Virulence in smut fungi: Insights from evolutionary comparative genomics

Dissertation

zur
Erlangung des Doktorgrades
der Naturwissenschaften
(Dr. rer. nat.)

Dem Fachbereich Biologie
der Philipps-Universität Marburg
vorgelegt von

Gabriel Schweizer
aus Kemnath

Marburg/Lahn, 2015

Die Untersuchungen zur vorliegenden Arbeit wurden von Oktober 2011 bis September 2015 unter der Betreuung von Herrn Dr. Julien Y. Dutheil und Frau Prof. Dr. Regine Kahmann in Marburg am Max-Planck-Institut für terrestrische Mikrobiologie in der Abteilung Organismische Interaktionen durchgeführt.

Vom Fachbereich Biologie
der Philipps-Universität Marburg als Dissertation
angenommen am: 26. November 2015

Erstgutachter: Frau Prof. Dr. Regine Kahmann
Zweitgutachter: Frau Prof. Dr. Anke Becker

Tag der mündlichen Prüfung: 14. Dezember 2015

Publications

Pletzer D, Schweizer G and Weingart H (2014): AraC/XylS family stress response regulators Rob, SoxS, PliA, and OpiA in the fire blight pathogen *Erwinia amylovora*. JOURNAL OF BACTERIOLOGY 196(17), 3098-3110.

Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, Zuccaro A, Reissmann S and Kahmann R (2015): Fungal effectors and plant susceptibility. ANNUAL REVIEW OF PLANT BIOLGOY 66, 513-545.

Schuster M, Schweizer G, Reissmann S and Kahmann R (2015): Genome editing in *Ustilago maydis* using the CRISPR-Cas system. FUNGAL GENETICS AND BIOLOGY [in press].

Dutheil JY, Mannhaupt G, Schweizer G, Sieber CMK, Münsterkötter M, Güldener U, Schirawski J and Kahmann R: The evolution of virulence clusters in smut fungi [in preparation].

Erklärung

Ich versichere, dass ich meine Dissertation mit dem Titel „Virulence in smut fungi: Insights from evolutionary comparative genomics“ selbstständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfsmittel bedient habe.
Diese Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den 29. September 2015

Gabriel Schweizer

In all things of nature there is something of the marvelous.
ARISTOTELES (384 – 322 BC)

Contents

SUMMARY

Fungi and plants show a long history of co-evolution since about 400 million years. This lead to the development of diverse types of interactions which include for example parasitism, in which fungi reduce the fitness of their host. Parasitic fungi can establish biotrophic interactions, which require living plant tissues for successful colonization. To establish biotrophy, fungi secret effectors, which are proteins that prevent or mediate plant immune responses. They can also contribute to virulence by changing the host physiology towards the needs of the pathogen. Effectors and their plant targets evolve in a molecular arms race, where both pathogen and plants evolve new effectors and plant interactors, respectively. In this process, single nucleotide polymorphisms and species-specific orphan genes can play an important role.

Smut fungi (order: *Ustilaginales*) are biotrophic pathogens, which parasitize mostly sweet grasses, including wheat, oat, barley, maize, sugar cane and *Sorghum* grass. The genomes of five related species with different host plants or colonization strategies (*Ustilago hordei*, *Ustilago maydis*, *Sporisorium scitamineum*, *Sporisorium reilianum* f. sp. *zeae* and *S. reilianum* f. sp. *sorghi*) were sequenced. Furthermore, methods allowing geneitc manipulations were developed, which makes this group of smut fungi an interesting model system for studying virulence and/or host specificity.

The aim of the present work was to investigated to which extent positively selected or species-specific effectors contribute to virulence of the respective species. To detect positive selection, families of homologous proteins were built. Positive selection was then inferred by applying a non-homogenous branch model of sequence evolution. Most genes under positive selection were found in both formae speciales of *S. reilianum.* A role in virulence could be shown for *sr10529* in *S. reilianum* f. sp. *zeae.* This gene is orthologous to *pit2* of *U. maydis*, where it encodes an inhibitor of cysteine proteases. To get insights in differences in the inhibition of maize cysteine proteases by Pit2-orthologues, a yeast-2-hybrid assay was conducted In contrast to the expectaion that Pit2-orthologues of maize pathogens can better interact with maize cysteine proteases compared to Pit2 of the *Sorghum* pathogen, no host/pathogen-specific interaction could be observed. Besides this, a contribution to virulence could be demonstrated for three gene clusters containing positively selected genes in*S. reilianum* f. sp. *zeae.*

Besides positively selected genes, species-specific orphan genes were bioinformatically identified. Most candidates could be detected in *Pseudocyma flocculosa.* Deletion of the orphan gene *um02193* in *U. maydis* did not reveal a contribution to virulence for this protein.

ZUSAMMENFASSUNG

Pilze und Pflanzen teilen eine lange Koevolution seit ungefähr 400 Millionen Jahren. Dabei entwickelten sich unterschiedliche Arten von Interaktionen wie zum Beispiel Parasitismus, bei dem Pilze auf Kosten ihrer Wirtspflanze leben. Parasitäre Pilze gehen dabei beispielsweise biotrophe Interaktionen ein, bei denen lebendes Pflanzengewebe für die Besiedelung des Wirts erforderlich ist. Zur erfolgreichen Etablierung biotropher Wechselwirkungen sekretieren Pilze Proteine, die als Effektoren Immunantworten verhindern oder abschwächen oder zur Virulenz beitragen, indem sie die Wirtsphysiologie zu Gunsten des Pathogens verändern. Effektoren und ihre pflanzlichen Interaktionspartner evolvieren in einem molekularen Wettrüsten, bei dem Pathogen und Wirtspflanze kontinuierlich neue Effektoren und Interaktionspartner evolvieren, um Virulenz bzw. Abwehr Sicher zu stellen. In diesen Prozessen können Einzelnukleotid-Polymorphismen unter positiver Selektion und artspezifische Gene eine wichtige Rolle spielen.

Brandpilze (Ordnung: *Ustilaginales*) sind biotrophe Pathogene, die vorwiegend Gräser befallen, darunter auch Weizen, Hafer, Gerste, Mais, Zuckerrohr und Sorghumgras. In der Vergangenheit wurden die Genome von fünf verwandten Arten mit unterschiedlichen Wirtspflanzen oder Besiedelungsstrategien sequenziert (*Ustilago hordei*, *Ustilago maydis*, *Sporisorium scitamineum*, *Sporisorium reilianum* f. sp. *zeae* und *S. reilianum* f. sp. *sorghi*). Außerdem wurden Methoden zur genetischen Manipulation entwickelt, was diese Gruppe ideal für Studien zur Virulenz und/oder Wirtsspezifität macht.

Ziel dieser Arbeit war es zu untersuchen, inwiefern positiv selektierte oder artspezifische Effektoren zur Virulenz beitragen. Um positive Selektion zu detektieren, wurden Familien von homologen Proteinen gebildet. Positive Selektion wurde unter Verwenundg eines nicht-homogenen Modells für die Evolution von Nukleotidsequenzen vorhergesagt. Die meisten Gene unter positiver Selektion wurden in den beiden formae speciales von *S. reilianum* detektiert. Ein Beitrag zur Virulenz konnte für *sr10529* aus *S. reilianum* f. sp. *zeae* gezeigt werden. Dieses Gen ist ein Ortholog zu *pit2* aus *U. maydis*, welches einen Inhibitor von Cysteinproteasen kodiert. Um Einblicke in mögliche Unterschiede in der Inhibition von Cysteinproteasen in Mais durch Pit2-Orthologe zu erhalten, wurde ein Hefe-2-Hybrid System verwendet. Im Gengensatz zu der Erwartung, dass Pit2-Orthologe aus Maispathogenen mit Cysteinproteasen aus Mais besser interagiren als Pit2 aus dem *Sorghum* pathogen, zeigten sich keine Wirt/Parasit-spezifischen Interaktionen. Daneben konnte eine Rolle in der Virulenz von *S. reilianum* f. sp. *zeae* für drei Gencluster, die positiv selektierte Gene enthalten, gezeigt werden.

Neben positiver Selection wurden artspezifische Gene bioinformatisch identifiziert. Dabei wurden die meisten Kandidaten in *Pseudozyma flocculosa* gefunden. Die Deletion des Kandidatengenes *um02193* in U. maydis konnte keinen Beitrag zur Virulenz zeigen.

ABBREVIATIONS

AD	Gal4-activation domain
BD	Gal4-binding domain
bp	Base pair(s)
Δ	Deletion
d_N	Rate of non-synonymous mutations
d_S	Rate of synonymous mutations
DTT	Dithiothreitol
ETI	Effector-triggered immunity
$H_2O_{bid.}$	Double distilled water
HA	Hemagglutinin
kDA	Kilodalton
LysM	Lysin-motiv
MAMP	Microbe-associated molecular pattern
MTI	MAMP-triggered immunity
OD_{600}	Optical density at 600 nm
NB-LRR	Nucleotide-binding leucine-rich receptor
ω	Ratio of the rates of non-synonymous and synonymous mutations (d_N/d_S)
PLCP	Papain-like cysteine protease
PRR	Pattern recognition receptor
SDS-PAGE	SDS-Polyacrylamid-Gelelectrophoresis

1. INTRODUCTION

Plants and fungi show a long lasting history of co-evolution over the last 400 million years. Analyses of ribosomal RNAs and fossil records suggested that the development of pioneering land plants was already supported by associations with symbiotic fungi (Remy *et al.*, 1994; Gehring *et al.*, 1996). Besides this, a whole range of forms of interactions evolved. Today, a widespread type is mutualism where both plants and fungi benefit (Parniske, 2008). At the other end of the spectrum are plant pathogenic fungi. They can greatly reduce plant fitness, which has negative impacts on agricultural productions (Fisher *et al.*, 2012). Fungal pathogens can also contribute to plant biodiversity by reducing the fitness of otherwise dominant individuals (Bagchi *et al.*, 2014). Among pathogenic fungi, a variety of colonization and interaction strategies evolved. Some species penetrate only epidermal cell layers, whereas others grow systemically in the infected site and adopt an intra- or extracellular mode of growth. Plant pathogenic fungi can be necrotrophic, meaning that fungi kill the host plant and feed on dead plant tissue, biotrophic, where fungi depend on living plants or a combination of both, where fungi switch from an initial biotrophic to a later necrotrophic stage during plant colonization. Interactions vary also in the degree of specificity: some pathogens are generalists and can parasitize hundreds of plant species (for instance, the grey mold *Botrytis cinerea*), whereas others are specialists and capable of colonizing only one host species (for example, the powdery mildew *Blumeria graminis*) (Dean *et al.*, 2012). All forms of interactions are mediated by fine tuned, multilayered molecular interplays between plants and fungi.

1.1 Molecular basis of plant-fungus interactions

To protect efficiently against pathogens, plants have evolved receptors that allow the recognition of microbes. As reliable perception is crucial for plant survival, these pattern recognition receptors (PRR) evolved to target microbial surface molecules, which are highly conserved and essential for survival. Importantly, these molecules are absent in plants, as their presence would lead to self-stimulated immune responses. This class of molecules is termed microbe-associated molecular pattern (MAMP). Recognition of MAMPs by PRRs initiates a first layer of defense reactions named MAMP-triggered immunity (MTI) (Dodds & Rathjen, 2010). PRRs occur in two variants: receptor-like kinases (RLK) and receptor-like proteins (RLP). PRR signaling elicits rapid plant responses, which involve activation of ion channels, production of reactive oxygen species, activation of defense-related mitogen-associated protein kinase (MAPK) cascades and transcriptional reprogramming as well as later responses like plant hormone biosynthesis and callose deposition at infection sites (Boller & Felix 2009; Macho & Zipfel, 2014). In fungi, the cell wall component chitin was found to serve as

MAMP: Suspension-cultured tomato cells showed an alkanization response upon treatment with yeast cell wall fractions (Felix *et al.*, 1993). In *Arabidopsis thaliana*, chitin oligomers are perceived by the LysM-RLK receptor chitin elicitor receptor kinase 1 (CERK1) through three extracelluar LysM-domains (Miya *et al.*, 2007; Liu *et al.*, 2012). Studies in rice revealed that the LysM protein chitin elicitor-binding protein (CEBiP) belonging to the RLP-PRRs is in addition to CERK1 needed for chitin recognition (Shimizu *et al.*, 2010). MTI is generally sufficient to protect the plant against non-adapted pathogens, a phenomenon known as non-host resistance. On the contrary, adapted pathogens can prevent, reduce or cope with MTI responses in a second, intracellular layer of interaction through the secretion of effector molecules (Dodds & Rathjen, 2010).

Effectors can be recognized by intracellular receptors belonging to the nucleotide-binding leucine-rich receptor (NB-LRR) class, whose activation leads to effector-triggered immunity (ETI). ETI and MTI responses are similar but ETI is qualitatively stronger and faster and often leads to a type of localized cell death called hypersensitive response (Dodds & Rathjen, 2010). A simple way of interaction between effectors and their plant targets was proposed in the "gene-for-gene" model, where the product of an effector gene (in this case termed *avirulence* or *avr* gene) interacts directly with the product of a plant *resistance* gene (*R* gene), thereby triggering a hypersensitive response (Flor, 1971). The idea of direct interactions between effectors and targets has been broadened by the "guard and decoy" as well as the "bait and switch" model (van der Hoorn & Kamoun, 2008; Collier & Moffett, 2009). In addition, it is conceivable that plants do not only sense effector molecules themselves but also their action. This could for example include recognition of changes of plant hormone levels. While some effectors allow to avoid or cope with MTI and ETI responses, others fulfill crucial virulence functions. Effectors may be toxic compounds that kill the host plant (in fungi with a necrotrophic phase in their life cycle). They may also be secreted and/or translocated proteins that change physiological and metabolic states of host cells towards the needs of pathogens by degrading, modifying, inhibiting or altering the activity of plant targets (in fungi with biotrophic life styles). In addition, they can aid in plant penetration or spore dispersal and shield hyphae on the plant surface and at penetration sites (Lo Presti *et al.*, 2015). Finally, the successful plant colonization of a pathogen depends not only on effectors and their plant targets, but also on environmental conditions (Hua, 2013) and likely on the composition of the phyllosphere surrounding the pathogen (Vorholt, 2012; Rovenich *et al.*, 2014).

1.2 Evolution of effector genes

Plants and fungal pathogens co-evolve antagonistically. In a simple model, this takes place in three steps: First, an adapted pathogen attacks and colonizes a host plant, which causes a reduction of plant fitness. This favors the selection of novel host defense strategies, leading to a spread of the corresponding genes in the plant population. In consequence, this leads to a reduction of pathogen adaptation and fitness. In turn, this selects for novel virulence factors and results in an increase of frequency of underlying genes in the pathogen population. Such ongoing adaptations and counter-adaptations are typically compared to an arms race (van Valen, 1973). Due to their essential function in the survival of microbes, MAMPs are unlikely to evolve and PRRs are selected for the reliable recognition of specific MAMPs. In contrast, effectors and their plant interactors can be highly variable, suggesting that the arms race operates on the level of pathogen effector molecules and their host targets (Figure 1.1).

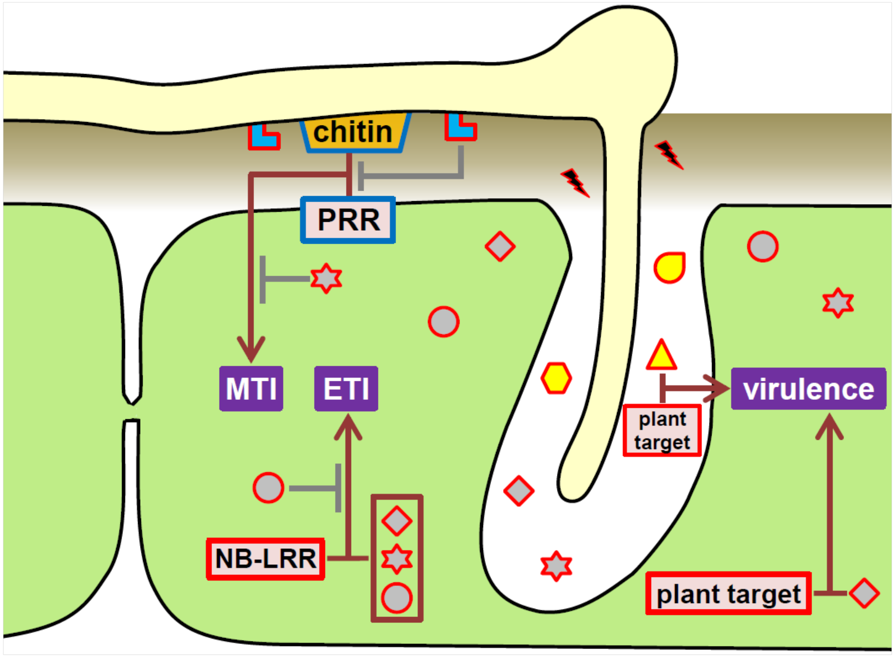


Figure 1.1: Plant and fungal molecules governing interaction. An invading fungal hypha is shown in light yellow. The plant cuticula is depicted in brown and the plant cytoplasm is filled in green. The white space between fungal hypha and plant cytoplasm represents the apoplastic space (enlarged). Black lines indicate fungal and plant plasma membranes. Plant and fungal cell walls are not shown. Blue borders of molecules (chitin and PRR) indicate low evolvability. Red molecule borders (NB-LRR, plant targets and effectors) depict fast and frequent changes, indicating that these molecules are engaged in an arms race. Effectors attached to the fungal cell wall are colored blue, effectors aiding in plant penetration are shown in black, apoplastic effectors are filled yellow and cytoplasmic effectors are colored grey. Effectors with different functions are drawn in different molecule shapes. Brown arrows show typical results of interactions of plant and fungal molecules, whereas grey lines indicate inhibition of such outcomes. PRR, pattern recognition receptor; NB-LRR, nucleotide-binding leucine-rich receptor; MTI, MAMP-triggered immunity; ETI, effector-triggered immunity. See main text for more details. Figure template was taken from Djamei & Kahmann (2012).

Since effector genes are under constant innovation pressure, their distribution within genomes evolved towards localization in niches that have high mutation rates and foster rapid adaptations. These parts of the genome are considered as "evolutionary cradles" for effectors (Croll & McDonald, 2012). For example, they involve gene-sparse regions, where repeat-rich transposon islands dominate and effectors are enriched in these regions. This trend has been discovered in the plant pathogenic oomycete genus *Phytophthora* and is most distinctive in *P. infestans* (Haas *et al.*, 2009). A similar trend was also found in *Leptosphaeria maculans* (Rouxel *et al.*, 2011). In *Magnaporthe oryzae*, effectors are often located in subtelomeric regions, which tend to evolve at higher mutation and recombination rates compared to the rest of the genome (Orbach *et al.*, 2000). Several plant pathogenic fungi like *Nectria haematococca*, *Fusarium oxysporum* and *Mycosphaerella graminicola* (*Zymoseptoria tritici*) have evolved conditionally dispensable chromosomes, which often harbor effector genes (Han *et al.*, 2001; Ma *et al.*, 2010; Stukenbrock *et al.*, 2010). Finally, expansions and contractions of gene families frequently participate in effector gene evolution. Family enlargement could set the playground for the evolution of new effector functions since new paralogous sequences are free to obtain novel characteristics, whereas ancestral sequences retain their current role. In contrast, family shrinking could be a sign for avoidance of host recognition. Expansions of families encoding secreted proteins were for instance observed in *Puccinia graminis* f. sp. *tritici* and *Melapsora lini-populina* (Duplessis *et al.*, 2011). *In silico*-analyses unraveled that homologues of the *Cladosporium fulvum*-virulence factor ecp2 (Hce2) are members of an ancient fungal family, which went through several lineage-specific family expansions and contractions (Stergiopoulos *et al.*, 2012).

1.2.1 Single nucleotide polymorphisms and positive selection

A simple and frequent way of creating novel alleles in a molecular arms race are single nucleotide polymorphisms (SNP). Their rate of occurrence varies among species and strains and also along chromosomes (Baer *et al.*, 2007). In protein coding sequences, SNP come in two flavors: they can leave the encoded amino acid unchanged (synonymous mutation) or they change the corresponding amino acid (non-synonymous mutation). Without selection, synonymous and non-synonymous mutations are detected at a specific rate that is determined by the structure of the genetic code. In this neutral scenario, the ratio of the non-synonymous and the synonymous mutation rate $\mathrm{d_N}/\mathrm{d_S}$ (ω) is considered to be 1. Deviations from this ratio are interpreted as selection. An excess of synonymous mutations leads to an $\omega < 1$, which is interpreted as negative (purifying) selection, meaning that deleterious mutations are counter-selected and that a protein is likely to keep its current function. In contrast, an excess of non-synonymous mutations leads to an $\omega > 1$ and is a sign of positive selection, which indicates that adaptive mutations are favored and that a protein is evolving a new function, a higher efficiency for its current role or adapts to changes of its target to maintain the function.

To detect genes showing signs of positive selection, models of sequence evolution are applied. Evidence for positive selection is considered when a model that allows sites with an $\omega > 1$ fits data significantly better than a model allowing sites with varying ω between 0 and 1 (Nielsen, 2005). Three main approaches were developed for the detection of positive selection. They assume heterogeneous substitution processes in space (site models), in time (branch models) or both (branch-site model). Site models are applied to population data and are used to identify specific sites of proteins under positive selection. Initial methods estimate the number of non-synonymous and synonymous mutations per site between two sequences (Li *et al.*, 1985; Nei & Gojobori, 1986). This model assumes that all sites in a protein are under the same selection pressure and share the same underlying d_N/d_S ratio. However, this hypothesis is unrealistic as different sites in a protein have different structural and functional roles and thus are expected to be under different selection pressures. Therefore, this approach was further developed by allowing variable selection intensities among different sites (Nielsen & Yang, 1998; Yang *et al.*, 2000). Branch models are applied to infer positive selection between different species. In this approach, ω varies between the branches of a phylogenetic tree. This allows the detection of positive selection acting on certain lineages (Yang, 1998; Yang & Nielsen, 1998). The algorithm reported in these studies requires that the user *a priori* defines branches among which similar selection pressures are assumed. This approach was further developed in such a way that scanning for positive selection between branches does not depend on *a priori* assumptions (Dutheil *et al.*, 2012). Branch models do not highlight specific sites of proteins under positive selection. Branch-site models allow ω to vary both among protein sites and branches of a phylogenetic tree. In this method, branches are *a priori* divided in 'foreground' and 'background' branches. A likelihood ratio test is used to compare a model that allows positive selection on the foreground branches with a model that does not allow positive selection (Yang *et al.*, 2005; Zhang *et al.*, 2005).

Inferring differences in selection pressures within and between sequences is commonly used to detect genes involved in adaptation processes. Many studies focus on pathogen - host systems, because beneficial mutations are expected to occur frequently in these environments (Nielsen & Yang 1998; Aguileta *et al.*, 2009; Aguileta *et al.*, 2010). It is also applied for the identification of putative effector genes of plant pathogens and their potential targets. For example, an early investigation of seven families of bacterial type III-secreted proteins showed that they contain members under positive selection (Rohmer *et al.*, 2004). To identify potential effectors of *P. infestans*, Liu *et al.* (2005) used a library of expressed sequence tags obtained from infection stages and identified scr74, which belongs to a highly polymorphic family. Maximum likelihood analysis showed that these polymorphisms are likely created by positive selection. A population study of 123 *Phaeosphaeria nodorum*-strains from 8 geographical origins unraveled signs of positive selection in the host-specific *toxA* gene, suggesting an adaption of each strain to its local host (Stukenbrock & McDonald, 2007). The wheat pathogen *Z. tritici* and its relatives parasitizing wild grasses, *Z.*

pseudotritici and *Z. ardabiliae*, were used to demonstrate that adaptation to a new host is accompanied by positive selection (Stukenbrock *et al.*, 2011). Moreover, a survey of families consisting of paralogous small secreted proteins in *Melampsora larici-poulina* uncovered potential effector genes by employing a positive selection analysis (Hacquard *et al.*, 2012). Molecularly investigated are also the selection pressures shaping the *avr-Pita* gene of the hemibiotroph *Magnaporthe oryzae* (Huang *et al.*, 2014). An intriguing study by Dong and colleagues (2014) could trace back by positive selection analysis the mutations underlying a recent host shift of *Phytophthora mirabilis* in both the pathogen and the respective host plants. Another study addressing positive selection occurring on the plant site found chitinases as targets of molecular selection in *Arabis* species. Surprisingly, amino acid-changing mutations were overrepresented in the active site cleft, which suggested that fungi protect against this plant enzyme through inhibition (Bishop *et al.*, 2000). Together, these studies illustrate the importance of SNPs under positive selection in plant-pathogen systems.

1.2.2 Orphan genes

Another major source of genetic innovation is the *de novo* creation of new genes, rather than the modification of existing ones. Such genes appear only in the group of decendants of the species in which this innovation occurred. These genes are often termed orphan genes, since they lack a (detectable) orthologous gene in other genomes. Orphan genes can constitute a significant proportion of a genome. For instance, a study in animal genomes revealed that between 10 % and 20 % of all genes are orphans (Khalturin *et al.*, 2009). Due to their restricted distributions, orphan genes are thought to be involved in lineage-specific characteristics (Tautz & Domazet-Lošo, 2011). These could for example include the adaptation of a plant pathogen to a specific host. An association between pathogenesis and orphan genes was proposed in the tree pathogenic species *Heterobasidion irregulare*. In this species, virulence QTL regions showed a significant enrichment in transposable elements, orphan genes and genes encoding proteins with a secretion prediction (Olson *et al.*, 2012). About one third of all genes are identified as orphans in the genome of the devastating necrotrophic fungal pathogen *Macrophomina phaseolina* (Islam *et al.*, 2012). Orphan genes are also found in symbiotic fungi. The genome analysis of the arbruscular mycorrhizal species *Rhizophagus irregularis* uncovered lineage-specific gene families of mycorrhiza-induced small secreted proteins (MiSSPs) (Tisserant *et al.*, 2013).

Two main models describe the possible birth of orphan genes: one postulates that after a gene duplication event, one copy acquires a new function and diverges to such an extent that homology between this and the founder gene cannot be detected in scans for similarities between sequences (Tautz & Domazet-Lošo, 2011). This process is known as neofunctionalization (Ohno, 1970). Although this model is in line with proposing that gene duplications are the major source of creating novelty (Ohno, 1970), it has some shortcomings. First, it is challenging to propose an evolutionary process that would affect only one gene copy and allow the accumulation of adaptive mutations while the second copy maintains its ancestral function (Lynch & Katju, 2004). Second, diversification beyond the detection limits of BLAST would require substitutions at the entire length of a gene. However, many genes code for proteins containing functional domains which cannot be easily mutated, and simulations have shown that even small conserved motifs are sufficient to detect homologous sequences, thereby discarding them in an orphan detection pipeline (Albá & Castresana, 2007). These constraints in the duplication model would cease if the duplication is associated with a rearrangement or a transposon insertion, which would alter the duplicated gene dramatically (Tautz & Domazet-Lošo, 2011). Transposable elements in protein coding genes are for example found in humans (Nekrutenko & Li, 2001). A different model suggests that two gene copies evolve by innovation, amplification and divergence. This concept proposes that an ancestral protein with a major and a minor function optimizes the minor function upon duplication, thereby decoupling the evolution of the ancestral and the duplicated sequence. This evolutionary process is known as subfunctionalization (Lynch & Force, 2000) and was demonstrated experimentally in *Salmonella enterica* using a histidine biosynthetic enzyme (Näsvall *et al.*, 2012).

Another, truly *de novo* origin of orphan genes is the evolution of protein coding sequences from non-coding regions. This could happen by random combinations of transcription initiation sites, splice sites, polyadenylation sites and/or regulatory regions which together could ensure the formation of functional transcripts (Tautz & Domazet-Lošo, 2011). Since this event is likely rare, it has been considered to be unimportant for the evolution of new genetic information (Jacob, 1977). However, there are examples for this scenario of gene birth: In *Saccharomyces cerevisiae*, the genes *BSC4* and *MDF1* were identified as *de novo* evolving. Bsc4p is a protein involved in DNA repair and Mdf1p promotes vegetative growth by binding MATα2 in rich medium (Cai *et al.*, 2008; Li *et al.*, 2010a). *CLLU1*, *C22ORF45* and *DNAH10OS* are characterized as human-specific genes that show synteny to non-transcribed regions in other primates. Proteomics demonstrated that these genes encode proteins and a putative role for CLLU1 was suggested in chronic lymphocytic leukaemia (Knowles & McLysaght, 2009). The human six-exon gene *FLJ33706* evolved from a non-coding region that is conserved in eutherian mammals. The first exon and some splice junctions were created through an *Alu* element insertion. Increased expression of this gene was observed in Alzheimer's disease brain samples (Li *et al.*, 2010b). In *Drosophila melanogaster*, the gene *Sdic* contains a coding exon that shows a history as intronic sequence. It codes for a sperm-specific dynein intermediate (Nurminksy *et al.*, 1998). Finally, the gene *Poldi* in *Mus musculus* is specifically expressed in testis. Although short open reading frames can be detected, the gene likely acts as non-coding RNA. A deletion of this gene reduces the mobility of sperm (Heinen *et al.*, 2009).

Despite the origins of orphans discussed here, one could speculate that orphan genes evolve from horizontally transferred genes, gene fusions or fissions, shifts in open reading frames or exon shuffling events.

1.3 Smut fungi as model organisms for biotrophic plant pathogens

A wide range of fungal taxonomic groups contains species infecting flowers of host plants (Ngugi & Scherm, 2006). Among those, an important group are the Basidiomycete smut fungi (Order: *Ustilaginales*), which comprise more than 2,500 described species. All of them have coupled their sexual reproduction to a biotrophic plant parasitic stage (Martínez-Espinoza *et al.*, 2002). Typically, they show a narrow host range and most members parasitize only one host species. Smut fungi can infect around 4,000 species of angiosperms, but are predominately found on sweet grasses (*Poaceae*) which include important crops like maize, sorghum, barley, wheat, oat and sugar cane. A prominent characteristic of smut fungi infections is the massive production of black teliospores which partly or completely replaces seeds and floral organs. In addition, some species can cause macroscopic symptoms in form of tumors or phyllody on discrete parts of a plant (Vánky, 2012).

The corn smut causing species *Ustilago maydis* serves as popular model organism for biotrophic fungal pathogens as well as fungal cell biology (Steinberg & Perez-Martin, 2008; Brefort *et al.*, 2009). It can be cultivated in artificial media as saprotrophic yeast-like budding cells termed sporidia. Strains have been engineered to develop filamentous growth on plates without the need of a compatible mating partner (Brachmann *et al.*, 2001). Constructing solopathogenic haploid strains (*i.e.* strains able to infect plants in absence of a compatible mating type) greatly advanced research on this organism (Bölker *et al.*, 1995; Kämper *et al.*, 2006). Today, PCR-based, FLP-based and CRISPR/Cas-based systems enabling genetic manipulations are established in *U. maydis* (Kämper, 2004; Khrunyk *et al.*, 2010; Schuster *et al.*, in press). Community efforts lead to the deciphering of the genome sequence and reverse genetic approaches highlighted the importance of clustered effector genes with crucial contributions to pathogenicity (Kämper *et al.*, 2006). Together with its short life cycle of three to four weeks under greenhouse conditions, these achievements made *U. maydis* a popular model organism and contributed to its listing upon the 10 most important fungal pathogens, despite not being a highly relevant pathogen in agricultural environments (Dean *et al.*, 2012).

1.3.1 The life cycle of smut fungi

Since the life cycle of smut fungi is mostly investigated in *U. maydis*, it is presented as an example here. *U. maydis* undergoes a dramatic switch of cell morphology during its life cycle: the growth form changes from yeast-like budding to dikaryotic filaments which are able to penetrate and proliferate in planta (Figure 1.2). This switch is governed by a tetrapolar mating system, which consists of the biallelic *a* locus and the multiallelic *b* locus (Kahmann & Schirawski, 2007). The *a* locus harbors a pheromone receptor system that allows haploid sporidia of opposite *a* mating types to sense each other, thereby stimulating the formation of conjugation tubes and eventually fusion (Banuett & Herskowitz, 1989) (Figure 1.2A). Recognition of the pheromone is transmitted via a c-AMP-dependent protein kinase A (PKA) and a mitogen-activated protein kinase (MAPK) pathway. Both pathways converge on the key transcription factor Prf1, which activates transcription of a large set of genes, including the *a* mating type genes *mfa1* and *pra1* as well as the *b* mating type genes (Brefort *et al.*, 2009).

The *b* locus encodes a pair of homeodomain transcription factors, termed bE and bW. These two proteins can dimerize if they are derived from different alleles. The heterodimeric bE/bW complex triggers formation of filaments and subsequent pathogenic development. Filaments show tip-directed growth and accumulation of the cytoplasm in the tip cell. Older, vacuolated parts of the filament are separated by septa (Brefort *et al.*, 2009) (Figure 1.2B). On the plant surface, hyphae stop their tip growth and develop non-melanized appressoria (Figure 1.2C). The formation of appressoria requires the recognition of chemical and physical cues (Mendoza-Mendoza *et al.*, 2009). The perception of hydrophobic surface involves Sho1 and the mucin transmembrane protein Msb2, which activate MAP kinase signaling important for pathogenic development (Lanver *et al.*, 2010).

Upon penetration, the plant plasma membrane invaginates and forms a tight interface between plant and fungus (Figure 1.2D). This interaction zone is used to exchange nutrients and signals and hence plays a crucial role in establishing the biotrophic phase of *U. maydis*. Colonization of the host plant is aided by the secretion of effector proteins, which change the plant physiology to favorable conditions for the proliferation of *U. maydis*. After penetration, *U. maydis* grows initially intercellularly in the epidermal layer. Later, hyphae of *U. maydis* are found mostly extracellular in mesophyll tissue and in vascular bundles (Figure 1.2E). Karyogamy and tumor formation begin around six days post infection. In tumors, fungal hyphae fragment and differentiate into melanized diploid teliospores (Banuett & Herskowitz, 1996) (Figure 1.2F). Recent work has uncovered a central role of the WOPR-domain protein ROS1 in these processes. Upon deletion of *ros1*, tumors are still induced, but *U. maydis* fails to undergo karyogamy and to start the spore differentiation program (M. Tollot & R. Kahmann, unpublished). Fully developed symptoms are typically scored twelve days after infection (Kämper *et al.*, 2006). After tumors have dried and broken up, spores are released. Under favorable conditions, spores germinate and the nuclei undergo meiosis resulting in haploid cells, which completes the life cycle of *U. maydis* (Figure 1.2G).

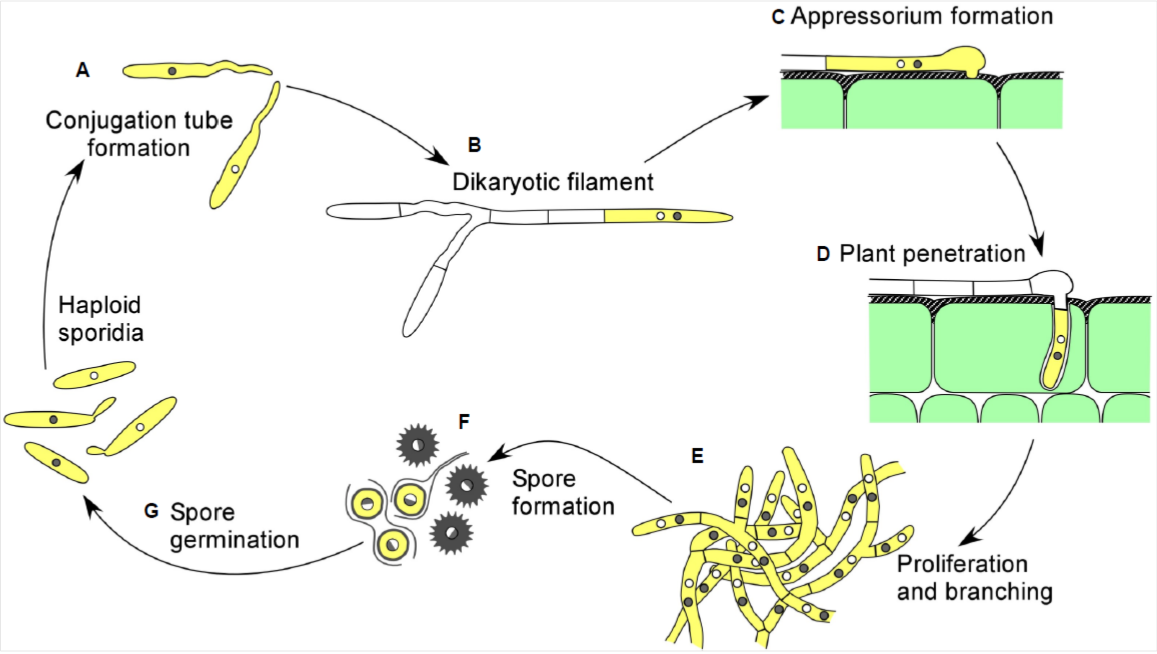


Figure 1.2: Life cycle of smut fungi exemplified by *U. maydis*. The life cycle involves a crucial switch from yeast-like saprotrophic to filamentous pathogenic growth. See main text for more details (Source: Kämper*et al.*, 2006).

1.3.2 Genomic features of smut fungi

Within the last decade, the genomes of five related smut fungi with different host plants have been sequenced: *Ustilago hordei* infecting barley (Laurie *et al.*, 2012), *Ustilago maydis* growing on maize and its wild ancestor teosinte (Kämper *et al.*, 2006), *Sporisorium scitamineum* parasitizing on sugarcane (Que *et al.*, 2014; Taniguti *et al.*, 2015; Dutheil *et al.*, in preparation), *Sporisorium reilianum* f. sp. *zeae* also growing on maize (Schirawski *et al.*, 2010) and *S. reilianum* f. sp. *sorghi* infecting sorghum grass (G. Mannhaupt & R. Kahmann, unpublished) (Figure 1.3).

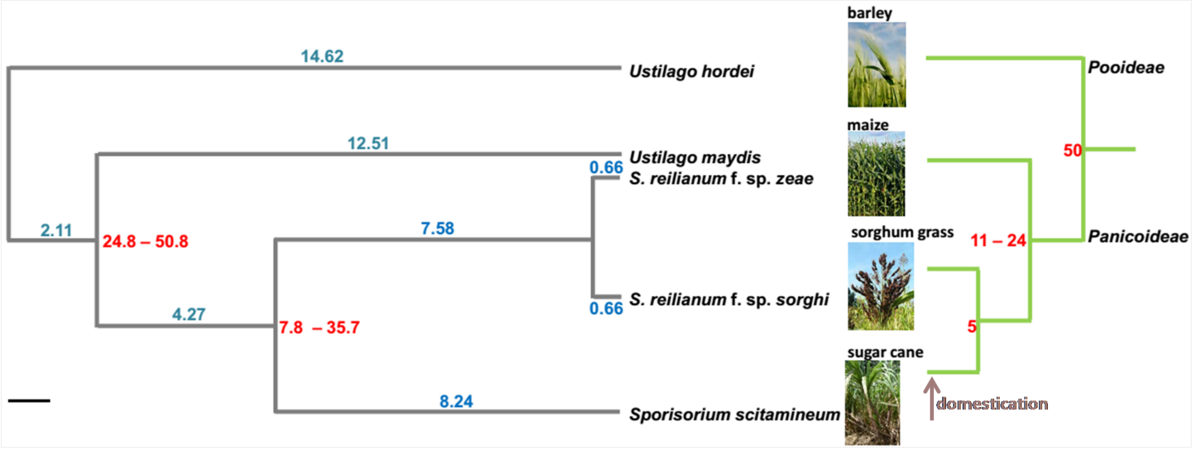


Figure 1.3: Relationship of five smut fungi species and their host plants. Grey and green lines indicate phylogenies of fungi and plants, respectively. Plant phylogeny is not drawn to scale. Blue numbers on the fungal tree represents the nucleotide divergence (in %) between each species (kindly provided by J. Y. Dutheil). Scale bare represents 1 % of nucleotide divergence. Red numbers at branch nodes show estimates of divergence times in myr according to Munkacsi *et al.* (2007). The arrow depicts the beginning of domestication about 10,000 years ago. Pictures of host plants were obtained from Wikipedia: The free encyclopedia.

Deciphering the genome sequences of these smut species exhibited features that are distinct from other filamentous plant pathogens. First, their genome size lies only between 18.48 and 21.15 Mbp, which represents the bottom end of the range of genome sizes in filamentous plant pathogens (Raffaele & Kamoun, 2012). The small genome size can be attributed to the low amount of repetitive elements in these genomes (Table 1.1). A relative exemption to this is *U. hordei*, where the fraction of repetitive elements is about doubled compared to the other species. However, this fraction is still small compared to other filamentous plant pathogens (Raffaele & Kamoun, 2012). Currently, it is not clear, how spreading of repetitive elements is limited in smuts. Some of their genomes show signatures similar to those created by repeat-induced point mutations (RIP) (Laurie *et al.*, 2012; Dutheil *et al.*, in preparation; G. Schweizer, J. Y. Dutheil and R. Kahmann, unpublished). However, it remains to be elucidated to what extent RIP could potentially contribute to limiting these elements in the genome. Moreover, the majority of protein coding genes in smut fungi are devoid of introns (Table 1.1), which contributes to their compact genome sizes. Finally, large scale genome duplications could not be detected (Kämper *et al.*, 2006).

Table 1.1: Characteristics of genomes of five sequenced smut fungi

Species	**Genome size [Mbp]^1**	**Number of proteins^1**	**Secreted proteins^2**	**Repetitive elements [%]^3**	**Genes without introns [%]^1**
U. hordei	21.15	7,113	523 (254)[a]	16.45	62.5
U. maydis	19.67	6,787	596 (308)[a]	6.70	72.3
S. scitamineum	19.63	6,693	602 (324)[b]	6.68	67.7
S. reilianum f. sp. *zeae*	18.48	6,673	602 (310)[a]	8.26	70.9
S. reilianum f. sp. *sorghi*	18.70	6,674	615 (319)[b]	not yet determined	65.2

[1] According to annotation information as of November 2011, which was used in this study
[2] According to predictions of SignalP 4.0
[3] According to Dutheil *et al.* (in preparation)
[a] The first value indicates the total number of predicted secreted proteins. Values in parenthesis indicate number of proteins without predictable functional domain according to Lo Presti *et al.* (2015)
[b] The first value indicates the total number of predicted secreted proteins. Values in parenthesis indicate number of proteins without predictable functional domain (G. Mannhaupt, personal communication)

A second distinct characteristic of the sequenced smut genomes is the clustering of genes encoding putative effector proteins (Kämper *et al.*, 2006; Schirawski *et al.*, 2010; Dutheil *et al.*, in preparation). The presence of such clusters is likely a consequence of the low amount of transposable elements: a duplication event was not followed by translocations (Dutheil et al., in preparation). Consequently, clustering in *U. hordei* was found to be less compact (Laurie *et al.*, 2012). In *U. maydis*, initially twelve clusters compromising three to 26 genes could be identified. They encode predicted secreted proteins for which a functional prediction is not possible. Expression profiling revealed that most genes in these clusters are induced *in planta*. The individual deletion of five clusters had an effect on virulence, which ranged from a complete lack of pathogenicity to hypervirulence (Kämper *et al.*, 2006). Gene clusters in *S. reilianum* f. sp. *zeae* were identified due to the remarkably high synteny of its genome to *U. maydis*. Gene clusters showed a low degree of similarity in overall higher conserved regions, indicating that they have rapidly evolved (Schirawski *et al.*, 2010).

The five species considered here harbor a similar number of gene models and genes encoding secreted proteins, of which about half have no predictable function (Table 1.1). In the present study, all predicted secreted proteins are considered as potential effectors. Efforts to elucidate the biological importance of effectors resulted in the functional characterization of three translocated (Cmu1, Tin2, See1) and two apoplastic effectors (Pep1, Pit2) in *U. maydis*. Cmu1 functions as chorismate mutase in the shikimate pathway and converts chorismate to prephenate, thereby reducing the pool of chorismate available for the synthesis of salicylic acid (Djamei *et al.*, 2011). Tin2 stabilizes the maize protein kinase ZmTKK1 by masking a degradation motif. Since ZmTKK1 acivates anthocyanin biosynthesis pathways, it has been speculated that the binding of ZmTKK1 by Tin2 directs metabolites to the anthocyanin pathway and lowers the amount of metabolites available for other defense reactions (Tanaka *et al.*, 2014). Redkar and colleagues (2015) could show that See1 is specifically required for reactivation of plant DNA synthesis in leaf cells by interacting with a SGT1 homolog. This interaction interferes with the phosphorylation of SGT1 and is important for leaf tumor formation. Pep1 was shown to function as inhibitor of plant peroxidases, thereby efficiently preventing peroxidase-driven oxidative burst and suppressing early defense responses of maize (Hemetsberger *et al.*, 2012). Pit2 was identified as another enzymatic inhibitor. It reduces the activity of a group of salicylic acid-induced papain-like cysteine proteases, which suppresses host immunity reactions (Müller *et al.*, 2013).

In summary, the availability of five annotated genomes of related smut fungi, the typically narrow host range and their amenability to genetic manipulations make smut fungi a particularly interesting model to investigate genes contributing to virulence and potentially host specificity.

1.4 Aim of the thesis

The aim of this work was to determine to what extend positive selection and *de novo* gene creation contribute to the evolution of virulence in smut fungi. To identify candidate genes for both groups, the genomes of five related smuts were employed for comparative pathogenomics analyses. In the next step, identified candidate genes were assessed for their contribution to virulence by creating deletion mutants.

2. RESULTS

2.1 Detection of species-specific genes and their contribution to virulence

2.1.1 Defining criteria to infer orphan genes

The first step was to find clustering criteria for coverage and identity that allow the inference of orphan genes in the five genomes of *U. hordei*, *U. maydis*, *S. scitamineum*, *S. reilianum* f. sp. *zeae* and *S. reilianum* f. sp. *sorghi*. Families were considered to consist of orphan genes, if they contain only members of one species. In this way, paralogous members in one family were also considered as orphan genes. As a result, settings of 5 % for coverage and identity still leave proteins that do not group in one family (Figure 2.1). Increasing the cutoffs for coverage and identity leads to the detection of more orphan genes, as proteins cannot be grouped in one family under stricter settings (Figure 2.1). However, higher thresholds would also increase the number of false positive detections, since looser criteria would allow the clustering in one family. Hence, the inference of orphan genes was carried out with settings of 5 % of both coverage and identity.

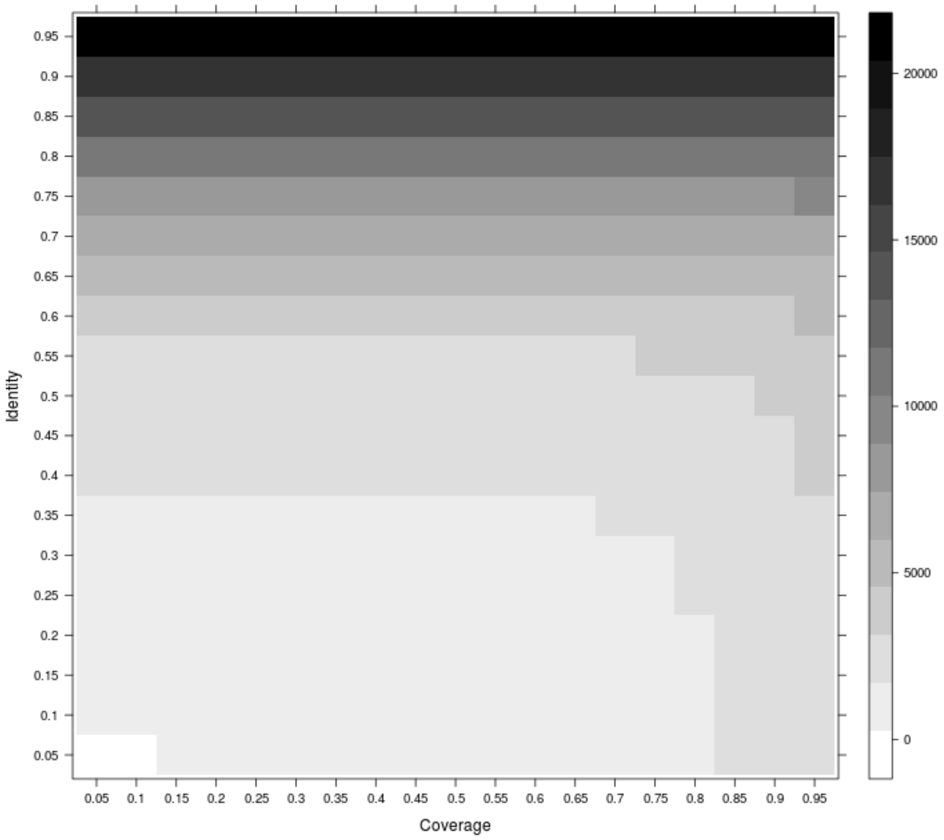


Figure 2.1: Identification of orphan genes in smut genomes. Shades of grey indicate the number of families which have only members of one species according to different settings for coverage and identity (scale on the right).

To infer orphan genes on a more reliable basis, the genomes of the related human pathogen *Malassezia globosa* (Xu *et al.*, 2007), the smut pathogen of dicot *Persicaria* species *Melanopsichium pennsylvanicum* (Sharma *et al.*, 2014) and the bio-control agent *Pseudozyma flocculosa* (Lefebvre *et al.*, 2013) were included. Initially, the proteome of all eight genomes was used to perform an all-agains-all blastp search. To build families of homologues, SiLiX was employed with settings of 5 % for both coverage and identity. In this way, 608 families could be built, of which 492 contained only members in one species.

2.1.2 Orphan genes are particular prominent in *P. flocculosa* and rarely in the pathovariants of *S. reilianum*

In the next step, all proteins of families with members in one species only were used to run a tblastn search against the non-redundant data base of the National Center for Biotechnology Information (NCBI). After this step, only sequences that did not have a hit except in their own genome were considered (e-value cutoff: 0.001) (Table 2.1). It turned out that most orphan genes are found in *P. focculosa*, whereas the *Sporisorium* species showed the lowest number of orphan genes (Table 2.1). All orphan genes found with this approach are listed in Table 6.1 (Supplementary Information).

Table 2.1: Total number of orphan genes and number of predicted secreted orphan proteins in eight related fungal species

Species	**Total number of orphan genes**	**Number of predicted secreted orphan proteins**
M. globosa	52	4
P. flocculosa	108	28
M. pennsylvanicum	89	4
U. hordei	73	4
U. maydis	87	7
S.scitamineum	21	1
S. reilianum f. sp. *zeae*	8	1
S. reilianum f. sp. *sorghi*	0	0

In *M. globosa*, *U. maydis* and *S. scitamineum*, one family consisting of two paralogous orphan genes could be found. Additionally, two families consisting of two paralogues could be found in *M. pennsylvanicum* and *P. flocculosa.* Such paralogous orphan genes were not found in *U. hordei* or the two pathovariants of *S. reilianum.* RNAseq data obtained from *U. maydis* infected maize plants (6 days post infection; M. Tollot and R. Kahmann, unpublished) demonstrated expression of 77 orphan genes (out of 87) in at least one replicate, indicating that these orphans correspond to valid gene models. None of the *U. maydis* orphan proteins was found to be potentially unconventionally secreted (data set: K. Schipper, T. Brefort, M. Mann and B. Macek, unpublished). Using the synteny browser of the MIPS *Ustilago maydis* Genome Database (version 2.0) revealed that 11 of the 87 orphan genes in *U. maydis* are located in a genomic region which is syntenic compared to *S. reilianum* f. sp. *zeae*. In addition, the syntenic region in *S. reilianum* f. sp. *zeae* encodes a gene in the region homolgous to an *U. maydis*-orphan gene. This could indicate that genes in this regions are homologous between *U. maydis* and *S. reilianum* f. sp. *zeae*, but diverged rapidly and cannot their homolgy cannot be recognized anymore. In addition, 32 orphan genes of *U. mayids* are located in a syntenic region compared with *S. reilianum* f. sp. *zeae*, but in these cases, no gene is annotated in *S. reilianum* in the region homolgous to the orpohan in *U. maydis*. Therefore, one could speculate that these *U. maydis* orphan genes originated *de novo*. To evaluate the importance of orphan genes in *U. maydis*, population data of 20 Mexican and two US-American *U. maydis* isolates (G. Schweizer, J. Y. Dutheil, N. Rössel and R. Kahmann, unpublished) were scanned for the presence of orphan genes. 60 candidates could be recovered from multiple genome alignments. These 60 orphan genes were all present in all 22 isolates, indicating that they play an important role in the fitness of *U. maydis*. Interestingly, one orphan gene (*um03039*) was found under positive selection in the population data, which was not predicted to encode a secreted protein. Since it was shown that conserved proteins tend to be longer compared to poorly conserved proteins (Lipman *et al.,* 2002), it was tested whether orphan genes in *U. maydis* differ in length compared to non-orphan genes. As a result, they tend to be significantly shorter (median: 469 bp *vs.* 1526 bp; p-value $= 2.2 \cdot 10^{-16}$; Wilcoxon rank sum test).

2.1.3 Contribution to virulence of predicted secreted orphan proteins in *U. maydis*

Out of 87 orphan genes in *U. maydis*, 7 encode a predicted secreted protein. Of those, *um02193* is part of the previously identified cluster 5A, whose deletion did not affect virulence in seedling infections (Kämper *et al.*, 2006). Since some effectors of *U. maydis* are known to act in an organ-specific manner (Skibbe *et al.*, 2010; Schilling *et al.*, 2014), the effect of the single deletion on tassel infections was tested. However, no contribution to virulence could be observed (Figure 2.2).

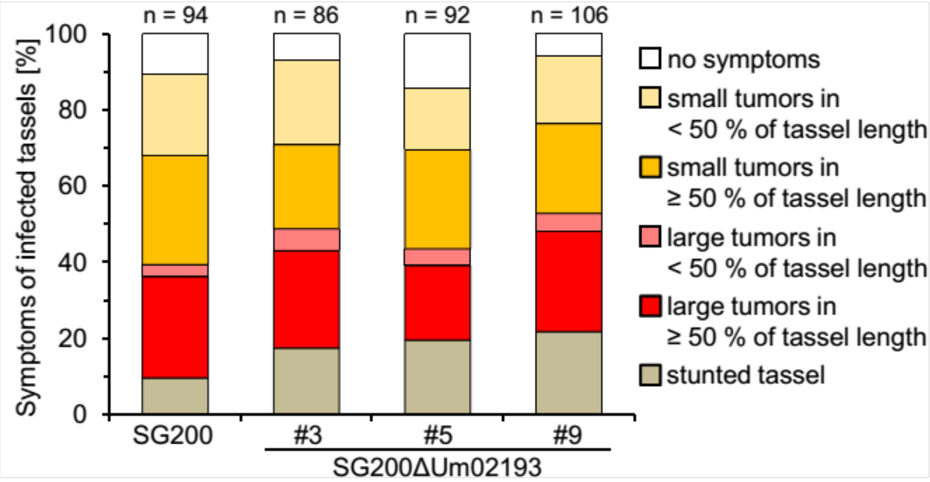


Figure 2.2: Deletion of the orphan gene *um02193* in the solopathogenic strain SG200 does not affect virulence in tassels. Gaspe Flint plants were either infected with the strain SG200 or with one of three independent deletion mutants thereof as indicated below each bar. Symptoms were recorded 10 days post infection according to severeness as shown on the right. The result is presented as mean of three independent experiments in relation to the total number of infected plants (n).

The orphan gene *um11980* is part of the *U. maydis*-cluster 13_10, whose deletion had no effect on virulence in seedling infections (K. Münch and R. Kahmann, unpublished).

2.2 Detection of positively selected genes and their contribution to virulence

2.2.1 Defining criteria to infer homologous sequences

The aim of this analysis was to identify homologous genes in the five smut fungi *U. hordei*, *U. maydis*, *S. scitamineum*, *S. reilianum* f. sp. *zeae* and *S. reilianum* f. sp. *sorghi*, which show signs of positive selection. Such genes are considered to have a possible role in the adaptation to the respective host plant. To define families of homologous sequences, a blastp search was conducted. All proteins served as query and data base (all-against-all search). The result was used to build families employing SiLiX (Miele *et al.*, 2011), which relies on degrees of coverage and identity as clustering criteria. Since the focus of this analysis was to identify positive selection between orthologous genes, the aim of the clustering step was to maximize the number of families containing one member per species (core families). In this way, paralogous members were avoided. To determine settings yielding the maximum number of core families, a range between 5 % and 95 % of coverage and identity was tested. As a result, an identity of 40 % and coverage between 5 % and 45 % lead to the maximum number of core families (5394, Figure 2.3). Stricter settings (*i.e.* higher identity and coverage cutoffs) lead to less core families, because more diverse sequences cannot be captured in one family anymore (Figure 2.3). Looser criteria (*i.e.* lower values for identity and coverage) also reduced the number of core families, because sequences are allowed to cluster in a single family, although the sequences are not necessarily related (Figure 2.3).

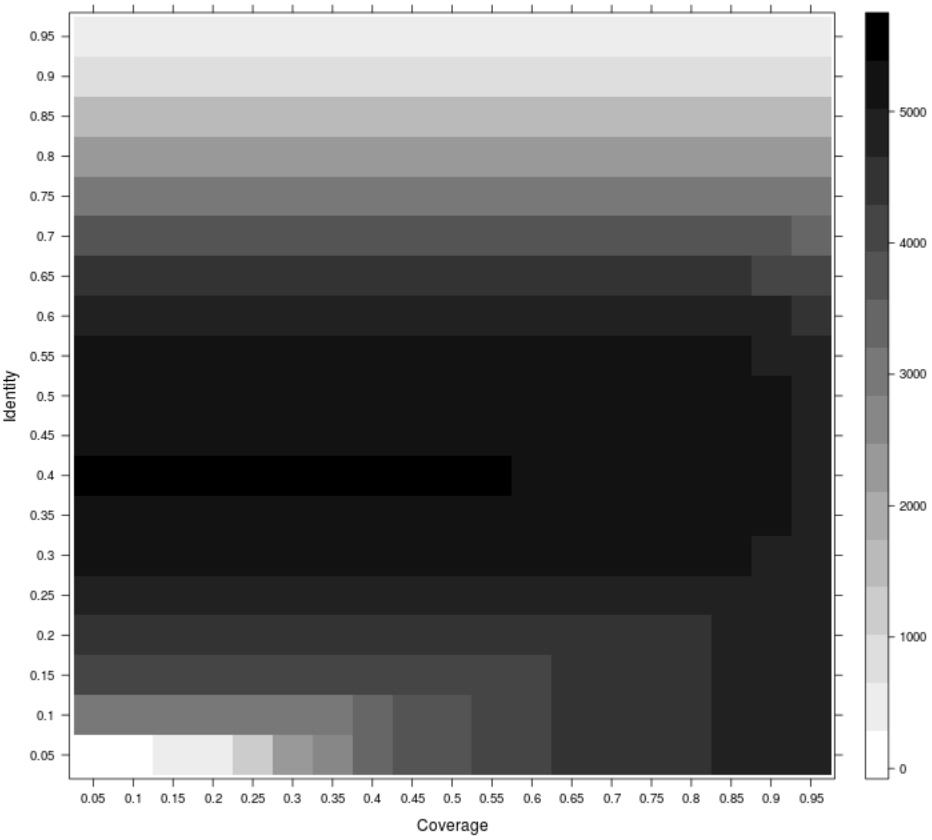


Figure 2.3: Core families in relation to varying settings for identity and coverage. Shades of grey indicate the number of core families according to the scale on the right. The maximal number of core families can be obtained with a coverage between 5 % and 45 % and an identity of 40 %.

Furthermore, it turned out that settings with 40 % identity and 80 % coverage still lead to 5326 core families. Since sequences in a family are more similar using these criteria, the members can be aligned with more confidence, which is important for the detection of positive selection. Building families based on these thresholds had a cost of only 68 core families; therefore, the following analysis was done using these stricter settings. In this way, a total of 8761 families could be inferred. Among those, 5254 families have one member in each species, 55 families have two members in each species, 8 families have three members in each species, and 5 families have four members in each species. 2 families have five members in each species and 2 families have six members in each species. Among these core families, there are 245 families comprising only members that are predicted to be secreted. In total, there were 5411 families where each species was represented at least once. Among those, 257 families consist only of members showing a prediction for secretion. Given that each species encodes on average 588 predicted secreted proteins, these results suggest that about 45 % of all potential effector proteins are conserved across the five species.

Interestingly, several species-specific family expantions could be found in *U. hordei*. There were 17 families which encompassed between 5 and 25 members in *U. hordei* but had no members in other species (Table 2.2). Moreover, 1 family contained 62 members in *U. hordei* and 1 *U. maydis*-protein, 1 family incorporated 8 members in *U. hordei* and 1 member in *S. scitamineum* and 1 family consisted of 6 members of *U. hordei* and 1 member of each *Sporisorium* species (Table 2.2).

Table 2.2: Species-specific family expansions in *U. hordei*

Family	**Members in each species**				
	U. hordei	***U. maydis***	***S. scitamineum***	***S. reilianum* f. sp. *zeae***	***S. reilianum* f. sp. *sorghi***
FAM006490	8	0	0	0	0
FAM006493	25	0	0	0	0
FAM006496	15	0	0	0	0
FAM006499	7	0	0	0	0
FAM006501	8	0	0	0	0
FAM006503	7	0	0	0	0
FAM006506	5	0	0	0	0
FAM006511	13	0	0	0	0
FAM006513	9	0	0	0	0
FAM006522	8	0	0	0	0
FAM006524	8	0	0	0	0
FAM006525	10	0	0	0	0
FAM006528	9	0	0	0	0
FAM006539	5	0	0	0	0
FAM006542	11	0	0	0	0
FAM006543	9	0	0	0	0
FAM006461	62	1	0	0	0
FAM006426	8	0	1	0	0
FAM003338	6	0	1	1	1

Notably, only four of such *U. hordei*-families contained each two members which are located adjacently in the genome. This suggests that duplication events were accompanied by subsequent translocations, which may be attributed to the higher content of repetitive elements in *U. hordei* compared to the other species considered here. Only 3 families consisted exclusively of members of the two maize parasites *U. maydis* and *S. reilianum* f. sp. *zeae*. No evidence for horizontal gene transfer between these species could be detected. Regardless of their composition, 6205 families enclosed at least three members. These families were used for the ensuing detection of positive selection.

2.2.2 Positively selected genes are particularly enriched in the pathovariants of *S. reilianum*

All families of homologous proteins with at least 3 members were aligned and pyhlogentically analyzed (PhyML 3.0). Next, a non-homogeneous model of sequence evolution allowing ω to vary along the phylogeny (Nielsen & Yang, 1998; Romiguier *et al.*, 2012) was applied to scan for positive selection. It turned out that genes showing signs of positive selection are particularly enriched in *S. reilianum* f. sp. *zeae* and *S. reilianum* f. sp. *sorghi*. In addition, a substantial number of candidates was also found in *U. hordei*, but only very few in *U. maydis* and *S. scitamineum* (Table 2.3). A complete list of identified positively selected genes is provided in Table 6.2 (Supplementary Information).

Table 2.3: Total number and number of predicted secreted proteins found under positive selection in five smut genomes

Species	**Total number of proteins under positive selection**	**Number of predicted secreted proteins under positive selection**
U. hordei	49	22
U. maydis	2	0
S. scitamineum	7	0
S. reilianum f. sp. *zeae*	84	25
S. reilianum f. sp. *sorghi*	111	27

Notably, all genes found under positive selection in the two pathovariants of *S. reilianum*, in *S. scitamineum* and in *U. maydis* share orthologous members in the other species, indicating positive selection between species. On the contrary, genes with signs of positive selection in *U. hordei* belong largely (36 out of 49 genes) to families showing species-specific expansions. This indicates selection between paralogous genes in *U. hordei* and could suggest a species-specific way of adapting to its host. Predicted secreted proteins are significantly over-represented in the group of proteins under positive selection compared to the group of proteins not found under positive selection. In *U. hordei*, 44.9 % of proteins showing signs of positive selection are predicted to be secreted, whereas only 7.1 % of proteins not found under positive selection have a secretion prediction (p-value = $5.72 \cdot 10^{-13}$; Fisher's exact test). In *S. reilianum* f. sp. *zeae*, the fraction of predicted secreted proteins among those with signs of positive selection is 29.8 % compared to 8.8 % predicted secreted proteins of not positively selected proteins (p-value = $4.34 \cdot 10^{-8}$). Similar values were found for *S. reilianum* f. sp. *sorghi* (24.3 % compared to 9.0 %; p-value = $1.77 \cdot 10^{-6}$). These results highlight the importance of positively selected and secreted proteins in responses to environmental changes. In *U. maydis* and *S. scitamineum*, no gene under positive selection was found to encode a predicted secreted protein. This can be attributed to the overall small number of genes showing signs of positive selection.

Because effector proteins are often considered to be small, it was tested whether genes under positive selection are shorter than gene that are not under positive selection. It turned out that genes under positive selection are significantly shorter in *U. hordei* (median 661 bp *vs.* 1449 bp) and in *S. reilianum* f. sp. *zeae* (median 1269 bp *vs.* 1539 bp), but not in *S. reilianum* f. sp. *sorghi* (median 1326 bp *vs.* 1534 bp) (Figure 2.4).

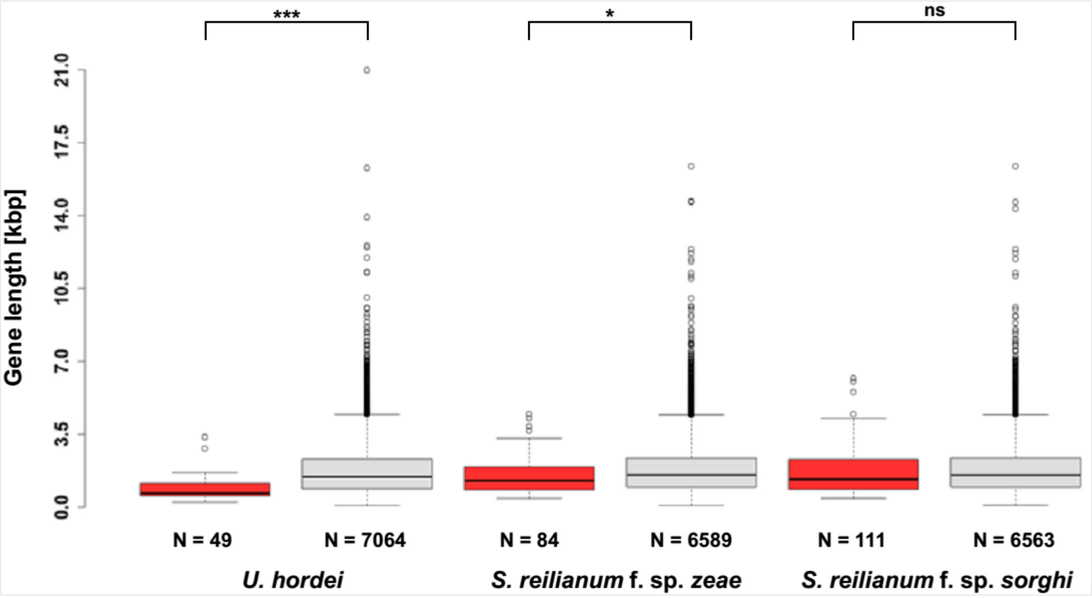


Figure 2.4: Lengths of genes under positive selection (red boxes) compared to genes not under positive selection (grey boxes) in *U. hordei*, *S. reilianum* f. sp. *zeae* and *S. reilianum* f. sp. *sorghi*. Data are represented in form of a box plot, where the top and bottom of the boxes indicate the first and third quartile, respectively. The thick middle line represents the median. Whiskers show data points within the 1.5 interquartile ranges. Open circles denote data points exceeding this range. N equals the number of genes in each category. Asterisks indicate levels of significance: ***, $p < 0.001$; *, $p < 0.05$; ns, not significant (Wilcoxon rank sum test).

Since it is known that effector genes of smut fungi are often found in clusters, it was tested if genes with signs of positive selection tend to locate in such clusters. For this approach, the clusters defined as in Dutheil *et al.* (in preparation) were used. In *S. reilianum* f. sp. *zeae*, 9 genes clusters were inferred. 1 cluster contains three genes under positive selection; 2 clusters harbor 2 genes each with signs of positive selection and 2 clusters possess one gene each under positive selection. Together, this lead to a significant trend of localization of genes under positive selection in clusters (p-value $= 4.54 \cdot 10^{-6}$; Fisher's exact test). In line with this result is the observation that no significant tendency for localization of genes under positive selection towards telomeres, as was described in other organisms, could be found in *S. reilianum* f. sp. *zeae* (p-value $= 0.07$; Wilcoxon rank sum test). Note that similar analyses for *U. hordei* and *S. reilianum* f. sp. *sorghi* are not possible, because genome information is only available on the level of contigs.

Likely due to the higher amount of transposable elements, no significant gene clusters could be detected in *U. hordei*. Since it is known that transposable elements contribute to gene family expansions, it was assessed whether genes under positive selection in *U. hordei* are physically closer to transposable elements compared to genes not harboring signs of positive selection. The minimal distance between uncharacterized interspersed repeats and genes under positive selection was significantly shorter than the distance to genes not under positive selection (median 166.5 bp *vs.* 2027 bp; p-value $= 1.79 \cdot 10^{-9}$; Wilcoxon rank sum test with Bonferroni correction). Intriguingly, candidate genes under positive selection were significantly more distant to low complexity regions than genes not found under positive selection (median $\approx$ 19300 bp *vs.* $\approx$ 9800 bp; p-value = $2.77 \cdot 10^{-3}$. Together, these results suggest that interspersed repeats, but not low complexity regions, could be involved in enhancing rates of non-synonymous mutations in *U. hordei.*

2.2.3 Virulence phenotypes of deletion mutants of predicted secreted positively selected genes in *S. reilianum* f. sp. *zeae*

To assess a potential contribution to virulence of genes found under positive selection and harboring a secretion prediction, deletion mutant strains were created by replacing the gene of interest with a resistance cassette via homologous recombination. Deleted single candidate genes and virulence phenotypes of the respective deletion mutants are summarized in Table 2.4.

Table 2.4: Genes under positive selection in *S. reilianum* f. sp. *zeae* that were individually deleted in JS161 and the according virulence phenotype

Gene	**Description**	ω	**Paralogues**[1]	**Virulence phenotype**
sr10529 (*srmpit2*)	conserved hypothetical protein	31.147	0	almost apathogenic
sr10059	conserved hypothetical *Ustilaginaceae*- specific protein	6.539	0	virulence not affected
sr10182	conserved hypothetical protein	1.575	12	virulence not affected
sr12968	conserved hypothetical protein	37.901	0	virulence not affected
sr14558	conserved hypothetical protein	24.355	0	virulence not affected
sr14944	conserved hypothetical *Ustilaginaceae*-specific protein	4.305	0	virulence not affected
sr14347	conserved hypothetical protein	544.340	5	virulence not affected
sr12897	conserved hypothetical protein	infinite	0	virulence not affected

[1] e-Value cutoff: 0.001

The solopathogenic strain JS161 of *S. reilianum* f. sp. *zeae* was used, because it penetrates and grows within host plants without a compatible mating partner, thereby eliminating the need to generate deletion mutants in both compatible strains (Schirawski *et al.*, 2010). The genotype of strains growing on selection media was verified by Southern analysis. JS161 and its deletion mutant derivatives were grown in $\mathrm{YEPS_L}$ until exponential growth was reached and cultures were adjusted to $\mathrm{OD}_{600} = 1$. Such cultures were used for needle infections of 1 week old maize seedlings of the dwarf variety Gaspe Flint. Virulence symptoms were recorded nine weeks post infection. A prominent symptom of maize infected by *S. reilianum* f. sp. *zeae* is the occurrence of phyllody, *i.e.* the development of leave-like structures instead of female flower organs (Figure 2.5).

Figure 2.5: Virulence phenotype of JS161 in Gaspe Flint. Cob phenotypes of mock-infected (left) and *S. reilianum* f. sp. *zeae*-infected (right) Gaspe Flint maize plants 9 week after treatment. A prominent symptom is the development of phyllody in female flower parts (pictures kindly provided by K. Münch)

Spore development after after infections with JS161 is rarely observed. In contrast, spore development is prominent when compatible strains are crossed (Schirawski *et al.*, 2010; Ghareeb *et al.*, 2011; Zuther *et al.*, 2012). K. Münch kindly created all deletion mutants in the solopathogenic strain JS161 and provided virulence data obtained with these strains.

The gene *sr10529* is orthologous to *pit2* of *U. maydis*, which was demonstrated to be essential for virulence by acting as inhibitor of a group of salicylic acid-induced cysteine proteases (Döhlemann *et al.*, 2011; Müller *et al.*, 2013). The deletion of *sr10529* (*srmpit2*) in JS161 resulted in an almost complete absence of phyllody formation (Figure 2.6.). In contrast to the striking contribution of *srmpit2* to virulence, individual deletions of seven other genes that show signatures of positive selection and encode a predicted secreted protein did not reveal a role in pathogenicity for these genes (Figure 2.6). This finding could indicate that these genes are needed under conditions not tested here or are involved in adaptation processes unrelated to host plant colonization.

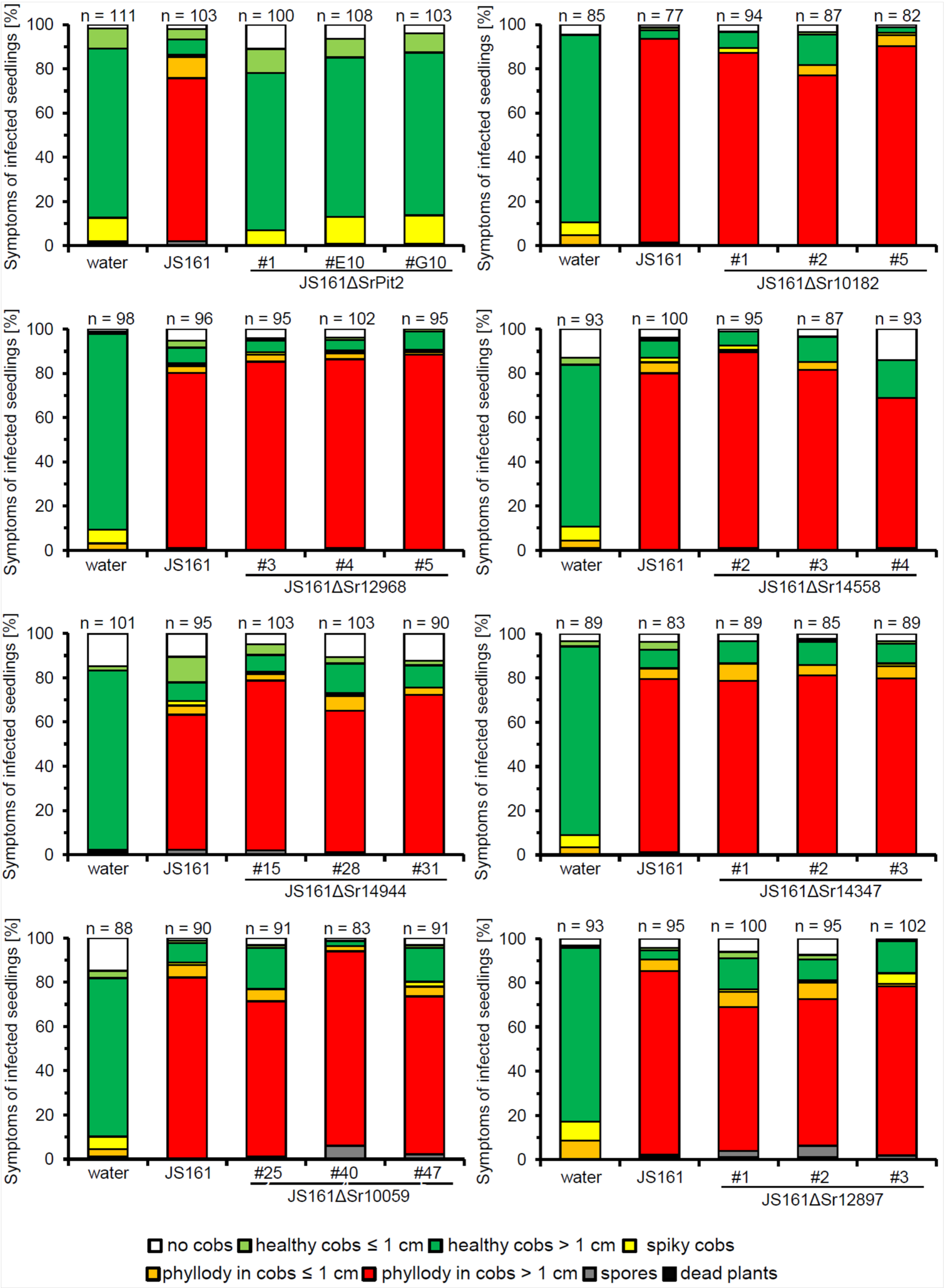


Figure 2.6: Virulence phenotype of single deletion mutants of positively selected in *S. reilianum* f. sp. *zeae*. Gaspe Flint maize plants were infected with water, the solopathogenic strain JS161 or three independent deletion mutants of *srPit2* (left panel in first row), *sr10182* (right panel in first row), *sr12968* (left panel in second row), *sr14558* (right panel in second row), *sr14944* (left panel in third row), *sr14347* (right panel in third row), *sr10059* (left panel in fourth row) or *sr12897* (right panel in fourth row) in JS161. The respective strain is indicated below each bar. Symptoms were scored 9 weeks post infection and categorized according to severeness as illustrated in the legend below the bar plot. The results are shown as mean of three independent experiments in relation to the total number of infected plants (n).

Since the solopathogenic strain JS161 is reduced in virulence compared to compatible wild type strains, the contribution of *srmpit2* to virulence was tested in crossings of compatible strains. A strong reduction in virulence could be observed. Notably, deletion mutant strains were unable to produce spores (Figure 2.7).

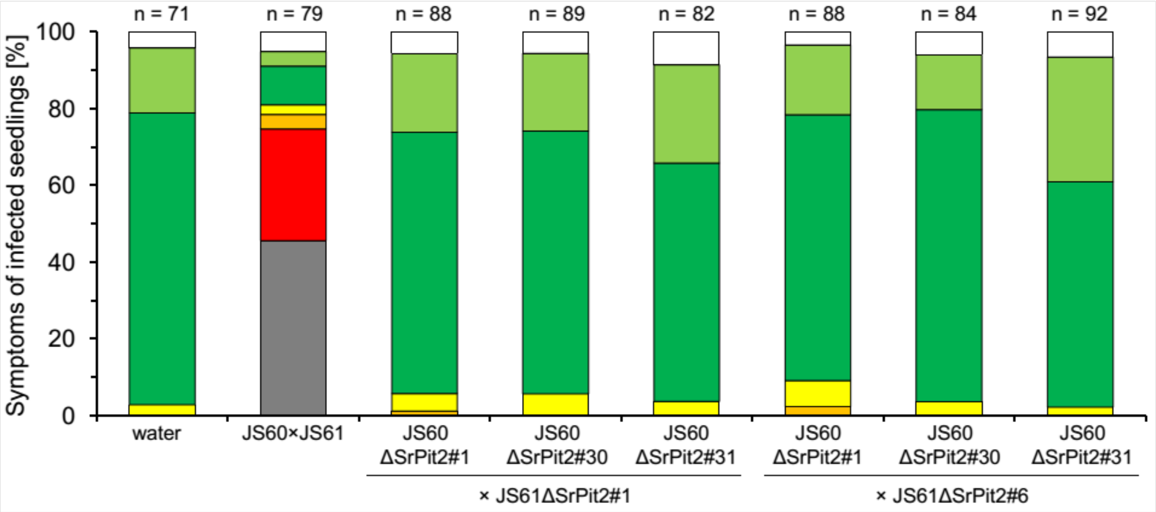


Figure 2.7: Deletion of *srPit2* leads to a severe reduction of virulence in crossings of compatible strains. Plants were either infected with water, crossings of compatible *S. reilianum.* f. sp. *zeae* wild type mating partners (JS60 × JS61) or with crossings of deletion mutants thereof as indicated below each bar. Symptoms were categorized according to severeness as shown in the legend of Figure 2.6. The result is represented as mean of three independent experiments in relation to the total number of infected plants (n).

While some genes found under positive selection are located solitary, others are part of a cluster of genes encoding predicted secreted proteins. Since such gene clusters are composed of families of paralogous genes (Kämper *et al.*, 2006; Schirawski *et al.*, 2010; Dutheil *et al.*, in preparation), it was decided to assess a role in virulence by deleting entire clusters. Contributions to pathogenicity of each cluster gene can then be monitored by individual complementations. In total, six gene clusters containing positively selected genes were deleted and their contribution to virulence was assessed in Gaspe Flint infections. The positively selected genes residing in clusters, phenotypes of deletion mutants and phenotypes of deletion mutants of orthologous regions in *U. maydis* are summarized in Table 2.5. Homologous relationships are for each cluster shown in Supplementary Figure 6.1.

Table 2.5: Genes under positive selection in *S. reilianum* f. sp. *zeae* residing in gene clusters and virulence phenotype of cluster deletions

Cluster in *S.* reilianum f. sp. *zeae*	**Genes under positive selection**	**Description**	ω	**Virulence phenotype**	**Paralogues outsite cluster**[1]	**Cluster in *U. maydis***	**Virulence phenotype**
1-32[a] (*sr12084* - *sr12087*; 4 genes)	*sr12084*[s]	conserved hypothetical *Ustilaginaceae*-specific protein	infinite	hypervirulent	0	I-1	virulence not affected
5-18[a] (*sr13421* - *sr13413*; 9 genes)	*sr13419*[s]	conserved hypothetical *Ustilaginaceae*-specific protein	1.245	hypervirulent	0		virulence not affected[a]
	sr13415[s]	hypothetical protein	1.146				
10-15[a] (*sr11226* - *sr11240*; 13 genes)	*sr11233*[s]	conserved hypothetical *Ustilaginaceae*-specific protein	5.230	virulence strongly reduced	0	10A	virulence reduced[b]
	sr11237[s]	conserved hypothetical protein	10.783				
	sr11239.2[ns]	conserved hypothetical protein	5.956				
	sr11240[ns]	conserved hypothetical protein	5.956				
20-15[a] (*sr16549* - *sr16561*; 13 genes)	*sr16550*[s]	conserved hyptothetical *Ustilaginaceae*-specific protein	infinite	virulence unaffected	5		virulence reduced[a]
	sr16553[s]	conserved hypothetical *Ustilago*-specific protein	3.223				
	sr16556[ns]	conserved hypothetical *Ustilago*-specific protein	3.235				
	sr16558[s]	conserved hypothetical *Ustilaginaceae*-specific protein	2.878				
12-15[a] (*sr15149* - *sr15146*; 4 genes)	*sr15149*[s]	conserved hypothetical protein	4.218	virulence unaffected	3		virulence unaffected[c]
	sr15147[s]	conserved hypothetical protein	31.726				
2-21[a] (*sr10308* - *sr10318*; 11 genes)	*sr10317*[s]	conserved hyothetical *Ustilaginaceae*-specific protein	3.533	virulence unaffected	0	2A	hypervirulent[b]

[1] e-Value cutoff: 0.001

[a] Reported in Schirawski *et al.*, 2010

[b] Reported in Kämper *et al.*, 2006

[c] K. Münch and R. Kahmann, unpublished

[s] Encodes a predicted secreted protein according to SignalP 4.0

[ns] Encodes not a predicted secreted protein according to SignalP 4.0

Interestingly, deletion of the clusters 1-32 (containing the positively selected gene *sr12084*) and 5-18 (containing the positively selected genes *sr13419* and *sr13415*) lead to a hypervirulent phenotype (Figure 2.8). Deleting cluster I-1 (homologous to cluster 1-32 in *S. reilianum* f. sp. *zeae*) in the solopathogenic *U. maydis*-strain SG200 did not affect virulence (Figure 2.9; deletion strains kindly provided by N. Rössel). Similarly, deletion of the homologous region of cluster 5-18 in SG200 did not have an impact on virulence (Schirawski *et al.*, 2010). Furthermore, deletion of cluster 10-15, which contains the positively selected gens *sr11233*, *sr11237*, *sr11239.2* and *sr11240*, lead to a strong reduction in virulence (Figure 2.8). In contrary, deleting the clusters 20-15 (with the positively selected genes *sr16550*, *sr16553*, *sr16556* and *sr16558*), 12-15 (containing the positively selected genes *sr15149* and *sr15147*) and 2-21 (with the positively selected gene *sr10317* had no effect on virulence (Figure 2.8).

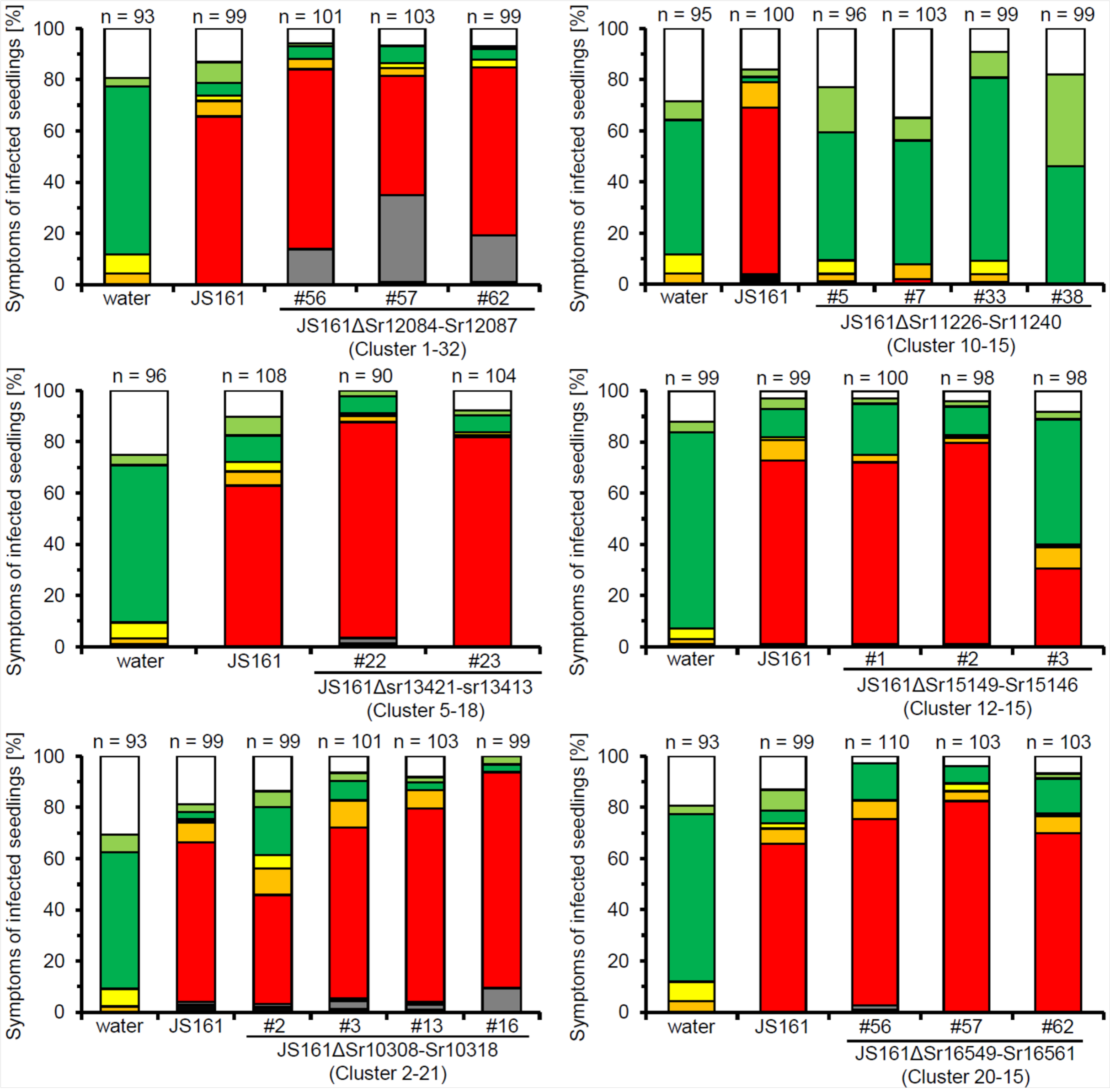


Figure 2.8: Virulence phenotype of deletion mutants of gene clusters containing positively selected genes in *S. reilianum* f. sp. *zeae*. Gaspe Flint maize plants were infected with water, the solopathogenic strain JS161 or three independent deletion mutants of cluster 1-32 (left panel in top row), cluster 5-18 (left panel in middle row), cluster 2-21 (left panel in last row), cluster 10-15 (right panel in first row), cluster 12-15 (right panel in middle row) and cluster 20-15 (right panel in last row) as indicated below each bar. Symptoms were categorized according to severeness as shown in the legend of Figure 2.6. The results are represented as mean of three independent experiments in relation to the total number of infected plants (n).

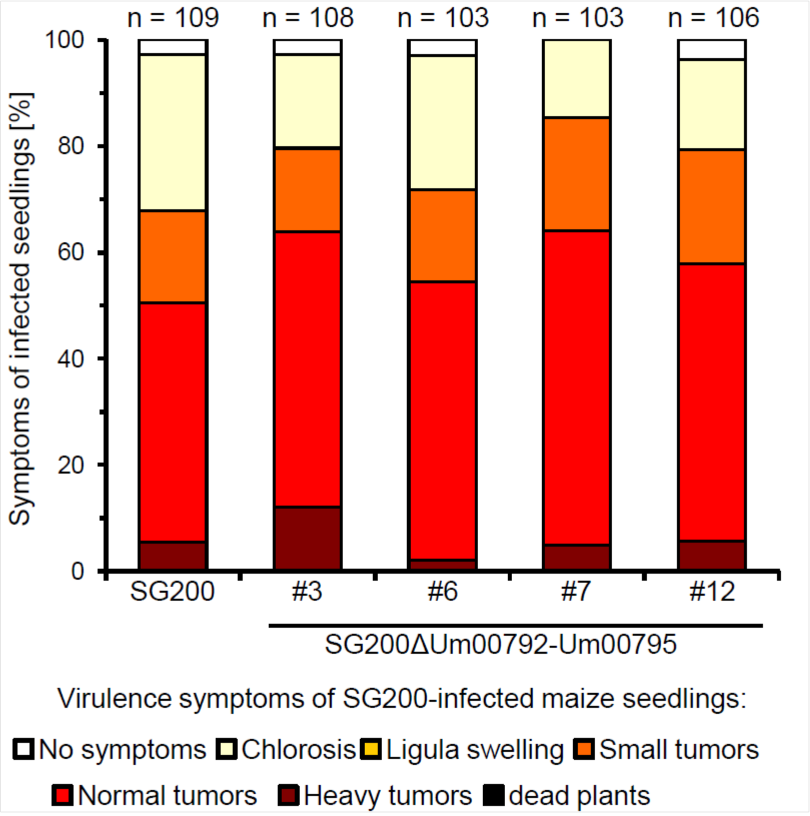


Figure 2.9: Virulence phenotype of the deletion of cluster Um00792-Um00795 in *U. mayids*. Early Golden Bantam maize plants were infected with water, the solopathogenic strain SG200 or three independent deletion mutants of cluster Um00792-Um00795 as indicated below each bar. Symptoms were categorized according to severeness as shown in the legend below the bar plot. The results are represented as mean of three independent experiments in relation to the total number of infected plants (n).

2.3 Potential role of SrPit2 in host specificity

2.3.1 Yeast-2-hybrid interaction analysis of salicylic acid-induced maize papain-like cysteine proteases and Pit2-orthologues

The *U. maydis*-ortholog of SrMPit2, UmPit2 (*um01375*) was shown to have an important function as inhibitor of a class of salicylic acid-induced papain-like cysteine proteases (PLCPs) (Müller *et al.*, 2013). Therefore, it was tempting to speculate that SrMPit2 could inhibit efficiently PLCPs of maize, but not of *Sorghum* (and *vice versa*) and could therefore play a role in host specificity. To get indications of a potentially differential interaction, a yeast-2-hybrid analysis with maize PLCPs as prey and with Pit2 orthologues of *U. maydis*, *S. reilianum* f. sp. *zeae* and *S. reilianum* f. sp. *sorghi* as bait was performed. Previously, it has been shown that UmPit2 can interact in a yeast-2-hybrid assay with the maize PLCPs corn cysteine protease 2 (CP2), CP1A, CP1B and xyleme cysteine protease 2 (XCP2). Furthermore, *in vitro* experiments demonstrated that UmPit2 can inhibit CP2, CP1A, CP1B and XCP2 (Müller *et al.,* 2013). Therefore, CP1A, CP1B and XCP2 were tested for differences in interaction with Pit2-orthologues in a yeast-2-hybrid screen. Since CP2 showed auto activation, it was not included here.

Plasmids carrying either one of the three PLCPs (prey plasmid) or one of the Pit2-orthologues (bait plasmid) were co-transformed in the yeast strain AH109. Single colonies were used to inoculate a SD liquid culture. Cells were grown until the exponential phase was reached and cell densities were adjusted with H_2O to $OD_{600} = 1$. 6 µL of 1:10 serial dilutions were dropped on on low stringency medium (growth control) or high stringency medium to test for interaction. To be able to detect also small differences in interaction, high stringency plates were supplemented with 0.1 mM or 0.5 mM 3-AT, which is a competitive inhibitor of HIS3 (Brennan & Struhl, 1980). Growth was monitored for three days at 28°C.

As a result, better growth could be observed for yeast strains expressing *pit2* of *U. maydis* and *S. reilianum* f. sp. *sorghi* compared to strains expressing *pit2* of *S. reilianum* f. sp. *zeae.* This observation was found for all interactions with CP1A, CP1B and XCP2 (Figure 2.10). In addition, the effect became more prominent when 0.1 mM or 0.5 mM 3-AT were added (Figure 2.10). This result is in contrast with the hypothesis that Pit2 of the maize pathogens *U. maydis* and *S. reilianum* f. sp. *zeae* can interact stronger with maize PLCPs than Pit2 of the *Sorghum* pathogen *S. reilianum* f. sp. *sorghi.* Whether a difference could be observed when PLCPs of *Sorghum* are tested for interaction or when an alternative readout is applied (*e.g.* degradation of ONPG) remains to be investigated.

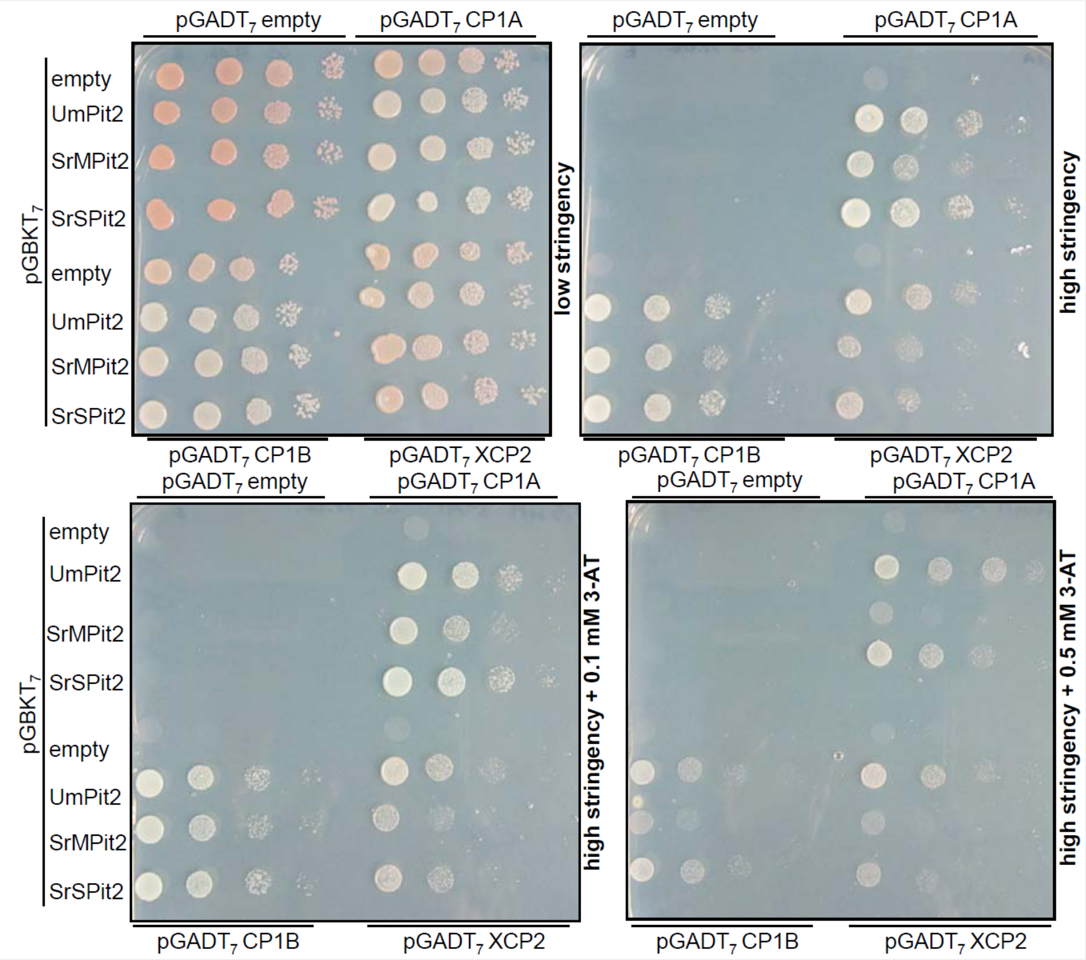


Figure 2.10: Yeast-2-hybrid analysis of interactions between Pit2 proteins of *U. mayids* and the two pathovariants of *S. reilianum* and maize papain-like cysteine proteases. Plasmid combinations are indicated at the left, top and bottom of each panel. Growth conditions are shown on the right site of each panel. Expressing *pit2* of *S. reilianum* f. sp. *zeae* resulted in a weaker growth compared with expression of *pit2* of *S. reilianum* f. sp. *sorghi* or *U. mayids*. UmPit2, *pit2* of *U. mayids*; SrMPit2, *pit2* of *S. reilianum* f. sp. *zeae*; SrSPit2, *pit2* of *S. reilianum* f. sp. *sorghi*. The figure is representative of two independent experiments.

To investigate whether differences in protein levels could explain the observed growth pattern, cells were disrupted by alkaline cell lysis and proteins were precipitated by adding trichloroacetic acid. Proteins were separated by SDS-PAGE. Pit2-orthologues and CP1A, CP1B or XCP2 were immunologically detected by Western blot using α-HA antibodies for the detection of PLCPs and α-c-myc antibodies for the detection of Pit2. A horse radish peroxidase-conjugated secondary antibody was used for chemo-luminescence emission. No differences in protein levels were observed (Figure 2.11).

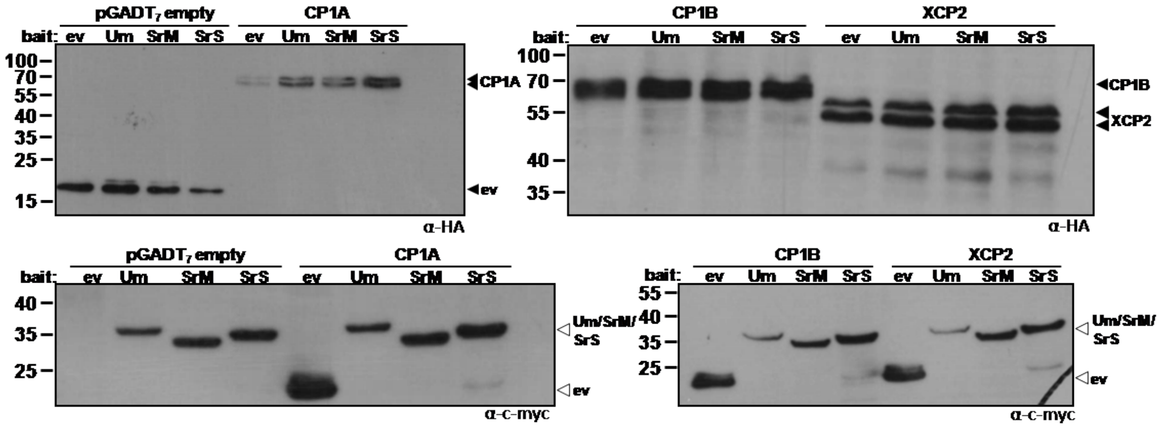
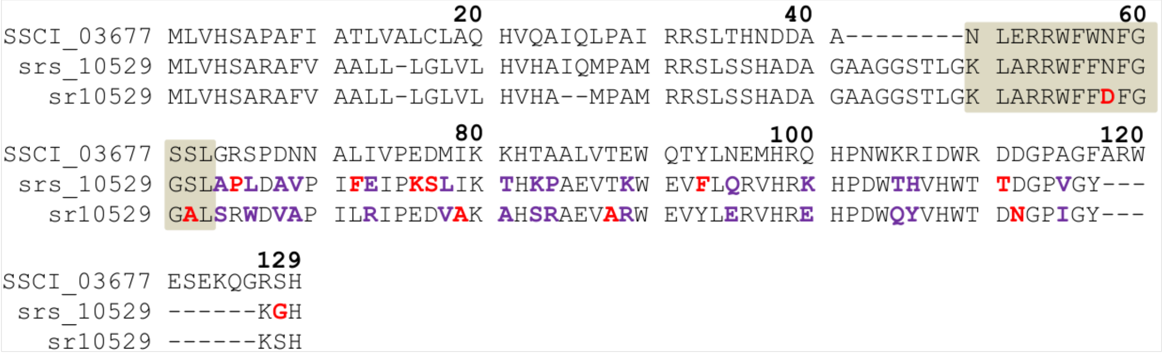


Figure 2.11: Western blot analysis of protein levels in yeast-2-hybrid interactions. The upper part shows α-HA detection of prey fusion proteins (empty vector, CP1A, CP1B or XCP2) in combination with different bait proteins (ev, empty vector; Um, UmPit2; SrM, SrMPit2; Srs, SrSPit2) as indicated above the blot. Filled arrows depict detected proteins. The lower part shows α-c-myc detection of bait fusion proteins as indicated above the blot. Open arrow heads indicate detected protein. Numbers onf the left site of each blot indicate the size standard in kDa.

2.3.2 The C-terminus of Pit2 of *S. reilianum* shows signs of positive selection

To identify residues of Pit2 of the two formae speciales in *S. reilianum* which are under positive selection and are therefore likely involved in the interaction with PLCPs, the branch-site model of PAML (Yang, 2007) was employed. Both *S. reilianum*-branches were defined as foreground. As a result, two residues of the previously identified conserved PID14 domain, which is in *U. maydis* sufficient for inhibition of PLCPs (Müller *et al.*, 2013), are under positive selection (Figure 2.12). Furthermore, additional residues under positive selection are scattered along the entire C-terminus (Figure 2.12).



```
                                20                                40                          60
SSCI_03677 MLVHSAPAFI ATLVALCLAQ HVQAIQLPAI RRSLTHNDDA A--------N LERRWFWNFG
 srs_10529 MLVHSARAFV AALL-LGLVL HVHAIQMPAM RRSLSSHADA GAAGGSTLGK LARRWFFNFG
   sr10529 MLVHSARAFV AALL-LGLVL HVHA--MPAM RRSLSSHADA GAAGGSTLGK LARRWFFDFG

                                80                               100                         120
SSCI_03677 SSLGRSPDNN ALIVPEDMIK KHTAALVTEW QTYLNEMHRQ HPNWKRIDWR DDGPAGFARW
 srs_10529 GSLAPLDAVP IFEIPKSLIK THKPAEVTKW EVFLQRVHRK HPDWTHVHWT TDGPVGY---
   sr10529 GALSRWDVAP ILRIPEDVAK AHSRAEVARW EVYLERVHRE HPDWQYVHWT DNGPIGY---

                  129
SSCI_03677 ESEKQGRSH
 srs_10529 ------KGH
   sr10529 ------KSH
```

Figure 2.12: Distribution of positively selected sites in Pit2. Alignment of Pit2-orthologues of *S. scitamineum* (top line), *S. reilianum* f. sp. *sorghi* (middle line) and *S. reilianum* f. sp. *zeae* (lower line) indicates residues under positive selection. Red amino acids indicate positive selection in the respective species and purple residues indicate positive selection in one or both species. The brown shaded area is homologous to the previously identified conserved PID14 domin, which is sufficient for inhibition of cysteine proteases (Müller *et al.*, 2013). Bold numbers indicated alignment sites.

2.3.3 One papain-like cysteine protease of *Sorghum bicolor* is under positive selection

The arms race model proposes an antagonistic co-evolution of pathogens an hosts. Therefore, the target of an effector should be under positive selection. To test whether PLCPs of maize and *Sorghum* show signs of positive selection, all proteins of maize and *Sorghum* which are homologous to the five previously identified SA-induced PLCPs of maize (van der Linde *et al.*, 2012) were used as data set. Inference of homology and scanning for positive selection was done as described in chapter 2.2.1. Only one PLCP of *S. bicolor* showed sings of positive selection ($\omega = 3.2$; Figure 2.13). Whether this accelerated evolution is due to the interaction with inhibitors of pathogens remains to be elucidated.

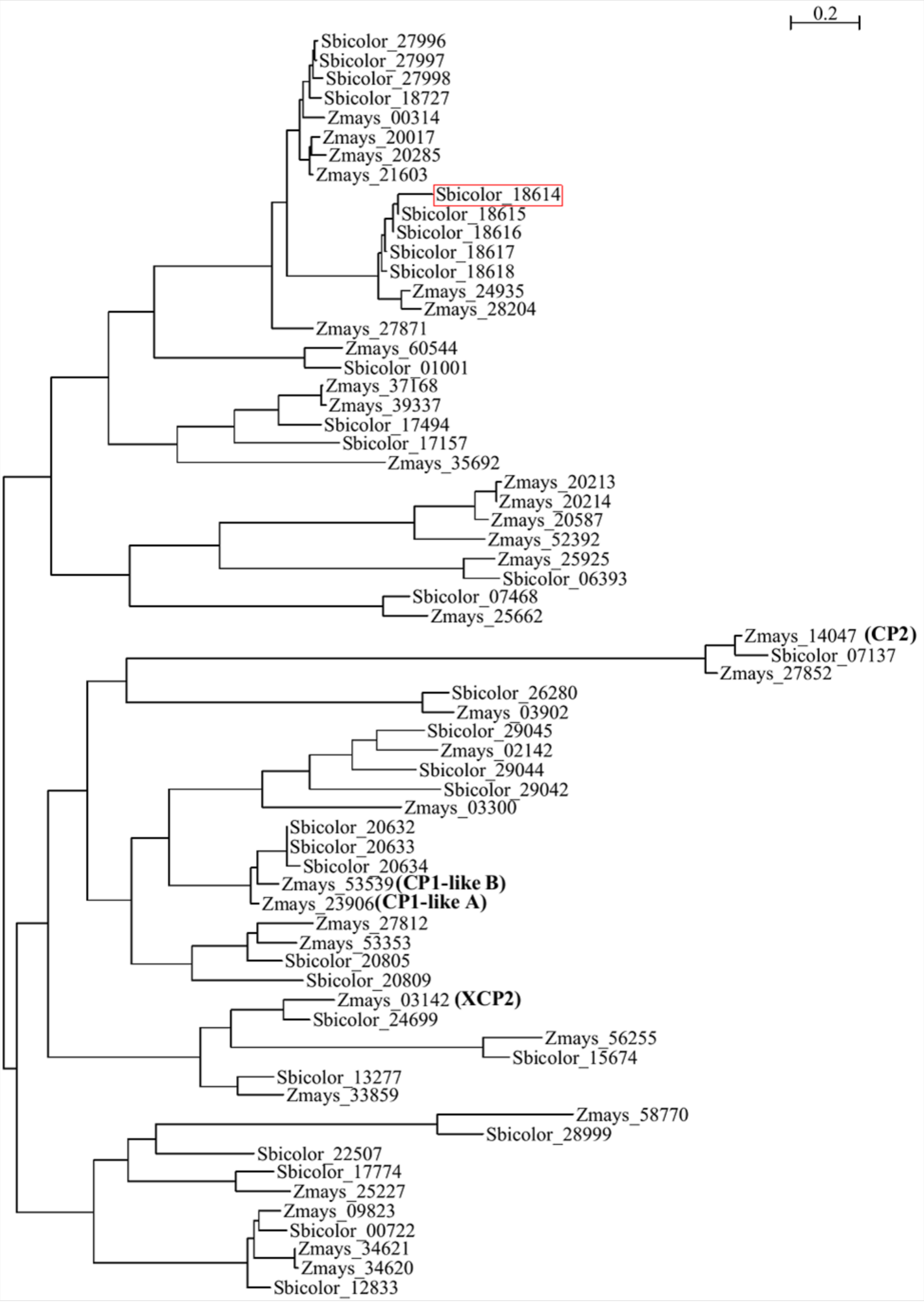


Figure 2.13: Phylogeny of PLCPs of maize (indicated by proteins named Zmays) and *S. bicolor* (indicated by proteins named Sbicolor). Positive selection was only inferred in Sbicolor_18614 as indicated by the red frame. Scale bar represents branch length as indicated.

3. DISCUSSION

In the present work, comparative pathogenomics of five related smut fungi with distinct host plants (*U. hordei*, *U. maydis*, *S. scitamineum*, *S. reilianum* f. sp. *zeae* and *S. reilianum* f. sp. *sorghi*) was used to investigate to which extent genes under positive selection and species-specific gene birth contribute to the evolution of virulence in smut fungi. Candidate genes from both evolutionary scenarios were assessed for their function in pathogenicity.

3.1 Orphan genes and their contribution to virulence in smut fungi

To assess whether orphan genes play a role in virulence of grass pathogenic smut fungi, the genomes of *U. hordei*, *U. maydis*, *S. scitamineum*, *S. reilianum* f. sp. *zeae* and *S. reilianum* f. sp. *sorghi* were compared. To obtain more reliable results, genomes of *M. globosa*, *P. flocculosa* and *M. pennsylvanicum* were in addition used in an all-against-all blastp search. Orphan genes resulting from this step were further verified by using them as query in a tblastn search against the non-redundant database of NCBI.

As a result, most orphan genes (in total 108 orphan genes among which 28 encode a predicted secreted protein) were found in *P. flocculosa*. This species is presumed to have lost its ability to colonize plants and is used as biocontrol agent (Lefebvre *et al.*, 2013). If *P. flocculosa* is indeed unable to infect plants, this finding would indicate that orphan genes are not specifically involved in pathogenicity. To explain the relative high number of orphan genes one could propose that the non-pathogenic life style of *P. flocculosa* might release evolutionary constraints on effector genes, since they do not need to evolve in response to changes of host targets any more. Therefore, previous effector genes of *P. flocculosa* might be free to accumulate mutations, which could diversify them to such an extent that orthologous genes are not detectable anymore. However, this neutral evolution would likely lead to the formation of pseudogenes, which should be frequently lost due to genetic drift (Francino, 2005) and should therefore not be detected in the genome. Thus, it is likely that the detected orphan genes still have a function. For example, they could be involved in the antagonistic interaction between *P. flocculosa* and powdery mildews, which is a unique trait of this species. This antagonism could be carried out at the level of secreted orphan proteins. Cytoplasmic orphan proteins could also contribute to this trait by acting in synthesis pathways of secondary metabolites.

A comparable number of orphan genes was found in the plant pathogens *M. pennsylvanicum* (in total 89 candidates among which 4 encode predicted secreted proteins), *U. maydis* (87/7) and *U. hordei* (73/4), which suggests that the adaptation of *M. pennsylvanicum* to a dicot host is not predominantly carried out by *de novo* gene birth.

Population data of 22 *U. maydis* strains showed that most orphan genes of this species are present in all 22 strains. This indicates that orphan genes contribute to important functions in natural environments. If their presence would be neutral, one would expect that more *U. maydis* strains lost these genes. So far, only one of the orphan genes discovered in the present study (*um12217*) was shown to play a role in virulence (Schilling *et al.*, 2014). The contribution of other candidates remains to be elucidated. It is attractive to propose a contribution of orphan genes to the formation of leaf tumors, which is a unique characteristic of *U. maydis*. This could be investigated by genome comparisons with the related species *Ustilago trichophora*, which can also induce leaf tumors on its host plant *Echinochloa sp.* (Tsukamot *et al.*, 1999). Detecting homologous sequences to the *U. maydis* orphans would dismiss the idea of true orphan genes, but could reveal insights in leaf tumor formation and potentially convergent evolution. In addition, it would be interesting to express orphan genes of *U. maydis* in *S. reilianum* f. sp. *zeae* and to assess whether such transformants might then induce tumors also in leaves or stems of infected maize plants.

The human pathogen *M. globosa* shows a lower number of orphan genes compared to the plant and non-pathogenic species in this group (in total 52 candidate genes of which 4 encode secreted proteins). If the reduction of orphan genes is related to its life style as human pathogen and whether the inferred orphan genes are involved in pathogenic traits remains to be elucidated. Due to their lower divergence, only few orphan genes could be detected in *Sporisorium* species (21/1, 8/1 and 0, respectively). The low number of total orphan genes and candidates encoding predicted secreted proteins suggests that orphan genes do not play a major role in plant colonization of these species.

3.2 Genes under positive selection in smut fungi

Besides creating new genes, polymorphisms in existing alleles are another source of genetic novelty. To detect genes under positive selection in *U. hordei*, *U. maydis*, *S. scitamineum*, *S. reilianum* f. sp. *zeae* and *S. reilianum* f. sp. *sorghi*, families of homologous sequences were built and a non-homogenous branch model of sequence evolution (Nielsen & Yang, 1998) was applied. It turned out that most candidate genes under positive selection were found in the two very closely related pathovariants *S. reilianum* f. sp. *zeae* (in total 84 genes) and *S. reilianum* f. sp. *sorghi* (in total 111 genes). In contrast, only few genes under positive selection were detected in the more distantly related species *S. scitamineum* (7 genes) and *U. maydis* (2 genes). However, 49 genes were found under positive selection in *U. hordei*, the most distantly related species in the group of smuts compared here. This contradictory finding can be explained by proposing different origins of genes under positive selection in these species. In *U. hordei*, several species-specific expansions of families could be inferred and 36 genes under positive selection are found in such families. Therefore, positive selection in *U. hordei* is predominantly found between paralogous genes. In contrast, genes under positive selection in the other four species belong to families of orthologous sequences. Because speciation times of *U. maydis* and *S. scitamineum* are large, many synonymous mutations could accumulate over time. This leads to saturation of the codon evolution model, as the high number of synonymous mutations results in $\mathrm{d_N/d_S}$ ratios lower than 1, even if non-synonymous mutations have occurred. Since the two pathovariants of *S. reilianum* are more than 98 % identical on the nucleotide level, saturation does not occur and most genes under positive selection are found in these species (84 and 111 candidates, respectively). The very high level of sequence diversity also prevented detection of positive selection in orthologues of *avirulence homolog* (*Avh*) genes in *Phytophthora sojae* and *Phytophthora ramorum*. Therefore, the authors restricted the inference of positive selection to paralogous sequences (Jiang *et al.*, 2008).

In *U. hordei* and the two formae speciales of *S. reilianum*, genes encoding predicted secreted proteins were found to be about three times over-represented in the set of positively selected genes compared to the set of genes without signs of accelerated evolution. This finding illustrates that secreted proteins are more likely under positive selection, consistent with their potential role in host interactions. This trend was also reported in other plant pathogenic fungi like poplar leaf rusts (Joly *et al.*, 2010) and the barley and wheat powdery mildews (Hacquard *et al.*, 2013; Wicker *et al.*, 2013). Together, these results highlight the importance of positive selection of putative effectors for successful plant colonization.

Although the group of positively selected genes is enriched for predicted secreted proteins, the majority of positively selected genes in *U. hordei*, *S. reilianum* f. sp. *zeae* and *S. reilianum* f. sp. *sorghi* does not encode secreted proteins. Despite wrong annotations, false negative predictions and the possibility of unconventional secretion, the respective proteins could be truly functioning in the cytoplasm. This role could for example include the production of secondary metabolites, which have been shown to be associated with host plant penetration and establishment of biotrophy (O'Connell *et al.*, 2012). Cytoplasmic proteins could also play important roles apart from pathogenicity. Such traits could for example involve the efficient competition with microorganism colonizing the same host plant. Since plants are colonized by numerous microorganisms (Vorholt, 2012), it is likely that smut fungi have to compete with a variety of microbial species. For example, smut fungi are likely suppressing the presence of pathogens with a necrotrophic life style. Necrotrophic fungi kill their host plant during colonization, which is in strong contrast to biotrophic interactions established by smut fungi. Initial experiments revealed a transcriptional and metabolic response of *U. maydis* to the presence of the hemibiotroph *Fusarium verticillioides* in axenic culture (Jonkers *et al.*, 2012). Interestingly, *U. maydis* and *F. verticillioides* are in nature frequently found in the same maize plant and even in the same tissue (Pan *et al.*, 2008). Maize plants co-infected with *U. maydis* and an endophytic strain of *F. verticillioides* grow larger compared to plants infected only with *U. maydis*. This suggests that *F. verticillioides* reduces pathogenicity of *U. maydis*. Furthermore, biomass of *U. maydis* was 20- to 60-fold lower in co-inoculation experiments compared to its biomass when infected alone (Rodriguez Estrada *et al.*, 2012).

Fungi can also serve as hosts for viral infections (Ghabrial & Suzuki, 2009; Drinnenberg *et al.*, 2011) and it is likely that targets of such infections are also found under positive selection. Furthermore, one could speculate that targets of natural and anthropogenic fungicides are also under positive selection.

Positively selected genes can also contribute to speciation processes. The cocktail of effectors allowing adaptation to a certain host plant can vary between strains. This could lead to strain isolation on different hosts. Hence, they can be drivers of ecological specialization and eventually speciation, because hybrids of strains adapted to different host plant might be less fit on either plant (Giraud et al., 2010). It would be interesting to elucidate whether this is true for the two S. reilianum strains. If hybridization of *S. reilianum* f. sp. *zeae* and *S. reilianum* f. sp. *sorghi* leads to offsprings that are less fit on either maize or *Sorghum* compared to their parents, this would suggest that hybrids are counter selected. Despite effectors, genes underlying mating, fusion of hyphae or dikaryon formation and maintenance can also contribute to speciation. It is conceivable that this trait is also carried out by positively selected genes, because it is likely that changes in protein structure play a role in ensuring incompatibility between strains.

3.3 Virulence function of putative effector proteins under positive selection

Genes showing signs of positive selection are assumed to play important roles in adaptations to the environment of an organism. Positive selection occurring in plant pathogens is attributed to an arms race carried out between effectors and their plant targets (Bishop *et al.*, 2000; Rohmer *et al.*, 2004; Liu *et al.* 2005; Stukenbrock & McDonald, 2007; Hacquard *et al.*, 2012). Therefore, secreted proteins found under positive selection are expected to play important roles in plant colonization.

In the present study, the contribution to virulence of single positively selected genes as well as gene clusters containing positively selected members was assessed by creating deletion mutants. Most genes under positive selection were found in the two pathovariants of *S. reilianum*. A solopathogenic strain of *S. reilianum* f. sp. *zeae* (JS161) was created previously, which eases genetic manipulations (Schirawski *et al.*, 2010). Therefore, this strain was used to elucidate virulence functions of potential effectors. Among strains carrying single gene deletions, only one candidate (*sr10529*; *srmpit2*) showed a strong contribution to virulence. This was also observed when deletion mutants of compatible wild type *S. reilianum* f. sp. *zeae* strains were analyzed. Importantly, no spore formation could be observed in deletion strains. *srmpit2* is homologous to the well characterized gene *pit2* of *U. maydis*, where it encodes an inhibitor of a group of salicylic acid-induced papain-like cysteine proteases of maize (Müller *et al.*, 2013). *pit2* deletion mutants in the solopathogenic *U. maydis* strain SG200 show also a drastic reduction of virulence (Döhlemann *et al.*, 2011). Interestingly, a recent study revealed that *pit2* is also found in *U. hordei* and *M. pennsylvanicum*, but absent in *M. globosa* and *Ceraceosorus bombacis*, an early lineage of smut fungi and a pathogen of cotton trees (Sharma *et al.*, 2015). This advances the idea that *pit2* was gained in the ancestor of *U. hordei* and *M. pennsylvanicum* and plays an important role in virulence in descendent species. Intriguingly, the effector proteins Tin3 and Stp1 of *U. maydis* are *in vitro* also capable of inhibiting the same class of PLCPs as Pit2. However, their overall effect on virulence upon deletion varies greatly (N. Neidig, T. Brefort and R. Kahmann, unpublished; K. Schipper, L. Liang and R. Kahmann, unpublished). Besides secreting effector proteins, *U. maydis* employs another strategy to inhibit PLCPs. Infections with SG200 were shown to transcriptionally activate a maize cystatine, which also inhibits PLCPs (van der Linde *et al.*, 2012). PLCPs were also identified as effector targets in other plant pathogens. For example, the tomato pathogen *Cladosprium fulvum* secrets the effector Avr2, which inhibits the apoplastic PLCP Rcr3 of tomato (Rooney *et al.*, 2005). Moreover, the nematode *Globodera rostochiensis* pathotype Ro-1-Mierenbos employs the effector Gr-VAP1 to also inhibit Rcr3 (Lozano-Torres *et al.*, 2012). Together, these results illustrate the importance of inhibiting PLCPs as prerequisite for pathogenicity in a variety of species.

In contrast to the striking virulence phenotype of *srmpit2* deletion mutants, seven deletions of single positively selected genes did not alter virulence. This finding can be in part attributed to the presence of paralogues: *sr10182* has 12 and *sr14347* has five paralogues. The other genes (*sr10059*, *sr12968*, *sr14558*, *sr14944* and *sr12897*) do not have paralogues (blastp e-value cutoff: 0.001). Nevertheless, these could be genes with redundant functions by acting in the same pathways that are targeted by other effectors. It could also be that effects on virulence are not apparent in the solopathogenic strain used here. This strain causes significantly reduced disease symptoms compared to crossings of compatible wild type strains. In particular, spore formation is only very rarely observed in JS161 (Schirawski *et al.*, 2010). In addition, a virulence phenotype could only be detectable in changes of plant physiological states. This was for example observed in deletion mutants of the chorismate mutase of *U. maydis* (*cmu1*). *cmu1* deletion mutants have only a weak macroscopic virulence phenotype, but infected maize plants show greatly elevated levels of salicylic acid, a plant hormone typically acting against biotrophs (Djamei *et al.*, 2011).

Some of the positively selected genes are located in clusters. Clusters contain paralogous sequences, which could have similar functions (Kämper *et al.*, 2006; Schirawski *et al.*, 2010; Dutheil *et al.*, in preparation). Therefore, entire cluster deletions were created in a first step. The contribution to virulence of each gene could then be assessed by complementing the cluster deletion strain with individual genes. Deletion of cluster 10-15 (*sr11226*-*sr11240*) in JS161 resulted in an almost complete absence of virulence. This cluster is homologous to the *U. maydis*-cluster 10A, whose deletion in SG200 also leads to a great decrease of virulence symptoms (Kämper *et al.*, 2006). Homologues of the positively selected genes in *S. reilianum* f. sp. *zeae* (*sr11233*, *sr11237*, *sr11239.2* and *sr11240*) are also found under positive selection in *S. reilianum* f. sp. *sorghi*. This result suggests that the virulence function of this cluster is conserved across smut fungi. To which extent each gene under positive selection contributes to the virulence phenotype of the cluster deletion remains to be investigated. Deletion of cluster 5-18 (*sr13421*-*sr13413*), which contains the positively selected genes *sr13419* and *sr13415* as well as deleteion of cluster 1-32 (*sr12084*-*sr12087*), which contains the positively selected gene *sr1084* resulted in increased virulence. Notably, deletions of regions in *U. maydis* homologous to these clusters did not affect virulence (Schirawski *et al.*, 2010; this work). This could suggest that the virulence regulating function of this group of effectors is ancestral and specifically lost in *U. maydis*. It is tempting to speculate that this loss could add to the ability of *U. maydis* of inducing tumors on all aerial parts of the plant. Alternatively, the new virulence function was gained in *S. scitamineum* after the speciation from *U. maydis* or even only recently in *S. reilianum*, potentially accompanied by the occurrence of positive selection. This is supported by the finding that the genes *srs_ 13419* and *srs_ 13415* of *S. reilianum* f. sp. *sorghi*, which are homologous to the positively selected cluster genes in *S. reilianum* f. sp. *zeae*, are also found to be under positive selection. Notably, cluster 5-18 of *S. reilianum* is enlarged compared to the homologous region in *U. maydis*, which could also explain a gain of a function in virulence for this cluster. However, five out of nine genes of this cluster are not predicted to encode secreted proteins; hence, this region was in a recent analysis not considered to constitute a cluster (Dutheil *et al.*, in preparation).

An opposite situation is found for deletion mutants of cluster 20-15 (*sr16549*-*sr16561*), which contains the positively selected genes *sr16550*, *sr16553*, *sr16556* and *sr16558*. Its deletion did not affect virulence, but deletion of the homologous region in *U. maydis* lead to attenuated pathogenicity (Schirawski *et al.*, 2010). This suggests that the effectors in cluster 20-15 are not required for seedling infections. Moreover, deletions of cluster 2-21 (*sr10308*-*sr10318*) did also not alter virulence, but deletion mutants of the orthologous cluster 2A in *U. maydis* resulted in a hypervirulent phenotype (Kämper *et al.*, 2006). This observation was attributed to the presence of the two genes *um01239* and *um01240*, which encode Virulence control protein 1 (Vcp1) and Vcp2 (K. Heidrich, A. Djamei and R. Kahmann, unpublished). The hypervirulent phenotype observed in *U. maydis* was explained by proposing an Avr-like function for Vcp1 and Vcp2, which means that the presence of these effectors attenuates virulence. Alternatively, Vcp1 and Vcp2 could actively restrict the proliferation of *U. maydis*, as heavy colonization might interfere with biotrophic development. The gene *sr10317* is found under positive selection in *S. reilianum* f. sp. *zeae* and is homologous to Vcp1 and Vcp2 in *U. maydis*. This could suggest that *sr10317* evolves towards escaping host recognition. Deleting cluster 2-21 including *sr10317* in JS161 did not affect virulence, which suggests that the encoded effectors do not play a virulence role under green house conditions. Alternatively, if the hypervirulent phenotype in *U. maydis* is explained by regulation of colonization, the missing virulence phenotype of infections with cluster deletion mutants in *S. reilianum* f. sp. *zeae* would mean that the regulatory effect on growth is not required in this system. It would be interesting to elucidate whether the change of function occurred already after the speciation from *U. maydis* (and is therefore also found in *S. scitamineum*) or only recently in *S. reilianum.* Given that the virulence function is lost after the split with *U. maydis*, it is not clear if and how the change in function is compensated in *Sporisorium* species. On the other hand, one could propose a gain of virulence regulation in *U. maydis*, which could in part explain its ability to induce virulence symptoms on all aerial parts of its host plant.

Deletion of cluster 12-15 (*sr15146-sr15149*) which contains the positively selected genes *sr15147* and *sr15149* did not reveal a contribution to virulence. Deleting the homologous region in *U. maydis* had also no effect on virulence (K. Münch and R. Kahmann, unpublished). The observation that entire cluster deletions do not affect virulence was previously reported (Kämper *et al.*, 2006; Schirawski *et al.*, 2010). These studies used seedling infections to elucidate a virulence function for cluster genes. Although no virulence phenotype could be observed in these experiments, it could well be that some effectors contribute to virulence in specific organs. Indeed, recent studies elucidated that some effectors of *U. maydis* function in an organ-specific manner (Skibbe *et al.*, 2010; Schilling *et al.*, 2014; Redkar *et al.*, 2015). This suggests that deletion mutants show a virulence phenotype only in certain organs and potentially only in specific tissues. Furthermore, these secreted molecules are potentially not contributing to virulence or are not needed under greenhouse conditions. It could also be that some effectors are obsolete for infections of domesticated maize plants in agricultural environments by *U. maydis* and *S. reilianum* f. sp. *zeae.* This may be explained by the intriguing speculation that these effectors were necessary for the successful colonization of teosinte, the wild progenitor of maize, but are dispensable for the infection of modern maize varieties. In addition, secreted proteins under positive selection can have a role not directly related to pathogenicity, like competing with other microbes present on the host plant.

A recent study addressing positively selected genes in *Z. tritici* could show that positively selected genes contribute to virulence in this species (Poppe *et al.*, 2015). In other cases, the investigated organisms where positively selected genes were identified cannot be grown in axenic culture or are not accessible for stable genetic manipulations. Some studies try to circumvent this problem by employing indirect approaches. For example, Kemen *et al.* (2011) expressed candidate effector genes of the *Arabidopsis*-pathogen *Albugo laibachii* in *Pseudomonas syringae* pv. tomato DC3000. To assess a potential virulence function of effectors in this heterologous system, growth rates were monitored. Another study used

Agrobacterium-infiltration for expression of candidate genes of *Phytophthora sojae* in *Nicotiana benthamiana*, where a hypersensitive response reaction could be detected (Win *et al.*, 2007). Although these studies made a contribution of positively selected candidate effectors to virulence likely, they allow only limited conclusions about the natural infection situation. Therefore, a global picture showing which fraction of positively selected genes contributes to virulence is missing.

3.4 Do putative effector proteins under positive selection play a role in host specificity?

It is conceivable that effector proteins not only play roles in coping with plant immune responses or in virulence, but also in determining the host range of fungal plant pathogens. The effector gene *srmpit2* shows strong signs of positive selection ($\omega = 31.147$) and its deletion lead to a strong reduction in virulence. One could speculate that SrMPit2 and its homolog in *S. reilianum* f. sp. *sorghi* SrSPit2 contribute to host specificity by efficiently inhibiting cysteine proteases of maize and *Sorghum*, respectively, but less efficiently of their non-host plants (Figure 3.1).

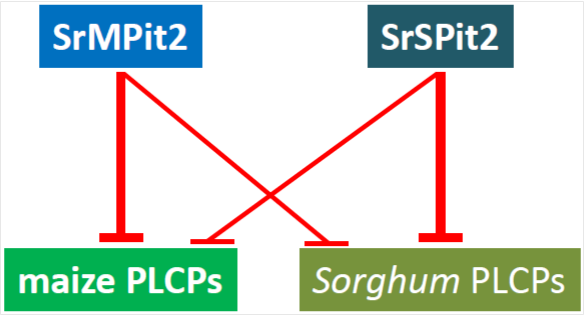


Figure 3.1: Model explaining a potential role for Pit2-orthologues in host specificity. Pit2 of *S. reilianum* f. sp. *zeae* (SrMPit2) can efficiently inhibit papain-like cysteine proteases (PLCPs) of its host plant maize (indicated by a thick red line). Similarly, Pit2 of *S. reilianum* f. sp. *sorghi* (SrSPit2) efficiently inhibits PLCPs of its host *Sorghum* (indicated by a thick red line). However, they inhibit PLCPs of their respective non-host plants to less extent (indicated by thin red lines).

To address this idea, the strength of interactions between the maize PLCPs XCP2, CP1A and CP1B and the Pit2-orthologues of *U. maydis* (UmPit2), *S. reilianum* f. sp. *zeae* (SrMPit2) and *S. reilianum* f. sp. *sorghi* (SrSPit2) was determined using a yeast-2-hybrid system. Surprisingly, the interaction between maize PLCPs and SrSPit2 was stronger than the interaction between maize PLCPs and SrMPit2. These preliminary results need to be corroborated by *in vitro* inhibition assays using purified proteins. If consistent results can be obtained, this finding could be explained by assuming that successful plant colonization requires fine meditated inhibition of PLCPs and that a too strong inhibition counteracts biotrophic establishment of *S. reilianum.* Alternatively, the result obtained in the yeast-2-hybrid analysis does not reflect the true inhibition capability of SrMPit2 and SrsPit2 of PLCPs *in planta.* To address the contribution of SrMPit2 and SrSPit2 to host specificity strains in which the two alleles of *pit2* are exchanged between *S. reilianum* f. sp. *zeae* and *S. reilianum* f. sp. *sorghi* are currently generated. If Pit2 plays a role in host specificity, strains expressing the endogenous *pit2* gene are expected to cause more virulence symptoms than strains expressing the orthologous *pit2* gene.

Interestingly, a contribution of positively selected cysteine protease inhibitors to host specialization was reported in *Phytophthora mirabilis* and *Phytophthora infestans.* Their orthologous effectors epiC1 and PmepiC1 inhibit efficiently the PLCPs of their respective host plants *Solanum sp.* (RCR3) and *Mirabilis jalapa* (MRP2), RCR3 and MRP2, but not the PLCPs of the respective non-host plant. Although the differences in inhibition efficiency were only shown *in vitro*, this suggests a role in host specificity for inhibitors of PLCPs in *Phytophthora sp.* (Dong *et al.*, 2014).

Besides interacting with important host targets, effectors contributing to host specificity are expected to be expressed at initial stages of infection, because it seems likely that the success of a plant-pathogen interaction is decided early. The recent finding that *pit2* in *U. maydis* is Msb2/Sho1-dependently induced *in vitro* by hydrophobic surfaces and/or hydroxy fatty acids suggests that Pit2 is indeed already needed early on for a successful plant colonization (Lanver *et al.*, 2014). This further supports the idea that Pit2-orthologs might contribute to host specialization.

A role of effector proteins under positive selection in host specialization was found in the wheat pathogen *Z. tritici.* Individual deletions of two effector genes in *Z. tritici* resulted in reduced pycnidia formation. One of the deletion mutants was complemented when expressing the orthologous gene of *Zymoseptoria ardabiliae* in this strain. In contrast, expressing orthologous genes of *Zymoseptoria pseudotritici* did not restore pycnidia formation (Poppe *et al.*,2015). This finding suggests that positively selected effectors can play a role in host specificity of plant pathogenic fungi.

3.5 Limitations of detecting positive selection as approach to identify virulence factors

In order to infer positive selection, families of homologous proteins were composed based on sequence identity and coverage cutoffs. Obtaining precise alignments is crucial for reliable reconstruction of phylogenies and eventually for the detection of positive selection (Schneider *et al.*, 2009; Jordan & Goldman, 2012). Since similar sequences can be aligned with higher confidence, rather strict settings for detection of homologies were applied. In addition, only alignment sites consistently found by two independent alignment programs were considered.

In this way, false positive predictions are avoided. However, this could be at the cost of missing biologically important candidate genes, because they were not considered for analysis under these settings, but could nevertheless play important roles in pathogenicity.

Scanning for positive selection as approach for the detection of virulence factors has another shortcoming: the inferred ω-value is an average value for a branch in a phylogeny. If a gene encodes a protein consisting of two domains, where one is under positive and one is under purifying selection, the global ω-value can be lower than or equal to 1, even though one domain shows an access of non-synonymous mutations. In addition, genes that begin to accumulate non-synonymous mutations are also missed in this approach, because the according ω-value does not yet exceed 1. Likewise, genes that have been under positive selection, but were subsequently under purifying selection, could show an ω-value below 1, and these genes are also not displayed as positively selected.

The used data set consists of very closely related species (the two pathovariants of *S. reilianum* show a genome identity of more than 98 %), but also of distantly related species (the genome identity between *U. hordei* and *S. reilianum* f. sp. *zeae* is only about 70 %). The largest number of positively selected genes was found in the pathovariants of *S. reilianum*, whereas almost no candidate genes were detected in *S. scitamineum* and *U. maydis*. This finding can be attributed to model saturation. It has been suggested that this problem could be solved by distinguishing radical and conservative amino acid replacements according to the physico-chemical properties of the original and substituted amino acid (Hughes *et al.*, 1990). The ratio of radical *vs.* conservative amino acid substitutions could also be used to scan for proteins involved in adaptation processes, because radical replacements are more likely to change or improve the function of a protein. However, this approach also has drawbacks, because the ratio is not only shaped by selection, but for example also by codon usage bias or transition *vs.* transversion ratios (Dagan *et al.*, 2002). In addition, even small changes of amino acid characteristics can provide a fitness advantage, but they will not be considered in this approach. Alternatively, population data of each species could be employed for the detection of positive selection and selection pressures of orthologous genes could be contrasted. Moreover, it would be interesting to have sequence data of additional species branching between *U. hordei*, *U. maydis* and *S. scitamineum* in a phylogenetic tree, because this sequence information would help to reduce the problem of large nucleotide divergences between species.

The present study aimed to identify single nucleotide polymorphisms under positive selection between homologous genes. Such genes are thought to evolve in an ‘arms race’. Despite molecular arms races, effectors and their plant targets can also be engaged in ‘trench warfare’. In this model, alleles of effectors and plant targets do not change over time (as in an arms race), but are preserved in a population in oscillating frequencies. In a simple model, which assumes that host plant resistance is carried out by *R* genes and that biotrophic pathogens are virulent if either an *avr* gene or its cognate *R* gene is missing, rare alleles are advantageous. If a plant genome contains an efficient *R* gene against pathogens, this allele will spread in the plant population. In turn, pathogens encoding a corresponding *avr* gene will show reduced fitness, because presence of the *avr* gene induces cell death. Therefore, pathogens in a population that do not encode that *avr* gene are favored. The low frequency of this *avr* gene in a population leads to a lower frequency of the cognate *R* in the host population, because its presence does not confer an advantage. If the frequency of a *R* gene is low, the low fraction of pathogens in a population which encode the corresponding *avr* gene have an advantage. In this way, resistance and susceptible alleles can persist over millions of years (Brown & Tellier, 2011). This type of selection was for example described in the *Puccinia chondrillina* - *Chondrilla juncea* (nakedweed) pathosystem (Chaboudez & Burdon, 1995), for *Linum – Melampsora* interactions (Thrall *et al.*, 2012) and for the *rpm1* locus of *Arabidopsis* species (Stahl *et al.*, 1infinite). While 'trench warfare' maybe common in natural ecosystems, agricultural habitats likely underlie selection of the 'arms race' model type, because susceptible plant cultivars will not be planted in fields anymore (Brown & Tellier, 2011). The smut fungi investigated in the study presented here are pathogens in agricultural ecosystems. This indicates that 'arms race' should predominate selection for pathogenicity and resistance. Nevertheless, it cannot be excluded that some effectors evolve according to 'trench warfare'.

This study aimed to uncover protein coding genes contributing to virulence. However, differences in virulence and potentially host specificity are not necessarily reflected at the level of protein sequences, but can also evolve at the level of regulation of gene expressions. Studies from model organisms like yeast and *Drosophila* showed that natural selection also affects large parts of the non-coding genome (Emerson *et al.*, 2010; Haddrill *et al.*, 2008). In a recent study, Rech *et al.* (2014) employed eight strains of the hemibiotrophic maize pathogen *Colletotrichum graminicola* to highlight selective pressures acting on different genomic regions. Five non-coding regions were distinguished: 5' and 3' untranslated regions (UTRs), 5' and 3' up- and downstream regions, respectively, and introns. Selection in these regions was identified by comparing the rate of nucleotide substitutions in these regions and the rate of synonymous substitutions in adjacent coding regions. It turned out that 3'-UTRs with signs of positive selection are predominantly associated with predicted secreted proteins, which could function as effectors. Therefore, strains encoding an identical effector gene, but with regulatory sequences under positive selection could still show variations in virulence. It would be interesting to elucidate whether similar patterns can be also found in smut fungi.

4. MATERIAL AND METHODS

4.1 Materials

4.1.1 Chemicals

All chemicals used in this study were obtained in the desired purity mainly from the companies Difco (Augsburg), Merck (Darmstadt), Roth (Karlsruhe) and Sigma-Aldrich (Deisenhofen).

4.1.2 Media and buffer

All media and buffers used in this study are listed under the respective method. If required, media and buffer were autoclaved (5 min, 121°C) or, in case of heat-sensitive material, filter sterilized (pore size: 0.2 μm; Life Technologies, Darmstadt).

4.1.3 Enzymes and antibodies

Restriction enzymes were obtained from New England Biolabs (Frankfurt). Phusion DNA polymerase F-530L (Thermofisher Scientific, Braunschweig) was used for polymerase chain reaction (PCR). Ligation of DNA molecules was performed using T4 DNA ligase (Roche, Mannheim). Digestion of fungal cell wall was done using Novocyme 234 (Novo Nordisc; Copenhagen, Denmark). Primary antibodies were obtained from Sigma-Aldrich and horse radish peroxidase-conjugated secondary antibodies from Cell Signaling Technology (Danver, USA).

4.1.4 Kits

Purifying of DNA fragments and PCR products from agarose gels was done with the Wizard SV Gel and PCR Clean-Up System (Promega, Mannheim). Plasmids were purified by employing the QIAprep Mini Plasmid Kit (Quiagen,, Hilden). Special kits are mentioned with the respective methods.

4.2 Cell culture

4.2.1 Cultivation of *Escherichia coli*

E. coli strains were grown in dYT at 37°C and 200 rmp. Antibiotics were added as selection marker when needed (Ampicillin, 100 μg/mL; Kanamycin 40 μg/mL).

dYT liquid medium: 1.6 % (w/v) Trypton-Pepton
1.0 % (w/v) Yeast Extract
0.5 % (w/v) NaCl
in $H_2O_{bid.}$

YT solid medium: 0.8 % (w/v) Trypton-Pepton
0.5 % (w/v) Yeast Extract
0.5 % (w/v) NaCl
1.3 % (w/v) Bactoagar
in $H_2O_{bid.}$

4.2.2 Cultivation of *Saccharomyces cerevisiae* and yeast-2-hybrid analysis

S. cerevisiae was grown at 28°C, (liquid cultures with shaking: 200 rpm) under aerobic conditions. YEPD was used as complete medium to maintain strains. Auxotrophic mutant strains were grown in SD medium supplemented with the respective amino acids.

YEPD medium: 2.0 % (w/v) Pepton
1 % (w/v) Yeast Extract
[solid medium: 2 % (w/v) Bactoagar]
in $H_2O_{bid.}$

SD medium: 0.67 % (w/v) Yeast Nitrogen Base w/o amino acids
0.16 % (w/v) DO supplements w/o adenine, histidine, leucine and tryptophan (Clontech)
2.0 % glucose (after autoclaving)
[solid medium: 2 % (w/v) Bactoagar]
in $H_2O_{bid.}$

For yeast-2-hybrid analyses, a single *S. cerevisiae* colony was grown over night in SD medium at 28°C, 200 rpm. This preculture was used to inoculate a main culture at an OD_{600} of 0.1 in SD medium. This culture was grown to an OD_{600} of 0.4 to 0.7 and adjusted with $H_2O_{bid.}$ to an OD_{600} of 1.0. This cell suspension was used for serial 1:10 dilutions up to 1:1000 and 6 μL of each dilution were dropped on SD solid medium, which selected for the presence of the transformed plasmids (low stringency; SD without leucine and without tryptophan) or selected in addition for interaction (high stringency; SD without leucin, tryptophan, adenine and histitdine). To further increase stringency, 0.1 mM 3-amin-otriazole (3-AT) or 0.5 mM 3-AT were added. Incubation occured at 28°C. The result was documented after 3 days by photography.

4.2.3 Cultivation of *Ustilago maydis* and *Sporisorium reilianum*

U. maydis was grown in YEPS$_L$ liquid medium at 28°C and 200 rpm. Potato-Dextrose (PD) agar plates were used as solid medium. Selection markers were added when needed (Hygromycin: 200 µg/mL; Geneticin: 50 µg/mL). Glycerol stocks for long term storage at -80°C were created by mixing a dense liquid culture with NSY-Glycerin (1:1).

YEPS$_L$:	1.0 % (w/v) Yeast Extract
(modified from Tsukuda *et al.*, 1988)	1.0 % (w/v) Peptone
	1.0 % (w/v) Saccharose
	in $H_2O_{bid.}$
PD solid medium:	3.9 % (w/v) Potato Dextrose Agar
	1.0 % (v/v) 1 M Tris-HCl (pH 8.0)
	in $H_2O_{bid.}$
NSY-Glycerin:	0.8 % (w/v) Nutrient Broth
	0.1 % (w/v) Yeast Extract
	0.5 % (w/v) Saccharose
	69.6 % (v/v) Glycerin
	in $H_2O_{bid.}$

4.2.4 Estimation of cell density

Cell densities of liquid cultures were estimated by using a photometer (Ultrospec 3000pro, Biochrom) at a wave length of 600 nm (OD$_{600}$). To ensure measurements of linear scale, cultures were diluted to an OD$_{600}$ between 0.3 and 0.8 when needed. The respective medium was used as reference. Cultures of *U. maydis* and *S. cerevisiae* with an OD$_{600}$ of 1.0 have about $1 - 5 \cdot 10^7$ cells/mL.

4.3 Strains, oligonucleotides and plasmids

4.3.1 *E. coli* strains

All experiments were performed with the strain TOP10 (Invitrogen, Karlsruhe), which is a derivative of *E. coli* K12. It has the following genotype:
F$^-$, *mcr*A, Δ(*mrr*-*hsd*RMS-*mcr*BC), ϕ80lacZΔM15, Δ*lac*X74, *deo*R, *rec*A1, *ara*D139, Δ(*ara-leu*)7697, *gal*U, *galK*, *rps*L(StrR), *end*A1, *nup*G

4.3.2 *S. cerevisiae* strains

The yeast-2-hybrid experiments were performed with the strain AH109 (Clonetech). It has the following genotype:
*MAT**a** trp1-901 leu2-3 ura3-52 his3-200 gal4Δ gal80Δ LYS2::*$GAL1_{UAS}$*-*$GAL1_{TATA}$*-HIS3* $GAL2_{UAS}$*-*$GAL1_{TATA}$*-ADE2 URA3::*$MEL1_{UAS}$*-*$MEL1_{TATA}$*-lacZ*

4.3.3 *U. maydis* and *S. reilianum* strains

All strains used in this study are listed in Table 4.1. Deletion mutants were created by replacing the gene(s) of interest with a resistance cassette according to Kämper (2004). All strains were verified by Southern analysis.

Table 4.1: Strains of *U. maydis* and *S. reilianum* f. sp. *zeae* used in this study

Strain	Genotype	Resistance[1]	Reference/Source
SG200	*a1:mfa2 bE1 bW2*	P	Kämper *et al.*, 2006
SG200ΔUm00792-Um00795	*a1:mfa2 bE1 BW2 um00792-um00795::hyg*	P, H	N. Rössel
SG200ΔUm02193	*a1:mfa2 bE1 bW2 um02193::hyg*	P, H	this work
S. reilianum f. sp. *zeae* SRZ 5-1 (JS60; sequenced)	*a2 b2*	-	Schirawski *et al.*, 2005
S. reilianum f. sp. *zeae* SRZ 5-2 (JS61)	*a1 b1*	-	Schirawski *et al.*, 2005
JS60ΔSrPit2	*a2 b2 sr10529::G418*	G	this work
JS61ΔSrPit2	*a1 b1 sr10529::G418*	G	this work
JS161	*a1 mfa2.1 bW1 bE2*	P	Schirawski *et al.*, 2010
JS161ΔSrPit2	*a1 mfa2.1 bW1 bE2 sr10529::hyg*	P, H	K. Münch
JS161Δ Sr10059	*a1 mfa2.1 bW1 bE2 sr10529::G418*	P, G	K. Münch
JS161ΔSr10182	*a1 mfa2.1 bW1 bE2 sr10182::G418*	P, G	K. Münch
JS161ΔSr12968	*a1 mfa2.1 bW1 bE2 sr12968::G418*	P, G	K. Münch
JS161ΔSr14558	*a1 mfa2.1 bW1 bE2 sr14558::G418*	P, G	K. Münch
JS161ΔSr14944	*a1 mfa2.1 bW1 bE2 sr14944::G418*	P, G	K. Münch
JS161ΔSr14347	*a1 mfa2.1 bW1 bE2 sr14347::G418*	P, G	K. Münch
JS161ΔSr12897	*a1 mfa2.1 bW1 bE2 sr12897::G418*	P, G	K. Münch
JS161ΔSr12084-Sr12087	*a1 mfa2.1 bW1 bE2 sr12084-sr12087::G418*	P, G	K. Münch
JS161ΔSr13421-Sr13413	*a1 mfa2.1 bW1 bE2 sr13421-sr13413::G418*	P, G	K. Münch
JS161ΔSr11226-Sr11240	*a1 mfa2.1 bW1 bE2 sr11226-sr11240::G418*	P, G	K. Münch
JS161ΔSr15149-Sr15147	*a1 mfa2.1 bW1 bE2 sr15149-sr15147::G418*	P, G	K. Münch
Js161ΔSr16549-Sr16561	*a1 mfa2.1 bW1 bE2 sr16549-sr16561::G418*	P, G	K. Münch
JS161ΔSr10308-Sr10318	*a1 mfa2.1 bW1 bE2 sr103108-sr10318::G418*	P, G	K. Münch

[1] P, Phleomycin; H, Hygromycin; G, Geneticin

4.3.4 Varieties of maize

Maize infections with *S. reilianum* and tassel infections with *U. maydis* strains were done using the dwarf variety 'Gaspe Flint'. For other *U. maydis* infections, the sweet corn variety 'Early Golden Bantam' was used (Olds Seed Company, Madison, USA).

4.3.5 Oligonucleotides

All oligonucleotides used in this study are listed in Table 4.2. They were ordered in the quality 'salt free' from Eurofins MWG Operon (Ebersberg). They were used for plasmid construction and sequencing.

Table 4.2: List of oligonucleotides used in this study

Name	**Sequence (5’ - 3’)[1]**	**Use[2]**
um02193 LB fwd	CTTGCCTTGGAGCTTGTTGG	amplifies upstream region of *um02193* [F]
um02193 LB rev	GATCGGCCATCTAGGCCAGTAGATCCTTGCCTCTTGC	amplifies upstream region of *um02193*; adds *Sfi*I site [R]
um02193 RB fwd	GATCGGCCTGAGTGGCCAAAGGAGGGTGTACAACTCC	amplifies downstream region of *um02193*; adds *Sfi*I site [F]
um02193 RB rev	GGCAACCGTTAACTTGAAGC	amplifies downstream region of *um02193* [R]
um01375_Y2H_f2	ACTGCCCGGGTATTCCGGTGCGTCGATCGCTC	amplifies *um01375* (*Umpit2*) w/o signal peptide; adds *Xma*I site [F]
um01375_Y2H_r	AGTCGGATCCTTATTCCCAGATGACCACATCTCC	amplifies *um01375* (*Umpit2*); adds *Bam*HI site [R]
sr10529_Y2H_f2	AGTCCATATGCTGGTCCACTCGGCGCG	amplifies *sr10529* (*SrMpit2*) w/o signal peptide; adds *Nde*I site [F]
sr10529_Y2H_r	AGTCCCCGGGTTAGTGGCTCTTGTACCCAATG	amplifies *sr10529* (*SrMpit2*); adds *Xma*I site [R]
srs10529_Y2H_f2	ACTGCCCGGGTATACAGATGCCAGCCATGC	amplifies *srs10529* (*SrSPit2*) w/o signal peptide; adds *Xma*I site [F]
srs10529_Y2H_r	AGTCGGATCCTTAGTGGCCCTTGTACCCAAC	amplifies *srs10529* (*SrSPit2*); adds *Bam*HI site [R]
sr10529 lb fw	GTTGAACTCGACATGGTGCG	amplifies upstream region of *sr10529* [F]
sr 10529 lb rev	GATCGGCCATCTAGGCCAAACGGATCGGCATGCAAGG	amplifies upstream region of *sr10529*; adds *Sfi*I site [R]
sr 10529 rb fw	GATCGGCCTGAGTGGCCTGTTTCGCATCATCGCACGG	amplifies downstream region of *sr10529*; adds *Sfi*I site [F]
sr 10529 rb rev	CTTTCCGTCACTCACTTCGC	amplifes downstream region of *sr10529* [R]
sr10059_lb_fw	GTAACGCCAGGGTTTTCCCAGTCACGACGAATATTTGGGAAGGCTCGAGCCA	amplifies upstream region of *sr10059*; adds *Ssp*I site [F]
sr10059_lb_rv	GAACTCGCTGGTAGTTACCACGTTCGGCCATCTAGGCCGGTGGATCGAAGGGAAATCG	amplifies upstream region of *sr10059*; restores *Sfi*I site [R]
sr10059_rb_fw	TCCGATGATAAGCTGTCAAACATGAGGCCTGAGTGGCCTGTAAACCAGGCTGCTGCAC	amplifies downstream region of *sr10059*; restores *SfiI* site [F]
sr10059_rb_rv	GCGGATAACAATTTCACACAGGAAACAGCAATATTCAGCAAAGATGAAGGGTC	amplifies downstream region of *sr10059*; adds *Ssp*I site [R]
sr10182_lb_fw	GTAACGCCAGGGTTTTCCCAGTCACGACGAATATTGCAGCATGCGAAGGTTG	amplifies upstream region of *sr10182*; adds *Ssp*I site [F]
sr10182_lb_rv	GAACTCGCTGGTAGTTACCACGTTCGGCCATCTAGGCCTGTCGGTGTCGTCTAGAGAG	amplifies upstream region of *sr10182*; restores *Sfi*I site [R]
sr10182_rb_fw	TCCGATGATAAGCTGTCAAACATGAGGCCTGAGTGGCCATGGGTCTTGCTCGGTTTCC	amplifies downstream region of *sr10182*; restores *Sfi*I site [F]

sr10182_rb_rv	GCGGATAACAATTTCACACAGGAAACAGCAATATTTCGGCAGCATCGCACGA	amplifies downstream region of *sr10182*; adds *Ssp*I site [R]
sr12968_lb_fw	GTAACGCCAGGGTTTTCCCAGTCACGACGAATATTGACGCCTCGAGGCCTTC	amplifies upstream region of *sr12968*; adds *Ssp*I site [F]
sr12968_lb_rv	GAACTCGCTGGTAGTTACCACGTTCGGCCATCTAGGCCTGGAAGAGAATGAGAGGAGG	amplifies upstream region of *sr12968*; restores *Sfi*I site [R]
sr12968_rb_fw	TCCGATGATAAGCTGTCAAACATGAGGCCTGAGTGGCCGTCACACCACGACGCTTCAC	amplifies downstream region of *sr12968*; restores *Sfi*I site [F]
sr12968_rb_rv	GCGGATAACAATTTCACACAGGAAACAGCAATATTTGGGAGACCAAGGCTCG	amplifies downstream region of *sr12968*; adds *Sfi*I site [R]
sr14558_lb_fw	GTAACGCCAGGGTTTTCCCAGTCACGACGAATATTTCATGTTTGCAGGGTCG	amplifies upstream region of *sr14558*; adds *Ssp*I site [F]
sr14558_lb_rv	GAACTCGCTGGTAGTTACCACGTTCGGCCATCTAGGCCGGAGAAGCTTGCTTGATGTG	amplifies upstream region of *sr14558*; restores *Sfi*I site [R]
sr14558_rb_fw	TCCGATGATAAGCTGTCAAACATGAGGCCTGAGTGGCCCGCTCTTTGTCATTTCCTGC	amplifies downstream region of*sr14558*; restores *Sfi*I site [F]
sr14558_rb_rv	GCGGATAACAATTTCACACAGGAAACAGCAATATTCAGTGCTCTCCCGTTTC	amplifies downstream region of *sr14558*; adds *Ssp*I site [R]
sr14944_lb_fw	GTAACGCCAGGGTTTTCCCAGTCACGACGAATATTCTTGGCATCCTGGTCTG	amplifies upstream region of *sr14944*; adds *Ssp*I site [F]
sr14944_lb_rv	GAACTCGCTGGTAGTTACCACGTTCGGCCATCTAGGCCGCGCTTGTGCTAGGTGAAAG	amplifies upstream region of *sr14944*; restores *Sfi*I site [R]
sr14944_rb_fw	TCCGATGATAAGCTGTCAAACATGAGGCCTGAGTGGCCGACAACGCGCAGGTTTCATC	amplifies downstream region of *sr14944*; restores *Sfi*I site [F]
sr14944_rb_rv	GCGGATAACAATTTCACACAGGAAACAGCAATATTTCGGGTTGCAAAGAAGC	amplifies downstream region of *sr14944*; adds *Sspi* site [R]
sr 14347 lb fw	AAAGCCAAGGTTACGACAGC	amplifies upstream region of *sr14347* [F]
sr 14347 lb rev	GATCGGCCATCTAGGCCCAATGTGGTTGTAGGACGGC	amplifies upstream region of *sr14347* [R]
sr 14347 rb fw	GATCGGCCTGAGTGCCTGTTCGATGCGACTTGTCGG	amplifies downstream region of *14347* [F]
sr 14347 rb rev	GTCCGTCGACATTTCATCGC	amplifies downstream region of *sr14347* [R]
sr12897 lb fw2	GTAACGCCAGGGTTTTCCCAGTCACGACGAATATT	amplifies upstream region of *sr12897*; adds *Ssp*I site [F]

	CGACAAGACGACCACCT	
sr12897 lb rv2	GAACTCGCTGGTAGTTACCACGTTCGGCCATCTAGGCCCTTGAACAGAGTAGGATTGG	amplifies upstream region of *sr12897*; restores *Sfi*I site [R]
sr12897 rb fw	TCCGATGATAAGCTGTCAAACATGAGGCCTGAGTGGCCGGCTTTCCAGCTCTCGTTGG	amplifies downstream region of *sr12897*; restores *Sfi*I site [F]
sr12897 rb rv	GCGGATAACAATTTCACACAGGAAACAGCAATATTCAGCCGCTGCCACTTCT	amplifies downstream region of *sr12897*; adds *Ssp*I site [R]
sr12084 lb fw	GTAACGCCAGGGTTTTCCCAGTCACGACGACTCGTTCAGCGCCTTTAGC	amplifies upstream region of *sr12084* [F]
sr12084 lb rv	GCGGCCGCAATTGTCACGCCATGGTGGCCATCTAGGCCAAGAATCGAGGGGGAAACAG	amplifies upstream region of *sr12084*; restores *Sfi*I site [R]
sr12087 rb fw	CTGTAGGAGTGCGGCCGCATTAATAGGCCTGAGTGGCCATACTGTCGGGTGCCAATGC	amplifies downstream region of *sr12087*; adds *Ssp*I site [F]
sr12087 rb rv	GCGGATAACAATTTCACACAGGAAACAGCGAGTCTCAACGTTGCTCTGC	amplifies downstream region of *sr12087* [R]
sr13421 lb fw	GTAACGCCAGGGTTTTCCCAGTCACGACGGTCTCGTGCAGCAATCAACG	amplifies upstream region of *sr13421* [F]
sr13421 lb rv	GCGGCCGCAATTGTCACGCCATGGTGGCCATCTAGGCCGCCTGACGATGCTTTCTTGG	amplifies upstream region of *sr13421*; resotres *Sfi*I site [R]
sr13413 rb fw	CTGTAGGAGTGCGGCCGCATTAATAGGCCTGAGTGGCCACTTTGTTTGCGGCACAAGG	amplifies downstream region of *sr13413*; restores *Sfi*I site [F]
sr13413 rb rv	GCGGATAACAATTTCACACAGGAAACAGCTTGCTGATCAGCCCTTCGAC	amplifies downstream region of *sr13413* [R]
sr11226 lb fw	GTAACGCCAGGGTTTTCCCAGTCACGACGATGGACGTATGCGGATCTGC	amplifies upstream region of *sr11226* [F]
sr11226 lb rv	GCGGCCGCAATTGTCACGCCATGGTGGCCATCTAGGCCAGTGTGTCGTTCGAGGTAGG	amplifies upstream region of *sr11226*; restores *Sfi*I site [R]
sr11240 rb fw	CTGTAGGAGTGCGGCCGCATTAATAGGCCTGAGTGGCCTGGGAGAGTCAGCGTGATTC	amplifies downstream region of *sr11240*; restores *Sfi*I site [F]
sr11240 rb rv	GCGGATAACAATTTCACACAGGAAACAGCTGAGAGCAGAGACGAGATCG	amplifies downstream region of *sr11240* [R]

sr15149G lb rv	GCGGCCGCAATTGTCACGCCATGGTGGCCATCTAGGCCCGTACTCTCGAAACGTTAGC	amplifies upstream region of *sr15149*; restores *Sfi*I site [R]
sr15149_lb_fw	GTAACGCCAGGGTTTTCCCAGTCACGACGAATATTCTGGGACGCCCAACACG	amplifies upstream region of *sr15149*; adds *Ssp*I site [F]
sr1514 rb fw	CTGTAGGAGTGCGGCCGCATTAATAGGCCTGAGTGGCCCCTGGCTGTCGAGTGATTTG	amplifies downstream region of*15146*; restores *Sfi*I site [F]
sr15146 rb rv	GCGGATAACAATTTCACACAGGAAACAGCTTGGAAGCGAGCTTCGCAAG	amplifies downstream region of *sr15146* [R]
sr16549 lb fw	GTAACGCCAGGGTTTTCCCAGTCACGACGAGAACGTCTGCAGCGTGTTG	amplifies upstream region of *sr16549* [F]
sr16549 lb rv	GCGGCCGCAATTGTCACGCCATGGTGGCCATCTAGGCCGCTTGGTGGGATGTTCTTGC	amplifies upstream region of *sr16549*; restores *Sfi*I site [R]
sr16561 rb fw	CTGTAGGAGTGCGGCCGCATTAATAGGCCTGAGTGGCCCGCCTCATGTTCACATGTCG	amplifies downstream region of *sr16561*; restores *Sfi*I site [F]
sr16561 rb rv	GCGGATAACAATTTCACACAGGAAACAGCGTTCCACTTTGCCAACATCC	amplifies downstream region of *sr16561* [R]
sr10308 lb fw	GTAACGCCAGGGTTTTCCCAGTCACGACGCTGCGCAGCATGTGTTGAAC	amplifies upstream region of *sr10308* [F]
sr10308 lb rv	GCGGCCGCAATTGTCACGCCATGGTGGCCATCTAGGCCTCGGATGCGGAAGAGGATGG	amplifies upstream region of *sr10308*; restores *Sfi*I site [R]
sr10318 rb fw	CTGTAGGAGTGCGGCCGCATTAATAGGCCGAGTGGCCTACTTTGAAGGGAGGTGTCC	amplifies downstream region of *sr10318*; restores *Sfi*I site [F]
sr10318 rb rv	GCGGATAACAATTTCACACAGGAAACAGCTGGTATCGCACTCGACGAGC	amplifies downstream region of *sr10318* [R]

---

[1] Sequence is shown in 5'-3' orientation; restriction sites are underlined

[2] Oligonucleotides hybridize either with the senes strand [R] or the complementary strand [F]

4.3.6 Plasmids

All plasmids created in this work are described here. Plasmid sequences were verified by sequencing (Eurofins MWG Operon, Ebersberg). pGBKT$_7$ and its derivatives confer Kanamycin resistance while all other plasmids confer Ampicillin resistance.

pTOPO_sr10529_Genet This plasmid carries the transformation construct for deleting *sr10529* in the *S. reilianum* f. sp. *zeae* wild type strains JS60 and JS61. It was created by replacing the Hygromycin cassette in pTV1 (T. Vellmer, personal communication) with the Geneticin resistance cassette of pUMa1057 (Baumann *et al.*, 2012). In this way, the Geneticin cassette is flanked by the up- and downstream regions of *sr10529*.

pGADT$_7$(ClonTech; Saint-Germain-en-Laye, France) This plasmid contains a GAL4 activation domain followed by an HA-epitope. It was used to created N-terminal GAL4AD-HA-fustions, which were tested for interaction with different preys in yeast-2-hybrid analysis. The plasmid carries the LEU2 auxotrophy marker.

pGADT$_7$-CP1A, pGADT$_7$-CP1B and pGADT$_7$-XCP2 These plasmids were published previously (Müller *et al.*, 2013) and contain the maize cysteine proteases CP1A, CP1B and XCP2, respectively, N-terminally fused to GAL4AD-HA (without activation domain).

pGBKT$_7$ (ClonTech; Saint-Germain-en-Laye, France) This plasmid encodes the GAL4 binding domain followed by a c-myc epitope. This plasmid was used for N-terminal fusions of GAL4BD-c-myc with different genes that should be tested for interaction in a yeast-2-hybrid analysis (bait plasmid). It contains the TRP1 auxotrophy marker.

pGBKT$_7$-um01375 This plasmid was used for the expression of *um01375* without signal peptide in a yeast-2-hybrid screen. Primers um01375_Y2H_f2 and um01375_Y2H_r were used to amplify a 0.3 kbp fragment from *U. maydis um01375* by PCR. The resulting fragment was digested with *Xma*I and *Bam*HI and ligated with the 7.3 kb *Xma*I/*Bam*HI fragment of pGBKT$_7$.

pGBKT$_7$-sr10529 This plasmid was used for the expression of *sr10529* without signal peptide in a yeast-2-hybrid screen. Primers sr10529_Y2H_f2 and sr10529_Y2H_r were used to amplify a 0.3 kbp fragment from the *S. reilianum* f. sp. *zeae* (JS60) gene *sr10529* by PCR. The resulting fragment was digested with *Xma*I and *Nde*I and ligated with the 7.3 kb *Xma*I/*Nde*I fragment of pGBKT$_7$.

pGBKT$_7$-srs10529 This plasmid was used for the expression of *srs10529* without signal peptide in a yeast-2-hybrid screen. Primers srs10529_Y2H_f2 and srs10529_Y2H_r were used to amplify a 0.3 kbp fragment from the *S. reilianum* f. sp. *sorghi* (JS678) gene *srs10529* by PCR. The resulting fragment was digested with *Xma*I and *Bam*HI and ligated with the 7.3 kbp *Xma*I/*Bam*HI fragment of pGBKT$_7$.

pTOPO Δsr10529 Hyg This plasmid contains the deletion construct consisting of up- and downstream flanking regions of *sr10529* and a Hygromycin resistance cassette. It was used to generate deletion mutants of *sr10529* in JS161. The plasmid was constructed by T. Vellmer.

pRS426 Δsr10059 Gen This plasmid contains a deletion construct consisting of up- and downstream flanking regions of *sr10059* and a Geneticin (G418) resistance cassette. It was used to generate deletion mutants of *sr10059* in JS161. The plasmid was kindly constructed by K. Münch.

pRS426 Δsr10182 Gen This plasmid contains a deletion construct consisting of up- and downstream flanking regions of *sr10182* and a Geneticin resistance cassette. It was used to generate deletion mutants of *sr10182* in JS161. The plasmid was kindly constructed by K. Münch.

pRS426 Δsr12968 Gen This plasmid contains a deletion construct consisting of up- and downstream flanking regions of *sr12968* and a Geneticin resistance cassette. It was used to generate deletion mutants of *sr12968* in JS161. The plasmid was kindly constructed by K. Münch.

pRS426 Δsr14558 Gen This plasmid contains a deletion construct consisting of up- and downstream flanking regions of *sr14558* and a Geneticin resistance cassette. It was used to generate deletion mutants of *sr14558* in JS161. The plasmid was kindly provided by K. Münch.

pRS426 Δsr14944 Gen This plasmid contains a deletion construct consisting of up- and downstream flanking regions of *sr14944* and a Geneticin resistance cassette. It was used to generate deletion mutants of *sr14944* in JS161. The plasmid was kindly constructed by K. Münch.

pTOPO Δsr14347 Gen This plasmid contains a deletion construct consisting of up- and downstream flanking regions of *sr14347* and a Geneticin resistance cassette. It was used to generate deletion mutants of *sr14347* in JS161. The plasmid was kindly constructed by K. Münch.

pRS426 Δsr12897 Gen This plasmid contains a deletion construct consisting of up- and downstream flanking regions of *sr12897* and a Geneticin resistance cassette. It was used to generate deletion mutants of *sr128897* in JS161. The plasmid was kindly created by K. Münch.

pRS426 Δsr12084-12087 Gen This plasmid contains a deletion construct consisting of the upstream region of *sr12084*, the downstream region of *12087* and a Geneticin resistance cassette. It was used to generate deletion mutants of cluster 1-32 (*sr12084-sr12087*) in JS161. The plasmid was kindly provided by K. Münch.

pRS426 Δsr13421-13413 Gen This plasmid contains a deletion construct consisting of the upstream region of *sr13421*, the downstream region of *13413* and a Geneticin resistance cassette. It was used to generate deletion mutants of cluster 5-18 (*sr13421-sr13313*) in JS161. The plasmid was kindly provided by K. Münch.

pRS426 Δsr11226-11240 Gen This plasmid contains a deletion construct consisting of the upstream region of *sr11226*, the downstream region of *11240* and a Geneticin resistance cassette. It was used to generate deletion mutants of cluster 10-15 (*sr11226-sr11240*) in JS161. The plasmid was kindly created by K. Münch.

pRS426 Δsr15149-15146 Gen This plasmid contains a deletion construct consisting of the upstream region of *sr15149*, the downstream region of *15146* and a Geneticin resistance cassette. It was used to generate deletion mutants of cluster 12-15 (*sr15149-sr15146)* in JS161. The plasmid was kindly provided by K. Münch.

pRS426 Δsr16549-16561 Gen This plasmid contains a deletion construct consisting of the upstream region of *sr16549*, the downstream region of *16561* and a Geneticin resistance cassette. It was used to generate deletion mutants of cluster 20-15 (*sr16549-sr16561*) in JS161. The plasmid was kindly provided by K. Münch.

pRS426 Δsr10308-10318 Gen This plasmid contains a deletion construct consisting of the upstream region of *sr10308*, the downstream region of *10318* and a Geneticin resistance cassette. It was used to generate deletion mutants of cluster 2-21 (*sr10308-sr10318)* in JS161. The plasmid was kindly provided by K. Münch.

4.4 Microbiological methods

4.4.1 Infections of *Z. mayds* with *U. maydis* and *S. reilianum*

To assess virulence of *U. maydis* and *S. reilianum* wild type and mutant strains, a preculture was grown in at test tube in $\mathrm{YEPS_L}$ over night at 28°C with shaking. A main culture was then inoculated with an $\mathrm{OD_{600}}$ of 0.1 and incubated at 28°, 200 rpm until an $\mathrm{OD_{600}}$ was reached (around 6 h of incubation). Cell culture was harvested by centrifugation (3500 rpm, 10 min, room temperature; Biofuge Stratos, Heraeus) and adjusted to an $\mathrm{OD_{600}}$ of 1.0 using sterile water. Compatible wild type strains were mixed 1:1 (v/v) prior to infections. Four maize seeds were planted in one flowerpot containing 'Frühstorfer Pikiererde' of type 'P' or 'H' and daily watered. Seven days old (in case of tassel infections: 15 days old) maize seedlings were infected by injecting cell suspensions in the leave whorl. The infection site was about 3 cm above the basal meristem. Plants were grown in a green house with 28°C (day) and 20°C (night). During the day phase, light intensity was a least 28,000 lux (with additional sun shine up to 90,000 lux). Relative humidity was between 40 % and 60 %. Symptom scoring was done 12 days post infection (dpi) in case of *U. maydis* seedling infections, 10 dpi (in case of *U. maydis* tassel infections) or 9 weeks post infection (in case of *S. reilianum* seedling infections. Symptom scoring occurred according to the categories described in Tables 4.3, 4.4 and 4.5, respectively. Routinely, each strain was infected in three independent experiments with 30 - 40 plants (10 flower pots). Data are presented as mean in relation to the total number of plants.

Table 4.3: Categories of virulence symptoms of *U. maydis* maize seedling infection (according to Kämper *et al.*, 2006)

Plant symptom	**Description**
No symptoms	No symptoms observable
Chlorosis / necrosis	Plant shows discoloration
Ligula swelling	Plant shows weak swellings at the ligula
Small tumors	Small tumors ($<$ 1 mm on leaves or very few tumors ($>$ 1 mm)
Normal tumors	Tumors on leaves and/or stem
Heavy tumors	Tumors on base of stem and/or change of growth axis
Dead	Plant died due to the infection

Table 4.4: Categories of virulence symptoms of *U. maydis* maize tassel infections (modified from Walbot & Skibbe, 2010)

Plant symptom	**Description**
No tassel	No tessel development was observed
Small tumors in $<$ 50 % of tassel length	Small tumors ($<$ 1 mm) in less than half of the tassel length
Small tumors in $\geq$ 50 % of tassel length	Small tumors ($<$ 1 mm) in at least half of the tassel length
Large tumors in $<$ 50 % of tassel length	Large tumors ($\geq$ 1 mm) in less than half of the tassel length
Large tumors in $\geq$ 50 % of tassel length	Large tumors ($\geq$ 1 mm) in at least half of the tassel length
Stunted tassel	Plant did not develop mature tassels

Table 4.5: Categories of virulence symptoms of *S. reilianum* f. sp. *zeae* maize seedling infections (modified from Ghareeb *et al.*, 2011)

Plant symptom	**Description**
No cobs	The maize plant did not develop cops
Healthy cobs $\leq$ 1 cm	The cob does not show virulence symptoms and its length is $\leq$ 1 cm
Healthy cobs $>$ 1 cm	The cob does not show virulence symptoms and its length is $>$ 1 cm
Spiky cobs	The cob shows spiky development, but no phyllody
Phyllody in cobs $\leq$ 1 cm	The cob developed phyllody and its length is $\leq$ 1 cm
Phyllody in cobs $>$ 1 cm	The cob developed pyhllody and its length is $>$ 1 cm
Spores	Spore formation was observed
Dead plants	The plant died due to the infection

4.4.2 Rubidium-chloride mediated transformation of *E. coli*

This protocol is modified after Cohen *et al.* (1972). In order to obtain chemo-competent *E. coli* strains, 100 mL dYT medium was supplemented with 10 mM $MgCl_2$ and 10 mM

$MgSO_4$. This medium was used for inoculation with 1 mL of an over night culture. This main culture was grown at 37°C and 200 rpm to an OD_{600} of 0.5 and harvested by centrifugation (15 min, 3000 rmp, 4°C; Biofuge Stratos, Heraeus). Cell pellet was resuspended in 33 mL ice cold RF-1 solution and incubated for 30 min to 60 min on ice. After another centrifugation step (15 min, 3000 rmp, 4°C; Biofuge Stratos, Heraeus), cells were resuspended in 5 mL ice cold RF-2 solution and incubated 15 min on ice. 50 µL of this cell suspension (1 aliquot) were shock frozen in liquid nitrogen and stored at -80°C.

RF-1 solution:	100 mM RbCl
	50 mM $MnCl_2 \cdot 2\ H_2O$
	30 mM Potassium acetate
	10 mM $CaCl_2 \cdot 2\ H_2O$
	15 % (v/v) Glycerin
	in $H_2O_{bid.}$
	adjust pH to 5.8 with acetate; filter sterilized

RF-2 solution:	10 mM 3-(*N*-morpholino)propanesulfonic acid (MOPS)
	10 mM RbCl
	75 mM $CaCl_2 \cdot 2\ H_2O$
	15 % (v/v) Glycerin
	in $H_2O_{bid.}$
	adjust pH to 5.8 with NaOH; filter sterilized

For transformations, one aliquot of cells was thawed on ice and mixed with up to 20 µL plasmid solution and 20 min incubated on ice. After a heat shock (1 min, 42°C), 150 mL dYT were added and the cell suspension was incubated at 37°C, 950 rpm for 45 min in a Thermomixer (Eppendorf). After that, the transformation mix was spread on plates with the respective antibiotic and incubated over night at 37°C.

4.4.3 Transformation of *S. cerevisiae*

S. cerevisiae was grown in 50 mL YEPD to an OD_{600} of 0.6, harvested (2000 rmp, 3 min; Biofuge Stratos, Heraeus) and washed in sterile $H_2O_{bid.}$. Cells were then resuspended in 10 mL SORB, again centrifuged and resuspended in 360 µL SORB. To this solution, 40 µL of heat denatured DNA-solution (10 mg/mL salmon sperm DNA; Invitrogen) were added. Aliquots of 15 µL were stored at -80°C for a couple of months.

For transformation, one aliquot was thawed and mixed with up to 10 µL of plasmid solution and 6 volumes of PEG and incubated for 30 min at room temperature. After a heat shock (15 min , 42°C), cells were centrifuged and washed once with YEPD. Cell suspension was then spread on SD plates with the respective selection marker and incubated at 28°C.

Colonies could be harvested after 2 - 3 days.

SORB: 100 mM Lithium acetate
1 mM Na_2-EDTA
1 M Sorbitol
in 10 mM Tris-HCl, pH 8.0 (filter sterilized)

PEG: 100 mM Lithium acetate
1 mM Na_2-EDTA
40 % (w/v) PEG 3350
in 10 mM Tris-HCl, pH 8.0 (filter sterilized)

4.4.4 Transformation of *U. maydis* and *S. reilianum*

For transformation, a protocol modified from Schulz *et al.* (1990) and Gillissen *et al.* (1992) was used. A cell culture was grown in 50 mL $YEPS_L$ (28°C, 200 rpm) until an OD_{600} of 0.8 - 1.0 was reached, harvested (3500 rpm, 5 min, room temperature; Biofuge Stratos; Heraeus) and resuspended in 25 mL SCS. After another centrifugation (3500 rpm, 5 min, room temperature), cells were resuspended in 2 mL Novozyme solution (2.5 mg/mL SCS; filter sterilized) and incubated at room temperature, until about 80 % of the cells begun to protoplast (about 2 - 10 min). The protoplast formation was monitored microscopically. The reaction was stopped by adding 20 mLSCS. After centrifugation (10 min, 2300 rpm, room temperature), cells were resuspended carefully in 20 mL SCS and again centrifuged (10 min, 2300 rpm, room temperature). Cells were resuspended in 10 mL SCS and after centrifugation (10 min, 2300 rpm, room temperature) resuspended in 20 mL STC. After centrifugation (10 min, 2400 rpm, room temperature), cells were resuspended in 0.5 mL ice cold STC. Aliquots (70 µL) were used immediately or stored at -80°C for several months.
For integrative transformations, one aliquot was thawed on ice, mixed with up to 10 µL linearized DNA (in total up to 5 µg) and 1 µL Heparin solution (10 mg/mL) and incubated 10 min on ice. Adding 0.5 mL STC/PEG was followed by another incubation for 15 min on ice. The total transformation mix was spread on a RegAgar plate containing double concentrated antibiotics. Plates were grown at 28°C and colonies were harvested after 4 - 7 days. They were plated on PD plates containing the respective antibiotic. Potential transformants were verified by Southern analysis.

SCS solution: 1 M Sorbitol
20 mM Sodium acetate
in $H_2O_{bid.}$ (pH 5.8; autoclaved)

STC solution: 1 M Sorbitol
10 mM Tris-HCl (pH 7.5)
100 mM $CaCl_2$
in $H_2O_{bid.}$ (autoclaved)

STC/PEG: 40 % (w/v) PEG 3350 in STC (filter sterilized)

RegAgar: 1.0 % (w/v) Yeast Extract
2.0 % (w/v) Bactp-Pepton
2.0 % (w/v) Sucrose
1 M Sorbitol
1.5 % (w/v) Bactoagar
in $H_2O_{bid.}$ (autoclaved)

4.5 Methods of molecular biology

4.5.1 Methods of *in vitro* modifications of nucleic acids

4.5.1.1 Restriction of DNA

Restrictions of DNA were carried out via type II endonucleases (NEB, Frankfurt) for 2 - 16 h at the enzyme-specific optimal temperature. A typical reaction mix was set up as follows:

X µL DNA (0.1 - 5 µg)
5 µL Enzyme-specific 10x buffer (NEB)
5 µL BSA (if required; NEB)
0.5 - 1 U Restriction endonuclease
ad 50 µL $H_2O_{bid.}$

4.5.1.2 Ligation of DNA fragments

To ligate DNA fragments, T4 ligase (Roche, Mannheim) was used. Ligations of a linearized vector and a DNA fragment were prepared in such a way that the fragment was present 3 times more than the vector. Ligations with more than two fragments were usually carried out in an equal molar ratio of all fragments. Typically, ligations were done in a total volume of 10 µL with 1 U T4 DNA ligase at 16°C over night.

4.5.1.3 Polymerase chain reaction

To amplify DNA fragmetns, the polymerase chain reaction (PCR) was used. For all reactions, Phusion Polymerase was employed. The PCR cycle consisted typically of the following steps: Initial denaturation (98°C, 1 min), denaturation (98°C, 10 sec), annealing (55°C, 30 sec), elongation (72°C, 30 sec), finale elongation (72°C, 10 min). In all cases, 35 cycles were run. The annealing temperature and elongation duration was adjusted to the primers used and to the length of the fragment (30 sec / 1 kpb), respectively. PCR reactions were performed in a Peqstar 96 universal gradient thermo cycler (Peqlab, Erlangen) or in a T Personal Thermocycler (Biometra, Göttingen). A typical reaction mix was set up as follows:

10.0 µL	5x HF-Buffer (Finnzymes)
1.5 µL	DMSO
0.4 µL	dNTPs (1:1:1:1 ratio)
1.0 µL	Oligonucleotide 1 (100 pM/µL)
1.0 µL	Oligonucleotide 2 (100 pM/µL)
1.0 µL	gDNA (as template)
0.5 µL	Phusion DNA polymerase F-530L
34.6 µL	$H_2O_{bid.}$

4.5.2 Isolation of nucleic acids

4.5.2.1 Isolation of plasmids from *E. coli*

Plasmids were isolated from a densely grown culture using the QIAprep Mini Plasmid Kit (Quiagen, Hilden) with 1.5 mL - 2.0 mL of culture.Plasmids were eluted with 30 µL - 50 µL $H_2O_{bid.}$. Yields were typically around 250 ng/µL

4.5.2.2 Isolation of genomic DNA from *U. maydis* and *S. reilianum*

The used approach is modified from Hoffman & Winston (1987). 4 mL of a dense over night culture in $YEPS_L$ were together with 200 µL (around 0.3 g) of glass beads centrifuged (5 min, 13.000 rpm, room temperature; Biofuge Pico, Heraeus). The pellet was resuspended in 500 µL *Ustilago*-lysis buffer and 500 µL TE-Phenol/Chloroform. Samples were shaken 15 min on a Vibrax VXR shaker (IKA, Staufen) at 1500 rpm. After centrifugation (20 min, 13.000 rpm, room temperature), which separates the phases, 400 µL of the supernatant were mixed with 1 mL 70 % (v/v) ethanol. After centrifugation (15 min, 13.000 rpm, room temperature; Biofuge Pico, Heraeus), the pellet was washed once with 500 µL 70 % (v/v) ethanol (5 min, 13.000 rpm, room temperature) and solved in 30 µL TE/RNase A (50:1) at 55°C for 15 min in a Thermomixer (Eppendorf). DNA was stored up to several months at -20°C.

Ustilago-lysis-buffer: 50 mM Na_2-EDTA
1 % (w/v) SDS
in 50 mM Tris-HCl (pH 7.5)

TE-Phenol/Chloroform: 1:1 mixture of phenol (equilibrated with TE-buffer)
and chloroform

TE-buffer: 1 mM Na_2-EDTA
in 10 mM Tris-HCl (pH 8.0)

4.5.3 Separation and detection of nucleic acids

4.5.3.1 Agarose-Gelelectrophoresis

DNA fragments were separated according to their size in an electric field, where DNA migrates due to its negative charge to the anode. The agarose concentration varied between 0.8 % and 1.0 % (w/v) in TAE-buffer according to the fragment length (shorter fragments were run with higher concentrations). Prior to use, agarose was supplemented with ethidium bromide (1 mg/mL). TAE served as buffer in the running chambers. DNA was mixed with loading buffer and transferred to the gel. Electrophoresis was run at 80 mA to 150 mA until the desired separation grade was reached. As size standards, either the 1 kb ladder (0.5 kb – 10 kb; NEB, Frankfurt) or the 100 bp ladder (0.1 kb – 1.5 kb; NEB, Frankfurt) was used. DNA was detected under UV light (254 nm). Photographs for documentation were taken with the BioDoc-IT-system; UVP).

50x TAE-Buffer: 2 M Tris-Base
2 M acetate
50 mM Na_2-EDTA
in $H_2O_{bid.}$

6x Loading buffer: 50 % (w/v) Saccharose
0.01 % (w/v) Bromphenol blue
in TE-buffer

4.5.3.2 Southern analysis

Genomic DNA was isolated from *U. maydis* and *S. reilianum* transformants as described in chapter 4.5.2.2. Around 5 µg of DNA were used for restrictions. Enzymes were chosen in such a way that the transformed constructs alter the number and/or length compared to the wild type locus. Restriction occurred over night at the enzyme-specific temperature. Restricted DNA was separated via agarose-gelelectrophoresis (TAE gel, 80 mA) and transferred to a nylon membrane with a method modified after Southern (1975). Prior to transfer, the gel was incubated in 0.25 M HCl for 20 min - 30 min, leading to depurination. After that, the gel was 15 min equilibrated in 0.4 M NaOH. Transfer to a positively charged Nylon Membrane (Roche, Mannheim) was carried out by using capillary forces created by a stack of paper towels and 0.4 M NaOH as transfer buffer. Due to this flux, DNA fragments are eluted from the gel and bind to the Nylon membrane. Transfer was done at room temperature over night (typically around 16 h).

To detect DNA fragments, probes were generated by PCR using the PCR DIG labeling mix (Roche, Mannheim). The recombination flanks (ca. 1 kb) served as template. The PCR products were purified from an agarose gel, eluted in 50 µL H_2O and mixed with 30 mL Southern Hybridization Buffer. Prior to use, the probe was denatured at 99°C for 20 min. Membranes were pre-hybridized with Southern Hybridization Buffer at 65°C for 30 min - 120 min and subsequently replaced by the denatured probe. Hybridization occurred for at least one day in a hybridization oven at 65°C under constant slow turning. The membrane was then washed twice with Southern Wash Buffer for 20 min at 65°C. After washing with DIG Wash Buffer (5 min, room temperature), the membrane was incubated in 20 mL - 30 mL DIG II buffer for 30 min - 60 min at room temperature. In this way, non-hybridized areas of the membrane were masked. The membrane was then incubated with 10 mL Antibody Solution for 30 min at room temperature. The antibody is covalently coupled to an alkaline phosphatase. After washing twice with DIG Wash Buffer (15 min, room temperature), the membrane was equilibrated with 30 mL DIG III buffer (5 min, room temperature). After incubation in 10 mL CDP Star Solution (5 min, room temperature), which serves as substrate for the phosphatase, excess solution was removed and the membrane was sealed in a plastic bag for further incubation (15 min, 37°C). For detection of a luminescence signal, the membrane was together with an X-ray film (Medical X-Ray Screen Film Blue Sensitive; CEA, Hamburg) placed in a film cassette. The signal was typically monitored for 10 min - 30 min and detected by developing the film in an x-ray film developer machine (QX-60; Konica or AGFA CP 1000; Mortsel, Belgium).

Na-Phosphate Buffer:	Solution 1: 1 M Na_2HPO_4 in $H_2O_{bid.}$ Solution 2: 1 M $NaH_2PO_4 \cdot H_2O$ in $H_2O_{bid.}$ Mix solution 1 and 2 (ratio ca. 4:1) (pH 7.0)
Southern Hybridization Buffer:	7.0 % (w/v) SDS in 0.5 M Na-Phosphate Buffer
Southern Wash Buffer:	1.0 % (w/v) SDS in 0.1 M Na-Phosphate Buffer
DIG I Buffer:	0.1 M Maleic acid 0.15 M NaCl in $H_2O_{bid.}$ adjust pH to 7.5 (with NaOH); autoclaved
DIG Wash Buffer:	0.3 % (v/v) Tween-20 in DIG I Buffer
DIG II Buffer:	1.0 % (w/v) Powdered Milk in DIG I Buffer
DIG III Buffer:	0.1 M NaCl 0.05 M $MgCl_2 \cdot 6\ H_2O$ in $H_2O_{bid.}$, adjust pH to 9.5 (with 1 M Tris-HCl)
Antibody Solution:	1 µL Anti-DIG antibody (Anti Digoxigenin Fab Fragment; Roche) in 10 mL DIG II Buffer
CDP Star Solution:	100 µL CDP Star (Roche) in 10 mL DIG II Buffer

4.6 Isolation and detection of proteins in yeast-2-hybrid analysis

Strains of *S. cerevisiae* used for yeast-2-hybrid analysis were grown in SD medium to an OD_{600} of 0.4 to 0.7 and adjusted to an OD_{600} of 1.0. 1 mL of this cell suspension was mixed with 150 µL Alkaline Lyses Buffer and incubated for 10 min on ice. After adding 150 µL of 55 % (v/v) trichloroacetic acid and another incubation step (10 min on ice), cell debris was pelleted by centrifugation (10 min, 13.000 rpm, room temperature). The pellet was resuspended in 100 µL HU-Buffer. Samples were heated (10 min, 65°C), spun (3 min, 13.000 rpm, room temperature) and 10 µL were used for separation by Sodiumdodecylsulfate-

Polyacrylamide-Gelelectrophoresis (SDS-PAGE).
Separation of protein samples was performed using SDS-PAGE with a method modified from Laemmli (1970). In this method, all proteins get a constant negative charge via the binding to SDS. This allows separation in an electric field. Chambers (Mini Protean System; Bio-Rad, München) were filled with SDS-Running Buffer. Gels were composed of a stacking gel and a separation gel. The stacking gel is used to concentrate the proteins in one layer prior to entering the separation gel. The separation gel separates the proteins in a polyacrylamide matrix according to their size, so that smaller proteins run faster. Separation was performed at 40 mA/gel. Protein mass was assessed by using a stained mixture of standard proteins (15 kDa – 170 kDa; Prestained Page Ruler; Fermentas, St. Leon-Roth).
Proteins were detected by an immunological assay using chemoluminescence. Proteins were transferred from the gel to a PVDF-membrane (GE Healthcare, München) with the transfer system Transfer-Blot Turbo (Bio-Rad, München). The membrane was activated by covering it shortly with methanol prior to use. The 'Mixed Protein Sizes' program (7 min) was used for blotting according to the manufacturer's instructions. The transferred proteins were immunologically detected. The membrane was incubated for 1 h at room temperature or over night at 4°C in Blocking Solution. After short washing with TBS-T, the membrane was incubated with antibody solution containing the primary antibody (mouse anti-HA, product number: #H9658, diluted 1:5000 or mouse anti-c-Myc, product number: #M5546, diluted 1:3000; both obtained from Sigma-Aldrich, Deisenhofen) at 4°C over night or for 1 h at room temperature with constant slow shaking. After washing three times with TBS-T for 15 min each, the membrane was incubated in Antibody Solution containing the secondary antibody (horse anti-mouse IgG, horse radish peroxidase (HRP)-linked, product number: 7076S, diluted 1:10000; Cell Signaling Technology, Danver, USA) for 1 h at room temperature with constant slow shaking. After washing three times with TBS-T for 15 min each, the membrane was transferred to a plastic bag and incubated with ECL (GE Healthcare, München), which serves as substrate for the HRP, for 5 min at room temperature. After removing excess solution, the membrane was sealed in a plastic bag and together with an X-ray film (Medical X-Ray Screen Film Blue Sensitive; CEA, Hamburg) placed in a film cassette. The signal was typically monitored for 2 min - 15 min and detected by developing the film in an x-ray film developer machine (QX-60; Konica or AGFA CP 1000; Mortsel, Belgium).

Alkaline Lyses Buffer:	2 M NaOH mixed with 2-Mercaptoethanol (ratio 12.3:1)
	prepared freshly prior of each experiment
HU Buffer:	8 M urea
	5 % (w/v) SDS
	200 mM $NaHPO_4$ (pH 6.8)
	0.1 mM EDTA
	0.1 % (w/v) Bromphenol blue
	in $H_2O_{bid.}$
	15 mg/mL DTT were added prior to use
SDS Running Buffer:	192 mM Glycine
	0.1 % (w/v) SDS
	in 25 mM Tris-HCl (pH 8.3)
Stacking Gel:	5 % (v/v) Acrylamid
	0.1 % (w/v) SDS
	in 125 mM Tris-HCl (pH 6.8)
to start polymerization:	0.1 % (w/v) Ammonium persulfate (APS)
	0.05 % (v/v) Tetramethylethylenediamine (TEMED)
Separation Gel:	12 % (v/v) Acrylamid
	0.1 % (w/v) SDS
	in 375 mM Tris-HCl (pH 8.8)
to start polymerization:	0.1 % (w/v) APS
	0.05 % (v/v) TEMED
TBS-T:	150 mM NaCl
	0.1 % (v/v) Tween20
	in 50 mM Tris-HCl (pH 7.5)
Blocking Solution:	10 % (w/v) Powdery milk in TBS-T
Antibody Solution:	diluted antibodies in 1 % (w/v) Powdery milk in TBST

4.7 Bioinformatic methods

4.7.1 Positively selected genes

4.7.1.1 Detection of positive selection between species

For this task it is important to build families of similar proteins, because obtaining reliable alignments is crucial for scanning for positive selection. Proteins of each species (*U. hordei*: 7,113 proteins, *U. maydis*: 6,787, *S. scitamineum*: 6,693, *S. reilianum* f. sp. *zeae*: 6,673, *S. reilianum* f. sp. *sorghi*: 6,674; all annotations as of November 2011) were used to perform a local blastp search (Altschul *et al.*, 1990), where the protein set served both as query and database (all-against-all search). To identify settings for coverage and identity, which lead to the maximum number of core families (i.e. families that have an equal number of members from each species), SiLiX (Miele *et al.*, 2011) was run with a range for coverage and identity between 5 % and 95 % (in 5 % steps).

Families with at least two members have been aligned using two alignment programs: MACSE, which aligns sequences on a codon level and accounts for frame shifts, produces both an amino acid and a nucleotide alignment (Ranwez *et al.*, 2011) and Prank, which considers insertions and deletions and also produces alignments based on amino acids and on nucleotides (Löytynoja & Goldman, 2008). The consensus alignment of the two software was determined using AlnScore, which is part of the Bio++ program suite (Dutheil *et al.*, 2006; Guéguen *et al.*, 2013). The nucleotide alignment created by Prank served as a reference. Consistent alignment sites with a maximum of 30 % gaps were used to translate the consensus nucleotide alignment to amino acid sequences using SeqMan, which is also part of the Bio++ programs. For families with at least three members, the translated sequences were used to create phylogenetic trees using PhyML 3.0 (Guindon *et al.*, 2010). Settings for PhyML were as follows: 'aa' (for amino acid sequences) was set as data type, a minimum parsimony starting tree was used (-p), the amino acid based default model LG was applied, character frequency option (-f) was set to m, distribution (-a) was set to e (maximum likelihood), number of relative substitution categories (-c) was set to 4 and tree topology search option (-s) was set to best (best of NNI and SPR search).

In the next step, BppML (Dutheil & Boussau, 2008) was used to fit the non-homogenous codon model YN98 (Nielsen & Yang, 1998), which allows the estimation of one ω value (d_N/d_S), one κ value (transversions/transitions) and the branch length for each branch of a phylogenetic tree. MapNH (Romiguier *et al.*, 2012) was used for mapping substituions. This was done by computing the liklehood for a homogenous substituion process *vs.* the alternative model of a heterogenous subsitution process. PartNH (Dutheil *et al.*, 2012) was used to build partitions of a phylogenitc tree according to the mapped parameters, but without any *a priori* assumptions. Model 'free' allows differing parameters between neighboring branches of the tree whereas model 'join' assumes that the parameters of neighboring branches are shared between them.

Since the goal of this analysis was to identify genes putatively contributing to virulence, the candidate set of genes under positive selection has been scanned for genes encoding predicted secreted proteins with SignalP 4.0 (Petersen *et al.*, 2011). Proteins were considered as predicted to be secreted if SignalP 4.0 indicates secretion and the absence of transmembrane domains.

4.7.1.2 Detection of positive selection in *U. maydis* population data

The genomes of 20 *U. maydis* strains-originating from Mexico were sequenced using Illumina paired end sequencing with a read length of 100 bp and a 100-fold coverage. A *de novo* assembly was performed using SOAPdenovo2 (Luo *et al.*, 2012). In order to estimate the ideal kmer-length for each strain, an assembly was performed using kmer lengths between 15 and 127. Since filtering of the reads (trimming the first 10 bp and requiring a minimum quality score of 35 with an exception at maximal 10 positions) did not improve the N_{50} value, a filtering step was not included. The assembly with the kmer size yielding the best N_{50} of contigs was used for further analysis.

Next, a multiple genome alignment with the assembled strains and the reference strain (*U. maydis* 521) as well as calling single nucleotide polymorphisms (SNPs) was conducted using MultiZ (Blanchette *et al.*, 2004). From this alignment, open reading frames were extracted with MafFilter (Dutheil *et al.*, 2014) according to the reference strain. The phylogeny of orthologous genes was inferred using PhyML 3.0 (Guindon *et al.*, 2010). Settings for PhyML were used as described before. Positive selection was inferred by applying a branch model (Yang & Nielsen, 1998) implemented in PAML4 (Yang, 2007) by comparing the M1a model (negative or neutral selection) with the M2a model (positive selection). Significant differences between the obtained maximum likelihood values of both models were detected using a χ^2 test. To infer potential effector proteins, prediction of secretion was done using SignalP 4.0 (Petersen *et al.*, 2011). Proteins were considered as predicted to be secreted if SignalP 4.0 indicates secretion and the absence of transmembrane domains.

4.7.1.3 Detection of positive selection in cysteine proteases of maize and *Sorghum*

Following the arms race model, not only effectors, but also their plant targets should be under positive selection. To detect cysteine proteases under positive selection in maize and *Sorghum*, the salicylic acid (SA)-induced maize cysteine proteases CP1-like A, CP1-like B, CatB3-like, XCP2 and CP2-like, which were identified in SA-infiltrated apoplastic fluid (van der Linde *et al.*, 2012) and which can be – with the exception of CatB3-like – inhibited by Pit2 (Müller *et al.*, 2013), where used as query to identify their closest homologues in

Sorghum bicolor in a blastp search. The best hits in *S. bicolor* and the original query files of CP1-like A, CP1-like B, CatB3-like, XCP2 and CP2-like were used as queries for a blastp search against the maize and *Sorghum* proteome. Building families of homologous sequences and detection of positive selection was done as described in chapter 4.7.1.1. The inference of families was done with an identity of 40 % and a coverage of 80 %.

4.7.1.4 Detection of sites of Pit2 under positive selection in *S. reilianum*

To infer which sites of Pit2 are under positive selection in the two pathovariants of *S. reilianum*, a branch-site model of PAML4 (Yang, 2007), which allows more than one d_N/d_S ratio per branch, was used. *S. reilianum* f. sp. *zeae* and *S. reilianum* f. sp. *sorghi* were *a priori* defined as foreground branches.

4.7.1.5 Parameters of positively selected genes between species

Tendency for cluster localization. Since many effectors are located in clusters, it was tested whether positively selected genes tend to reside in clusters. This was done by contrasting the fraction of positively selected genes residing in clusters with the fraction of not positively selected genes residing in clusters with Fisher's Exact Test. Cluster definitions were taken from Dutheil *et al.* (in preparation).

Localization within chromosomes. In some species, it has been described that effector genes tend to locate towards telomeres. To test whether this is also true in smut fungi, the relative physical distance to telomeres has been computed for each gene. The minimal distance was estimated by the distance between the midpoint of each gene and the closer telomere. This distance was divided by the length of the respective chromosome. In this way, the location relative to telomeres could be obtained. For example, if a gene is located in the proximity of a centromere, the relative distance to the closer telomere would be around 0.5. The relative distances to telomeres have been contrasted between genes showing signs of positive selection and genes not under positive selection with the Wilcoxon Rank-Sum Test.

Distance to repetitive elements. Since *U. hordei* shows the highest content of repetitive elements in the group of smut fungi investigated here, it was tested whether genes under positive selection are closer located to repetitive elements. For this analysis, only elements with at least 10 copies were considered. These included: BEL, Copia, DNA transposon, EnSpm, ERV1, Gypsy, Harbinger, hAT, Jockey, L1, Low complexity, MuDR, Polinton, Pseudogene, R1, Simple repeat, SINE2/tRNA, Sola and uncharacterized Interspersed repeats. The closest distance between each gene and each of the repetitive elements was computed. The result was contrasted for genes under positive selection and genes not under positive selection by the Wilcoxon Rank-Sum Test. To account for different copy numbers of the repetitive elements, the resulting p-value was multiplied with the copy number of each repetitive element (Bonferroni correction). P-values ≤ 0.05 after this correction step were considered significant.

4.7.2 Detection of orphan genes

For the prediction of orphan genes, the five smut genomes described before were employed. In addition, the genome of *M. globosa* (4,283 proteins), *M. pennsylvanicum* (6,280) and *P. flocculosa* (6,877) were used. A local blastp search was performed using all proteins as query and data base, respectively. The output was used to create families of homologous sequences using SiLiX (Miele *et al.*, 2011). To identify settings for coverage and identity, which lead to a confident prediction of orphan genes, a range of settings between 5 % and 95 % (in 5 % steps) was tested. To account for potential homologues outside of the 8 genomes initially used and to account for annotation errors, the orphan genes obtained with an identity and coverage of 5 % were used as query for a tblastn search against the non-redundant nucleotide data base of the National Center for Biotechnology Information (NCBI). Candidates were discarded, if a hit outside their own genome was found with an e-Value ≤ 0.001. Prediction of secretion was done with SignalP 4.0 (Petersen *et al.*, 2011). Proteins were considered as predicted to be secreted if SignalP 4.0 indicates secretion and the absence of transmembrane domains.

4.7.3 Horizontal gene transfer

To detect a potential horizontal gene transfer between *U. maydis* and *S. reilianum* f. sp. *zeae*, all families which contained one member of *U. hordei*, *U. maydis*, *S. scitamineum*, *S. reilianum* f. sp. *zeae* and *S. reilianum* f. sp. *sorghi* were considered. In addition, analysis was restricted to families whose phylogenetic tree had only branch lengths > 0.001. In this way, 3,402 families were kept for further analysis. Remaining trees were rerooted by using the *U. hordei* branch as out group. Next, cophenetic distances for each phylogeny were computed using the R library APE (Paradis et al., 2004). This distances were used as signs for potential horizontal gene transfers.

4.7.4 Inferring syntenic regions between *U. maydis* and *S. reilianum* f. sp. *zeae*

To assess whether genomic regions of *U. maydis* and *S. reilianum* f. sp. *zeae*, the Synteny Viewer of MIPS, hosted by the Helmhotz Zentrum München, Institute for Bioinformatics and Systems Biology was employed. Genes of interest were considered to locate in a syntenic region, if the two up- and downstream neighboring genes are syntenic and maximal one non-syntenic gene was found between these two adjacent genes. The Synteny Viewer can

be found at
http://mips.helmholtz-muenchen.de/gbrowse2/cgi-bin/gbrowse_syn/ust_um_uh_sr/

5. BIBLIOGRAPHY

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990): Basic local alignment search tool. JOURNAL OF MOLECULAR BIOLOGY, 215(3), 403-410.

Aguileta G, Refrégier G, Yockteng R, Fournier E and Giraud T (2009): Rapidly evolving genes in pathogens: Methods for detecting positive selection and examples among fungi, bacteria, viruses and protists. INFECTION, GENETICS AND EVOLUTION 9, 656-670.

Aguileta G, Lengelle J, Marthey S, Chiapello H, Rodolphe F, Gendrault A, Yockteng R, Vercken E, Devier B, Fontaine MC, Wincker P, Dossat C, Cruaud C, Couloux A and Giraud T (2010): Finding candidate genes under positive selection in non-model species: examples of genes involved in host specialization in pathogens. MOLECULAR ECOLOGY 19, 292-306.

Albá MM and Castresana J (2007): On homology searches by protein Blast and the characterization of the age of genes. BMC EVOLUTIONARY BIOLOGY 7:53.

Baer CF, Miyamoto MM and Denver DR (2007): Mutation rate variation in multicellular eukaryotes: causes and consequences. NATURE REVIEWS GENETICS 8, 619-631.

Bagchi R, Gallery RE, Gripenberg S, Gurr SJ, Narayan L, Addis CE, Freckleton RP and Lewis OT (2014): Pathogens and insect herbivores drive rainforest plant diversity and composition. NATURE 506, 85-88.

Banuett F and Herskowitz I (1989): Different a alleles of *Ustilago maydis* are necessary for maintenance of filamentous growth but not for meiosis. PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES USA 86, 5878-5882.

Banuett F and Herskowitz I (1996): Discrete developmental stages during teliospore formation in the corn smut fungus *Ustilago maydis*. DEVELOPMENT 122, 2965–2976.

Baumann S, Pohlmann T, Jungbluth M, Brachmann A and Feldbrügge M (2012): Kinesin-3 and dynein mediate microtubule-dependent co-transport of mRNPs and endosomes. JOURNAL OF CELL SCIENCE 125, 2740-2752.

Bishop JG, Dean AM and Mitchell-Olds T (2000): Rapid evolution in plant chitinases: Molecular targets of selection in plant-pathogen coevolution. textscProceedings of the National Academy of Sciences USA 97(10), 5322-5327.

Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AFA, Roskin KM, Baertsch R, Bosenbloom K, Clawson H, Green ED, Haussler D, and Miller W (2004): Aligning multiple genome sequences with the threaded blockset aligner. GENOME RESEARCH 14, 708-715.

Bölker M, Genin S, Lehmler C and Kahmann R (1995): Genetic regulation of mating and dimorphism in *Ustilago maydis*. CANADIAN JOURNAL OF BOTANY 73, 320-325.

Boller T and Felix G (2009): A Renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. Annual Review of Plant Biology 60, 379-406.

Brachmann A, Weinzierl G, Kämper J and Kahmann R (2001): Identification of genes in the bW/bE regulatory cascade in *Ustilago maydis*. Molecular Microbiology 42(4), 1047-1063.

Brefort T, Döhlemann G, Mendoza-Mendoza A, Reissmann S, Djamei A and Kahmann R (2009): *Ustilago maydis* as a pathogen. Annual Reviews of Phytopathology 47, 423-445.

Brefort T, Tanaka S, Neidig N, Doehlemann G, Vincon V and Kahmann R (2014): Characterization of the largest effector gene cluster of *Ustilago maydis*. PLOS Pathogens 10(7), e1003866.

Brennan MB and Struhl K (1980): Mechanisms of increasing expression of a yest gene in *Escherichia coli*. Journal of Molecular Biology 136(3), 333-338.

Brown JKM and Tellier A (2011): Plant-parasite coevolution: bridging the gap between genetics and ecology. Annual Reviews of Phytopathology 49, 345-367.

Cai JJ, Zhao R, Jiang H and Wang W (2008): *De novo* origination of a new protein-coding gene in *Saccharomyces cerevisiae*. Genetics 179, 487-496.

Chabouz P and Burdon JJ (1995): Frequency-dependent selection in a wild plant-pathosystem. Oecologia 102, 490-493.

Cohen SN, Chang AC and Hsu L (1972): Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. Proceedings of the National Academy of Sciences USA 69, 2110-2114.

Collier SM and Moffett P (2009): NB-LRRs work a “bait and switch” on pathogens. Trends in Plant Sciences 14(10), 521-529.

Croll D and McDonald BA (2012): The accessory genome as a cradle for adaptive evolution in pathogens. PLOS Pathogens 8(4), e1002608.

Dean R, van Kan JA, Pretorius ZA, Hammond-Kosack KE, Di Pietro A, Spanu PD, Rudd JJ, Dickman M, Kahmann R, Ellis J and Foster GD (2012): The Top 10 fungal pathogens in molecular plant pathology. Molecular Plant Pathology 13(4), 414-430.

Dagan T, Talmor Y and Graur D (2002): Ratios of radical to conservative amino acid replacements are affected by mutational and compositional factors and may not be indicative of positive Darwinian selection. Molecular Biology and Evolution 19(7), 1022-1025.

Djamei A, Schipper K, Rabe F, Gosh A, Vincon V, Kahnt J, Osorio S, Tohge T, Fernie AR, Feussner I, Feussner K, Meinicke P, Stierhof Y-D, Schwarz H, Macek B, Mann M and Kahmann R (2011): Metabolic priming by a secreted fungal effector. Nature 478, 395-398.

Djamei A and Kahmann R (2012): *Ustilago maydis*: Dissecting the molecular interface between pathogen and plant. PLOS Pathogens 8(11), e1002955.

Dodds PN and Rathjen JP (2010): Plant immunity: towards an integrated view of plant–pathogen interactions. Nature Reviews Genetics 11, 539-548.

Doehlemann G, van der Linde K, Aßmann D, Schwammbach D, Hof A, Mohanty A, Jackson D and Kahmann R (2009): Pep1, a secreted effector protein of *Ustilago maydis*, is required for successful invasion of plant cells. PLOS Pathogens 5(2), e1000290.

Doehlemann G, Reissmann S, Aßmann D, Fleckenstein M and Kahmann R (2011): Two linked genes encoding a secreted effector and a membrane protein are essential for *Ustilago maydis*-maize induced tumor formation. Molecular Microbiology 81(3), 751-766.

Dong S, Stam R, Cano LM, Sklenar J, Yoshida K, Bozkurt TO, Oliva R, Liu Z, Tian M, Win J, Banfield MJ, Jones AME, van der Hoorn RAL and Kamoun S (2014): Effector specialization in a lineage of the Irish potato famine pathogen. Science 343(6170), 552-555.

Drinnenberg IA, Fink GR and Bartel DP (2011): Compatibility with Killer explains the raise of RNAi-deficient fungi. Science 333(6049), 1592.

Duplessis S, Cuomo CA, Lin Y-C, Aerts A, Tisserant E, Veneault-Fourrey C, Joly DL, Hacquard S, Amselem J, Cantarel BL, Chiu R, Coutinho PM, Feau N, Field M, Frey P, Gelhaye E, Goldberg J, Grabherr MG, Kodira CD, Kohler A, Kües U, Lindquist EA, Lucas SM, Mago R, Mauceli E, Morin E, Murat C, Pangilinan JS, Park R, Pearson M, Quesneville H, Rouhier N, Sakthikumar S, Salamov AA, Schmutz J, Selles B, Shapiro H, Tanguay P, Tuskan GA, Henrissat B, van de Peer Y, Rouzé P, Ellis JG, Dodds PN, Schein JE, Zhong S, Hamelin RC, Grigoriev IV, Szabo LJ and Martin F (2011): Obligate biotrophy features unraveled by the genomic analysis of rust fungi. Proceedings of the National Academy of Sciences USA 108(22), 9166-9171.

Dutheil JY, Gaillard S, Bazin E, Glémin S, Ranwez V, Galtier N and Belkhir K (2006): Bio++: a set of C++ libraries for sequence analysis, phylogenetics, molecular evolution and population genetcis. BMC Bioinformatics 7, 188.

Dutheil JY and Boussau B (2008): Non-homogeneous models of sequence evolution in the Bio++ suite of libraries and programs. BMC Evolutionary Biology 8:255.

Dutheil JY, Galtier N, Romiguier J, Douzery EJP, Ranwez V and Boussau B (2012): Efficient selection of branch-specific models of sequence evolution. Molecular Biology and Evolution 29(7), 1861-1874.

Dutheil JY, Gaillard S and Stukenbrock EH (2014): MafFilter: A highly flexible and extensible multiple genome alignment files processor. BMC Genomics 15, 53.

Dutheil JY, Mannhaupt G, Schweizer G, Sieber CMK, Münsterkötter M, Güldener U, Schirawski J and Kahmann R: The evolution of virulence clusters in smut fungi [in preparation].

Emerson JJ, Hsieh L-C, Sung H-M, Wang T-Y, Huang C-J, Lu HH-S, Lu M-YJ, Wu S-H, Li W-H (2010): Natural selection acts on cis and trans regulation in yeasts. GENOME RESEARCH 20, 826-836.

Felix G, Regenass M and Boller T (1993): Specific perception of subnanomolar concentrations of chitin fragments by tomato cells: induction of extracellular alkalinization, changes in protein phosphorylation, and establishment of a refractory state. PLANT JOURNAL 4,307-316.

Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL and Gurr SJ (2012): Emerging fungal threats to animal, plant and ecosystem health. NATURE 484, 186-194.

Flor HH (1971): Current status of the gene-for-gene concept. ANNUAL REVIEW OF PHYTOPATHOLOGY 9, 275-296.

Francino MP (2005): An adaptive radiation model for the origin of new gene functions. NATURE GENETICS 37(6), 573-577.

Gehrig H, Schüssler A and Kluge M (1996): *Geosiphon pyriforme*, a fungus forming endocytobiosis with *Nostoc* (cyanobacteria), is an ancestral member of the *Glomales*: evidence by SSU rRNA analysis. JOURNAL OF MOLECULAR EVOLUTION 43(1), 71-81.

Ghabrial SA and Suzuki N (2009): Viruses of plant pathogenic fungi. ANNUAL REVIEW OF PHYTOPATHOLOGY 47, 353-384.

Ghareeb HA, Becker A, Iven T, Feussner I and Schirawski J (2011): *Sporisorium reilianum* infection changes inflorescence and branching architecture of maize. PLANT PHYSIOLOGY 156, 2037-2052.

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

Giraud T, Gladieux P and Gavrilets S (2010): Linking the emergence of fungal plant diseases with ecological speciation. TRENDS IN ECOLOGY AND EVOLUTION 25, 387-395.

Guéguen L, Gaillard S, Boussau B, Gouy M, Groussin M, Rochette NC, Bigot T, Fournier D, Pouyet F, Cahais V, Bernard A, Scornavacca C, Nabholz B, Haudry A, Dachary L, Galtier N, Belkhir K and Dutheil JY (2013): Bio++: Efficient extensible libraries and tools for computational molecular evolution. MOLECULAR BIOLOGY AND EVOLUTION 30(8), 1745-1750.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W and Gascuel O (2010): New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. SYSTEMATIC BIOLOGY 59(3), 307-321.

Haas BJ, Kamoun S, Zody MC, Jiang RHY, Handsaker RE, Cano LM, Grabherr M, Kodira CD, Raffaele S, Torto-Alalibo T, Bozkurt TO, Ah-Fong AMV, Alvarado L,

Anderson VL, Armstrong MR, Avrova A, Baxter L, Beynon J, Boevink PC, Bollmann SR, Bos JIB, Bulone V, Cai G, Cakir C, Carrington JC, Chawner M, Conti L, Costanzo S, Ewan R, Fahlgren N, Fischbach MA, Fugelstad J, Gilroy EM, Gnerre S, Green PJ, Grenville-Briggs L-J, Griffith J, Grünwald NJ, Horn K, Horner NR, Hu C-H, Huitema E, Jeong D-H, Jones AME, Jones JDG, Jones RW, Karlsson EK, Kunjeti SG, Lamour K, Liu Z, Ma LJ, MacLean D, Chibucos MC, McDonald H, McWalters J, Meijer HJG, Morgan W, Morris PF, Munro CA, O'Neill K, Ospina-Giraldo M, Pinzón A, Pritchard L, Ramsahoye B, Ren Q, Restrepo S, Roy S, Sadanandom A, Savidor A, Schornack S, Schwartz DC, Schumann UD, Schwessinger B, Seyer L, Sharpe T, Silvar C, Song J, Studholme DJ, Sykes S, Thines M, van de Vondervoort PJI, Phuntumart V, Wawra S, Weide R, Win J, Young C, Zhou S, Fry W, Meyers BC, van West P, Ristaino J, Govers F, Birch PRJ, Whisson SC, Judelson HS and Nusbaum C (2009): Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans.* NATURE 461, 393-398.

Han Y, Liu X, Benny U, Kistler HC and van Etten HD (2001): Genes determining pathogenicity to pea are clustered on a supernumerary chromosome in the fungal plant pathogen *Nectaria haematococca.* PLANT JOURNAL 25, 305-3014.

Hacquard S, Joly DL, Lin Y-C, Tisserant E, Feau N, Delaruelle C, Legué V, Kohler A, Tanguay P, Petre B, Frey P, van de Peer Y, Rouze P, Martin F, Hamelin RC and Duplessis S (2012): A comprehensive analysis of genes encoding small secreted proteins identifies candidate effectors in *Melampsora larici-populina* (Poplar leaf rust). MOLECULAR PLANT-MICROBE INTERACTIONS 25(3), 279-293.

Hacquard S, Kracher B, Maekawa T, Vernaldi S, Schulze-Lefert P and van Themaat EVL (2013): Mosaic genome structure of the barley powdery mildew pathogen and conservation of transcriptional programs in divergent hosts. PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES USA 110(24), E2219-E2228.

Haddrill PR, Bachtrog D and Adolfatto P (2008): Positive and negative selection on non-coding DNA in *Drosophila simulans.* MOLECULAR BIOLOGY AND EVOLUTION 25, 1825-1834.

Heinen TJAJ, Staubach F, Hämling D and Tautz D (2009): Emergence of a new gene from an intergenic region. CURRENT BIOLOGY 19, 1527-1531.

Hemetsberger C, Herrberger C, Zechmann B, Hillmer M and Doehlemann G (2012): The *Ustilago maydis* effector Pep1 suppresses plant immunity by inhibition of host peroxidase activity. PLOS PATHOGENS 8(5), e1002684.

Hemetsberger C, Mueller AN, Matei A, Herrberger C, Hensel C, Kumlehn J, Mishra B, Sharma R, Thines M, Hückelhoven R and Doehlemann G (2015): The fungal core effector Pep1 is conserved across smuts of dicots and monocots. NEW PHYTOLOGIST 206, 1116-1126.

Hoffman CS and Winston F (1987): A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *E. coli.* GENE 57, 267-272.

Hua J (2013): Modulation of plant immunity by light, circadian rhythm, and temperature. CURrent Opinion in Plant Biology 16, 40-413.

Huang J, Si W, Deng Q, Li P and Yang S (2014): Rapid evolution of avirulence genes in rice blast fungus *Magnaporthe oryzae.* BMC Genetics 15:45.

Hughes AL, Ota T and Nei M (1990): Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of Class I Major-Histocompatibility-Complex molecules. Molecular Biology and Evolution 7(6), 515-524.

Islam MS, Haque MS, Islam MM, Emdad EM, Halim A, Hossen QMM, Hossain MZ, Ahmed B, Rahim S, Rahman MS, Alam MM, Hou S, Wan X, Saito JA and Alam M (2012): Tools to kill: Genome of one of the most destructive plant pathogenic fungi *Macrophomina phaseolina.* BMC Genomics 13:493.

Jacob F (1977): Evolution and tinkering. Science 196(4295), 1161-1166.

Jiang RHY, Tripathy S, Govers F and Brett MT (2008): RXLR effector reservoir in two *Phytophthora* species is dominated by a single rapidly evolving superfamily with more than 700 members. Proceedings of the National Academy of Sciences USA 105(12), 4874-4879.

Joly DL, Feau N, Tanguay P, Hamelin RC (2010): Comparative analysis of secreted protein evolution using expressed sequence tags from four popular leaf rusts (*Melampsora ssp.*). BMC Genomics 11:422.

Jonkers W, Rodriguez Estrada AE, Lee K, Breakspear A, May G and Kistlera HC (2012): Metabolome and trancriptome of the interaction between *Ustilago maydis* and *Fusarium verticillioides in virto.* Applied and Environmental Microbiology 78(10), 3656-3667.

Jordan G and Goldman N (2012): The effects of alignment error and alignment filtering on the sitewise detection of positive selection. Molecular Biology and Evolution 29(4), 1125-1139.

Kahmann R and Schirawski J (2007): Mating in the smut fungi: from *a* to *b* to the downstream cascades. In: Heitman J, Kronstad JW, Taylor JW and Casselton LA (Eds.): Sex in fungi: molecular determination and evolutionary implications. ASM Press, Washington DC.

Kämper J (2004): A PCR-based system for highly efficient generation of gene replacement mutants in *Ustilago maydis.* Molecular Genetics and Genomics 271, 103-110.

Kämper J, Kahmann R, Bölker M, Ma L-J, Brefort T, Saville BJ, Banuett F, Kronstad JW, Gold SE, Müller O, Perlin MH, Wösten HAB, de Vries R, Ruiz-Herrera J, Reynaga-Peña CG, Snetselaar K, McCann M, Pérez-Martin J, Feldbrügge M, Basse CW, Steinberg G, Ibeas JI, Holloman W, Guzman P, Farman M, Stajich JE, Sentandreu R, González-Prieto JM, Kennell JC, Molina L, Schirawski J, Mendoza-Mendoza A, Greilinger D, Münch K, Rössel N, Scherer M, Vranes M, Ladendorf O, Vincon V, Fuchs U, Sandrock B, Meng S,Ho ECH, Cahill MJ, Boyce KJ, Klose J, Klosterman SJ, Deelstra HJ, Ortiz-Castellanos L, Li W, Sanchez-Alonso P, Schreier PH, Häuser-Hahn I, Vaupel M, Koopmann E, Friedrich G, Voss H, Schlüter, Margolis J, Platt D, Swimmer C, Gnirke A, Chen F, Vysotskaia V, Mannhaupt G, Güldener U, Münsterkötter

M, Haase D, Oesterheld M, Mewes H-W, Mauceli EW, DeCaprio D, Wade CM, Butler J, Young S, Jaffe DB, Calvo S, Nusbaum C, Galagan J and Birren BW (2006): Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. NATURE 444, 97-101.

Kemen E, Gardinger A, Schultz-Larsen T, Kemen AC, Balmuth AL, Robert-Seilaniantz A, Bailey K, Holub E, Studholme DJ, MacLean D and Jones JDG (2011): Gene gain and loss during evolution of obligate parasitism in the white rust pathogen of *Arabidopsis thaliana*. PLOS BIOLOGY 9(7), e1001094.

Khaltruin K, Hemmrich G, Fraune S, Augustin R, and Bosch TCG (2009): More than just orphans: are taxonomically-restricted genes important in evolution? TRENDS IN GENETICS 25(9), 404-413.

Khrunyk Y, Münch K, Schipper K, Lupas AN and Kahmann R (2010): The use of FLP-mediated recombination for the functional analysis of an effector gene family in the biotrophic smut fungus *Ustilago maydis*. NEW PHYTOLOGIST 187, 957-968.

Knowles DG and McLysaght A (2009): Recent *de novo* origin of human proteincoding genes. GENOME RESEARCH 19, 1752-1759.

Laemmli UK (1970): Cleavage of structural proteins during the assembly of the head of bacteriophage T4. NATURE 227, 680-685.

Lanver D, Mendoza-Mendoza A, Brachmann A and Kahmann R (2010): Sho1 and Msb2-related proteins regulate appressorium development in the smut fungus *Ustilago maydis*. PLANT CELL 22, 2085-2101.

Lanver D, Berndt P, Tollot M, Naik V, Vranes M, Warmann T, Münch K, Rössel N and Kahmann R (2014): Plant surface cues prime *Ustilago maydis* for biotrophic development. PLOS PATHOGENS 10(7), e1004272.

Laurie JD, Ali S, Linning R, Mannhaupt G, Wong P, Güldener U, Münsterkötter M, Moore R, Kahmann R, Bakkeren G and Schirawski J (2012): Genome comparison of barley and maize smut fungi reveals targeted loss of RNA silencing components and species-specific presence of transposable elements. PLANT CELL 24, 1733-1745.

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. PLANT CELL 25, 1946–1959.

Li D, Dong Y, Jiang Y, Jiang H, Cai J and Wang W (2010a): A *de novo* originated gene depresses budding yeast mating pathway and is repressed by the protein encoded by its antisense strand. CELL RESEARCH 20, 408-420.

Li C-Y, Zhang Y, Wang Z, Zhang Y, Cao C, Zhang P-W, Lu S-J, Li X-M, Yu Q, Zheng X, Du Q, Uhl GR, Liu Q-R and Wei L (2010b): A human-specific *de novo* protein-coding gene associated with human brain functions. PLOS COMPUTATIONAL BIOLOGY 6(3), e100034.

Li W-H, Wu C-I and Luo C-C (1985): A new method for estimating synonymous and nonsynonymous rates of nucleotide substitutions considering the relative likelihood of nucleotide and codon changes. MOLECULAR BIOLOGY AND EVOLUTION 2, 150-174.

Lipman DJ, Souvorov A, Koonin EV, Panchenko AR and Tatusova TA (2002): The relationship of protein conservation and sequence length. BMC EVOLUTIONARY BIOLOGY 2:20.

Liu Z, Bos JIB, Armstrong M, Whisson SC, da Cunha L, Torto-Alalibo T, Win J, Avrova AO, Wright F, Birch PRJ and Kamoun S (2005): Patterns of Diversifying Selection in the phytotoxin-like *scr74* gene family of *Phytophthora infestans*. MOLECULAR BIOLOGY AND EVOLUTION 22(3), 659-672.

Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, Zuccaro A, Reissmann S and Kahmann R (2015): Fungal effectors and plant susceptibility. ANNUAL REVIEWS OF PLANT BIOLOGY 66, 513-545.

Löytynoja A and Goldman N (2008): Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. SCIENCE 320(5883), 1632-1635.

Lozana-Torres JS, Wilbers RHP, Gawronski P, Boshoven JC, Finkers-Tomczak A, Cordewener JHG, America AHP, Overmars HA, van't Klooster JW, Baranowski L, Sobczak M, Ilyas M, van der Hoorn RAL, Schots A, de Wit PJGM, Bakker J, Goverse A and Smant G (2012): Dual resistance mediated by the immune receptor Cf-2 in tomato requires a common virulence target of a fungus and a nematode. PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES USA 109(25), 10119-10124.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu S-M, Peng S, Xioqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam T-W and Wang J (2012): SOAPdenovo2: an empirically improved memory-efficient short read de novo assembler. GIGASCIENCE 1, 18.

Lynch M and Force A (2000): The probability of duplicate gene preservation by subfunctionalization. GENETICS 154(1), 459-473.

Lynch M and Katju V (2004): The altered evolutionary trajectories of gene duplicates. TRENDS IN GENETICS 20(11), 544-549.

Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, Di Pietro A, Dufresne M, Freitag M, Grabherr M, Henrissat B, Houterman PM, Kang S, Shim W-B, Woloshuk C, Xie X, Xu J-R, Antoniw J, Baker SE, Bluhm BH, Breakspear A, Brown DW, Butchko RAE, Chapman S, Coulson R, Coutinho PM, Danchin EGJ, Diener A, Gale LR, Gardiner DM, Goff S, Hammond-Kosack KE, Hilburn K, Hua-Van A, Jonkers W, Kazan K, Kodira CD, Koehrsen M, Kumar L, Lee Y-H, Li L, Manners JM, Miranda-Saavedra D, Mukherjee M, Park G, Park J, Park S-J, Proctor RH, Regev A, Ruiz-Roldan MC, Sain D, Sakthikumar S, Sykes S, Schwartz DC, Turgeon BG, Wapinski I, Yoder O, Young S, Zeng Q, Zhou S, Galagan J, Cuomo CA, Kistler HC and Rep M (2010): Comparative genomics reveals mobile pathogenicity chromosomes in

Fusarium. NATURE 464, 367-373.

Martínez-Espionza AD, García-Pedrajas MD and Gold SE (2002): The *Ustilaginales* as pests and model systems. FUNGAL GENETICS AND BIOLOGY 35(1), 1-20.

Mendoza-Mendoza A, Berndt P, Djamei A, Weise C, Linne U, Marahiel M, Vraneš M, Kämper J and Kahmann R (2009): Physical-chemical plant derived signals induce differentiation in *Ustilago maydis* 71(4), 895-911.

Miele V, Penel S and Duret L (2011): Ultra-fast sequencing clustering from similarity networks with SiLiX. BMC BIOINFORMATICS 12, 116.

Miya A, Albert P, Shinya T, Desaki Y, Ichimura K, Shirasu K, Narusaka Y, Kawakami N, Kaku H and Shibuya N (2007): CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in *Arabidopsis.* PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES USA 104(49), 19613-19618.

Müller AN, Ziemann S, Treitschke S, Aßmann D and Dohelemann G (2013): Compatibility in the *Ustilago maydis*-maize interaction requires inhibition of host cysteine proteases by the fungal effector Pit2. PLOS PATHOGENS 9(2), e1003177.

Munkacsi AB, Stoxen S and May G (2007): Domestication of maize, sorghum, and sugar cane did not drive the divergence of their smut pathogens. EVOLUTION 61(2), 388-403.

Näsvall J, Sun L, Roth JR and Andersson DI (2012): Real-time evolution of new genes by innovation, amplification, and divergence. SCIENCE 338(6105), 384-387.

Nei M and Gojobori T (1986): Simple methods for estimating the number of synonymous and non-synonymous nucleotide substitutions. MOLECULAR BIOLOGY AND EVOLUTION 3, 418-426.

Nekrutenko A and Li WH (2001): Transposable elements are found in a large number of human protein-coding genes. TRENDS IN GENETICS 17, 619-621.

Ngugi HK and Scherm H (2006): Biology of flower-infecting fungi. ANNUAL REVIEWS OF PHYTOPATHOLOGY 44, 261-282.

Nielsen R (2005): Molecular signatures of natural selection. ANNUAL REVIEW OF GENETICS 39, 197-218.

Nielsen R and Yang ZH (1998): Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. GENETICS 168, 2373-2782.

Nurminsky DI, Nurminskaya MV, De Aguiar D and Hartl DL (1998): Selective sweep of a newly evolved sperm-specific gene in *Drosophila.* NATURE 396, 572-575.

O'Connell RJ, Thon MR, Hacquard S, Amyotte SG, KleemannJ, Torres MF, Damm U, Buiate EA, Epstein L, Alkan N, Altmüller J, Alvarado-Balderrama L, Bauser CA,

Becker C, Birren BW, Chen Z, Choi J, Crouch JA, Duvick JP, Farman MA, Gan P, Heiman D, Henrissat B, Howard RJ, Kabbage M, Koch C, Kracher B, Kubo Y, Law AD, Lebrun M-H, Lee Y-H, Miyara I, Moore N, Neumann U, Nordström K, Panaccione DG, Panstruga R, Place M, Proctor RH, Prusky D, Rech G, Reinhardt R, Rollins JA, Rounsley S, Schardl CL, Schwartz DC, Shenoy N, Shirasu K, Sikhakolli UR, Stüber K, Sukno SA, Sweigard JA, Takano Y, Takahara H, Trail F, van der Does HC, Voll LM, Will I, Young S, Zeng Q, Zhang J, Zhou S, Dickman MB, Schulze-Lefert P, van Themaat EVL, Ma L-J and Vaillancourt LJ (2012): Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. NATURE GENETICS 44, 1060-1065.

Ohno S (1970): Evolution by gene duplication. SPRINGER, New York.

Olson Å, Aerts A, Asiegbu F, Belbahri L, Bouzid O, Broberg A, Canbäck B, Coutinho PM, Cullen D, Dalman K, Deflorio G, van Diepen LTA, Dunand C, Duplessis S, Durling M, Gonthier P, Grimwood J, Fossdal CG, Hansson D, Henrissat B, Hietala A, Himmelstrand K, Hoffmeister D, Högberg N, James TY, Karlsson M, Kohler A, Kües U, Lee Y-H, Lin Y-C, Lind M, Lindquist E, Lombard V, Lucas S, Lundén K, Morin E, Murat C, Park J, Raffaello T, Rouzé P, Salamov A, Schmutz J, Solheim H, Ståhlberg J, Vêlëz H, de Vries RP, Wiebenga A, Woodward S, Yakovlev I, Garbelotto M, Martin F, Grigoriev IV and Stenlid J (2012): Insight into trade-off between wood decay and parasitism from the genome of a fungal forest pathogen. NEW PHYTOLOGIST 194, 1001-1013.

Orbach MJ, Farrall L, Sweigard JA, Chumley FG and Valent B (2000): A telomeric avirulence gene determines efficacy for the rice blast resistance gene *Pi-ta*. PLANT CELL 12, 2019-2032.

Pan JJ, Baumgarten AM and May G (2008): Effects of host plant environment and *Ustilago maydis* infection on the fungal endophyte community of maize (*Zea mays*).**New Phytologist** 178, 147-156.

Paradis E, Claude J and Strimmer K (2004): APE: analysis of phylogenetics and evolution in R language. BIOINFORMATICS 20, 289-290.

Parniske M (2008): Arbuscular mycorrhiza: the mother of plant root endosymbioses. NATURE REVIEWS MICROBIOLOGY 6, 763-775.

Petersen TN, Brunak S, von Heijne G and Nielsen H (2011): SignalP 4.0: Discriminating signal peptides from transmembrane regions. NATURE METHODS 8, 785-786.

Poppe S, Dorsheimer L, Happel P and Stukenbrock EH: Rapidly evolving genes are key players in host specialization and virulence of the fungal wheat pathogen *Zymoseptoria tritici* (synonym *Mycosphaerella graminicola*). PLOS PATHOGENS 11(7), e1005055.

Que Y, Liping Xu L, Wu Q, Liu Y, Ling H, Liu Y, Zhang Y, Guo J, Su Y, Chen J, Wang S and Zhang C (2014): Genome sequencing of *Sporisorium scitamineum* provides insights into the pathogenic mechanisms of sugarcane smut. BMC GENOMICS 15:996.

Raffaele S and Kamoun S (2012): Genome evolution in filamentous plant pathogens: why bigger can be better. NATURE REVIEWS MICROBIOLOGY 10, 417-430.

Ranwez V, Harispe S, Delsuc F and Douzery EJP (2011): MACSE: Multiple alignment of Coding SEqeunces accounting for frame shifts and stop codons. PLOS ONE 6(9), e22594.

Rech GE, Sanz-Martin JM, Aisimova M, Sukno SA and Thon MR (2014): Natural selection on coding and noncoding DNA sequences is associated with virulence genes in a plant pathogenic fungus. GENOME BIOLOGY AND EVOLUTION 6(9), 2368-2379.

Redkar A, Hoser R, Schilling L, Zechmann B, Krzymowska M, Walbot Vand Doehlemann G (2015): A secreted effector protein of *Ustilago maydis* guides maize leaf cells to form tumors. PLANT CELL 27, 1332-1351.

Remy W, Taylor TN, Hass H and Kerp H (1994): Four hundred-million-year-old vesicular arbuscular mycorrhizae. PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES USA 91(25), 11841-11843.

Rodriguez Estrada AE, Jonkers W Kistler HC and May G (2012): Interactions between *Fusarium verticillioides*, *Ustilago maydis* and *Zea mays*: An endophyte, a pathogen, and their shared plant host. **Fungal Genetics and Biology** 49, 578-587.

Rohmer L, Guttman DS and Dangl JL (2004): Diverse evolutionary mechanisms shape the type III effector virulence facotr repertoire in the plant pathogen *Pseudomonas syringae*. GENETICS 167, 1341-1360.

Romiguier J, Figuet E, Galtier N, Douzery EJP, Boussau B, Dutheil JY and Ranwez V (2012): Fast and robust characterization of time-heterogeneous sequence evolutionary processes using substitution mapping. PLOS ONE 7(3), e33852.

Rooney HCE, van't Klooster JW, van der Horn RAL, Joosten MHAJ, Jones JDG and de Wit PJGM (2005): *Cladosporium* Avr2 inhibits tomato Rcr3 protease required for Cf-2-dependent disease resistance. SCIENCE 308(5729), 1783-1786.

Rouxel T, Grandaubert J, Hane JK, Hoede C, van de Wouw AP, Couloux A, Dominguez V, Anthouard V, Bally P, Bourras S, Cozijnsen AJ, Ciuffetti LM, Degrave A, Dilmaghani A, Duret L, Fudal I, Goodwin SB, Gout L, Glaser N, Linglin J, Kema GHJ, Lapalu N, Lawrence CB, May K, Meyer M, Ollivier B, Poulain J, Schoch CL, Simon A, Spatafora JW, Stachowiak A, Turgeon BG, Tyler BM, Vincent D, Weissenbach J, Amselem J, Quesneville H, Oliver RP, Wincker P, Balesdent M-H and Howlett BJ (2011): Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by repeat-induced point mutations. NATURE COMMUNICATIONS 2:202.

Rovenich H, Boshoven JC and Thomma BPHJ (2014): Filamentous pathogen effector functions: of pathogens, hosts and microbiomes. CURRENT OPINION IN PLANT BIOLOGY 20, 96-103.

Schilling L, Matei A, Redkar A, Walbot V and Doehlemann G (2014): Virulence of the maize smut fungus *Ustilago maydis* is shaped by organ-specific effectors. MOLECULAR PLANT

Pathology 15(8), 780-789.

Schirawski J, Heinze B, Wagenknecht M and Kahmann R (2005): Mating type loci of *Sporisorium reilianum*: Novel pattern with three *a* and multiple *b* specificities. Eukaryotic Cell 4, 1317-1327.

Schirawski J, Mannhaupt G, Münch K, Brefort T, Schipper K, Doehlemann G, Di Stasio M, Rössel N, Mendoza-Mendoza A, Pester D, Müller O, Winterberg B, Meyer E, Ghareeb H, Wollenberg T, Münsterkötter M, Wong P, Walter M, Stukenbrock E, Güldener U and Kahmann R (2010): Pathogenicity determinants in smut fungi revealed by genome comparison. Science 330(6010), 1546-1548.

Schneider A, Souvorov A, Sabath N, Landan G, Gonnet GH and Graur D (2009): Estimates of positive Darwinian selection are inflated by errors in sequencing, annotation and alignment. Genome Biology and Evolution 1, 114-118.

Schulz B, Banuett F, Dahl M, Schlesinger R, Schäfer W, Martin T, Herskowitz I and 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.

Schuster M, Schweizer G, Reissmann S and Kahmann R: Genome editing in *Ustilago maydis* using the CRISPR-Cas system. Fungal Genetics and Biology [in press].

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, 2034–2049.

Sharma R, Xia X, Riess K, Bauer R and Thines M (2015): Comparative genomics including the early-diverging smut fungus *Ceraceosorus bombacis* reveals signatures of parallel evolution within plant and animal pathogens of fungi and oomycetes. Genome Biology and Evolution [in press].

Shimizu T, Nakano T, Takamizawa D, Desaki Y, Ishii-Minami N, Nishizawa Y, Minami E, Okada K, Yamane H, Kaku H and Shibuya N (2010): Two LysM receptor molecules, CEBiP and OsCERK1, cooperatively regulate chitin elicitor signaling in rice. Plant Journal 64(2), 204-214.

Skibbe DS, Doehlemann G, Fernandes J and Walbot V (2010): Maize tumors caused by *Ustilago maydis* require organ-specific genes in host and pathogen. Science 328(5974), 89-92.

Staats M, van Baarlen P, Schouten A, van Kan JAL and Bakker FT (2007): Positive selection in phytotoxic protein-encoding genes of *Botrytis* species. Fungal Genetics and Biology 44, 52-63.

Stahl EA, Dwyer G, Mauricio R, Kreitman M and Bergelson J (1infinite): Dynamics of disease resistance polymorphism at the Rpm1 locus of *Arabidopsis*. Nature 400, 667-671.

Sothern EM (1975): Detection of specific sequences among DNA fragments separated by gel electrophoresis. Journal of Molecular Biology 98, 503-517.

Steinberg G and Perez-Martin J (2008): *Ustilago maydis*, a new fungal model system for cell biology. Trends in Cell Biology 18(2), 61-67.

Stergiopoulos I, Kourmpetis YAI, Slot JC, Bakker FT, De Wit PJGM and Rokas A (2012): *In silico* characterization and molecular evolutionary analysis of a novel superfamily of fungal effector proteins. Molecular Biology and Evolution 29(11), 3371-3384.

Stukenbrock EH and McDonald BA (2007): Geographical variation and positive diversifying selection in the host-specific toxin SnToxA. Molecular Plant Pathology 8, 321-332.

Stukenbrock EH, Jørgensen FG, Zala M, Hansen TT, McDonald BA and Schierup MH (2010): Whole-genome and chromosome evolution associated with host adaptation and speciation of the wheat pathogen *Mycosphaerella graminicola.* PLoS Genetics 6(12), e1001189.

Stukenbrock EH, Bataillon T, Dutheil JY, Hansen TT, Li R, Zala M, McDonal BA, Wang J and Schierup MH (2011): The making of a new pathogen: Insights from comparative population genomics of the domesticated wheat pathogen *Mycosphaerella graminicola* and its wild sister species. Genome Research 21, 2157-2166.

Tanaka S, Brefort T, Neidig N, Djamei A, Kahnt J, Vermerris W, Koenig S, Feussner K, Feussner I and Kahmann R (2014): A secreted *Ustilago maydis* effector promotes virulence by targeting anthocyanin biosynthesis in maize. eLife 3, e013755.

Taniguti LM, Schaker, PDC, Benevenuto J, Peters LP, Carvalho G, Palhares A, Quecine MC, Nunes FRS, Kmit MCP, Wai A, Hausner G, Aitken KS, Berkman PJ, Fraser JA, Moolhuijzen, PM, Coutinho LL, Creste S, Vieira MLC, Kitajima JP, Monteiro-Vitorello CB (2015): Complete genome sequence of *Sporisorium scitamineum* and biotrophic interaction transcriptome with sugarcane. PLOS ONE 10(6), e0129318.

Thrall PH, Laine A-L, Ravensale M, Nemri A, Dodds PN, Barrett LG and Burdon JJ (2012): Rapid genetic change underpins antagonistic coevolution in a natural host-pathogen metapopulation. Ecology Letters 15, 425-435.

Tautz D and Domazet-Lošo T (2011): The evolutionary origin of oprhan genes. Nature Reviews Genetics 12, 692-702.

Tisserant E, Malbreil M, Kuo A, Kohler A, Symeonidi A, Balestrini R, Charron P, Duensing N, dit Frey NF, Gianinazzi-Pearson V, Gilbert LB, Handa Y, Herr JR, Hijri M, Koul R, Kawaguchi M, Krajinski F, Lammers PJ, Masclaux FG, Murat C, Morina E, Ndikumana S, Pagni M, Petitpierre D, Requena N, Rosikiewicz P, Riley R, Saito K, Clemente HS, Shapiro H, van Tuinen D, Bécard G, Bonfante P, Paszkowski U, Shachar-Hill YY, Tuskan GA, Young JPW, Sanders IR, Henrissat B, Rensing SA, Grigoriev IV, Corradi N, Roux C and Martin F (2013): Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. Proceedings of the National Academy of Sciences USA 110(50), 20117-20122.

Tsukamot H, Tsuda M and Fujimori T (1999): Survey and evaluation of *Ustilago trichophora* as bioherbicidal agent for *Echinochloa* species. Annals of the Phytopathological Society of Japan 65, 537-542.

Tsukuda T, Carleton S, Fotheringham S, and Holloman WK (1988): Isolation and characterization of an autonomously replicating sequence from *Ustilago maydis.* Molecular and Cellular Biology 8, 3703-3709.

van der Linde K, Hemetsberger C, Kastner C, Kaschani F, van der Hoorn RAL, Kumlehn J and Doehlemann G (2012): A maize cystatin suppresses host immunity by inhibiting apoplasitc cysteine proteases. Plant Cell 24, 1285-1300.S

Vânky K (2012): Smut fungi of the world. APS Press.

van Valen L (1973): A new evolutionary law. Evolutionary Theory 1, 1-30.

Vorholt JA (2012): Microbial life in the phyllosphere. Nature Reviews Microbiology 10, 828-840.

Walbot V and Skibbe DS (2010): Maize host requirements for *Ustilago maydis* tumor induction. Sexual Plant Reproduction 23, 1-13.

Wicker T, Oberhaensli S, Parlange F, Buchmann JP, Shatalina M, Roffler S, Roi B-D, Doležel J, Šimková H, Schulze-Lefert P, Spanu PD, Bruggmann R, Amselem J, Quesneville H, van Themaat EVL, Paape T, Shimizu KK and Keller B (2013): The wheat powdery mildew genome shows the unique evolution of an obligate biotroph. Nature Genetics 45, 1092-1096.

Win J, Morgan W, Bos J, Krasileva KV, Cano LM, Chaparro-Garcia A, Ammar R, Staskawicz BJ and Kamoun S (2007): Adaptive evolution has targeted the C-terminal domain of the RxLR effectors of plant pathogenic oomycetes. Plant Cell 19, 2349-2369.

Xu J, Saunders CW, Hu P, Grant RA, Boekhout T, Kuramae EE, Kronstad JW, DeAngelis YM, Reeder NL, Johnstone KR, Leland M, Fieno AM, Begley WM, Sun Y, Lacey MP, Chaudhary T, Keough T, Chu L, Sears R, Yuan B and Dawson Jr, TL (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 USA 104(47), 18730–18735.

Yang Z (1998): Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Molecular Biology and Evolution 15(5), 568-573.

Yang Z (2007): PALM4: a program package for phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24, 1586-1591.

Yang Z and Nielsen R (1998): Synonymous and nonsynonymous rate variation in nuclear genes of mammals. Journal of Molecular Evolution 46, 409-418.

Yang Z, Nielsen R, Goldman N and Krabbe Pedersen A-M (2000): Codon-substitution models for heterogeneous selection pressure at amino acid sites. GENETICS 155, 431-449.

Yang Z, Wong WSW and Nielsen R (2005): Bayes empirical Bayes inference of amino acid sites under positive selection. MOLECULAR BIOLOGY AND EVOLUTION 22(4), 1107-1118.

Zhang J, Nielsen R and Yang Z (2005): Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. MOLECULAR BIOLOGY AND EVOLUTION 22(12), 2472-2479.

Zuther K, Kahnt J, Utermark J, Imkampe J, Uhse S and Schirawski J (2012): Host specificity of *Sporisorium reilianum* is tightly linked to generation of the phytoalexin luteolinidin by *Sorghum bicolor*. MOLECULAR PLANT-MICROBE INTERACTIONS 25, 1230-1237.

6. SUPPLEMENTARY INFORMATION

All orphan genes in *M. globosa*, *P. flocculosa*, *M. pennsylvanicum*, *U. hordei*, *U. maydis*, *S. scitamineum* and *S. reilianum* f. sp. *zeae*, which were detected in the present study are listed in Table 6.1.

Table 6.1: List of orphan genes in eight fungal species

Species	**Gene**	**Description**	**Secretion**	**Synteny**[1]
P. flocculosa	gi|521581819|gb|EPQ25726.1|	NA	no	NA
P. flocculosa	gi|521581874|gb|EPQ25773.1|	NA	no	NA
P. flocculosa	gi|521581969|gb|EPQ25853.1|	NA	no	NA
P. flocculosa	gi|521582005|gb|EPQ25889.1|	NA	no	NA
P. flocculosa	gi|521582197|gb|EPQ26065.1|	NA	no	NA
P. flocculosa	gi|521582202|gb|EPQ26070.1|	NA	yes	NA
P. flocculosa	gi|521582272|gb|EPQ26140.1|	NA	yes	NA
P. flocculosa	gi|521582300|gb|EPQ26168.1|	NA	no	NA
P. flocculosa	gi|521582306|gb|EPQ26174.1|	NA	no	NA
P. flocculosa	gi|521582348|gb|EPQ26204.1|	NA	yes	NA
P. flocculosa	gi|521582423|gb|EPQ26279.1|	NA	no	NA
P. flocculosa	gi|521582594|gb|EPQ26438.1|	NA	yes	NA
P. flocculosa	gi|521582673|gb|EPQ26501.1|	NA	no	NA
P. flocculosa	gi|521582738|gb|EPQ26566.1|	NA	yes	NA
P. flocculosa	gi|521582805|gb|EPQ26633.1|	NA	no	NA
P. flocculosa	gi|521582806|gb|EPQ26634.1|	NA	yes	NA
P. flocculosa	gi|521582844|gb|EPQ26663.1|	NA	no	NA
P. flocculosa	gi|521582918|gb|EPQ26737.1|	NA	no	NA
P. flocculosa	gi|521582949|gb|EPQ26768.1|	NA	no	NA
P. flocculosa	gi|521583031|gb|EPQ26839.1|	NA	yes	NA
P. flocculosa	gi|521583056|gb|EPQ26864.1|	NA	no	NA
P. flocculosa	gi|521583171|gb|EPQ26979.1|	NA	no	NA
P. flocculosa	gi|521583228|gb|EPQ27028.1|	NA	no	NA
P. flocculosa	gi|521583258|gb|EPQ27058.1|	NA	yes	NA
P. flocculosa	gi|521583294|gb|EPQ27094.1|	NA	no	NA
P. flocculosa	gi|521583298|gb|EPQ27098.1|	NA	no	NA
P. flocculosa	gi|521583305|gb|EPQ27105.1|	NA	no	NA
P. flocculosa	gi|521583345|gb|EPQ27145.1|	NA	yes	NA
P. flocculosa	gi|521583352|gb|EPQ27152.1|	NA	no	NA
P. flocculosa	gi|521583360|gb|EPQ27160.1|	NA	no	NA
P. flocculosa	gi|521583431|gb|EPQ27216.1|	NA	no	NA
P. flocculosa	gi|521583522|gb|EPQ27307.1|	NA	no	NA
P. flocculosa	gi|521583543|gb|EPQ27328.1|	NA	no	NA
P. flocculosa	gi|521583610|gb|EPQ27389.1|	NA	no	NA

P. flocculosa	gi|521583654|gb|EPQ27433.1|	NA	no	NA
P. flocculosa	gi|521583696|gb|EPQ27475.1|	NA	no	NA
P. flocculosa	gi|521583714|gb|EPQ27493.1|	NA	no	NA
P. flocculosa	gi|521583813|gb|EPQ27576.1|	NA	no	NA
P. flocculosa	gi|521583887|gb|EPQ27650.1|	NA	yes	NA
P. flocculosa	gi|521583986|gb|EPQ27749.1|	NA	no	NA
P. flocculosa	gi|521584029|gb|EPQ27780.1|	NA	no	NA
P. flocculosa	gi|521584030|gb|EPQ27781.1|	NA	no	NA
P. flocculosa	gi|521584035|gb|EPQ27786.1|	NA	no	NA
P. flocculosa	gi|521584067|gb|EPQ27818.1|	NA	no	NA
P. flocculosa	gi|521584165|gb|EPQ27916.1|	NA	no	NA
P. flocculosa	gi|521584189|gb|EPQ27940.1|	NA	yes	NA
P. flocculosa	gi|521584190|gb|EPQ27941.1|	NA	no	NA
P. flocculosa	gi|521584193|gb|EPQ27944.1|	NA	yes	NA
P. flocculosa	gi|521584251|gb|EPQ27988.1|	NA	no	NA
P. flocculosa	gi|521584258|gb|EPQ27995.1|	NA	yes	NA
P. flocculosa	gi|521584267|gb|EPQ28004.1|	NA	no	NA
P. flocculosa	gi|521584352|gb|EPQ28089.1|	NA	no	NA
P. flocculosa	gi|521584515|gb|EPQ28249.1|	NA	no	NA
P. flocculosa	gi|521584538|gb|EPQ28272.1|	NA	no	NA
P. flocculosa	gi|521584570|gb|EPQ28304.1|	NA	no	NA
P. flocculosa	gi|521584643|gb|EPQ28377.1|	NA	no	NA
P. flocculosa	gi|521584669|gb|EPQ28403.1|	NA	yes	NA
P. flocculosa	gi|521584672|gb|EPQ28406.1|	NA	yes	NA
P. flocculosa	gi|521584804|gb|EPQ28525.1|	NA	yes	NA
P. flocculosa	gi|521584858|gb|EPQ28579.1|	NA	no	NA
P. flocculosa	gi|521584864|gb|EPQ28585.1|	NA	yes	NA
P. flocculosa	gi|521584879|gb|EPQ28600.1|	NA	no	NA
P. flocculosa	gi|521585060|gb|EPQ28762.1|	NA	no	NA
P. flocculosa	gi|521585079|gb|EPQ28781.1|	NA	yes	NA
P. flocculosa	gi|521585127|gb|EPQ28829.1|	NA	no	NA
P. flocculosa	gi|521585182|gb|EPQ28884.1|	NA	no	NA
P. flocculosa	gi|521585438|gb|EPQ29127.1|	NA	no	NA
P. flocculosa	gi|521585614|gb|EPQ29291.1|	NA	yes	NA
P. flocculosa	gi|521585615|gb|EPQ29292.1|	NA	yes	NA
P. flocculosa	gi|521585630|gb|EPQ29307.1|	NA	no	NA
P. flocculosa	gi|521585644|gb|EPQ29321.1|	NA	no	NA
P. flocculosa	gi|521585651|gb|EPQ29328.1|	NA	no	NA
P. flocculosa	gi|521585770|gb|EPQ29447.1|	NA	no	NA
P. flocculosa	gi|521585784|gb|EPQ29461.1|	NA	yes	NA
P. flocculosa	gi|521585980|gb|EPQ29640.1|	NA	no	NA
P. flocculosa	gi|521586029|gb|EPQ29689.1|	NA	no	NA
P. flocculosa	gi|521586033|gb|EPQ29693.1|	NA	no	NA
P. flocculosa	gi|521586087|gb|EPQ29747.1|	NA	yes	NA
P. flocculosa	gi|521586188|gb|EPQ29835.1|	NA	no	NA
P. flocculosa	gi|521586309|gb|EPQ29956.1|	NA	no	NA
P. flocculosa	gi|521586366|gb|EPQ30013.1|	NA	no	NA

P. flocculosa	gi|521586427|gb|EPQ30059.1|	NA	no	NA
P. flocculosa	gi|521586578|gb|EPQ30210.1|	NA	no	NA
P. flocculosa	gi|521586593|gb|EPQ30225.1|	NA	no	NA
P. flocculosa	gi|521586642|gb|EPQ30274.1|	NA	yes	NA
P. flocculosa	gi|521586726|gb|EPQ30336.1|	NA	yes	NA
P. flocculosa	gi|521586831|gb|EPQ30441.1|	NA	no	NA
P. flocculosa	gi|521586919|gb|EPQ30529.1|	NA	no	NA
P. flocculosa	gi|521587081|gb|EPQ30671.1|	NA	no	NA
P. flocculosa	gi|521587150|gb|EPQ30740.1|	NA	no	NA
P. flocculosa	gi|521587281|gb|EPQ30871.1|	NA	no	NA
P. flocculosa	gi|521587578|gb|EPQ31143.1|	NA	no	NA
P. flocculosa	gi|521587705|gb|EPQ31270.1|	NA	yes	NA
P. flocculosa	gi|521587858|gb|EPQ31418.1|	NA	no	NA
P. flocculosa	gi|521587869|gb|EPQ31429.1|	NA	no	NA
P. flocculosa	gi|521587875|gb|EPQ31435.1|	NA	yes	NA
P. flocculosa	gi|521587878|gb|EPQ31438.1|	NA	no	NA
P. flocculosa	gi|521588008|gb|EPQ31568.1|	NA	no	NA
P. flocculosa	gi|521588059|gb|EPQ31619.1|	NA	yes	NA
P. flocculosa	gi|521588434|gb|EPQ31968.1|	NA	no	NA
P. flocculosa	gi|521588533|gb|EPQ32067.1|	NA	yes	NA
P. flocculosa	gi|521588551|gb|EPQ32085.1|	NA	no	NA
P. flocculosa	gi|521588748|gb|EPQ32282.1|	NA	no	NA
P. flocculosa	gi|521588779|gb|EPQ32313.1|	NA	no	NA
P. flocculosa	gi|521588834|gb|EPQ32368.1|	NA	no	NA
P. flocculosa	gi|521588867|gb|EPQ32401.1|	NA	no	NA
P. flocculosa	gi|521588877|gb|EPQ32411.1|	NA	no	NA
P. flocculosa	gi|521588937|gb|EPQ32471.1|	NA	no	NA
M. globosa	jgi|Malgl1|1115|MGL_1114	NA	no	NA
M. globosa	jgi|Malgl1|1137|MGL_1136	NA	no	NA
M. globosa	jgi|Malgl1|1229|MGL_1228	NA	no	NA
M. globosa	jgi|Malgl1|1316|MGL_1315	NA	no	NA
M. globosa	jgi|Malgl1|1358|MGL_1357	NA	no	NA
M. globosa	jgi|Malgl1|1384|MGL_1383	NA	no	NA
M. globosa	jgi|Malgl1|1424|MGL_1423	NA	no	NA
M. globosa	jgi|Malgl1|1472|MGL_1471	NA	no	NA
M. globosa	jgi|Malgl1|1473|MGL_1472	NA	no	NA
M. globosa	jgi|Malgl1|1493|MGL_1492	NA	no	NA
M. globosa	jgi|Malgl1|176|MGL_0175	NA	no	NA
M. globosa	jgi|Malgl1|18|MGL_0017	NA	no	NA
M. globosa	jgi|Malgl1|2077|MGL_2076	NA	no	NA
M. globosa	jgi|Malgl1|2165|MGL_2164	NA	no	NA
M. globosa	jgi|Malgl1|2332|MGL_2331	NA	no	NA
M. globosa	jgi|Malgl1|2384|MGL_2383	NA	no	NA
M. globosa	jgi|Malgl1|2480|MGL_2479	NA	no	NA
M. globosa	jgi|Malgl1|2604|MGL_2603	NA	no	NA
M. globosa	jgi|Malgl1|2674|MGL_2673	NA	yes	NA
M. globosa	jgi|Malgl1|2689|MGL_2688	NA	no	NA

M. globosa	jgi|Malgl1|2799|MGL_2798	NA	no	NA
M. globosa	jgi|Malgl1|2821|MGL_2820	NA	no	NA
M. globosa	jgi|Malgl1|2860|MGL_2859	NA	no	NA
M. globosa	jgi|Malgl1|2983|MGL_2982	NA	no	NA
M. globosa	jgi|Malgl1|2999|MGL_2998	NA	yes	NA
M. globosa	jgi|Malgl1|3098|MGL_3097	NA	no	NA
M. globosa	jgi|Malgl1|3099|MGL_3098	NA	no	NA
M. globosa	jgi|Malgl1|3333|MGL_3332	NA	no	NA
M. globosa	jgi|Malgl1|3500|MGL_3499	NA	no	NA
M. globosa	jgi|Malgl1|3509|MGL_3508	NA	no	NA
M. globosa	jgi|Malgl1|3697|MGL_3696	NA	no	NA
M. globosa	jgi|Malgl1|3726|MGL_3725	NA	no	NA
M. globosa	jgi|Malgl1|3855|MGL_3854	NA	no	NA
M. globosa	jgi|Malgl1|3863|MGL_3862	NA	no	NA
M. globosa	jgi|Malgl1|3884|MGL_3883	NA	no	NA
M. globosa	jgi|Malgl1|3902|MGL_3901	NA	no	NA
M. globosa	jgi|Malgl1|3984|MGL_3983	NA	no	NA
M. globosa	jgi|Malgl1|4190|MGL_4189	NA	yes	NA
M. globosa	jgi|Malgl1|4248|MGL_4247	NA	no	NA
M. globosa	jgi|Malgl1|4280|MGL_4279	NA	no	NA
M. globosa	jgi|Malgl1|4281|MGL_4280	NA	no	NA
M. globosa	jgi|Malgl1|4284|MGL_4283	NA	no	NA
M. globosa	jgi|Malgl1|4285|MGL_4284	NA	no	NA
M. globosa	jgi|Malgl1|515|MGL_0514	NA	no	NA
M. globosa	jgi|Malgl1|560|MGL_0559	NA	no	NA
M. globosa	jgi|Malgl1|635|MGL_0634	NA	no	NA
M. globosa	jgi|Malgl1|663|MGL_0662	NA	no	NA
M. globosa	jgi|Malgl1|887|MGL_0886	NA	no	NA
M. globosa	jgi|Malgl1|891|MGL_0890	NA	no	NA
M. globosa	jgi|Malgl1|945|MGL_0944	NA	no	NA
M. globosa	jgi|Malgl1|948|MGL_0947	NA	no	NA
M. globosa	jgi|Malgl1|969|MGL_0968	NA	yes	NA
M. pennsylvanicum	mp00071	uncharacterized protein	no	NA
M. pennsylvanicum	mp00099	uncharacterized protein	no	NA
M. pennsylvanicum	mp00371	uncharacterized protein	no	NA
M. pennsylvanicum	mp00381	uncharacterized protein	no	NA
M. pennsylvanicum	mp00818	uncharacterized protein	yes	NA
M. pennsylvanicum	mp00858	uncharacterized protein	no	NA
M. pennsylvanicum	mp00895	uncharacterized protein	no	NA
M. pennsylvanicum	mp01750	uncharacterized protein	no	NA
M. pennsylvanicum	mp02293	uncharacterized protein	no	NA
M. pennsylvanicum	mp02658	uncharacterized protein	no	NA
M. pennsylvanicum	mp02770	uncharacterized protein	no	NA
M. pennsylvanicum	mp03081	uncharacterized protein	no	NA
M. pennsylvanicum	mp03265	uncharacterized protein	no	NA
M. pennsylvanicum	mp03288	uncharacterized protein	no	NA
M. pennsylvanicum	mp03427	uncharacterized protein	yes	NA

M. pennsylvanicum	mp03669	uncharacterized protein	no	NA
M. pennsylvanicum	mp04410	uncharacterized protein	yes	NA
M. pennsylvanicum	mp05484	uncharacterized protein	no	NA
M. pennsylvanicum	mp05636	uncharacterized protein	no	NA
M. pennsylvanicum	mp05672	uncharacterized protein	no	NA
M. pennsylvanicum	mp06007	uncharacterized protein	no	NA
M. pennsylvanicum	mp06011	uncharacterized protein	no	NA
M. pennsylvanicum	mp06013	uncharacterized protein	no	NA
M. pennsylvanicum	mp06014	uncharacterized protein	no	NA
M. pennsylvanicum	mp06015	uncharacterized protein	no	NA
M. pennsylvanicum	mp06019	uncharacterized protein	no	NA
M. pennsylvanicum	mp06024	uncharacterized protein	no	NA
M. pennsylvanicum	mp06028	uncharacterized protein	no	NA
M. pennsylvanicum	mp06029	uncharacterized protein	no	NA
M. pennsylvanicum	mp06033	uncharacterized protein	no	NA
M. pennsylvanicum	mp06035	uncharacterized protein	no	NA
M. pennsylvanicum	mp06043	uncharacterized protein	no	NA
M. pennsylvanicum	mp06044	uncharacterized protein	no	NA
M. pennsylvanicum	mp06047	uncharacterized protein	no	NA
M. pennsylvanicum	mp06054	uncharacterized protein	no	NA
M. pennsylvanicum	mp06056	uncharacterized protein	no	NA
M. pennsylvanicum	mp06058	uncharacterized protein	no	NA
M. pennsylvanicum	mp06059	uncharacterized protein	no	NA
M. pennsylvanicum	mp06068	uncharacterized protein	no	NA
M. pennsylvanicum	mp06070	uncharacterized protein	no	NA
M. pennsylvanicum	mp06072	uncharacterized protein	no	NA
M. pennsylvanicum	mp06086	uncharacterized protein	no	NA
M. pennsylvanicum	mp06101	uncharacterized protein	no	NA
M. pennsylvanicum	mp06103	uncharacterized protein	no	NA
M. pennsylvanicum	mp06109	uncharacterized protein	no	NA
M. pennsylvanicum	mp06110	uncharacterized protein	no	NA
M. pennsylvanicum	mp06111	uncharacterized protein	no	NA
M. pennsylvanicum	mp06112	uncharacterized protein	no	NA
M. pennsylvanicum	mp06119	uncharacterized protein	no	NA
M. pennsylvanicum	mp06121	uncharacterized protein	no	NA
M. pennsylvanicum	mp06124	uncharacterized protein	no	NA
M. pennsylvanicum	mp06128	uncharacterized protein	no	NA
M. pennsylvanicum	mp06135	uncharacterized protein	no	NA
M. pennsylvanicum	mp06139	uncharacterized protein	no	NA
M. pennsylvanicum	mp06146	uncharacterized protein	no	NA
M. pennsylvanicum	mp06158	uncharacterized protein	no	NA
M. pennsylvanicum	mp06166	uncharacterized protein	no	NA
M. pennsylvanicum	mp06171	uncharacterized protein	no	NA
M. pennsylvanicum	mp06172	uncharacterized protein	no	NA
M. pennsylvanicum	mp06184	uncharacterized protein	no	NA
M. pennsylvanicum	mp06185	uncharacterized protein	no	NA
M. pennsylvanicum	mp06191	uncharacterized protein	no	NA

M. pennsylvanicum	mp06198	uncharacterized protein	no	NA
M. pennsylvanicum	mp06200	uncharacterized protein	no	NA
M. pennsylvanicum	mp06201	uncharacterized protein	no	NA
M. pennsylvanicum	mp06202	uncharacterized protein	no	NA
M. pennsylvanicum	mp06204	uncharacterized protein	no	NA
M. pennsylvanicum	mp06217	uncharacterized protein	no	NA
M. pennsylvanicum	mp06221	uncharacterized protein	no	NA
M. pennsylvanicum	mp06226	uncharacterized protein	no	NA
M. pennsylvanicum	mp06231	uncharacterized protein	no	NA
M. pennsylvanicum	mp06240	uncharacterized protein	no	NA
M. pennsylvanicum	mp06243	uncharacterized protein	no	NA
M. pennsylvanicum	mp06247	uncharacterized protein	no	NA
M. pennsylvanicum	mp06265	uncharacterized protein	no	NA
M. pennsylvanicum	mp06269	uncharacterized protein	no	NA
M. pennsylvanicum	mp06276	uncharacterized protein	no	NA
M. pennsylvanicum	mp06280	uncharacterized protein	no	NA
M. pennsylvanicum	mp06282	uncharacterized protein	no	NA
M. pennsylvanicum	mp06296	uncharacterized protein	no	NA
M. pennsylvanicum	mp06301	uncharacterized protein	no	NA
M. pennsylvanicum	mp06305	uncharacterized protein	no	NA
M. pennsylvanicum	mp06317	uncharacterized protein	no	NA
M. pennsylvanicum	mp06326	uncharacterized protein	no	NA
M. pennsylvanicum	mp06328	uncharacterized protein	no	NA
M. pennsylvanicum	mp06333	uncharacterized protein	no	NA
M. pennsylvanicum	mp06336	uncharacterized protein	no	NA
M. pennsylvanicum	mp06349	uncharacterized protein	yes	NA
M. pennsylvanicum	mp06353	uncharacterized protein	no	NA
S. reilianum f. sp. *zeae*	sr10613	hypothetical protein	no	NA
S. reilianum f. sp. *zeae*	sr10788.2	hypothetical protein	no	NA
S. reilianum f. sp. *zeae*	sr16736	hypothetical protein	no	NA
S. reilianum f. sp. *zeae*	sr17069	hypothetical protein	no	NA
S. reilianum f. sp. *zeae*	sr17257	hypothetical protein	no	NA
S. reilianum f. sp. *zeae*	sr17262	hypothetical protein	no	NA
S. reilianum f. sp. *zeae*	sr17427	hypothetical protein	yes	NA
S. reilianum f. sp. *zeae*	sr17432	hypothetical protein	no	NA
S. scitamineum	SSCI_00006	hypothetical protein	no	NA
S. scitamineum	SSCI_00007	hypothetical protein	no	NA
S. scitamineum	SSCI_00654	hypothetical protein	no	NA
S. scitamineum	SSCI_01142	hypothetical protein	no	NA

S. scitamineum	SSCI_01522	hypothetical protein	yes	NA
S. scitamineum	SSCI_01820	hypothetical protein	no	NA
S. scitamineum	SSCI_02260	hypothetical protein	no	NA
S. scitamineum	SSCI_02665	hypothetical protein	no	NA
S. scitamineum	SSCI_03331	hypothetical protein	no	NA
S. scitamineum	SSCI_03442	conserved hypothetical protein	no	NA
S. scitamineum	SSCI_03446	hypothetical protein	no	NA
S. scitamineum	SSCI_03514	hypothetical protein	no	NA
S. scitamineum	SSCI_03817	hypothetical protein	no	NA
S. scitamineum	SSCI_03843	hypothetical protein	no	NA
S. scitamineum	SSCI_03844	hypothetical protein	no	NA
S. scitamineum	SSCI_03850	hypothetical protein	no	NA
S. scitamineum	SSCI_03851	hypothetical protein	no	NA
S. scitamineum	SSCI_03924	conserved hypothetical protein	no	NA
S. scitamineum	SSCI_04055	hypothetical protein	no	NA
S. scitamineum	SSCI_05565	hypothetical protein	no	NA
S. scitamineum	SSCI_05582	hypothetical protein	no	NA
U. hordei	UH_00606	hypothetical protein	no	NA
U. hordei	UH_01152	hypothetical protein	no	NA
U. hordei	UH_01204	hypothetical protein	yes	NA
U. hordei	UH_01288	hypothetical protein	no	NA
U. hordei	UH_01318	hypothetical protein	no	NA
U. hordei	UH_01444	hypothetical protein	no	NA
U. hordei	UH_01578	hypothetical protein	no	NA
U. hordei	UH_01728	hypothetical protein	no	NA
U. hordei	UH_01931	hypothetical protein	no	NA
U. hordei	UH_02069	hypothetical protein	no	NA
U. hordei	UH_02299	hypothetical protein	no	NA
U. hordei	UH_02474	hypothetical protein	no	NA
U. hordei	UH_02679	hypothetical protein	no	NA
U. hordei	UH_02860	hypothetical protein	no	NA
U. hordei	UH_02908	hypothetical protein	no	NA
U. hordei	UH_02914	hypothetical protein	no	NA
U. hordei	UH_03063	hypothetical protein	no	NA
U. hordei	UH_03141	hypothetical protein	no	NA
U. hordei	UH_03159	hypothetical protein	no	NA
U. hordei	UH_03178	hypothetical protein	no	NA
U. hordei	UH_03267	hypothetical protein	no	NA
U. hordei	UH_03368	hypothetical protein	no	NA
U. hordei	UH_03501	hypothetical protein	no	NA
U. hordei	UH_04023	hypothetical protein	no	NA
U. hordei	UH_04081	hypothetical protein	no	NA
U. hordei	UH_04117	hypothetical protein	no	NA
U. hordei	UH_04314	hypothetical protein	yes	NA
U. hordei	UH_04574	hypothetical protein	no	NA

U. hordei	UH_04772	hypothetical protein	no	NA
U. hordei	UH_05269	hypothetical protein	no	NA
U. hordei	UH_05480	hypothetical protein	no	NA
U. hordei	UH_05524	hypothetical protein	no	NA
U. hordei	UH_05543	hypothetical protein	no	NA
U. hordei	UH_05554	hypothetical protein	no	NA
U. hordei	UH_05559	hypothetical protein	no	NA
U. hordei	UH_05722	hypothetical protein	no	NA
U. hordei	UH_05770	hypothetical protein	no	NA
U. hordei	UH_06049	hypothetical protein	yes	NA
U. hordei	UH_06080	hypothetical protein	no	NA
U. hordei	UH_06111	hypothetical protein	no	NA
U. hordei	UH_06222	hypothetical protein	no	NA
U. hordei	UH_06370	hypothetical protein	no	NA
U. hordei	UH_06474	hypothetical protein	no	NA
U. hordei	UH_06851	hypothetical protein	no	NA
U. hordei	UH_06927	hypothetical protein	no	NA
U. hordei	UH_07217	hypothetical protein	no	NA
U. hordei	UH_07332	hypothetical protein	yes	NA
U. hordei	UH_07359	hypothetical protein	no	NA
U. hordei	UH_07378	hypothetical protein	no	NA
U. hordei	UH_07468	hypothetical protein	no	NA
U. hordei	UH_07661	hypothetical protein	no	NA
U. hordei	UH_07678	hypothetical protein	no	NA
U. hordei	UH_07691	hypothetical protein	no	NA
U. hordei	UH_07829	hypothetical protein	no	NA
U. hordei	UH_07907	hypothetical protein	no	NA
U. hordei	UH_08083	hypothetical protein	no	NA
U. hordei	UH_08258	hypothetical protein	no	NA
U. hordei	UH_08348	hypothetical protein	no	NA
U. hordei	UH_08559	hypothetical protein	no	NA
U. hordei	UH_08652	hypothetical protein	no	NA
U. hordei	UH_08839	hypothetical protein	no	NA
U. hordei	UH_08887	hypothetical protein	no	NA
U. hordei	UH_12255	hypothetical protein	no	NA
U. hordei	UH_13141	hypothetical protein	no	NA
U. hordei	UH_13467	hypothetical protein	no	NA
U. hordei	UH_13888	hypothetical protein	no	NA
U. hordei	UH_14430	hypothetical protein	no	NA
U. hordei	UH_15370	hypothetical protein	no	NA
U. hordei	UH_15756	hypothetical protein	no	NA
U. hordei	UH_16081	hypothetical protein	no	NA
U. hordei	UH_16255	hypothetical protein	no	NA
U. hordei	UH_16407	hypothetical protein	no	NA
U. hordei	UH_16797	hypothetical protein	no	NA
U. maydis	*um00052*	putative protein	no	sr11383
U. maydis	*um00187*	hypothetical protein	yes	not syntenic

U. maydis	*um00276*	hypothetical protein	no	syntenic
U. maydis	*um00278*	hypothetical protein	no	syntenic
U. maydis	*um00731*	hypothetical protein	no	syntenic
U. maydis	*um00858*	hypothetical protein	no	sr12146
U. maydis	*um01024*	hypothetical protein	no	not syntenic
U. maydis	*um01041*	hypothetical protein	no	syntenic
U. maydis	*um01047*	hypothetical protein	no	syntenic
U. maydis	*um01053*	hypothetical protein	no	syntenic
U. maydis	*um01125*	hypothetical protein	no	syntenic
U. maydis	*um01226*	hypothetical protein	no	syntenic
U. maydis	*um01389*	hypothetical protein	no	syntenic
U. maydis	*um01455*	hypothetical protein	no	syntenic
U. maydis	*um02112*	hypothetical protein	no	syntenic
U. maydis	*um02193*	hypothetical protein	yes	sr10767
U. maydis	*um02251*	conserved hypothetical *Ustilago*-specific protein	no	not syntenic
U. maydis	*um02313*	hypothetical protein	no	syntenic
U. maydis	*um02756*	hypothetical protein	no	sr17069
U. maydis	*um03039*	hypothetical protein	no	not syntenic
U. maydis	*um03047*	hypothetical protein	no	not syntenic
U. maydis	*um03196*	hypothetical protein	no	not syntenic
U. maydis	*um03388*	hypothetical protein	no	not syntenic
U. maydis	*um03472*	hypothetical protein	no	syntenic
U. maydis	*um03535*	hypothetical protein	no	syntenic
U. maydis	*um03562*	hypothetical protein	no	not syntenic
U. maydis	*um03797*	hypothetical protein	no	syntenic
U. maydis	*um03932*	hypothetical protein	no	sr14837
U. maydis	*um04087*	hypothetical protein	no	not syntenic
U. maydis	*um04120*	hypothetical protein	no	not syntenic
U. maydis	*um04286*	hypothetical protein	no	not syntenic
U. maydis	*um04490*	hypothetical protein	no	not syntenic
U. maydis	*um04772*	hypothetical protein	no	syntenic
U. maydis	*um04929*	conserved hypothetical protein	no	syntenic
U. maydis	*um04958*	hypothetical protein	no	syntenic
U. maydis	*um04968*	hypothetical protein	no	not syntenic
U. maydis	*um05095*	hypothetical protein	no	not syntenic
U. maydis	*um05155*	hypothetical protein	no	syntenic
U. maydis	*um05342*	hypothetical protein	no	syntenic
U. maydis	*um05350*	hypothetical protein	no	syntenic
U. maydis	*um05397*	putative protein	no	syntenic
U. maydis	*um05399*	hypothetical protein	no	not syntenic
U. maydis	*um05517*	hypothetical protein	no	syntenic
U. maydis	*um05573*	hypothetical protein	no	syntenic
U. maydis	*um05772*	hypothetical protein	no	syntenic
U. maydis	*um05976*	hypothetical protein	no	not syntenic
U. maydis	*um06357.2*	putative protein	no	not syntenic

U. maydis	*um06384*	hypothetical protein	no	syntenic
U. maydis	*um10045*	putative protein	no	not syntenic
U. maydis	*um10078*	hypothetical protein	no	not syntenic
U. maydis	*um10171*	hypothetical protein	no	sr13001
U. maydis	*um10264*	hypothetical protein	no	sr10578
U. maydis	*um10328*	hypothetical protein	no	syntenic
U. maydis	*um10358*	hypothetical protein	no	syntenic
U. maydis	*um10495*	putative protein	no	syntenic
U. maydis	*um10853*	putative protein	no	syntenic
U. maydis	*um10950*	hypothetical protein	no	not syntenic
U. maydis	*um11061*	hypothetical protein	no	not syntenic
U. maydis	*um11072*	hypothetical protein	no	syntenic
U. maydis	*um11081*	putative protein (C-terminal fragment)	no	syntenic
U. maydis	*um11082*	hypothetical protein	no	sr17427
U. maydis	*um11094*	hypothetical protein	yes	syntenic
U. maydis	*um11146*	hypothetical protein	no	syntenic
U. maydis	*um11237*	putative protein	no	sr16847
U. maydis	*um11251*	putative protein	no	sr17065
U. maydis	*um11332*	putative protein	no	syntenic
U. maydis	*um11371*	putative protein	no	no information
U. maydis	*um11387*	hypothetical protein	no	syntenic
U. maydis	*um11397*	putative protein	no	syntenic
U. maydis	*um11524*	conserved hypothetical protein	no	syntenic
U. maydis	*um11639*	hypothetical protein	yes	syntenic
U. maydis	*um11775*	putative protein	no	not syntenic
U. maydis	*um11809*	putative protein	no	syntenic
U. maydis	*um11813*	putative protein	yes	not syntenic
U. maydis	*um11835*	conserved hypothetical protein	no	not syntenic
U. maydis	*um11873*	hypothetical protein	no	not syntenic
U. maydis	*um11891*	conserved hypothetical protein	no	not syntenic
U. maydis	*um11900*	putative protein	no	syntenic
U. maydis	*um11980*	putative protein	yes	not syntenic
U. maydis	*um12092*	hypothetical protein	no	syntenic
U. maydis	*um12098*	putative protein	no	not syntenic
U. maydis	*um12156*	putative protein	no	syntenic
U. maydis	*um12217*	hypothetical protein	yes	not syntenic
U. maydis	*um12235*	hypothetical protein	no	not syntenic
U. maydis	*um12275*	hypothetical protein	no	sr15526
U. maydis	*um12319*	hypothetical protein	no	not syntenic
U. maydis	*um12339*	hypothetical protein	no	not syntenic

---

[1] Synteny was only compared between regions in *U. maydis* that contain an orphan gene and the homologous region in *S. reilianum* f. sp. *zeae*

All genes with signs of positive selection in *U. hordei*, *U. maydis*, *S. scitamineumS. reilianum* f. sp. *zeae* and *S. reilianum* f. sp. *sorghi*, which were detected in the present study are listed in Table 6.2.

Table 6.2: List of genes under positive selection in five related smut fungi

Species	**Gene**	**Description**	ω_{free}	ω_{join}	**Secretion**
S. reilianum f. sp. *zeae*	*sr00093*	hypothetical protein	1.71197	1.71197	no
S. reilianum f. sp. *zeae*	*sr00846.2*	conserved hypothetical protein	3.07138	≤ 1	yes
S. reilianum f. sp. *zeae*	*sr06404*	related to ARO80 - positive transcription regulator of ARO9 and ARO10	≤ 1	1.55914	no
S. reilianum f. sp. *zeae*	*sr06452*	related to NADH-dependent flavin oxidoreductase	1.40228	1.40228	no
S. reilianum f. sp. *zeae*	*sr10057*	conserved hypothetical protein	1.17126	1.17126	yes
S. reilianum f. sp. *zeae*	*sr10059*	conserved hypothetical *Ustilagiceae*-specific protein	6.53881	6.53881	yes
S. reilianum f. sp. *zeae*	*sr10206*	conserved hypothetical protein	1.24279	infinite	no
S. reilianum f. sp. *zeae*	*sr10317*	conserved hypothetical *Ustilaginaceae*-specific protein	3.53338	3.53338	yes
S. reilianum f. sp. *zeae*	*sr10529*	conserved hypothetical protein	31.1469	31.1469	yes
S. reilianum f. sp. *zeae*	*sr10680*	probable methylglutaconyl-coa hydratase; mitochondrial precursor	1.53935	infinite	no
S. reilianum f. sp. *zeae*	*sr10916.2*	conserved hypothetical *Ustilaginaceae*-specific protein	1.45001	1.45001	yes
S. reilianum f. sp. *zeae*	*sr11005*	conserved hypothetical protein	1.02783	1.15079	no
S. reilianum f. sp. *zeae*	*sr11154*	conserved hypothetical protein	1.20855	≤ 1	no
S. reilianum f. sp. *zeae*	*sr11233*	conserved hypothetical *Ustilaginaceae*-specific protein	5.22937	5.22937	yes
S. reilianum f. sp. *zeae*	*sr11237*	conserved hypothetical protein	10.7826	10.7826	yes
S. reilianum f. sp. *zeae*	*sr11239.2*	conserved hypothetical protein	5.95563	3.93728	no
S. reilianum f. sp. *zeae*	*sr11240*	conserved hypothetical protein	5.95563	≤ 1	no
S. reilianum f. sp. *zeae*	*sr11254*	conserved hypothetical protein	2.24726	infinite	no
S. reilianum f. sp. *zeae*	*sr11259*	conserved hypothetical protein	2.30171	≤ 1	no
S. reilianum f. sp. *zeae*	*sr11400*	conserved hypothetical *Ustilaginaceae*-specific protein	1.13351	1.13351	yes
S. reilianum f. sp. *zeae*	*sr11471*	related to Serine protease	2.66863	infinite	no
S. reilianum f. sp. *zeae*	*sr11541*	conserved hypothetical protein	≤ 1	1.80241	no
S. reilianum f. sp. *zeae*	*sr11929*	conserved hypothetical protein	2.52875	2.52875	no
S. reilianum f. sp. *zeae*	*sr12013*	probable FEN2 - Pantothenate permease	1.12702	≤ 1	no
S. reilianum f. sp. *zeae*	*sr12126*	related to FRE6 - Ferric reductase	≤ 1	1.67466	no
S. reilianum f. sp. *zeae*	*sr12173*	related to YEA4 - uridine diphosphate-N-acetylglucosamine	1.39832	1.39816	no
S. reilianum f. sp. *zeae*	*sr12295*	conserved hypothetical protein	1.19001	1.19001	no
S. reilianum f. sp. *zeae*	*sr12344*	conserved hypothetical protein	1.80458	≤ 1	no

S. reilianum f. sp. *zeae*	*sr12419*	conserved hypothetical protein	1.01299	≤ 1	no
S. reilianum f. sp. *zeae*	*sr12427*	hypothetical protein	1.06862	1.06862	no
S. reilianum f. sp. *zeae*	*sr12501*	related to Enoyl-CoA hydratase	1.53935	≤ 1	no
S. reilianum f. sp. *zeae*	*sr12655*	related to NCR1 - transmembrane glycoprotein; involved in sphingolipid metabolism	1.60055	≤ 1	no
S. reilianum f. sp. *zeae*	*sr12770*	conserved hypothetical protein	1.55237	1.55237	no
S. reilianum f. sp. *zeae*	*sr12844*	conserved hypothetical protein	2.67856	2.67856	no
S. reilianum f. sp. *zeae*	*sr12972*	conserved hypothetical protein	1.76216	1.76216	no
S. reilianum f. sp. *zeae*	*sr13080*	conserved hypothetical protein	≤ 1	782.558	no
S. reilianum f. sp. *zeae*	*sr13334*	conserved hypothetical protein	2.26451	2.26451	no
S. reilianum f. sp. *zeae*	*sr13398*	conserved hypothetical protein	1.48512	≤ 1	no
S. reilianum f. sp. *zeae*	*sr13409*	conserved hypothetical protein	≤ 1	1.10732	no
S. reilianum f. sp. *zeae*	*sr13412*	related to membrane protein Dik6	1.48676	1.48676	no
S. reilianum f. sp. *zeae*	*sr13415*	hypothetical protein	1.14552	1.14552	yes
S. reilianum f. sp. *zeae*	*sr13419*	conserved hypothetical *Ustilagnaceae*-specific protein	1.24534	1.24534	yes
S. reilianum f. sp. *zeae*	*sr13490*	conserved hypothetical protein	2.74755	2.74755	yes
S. reilianum f. sp. *zeae*	*sr13496*	conserved hypothetical *Ustilaginaceae*-specific protein	2.17684	2.17684	yes
S. reilianum f. sp. *zeae*	*sr13671*	conserved hypothetical protein	1.05028	1.05028	no
S. reilianum f. sp. *zeae*	*sr13869*	conserved hypothetical protein	1.33366	1.33618	yes
S. reilianum f. sp. *zeae*	*sr14022*	conserved hypothetical protein	2.59006	2.59006	no
S. reilianum f. sp. *zeae*	*sr14083*	conserved hypothetical *Ustilaginaceae*-specific protein	1.11394	1.11394	yes
S. reilianum f. sp. *zeae*	*sr14347*	conserved hypothetical protein	544.37	544.37	yes
S. reilianum f. sp. *zeae*	*sr14406*	related to putative monooxygenase	≤ 1	688.034	no
S. reilianum f. sp. *zeae*	*sr14463*	conserved hypothetical protein	≤ 1	665.329	no
S. reilianum f. sp. *zeae*	*sr14558*	conserved hypothetical protein	24.355	24.355	yes
S. reilianum f. sp. *zeae*	*sr14682*	conserved hypothetical protein	≤ 1	3.1438	no
S. reilianum f. sp. *zeae*	*sr14768*	related to UTP6 - U3 snoRNP protein	≤ 1	1.20122	no
S. reilianum f. sp. *zeae*	*sr14937*	probable beta-glucosidase	1.08064	≤ 1	yes
S. reilianum f. sp. *zeae*	*sr14939*	conserved hypothetical protein	1.80117	1.80117	no
S. reilianum f. sp. *zeae*	*sr14944*	conserved hypothetical *Ustilaginaceae*-specific protein	4.30527	4.30527	yes
S. reilianum f. sp. *zeae*	*sr15147*	conserved hypothetical protein	31.7216	31.7216	yes
S. reilianum f. sp. *zeae*	*sr15149*	conserved hypothetical protein	4.2176	4.2176	yes

S. reilianum f. sp. *zeae*	*sr15192*	conserved hypothetical protein	≤ 1	101.265	no
S. reilianum f. sp. *zeae*	*sr15203*	conserved hypothetical protein	1.61396	1.61396	no
S. reilianum f. sp. *zeae*	*sr15229*	conserved hypothetical protein	1.11082	1.11082	yes
S. reilianum f. sp. *zeae*	*sr15255*	conserved hypothetical protein	≤ 1	498.94	no
S. reilianum f. sp. *zeae*	*sr15500*	related to APP1 - Actin Patch Protein	≤ 1	4.21556	no
S. reilianum f. sp. *zeae*	*sr15591*	related to TPT1 - tRNA 2‘-phosphotransferase	1.00706	≤ 1	no
S. reilianum f. sp. *zeae*	*sr15671*	conserved hypothetical protein	1.88713	1.88713	no
S. reilianum f. sp. *zeae*	*sr15716*	conserved hypothetical protein	≤ 1	337.43	no
S. reilianum f. sp. *zeae*	*sr15939*	conserved hypothetical protein	2.47531	infinite	no
S. reilianum f. sp. *zeae*	*sr15970*	conserved hypothetical protein	≤ 1	604.175	no
S. reilianum f. sp. *zeae*	*sr16021*	conserved hypothetical protein	2.51846	≤ 1	no
S. reilianum f. sp. *zeae*	*sr16116*	probable acyl transferase-like protein	1.02694	≤ 1	no
S. reilianum f. sp. *zeae*	*sr16119*	conserved hypothetical protein	2.11268	2.11268	no
S. reilianum f. sp. *zeae*	*sr16120*	conserved hypothetical protein	1.9598	1.9598	no
S. reilianum f. sp. *zeae*	*sr16122*	related to Cytochrome P450	1.13438	1.06102	no
S. reilianum f. sp. *zeae*	*sr16135*	probable ATP18 - subunit i/j of the mitochondrial F1F0-ATP synthase	1.13546	≤ 1	no
S. reilianum f. sp. *zeae*	*sr16278*	conserved hypothetical protein	1.039	1.041	no
S. reilianum f. sp. *zeae*	*sr16520*	conserved hypothetical protein	7.70556	7.70556	no
S. reilianum f. sp. *zeae*	*sr16553*	conserved hypothetical Ustilago-specific protein	3.22326	3.22326	yes
S. reilianum f. sp. *zeae*	*sr16556*	conserved hypothetical Ustilago-specific protein	3.23532	3.23532	no
S. reilianum f. sp. *zeae*	*sr16558*	conserved hypothetical *Ustilaginaceae*-specific protein	2.87797	2.87797	yes
S. reilianum f. sp. *zeae*	*sr16642*	conserved hypothetical *Ustilaginceae*-specific protein	1.02783	≤ 1	no
S. reilianum f. sp. *zeae*	*sr16650*	probable HMG-box transcription factor	40.4119	40.4119	no
S. reilianum f. sp. *zeae*	*sr16859*	conserved hypothetical protein	≤ 1	1.87546	no
S. reilianum f. sp. *zeae*	*sr20006*	conserved hypothetical *Ustilaginaceae*-specific protein	10.6794	10.6794	yes
S. reilianum f. sp. *sorghi*	*srs_ 00093*	hypothetical protein	1.71197	1.71197	no
S. reilianum f. sp. *sorghi*	*srs_ 00846*	conserved hypothetical protein	3.07138	infinite	yes
S. reilianum f. sp. *sorghi*	*srs_ 06410*	conserved hypothetical protein	≤ 1	29.18	no
S. reilianum f. sp. *sorghi*	*srs_ 06431*	conserved hypothetical protein	475.801	475.801	no
S. reilianum f. sp. *sorghi*	*srs_ 06452*	related to NADH-dependent flavin oxidoreductase	1.40228	1.40228	no
S. reilianum f. sp. *sorghi*	*srs_ 10057*	conserved hypothetical protein	1.17126	1.17126	yes

S. reilianum f. sp. *sorghi*	*srs_10059*	conserved hypothetical protein	6.53881	6.53881	yes
S. reilianum f. sp. *sorghi*	*srs_10087*	conserved hypothetical protein	1.56546	1.56546	no
S. reilianum f. sp. *sorghi*	*srs_10182*	conserved hypothetical protein	1.57473	1.57473	yes
S. reilianum f. sp. *sorghi*	*srs_10206*	conserved hypothetical protein	1.24279	≤ 1	no
S. reilianum f. sp. *sorghi*	*srs_10317*	conserved hypothetical protein	3.53338	3.53338	yes
S. reilianum f. sp. *sorghi*	*srs_10515*	conserved hypothetical protein	2.27231	2.37006	no
S. reilianum f. sp. *sorghi*	*srs_10529*	conserved hypothetical protein	31.1469	31.1469	yes
S. reilianum f. sp. *sorghi*	*srs_10680*	probable methylglutaconyl-coa hydratase; mitochondrial precursor	1.53935	infinite	no
S. reilianum f. sp. *sorghi*	*srs_10775*	conserved hypothetical protein	≤ 1	1.47161	no
S. reilianum f. sp. *sorghi*	*srs_10916*	conserved hypothetical protein	1.45001	1.45001	yes
S. reilianum f. sp. *sorghi*	*srs_11154*	conserved hypothetical protein	1.20855	infinite	no
S. reilianum f. sp. *sorghi*	*srs_11233*	conserved hypothetical protein	5.22937	5.22937	yes
S. reilianum f. sp. *sorghi*	*srs_11237*	conserved hypothetical protein	10.7826	10.7826	yes
S. reilianum f. sp. *sorghi*	*srs_11239*	conserved hypothetical protein	5.95563	3.93728	no
S. reilianum f. sp. *sorghi*	*srs_11240*	conserved hypothetical protein	5.95563	infinite	no
S. reilianum f. sp. *sorghi*	*srs_11254*	conserved hypothetical protein	2.24726	≤ 1	no
S. reilianum f. sp. *sorghi*	*srs_11259*	conserved hypothetical protein	2.30171	infinite	no
S. reilianum f. sp. *sorghi*	*srs_11400*	conserved hypothetical protein	1.13351	1.13351	yes
S. reilianum f. sp. *sorghi*	*srs_11612*	conserved hypothetical protein	9.27278	9.27278	no
S. reilianum f. sp. *sorghi*	*srs_11638*	conserved hypothetical protein	1.63895	1.63895	no
S. reilianum f. sp. *sorghi*	*srs_11756*	conserved hypothetical protein	2.60354	≤ 1	no
S. reilianum f. sp. *sorghi*	*srs_11791*	conserved hypothetical protein	1.08237	1.08237	no
S. reilianum f. sp. *sorghi*	*srs_11797*	conserved hypothetical protein	≤ 1	1.83438	no
S. reilianum f. sp. *sorghi*	*srs_11929*	conserved hypothetical protein	2.52875	2.52875	no
S. reilianum f. sp. *sorghi*	*srs_12108*	conserved hypothetical protein	≤ 1	1.95946	no
S. reilianum f. sp. *sorghi*	*srs_12173*	related to YEA4 - uridine diphosphate-N-acetylglucosamine	1.39832	1.39816	no
S. reilianum f. sp. *sorghi*	*srs_12269*	conserved hypothetical protein	1.23926	1.23926	no
S. reilianum f. sp. *sorghi*	*srs_12270*	related to nucleoside hydrolase	6.41352	6.41352	no
S. reilianum f. sp. *sorghi*	*srs_12295*	conserved hypothetical protein	1.19001	1.19001	no
S. reilianum f. sp. *sorghi*	*srs_12344*	conserved hypothetical protein	1.80458	28.4209	no
S. reilianum f. sp. *sorghi*	*srs_12419*	conserved hypothetical protein	1.01299	≤ 1	no

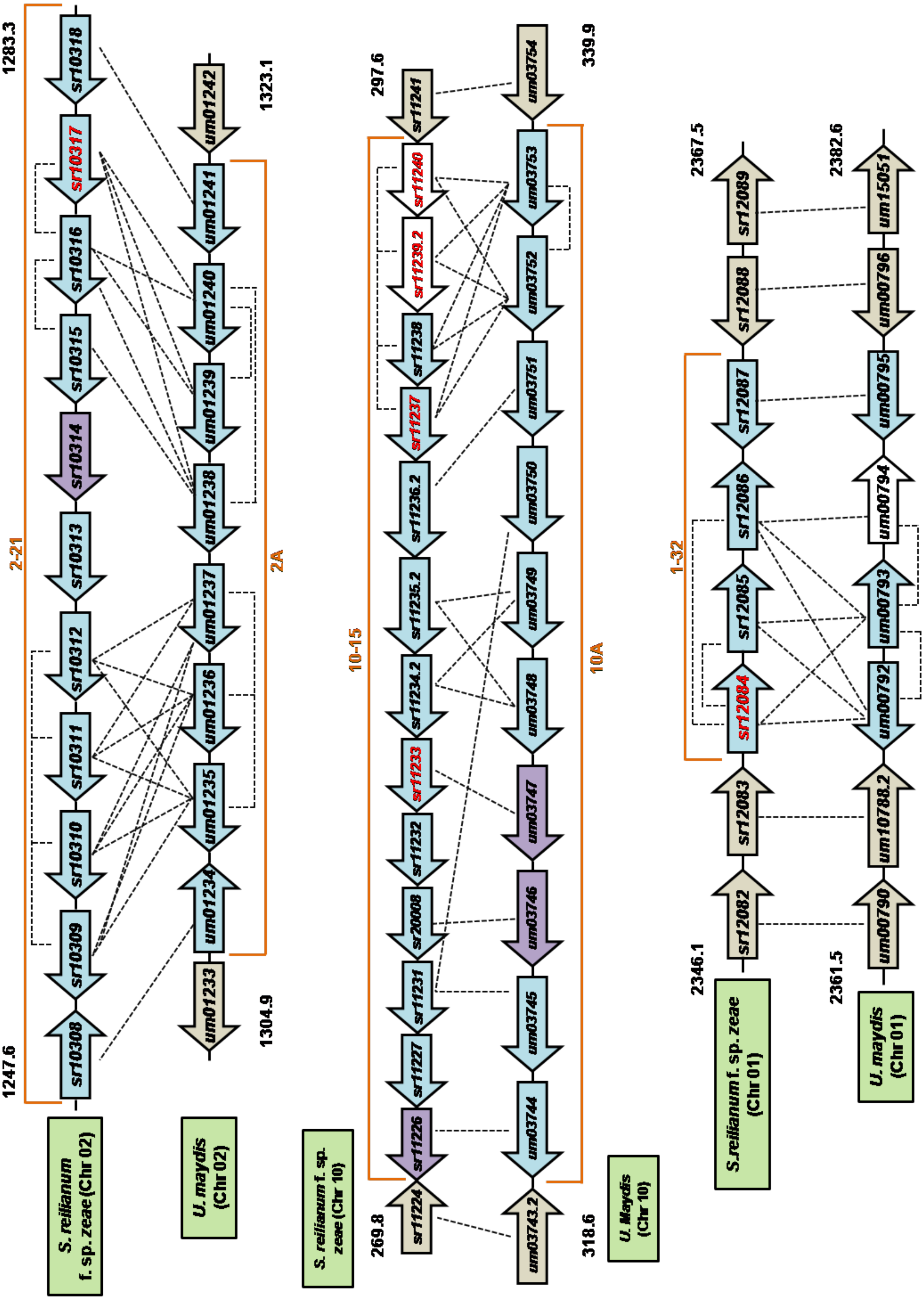
S. reilianum f. sp. *sorghi*	*srs_12427*	hypothetical protein	1.06862	1.06862	no
S. reilianum f. sp. *sorghi*	*srs_12456*	probable KRE2 - alpha-1;2-mannosyltransferase	1.50484	5.96477	no
S. reilianum f. sp. *sorghi*	*srs_12501*	related to Enoyl-CoA hydratase	1.53935	≤ 1	no
S. reilianum f. sp. *sorghi*	*srs_12655*	related to NCR1 - transmembrane glycoprotein; involved in sphingolipid metabolism	1.60055	infinite	no
S. reilianum f. sp. *sorghi*	*srs_12757*	conserved hypothetical protein	≤ 1	707.986	no
S. reilianum f. sp. *sorghi*	*srs_12770*	conserved hypothetical protein	1.55237	1.55237	no
S. reilianum f. sp. *sorghi*	*srs_12776*	conserved hypothetical protein	≤ 1	2.12026	no
S. reilianum f. sp. *sorghi*	*srs_12877*	related to molybdopterin synthase large subunit	1.15798	≤ 1	no
S. reilianum f. sp. *sorghi*	*srs_12957*	probable mfs-multidrug-resistance transporter	≤ 1	907.41	no
S. reilianum f. sp. *sorghi*	*srs_12968*	conserved hypothetical protein	37.9007	37.9007	yes
S. reilianum f. sp. *sorghi*	*srs_12970*	conserved hypothetical protein	≤ 1	1.72837	no
S. reilianum f. sp. *sorghi*	*srs_12972*	conserved hypothetical protein	1.76216	1.76216	no
S. reilianum f. sp. *sorghi*	*srs_13056*	related to MDR1 - Mac1p interacting protein	2.27469	2.15835	no
S. reilianum f. sp. *sorghi*	*srs_13060*	conserved hypothetical protein	≤ 1	1.29293	yes
S. reilianum f. sp. *sorghi*	*srs_13205*	conserved hypothetical protein	≤ 1	1.08661	no
S. reilianum f. sp. *sorghi*	*srs_13324*	conserved hypothetical protein	≤ 1	1.5301	no
S. reilianum f. sp. *sorghi*	*srs_13334*	conserved hypothetical protein	2.26451	2.26451	no
S. reilianum f. sp. *sorghi*	*srs_13398*	conserved hypothetical protein	1.48512	≤ 1	no
S. reilianum f. sp. *sorghi*	*srs_13412*	related to membrane protein Dik6	1.48676	1.48676	no
S. reilianum f. sp. *sorghi*	*srs_13415*	hypothetical protein	1.14552	1.14552	yes
S. reilianum f. sp. *sorghi*	*srs_13419*	conserved hypothetical protein	1.24534	1.24534	yes
S. reilianum f. sp. *sorghi*	*srs_13490*	conserved hypothetical protein	2.74755	2.74755	yes
S. reilianum f. sp. *sorghi*	*srs_13496*	conserved hypothetical protein	2.17684	2.17684	yes
S. reilianum f. sp. *sorghi*	*srs_13546*	conserved hypothetical protein	≤ 1	1.54943	no
S. reilianum f. sp. *sorghi*	*srs_13661*	related to TAF2 - component of TFIID complex	876.665	876.665	no
S. reilianum f. sp. *sorghi*	*srs_13671*	conserved hypothetical protein	1.05028	1.05028	no
S. reilianum f. sp. *sorghi*	*srs_13781*	related to molybdenum cofactor biosynthetic protein	1.41492	1.41492	no
S. reilianum f. sp. *sorghi*	*srs_13869*	conserved hypothetical protein	1.33366	1.33618	yes
S. reilianum f. sp. *sorghi*	*srs_13946*	related to SHP1 - potential regulatory subunit for Glc7p	≤ 1	1.12427	no
S. reilianum f. sp. *sorghi*	*srs_14083*	conserved hypothetical protein	1.11394	1.11394	yes
S. reilianum f. sp. *sorghi*	*srs_14529*	related to Zinc finger protein SFP1	4.77608	4.77608	no

S. reilianum f. sp. *sorghi*	*srs_14785*	conserved hypothetical protein	63.1724	63.1724	no
S. reilianum f. sp. *sorghi*	*srs_14894*	conserved hypothetical protein	≤ 1	9.49286	no
S. reilianum f. sp. *sorghi*	*srs_14937*	probable beta-glucosidase	1.08064	infinite	yes
S. reilianum f. sp. *sorghi*	*srs_14939*	conserved hypothetical protein	1.80117	1.80117	no
S. reilianum f. sp. *sorghi*	*srs_14944*	conserved hypothetical protein	4.30527	4.30527	yes
S. reilianum f. sp. *sorghi*	*srs_14996*	probable PHO8 - repressible alkaline phosphatase vacuolar	≤ 1	677.664	yes
S. reilianum f. sp. *sorghi*	*srs_15033*	conserved hypothetical protein	1.56395	1.56395	no
S. reilianum f. sp. *sorghi*	*srs_15122*	conserved hypothetical protein	≤ 1	7.4706	no
S. reilianum f. sp. *sorghi*	*srs_15147*	conserved hypothetical protein	31.7216	31.7216	yes
S. reilianum f. sp. *sorghi*	*srs_15149*	conserved hypothetical protein	4.2176	4.2176	yes
S. reilianum f. sp. *sorghi*	*srs_15203*	conserved hypothetical protein	1.61396	1.61396	no
S. reilianum f. sp. *sorghi*	*srs_15229*	conserved hypothetical protein	1.11082	1.11082	yes
S. reilianum f. sp. *sorghi*	*srs_15315*	conserved hypothetical protein	≤ 1	1.57954	no
S. reilianum f. sp. *sorghi*	*srs_15329*	related to IST2 - Plasma membrane protein that may be involved in osmotolerance	446.687	446.687	no
S. reilianum f. sp. *sorghi*	*srs_15591*	related to TPT1 - tRNA 1 2‘-phosphotransferase	1.00706	≤ 1	no
S. reilianum f. sp. *sorghi*	*srs_15619*	conserved hypothetical protein	1.91947	1.91947	no
S. reilianum f. sp. *sorghi*	*srs_15654*	probable small nuclear ribonucleoprotein chain D2	≤ 1	407.237	no
S. reilianum f. sp. *sorghi*	*srs_15918*	related to glyoxylate/hydroxypyruvate reductase	≤ 1	1.79065	no
S. reilianum f. sp. *sorghi*	*srs_15939*	conserved hypothetical protein	2.47531	≤ 1	no
S. reilianum f. sp. *sorghi*	*srs_15961*	conserved hypothetical protein	2.25348	12.4355	no
S. reilianum f. sp. *sorghi*	*srs_16021*	conserved hypothetical protein	2.51846	infinite	no
S. reilianum f. sp. *sorghi*	*srs_16116*	probable acyl transferase-like protein	1.02694	≤ 1	no
S. reilianum f. sp. *sorghi*	*srs_16120*	conserved hypothetical protein	1.9598	1.9598	no
S. reilianum f. sp. *sorghi*	*srs_16122*	related to Cytochrome P450	1.13438	≤ 1	no
S. reilianum f. sp. *sorghi*	*srs_16129*	conserved hypothetical protein	≤ 1	537.788	no
S. reilianum f. sp. *sorghi*	*srs_16135*	probable ATP18 - subunit i/j of the mitochondrial F1F0-ATP synthase	1.13546	≤ 1	no
S. reilianum f. sp. *sorghi*	*srs_16157*	conserved hypothetical protein	≤ 1	463.206	no
S. reilianum f. sp. *sorghi*	*srs_16278*	conserved hypothetical protein	1.039	1.041	no
S. reilianum f. sp. *sorghi*	*srs_16371*	conserved hypothetical protein	≤ 1	5.99124	no
S. reilianum f. sp. *sorghi*	*srs_16377*	conserved hypothetical protein	≤ 1	1.50671	no
S. reilianum f. sp. *sorghi*	*srs_16449*	conserved hypothetical protein	997.171	infinite	no

S. reilianum f. sp. *sorghi*	*srs_16466*	probable Myp1 protein	≤ 1	3.90275	no
S. reilianum f. sp. *sorghi*	*srs_16553*	conserved hypothetical protein	3.22326	3.22326	yes
S. reilianum f. sp. *sorghi*	*srs_16556*	conserved hypothetical Ustilago-specific protein	3.23532	3.23532	no
S. reilianum f. sp. *sorghi*	*srs_16558*	conserved hypothetical protein	2.87797	2.87797	yes
S. reilianum f. sp. *sorghi*	*srs_16588*	conserved hypothetical protein	≤ 1	357.195	no
S. reilianum f. sp. *sorghi*	*srs_16642*	conserved hypothetical protein	1.02783	≤ 1	no
S. reilianum f. sp. *sorghi*	*srs_16863*	conserved hypothetical protein	≤ 1	1.03443	no
S. reilianum f. sp. *sorghi*	*srs_20006*	conserved hypothetical protein	10.6794	10.6794	yes
S. reilianum f. sp. *sorghi*	*srs_25015*	conserved hypothetical protein	1.02783	infinite	no
S. reilianum f. sp. *sorghi*	*srs_25027*	conserved hypothetical protein	1.02783	≤ 1	no
S. reilianum f. sp. *sorghi*	*srs_25043*	conserved hypothetical protein	1.02783	1.11774	no
S. reilianum f. sp. *sorghi*	*srs_25047*	conserved hypothetical protein	1.02783	infinite	no
S. scitamineum	*SSCI_00812*	related to TPT1 - tRNA 1 2‘-phosphotransferase	1.00706	≤ 1	no
S. scitamineum	*SSCI_01369*	conserved hypothetical protein	≤ 1	15.104	no
S. scitamineum	*SSCI_02435*	related to carbonic anhydrase	366.235	823.893	no
S. scitamineum	*SSCI_03055*	related to molybdopterin synthase large subunit	1.15798	≤ 1	no
S. scitamineum	*SSCI_04694*	conserved hypothetical protein	2.86628	2.86628	no
S. scitamineum	*SSCI_05503*	conserved hypothetical protein	1.30406	1.30406	no
S. scitamineum	*SSCI_06326*	probable RPL40A - Ubiquitin	5.02301	5.02301	no
U. hordei	*UH_00833*	conserved hypothetical protein	≤ 1	941.957	no
U. hordei	*UH_01988*	conserved hypothetical *Ustilaginaceae*-specific protein (N-terminal fragment)	≤ 1	1.14629	no
U. hordei	*UH_02987*	hypothetical protein	1.59605	1.10372	no
U. hordei	*UH_03016*	hypothetical protein	1.07083	≤ 1	no
U. hordei	*UH_03570*	conserved hypothetical protein	1.48512	1.64585	no
U. hordei	*UH_04676*	related to Mig1 protein	1.35023	1.35023	yes
U. hordei	*UH_04736*	related to Mig1 protein; induced during biotrophic phase	1.58458	1.20234	yes
U. hordei	*UH_04922*	related to Mig1 protein	1.35023	1.35023	yes
U. hordei	*UH_04923*	related to Mig1 protein	1.35023	1.35023	yes
U. hordei	*UH_04990*	related to Mig1 - Mig1 protein; induced during biotrophic phase	1.14956	1.14956	yes
U. hordei	*UH_05685*	probable CPR1 - cyclophilin (peptidylprolyl isomerase)	≤ 1	1.29869	no
U. hordei	*UH_06051*	related to Mig1 protein; induced during biotrophic phase	1.58458	1.97029	yes

U. hordei	*UH_06234*	related to Mig1 protein; induced during biotrophic phase	1.58458	1.48227	yes
U. hordei	*UH_06285*	hypothetical protein	3.34085	3.8597	no
U. hordei	*UH_06702*	related to Mig1 protein; induced during biotrophic phase	1.58458	1.48227	yes
U. hordei	*UH_06803*	hypothetical protein	1.14956	1.14956	yes
U. hordei	*UH_06868*	conserved hypothetical protein	1.83929	1.83929	no
U. hordei	*UH_07399*	conserved hypothetical protein	1.48512	1.64585	no
U. hordei	*UH_08252*	related to Mig1 protein; induced during biotrophic phase	1.58458	1.97029	yes
U. hordei	*UH_08758*	related to Mig2-2	4.15392	4.15392	yes
U. hordei	*UH_08826*	related to Mig1 protein; induced during biotrophic phase	1.58458	1.48227	yes
U. hordei	*UH_08960*	conserved hypothetical protein	1.14956	1.14956	yes
U. hordei	*UH_12520*	hypothetical protein	1.07083	≤ 1	no
U. hordei	*UH_12644*	related to Mig2 protein	4.15392	4.15392	yes
U. hordei	*UH_12937*	hypothetical protein	≤ 1	1.30498	no
U. hordei	*UH_13170*	conserved hypothetical *Ustilaginaceae*-specific protein	2.42678	2.42678	yes
U. hordei	*UH_13752*	hypothetical protein	≤ 1	1.05011	no
U. hordei	*UH_14039*	conserved hypothetical protein	1.346	1.346	no
U. hordei	*UH_14091*	hypothetical protein	3.05769	≤ 1	no
U. hordei	*UH_14097*	hypothetical protein	3.05769	≤ 1	no
U. hordei	*UH_14101*	hypothetical protein	3.05769	infinite	no
U. hordei	*UH_14108*	conserved hypothetical protein	1.23253	1.23253	yes
U. hordei	*UH_14114*	conserved hypothetical protein	1.23253	1.23253	yes
U. hordei	*UH_14135*	hypothetical protein	3.05769	≤ 1	yes
U. hordei	*UH_14141*	hypothetical protein	3.05769	infinite	no
U. hordei	*UH_14142*	hypothetical protein	1.346	1.346	no
U. hordei	*UH_14166*	hypothetical protein	3.05769	infinite	no
U. hordei	*UH_14168*	hypothetical protein	1.23253	1.23253	no
U. hordei	*UH_14176*	hypothetical protein	1.346	1.346	no
U. hordei	*UH_14447*	hypothetical protein	1.07083	1.82423	no
U. hordei	*UH_14451*	hypothetical protein	≤ 1	9.27024	no
U. hordei	*UH_14558*	hypothetical protein	1.20794	1.20105	no
U. hordei	*UH_14700*	hypothetical protein (N-terminal fragment)	≤ 1	1.14629	no

U. hordei	*UH_14902*	related to Mig2	4.15392	4.15392	yes
U. hordei	*UH_15096*	hypothetical protein	1.20794	1.20105	no
U. hordei	*UH_15387*	hypothetical protein	2.42678	2.42678	yes
U. hordei	*UH_15725*	hypothetical protein	2.42678	2.42678	yes
U. hordei	*UH_15858*	hypothetical protein	2.42678	2.42678	yes
U. hordei	*UH_16658*	hypothetical protein	≤ 1	1.30498	no
U. maydis	*um03440*	conserved hypothetical Ustilago-specific protein	9.03615	9.03615	no
U. maydis	*um05426*	conserved hypothetical Ustilago-specific protein	9.03615	9.03615	no


S. reilianum f. sp. zeae (Chr 02)
U. maydis (Chr 02)
1247.6
1283.3
1304.9
1323.1
2-21
2A
sr10308
sr10309
sr10310
sr10311
sr10312
sr10313
sr10314
sr10315
sr10316
sr10317
sr10318
um01233
um01234
um01235
um01236
um01237
um01238
um01239
um01240
um01241
um01242
S. reilianum f. sp. zeae (Chr 10)
U. Maydis (Chr 10)
269.8
297.6
318.6
339.9
10-15
10A
sr11224
sr11226
sr11227
sr11231
sr20008
sr11232
sr11233
sr11234.2
sr11235.2
sr11236.2
sr11237
sr11238
sr11239.2
sr11240
sr11241
um03743.2
um03744
um03745
um03746
um03747
um03748
um03749
um03750
um03751
um03752
um03753
um03754
S.reilianum f. sp. zeae (Chr 01)
U. maydis (Chr 01)
2346.1
2367.5
2361.5
2382.6
1-32
sr12082
sr12083
sr12084
sr12085
sr12086
sr12087
sr12088
sr12089
um00790
um10788.2
um00792
um00793
um00794
um00795
um00796
um15051

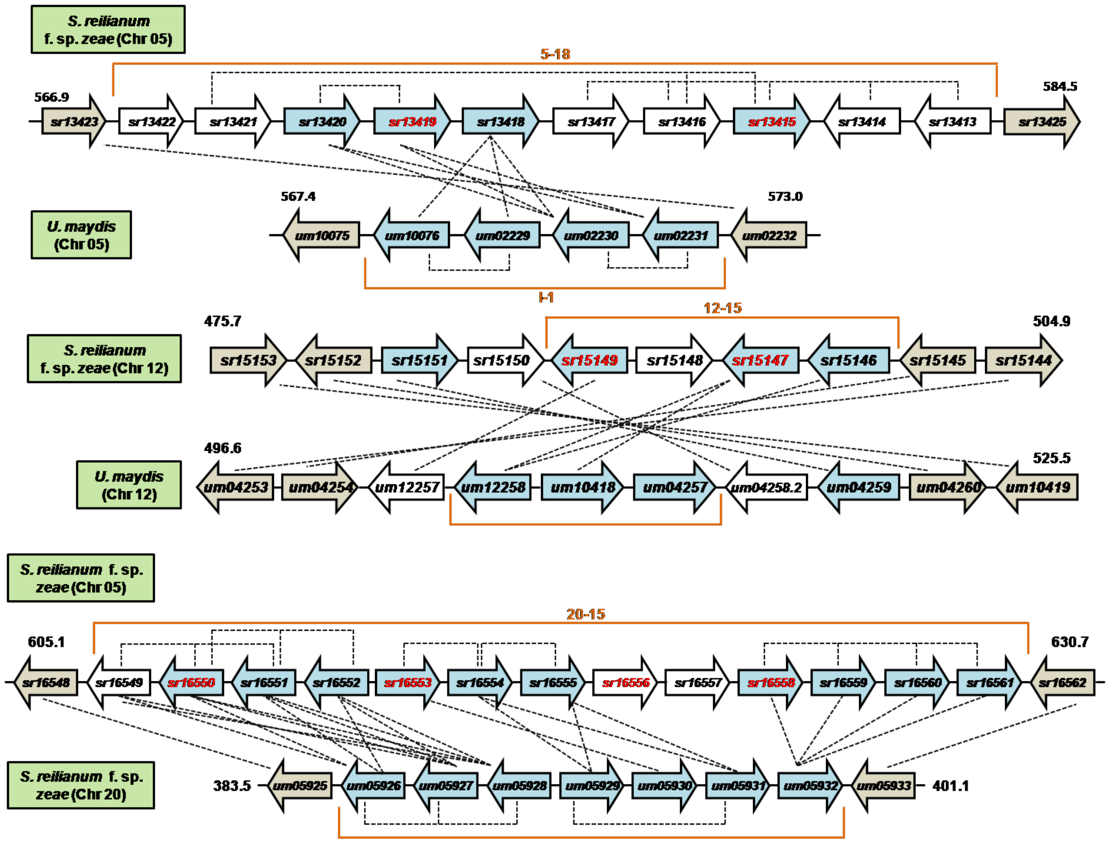


Figure 6.1: Relationships between gene clusters 1-32, 5-18, 10-15, 20-15, 12-15 and 2-21 of *S. reilianum* f. sp. *zeae* and homologous regions in *U. maydis*. The genomic region of *S. reilianum* f. sp. *zeae* is shown on the upper part of each cluster comparison.Arrows indicate the orientation of genes and lines between arrows symbolize intergenic regions (not drawn to scale). Blue arrows indicate genes encoding a predicted secreted protein. Purple arrows indicate an ambiguous prediction of secretion. Brown arrows show genes adjacent to the gene cluster encoding secreted proteins. The red gene name indicates positive selection of this gene. Numbers show the position on the chromosome in kbp. Dashed lines illustrate homology according to the all-agains-all blast search with an e-Value cutoff of 0.001. Paralogues and orthologues which are not part of the genomic region shown here as well has hits only found in one direction are not indicated. Orange brackets mark a cluster described previously (Kämper *et al.*, 2006; Schirawski *et al.*, 2010).

ACKNOWLEDGMENTS

With great pleasure, I would like to express my deep gratitude to all who contributed in one way or the other to this thesis:

my supervisors Prof. Dr. Regine Kahmann and Dr. Julien Y. Dutheil for their suggestions, support and encouragement throughout the development of my PhD project. I am grateful to Regine Kahmann for giving me the freedom to enter the world of comparative genomics and to touch on related research questions. I sincerely thank Julien for guiding me on this journey and for assisting me in transforming my perception of bioinformatics from a closed book to something useful and eventually something enjoyable.

the members of my advisory thesis committee Prof. Dr. Eva H. Stukenbrock, Prof. Dr. Anke Becker and Prof. Dr. Hans-Ulrich Mösch for their suggestions and guidance.
I thank Anke Becker also for serving as reviewer of this thesis and Hans-Ulrich Mösch and Prof. Dr. Alfred Batschauer for agreeing to be members of the examination committee.

all present and former lab members for lots of motivation, support and discussion. I especially thank Karin Münch for a wonderful lab neighborhood and for introducing me to the molecular biology of smut fungi. I am very grateful for numerous strains and virulence assays created by her. They are a valuable part of this thesis!

the International Max Planck Research School for Environmental, Cellular and Molecular Microbiology for support and its members for discussions and exchange of ideas.

all participants of the ’Genomics Journal Club of the MPI Marburg’, especially Eva and Julien for hosting it. I really enjoyed interesting and helpful discussions.

my family and friends for their encouragement and for sharing the sacrifices associated with a PhD project, even without experiencing the joyful sites of it.

CURRICULUM VITAE