Identification of host-specific virulence factors of the smut fungus *Sporisorium reilianum* by genotype to phenotype mapping

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der
RWTH Aachen University zur Erlangung des akademischen Grades einer
Doktorin der Naturwissenschaften

vorgelegt von

Nilam Nayan Borah M. Sc.

aus Jorhat, Indien

Berichter: Prof. Dr. Jan Schirawski
Prof. Dr. Lars Mathias Blank

Tag der mündlichen Prüfung: 05.11.2018

Diese Dissertation ist auf den Internetseiten der Universitätsbibliothek verfügbar.

Part of this dissertation has already been published. Thus, wording and figure belonging to the introduction part and the published article in part can be identical.

Teile dieser Dissertation wurden bereits veröffentlicht. Deshalb können Wortwahl und Abbildungen in dieser Dissertation mit denen der Publikationen teilweise identisch sein.

Students that worked under this project:
1. Patrick Schwinges - worked as HiWi and helped in inoculating the plants. Also contributed to isolate *a1b1* mating population.
2. David Weniger - a1b1 population isolation

Students supervised that worked under this project:
1. Melissa Romich - worked as HiWi, contributed in inoculating the plants and phenotyping the disease rates.
2. Guido Henning - worked as HiWi and helped in inoculating the plants.
3. Ann-Christin Böttcher - worked as HiWi and contributed to isolate *a1b1* mating population.
4. Hana Adamova - worked as HiWi and contributed in making stock samples.
5. Alan Midani - Masters practical, 'Microscopic investigation of plant infection capacities of non-virulent SRSZ hybrids'.
6. Tobias Reinicke - Bachelor work, 'Correlation of presence or expression of a putative host determining gene cluster to host-specific virulence of *Sporisorium reilianum* segregants'.
7. Tom Spilker - Masters practical, 'Microscopy of avirulent SRSZ hybrids'.

Conferences attended:

1. 2018: 14th European Conference on Fungal Genetics, Haifa, Israel - Poster presentation
2. 2016: 13th European Conference on Fungal Genetics, Paris, France - Poster presentation
3. 2015: Annual Conference of the Association for General and Applied Microbiology, Marburg, Germany - Poster presentation

I dedicate my Phd dissertation to my parents ("Ma" & "Deta").

"And, when you want something, all the universe conspires in helping you to achieve it."

— Paulo Coelho, The Alchemist

Contents

Figure list

Table list

Summary

The two host-adapted varieties of the smut fungus *Sporisorium reilianum* (namely *S. reilianum* f. sp. *zeae*, SRZ, and *S. reilianum* f. sp. *reilianum*, SRS) produce spores either on maize or on sorghum. For plant infection, mating compatible haploid sporidia need to fuse and form infectious dikaryotic filaments that infect the plant at seedling stage and cause disease phenotype in the form of spore development in male and female inflorescences of the plant. Despite both *formae speciales* being very closely related on the genomic level why they differ in terms of host selection remains elusive. To know the host determining factors we generated sexual spores by crossing the compatible SRZ strain SRZ1_5-2 (mating type *a1b1*) with the SRS strain SRS2_H2-7 (mating type *a2b6*). Meiotic progeny (SRSZ) with the mating type *a1b1* were selected and tested for virulence on sorghum after mating with SRS_H2-7. Virulence assays showed that the progeny were either non-virulent or showed various degrees of disease phenotype. We selected 110 non-virulent offspring, 50 offspring with full virulence and 29 offspring with intermediate virulence on sorghum for genotype analysis. Genomic DNA of 191 strains (189 offspring and 2 parental strains) was isolated and sequenced using Illumina technology with a read length of 125 nt in paired end. Mapping of the reads against the two assembled sister parental genomes showed that parental origin of chromosomal regions could be unambiguously assigned for all SRSZ strains. Interestingly, few strains contained partially duplicated genomic regions, i.e. carrying the same chromosomal fragments from both parents. Correlation of phenotype with parental origin of genomic loci revealed a region of 35 genes in the left arm of chromosome 7 potentially associated with the virulence phenotype on sorghum. Analysis of this region revealed a cluster of 9 genes that were expressed *in planta*, showed low sequence conservation and coded for proteins carrying predicted secretion signal peptides. Deletion of this gene cluster in *forma specialis* SRS and subsequent virulence analysis showed that this cluster is essential for the spore forming capacity of SRS in sorghum. These results indicate the power of classical genetics combined with genotype to phenotype association analysis using Next generation sequencing approaches for the discovery of genes involved in determining host specificity in *S. reilianum*.

Zusammenfassung

Die beiden hostadaptierten Varietäten des Brandpilzes *Sporisorium reilianum* (*S. reilianum* f. sp. *zeae*, SRZ und *S. reilianum* f. sp. *reilianum*, SRS) produzieren Sporen entweder auf Mais oder auf Hirse. Für eine Pflanzeninfektion müssen paartaugliche haploide Sporidien verschmelzen und infektiöse dikaryotische Filamente bilden, die die Pflanze im Keimlingsstadium infizieren und einen Krankheitsphänotyp in Form von Sporenentwicklung in den Blütenständen der Pflanze verursachen. Obwohl beide Varietäten auf genomischer Ebene sehr eng verwandt sind, sind die molekularen Ursachen der Wirtsselektion unbekannt. Um die wirtsbestimmenden Faktoren zu identifizieren, erzeugten wir Geschlechtssporen, indem wir den kompatiblen SRZ-Stamm SRZ1_5-2 (Kreuzungstyp *a1b1*) mit dem SRS-Stamm SRS2_H2-7 (Kreuzungstyp *a2b6*) kreuzten. Meiotische Nachkommen (SRSZ) mit dem Paarungstyp *a1b1* wurden ausgewählt und auf Virulenz auf Hirse nach Paarung mit SRS_H2-7 getestet. Die Virulenz-tests zeigten, dass die Nachkommen entweder nicht-virulent waren oder verschiedene Grade des Krankheitsphänotyps zeigten. Wir wählten 110 nicht-virulente Nachkommen, 50 Nachkommen mit voller Virulenz und 29 Nachkommen mit intermediärer Virulenz auf Hirse für die Genotypanalyse aus. Genomische DNA von 191 Stämmen (189 Nachkommen und 2 Elternstämme) wurde isoliert und unter Verwendung der Illumina-Technologie mit einer Leselänge von 125 nt (paired-end) sequenziert. Die Kartierung der Reads gegen die beiden zusammengesetzten Schwester-Elterngenome zeigte, dass der Ursprung der chromosomalen Regionen für alle SRSZ-Stämme eindeutig zugeordnet werden konnte. Interessanterweise enthielten wenige Stämme teilweise duplizierte genomische Regionen, d. H. Sie trugen die gleichen chromosomalen Fragmente von beiden Eltern. Die Korrelation des Phänotyps mit dem Ursprung der genomischen Loci der Eltern ergab eine Region von 35 Genen im linken Arm des Chromosoms 7, die möglicherweise mit dem Virulenzphänotyp auf Hirse assoziiert ist. Die Analyse dieser Region ergab einen Cluster von 9 Genen, die in Hirse hochreguliert sind, nur eine geringe Sequenzkonservierung zwischen den Varietäten zeigen und für Proteine kodieren, die vorhergesagte Sekretionssignalpeptide tragen. Die Deletion dieses Genclusters in SRS und anschließende Virulenzanalyse auf Hirse zeigte, dass die Abwesenheit dieses Clusters die Sporenbildungsfähigkeit von SRS in Hirse aufhebt. Diese Ergebnisse zeigen, dass klassische Genetik kombiniert mit Phänotyp-Assoziationsanalyse unter Verwendung von NGS Methoden zur Identifizierung von Wirtsspezifitätsfaktoren in *S. reilianum* geeignet ist.

1 Introduction

1.1 Fungi cause various crop loss worldwide

Fungi cause serious crop diseases worldwide. The fact that there are more plant pathogenic fungi than there are plant pathogenic bacteria or viruses makes them a major threat to crop loss. The scenario is further complicated by an increased rate of invasion of new territories by fungi due to various factors including changes in agricultural practices, habitat disturbances, introduction from other areas into new ones, and global climate change. Fungi along with oomycetes turn out to be the most widespread "crop pest and pathogen (CPP)" [1]. It has been found that even though dispersal increases with host-range overall, fungi have the narrowest host range but are the most widely dispersed group [1]. More interestingly, a result showed that plant pathogenic fungi are marching towards the poles at a rate of 8 km pa thereby creating more challenges for disease control in northern hemisphere [2].

World food security is seriously threatened by fungal infections, with many nations underestimating the burden of pathogens in their farmers' fields. It has been shown that if we can save the crops lost to even low-level diseases, we could feed 8.5 percent of the 7 billion humans alive which is nearly 600 million people [3]. Therefore, elucidating how fungi overcome the plant's defense systems is a critical step for the development of resistant cultivars or efficient plant protection methods.

1.2 *Sporisorium reilianum* causes head smut on maize and sorghum

Smut fungi basically comprise a group of basidiomycetes falling under order Ustilaginales. The name "smut" comes from the fact that they produce black mass of spores that resemble smut called sori that are readily dispersed by wind. An interesting characteristic of smuts is the very limited host range, with most of them having between one and three different host plants [4]. Smut plant disease primarily affects grasses such as maize, wheat, sugarcane and sorghum. The most important diseases are caused by species from the genera *Ustilago*, *Tilletia* and *Sporisorium*. Maize and sorghum are two important crop plants that are challenged by fun-

gal smut disease. Particularly, one member of the smut fungi, *Sporisorium reilianum* (Kühn) Langdon & Fullerton causes head smut of maize and sorghum (**Figure 1**).
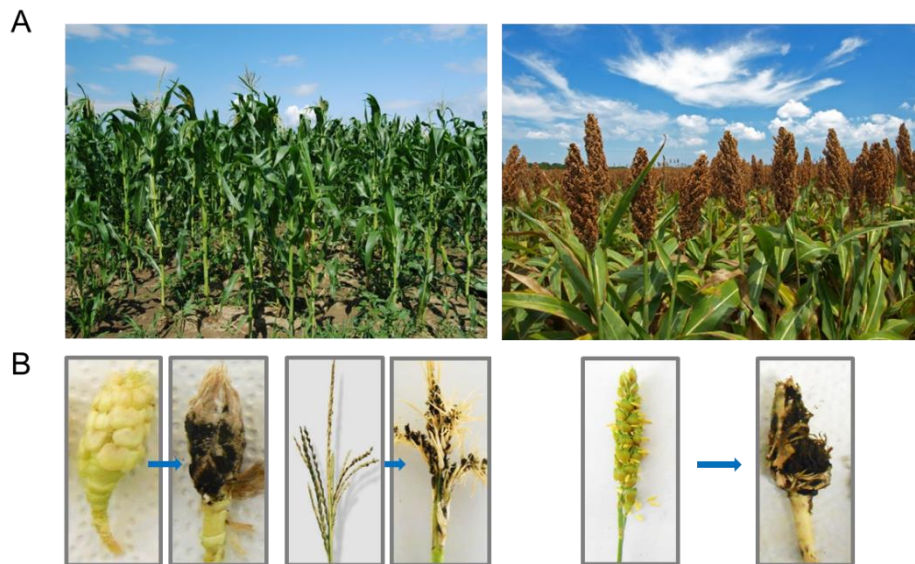
Figure 1: *Sporisorium reilianum* infects both maize and sorghum. **A**, field production of healthy maize (left) and sorghum (right) (source: internet). **B**, infected maize inflorescence (male and female) is turning into black mass of sori (left) and head smut of sorghum panicle (right).

1.3 The Life cycle of *S. reilianum*

S. reilianum is a facultative biotrophic fungus that can live saprotrophically as haploid yeast cells (sporidia) but depends on colonization of a host plant for sexual reproduction (**Figure 2**). Accordingly, plant infection is preceded by a mating reaction, during which two compatible sporidia with different mating type recognize each other, grow conjugation tubes that fuse, and form a dikaryotic filament [5]. The dikaryotic filament is able to penetrate the plant surface and colonize and multiply in different plant tissues. The hyphae accumulate in the bundle sheath cells and then follow the vascular system towards the shoot apical meristem of the plant [6-8]. At flowering time, emerging inflorescences are heavily colonized and partially or completely replaced by fungal sori, in which fungal nuclear fusion and formation of masses of melanized diploid teliospores takes place [9]. Sometimes no sorus is present and instead, the

inflorescences have a structure known as phyllody, showing a leafy-like morphology. The spores can be dispersed by several vectors, such as water and wind, and rest in the soil where they serve as inoculum for new plants. When spores germinate; they undergo meiosis and form haploid sporidia of four different mating types.
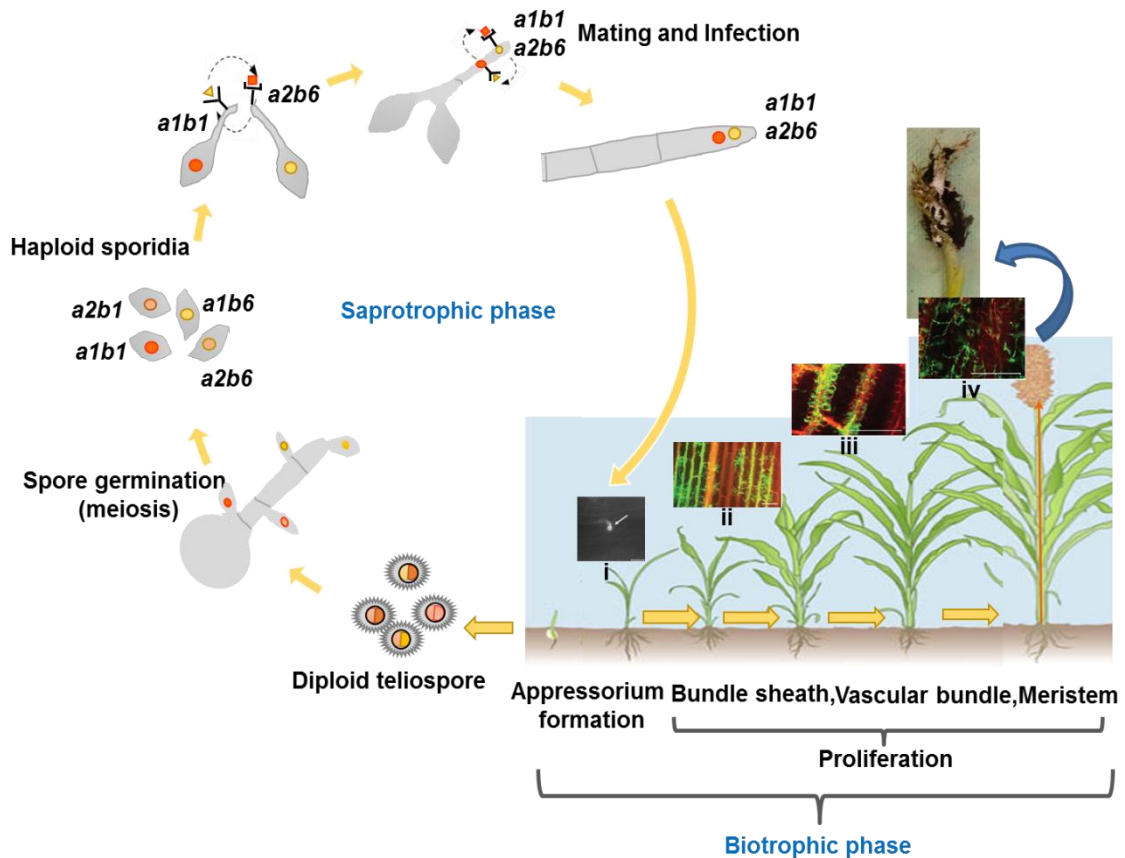
Figure 2:   Life cycle of *Sporisorium reilianum*. The teliospores of *S. reilianum* formed in the inflorescences of sorghum germinate under favorable conditions to form sporidia of different mating types. Two haploid sporidia with compatible mating types come near to each other and form conjugation tubes that lead to dikaryotic filament formation. This filament penetrates the plant surface through the formation of an appressorium (i). The fungus then proliferates through bundle sheath cells (ii), vascular bundles (iii) and nodes reaching the apical meristem (iv). During flowering time, when the inflorescence emerges, it is completely filled with spores. (Source of ii, iii and iv: [8]).

1.4 Successful mating is a pre-requisite for plant infection

The mating type of *S. reilianum* is determined by two different mating type loci, *a* and *b*, that freely segregate because they are located on different chromosomes [5,7]. The *a* locus encodes a pheromone-receptor system that allows partner recognition, while the *b* locus encodes a pair of homeodomain proteins that dimerize to form an active transcription factor only when the subunits are encoded on different alleles. The active transcription factor is necessary to induce transcription of genes essential for plant colonization. For mating and plant infection to occur, haploid sporidia need to be different in both *a* and *b* mating type loci.

1.5 *S. relianum* exists in two host-adapted varieties

S. reilianum exists in two *formae speciales* with different host preferences [10]. One of them, *S. reilianum* f. sp. *reilianum* (SRS), was isolated from sorghum and is highly virulent in this plant but does not produce spores on maize. *S. reilianum* f. sp. *zeae* (SRZ), was isolated from maize and does not cause disease on sorghum. SRS however can produce intermediate phyllody phenotype on maize without producing spores. Both *formae speciales* are able to penetrate and multiply in the leaves of both hosts [8]. In sorghum, the hyphae of SRS reach the apical meristems, whereas the hyphae of SRZ do not. SRZ strongly induces several defence responses in sorghum, such as the generation of H_2O_2, callose and phytoalexins, whereas the hyphae of SRS do not [8,10]. In maize, both SRS and SRZ are able to spread through the plant to the apical meristem. SRZ succumbs to plant defense after sorghum penetration, whereas SRS proliferates in a relatively undisturbed manner, but non-efficiently, on maize [8].

1.6 The concept of host specificity is intriguing

This part of the introduction partially belongs to our recently published review [11]. Plant pathologists have always been intrigued by the concept of host specificity. The questions that have shaped and will shape research in plant-fungus interaction are- why can some fungal plant pathogens cause disease only in one specific host and not in others; and how can a fungal pathogen of a particular plant adapt and switch to a new host and thereby become a new pathogen for the new host. The term host specificity refers to the capability of some fungal species or some members of one fungal species to cause disease only on particular plant species or only on

some members of a specific plant species. Molecular models have been developed that explain the basis of host specificity, like the gene-for-gene hypothesis developed by Henry H. Flor following his careful observations of the interactions of flax with flax rust [12]. According to this hypothesis, incompatible interactions result from the presence of resistance (R) proteins in the particular plants that recognize specific avirulence (AVR) proteins of the fungus, which results in successful plant defense against the pathogen. All interactions of plants carrying a particular R gene with pathogens carrying the corresponding AVR gene would be incompatible. This model has been refined by the guard and later the decoy models that predict that the AVR-R interaction is indirect, with the R protein guarding a target of the AVR protein, or guarding a decoy, a functionally inactive version of the target [12-14]. These models have led to a tremendous increase in the understanding of plant pathogen interactions. The second great advancement in this understanding is the discovery of effectors – small, secreted, variable, genome-encoded proteins of the pathogens – that are essential to modulate the plant response [13,15,16]. The realization that effectors can have virulence and/or avirulence (AVR) functions elegantly links the two concepts and explains the multitude of observed species-specific effector genes on the pathogen side as well as the multitude of receptor genes with the role of R proteins on the plant side, by an evolutionary arms race [16,17].

While we have a fairly good understanding of host-adaptation in some systems, in others the mechanism is still unknown. Even if the mechanism can be inferred, the identification of the participating factors is still quite challenging. For example, after knowledge that the *Ustilago hordei*-barley interaction is governed by AVR-R gene interactions, it took 32 years to clone an 80-kb region containing AVR1, and an additional 10 years to identify the gene encoding AVR1 [18,19]. Therefore, powerful methods are needed for the identification of host specificity factors to unravel the molecular basis of host adaptation of phytopathogenic fungi to their plant hosts. Plenty of methods have been developed that help in the identification of host specificity factors (**Figure 3**).
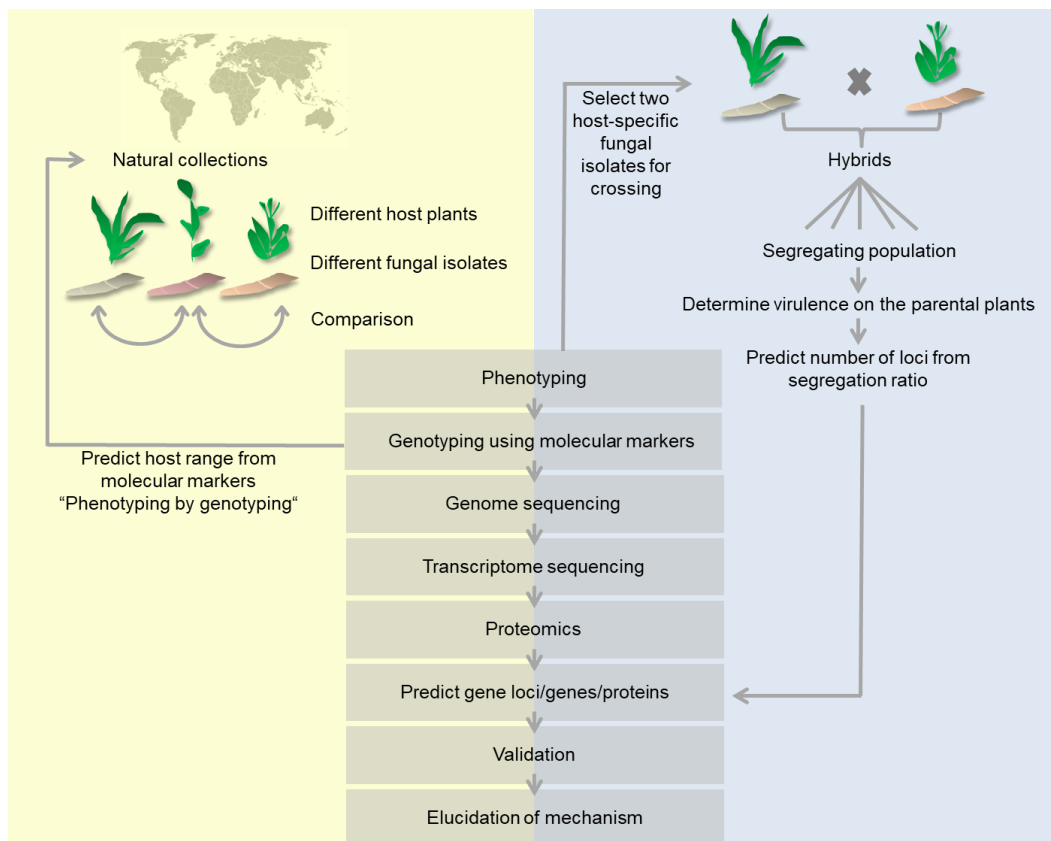
Figure 3: Graphical representation of comparative methods used to determine molecular host specificity determinants. Methods starting with a collection of natural fungal isolates (yellow background) are contrasted to methods starting with segregating populations of defined strains (grey background). Both starting materials use the same set of analyzing methods. (Source of figure:[11]).

While most molecular methods are excellent for describing molecular differences between host-selective strains, few are suited for rapid comparative identification of genes involved in host specificity. With the development of the next-generation sequencing technologies, at least the relatively rapid prediction of potential target genes now becomes possible. However, prediction alone is not enough. Without validation of functional involvement in host specificity, the underlying functional mechanisms cannot be resolved. Resolving functional mechanisms is still a slow and challenging task, involving rigorous scientific work and extensive experimentation. In most cases, in spite of great progress in the description of molecular differences of host-specific strains and in the prediction of genes possibly involved in host selection, we are still far from understanding how host-specific infection is achieved.

1.6.1 <u>A well-defined pathosystem is the basis</u>

Any comparative molecular analysis is dependent on a well-characterized pathosystem. It is therefore not astounding that many studies collected and phenotypically characterized the available biodiversity of natural fungal isolates. Fungal isolates are collected from different parts of the world e.g. from fields of host plants that are infected by the fungus [20,21], or procured from various laboratories and culture collections [22]. Depending on the necessity and the type of infection, fungal specimens can be isolated from infected leaves, stems or inflorescences [23]. After collection, the natural isolates are phenotypically characterized either by determining infection capability on different host plants coupled to microscopic analysis and analysis of the disease phenotype [24-27], or by the characterization of physiological traits, like their ability to produce or degrade various chemicals [28-30]. This way, various pathosystems were defined, like the *formae speciales* of *Fusarium oxysporum*, of *Puccinia*, and of *Alternaria*.

A well-defined pathosystem is the basis of any follow-up comparative analysis. From this pool of data, two different approaches are followed. In one approach, two isolates with different host specificity are crossed or hybridized, generating a segregating population with differential capabilities to infect the one or the other host plant. This segregating population is then analyzed to define genomic loci associated with differential infection capabilities. In the other approach, the complete pool of natural isolates is used for phenotypic and molecular analyses to link genomic markers to phenotypic traits (**Figure 3**). The approach using segregating populations generated by hybridization of host-specific isolates will be treated in a separate section below. In most examples, however, the approach using natural populations was followed.

1.6.2 <u>Molecular markers have an important role in genotyping</u>

To link genomic loci or markers to the host infection specificity phenotype, many different methods were developed that allow a molecular comparative characterization of the different isolates. These methods include analysis of restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD), micro- and minisatellites, mitochondrial haplotype, internal transcribed spacer (ITS) regions, as well as of complete proteomes and genome sequences.

A genome-wide and less biased comparative approach is the use of amplified fragment length polymorphism (AFLP) [31]. In this technique, the genomic DNA is first digested with

two different restriction enzymes, a frequent cutter and a rare cutter. Then specific adaptors are ligated that allow the PCR-amplification of restriction fragments. The number of amplified restriction fragments is reduced by extending the primers by 1 or 2 nt into the restriction fragment. To further reduce the number of amplified restriction fragments, a second PCR with labeled primers follows that extend by 3 to 4 nt into the restriction fragment. In this PCR, hybrid fragments with one rare-cutter adaptor and one frequent-cutter adaptor are preferentially amplified because two frequent-cutter adaptor primers form secondary structures that hamper product amplification. Primers with different extensions are tested to obtain about 50 to 120 amplified fragments, whose template-dependent lengths can then be compared by gel electrophoresis and label detection. AFLP has been successfully used to differentiate host-specific fungal isolates such as *P. infestans* isolates collected from cultivated potatoes and the native wild Solanum spp., *S. demissum* and *S. xendinense* in the Toluca Valley of central Mexico and *Magnaporthe grisea* isolates associated primarily from perennial ryegrass and kikuyugrass in golf courses in California [32,33]. Host-specific groups could also be defined according to the AFLP pattern of various endophytic fungi following initial morphological identification [34]. In *P. infestans*, AFLP was used to determine host specificity, race composition, and race variation [23]. *B. graminis* has been classified into eight *formae speciales* with a diverse host. rDNA ITS region and β-tubulin gene-based phylogenetic analysis showed grouping of isolates according to their principal host genus, which was further supported by AFLP analysis [35]. AFLP data can also be used for the generation of phylogenetic analysis, revealing the origin of host-specific taxa [22].

A special form of AFLP is the investigation of the mitochondrial DNA (mtDNA) haplotype. In this technique, certain regions in mitochondrial genomes are first PCR-amplified and then subjected to restriction enzyme digestion. mtDNA haplotype analysis has been successfully used to characterize different isolates of *P. infestans* [33,36]. The result can reveal whether the population structure is of monomorphic or polymorphic nature. The mtDNA haplotype has also been investigated in various geographical isolates of *Colletotrichum orbiculare* having a common host and in various other species of Colletotrichum [37]. The data helped to link host specificity with the haplotype pattern, thereby suggesting a way to recognize different host-specific isolates.

Random amplified polymorphic DNA (RAPD) is a PCR-based method that does not rely on restriction enzymes. A 10mer primer of arbitrary sequence is used to amplify random segments of genomic DNA. This method has been widely used to identify and isolate molecular markers specific to a particular fungus, thereby helping in diagnosis of fungal infection in symp-

tomless plants [38,39]. RAPD was used to characterize the genetic differentiation and correlation with host specificity among *Alternaria* spp. that cause brown spots on different Citrus spp. [40,41]. In *F. oxysporum*, a robust RAPD protocol was developed to identify economically important strains infecting specific hosts [42]. The RAPD technique was also used to discriminate between isolates of different host-specialized *Rynchosporium* species [43]. In *U. hordei*, RAPD was used in combination with microsatellites to assess genetic variation among different isolates in Tibetan areas of China [44]. In addition, RAPD has also been used as a tool to differentiate *formae speciales* of *Microbotryum violaceum* and to characterize the genetic diversity of the Italian population of *Ceratocystis fimbriata* f. sp. *platani* [45,46]. Although RAPD is a more unbiased technique covering a larger part of the fungal genome, the method has issues with reproducibility. Therefore, it is usually only used in combination with other markers, like micro- or mini-satellites.

Satellite DNA was first observed as a less-dense band of DNA clearly separated from the bulk of chromosomal DNA during density-gradient centrifugation. Microsatellites or simple sequence repeats (SSR), and their longer cousins the minisatellites consist of AT-rich repetitive DNA that seems to have a higher mutation rate resulting in a larger genetic diversity than other genomic regions. This genetic diversity can be used to discriminate different fungal isolates by PCR-amplifying the repetitive DNA [47]. Microsatellites have been found to be more informative for genotyping isolates from different hosts of the necrotrophic fungus *Botrytis cinerea* than RFLP patterns of the ADP-ATP translocase and nitrate reductase genes or MSB2 minisatellite sequence data [48]. In a study of the anther smut fungus *M. violaceum*, genetic diversity in sympatric, parapatric and allopatric populations of two host species was found using four polymorphic microsatellite regions [49]. In *Pyrenophora semeniperda*, that was shown to lack host specialization, weak yet significant population genetic structures as a function of host species could be observed by the use of seven polymorphic microsatellite loci [50]. Thus, the investigation of the diversity in microsatellite loci is a very sensitive tool to discriminate between different fungal isolates and to visualize even weak associations. However, microsatellite diversity has not yet been shown to be causally related to host specificity.

Other repetitive elements like retrotransposons or transposable elements have also been used for the characterization of host-specific fungal isolates. When using the reverse transcriptase gene of the LTR-retrotransposon CfT-l from *Cladosporium fulvum* as a probe, different *formae speciales* of *F. oxysporum* could clearly be differentiated [51]. PCR-detection of two transposable elements revealed different population structures of *B. cinerea* on a variety of different host plants [52].

It seems that a large number of different techniques have been developed that allow for molecular comparison of different organisms. Of the described techniques, AFLP and microsatellites seem to be the most sensitive methods to molecularly discriminate fungal isolates for which no or only very little sequence information is available. Most studies using these methods did not come up with a clear genomic link to host specificity of the investigated fungal strains. This means that other, even more precise techniques are needed to decipher the molecular basis of host specificity. Here, the field profited tremendously from the development of the next-generation sequencing (NGS) techniques that now allow cost-effective genome sequencing and data assembly of even large fungal genomes. Since their availability, the use of -omics (genomics, transcriptomics, and proteomics) approaches highly outnumber the use of the classical comparative approaches described above.

1.6.3   <u>Whole genome sequence comparison has vast potential</u>

One of the first eukaryotic genomes sequenced using next-generation sequencing techniques was that of *Sporisorium reilianum* f. sp. *zeae*, a close relative of *Ustilago maydis*. Both cause smut disease of maize but induce quite distinct symptoms. Genome comparison revealed that both genomes were highly syntenic but contained so called "divergence clusters" containing genes with below-average sequence conservation between the two organisms [7,53]. Within these divergence clusters, a high percentage of genes encoded proteins containing predicted secretion signal peptides and were thought to be involved in interaction with the plant, thereby explaining their increased evolution rate. Deletion analysis of complete cluster regions in *U. maydis* confirmed the role in virulence for four of six randomly selected clusters [7,53]. Genome sequencing of the related barley smut pathogen *U. hordei* allowed comparison with an organism showing a similar infection strategy as *S. reilianum* and being virulent on a different host plant. The three-way genome comparison revealed among others that most of the weakly conserved genes present in the *S. reilianum/U. maydis* divergence clusters have weakly conserved homologs in *U. hordei* [7,53], which supports the proposition that these proteins could play a function in adaptation to different hosts or lifestyles. Sequencing of the *Sporisorium scitamineum* genome allowed a four-genome comparison of effector genes [54]. This comparison revealed that evolution of effector-encoding clusters is driven by tandem gene duplication and the activity of transposable elements, and supports the drawn conclusions resulting from analysis of the *U. hordei* genome [7,54]. Thus, genome comparison of smut fungi so far resulted in lists of genes potentially encoding host specificity factors.

Interesting gene candidates with a suspected role in host specificity were also obtained when comparing the genomes of the closely related species *C. graminicola* and *C. sublineola* [55]. The main differences of the otherwise very similar genomes were found in genes for biosynthesis of specialized secondary metabolites and of small -secreted protein effectors. However, whether these genes indeed contribute to host selection has yet to be tested. Key enzymes of fungal secondary metabolism and effector proteins were identified as potential host specificity factors by genome comparison of *Rhynchosporium* species [43], and of members of the *Fusarium fujikuroi* species complex, where species-specific and isolate-specific differentiation in secondary metabolite- producing genes both in composition and expression were detected [56]. The comparison of whole genome sequences of four different strains of the *C. acutatum* species complex showed that changes in gene content were related to changes in host range with lineage-specific gene losses and gene family expansions [57]. Gene loss was also suggested as a cause of fungal adaptation to a new dicot host after a host jump from a monocot plant by the smut fungus *Melanopsichium pennsylvanicum*. When compared to the genomes of three other smut fungi, *M. pennsylvanicum* was found to lack putative effector genes [58]. In addition to putative effectors, comparative whole genome and transcriptome analyses of *Lasiodiplodia theobromae and* five other Botryosphaeriaceae pathogens causing opportunistic infections in woody plants identified two in-planta expressed lignocellulose genes, whose overexpression increased virulence of the pathogens [59].

In a genome comparison study of the three species *Fusarium graminearum*, *Fusarium verticillioides* and *F. oxysporum* f. sp. *lycopersici* (Fol) lineage specific (LS) genomic regions and LS chromosomes were discovered [60]. The authors of the study could experimentally prove that the presence of Fol LS chromosome 14 provides specificity for Fusarium adaptation towards tomato, limiting the search for host specificity factors to a single chromosome. This information was used in a recent study where the genomes of three legume-infecting *formae speciales* of *F. oxysporum* were compared to the genomes of the tomato-infecting Fol and the pea-infecting *Fusarium solani*. Combining comparative genome analysis with predicted LS gene content and in-planta transcription analysis revealed four candidate effectors conserved among legume-infecting *formae speciales* [61]. Clustering of presence/absence patterns of candidate effector genes following whole genome sequencing of five different *formae speciales* of *F. oxysporum* showed clustering of members of the same forma specialis [62] suggesting that effectors contribute to host specificity. This confirms and extends earlier reports of an association of the three secreted in xylem (SIX) effectors SIX1, SIX2 and SIX3 with tomato-infecting

isolates, and of the suitability of the presence/absence determination of SIX1 to SIX5 as a robust method to differentiate different *formae speciales* [63].

For the identification of effectors, sequencing of RNA or cloning of expressed sequence tags (ESTs) can be very helpful. Sequencing of about 2000 cloned ESTs from C. lentis-infected lentil leaf tissues enabled annotation of 15 candidate effectors. Infection stage-specific gene expression was observed for the candidate effectors. One candidate effector, CICE6, was found to carry a single nucleotide polymorphism (SNP) that could be used efficiently to differentiate between two pathogenic races of *C. lentis* [64]. Comparison between transcriptional profiles of two races of *F. oxysporum* f. sp. *cubense* revealed a remarkably different gene expression profile in response to host cell wall [59] suggesting that differences in gene expression could contribute to host specificity.

Without being exhaustive, this enumeration already shows that the field profited tremendously from the development of next-generation sequencing techniques. Genome comparison that is sometimes coupled to a comparison of the transcriptional profile, often resulted in the identification of genomic regions, and sometimes even of genes, that are associated with adaptation to a specific host and are therefore strong candidates for the determination of host specificity.

1.6.4   Comparing complete proteomes is another well used method

An alternative approach to genome comparison and transcriptome analysis for the identification of proteins involved in host specificity determination is the direct comparative analysis of the fungal proteomes. Several studies compared the proteomes of different isolates or species with the idea to identify proteins crucial for host specificity. Tandem mass spectrometry was used as a tool to compare the proteomes of hyphae and germinating cysts of two closely related oomycetes, *P. pisi* and *P. sojae*, that cause disease on pea and soybean, respectively [65]. A global proteomic comparison of mycelium and germinating cysts was done in two other oomycete plant pathogens, *P. ramorum* and *P. sojae* [66]. A proteome comparison was also done with uredospores from two different populations of the rust fungus *Puccinia psidii* isolated from eucalyptus leaves and guava fruits [67]. Mycelial proteins from isolates of the brown rot fungus *Monilinia laxa* were obtained from apples and apricots, and were separated by 2-D gel electrophoresis, followed by LC-MS/MS of identified differentially expressed proteins [68]. In all of these studies, clear differences in the fungal proteomes were observed, and detected species- or

population-specific proteins were suspected to have a role in host adaptation. However, lists were long, and a causal connection of identified proteins to host specificity has not yet been shown.

Global comparative approaches have brought the scientific community much closer to the goal of identifying host specificity factors by providing lists with genes or proteins that have a high probability of being involved in host adaptation. The follow-up of these approaches would now consist of testing individual high-probability candidates for their specific contribution to host selection. This can be challenging, especially if genetic tools for the investigated system are not yet available, or if lists contain several dozens of putative candidates. Success of the follow-up experiments will highly depend on the predictive quality of these lists, which will increase with closer relatedness of the compared pathogen genomes.

1.6.5   Investigating segregating population as a mean to identify host specific regions

One way to overcome the problem of having to test candidates from long lists of putative host specificity factors that have a certain possibility of being involved in host adaptation is to investigate a well-defined population that segregates for the host infection phenotype. The easiest method to obtain such a well-defined population is to generate it by targeted hybridization experiments or genetic crosses of closely related fungal isolates that differ in their infection phenotype.

Hybrids have been investigated before the knowledge of genome sequences. Mating compatible isolates of *Colletotrichum gloeosporioides* with different host specificities were crossed and ascospore offspring were analyzed for virulence and RFLP patterns. However, no correlation between pathogenicity to the parental hosts and presence/absence of any RFLP marker could be found [69]. In a comparative study of various host-specific Ascochyta species, interspecific crosses could be obtained, and offspring analysis showed that AFLP markers segregated freely. While not being able to assign specific AFLP markers to host infection capacity, it was speculated that host specificity may contribute to speciation [70]. F1 progeny of a cross of two *Leptosphaeria maculans i*solates that differed in their capacity to cause disease on *Brassica juncea*, were used to create a genetic map of different AFLP and RAPD markers as well as the mating type and a host specificity locus that could be placed on the end of chromosome 9 [71]. Hybrids of *Magnaporthe oryzae* isolates from rice and wheat were generated and the re-

sultant F1 population was tested on wheat for pathogenicity. The segregation ratio of avirulent to virulent offspring provides information of the number of loci involved. In this case, a ratio of 7:1 led to the prediction of three loci being involved in avirulence of the *M. oryzae* rice isolate on wheat. Allelism tests could associate two known loci, Pwt2 to papilla formation and Pwt1 for hypersensitive reaction. The third locus did not correlate with any known loci and was therefore named Pwt5 [72]. In a F1 population of a cross between wheat- and foxtail millet-pathogenic *Magnaporthe grisea*, virulent isolates segregated in a 1:1 ratio on foxtail millet cultivars Beni-awa and Oke-awa but not on cultivar Kariwano-zairai, suggesting that specificity of *M. grisea* toward foxtail millet is governed by cultivar-dependent genetic mechanisms like gene-for-gene interactions [73].

Since these studies were all conducted without the knowledge of the genome sequences, it is conceivable that the combination of genome sequencing and the investigation of defined segregating populations will either allow much faster genome map-based cloning or, through the sequencing of complete populations, allow high resolution single nucleotide polymorphism (SNP) mapping of host specificity traits. Creating genetic linkage maps in fungi is an underdeveloped but powerful field that is expected to gain future attention in the light of whole genome sequencing [74]. In this light we have started to analyze hybrids of two compatible *formae speciales* of the head smut fungus *S. reilianum*, *S. reilianum* f. sp. *zeae* (SRZ) and *S. reilianum* f. sp. *reilianum* (SRS) that either infect maize (SRZ) or sorghum (SRS). On each other's hosts, both fungi can colonize but do not cause smut disease [8]. A population of meiotic progeny (SRSZ) of a mating event between SRS and SRZ was generated and of each individual strain the virulence potential on sorghum was tested that varied greatly between individuals. To associate the virulence phenotype on sorghum to particular genomic-regions, about 189 strains were selected for genome sequencing. Mapping of sequencing reads to the parental genomes confirmed the presence of mosaic genomes in the offspring. A detailed genome analysis will show whether specific parental genomic regions can be associated with the virulence phenotype on sorghum (**Figure 4**). As soon as such an associated region is known, knowledge of the genome sequence will allow direct prediction of candidate genes for validation of their role in determining host specificity.

1.6.6 Functional validation of factors contributing to host specificity

Generating lists of high potential candidates for an involvement in host specificity determination is already a great step forward. However, the goal is to show that a given candidate has indeed a role in host specificity. Validation of a functional involvement of a specific candidate gene could be done by the generation of gene deletion or overexpression strains and monitoring of the mutant's host preference. As mentioned above, this can be a challenging task. In a few cases, functional validation was done and showed a clear relationship of the identified genes with host specificity.
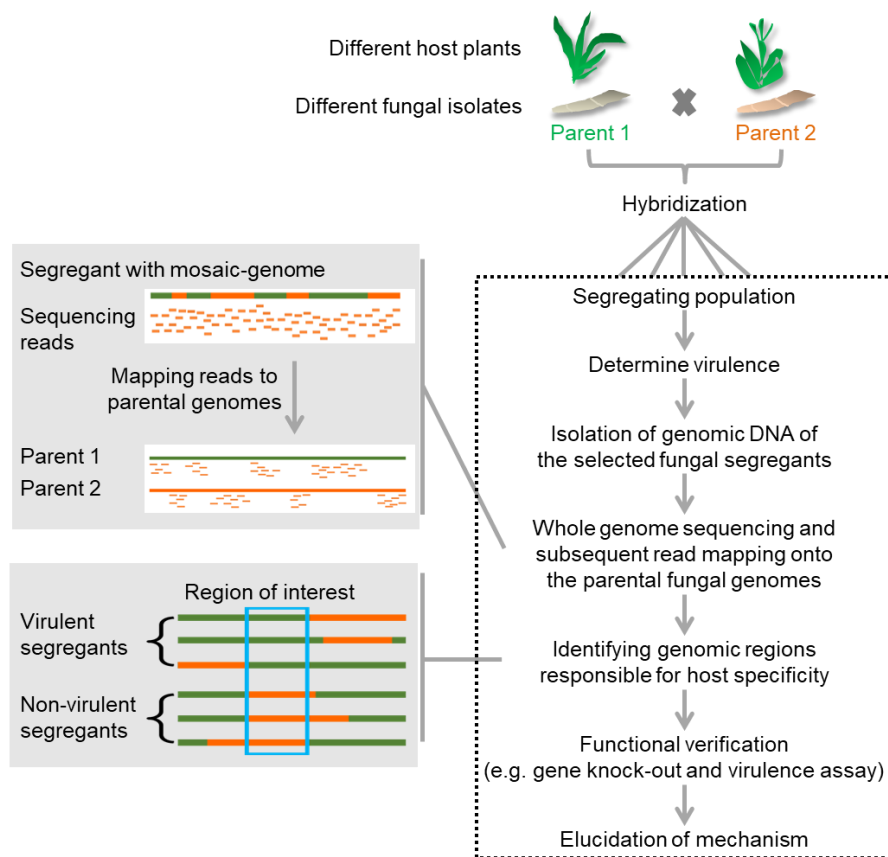
Figure 4: Strategy for identification of host specificity factors using segregating populations of a hybridization event between two host-specific individuals. Virulence capacity of each offspring on one or both hosts is individually determined. Genome sequencing of fully virulent and avirulent offspring reveals parental origin of the mosaic genomes. Associating the parental

origin of specific genomic loci to the virulence phenotype should lead to identification of genomic regions linked to host-specific virulence. (Source of figure:[11]).

In many cases, genes involved in host specificity were identified as a part of classical Avr-R gene interactions. For example, using crosses of *M. grisea* strains infecting different grass species, single genes that determine specificity towards the host weeping lovegrass (Eragrostis curvula) have been identified [75] and were shown to be part of a gene family. The inability to infect weeping lovegrass could be associated with one member of the gene family, PWL2, where frequently occurring loss-of-function mutations of PWL2 led to spontaneous pathogenic mutants [76]. Thus, PWL2 has all the characteristics of a classical avirulence gene. AVR1-CO39 of *M. oryzae* is another avirulence gene involved in host specificity. The transfer of AVR1-CO39 to a rice-pathogenic isolate resulted in transformants unable to cause disease on the rice cultivar CO39, while the rice cultivar 51583 that lacks the resistance gene Pi-CO39(t) could still be infected [77].

In other cases, lacking complementation of deleted virulence genes by orthologous genes from related species suggested a role in host-specific virulence. For example, the SIX1 gene of *F. oxysporum* f. sp. *lycopersici* (Fol) was found to be necessary for full virulence of Fol on susceptible tomato but was also recognized by the I-3 resistance gene, leading to avirulence on I-3 tomato lines [78]. A homolog of SIX1 was found to be responsible for virulence of *F. oxysporum* f. sp. *conglutinans* (Foc) on cabbage [99]. Interestingly, virulence of the Foc-SIX1 deletion mutant on cabbage could be restored by reintroduction of SIX1 of Foc but not of Fol. This suggested a host-specific virulence role for Fol-SIX1 [71]. Similarly, virulence of mutants of the wheat pathogen *Zymoseptoria tritici* carrying deletions in the Zt80707 or the Zt89160 virulence gene, could be complemented with the respective genes from *Z. tritici*, but not with the respective orthologs from *Z. pseudo tritici* or *Z. ardabiliae* suggesting that the identified genes are involved in host-specific disease development.

In fungi where host specificity is governed by host selective toxins, validation was done by heterologous expression of the toxin biosynthesis gene in a related species. For example, in the strawberry pathotype of *A. alternata*, transformation-mediated loss of the 1.05-Mb conditionally dispensable chromosome encoding all known toxin biosynthesis genes led to non-pathogenicity [79], which validated the role of the dispensable chromosome in host-specific pathogenicity. In *Alternaria citri*, mutation of a gene encoding an endo poly galacturonase led to a reduction in virulence, while virulence was unchanged in mutants of *A. alternata* rough lemon

pathotype lacking the same gene [80]. This showed that a cell-wall degrading enzyme contributes to virulence of one but not of the other pathotype. In contrast, PbToxB of the bromegrass pathogen *Pyrenophora bromi*, a homolog of the known host selective toxin PtrToxB from the wheat pathogen *Pyrenophora tritici-repentis*, was unexpectedly shown not to be toxic on bromegrass but on wheat, indicating that *P. bromi* has the potential to become a wheat pathogen [81].

These few examples and the surprises they contained illustrate the importance of validation experiments for genes suspected to have a role in host specificity. Without knowledge of their contribution to host-specific virulence, it is impossible to generate meaningful hypotheses about potential mechanisms.

1.6.7   Functional mechanisms involved in host specificity

Knowing the mechanisms that govern host-specific virulence capacities is the ultimate goal in plant-pathogen interaction research. The gained understanding allows generating novel plant protection strategies and a more-and-more reliable prediction of the danger of a particular fungus to be involved in future disease outbreaks in novel hosts. In a few cases, more than just the virulence/avirulence-causing gene of the pathogen is known, which allows a better understanding of the basic mechanisms involved in host specific interactions.

In *P. infestans*, virulence towards potato was shown to depend on the function of the RXLR effector AVR3a in inhibiting enzyme activity of the host ubiquitin proteasome system (UPS) [82]. When studying host adaptation of *P. infestans* and its *Mirabilis jalapa*-infecting sister species *P. mirabilis*, a single amino-acid polymorphism was identified in the host protease and a corresponding single amino-acid change in the pathogen effector as being responsible for virulence of their respective host plants, explaining ecological diversification. In case of *A. alternata*, the contribution of toxins to host-specific virulence success has been thoroughly proven. This does unfortunately not yet explain, why a particular toxin would allow virulence only on a particular host plant. For the ACR toxin necessary for virulence of the *A. alternata* rough lemon pathotype, part of the functional mechanism was unraveled. Toxicity on rough citrus (*Citrus jambhiri* Lush.) was found to depend on differential post-transcriptional processing of transcripts of the mitochondrial ACRS (ACR-toxin sensitivity) gene, which is present in both toxin-sensitive and toxin-insensitive citrus but processed to shorter transcripts in mitochondria of insensitive plants [83]. This work showed that host specificity of the rough lemon pathotype of *A. alternata* towards its host is due to altered mitochondrial RNA processing. In *Alternaria*

brassicicola, that causes black spot on Brassica plants, the essential contribution of the AB-toxin to virulence is well known. It was shown that a host-derived factor, an oligosaccharide of 1.3 kDa, secreted from the plant just after *A. brassicicola* spore germination was necessary to induce AB-toxin production [84] showing the involvement of a host-derived factor in production of host-selective toxins.

In spite of significant progress, the battle for understanding principles of host selection in fungal plant-pathogen interaction is not yet won. A variety of comparative methods is available to help in the identification of host specificity factors. Intelligent combination of classical genetics and next-generation sequencing under consideration of gene expression changes of both the pathogen and the host may be needed to unravel the mechanistic basis of host specificity of plant pathogenic fungi.

1.7 <u>Aim of this study</u>

The objective of this study is to identify host specificity determinants of two *S. reilianum* varieties- SRS and SRZ. The approach is by using classical genetics, next generation sequencing and whole genome analysis for the identification of fungal factors contributing to host preference of the two *formae speciales* of *S. reilianum* that infect either sorghum or maize. At the beginning, available *S. reilianum a1b1*-mapping population should be enlarged. This should be followed by identifying strongly virulent and non-virulent offspring by virulence analysis on sorghum, also subsequent virulence analysis on maize. Two other measures of phenotype, phytoalexin responses by sorghum plant and colony morphology of the offspring, should be evaluated. The re-sequencing of the genomes of the selected offspring using Illumina NextSeq should be performed. Analyzing parental origin of all chromosomal regions in the virulent and non-virulent offspring by re-sequencing and identification of parental genomic region needed to be done. Identifying the chromosomal regions linked to non-virulence or virulence should give prediction on region of interest. The approach allows precise and unbiased chromosomal mapping of host specificity candidates. Candidate mapping should be followed by functional analysis of selected candidate genes through virulence analysis of deletion strains.

2 Materials and Methods

2.1 Materials

2.1.1   Sorghum and maize plants

Seeds of healthy flowering cultivar *Sorghum bicolor* var Tall Polish were obtained from Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben. Seeds of *Zea mays* cultivar Gaspe Flint were obtained from Prof. Regine Kahmann, Max Planck Institute for Ter-restrial Microbiology, Marburg. Maize and sorghum seeds were grown, respectively, for 7 and 14 days, in greenhouse conditions: day conditions- 28°C, 40% relative humidity and minimum 28000 Lux light with additional 90000 Lux sun radiation for 15 h and night conditions: 20°C and 60% relative humidity for 9 h.

2.1.2   Fungal strains

The compatible wild-type *Sporisorium reilianum* strains SRZ1_5-2 (*a1b1*) and SRZ2_5-1 (*a2b2*), originally isolated from maize [7], and SRS1_H2-8 (*a1b1*) and SRS2_H2-7 (*a2b6*), isolated from sorghum [10], were used in this study. The strains were maintained at -80°C in NSY glycerol medium (8 g/l nutrient broth, 10 g/l yeast extract, 5 g/l sucrose, 696 ml/l glycerol) or on potato dextrose agar plates at 4°C for no longer than 1 week. All the offspring SRSZ strains used in this study have been selected after a mating test against tester strains - SRZ1_5-2, SRZ2_5-1, SRS2_H2-7 and SRS1_H2-8. These offspring have the same progenitor strains - SRS2_H2-7 and SRZ1_5-2, have the same mating type (- *a1b1*) and they are listed below (table 1).

Table 1:   List of the generated SRSZ strains with mating type *a1b1* used for plant infection.

SRSZ strain	SRSZ strain	SRSZ strain	SRSZ strain	SRSZ strain	SRSZ strain	SRSZ strain
TW74	TW156	TW242	TW301	TW356	TW423	TW486

TW75	TW157	TW243	TW302	TW358	TW424	TW487
TW76	TW159	TW244	TW303	TW359	TW425	TW488
TW77	TW160	TW245	TW305	TW360	TW426	TW489
TW78	TW161	TW246	TW305	TW361	TW427	TW490
TW79	TW162	TW247	TW306	TW362	TW428	TW491
TW80	TW163	TW248	TW307	TW363	TW430	TW492
TW81	TW164	TW249	TW308	TW364	TW431	TW493
TW83	TW165	TW250	TW308	TW366	TW432	TW495
TW84	TW166	TW251	TW309	TW368	TW433	TW496
TW85	TW167	TW252	TW309	TW369	TW434	TW497
TW86	TW168	TW253	TW310	TW371	TW435	TW499
TW87	TW169	TW254	TW311	TW372	TW436	TW500
TW88	TW170	TW255	TW312	TW375	TW437	TW501
TW89	TW171	TW256	TW313	TW376	TW438	TW502
TW91	TW173	TW257	TW314	TW378	TW439	TW503
TW97	TW174	TW258	TW314	TW379	TW441	TW505
TW98	TW175	TW259	TW315	TW380	TW442	TW506
TW99	TW177	TW260	TW315	TW381	TW444	TW507
TW101	TW179	TW261	TW316	TW382	TW445	TW508
TW102	TW180	TW262	TW317	TW383	TW446	TW510
TW103	TW181	TW263	TW318	TW384	TW447	TW511
TW106	TW182	TW264	TW319	TW385	TW448	TW512
TW107	TW193	TW265	TW319	TW386	TW449	TW513
TW110	TW194	TW266	TW320	TW387	TW450	TW514
TW112	TW195	TW268	TW322	TW388	TW451	TW516
TW113	TW196	TW269	TW322	TW389	TW452	TW517
TW114	TW200	TW271	TW323	TW390	TW453	TW518
TW118	TW201	TW272	TW323	TW391	TW455	TW520
TW124	TW204	TW273	TW324	TW392	TW457	TW521
TW125	TW207	TW274	TW326	TW393	TW458	TW522
TW126	TW208	TW275	TW327	TW395	TW459	TW524
TW127	TW209	TW276	TW328	TW396	TW460	TW526
TW128	TW210	TW277	TW329	TW397	TW461	TW528
TW129	TW211	TW278	TW330	TW398	TW462	TW529
TW130	TW212	TW279	TW331	TW399	TW463	TW530
TW131	TW215	TW280	TW332	TW400	TW464	TW531
TW132	TW216	TW281	TW333	TW401	TW465	TW532
TW133	TW221	TW282	TW334	TW402	TW466	TW533
TW134	TW222	TW283	TW335	TW403	TW467	TW534
TW135	TW224	TW284	TW336	TW404	TW469	TW535
TW136	TW225	TW285	TW337	TW406	TW470	TW536
TW137	TW226	TW286	TW338	TW407	TW471	TW537
TW138	TW227	TW287	TW339	TW408	TW472	TW538

TW139	TW228	TW288	TW340	TW409	TW473	TW539
TW141	TW229	TW289	TW341	TW410	TW474	TW540
TW142	TW230	TW290	TW342	TW412	TW475	TW542
TW143	TW233	TW291	TW343	TW413	TW477	TW545
TW144	TW234	TW293	TW345	TW414	TW478	TW547
TW145	TW235	TW294	TW347	TW415	TW479	TW550
TW146	TW236	TW295	TW348	TW416	TW480	
TW147	TW237	TW296	TW349	TW417	TW481	
TW149	TW238	TW297	TW350	TW418	TW482	
TW150	TW239	TW298	TW351	TW419	TW483	
TW154	TW240	TW299	TW352	TW420	TW484	
TW155	TW241	TW300	TW353	TW421	TW485	

Cluster 7_11 deletion strains were generated as described in chapter 2.1.12 and are listed in Table 2.

Table 2:   List of generated *S. reilianum* cluster 7_11 deletion strains.

Strains	Genotype	Progenitor strains
*Δ7_11*H2-7a1	*a2b6 Δ7_11*	SRS2_H2-7
*Δ7_11*H2-7b1	*a2b6 Δ7_11*	SRS2_H2-7
*Δ7_11*H2-8a1	*a1b1 Δ7_11*	SRS1_H2-8

2.1.3   **Plasmid vector**

The plasmid vector used in this study is p*BS-hhn*. It has a 1.82 bp- hygromycin resistance cassette [85] that can be cut out using *Sfi*I restriction digest..

2.1.4   **Oligonucleotides**

The oligonucleotides designed and used are listed below:

Table 3:   List of all the oligonucleotide primers used in this study

Name	Sequence	Purpose

oNB001	GTATGCGAACCGGACCAAGC	For LF amplification of cluster 7_11
oNB002	ATCTAGGCCATCTAGGCCACGATCCGCA ACCGAAGACC	For LF amplification of cluster 7_11
oNNB003	TATAGGCCTGAGTGGCCTGCCTGCTCAA CTGTACC	For RF amplification of cluster 7_11
oNB004	ATCTGACCGCTCGTTTGC	For RF amplification of cluster 7_11
oNB005	TCAGCTGGACGTTCTATGG	For LF amplification of cluster 7_11
oNB006	ACGTATCTGGAGCGATACC	For RF amplification of cluster 7_11
oNB007	GCTCTCGGGATAGGTGTAAGC	From inside of 13897 gene for verification of cluster 7_11 presence
oNB008	GATGTAGCAGCTTGGCTTGG	From inside of 13906 gene for verification of cluster 7_11 presence
oNB009	CTCGCAACGCGGCTCATGTCATC	For LF amplification of cluster 7_11
oNB010	ATCTAGGCCATCTAGGCCAGTCGTCCGC CACGAAACTAC	For LF amplification of cluster 7_11
oNB011	TATAGGCCTGAGTGGCCCACGCCTCTCG CTCATCACAACAATC	For RF amplification of cluster 7_11
oNB012	TTCGAGCTCTACACGTGGTACATC	For RF amplification of cluster 7_11
oNB013	CAAGCTCAGCTGGACGTTCTATGG	For nested PCR to amplify deletion construct
oNB014	CTACGCTCGTGCTGTCGTATCTC	For nested PCR to amplify deletion construct
oNB015	TCTAGGCCTCACTGGCCCTCGCA	For LF amplification of cluster 7_11
oNB016	ACTCGGCCTCGCTGGCCTTCGAG	For RF amplification of cluster 7_11

oNB017	GCAGGGCACAGGTGAAGAAG	From inside the cluster 7_11 to differentiate SRS from SRZ
oNB018	GCAACGCGCAAGAGGCTTTC	From inside the cluster 7_11 to differentiate SRS from SRZ
oNB019	TCAGGGATGGGCCTTTGACG	To amplify gene 13901
oNB020	CACGGGAGCTCACTGGAATC	To amplify gene 13901
oNB021	AGCCTGTTGCGATATCTTGG	To amplify gene 13903
oNB022	GGTCATTCCCTTCCTGTTTC	To amplify gene 13903
oNB023	CATGAGCATGCCCTGCCCCTAGACCGGA AGAGAAGAGGGTG	For RF amplification of cluster 7_11
oNB024	CCACTTTGCGCCGTCCCAGCATACGATC CGCAACGAAGA	For LF amplification of cluster 7_11
oNB025	AATATTAAGCTTGGTACCGAGCTCGTCA GCTGGACGTTCTATGGGCTGGCGGTGAA	For LF amplification of cluster 7_11
oNB026	ACACTGGCGGCCGTTACTAGTGGATCAC GTATCTGGAAGCGATACCAGCAAGCCTT	For RF amplification of cluster 7_11
oNB027	ATCTCAGGGATGGGCCTTTGACG	For amplification of a region inside SRS cluster 7_11 for cloning into pJET1.2
oNB028	ATCATGAAGCTACACAAGGCCGTTCTCC	For amplification of a region inside SRS cluster 7_11 for cloning into pJET1.2
oNB029	CACCTCCGAATCGGAAGCCAAAC	To verify luster 7_11KO mutants
oVT022	TCCGAGGGCAAAGGAATAG	To verify cluster 7_11KO mutants
oVT023	AACTTCCAGCCACCCATTCC	To verify cluster 7_11 KO mutants

2.1.5 Enzymes

The enzymes used in this study are listed below (Table4).

Table 4: List of enzymes used in this study.

Enymes	Supplier
Lysozyme	Merk
Novozyme 234	Novo Nordisc
Restrictionenzymes	New England Biolabs GmbH
Phusion® High-Fidelity DNA Polymerase	New England Biolabs GmbH
T4-DNA-Ligase	Roche, Mannheim
Taq-Polymerase Fermentas (St. Leon-Rot)	Fermentas
RNAse A	Macherey-Nagel
Q5® High-Fidelity DNA Polymerase	New England Biolabs GmbH

2.1.6 Commercial kits

The commercial kits used in this study are listed below (Table5).

Table 5: List of all the kits used in this study.

Kit name	Supplier
NucleoSpin® Gel and PCR Clean-up	Macherey-Nagel
MinElute Gel Extraction Kit	Qiagen
DNeasy plant mini kit	Qiagen
NucleoSpin® Plasmid	Macherey-Nagel
Quick Ligation™ Kit	New England Biolabs GmbH

2.1.7 Media and reagents

The reagents used to prepare the media are listed below with the company names (Table6)

Table 6:   List of all the reagents used in this study.

Media	Components [concentration]	Supplier
YEPSL Medium Modified from (Tsukuda et al., 1988)	Tryptone (10 g/l) Yeast extract (10 g/l) Sucrose (3.10 g/l)	Roth Merck Roth
PD Medium	Potato Dextrose Broth (24 g/l) Agar (20 g/l) only for solid medium	Roth Applichem
Regeneration Medium	Tryptone (10 g/l) Yeast extract (10 g/l) Sucrose (10 g/l) Sorbitol (182g/l) Agar (20 g/l) only for solid medium	Roth Merck Roth Applichem
Water Agar Medium	Agar (10 g/l)	Applichem
LB medium	Tryptone (10 g/l) Yeast extract (5 ml/l) NaCl (10 g/l) Agar (20 g/l) only for solid medium pH 7.0	Roth Merck Roth Applichem
SCS solution	 Na-Citrate (20 mM) Sorbitol (1 M) The solution was filter sterilized	Fluka
STC Solution	 Tris-HCl (10 mM) , pH 7.5 Na-Citrate (20 mM) , pH 5.8 Sorbitol (1 M) The solution was filter sterilized	Roth Fluka
STC/PEG	STC (15 ml) PEG4000 (10 g)	 Roth

RFI Solution		Biochema
	RbCl (100 mM)	Merck
	MnCl$_2$ x 4 H$_2$O (50 mM)	Roth
	K-Ac (30 mM)	Roth
	CaCl$_2$ x 2 H$_2$O (10 mM)	Roth
	Glycerin in ddH$_2$O (15%)	
	pH 5.8 Filter sterilized	
RFII Solution		Biochema
	RbCl (10 mM)	Roth
	MOPS (10 mM)	Roth
	CaCl$_2$ x 2 H$_2$O (75 mM)	Roth
	Glycerin in ddH$_2$O (15%)	
	pH= 5.8 Filter sterilized	
TE-RNasebuffer	TE buffer (100 μl)	Machery-Nagel
	RNase A (20 mg/ml)	
1 M Tris-HCl pH 8		Roth
	Tris-HCl (563 mM)	Roth
	Tris-Base (437 mM)	
UstilagoLysis Buffer		Roth
	Tris-HCl (50 mM) , pH 7.5	Roth
	Na-EDTA (50 mM)	Roth
	SDS (1 %)	
TE-Phenol-Chloroform	Phenol (0.5 Vol)	Roth
	Chloroform (0.5 Vol)	Roth
	EqualibrateTE	
1X TAE Buffer		Roth
	Tris-Acetate (40 mM)	Roth
	Na-EDTA (1 mM)	
0.5X TBE Buffer		Roth
	Tris-Borate (50 mM) , pH 7.9	Roth
	Na-EDTA (1 mM)	
6X Loading Buffer	Sucrose (50%)	Roth
	Bromphenolblue (0.25%)	Merck
		Roth

		Roth
	HCl (10 mM) , pH 7.9 Na-EDTA (1 mM)	Roth
10X Phosphate Buffered Saline 10x (stock)	NaCl (180 g/l) KCl (2 g/l) Na$_2$HPO$_4$ (14.4 g/l) KH$_2$PO$_4$ (2.4 g/l) PBS (working solution)- Dilute 1:10 in ddH$_2$0	Roth Roth Roth Roth
Calcofluor (stock)	Fluorescent Brightener 28(Sigma F-3543) in DMSO (10 mg/ml) Keep in the dark Store at –20°C Calcofluor (staining solution)- Dilute stock 1:100 in 0.2 M Tris pH 7.8 Keep in the dark, Store at 4 °C	Sigma Aldrich
WGA-Alexa Fluor 488 Propidium Iodide (Stock)	Propidium iodide in PBS (10 mg/ml) Store at 4°C 1 mg/ml WGA-AF488 in H$_2$O Store at 4°C in the dark	Sigma Aldrich Sigma Aldrich
WGA-AF 488 (staining solution)	Propidium iodide (2 μg/ml) WGA(Wheat Germ Agglutinin)- AF(Alexa flour) (10 μg/ml) In 1x PBS (pH 7.8)	Sigma Aldrich
Aniline blue staining solution	Aniline blue in PBS (0.05 g/l) pH 8.2	Sigma Aldrich

2.2 Methods

2.2.1 Generation of an *a1b1* SRSZ mapping population

2.2.1.1 Isolation of independent SRSZ strains

The teliospores of the combined SRS H2-7 and SRZ 5-2 infected on *Sorghum bicolor* var Tall Polish were mixed with sterile distilled water, diluted and spread onto Potato dextrose agar (PDA) plates and germinated approximately for 4-5 days at 28^0C. Each spore resulted in a colony of meiosis products namely haploid sporidia (having a mixture of all the four possible mating types, *a1b1*, *a2b6*, *a1b6*, *a2b1*) (**Figure 5**). In order to isolate single sporidium, colonies were spread onto PDA. The single colonies emerged from it were again streaked onto another PDA plate. This way one colony was harvested from every spore.
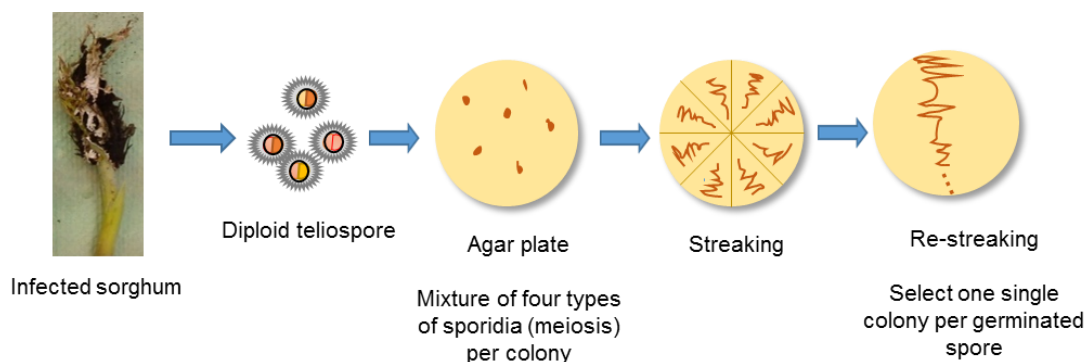
Figure 5:   Schematic diagram showing how pure colonies were isolated.

2.2.1.2 Mating assay

In order to identify strains of mating type *a1b1*, all strains were first tested for mating with an *a2b6* strain, before testing 4 different mating with *a1b1* (SRZ1_5-2), *a2b2* (SRZ2_5-1), *a2b6* (SRS2_H2-7) and *a1b6* (SRS1_H2-8), mating partners. The isolated single colonies were inoculated in Potato dextrose broth from PDA plates and shaken at $28\ ^0$C for 1-2 days at 200 rpm. SRS strain having *a2b6* background (SRS2_H2-7) was also grown in broth in the same condition to use as a mating partner. SRSZ colonies were distributed into 96 well plates and

labeled properly. After a centrifugation step, supernatants were removed carefully with pipette and 10 µl of tester *a2b6* strain was added to each well. After another round of centrifugation, the supernatants were removed again, and 10 µl of sterile distilled water added. With the help of a 96 well steel stamp these mixtures were stamped onto a 2% 12×12cm water agar plate. Plates were sealed with Para film and incubated for 1-4 days at room temperature. A magnifying laboratory microscope (Zeiss Stemi 2000C) was used at 100× magnification to check positive mating i.e. hydrophobic aerial hyphae moved slightly upon gentle blowing onto the open plate. If positive result was seen the strain was maintained in PDA plates. In a next step to have better confirmation, those selected positive strains were again grown in broth in the condition mentioned above and this time tested against all four possible tester mating type partners *a1b1*, *a2b6*, *a1b6*, *a2b1*. The four tester SRS strains with mating types were also inoculated and also grown in the same condition as mentioned above. After a centrifugation step supernatant were removed and tester strains were added in such a way that each SRSZ inoculum is mated against all four tester strains. After a subsequent centrifugation step the supernatants were removed again and strain mixtures were re-suspended in sterile distilled water. With the help of a 96 well steel stamp these mixtures were stamped onto a 2% 12×12cm water agar plate. Plates were sealed with Para film and incubated for 1-4 days at room temperature. A magnifying laboratory microscope was used at 100× magnification to check positive mating. SRSZ that showed positive results only with *a2b6* mating type (thus having an *a1b1* background) were maintained on fresh PDA plates and stored at -80 ^{0}C

2.2.2 <u>Sorghum and maize plant infection</u>

2.2.2.1 <u>Growing the microorganisms</u>

3-4 days prior to infection SRSZ strains were streaked from -80^{0}C onto 2% PDA plates. They were incubated at -28^{0}C for 3-4 days until colonies were visible.

2.2.2.2 <u>Growing the sorghum and maize plants</u>

For sorghum (*Sorghum bicolor* var Tall Polish), seeds were sown 14 days before infection. Sorghum seeds were put into 11× 7 well trays (total 77 wells) with ED73 type (BalsterEinheitserdewerk GmbH) of soil. The trays were filled with soil and the small sorghum seeds were

pushed gently into the soil 2-4 cm deep using index finger inside each well. The soil was watered until uptake limit of soil was reached. This way we created bonsai type of sorghum plants and tried to overcome the problem of the available resource (mainly space) as we required large number of plants (approximately 400 SRSZ strains $\times$ 11 plants =4400 plants in first round of infection and 200 SRSZ strains $\times$11 plants = 2200 plants in three biological replicates= 6600 plants). Plants were watered by applying up to 3 cm water into the table below the trays whenever seen needed. The plants were checked at least every two days and those plants that were weak or dying out were removed. In reality, more number of seeds was sawed than required as few seeds never germinated, few seedlings were too weak, and few died out. For seed production purpose, one seed is put every 13 cm diameter pots filled with ED73 soil type. Seeds were collected after plants fully developed.

For maize (*Zea mais* var Gaspe Flint), seeds were sawed 7 days before infection in 13 cm diameter pots filled with ED73 soil type. Seeds were pushed into the soil about 5-9 cm deep with index finger. The soil was watered until uptake limit of soil was reached. The plants that looked weak were removed before infection. Each pot contained 4-5 plants for infection. Pots for seed production contained up to 2 plants per pot. Like in sorghum the plants were watered whenever seen needed into the tables below the pots and they were checked at least every two days.

The greenhouse was set to supply light when the outdoor light intensity was below 900 µmol (350 µmol including 60% shading coefficient from the lamps) for 16 h per day. During day time the greenhouse regulated the temperature to 26°C while the temperature was dropped to 22°C at night time. Plants were treated once per week with Steinernema feltiae (Sautter und Stepper GmbH) and every two weeks with Amblyseius barkeri/ cucumeris and Acublyseius (Sautter und Stepper GmbH).

2.2.2.3   Liquid culture of *Sporisorium reilianum* for infection

One day before infection (morning), the streaked-out colonies were inoculated in 2 ml of YEPS Light medium each to make a preculture. They were then shaken at 28^0C with 200 rpm for 8-10 hours. In the evening of the same day, the OD_{600nm} of the preculture was measured and the main culture of 50 ml potato dextrose liquid medium (PD) was inoculated in such a way that the next day morning the OD_{600nm} was 0.6-0.8. The main culture was grown overnight at 28^0C and 200 rpm shaking overnight. On the day of infection (morning), the OD_{600nm} of the culture

was measured which was in between 0.6-0.8. The cultures were pelleted by centrifugation (Heraeus Multifuge X3R, Thermo Scientific) at 3500 rpm for 5 min. The supernatant was decanted and the cell pellet was re-suspended in sterile H_2O to a final OD_{600nm} of 2. Before infection SRSZ strains were mixed at 1:1 volume ratio with their compatible mating partners (for sorghum infection; with SRS2_H2-7, for maize infection; with SRZ2_5-1 or SRS2_H2-7). 5 µL of the mixed inoculum was dropped onto water agar plates and the plates were sealed with Para film and incubated at RT overnight. This was done to check the production of hyphae under microscope to confirm mating compatibility of the strains.

2.2.2.4   <u>Inoculation of sorghum and maize seedlings</u>

Suspension cultures of SRZ1_5-2 and SRZ2_5-1, or SRS1_H2-8 and SRS2_H2-7 were mixed in a ratio 1:1 and the mixture was syringe-inoculated into the leaf whorls of 7-day old maize and 14-day old sorghum seedlings as described in [86]. These two were used as control strains for every infection performed. Similarly, SRSZ strains were mixed in a 1:1 ratio with SRS2_H2-7 each and used for infecting sorghum seedlings. For maize infection, SRSZ strains were mated with SRZ2_5-1 and few selected SRSZ strains were also mated with SRS2_H2-7. For infection a syringe was used to penetrate the leaves about 1 cm above the soil and the inoculum was applied into the middle of the leaf whorl. The whorl was filled completely until a drop of the inoculum oozed out on top. After infection the plants were not watered for one day. 11 sorghum seedlings and 10 maize seedlings were used per strain combination.
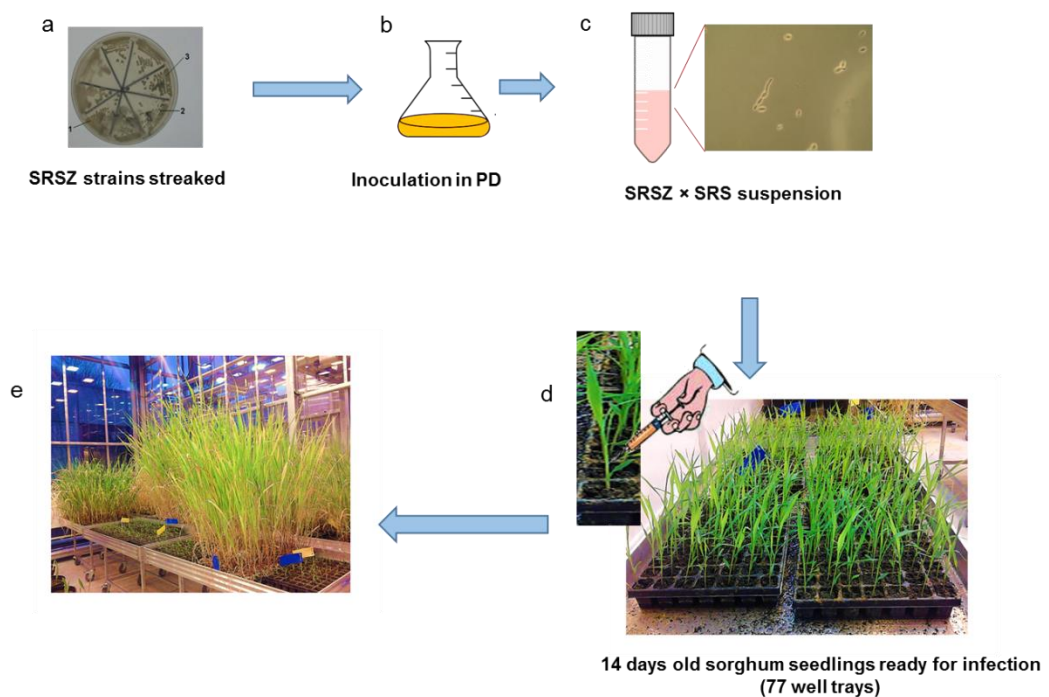
Figure 6: Schematic diagram to show various steps on infecting a sorghum seedling. a, streaked out fungal colonies on PDA plates. b, main culture of the fungal stains in 50 ml of PD liquid medium. c, suspension culture of fungal strain mixed in 1:1 ratio with compatible mating partner. d, two weeks old sorghum plants in green house ready for syringe inoculation. e, about six weeks old sorghum plants starting to form inflorescence in green house.

2.2.3 Phenotyping on sorghum pathogenicity test

Sorghum plants infected with *S. reilianum* strains were evaluated after flowering. It took around 2.5 months (10 weeks) until the flowering occurred. In case of plants with sorus instead of inflorescence, the phenotype was seen a little earlier (just over 2 months or 8 weeks). The disease phenotype was characterized into 5 categories namely (1) healthy, (2) leafy, (3) healthy+ spores, (4) leafy+ spores, (5) one sorus instead of inflorescence/ spore (see Fig. 13).

2.2.4   Phenotyping on maize pathogenicity test

Plants infected with *S. reilianum* strains were evaluated 8 to 10 weeks after the inoculation**.** They were categorized as presence or absence of spores. For each strain, all the spore forming inflorescence (in a total of 8-10 plants) were counted as spore or spore+leafy phenotype and included under total spore formation category. The percentage was calculated by dividing the number of total spore forming inflorescence by the total number of inflorescence and multiplied by 100.

2.2.5   Microscopic characterization of plant infection

2.2.5.1   Calcofluor-staining

Calcofluor staining was used to stain the fungal cells on the surface of the leaf, which makes it possible to observe infection structures (appressoria). Sorghum and maize leaf samples (1 dpi) were cut to fit into 2 ml Eppendorf tubes with around a fourth of the area being above and the rest below the injection hole. The tubes were filled with ethanol and samples were soaked overnight. Then the cut pieces were washed in water, to remove ethanol and make the samples less brittle. Afterwards they were put into 1% tween 20 for roughly 30 sec to make the hydrophobic surface of the leaf more accessible for the staining solution. The tween 20 was washed away with water and the leaves were immersed into the staining solution (see Table 6) for 30 secs before the solution was washed away with water again. The samples were put (leaf surface facing up) on slides and coverslips were put on top. Samples were analyzed with florescence microscope (Leica DM600 B). For calcofluor stain EX BP 395-440, BS FT 460, EM BP 470 was used.

2.2.5.2   WGA-Alexa Fluor 488 / Propidium Iodide Double-staining

Sorghum and maize leaf samples (3/4 dpi) were cut to fit into 2 ml Eppendorf with around a fourth of the area being above and the rest below the injection hole. The tubes were filled with ethanol and samples were soaked overnight. For staining, WGA-Alexa Fluor 488 [87] method was followed with slight modification prior to analysis by fluorescence microscopy. The samples were incubated in 10 % KOH at 95 °C for 90 minutes. Then they were washed

in phosphate-buffered saline (PBS). The PBS was discarded and 2 ml of a 1:1 mixture of the staining solution (see Table 6) (WGA-AF488 and propidium iodide) was added. The leaves were then kept in the dark for 30 min before vacuum-infiltrating them twice for 1 min each. After that, the samples were washed and destained in PBS, the midvein was cut out and the samples were put on a slide for microscopy. A drop of PBS was added, and a coverslip put onto it. Then the samples were sealed with nail polish and stored at 4 °C. A Leica DM600 B microscope equipped with filters FITC EX BP 475/40, BS FT 500, EM BP 530/50 was used for detection of WGA-Alexa Fluor, and Cy3 EX BP 545/25, BS FT 570, EM BP 605/70 for detection of propidium iodide.

2.2.5.3   <u>Callose-staining with aniline blue</u>

For detection of callose, infected leaves were collected at 1 and 2 days after inoculation and soaked overnight in ethanol. The leaf samples were washed in sodium phosphate buffer before being put into aniline blue staining solution (0.005%) (see Table 6). They were incubated for one hour and afterwards washed three times for 5 min in 50mM sodium phosphate buffer, pH 8.2. An area reaching from 0.5 cm above and 1 cm below the injection hole was cut out as the microscopy sample. Then the midvein was cut out, the samples were put on slides with sodium phosphate buffer, a coverslip was added and sealed with nail polish. Under the fluorescence microscope (Leica DM6000 B), 8 areas of 80,000 µm² were selected equally spread around the sample. 2 areas were above, 2 on roughly the same height as and 4 areas below the injection hole. The callose depositions in these areas were counted. For aniline blue stain BFP EX BP 365/12, BS FT395, EM LP397 was used.

2.2.6   <u>Colony morphology determination</u>

To study the variation in colony color and texture the SRSZ strains were first streaked on PDA plates and incubated at 28 °C for four days and evaluated. Pictures of the colonies were taken and analyzed in ImageJ NIH software. Firstly, the colored colonies were converted to grayscale and an area of about 780 sq mm was selected at the center of the colony. The imaging software "measurement" tool gave a maximum and minimum score of the grey intensity along with their mean values. This was repeated for all the strains plus the control WTs. To study the texture variations in the colonies, the streaked out colonies were categorized into 6 categories

depending on their appearance into- (1) "WT", (2) "WT with wrinkled periphery", (3) "wrinkled", (4) "granulated", (5) "smooth" and (6) "slimy" (see Fig. 31).

2.2.7   *S. reilianum* protoplast preparation

Protoplast preparation and transformation of *S. reilianum* were done as described in [86] with some modifications. Protoplasts were used so that it enables the introduction of heterologous DNA into fungal cells by transformation. To make protoplasts of SRS1_H2-8 and SRS2_H2-7, strains were streaked out on PD plates from -80^{0}C and incubated at 28^{0}C for 3-4 days until colonies were visible. The precultures were prepared in 2 ml YEPS light medium by inoculating single colony. The precultures were grown for 8-12 h at 28°C at 200 rpm. 100 ml YEPS Light main cultures were inoculated with the precultures and the main cultures were grown overnight at 28°C at 200 rpm to an OD$_{600nm}$ of 0.6-0.8. The cultures were checked under microscope for possible contamination. Cells were pelleted by centrifugation at 3500 rpm for 5 min. The supernatants were discarded, and the cell pellets were re-suspended in 50 ml SCS (see materials section for composition). Centrifugation at 3500 rpm for 10 min produced cell pellets. The pellets were re-suspended in 2 ml novozyme solution each (Novo Nordisc, Copenhagen, Denmark) and incubated for 5-10 min at RT. Cells were checked for protoplastation under the microscope until about 50 % of the cells are beginning to protoplast. The protocol was continued if 50% of all cells formed round protoplast, otherwise the solutions were incubated further and checked after 5 min. All steps after protoplastation were performed with extreme care to not destroy the protoplast. Protoplastation was stopped by adding 20 ml SCS buffer and centrifugation at 2300 rpm for 15 min. Supernatants were discarded. The pellets were washed twice with 20 ml SCS and once with 20 ml STC (see Table 6), by re-suspending the pellets and centrifuging at 2300 rpm for 25 min for each washing step and at 280 rpm after STC step. Washed protoplasts were re-suspended in 500 µL cold STC buffer and aliquoted into 70 µL portions. Aliquots were stored at -80°C until further use.

2.2.8   Preparation of Rubidium chloride competent *Escherichia coli*

Competent *E. coli* cells are vital for later transformation. Chemically competent cells were made using the method described in [88] with minor modifications. A single *E. coli* colony was used to inoculate 10 ml LB supplemented with 0.5 mM CaSO$_4$ and 0.5 mM MgSO$_4$. This

preculture was grown overnight at 37^0C at 200 rpm. A main culture of 100 ml was inoculated with 1 ml of the preculture and again grown at 37^0C at 200 rpm upto an OD_{600nm} of 0.5-0.6. The next day the culture was pelleted by centrifugation (Heraeus Multifuge X3R, Thermo Scientific) at 300 rpm for 15 min at 4^0C. The supernatant was discarded, and the pellet was re-suspended in 33 ml ice cold RFI solution (see Table6). The suspension was then incubated for one hour in ice. Cells were pelleted at 3000 rpm for 15 min at 4^0C. The supernatant was discarded, and the pellet was re-suspended in 330 µl ice cold RFII solution (see materials section for composition). The suspension was then incubated for 15 min in ice. The cell suspension was aliquoted in 30 µl portions each. They were stocked at -80^0C. Competent cells were 6× concentrated.

2.2.9   Transformation of chemically competent *E. coli* cells

The method described in [Cohen *et al.* 1972] was adopted with minor modifications. First competent cells were thawed on ice from -80^0C stock. The thawed cells were resuspended in in 270 µl ice cold RFII solution. 50 µl of cell suspension was used for one transformation. 2-5 µl of DNA was added to the cell suspension in ice and the mixture was incubated for 20 min in ice. Cells were given heat shock at 42^0C for 45 sec and subsequent cooling on ice. Cells were recovered by adding 250 µl of LB medium and incubating for 1 h at 37^0C with slight agitation. 50 µl of cells were plated on LB plate having 100 µg/ml Ampicillin. Cells were grown overnight at 37°C. Next day single colonies were picked out using sterile toothpicks inoculated in LB medium plus Ampicilin to isolate plasmids further.

2.2.10   Plasmid isolation

Plasmids were isolated using Nucleospin® Plasmid kit (Macherey-Nagel) as explained in the provided manual.

2.2.11   Gel electrophoresis of Nucleic acids

Nucleic acids were analyzed and visualized by electrophoretical separation on a 0.8-2% TAE or TBE agarose gel (see Table 6). The proper weight of agarose was added in 1X TAE or 0.5X TBE buffers and boiled until agarose was completely melted. Roti stain was added to a

final concentration 1 µg/ml. The gel was cooled to ~60°C and then was poured in closed cassette with the appropriate comb. After solidification, the gel was placed in electrophoresis tank filled with the same buffer as the gel (1X TAE or 0.5X TBE buffers). DNA samples were mixed with loading buffer and loaded in the wells along with 5 µl DNA 2-log ladder (NEB, Frankfurt). The samples were electrophoretically separated with 70-150 V for 0.3-4 h.

2.2.12   Gene deletion in *S. reilianum*

Clonemanager Professional suite software version 8 from Sci-Ed was used to analyze plasmid sequences, design primers, perform restriction digests and perform PCR reactions. In addition, deletion constructs suitable for homologous recombination in order to knock out gene(s) in *S. reilianum* were constructed *in silico* using it.

For gene deletion in *S. reilianum*, by double homologous recombination [85] (**Figure 7**), a reporter cassette that is surrounded by a left and a right flank, called the deletion construct is generated. The flanks are homologous DNA sequences to regions left and right to the target gene. Flanks are should be about 1000 bps in size. When introducing the deletion construct into *S. reilianum* cells, the target gene gets exchanged by the reporter cassette and thus the gene is deleted.
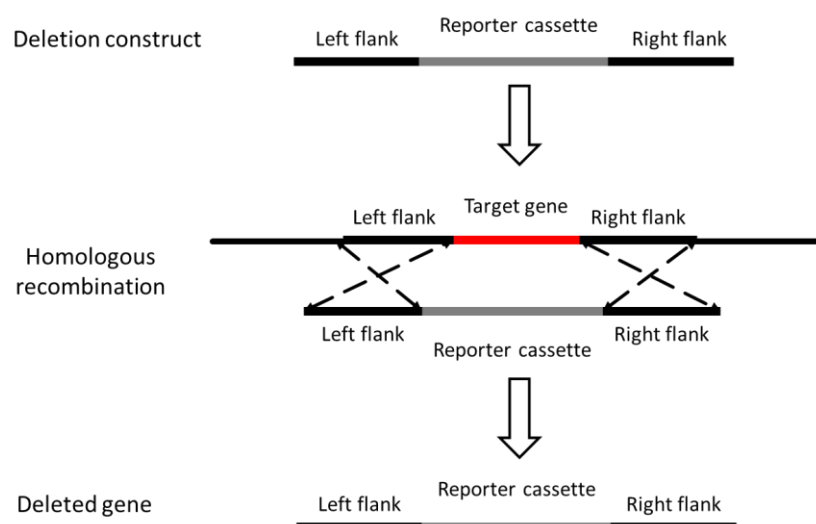
Figure 7:   Deletion of a target gene using homologous recombination. Deletion construct (top) con-sists of flanking region (black bars) flanking a reporter cassette (grey bar). Via homologous recombination (middle) the target gene (red bar) gets exchanged with the reporter

cassette. The reporter cassette gets inserted into the genome (black line) at the position of the gene, leading to gene deletion (bottom).

2.2.12.1   PCR amplification of flanking regions surrounding the region to be deleted

For the construction of the deletion construct, regions of about 1 kb surrounding the 7_11 cluster genes were amplified. Specific primers that carry a distinct non-palindromic *Sfi*I recognition site were used for PCR (see Table 3). PCR was performed on genomic DNA using Phusion® polymerase (NEB) to amplify the flanks. The amplification was done using Personal thermocycler (Biometra) and the cycling conditions were set depending on polymerase, size of amplification and individual primer annealing temperature:

Programs to amplify flanks for knock-out construct with Phusion®

Denaturation	98°C	0:40 min
Denaturation	98°C	0:20 min
Anealing	x°C	0:30 min
Extension	72°C	1 min
Final Extension	72°C	10 min
35 Cycles		

2.2.12.2   PCR purification of the flanks

The amplified products from the flanks were purified using Nucleospin® Gel and PCR Clean up kit (Macherey-Nagel) as described in the manual. 1 µl of each sample was checked on a 1% Agarose gel. The amount of DNA was measured by NanoDrop Spectrophotometer (Thermo Scientific).

2.2.12.3   Digestion of the flanks and the resistance cassette using *Sfi*I

Both the flanks were amplified using primers having distinct non-palindromic *Sfi*I recognition sites. Those sites were cut with *Sfi*I (NEB, Frankfurt) that cuts at specific sites with sequence GGCCN|NNN|NGGCC producing unique sticky ends. The purified flanks were digested

with *Sfi*I for 2-4 h at 50 ^{0}C. A hygromycin resistance cassette was cut from p^{BS}-*hhn* plasmid [85] with *Sfi*I. The digest was directly separated by electrophoresis on a 1 % agarose gel. The 1871 bp of hph resistance cassette was sliced out form the gel with MinElute Gel Extraction Kit (Qiagen). The concentration of purified products was measured using Nanodrop Spectrophotometer (Thermo Scientific). The *Sfi*I sites in the primers to amplify the flanks were designed in such a way that the left flank only ligates with the left border of hygromycin cassette while the right flank ligates only with the right side of hygromycin resistance cassette

Reaction for digestion with *Sfi*I

0.5-1 µg	DNA
2 µl	10×NEB buffer 2
0.2 µl	BSA
0.5-2 U	*Sfi*I
Fill to 20 µl	Water

2.2.12.4   DNA ligation

The Ligation of DNA fragments (Left flank+ hygromycin cassettes+ right flank) was performed using T4 DNA Ligase (NEB) at 16 ^{0}C for 8 h. The DNA fragments were mixed in 1:2:1 (left flank: resistance cassette: right flank) ratio.

Reaction for ligation

50-500 ng	digested flanks and cassette
3 µl	10× T4 DNA Ligase buffer
1 U	T4 DNA Ligase
Add to 30 µl	Water

The desired ligation product of 4.4 kb was sliced out from the gel using Nucleospin© Gel elution kit (Macherey-Nagel).

2.2.12.5   Nested PCR to create final deletion construct

The ligation product of 4.4 kb was used directly as template for nested PCR using nested primers (see Table 3). PCR was performed using Q5® High-Fidelity DNA Polymerase (NEB). The final product size expected was 4.2 kb.

Programs to amplify the knock-out construct with Q5® High-Fidelity DNA Polymerase

Denaturation	98°C	0:30 min	
Denaturation	98°C	0:10 min	
Anealing	x°C	0:30 min	27 cycles
Extension	72°C	2.5 min	
Final Extension	72°C	2 min	

2.2.12.6   Confirmation of knock-out by PCR and gel electrophoresis

Colonies from each mating type were picked from plates having hygromycin. The genomic DNA of these mutants were isolated and used as template to check by PCR. First a PCR with primer pair oNB017 and oNB018 that amplfied a product inside the cluster of size 534 bps was performed. The samples that did not give a product in the first PCR were used for a second PCR with primer pair oVT022 and oNB012 which amplified the hygromycin cassette and right flank of the deletion construct with a product size of 1.56 kb. These strains were our knock-out mutants.

2.2.13   Genomic DNA extraction from *S. reilianum* offspring

All the *Sporisorium* strains used for genomic DNA isolation were streaked out to PD plates from -80^0C stocks and incubated at 28^0C for 3-4 days. After visible colonies appeared they were grown for two days in 3.5 ml YEPS Light at 28°C and 200 rpm to OD_{600nm} greater than 1.0. Around 4 µL of each culture was checked under microscope for possible contamination after first day of growth and immediately before isolation. In 2 ml eppendorf tubes about 0.3 g glass beads were added. 2 ml of the grown culture were added to individual eppendorf tubes with glass beads and centrifuged for 2 min at 13.000 rpm. Supernatant was removed completely from each sample completely. The last 1.5 ml leftover culture was added to each tube

and centrifuged again for 2 min at 13000 rpm. The supernatants were removed. Clearly visible pellet was seen at the bottom of the eppendorf tubes. The samples were frozen at -20^0C for at least 20 min or overnight. For the probable knock-out *Sporisorium* strains, genomic DNA was isolated using standard phenol-chloroform extraction method. The pelleted strains kept at -20^0C were taken out and immediately 600 µL of freshly prepared TE-Phenol-Chloroform (Table 6) and 500 µL of Ustilago lysis buffer (Table 6) were added to each of them. They were put on the vibrax for 10 min at max speed. Then the samples were centrifuged for 25 min at 13.000 rpm (Biofuge Pico). The tubes were removed carefully so as not to disturb the phases. Carefully the supernatants right above the interphase were taken out and transferred into new eppendorf tubes. 900 µL of 100% ethanol was added to each and inverted carefully. After 1 min of incubation, they were then centrifuged for 15 min at 13.000 rpm. The supernatants were discarded again and was washed the pellet with 70% ethanol and centrifuged again for 15 min at 13.000 rpm. This step was repeated two times. After removing the supernatant; the pellets were dried until the ethanol smell was gone. The pellets were dissolved later in 50 µL of TE-RNase buffer and incubated at 50^0C for 10 min. 1 µL of each sample was checked on agarose gel for quality since a photometric measurement by NanoDrop of phenol-extracted DNA led to false values.

For all the SRSZ strains and SRS2_H2-7 and SRZ1_5-2 strains used for genome re-sequencing, genomic DNA was extracted with QiagenRNeasy plant mini kit as described in the kit manual. This resulted in 30 µL of DNA with minimum concentration of 50 ng µL^{-1} measured with Nanodrop (GE Nanovue plus) and the A260/280 ratio of 1.7-1.8. A 1% agarose gel was also run to check their quality. A total of 192 samples were sent at RT to GATC (Eurofins GATC Biotech GmbH, Germany) for library construction and Illumina sequencing. Three samples were later on not considered for further analysis (due to lack of needed quality). Raw reads were obtained from GATC for the 189 strains that included the two parental genomes of SRS2_H2-7 and SRZ1_5-2.

2.2.14   DNA re-sequencing analysis by CLC genomics workbench 7.5

CLC Genomics Workbench is a tool to analyze and visualize next generation sequencing (NGS) data. The workbench is a good tool to generate custom workflows, and accelerate data analysis by combining quality control steps, adapter trimming, read mapping, variant detection, and multiple filtering and annotation steps (https://www.qiagenbioinformatics.com/products/clc-genomics-workbench/).

2.2.14.1 *De novo* assembly of SRS H2-7 and SRZ 5-2

To assemble the sequenced parental strains SRS2_ H2-7 and SRZ1_5-2, the raw reads were imported to the workbench after they were unzipped. To import, import-> standard import-> select the files-> next-> save. To de novo assemble, Toolbox-> NGS core tools-> Trim sequences (Trim adapter list TruSeq3-PE-2-1, discard reads below length=15) -> Save->select the trimmed file-> De Novo Sequencing-> De Novo Assemble (Mismatch cost=2, insertion cost=3, Deletion cost=3, Length fraction=0.5, Similarity fraction=0.8, minimum contig length=1000, word size=45, bubble size= 9) -> Save (**Figure 8**).
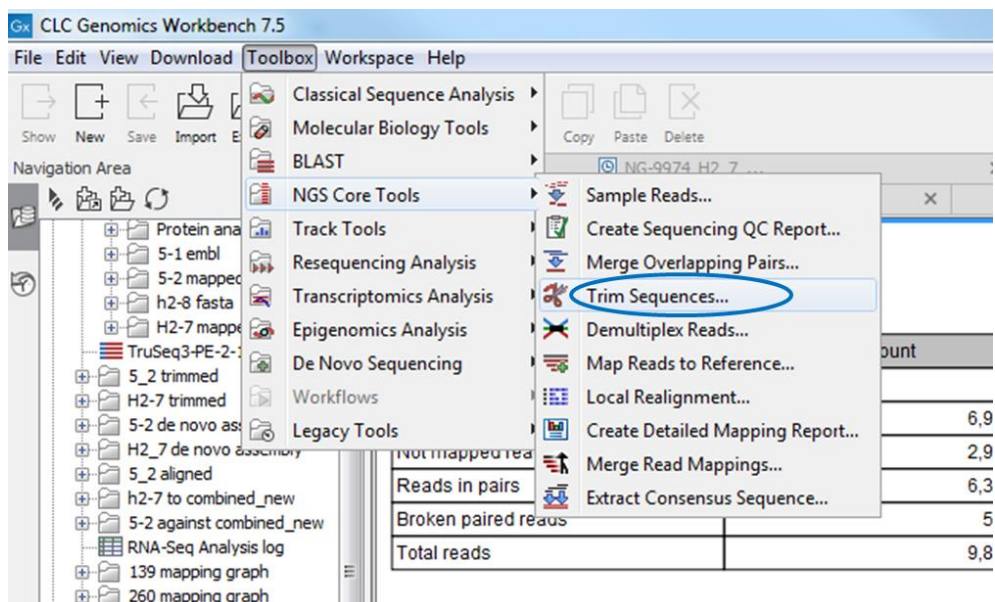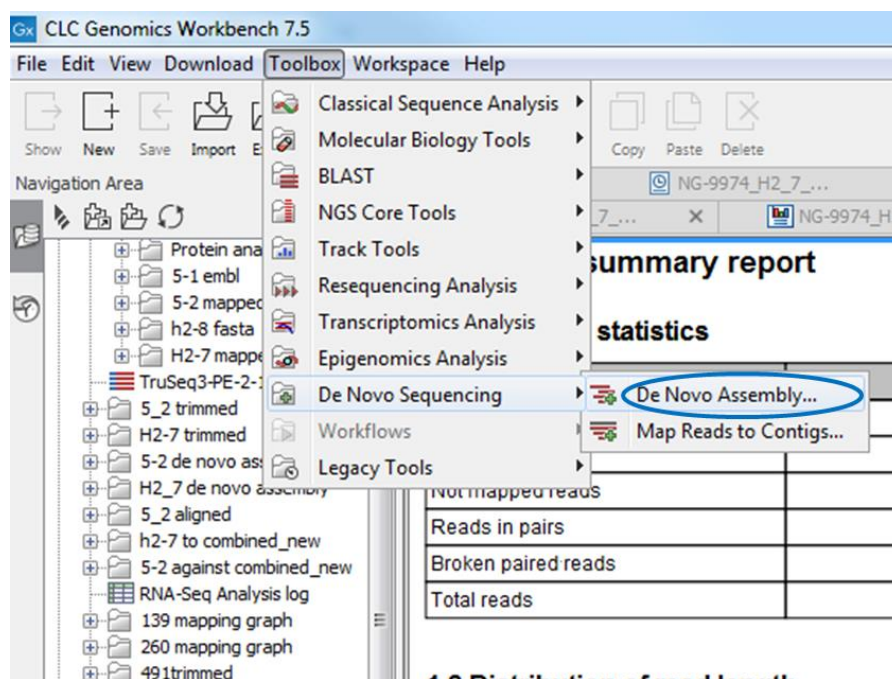
Figure 8: Screenshot from CLC genomics workbench 7.5. The upper picture shows the way to trim the files. The lower picture shows the way to de novo assemble the raw reads.

2.2.14.2 Mapping reads to reference

To perform standalone mapping of SRS2_ H2-7 reads to SRS1_ H2-8 as reference and SRZ1_5-2 reads to SRZ2_ 5-1 as reference, the raw reads files were first imported to workbench. In order to import, import -> standard import -> select the files -> next -> save. They were first trimmed for quality control, Toolbox -> NGS core tools -> Trim sequences (Trim adapter list TruSeq3-PE-2-1, discard reads below length=15) -> Save. For mapping, Toolbox -> NGS core tools -> map reads to reference -> select file -> (no masking, map randomly, global alignment, Mismatch cost=2, insertion cost=3, Deletion cost=3, Length fraction=0.5, Similarity fraction=0.8, intergenic region/non intergenic region) -> save result.
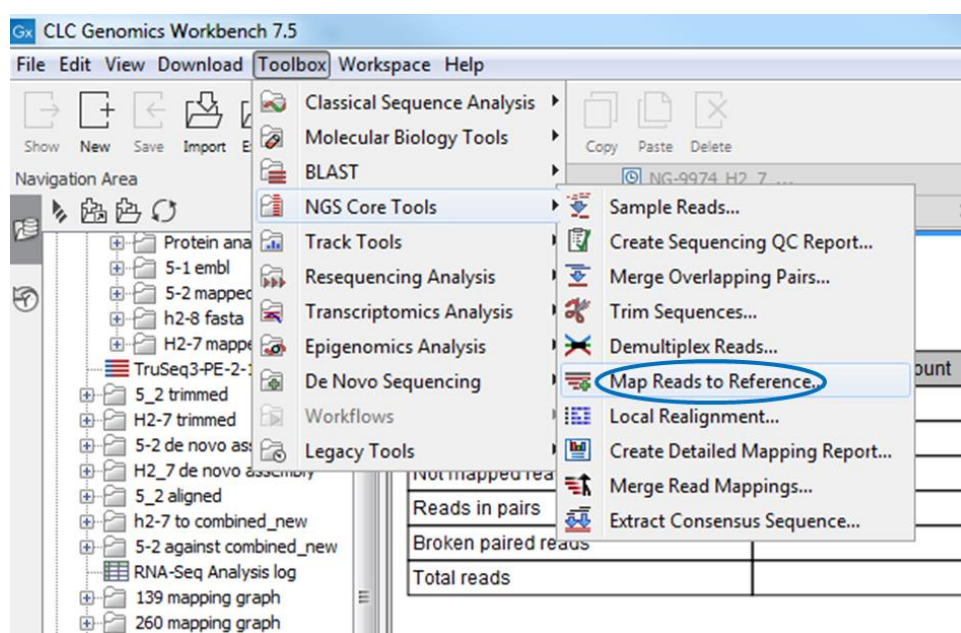
Figure 9: Screenshot from CLC genomics workbench 7.5. The pathway to perform mapping of reads to reference is shown.

2.2.14.3 Importing the combined SRS H2-8 and SRZ 5-1 annotated file

The SRZ2_ 5-1 annotated genome was searched in Pedant database. To import the combined annotated files (combined by Dr. Emad Albarouki) of the two sister parents to the workbench, import -> standard import -> select the file (*S. reili*_H2-8&5-1_CDS) -> next -> save.

2.2.14.4 Importing the SRSZ offspring raw read files

Due to space limitation the files were imported one at a time to the workbench and when the mapping was over it was deleted to import the next file. The raw read files were first unzipped. In order to import, import -> standard import -> select the files -> next -> save. They were first trimmed for quality control, Toolbox -> NGS core tools -> Trim sequences (Trim adapter list TruSeq3-PE-2-1, discard reads below length=15) -> Save.

2.2.14.5 Mapping of the SRSZ reads against combined SRS H2-8 and SRZ 5-1

The trimmed reads were considered as transcript reads and a transcriptomic analysis was performed as follows, Tooolbox ->Transcriptomics analysis -> RNA–seq analysis (S. reili_H2-8&5-1(Genome) as Reference sequence and S.reili_H2-8&5-1_CDS as Gene track) (Map to gene regions only, Mismatch cost=2, insertion cost=3, Deletion cost=3, Length fraction=0.8, Similarity fraction=0.8, maximum number of hits for a read=10, strand specific=both, global alignment) -> save (**Figure 10**). This gave the coverage analysis of the reads against the combined parents onto each gene of the combined 23+23 chromosomes.
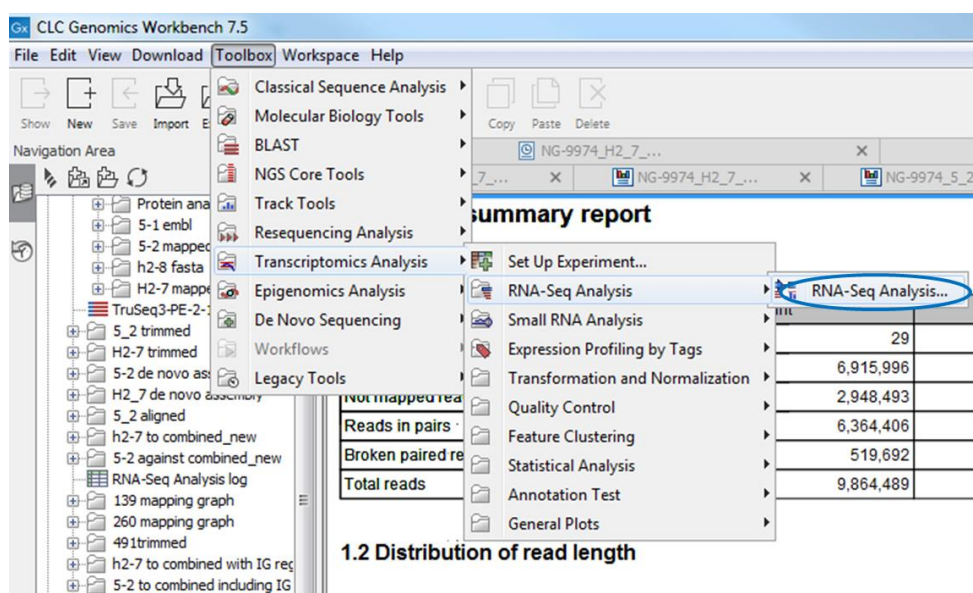
Figure 10:   Screenshot of CLC genomics workbench 7.5. The picture depicts the way to perform RNA seq analysis of the raw reads.

2.2.14.6 Exporting the data out of the workbench

The mapped file is exported out of the workbench as, select the file -> right click -> export -> excel 2010 format -> save. This gave an excel sheet with the following data: Name of all the genes from both the parental chromosomes arranged in a linear fashion, chromosome numbers, expression value, gene length, RPKM, unique gene reads, total gene reads.

2.2.15   *In silico* analysis of cluster 7_11

2.2.15.1   Amino acid sequence identity comparison between SRS and SRZ

To compare amino acid sequence between the SRS H2-8 and SRZ 5-1 SIMAP values were compared against each other. Similarity Matrix of Proteins (SIMAP) database (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3965014/) provides a comprehensive and up-to-date pre-calculation of the protein sequence similarity matrix, sequence-based features and sequence clusters.

2.2.15.2   Prediction of secreted proteins

The following bioinformatics tools were used to predict the secretory nature of the proteins:

SignalP: The Signal Peptide Prediction tools can be used to find secretory signal peptides in protein sequences. Signal peptides target proteins to the extracellular environment either through direct plasmamembrane translocation in prokaryotes or are routed through the endoplasmatic reticulum in eukaryotic cells. The signal peptide is removed from the resulting mature protein during translocation across the membrane (http://www.cbs.dtu.dk/services/SignalP/versions.php). D-score (discrimination score) is a weighted average of the mean S (signal peptide score) and the max Y (combined cleavage site score) scores. This is the score that is used to discriminate signal peptides from non-signal peptides.

Tools	Purpose	Guideline for SP
SignalP3.0 NN	prediction of signal peptides	D score >0.43

Phobius: This server is for prediction of transmembrane topology and signal peptides from the amino acid sequence of a protein. The program takes proteins in FASTA format. For the long format (default) output, Phobius gives a list of the location of the predicted transmembrane helices, the predicted location of the intervening loop regions and signal peptide. If the whole sequence is labeled as cytoplasmic or noncytoplasmic, the prediction is that it contains no membrane helices. In the short output format one line is produced for each protein with no graphics. SP": Y/N   indicator   if   a   signal   peptide   was   predicted   or   not (http://phobius.sbc.su.se/instructions.html).

TMHMM2.0: This program is for prediction of transmembrane helices in proteins. The program takes proteins in FASTA format. The output long format (default) gives some statistics and a list of the location of the predicted transmembrane helices and the predicted location of the intervening loop regions, if the whole sequence is labeled as inside or outside, the prediction is that it contains no membrane helices (https://hpc.nih.gov/apps/tmhmm/TMHMM2.0.html).

WoLFPsort: It is a program for protein subcellular localization prediction (https://wolfpsort.hgc.jp/aboutWoLF_PSORT.html.en).

LocSigDB: LocSigDB is a manually curated database of experimental localization signals for eight distinct subcellular locations; primarily in a eukaryotic cell with brief coverage of bacterial proteins (http://genome.unmc.edu/LocSigDB/).

Nucpred: It analyses a eukaryotic protein sequence and predicts if the protein spends at least some time in the nucleus or no time in the nucleus (https://nucpred.bioinfo.se/nucpred/).

Tools	Purpose	Guideline for SP
WoLFPsort	protein Subcellular Localization Prediction	"Extr" major
Phobius	signal peptide predictor	"Y" for SP

	Purpose	Cutoff for non SP
TMHMM2.0	to run the transmembrane helix prediction	>2
Phobius	signal peptide predictor	>2
LocSigDB	database of experimental localization signals	ER retention signalpresent
Nucpred	predicting Nuclear Localization of Proteins	Threshold>0.8

2.2.15.3 Structural modeling

Structural modeling of predicted proteins utilized Swiss-Model (Swiss institute of bioinformatics). It is a structural bioinformatics web-server dedicated to homology modeling of protein 3D structures. The SWISS-MODEL pipeline comprises the following main steps that are

involved in building a homology model of a given protein structure (https://swissmodel.expasy.org/):Identification of structural template(s). BLAST and HHblits are used to identify templates. The templates are stored in the SWISS-MODEL Template Library (SMTL), which is derived from PDB.Alignment of target sequence and template structure(s).Model building and energy minimization

2.2.15.4 Prediction of intrinsic disorder

To analyze the residue level of disorder propensity of cluster 7_11, four intrinsic disorder predictors were used: PONDR® VLXT, PONDR® XL1_XT, PONDR® VL3-BA and PONDR® VSL2 (http://www.pondr.com/pondr-tut2.html). PONDRs are typically feed forward neural networks that use sequence attributes taken over windows of 9 to 21 amino acids. These attributes, such as the fractional composition of particular amino acids, hydropathy, or sequence complexity, are averaged over these windows and the values are used to train the neural network during predictor construction.

ANCHOR algorithm (http://anchor.enzim.hu/) was used to predict protein binding regions that are disordered in isolation but can undergo disorder-to-order transition upon binding. This computational tool finds segments within disorder regions that cannot form stable intrachain interactions to fold on their own but are likely to gain stabilizing energy by interacting with a globular protein partner.

2.2.15.5 Additional Bioinformatic Analyses

Alignment of nucleotide and/or amino acid sequences to find regions of similarity between such biological sequences employed Basic Local Alignment Search Tool (BLAST; https://blast.ncbi.nlm.nih.gov/Blast.cgi).

3   Results

3.1   Phenotyping

3.1.1   Analysis of pathogenicity of F1 progeny (SRSZ) on sorghum (using SRS2_H2-7 as mating partner) reveals various degree of disease phenotype

To test their virulence on sorghum, infection of *Sorghum bicolor* 'Tall Polish' with the two controls SRS2_ H2-7 (mating type *a2b6*) × SRS1_ H2-8 (mating type *a1b1*) (hereafter re-ferred to as SRS) and SRZ2_ 5-1(mating type *a2b2*) × SRZ1_5-2 (mating type *a1b1*) (hereafter referred to as SRZ) was carried out. We used these two compatible mixtures and inoculated 14-days old seedlings of sorghum. Disease symptoms were scored at flowering time. Inoculation of SRS led to 100% of the plants with spores and inoculation of SRZ showed 100% of the plants with healthy phenotype (**Figure 11**).
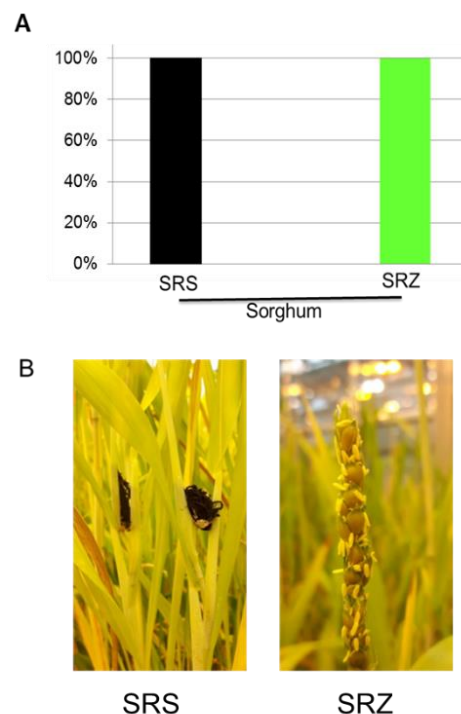
Figure 11: Symptoms of *Sporisorium reilianum* infection. **A**, Seedlings of Sorghum bicolor 'Tall Polish' were inoculated with a mixture of the compatible haploid *S. reilianum* strains

SRS1_H2-8 and SRS2_H2-7 (SRS) and SRZ1_5-2 and SRZ2_5-1 (SRZ). Symptoms were scored in the inflorescences at 12 to 16 weeks after inoculation. For each round of infection 11 sorghum plants were inoculoated with SRS and SRZ inoculum. This was repeated for thre rounds. **B**, Inflorescence symptoms of *S. reilianum* infection of sorghum. Inoculation of sorghum seedlings with SRS replaced the inflorescence by a fungal sorus containing masses of black spores (left), while while inoculation with SRZ led to healthy panicles (right).

A cross between SRZ1_5-2 (mating type *a1b1*) × SRS2_ H2-7 (mating type *a2b6*) was done and used for sorghum infection before I joined the laboratory. After evaluation it was seen that ~1% of the plants could produce spores. To increase the number of strains for the mapping population, we made use of spores that were generated during this cross-variety infection experiment. To generate cross-variety offspring, the diploid teliospores were spread and cultivated on nutrient agar where they germinated and formed small colonies composed of haploid sporidia of different mating type (see methods section for details). They were haploid basidiospores that represented different segregants of the parental strains' genetic pool. Individual sporidia were isolated by streaking to dilution on nutrient plates. Forming colonies were composed of mitotically multiplying haploid sporidia that carried the same mating type (which could be any among *a1b1*, *a1b6*, *a2b1*, or *a2b6*). The segregants with mating type *a1b1* (hereafter referred to as SRSZ) were selected by mating test (see methods section for details). All the SRSZ segregants were named starting with the initial "TW" (for example TW1, TW2 and so on). We used SRS2_H2-7 (mating type *a2b6*) as a mating partner to analyze the SRSZ virulence capacity.

For plant infection, cultures of each SRSZ offspring (*a1b1*) were mixed in a one-to-one ratio with a culture of the strain SRS2_H2-7 (*a2b6*). The strain mixture was inoculated into the leaf whorl of at least eleven sorghum seedlings, which were subsequently bred until they formed flowers (around three months) and analyzed for signs of smut disease. There were 386 SRSZ offspring that were tested on sorghum (See the material section for the list). We have used 11 sorghum seedlings per strain grown in a tray having 11×7 wells (**Figure 12**). The plants were thus shaped to form bonsai like growth in the limited amount of soil provided. This saved a lot of space as in normal condition they can grow up to 5 feet height.
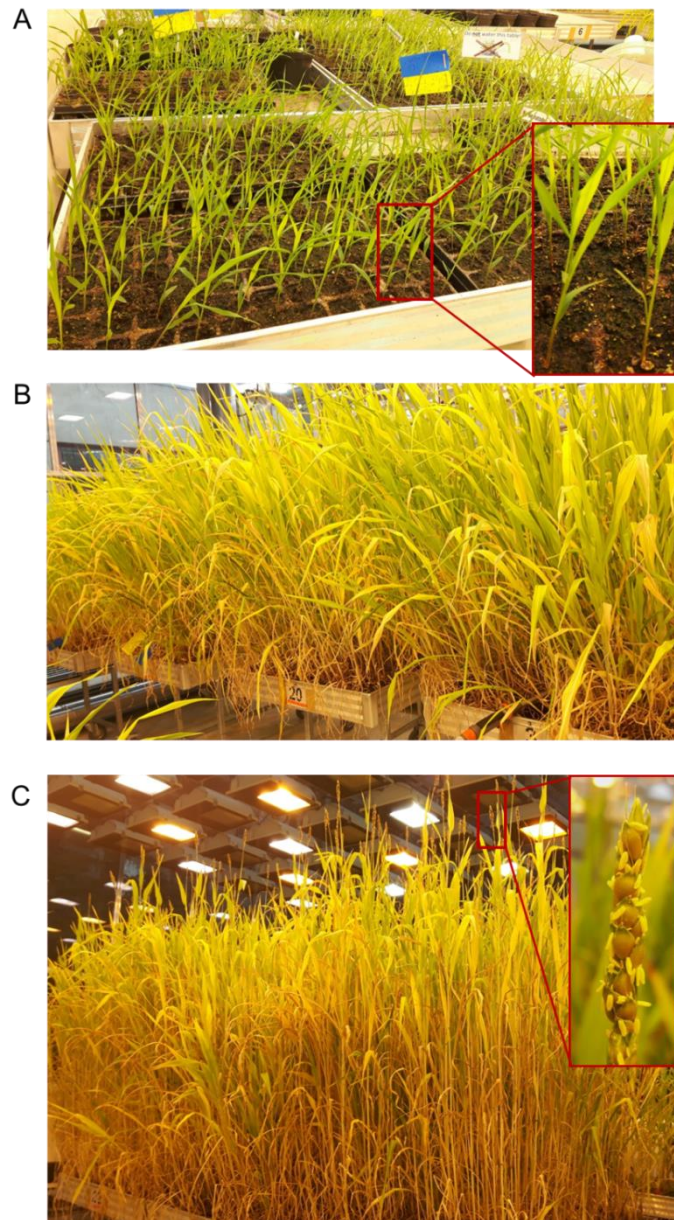
Figure 12: Sorghum plants inside greenhouse condition. **A**, two - week old seedlings ready for inoculation; there are 11 seedlings per row used for each strain, inset view: two seedlings at 3 – leaf stage. **B**, one-month old sorghum plants as seen to grow under bonsai like condition. **C**. three months old flowering plants ready for evaluation, inset view: one healthy panicle.

In some cases, the sorghum plants that did not survive. Some seeds couldn't germinate, or some plants died before inoculation. In those cases, the experiment was repeated.

We could observe 5 different disease phenotypes of SRSZ (*a1b1*) × SRS2_H2-7 (*a2b6*) infection on sorghum (**Figure 13**). In the two extreme cases we saw disease symptoms similar to the two control strains of SRS and SRZ. They were healthy and spore phenotypes. In addition, three intermediate phenotypes were seen. In one case, the inflorescence formed phyllody like structures and was named as "leafy", another kind showed leafy phenotype mixed with spores and was named as "leafy+spore" and the last ones were healthy phenotype mixed with spores and named as "healthy+spore".

Healthy   Leafy   Healthy+Spore   Leafy+Spore   Spore

Figure 13: Inflorescence symptoms of *S. reilianum* SRSZ infection of sorghum. Inoculation of sorghum seedlings of SRSZ (a1b1) mated with compatible SRS2_H2-7 (a2b6) led to five different kinds of disease symptoms - Healthy, leafy, healthy+spore, leafy+spore and spore. Scale bar: 1 cm.

The graphical representation of the result (as % of virulence capacity) of the first round of infection on sorghum is shown in **Figure 14**. The final result showed 145 strains were completely healthy, 173 strains were intermediate and 68 strains were 100% spore phenotype on sorghum. That formed roughly 37.5% of the total SRSZ strains tested to be healthy, 44.8% intermediate and 17.6% as 100% spore in the first round of sorghum infection.
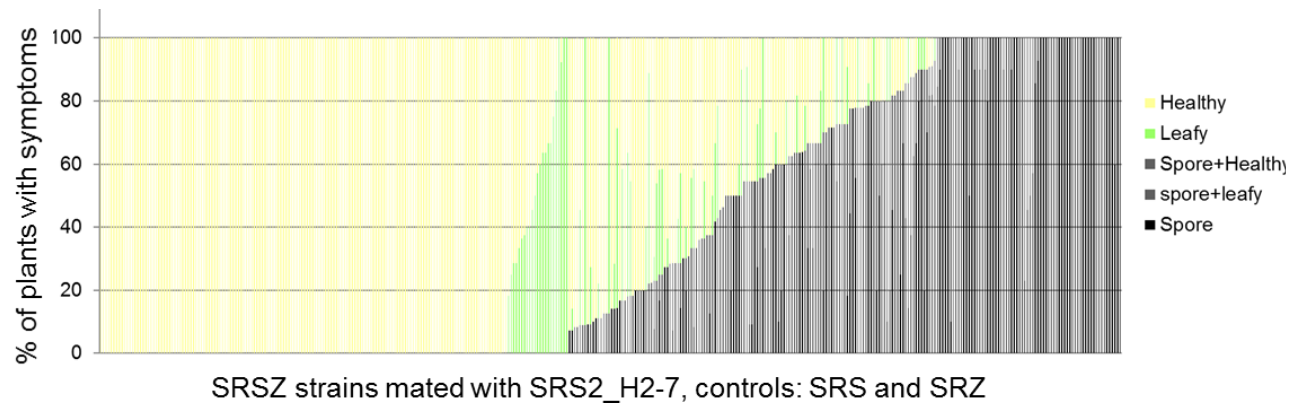
Figure 14: SRSZ segregants show phenotypic virulence differences on sorghum. Different capacities to form spores on sorghum is seen above. The y-axis represents the percentage of the sorghum plants (11 sorghum plants per strain) that show various disease phenotypes, and the x-axis has all the tested SRSZ strains. The strains are sorted by severity of plant infection. The symptoms were scored 12 to 16 weeks after inoculation. The list of the SRSZ strains is in the materials section. See supplement for the evaluation list (**Table 5s**).

We selected 123 healthy and 67 offspring showing the full spore phenotype for a second and third round of infection on sorghum. After three biological replicates were tested on sorghum, the final result was 110 healthy offspring, 50 offspring with 100% spore and 29 offspring with intermediate phenotype on sorghum (**Figure 15** A). These total 189 SRSZ strains along with parental SRZ1_5-2 and SRS2_ H2-7 were further used for genotypic analysis (see section 3.2).

A

B

189 SRSZ offspring mated with SRS2_H2-7, controls: SRS and SRZ
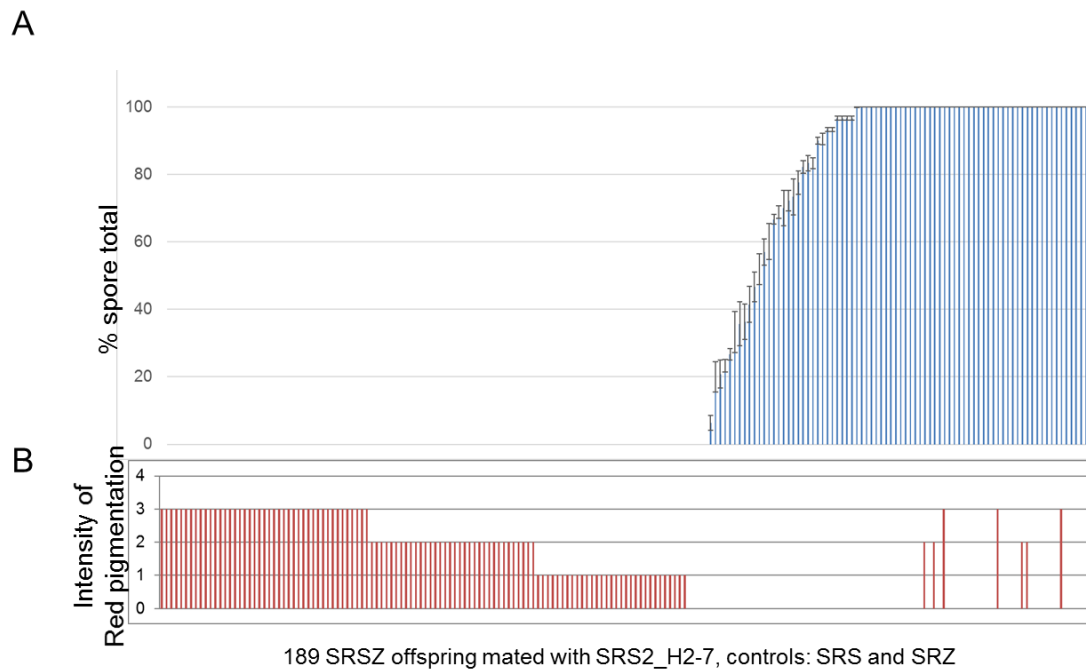
Figure 15: Three biological replicates of SRSZ mated with SRS2_H2-7 infection result on sorghum. A. A total of 47 SRSZ offspring produced 100% spore, 30 showed intermediate phenotype and 100 were healthy and didn't produce any spores. The symptoms were scored 12 to 16 weeks after inoculation. Data shown are means of three inoculation experiments. For each experiment, eleven plants were used. Error bars show standard error of the means (Table 3s). **B.** Phytoalexin response of the sorghum leaves at 4-5 dpi inoculated with SRSZ × SRS2_H2-7. The intensity of the red pigmentation is quantified from scale 0 to 3 (0 showing no pigmentation and 3 showing maximum pigmentation). Most of the phytoalexin evaluations were performed twice and for few strains thrice (for text see section 3.1.5). See supplement for the evaluation lists used in A and B (**Table 4s**).

3.1.2   Most of the SRSZ strains can form spore on maize (using SRZ2_5-1 as mating partner)

182 out of selected 189 SRSZ strains were tested for their spore formation capacity on maize after crossing with SRZ2_5-1 (mating type *a2b2*). Like in sorghum, we used a mixture of the compatible haploid strains of SRSZ (*a1b1*) × SRZ2_5-1 (*a2b2*) in one-to-one ratio to inoculate 7-day old seedlings of *Zea mays* 'Gaspe Flint'. Disease symptoms were scored at flowering

time which was after about 2-2.5 months. As control strains we used SRS and SRZ. We could observe six different disease phenotypes of SRSZ (*a1b1*) × SRZ2_5-1 (*a2b2*) infection on maize (**Figure 16**). Control SRS did not produce any spores (neither in male nor in female inflorescences) but showed healthy (male and female inflorescences or leafy female phenotype. SRZ control on the other hand produced spores in female inflorescence (spore and spore+leafy) along with leafy female and healthy (male and female inflorescence). In addition, there were quite a few SRSZ strains that could produce spores on male inflorescences as well (TW298, TW336, TW442, TW547, TW79, TW477, TW107, TW182, TW235, TW412, TW439,TW345, TW239, TW466, TW78, TW311, TW343, TW285, TW418, TW449, TW451, TW150 and TW337) and a few showed leafy male inflorescences (TW87, TW438, TW322, TW368, TW229, TW281, TW349, TW484, TW233, TW309, TW488, TW408 and TW208).
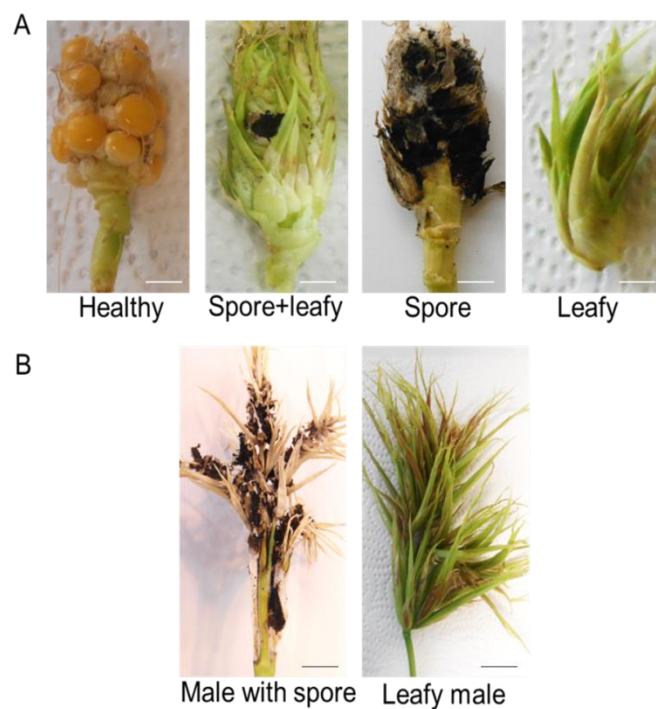
Figure 16: Inflorescence symptoms of *S. reilianum* SRSZ infection of maize. Inoculation of sorghum seedlings of SRSZ (a1b1) mated with compatible SRZ2_5-1 (a2b2) led to six different kinds of disease symptoms – **A**, Healthy, Spore+leafy, Spore, and Leafy (all female inflorescences) and **B**, Male with spore and Leafy male inflorescence. Scale bar: 1 cm.

Out of 182 SRSZ strains tested, 148 could form spores (total spore male+ female) in various degrees and 34 did not produce any spores on maize (**Figure 17**). From this evaluation it was found that 33 out of 34 strains that did not produce spores in maize also did not produce spore in sorghum. However, except for five strains, most showed at least leafy phenotype in maize. The five strains that didn't produce spore either in sorghum or maize could however mate successfully. Their *in-planta* proliferation capacity was checked by microscopy in the following experiments (see section 3.1.4).
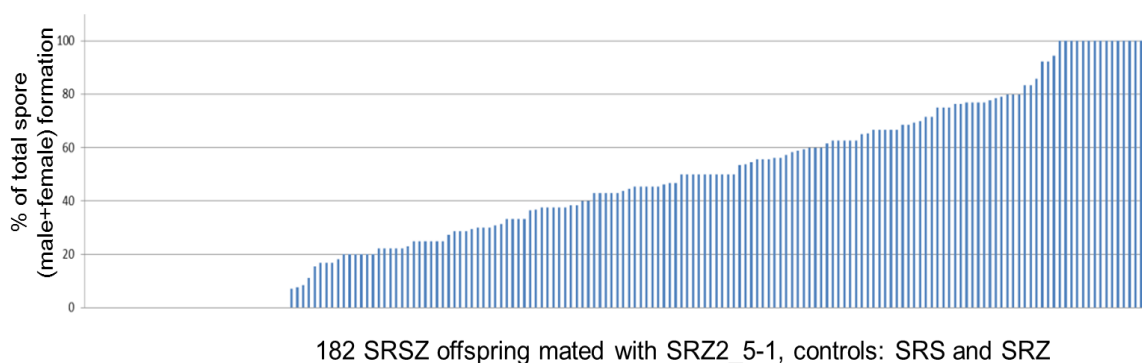
Figure 17: SRSZ segregants show phenotypic virulence differences on maize. Different capacities to form spores on maize is seen above. The y-axis represents the percentage of the total spore formation combining both male and female inflorescence (8-10 maize plants per strain) and x-axis has all the tested SRSZ strains. For each strain, all the spore forming inflorescence (in a total of 8-10 plants) were counted as spore or spore+leafy phenotype and included under total spore formation category. The percentage was calculated by dividing the number of total spore forming inflorescence by the total number of inflorescences and multiplied by 100. The symptoms were scored 8 to 10 weeks after inoculation. See annex for evaluation list (**Table 8s**).

3.1.3   Most of the fully virulent SRSZ strains on sorghum show various degrees of virulence on maize (using SRS2_H2-7 as mating partner)

We wanted to know the spore forming capacity of those SRSZ strains on maize that could previously produce 100% spores in sorghum (in all three biological replicates) on mating with SRS2_H2-7. Out of 47 total 100% spore producing SRSZ in sorghum, we tested 28 in maize

after mating with SRS2_H2-7. The controls used were - SRS, SRZ and a combination of SRZ1_5-2 (*a1b1*) × SRS2_H2-7 (*a2b6*). 10 strains could produce spores (total spore male and female) and rest 18 strains didn't produce any spores. The control SRZ1_5-2 (*a1b1*) × SRS2_H2-7 formed spores in between SRS (no spore) and SRZ (**Figure 18**).
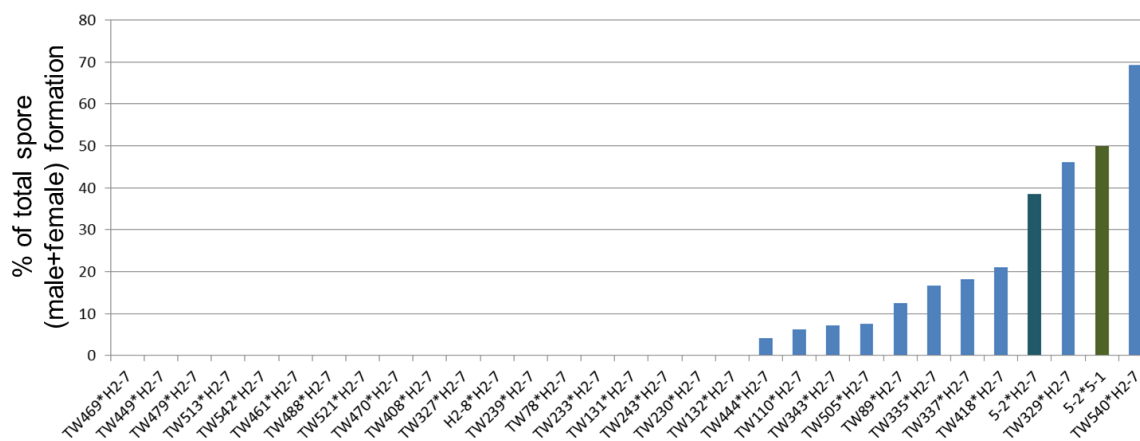
182 SRSZ offspring mated with SRS2_H2-7, controls: SRS, SRZ, SRZ1_5-2 × SRS2_H2-7

Figure 18: SRSZ segregants show phenotypic virulence differences on maize. Different capacities to form spores on maize is seen above. The y-axis represents the percentage of the total spore formation combining both male and female inflorescence (8-10 maize plants) were counted as spore or spore+leafy phenotype and included under total spore formation category. The percentage was calculated by dividing the number of total spore forming inflorescence by the total number of inflorescence and multiplied by 100. The symptoms were scored 8 to 10 weeks after inoculation. See supplement for the evaluation list. (**Table 9s**).

3.1.4   Few strains avirulent on both sorghum and maize stop before reaching vascular bundles

Five SRSZ strains that were unable to produce spores in both sorghum (mating partner SRS2_H2-7) and maize (mating partner SRZ2_5-1) were checked for their *in planta* proliferation capability inside sorghum and maize. These are - TW260, TW491, TW376, TW280 and TW342.

For maize the aviruelnt SRSZ strains were mated with compatible mating partner SRS2_H2-7 and for sorghum with SRZ2_5-1. Leaf samples were collected 1 dpi from both to check for appressoria formation. All the strains could successfully form appressoria on maize

leaves shown using calcofluor staining (**Figure 19**). The same was seen in the case of sorghum plants.
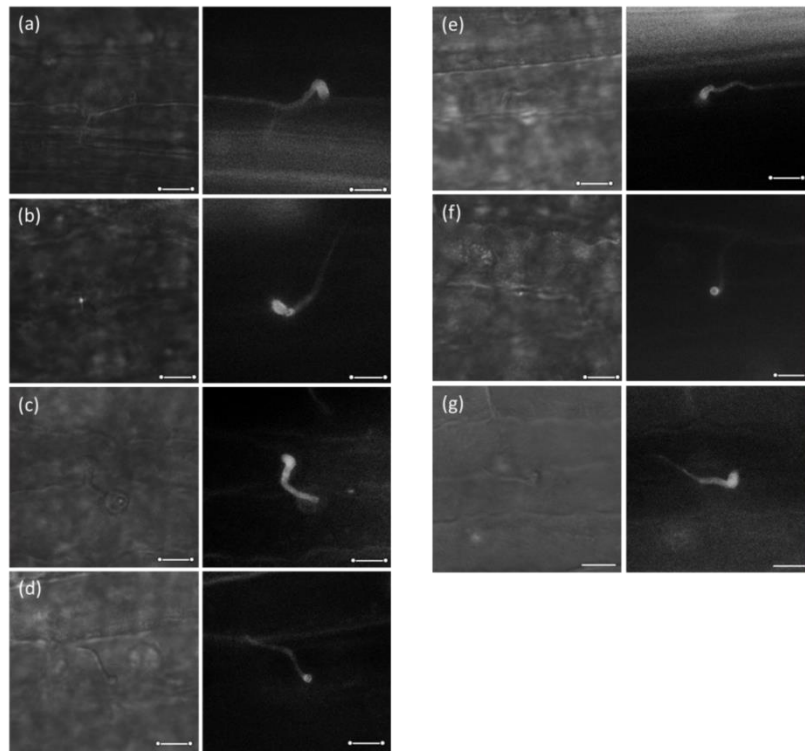
Figure 19: Fluorescence microscopy of calcofluor stained SRSZ fungal strains on *Zea mays* leaf surface at 1 dpi. The five avirulent strains TW260, TW491, TW376, TW280 and TW342 (**a**, **b**, **c**, **d** and **e** respectively) could form appressoria successfully on leaf surface. Controls are SRS (**f**) and SRZ (**g**). Scale bar = 10 μm. (Picture courtesy: Alan Midani)

To further investigate the behavior of these avirulent SRSZ during plant colonization, an analysis of the strains in maize and sorghum was performed by fluorescence microscopy of stained samples taken from leaf. For staining, a combination of Propidium iodide / WGA-Alexa Fluor 488 -staining was used. WGA-Alexa Fluor 488 stains fungal structures in green, while propidium iodide stains plant cells and dead fungal hyphae in red. Leaf samples were collected 3 dpi from both sorghum and maize. At 3 dpi, the strains had efficiently spread into maize leaves and hyphae were found colonizing bundle sheath cells (**Figure 20**). However, they did not proliferate inside the vascular bundles as did the SRZ control. They were more prominent in the mesophyll areas in between neighboring vascular bundles. They also as expected could

not reach apical meristems. This indicated that these avirulent SRSZ strains stopped before colonizing vascular bundles in maize.
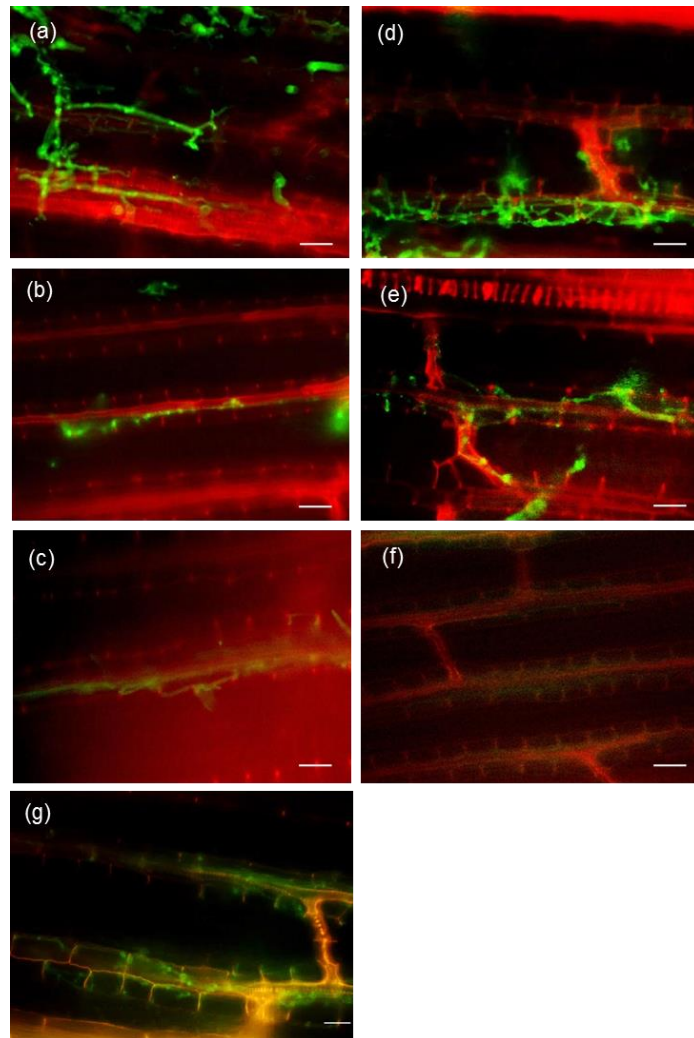
Figure 20: Fluorescence microscopy of fungal structures inside *Zea mays* seedling leaves, 3 dpi. The five avirulent strains TW260, TW491, TW376, TW280 and TW342 (**a**, **b**, **c**, **d** and **e** respectively) could proliferate up to the bundle sheath cells but stop before entering the vascular bundles. Controls: SRZ (**g**) and H$_2$O inoculated maize seedling (**f**). Green: WGA-Alexa Fluor 488 (fungus); Red: Propidium Iodide (Plant). Scale bar: 25 μm, 40× magnification. (Picture courtesy: Tom Spilker)

As seen in case of maize, fluorescence microscopy of avirulant SRSZ strains in sorghum leaves at 3 dpi also showed similar results. The hyphae were found colonizing mesophyll cells

without being able to grow inside vascular bundles (**Figure 21**). It could be that the plant recognizes the fungus early and need some time to utilize its effective mechanisms to prevent it from growing inside the vascular bundles.
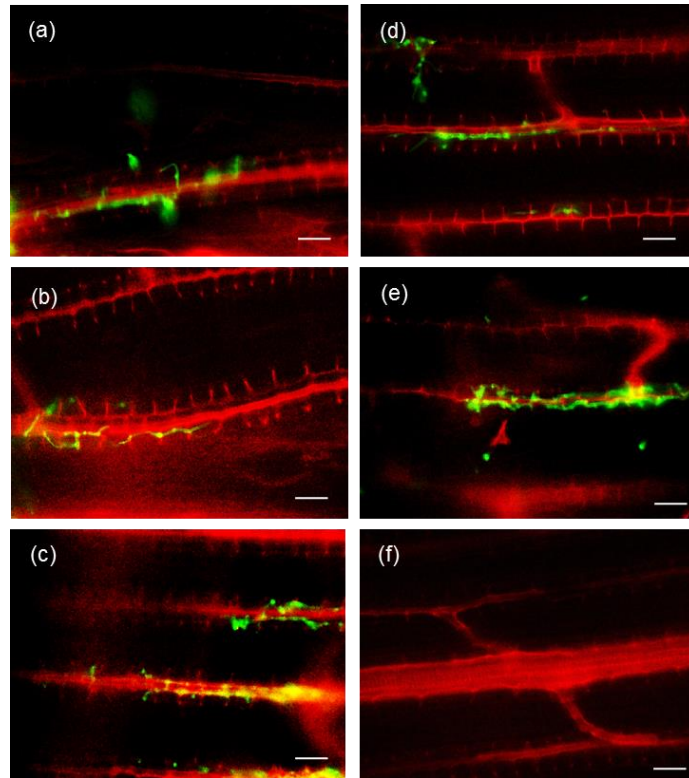
Figure 21: Fluorescence microscopy of fungal structures inside *Sorghum bicolor* 'Tall Polish' seed-ling leaves, 3 dpi. The five avirulent strains TW260, TW491, TW376, TW280 and TW342 (**a**, **b**, **c**, **d** and **e** respectively) could proliferate up to the bundle sheath cells but stop before entering the vascular bundles. Control: H_2O inoculated maize seedling (**f**). Green: WGA-Alexa Fluor 488 (fungus); Red: Propidium Iodide (Plant). Scale bar: 25 μm, 40× magnification. (Picture courtesy: Tom Spilker)

To further investigate the development of the fungus inside the plant, samples of meristems from infected host plants at 10 dpi were taken stained and inspected under the fluorescence microscope. The meristems of all plants inoculated with the tested strains showed the same result, regardless of the host. As seen in **Figure 22**, the meristems showed no sign of fun-

gal cells growing inside or around them. This result was also expected as these strains did not form any spores in any of the hosts.
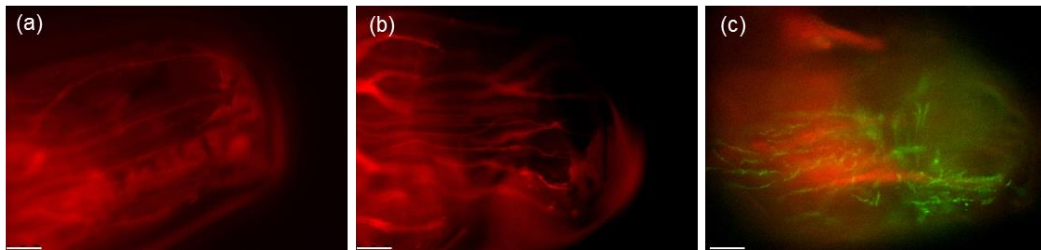
Figure 22: Fluorescence microscopy of meristem tips of (**a**), maize and (**b**), sorghum plants after infection with avirulent SRSZ strains at 10 dpi (TW260 in picture above, the same observation was found in all the rest strains TW491, TW376, TW280 and TW342). Control: SRS in sorghum (**c**). Plant tissue was stained red with propidium iodide and fungal tissue stained green with WGA-AF488. Scale bar: 250 μm, 10× magnification. (Picture courtesy: Tom Spilker)

3.1.5 The phytoalexin production to a large extent is associated with disease phenotype

It is known that sorghum produces red spots, previously identified as the production of phytoalexins luteolinidin and apigenidin when infected with SRZ, but not with SRS [10]. We wanted to know the phytoalexin production on sorghum leaves after inoculation with the SRSZ strains (mating partner SRS2_H2-7).

As expected, sorghum leaves inoculated with control SRS did not show signs of red spots at 4 dpi. Only visible was a weak chlorosis near the inoculation point (**Figure 23**). However, sorghum seedlings inoculated with SRZ showed red spots covering the point of inoculation at 4 dpi.

Figure 23: Phytoalexin production in sorghum. Sorghum leaves infected with SRZ show the emer-gence of phytoalexins visible by their red color at 4 dpi (right side picture), which gets more intense at later time points but no phytoalexin production on infection with SRS (left side picture). The phytoalexin production was scored at 4 dpi. Scale bar: 0.5 cm.

All the 189 SRSZ strains were tested for production of phytoalexin. The phytoalexin for-mation was quantified as appeared to naked eye from 0 to 3, 0 being no red pigmentation and 3 being the highest pigmentation (**Figure 24**).
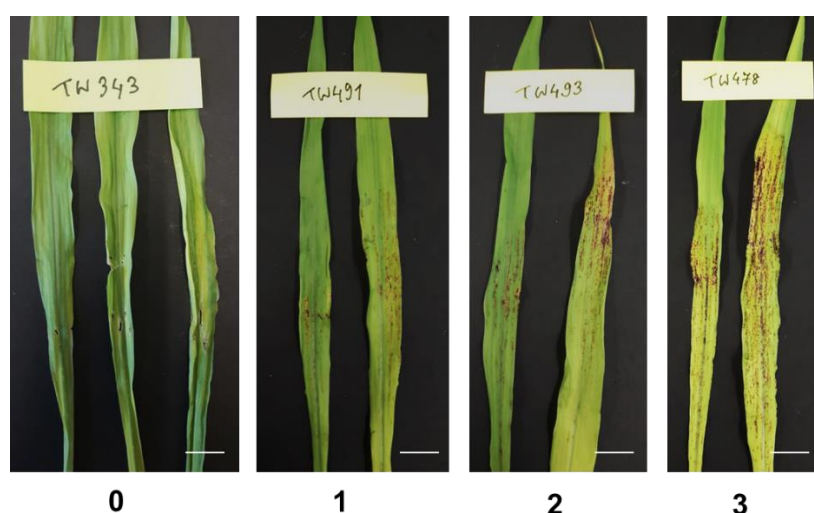
Figure 24: Representative figure showing various intensities of red pigmentation of phytoalex-in formation upon infection with SRSZ strains mated with SRS2_H2-7 in sorghum. **From left to right:** TW343 shows no phytoalexin formation, only mild chlorosis and quantified as '0'.

TW491 showing mild red pigmentation, quantified as '1'. TW493 showing higher pigmentation than TW491, quantified as '2' and TW478 showing the highest pigmentaion, quantified as '3'. The phytoalexin production was scored at 4 dpi. Scale bar: 1 cm.

As expected most strains that were producing spores on sorghum did not show any phytoalexin production but mild chlorosis and most strains that couldn't produce spores later on in sorghum were seen to be resisted by plant defense response in the form of phytoalexin production (**Figure 25**, **Figure 15** B).
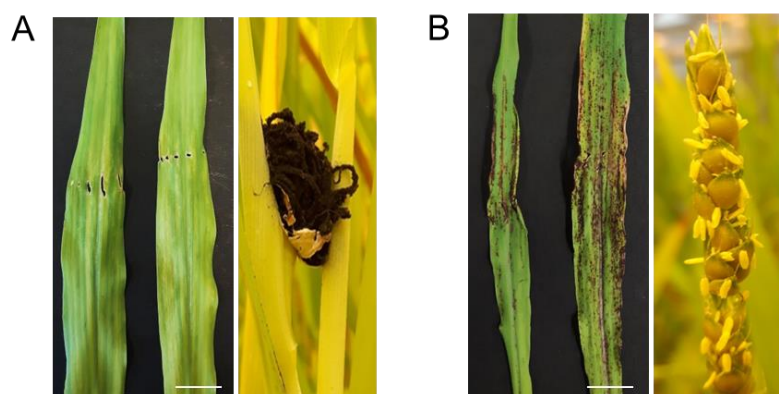
Figure 25: Phytoalexin production by SRSZ strains. **A**, the virulent SRSZ strains (see Table 16s) did not induce phytoalexin production in sorghum (except 7 strains) and **B**, the avirulent SRSZ strains were restricted by strong plant defense in the form of phytoalexin prodution (except 3 strains). The phytoalexin production was scored at 4 dpi. Scale bar: 1 cm for left picture of A and B.

There were 10 SRSZ strains that showed anomaly in terms of correlation of phytoalexin production with respect to disease phenotype. These were: TW413, TW331, TW445, TW161, TW177, TW311, TW212, TW513, TW160 and TW128. In 7 of them even after inducing phytoalexin production at 4 dpi, they could produce spores at the time of flowering. On the other hand, in 3 of them, even though there was no phytoalexin production as plant defense mechanism, they were not able to produce spores (**Figure 15** B).

3.1.6   The amount of callose depositions in sorghum can differ regardless of phytoalexin responses or virulence

Another way of plant defense mechanism is by callose deposition [89]. Deposition of callose, a (1,3)-β-glucan polymer, in the form of cell wall thickenings called papillae, at site of wall penetration is a common response by plants upon fungal attack [89]. In *Arabidopsis thaliana* callose can strongly support penetration resistance when deposited in elevated amounts at early time points of infection by Powdery Mildew [90]. Callose deposition is mostly seen at 1 or 2 dpi using aniline blue staining, which stains callose in fluorescent blue under UV-illumination. We wanted to know if there existed any correlation between callose deposition and of virulence of SRSZ strains.

As control SRS and SRZ were used. In sorghum leaves inoculated with SRS, the presence of fluorescence was rare as compared to plants inoculated with SRZ at 2 dpi (**Figure 26**). This showed involvement of callose in the defense response against SRZ in host plant. Now, six avirulent SRSZ strains were tested for callose deposition in sorghum leaves. Except for TW413 (phytoalexin was not produced on its inoculation in sorghum), others were resisted by plant defense via phytoalexin production. The results can be seen below (**Figure 27**, **Table 7**).
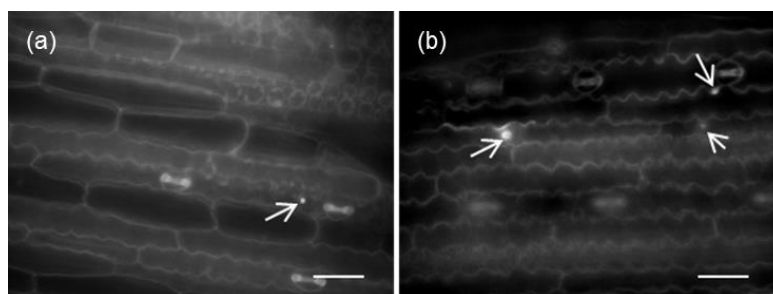
Figure 26: Black and white pictures showing callose deposition in infected plants with SRS and SRZ. Sorghum and leaves were collected at 2 dpi. Sorghum infected with SRS showed no or weak callose deposition (**a**), while strong levels of callose were observed for SRZ (**b**), White arrows indicate callose deposition, Scale bars: 20 μm.
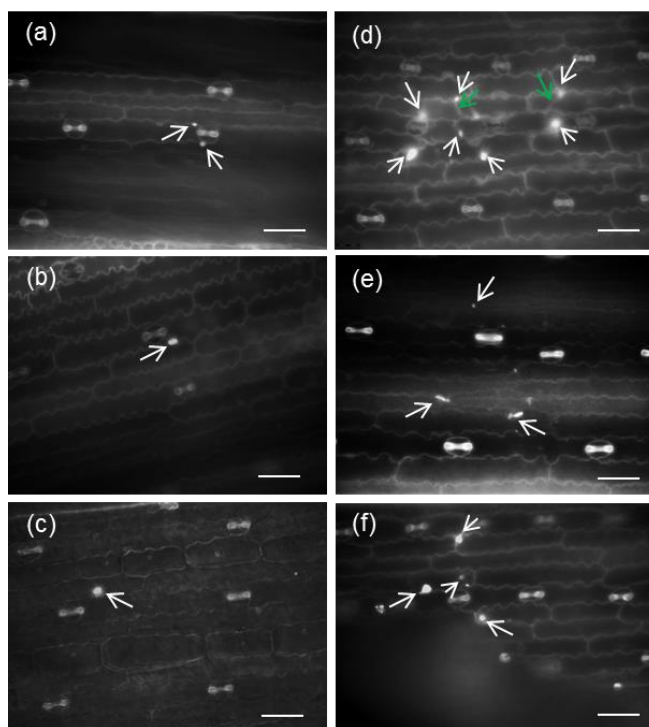
Figure 27: Black and white pictures showing callose deposition in infected plants with SRSZ. Sorghum and leaves were collected at 2 dpi. The induction of callose depositions can either be on a low level as in SRS, TW342 and TW491 (**b** and **c** respectively), on a higher level like SRZ, TW289, and TW229 (**d** and **f** respectively) or in between, TW260 and TW413 (**a** and **e** respectively). White arrows indicate callose deposition, green arrows show fungal hyphae. Scale bar: 20 μm. (Picture courtesy: Tom Spilker)

For a better statistical comparison of the callose depositions, the average numbers with their respective standard deviations were calculated for each strain. These values were then compared via a t-test to see whether there is a significant difference between the tested hybrid strain and the two wild types. As expected, the number of callose depositions is much higher in the SRZ control than in the SRS control. The pure numbers of callose depositions showed clear differences between the strains that could also be proven with statistics (**Table 7**). The strains TW229 and TW280 showed an amount like SRZ, while the number in TW342 and TW491 was much smaller, like SRS. TW413 and TW260 seemed to have values in between (**Figure 28**). Table showed that most tested SRSZ strains showed one significant difference to one of the control strains. TW229, TW260 and TW280 showed a significant difference to SRS, while TW342 and TW491 showed a significant difference to SRZ. TW413 showed no significant difference to both control strains and seemed to sit in the middle. On a one-sided test with the

same significance threshold of 0.95 however, it would surpass the critical t-value of 2.131 for SRZ. This showed that the induction of callose depositions can either be on a low-level like in SRS, on a high level like SRZ, or in between. The amount of callose depositions is most likely not linked to virulence, as the avirulent TW342 and TW491 showed similar results to SRS whereas avirulent TW289 and TW229 showed similar result to SRZ (**Figure 28**). Also, phytoalexin response seemed to have no effect on callose deposition. TW413 that did not induce phytoalexin showed an only small level of callose deposition as compared to TW260 that could induce phytoalexin production.

Table 7: Statistical evaluation of the number of callose depositions in sorghum leaves infected with SRSZ by means of a t-test comparing each strain to controls SRS and SRZ. Values that exceed the critical t-value for a two-sided significance level of 0.95 ($|t| > 2.776$) are marked blue (Data courtesy: Tom Spilker)

Strains	Average number/area	S	t-value (SRS)	t-value (SRZ)
SRS	1.958	0.065	0	-6.673
SRZ	6.333	0.926	6.663	0
TW229	5.875	0.935	5.940	-0.603
TW260	4.625	1.274	3.224	-1.878
TW280	6.917	2.363	3.502	0.399
TW342	1.708	0.236	-0.621	-8.383
TW413	3.333	1.838	1.220	-2.525
TW491	2.125	0.177	0.426	-7.731
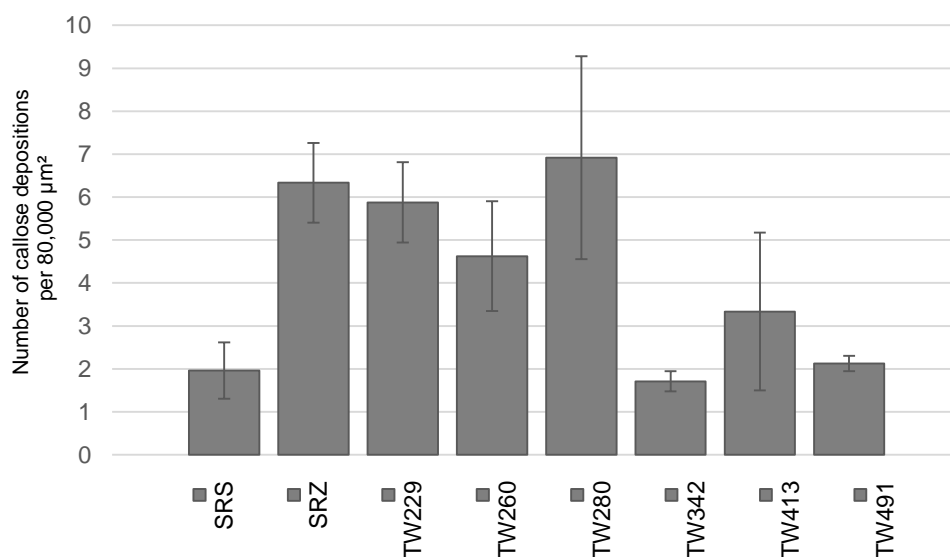
Figure 28: Average number of callose depositions per 80,000 μm² in Sorghum leaves infected with SRSZ mated with SRS2_H2-7. Leaf samples were collected at 2 dpi. Controls: SRS and SRZ. (Data courtesy: Tom Spilker)

3.1.7   Colony morphology of haploid yeast cells of SRSZ vary in coloration and texture

S. reilianum is a facultative biotrophic fungus that can live saprotrophically as haploid yeast cells (sporidia). We have noticed that there was a difference in colony morphology in terms of its coloration and texture when grown in potato dextrose agar medium. We wanted to measure this difference for all 189 SRSZ strains as compared to control strains- SRS2_ H2-7, SRS1_ H2-8, SRZ2_ 5-1 and SRZ1_5-2. For this purpose, the streaked out SRSZ strains from -80^0C stocks were grown at 28^0C for 4 days and then used for analysis.

The images captured were converted into black & white and by use of the software ImageJ (NIH), the intensity of gray scale coloration was measured for each strain (**Figure 29**). An area of about 700-780 sq. mm at the center of each colony was selected. The mean of intensity of gray coloration was calculated by the software.

area	mean	min	max
784	168.102	153	181

area	mean	min	max
780	124.87	102	152
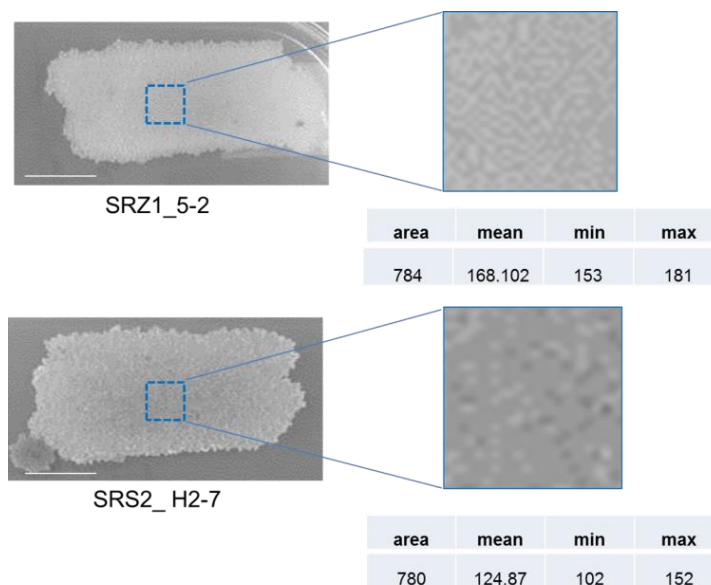
Figure 29: Representative figure to show measurement of gray scale coloration of two control strains - SRZ1_5-2 and SRS2_ H2-7 growing in potato dextrose agar using software ImageJ (NIH). The measurement was done 4 days after streaking of the strain. The area at the center of about 780 sq. mm gave maximum and minimum intensity and the mean of the two was used as a final value. As can be seen darker the coloration, lower is the mean value. Scale bar: 1 cm.

For SRZ controls intensity of gray scale coloration was- 168.10 and 166.15 (SRZ1_5-2 and SRZ2_ 5-1 respectively). For control SRS it was- 137.72 and 124.87 (SRS1_ H2-8 and SRS2_ H2-7 respectively). All the other SRSZ strains showed values higher, in between or lower than the control strains (**Figure 30**).

Figure 30: Intensity of gray scale coloration of SRSZ colonies on potato dextrose agar medium. The two orange bars represent SRS controls - SRS2_ H2-7, SRS1_ H2-8 and two green bars represent SRZ controls - SRZ2_ 5-1 and SRZ1_5-2. Y-axis is showing mean intensity of an

area of about 780 sq. mm at the center of the colony. Lower the value, darker is the coloration and vice versa. Pictures were taken at 4 days after streaking. Software used for quantification: ImageJ (NIH). See supplement for the evaluation list (**Table 6s**).

For checking the colony texture, the strains were streaked out on potato dextrose agar plates and classified by looking on various textures as visible to naked eyes. We classified them to six categories based on the type of texture observed (**Figure 31**). The texture shown by the controls SRS2_ H2-7, SRS1_ H2-8, SRZ2_ 5-1 and SRZ1_5-2 were named as "WT". The rest five were- "WT with wrinkled periphery", "wrinkled", "granulated", "smooth" and "slimy". Most of the strains showed wrinkled followed by WT texture.
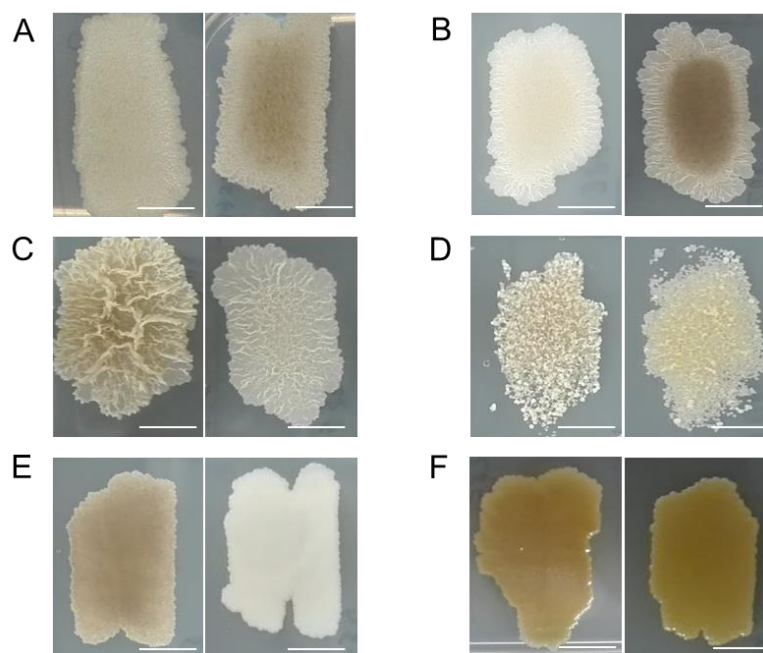
Figure 31: Texture of SRSZ colonies on potato dextrose agar medium. **A**, the four control strains showed WT texture. **B**, WT with wrinkled periphery. **C**, wrinkled. **D**, granulated. **E**, smooth and **F**, slimy texture. Pictures were taken at 4 days after streaking. Scale bar: 1 cm. See annex for evaluation list (**Figure 7s**).

3.2 Genotyping

3.2.1   *De novo* assembly of parental SRS2_ H2-7 and SRZ1_5-2 show the average genome size to be 18.368 Mbp

For genotype analysis, genomic DNA of all selected 189 SRSZ strains and two parental strains SRS2_ H2-7 and SRZ1_5-2 was isolated. The strains were grown in potato dextrose liquid medium and genomic DNA was isolated (see methods section for details). Gel electro-phoresis was performed to check DNA integrity and concentration of >100nanogram/microliter. The A260/280 ratio was 1.8 for all the samples (**Figure 32**).
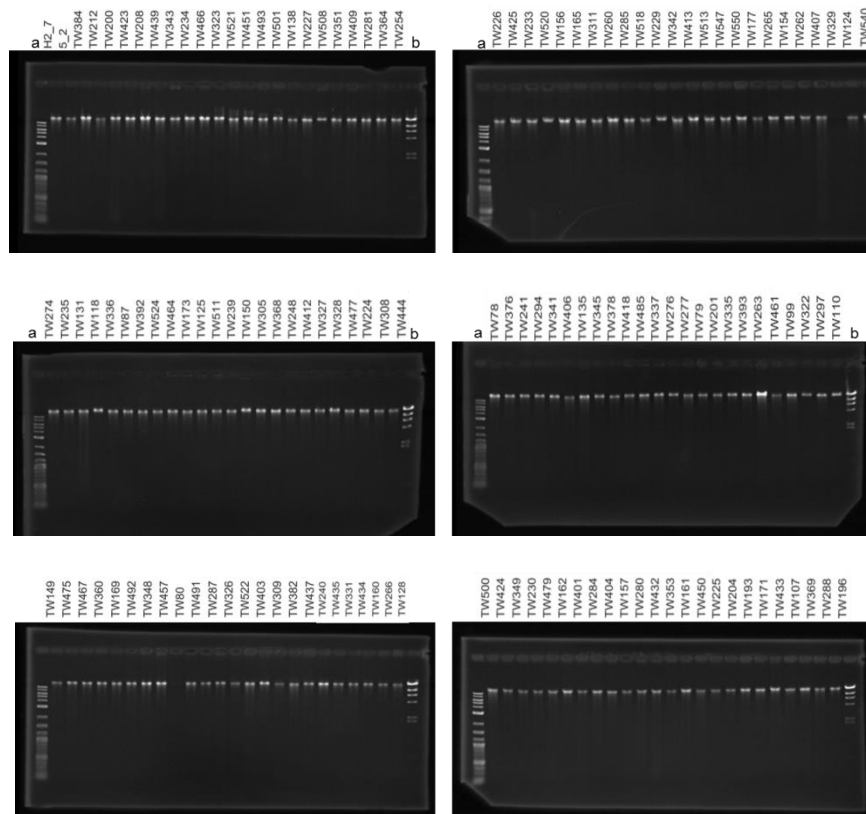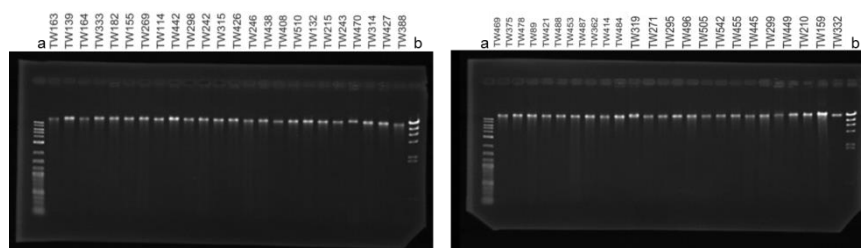
Figure 32: Quality control of SRSZ strains and two parental strains SRS2_ H2-7 and SRZ1_5-2. QC was performed through agarose gel electrophoresis and following DNA visualization through ethidium bromide staining. DNA ladders used: a, Quick Load 2-Log DNA Ladder Mix and b, Lambda Hind III. For the strains showing no band or poor-quality band, genomic DNA isolation was repeated.

Sequencing of the genomes was carried out by GATC (Eurofins GATC Biotech GmbH, Germany) using Illumina NextSeq high-output (Illumina HiSeq 125 bp paired end) platform. For the parental genomes, a total of 6,258,090 and 5,044,818 reads (paired reads, ~ 100 x depths) were generated and the read length was 126nts and 126 nts for SRS2_ H2-7 and SRZ1_5-2 respectively. Discarding low quality reads resulted in 6,243,200 and 5,009,390 reads for SRS2_ H2-7 and SRZ1_5-2 respectively which corresponds to approximately 99.5x sequence depth on average.

These reads were assembled de novo using CLC genomics workbench 7.5 (minimum contig length= 1000, mismatch cost= 2, insertion cost= 3, deletion cost= 3, length fraction= 0.5, similarity fraction= 0.8). The assembly process resulted in 525 contigs with total consensus genome size of 18,394,268 bp for SRS2_ H2-7 and 619 contigs with total consensus genome size of 18,343,169 bp for SRZ1_5-2 (at mismatch cost=2, insertion cost=3, deletion cost=3, length fraction=0.5 and similarity fraction=0.8). The maximum contig size was 266,546 bp, and minimum contig size was 1,018 bp (including scaffolded regions) for SRS2_ H2-7 and the maximum contig size was 217,849 bp, and minimum contig size was 1,016 bp (including scaffolded regions) for SRZ1_5-2. The average size of the contig was 35,036 bp and the N50 value of contigs was 59,840 Kb for SRS2_ H2-7. The average size of the contig was 29,633 bp and the N50 value of contigs was 54,118 Kb for SRZ1_5-2 (**Table 8**). Previously the genome size of SRZ2_ 5-1 genome was reported to be 18.7 Mb (coverage 97%) [7]. The genome of *S. reilianum* (SRZ2_ 5-1) is organized in 23 chromosomes, of which chromosome 1 being the longest in size and chromosome 23 the shortest.

Table 8: General statistical features of SRS2_ H2-7 and SRZ1_5-2 genomes after assembly using CLC genomics workbench 7.5.

Summary statistics	SRS H2-7	SRZ 5-2
Total reads	21,635,738	9,864,489
Average read length (bp)	123.19	123.21
Total genome size (bp)	18,394,268	18,343,169
Average contig length (bp)	35,036	29,633
Number of contigs	525	619
N50	59,840	54,118

3.2.2   <u>Chromosomal regions can be unambiguously assigned for all SRSZ strains</u>

We wanted to assess the parental origin of chromosomal regions of all the 189 SRSZ stains. CLC genomics workbench 7.5 was used for all the analysis hereafter. These strains had mosaic genomes (parts coming from both the parents SRS2_ H2-7 and SRZ1_5-2). In order to check the regions originating from both the parents and make an association to the virulence phenotype, the mapping of the SRSZ strains to the parental genomes needed to be done. However, the assembled parental genomes of SRS2_ H2-7 and SRZ1_5-2 were of not so high quality (N50 values being 59,840 and 54,118 respectively). The assembly and annotation of two sister genomes of SRS2_ H2-7 and SRZ1_5-2 namely SRS1_ H2-8 and SRZ2_ 5-1 were done previously [7]. Due to this availability, we were able to map the reads of SRSZ strains against sister SRS1_ H2-8 and SRZ2_ 5-1 reference genomes. The genomes of the original parents to the sister genomes are highly similar (Table 9).The standalone mapping of SRS2_ H2-7 reads to SRS1_ H2-8 as reference showed 23.20% reads as not mapped and mapping of SRZ1_5-2 reads to SRZ2_ 5-1 as reference showed 3.25% reads as not mapped (Mismatch cost=2, insertion cost =3, deletion cost= 3, length fraction= 0.5, similarity fraction= 0.8 and ignorance= non-specific matches, global alignment) (**Table 9**). 76.80% of the reference was covered by SRS2_ H2-7 reads, in which 23.13% (616,389,637 bases in 18.52 Mbp) of bases consisted of unknown se-

quence "N". Similarly, 96.75% of the reference was covered by SRZ 5-2 reads, in which 2.86% (34,792,662 bases in 18.476 Mbp) of bases consisted of unknown sequence "N".

Table 9: Mapping statistics of standalone mapping of SRS2_ H2-7 reads to SRS1_ H2-8 as reference and of SRZ1_5-2 reads to SRZ2_ 5-1 as reference.

Mapping statistics	SRS2_ H2-7 against SRS1_ H2-8	SRZ1_5-2against SRZ2_ 5-1
Number of mapped reads	16,616,984	9,544,082
Mapped reads(%)	76.80%	96.75%
Unmapped reads(%)	23.20%	3.25%

We can see from Figure 34A and Figure 35A that mapping of all 23 chromosomal reads of SRS2_H2-7 and SRZ1_5-2 showed very high coverage (93.53% and 97.75% respectively) against SRS1-H2-8 and SRZ2_5-1 respectively. This result showed the sufficient similarity to use already annotated SRS1_ H2-8 and SRZ2_ 5-1 genomes as reference genome for all the SRSZ strains although they were originated from a cross between SRS2_ H2-7 x SRZ1_5-2. The genomes of SRS1_ H2-8 and SRZ2_ 5-1 were combined and used as single reference file to map all the reads from SRSZ strains. This means reads from a single SRSZ strain was mapped against a combined SRS1_ H2-8 and SRZ2_ 5-1 reference.

As a starting point for the mapping analysis, we began with the mapping of SRS2_ H2-7 and SRZ1_5-2 against combined reference (**Figure 33**). The mapping was performed where the reads were considered as transcripts and an RNAseq analysis was done against the combined annotated reference (at mismatch cost=2, insertion cost=3, deletion cost=3, length fraction=0.98 and similarity fraction=0.98, global alignment). This was done to mimic the reads as transcripts that should give us the genomic loci or gene identification easily once we export the data out of CLC genomics workbench.
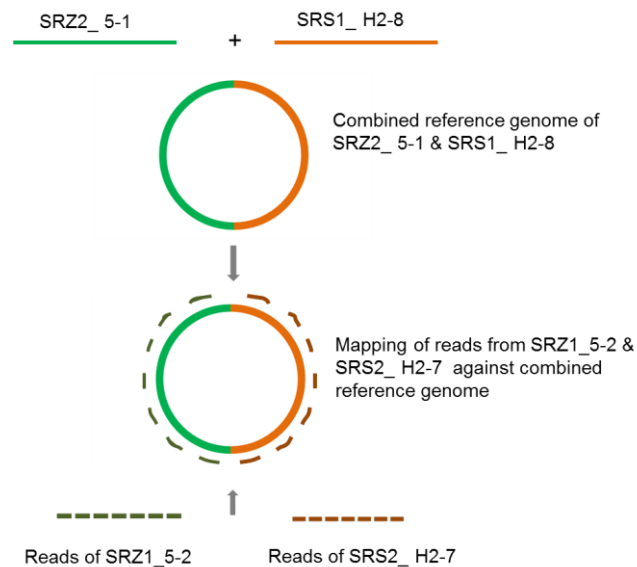
Figure 33: Schematic representation of mapping of reads from of SRS2_ H2-7 and SRZ1_5-2 against combined reference of SRS1_ H2-8 and SRZ2_ 5-1. In ideal case 50% of the reference should be covered by reads from one strain and 50% by reads from the second strain (excluding intergenic region).

Two kinds of parameters were chosen, in first case the intergenic regions were considered whereas in the next the intergenic regions were unmarked. The result showed that in case of SRS2_ H2-7 (including intergenic region), 93.53% of the reads mapped (in pairs and broken pairs) and 6.46% reads didn't map to the combined reference genome (**Table 10**). On the other hand, in case of SRZ1_5-2 (including intergenic region), 97.75% of the reads mapped (in pairs and broken pairs) and 2.26% reads didn't map to the combined reference genome (**Table 11**). When we didn't consider mapping against intergenic region, in case of SRS2_ H2-7 (excluding intergenic region), 50.17% of the reads mapped (in pairs and broken pairs) and 49.83% reads didn't map to the combined reference genome (**Table 10**). In case of SRZ1_5-2 (excluding intergenic region), 47.82% of the reads mapped (in pairs and broken pairs) and 52.18% reads didn't map to the combined reference genome (**Table 11**).

Table 10: Mapping statistics of mapping of reads from SRS2_ H2-7 against combined reference of SRS1_ H2-8 and SRZ2_ 5-1 genomes.

Mapping statistics	SRS2_ H2-7 against combined reference (including intergenic region)		SRS2_ H2-7 against combined reference (excluding intergenic region)	
Reads mapped in pairs	18,392,086	85.21%	7,968,058	36.92%
Reads mapped in broken pairs	1,796,811	8.32%	2,860,711	13.25%
Reads not mapped	1,395,219	6.46%	10,755,347	49.83%

Table 11: Mapping statistics of mapping of reads from SRZ1_5-2 against combined reference of SRS1_ H2-8 and SRZ2_ 5-1 genomes.

Mapping statistics	SRZ1_5-2 against combined reference (including intergenic region)		SRZ1_5-2 against combined reference (excluding intergenic region)	
Reads mapped in pairs	8,988,160	91.48%	3,465,630	35.27%
Reads mapped in broken pairs	615,960	6.27%	1,233,145	12.55%
Reads not mapped	221,650	2.26%	5,126,995	52.18%

The figure following shows mapping graph of all the 23 chromosomes. We could see clearly that the reads from SRS2_ H2-7 showed higher coverage against SRS1_ H2-8 chromosomes and not against SRZ2_ 5-1 (**Figure 34**). Likewise the reads from SRZ1_5-2

showed higher coverage against SRZ2_ 5-1 chromosomes and not against SRS1_ H2-8 chromosomes (**Figure 35**).

Figure 34: Mapping graph of all the 23 chromosomes of SRS2_ H2-7 against combined reference of SRS1_ H2-8 and SRZ2_ 5-1 genomes. **A**, the reads showed very high coverage against the 23 chromosomes of SRS1_ H2-8 but **B**, showed no/very little coverage against 23 chromosomes of SRZ2_ 5-1. The coverage graph is colored in blue and was set to the same scale for all chromosomes.
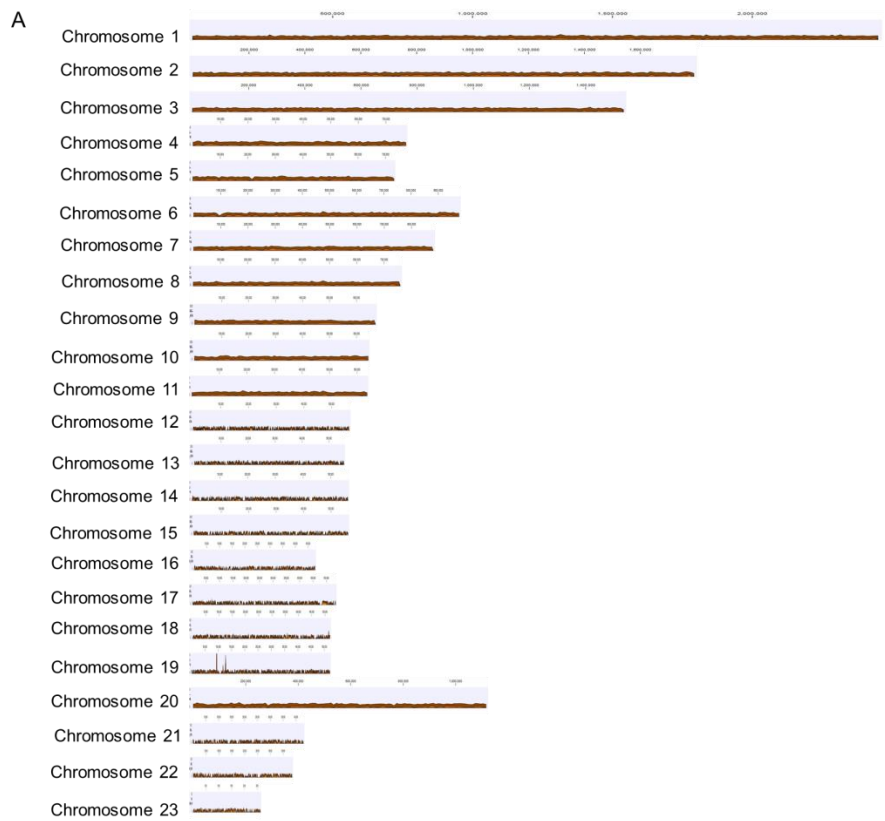
Figure 35: Mapping graph of all the 23 chromosomes of SRZ1_5-2 against combined reference of SRS1_ H2-8 and SRZ2_ 5-1 genomes. **A**, the reads showed very high coverage against the 23 chromosomes of SRZ2_ 5-1 but **B**, showed no/very little coverage against 23 chromosomes of SRS1_ H2-8. The coverage graph is colored in orange.

Afterwards all the 189 SRSZ strain reads were mapped individually against the combined reference (**Figure 36**).

Figure 36: Schematic representation of mapping of reads of SRSZ strains against combined reference of SRS1_ H2-8 and SRZ2_ 5-1. SRSZ genomes are mosaic having regions from both the parents in various ratios. The reads obtained should map against the parental genomes of one of the parents for each region.

Since the reads were considered as transcripts and RNAseq analysis was performed, the output was in the form of expression value, RPKM, unique gene reads, and total gene reads (**Figure 37**). The unique gene reads were exported as excel sheet out of CLC genomics work-bench 7.5 for further analysis.

Name	Chromosome	Gene number	Expression value	Gene length	RPKM	Unique gene reads	Total gene reads
mRNA:sr00846.2	Chr.01	1	59	940	42.940	59	59
mRNA:sr00845	Chr.01	2	122	1134	73.601	122	122
mRNA:sr00843	Chr.01	3	151	1413	73.109	151	151
mRNA:sr00842.2	Chr.01	4	84	993	57.872	84	84
mRNA:sr00841.2	Chr.01	5	96	1062	61.842	96	96
mRNA:sr00840.2	Chr.01	6	123	1325	63.507	123	123
mRNA:sr00838.2	Chr.01	7	321	2506	87.632	321	321

Figure 37: An example of a part of an excel sheet made from mapping of a SRSZ strain against combined parental reference. The reads from SRSZ were considered as transcripts and a transcriptome analysis was performed. This gave the genomic loci of the reference in the form of expressed transcripts in the very first column in left side. This was followed by the chromosome numbers arranged linearly. Expression values showed the coverage of the reads against that specific transcript region which is basically the unique gene reads.

The result showed that parental origin of chromosomal regions could be unambiguously assigned for all SRSZ strains. This means, in those regions of the chromosome where the reads of one SRSZ strain mapped against one of the reference genome had null or very less reads mapping to the other reference genome (**Figure 38**, **Figure 39**).

Name	Chromosome	Gene number	Gene length	Unique gene reads H2-7	Unique gene reads 5-2	Unique gene reads TW78	Unique gene reads TW79
sr00846.2	5-1-Chr.01	1	940	0	211	262	267
sr00845	5-1-Chr.01	2	1134	0	188	226	222
sr00843	5-1-Chr.01	3	1413	0	248	346	302
sr00842.2	5-1-Chr.01	4	993	0	195	199	253
sr00841.2	5-1-Chr.01	5	1062	0	171	232	274
sr00840.2	5-1-Chr.01	6	1325	0	232	310	321
sr00838.2	5-1-Chr.01	7	2506	0	397	422	481
sr20011	5-1-Chr.01	8	852	0	112	149	146
sr06455	5-1-Chr.01	9	1236	0	185	229	259
sr06454	5-1-Chr.01	10	282	0	73	98	100
SRS1_00846	H2-8-chr01	1	943	475	0	0	0
SRS1_00845	H2-8-chr01	2	1164	513	2	0	1
SRS1_00843	H2-8-chr01	3	1413	553	0	0	0
SRS1_00842	H2-8-chr01	4	993	406	0	0	2
SRS1_00841	H2-8-chr01	5	1062	378	0	0	0
SRS1_00840	H2-8-chr01	6	1325	598	0	0	1
SRS1_00838	H2-8-chr01	7	2470	848	0	0	0
SRS1_20011	H2-8-chr01	8	852	284	0	0	0
SRS1_06455	H2-8-chr01	9	1236	506	0	0	0
SRS1_06454	H2-8-chr01	10	279	170	0	0	0

Figure 38: An example of two parts of an excel sheet made out of mapping of SRSZ strains against combined parental reference. The unique gene reads can be seen for two parents SRS2_ H2-7 and SRZ1_5-2 along with two SRSZ strains TW78 and TW79 as examples. The complete sheet has all the 189 SRSZ strains. As can be seen, SRS2_ H2-7 showed no coverage against parental SRZ2_ 5-1 (in upper half) but showed high coverage against SRS1_ H2-8 (lower half). The reverse was true for SRZ1_5-2. For SRSZ strains the parts shown in the example showed high coverage against SRZ2_ 5-1. This means genomic regions of the SRSZ strains can be assigned exclusively to one of the parents for a particular genomic region.
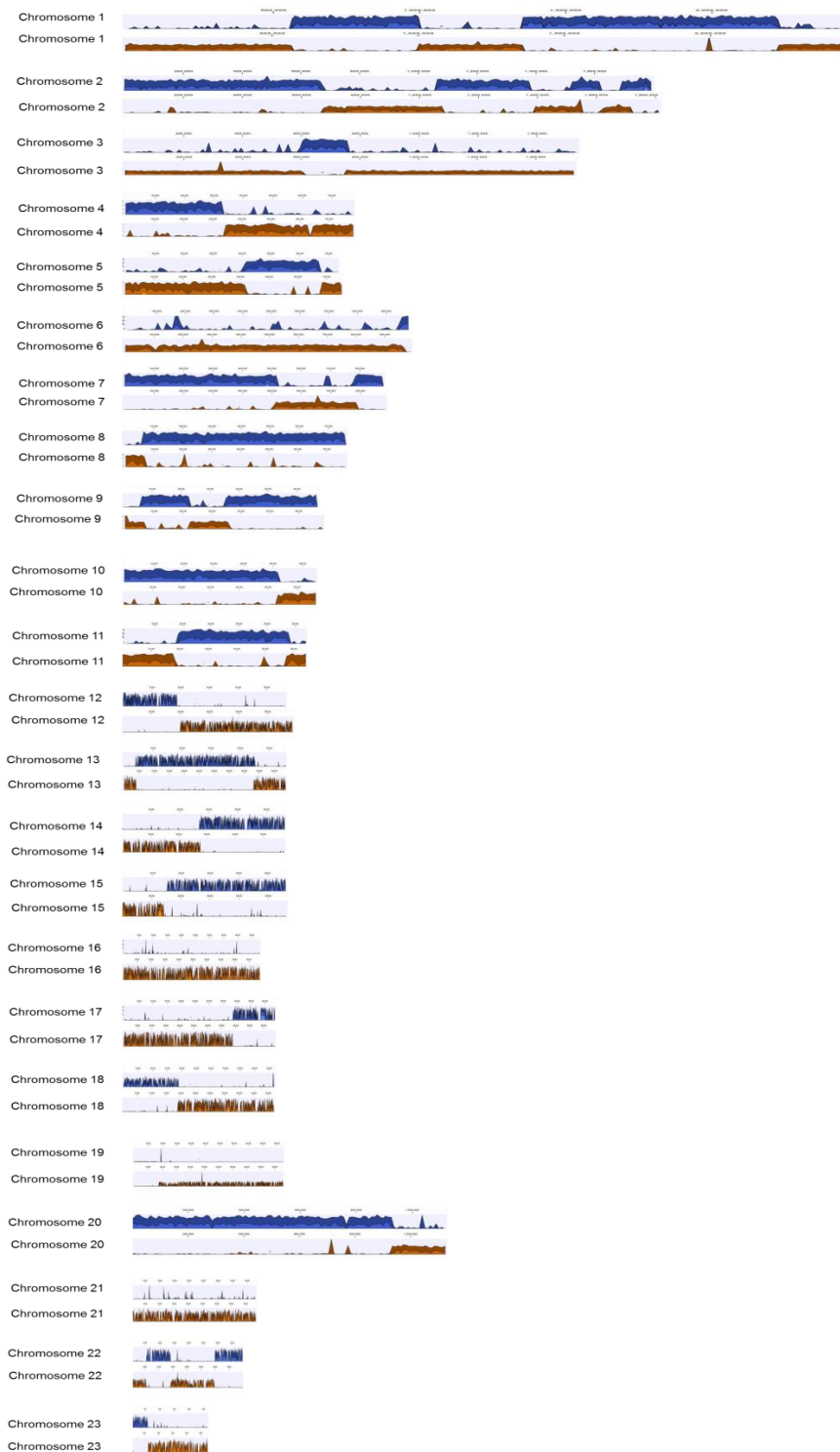
Figure 39: Representative mapping graph of TW260 SRSZ strain. All the 23 chromosomes mapped against combined reference of SRS1_ H2-8 and SRZ2_ 5-1 genomes. As can be seen for each chromosome the region that is showing high coverage against one parent shows no or

very less coverage to the other parent (blue color showing coverage against SRZ2_ 5-1 and orange color showing coverage against SRS1_ H2-8 parent) depicting cross over points clearly.

3.2.3 <u>Many strains contain partially duplicated genomic regions</u>

For few SRSZ strains it was observed that they contained chromosomal regions from both the parents. This was observed for strains TW139, TW287, TW285, TW233 and TW230. In case of TW139 as can be seen from the following **Figure 40**, at chromosome 16 (and/or chromosome 15) the offspring showed high coverage against both the parents suggesting that this chromosome is present in two copies coming from both.
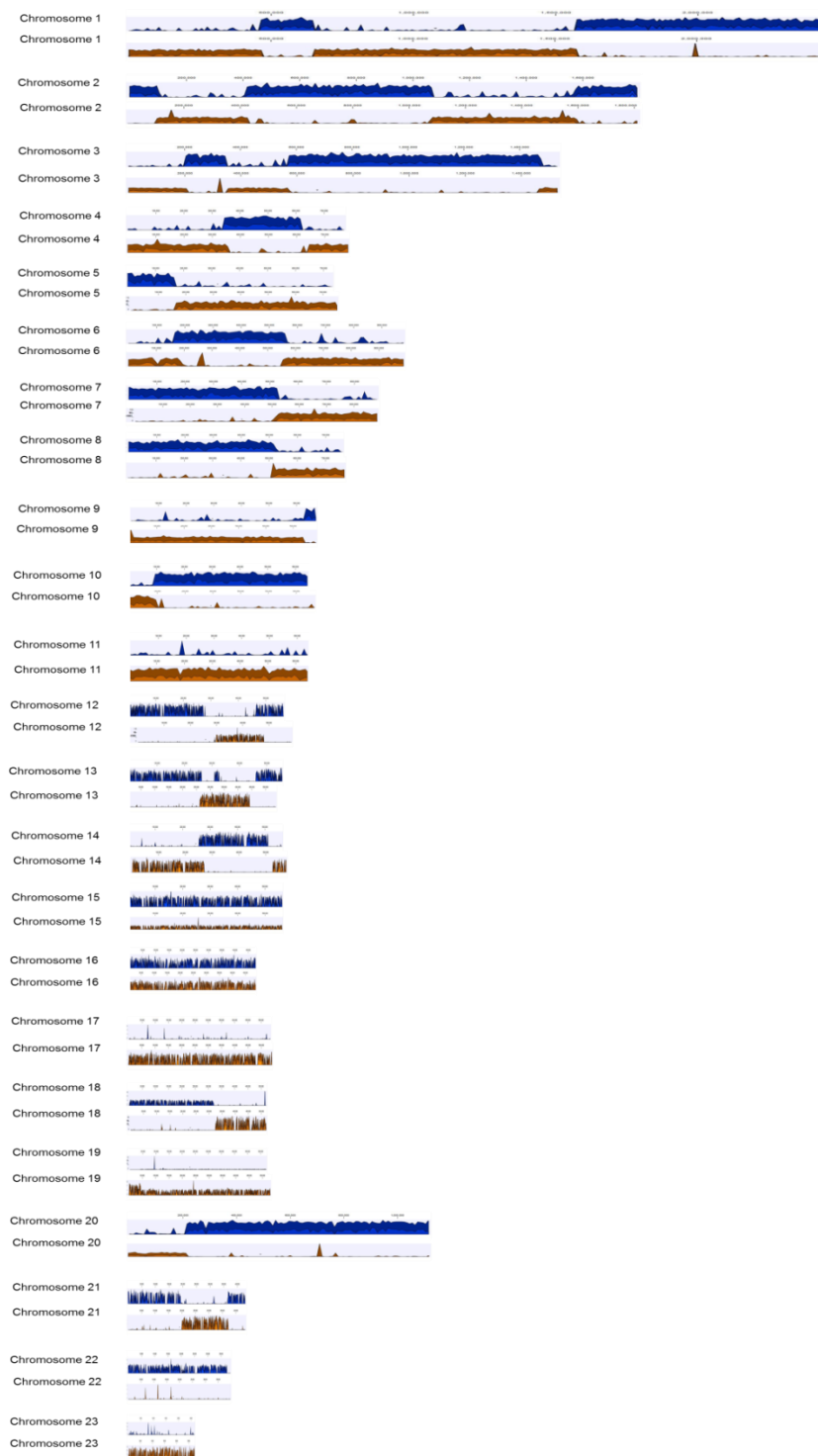
Figure 40: Representative mapping graph of TW139 SRSZ strain. All the 23 chromosomes mapped against combined reference of SRS1_ H2-8 and SRZ2_ 5-1 genomes. As can be seen for each chromosome the region that is showing high coverage against one parent shows no or no or very less coverage to the other parent (blue color showing coverage against SRZ2_ 5-1 and orange color showing coverage against SRS1_ H2-8 parent) depicting cross over points

clearly. However, at chromosome 16 the reads mapped against both the parental genomes suggesting presence of duplicated region in the offspring.

3.2.4 Correlation of phenotype with parental origin of genomic loci reveal a 'region of interest' (ROI)

By associating virulence analysis of the SRSZ strains on sorghum with parental origin of the genomic regions, a region of interest (ROI) covering 35 genes in the left arm of chromosome 7 was detected (**Figure 41**). These 35 genes were found to be of SRS1_ H2-8 origin in 77 out of 79 sorghum spore producing SRSZ offspring. This means the reads from these strains showed high coverage at this region against SRS1_ H2-8 parent and no coverage against SRZ2_ 5-1 parent. This suggested its importance as a genomic region needed for SRS forma specialis to produce spores in sorghum. Two strains (TW243 and TW451) however showed no coverage against SRS1_ H2-8 at this position despite being able to produce spores in sorghum; instead they showed read coverage against SRZ2_ 5-1 parent (**Figure 41**).

~70 Kb   ~0.60 Mb

Length of chromosome 7

TW421
TW131
TW78
TW319
TW233
TW466
TW239
TW404
TW378
TW89
TW427
TW309
TW132
TW484
TW110
TW522
TW315
TW345
TW332
TW432
TW375
TW392
TW208
TW299
TW276
TW159
TW163
TW226

~70 bp

~0.60 Mb

Length of chromosome 7

	TW329
	TW212
	TW210
	TW171
	TW285
	TW540
	TW521
	TW488
	TW479
	TW469
	TW461
	TW343
	TW311
	TW241
	TW177
	TW173
	TW161
	TW118
	TW470
	TW444
	TW230

~70 bp   ~0.60 Mb

Length of chromosome 7

Figure 41: The mapping graphs of 79 SRSZ strains producing spores in sorghum at chromosome 7. The strains showed high read coverage against SRS1_ H2-8 parent at a region which is on extreme left and consisted of 35 genes. This region was identified as potential region of interest (ROI) and colored in orange. 2 SRSZ strains (TW243 and TW451) producing spores in

sorghum showed no read coverage against SRS1_ H2-8 parent as against all other virulent strains colored in green.

It was observed that out of the 35 genes in the ROI, there was a region of 9 genes which exclusively showed high coverage against SRZ2_ 5-1 in 87% in the avirulent SRSZ offspring (96 out of 110 healthy SRSZ strains on sorghum) (

Figure 42). We also did not observe any recombination event inside this region. This strengthened our hypothesis of involvement of this region exclusively in spore formation in sorghum for *formae specialis* SRS and particularly the region of 9 genes.
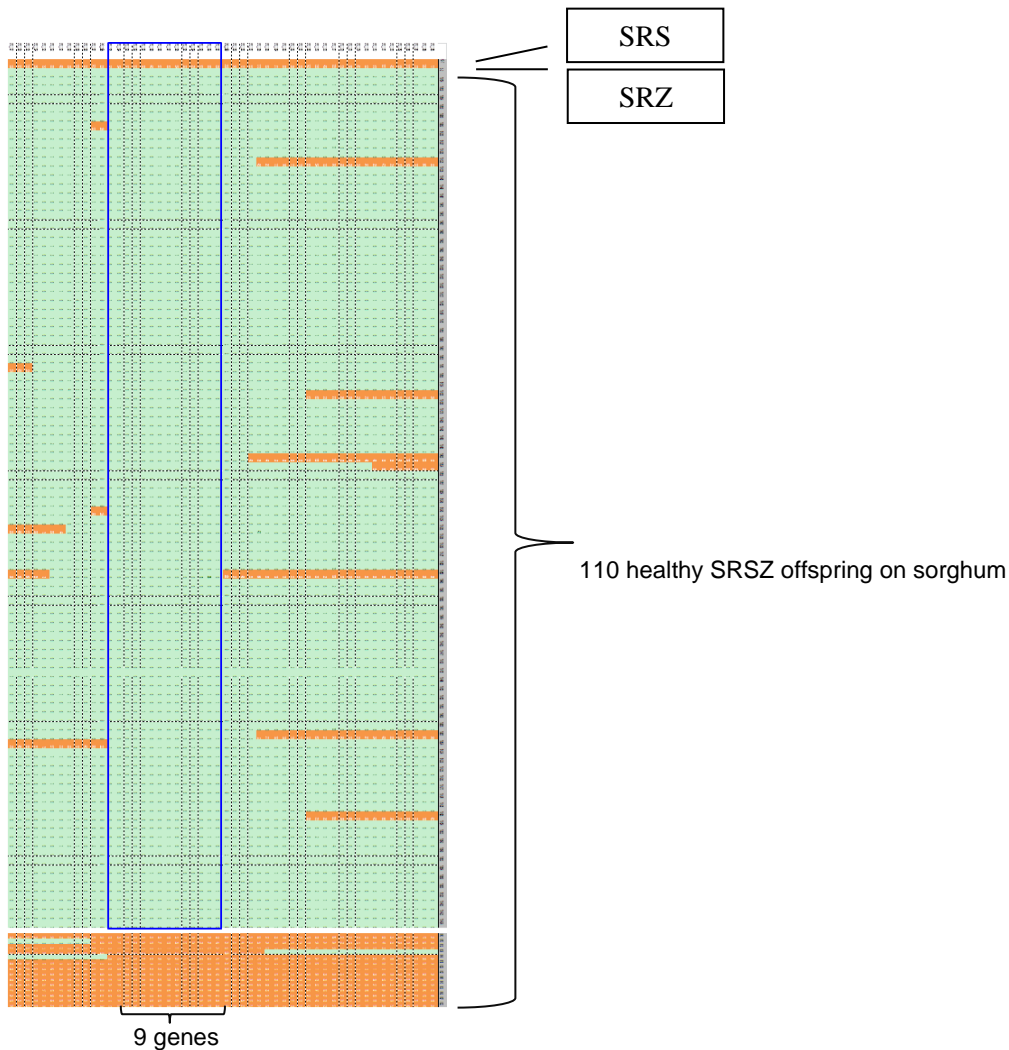
Figure 42: Coverage analysis of 110 SRSZ strains avirulent in sorghum. Out of all the 35 genes of ROI, 9 genes showed coverage exclusively against SRZ2_ 5-1 (green area inside the blue box). 14 SRSZ strains avirulent in sorghum showed no read coverage against SRZ2_ 5-1 parent as against all other avirulent strains colored most part in orange.

3.3 *In silico* analysis and functional characterization

3.3.1 A cluster of genes within ROI shows up as a 'diversity cluster'

We wanted to further investigate the nature of the 9 proteins that we have identified as potential candidate for virulence in sorghum thereby making them hot spot region for host specificity. These 9 genes were previously identified as a part of a divergence cluster and was named cluster 7_11 [7]. Hereafter we refer to this region as cluster 7_11.

Fungal effectors are known to play the most important role as virulence factors. There are various properties of these effector proteins that have been described in many researches over the years. One such property is "effector evolution". Effector genes are constantly under selection pressure to evade plant immunity [16]. It is very likely that they are one of the least conserved regions in the genome. To check this, we compared the amino acid identity of all the 35 proteins (that also includes cluster 7_11) in the ROI. Another important feature of effector like proteins is that they are "secretory in nature" as they have to be delivered effectively into the host plant [16]. We checked the secretory nature of the 35 proteins in ROI.

3.3.1.1 Gene-by-gene comparison of the diversity cluster 7_11 between SRS and SRZ

All the 9 genes in the cluster 7_11 are uncharacterized proteins and have homologs in *Ustilago maydis* (**Table 12**). They are syntenic in nature between SRS and SRZ. There is however present one transposase in SRZ in between genes 13898 and 13900 and named sr13899. This gene is found in chromosome 21 in SRS named srs25049 (**Figure 43**).

Table 12: Names and annotation of 9 genes within ROI that was detected as exclusively associated with spore formation for forma specialis SRS in sorghum

SRZ genes	Annotation	SRS gene	Annotation	*U. maydis* homolog of *S. reilianum* gene	Annotation *U. maydis* gene
sr13897	conserved hypothetical Ustilaginaceae_specific protein [*Sporisorium reilianum* SRZ2]	srs_13897	uncharacterized protein SRS1_13897	um02851	putative protein
sr13898	conserved hypothetical Ustilaginaceae_specific protein [*Sporisorium reilianum* SRZ2]	srs_13898	uncharacterized protein SRS1_13898	-	putative protein
sr13900	hypothetical protein sr13900	srs_13900	uncharacterized protein SRS1_13900	-	-
sr13901	uncharacterized protein SRS1_13901	srs_13901	hypothetical protein sr13901	-	-
sr13902	Conserved hypothetical Ustilaginaceae_specific protein	srs_13902	uncharacterized protein SRS1_13902	um02851	putative protein
sr13903	conserved hypothetical Ustilaginaceae_specific protein	srs_13903	uncharacterized protein SRS1_13903	um02852	conserved hypothetical Ustilago-specific protein
sr13904	conserved hypothetical Ustilaginaceae_specific protein	srs_13904	uncharacterized protein SRS1_13904	um02853	conserved hypothetical Ustilago-specific protein
sr13905	conserved hypothetical Ustilaginaceae_specific protein	srs_13905	uncharacterized protein SRS1_13905	um02854	hypothetical protein
sr13906	conserved hypothetical Ustilaginaceae_specific protein	srs_13906	uncharacterized protein SRS1_13906	um11484	hypothetical protein
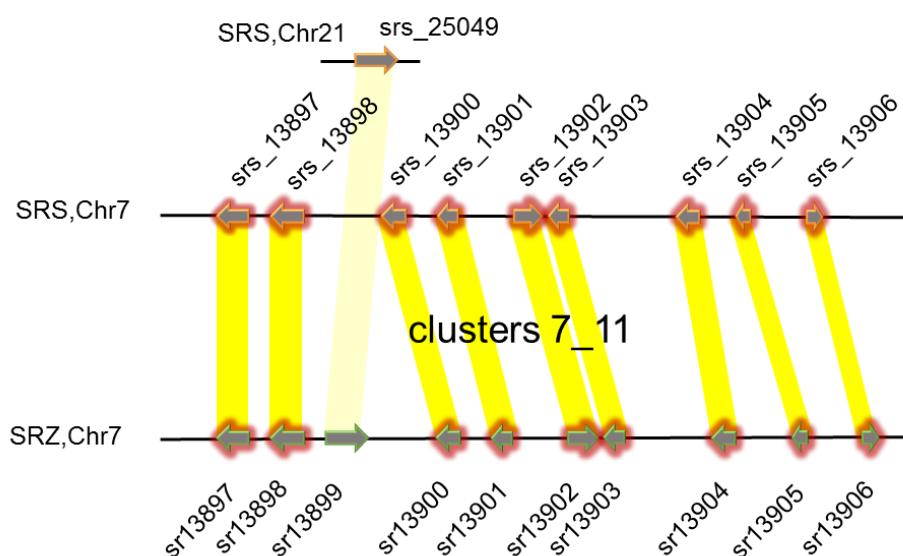
Figure 43: Gene-by-gene comparison of the diversity cluster 7_11 between SRS and SRZ. All the genes are syntenic in nature (shown by yellow lines). The sr13899 is a transposase and is present in chromosome 21 in case of SRS.

3.3.1.2   Secretory nature of the proteins of cluster 7_11

To analyze the nature of the proteins in the ROI we first checked if they were secretory in nature. Several bioinformatics tools were applied (seen below). Region of cluster 7_11 exclusively showed secretory nature of the proteins. They all passed the measures used to qualify them as probable secreted proteins (SP).

Initially we used Signal Peptide Prediction tool SignalP (see methods section for details). The D-score (discrimination score) cut off was 0.43. As can be seen from the **Figure 44** below, all the 9 genes showed value more than the cut off suggesting that they are secretory in nature. In addition, gene 13888 showed values higher than cut off.

D value — legend: SRZ, SRS

	13875	13876	13877	13878	13879	13880	13881	13882	13883	13884	13885	13886	13887	13888	13889	13890	13891	13892	13893	13894	13895	13896	13897	13898	13900	13901	13902	13903	13904	13905	13906	13907	13908	13909	13910	13911	13912
Phobius SP SRS	0	0	Y	0	0	0	0	0	0	0	Y	0	Y	0	Y	0	0	0	0	0	0	0	Y	Y	Y	Y	Y	Y	Y	Y	Y	0	0	0	0	0	0
Phobius SP SRZ	0	0	Y	0	0	0	0	0	0	0	Y	0	Y	0	Y	0	0	0	0	0	0	0	Y	Y	Y	Y	Y	Y	Y	Y	Y	0	0	0	0	0	0
Phobius TM SRS	0	0	0	1	0	0	0	0	0	0	0	0	0	6	1	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0
Phobius TM SRZ	0	0	0	1	0	0	0	0	0	0	0	0	0	6	1	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0
tmhmm TMH SRS	0	0	0	1	0	0	0	0	0	0	0	0	0	6	1	1	0	0	2	0	0	0	0	1	0	0	1	0	1	0	1	0	0	0	0	1	0
tmhmm TMH SRZ	0	0	0	1	0	0	0	0	0	0	0	0	0	6	1	1	0	0	2	0	0	0	0	1	0	0	1	0	1	0	2	0	0	0	0	1	0

Figure 44: *In silico* analysis to show secretory nature of the proteins of cluster 7_11. The analysis was performed for all 35 proteins in ROI. The 9 proteins of the cluster stood out as most significant. The SignalP tool showed D-value higher than 0.43 for them. Phobius tools predicted the 9 proteins to be secretory in nature.

In addition, other *in silico* prediction tools like Phobius and tmhmm results are shown in the figure. Phobius tool was used to analyze whether the proteins have signal peptide. It showed all the cluster 7_11 proteins to be yes ("Y") for SP. This tool also predicted transmembrane helicases. For a cut off value of $</=2$, all the 9 proteins of the cluster 7_11 showed value less than or equal to 2. tmhmm program is for prediction of transmembrane helices in proteins. This tool showed similar result as in Phobius. This combined result confirmed that the proteins of cluster 7_11 are secretory.

WolFPsort tool was used for protein subcellular localization prediction and showed that all the 9 proteins in the cluster are extracellular ("Extr major") in nature. LocSigDB tool was used to check whether they have any ER retention signal. For all the 9 proteins we saw no ER retention signal. Finally, Nucpred tool was used to check if the protein spends at least some time in the nucleus or no time in the nucleus. If the Nucpred score is less than 0.8, they don't spend time in nucleus. For all the 9 proteins the values were found to be less than 0.8.

3.3.1.3   Diversity in terms of amino acid identity in cluster 7_11

A comparison of amino acid sequences between SRS and SRZ in ROI showed that there was a region of 8 proteins that were less conserved as according to SIMAP values with respect to surrounding proteins (**Figure 45**). The SIMAP values were procured from the data of - Getrud Manhaupt, Ulrich Güldner and Martin Münsterkötter.
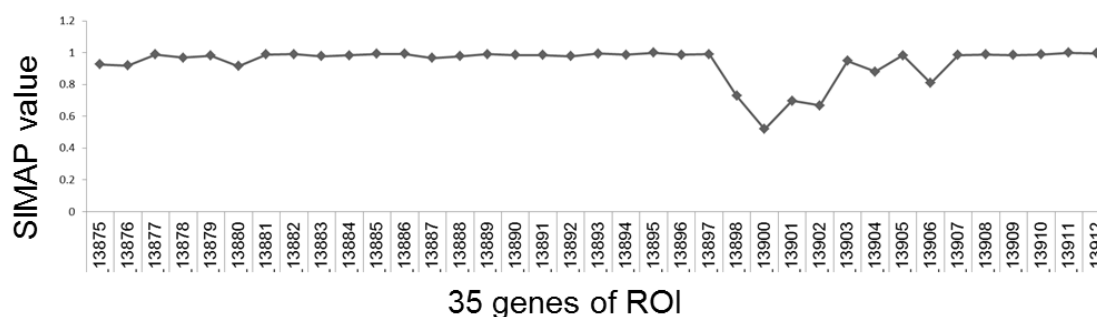
Figure 45:   Comparison of SIMAP values to check amino acid identity between SRS and SRZ proteins falling under ROI. 8 proteins showed less conservation and they fall under the region we found as potentially associated with virulence in sorghum. The SIMAP values were procured from the data of - Getrud Manhaupt, Ulrich Güldner and Martin Münsterkötter.

3.3.2   The genes in the cluster 7_11 are highly expressed *in planta* at 3 dpi in sorghum

Another aspect of considering small proteins as putative effectors is their *in planta* expression. They are predicted to show high expression level in the early stages of infection. An RNA seq was performed before I joined the lab. Upon careful analysis of the cluster7_11 genes; it was found that the expression levels of the genes in the cluster were relatively higher as compared to other genes in the ROI (in sorghum at 3 dpi) (**Figure 46**). In particular, the genes SRS-13898, 13901, 13904 13905 and 13906 and SRZ-13898, 13901, 13902, 13904, 13905 and 13906 were seen to show higher expression values.
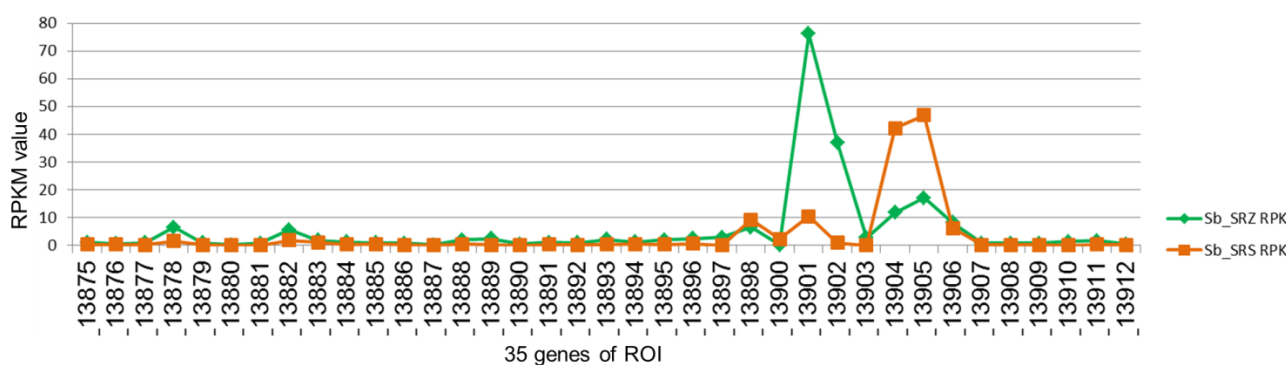
Figure 46: *In planta* expression analysis of the genes in ROI. As seen from RPKM values comparison, the cluster genes to be relatively higher expressed 3 dpi in sorghum. The RNA seq was performed before I joined the lab and is part of the doctoral thesis of Alana Poloni.

3.3.3 Deletion of cluster 7_11 in SRS

To functionally characterize the role of the cluster 7_11 in pathogenicity we decided to delete this cluster in SRS background and subsequently to do a virulence analysis on sorghum. Deletion of the cluster was achieved by replacement of the 9 genes in the cluster by a hygromycin resistance cassette achieved through homologous recombination. The transformants surviving on selection plates were verified by PCR (**Figure 47**). For this experiment, primer pair oNB010 & oNB024 was used to amplify left flank and primer pair oNB011 & oNB023 was used to amplify right flank of the cluster (see 2.2.12 section). Both the flanks were digested by *Sfi*I enzyme. The hygromycin resistance cassette was excised out using *Sfi*I enzyme from the vector pBS-hhn. Three fragments were ligated and the final product was amplified using nested primer pair oNB013 and oNB014 (see section 2.2.12). The resultant ligated product is called deletion construct. This final product was cut out after gel electrophoresis and put into a 200 µl eppendorf tube that has a filter with a hole created at its bottom. This was placed inside another eppendorf tube of 1.5 ml. After 2 min of spin the DNA was collected from the 1.5 ml tube. This DNA construct was then used to transform freshly prepared protoplasts of SRS and SRZ strains. The transformants were selected on Hyg^R plates.
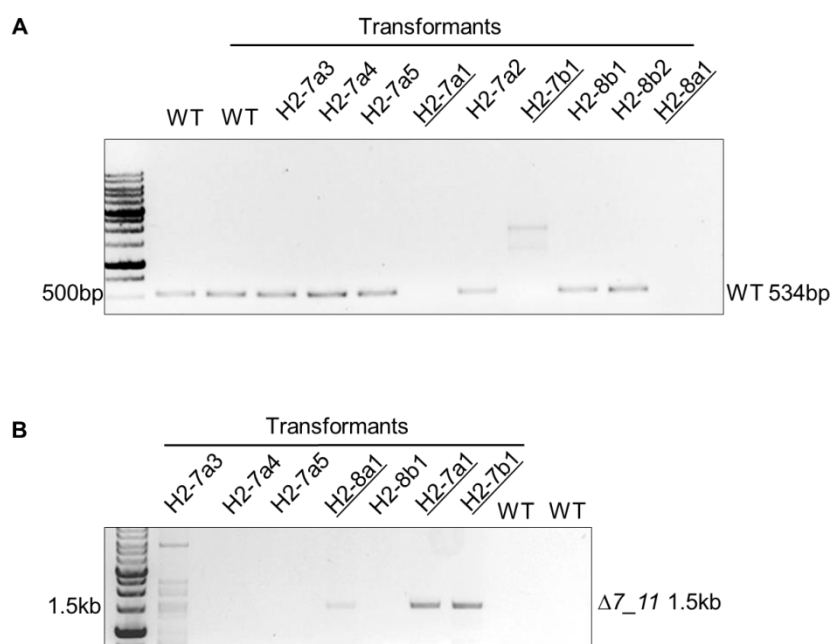
Figure 47: Confirmation of cluster 7_11 deletion mutants by PCR. Colonies from each mating type were picked from plates having hygromycin. The genomic DNA of these mutants was isolated and used as template to check by PCR. **A**, PCR with primer pair oNB017 and oNB018 that amplify a product inside the cluster of size 534 bps. **B**, The samples that did not give a product in the first PCR were used for a second PCR with primer pair oVT022 and oNB012 which amplify the hygromycin cassette and right flank of the deletion construct with a product size of 1.56 kb. WT: SRS2_ H2-7 and SRS1_ H2-8. The positive mutant strains are underlined.

3.3.4 Deletion of cluster 7_11 in SRS abolishes spore formation in sorghum

The created deletion mutants were used to test their capacity to form spores on sorghum. They were found positive for their mating ability on water agar plates. The sorghum plants inoculated with deletion strains were not able to produce any spore on sorghum whereas the WT SRS could successfully form spore (**Figure 48**). Moreover the strains also did not induce phytoalexin formation on sorghum leaves at 4 dpi. Virulence analysis of the cluster 7_11 deletion thus supports of containing effector like genes that are responsible for virulence of SRS *formae speciales* on sorghum.
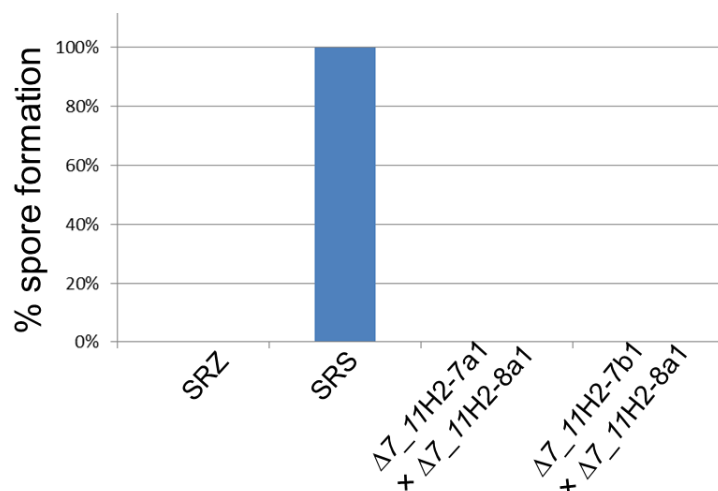
Figure 48: Virulence analysis of *S. reilianum* SRS *formae speciales* cluster 7_11 deletion strains on of Sorghum bicolor 'Tall Polish'. Virulence analysis was assessed for two combinations of mutant's pairs in comparison to WT. % spore formation capacity for SRS is 100% and SRZ is 0%. For the mutants combinations we couldn't observe any spore formation and the result was similar as in SRZ which is 0%. Number of plants used for each combination is 15. This experiment was performed only once.

3.3.5 Structural modeling of the cluster 7_11 proteins

In order to gain insight into the 3D models for proteins of the cluster, structural modeling of the predicted secretory proteins was done using Swiss-Model. Due to very less sequence conservation and small size (effector like property), the identification of structural templates using this tool was restricted. The software could identify only fragments here and there which matched to the proteins structurally. Alignment of the fragments of target sequence and template structure are shown below. For each protein coming from both SRS and SRZ three models were picked depending on the identity and coverage.

1a. Srs_13897

Name	Title	Identity	Method	Oligo state	Ligands
5of3.1.B	DNA primase large subunit PriL	21.88	X-ray, 2.9A	Heterohexamer	None
1l0f.1.A	Beta-lactamase	17.24	X-ray, 1.7A	Monomer	None
1r6a.1.A	Genome polyprotein	11.32	X-ray, 2.6A	Monomer	1×RVP
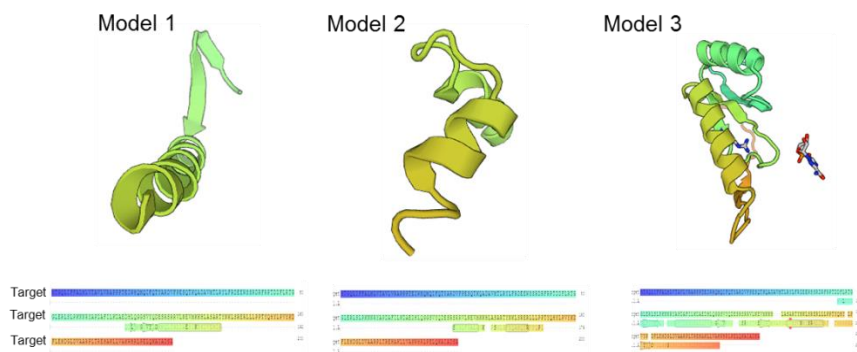
1b. Srz_13897

Name	Title	Identity	Method	Oligo state	Ligands
5of3.1.B	DNA primase large subunit PriL	21.88	X-ray, 2.9A	Heterohexamer	None
5fvl.1.A	Vacuolar protein sorting-associated protein 4	15.56	X-ray, 2.0A	Hetero-dimer	None
1r6a.1.A	Genome polyprotein	11.32	X-ray, 2.6A	Monomer	1×RVP
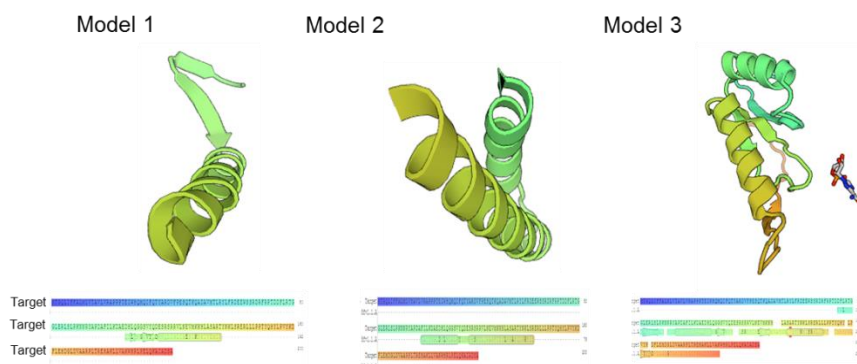
2a. Srs_13898

Name	Title	Identity	Method	Oligo state	Ligands
3csk.1.A	Probable dipeptidyl-peptidase 3	21.28	X-ray, 2.0A	Monomer	1 × ZN
3t6b.1.A	Dipeptidyl-peptidase 3	21.28	X-ray, 2.4A	Monomer	1 × VAL-VAL-TYR-PRO-TRP
2hiy.1.A	Hypothetical protein	20.59	X-ray, 1.4A	Homo- tetramer	None
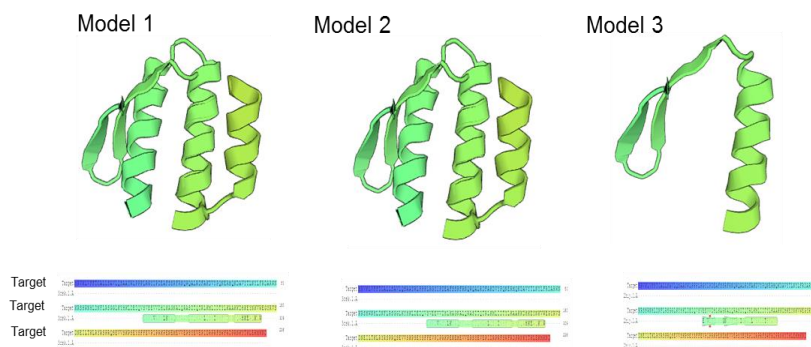
2b. Srz_13898

Name	Title	Identity	Method	Oligo state	Ligands
2b7v.1.A	Ds bRNA-specific editase 1	23.53	NMR	Monomer	None
2l2k.1.A	Adenosine deaminase	23.53	NMR	Monomer	None
2v6y.1.A	AAA family ATPASE, P60 KATANIN	22.50	X-ray, 2.4A	Monomer	1 × STR
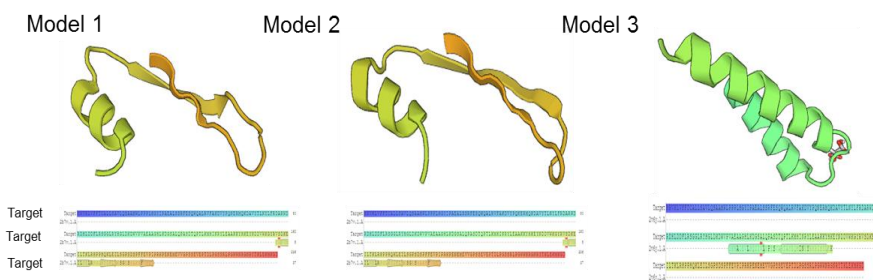
3a. Srs_13900

Name	Title	Identity	Method	Oligo state	Ligands
2jmp.1.A	Chromosomal replication initiator protein dnaA	39.29	NMR	Monomer	None
4pn9.1.A	CC-Hex2	34.48	X-ray, 2.2A	Homo-hexamer	none
2jsw.1.A	Talin-1	25.00	NMR	monomer	None

Model 1   Model 2   Model 3
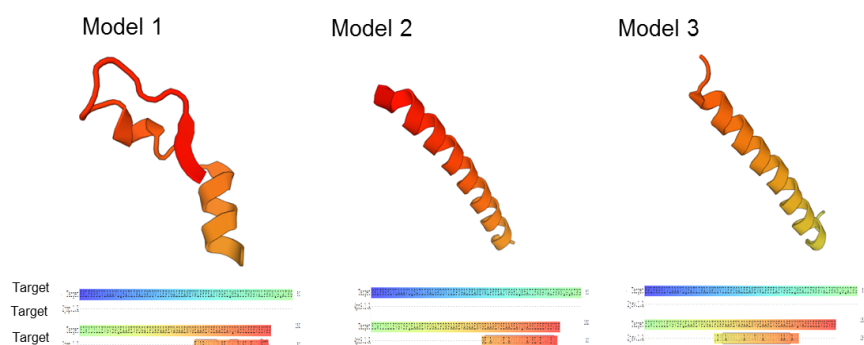
3b. Srz_13900

Name	Title	Identity	Method	Oligo state	Ligands
3ic3.1.B	Putative pyruvate dehydrogenase	22.86	X-ray, 1.8A	Homo-dimer	None
3t55.1.A	Indole-3-glycerol phosphate synthase	15.15	X-ray, 2.1A	Monomer	1x4RG
2ga1.1.A	Protein of unknown function DUF433	12.24	X-ray, 2.0A	Homo-dimer	None

Model 1   Model 2   Model 3
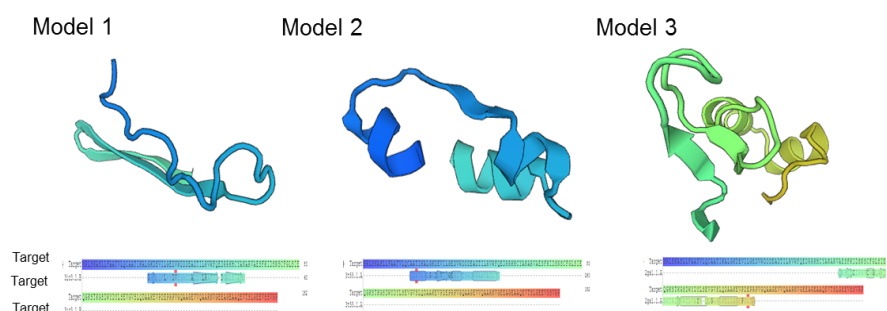
4a. Srs_13901

Name	Title	Identity	Method	Oligo state	Ligands
2vrx.1.B	Inner Centromere Protein A	36.00	X-ray, 1.9A	Hetero-dimer	None
5it7.4.A	KLLA0D06941p	32.26	EM	Monomer	None
5lyb.113.A	60S ribosomal protein L5	29.03	X-ray, 3.2A	monomer	None

Model 1   Model 2   Model 3
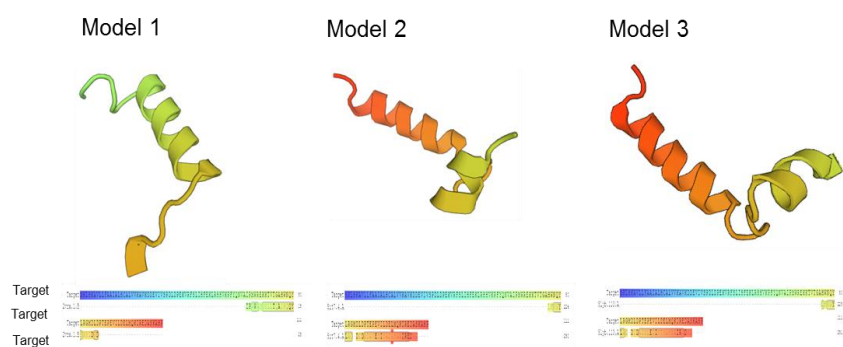
4b. Srz_13901

Name	Title	Identity	Method	Oligo state	Ligands
1ysh.1.F	40S ribosomal protein S13	35.48	EM, 9.5A	Hetero-trimer	None
3jbn.57.A	60S ribosomal protein uL18	26.92	EM	Monomer	None
3jah.4.A	uL18	23.21	EM	monomer	None

Model 1   Model 2   Model 3
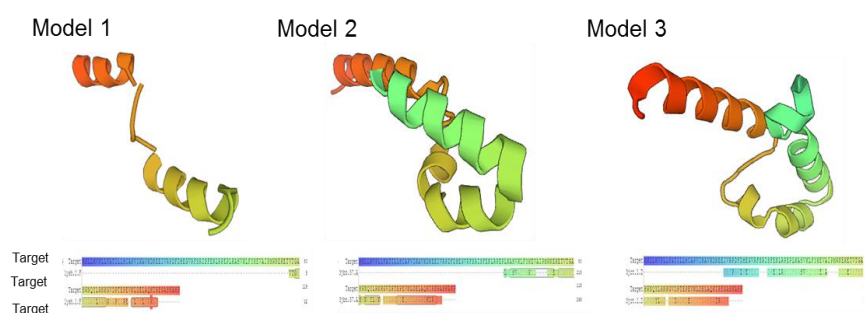
5a. Srs_13902

Name	Title	Identity	Method	Oligo state	Ligands
2vnd.1.A	ENDONUCLEASE I	23.33	X-ray, 1.7A	Monomer	1 × MG
2mmv.1.A	Cell division protein Zap A	20	NMR	Homo-dimer	None
1yar.1.O	Proteosome activator protein PA26	13.43	X-ray, 1.9A	Hetero-2 mer	None

Model 1   Model 2   Model 3
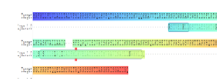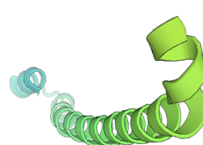
5b. Srz_13902

Name	Title	Identity	Method	Oligo state	Ligands
2vnd.1.A	ENDONUCLEASE I	26.67	X-ray, 1.7A	Monomer	1 × MG
2mju.1.A	Polypyrimidine tract-binding protein 2	20.00	NMR	Monomer	None
1z7q.1.3	Proteosome activator protein PA26	17.11	X-ray, 3.2A	Hetero-2-mer	None
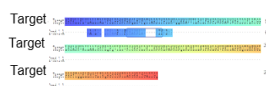
Model 1   Model 2   Model 3
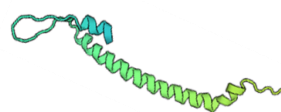
6a. Srs_13903

Name	Title	Identity	Method	Oligo state	Ligands
1dpj.1.B	Proteinase inhibitor IA3 peptide	27.59	X-ray, 1.8A	Hetero-dimer	1xNAG
4o8m.1.A	TRAP dicarboxylate transporter, DcP subunit	25.00	X-ray, 1.7A	Monomer	1x2Q2
4xws.1.A	OxyR	21.43	X-ray, 3.0A	Homo-dimer	None
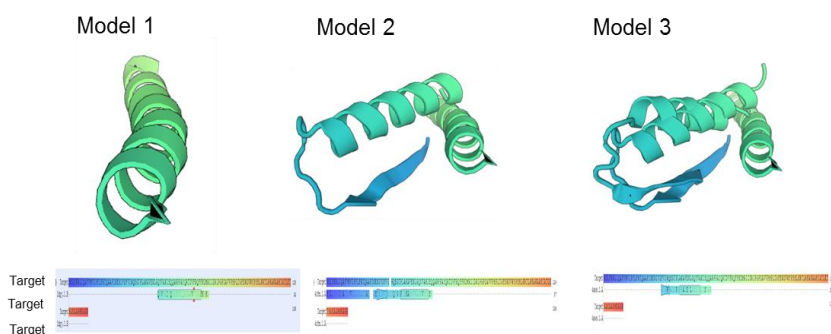
6b. Srz_13903

Name	Title	Identity	Method	Oligo state	Ligands
1dpj.1.B	Proteinase inhibitor IA3 peptide	27.59	X-ray, 1.8A	Hetero-dimer	1xNAG
4o8m.1.A	TRAP dicarboxylate transporter, DcP subunit	25.00	X-ray, 1.7A	Monomer	1x2Q2
4xws.1.A	OxyR	18.75	X-ray, 3.0A	Homo-dimer	None
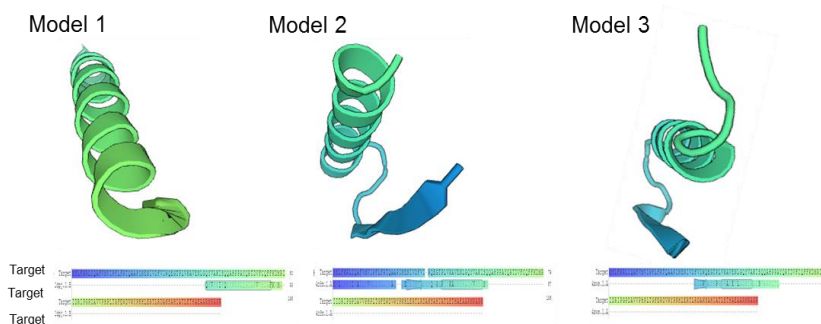
7a. Srs_13904

Name	Title	Identity	Method	Oligo state	Ligands
2dsy.1.A	Hypothetical protein TTHA0281	27.59	X-ray, 1.9A	Homo-tetramer	2xNHE, 1xMG
3lcn.1.A	Nuclear polyadenylated RNA-binding protein NAB2	26,09	X-ray, 2.0A	Hetero-dimer	2xZN
2n28.1.A	Protein Vpu	19.67	NMR	monomer	None

Model 1   Model 2   Model 3
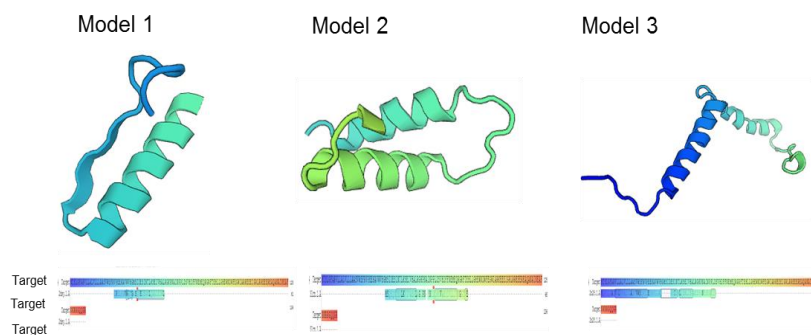
7b. Srz_13904

Name	Title	Identity	Method	Oligo state	Ligands
1anv.1.A	Adenovirus single-stranded DNA-binding protein	28.12	X-ray, 2.7A	Monomer	2xZN, 4xIUM
2waz.1.A	E2A DNA-binding protein	28.12	X-ray, 2.3A	Monomer	2xZN
3lcn.2.A	Nuclear polyadenylated RNA-binding protein NAB2	27.27	X-ray, 2.0A	Hetero-dimer	1xZN

Model 1   Model 2   Model 3
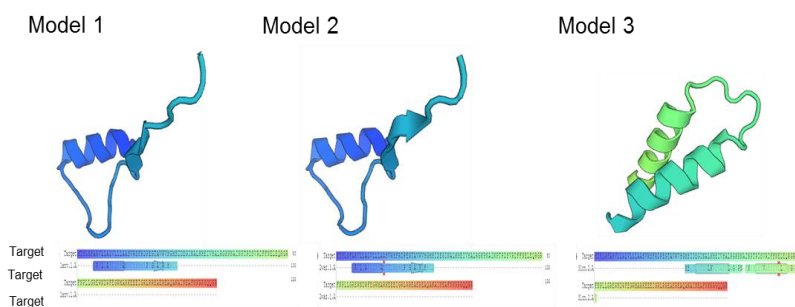
8a. Srs_13905

Name	Title	Identity	Method	Oligo state	Ligands
4z6y.1.D	Hamartin	33.33	X-ray, 2.8A	Hetero-tetramer	None
4bwd.2.A	Short coiled-coil protein	31.43	X-ray, 2.7A	Homo-dimer	None
4v7s.1.O	30S ribosomal protein S15	21.95	X-ray, 3.3A	Hetero-49-mer	10xMG, 1xTEL, 1xZN

Model 1   Model 2   Model 3
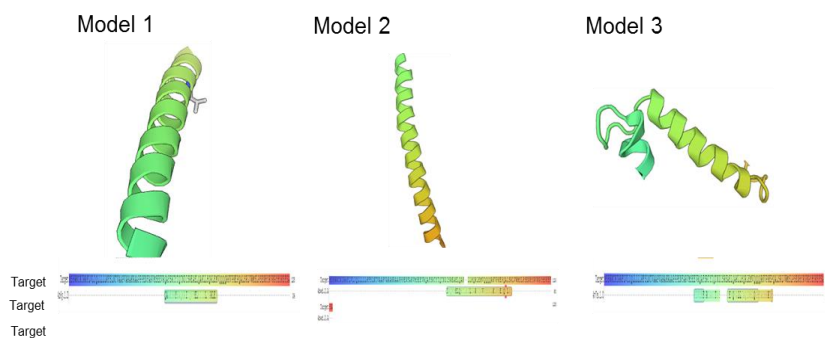
8b. Srz_13905

Name	Title	Identity	Method	Oligo state	Ligands
5ejc.1.B	Hamartin	33.33	X-ray, 3.1A	Hetero-trimmer	None
4bwd.2.A	Short coiled-coil protein	31.43	X-ray, 2.7A	Homo-dimer	None
4v7s.1.O	30S ribosomal protein S15	21.95	X-ray, 3.3A	Hetero-49-mer	10xMG, 1xTEL, 1xZN

Model 1   Model 2   Model 3
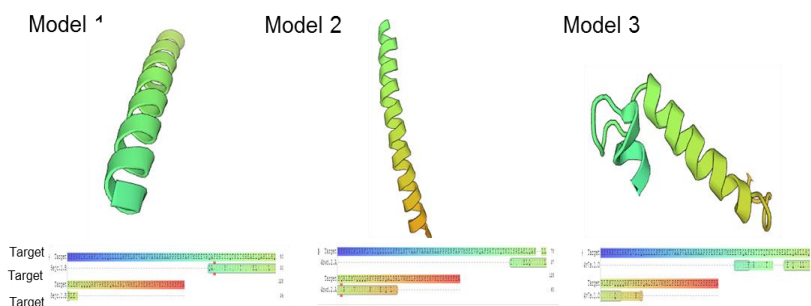
9a. Srs_13906

Name	Title	Identity	Method	Oligo state	Ligands
1vfc.1.C	Telomeric repeat binding factor 2	48.83	NMR	Monomer	None
2d9a.1.A	Myd-related protein B	36.00	NMR	Monomer	None
4c2m.1.D	DNA-directed RNA polymerase I subunit RPA14	33.33	X-ray, 2.8A	Hetero-15-mer	7xZN
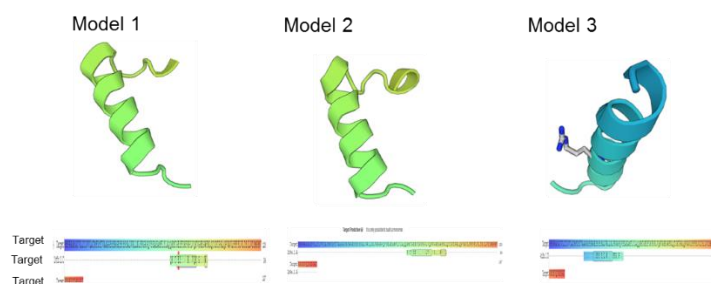
9b. Srz_13906

Name	Title	Identity	Method	Oligo state	Ligands
3wja.1.A	NADP-dependent malic enzyme	30.30	X-ray, 2.5A	Homo-tetramer	None
2qgx.4.A	Ubiquitin-conjugating enzyme E2 Q1	28.57	X-ray, 2.6A	monomer	None
4dhx.2.C	Enhancer of yellow 2 transcription factor homolog	26.67	X-ray, 2.1A	Hetero-trimer	None
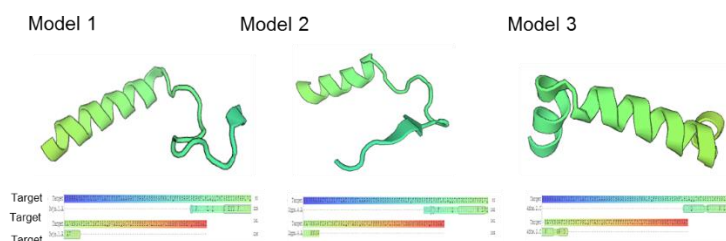
Figure 49: Structural modeling of Cluster 7_11 proteins. Three structural templates identified are given in tabular form along with their annotation, identity match, method used, oligo state and ligand. Three models for each protein aligning against the region of template are shown. 1a-Srs_13897 and 1b-Srz_13897, 2a-Srs_13898 and 2b-Srz_13898, 3a-Srs_13900 and 3b-Srz_13900, 4a-Srs_13901 and 4b-Srz_13901, 5a-Srs_13902 and 5b-Srz_13902, 6a-Srs_13903 and 6b-Srz_13903, 7a-Srs_13904 and 7b-Srz_13904, 8a-Srs_13905 and 8b-Srz_13905, 9a-Srs_13906 and 9b-Srz_13906.

3.3.6 Intrinsic disorder prediction of predicted secreted proteins in cluster 7_11 show them to have region of disordered residues

Intrinsically disordered protein regions are common in pathogenic microbes [91]. We analyzed the overall intrinsic disorder predisposition in our 9 predicted secretory proteins of the cluster. We used a set of established disorder predictors of PONDR family (PONDR® VLXT, PONDR® XL1_XT, PONDR® VL3-BA and PONDR® VSL2). We saw an impressive prevalence of intrinsic disorder in the predicted secretory proteins. In particular, PONDR® VLXT predicted Srs_13900, Srs_13901, Srs_13905, Srz_13898, Srz_13900, Srz_13901 and Srz_13905 as highly disordered having between 30% and 50% of disordered residues **(Table. 1s)**. Rest of the proteins was found to be moderately disordered having values between 10% and 30%. On average combining all the above 4 tools, the proteins were found to have disordered residues with Srs_13905 and Srz_13905 showing the highest percentage disordered residues of 54.4% and 56.2% respectively (**Table 1s**).

Disorder-based binding sites or molecular recognition features (MoRFs i.e., sites that are disordered in the unbound state and undergo disorder-to-order transition on interaction with the binding partners) are shown to be present in many effectors [92]. We used ANCHOR algorithm [93] to evaluate the presence of the disordered- based protein-protein interaction sites, MoRFs. Four SRS proteins (Srs_13898, Srs_13900, Srs_13902 and Srs_13905) showed positive for presence of such regions whereas only two SRZ proteins (Srz_13900 and Srz_13902) were having MoRFs (**Figure 3s).** This difference may contribute to their variation in the way of interacting with partners in regulation of functionality.

4 Discussion

The concept of host specificity as mentioned in the introduction has been a very interesting topic across all the scientific community. This area found its voice in various molecular marker associated studies which could successfully differentiate among different closely related species. However, it was lacking the precision of narrowing down to the genomic loci associated with a particular phenotype. The appearance of next generation sequencing platforms acted like a boon for this whole research area. Availability of various fast sequencing platforms enabled easy and fast sequencing of many organisms which was not possible a decade back. The need of the hour is however to decode this large pool of information and bring out as much data as possible that is hidden in these genomes. It is not enough to mention again and again the importance of comparative genomics to identify genomic regions for evolution, regulation and virulence to name a few. Not only in fungal biology; had this method found its immense use in all other biological systems.

In case of fungal biology however, one of the most interesting outcomes using comparative genomics came from the genome comparison between *S. reilianum* and *U. maydis* [7]. Identification of regions of low conservation and subsequent deletion to validate, paved pathway for many other related works in the following years. For example, comparative analysis of secreted proteins in phytopathogenic smut fungi and basidiomycetes has recently shown interesting outcomes [94]. Comparison among the basidiomycetes led to identification of various clusters and also strong correspondence between the secretomes of *Pseudozyma* species and smuts. Keeping in mind the importance of comparative genomics, we wanted to find out the host- specific virulence determinants in *S. reilianum*. *S. reilianum* that exists in two host adapted *formae speciales* infects either sorghum (f. sp. *reilianum*, or SRS) or maize (f. sp. *zeae*, or SRZ) is an excellent system to study host specificity in plant pathogenic fungi. At genome level SRS and SRZ are very similar. Also, there are no additional genes in the genome of SRZ that donot have homologs in SRS and thus can be immediate host specificity candidates. This emphasizes on the idea that host specificity determinants have homologs but likely differ in amino acid sequence. A comparison between SRS and SRZ shows that more than 90% of all the genes share an amino acid identity higher than 95% (**Figure 4s**). In order to have an approach which allows identification of virulence associated genes in an unbiased way and not solely based on genome comparison, we chose a combination of classical genetics, next generation sequencing and

whole genome analysis. We made use of the fact that the *S. reilianum* varieties can mate with each other when they differ at their *a* and *b* mating type loci. Use of meiotic recombinants to phenotype and subsequent association to genomic region – "phenotyping by genotyping" is not new to other species, although not so much prevalent in phytopathogenic fungal studies. In case of plants, genome wide association mapping of large population of soybean germplasm was done for seed protein and oil contents [95]. They could successfully identify SNP markers related to seed oil content and protein content. In case of fish evolution or disease resistance studies this approach has been used widely. In case of Rainbow trout fish, disease resistance traits were identified by association mapping after offspring from two different phenotypes (healthy and diseased) were checked for mortality rates [96]. Similarly, genomic basis of adaptive evolution was assessed in Threespine sticklebacks fish [97]. Genome wide set of loci associated with marine–freshwater divergence was thus identified. For this F2 population from freshwater and marine intercross was generated. These kinds of results already prove the importance and efficiency of the method in other biological systems. The future of these methods is brightened by not needing to sequence the whole genomes of species and thereby reducing the cost and time. Many new platforms such as association mapping using K-mers [98] and association mapping by pooled sequencing [99] are few of many such methods that can be used. Therefore, we can be hopeful about the usefulness of the approach that we had taken in order to identify host specificity regions in *S. reilianum*.

4.1 Can we also find association of other phenotypes apart from virulence?

The importance of phenotyping as the basis of associating genomic region linked to such phenotypes is immense. To begin our work, we started by scoring the disease phenotypes of all the offspring on sorghum. However, as our study progressed we saw few other kinds of phenotypical variance in the mapping population. One such variation was in the form of phytoalexin induction. Another was in the variation in colony morphology in terms of coloration and texture.

For assessing virulence capacity, we mated SRSZ strains with SRS2_ H2-7, thus all used strains were mated with a strain specialized on the host, they were injected. Thus, all genes necessary for virulence are present, at least in one copy originating from the nucleus from the mating strain. Whatever stops the spore formation in sorghum must have come from the SRZ genomic part of the SRSZ offspring. Virulence analysis revealed that the mapping population dis-

played a continuum of different virulence capacities on sorghum. Since the analyzed offspring showed varying degrees of virulence on sorghum, host specificity of *S. reilianum* is likely a multigenic trait. We observed at least five different kinds of phenotypes upon infection in sorghum (**Figure 13**, **Figure 14**). Similarly, when we infected maize seedling we also saw various degree of spore forming capacity of the offspring (**Figure 17**). In this case we used SRZ2_5-1 as mating partner. Since at least one copy of the genes necessery for virulence is present coming from the mating partner, the factors that hinder spore forming capacity must have come from the SRS part of the SRSZ offpring. Interestingly when we switched the mating partner in maize and used SRS2_ H2-7 as mating partner (like in case of sorghum), we saw the spore forming capacity went lower which is very likely as there exist only fraction of the SRZ genomic content present in SRSZ offspring (**Figure 18**). After trying to associate genomic region related to spore formation in sorghum by SRS, we found a region of 9 genes that segregate together and do not undergo any recombination in any of the offspring. This region was originally described as a region of divergence in genome comparison between *S. reilianum* and *U. maydis* [7]. This region was found to coming from SRS parent in 97.4% of the virulent offspring on sorghum (**Figure 41**) whereas the same region in 87.27% of the avirulent offspring on sorghum comes from SRZ parent. For those strains that deviated from this pattern of association, the phenotyping need to be repeated. Also there might be manual error while naming the vials at the time of genomic DNA isolation. However, if everything went correct this deviation would be interesting to at least know that this kind of association is not 100% in nature. This also emphasizes on the fact of presence of other regions of interest apart from this as a sole region. It is important to mention here that when we deleted this cluster of 9 genes there was no spore formation in sorghum whatsoever. However, we saw many offspring with intermediate phenotype before. This also shows way to future direction of comparing the genomes of those strains which are of the combinations - full spore vs. little spore, leafy vs. healthy or leafy vs. spore. This could give us hint on those regions in genome that might have some additive effect on the way to spore formation. On the other hand, it is also possible not to find any specific region in those combinations and therefore they may come out as not so big phenotypic pairs as spore vs. no spore combination that we have used.

Results showed that parental origin of chromosomal regions could be unambiguously assigned for all offspring (**Figure 39**). Interestingly, few strains contained partially duplicated genomic regions, i.e. carrying the same chromosomal fragments from both parents. This kind of meiosis-driven genome variation is having been considered as major driving force for gene and genome evolution in nature. They are often considered as meiotically unbalanced gametes.

Knowledge of these kinds of occurrence is seen in plant kingdom [100]. In fungi however this is not reported so commonly. Viable offspring having a novel chromosome originating from a large duplication gave rise to accessory chromosome the fungal wheat pathogen *Zymoseptoria tritici* [101].

We have shown before that sorghum infected with SRZ leads to the induction of phytoalexins that form readily visible red stains on plant leaves and fungal hyphae [10]. Phytoalexins are considered a good marker for resistance to biotic stress in sorghum [102]. The phytoalexin induced by SRZ in sorghum were identified as luteolinidin and apigeninidin, but only luteolinidin was able to restrict the growth of haploid *S. reilianum* cells *in vitro* [10]. In this study we observed that except for seven strains all the other spore forming offspring did not induce phytoalexin formation in sorghum (**Figure 15** B). Therefore, they were able to successfully colonize the host and finally lead to spore formation at the time of flowering. On the other hand, the avirulent strains (except three) were stopped by plant defense response in the form of phytoalexin production in sorghum and could not lead to spore formation.

So which defense response of the plant stops the fungal growth inside? Callose deposition is most likely not the reason as the amount of callose depositions in sorghum can differ regardless of phytoalexin responses or virulence. Both SRS and SRZ can induce callose deposition even if in different amounts upon infection on sorghum. We saw that in majority of the strains the typical phytoalexin is-linked-to-avirulence approach was found to be true; however, there were still at least five percent of the strains that deviated from this behavior.

Another feature we observed was the difference of colony morphology of the strains when grown on PDA medium. The morphology of a colony, representing one kind of microorganism is unique for every microorganism there exists. In case of bacteria variation in colony morphology and its relatedness to virulence is quite common. In case of *Yersinia enterolitica*, a gram-negative bacterium association of colony morphology with virulence was reported [103]. We wanted to know if there existed any pattern in the colony morphology variation in *S. reilianum* or there existed any genomic region responsible for it. *S. reilianum* is a facultative biotrophic fungus that can live saprotrophically as haploid yeast cells (sporidia). It was interesting how the offspring originating from the same species have various colony morphologies in terms of colour and texture. Therefore, we categorised the colonies based on their texture into six different kinds (**Figure 31**). While most of the strains showed texture similar to the WT there were others that differed. In addition, we also quantified the grey scale intensity of the colonies. It was observed that for most of the strains the value was higher as compared to both SRS and SRZ WT (**Figure 30**). The pattern showed a gradual increase in the value and seems like there

are more than one factor responsible for different colony morphologies in *S. reilianum*. It would be another interesting outcome to find out associated region in the genome of *S. reilianum* for such variations.

4.2 <u>What do we already know about the genes in cluster 7_11?</u>

As mentioned earlier the region of 9 genes identified above forms a part of a divergent cluster 7_11 previously identified [7]. We found that the genes here were more highly expressed *in planta* (**Figure 46**), showed the least amount of sequence conservation (**Figure 45**) and coded for proteins carrying predicted secretion signal peptides (**Figure 44**). Keeping in mind the importance of this region we decided to delete it in SRS background. Virulence analysis of the mutant in sorghum showed that there is no spore formation (**Figure 48**). This suggests the importance of the cluster in virulence in SRS. Since we did not observe any recombination event whatsoever inside of this cluster, it can be assumed that the whole cluster in general is associated with the disease phenotype. Cluster 7_11 thus identified to be a region of host specificity associated with virulence of SRS in sorghum.

6 of the genes present in the cluster are found to have homologs in *U. maydis*. The *U. maydis* genes – UMAG_02851 (homolog in *S. reilianum* sr13897 and sr13903), UMAG_02852 (homolog in *S. reilianum* sr13904) and UMAG_02853 (homolog in *S. reilianum* sr13905) were found to be up regulated 2-4 dpi whereas UMAG_02854 (homolog in *S. reilianum* sr13906) and UMAG_11484 (homolog in *S. reilianum* sr13907) were found to be up regulated 4-8 dpi [104]. Up regulation of genes early during *in planta* invasion is important for the successful fungal proliferation. This was similar in case of all the *S. reilianum* particularly the genes SRS-13898, 13901, 13904 and 13905 and SRZ-13898, 13901, 13902, 13904, 13905 and 13906 were seen to show higher expression values early *in planta*. Gene um02854 of *U. maydis* (homolog in *S. reilianum* Sr13905) has been shown to be up regulated 5 days after co- cultivation with *Fusarium verticillioide* than grown alone [105]. Thereby this gene might be involved in mechanism responsible for sensing and responding to each other's presence. They speculated that this gene (along with few other genes) may encode proteins involved in the defense against other microbes during *in planta* colonization. InterPro scan was used to predict functions or domains of 8 of the *S. reilianum* cluster genes (excluding sr1390) among other smut fungi [94]. Except for sr13904 and sr13905 where the prediction was coils, others had no InterPro prediction. In yet

another study um02851 (homolog in *S. reilianum* sr13902 and sr13897), um02852 (homolog in *S. reilianum* sr13903) and um02853 (homolog in *S. reilianum* sr13904) is shown to be positively regulated by GCN5 gene which is a histone acetyltransferases (HATs) [106]. All these studies point towards the importance of this particular region in smut fungi.

Protein structure homology modelling provides idea about the possible 3D structure of proteins based on information derived from known structures (templates) available. Due to the small protein sizes in the cluster; it was not possible to arrive on a model which covers the whole template(s) available. However, from our data we have observed that more similar the proteins were at sequence level, higher the chances that they hit similar templates in database (**Figure 49**). This indicates towards the structural similarity of proteins with higher SIMAP value conservations and variation at structural level among proteins of lowers SIMAP value conservation between SRS and SRZ. This might have role in variation in interaction with various ligands or receptors towards successful invasion of the host plant(s).

5   <u>References</u>

1. Daniel, B.; Timothy, H.; J., G.S. The global spread of crop pests and pathogens. *Global Ecology and Biogeography* **2014**, *23*, 1398-1407.
2. Bebber, D.; Ramotowski, M.A.T.; Gurr, S.J. Crop pests and pathogens move polewards in a warming world. *Nature Climate Change* **2013**, *3*, 985.
3. Fisher, M.C.; Henk, D.A.; Briggs, C.J.; Brownstein, J.S.; Madoff, L.C.; McCraw, S.L.; Gurr, S.J. Emerging fungal threats to animal, plant and ecosystem health. *Nature* **2012**, *484*, 186.
4. Bauer, R.; Begerow, D.; Oberwinkler, E.; Piepenbring, M.; Berbee, M.L. Ustilaginomycetes. In *Systematics and Evolution*, McLaughlin, D.J.; McLaughlin, E.G.; Lemke, P.A., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, **2001**; pp 57-83.
5. Schirawski, J.; Heinze, B.; Wagenknecht, M.; Kahmann, R. Mating type loci of *Sporisorium reilianum*: novel pattern with three a and multiple b specificities. *Eukaryotic Cell* **2005**, *4*, 1317-1327.
6. Martinez, C.; Roux, C.; Dargent, R. Biotrophic Development of Sporisorium reilianum f. sp. zeae in Vegetative Shoot Apex of Maize. *Phytopathology* **1999**, *89*, 247-253.
7. Schirawski, J.; Mannhaupt, G.; Munch, K.; Brefort, T.; Schipper, K.; Doehlemann, G.; Di Stasio, M.; Rossel, N.; Mendoza-Mendoza, A.; Pester, D.*, et al.* Pathogenicity determinants in smut fungi revealed by genome comparison. *Science* **2010**, *330*, 1546-1548.
8. Poloni, A.; Schirawski, J. Host specificity in *Sporisorium reilianum* is determined by distinct mechanisms in maize and sorghum. *Mol Plant Pathol* **2016**, *17*, 741-754.
9. Ghareeb, H.; Becker, A.; Iven, T.; Feussner, I.; Schirawski, J. *Sporisorium reilianum* Infection Changes Inflorescence and Branching Architectures of Maize. *Plant Physiology* **2011**, *156*, 2037-2052.
10. Zuther, K.; Kahnt, J.; Utermark, J.; Imkampe, J.; Uhse, S.; Schirawski, J. Host specificity of *Sporisorium reilianum* is tightly linked to generation of the phytoalexin luteolinidin by *Sorghum bicolor*. *Molecular Plant Microbe Interaciont* **2012**, *25*, 1230-1237.
11. Borah, N.; Albarouki, E.; Schirawski, J. Comparative methods for molecular determination of host-specificity factors in plant-pathogenic fungi. *International Journal of Molecular Sciences* **2018**, *19*, 863.
12. Flor, H.H. Current status of the gene-for-gene concept. *Annual Review of Phytopathology* **1971**, *9*, 275-296.
13. van der Hoorn, R.A.L.; Kamoun, S. From guard to decoy: a new model for perception of plant pathogen effectors. *The Plant Cell* **2008**, *20*, 2009-2017.
14. Van Der Biezen, E.A.; Jones, J.D.G. Plant disease-resistance proteins and the gene-for-gene concept. *Trends in Biochemical Science* **1998**, *23*, 454-456.
15. Chisholm, S.T.; Coaker, G.; Day, B.; Staskawicz, B.J. Host-microbe interactions: shaping the evolution of the plant immune response. *Cell* **2006**, *124*, 803-814.
16. de Jonge, R.; Bolton, M.D.; Thomma, B.P.H.J. How filamentous pathogens co-opt plants: the ins and outs of fungal effectors. *Current Opinion Plant Biology* **2011**, *14*, 400-406.
17. Petit-Houdenot, Y.; Fudal, I. Complex interactions between fungal avirulence genes and their corresponding plant resistance genes and consequences for disease resistance management. *Frontier Plant Science* **2017**, *8*.
18. Sidhu, G.; Person, C. Genetic control of virulence in *Ustilago hordei* iii. identification of genes for host resistance and demonstration of gene-for-gene relations. *Canadian Journal of Genetics and Cytology* **1972**, *14*, 209-213.
19. Linning, R.; Lin, D.; Lee, N.; Abdennadher, M.; Gaudet, D.; Thomas, P.; Mills, D.; Kronstad, J.W.; Bakkeren, G. Marker-based cloning of the region containing the

Uhavr1 avirulence gene from the basidiomycete barley pathogen *Ustilago hordei*. *Genetics* **2004**, *166*, 99-111.

20. Andrie, R.M.; Pandelova, I.; Ciuffetti, L.M. A combination of phenotypic and genotypic characterization strengthens *Pyrenophora tritici-repentis* race identification. *Phytopathology* **2007**, *97*, 694-701.

21. Chiapello, H.; Mallet, L.; Guerin, C.; Aguileta, G.; Amselem, J.; Kroj, T.; Ortega-Abboud, E.; Lebrun, M.H.; Henrissat, B.; Gendrault, A.*, et al.* Deciphering genome content and evolutionary relationships of isolates from the fungus *Magnaporthe oryzae* attacking different host plants. *Genome Biology and Evolution* **2015**, *7*, 2896-2912.

22. Baayen, R.P.; O'Donnell, K.; Bonants, P.J.M.; Cigelnik, E.; Kroon, L.P.N.M.; Roebroeck, E.J.A.; Waalwijk, C. Gene genealogies and AFLP analyses in the *Fusarium oxysporum* complex identify monophyletic and nonmonophyletic formae speciales causing wilt and rot disease. *Phytopathology* **2000**, *90*, 891-900.

23. Chen, C.-H.; Sheu, Z.-M.; Wang, T.-C. Host specificity and tomato-related race composition of *Phytophthora infestans* isolates in Taiwan during 2004 and 2005. *Plant Disease* **2008**, *92*, 751-755.

24. Kovacikova, E.; Kratka, J. Different manifestations of the pathogenity of some strains of *Fusarium oxysporum* f. sp. *pisi*. *Zentralblatt für Bakteriologie Naturwissen* **1979**, *134*, 159-166.

25. Heath, M.C. Host species specificity of the goldenrod rust fungus and the existence of rust resistance within some Goldenrod species. *Canadian Journal of Botany* **1992**, *70*, 2461-2466.

26. Yehuda, P.B.; Eilam, T.; Manisterski, J.; Shimoni, A.; Anikster, Y. Leaf rust on *Aegilops speltoides* caused by a new forma specialis of *Puccinia triticina*. *Phytopathology* **2004**, *94*, 94-101.

27. Bernstein, B.; Zehr, E.I.; Dean, R.A.; Shabi, E. Characteristics of colletotrichum from peach, apple, pecan, and other hosts. *Plant Disease* **1995**, *79*, 478-482.

28. Greco, M.; Patriarca, A.; Terminiello, L.; Fernandez Pinto, V.; Pose, G. Toxigenic Alternaria species from Argentinean blueberries. *International Journal of Food Microbiology* **2012**, *154*, 187-191.

29. Yasuda, K.; Kojima, M. The role of stress metabolites in establishing host-parasite specificity between sweet potato and *Ceratocystis fimbriata*, black rot fungus. *Agricultural and Biological Chemistry* **1986**, *50*, 1839-1846.

30. Masunaka, A.; Ohtani, K.; Peever, T.L.; Timmer, L.W.; Tsuge, T.; Yamamoto, M.; Yamamoto, H.; Akimitsu, K. An isolate of *Alternaria alternata* that is pathogenic to both tangerines and rough lemon and produces two host-selective toxins, ACT- and ACR-toxins. *Phytopathology* **2005**, *95*, 241-247.

31. Vos, P.; Hogers, R.; Bleeker, M.; Reijans, M.; van de Lee, T.; Hornes, M.; Frijters, A.; Pot, J.; Peleman, J.; Kuiper, M.*, et al.* AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **1995**, *23*, 4407-4414.

32. Douhan, G.W.; de la Cerda, K.A.; Huryn, K.L.; Greer, C.A.; Wong, F.P. Contrasting genetic structure between *Magnaporthe grisea* populations associated with the golf course turfgrasses *Lolium perenne* (perennial ryegrass) and *Pennisetum clandestinum* (kikuyugrass). *Phytopathology* **2011**, *101*, 85-91.

33. Flier, W.G.; Grünwald, N.J.; Kroon, L.P.N.M.; Sturbaum, A.K.; van den Bosch, T.B.M.; Garay-Serrano, E.; Lozoya-Saldaña, H.; Fry, W.E.; Turkensteen, L.J. The population structure of *Phytophthora infestans* from the toluca valley of central Mexico suggests genetic differentiation between populations from cultivated potato and wild Solanum spp. *Phytopathology* **2003**, *93*, 382-390.

34. Karimi, S.; Mirlohi, A.; Sabzalian, M.R.; Sayed Tabatabaei, B.E.; Sharifnabi, B. Molecular evidence for Neotyphodium fungal endophyte variation and specificity within host grass species. *Mycologia* **2012**, *104*, 1281-1290.

35. Wyand, R.A.; Brown, J.K. Genetic and forma specialis diversity in *Blumeria graminis* of cereals and its implications for host-pathogen co-evolution. *Molecular Plant Pathology* **2003**, *4*, 187-198.

36. Garry, G.; Forbes, G.A.; Salas, A.; Santa Cruz, M.; Perez, W.G.; Nelson, R.J. Genetic diversity and host differentiation among isolates of *Phytophthora infestans* from cultivated potato and wild solanaceous hosts in Peru. *Plant Pathology* **2005**, *54*, 740-748.

37. Liu, B.; Wasilwa, L.A.; Morelock, T.E.; O'Neill, N.R.; Correll, J.C. Comparison of *Colletotrichum orbiculare* and several allied Colletotrichum spp. for mtDNA RFLPs, intron RFLP and sequence variation, vegetative compatibility, and host specificity. *Phytopathology* **2007**, *97*, 1305-1314.

38. Rigotti, S.; Gindro, K.; Richter, H.; Viret, O. Characterization of molecular markers for specific and sensitive detection of *Botrytis cinerea* Pers.: Fr. in strawberry (*Fragariaxananassa duch*.) using PCR. *FEMS Microbiological Letters* **2002**, *209*, 169-174.

39. Mahmodi, F.; Kadir, J.B.; Puteh, A.; Pourdad, S.S.; Nasehi, A.; Soleimani, N. Genetic diversity and differentiation of Colletotrichum spp. isolates associated with Leguminosae using multigene loci, RAPD and ISSR. *Plant Pathology* **2014**, *30*, 10-24.

40. Peever, T.L.; Canihos, Y.; Olsen, L.; Ibanez, A.; Liu, Y.C.; Timmer, L.W. Population genetic structure and host specificity of Alternaria spp. causing brown spot of minneola tangelo and rough lemon in Florida. *Phytopathology* **1999**, *89*, 851-860.

41. Peever, T.L.; Olsen, L.; Ibanez, A.; Timmer, L.W. Genetic differentiation and host specificity among populations of Alternaria spp. causing brown spot of grapefruit and tangerine x grapefruit hybrids in Florida. *Phytopathology* **2000**, *90*, 407-414.

42. Lievens, B.; Claes, L.; Vakalounakis, D.J.; Vanachter, A.C.; Thomma, B.P. A robust identification and detection assay to discriminate the cucumber pathogens *Fusarium oxysporum* f. sp. *cucumerinum* and f. sp. *radicis-cucumerinum*. *Environtal Microbiology* **2007**, *9*, 2145-2161.

43. Penselin, D.; Munsterkotter, M.; Kirsten, S.; Felder, M.; Taudien, S.; Platzer, M.; Ashelford, K.; Paskiewicz, K.H.; Harrison, R.J.; Hughes, D.J*., et al.* Comparative genomics to explore phylogenetic relationship, cryptic sexual potential and host specificity of Rhynchosporium species on grasses. *BMC Genomics* **2016**, *17*, 953.

44. Zhong, Z.; Norvienyeku, J.; Chen, M.; Bao, J.; Lin, L.; Chen, L.; Lin, Y.; Wu, X.; Cai, Z.; Zhang, Q*., et al.* Directional Selection from Host Plants Is a Major Force Driving Host Specificity in Magnaporthe Species. *Science Report* **2016**, *6*, 25591.

45. Perlin, M.H.; Hughes, C.; Welch, J.; Akkaraju, S.; Steinecker, D.; Kumar, A.; Smith, B.; Garr, S.S.; Brown, S.A.; Andom, T. Molecular approaches to differentiate subpopulations or formae speciales of the fungal phytopathogen *Microbotryum violaceum*. *International Journal of Plant Science* **1997**, *158*, 568-574.

46. Santini, A.; Capretti, P. Analysis of the Italian population of *Ceratocystis fimbriata* f.sp. *platani* using RAPD and minisatellite markers. *Plant Pathology* **2000**, *49*, 461-467.

47. Freeman, S.; Katan, T.; Shabi, E. Characterization of *Colletotrichum gloeosporioides* isolates from avocado and almond fruits with molecular and pathogenicity tests. *Appl Environtal Microbiology* **1996**, *62*, 1014-1020.

48. Asadollahi, M.; Fekete, E.; Karaffa, L.; Flipphi, M.; Arnyasi, M.; Esmaeili, M.; Vaczy, K.Z.; Sandor, E. Comparison of *Botrytis cinerea* populations isolated from two open-field cultivated host plants. *Microbiological Research* **2013**, *168*, 379-388.

49. Van Putten, W.F.; Biere, A.; Van Damme, J.M.M. Host-related genetic differentiation in the anther smut fungus *Microbotryum violaceum* in sympatric, parapatric and allopatric populations of two host species *Silene latifolia* and *S. dioica*. *Journal of Evolutionary Biology* **2005**, *18*, 203-212.

50. Beckstead, J.; Meyer, S.E.; Ishizuka, T.S.; McEvoy, K.M.; Coleman, C.E. Lack of host specialization on winter annual grasses in the fungal seed bank pathogen *Pyrenophora semeniperda*. *PLoS One* **2016**, *11*, e0151058.

51. Dipietro, A.; Anaya, N.; Roncero, M.I.G. Occurrence of a retrotransposon-like sequence among different formae speciales and races of *Fusarium oxysporum*. *Mycolgical Research* **1994**, *98*, 993-996.

52. Samuel, S.; Veloukas, T.; Papavasileiou, A.; Karaoglanidis, G.S. Differences in frequency of transposable elements presence in *Botrytis cinerea* populations from several hosts in Greece. *Plant Disease* **2012**, *96*, 1286-1290.

53. Laurie, J.D.; Ali, S.; Linning, R.; Mannhaupt, G.; Wong, P.; Guldener, U.; Munsterkotter, M.; Moore, R.; Kahmann, R.; Bakkeren, G.*, et al.* Genome comparison of barley and maize smut fungi reveals targeted loss of RNA silencing components and species-specific presence of transposable elements. *Plant Cell* **2012**, *24*, 1733-1745.

54. Dutheil, J.Y.; Mannhaupt, G.; Schweizer, G.; C, M.K.S.; Munsterkotter, M.; Guldener, U.; Schirawski, J.; Kahmann, R. A tale of genome compartmentalization: the evolution of virulence lusters in smut fungi. *Genome Biology and Evolution* **2016**, *8*, 681-704.

55. Buiate, E.A.; Xavier, K.V.; Moore, N.; Torres, M.F.; Farman, M.L.; Schardl, C.L.; Vaillancourt, L.J. A comparative genomic analysis of putative pathogenicity genes in the host-specific sibling species *Colletotrichum graminicola* and *Colletotrichum sublineola*. *BMC Genomics* **2017**, *18*, 67.

56. Niehaus, E.M.; Munsterkotter, M.; Proctor, R.H.; Brown, D.W.; Sharon, A.; Idan, Y.; Oren-Young, L.; Sieber, C.M.; Novak, O.; Pencik, A.*, et al.* Comparative "Omics" of the *Fusarium fujikuroi* species complex highlights differences in genetic potential and metabolite synthesis. *Genome Biology and Evolution* **2016**, *8*, 3574-3599.

57. Baroncelli, R.; Amby, D.B.; Zapparata, A.; Sarrocco, S.; Vannacci, G.; Le Floch, G.; Harrison, R.J.; Holub, E.; Sukno, S.A.; Sreenivasaprasad, S.*, et al.* Gene family expansions and contractions are associated with host range in plant pathogens of the genus Colletotrichum. *BMC Genomics* **2016**, *17*, 555.

58. Sharma, R.; Mishra, B.; Runge, F.; Thines, M. 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* **2014**, *6*, 2034-2049.

59. Qin, S.; Ji, C.; Li, Y.; Wang, Z. Comparative transcriptomic analysis of race 1 and race 4 of *Fusarium oxysporum* f. sp. *cubense* induced with different carbon sources. *G3 (Bethesda)* **2017**, *7*, 2125-2138.

60. Ma, L.-J.; van der Does, H.C.; Borkovich, K.A.; Coleman, J.J.; Daboussi, M.-J.; Di Pietro, A.; Dufresne, M.; Freitag, M.; Grabherr, M.; Henrissat, B.*, et al.* Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. *Nature* **2010**, *464*, 367-373.

61. Williams, A.H.; Sharma, M.; Thatcher, L.F.; Azam, S.; Hane, J.K.; Sperschneider, J.; Kidd, B.N.; Anderson, J.P.; Ghosh, R.; Garg, G.*, et al.* Comparative genomics and prediction of conditionally dispensable sequences in legume-infecting *Fusarium oxysporum* formae speciales facilitates identification of candidate effectors. *BMC Genomics* **2016**, *17*.

62. van Dam, P.; Fokkens, L.; Schmidt, S.M.; Linmans, J.H.; Kistler, H.C.; Ma, L.J.; Rep, M. Effector profiles distinguish formae speciales of *Fusarium oxysporum*. *Environtal Microbiology* **2016**, *18*, 4087-4102.

63. Lievens, B.; van Baarlen, P.; Verreth, C.; van Kerckhove, S.; Rep, M.; Thomma, B.P. Evolutionary relationships between *Fusarium oxysporum* f. sp. *lycopersici* and F. oxysporum f. sp. radicis-lycopersici isolates inferred from mating type, elongation factor-1alpha and exopolygalacturonase sequences. *Mycological Research* **2009**, *113*, 1181-1191.

64. Bhadauria, V.; MacLachlan, R.; Pozniak, C.; Banniza, S. Candidate effectors contribute to race differentiation and virulence of the lentil anthracnose pathogen *Colletotrichum lentis*. *BMC Genomics* **2015**, *16*, 628.

65. Hosseini, S.; Resjo, S.; Liu, Y.; Durling, M.; Heyman, F.; Levander, F.; Liu, Y.; Elfstrand, M.; Funck Jensen, D.; Andreasson, E.*, et al.* Comparative proteomic analysis of hyphae and germinating cysts of *Phytophthora pisi* and *Phytophthora sojae*. *J Proteomics* **2015**, *117*, 24-40.

66. Savidor, A.; Donahoo, R.S.; Hurtado-Gonzales, O.; Land, M.L.; Shah, M.B.; Lamour, K.H.; McDonald, W.H. Cross-species global proteomics reveals conserved and unique processes in *Phytophthora sojae* and *Phytophthora ramorum*. *Molecular Cell Proteomics* **2008**, *7*, 1501-1516.

67. Quecine, M.C.; Leite, T.F.; Bini, A.P.; Regiani, T.; Franceschini, L.M.; Budzinski, I.G.F.; Marques, F.G.; Labate, M.T.V.; Guidetti-Gonzalez, S.; Moon, D.H.*, et al.* Label-free quantitative proteomic analysis of *Puccinia psidii* uredospores reveals differences of fungal populations infecting eucalyptus and guava. *Plos One* **2016**, *11*.

68. Bregar, O.; Mandelc, S.; Celar, F.; Javornik, B. Proteome analysis of the plant pathogenic fungus *Monilinia laxa* showing host specificity. *Food Technology and Biotechnology* **2012**, *50*, 326-333.

69. Cisar, C.R.; Spiegel, F.W.; TeBeest, D.O.; Trout, C. Evidence for mating between isolates of *Colletotrichum gloeosporioides* with different host specificities. *Current Genetics* **1994**, *25*, 330-335.

70. Hernandez-Bello, M.A.; Chilvers, M.I.; Akamatsu, H.; Peever, T.L. Host specificity of Ascochyta spp. infecting legumes of the Viciae and Cicerae tribes and pathogenicity of an interspecific hybrid. *Phytopathology* **2006**, *96*, 1148-1156.

71. Li, E.; Wang, G.; Xiao, J.; Ling, J.; Yang, Y.; Xie, B. A SIX1 homolog in *Fusarium oxysporum* f. sp. *conglutinans* is required for full virulence on cabbage. *PLoS One* **2016**, *11*, e0152273.

72. Tosa, Y.; Tamba, H.; Tanaka, K.; Mayama, S. Genetic analysis of host species specificity of *Magnaporthe oryzae* isolates from rice and wheat. *Phytopathology* **2006**, *96*, 480-484.

73. Murakami, J.; Tomita, R.; Kataoka, T.; Nakayashiki, H.; Tosa, Y.; Mayama, S. Analysis of host species specificity of *Magnaporthe grisea* toward foxtail millet using a genetic cross between isolates from wheat and foxtail millet. *Phytopathology* **2003**, *93*, 42-45.

74. Foulongne-Oriol, M. Genetic linkage mapping in fungi: current state, applications, and future trends. *Applied Microbiology and Biotechnology* **2012**, *95*, 891-904.

75. Valent, B.; Farrall, L.; Chumley, F.G. *Magnaporthe grisea* genes for pathogenicity and virulence identified through a series of backcrosses. *Genetics* **1991**, *127*, 87-101.

76. Sweigard, J.A.; Carroll, A.M.; Kang, S.; Farrall, L.; Chumley, F.G.; Valent, B. Identification, Cloning, and Characterization of Pwl2, a Gene for Host Species-Specificity in the Rice Blast Fungus. *Plant Cell* **1995**, *7*, 1221-1233.

77. Peyyala, R.; Farman, M.L. *Magnaporthe oryzae* isolates causing gray leaf spot of perennial ryegrass possess a functional copy of the AVR1-CO39 avirulence gene. *Molecular Plant Pathology* **2006**, *7*, 157-165.

78. Rep, M.; Meijer, M.; Houterman, P.M.; van der Does, H.C.; Cornelissen, B.J. *Fusarium oxysporum* evades I-3-mediated resistance without altering the matching avirulence gene. *Molecular Plant Microbe Interaction* **2005**, *18*, 15-23.

79. Hatta, R.; Ito, K.; Hosaki, Y.; Tanaka, T.; Tanaka, A.; Yamamoto, M.; Akimitsu, K.; Tsuge, T. A conditionally dispensable chromosome controls host-specific pathogenicity in the fungal plant pathogen *Alternaria alternata*. *Genetics* **2002**, *161*, 59-70.

80. Isshiki, A.; Akimitsu, K.; Yamamoto, M.; Yamamoto, H. Endopolygalacturonase is essential for citrus black rot caused by *Alternaria citri* but not brown spot caused by *Alternaria alternata*. *Molecular Plant Microbe Interaction* **2001**, *14*, 749-757.

81. Andrie, R.M.; Ciuffetti, L.M. *Pyrenophora bromi,* causal agent of brownspot of bromegrass, expresses a gene encoding a protein with homology and similar activity to Ptr ToxB, a host-selective toxin of wheat. *Molecular Plant Microbe Interaction* **2011**, *24*, 359-367.

82. Birch, P.R.; Armstrong, M.; Bos, J.; Boevink, P.; Gilroy, E.M.; Taylor, R.M.; Wawra, S.; Pritchard, L.; Conti, L.; Ewan, R.*, et al.* Towards understanding the virulence functions of RXLR effectors of the oomycete plant pathogen *Phytophthora infestans*. *Journal of Experimental Botany* **2009**, *60*, 1133-1140.

83. Ohtani, K.; Yamamoto, H.; Akimitsu, K. Sensitivity to *Alternaria alternata* toxin in citrus because of altered mitochondrial RNA processing. *Proeeding of National Academy of Sciences U S A* **2002**, *99*, 2439-2444.

84. Oka, K.; Akamatsu, H.; Kodama, M.; Nakajima, H.; Kawada, T.; Otani, H. Host-specific AB-toxin production by germinating spores of *Alternaria brassicicola* is induced by a host-derived oligosaccharide. *Physiology of Molecular Plant Pathology* **2005**, *66*, 12-19.

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

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

87. Doehlemann, G.; Wahl, R.; Vranes, M.; de Vries, R.P.; Kämper, J.; Kahmann, R. Establishment of compatibility in the *Ustilago maydis*/maize pathosystem. *Journal of Plant Physiology* **2008**, *165*, 29-40.

88. Cohen, S.N.; Chang, A.C.Y.; Hsu, L. Nonchromosomal Antibiotic Resistance in Bacteria: Genetic Transformation of *Escherichia coli* by R-Factor DNA. *Proceeding of National Academy of Sciences U S A* **1972**, *69*, 2110-2114

89. Voigt, C.A. Callose-mediated resistance to pathogenic intruders in plant defense-related papillae. *Frontier in Plant Scencei* **2014**, *5*.

90. Ellinger, D.; Naumann, M.; Falter, C.; Zwikowics, C.; Jamrow, T.; Manisseri, C.; Somerville, S.C.; Voigt, C.A. Elevated Early Callose Deposition Results in Complete Penetration Resistance to Powdery Mildew in Arabidopsis. *Plant Physiology* **2013**, *161*, 1433-1444.

91. Mohan, A.; Sullivan, W.J., Jr.; Radivojac, P.; Dunker, A.K.; Uversky, V.N. Intrinsic disorder in pathogenic and non-pathogenic microbes: discovering and analyzing the unfoldomes of early-branching eukaryotes. *Molecular Biosystem* **2008**, *4*, 328-340.

92. Kuppireddy, V.S.; Uversky, V.N.; Toh, S.S.; Tsai, M.C.; Beckerson, W.C.; Cahill, C.; Carman, B.; Perlin, M.H. Identification and initial characterization of the effectors of an anther smut fungus and potential host target proteins. *International Journal of Molecular Sciences* **2017**, *18(11),* pii: E2489.

93. Dosztanyi, Z.; Meszaros, B.; Simon, I. ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics* **2009**, *25*, 2745-2746.

94. Schuster, M.; Schweizer, G.; Kahmann, R. Comparative analyses of secreted proteins in plant pathogenic smut fungi and related basidiomycetes. *Fungal Genetics Biology* **2017,** *112*, 21-30.

95. Li, D.; Zhao, X.; Han, Y.; Li, W.; Xie, F. Genome-wide association mapping for seed protein and oil contents using a large panel of soybean accessions. *Genomics* **2018,** *111(1),* 90-95.

96. Campbell, N.R.; LaPatra, S.E.; Overturf, K.; Towner, R.; Narum, S.R. Association Mapping of Disease Resistance Traits in Rainbow Trout Using Restriction Site Associated DNA Sequencing. *G3: Genes|Genomes|Genetics* **2014**, *4*, 2473-2481.

97. Jones, F.C.; Grabherr, M.G.; Chan, Y.F.; Russell, P.; Mauceli, E.; Johnson, J.; Swofford, R.; Pirun, M.; Zody, M.C.; White, S*., et al.* The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **2012**, *484*, 55.

98. Rahman, A.; Hallgrímsdóttir, I.; Eisen, M.B.; Pachter, L. Association Mapping From Sequencing Reads Using K-mers. *eLIFE* **2017**.

99. Redmond, S.N.; Eiglmeier, K.; Mitri, C.; Markianos, K.; Guelbeogo, W.M.; Gneme, A.; Isaacs, A.T.; Coulibaly, B.; Brito-Fravallo, E.; Maslen, G*., et al.* Association mapping by pooled sequencing identifies TOLL 11 as a protective factor against *Plasmodium falciparum* in Anopheles gambiae. *BMC Genomics* **2015**, *16*, 779.

100. Cai, X.; Xu, S.S. Meiosis-Driven Genome Variation in Plants. *Current Genomics* **2007**, *8*, 151-161.

101. Croll, D.; Zala, M.; McDonald, B.A. Breakage-fusion-bridge cycles and large insertions contribute to the rapid evolution of accessory chromosomes in a fungal pathogen. *PLoS Genetics* **2013**, *9*, e1003567.

102. Dicko, M.H.; Gruppen, H.; Barro, C.; Traore, A.S.; van Berkel, W.J.H.; Voragen, A.G.J. Impact of Phenolic Compounds and Related Enzymes in Sorghum Varieties for Resistance and Susceptibility to Biotic and Abiotic Stresses. *Journal of Chemical Ecology* **2005**, *31*, 2671-2688.

103. Lazere, J.R.; Gemski, P. Association of colony morphology with virulence of *Yersinia enterocolitica*. *FEMS Microbiology Letters* **1983**, *17*, 121-126.

104. Lanver, D.; Müller, A.N.; Happel, P.; Schweizer, G.; Haas, F.B.; Franitza, M.; Pellegrin, C.; Reissmann, S.; Altmüller, J.; Rensing, S.A*., et al.* The Biotrophic Development of *Ustilago maydis* Studied by RNA-Seq Analysis. *The Plant Cell* **2018**, *30*, 300-323.

105. Jonkers, W.; Rodriguez Estrada, A.E.; Lee, K.; Breakspear, A.; May, G.; Kistler, H.C. Metabolome and Transcriptome of the Interaction between *Ustilago maydis* and *Fusarium verticillioides* In Vitro. *Applied and Environmental Microbiology* **2012**, *78*, 3656-3667.

106. Martinez-Soto, D.; Gonzalez-Prieto, J.M.; Ruiz-Herrera, J. Transcriptomic analysis of the GCN5 gene reveals mechanisms of the epigenetic regulation of virulence and morphogenesis in *Ustilago maydis*. *FEMS Yeast Research* **2015**, *15(6)*, pii: fov055.

6 <u>Supplements</u>

Table. 1s: Evaluating intrinsic disorder propensity of cluster 7_11 proteins. Disorder profiles were generated by PONDR® VLXT, PONDR® XL1_XT, PONDR® VL3-BA and PONDR® VSL2. The proteins were found to have disordered residues with Srs_13905 and Srz_13905 showing the highest percentage disordered residues of 54.4% and 56.2% respectively (in blue).

Genes	VLXT	XL1_XT	VL3-BA	VSL2	Average
Srs_13897	17.5	32.5	13.5	29.5	23.25
Srs_13898	29.66	41.53	29.66	36.86	34.42
Srs_13900	40.79	29.61	38.16	42.11	37.66
Srs_13901	31.53	27.03	11.71	13.51	20.94
Srs_13902	18.69	40.4	43.94	47.98	37.75
Srs_13903	18.38	25	9.56	17.65	17.64
Srs_13904	25.37	41.79	23.13	32.84	30.78
Srs_13905	40.8	46.4	53.6	76.8	54.4
Srs_13906	19.71	36.5	20.44	39.42	29.01
Srz_13897	17	33	13.5	30	23.35
Srz_13898	42.37	31.36	30.51	38.98	35.80
Srz_13900	37.5	26.32	36.84	42.11	35.69
Srz_13901	21.03	34.48	16.38	19.83	25.43
Srz_13902	23.76	38.12	38.12	46.04	36.51
Srz_13903	16.18	11.76	6.62	17.65	13.05
Srz_13904	26.32	39.85	25.56	38.35	32.52
Srz_13905	44.8	46.4	56.8	76.8	56.2

Srz_13906	19.15	43.36		24.11	51.77	34.57
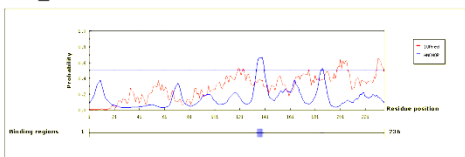
Figure. 3s: Predicted disordered binding regions. Four SRS proteins (Srs_13898, Srs_13900, Srs_13902 and Srs_13905) showed positive for presence of such regions whereas only two SRZ proteins (Srz_13900 and Srz_13902) were having MoRFs using ANCHOR algorithm. Blue graph: prediction profile calculated by ANCHOR, the disordered binding region prediction method; Red graph: prediction profile calculated by IUPred, a general disorder prediction method.
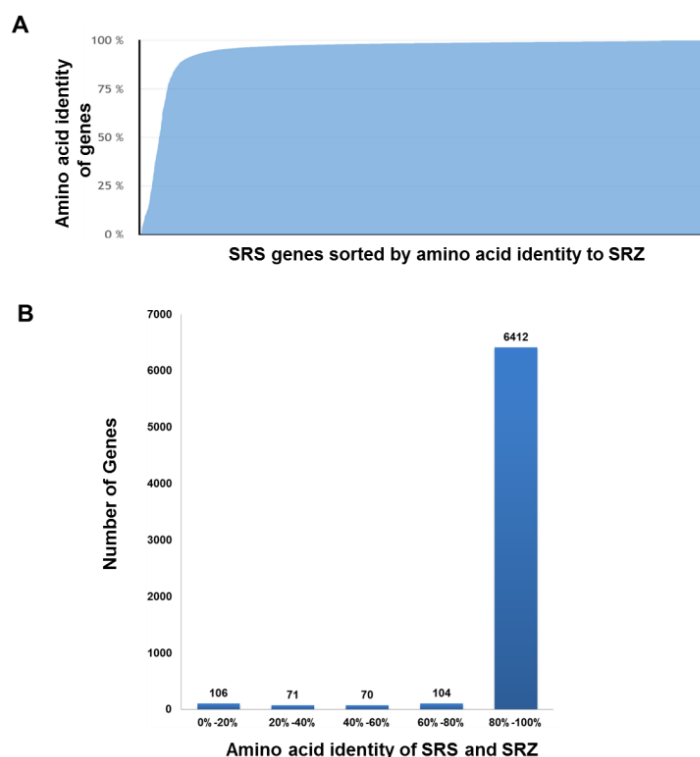
Figure. 4s: Genes of SRS and SRZ share high sequence identity. **A**, All genes from both *S. reilianum formae speciales* SRS and SRS are sorted according to amino acid identity based on SIMAP values. **B**, More than 90% of all genes (over 6000) have a similarity over 95%. The SIMAP values were procured from the data of the original work by - Getrud Manhaupt, Ulrich Güldner and Martin Münsterkötter. Tables provided by J. Schirawski.

Table. 2s: Table of abbreviations used in this study

Abbreviation	Description
Δ	Deletion
aa	Amino acid
bp	Base pair
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide
dpi	Days post inoculation
EDTA	Ethylenediaminetetraacetic acid
EM	Emission
EX	Excitation
f. sp.	Forma specialis

gDNA	Genomic deoxyribonucleic acid
h	Hour
LF	Left flank
max	Maximum
min	Minute
min	Minimum
mRNA	Messenger RNA
mRNA	Messenger RNA
OD	Optical density
PBS	Phosphate buffer saline
PCR	Polymerase chain reaction
PD(A)	Potato dextrose (agar)
RF	Right flank
ROI	Region of interest
RPKM	Reads per kilo base per million
rpm	Rounds per minute
SIMAP	Similarity matrix of proteins
sr	SRZ
SRS	*S. reilianum* f. sp. *reilianum*
SRZ	*S. reilianum* f. sp. *zeae*
TAE	Tris-Acetate + Na_2-EDTA
TBE	Tris-Borate + Na_2-EDTA
TE	Tris-EDTA
Tris	Tris(hydroxymethyl)aminomethan
WGA-AF	Wheat germ agglutinin conjugated with Alexa fluor
WT	Wild type
YEPS	Yeast extract peptone sucrose

Table. 3s: List of all 189 SRSZ strains used as three biological replicates for sorghum infection after mating with SRS2_H2-7. The spore total (spore, spore+leafy, spore+healthy) after three rounds of biological replicates are represented in percentage. Number of plants used for each strain = 11). Digits after decimal points are rounded up.

SRSZ strains	spore total	spore total	spore total	average	SD*	SE*
TW114	0	0	0	0	0	0
TW139	0	0	0	0	0	0
TW164	0	0	0	0	0	0
TW169	0	0	0	0	0	0

TW260	0	0	0	0	0	0
TW262	0	0	0	0	0	0
TW288	0	0	0	0	0	0
TW297	0	0	0	0	0	0
TW298	0	0	0	0	0	0
TW328	0	0	0	0	0	0
TW336	0	0	0	0	0	0
TW401	0	0	0	0	0	0
TW442	0	0	0	0	0	0
TW457	0	0	0	0	0	0
TW475	0	0	0	0	0	0
TW478	0	0	0	0	0	0
TW487	0	0	0	0	0	0
TW491	0	0	0	0	0	0
TW492	0	0	0	0	0	0
TW493	0	0	0	0	0	0
TW496	0	0	0	0	0	0
TW500	0	0	0	0	0	0
TW501	0	0	0	0	0	0
TW508	0	0	0	0	0	0
TW511	0	0	0	0	0	0
TW518	0	0	0	0	0	0
TW547	0	0	0	0	0	0
TW87	0	0	0	0	0	0
TW155	0	0	0	0	0	0
TW240	0	0	0	0	0	0
TW437	0	0	0	0	0	0
TW510	0	0	0	0	0	0
TW79	0	0	0	0	0	0
TW165	0	0	0	0	0	0
TW154	0	0	0	0	0	0
TW308	0	0	0	0	0	0
TW360	0	0	0	0	0	0
TW376	0	0	0	0	0	0
TW406	0	0	0	0	0	0
TW425	0	0	0	0	0	0
TW426	0	0	0	0	0	0
TW435	0	0	0	0	0	0
TW438	0	0	0	0	0	0
TW124	0	0	0	0	0	0
TW156	0	0	0	0	0	0
TW157	0	0	0	0	0	0
TW204	0	0	0	0	0	0
TW246	0	0	0	0	0	0

TW248	0	0	0	0	0	0
TW254	0	0	0	0	0	0
TW265	0	0	0	0	0	0
TW266	0	0	0	0	0	0
TW269	0	0	0	0	0	0
TW277	0	0	0	0	0	0
TW280	0	0	0	0	0	0
TW287	0	0	0	0	0	0
TW322	0	0	0	0	0	0
TW341	0	0	0	0	0	0
TW368	0	0	0	0	0	0
TW413	0	0	0	0	0	0
TW414	0	0	0	0	0	0
TW450	0	0	0	0	0	0
TW453	0	0	0	0	0	0
TW455	0	0	0	0	0	0
TW477	0	0	0	0	0	0
TW99	0	0	0	0	0	0
SRZ	0	0	0	0	0	0
TW107	0	0	0	0	0	0
TW125	0	0	0	0	0	0
TW135	0	0	0	0	0	0
TW138	0	0	0	0	0	0
TW182	0	0	0	0	0	0
TW193	0	0	0	0	0	0
TW200	0	0	0	0	0	0
TW201	0	0	0	0	0	0
TW215	0	0	0	0	0	0
TW224	0	0	0	0	0	0
TW229	0	0	0	0	0	0
TW235	0	0	0	0	0	0
TW263	0	0	0	0	0	0
TW271	0	0	0	0	0	0
TW274	0	0	0	0	0	0
TW281	0	0	0	0	0	0
TW295	0	0	0	0	0	0
TW314	0	0	0	0	0	0
TW333	0	0	0	0	0	0
TW342	0	0	0	0	0	0
TW349	0	0	0	0	0	0
TW353	0	0	0	0	0	0
TW364	0	0	0	0	0	0
TW369	0	0	0	0	0	0
TW388	0	0	0	0	0	0

TW393	0	0	0	0	0	0
TW403	0	0	0	0	0	0
TW409	0	0	0	0	0	0
TW412	0	0	0	0	0	0
TW423	0	0	0	0	0	0
TW424	0	0	0	0	0	0
TW433	0	0	0	0	0	0
TW439	0	0	0	0	0	0
TW464	0	0	0	0	0	0
TW524	0	0	0	0	0	0
TW550	0	0	0	0	0	0
TW331	0	0	0	0	0	0
TW445	0	0	0	0	0	0
TW326	0	0	0	0	0	0
TW348	0	0	0	0	0	0
TW434	0	0	0	0	0	0
TW384	0	0	0	0	0	0
TW149	0	0	0	0	0	0
TW150	0	0	0	0	0	0
TW227	0	10	9	6	6	2
TW520	20	0	40	20	20	4
TW407	0	38	25	21	19	4
TW234	17	33	20	23	9	2
TW305	30	17	33	27	9	1
TW225	0	30	70	33	35	6
TW351	9	18	80	36	39	6
TW284	0	55	55	36	31	5
TW382	80	14	30	41	34	5
TW485	20	40	80	47	31	4
TW242	50	20	86	52	33	5
TW467	36	45	90	57	29	4
TW226	90	78	13	60	42	5
TW163	80	60	60	67	12	1
TW159	64	86	57	69	15	2
TW276	100	20	90	70	44	5
TW299	100	67	50	72	25	3
TW208	100	20	100	73	46	5
TW392	100	43	90	78	31	3
TW375	100	80	67	82	17	2
TW432	60	100	90	83	21	2
TW332	70	80	100	83	15	2
TW345	100	80	90	90	10	1
TW315	72	100	100	90	16	2
TW522	90	90	100	93	6	0.6

TW110	100	90	90	93	6	0.6
TW484	90	100	100	97	6	0.6
TW132	100	90	100	97	6	0.6
TW309	100	90	100	97	6	0.6
TW243	100	100	90	97	6	0.6
TW427	100	100	100	100	0.005	0.0005
TW89	100	100	100	100	0	0
TW378	100	100	100	100	0	0
TW404	100	100	100	100	0	0
TW239	100	100	100	100	0	0
TW466	100	100	100	100	0	0
TW233	100	100	100	100	0	0
TW319	100	100	100	100	0	0
TW78	100	100	100	100	0	0
TW131	100	100	100	100	0	0
TW421	100	100	100	100	0	0
TW230	100	100	100	100	0	0
TW444	100	100	100	100	0	0
TW470	100	100	100	100	0	0
TW118	100	100	100	100	0	0
TW161	100	100	100	100	0	0
TW173	100	100	100	100	0	0
TW177	100	100	100	100	0	0
TW241	100	100	100	100	0	0
TW311	100	100	100	100	0	0
TW343	100	100	100	100	0	0
TW461	100	100	100	100	0	0
TW469	100	100	100	100	0	0
TW479	100	100	100	100	0	0
TW488	100	100	100	100	0	0
TW521	100	100	100	100	0	0
TW540	100	100	100	100	0	0
TW285	100	100	100	100	0	0
TW171	100	100	100	100	0	0
TW210	100	100	100	100	0	0
TW212	100	100	100	100	0	0
TW329	100	100	100	100	0	0
TW418	100	100	100	100	0	0
TW449	100	100	100	100	0	0
TW451	100	100	100	100	0	0
TW513	100	100	100	100	0	0
TW160	100	100	100	100	0	0
TW196	100	100	100	100	0	0
TW323	100	100	100	100	0	0

TW337	100	100	100	100	0	0
TW408	100	100	100	100	0	0
TW294	100	100	100	100	0	0
SRS	100	100	100	100	0	0
TW128	100	100	100	100	0	0
TW162	100	100	100	100	0	0
TW327	100	100	100	100	0	0
TW335	100	100	100	100	0	0
TW362	100	100	100	100	0	0
TW505	100	100	100	100	0	0
TW542	100	100	100	100	0	0

*SD=standard deviation

*SE=standard error of mean

Table. 4s: List of all the phytoalexin quantification (*) of the 189 SRSZ strains.

SRSZ strain	phytoalexin	SRSZ strain	phytoalexin	SRSZ strain	phytoalexin	SRSZ strain	phytoalexin
TW114	3	TW201	3	TW265	2	TW163	0
TW139	3	TW215	3	TW266	1	TW159	0
TW164	1	TW224	3	TW269	1	TW276	0
TW169	1	TW229	1	TW277	3	TW299	0
TW260	2	TW235	1	TW280	3	TW208	0
TW262	3	TW263	3	TW287	1	TW392	0
TW288	1	TW271	1	TW322	2	TW375	0
TW297	3	TW274	2	TW341	1	TW432	0
TW298	1	TW281	2	TW368	2	TW332	0
TW328	3	TW295	2	TW413	0	TW345	0
TW336	1	TW314	3	TW414	3	TW315	0
TW401	3	TW333	2	TW450	1	TW522	0
TW442	2	TW342	1	TW453	2	TW110	0
TW457	3	TW349	2	TW455	3	TW484	0
TW475	2	TW353	3	TW477	1	TW132	0
TW478	3	TW364	2	TW99	3	TW309	0
TW487	3	TW369	3	SRZ	3	TW243	0
TW491	1	TW388	1	TW107	2	TW427	0
TW492	3	TW393	3	TW125	3	TW89	0
TW493	2	TW403	2	TW135	2	TW378	0

TW496	2	TW409	2	TW138	3	TW404	0
TW500	3	TW412	3	TW182	1	TW239	0
TW501	2	TW423	2	TW193	3	TW466	0
TW508	3	TW424	2	TW200	2	TW233	0
TW511	3	TW433	1	TW319	0	TW461	0
TW518	1	TW439	1	TW78	0	TW469	0
TW547	2	TW464	2	TW131	0	TW479	0
TW87	3	TW524	2	TW421	0	TW488	0
TW155	2	TW550	1	TW230	0	TW521	0
TW240	1	TW331	0	TW444	0	TW540	0
TW437	2	TW445	0	TW470	0	TW285	0
TW510	3	TW326	3	TW118	0	TW171	0
TW79	2	TW348	3	TW161	2	TW210	0
TW165	2	TW434	1	TW173	0	TW212	3
TW154	3	TW384	1	TW177	2	TW329	0
TW308	3	TW149	3	TW241	0	TW418	0
TW360	3	TW150	3	TW311	3	TW449	0
TW376	1	TW227	0	TW343	0	TW451	0
TW406	2	TW520	0	TW294	0	TW513	2
TW425	1	TW407	0	SRS	0	TW160	2
TW426	3	TW234	0	TW128	3	TW196	0
TW435	1	TW305	0	TW162	0	TW323	0
TW438	2	TW225	0	TW327	0	TW337	0
TW124	1	TW351	0	TW335	0	TW408	0
TW156	2	TW284	0	TW362	0	TW505	0
TW157	2	TW382	0	TW242	0	TW542	0
TW204	1	TW485	0	TW467	0	TW254	1
TW246	3	TW248	3	TW226	0		

* 0= no red pigmentation, 1&2 = intermediate pigmentation and 3 = the highest pigmentation

Table. 5s: List of all the SRSZ strains used to infect sorghum after mating with SRS2_H2-7. The evaluation is shown in percentage after dividing the number of inflorescences with a certain phenotype by the number of total inflorescences (*100), Number of plants used for each strain=11. The digits after decimal points are rounded up.

SRSZ strains	Spore	spore+leafy	Spore+Healthy	Leafy	Healthy	Spores total

TW103	0	0	0	0	100	0
TW107	0	0	0	0	100	0
TW114	0	0	0	0	100	0
TW125	0	0	0	0	100	0
TW126	0	0	0	0	100	0
TW130	0	0	0	0	100	0
TW133	0	0	0	0	100	0
TW135	0	0	0	0	100	0
TW137	0	0	0	0	100	0
TW138	0	0	0	0	100	0
TW139	0	0	0	0	100	0
TW142	0	0	0	0	100	0
TW143	0	0	0	0	100	0
TW144	0	0	0	0	100	0
TW145	0	0	0	0	100	0
TW146	0	0	0	0	100	0
TW147	0	0	0	0	100	0
TW149	0	0	0	0	100	0
TW150	0	0	0	0	100	0
TW154	0	0	0	0	100	0
TW155	0	0	0	0	100	0
TW156	0	0	0	0	100	0
TW157	0	0	0	0	100	0
TW550	0	0	0	0	100	0
TW164	0	0	0	0	100	0
TW165	0	0	0	0	100	0
TW167	0	0	0	0	100	0
TW169	0	0	0	0	100	0
TW249	0	0	0	0	100	0
TW326	0	0	0	0	100	0
TW368	0	0	0	0	100	0
TW369	0	0	0	0	100	0
TW388	0	0	0	0	100	0
TW398	0	0	0	0	100	0
TW426	0	0	0	0	100	0
TW433	0	0	0	0	100	0
TW434	0	0	0	0	100	0
TW435	0	0	0	0	100	0
TW437	0	0	0	0	100	0
TW438	0	0	0	0	100	0
TW439	0	0	0	0	100	0
TW442	0	0	0	0	100	0
TW447	0	0	0	0	100	0

TW450	0	0	0	0	100	0
TW453	0	0	0	0	100	0
TW455	0	0	0	0	100	0
TW457	0	0	0	0	100	0
TW458	0	0	0	0	100	0
TW464	0	0	0	0	100	0
TW472	0	0	0	0	100	0
TW473	0	0	0	0	100	0
TW475	0	0	0	0	100	0
TW477	0	0	0	0	100	0
TW478	0	0	0	0	100	0
TW487	0	0	0	0	100	0
TW491	0	0	0	0	100	0
TW492	0	0	0	0	100	0
TW493	0	0	0	0	100	0
TW496	0	0	0	0	100	0
TW500	0	0	0	0	100	0
TW501	0	0	0	0	100	0
TW506	0	0	0	0	100	0
TW508	0	0	0	0	100	0
TW510	0	0	0	0	100	0
TW511	0	0	0	0	100	0
TW512	0	0	0	0	100	0
TW518	0	0	0	0	100	0
TW524	0	0	0	0	100	0
TW534	0	0	0	0	100	0
TW74	0	0	0	0	100	0
TW75	0	0	0	0	100	0
TW76	0	0	0	0	100	0
TW79	0	0	0	0	100	0
TW83	0	0	0	0	100	0
TW85	0	0	0	0	100	0
TW87	0	0	0	0	100	0
TW88	0	0	0	0	100	0
TW97	0	0	0	0	100	0
TW99	0	0	0	0	100	0
TW547	0	0	0	0	100	0
TW 305	0	0	0	0	100	0
TW 308	0	0	0	0	100	0
TW 309	0	0	0	0	100	0
TW 314	0	0	0	0	100	0
TW 318	0	0	0	0	100	0
TW 319	0	0	0	0	100	0

TW 322	0	0	0	0	100	0
TW 323	0	0	0	0	100	0
TW193	0	0	0	0	100	0
TW200	0	0	0	0	100	0
TW201	0	0	0	0	100	0
TW204	0	0	0	0	100	0
TW207	0	0	0	0	100	0
TW209	0	0	0	0	100	0
TW215	0	0	0	0	100	0
TW216	0	0	0	0	100	0
TW224	0	0	0	0	100	0
TW229	0	0	0	0	100	0
TW235	0	0	0	0	100	0
TW240	0	0	0	0	100	0
TW246	0	0	0	0	100	0
TW247	0	0	0	0	100	0
TW248	0	0	0	0	100	0
TW254	0	0	0	0	100	0
TW257	0	0	0	0	100	0
TW260	0	0	0	0	100	0
TW262	0	0	0	0	100	0
TW263	0	0	0	0	100	0
TW265	0	0	0	0	100	0
TW266	0	0	0	0	100	0
TW268	0	0	0	0	100	0
TW269	0	0	0	0	100	0
TW274	0	0	0	0	100	0
TW277	0	0	0	0	100	0
TW280	0	0	0	0	100	0
TW281	0	0	0	0	100	0
TW282	0	0	0	0	100	0
TW288	0	0	0	0	100	0
TW295	0	0	0	0	100	0
TW297	0	0	0	0	100	0
TW328	0	0	0	0	100	0
TW333	0	0	0	0	100	0
TW336	0	0	0	0	100	0
TW341	0	0	0	0	100	0
TW342	0	0	0	0	100	0
TW348	0	0	0	0	100	0
TW349	0	0	0	0	100	0
TW353	0	0	0	0	100	0
TW360	0	0	0	0	100	0

TW364	0	0	0	0	100	0
TW372	0	0	0	0	100	0
TW376	0	0	0	0	100	0
TW384	0	0	0	0	100	0
TW393	0	0	0	0	100	0
TW397	0	0	0	0	100	0
TW401	0	0	0	0	100	0
TW403	0	0	0	0	100	0
TW406	0	0	0	0	100	0
TW409	0	0	0	0	100	0
TW412	0	0	0	0	100	0
TW413	0	0	0	0	100	0
TW423	0	0	0	0	100	0
TW424	0	0	0	0	100	0
TW425	0	0	0	0	100	0
TW428	0	0	0	0	100	0
TW385	0	0	0	18	82	0
TW244	0	0	0	25	75	0
TW340	0	0	0	29	71	0
TW399	0	0	0	29	71	0
TW462	0	0	0	33	67	0
TW379	0	0	0	36	64	0
TW136	0	0	0	38	63	0
TW245	0	0	0	40	60	0
TW363	0	0	0	40	60	0
TW250	0	0	0	46	55	0
TW535	0	0	0	50	50	0
TW536	0	0	0	57	43	0
TW533	0	0	0	60	40	0
TW441	0	0	0	64	36	0
TW531	0	0	0	64	36	0
TW517	0	0	0	67	33	0
TW283	0	0	0	67	33	0
TW331	0	0	0	75	25	0
TW225	0	0	0	83	17	0
TW445	0	0	0	100	0	0
TW227	0	0	0	92	8	0
TW284	0	0	0	100	0	0
TW407	0	0	0	100	0	0
TW182	0	0	0	0	100	0
TW124	0	0	0	0	100	0
TW271	0	0	0	0	100	0
TW287	0	0	0	0	100	0

TW414	0	0	0	0	100	0
TW308	0	0	0	0	100	0
TW322	0	0	0	0	100	0
TW314	0	0	0	0	100	0
TW298	0	0	0	0	100	0
TW166	7	0	0	0	93	7
TW 303	7	0	0	7	86	7
TW106	0	0	8	0	92	8
TW279	8	0	0	0	92	8
TW351	0	9	0	36	55	9
TW168	9	0	0	0	91	9
TW416	0	9	0	91	0	9
TW452	9	0	0	0	91	9
TW446	0	9	0	18	73	9
TW77	10	0	0	0	90	10
TW91	11	0	0	0	89	11
TW86	0	11	0	11	78	11
TW431	0	11	0	0	89	11
TW273	0	13	0	0	88	13
TW389	13	0	0	0	88	13
TW539	0	13	0	88	0	13
TW134	14	0	0	0	86	14
TW334	14	0	0	14	71	14
TW490	14	0	0	57	29	14
TW234	17	0	0	0	83	17
TW253	0	17	0	42	42	17
TW371	17	0	0	0	83	17
TW251	0	18	0	45	36	18
TW459	0	18	0	36	45	18
TW419	0	18	0	0	82	18
TW485	20	0	0	0	80	20
TW520	0	20	0	0	80	20
TW275	0	0	20	0	80	20
TW101	20	0	0	0	80	20
TW528	20	0	0	20	60	20
TW359	22	0	0	67	11	22
TW430	22	0	0	0	78	22
TW129	8	15	0	8	69	23
TW339	0	23	0	31	46	23
TW181	17	8	0	33	42	25
TW 307	0	25	0	33	42	25
TW514	27	0	0	0	73	27
TW516	27	0	0	9	64	27

TW386	0	14	14	0	71	28
TW293	7	21	0	0	71	29
TW127	29	0	0	0	71	29
TW395	0	29	0	14	57	29
TW538	14	14	0	29	43	29
TW113	30	0	0	0	70	30
TW305	20	10	0	10	60	30
TW 313	0	31	0	0	69	31
TW347	0	33	0	22	44	33
TW338	8	25	0	25	42	33
TW486	33	0	0	0	67	33
TW467	0	36	0	0	64	36
TW356	36	0	0	0	64	36
TW300	0	36	0	18	45	36
TW358	0	38	0	0	63	38
TW258	13	25	0	0	63	38
TW383	0	38	0	13	50	38
TW228	42	0	0	25	33	42
TW255	0	43	0	36	22	43
TW102	45	0	0	0	55	45
TW 324	46	0	0	0	54	46
TW242	50	0	0	0	50	50
TW350	0	50	0	0	50	50
TW396	20	30	0	0	50	50
TW301	50	0	0	0	50	50
TW 310	50	0	0	0	50	50
TW532	50	0	0	10	40	50
TW366	40	10	0	40	10	50
TW84	55	0	0	0	45	55
TW 306	0	55	0	36	9	55
TW481	9	45	0	0	45	55
TW499	9	45	0	0	45	55
TW526	55	0	0	0	45	55
TW471	27	27	0	18	27	55
TW170	0	56	0	22	22	56
TW174	56	0	0	44	0	56
TW489	33	22	0	0	44	56
TW211	57	0	0	0	43	57
TW390	57	0	0	0	43	57
TW98	58	0	0	0	42	58
TW432	60	0	0	10	30	60
TW480	10	50	0	0	40	60
TW222	20	40	0	0	40	60

TW302	60	0	0	0	40	60
TW507	0	60	0	20	20	60
TW495	38	25	0	0	38	63
TW410	63	0	0	0	38	63
TW259	64	0	0	0	36	64
TW237	0	64	0	18	18	64
TW159	64	0	0	0	36	64
TW420	64	0	0	0	36	64
TW195	0	64	0	14	21	64
TW141	33	33	0	0	33	67
TW 320	58	8	0	0	33	67
TW463	33	33	0	0	33	67
TW482	67	0	0	0	33	67
TW291	67	0	0	0	33	67
TW252	0	67	0	17	17	67
TW332	20	50	0	30	0	70
TW448	60	10	0	0	30	70
TW194	0	72	0	0	29	72
TW 315	72	0	0	0	29	72
TW315	72	0	0	0	29	72
TW81	55	18	0	27	0	73
TW261	73	0	0	0	27	73
TW330	0	73	0	27	0	73
TW483	0	73	0	0	27	73
TW502	18.	55	0	18	9	73
TW415	44	33	0	0	22	78
TW112	78	0	0	0	22	78
TW381	56	22	0	0	22	78
TW391	0	78	0	22	0	78
TW236	0	78	0	0	22	78
TW380	78	0	0	0	22	78
TW175	79	0	0	0	21	79
TW 316	79	0	0	7	14	79
TW163	80	0	0	0	20	80
TW382	0	80	0	20	0	80
TW503	20	60	0	0	20	80
TW 312	50	30	0	0	20	80
TW272	60	20	0	0	20	80
TW387	80	0	0	0	20	80
TW402	10	70	0	20	0	80
TW417	20	60	0	20	0	80
TW400	45	36	0	0	18	82
TW474	0	81	0	0	18	82

TW221	0	83	0	0	17	83
TW465	25	58	0	0	17	83
TW278	67	17	0	0	17	83
TW180	43	43	0	0	14	86
TW497	14	71	0	14	0	86
TW352	38	50	0	0	13	88
TW545	63	25	0	0	13	88
TW361	67	22	0	0	11	89
TW484	80	10	0	10	0	90
TW522	90	0	0	10	0	90
TW226	20	70	0	10	0	90
TW537	70	20	0	0	10	90
TW530	82	9	0	0	9	91
TW256	82	9	0	0	9	91
TW290	77	14	0	7	0	93
TW179	85	15	0	0	0	100
TW110	90	10	0	0	0	100
TW131	100	0	0	0	0	100
TW132	100	0	0	0	0	100
TW171	100	0	0	0	0	100
TW375	10	90	0	0	0	100
TW392	0	100	0	0	0	100
TW444	100	0	0	0	0	100
TW466	90	10	0	0	0	100
TW470	100	0	0	0	0	100
TW479	100	0	0	0	0	100
TW521	100	0	0	0	0	100
TW78	50	50	0	0	0	100
TW89	100	0	0	0	0	100
TW208	90	10	0	0	0	100
TW230	100	0	0	0	0	100
TW233	90	10	0	0	0	100
TW241	100	0	0	0	0	100
TW243	90	10	0	0	0	100
TW276	80	20	0	0	0	100
TW294	100	0	0	0	0	100
TW327	100	0	0	0	0	100
TW329	100	0	0	0	0	100
TW335	100	0	0	0	0	100
TW345	100	0	0	0	0	100
TW404	90	10	0	0	0	100
TW408	100	0	0	0	0	100
TW418	100	0	0	0	0	100

TW421	90	10	0	0	0	100
TW80	100	0	0	0	0	100
TW285	100	0	0	0	0	100
TW 317	0	100	0	0	0	100
TW289	0	100	0	0	0	100
TW296	23	77	0	0	0	100
TW460	45	55	0	0	0	100
TW264	50	50	0	0	0	100
TW529	57	43	0	0	0	100
TW286	86	14	0	0	0	100
TW238	93	7	0	0	0	100
TW118	100	0	0	0	0	100
TW128	100	0	0	0	0	100
TW160	100	0	0	0	0	100
TW161	100	0	0	0	0	100
TW162	100	0	0	0	0	100
TW177	100	0	0	0	0	100
TW239	100	0	0	0	0	100
TW436	100	0	0	0	0	100
TW449	100	0	0	0	0	100
TW451	100	0	0	0	0	100
TW461	100	0	0	0	0	100
TW469	100	0	0	0	0	100
TW488	100	0	0	0	0	100
TW505	100	0	0	0	0	100
TW513	100	0	0	0	0	100
TW540	100	0	0	0	0	100
TW542	100	0	0	0	0	100
TW311	100	0	0	0	0	100
TW196	100	0	0	0	0	100
TW210	100	0	0	0	0	100
TW212	100	0	0	0	0	100
TW299	100	0	0	0	0	100
TW337	100	0	0	0	0	100
TW343	100	0	0	0	0	100
TW362	100	0	0	0	0	100
TW378	100	0	0	0	0	100
TW427	100	0	0	0	0	100
TW319	100	0	0	0	0	100
TW309	60	40	0	0	0	100
TW173	100	0	0	0	0	100
TW323	100	0	0	0	0	100

Table. 6s: List of quantification (*) of grey scale measurement of 187 SRSZ strains.

sample	mean	sample	mean	sample	mean	sample	mean
TW107	161.96	TW200	165.90	TW277	180.81	TW342	158.20
TW110	82.26	TW201	201.50	TW280	165.36	TW343	132.92
TW114	174.06	TW204	99.44	TW281	200.47	TW345	153.95
TW118	166.87	TW208	98.34	TW284	188.79	TW348	153.47
TW124	179.81	TW210	176.95	TW285	97.01	TW349	171.60
TW125	93.15	TW212	171.97	TW287	103.21	TW351	161.45
TW128	186.72	TW215	123.92	TW288	208.77	TW353	189.10
TW131	164.66	TW224	117.62	TW294	171.72	TW360	186.65
TW132	130.60	TW226	226.43	TW295	123.20	TW362	176.59
TW135	83.66	TW227	173.34	TW297	203.87	TW364	189.64
TW138	138.07	TW229	132.95	TW298	128.95	TW368	195.68
TW139	181.21	TW230	213.75	TW299	147.06	TW369	157.48
TW149	158.97	TW233	135.64	TW305	168.16	TW375	178.74
TW150	190.81	TW234	204.86	TW308	178.13	TW376	188.04
TW154	203.13	TW235	153.82	TW309	160.95	TW378	186.09
TW155	158.96	TW239	196.16	TW311	144.99	TW382	221.05
TW156	162.51	TW240	154.01	TW314	185.83	TW384	218.81
TW157	188.27	TW241	117.24	TW315	208.33	TW388	202.49
TW159	189.80	TW242	178.30	TW319	197.75	TW392	201.88
TW160	189.00	TW243	151.08	TW322	165.11	TW393	226.35
TW161	181.05	TW246	124.96	TW323	187.77	TW401	215.21
TW162	188.20	TW248	170.10	TW326	208.32	TW403	201.36
TW163	168.41	TW254	175.07	TW327	147.32	TW404	171.51
TW164	207.14	TW260	107.21	TW328	182.64	TW406	189.40
TW165	179.14	TW262	201.65	TW329	190.83	TW408	172.04
TW169	158.07	TW263	144.40	TW331	172.59	TW409	180.84
TW171	124.62	TW265	105.07	TW332	181.32	TW412	175.05
TW173	142.46	TW266	147.27	TW333	173.44	TW413	199.22
TW177	222.77	TW269	187.82	TW335	215.14	TW414	139.69
TW182	212.86	TW271	192.83	TW336	219.07	TW418	174.33
TW193	201.52	TW274	182.58	TW337	213.69	TW421	161.56
TW196	75.46	TW276	190.04	TW341	189.70	TW423	190.78
TW487	96.20	TW500	207.14	TW513	132.66	TW540	141.01
TW488	184.84	TW501	100.48	TW518	121.38	TW542	120.32
TW491	191.04	TW505	99.38	TW520	201.57	TW547	191.85
TW492	101.63	TW508	182.47	TW521	217.61	TW550	112.09
TW493	134.41	TW510	198.43	TW522	173.22	TW78	147.31
TW496	184.39	TW511	204.90	TW524	77.93	TW79	156.39

 Nilam Nayan Borah

TW424	214.10	TW442	146.36	TW466	164.21	TW89	198.45
TW425	189.36	TW444	161.07	TW467	151.47	TW99	77.45
TW426	186.54	TW445	126.86	TW469	211.80	SRS H2-7	124.87
TW427	187.09	TW449	138.25	TW470	160.06	SRS H2-8	137.72
TW432	188.10	TW450	162.47	TW475	105.81	SRZ 5-1	166.15
TW433	163.24	TW451	144.11	TW477	128.11	SRZ 5-2	168.10
TW434	179.25	TW453	173.10	TW478	107.17		
TW435	176.33	TW455	133.96	TW479	154.38		
TW437	169.44	TW457	155.39	TW484	177.81		
TW438	138.95	TW461	113.46	TW485	123.09		
TW439	176.35	TW464	180.27	TW87	172.79		

*The imaging software ImageJ "measurement" tool gave a maximum and minimum score of the grey intensity along with their mean values for an area of about 780 sq mm.

Table. 7s: List of colony texture categorization (*) of 187 SRSZ strains

sample	colony morphology	sample	colony morphology	sample	colony morphology	sample	colony morphology
TW107	5	TW200	3	TW277	2	TW342	3
TW110	6	TW201	3	TW280	2	TW343	2
TW114	1	TW204	1	TW281	5	TW345	3
TW118	4	TW208	2	TW284	4	TW348	4
TW124	3	TW210	1	TW285	6	TW349	3
TW125	6	TW212	6	TW287	1	TW351	6
TW128	4	TW215	1	TW288	3	TW353	5
TW131	3	TW224	4	TW294	2	TW360	1
TW132	5	TW226	5	TW295	1	TW362	3
TW135	4	TW227	4	TW297	2	TW364	5
TW138	1	TW229	1	TW298	5	TW368	3
TW139	4	TW230	3	TW299	4	TW369	4
TW149	5	TW233	5	TW305	3	TW375	3
TW150	1	TW234	5	TW308	3	TW376	4
TW154	3	TW235	6	TW309	5	TW378	1
TW155	1	TW239	2	TW311	6	TW382	5
TW156	1	TW240	3	TW314	1	TW384	4
TW157	4	TW241	6	TW315	3	TW388	5
TW159	4	TW242	1	TW319	3	TW392	1
TW160	4	TW243	1	TW322	2	TW393	3
TW161	4	TW246	1	TW323	2	TW401	3
TW162	1	TW248	5	TW326	2	TW403	4
TW163	4	TW254	3	TW327	3	TW404	5
TW164	3	TW260	2	TW328	1	TW406	4

TW165	3	TW262	6	TW329	5	TW408	5
TW169	1	TW263	1	TW331	4	TW409	5
TW171	3	TW265	6	TW332	4	TW412	2
TW173	5	TW266	5	TW333	3	TW413	3
TW177	1	TW269	1	TW335	2	TW414	3
TW182	5	TW271	3	TW336	5	TW418	1
TW193	2	TW274	3	TW337	3	TW421	3
TW196	3	TW276	1	TW341	3	TW423	3
TW487	6	TW424	3	TW442	2	TW542	5
TW488	5	TW425	1	TW444	6	TW547	3
TW491	4	TW426	1	TW445	1	TW550	1
TW492	3	TW427	5	TW449	1	TW78	5
TW493	1	TW432	1	TW450	3	TW79	5
TW496	2	TW433	5	TW451	5	TW87	5
TW500	1	TW434	1	TW453	1	TW89	5
TW501	1	TW435	4	TW469	3	TW99	1
TW505	4	TW437	4	TW470	1	SRS H2-7	1
TW508	5	TW438	5	TW475	3	SRS H2-8	1
TW510	2	TW439	5	TW477	1	SRZ 5-1	1
TW511	1	TW455	1	TW478	1	SRZ 5-2	1
TW513	2	TW457	3	TW479	4		
TW518	1	TW461	6	TW484	1		
TW520	3	TW464	3	TW485	1		
TW521	4	TW466	1	TW524	3		
TW522	3	TW467	4	TW540	1		

*1. =WT, 2. =WT with wrinkled periphery, 3. =Wrinkled, 4. =Granulated, 5. =Smooth, 6. =Slimy

Table. 8s: List of SRSZ strains used to infect maize after mating with SRZ2_ 5-1. The percentage of total spore (male+female inflorescence) was calculated by dividing the number of total spore forming inflorescence by the total number of inflorescence *100. The symptoms were scored 8 to 10 weeks after inoculation.

SRSZ strain	%male and female spores	SRSZ strain	%male and female spores	SRSZ strain	%male and female spores	SRSZ strain	%male and female spores
TW114	0	TW522	20	TW461	45	TW79	75

TW164	0	TW432	22	TW378	46	TW156	75
TW260	0	TW297	22	TW315	47	TW505	76
TW288	0	TW524	22	TW294	47	TW345	76
TW401	0	TW484	22	TW243	50	TW511	77
TW475	0	TW226	22	TW470	50	TW477	77
TW487	0	TW132	23	TW173	50	TW125	77
TW491	0	TW234	25	TW162	50	TW550	77
TW496	0	TW521	25	TW501	50	TW298	78
TW500	0	TW154	25	TW265	50	TW182	79
TW437	0	TW 308	25	TW99	50	TW442	79
TW165	0	TW351	25	TW276	50	TW262	80
TW376	0	TW233	25	TW375	50	SRZ	80
TW406	0	TW230	27	TW311	50	TW392	80
TW435	0	TW200	29	TW362	53	TW314	83
TW157	0	TW439	29	TW413	54	TW349	83
TW246	0	TW319	29	TW239	55	TW208	86
TW254	0	TW513	29	TW281	56	TW478	92
TW277	0	TW407	30	TW242	56	TW449	92
TW280	0	TW547	30	TW348	56	TW285	94
TW368	0	TW384	30	TW89	56	TW124	100
TW453	0	TW271	31	TW444	56	TW287	100
TW455	0	TW204	31	TW404	57	TW540	100
TW193	0	TW169	33	TW418	58	TW336	100
TW274	0	TW240	33	TW87	59	TW457	100
TW333	0	TW248	33	TW295	59	TW360	100
TW342	0	TW235	33	TW128	60	TW341	100
TW353	0	TW155	36	TW493	60	TW107	100
TW364	0	TW510	37	TW332	60	TW224	100
TW388	0	TW414	38	TW110	62	TW263	100
TW393	0	TW450	38	TW328	63	TW423	100
TW403	0	TW138	38	TW369	63	TW464	100
TW433	0	TW149	38	TW412	63	TW466	100
TW227	0	TW131	38	TW485	63	TW78	100
SRS	0	TW215	38	TW171	63	TW421	100
TW424	7	TW322	38	TW518	65	TW299	100
TW467	8	TW 309	40	TW229	65	TW329	100
TW426	8	TW 323	40	TW438	67		
TW520	11	TW508	43	TW266	67		
TW163	15	TW343	43	TW135	67		
TW492	17	TW479	43	TW542	67		
TW201	17	TW488	43	TW118	67		
TW382	17	TW210	43	TW427	68		
TW469	18	TW150	44	TW337	68		
TW408	20	TW335	44	TW451	69		

TW269	20	TW409	45	TW425	70
TW 305	20	TW326	45	TW161	71
TW212	20	TW177	45	TW241	71
TW327	20	TW434	45	TW139	75

Table. 9s: List of SRSZ strains used to infect maize after mating with SRS2_ H2-7 the evaluation was done similar as mentioned in table20s.

SRSZ strains	%male and female spores	SRSZ strains	%male and female spores
TW469	0	TW233	0
TW449	0	TW131	0
TW479	0	TW243	0
TW513	0	TW230	0
TW542	0	TW132	0
TW461	0	TW444	4
TW488	0	TW110	6
TW521	0	TW343	7
TW470	0	TW505	8
TW408	0	TW89	13
TW327	0	TW335	17
SRS1_ H2-8	0	TW337	18
TW239	0	TW418	21
TW78	0	SRZ1_5-2	38
SRZ	50	TW329	46
TW540	69		

7 <u>Acknowledgements</u>

8 <u>Resume</u>

Personal data

Name:	Nilam Nayan Borah
Birthday:	25.12.1986
Birthplace:	Assam, India
Address (present):	Siedlerweg19, Rheinberg 47495
Contact:	nilamborah@gmail.com, +49-17672925244
Nationality:	Indian
Status:	Unmarried

Education

05.2014-05.2018	Research scholar, iAMB, RWTH, Aachen, Germany
07.2011-02.2014	Research scholar (plant pathology), NIPGR, India
07.2009-04.2011	MSc.(Molecular biology and biotechnology), Tezpur Central University, India, 1st class with distinction (gold medal)
07.2006-06.2009	BSc. (Biotechnology), St. Anthony's college, India, 1st class
04.2003-03.2005	Higher secondary (Science), J.B. College, India, 1st division
01.1991-01.2003	Balya Bhawan School, India, 1st division with distinction

Conference Papers

2018:	Molecular methods of comparative host specificity determination in plant-pathogenic fungi- International J. of Molecular sciences
2018:	14th European Conference on Fungal Genetics, Haifa,Israel-Poster presentation
2016:	13th European Conference on Fungal Genetics, Paris, France- Poster presentation
2015:	Annual Conference of the Association for General and Applied Microbiology, Marburg, Germany- Poster presentation
2013:	International conference on Plant biotechnology, molecular medicine and human health, University of Delhi- Poster presentation
2014:	Comprehensive genome-wide analysis reveals different classes of enigmatic old yellow enzyme in fungi) – Scientific reports

Language

Assamese (Native), Hindi (Fluent), English (Fluent), German (Basic/Intermediate), Bengali (Fluent), Punjabi (Basic)

9   <u>Eidesstattliche Erklärung</u>

**Faculty of
Mathematics, Computer Science
and Natural Sciences** | **RWTH**AACHEN
UNIVERSITY

Ich erkläre eidesstattlich, dass ich die vorliegende Dissertation selbständig verfasst und alle in Anspruch genommenen Hilfen in der Dissertation angegeben habe. Des Weiteren erkläre ich, dass die vorliegende Dissertation nicht bereits als Diplomarbeit oder vergleichbare Prüfungsarbeit verwendet worden ist.

Aachen,den   *Unterschrift*

(Nilam Nayan Borah)