L. Frantzeskakis / L. Pluecker	Moving from the old plasmids to the Matchmaker plasmids
pUMa2631	pGAD-DS_Tth_bE2	L. Plücker /V. Göhre	MatchMaker vector with corresponding construct
pUMa2632	pGBKT7-bE2	L. Plücker /V. Göhre	MatchMaker vector with corresponding construct
pUMa2698	pEarlyGATE-103	L. Frantzeskakis	Gateway vector for plant transformations, brought from the Day Lab MSU
pUMa2730	TtbW2-pGAD-DS	L. Plücker /V. Göhre	MatchMaker vector with corresponding construct
pUMa2731	TtbW2-pGBK-T7	L. Plücker /V. Göhre	MatchMaker vector with corresponding construct
pUMa2780	pINVITROGEN_ttNLP1	L. Frantzeskakis	Synthesized sequence of TtNLP1 from GeneArt
pUMa2781	pINVITROGEN_ttNLP2	L. Frantzeskakis	Synthesized sequence of TtNLP1 from GeneArt
pUMa2782	pENTR-NLP1STOP	L. Frantzeskakis	pENTR Gateway vector with corresponding NLP construct
pUMa2783	pENTR-NLP1NOSTOP	L. Frantzeskakis	pENTR Gateway vector with corresponding NLP construct
pUMa2784	pENTR-NLP2STOP	L. Frantzeskakis	pENTR Gateway vector with corresponding NLP construct
pUMa2785	pENTR-NLP2NOSTOP	L. Frantzeskakis	pENTR Gateway vector with corresponding NLP

			construct
pUMa2786	pEG103-NLP1STOP	L. Frantzeskakis	TtNLP1 cloned in pUma2698
pUMa2787	pEG103-NLP1-GFP	L. Frantzeskakis	TtNLP1 without stop codon cloned in pUma2698
pUMa2788	pEG103-NLP2STOP	L. Frantzeskakis	TtNLP2 cloned in pUma2698
pUMa2789	pEG103-NLP2-GFP	L. Frantzeskakis	TtNLP2 without stop codon cloned in pUma2698
pUMa2796	pEG103-PsojNIP1STOP	L. Frantzeskakis	PsojNIP1 cloned in pUma2698
pUMa2810	TtbE1_var_pGAD	L. Plücker /L. Frantzeskakis	MatchMaker vector with corresponding construct
pUMa2811	TtbE1_var_pGBKT	L. Plücker /L. Frantzeskakis	MatchMaker vector with corresponding construct
pUMa2812	TtbE2_var_pGAD	L. Plücker /L. Frantzeskakis	MatchMaker vector with corresponding construct
pUMa2813	TtbE2_var_pGBKT	L. Plücker /L. Frantzeskakis	MatchMaker vector with corresponding construct
pUMa2842	pEG103-PsojNIP_GFP	L. Frantzeskakis/L. Pluecker	PsojNIP1 with no stop codon cloned in pUma2698
pUMa2871	pGAD_TtbE1conserved_domain	L. Frantzeskakis	MatchMaker vector with corresponding construct
pUMa2872	pGAD_TtbE2conserved_domain	L. Frantzeskakis	MatchMaker vector with corresponding construct
pUMa2873	pGAD_UmbE1variable_domain	L. Frantzeskakis	MatchMaker vector with corresponding construct
pUMa2874	pGAD_UmbE1conserved_domain	L. Frantzeskakis	MatchMaker vector with corresponding construct
pUMa2875	pGAD_UmbE2_conserved_domain	L. Frantzeskakis	MatchMaker vector with corresponding construct
pUMa2876	pGBK_TtbE1conserved_domain	L. Frantzeskakis	MatchMaker vector with corresponding construct
pUMa2877	pGBK_TtbE2conserved_domain	L. Frantzeskakis	MatchMaker vector with corresponding construct
pUMa2878	pGBK_UmbE1variable_domain	L. Frantzeskakis	MatchMaker vector with corresponding construct
pUMa2879	pGBK_UmbE1conserved_domain	L. Frantzeskakis	MatchMaker vector with corresponding construct
pUMa2880	pGBK_UmbE2_conserved_domain	L. Frantzeskakis	MatchMaker vector with corresponding construct

4.1.16 Strain list

ID	Species/Strain
UMa559	*Tilletiopsis pallescens*
UMa687	*Thecaphora seminis-convolvuli "Tsc 4e5-1"*
UMa1387	*Agrobacterium tumefaciens GV3101 pMP90RK*
UMa1541	*Thecaphora thlaspeos LF1 (a1b1)*
UMa1553	*Thecaphora oxalidis 4-1*
UMa2020	*Thecaphora thlaspeos LF2 (a2b2)*

4.2 Bioinformatics and computational methods

4.2.1 General purpose software

For preparation of plasmid maps and primer design the software package CloneManager 9 (Sci-Ed software) was used. For the alignments and trees as presented in the figures of this theses CLC Main Workbench 7 was used (Qiagen).

4.2.2 Genome sequencing and assembly

Sequencing and library preparation was done by the Sequencing Service of the Ludwig Maximilian University of Munich, Germany and in the Max Planck Genome Centre Cologne, Germany.

The short read dataset was generated by using the Illumina Mi-Seq platform and yielded 5433377 300bp paired reads. 4506766 passed the quality control and error correction step performed with Trimmomatic v0.32 (Bolger *et al*., 2014). The long read dataset was generated by using the PacBio platform (P6-C4) and yielded 332950 single reads. After error correction with the PBcR version included in the Celera release 8.3rc2 (Myers, 2000), 107327 reads were used for scaffolding and gap filling.

Assembly of the short reads to contigs was done with IDBA-UD v1.1.1 (Peng *et al*., 2012), which outperformed CLC Genomic Workbench 7 (QIAGEN) and Velvet 1.2.10 (Zerbino & Birney, 2008). Scaffolding of the IDBA-UD contigs was done using SSPACE-Longreads v1.1 (Boetzer & Pirovano, 2014) and the corrected PacBio reads. Gaps were filled using GapFiller v1-10 and PbSuite v14.7.14 (Boetzer & Pirovano, 2012; English *et al*., 2012).

Average coverage of the final assembly by short and long reads was estimated using a custom script and SAMtools v1.2 (Li *et al*., 2009). For an estimation of the assembly completeness CEGMA v2.5 and BUSCO v1.1b1, were used (Parra *et al*., 2007; Simao *et al*., 2015).

RepeatMasker v4.0.5 was used to report and mask repetitive regions in the genome (Jurka *et al*., 2005; Tempel, 2012).

4.2.3 RNA dataset

For *T. thlaspeos LF1* and *T. seminis-convolvuli 4e51* an RNAseq dataset was prepared to support the annotation. For both species, RNA was extracted from a single culture growing in CM for 48h at 16^{o}C and a single library was sequenced.

Sequencing and library preparation was done by the BMFZ sequencing center of the Heinrich-Heine University of Düsseldorf and sequenced with the Hi-Seq Illumina platform resulting in

12903314 single-end reads. Error correction was performed using Trimmomatic v0.32 (Bolger *et al*., 2014). After error correction, 11640030 were left. Mapping the reads to the genome was performed using TopHat2 ver. 2.1.0 (Kim *et al*., 2013).

4.2.4 Structural annotation

Structural annotation of the genome was done using MAKER 2.31.8 (Cantarel *et al*., 2008; Holt & Yandell, 2011), according to the protocols described in (Campbell *et al*., 2014). The same pipeline was followed for *T. seminis-convolvuli*.

After the first round of MAKER using the following protein evidence mentioned in Table 9 and the RNAseq dataset, 397 gene models were manually curated and used to train AUGUSTUS ver 3.0.3 (Stanke & Morgenstern, 2005) and SNAP ver 2006-07-28. Subsequently 3 iterations were performed, generating a new prediction models for AUGUSTUS and SNAP in every run.

Table 9. Proteome evidence used for annotation

	Version/Date
Complete UniProtKB/Swiss-Prot data set*	June 2015
Ustilago maydis proteome	MUMDB June 2015
Ustilago hordei proteome	MUMDB June 2015
Malassezia globosa CBS 7966 proteome	NCBI June 2015
Malassezia sympodialis ATCC 42132 proteome	NCBI June 2015
Melanopsichium pennsylvanicum proteome	MUMDB June 2015
Pseudozyma antartica T34 proteome	NCBI June 2015
Pseudozyma flocculosa proteome	NCBI June 2015
Pseudozyma aphidis DSM 70725 proteome	NCBI June 2015
Pseudozyma brasiliensis GHG001 proteome	NCBI June 2015
Pseudozyma hubeiensis SY62 proteome	NCBI June 2015
Sporisorium scitamineum proteome	NCBI June 2015
Sporisorium reilianum proteome	MUMDB June 2015
Tilletiaria anomala UBC 951 proteome	NCBI June 2015

*ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/complete/

For visualization Artemis ver. 16.0.11 was used (Carver *et al*., 2012). In the final annotation set, 225 gene models were manually curated for wrong start/stop codons, overlapping CDS features and wrong splice sites. All manually curated genes have the flag “note=curated” in the final gff file.

4.2.5 Functional annotation

For assignment of functional domains in the predicted proteins, InterProScan version 5.15-54 was used. The versions of the databases used for the annotation are listed in Table 9.

Table 10. Databases used for functional annotation

Database	Version
Hamap	201502.04
ProDom	2006.1
PIRSF	3.01
PFAM	28.0
SMART	6.2
Gene3d	3.5.0
Coils	2.2.1
TIGRFAM	15.0
PRINTS	42.0
SUPERFAMILY	1.75
PrositePatterns	20.113
PANTHER	9.0
PrositeProfiles	20.113
SignalP-EUK	4.0
TMHMM	2.0c

For annotation of the CAZymes, HMMER 3.1b1 (Finn *et al*., 2011) together with the database dbCAN v3 (04-Aug-2013) was used. The results were parsed with the accompanying script in the database package. The cut-offs described in (Yin *et al*., 2012). Thorough descriptions of the families can be found in the website www.cazy.org (Cantarel *et al*., 2009).

For the annotation of the secretome SignalP 4.0 (Petersen *et al*., 2011), TMHMM 2.0c (Käll *et al*., 2004), Phobius 1.01 (Käll *et al*., 2007) and WolfsPORT 0.2 (Horton *et al*., 2007) were used. The pipeline followed uses the criteria described in (De Jonge, 2012). Using this method, in order for a protein to be considered as secreted, it has to have a predicted signal peptide by SignalP and WolfsPORT and no transmembrane domains by TMHMM and Phobius.

For the annotation and identification of secondary metabolism clusters Anti-SMASH 3.0.5 was used (Weber *et al*., 2015).

The calculation of cysteine content of every secreted protein was carried out using an R script.

Prediction of NLS signals was performed with NLStradamus ver. 1.8 (Nguyen Ba *et al*., 2009)

4.2.6 Orthology with other species

For the identification of orthologous gene groups Orthofinder v 0.2.5 was used (Emms & Kelly, 2015). The proteomes used, excluding the Uniprot database, are mentioned in the Table 8.

4.2.7 Multi-locus phylogeny

For the multi-locus phylogeny, the pipeline described in (Lefebvre *et al*., 2013) was adopted. Initially single copy orthologs between the 13 *Ustilaginomycotina* genomes of Table 9 and T. *thlaspeos LF1* were extracted using a custom Perl script from the Orthofinder results (Suppl. Data 2). Then an alignment was generated for every gene family using MAFFT 7.215 (--phylipout --maxiterate 1000 –localpair) (Katoh *et al*., 2002), and for every alignment a tree was generated using PhyML 3.0 (default options) (Le & Gascuel, 2008). The 1773 resulting trees were collapsed using Phylip Consense (version 3.69) (Felsenstein, 1989).

4.2.8 Synteny

Whole genome alignments were carried out with the program MUMmer ver. 3.23 (Kurtz *et al*., 2004).

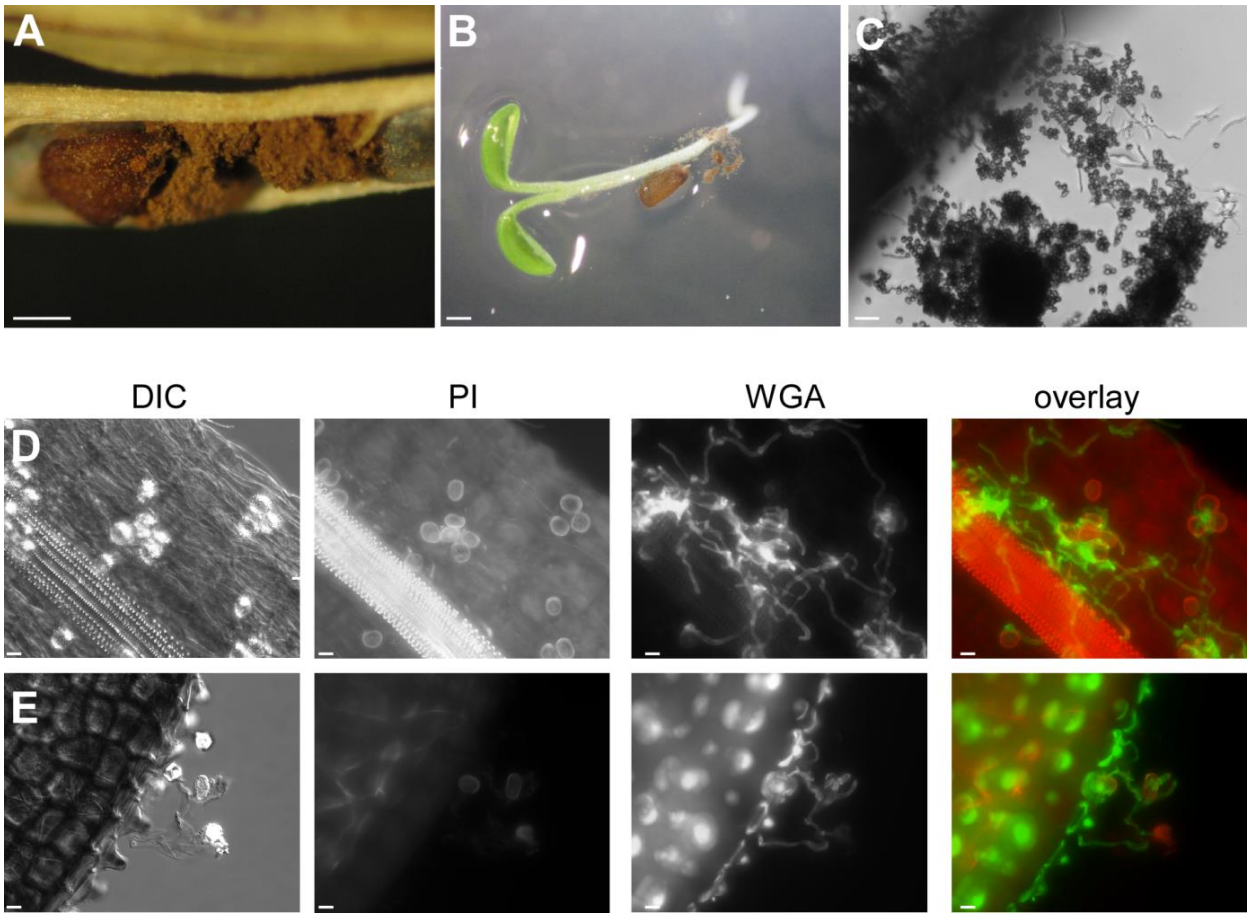
4.2.9 Expression of genes during culture

RNAseq reads mapping to annotation features (noted as mRNA in the gff file) were counted using htseq-count (Anders *et al*., 2015).

5 APPENDIX

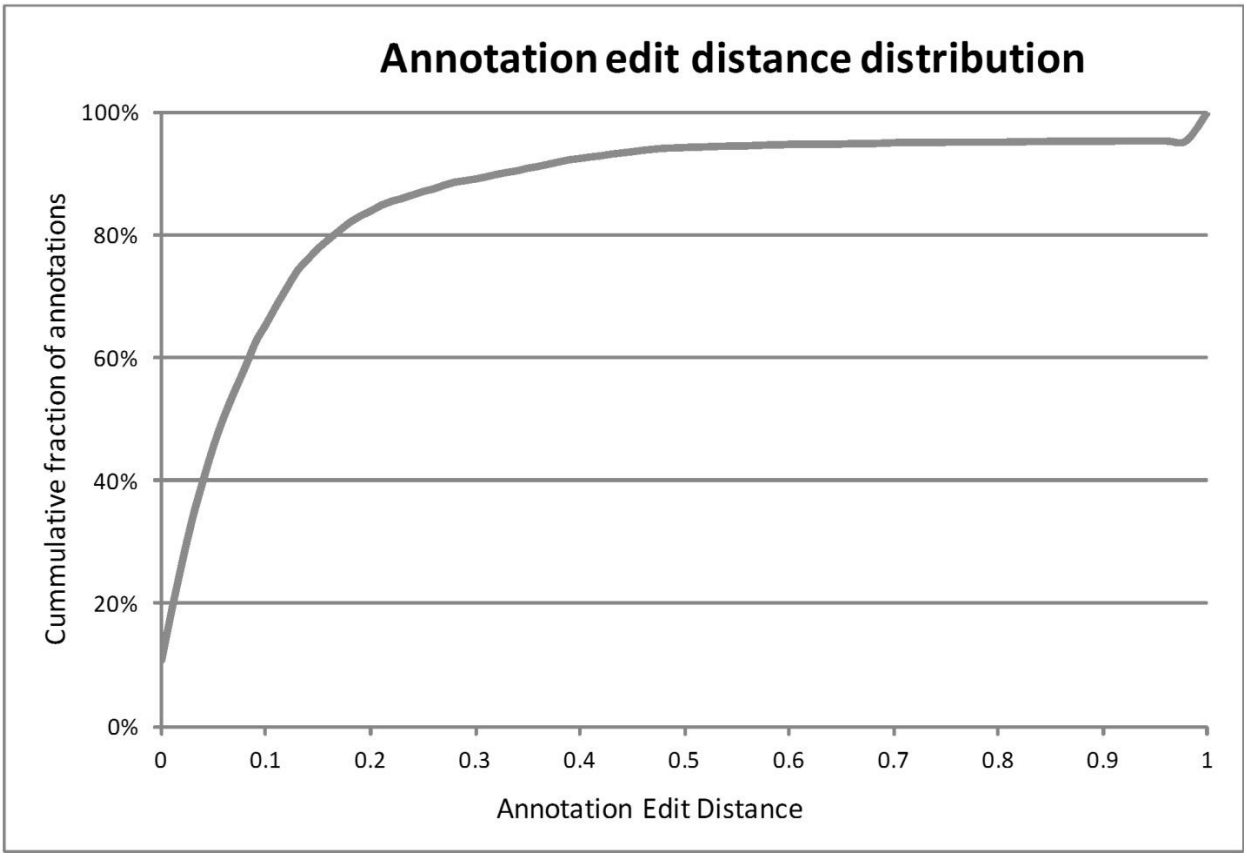
5.1 Supplementary Figures and Tables

Supplementary Figure 1



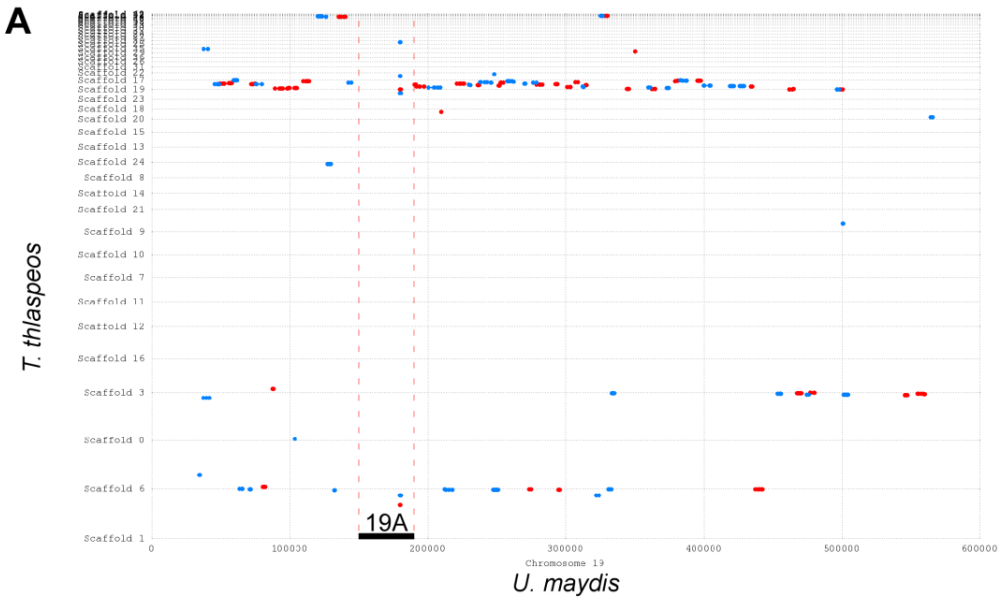
Suppl. Figure S1. Infected seeds of Ar. ciliata by *T. thlaspeos* are viable. Seeds from infected *Ar. ciliata* plants collected in nature are viable and germinate when plated in half strength MS medium (A and B, bar 0,5 mm). Germination of the seeds initiates germination of the attached spores (C, bar 50um). When seedlings were examined microscopically after WGA/PI staining, attached germinated spores were found in the hypocotyl (panel D, bar 10um), but also in the seed coat (panel E, bar 10um).

Supplementary Figure 2



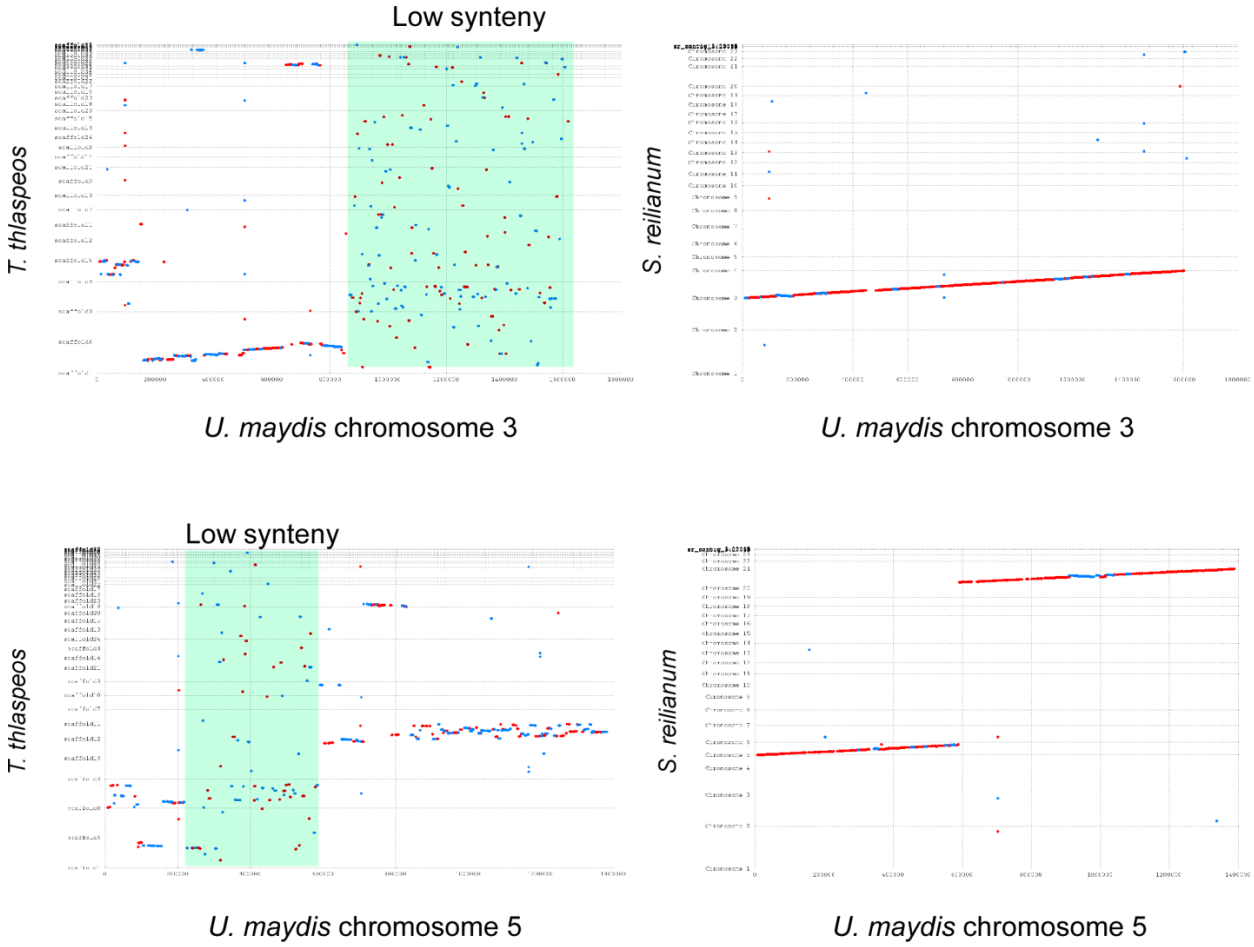
Suppl. Figure S2. Annotation edit distance distribution. For all gene models of the annotation version "tthl_lf1_finalv09_20151011_wo-mitgenome.functional_annotation_included_thtg_corr.gff" annotation edit distance was calculated based on the output of MAKER.

Supplementary Figure 3



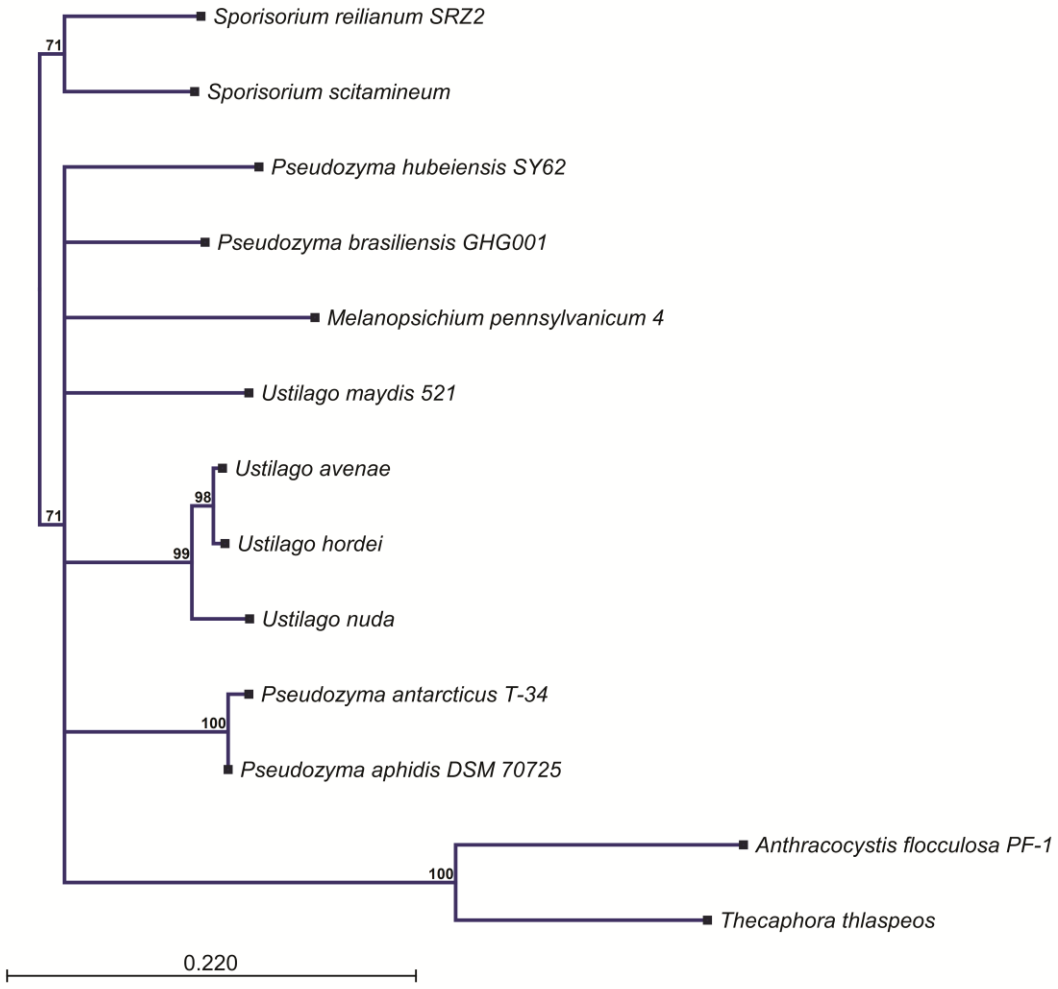
Suppl. Figure S3. Synteny between the *T. thlaspeos* scaffolds and the *U. maydis* chromosome 19. The genomic locus spanning from the gene UMAG_05294 (location 145921) to the gene UMAG_05319 (location 189148) (cluster 19A) harbors no homology to the genome of *T. thlaspeos* although the rest of the chromosome is homologous mainly to scaffolds 19 and 0. Blue dots indicate sense to sense alignment blocks are, and red dots indicate sense to anti-sense alignment blocks. Scaffolds in *T. thlaspeos* with a size smaller than 10 kb were removed.

Supplementary Figure 4



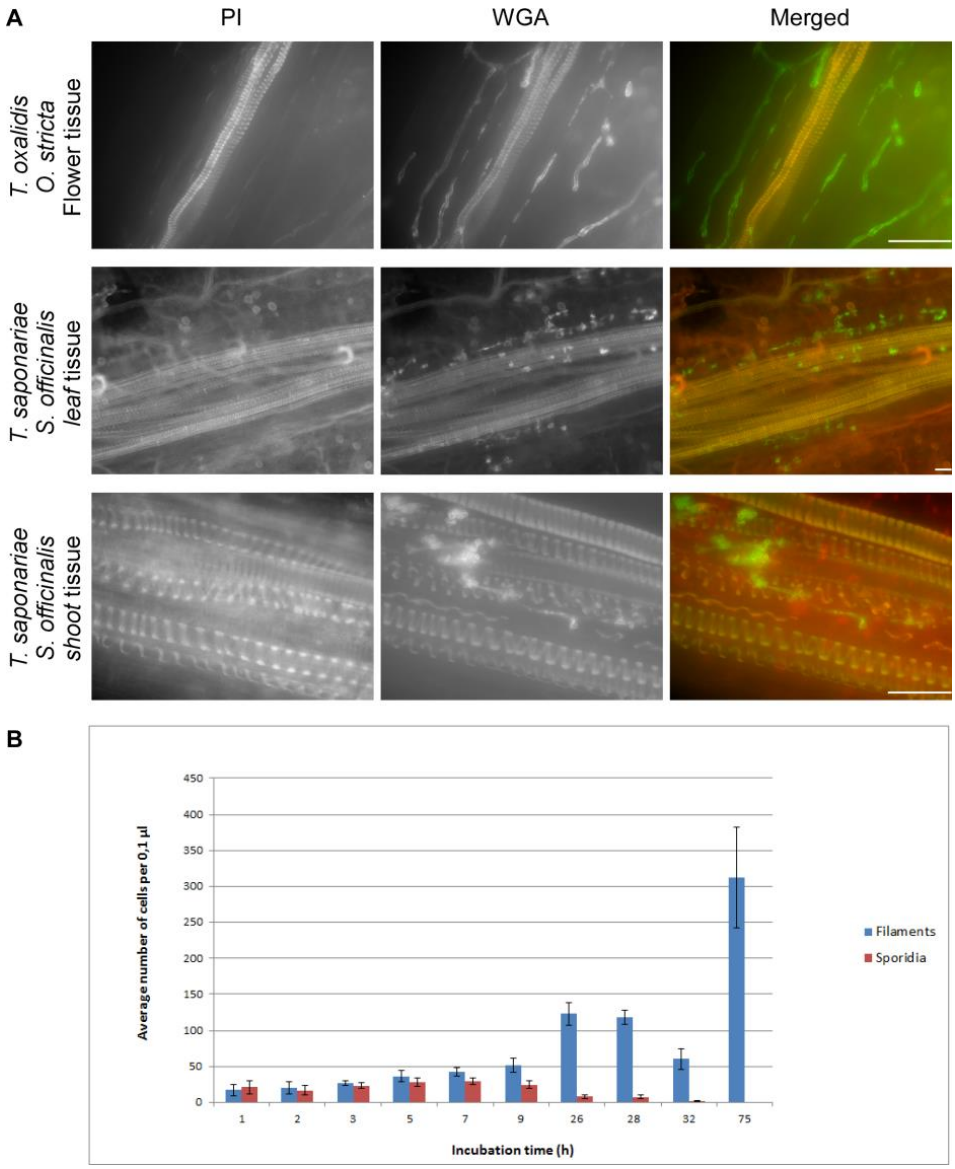
Suppl. Figure S4. Synteny between the *T. thlaspeos* scaffolds and the *U. maydis* chromosomes 3 and 5. Synteny is disrupted in areas highlighted in color boxes. Scaffolds in *T. thlaspeos* with a size smaller than 10 kb were removed, to aid visualization.

Supplementary Figure 5



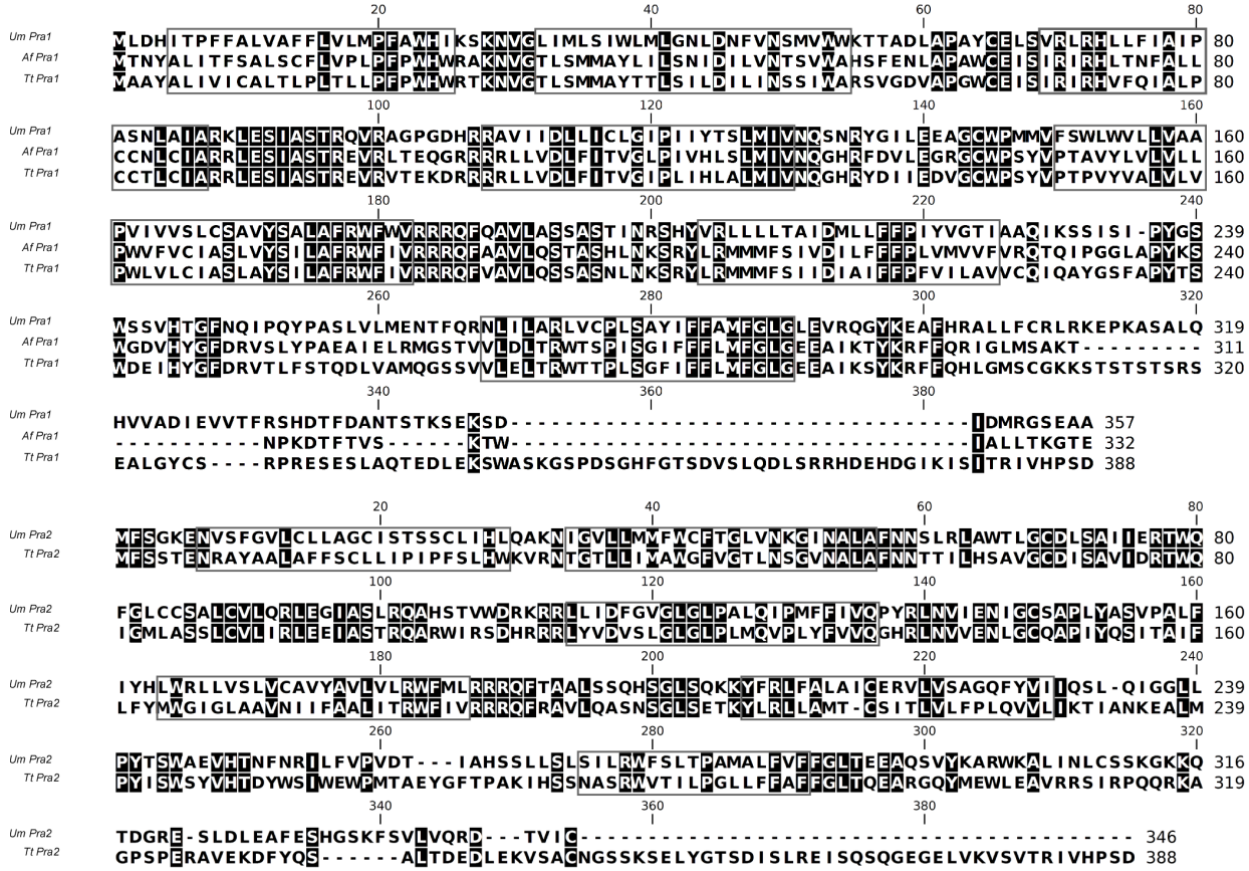
Suppl. Figure S5 Maximum likelihood phylogeny of TtPep1 and homologs. Maximum likelihood phylogeny tree was generated with CLC Main Workbench, after 1000 repetitions. Branches with bootstrap values lower than 60% were collapsed.

Supplementary Figure 6



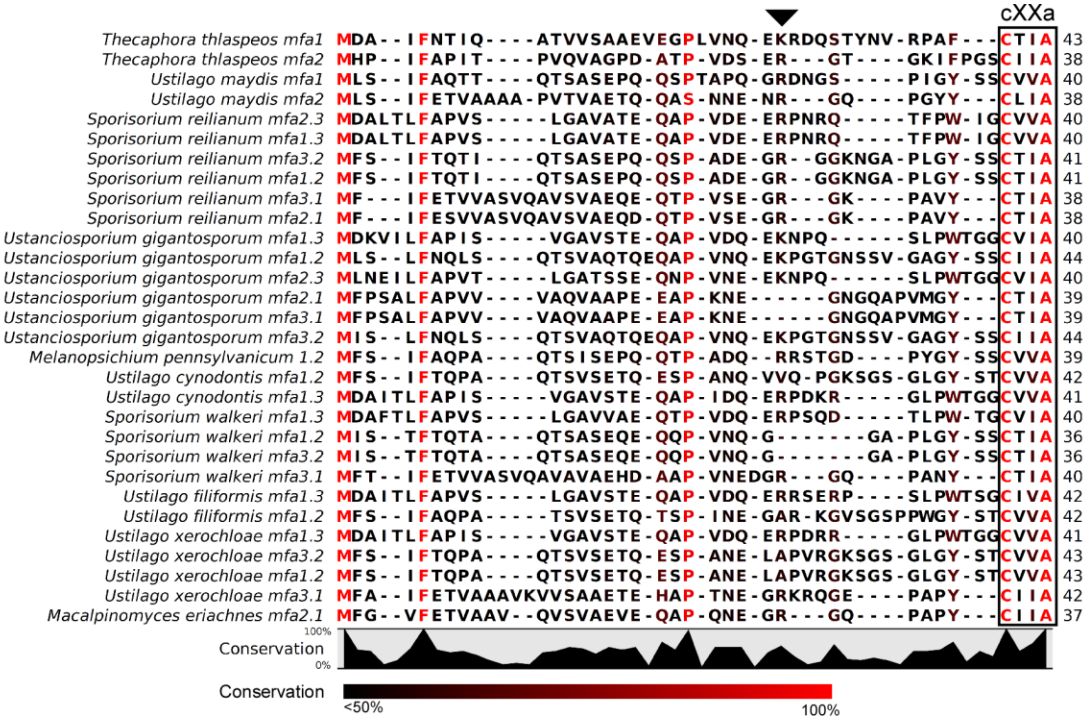
Suppl. Figure S6. Additional data on *T. saponariae, T. oxalidis and T. seminis-convolvuli*. (A) *T. oxalidis* and *T. saponariae* infected host tissues were sampled and examined for fungal growth. In both cases fungal hyphae was detected growing close to the vascular bundle as in *T. thlaspeos* (Bar = 50µm). (B) *T. seminis-convolvuli* cells growing in a CM liquid culture were counted with a Neubauer chamber. Although conidia take up half of the initial cell population, after prolonged incubation conidiation stops and all cells are switched to filaments. Bars show standard deviation of one sample.

Supplementary Figure 7



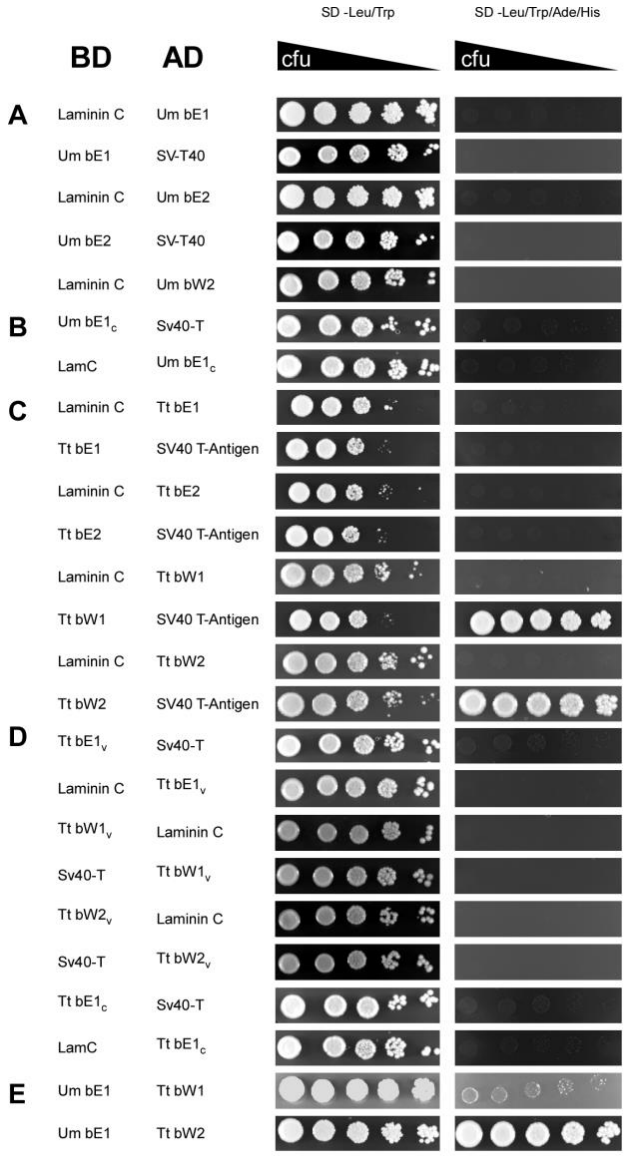
Suppl. Figure S7. Alignment of the pheromone receptor Pra1 and Pra2 of *T. thlaspeos*.. Aminoacid alignments of the *T. thlaspeos* Pra1 with the corresponding *U. maydis* and *A. flocculosa* homologues (UMAG_02383 and PFL1_01711 respectively) and *T. thlaspeos* Pra2 with the corresponding *U. maydis* homologue (AAA99768.1). Grey boxes show the transmembrane domains as predicted by TMHMM Server v 2.0 based on the UmPra1 and UmPra2 aminoacid sequences. Highlighted squares show conserved aminoacids.

Supplementary Figure 8



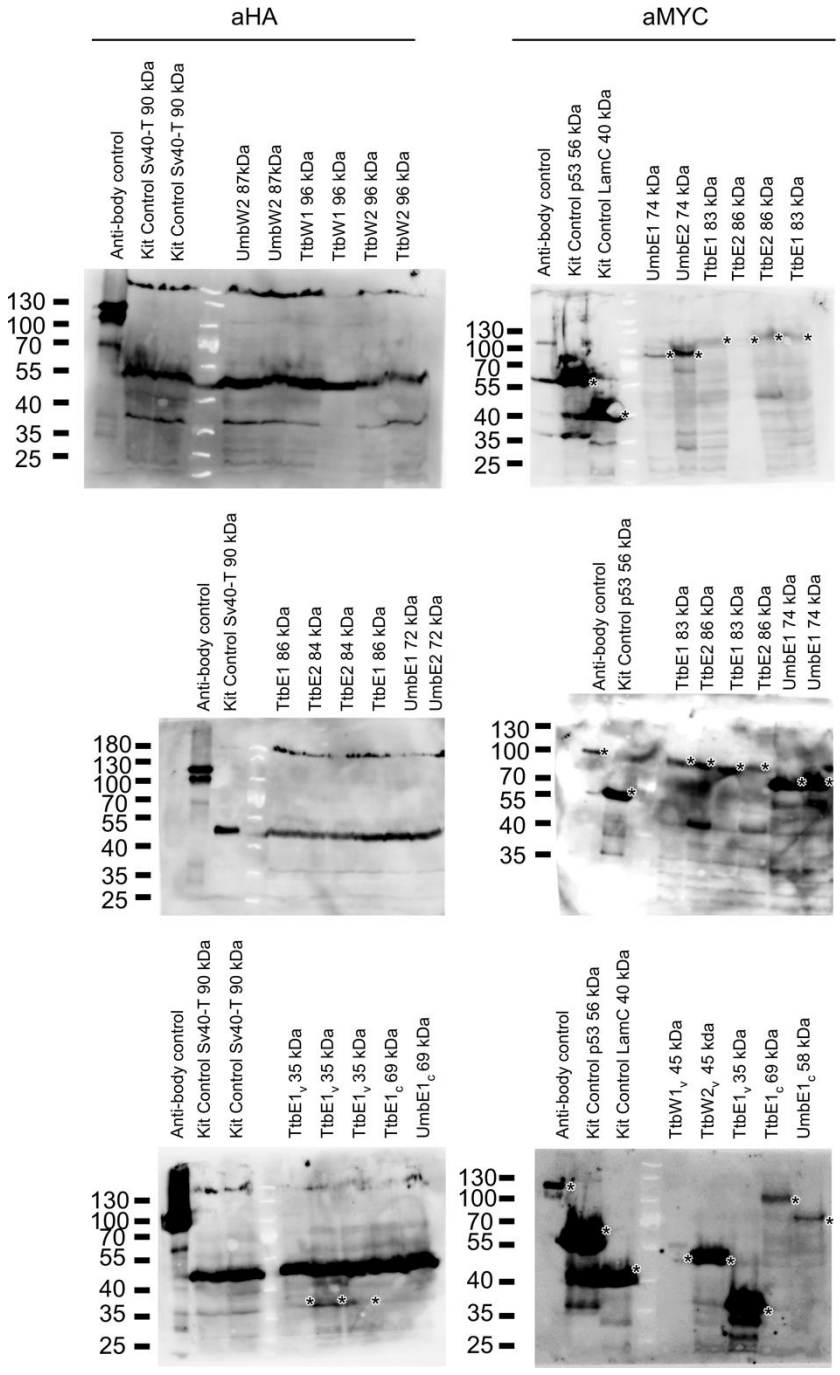
Suppl. Figure S8. Alignment of *T. thlaspeos* pheromone precursors with homologs of other smut species. Sequences of pheromone precursors included in the study (Kellner *et al*., 2011) were aligned with CLC Main Workbench. Filled triangle shows putative cleavage site.

Supplementary Figure 9



Suppl. Figure S9. Control interactions in the Clonetech MatchMaker system and supplementary combinations to the ones presented in Figure 14. Control interactions for the *U*. *maydis* bE1, bE2 and bW2. (A). Control interactions for conserved domain of *U*. *maydis* bE1. (B) Control interactions for TtbE1, TtbE2, TtbW1, TtbW2. TtbW contains an activation domain as the UmbW and it was not used in the vector pGBK which contains the GAL binding domain. (C). Control interactions for the variable and the conserved domain truncated constructs. (D). *U*. *maydis* bE1 interacts with TtbW, suggesting that interspecific interactions might also be possible (E).

Supplementary Figure 10

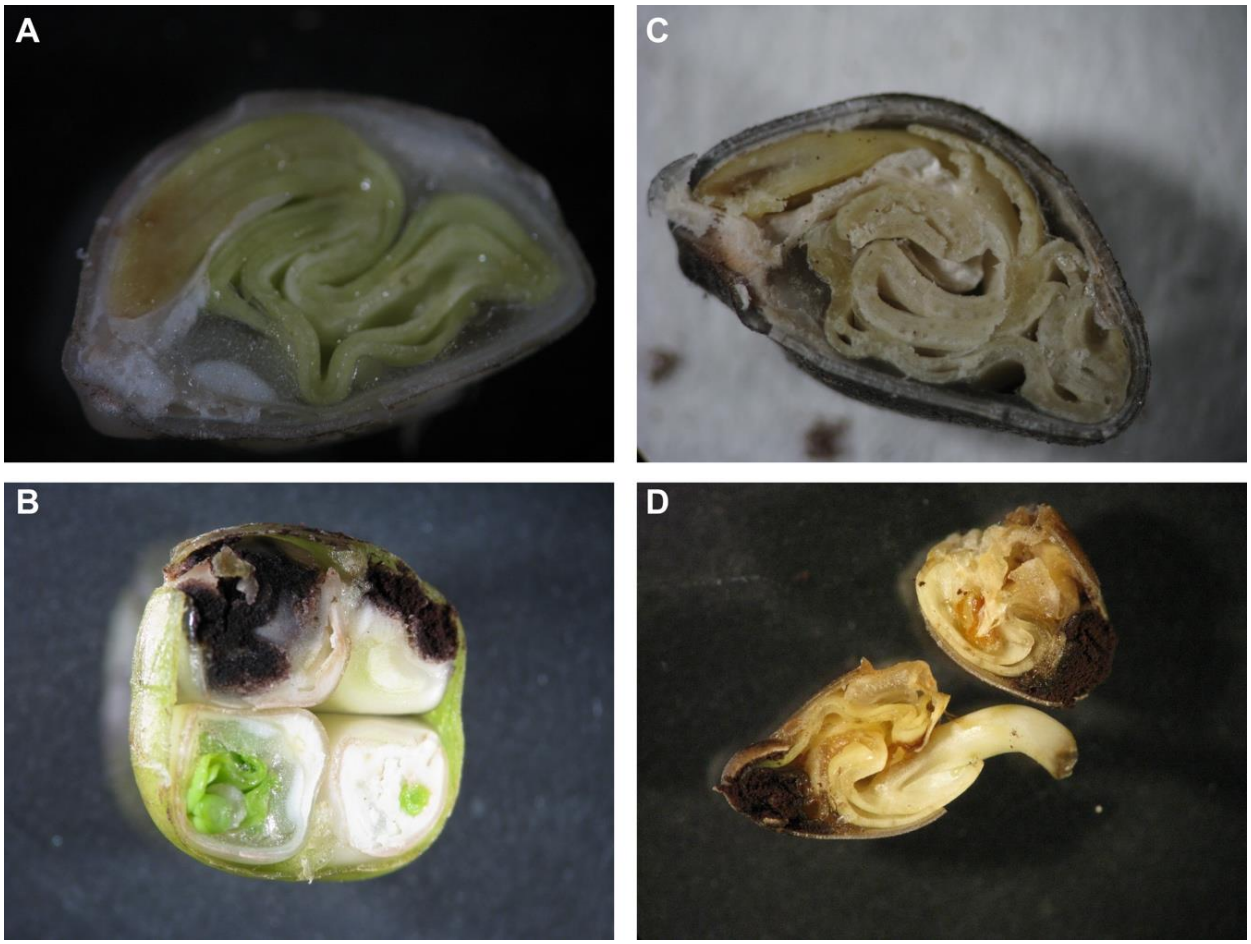


Suppl. Figure S10. Western blots of the Y2H interactions shown in Figure 14. Asterisks show expected sizes.

Supplementary Figure 11

Suppl. Figure S11. *bbs* sequences in *T. thlaspeos*. In (A) the relative position of the *bbs* elements to the *rbf1* homolog of *T. thlaspeos* is presented. In (B) alignments between the *T. thlaspeos* and *U. maydis bbs* elements is presented.

Supplementary Figure 12



Suppl. Figure S12. Healthy and infected seeds of *C. sepium* by *T. seminis-convolvuli*. Healthy immature (A) and dry (C) seed. In freshly set seeds of *C. sepium* (B) infection doesn't abolish seed formation, but allows the development of viable seeds which are able to germinate when placed in wet Whatman paper (D).

5.2 References

Albert I, Böhm H, Albert M, Feiler CE, Imkampe J, Wallmeroth N, Brancato C, Raaymakers TM, Oome S, Zhang H, *et al.* 2015. An RLP23–SOBIR1–BAK1 complex mediates NLP-triggered immunity. *Nature Plants* **1**: 15140.

Alderman SC. **2007**. Occurrence of Ustilago striiformis in Dactylis glomerata Seed Production Fields in Oregon. *Plant Health Progress*.

Alfaro M, Oguiza JA, Ramírez L, Pisabarro AG. **2014**. Comparative analysis of secretomes in basidiomycete fungi. *Journal of proteomics* **102**: 28–43.

Ali S, Laurie JD, Linning R, Cervantes-Chávez JA, Gaudet D, Bakkeren G. **2014**. An Immunity-Triggering Effector from the Barley Smut Fungus Ustilago hordei Resides in an Ustilaginaceae-Specific Cluster Bearing Signs of Transposable Element-Assisted Evolution. (J-R Xu, Ed.). *PLoS pathogens* **10**: e1004223.

Anders S, Pyl PT, Huber W. **2015**. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166–169.

Andrade O, Muñoz G, Galdames R, Durán P, Honorato R. **2004**. Characterization, In Vitro Culture, and Molecular Analysis of Thecaphora solani, the Causal Agent of Potato Smut. *Phytopathology* **94**: 875–82.

Bahn YS, Cox GM, Perfect JR, Heitman J. **2005**. Carbonic anhydrase and CO2 sensing during Cryptococcus neoformans growth, differentiation, and virulence. *Current Biology* **15**: 2013–2020.

Bai Y, Müller DB, Srinivas G, Garrido-Oter R, Potthoff E, Rott M, Dombrowski N, Münch PC, Spaepen S, Remus-Emsermann M, *et al.* 2015. Functional overlap of the Arabidopsis leaf and root microbiota. *Nature* **528**: 364–369.

Bailey B. **1995**. Purification of a Protein from Culture Filtrates of Fusarium oxysporum that Induces Ethylene and Necrosis in Leaves of Erythroxylum coca. *Phytopathology* **85**: 1250.

Bailey B a, Jennings JC, Anderson JD. **1997**. The 24-kDa protein from Fusarium oxysporum f.sp. erythroxyli: occurrence in related fungi and the effect of growth medium on its production. *Canadian journal of microbiology* **43**: 45–55.

Batts CC V. **1955**. Observations on the infection of wheat by loose smut (Ustilago tritici (Pres.) Rostr.). *Transactions of the British Mycological Society* **38**: 465–475, IN4–IN7.

Bauchwitz R, Holloman WK. **1990**. Isolation of the REC2 gene controlling recombination in Ustilago maydis. *Gene* **96**: 285–288.

Baumgarten AM, Suresh J, May G, Phillips RL. **2007**. Mapping QTLs contributing to Ustilago maydis resistance in specific plant tissues of maize. *Theoretical and Applied Genetics* **114**: 1229–1238.

Bayram E, Senturk M, Irfan Kufrevioglu O, Supuran CT. **2008**. In vitro inhibition of salicylic acid derivatives on human cytosolic carbonic anhydrase isozymes I and II. *Bioorganic and Medicinal*

Chemistry **16**: 9101–9105.

Bebber DP, Ramotowski M a. T, Gurr SJ. **2013**. Crop pests and pathogens move polewards in a warming world. *Nature Climate Change* **3**: 985–988.

Becht P, König J, Feldbrügge M. **2006**. The RNA-binding protein Rrm4 is essential for polarity in Ustilago maydis and shuttles along microtubules. *Journal of cell science* **119**: 4964–73.

Begerow D, Bauer R, Boekhout T. **2000**. Phylogenetic placements of ustilaginomycetous anamorphs as deduced from nuclear LSU rDNA sequences. *Mycological Research* **104**: 53–60.

Begerow D, Goker M, Lutz M, Stoll M. **2004**. On the evolution of smut fungi on their hosts. Frontiers in Basidiomycote Mycology. IHW-Verlag, .

Begerow D, Schaefer AM, Kellner R, Yurkov A, Kemler M, Oberwinkler F, Bauer R. **2014**. 11 Ustilagomycotina. The Mycota: Systematics and Evolution Part A VII.

Begerow D, Stoll M, Bauer R. **2006**. A phylogenetic hypothesis of Ustilaginomycotina based on multiple gene analyses and morphological data. *Mycologia* **98**: 906–16.

Berruti A, Lumini E, Balestrini R, Bianciotto V. **2016**. Arbuscular Mycorrhizal Fungi as Natural Biofertilizers: Let's Benefit from Past Successes. *Frontiers in microbiology* **6**: 1–13.

Bhavsar AP, Brown NF, Stoepel J, Wiermer M, Martin DDO, Hsu KJ, Imami K, Ross CJ, Hayden MR, Foster LJ, *et al.* 2013. The Salmonella Type III Effector SspH2 Specifically Exploits the NLR Co-chaperone Activity of SGT1 to Subvert Immunity. *PLoS Pathogens* **9**: 1–12.

Boetzer M, Pirovano W. **2012**. Toward almost closed genomes with GapFiller. *Genome biology* **13**: R56.

Boetzer M, Pirovano W. **2014**. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC bioinformatics* **15**: 211.

Böhm H, Albert I, Oome S, Raaymakers TM, Van den Ackerveken G, Nürnberger T. **2014**. A Conserved Peptide Pattern from a Widespread Microbial Virulence Factor Triggers Pattern-Induced Immunity in Arabidopsis. *PLoS Pathogens* **10**: e1004491.

Bolger AM, Lohse M, Usadel B. **2014**. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*: btu170–.

Bölker M, Basse CW, Schirawski J. **2008**. Ustilago maydis secondary metabolism-from genomics to biochemistry. *Fungal genetics and biology* **45 Suppl 1**: S88–93.

Bölker M, Genin S, Lehmler C, Kahmann R. **1995**. Genetic regulation of mating and dimorphism in Ustilago maydis. **73**: 6.

Bölker M, Urban M, Kahmann R. **1992**. The a mating type locus of U. maydis specifies cell signaling components. *Cell* **68**: 441–50.

Boyce KJ, Chang H, Souza CAD, Kronstad JW. **2005**. An Ustilago maydis Septin Is Required for

Filamentous Growth in Culture and for Full Symptom Development on Maize An Ustilago maydis Septin Is Required for Filamentous Growth in Culture and for Full Symptom Development on Maize. *Society* **4**: 2044–2056.

Brachmann A, Weinzierl G, Kämper J, Kahmann R. **2001**. Identification of genes in the bW/bE regulatory cascade in Ustilago maydis. *Molecular microbiology* **42**: 1047–63.

Brefort T, Doehlemann G, Mendoza-Mendoza A, Reissmann S, Djamei A, Kahmann R. **2009**. Ustilago maydis as a Pathogen. *Annual review of phytopathology* **47**: 423–45.

Brefort T, Tanaka S, Neidig N, Doehlemann G, Vincon V, Kahmann R. **2014**. Characterization of the Largest Effector Gene Cluster of Ustilago maydis. (J-R Xu, Ed.). *PLoS pathogens* **10**: e1003866.

Brooks DM, Hernández-Guzmán G, Kloek AP, Alarcón-Chaidez F, Sreedharan A, Rangaswamy V, Peñaloza-Vázquez A, Bender CL, Kunkel BN. **2004**. Identification and characterization of a well-defined series of coronatine biosynthetic mutants of Pseudomonas syringae pv. tomato DC3000. *Molecular plant-microbe interactions : MPMI* **17**: 162–174.

Brown a J, Casselton L a. **2001**. Mating in mushrooms: increasing the chances but prolonging the affair. *Trends in genetics : TIG* **17**: 393–400.

Bulgarelli D, Rott M, Schlaeppi K, Ver Loren van Themaat E, Ahmadinejad N, Assenza F, Rauf P, Huettel B, Reinhardt R, Schmelzer E, *et al.* 2012. Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. *Nature* **488**: 91–95.

Cabral A, Oome S, Sander N, Küfner I, Nürnberger T, Van den Ackerveken G. **2012**. Nontoxic Nep1-Like Proteins of the Downy Mildew Pathogen Hyaloperonospora arabidopsidis: Repression of Necrosis-Inducing Activity by a Surface-Exposed Region. *Molecular Plant-Microbe Interactions* **25**: 697–708.

Caillaud MC, Asai S, Rallapalli G, Piquerez S, Fabro G, Jones JDG. **2013**. A Downy Mildew Effector Attenuates Salicylic Acid-Triggered Immunity in Arabidopsis by Interacting with the Host Mediator Complex. *PLoS Biology* **11**.

Campbell MS, Holt C, Moore B, Yandell M. **2014**. *Genome Annotation and Curation Using MAKER and MAKER-P*.

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. **2009**. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic acids research* **37**: D233–8.

Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado A, Yandell M. **2008**. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome research* **18**: 188–96.

Cantu D, Segovia V, MacLean D, Bayles R, Chen X, Kamoun S, Dubcovsky J, Saunders DG, Uauy C. **2013**. Genome analyses of the wheat yellow (stripe) rust pathogen Puccinia striiformis f. sp. tritici reveal polymorphic and haustorial expressed secreted proteins as candidate effectors. *BMC Genomics* **14**: 270.

Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. **2012**. Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* **28**: 464–469.

Casselton LA, Olesnicky NS. **1998**. Molecular genetics of mating recognition in basidiomycete fungi. *Microbiology and molecular biology reviews : MMBR* **62**: 55–70.

Christensen JJ. **1963**. *Corn smut caused by Ustilago maydis*. St. Paul, Minnesota, USA: APS PRESS.

Conforto C, Cazón I, Fernández FD, Marinelli A, Oddino C, Rago AM. **2013**. Molecular sequence data of Thecaphora frezii affecting peanut crops in Argentina. *European Journal of Plant Pathology* **137**: 663–666.

Cottier F, Leewattanapasuk W, Kemp LR, Murphy M, Supuran CT, Kurzai O, Mühlschlegel FA. **2013**. Carbonic anhydrase regulation and CO2 sensing in the fungal pathogen Candida glabrata involves a novel Rca1p ortholog. *Bioorganic and Medicinal Chemistry* **21**: 1549–1554.

Cui H, Wang Y, Xue L, Chu J, Yan C, Fu J, Chen M, Innes RW, Zhou JM. **2010**. Pseudomonas syringae Effector Protein AvrB Perturbs Arabidopsis Hormone Signaling by Activating MAP Kinase 4. *Cell Host and Microbe* **7**: 164–175.

Dangl JL, Jones JDG. **2001**. Plant pathogens and integrated defence responses to infection. *Nature* **411**.

Daub ME, Briggs SP. **1983**. Changes in tobacco cell membrane composition and structure caused by cercosporin. *Plant physiology* **71**: 763–766.

DebRoy S, Thilmony R, Kwack Y-B, Nomura K, He S. **2004**. A family of conserved bacterial effectors inhibits salicylic acid-mediated basal immunity and promotes disease necrosis in plants. *Proceedings of the National Academy of Sciences* **101**: 9927–9932.

Denchev CM, YoungJoon C, SeungKyu L, HyeonDong S. **2007**. New records of smut fungi from Korea. 1. *Mycotaxon* **100**: 67–72.

Dietrich FS, Voegeli S, Brachat S, Lerch A, Gates K, Steiner S, Mohr C, Pöhlmann R, Luedi P, Choi S, *et al.* 2004. The Ashbya gossypii genome as a tool for mapping the ancient Saccharomyces cerevisiae genome. *Science (New York, N.Y.)* **304**: 304–7.

Djamei A, Schipper K, Rabe F, Ghosh A, Vincon V, Kahnt J, Osorio S, Tohge T, Fernie AR, Feussner I, *et al.* 2011. Metabolic priming by a secreted fungal effector. *Nature* **478**: 395–8.

Doehlemann G, van der Linde K, Assmann D, Schwammbach D, Hof A, Mohanty A, Jackson D, Kahmann R. **2009**. Pep1, a secreted effector protein of Ustilago maydis, is required for successful invasion of plant cells. *PLoS pathogens* **5**: e1000290.

Dong S, Kong G, Qutob D, Yu X, Tang J, Kang J, Dai T, Wang H, Gijzen M, Wang Y. **2012**. The NLP Toxin Family in Phytophthora sojae Includes Rapidly Evolving Groups That Lack Necrosis-Inducing Activity. *Molecular Plant-Microbe Interactions* **25**: 896–909.

Earley KW, Haag JR, Pontes O, Opper K, Juehne T, Song K, Pikaard CS. **2006**. Gateway-compatible

vectors for plant functional genomics and proteomics. *Plant Journal* **45**: 616–629.

Eastmond PJ, Van Dijken AJH, Spielman M, Kerr A, Tissier AF, Dickinson HG, Jones JDG, Smeekens SC, Graham IA. **2002**. Trehalose-6-phosphate synthase 1, which catalyses the first step in trehalose synthesis, is essential for Arabidopsis embryo maturation. *Plant Journal* **29**: 225–235.

Eilbeck K, Moore B, Holt C, Yandell M. **2009**. Quantitative measures for the management and comparison of annotated genomes. *BMC Bioinformatics* **10**: 67.

Elleuche S, Pöggeler S. **2009a**. Evolution of carbonic anhydrases in fungi. *Current Genetics* **55**: 211–222.

Elleuche S, Pöggeler S. **2009b**. β-carbonic anhydrases play a role in fruiting body development and ascospore germination in the filamentous fungus Sordaria macrospora. *PLoS ONE* **4**: e5177.

Emms DM, Kelly S. **2015**. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* **16**: 157.

English AC, Richards S, Han Y, Wang M, Vee V, Qu J, Qin X, Muzny DM, Reid JG, Worley KC, *et al.* 2012. Mind the Gap: Upgrading Genomes with Pacific Biosciences RS Long-Read Sequencing Technology. *PLoS ONE* **7**: 1–12.

Evelin H, Kapoor R, Giri B. **2009**. Arbuscular mycorrhizal fungi in alleviation of salt stress: A review. *Annals of Botany* **104**: 1263–1280.

Faino L, Seidl MF, Datema E, van den Berg GCM, Janssen A, Wittenberg AHJ, Thomma BPHJ. **2015**. Single-Molecule Real-Time Sequencing Combined with Optical Mapping Yields Completely Finished Fungal Genome. *mBio* **6**: e00936–15.

Faino L, Thomma BPHJ. **2014**. Get your high-quality low-cost genome sequence. *Trends in plant science* **19**: 288–291.

Feldbrügge M, Kämper J, Steinberg G, Kahmann R. **2004**. Regulation of mating and pathogenic development in Ustilago maydis. *Current opinion in microbiology* **7**: 666–72.

Felsenstein J. **1989**. PHYLIP-phylogeny inference package (version 3.2). *Cladistics*.

Finn RD, Clements J, Eddy SR. **2011**. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research* **39**: W29–W37.

Fisch KM, Gillaspy a. F, Gipson M, Henrikson JC, Hoover a. R, Jackson L, Najar FZ, Wägele H, Cichewicz RH. **2009**. Chemical induction of silent biosynthetic pathway transcription in Aspergillus niger. *Journal of Industrial Microbiology & Biotechnology* **36**: 1199–1213.

Flor-parra I, Vranes M. **2006**. Biz1 , a Zinc Finger Protein Required for Plant Invasion by Ustilago maydis , Regulates the Levels of a Mitotic Cyclin. *The Plant cell* **18**: 2369–2387.

Frantisek B. **1916**. Die Pilze Böhmens. II. Teil. Brandpilze (Hemibasidii).14.

Freeman S, Rodriguez RJ. **1993**. Genetic Conversion of a Fungal Plant Pathogen to a Nonpathogenic,

Endophytic Mutualist. *Science* **260**: 75–78.

Gafni A, Calderon CE, Harris R, Buxdorf K, Dafa-Berger A, Zeilinger-Reichert E, Levy M. **2015**. Biological control of the cucurbit powdery mildew pathogen Podosphaera xanthii by means of the epiphytic fungus Pseudozyma aphidis and parasitism as a mode of action. *Frontiers in Plant Science* **6**: 1–11.

Gan P, Ikeda K, Irieda H, Narusaka M, O'Connell RJ, Narusaka Y, Takano Y, Kubo Y, Shirasu K. **2013**. Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of Colletotrichum fungi. *New Phytologist* **197**: 1236–1249.

García E, Alonso Á, Platas G, Sacristán S. **2012**. The endophytic mycobiota of Arabidopsis thaliana. *Fungal Diversity* **60**: 71–89.

Garre V, Tenberge KB, Eising R. **1998**. Secretion of a Fungal Extracellular Catalase by Claviceps purpurea During Infection of Rye: Putative Role in Pathogenicity and Suppression of Host Defense. *Phytopathology* **88**: 744–753.

Geiser E, Przybilla SK, Friedrich A, Buckel W, Wierckx N, Blank LM, Bölker M. **2016**. Ustilago maydis produces itaconic acid via the unusual intermediate trans-aconitate. *Microbial Biotechnology* **9**: 116–126.

Ghareeb H, Becker A, Iven T, Feussner I, Schirawski J. **2011**. Sporisorium reilianum infection changes inflorescence and branching architectures of maize. *Plant physiology* **156**: 2037–52.

Gilchrist DG. **1976**. Production and Nature of a Host-Specific Toxin from Alternaria alternata f. sp. lycopersici. *Phytopathology* **66**: 165.

Gillissen B, Bergemann J, Sandmann C, Schroeer B, Bölker M, Kahmann R. **1992**. A two-component regulatory system for self/non-self recognition in Ustilago maydis. *Cell* **68**: 647–657.

Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, Pretty J, Robinson S, Thomas SM, Toulmin C. **2010**. Food Security: The Challenge of Feeding 9 Billion People. *Science* **327**: 812–818.

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

Grewal TS, Rossnagel BG, Bakkeren G, Scoles GJ. **2008**. Identification of resistance genes to barley covered smut and mapping of the Ruh1 gene using Ustilago hordei strains with defined avirulence genes. *Canadian Journal of Plant Pathology* **30**: 277–284.

Guyon K, Balagué C, Roby D, Raffaele S. **2014**. Secretome analysis reveals effector candidates associated with broad host range necrotrophy in the fungal plant pathogen Sclerotinia sclerotiorum. *BMC Genomics* **15**: 336.

Hane JK, Rouxel T, Howlett BJ, Kema GH, Goodwin SB, Oliver RP. **2011**. A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi. *Genome Biology* **12**: R45.

Harman GE, Howell CR, Viterbo A, Chet I, Lorito M. **2004**. Trichoderma species--opportunistic, avirulent plant symbionts. *Nature reviews. Microbiology* **2**: 43–56.

Heimel K, Scherer M, Vranes M, Wahl R, Pothiratana C, Schuler D, Vincon V, Finkernagel F, Flor-Parra I, Kämper J. **2010**. The transcription factor rbf1 is the master regulator for b-mating type controlled pathogenic development in Ustilago maydis. *PLoS Pathogens* **6**: 17–18.

Heller J, Tudzynski P. **2011**. Reactive oxygen species in phytopathogenic fungi: signaling, development, and disease. *Annual review of phytopathology* **49**: 369–390.

Hemetsberger C, Herrberger C, Zechmann B, Hillmer M, Doehlemann G. **2012**. The Ustilago maydis effector Pep1 suppresses plant immunity by inhibition of host peroxidase activity. *PLoS Pathogens* **8**.

Hemetsberger C, Mueller N, Matei A, Herrberger C, Doehlemann G. **2015**. The fungal core effector Pep1 is conserved across smuts of dicots and monocots. *New Phytologist*.

Hewald S, Josephs K, Bölker M, Bo M. **2005**. Genetic Analysis of Biosurfactant Production in Ustilago maydis Genetic Analysis of Biosurfactant Production in Ustilago maydis. *Applied and environmental microbiology* **71**: 3033–3040.

Hibbett DS, Stajich J, Spatafora JW. **2013**. Toward genome-enabled mycology. *Mycologia*: 13–196–.

Hiruma K, Gerlach N, Sacristán S, Nakano RT, Hacquard S, Kracher B, Neumann U, Ramírez D, Bucher M, O'Connell RJ, *et al.* 2016. Root Endophyte Colletotrichum tofieldiae Confers Plant Fitness Benefits that Are Phosphate Status Dependent. *Cell*: 1–11.

Holliday R. **1961a**. The genetics of Ustilago maydis. *Genet. Res* **2**: 204–230.

Holliday R. **1961b**. Induced mitotic crossing-over in Ustilago maydis. *Genet. Res. Camb.* **2**: 231–248.

Holliday R. **1974**. Ustilago maydis. In Handbook of genetics, R. C. King ed.575–595.

Holt C, Yandell M. **2011**. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**: 491.

van der Hoorn RAL, Kamoun S. **2008**. From Guard to Decoy: a new model for perception of plant pathogen effectors. *The Plant cell* **20**: 2009–17.

Horton MW, Bodenhausen N, Beilsmith K, Meng D, Muegge BD, Subramanian S, Vetter MM, Vilhjálmsson BJ, Nordborg M, Gordon JI, *et al.* 2014. Genome-wide association study of Arabidopsis thaliana leaf microbial community. *Nature Communications* **5**: 5320.

Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K. **2007**. WoLF PSORT: Protein localization predictor. *Nucleic Acids Research* **35**: 585–587.

Ingold CT. **1983**. The basidium in Ustilago. *Transactions of the British Mycological Society* **81**: 573–584.

Ingold CT. **1987**. Aerial sporidia of Ustilago hypodytes and of Sorosporium saponariae. *Transactions of the British Mycological Society* **89**: 471–475.

Ingold CT. **1988**. Teliospore germination in some species of Ustilaginaceae. *Transactions of the British Mycological Society* **91**: 505–514.

Ingold CT. **1989**. Basidium development in some species of Ustilago. *Mycological Research* **93**: 405–412.

Ingold CT. **1997**. The basidium of Tilletia and its evolution. *Mycologist* **11**: 10–12.

Ingold CT. **1998**. The basidium in Ustilaginaceae. *Mycologist* **12**.

Jones JDG, Dangl JL. **2006**. The plant immune system. *Nature* **444**: 323–9.

De Jonge R. **2012**. In silico identification and characterization of effector catalogs. *Methods in molecular biology (Clifton, N.J.)* **835**: 75–96.

Jurka J, Kapitonov V V, Pavlicek a, Klonowski P, Kohany O, Walichiewicz J. **2005**. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and genome research* **110**: 462–7.

Käll L, Krogh A, Sonnhammer ELL. **2004**. A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology* **338**: 1027–1036.

Käll L, Krogh A, Sonnhammer ELL. **2007**. Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server. *Nucleic Acids Research* **35**: 429–432.

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

Kämper J, Reichmann M, Romeis T, Bölker M, Kahmann R. **1995**. Multiallelic recognition: Nonself-dependent dimerization of the bE and bW homeodomain proteins in ustilago maydis. *Cell* **81**: 73–83.

Katoh K, Misawa K, Kuma K, Miyata T. **2002**. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research* **30**: 3059–3066.

Kellner R, Vollmeister E, Feldbrügge M, Begerow D. **2011**. Interspecific sex in grass smuts and the genetic diversity of their pheromone-receptor system. *PLoS genetics* **7**: e1002436.

Kersey PJ, Allen JE, Armean I, Boddu S, Bolt BJ, Carvalho-Silva D, Christensen M, Davis P, Falin LJ, Grabmueller C, *et al.* 2015. Ensembl Genomes 2016: more genomes, more complexity. *Nucleic Acids Research* **44**: 574–580.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. **2013**. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* **14**: R36.

Kitagawa K, Skowyra D, Elledge SJ, Harper JW, Hieter P. **1999**. SGT1 encodes an essential component of the yeast kinetochore assembly pathway and a novel subunit of the SCF ubiquitin ligase complex. *Molecular Cell* **4**: 21–33.

Kleemann J, Rincon-Rivera LJ, Takahara H, Neumann U, van Themaat EVL, van der Does HC, Hacquard S, St??ber K, Will I, Schmalenbach W, *et al.* 2012. Sequential delivery of host-induced virulence effectors by appressoria and intracellular hyphae of the phytopathogen colletotrichum higginsianum. *PLoS Pathogens* **8**.

Kloppholz S, Kuhn H, Requena N. **2011**. A secreted fungal effector of glomus intraradices promotes symbiotic biotrophy. *Current Biology* **21**: 1204–1209.

Kmiec EB, Holloman WK. **1986**. Homologous pairing of DNA molecules by ustilago rec1 protein is promoted by sequences of Z-DNA. *Cell* **44**: 545–554.

Knox RE, Campbell HL, Clarke FR, Menzies JG, Popovic Z, Procunier JD, Clarke JM, DePauw RM, Cuthbert RD, Somers DJ. **2014**. Quantitative trait loci for resistance in wheat (Triticum aestivum) to Ustilago tritici. *Canadian Journal of Plant Pathology* **36**: 187–201.

Kokontis J, Ruddat M. **1986**. Promotion of hyphal growth in Ustilago violacea by host factors from Silene alba. *Archives of Microbiology*: 302–306.

Kono Y, Daly JM. **1979**. Characterization of the host-specific pathotoxin produced by Helminthosporium maydis, race T, affecting corn with Texas male sterile cytoplasm. *Bioorganic Chemistry* **8**: 391–397.

Kothe E. **1996**. Tetrapolar fungal mating types: Sexes by the thousands. *FEMS Microbiology Reviews* **18**: 65–87.

Kronstad JW. **1997**. Pathogenesis and Sexual Development of the Smut Fungi. In: Stacey G,, In: Keen NT, eds. Plant-Microbe Interactions. Boston, MA: Springer US, 141–186.

Kruzel EK, Hull CM. **2010**. Establishing an unusual cell type: how to make a dikaryon. *Current opinion in microbiology* **13**: 706–11.

Kubicek CP, Starr TL, Glass NL. **2014**. Plant cell wall-degrading enzymes and their secretion in plant-pathogenic fungi. *Annual review of phytopathology* **52**: 427–51.

Kulakovskaya T V., Shashkov AS, Kulakovskaya E V., Golubev WI. **2005**. Ustilagic acid secretion by Pseudozyma fusiformata strains. *FEMS Yeast Research* **5**: 919–923.

Kuo H-C, Hui S, Choi J, Asiegbu FO, Valkonen JPT, Lee Y-H. **2014**. Secret lifestyles of Neurospora crassa. *Scientific reports* **4**: 5135.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. **2004**. Versatile and open software for comparing large genomes. *Genome biology* **5**: R12.

Lanver D, Berndt P, Tollot M, Naik V, Vranes M, Warmann T, Münch K, Rössel N, Kahmann R. **2014**. Plant Surface Cues Prime Ustilago maydis for Biotrophic Development. (NI Naqvi, Ed.). *PLoS Pathogens* **10**: e1004272.

Laurie JD, 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. *The Plant cell* **24**: 1–14.

Le SQ, Gascuel O. **2008**. An improved general amino acid replacement matrix. *Molecular Biology and Evolution* **25**: 1307–1320.

Leaper S, Resnick MA, Holliday R. **1980**. Repair of double-strand breaks and lethal damage in DNA of Ustilago maydis. *Genetical Research* **35**: 291.

Lefebvre F, Joly DL, Labbé C, Teichmann B, Linning R, Belzile F, Bakkeren G, Bélanger RR. **2013**. The transition from a phytopathogenic smut ancestor to an anamorphic biocontrol agent deciphered by comparative whole-genome analysis. *The Plant cell* **25**: 1946–59.

Lehneck R, Pöggeler S. **2014**. A matter of structure: structural comparison of fungal carbonic anhydrases. *Applied Microbiology and Biotechnology*: 8433–8441.

Lemieux RU, Thorn JA, Brice C, Haskins RH. **1951**. BIOCHEMISTRY OF THE USTILAGINALES: II. ISOLATION AND PARTIAL CHARACTERIZATION OF USTILAGIC ACID. *Canadian Journal of Chemistry* **29**: 409–414.

Leuzinger K, Dent M, Hurtado J, Stahnke J, Lai H, Zhou X, Chen Q. **2013**. Efficient Agroinfiltration of Plants for High-level Transient Expression of Recombinant Proteins. *Journal of Visualized Experiments*.

Lévesque CA, Brouwer H, Cano L, Hamilton JP, Holt C, Huitema E, Raffaele S, Robideau GP, 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.

Levine A, Tenhaken R, Dixon R, Lamb C. **1994**. H2O2 from the oxidative burst orchestrates the plant hypersensitive disease resistance response. *Cell* **79**: 583–593.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. **2009**. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Little CR, Perumal R, Tesso T, Prom LK, Odvody GN, Magill CW. **2012**. Sorghum Pathology and Biotechnology - A Fungal Disease Perspective : Part I . Grain Mold , Head Smut , and Ergot. *European Journal of Plant Sciences and Biotechnology* **6**: 10–30.

Liu Z, Liu Y, Shi L, Yang S, Shen L, Yu H, Wang R, Wen J, Tang Q, Hussain A, *et al.* 2016. SGT1 is required in PcINF1/SRC2-1 induced pepper defense response by interacting with SRC2-1. *Scientific Reports* **6**: 21651.

Lobell DB, Schlenker W, Costa-Roberts J. **2011**. Climate Trends and Global Crop Production Since 1980. *Science* **333**: 616–620.

Long SP, Ainsworth EA, Leakey ADB, Ort DR, No J. **2006**. Food for Thought : Lower-Than-Expected crop yield stimulation with rising CO2 concentrations. *Science* **312**: 1918–1921.

Lovely C Ben, Aulakh KB, Perlin MH. **2011**. Role of Hsl7 in morphology and pathogenicity and its interaction with other signaling components in the plant pathogen Ustilago maydis. *Eukaryotic cell* **10**: 869–83.

Lubadde G, Tongoona P, Derera J, Sibiya J. **2014**. Major pearl millet diseases and their effects on-farm grain yield in Uganda. *African Journal of Agricultural Research* **9**: 2911–2918.

Lucas JA, Hawkins NJ, Fraaije BA. **2015**. *The Evolution of Fungicide Resistance*. Elsevier.

Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, Malfatti S, Tremblay J, Engelbrektson A, Kunin V, del Rio TG, *et al.* 2012. Defining the core Arabidopsis thaliana root microbiome. *Nature* **488**: 86–90.

Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro A, Dufresne M, Freitag M, Grabherr M, Henrissat B, *et al.* 2010. Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. *Nature* **464**: 367–73.

Ma Z, Song T, Zhu L, Ye W, Wang Y, Shao Y, Dong S, Zhang Z, Dou D, Zheng X, *et al.* 2015. A Phytophthora sojae Glycoside Hydrolase 12 Protein Is a Major Virulence Factor during Soybean Infection and Is Recognized as a PAMP. *The Plant cell* **27**: 2057–72.

Malik MMS, Batts CCV. **1960**. The development of loose smut (Ustilago nuda) in the barley plant, with observations on spore formation in nature and in culture. *Transactions of the British Mycological Society* **43**: 126–IN9.

Mattinen L, Tshuikina M, Pirhonen M. **2004**. Identification and Characterization of Nip, Necrosis-Inducing Virulence Protein of Erwinia carotovora subsp. carotovora. *Molecular plant-microbe interactions : MPMI* **17**: 1366–1375.

Medema MH, Blin K, Cimermancic P, De Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R. **2011**. AntiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Research* **39**: 339–346.

Menardo F, Praz C, Wyder S, Bourras S. A, McNally KE, Parlange F, Riba A, Roffler S, Schaefer L, Shimizu KK, *et al.* 2015. Hybridization of powdery mildew strains gives raise to pathogens on novel agricultural crop species. *Nature genetics*: 1–24.

Mendoza-Mendoza A, Berndt P, Djamei A, Weise C, Linne U, Marahiel M, Vranes M, Kämper J, Kahmann R, Vraneš M, *et al.* 2009. Physical-chemical plant-derived signals induce differentiation in Ustilago maydis. *Molecular microbiology* **71**: 895–911.

Mohnen D. **2008**. Pectin structure and biosynthesis. *Current Opinion in Plant Biology* **11**: 266–277.

Molina L, Kahmann R. **2007**. An Ustilago maydis Gene Involved in H2O2 Detoxification Is Required for Virulence. *THE PLANT CELL ONLINE* **19**: 2293–2309.

Molinier J, Ries G, Zipfel C, Hohn B. **2006**. Transgeneration memory of stress in plants. *Nature* **442**: 1046–1049.

Morita T, Koike H, Hagiwara H, Ito E, Machida M, Sato S, Habe H, Kitamoto D. **2014**. Genome and transcriptome analysis of the basidiomycetous yeast Pseudozyma antarctica producing extracellular glycolipids, mannosylerythritol lipids. *PLoS ONE* **9**: 1–12.

Motteram J, Küfner I, Deller S, Brunner F, Hammond-Kosack KE, Nürnberger T, Rudd JJ. **2009**. Molecular characterization and functional analysis of MgNLP, the sole NPP1 domain-containing protein, from the fungal wheat leaf pathogen Mycosphaerella graminicola. *Molecular plant-microbe*

interactions : MPMI **22**: 790–799.

Mueller O, Kahmann R, Aguilar G, Trejo-Aguilar B, Wu A, de Vries RP. **2008**. The secretome of the maize pathogen Ustilago maydis. *Fungal Genetics and Biology* **45**.

Mueller AN, Ziemann S, Treitschke S, Aßmann D, Doehlemann G. **2013**. Compatibility in the Ustilago maydis-Maize Interaction Requires Inhibition of Host Cysteine Proteases by the Fungal Effector Pit2. *PLoS Pathogens* **9**.

Myers EW. **2000**. A Whole-Genome Assembly of Drosophila. *Science* **287**: 2196–2204.

Nemri A, Saunders DGO, Anderson C, Upadhyaya NM, Win J, Lawrence GJ, Jones D a, Kamoun S, Ellis JG, Dodds PN. **2014**. The genome sequence and effector complement of the flax rust pathogen Melampsora lini. *Frontiers in plant science* **5**: 98.

Nesi N, Jond C, Debeaujon I, Caboche M, Lepiniec L. **2001**. The Arabidopsis TT2 gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. *The Plant cell* **13**: 2099–2114.

Neveu B, Labbé C, Bélanger RR. **2007**. GFP technology for the study of biocontrol agents in tritrophic interactions: A case study with Pseudozyma flocculosa. *Journal of Microbiological Methods* **68**: 275–281.

Newcombe G, Stirling B, McDonald S, Bradshaw HD. **2000**. Melampsora × columbiana, a natural hybrid of M. medusae and M. occidentalis. *Mycological Research* **104**: 261–274.

Ngugi HK, King SB, Abayo GO, Reddy YVR. **2002**. Prevalence, Incidence, and Severity of Sorghum Diseases in Western Kenya. *Plant Disease* **86**: 65–70.

Nguyen Ba AN, Pogoutse A, Provart N, Moses AM. **2009**. NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. *BMC bioinformatics* **10**: 202.

O'Connell RJ, Thon MR, Hacquard S, Amyotte SG, Kleemann J, Torres MF, 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. *Nature genetics* **44**: 1060–5.

O'Donnell KL, McLaughlin DJ. **1984**. Ultrastructure of Meiosis in Ustilago maydis. *Mycologia* **76**: 468.

Oerke E-C. **2006**. Crop losses to pests. *The Journal of Agricultural Science* **144**: 31.

Oerke EC, Dehne HW. **2004**. Safeguarding production - Losses in major crops and the role of crop protection. *Crop Protection* **23**: 275–285.

El Oirdi M, Bouarab K. **2007**. Plant signalling components EDS1 and SGT1 enhance disease caused by the necrotrophic pathogen Botrytis cinerea. *New Phytologist* **175**: 131–139.

Oliveira JV de C, Borges TA, dos Santos RAC, Freitas LFD, Rosa CA, Goldman GH, Ria??o-Pach??n DM. **2014**. Pseudozyma brasiliensis sp. nov., a xylanolytic, ustilaginomycetous yeast species isolated from an insect pest of sugarcane roots. *International Journal of Systematic and Evolutionary Microbiology* **64**: 2159–2168.

Omacini M, Chaneton EJ, Ghersa CM, Muller CB. **2001**. Symbiotic fungal endophytes control insect host-parasite interaction webs. *Nature* **409**: 78–81.

Oome S, Van den Ackerveken G. **2014**. Comparative and functional analysis of the widely occurring family of Nep1-like proteins. *Molecular plant-microbe interactions : MPMI* **27**: 1–51.

Ottmann C, Luberacki B, Küfner I, Koch W, Brunner F, Weyand M, Mattinen L, Pirhonen M, Anderluh G, Seitz HU, *et al.* 2009. A common toxin fold mediates microbial attack and plant defense. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 10359–10364.

Palmer JM, Keller NP. **2010**. Secondary metabolism in fungi: Does chromosomal location matter? *Current Opinion in Microbiology* **13**: 431–436.

Parlak Y. **1981**. Seed-borne Pathogens on Wheat (Particularly Smuts) in Turkey 1. *EPPO Bulletin* **11**: 83–86.

Parlevliet J. **2002**. Durability of resistance against fungal, bacterial and viral pathogens; present situation. *Euphytica* **124**: 147–156.

Parra G, Bradnam K, Korf I. **2007**. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics (Oxford, England)* **23**: 1061–7.

Peart JR, Lu R, Sadanandom A, Malcuit I, Moffett P, Brice DC, Schauser L, Jaggard D a W, Xiao S, Coleman MJ, *et al.* 2002. Ubiquitin ligase-associated protein SGT1 is required for host and nonhost disease resistance in plants. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 10865–10869.

Pedro H, Maheswari U, Urban M, Irvine AG, Cuzick A, McDowall MD, Staines DM, Kulesha E, Hammond-Kosack KE, Kersey PJ. **2015**. PhytoPath: an integrative resource for plant pathogen genomics. *Nucleic acids research* **44**: gkv1052.

Pellegrin C, Morin E, Martin FM, Veneault-Fourrey C. **2015**. Comparative analysis of secretomes from ectomycorrhizal fungi with an emphasis on small-secreted proteins. *Frontiers in Microbiology* **6**: 1–15.

Peng Y, Leung HCM, Yiu SM, Chin FYL. **2012**. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics (Oxford, England)* **28**: 1420–8.

Perkins DD, Turner BC. **1988**. Neurospora from natural populations: Toward the Population Biology of a Hapoid Eukaryote. *Experimental Mycology* **12**: 91–131.

Petersen TN, Brunak S, von Heijne G, Nielsen H. **2011**. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods* **8**: 785–6.

Petutschnig EK, Jones AME, Serazetdinova L, Lipka U, Lipka V. **2010**. The Lysin Motif Receptor-like Kinase (LysM-RLK) CERK1 Is a Major Chitin-binding Protein in Arabidopsis thaliana and Subject to Chitin-induced Phosphorylation. *Journal of Biological Chemistry* **285**: 28902–28911.

Philippsen P, Kaufmann A, Schmitz HP. **2005**. Homologues of yeast polarity genes control the development of multinucleated hyphae in Ashbya gossypii. *Current Opinion in Microbiology* **8**: 370–377.

Plett JM, Kemppainen M, Kale SD, Kohler A, Legu?? V, Brun A, Tyler BM, Pardo AG, Martin F. **2011**. A secreted effector protein of laccaria bicolor is required for symbiosis development. *Current Biology* **21**: 1197–1203.

Pothiratana C. **2007**. Functional characterization of the homeodomain transcription factor Hdp1 in Ustilago maydis.

Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, Zuccaro A, Reissmann S, Kahmann R. **2015**. Fungal Effectors and Plant Susceptibility. *Annual Review of Plant Biology* **66**: 513–545.

Puhalla JE. **1968**. COMPATIBILITY REACTIONS ON SOLID MEDIUM AND INTERSTRAIN INHIBITION IN USTILAGO MAYDIS. *Genetics* **60**.

Quadbeck-Seeger C, Wanner G, Huber S, Kahmann R, Kämper J. **2000**. A protein with similarity to the human retinoblastoma binding protein 2 acts specifically as a repressor for genes regulated by the b mating type locus in Ustilago maydis. *Molecular Microbiology* **38**: 154–166.

Que Y, Xu L, Wu Q, Liu Y, Ling H, Liu Y, Zhang Y, Guo J, Su Y, Chen J, *et al.* 2014. Genome sequencing of Sporisorium scitamineum provides insights into the pathogenic mechanisms of sugarcane smut. *BMC genomics* **15**: 996.

Qutob D, Kamoun S, Gijzen M. **2002**. Expression of a Phytophthora sojae necrosis-inducing protein occurs during transition from biotrophy to necrotrophy. *Plant Journal* **32**: 361–373.

Raeder U, Broda P. **1985**. Rapid preparation of DNA from filamentous fungi. *Letters in Applied Microbiology* **1**: 17–20.

Raffaele S, Kamoun S. **2012**. Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature reviews. Microbiology* **10**: 417–30.

Raffaele S, Win J, Cano LM, 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.

Redkar A, Hoser R, Schilling L, Zechmann B, Krzymowska M, Walbot V, Doehlemann G. **2015a**. A Secreted Effector Protein of Ustilago maydis Guides Maize Leaf Cells to Form Tumors. *The Plant cell* **27**: 1332–51.

Redkar A, Villajuana- Bonequi M, Doehlemann G. **2015b**. Conservation of the *Ustilago maydis* effector See1 in related smuts. *Plant Signaling & Behavior* **10**: e1086855.

Redman RS, Dunigan DD, Rodriguez RJ. **2001**. Fungal symbiosis from mutualism to parasitism: Who controls the outcome, host or invader? *New Phytologist* **151**: 705–716.

Redman RS, Ranson JC, Rodriguez RJ. **1999**. Conversion of the Pathogenic Fungus to a Nonpathogenic, Endophytic Mutualist by Gene Disruption. *Molecular Plant-Microbe Interactions* **12**: 969–975.

Roelants SLKW, De Maeseneire SL, Ciesielska K, Van Bogaert INA, Soetaert W. **2014**. Biosurfactant gene clusters in eukaryotes: Regulation and biotechnological potential. *Applied Microbiology and*

Biotechnology **98**: 3449–3461.

Roets F, Dreyer LL, Wingfield MJ, Begerow D. **2008**. Thecaphora capensis sp. nov., an unusual new anther smut on Oxalis in South Africa. *Persoonia* **21**: 147–52.

Romeis T, Brachmann A, Kahmann R, Kämper J. **2000**. Identification of a target gene for the bE-bW homeodomain protein complex in Ustilago maydis. *Molecular microbiology* **37**: 54–66.

Rudd J, Kanyuka K, Hassani-Pak K, Derbyshire M, Andongabo A, Devonshire J, Lysenko A, Saqi M, Desai N, Powers S, *et al.* 2015. Transcriptome and metabolite profiling the infection cycle of Zymoseptoria tritici on wheat (Triticum aestivum) reveals a biphasic interaction with plant immunity involving differential pathogen chromosomal contributions, and a variation on the hemibiotro. *Plant Physiology* **167**: pp.114.255927.

Sambrook J, Fritsch EF, Maniatis T. **1989**. *Molecular cloning*. New York: Cold spring harbor laboratory press.

Santhanam P, Esse P Van, Albert I, Faino L, Nürnberger T, Thomma BPHJ. **2012**. Evidence for Functional Diversification within a Fungal NEP1-Like Protein Family. *Molecular Plant-Microbe Interactions* **26**: 121010102104007.

Saunders DGO, Win J, Cano LM, Szabo LJ, Kamoun S, Raffaele S. **2012**. Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. (JE Stajich, Ed.). *PloS one* **7**: e29847.

Scheller HV, Ulvskov P. **2010**. Hemicelluloses. *Annual Review of Plant Biology* **61**: 263–289.

Schirawski J, Heinze B, Wagenknecht M, Kahmann R. **2005**. Mating type loci of Sporisorium reilianum: novel pattern with three a and multiple b specificities. *Eukaryotic cell* **4**: 1317–27.

Schirawski J, Mannhaupt G, Brefort T, Doehlemann G, Stasio M Di, Mendoza-mendoza A, Pester D, Ghareeb H, Wollenberg T, Wong P, *et al.* 2010. Pathogenicity determinants in smut fungi revealed by genome comparison. *Science* **330**: 1546–8.

Schouten A, Van Baarlen P, Van Kan J a L. **2008**. Phytotoxic Nep1-like proteins from the necrotrophic fungus Botrytis cinerea associate with membranes and the nucleus of plant cells. *New Phytologist* **177**: 493–505.

Schouten A, Tenberge KB, Vermeer J, Stewart J, Wagemakers L, Williamson B, Van Kan JAL. **2002**. Functional analysis of an extracellular catalase of Botrytis cinerea. *Molecular Plant Pathology* **3**: 227–238.

Schulz B, Banuett F, Dahl M, Schlesinger R, Schäfer W, Martin T, Herskowitz I, Kahmann R. **1990**. The b alleles of U. maydis, whose combinations program pathogenic development, code for polypeptides containing a homeodomain-related motif. *Cell* **60**: 295–306.

Scott B, Eaton CJ. **2008**. Role of reactive oxygen species in fungal cellular differentiations. *Current Opinion in Microbiology* **11**: 488–493.

Sharma R, Mishra B, Runge F, Thines M. **2014**. Gene Loss Rather Than Gene Gain Is Associated with a Host Jump from Monocots to Dicots in the Smut Fungus Melanopsichium pennsylvanicum. *Genome biology and evolution* **6**: 2034–49.

Simao F a., Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM, Simão F a, Waterhouse RM, Ioannidis P, Kriventseva E V. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **3**: 1–3.

Singh RP, Hodson DP, Huerta-Espino J, Jin Y, Bhavani S, Njau P, Herrera-Foessel S, Singh PK, Singh S, Govindan V. **2011**. The Emergence of Ug99 Races of the Stem Rust Fungus is a Threat to World Wheat Production. *Annual review of phytopathology* **49**: 465–481.

Skamnioti P, Henderson C, Zhang Z, Robinson Z, Gurr SJ. **2007**. A novel role for catalase B in the maintenance of fungal cell-wall integrity during host invasion in the rice blast fungus Magnaporthe grisea. *Molecular plant-microbe interactions : MPMI* **20**: 568–580.

Skibbe DS, Doehlemann G, Fernandes J, Walbot V. **2010**. Maize tumors caused by Ustilago maydis require organ-specific genes in host and pathogen. *Science* **328**: 89–92.

Slaughter A, Daniel X, Flors V, Luna E, Hohn B, Mauch-Mani B. **2012**. Descendants of Primed Arabidopsis Plants Exhibit Resistance to Biotic Stress. *Plant Physiology* **158**: 835–843.

Slaymaker DH, Navarre D a, Clark D, del Pozo O, Martin GB, Klessig DF. **2002**. The tobacco salicylic acid-binding protein 3 (SABP3) is the chloroplast carbonic anhydrase, which exhibits antioxidant activity and plays a role in the hypersensitive defense response. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 11640–11645.

Smith IM, Dunez J, Phillips DH, Lelliott RA, Archer SA. **2009**. *European Handbook of Plant Diseases*.

Smith SE, Smith FA. **2012**. Fresh perspectives on the roles of arbuscular mycorrhizal fungi in plant nutrition and growth. *Mycologia* **104**: 1–13.

Snetselaar K, Mims CW. **1994**. Light and electron microscopy of Ustilago maydis hyphae in maize. *Mycological Research* **98**: 347–355.

Soanes D, Richards TA. **2014**. Horizontal Gene Transfer in Eukaryotic Plant Pathogens. *Annual review of phytopathology* **52**: 583–614.

Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stüber K, Ver Loren van Themaat E, Brown JKM, Butcher SA, Gurr SJ, *et al.* 2010. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science (New York, N.Y.)* **330**: 1543–6.

Spellig T, Bölker M, Lottspeich F, Frank RW, Kahmann R. **1994**. Pheromones trigger filamentous growth in Ustilago maydis. *The EMBO journal* **13**: 1620–7.

Sperschneider J, Dodds PN, Gardiner DM, Manners JM, Singh KB, Taylor JM. **2015a**. Advances and challenges in computational prediction of effectors from plant pathogenic fungi. (DC Sheppard, Ed.). *PLoS pathogens* **11**: e1004806.

Sperschneider J, Williams AH, Hane JK, Singh KB, Taylor JM. **2015b**. Evaluation of Secretion Prediction Highlights Differing Approaches Needed for Oomycete and Fungal Effectors. *Frontiers in Plant Science* **6**: 1–14.

Stajich JE, Berbee ML, Blackwell M, Hibbett DS, James TY, Spatafora JW, Taylor JW. **2009**. The fungi. *Current biology : CB* **19**: R840–5.

Stanke M, Morgenstern B. **2005**. AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research* **33**: 465–467.

Stergiopoulos I, Collemare J, Mehrabi R, De Wit PJGM. **2013**. Phytotoxic secondary metabolites and peptides produced by plant pathogenic Dothideomycete fungi. *FEMS Microbiology Reviews* **37**: 67–93.

Stock J, Sarkari P, Kreibich S, Brefort T, Feldbrügge M, Schipper K. **2012**. Applying unconventional secretion of the endochitinase Cts1 to export heterologous proteins in Ustilago maydis. *Journal of biotechnology*: 1–12.

Stoll M, Begerow D, Oberwinkler F. **2005**. Molecular phylogeny of Ustilago, Sporisorium and related taxa based on combined analysis of rDNA sequences. *Mycological Research* **3**: 342–356.

Stukenbrock EH, Croll D. **2014**. The evolving fungal genome. *Fungal Biology Reviews* **28**: 1–12.

Sundar, AR; Barnabas, EL; Malathi, P; Viswanathan R. **1989**. A Mini-Review on Smut Disease of Sugarcane Caused by Sporisorium scitamineum. *Journal of Chemical Information and Modeling* **53**: 160.

Szabó Z, Tönnis M, Kessler H, Feldbrügge M. **2002**. Structure-function analysis of lipopeptide pheromones from the plant pathogen Ustilago maydis. *Molecular genetics and genomics : MGG* **268**: 362–70.

Tanaka S, Brefort T, Neidig N, Djamei A, Kahnt J, Vermerris W, Koenig S, Feussner K, Feussner I, Kahmann R. **2014**. A secreted Ustilago maydis effector promotes virulence by targeting anthocyanin biosynthesis in maize. *eLife* **2014**: 1–27.

Taniguti LM, Schaker PDC, Benevenuto J, Peters LP, Carvalho G, Palhares A, Quecine MC, Nunes FRS, Kmit MCP, Wai A, *et al.* 2015. Complete Genome Sequence of Sporisorium scitamineum and Biotrophic Interaction Transcriptome with Sugarcane (J-F Pombert, Ed.). *PLOS ONE* **10**: e0129318.

Teichmann B, Labbé C, Lefebvre F, Bölker M, Linne U, Bélanger RR. **2011**. Identification of a biosynthesis gene cluster for flocculosin a cellobiose lipid produced by the biocontrol agent Pseudozyma flocculosa. *Molecular Microbiology* **79**: 1483–1495.

Teichmann B, Linne U, Hewald S, Marahiel M a., Bölker M. **2007**. A biosynthetic gene cluster for a secreted cellobiose lipid with antifungal activity from Ustilago maydis. *Molecular Microbiology* **66**: 525–533.

Teichmann B, Liu L, Schink KO, Bölker M. **2010**. Activation of the ustilagic acid biosynthesis gene cluster in Ustilago maydis by the C2H2 zinc finger transcription factor Rua1. *Applied and environmental microbiology* **76**: 2633–2640.

Tempel S. **2012**. Using and Understanding RepeatMasker. In: Bigot Y, ed. Methods in Molecular Biology. Methods in microbiology vol. 859 Mobile Genetic Elements: Protocols and Genomic Applications. Totowa, NJ: Humana Press, .

Thomma BPHJ, Seidl MF, Shi-Kunne X, Cook DE, Bolton MD, van Kan JAL, Faino L. **2015**. Mind the gap; seven reasons to close fragmented genome assemblies. *Fungal Genetics and Biology* **90**: 24–30.

Thompson NM, Epplin FM, Edwards JT, Hunger RM. **2014**. Economics of foliar fungicides for hard red winter wheat in the USA southern Great Plains. *Crop Protection* **59**: 1–6.

Tsuge T, Harimoto Y, Akimitsu K, Ohtani K, Kodama M, Akagi Y, Egusa M, Yamamoto M, Otani H. **2013**. Host-selective toxins produced by the plant pathogenic fungus Alternaria alternata. *FEMS Microbiology Reviews* **37**: 44–66.

Tubiello FN, Soussana J-F, Howden SM. **2007**. Crop and pasture response to climate change. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 19686–90.

Tyler BM, Sucheta Tripathy, Zhang X, Dehal P, Jiang RHY, Aerts A, Arredondo FD, Baxter L, Bensasson D, Beynon JL, *et al.* 2006. Phytophthora Genome Sequences Uncover Evolutionary Origins and Mechanisms of Pathogenesis. *Science* **313**: 1261–1266.

Urban M, Kahmann R, Bölker M. **1996**. The biallelic a mating type locus of Ustilago maydis: remnants of an additional pheromone gene indicate evolution from a multiallelic ancestor. *Molecular & general genetics : MGG* **250**: 414–20.

Vanky K. **1999**. The new classifacatory system for smut fungi and two new genera. *Mycotaxon* **70**: 35–49.

Vanky K. **2012**. *Smut fungi of the world*. APS PRESS.

Vánky K. **1994**. European smut fungi.

Vánky K, Lutz M. **2007**. Revision of some Thecaphora species (Ustilaginomycotina) on Caryophyllaceae. *Mycological research* **111**: 1207–19.

Vánky K, Lutz M, Bauer R. **2007**. About the genus Thecaphora (Glomosporiaceae) and its new synonyms. *Mycological Progress* **7**: 31–39.

Vasighzadeh A, Zafari D, Selçuk F, Hüseyin E, Kurşat M, Lutz M, Piątek M. **2014**. Discovery of Thecaphora schwarzmaniana on Rheum ribes in Iran and Turkey: implications for the diversity and phylogeny of leaf smuts on rhubarbs. *Mycological Progress*.

Wahl V, Ponnu J, Schlereth A, Arrivault S, Langenecker T, Franke A, Feil R, Lunn JE, Stitt M, Schmid M. **2013**. Regulation of Flowering by Trehalose-6-Phosphate Signaling in Arabidopsis thaliana. *Science* **339**: 704–707.

Wahl R, Zahiri A, Kämper J. **2010**. The Ustilago maydis b mating type locus controls hyphal proliferation and expression of secreted virulence factors in planta. *Molecular microbiology* **75**: 208–20.

Walton JD. **1994**. Deconstructing the Cell Wall. *Plant physiology* **104**: 1113–1118.

Walton JD. **1996**. Host-Selective Toxins: Agents of Compatibility. *the Plant Cell Online* **8**: 1723–1733.

Wang Q-M, Begerow D, Groenewald M, Liu X-Z, Theelen B, Bai F-Y, Boekhout T. **2015**. Multigene phylogeny and taxonomic revision of yeasts and related fungi in the Ustilaginomycotina. *Studies in Mycology* **81**: 55–83.

Wang K, Uppalapati SR, Zhu X, Dinesh-Kumar SP, Mysore KS. **2010**. SGT1 positively regulates the process of plant cell death during both compatible and incompatible plant-pathogen interactions. *Molecular Plant Pathology* **11**: 597–611.

Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach M a., Muller R, Wohlleben W, *et al.* 2015. antiSMASH 3.0--a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research* **43**: 1–7.

Weiberg A, Wang M, Lin F-MF-M, Zhao H, Zhang Z, Kaloshian I, Huang H-DH-D, Jin H. **2013**. Fungal small RNAs suppress plant immunity by hijacking host RNA interference pathways. *Science* **342**: 118–23.

Weigel D, Glazebrook J. **2006**. Transformation of Agrobacterium Using Electroporation. *Cold Spring Harbor Protocols* **2006**: pdb.prot4665–pdb.prot4665.

Wiik L, Rosenqvist H. **2010**. The economics of fungicide use in winter wheat in southern Sweden. *Crop Protection* **29**: 11–19.

Williams AH, Sharma M, Thatcher LF, Azam S, Hane JK, Sperschneider J, Kidd BN, Anderson JP, Ghosh R, Garg G, *et al.* 2016. Comparative genomics and prediction of conditionally dispensable sequences in legume–infecting Fusarium oxysporum formae speciales facilitates identification of candidate effectors. *BMC Genomics* **17**: 191.

de Wit PJGM, van der Burgt A, Ökmen B, Stergiopoulos I, Abd-Elsalam K a., Aerts AL, Bahkali AH, Beenen HG, Chettri P, Cox MP, *et al.* 2012. The Genomes of the Fungal Plant Pathogens Cladosporium fulvum and Dothistroma septosporum Reveal Adaptation to Different Hosts and Lifestyles But Also Signatures of Common Ancestry. *PLoS Genetics* **8**.

Wollenberg T, Schirawski J. **2014**. Comparative genomics of plant fungal pathogens: the Ustilago-Sporisorium paradigm. (WE Goldman, Ed.). *PLoS pathogens* **10**: e1004218.

Xu J, Saunders CW, Hu P, Grant R a, Boekhout T, Kuramae EE, Kronstad JW, Deangelis YM, Reeder NL, Johnstone KR, *et al.* 2007. Dandruff-associated Malassezia genomes reveal convergent and divergent virulence traits shared with plant and human fungal pathogens. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 18730–5.

Yan M, Zhu G, Lin S, Xian X, Chang C, Xi P, Shen W, Huang W, Cai E, Jiang Z, *et al.* 2016. The mating-type locus b of the sugarcane smut Sporisorium scitamineum is essential for mating, filamentous growth and pathogenicity. *Fungal genetics and biology : FG & B* **86**: 1–8.

Yang F. **2015**. Genome-wide analysis of small RNAs in the wheat pathogenic fungus Zymoseptoria tritici.

Fungal Biology **119**: 631–640.

Yang XL, Awakawa T, Wakimoto T, Abe I. **2013**. Induced production of the novel glycolipid ustilagic acid C in the plant pathogen Ustilago maydis. *Tetrahedron Letters* **54**: 3655–3657.

Yee AR, Kronstad JW. **1998**. Dual Sets of Chimeric Alleles Identify Specificity Sequences for the bE and bW Mating and Pathogenicity Genes of Ustilago maydis. *Molecular and cellular biology* **18**: 221–232.

Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. **2012**. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic acids research* **40**: W445–51.

Zerbino DR, Birney E. **2008**. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research* **18**: 821–9.

Zhang Z, Henderson C, Gurr SJ. **2004**. Blumeria graminis secretes an extracellular catalase during infection of barley: Potential role in suppression of host defence. *Molecular Plant Pathology* **5**: 537–547.

Zhao Z, Liu H, Wang C, Xu J. **2014**. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC genomics* **15**: 6.

Zhao X, Ye J, Wei L, Zhang N, Xing Y, Zuo W, Chao Q, Tan G, Xu M. **2015**. Inhibition of the spread of endophytic Sporisorium reilianum renders maize resistance to head smut. *The Crop Journal* **3**: 87–95.

Zuo W, Chao Q, Zhang N, Ye J, Tan G, Li B, Xing Y, Zhang B, Liu H, Fengler K a, *et al.* 2014. A maize wall-associated kinase confers quantitative resistance to head smut. *Nature Genetics* **47**: 151–157.

5.3 Figure and Table list

5.3.1 Figures

Figure 1. Occurrence of smut diseases in several parts of the world. 5
Figure 2. The *b* locus of *U. maydis*. 7
Figure 3. Symptoms of *T. thlaspeos* infection on *Ar. hirsuta* and *Ar. ciliata*. 13
Figure 4. Maturation of teliospores in *Ar. ciliata* infected siliques. 14
Figure 5. Germination of *T. thlaspeos* teliospores. 16
Figure 6. Staining of infected host plants from the field collections or lab infection assays. 18
Figure 7. Germination, growth and symptoms of three *Thecaphora* species. 21
Figure 8. Overview of the de-novo genome assembly pipeline used. 26
Figure 10. Phylogeny and ortholog clustering with the *T. thlaspeos* genome. 27
Figure 11. Inter- and intra-species synteny of the a locus of *T. thlaspeos*. 30
Figure 12. Phylogeny and expression of *T. thlaspeos* mating genes. 31
Figure 13. Alignments of the TtbE and TtbW. 33
Figure 14. Yeast-two-hybrid analysis of the interactions between the bE and bW proteins 34
Figure 15. Phylogeny of the *T. thlaspeos* a-L-Rhamnosidase (THTG_02733) 36
Figure 16. Loss of the flocculosin cluster in *T. thlaspeos*. 37
Figure 17. Novel elements in the genome of *T. thlaspeos* – α-CA. 41
Figure 18. The secreted catalase THTG_00174 of *T. thlaspeos*. 42
Figure 19. Necrosis inducing-like proteins in *T. thlaspeos*. 44
Figure 20. A proposed life cycle for *T. thlaspeos*. 53

5.3.2 Tables

Table 1. Collections of *T. thlaspeos* 12
Table 2. Infection experiments with *T. thlaspeos* 17
Table 3. Additional species collection 19
Table 4. Synopsis of the genomic assembly 23
Table 5. BUSCO analysis results 24
Table 6. Percent of non-orthologs with other sequenced *Ustilaginomycotina* per species 29
Table 7. Secretome prediction 39
Table 8. Completeness analysis for the *T. seminis-convolvuli* genome 45
Table 9. Proteome evidence used for annotation 68
Table 10. Databases used for functional annotation 69

5.3.3 Supplementary figures

Suppl. Figure S1. Infected seeds of *Ar. ciliata* by *T. thlaspeos* are viable.
Suppl. Figure S2. Annotation edit distance distribution
Suppl. Figure S3. Synteny between the *T. thlaspeos* scaffolds and the *U. maydis* chromosome 19
Suppl. Figure S4. Synteny between the *T. thlaspeos* scaffolds and the *U. maydis* chromosomes 3 and 5
Suppl. Figure S5. Maximum likelihood phylogeny of TtPep1 and homologs.
Suppl. Figure S6. Additional data on *T. saponariae, T. oxalidis* and *T. seminis-convolvuli*
Suppl. Figure S7. Alignment of the pheromone receptor Pra1 and Pra2 of *T. thlaspeos*
Suppl. Figure S8. Alignment of *T. thlaspeos* pheromone precursors with homologs of other smut species.
Suppl. Figure S9. Control interactions in the Clonetech MatchMaker system and supplementary combinations to the ones presented in Figure 14
Suppl. Figure S10. Western blots of the Y2H interactions shown in Figure 14
Suppl. Figure S11. *bbs* sequences in *T. thlaspeos*.
Suppl. Figure S12. Healthy and infected seeds of *C. sepium* by *T. seminis-convolvuli*.

5.3.4 Supplementary tables

Table S1. Media used for teliospore germination based on literature search
Table S2. Media used to initiate teliospore germination of *T. thlaspeos*
Table S3. Analysis of the repetitive content of *T. thlaspeos* LF1 genome
Table S4. BUSCO results extended with more species
Table S5. CEGMA analysis results
Table S6. Ortholog genes to *U. maydis* known effectors
Table S7. Genes that have no homologs to other smut fungi and are secreted
Table S8. Genes involved in mating
Table S9. AntiSMASH 3.0.5 prediction of secondary metabolite synthesis genes in *T. thlaspeos*
Table S10. Complete list with putative secreted proteins as predicted with the (De Jonge,2012) method
Table S11. Complete list with putative secreted proteins as predicted by SignalP 4.0 and filtered with TMHMM 2.0c
Table S12. Recovery of known effector recovery in *U. maydis* in the predicted secretome
Table S13. Parsed Table of annotations for the putative secreted proteins
Table S14. Nuclear localization signal predition on putatively secreted proteins
Table S15. Common functional domains between *U. maydis, S. reilianum, U. hordei, S. scitamineum, T. thlaspeos, M. pennsylvanicum*
Table S16. Carbohydrate active enzymes of 14 smut fungi
Table S17. *T. thlaspeos* genes with no homologs to other sequenced smut fungi
Table S18. Count of RNAseq reads mapped to annotation features

5.3.5 Supplementary Data

Supplementary data 1	Fasta files of large ribosomal subunit used for identification of collected spores, hosts and their respective alignments to reference sequences.
Supplementary data 2	Raw results of OrthoFinder analysis
Supplementary data 3	Results of Anti-Smash 3
Supplementary data 4	Functional annotation derived by InterPro
Supplementary data 5	Genome of *T. thlaspeos LF1*
Supplementary data 6	Genome of *T. thlaspeos LF2*
Supplementary data 7	Genome of *T. seminis-convolvuli 4e5-1*
Supplementary data 8	Annotation file (gff) for *T. thlaspeos LF1*
Supplementary data 9	Annotation file (gff) for *T. seminis-convolvuli 4e5-1*
Supplementary data 10	MUMMER comparison of *T. seminis-convolvuli* with *T. thlaspeos*, *U. maydis* and *A. flocculosa*.

Movie 1	*T. thlaspeos* teliospores germinate when co-incubated with *A. thaliana* Col-0 seeds.
Movie 2	Growth of *T. thlaspeos* LF1 filaments in NM.
Movie 3	*T. seminis-convolvuli* produces conidia at the tips of its filaments on CM
Movie 4	*T. seminis-convolvuli* conidia on NM
Movie 5	*T. oxalidis* conidia on NM
Movie 6	Fusion of *T. thlaspeos* LF1 and LF2 filaments upon confrontation. Filaments were plates on NM plates.

5.4 Acknowledgements

My contribution to this almost Sisyphean project is for now over. It was a wonderful four years in the Institute of Microbiology in Düsseldorf and I have definitely learned and earned a lot. For this I am definitely grateful to Prof. Dr. Michael Feldbrügge for giving me the opportunity to work on *Thecaphora*.

I would also like to thank Dr. Vera Göhre for all her help and advice the last four years, as also to all the members of the "RABattack" group, the iGRAD-plant graduate school and the Institute of Microbiology.

Finally I would like to thank Prof. Dr. Ernst, for reviewing this thesis and PD Dr. Linka for participating in my committee.

Οτι και να γράψω, δυστυχώς είναι λίγο για να εκφράσω την ευγνωμοσύνη μου για όλους τους ανθρώπους που με βοήθησαν σε αυτή τη διαδρομή. Παρ' όλα αυτα:

Η εργασία αυτή αφιερώνεται στον πατέρα μου Ιωάννη Φραντζεσκάκη, τη μητέρα μου Θεοδώρα Χάδου και την αδερφή μου Βάσω, για την υποστήριξη και αγάπη τους. «Όπως είπαμε!»

Επίσης θα ήθελα ειδικά να ευχαριστήσω την Κέιτη Κουρβίλ για την συνεχή ψυχολογική υποστήριξη και πάρα πολλές στιγμές υπομονής που περάσαμε μαζι στην Λίντενμανστρασε του Ντύσσελντορφ.

Λάμπρος Φραντζεσκάκης 17/05/2016

5.5 CV

Lamprinos Frantzeskakis

lamprinos.frantzeskakis@hhu.de

EDUCATION

Heinrich-Heine-Universität Düsseldorf, iGRAD-plant graduate school 2012 - 2016

Düsseldorf, Germany

Ph.D. in Natural Sciences

Thesis: The genome of the fungal plant pathogen *Thecaphora thlaspeos*

Rheinische Friedrich-Wilhelms-Universität Bonn 2010 - 2012

Bonn, Germany

M.Sc. Plant Sciences – GPA 1.7/1.0

Major subjects: Plant Physiology, Cell Biology and Biotechnology

Thesis: VirD2-TALeffector (D2TALe) mediated targeted genome insertions and editing in plants (Grade awarded 1.3)

Agricultural University of Athens 2003 - 2010

Athens, Greece

M.Sc. Agricultural Science – GPA 6.89/10

Major subjects: Agronomy, Plant Breeding

Thesis: The yield potential of grain sorghum in Evros prefecture (Grade awarded 10/10)

WORK EXPERIENCE

Student Assistant September 2011 – August 2012

IZMB - Institute of Cellular and Molecular Botany (Bonn, Germany)

Internship February 2011 – April 2011

IMBIO - Institute of Molecular Physiology and Biotechnology of Plants (Bonn, Germany)

Student Assistant January 2010 – August 2010

Crop Science Lab, Agricultural University of Athens (Athens, Greece)

Internship September 2008 – December 2008

Enza Zaden (Enkhuizen, Netherlands)

Internship 2007 – 2009

Plant Breeding and Biometry Lab, Agricultural university of Athens (Athens, Greece)

PEER REVIEWED PUBLICATIONS - PATENTS

- Hommelsheim, C. M., Frantzeskakis, L., Huang, M., Ulker, B. (2014). PCR amplification of repetitive DNA: a limitation to genome editing technologies and many other applications. Scientific Reports, 4, 5052. doi:10.1038/srep05052
- Ulker, B., Hommelsheim, C. M., Berson, T., Thomas, S., Chandrasekar, B., Olcay, A. C., Berendzen, K. W., Frantzeskakis, L. (2012). Reevaluation of the Reliability and Usefulness of the Somatic Homologous Recombination Reporter Lines. The Plant Cell, 1–11. doi:10.1105/tpc.112.100404
- Hommelsheim, C. M., Frantzeskakis, L., Ulker, B. (2014) Improved gene targeting and nucleic acid carrier molecule, in particular for use in plants. EP 2796558 A1

POSTERS

- Frantzeskakis L, Courville K, Baltes S, Kellner R, Brachmann A, Feldbrügge M, Göhre V, *Thecaphora thlaspeos*, a smut pathogen that colonizes the model plant *Arabidopsis thaliana*, 28th Fungal Genetics Conference, 17-22 March 2014 Pasadena, CA, USA
- Frantzeskakis L, Courville K, Baltes S, Kellner R, Brachmann A, Göhre V, Feldbrügge M, Thecaphora thlaspeos : A smut pathogen of Arabis ciliata, XVI International Congress on Molecular Plant-Microbe Interactions, 6-10 July 2014 Rhodes, Greece

WORKSHOPS

- RNA sequencing tools, ICER Michigan State University (2015)
- Advanced shell scripting, ICER Michigan State University (2015)
- Joint EMBL-EBI/Wellcome Trust Summer School in Bioinformatics (2013)

SCHOLARSHIPS - AWARDS

Genetics Society of America - GSA 2003

Travel award for participating in the 28th Fungal Genetics Conference

PROFESSIONAL SOCIETIES

International Society for Plant Microbe Interactions 2014