12-2018

Molecular Basis of Pathogenesis and Host Determination in Cercospora sojina: from Phenotypic to Genotypic Patterns

Wagner Calegari Fagundes
University of Arkansas, Fayetteville

Molecular Basis of Pathogenesis and Host Determination in *Cercospora sojina*: from Phenotypic to Genotypic Patterns

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Plant Pathology

by

Wagner Calegari Fagundes
Pontifical Catholic University of Rio Grande do Sul
Bachelor of Science in Biology, 2016

December 2018
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

_____
Burton H. Bluhm, Ph.D.
Thesis Director

_____
John Rupe, Ph.D.
Committee Member

_____
Ainong Shi, Ph.D.
Committee Member

Abstract

Frogeye leaf spot (FLS), caused by *Cercospora sojina*, is an important and recurrent disease of soybean in many production regions. Genetic resistance is potentially one of the most cost-effective and sustainable strategies to control FLS. However, *C. sojina* has already demonstrated the ability to overcome resistance conveyed by single R-genes (resistance genes) of soybeans, followed by the emergence of new physiological races. Although understanding population genomics and the virulence gene inventories in fungal plant pathogens is extremely important to improve disease control measures, studies regarding host specificity and pathogenesis in *C. sojina* are very limited. Therefore, the overarching goal of this study was to elucidate the genetic and molecular basis of race specificity, and pathogenesis in general, in *C. sojina*. To this end, a bulk-sequencing analysis was performed on two subcollections of *C. sojina* classified by differential infection responses (virulence or avirulence) on cultivars Blackhawk and Hood followed by mapping to the recently assembled *C. sojina* strain 2.2.3 reference genome. From the 18004 SNPs identified among the two subcollections, 75 SNPs showed an F_{st} > 0.2 and were localized within three distinct loci of the *C. sojina* genome, which harbored genes implicated in oxidative stress and pathogenesis. Unusual genomic architectures were also observed in these regions, possibly resulting from InDels or duplications in the *C. sojina* genome. Further SNP annotation analysis also identified candidate effector genes under positive selection pressure ($dN/dS > 1.0$), including two genes potentially restricted to the *Cercospora* genus. Intriguingly, *C.* cf. *flagellaris* isolates causing FLS-like lesions and *C. sojina* isolates virulent on cultivar Davis were also identified within the collection of fungal isolates, which underscores the importance of better understanding host specificity in the *C. sojina* and *Cercospora* spp. general. Altogether, this study provided key resources to unravel the genetics

and genomics of race specificity and pathogenesis in *C. sojina,* and augmented long-term efforts to improve FLS resistance in soybeans through breeding and genetic engineering approaches.

Acknowledgements

First of all, I would like to thank my primary advisor, Dr. Bluhm, for his time, for his valuable discussions inside and outside the lab and for the many opportunities this degree and his lab could give me – thank you very much. Besides him, I would like to give a special thanks to Dr. Rupe and Dr. Shi for guidance throughout my studies, which helped to make this study possible. Additionally, I would like to thank Alex Zaccaron for the great bioinformatics support, Marcio Zaccaron for technical support in the first steps of this project, and the entire Bluhm lab for discussions, support and valuable memories – thank you all for making the workspace even more enjoyable and productive.

Many thanks to all my family and friends who made me feel we were close and therefore supported me (a lot) through this journey.

Table of Contents

List of Tables

List of Figures

I. Chapter 1: Introduction

1.1 Literature review

Frogeye leaf spot (FLS) of soybean: background and epidemiology

Foliar diseases cause extensive economic impacts on crop production worldwide, and fungal pathogens are the predominant causal agents of such plant diseases (Fisher et al., 2012; Strange & Scott, 2005). In soybeans, frogeye leaf spot (FLS) caused by the fungal pathogen *Cercospora sojina* is one of the most important and common diseases in many soybean production areas throughout the world, frequently causing yield reductions from 10 to 60% (Mian, Mengistu, Wiebold, Shannon, & Wrather, 2009). FLS was first reported in Japan in 1915 and in the United States as early as 1920s (Mian et al., 2009; Phillips & Boerma, 1981). Since then, FLS has been documented in at least 27 countries throughout North and South America, Europe, Africa and Asia (Crous & Braun, 2003).

In areas with tropical or sub-tropical climates, where high humidity and warm temperatures predominate, FLS can cause severe production losses. For example, Argentina suffered severe outbreaks of FLS in the growing season of 1999/2000 and later in 2009/2010, causing losses estimated from 25 to 48% in susceptible cultivars (Carmona, Scandiani, & Luque, 2009; Ploper et al., 2001). Constant rainfall and warm temperatures in the Pampean region of Argentina favored the FLS incidence of 100% in the soybean cultivation areas (Carmona et al., 2009).

In the United States, FLS has historically occurred primarily in southern and midwestern soybean production regions due to the warm and humid conditions. However, *C. sojina* has recently moved to northern parts of the United States, with reports of FLS in Iowa, Wisconsin and Ohio (Cruz & Dorrance, 2009; Mengistu, Kurtzweil, & Grau, 2002; Yang, Uphoff, &

Sanogo, 2001). In 2009, FLS was responsible for the estimated loss of 7.5 million bushels of soybean among 28 US states (Koenning & Wrather, 2010). More recently, in 2017, yield suppression caused by FLS was around 9.27 million bushels in just 16 US states (Allen et al., 2018).

FLS manifests primarily on foliage of soybean plants, although seeds, pods, and stems can also become infected (Sherwin & Kreitlow, 1952). Common symptoms in the initial stages of the disease include small circular to angular dark-brownish spots ranging from 1 to 5mm, possibly with the presence of lighter centers (Grau, Dorrance, Bond, & Russin, 2004; Mian, Missaoui, Walker, Phillips, & Boerma, 2008). Young leaves that are not fully developed are highly susceptible, while fully expanded leaves are more resistant to invasion (Phillips, 1999). As the disease develops, lesions may merge and coalesce, forming irregular brown spots with dark reddish margins (Lehman, 1928; Phillips, 1999). From the center of the lesions, clusters of darkly-pigmented conidiophores (52-120µm x 4-4.5µm) can emerge on either side of the leaf, but tend to be more pronounced on the adaxial surface (Lehman, 1928). When 50% or more of leaf surface area is affected, lesions may cause premature defoliation and/or reduction in photosynthetic leaf area, leading to decreases in yield (Mian, Missaoui, Walker, Phillips, & Boerma, 2008). In extremely favorable conditions, the disease can spread into other plant tissues, resulting in long, narrow lesions on stems as well as elongate and slightly sunken reddish spots on pods (Mian, Missaoui, Walker, Phillips, & Boerma, 2008). *C. sojina* may also infect seeds by penetrating the pod wall, causing symptoms of light to dark grey or brown spots that may cause the seed coats to crack or flake (Phillips, 1999). Heavily infested seeds have poor germination, and the percentage of germination may be inversely related to the extent of symptomatic spots on the seed surface (Phillips, 1999; Sherwin & Kreitlow, 1952). Planting inconspicuously infested

seeds can lead to the emergence of weak seedlings with lesions on the cotyledons, which produce inoculum that may subsequently infect young leaves (Mian, Missaoui, Walker, Phillips, & Boerma, 2008).

Frogeye leaf spot is a polycyclic disease that can be prevalent throughout the growing season (Kim et al., 2013; Laviolette, Athow, Probst, Wilcox, & Abney, 1970). Consistent to many fungal pathogens, *C. sojina* favors warm (25-30°C) and humid (>90%) conditions, and can rapidly sporulate within 48h of the first visible symptoms (Mian, Missaoui, Walker, Phillips, & Boerma, 2008). Additionally, *C. sojina* can survive below 0°C in infested seeds and soybean residues (Cruz & Dorrance, 2009; Zhang, 2012). Overwintering spores can increase the inoculum load in successive growing seasons when conditions become favorable, spurring new epidemics when control measures are not properly adopted (Cruz & Dorrance, 2009).

FLS is currently controlled with foliar fungicides, seed treatments, crop rotation, biological control agents, and resistant cultivars (Nascimento et al., 2014; Pham et al., 2015; Simonetti et al., 2012; Tonelli & Fabra, 2014). Cultivars with qualitative genetic resistance have been the most cost-effective means to control FLS, and three key resistance genes (R-genes), namely *Rcs1* (Athow & Probst, 1952), *Rcs2* (Probst & Athow, 1964) and *Rcs3* (Mian, Wang, Phillips, Alvernaz, & Boerma, 1999) have been identified and deployed on fields. Although genetic resistance can be effective, the selection pressure it imposed on some *C. sojina* populations resulted in strains that overcame genetic resistance, leading to the emergence of new physiological races (Pham et al., 2015). Additionally, *C. sojina* can develop resistance to quinone outside inhibitor (QoI) fungicides, which further hinders effective FLS disease management (FRAC 2011; Zeng et al. 2015; Zhang and Bradley 2017; Zhang, Newman, and Bradley 2012).

Biology and race designations of *Cercospora sojina*

The genus *Cercospora* Fresen. (Mycosphaerellaceae, Ascomycota) is globally distributed and contains many destructive plant pathogens (Groenewald et al., 2013; Soares et al., 2015). Most members of this genus are predicted to have *Mycosphaerella* teleomorphs, as confirmed by the analyses of internal transcribed spacer (ITS) regions of ribosomal DNA (Goodwin, Dunkle, & Zismann, 2001). With more than 3,000 described species (Pollack, 1987), only two *Cercospora* species were historically reported to infect soybeans: *Cercospora kikuchii* and *Cercospora sojina* (Phillips, 1999; Soares et al., 2015). However, more recently, two other species have been described to be associated with Cercospora leaf blight and purple seed stain of soybeans, namely *Cercospora* cf. *flagellaris* and *Cercospora* cf. *sigesbeckiae* (Soares et al., 2015; Albu, Schneider, Price, & Doyle, 2016).

Even though no sexual stage has yet been observed for *C. sojina*, a recent characterization of sexual reproduction with a relatively equal distribution of mating types loci in field specimens (Kim et al., 2013), as well as the recognition of *Mycosphaerella* teleomorphs within the *Cercospora* genus (Goodwin et al., 2001), suggest that cryptic sexual reproduction is probably occurring within *C. sojina* species and, therefore, may be contributing to the overall genetic diversity of this pathogen (Kim et al., 2013).

During growth in laboratory conditions, *C. sojina* forms a typical darkly pigmented mycelium, which differs considerably from other *Cercospora* species that infect soybeans (Yeh & Sinclair, 1980). Conidia of *C. sojina* are generally septate, hyaline and slender, measuring around 5 to 7µm × 39 to 70µm and often emerge from infested plant residues and seeds (Mian, Missaoui, Walker, Phillips, & Boerma, 2008). Conidiophores may produce 1-11 asexual conidia, varying in size and shape depending on the substrate in which the fungus is growing (Lehman,

1928). Conidia can germinate on infected tissue in about 1 hour if moisture is present, producing hyphae within 18 hours in tap water at 25°C and still being viable even after 3 months of deposition on a dry leaf tissue (Lehman, 1928; Mian, Missaoui, Walker, Phillips, & Boerma, 2008). As observed for many other plant pathogenic fungi, conidia can be dispersed by air or rain splash and cause secondary infections during the growing season if environmental and host conditions are favorable (Laviolette et al., 1970; Mian, Missaoui, Walker, Phillips, & Boerma, 2008). Regarding infection, *C. sojina* possesses a distinct strategy compared to many fungi; rather than forming appressoria to penetrate host issues, *C. sojina* infects its hosts by forming branched hyphae that penetrate open stomata (Luo et al., 2017).

Cercosporin is a photoactivated toxin produced by *Cercospora* species that plays an important role in virulence. Upon absorption of visible light, cercosporin transfers energy to reactive oxygen species (ROS) as singlet oxygens ($^1O_2$) and superoxide radicals (O_2) (Beseli, Noar, & Daub, 2015; Newman & Townsend, 2016), leading to target cell damages due to the high toxicity of ROS (Daub, 1981; Daub & Chung, 2007). Some earlier reports suggested that *C. sojina* may have lost the ability to produce cercosporin (Chupp, 1954; Goodwin et al., 2001). However, the gene cluster underlying cercosporin biosynthesis (CTB cluster) has been identified in some *C. sojina* specimens (Chen, Lee, Daub, & Chung, 2007; Luo et al., 2017). Thus, important questions remain to be answered regarding how pathogenesis is activated and deployed in *C. sojina*.

The infection phenotypes of pathogens that vary in their pattern of compatible (virulence) or incompatible (avirulence) reactions on a set of host plant cultivars termed “differentials” are referred to as races (Flor, 1971; Keen, 1990). Historically, the deployment of resistant soybean cultivars kept FLS under control in the US until the of new races emerged, as was the case for

the R-genes *Rcs*1 and *Rcs*2 and the races 1 and 2 in the late 1950s (Athow & Probst, 1952; Phillips & Boerma, 1981), races 3 and 4 in the mid 1960s (Ross, 1968), and the race 5 in the late 1970s (Phillips & Boerma, 1981). Since then, several races of *C. sojina* have been reported worldwide: 22 in Brazil (Gravina et al., 2004; Yorinori, 1992); 14 races in China (Ma & Li, 1997) and 11 races throughout the US, as proposed by Mian et al. (2008) (Table S2). The *Rcs3* gene found in the cultivar "Davis" has been described to confer resistance to race 5 and all other races reported in the US so far (Missaoui, Ha, Phillips, & Boerma, 2007; Missaoui, Phillips, & Boerma, 2007). However, *C. sojina* race determinations have only been performed by phenotypic reactions on a set of genetically diverse soybean cultivars, which hinders a more precise classification of races due to the lack of a universally accepted set of soybean differential cultivars (Mian, Missaoui, Walker, Phillips, & Boerma, 2008). As an example, using the same set of 12 soybean differential cultivars as used by Mian et al. (2008), another study showed a differing number of proposed *C. sojina* races when classifying fifty *C. sojina* specimens collected on Ohio fields (Cruz & Dorrance, 2009). Additionally, it is still very unclear how *C. sojina* races have evolved and adapted for host specificity. Therefore, the need is evident the for a more profound understanding of the genetic basis of *C. sojina* pathogenesis and race population structure.

Selective pressure on pathosystems and the breakage of genetic resistance

Plants and their surrounding pathogens have been in a constant co-evolutionary arms race for millions of years in a fascinating battle of genomic diversifications and population adaptation. This antagonist, co-evolutionary selective pressure in pathosystems, had its first description by Flor in the 1940s, in which the gene-for-gene (GFG) model between a plant host

(flax) and its pathogen (the flax rust fungus *Melampsora limi*) was unraveled (Flor, 1971). Although not completely elucidated, the *C. sojina* – soybean pathosystem looks to fall into the GFG disease model, in which single resistance genes in soybean genotypes and the regular emergence of *C. sojina* races are in a permanent arms race.

The qualitative GFG model suggests a constant and strong selection for polymorphisms at pathogen virulence loci that allow these pathogens to escape recognition by host immune defenses, which are generally triggered by resistance proteins that counterattack pathogen invasion (Barrett et al., 2009). In this co-evolutionary fight, membrane-associated or cytosolic proteins encoded by plant R-genes can detect directly or indirectly the presence of avirulence (Avr) proteins of a pathogen. This detection event subsequently triggers downstream immune defenses in the form of the hypersensitive response (HR), oxidative burst, and crosstalk of plant hormones, collectively preventing pathogens from spreading past the infection site (Boyd, Ridout, O'Sullivan, Leach, & Leung, 2013; Jones & Dangl, 2006). On the pathogen side, *Avr* genes generally encode generally small, secreted virulence proteins with no sequence homology known as effectors (Lo Presti et al., 2015). These small molecules can modulate host cell structure and plant metabolism, and therefore allow pathogens to evade host defense responses through diverse mechanisms, sometimes in a highly specific manner (Raffaele & Kamoun, 2012; Stergiopoulos & de Wit, 2009). Recently, Luo et al. (2017) identified 233 candidate effectors in the *C. sojina* secretome, 80 of which were differentially expressed during starvation treatments that mimicked the plant infection environment. Additionally, they noted that 13 of these candidate effectors suppressed BAX-triggered programmed cell death (BT-PSD) in *Nicotiana benthamiana* leaves during transient expression (Luo et al., 2017). Hence, *C. sojina* seems to

employ a wide array of effector proteins to suppress host defenses and promote successful infection, which may explain the constant emergence of new physiological races.

Currently, it is well established that effectors play important roles in host-pathogen interactions and can impact the outcome of an infection in both positive or negative ways, depending on the host genotype (Sánchez-Vallet, Fouché, et al., 2018). The effector complement of a pathogen is a major determinant of host specialization (Hartmann, Sánchez-Vallet, McDonald, & Croll, 2017; Poppe, Dorsheimer, Happel, & Stukenbrock, 2015), and as a result, genes encoding effectors are often rapidly evolving and can be the targets of changing selection pressures (Brown & Tellier, 2011; Sánchez-Vallet, Fouché, et al., 2018). Underlying these protein selective pressures, new recognition specificities of R-proteins and novel *Avr* protein features may be generated through genomic sequence diversification, deletions, or loss-of-function mutations (Barrett et al., 2009; Stukenbrock & McDonald, 2009).

Genomic variability in the coevolution of plants and pathogens

Throughout the coevolution of plants and pathogens, genomic variation has greatly impacted disease epidemiology, in which host resistance polymorphisms can dampen pathogen infection and virulence polymorphisms can determine host range (Karasov, Horton, & Bergelson, 2014). The genomes of fungal plant pathogens, like many other eukaryotes, evolve via point mutations as Single Nucleotide Polymorphisms (SNPs); insertions/deletions events (InDels); transposable elements (TEs), and other chromosomal rearrangements (Genissel, Confais, Lebrun, & Gout, 2017; Plissonneau et al., 2017). Genomic variability is often much more abundant in regions harboring genes encoding products associated with pathogenesis (virulence genes; toxins; secondary metabolites). Due to their possible impact in host

recognition, these regions play a crucial role in the evolution of plant pathogens (Karasov et al., 2014; Poppe et al., 2015; Stukenbrock & McDonald, 2009).

In fungal plant pathogens, avirulence (*Avr*) genes and effector genes are often located in dispensable parts of the fungal genomes characterized by accelerated rates of evolution and a higher abundance of TEs and point mutations (Lo Presti et al., 2015; Raffaele & Kamoun, 2012; Sánchez-Vallet, Fouché, et al., 2018). Such heterogeneity and compartmentalization across fungal genomes has been termed the "two-speed" genome, so named to describe the plasticity and virulence evolution in fungal plant pathogens (Dong, Raffaele, & Kamoun, 2015; Plissonneau et al., 2017). Examples of such genome architecture governing pathogen virulence are well described in the literature, such as for Avr-Pita in *Magnoporthe oryzae* (Orbach, Farrall, Sweigard, Chumley, & Valent, 2000) and the effector *AvrStb6* in *Zymoseptoria tritici* (Zhong et al., 2017); both of these genes are embedded in sub-telomeric chromosomal regions rich in TEs.

Studies of other plant pathogenic fungi have described how genomic variability and gain or loss-of-function variants shaped the coevolution of pathogenesis, often with a rapid breakdown of host resistance. For example, in the rice blast pathosystem, speciation of *Magnoporthe oryzae* as a pathogen of rice could have been driven by transposon-mediated gene loss (Couch et al., 2005; Huang, Si, Deng, Li, & Yang, 2014; Li et al., 2009). In oomycetes, sequence variation or deletion of effector genes was associated with an increase virulence in host genotypes carrying specific R-genes (Jiang & Tyler, 2012). A similar phenomenon was observed in fungal pathogens, including *Leptosphaeria maculans* (Ghanbarnia et al., 2015) and *Fusarium oxysporum* f. sp. *melonis* (Schmidt et al., 2015). Additionally, comparative analysis of a major effector encoded by the gene Zt_8_609 demonstrated that *Zymoseptoria tritici* most likely

evaded recognition by the wheat cultivar Toronit by adaptive loss of this effector gene (Hartmann, Sánchez-Vallet, McDonald, & Croll, 2017).

Considering these examples, there is strong evidence that SNPs, InDels, and other genomic structural variants are common in fungal genomes, and likely contribute to the evolution of virulence strategies. As such, they are extremely important factors to be analyzed in host-pathogen interactions (Genissel et al., 2017). In the modern sequencing era, Genome-Wide Association Studies (GWAS) of pathosystems have mostly focused on host resistance mechanisms to a wide range of pathogens, yet the analysis of diversifying virulence mechanisms underlying pathogens still being largely unknown (Bartoli & Roux, 2017; Sánchez-Vallet, Hartmann, Marcel, & Croll, 2018). Relatively few studies associating genomic variants to virulence traits in pathogens have been performed (Dalman et al., 2013; Gao et al., 2016; Guy et al., 2013; Pensec et al., 2015; Talas, Kalih, Miedaner, & McDonald, 2016). In many ways, the vast biological diversity of plant pathogens remains untapped as a resource to elucidate the genetic basis of pathogenesis. For FLS, identifying genes and genetic factors underlying race structure in *C. sojina* can provide insights into the correct geographical distribution of haplotypes and *Avr* genes, from which a more durable genetic resistance can be developed in soybeans.

Population genomics: understanding plant disease outbreaks at the population level

The genetic variation and local adaptation of plant pathogens in agricultural ecosystems differ from dynamics that occur in wild, natural ecosystems (Croll & McDonald, 2017; McDonald & Stukenbrock, 2016). In natural ecosystems, host populations consist of genetically diverse individuals that are heterogeneously distributed in space and time, which dilutes selection pressure on pathogen populations (Möller & Stukenbrock, 2017). On the other hand, in

managed, agricultural ecosystems, crops evolve through artificial selection, in which agriculturally desired traits (such as yield and genetic resistance) are favored (Möller & Stukenbrock, 2017). Every time a new selective agent such as an R-gene or a fungicide is introduced into the environment, new mutations in the corresponding avirulence (*Avr*) gene or in the fungicide target gene can ascend in the pathogen populations, leading to increased genetic variance for virulence or fungicide sensitivity (Stukenbrock & McDonald, 2009). In agricultural ecosystems, this continual selective pressure is known as the "boom-and-bust cycle", in which the "boom" happens when a newly deployed R-gene provides resistance to a specific pathogen race population, and it is easily and widely adopted; the "bust" occurs when the pathogen population evolves in regards to the selective pressure imposed by the host population and becomes virulent on the R-gene, leading to a wide spread loss of effectiveness of this particular genetic resistance (Brown & Tellier, 2011; McDonald, 2010).

The homogeneous environment of genetically uniform hosts in agricultural ecosystems enforces strong directional selection on the pathogen populations to adapt, leading to highly specialized agricultural pathogens and consequent large-scale yield losses (Lo Presti et al., 2015; Mohd-assaad, Mcdonald, & Croll, 2017; Möller & Stukenbrock, 2017). Examples of how pathogen populations rapidly evolve to overcome new host genetic resistance are well described in the literature, including the poplar rust fungus *Melampsora larici-populina* (Persoons et al., 2017) and the oilseed rape pathogen *Leptosphaeria maculans* 'brassicae' (Daverdin et al., 2012). In both examples, mutations on *Avr* genes and the consequent rise of new populations allowed these pathogens to evade host resistance and quickly replace ancestral pathogen populations.

Similar to GWAS, the increasing accessibility to whole-genome sequencing and the ability to identify and analyze millions of genetic variants such as SNPs and InDels throughout the

pathogens' genomes can address novel questions of evolutionary genomics and epidemiology. Molecular ecology studies have identified signatures of genomic selection in genes and traits underlying host-pathogen interactions (Grünwald, McDonald, & Milgroom, 2016; Plissonneau et al., 2017; Stukenbrock & McDonald, 2009). Nowadays, methods to assess population divergence such as the Fixation index (Wright's F_{st}) and signatures of genomic selective pressure on populations by the ratio of non-synonymous (*dN*) to synonymous (*dS*) nucleotide substitution rates (*dN/dS*) are more approachable and can bring valuable insights of adaptive pressure in both host and pathogen populations (Aguileta, Refregier, Yockteng, Fournier, & Giraud, 2009; Grünwald, McDonald, and Milgroom 2016; Meirmans & Hedrick, 2011; Plissonneau et al., 2017). Moreover, such population genomics approaches can identify genes under selective pressure without any prior knowledge of traits they may affect or their effects on the organismal fitness, such as effectors, fungicide or host genetic resistance (Grünwald, McDonald, and Milgroom 2016; Stukenbrock et al., 2011).

In summary, it is clearly important to better understand the pathogen population biology and the driving forces of selection in order to create effective and durable resistance breeding strategies (McDonald, 2015). However, even though host and pathogen populations constantly evolve, and molecular tools to dissect this coevolution are readily available, information about the population genetic structure underlying host specificity and adaptation are still very limited. In the case of FLS in the US, understanding host-driven selective effects on the *C. sojina* could elucidate crucial aspects of disease epidemiology and control.

1.2 Justification and project objectives

Fungal plant pathogens pose severe problems in major economical crops worldwide (Strange & Scott, 2005). In this context, frogeye leaf spot (FLS) is one of the most important and common diseases on soybean worldwide (Mian, Missaoui, Walker, Phillips, & Boerma, 2008). Genetic resistance is potentially one of the most cost-effective and sustainable strategies to control FLS. However, *C. sojina* has demonstrated an ability to overcome resistance conveyed by single R-genes (resistance genes) of soybean, followed by the emergence of new physiological races (Pham et al., 2015). Discrepancies in the current classification of *C. sojina* races and the large gap in the knowledge of characterized genes in the pathogen genome also hamper the correct identification and full understanding of mechanisms underlying pathogenesis and population genomics in this pathosystem. A detailed understanding of *C. sojina* genomics underlying host determination and virulence would lead not only to important insights regarding the biology and race population structure of this fungal pathogen, but would also accelerate efforts to develop new soybean germplasm with durable genetic resistance through conventional or molecular breeding approaches. Moreover, the identification and distribution of haplotypes and linked *Avr* genes underlying race and host virulence in *C. sojina* can be further explored to contribute for a more precise and durable genetic resistance in soybeans. Therefore, the overall objective of this project is to elucidate the molecular basis of race specificity and pathogenesis pathways in *Cercospora sojina*. To achieve this objective, the specific goals of this study are:

(i) Identify genomic variations between *C. sojina* isolates differing in infection phenotypes;

(ii) Identify genomic loci and genes with genetic variability between isolates of *C. sojina* differing in host virulence;

(iii) Identify candidate genes underlying pathogenesis and host specificity in *C. sojina.*

1.3 Citations

Aguileta, G., Refregier, G., Yockteng, R., Fournier, E., & Giraud, T. (2009). Rapidly evolving genes in pathogens: methods for detecting positive selection and examples among fungi, bacteria, viruses and protists. *Infection, Genetics and Evolution*, *9*(4), 656–670.

Albu, S., Schneider, R. W., Price, P., & Doyle, V. P. (2016). Cercospora cf. flagellaris and C. cf. sigesbeckiae are associated with Cercospora leaf blight and purple seed stain on soybean in North America. *Phytopathology*, *Epub ahead*, 1–10. https://doi.org/10.1094/PHYTO-12-15-0332-R

Allen, T. W., Damcone, J. P., Dufault, N. S., Faske, T. R., Hershman, D. E., Hollier, C. A., … Young, H. (2018). Southern United States Soybean Disease Loss Estimates for 2017. *Proceedings of the Southern Soybean Disease Workers (SSDW)*, (Table 2), 10–15.

Athow, K. L., & Probst, A. H. (1952). The inheritance of resistance to frog-eye leaf spot on Soybeans. *Phytopathology*, *42*, 660–662.

Barrett, L. G., Thrall, P. H., Dodds, P. N., Van Der Merwe, M., Linde, C. C., Lawrence, G. J., & Burdon, J. J. (2009). Diversity and evolution of effector loci in natural populations of the plant pathogen melampsora lini. *Molecular Biology and Evolution*, *26*(11), 2499–2513. https://doi.org/10.1093/molbev/msp166

Bartoli, C., & Roux, F. (2017). Genome-Wide Association Studies In Plant Pathosystems: Toward an Ecological Genomics Approach. *Frontiers in Plant Science*, *8*(May), 763. https://doi.org/10.3389/fpls.2017.00763

Beseli, A., Noar, R., & Daub, M. E. (2015). Characterization of cercospora nicotianae hypothetical proteins in cercosporin resistance. *PLoS ONE*, *10*(10), 1–24. https://doi.org/10.1371/journal.pone.0140676

Boyd, L. A., Ridout, C., O'Sullivan, D. M., Leach, J. E., & Leung, H. (2013). Plant-pathogen interactions: Disease resistance in modern agriculture. *Trends in Genetics*, *29*(4), 233–240. https://doi.org/10.1016/j.tig.2012.10.011

Brown, J. K. M., & Tellier, A. (2011). Plant-Parasite Coevolution: Bridging the Gap between Genetics and Ecology. *Annual Review of Phytopathology*, *49*(1), 345–367. https://doi.org/10.1146/annurev-phyto-072910-095301

Carmona, M. A., Scandiani, M., & Luque, A. (2009). Severe Outbreaks of Soybean Frogeye Leaf Spot Caused by Cercospora sojina in the Pampean Region, Argentina. *Plant Disease*, *93*(9), 966. https://doi.org/10.1094/PDIS-93-9-0966B

Chen, H., Lee, M. H., Daub, M. E., & Chung, K. R. (2007). Molecular analysis of the cercosporin biosynthetic gene cluster in Cercospora nicotianae. *Molecular Microbiology*, *64*(3), 755–770. https://doi.org/10.1111/j.1365-2958.2007.05689.x

Chupp, C. (1954). *A monograph of the fungus genus Cercospora*. New York: The Ronald Press.

Couch, B. C., Fudal, I., Lebrun, M. H., Tharreau, D., Valent, B., Van Kim, P., … Kohn, L. M. (2005). Origins of host-specific populations of the blast pathogen Magnaporthe oryzae in crop domestication with subsequent expansion of pandemic clones on rice and weeds of rice. *Genetics*, *170*(2), 613–630. https://doi.org/10.1534/genetics.105.041780

Croll, D., & McDonald, B. A. (2017). The genetic basis of local adaptation for pathogenic fungi in agricultural ecosystems. *Molecular Ecology*, *26*(7), 2027–2040. https://doi.org/10.1111/mec.13870

Crous, P. W., & Braun, U. (2003). *Mycosphaerella and its Anamorphs: 1. Names published in Cercospora and Passalora* (1st ed.). CBS Biodiversity Series.

Cruz, C. D., & Dorrance, a. E. (2009a). Characterization and Survival of Cercospora sojina in Ohio. *Plant Health Progress*, (April). https://doi.org/10.1094/PHP-2009-0512-03-RS

Cruz, C. D., & Dorrance, A. E. (2009b). Characterization and Survival of Cercospora sojina in Ohio. *Plant Health Progress*. https://doi.org/10.1094/PHP-2009-0512-03-RS

Dalman, K., Himmelstrand, K., Olson, Å., Lind, M., Brandström-Durling, M., & Stenlid, J. (2013). A Genome-Wide Association Study Identifies Genomic Regions for Virulence in the Non-Model Organism Heterobasidion annosum s.s. *PLoS ONE*, *8*(1). https://doi.org/10.1371/journal.pone.0053525

Daub, M. E. (1981). Cercosporin, a Photosensitizing Toxin from Cercospora Species. *Phytopathology* 72, 370-374.

Daub, M. E., & Chung, K.-R. (2007). Cercosporin: A Phytoactivated Toxin in Plant Disease. *APSnet Feature Articles*. https://doi.org/10.1094/APSnetFeature/2007-0207

Daverdin, G., Rouxel, T., Gout, L., Aubertot, J. N., Fudal, I., Meyer, M., … Balesdent, M. H. (2012). Genome Structure and Reproductive Behaviour Influence the Evolutionary Potential of a Fungal Phytopathogen. *PLoS Pathogens*, *8*(11). https://doi.org/10.1371/journal.ppat.1003020

Dong, S., Raffaele, S., & Kamoun, S. (2015). The two-speed genomes of filamentous pathogens: Waltz with plants. *Current Opinion in Genetics and Development*, *35*, 57–65. https://doi.org/10.1016/j.gde.2015.09.001

Fisher, M. C., Henk, D. A., Briggs, C. J., Brownstein, J. S., Madoff, L. C., McCraw, S. L., & Gurr, S. J. (2012). Emerging fungal threats to animal, plant and ecosystem health. *Nature*, *484*(7393), 186–194. https://doi.org/10.1038/nature10947

Flor, H. H. (1971). Current status of the gene-for-gene concept. *Annual Review of Phytopathology*, *9*, 275–296.

FRAC. (2011). List of Plant Pathogenic Organisms Resistant to Disease Control Agents. Retrieved February 8, 2017, from http://www.frac.info/

Gao, Y., Liu, Z., Faris, J. D., Richards, J., Brueggeman, R. S., Li, X., … Friesen, T. L. (2016). Validation of Genome-Wide Association Studies as a Tool to Identify Virulence Factors in *Parastagonospora nodorum*. *Phytopathology*, *106*(10), 1177–1185. https://doi.org/10.1094/PHYTO-02-16-0113-FI

Genissel, A., Confais, J., Lebrun, M.-H., & Gout, L. (2017). Association Genetics in Plant Pathogens: Minding the Gap between the Natural Variation and the Molecular Function. *Frontiers in Plant Science*, *8*(July), 8–11. https://doi.org/10.3389/fpls.2017.01301

Ghanbarnia, K., Fudal, I., Larkan, N. J., Links, M. G., Balesdent, M. H., Profotova, B., … Borhan, M. H. (2015). Rapid identification of the Leptosphaeria maculans avirulence gene AvrLm2 using an intraspecific comparative genomics approach. *Molecular Plant Pathology*, *16*(7), 699–709. https://doi.org/10.1111/mpp.12228

Goodwin, S. B., Dunkle, L. D., & Zismann, V. L. (2001). Phylogenetic Analysis of Cercospora and Mycosphaerella Based on the Internal Transcribed Spacer Region of Ribosomal DNA. *Phytopathology*, *91*(7), 648–658. https://doi.org/10.1094/PHYTO.2001.91.7.648

Grau, C. R., Dorrance, A. E., Bond, J., & Russin, J. S. (2004). Fungal diseases. In H. R. Boerma & J. E. Specht (Eds.), *Soybeans: improvement, production, and uses* (3rd ed., pp. 679–763).

Gravina, G. de A., Sediyama, C. S., Filho, S. M., Moreira, M. A., de Barros, E. G., & Cruz, C. D. (2004). Multivariate analysis of combining ability for soybean resistance to Cercospora sojina Hara. *Genetics and Molecular Biology*, *27*(3), 395–399. https://doi.org/10.1590/S1415-47572004000300015

Groenewald, J. Z., Nakashima, C., Nishikawa, J., Jama, A. N., Groenewald, M., Braun, U., & Crous, P. W. (2013). Species concepts in Cercospora. *Studies in Mycology*, *75*, 115–170. https://doi.org/10.3114/sim0012.

Grünwald, N. J., McDonald, B. A., & Milgroom, M. G. (2016). Population Genomics of Fungal and Oomycete Pathogens. *Annual Review of Phytopathology*, *54*(1), 323–346. https://doi.org/10.1146/annurev-phyto-080614-115913

Guy, E., Genissel, A., Hajri, A., Chabannes, M., David, P., Carrere, S., … Noel, L. D. (2013). Natural Genetic Variation of Xanthomonas campestris pv. campestris Pathogenicity on Arabidopsis Revealed by Association and Reverse Genetics. *MBio*, *4*(3), 1–14. https://doi.org/10.1128/mBio.00538-12

Hartmann, F. E., Sánchez-Vallet, A., McDonald, B. A., & Croll, D. (2017). A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. *ISME Journal*, *11*(5), 1189–1204. https://doi.org/10.1038/ismej.2016.196

Huang, J., Si, W., Deng, Q., Li, P., & Yang, S. (2014). Rapid evolution of avirulence genes in rice blast fungus Magnaporthe oryzae. *BMC Genetics*, *15*, 15–17. https://doi.org/10.1186/1471-2156-15-45

Jiang, R. H. Y., & Tyler, B. M. (2012). Mechanisms and Evolution of Virulence in Oomycetes. *Annual Review of Phytopathology*, *50*(1), 295–318. https://doi.org/10.1146/annurev-phyto-081211-172912

Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. *Nature*, *444*(7117), 323–329. https://doi.org/10.1038/nature05286

Karasov, T. L., Horton, M. W., & Bergelson, J. (2014). Genomic variability as a driver of plant–pathogen coevolution? *Current Opinion in Plant Biology*, *18*, 24–30. https://doi.org/10.1016/j.pbi.2013.12.003

Keen, N. T. (1990). Gene-for-Gene Complementarity in Plant-Pathogen Interactions. *Annual Review of Genetics*, (24), 447–463.

Kim, H., Newell, A. D., Cota-Sieckmeyer, R. G., Rupe, J. C., Fakhoury, A. M., & Bluhm, B. H. (2013). Mating-type distribution and genetic diversity of Cercospora sojina populations on soybean from Arkansas: evidence for potential sexual reproduction. *Phytopathology*, *103*(10), 1045–51. https://doi.org/10.1094/PHYTO-09-12-0229-R

Koenning, S. R., & Wrather, J. A. (2010). Suppression of Soybean Yield Potential in the Continental United States by Plant Diseases from 2006 to 2009. *Plant Health Progress*. https://doi.org/10.1094/PHP-2010-1122-01-RS

Laviolette, F., Athow, K., Probst, A., Wilcox, J., & Abney, T. (1970). Effect of bacterial pustule and frogeye leaf spot on yield of Clark soybeans. *Crop Science*, *10*(4), 514–517.

Lehman, S. G. (1928). Frogeye Leaf Spot of Soybean caused by Cercospora Diazu Miura. *Journal of Agricultural Research*, *36*(9), 811–833.

Li, W., Wang, B., Wu, J., Lu, G., Hu, Y., Zhang, X., … Zhou, B. (2009). The *Magnaporthe oryzae* Avirulence Gene *AvrPiz-t* Encodes a Predicted Secreted Protein That Triggers the Immunity in Rice Mediated by the Blast Resistance Gene *Piz-t*. *Molecular Plant-Microbe Interactions*, *22*(4), 411–420. https://doi.org/10.1094/MPMI-22-4-0411

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., … Kahmann, R. (2015). Fungal Effectors and Plant Susceptibility. *Annual Review of Plant Biology*, *66*(1), 513–545. https://doi.org/10.1146/annurev-arplant-043014-114623

Luo, X., Cao, J., Huang, J., Wang, Z., Guo, Z., Chen, Y., … Liu, J. (2017). Genome sequencing and comparative genomics reveal the potential pathogenic mechanism of Cercospora sojina Hara on soybean. *DNA Research*, *25*(March), 25–37. https://doi.org/10.1093/dnares/dsx035

Ma, S. M., & Li, B. Y. (1997). Primary report on the identification for physiological races of Cercospora sojina Hara in northeast China. *Acta Phytopathol.*, *27*(2).

McDonald, B. (2010). How can we achieve durable disease resistance in agricultural ecosystems? *New Phytologist*, *185*(1), 1–3. https://doi.org/10.1111/j.1469-8137.2009.03111.x

McDonald, B. A. (2015). How can research on pathogen population biology suggest disease management strategies? The example of barley scald (Rhynchosporium commune). *Plant Pathology*. https://doi.org/10.1111/ppa.12415

McDonald, B. A., & Stukenbrock, E. H. (2016). Rapid emergence of pathogens in agro-ecosystems: Global threats to agricultural sustainability and food security. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1709). https://doi.org/10.1098/rstb.2016.0026

Meirmans, P. G., & Hedrick, P. W. (2011). Assessing population structure: FST and related measures. *Molecular Ecology Resources*, *11*(1), 5–18. https://doi.org/10.1111/j.1755-0998.2010.02927.x

Mengistu, A., Kurtzweil, N. C., & Grau, C. R. (2002). First Report of Frogeye Leaf Spot (Cercospora sojina) in Wisconsin. *Plant Disease*, *86*(11), 1272. https://doi.org/10.1094/PDIS.2002.86.11.1272B

Mian, M. A. R., Missaoui, A. M., Walker, D. R., Phillips, D. V., & Boerma, H. R. (2008). Frogeye leaf spot of soybean: A review and proposed race designations for isolates of Cercospora sojina Hara. *Crop Science*. https://doi.org/10.2135/cropsci2007.08.0432

Mian, M. A. R., Wang, T., Phillips, D. V., Alvernaz, J., & Boerma, H. R. (1999). Molecular mapping of the Rcs3 gene for resistance to frogeye leaf spot in soybean. In *Crop Science* (Vol. 39, pp. 1687–1691).

Mian, R., Mengistu, A., Wiebold, W., Shannon, G., & Wrather, A. (2009). Identification of Soybean Genotypes Resistant to C ercospora sojina by Field Screening and Molecular Markers, *93*(4), 2003–2006.

Missaoui, A. M., Ha, B. K., Phillips, D. V., & Boerma, H. R. (2007). Single Nucleotide Polymorphism Detection of the Gene for Resistance to Frogeye Leaf Spot in Soybean. *Crop Science*, *47*(4), 1681. https://doi.org/10.2135/cropsci2006.11.0711

Missaoui, A. M., Phillips, D. V., & Boerma, H. R. (2007). DNA Marker Analysis of 'Davis' Soybean and Its Descendants for the 3 Gene Conferring Resistance to. *Crop Science*, *47*(3), 1263. https://doi.org/10.2135/cropsci2006.07.0472

Mohd-assaad, N., Mcdonald, B. A., & Croll, D. (2017). Genome-wide detection of genes under positive selection in worldwide populations of the barley scald pathogen. *Genome Biol. Evol*. 10(5):1315–1332. doi:10.1093/gbe/evy087

Möller, M., & Stukenbrock, E. H. (2017). Evolution and genome architecture in fungal plant pathogens. *Nature Reviews Microbiology*, *15*(12), 756–771. https://doi.org/10.1038/nrmicro.2017.76

Nascimento, K. J. T., Debona, D., Kelly, S., França, S., Gabriele, M., Gonçalves, M., … Rodrigues, F. Á. (2014). Soybean Resistance to Cercospora sojina Infection Is Reduced by Silicon. *Phytopathology*, *104*(11), 1183–1191. https://doi.org/10.1094/PHYTO-02-14-0047-R

Newman, A. G., & Townsend, C. A. (2016). Molecular Characterization of the Cercosporin Biosynthetic Pathway in the Fungal Plant Pathogen Cercospora nicotianae. *Journal of the American Chemical Society*, *138*(12), 4219–4228. https://doi.org/10.1021/jacs.6b00633

Orbach, M. J., Farrall, L., Sweigard, J. A., Chumley, F. G., & Valent, B. (2000). A Telomeric Avirulence Gene Determines Efficacy for the Rice Blast Resistance Gene Pi-ta. *The Plant Cell*, *12*(11), 2019. https://doi.org/10.2307/3871102

Pensec, F., Lebeau, A., Daunay, M. C., Chiroleu, F., Guidot, A., & Wicker, E. (2015). Towards the Identification of Type III Effectors Associated with *Ralstonia solanacearum* Virulence on Tomato and Eggplant. *Phytopathology*, *105*(12), 1529–1544. https://doi.org/10.1094/PHYTO-06-15-0140-R

Persoons, A., Hayden, K. J., Fabre, B., Frey, P., De Mita, S., Tellier, A., & Halkett, F. (2017). The escalatory Red Queen: Population extinction and replacement following arms race dynamics in poplar rust. *Molecular Ecology*, *26*(7), 1902–1918. https://doi.org/10.1111/mec.13980

Pham, A. T., Harris, D. K., Buck, J., Hoskins, A., Serrano, J., Abdel-Haleem, H., … Li, Z. (2015). Fine mapping and characterization of candidate genes that control resistance to Cercospora sojina K. Hara in two soybean germplasm accessions. *PLoS ONE*, *10*(5). https://doi.org/10.1371/journal.pone.0126753

Phillips, D., & Boerma, H. . (1981). Cercospora sojina Race 5: A Threat to Soybeans in the Southerneastern United States. *Phytopathology*, *71*, 334–336.

Phillips, D. V. (1999). Frogeye leaf spot. In G. L. Hartman, J. B. Sinclair, & J. C. Rupe (Eds.), *Compendium of Soybean Diseases* (4th ed., pp. 20–21). American Phytopathological Society.

Plissonneau, C., Benevenuto, J., Mohd-Assaad, N., Fouché, S., Hartmann, F. E., & Croll, D. (2017a). Using Population and Comparative Genomics to Understand the Genetic Basis of Effector-Driven Fungal Pathogen Evolution. *Frontiers in Plant Science*, *8*, 119. https://doi.org/10.3389/fpls.2017.00119

Plissonneau, C., Benevenuto, J., Mohd-Assaad, N., Fouché, S., Hartmann, F. E., & Croll, D. (2017b). Using Population and Comparative Genomics to Understand the Genetic Basis of Effector-Driven Fungal Pathogen Evolution. *Frontiers in Plant Science*, *8*, 119. https://doi.org/10.3389/fpls.2017.00119

Ploper, L. D., González, V., Gálvez, M. R., Devani, M. R., Ledesma, F., & Zamorano, M. A. (2001). Frogeye Leaf Spot of Soybean Caused by Cercospora sojina in Northwestern Argentina. *Plant Disease*, *85*(7), 801. https://doi.org/10.1094/PDIS.2001.85.7.801B

Pollack, F. (1987). An annotated compilation of *Cercospora* names. *Mycol.Mem.* (Vol. 12).

Poppe, S., Dorsheimer, L., Happel, P., & Stukenbrock, E. H. (2015). Rapidly Evolving Genes Are Key Players in Host Specialization and Virulence of the Fungal Wheat Pathogen Zymoseptoria tritici (Mycosphaerella graminicola). *PLoS Pathogens*, *11*(7), 1–21. https://doi.org/10.1371/journal.ppat.1005055

Probst, A. H., & Athow, K. L. (1964). Registration of Kent Soybean. *Crop Science*, (4), 240.

Raffaele, S., & Kamoun, S. (2012). Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Reviews Microbiology*, *10*(6), 417–430. https://doi.org/10.1038/nrmicro2790

Ross, J. T. (1968). Additional physiologic races of Cercospora sojina on soybean in North Carolina. *Phytopathology*, *58*, 708–709.

Sánchez-Vallet, A., Fouché, S., Fudal, I., Hartmann, F. E., Soyer, J. L., Tellier, A., & Croll, D. (2018). The Genome Biology of Effector Gene Evolution in Filamentous Plant Pathogens. *Annual Review of Phytopathology*, *56*(1), annurev-phyto-080516-035303. https://doi.org/10.1146/annurev-phyto-080516-035303

Sánchez-Vallet, A., Hartmann, F. E., Marcel, T. C., & Croll, D. (2018). Nature's genetic screens: using genome-wide association studies for effector discovery. *Molecular Plant Pathology*, *19*(1), 3–6. https://doi.org/10.1111/mpp.12592

Schmidt, S. M., Lukasiewicz, J., Farrer, R., van Dam, P., Bertoldo, C., & Rep, M. (2015). Comparative genomics of *Fusarium oxysporum* f. sp. *melonis* reveals the secreted protein recognized by the Fom-2 resistance gene in melon. *New Phytologist*, *209*(1), n/a--n/a. https://doi.org/10.1111/nph.13584

Scholthof, K. B. G. (2007). The disease triangle: Pathogens, the environment and society. *Nature Reviews Microbiology*, *5*(2), 152–156. https://doi.org/10.1038/nrmicro1596

Sherwin, H. S., & Kreitlow, K. W. (1952). Discoloration of soybean seeds by the frogeye fungus, Cercospora sojina. *Phytopathology*, *42*, 560–572.

Simonetti, E., Carmona, M. A., Scandiani, M. M., García, A. F., Luque, A. G., Correa, O. S., & Balestrasse, K. B. (2012). Evaluation of indigenous bacterial strains for biocontrol of the frogeye leaf spot of soya bean caused by Cercospora sojina. *Letters in Applied Microbiology*, *55*(2), 170–173. https://doi.org/10.1111/j.1472-765X.2012.03266.x

Soares, A. P. G., Guillin, E. A., Borges, L. L., Da Silva, A. C. T., De Almeida, Á. M. R., Grijalba, P. E., … De Oliveira, L. O. (2015). More Cercospora species infect soybeans across the Americas than meets the eye. *PLoS ONE*, *10*(8). https://doi.org/10.1371/journal.pone.0133495

Stergiopoulos, I., & de Wit, P. J. G. M. (2009). Fungal Effector Proteins. *Annual Review of Phytopathology*, *47*(1), 233–263. https://doi.org/10.1146/annurev.phyto.112408.132637

Strange, R. N., & Scott, P. R. (2005). Plant Disease: A Threat to Global Food Security. *Annual Review of Phytopathology*, *43*(1), 83–116. https://doi.org/10.1146/annurev.phyto.43.113004.133839

Stukenbrock, E. H., Bataillon, T., Dutheil, J. Y., Stukenbrock, E. H., Bataillon, T., Dutheil, J. Y., … Schierup, M. H. (2011). The making of a new pathogen : Insights from comparative population genomics of the domesticated wheat pathogen Mycosphaerella graminicola and its wild sister species, 2157–2166. https://doi.org/10.1101/gr.118851.110

Stukenbrock, E. H., & McDonald, B. A. (2009). Population genetics of fungal and oomycete effectors involved in gene-for-gene interactions. *Molecular Plant-Microbe Interactions : MPMI*, *22*(4), 371–80. https://doi.org/10.1094/MPMI-22-4-0371

Talas, F., Kalih, R., Miedaner, T., & McDonald, B. A. (2016). Genome-Wide Association Study Identifies Novel Candidate Genes for Aggressiveness, Deoxynivalenol Production, and Azole Sensitivity in Natural Field Populations of Fusarium graminearum. *Molecular Plant-Microbe Interactions*, *29*(1), MPMI-09-15-0218. https://doi.org/10.1094/MPMI-09-15-0218-R

Tonelli, M. L., & Fabra, A. (2014). The biocontrol agent Bacillus sp. CHEP5 primes the defense response against Cercospora sojina. *World Journal of Microbiology and Biotechnology*, *30*(9), 2503–2509. https://doi.org/10.1007/s11274-014-1675-3

Yang, X. B., Uphoff, M. D., & Sanogo, S. (2001). Outbreaks of Soybean Frogeye Leaf Spot in Iowa. *Plant Disease*, *85*(4), 443. https://doi.org/10.1094/PDIS.2001.85.4.443A

Yeh, C. C., & Sinclair, J. B. (1980). Sporulation and Variation in Size of Conidia and Conidiophores Among Five Isolates of Cercospora kikuchii. *Plant Disease*, *64*(April), 373–374.

Yorinori, J. T. (1992). Management of foliar fungal diseases in Brazil. In *Pest management in soybean* (pp. 185–193). London: Elsevier Applied Science.

Zeng, F., Arnao, E., Zhang, G., Olaya, G., Wullschleger, J., Sierotzki, H., … Bradley, C. A. (2015). Characterization of Quinone Outside Inhibitor Fungicide Resistance in *Cercospora sojina* and Development of Diagnostic Tools for its Identification. *Plant Disease*, *99*(4), 544–550. https://doi.org/10.1094/PDIS-05-14-0460-RE

Zhang, G. (2012). Cercospora Sojina : Over-Winter Survival and Fungicide Resistance. University of Illinois at Urbana-Champaign.

Zhang, G., & Bradley, C. A. (2017). Comparison of quinone outside inhibitor fungicide-resistant and -sensitive isolates of Cercospora sojina. *Crop Protection*, *94*, 59–63. https://doi.org/10.1016/j.cropro.2016.12.007

Zhang, G., Newman, M. A., & Bradley, C. A. (2012). First Report of the Soybean Frogeye Leaf Spot Fungus Cercospora sojina Resistant to Quinone Outside Inhibitor Fungicides in North America. *Plant Disease*, *96*(5), 767–767. https://doi.org/10.1094/PDIS-10-11-0915-PDN

Zhong, Z., Marcel, T. C., Hartmann, F. E., Ma, X., Plissonneau, C., Zala, M., Palma-Guerrero, J. (2017). A small secreted protein in Zymoseptoria tritici is responsible for avirulence on wheat cultivars carrying the Stb6 resistance gene. *New Phytologist*, *214*(2), 619–631. https://doi.org/10.1111/nph.14434

II. Chapter 2:

2.1. Abstract

Frogeye leaf spot (FLS), caused by the fungal pathogen *Cercospora sojina*, is one of the most common and recurrent soybean diseases worldwide. Among potential control measures, deployment of resistant cultivars could be the most cost-effective way to control FLS. However, *C. sojina* has already demonstrated the ability to overcome genetic resistance conveyed by single R-genes. In this study, the primary objective was to elucidate the genetic basis of host adaptation among subcollections of *C. sojina* that differ in infection phenotypes. A new reference genome assembly of *C. sojina* strain 2.2.3, coupled with a robust pathogenicity screen and a novel bulk-sequencing approach identified interesting genomic features among two subcollections (Hood-virulent and Hood-avirulent). From 18004 SNPs, 75 with F_{st} values greater than 0.2 were localized within three distinct genomic regions in the *C. sojina* genome. These regions contained genes possibly associated with pathogenesis, including a candidate effector. SNP annotation also identified additional candidate effectors with evidence of diversifying selection. In addition, phylogenetic analysis demonstrated that three isolates that were closely related to *C.* cf. *flagellaris*, although they caused symptoms indistinguishable from FLS lesions in the greenhouse assay. Altogether, the information provided here will augment efforts to improve genetic resistance against FLS in soybean through conventional and molecular breeding approaches.

2.2 Introduction

Frogeye leaf spot (FLS), caused by the fungal pathogen *Cercospora sojina* Hara, is a common and recurrent disease of soybean in many production areas around the world. The disease primarily manifests on soybean foliage, in which initial symptoms range from small circular to angular dark-brownish spots (Mian et al. 2008). As the disease develops and conditions still favorable, these leaf lesions may merge and coalesce, forming irregular brown spots with dark-reddish margins, sometimes with light centers (Lehman, 1928; Phillips, 1999). For this reason, these lesions may cause premature defoliation when covering up to 50% of the leaf surface area, leading to decreases in yield from 10 to 60% (Mian et al. 2008, 2009). In 2009, FLS was responsible for the estimated loss of 7.5 million soybean bushels amongst 28 U.S. states (Koenning & Wrather, 2010) and more recently, the soybean yield suppression caused by this pathogen increased to around 9.28 million bushels in 2017 (Allen et al., 2018).

Plants and surrounding pathogens are in a constant co-evolutionary arms race, in which the molecular products of plant resistance genes (R-genes) and pathogen avirulence genes (*Avr* genes) are in a constant battle. Proteins encoded by R-genes can detect direct or indirectly the presence of avirulence proteins of a pathogen and trigger downstream immune defenses as hypersensitive response (HR), oxidative burst and crosstalk of plant hormones, preventing the pathogen colonization in plant tissues (Boyd et al., 2013; Jones & Dangl, 2006). On the other hand, pathogens' *Avr* genes generally encode small, secreted proteins with no sequence homology known as effectors (Lo Presti et al., 2015). These small molecules can modulate host cell structure and therefore evade host defense responses through diverse mechanisms (Raffaele & Kamoun, 2012; Sánchez-Vallet, Hartmann, et al., 2018; Stergiopoulos & de Wit, 2009). This constant gene-for-gene (GFG) interaction, as firstly proposed by Flor in a 1955 paper (Flor,

1971), still being one of the main selection pressures imposed on pathogen populations at agricultural ecosystems, in which vast monocultures of host genotypes encoding specific major R-genes favors mutations on encoded effectors previously existent on the pathogen population and which can evade host recognition (Lo Presti et al., 2015; Raffaele & Kamoun, 2012). Over time and generations, this continuous selection pressure leaves genomic footprints of selection in both pathogen and host populations (Aguileta, Refregier, Yockteng, Fournier, & Giraud, 2009), which can be further studied, scanned and compared throughout the genome of individuals from different populations (Stukenbrock, 2013). Even though there is a constant evolution on both *Avr* and R-gene sides, the genetic structure underlying fungal plant pathogen host specificity and effector genomic variance still very limited.

In regards of GFG interaction, biotypes of pathogens that vary in their pattern of compatible (virulence) or incompatible (avirulence) reactions on a set of host plant cultivars differing in number and identity of resistance genes (termed "differentials") are referred to races (Flor, 1971; Keen, 1990). Although FLS seems to fall into the GFG, qualitative disease model, there are no universally accepted soybean differentials for FLS and several races of *C. sojina* have been already reported worldwide: 22 races in Brazil (Gravina et al., 2004; Yorinori, 1992); 14 races in China (Ma & Li, 1997) and 11 races through the US, as proposed by Mian et al. (2008). Discrepancies on the same current *C. sojina* race classification have also already been reported in the US. Using the same set of 12 differential soybean cultivars as Mian et al. (2008), a different number of proposed races was described by Cruz & Dorrance (2009) when classifying 50 *C. sojina* specimens isolated in Ohio soybean fields (Cruz & Dorrance, 2009).

The control of FLS has been accomplished by the use of fungicides, seed treatments, crop rotation and the use of resistance cultivars. Cultivars with qualitative genetic resistance have

been the most cost-effective means to control FLS, and three resistant genes (R-genes), namely *Rcs1* (Athow & Probst, 1952), *Rcs2* (Probst & Athow, 1964) and *Rcs3* (Mian et al. 1999) have been identified and employed on fields. The *Rcs3* gene found in the cultivar "Davis" has been described to confer resistance to race 5 and all other reported races in the US (Missaoui, Ha, et al., 2007; Missaoui, Phillips, et al., 2007). While these disease control measures can still be effective, the host-driven selective pressure imposed on some *C. sojina* populations resulted on the rise of isolates that have overcome the genetic resistance conveyed by single R-genes, leading to the emergence of new races (Pham et al., 2015). It was also observed that some *C. sojina* specimens could develop resistance to quinone outside inhibitor (QoI) fungicides, increasing barriers to an effective FLS disease management (FRAC 2011; Zeng et al. 2015; Zhang and Bradley 2017; Zhang, Newman, and Bradley 2012).

The existence of several *C. sojina* races requires a permanent search for new sources of resistance. The emergence of new races increases the chances for reshuffling and spreading of new virulent genes among the pathogen populations, enabling these pathogens to evade the recognition by R-proteins (resistance proteins) and consequently cause disease and damage even at host cells that were previously genetic resistant to FLS. Considering that the evaluation of resistance soybean lines is a time-consuming process, the better understanding of the biology and genetic variation of plant pathogens populations as *C. sojina* are extremely important for improvements in soybean resistance breeding programs (McDonald 2015; McDonald and Mundt 2016). Moreover, due to the discrepancies in the current classification of *C. sojina* races based on phenotypic reactions of differential soybean cultivars, a more profound understanding of the genetic and molecular basis of population structure and host specificity in *C. sojina* would help to correctly identify haplotypes of this pathogen, serving as a long-term effort to improve and

establish a more durable FLS resistance in soybean through conventional or marker-assisted selection approaches.

In this study, our primary objective was to elucidate the genetic basis of host adaptation on subcollections of *C. sojina* differing in infection phenotypes. Addressing this objective, a *de novo* genome assembly was obtained for *C. sojina* strain 2.2.3 and a robust race phenotyping screening of geographically diverse *C. sojina* isolates followed by a novel strategy of pooled sequencing analysis of identified subcollections were implemented.

2.3 Material and Methods

Fungal isolates and culture procedures

This study utilized a subset of *C. sojina* isolates previously obtained from symptomatic soybean leaves collected from various locations, and growing seasons, throughout the United States. All isolates were stored as mycelia in 30% (v/v) glycerol at -80 ºC. Geographical source served as criteria for selection of working isolates. A total of 240 isolates were selected (Table S1), cultured and maintained in V8 agar medium (Leslie and Summerell, 2006) at room temperature and constant darkness.

Race phenotype screening

To create subcollections of *C. sojina* isolates that differed in host virulence, race determination of 240 *C. sojina* isolates was performed based on a set of 6 of the 12 soybean differentials described by Mian et al. (2008): Davis, Tracy, Hood, Lincoln, Lee and Blackhawk. The reduced number of soybean cultivars chosen for this study differentiated most of the 11 described races of *C. sojina* in the US. The experimental design was composed of 4 plants per cultivar per isolate. Control treatment (sterile deionized water) was also composed by the same

number of plants. Soybean plants were grown in 24-cell trays at greenhouse benches (14h photoperiod; ± 25 ºC) until the first trifoliate leaf stage (V1) - approximately 20 days, at which point the plants were inoculated with *C. sojina* conidial suspensions or sterile water. Prior to inoculation, 5 to 7-days-old *C. sojina* culture plates were flooded with sterile deionized water and conidia were dislodged with a sterile cell spreader. Conidial suspensions were adjusted to concentrations of 1 x 10^5 to 3 x 10^5 conidia mL^{-1} and Tween 20 was added (0.003 vol/vol) to each suspension (Mian et al. 2009). The suspensions were atomized and applied to both adaxial and abaxial surface of leaves on each plant until runoff. Plants were kept in a dew chamber for 72 hours after-inoculation (hai) to maintain high humidity and an optimal infection environment. FLS symptoms were assessed 14 days-after inoculation (dai) with a qualitative score: plants demonstrating *C. sojina* lesions were classified as susceptible (score 0), while plants with no lesions were classified as resistant (score 1). Races were classified as proposed by Mian et al. (2008) (Table S2).

Nucleic acid extraction, PCR amplification and sequencing

For pool-sequencing, genomic DNA was isolated from 5 days-old fungal cultures grown in PDB medium (24 g/L Potato Dextrose Broth) using a modified cetyltrimethylammonium bromide (CTAB) method (Leslie and Summerell, 2006; Kim et al., 2010). Briefly, 1 mL of conidial suspension from each *C. sojina* isolate was inoculated into 50 mL PDB amended with carbenicilin (100 μg/mL). Cultures were incubated at room temperature in constant darkness with no agitation for 5 days. Mycelial suspensions were centrifuged and the tissue was washed twice with sterile deionized water. The samples were then freeze-dried for 72h, and the dried mycelia were powdered with 2-3 sterilized 3 mm glass beads in a TissueLyzer® (QIAGEN™)

for 2 minutes with a frequency of 30 cycles per second. DNA was extracted immediately from pulverized samples with CTAB as described by Kim et al. (2010). DNA samples were quantified with a fluorometer (PicoGreen®, Thermo Scientific™) and visualized on a 1% agarose gel. High quality (concentration of $\geq$ 30 ng/μL and absorbance ratio of A260 nm/ A280 nm between 1.8 and 2.0), RNA-free genomic DNA samples were pooled in equimolar concentrations. The two pools (subcollections) contained isolates from different US states and years of collection. Samples were submitted for library preparation and whole-genome resequencing at the Oklahoma Medical Research Foundation (Oklahoma City, OK, US), in which paired-end reads of 150 base pairs (2x150bp) were sequenced on an Illumina HiSeq3000 platform (Illumina Inc., San Diego, CA, USA). The isolates included in each subcollection are indicated in Table S1.

For *C. sojina* 2.2.3 genome sequencing, cultures were grown in V8 media and genomic DNA was also isolated with a modified cetyltrimethylammonium bromide (CTAB) method (Leslie & Summerell, 2006), and further purified with a Qiagen Genomic-tip 500/G column (QIAGEN, Germantown, MD, USA) following the manufacturer's recommendations. Genome sequencing was performed with a hybrid approach. For Illumina sequencing, high-quality, RNA-free DNA was submitted for library preparation (target insert size: 700 bp) and sequenced at BGI Americas (Cambridge, MA, USA) with an Illumina HiSeq 2000 platform (Illumina Inc., San Diego, CA, USA). For PacBio, two libraries were prepared (target size: 3–10 kb) and sequenced by the Yale Center for Genome Analysis (Orange, CT, USA) with a PacBio RS II Sequencing System (Pacific Biosciences, Menlo Park, CA, USA). For RNA sequencing, *C. sojina* 2.2.3 was grown on YEPD (5g/L Yeast Extract; 10g/L Peptone; 30g/L Dextrose) for 5 days. RNA was isolated and sequenced as described previously (Zaccaron & Bluhm, 2017). IonTorrent sequencing was performed with two Ion Chips (316 and 318 Kits v2) (Thermo Fisher Scientific).

Polymerase chain reactions (PCRs) were also performed on isolates of pool 1 and *C. sojina* 2.2.3 to amplify five nuclear gene regions for phylogenetic analysis (Groenewald et al. 2013; Soares et al. 2015). The PCR primers and conditions to amplify the *act*, *cal*, *his*, ITS and *tef* loci were followed as described previously (Groenewald et al. 2013; Soares et al. 2015). Detailed information about the primers used in this study are provided in Table S3. The PCR products were submitted to Genewiz (Cambrigde, MA, US) for single-pass Sanger sequencing.

Nuclear loci sequencing analysis and phylogeny reconstruction

The DNA sequences of five nuclear regions (*act*, *cal*, *his*, ITS and *tef*) of *C. sojina* isolates and other *Cercospora* species were used to assess sequence similarities and phylogenetic relationships. Prior to alignment, Sanger *ab1* files of *C. sojina* sequences were converted to *fastq* with the function *SeqIO.parse* within Biophyton v1.7 (Cock et al., 2009), and then trimmed with the function *trimfq* within *seqtk* v1.0-r68-dirty with error rate threshold of 0.01. Trimmed sequences were aligned with ClustalOmega v1.2.3 (Sievers et al., 2011) with default settings. Individual alignments were visualized and trimmed with Jalview v2.10.3b1 (Waterhouse, Procter, Martin, Clamp, & Barton, 2009). A maximum likelihood phylogenetic tree was constructed with RAxML v8.2.11 (Stamatakis, 2014) based on the concatenated alignment of the five nuclear loci with parameters adjusted to use the GTRGAMMA model of substitution, and to perform 100 rapid bootstrap analysis and search for the best-scoring tree. *Cercospora sojina* sequences were also queried in homology searches with BLASTn against the NCBI nucleotide database.

Genome assembly and gene prediction

The genome of *C. sojina* 2.2.3 was assembled with SPAdes v3.1 (Bankevich et al., 2012) based on a hybrid strategy that combined Illumina and PacBio sequencing technologies, as previously described (Zaccaron & Bluhm, 2017).

To predict the genes, RNA-seq reads were mapped to the *C. sojina* 2.2.3 genome assembly with GSNAP v2014-10-09 (Wu & Nacu, 2010), and transcripts were reconstructed with Cufflinks v2.2.1 (Trapnell et al., 2012). The reconstructed transcripts and protein sequences from *C. zeae-maydis* and *M. graminicola*, publicly available at JGI MycoCosm (http://genome.jgi.doe.gov/programs/fungi/index.jsf), were given as input to Maker v2.31.6 (Cantarel et al., 2008), in order to infer gene models directly from EST evidence (parameter *est2genome=1*) and protein homology (parameter *protein2genome=1*). The script *Maker2zff* implemented in Maker was used with default options to select gene models to train the *ab initio* predictors SNAP v2006-07-28 (Korf, 2004) and Augustus v3.0.2 (Stanke & Waack, 2003). SNAP was trained following the instructions provided by the software manual. To train Augustus, the protein sequences of the selected gene models were mapped to *C. sojina* 2.2.3 genome assembly with Scipio v1.4 (Keller, Odronitz, Stanke, Kollmar, & Waack, 2008). The mapped proteins were converted to a GenBank file using the auxiliary script *gff2gbSmallDNA.pl* implemented in Augustus, and submitted as a training dataset to WebAUGUSTUS website (http://bioinf.uni-greifswald.de/webaugustus) (Hoff & Stanke, 2013). After the *ab initio* predictors were trained, Maker was run again with settings adjusted to keep all predicted gene models (parameter *keep_preds=1*). Proteins containing signal peptide according to SignalP (Nielsen et al. 1997) and TargetP (Emanuelsson et al. 2007) were classified as secreted proteins. Secreted proteins were classified as candidate effectors either by EffectorP (Sperschneider et al.

2015) or when containing less than 300 aminoacids and more than 2% cysteine residues as predicted by EMBOSS pepstats (Chojnacki et al. 2017).

SNP identification and annotation

Quality of the obtained pool-sequencing data was analyzed with FastQC v0.11.7 (www.bioinformatics.babraham.ac.uk/projects/fastqc, last accessed June 25th 2018). The reads were mapped to the *C. sojina* 2.2.3 genome assembly with BWA mem v.0.7.12-r1039 (Heng Li & Durbin, 2010). Prior to read alignment, repetitive sequences of the reference genome were masked with RepeatMasker v.4.0.5 (http://www.repeatmasker.org, last accessed June 25th 2018) based on a custom repeat library created with RepeatScout v.1.0.5 (Price, Jones, & Pevzner, 2005).

From the aligned reads, SNPs were predicted with the packages GATK v4.0.1 (McKenna et al., 2010) and PoPoolation2 v1.201 (Kofler, Pandey, & Schlötterer, 2011). For GATK, the alignment was sorted and duplicated reads were marked with Picard v2.17.6 (broadinstitute.github.io/picard, last accessed June 25th 2018). Variants were called with the HaplotypeCaller module within GATK with the sample ploidy parameter adjusted to the number of individuals in each sequencing pool. Called variants were quality filtered (Table S4), and base quality score recalibration was performed. Then, HaplotypeCaller was run again to obtain a new set of variants. Low quality variants were filtered out, along with multiallelic SNPs and SNPs with allele frequency < 15% (Table S4). To predict SNPs with PoPoolation2, *mpileup* files were created with SAMTools v0.1.19 (Li et al. 2009), and synchronized files were built with the script *mpileup2sync.pl* within PoPoolation2 package with minimum base quality of 20. After that, SNPs as well as allele frequency differences, F_{st} values and significance of allele frequency difference based on the Fisher's exact test were calculated with the PoPoolation2 scripts *snp-*

frequency-diff.pl, *fst-sliding.pl* and *fisher-test.pl*, with minimum count of the minor allele of 6, minimum coverage of 50, maximum coverage of 4000, and sliding window length and step size equal to 1 kb. Downstream analyses were performed with the final set of SNPs predicted by both GATK and PoPoolation2. The effect of the SNPs was determined with SnpEff v4.3 (Cingolani et al., 2012) with a custom database of the *C. sojina* 2.2.3 predicted genes, as created with the module *build* within SnpEff.

Read coverage analysis

Genomic regions with a sudden drop of read coverage were identified with a sliding window of 500 bp and step size of 200 bp across the whole genome. For each window, an average coverage ratio (C_w) was calculated, which corresponded to the average coverage within the window normalized by the average coverage of the scaffold being analyzed. For each subcollection, the coverage for each position of the genome was calculated with the sub-command *genomecov* within bedtools v2.26 (Quinlan & Hall, 2010), and C_w values were determined with an R script. The absolute difference between C_w values of both subcollections was used to identify potential genomic regions more conserved in one subcollection than in the other. Windows which $C_w > 2$ in both subcollections were ignored, as they were likely repetitive regions.

2.4 Results

Race phenotype screening

The races of *C. sojina* isolates were determined with a greenhouse pathogenicity assay that utilized 6 of the 12 differential soybean cultivars described by Mian et al. (2008). Among the 240 *C. sojina* isolates screened in this study, 62.50% isolates infected all cultivars evaluated except Davis, which indicated the majority of the isolates belonged to one of the two most recently described US *C. sojina* races (R14 and R15). Additionally, four additional races were identified among the isolates (races 9, 11, 12, 13), and a small percentage of isolates were of indeterminate races ("unknowns") using the described race classification (Table 1). Isolates classified as "unknowns" could not have their race designated due to their reactions in one or more soybean cultivars that were distinct from those stated on the proposed race classification (Table S2). No observed correlations were identified between year, location and designed race.

The greenhouse assays to evaluate race yielded important information. Surprisingly, five isolates were identified that were virulent on cultivar Davis, a soybean cultivar harboring *Rcs3* widely regarded to be resistant to all known *C. sojina* races in the US (Fig 1 and Fig S1). Re-isolations from lesions and re-inoculations confirmed their virulent phenotypes. Results from other soybean cultivars allowed the establishment of distinct subcollections. Specifically, 15% of the *C. sojina* isolates in this study could not infect the cultivar Hood, but were virulent on cultivar Blackhawk (susceptible control) and avirulent on cultivar Davis (resistant control) while most of the isolates were virulent on cultivar Hood and Blackhawk, but avirulent on Davis. This phenotypic segregation into two subcollections delimited by reactions on Hood (Hood-virulent and Hood-avirulent collections; Figure S2) allowed further investigation of genetic variation

underlying host adaptation of *C. sojina* isolates spanning different locations and years of isolation.

Genome assembly and comparative genomics

The genome of *C. sojina* 2.2.3 was sequenced with Illumina and Pacific Biosciences (PacBio) technologies (Table S5). Illumina sequencing produced 90 M reads with an average length of 2 x 100 bp (approximate genome coverage of 580x), and PacBio sequencing produced 373,928 reads with an average length of 2,151 bp (approximate genome coverage of 25x). A hybrid assembly approach produced a genome assembly of 31 Mb organized into 634 scaffolds (Table 2). Compared with other published *C. sojina* genome assemblies; strains S9 (Zeng et al. 2017) and N1 (Luo et al.2017), *C. sojina* 2.2.3 and *C. sojina* S9 had similar assembly sizes and number of protein coding genes, but they were significantly smaller than the assembly of isolate N1 (Table 2 and Table 3). Pairwise whole genome alignments identified 10,493 SNPs between isolates 2.2.3 and N1, and 5,246 SNPs between isolates 2.2.3 and S9 (Fig S3).

Repetitive DNA analyses revealed that a small fraction of *C. sojina* 2.2.3 genome assembly (3.2%) was comprised of repeats. By comparison, less than 1% of the *C. sojina* S9 genome corresponded to repeats, while *C. sojina* N1 had approximately 27% (10 Mb) of its assembly covered by repetitive elements.

Ion Torrent RNA sequencing reads were mapped to the genome assembly of *C. sojina* 2.2.3 and 10,982 transcripts were reconstructed. The reconstructed transcripts along with protein sequences from closely related species were used as evidence, and 12,096 protein-encoding genes were predicted. From the predicted genes, 10,910 (90%) had at least one homologous sequence in the NCBI nr database (e-value < 1e-5), 9,243 genes (77%) received a Gene Ontology (GO) term attributed by Blast2GO, and 9,174 genes (75%) had a conserved domain

according to the InterPro database. Assessment of universal single-copy orthologs among members of the Ascomycota revealed that the *C. sojina* 2.2.3 genome assembly was 97.4% of complete, with 1.1% of genes missing.

Genomic loci differentiating between subcollections

GATK and PoPoolation2 identified 15,858 SNPs within pool 1 (Hood-avirulent) and 13,016 SNPs within pool 2 (Hood-virulent), totaling 18,004 SNPs differing among the two subcollections. Fixation index (F_{st}) values were determined for 17,851 SNPs (Fig 2A). Most of the identified SNPs (94%) had F_{st} values smaller than 0.1, which indicates weak or no difference between subcollection 1 and subcollection 2. However, 75 SNPs had F_{st} values greater than 0.2, suggesting considerable difference between both subcollections. From these 75 SNPs, 67 were located within three distinct genomic regions in the *C. sojina* genome (Fig 2B), with each locus spanning less than 2 kb.

Thirty-six SNPs with an $F_{st} > 0.2$ were located within the open reading frames (ORFs) of two predicted genes (Cs_10456 and Cs_10457) at the very end of scaffold251 (31 kb). Both of these genes encoded putative copper amine oxidases (CAOs), which are enzymes that catalyze the oxidative deamination of amines, including histamine and xenobiotic amines (Dawkes & Phillips, 2001). Other predicted genes in physically proximity to Cs_10456 and Cs_10457 included Cs_10453, which encoded a hypothetical secreted protein, Cs_10454 encoding a WD40 repeat-containing protein (Neer et al. 1994), and Cs_10455 encoding a putative RFT1 protein involved in oligosaccharide translocation (Helenius et al. 2002). Another nine SNPs with an $F_{st} > 0.2$ were located within scaffold504, which contained only 5 kb. Just one gene was predicted within this scaffold, Cs_11981, which encoded a putative basic leucine zipper (bZIP)

transcription factor. SNPs were either within or physically close (less than 210 bp) to the Cs_11981 ORF. Another 22 SNPs with an $F_{st} > 0.2$ were located on scaffold103. Thirteen of these SNPs were located at the 3' end of the ORF of Cs_07323, which encodes a putative heterokaryon incompatibility protein, and the nine remaining SNPs were located in the intergenic region between Cs_07323 and Cs_07324. Gene Cs_07324 encoded a small secreted protein (167 aa) with no conserved domains that was classified as a candidate effector. Homology searches with BLASTp against the NCBI nr database and the JGI MycoCosm revealed that only three fungal species with genomic sequences with homologs of Cs_07324: *C. zeae-maydis* (Cerzm1_93113; 75% identity), *C. beticola* (CB0940_07205; 91% identity) and *C. berteroae* (CBER1_11401; 94% identity), suggesting that genes similar to Cs_07324 might be restricted to species within the *Cercospora* genus. Other genes physically close to Cs_07324 included Cs_07321 that encoded a putative vacuolar ATP synthase subunit D; Cs_07322, which encoded a pre-mRNA-splicing factor homologous to syf1; and Cs_07325, which encoded a putative carboxypeptidase.

Although the SNPs described above had significant F_{st} values, further observations revealed that the coverage of the sequencing reads dropped significantly in the genomic regions where these SNPs were located (Fig S4), which suggested that these loci were not widely conserved across all *C. sojina* isolates sequenced. Based on this observation, the *C. sojina* genome was scanned with a sliding window to identify relatively short regions (500 bp) likely conserved in one but not in the other subcollection. However, the results failed to identify such genomic regions (Table S6 and Fig S5). The locus with the highest difference in coverage had C_w values ranging from 0.37 and 0.59, and was located around position 203,000 of scaffold20

(257 kb). Nonetheless, the coverage of each subcollection in this region was similar to the average coverage of the whole scaffold20 ($1.1 < C_w < 2.1$).

High nuclear sequence variability of a few *C. sojina* isolates

During GATK SNP calling, about 830,000 SNPs within subcollection 1 were filtered out due to low allele frequency (Table S4). This number was significantly higher than subcollection 2, for which about 2,200 SNPs were filtered out due to low allele frequency. This led to the hypothesis that a limited number of isolates in pool 1 had significant genomic variability compared to the other *C. sojina* isolates sequenced. To further explore this hypothesis, isolates within subcollection 1 were further analyzed for their genetic relatedness to *C. sojina* 2.2.3 and five other *Cercospora* species (*C.* cf. *flagellaris*; *C. beticola*; *C. kikuchii*; *C.* cf. *nicotianae*; and *C. zeae-maydis*) based on the concatenated alignment of five conserved nuclear loci (*act*, *cal*, *his*, ITS and *tef*). Surprisingly, a phylogenetic tree indicated that three of these putative *C. sojina* isolates (Csj ARCS_24, Csj ARCS_22, and Csj NT1-F9-2-1) did not group with *C. sojina*. Instead, they were more closely related to *C.* cf. *flagellaris*, another soybean pathogen (Fig 3). Homology searches with BLASTn also indicated that these three *C. sojina* isolates were genetically distant from other *C. sojina* isolates (Table S7). More precisely, among the five most homologous sequences of each one of the five nuclear loci analyzed in the NCBI nucleotide database, there was no sequence similar to *C. sojina*. Despite the taxonomic separation and some morphology distinctions of these three isolates from other *C. sojina* ones, they caused typical FLS lesions on soybean cultivars in the greenhouse assay.

C. sojina genes under selective pressure

The predicted SNPs were functionally annotated and their potential impact in the genome was assessed. The observed transition (Ts)/transversion (Tv) ratio was equal to 2.1. Out of the 18,004 SNPs, 9229 were located within transcribed regions of the genome, which included 1,259 within introns, and 7,884 within coding regions. A total of 2,850 genes had at least one SNP within their ORFs, and 2,515 genes had at least one SNP corresponding to a synonymous (*dN*) or non-synonymous (*dN*) mutation (Fig 4A). This ratio was calculated solely based on the number of missense (non-synonymous; d*N*) and synonymous SNPs (d*S*) within the ORF of each gene. From these genes, 1,185 had possible evidence of positive selection ($d_N > d_S$), and 1,060 had a possible evidence of negative selection ($d_N < d_S$). The d_N/d_S ratio observed across the whole genome was 0.94.

From the genes that had SNPs within their predicted ORFs, some presented a larger number of SNPs and stronger evidence of selective pressure (Fig 4B and C). The gene Cs_12082 contained the largest number of SNPs across the whole genome, with 51 SNPs within its 410 bp ORF (average of 124 SNPs per ORF kb) and had evidence of positive selection pressure (d_N/d_S = 5.5). This gene encoded a small secreted protein (123 aa), with no homologs in the NCBI nr database (e-value < 1e-5), lacking conserved domains, and was classified as a candidate effector.

Other candidate effectors that showed possible signatures of diversifying selection included Cs_09525 (d_N/d_S = 5; 26 SNPs per ORF kb), Cs_09831 (d_N/d_S = 5; 10 SNPs per ORF kb), and Cs_09832 (d_N/d_S = 4; 6 SNPs/ORF kb). Interestingly, BLASTp searches revealed that homologs of Cs_09525 were present only in two other *Cercospora* species: *C. beticola* (68% identity; CB0940_05810) and *C. berteroae* (67% identity; CBER1_04371). Although not classified as a candidate effector, another interesting gene was Cs_08358, which encoded a

putative secreted protein with no conserved domains. This high level of variability within this gene (56 SNPs per ORF kb) presented some of the strongest evidence of positive selective pressure in this study, with a d_N/d_S of 9.8. The gene with the highest d_N/d_S was Cs_08684, with $d_N = 15$ and $d_S = 1$. This gene encoded a protein containing the activator-binding domain KIX and was predicted to have transcription cofactor activity. Other genes rich in SNPs included the putative copper amine oxidase-encoding genes Cs_10456 and Cs_10457, with more than 60 SNPs per ORF kb. These genes also presented strong evidence of negative selective pressure, with $d_N/d_S < 0.14$. Other SNP-rich genes that also presented purifying selection footprints included a putative transcription factor Cs_11981 ($d_N/d_S = 0.12$; 30 SNPs/ORF kb), and Cs_07880 ($d_N/d_S = 0.08$; 28 SNPs/ORF kb), which encoded a putative mitochondrial superoxide dismutase [MnSOD].

2.5 Discussion

Cercospora sojina, the causal agent of frogeye leaf spot (FLS), is among the most recurrent and destructive fungal pathogens of soybean in many production areas throughout the world. Discrepancies about the current classification of *C. sojina* races, as well as the constant emergence of new infection genotypes (races) and consequent breakage of host genetic resistance conveyed by R-genes, make it necessary to understand pathogen genetics underlying the *C. sojina*-soybean pathosystem. In this study, genomic regions and possible candidate effector genes underlying *C. sojina* race specificity were elucidated through a reliable, cost-effective phenotyping and bulk-sequencing approach.

Knowledge of pathogen population structure and the genetic basis of host-pathogen specificity provide important and essential insights to improve disease management techniques and breeding programs. Until recently, most genomic studies of pathosystems had focused

primarily on the identification of molecular markers associated with resistance to a range of pathogens by the analysis of a diverse set of host populations, while the genetic variance underlying pathogens' virulence remained largely unknown (Armstrong et al., 2005; Bartoli & Roux, 2017). To fill this gap, some studies identifying SNPs associated with interesting phenotypes have been performed on the pathogen side (Gao et al., 2016; Guy et al., 2013; Hartmann et al., 2017; Pensec et al., 2015; Talas et al., 2016), but to our knowledge, this is the first report identifying candidate loci underlying race specificity and host determination in the diverse genus *Cercospora* through association and population genomics techniques of infection phenotype-divergent *C. sojina* subcollections.

Pathogen populations adapt to agricultural ecosystems differently than natural environments. The genetic uniformity of agricultural ecosystems imposes strong directional selection on pathogen genotypes (Croll and McDonald 2017; Stukenbrock and McDonald 2009). Since the first recorded occurrence of FLS in Japan in 1915 (Melchers, 1925), populations of *C. sojina* around the world have had countless opportunities to shuffle and diversify their genetic inventory. This genetic variation enabled a successful host specialization coupled with the emergence of new pathogen races, which consequently led to host resistance breakage as documented with the first two employed R-genes against FLS: *Rcs*1 and *Rcs*2 (Athow & Probst, 1952; Phillips & Boerma, 1981). Consistent with this concept, our results corroborated the assumption that *C. sojina* populations differing in host specificity, combined with possible cryptic sexual reproductions, have distinguishing genotypic features.

The qualitative, GFG-model of pathosystems imposes strong selection for polymorphisms at *Avr* loci that permit pathogens to escape host R-gene recognition. Such polymorphisms may emerge through deletions, insertions, loss-of-function mutations and

sequence diversification in (or close to) *Avr* genes (Barrett et al. 2009; Plissonneau et al. 2017; Stukenbrock and McDonald 2009). In addition to point mutations as SNPs, interesting changes in genomic architecture were observed among the two *C. sojina* subcollections. Although no genomic regions were identified with striking read coverage differences between the subcollections, differences in the proportions of sequencing reads near genomic regions harboring SNPs with high F_{st} values, especially close to the candidate effector Cs_07324, was still surprising. These differences suggested possible genomic rearrangements such as deletions, duplications, or recombination at these loci in some isolates of both subcollections. Regarding this finding, structural genomic rearrangements are known to impact the adaptive evolution of fungal plant pathogens on various hosts, leading to gains and losses of *Avr* genes, as well as duplications and neofunctionalizations (Coghlan et al. 2005; Plissonneau, Hartmann, and Croll 2018; Raffaele and Kamoun 2012; Stukenbrock and McDonald 2009). Hence, this finding suggests that mutations as SNPs may not be the only force shaping virulence specificity in *C. sojina*, and that other genomic rearrangements may have shaped the coevolution of the *C. sojina* – soybean pathosystem. Further studies exploring these genomic regions are needed to elucidate this interesting insight.

Mutations can leave footprints of selection in individual genomes, even within the same population, which can be assessed with classical approaches of population differentiation such as the Wright's fixation index (F_{st}) (Aguileta et al. 2009; Grünwald, McDonald, and Milgroom 2016; Plissonneau et al. 2017). Genomic regions with low F_{st} values between populations may indicate regions under negative selection that are conserved among the populations analyzed. On the other hand, regions with higher F_{st} values – and, therefore, higher divergence - may be under positive selection, and contain loci that may contribute to local adaptation (Grünwald,

McDonald, and Milgroom 2016). In this study, regions with possible footprints of selection were identified that differentiated the two *C. sojina* subcollections; 67 SNPs with $F_{st} > 0.2$ were distributed in three distinct regions within the *C. sojina* genome. In one of these regions, SNPs were near to a candidate effector possibly restricted to *Cercospora* species (Cs_07324). This finding not only demonstrated the effectiveness of the pipeline to search for differentiated regions, but it also identified genomic variability close to a candidate effector, a key potential determinant for host-pathogen adaptation (Lo Presti et al., 2015; Sánchez-Vallet, Hartmann, et al., 2018).

Other interesting findings included SNPs located either within or close to genes possibly involved in oxidative stress and pathogenesis. SNPs with $F_{st} > 0.2$ were found to be close or within to a putative basic leucine zipper (bZIP) transcription factor gene and also to two putative copper amine oxidases (CAOs) genes. bZIPs belong to the largest family of transcription factors from yeasts to mammals, and regulate the gene expression in diverse pathways and stress responses (Amoutzias et al., 2007). Homologs of this gene have regulated oxidative stress responses and detoxification of reactive oxygen species (ROS) in *Saccharomyces cerevisiae* (Jamieson, 1998) and in *Candida albicans* (Alarco & Raymond, 1999), besides being also involved in cell wall integrity, pathogenesis and virulence in the rice blast pathogen *Magnoporthe oryzae* and the anthracnose fungal pathogen *Colletotrichum gloeosporioides* (Guo et al., 2010; Li, Wu, Liu, & Zhang, 2017). Copper amine oxidases (CAOs) are a group of metalloenzymes found in both prokaryotes and eukaryotes that regulate nutrient metabolism, signaling, and development (Dawkes & Phillips, 2001). They catalyze the oxidation of amine groups using molecular dioxygen to form hydrogen peroxide (H_2O_2), a metabolite involved in host-pathogen interactions during compatible and incompatible interactions (Shetty et al., 2007;

Whittaker, 1999; Lamb & Dixon, 1997). Considering that creation of reactive oxygen species (ROS) by NADPH oxidases is characteristic of initial plant responses to pathogen infection (PAMP-triggered immunity; PTI) (Boyd et al., 2013; Jones & Dangl, 2006; Yoshioka, Bouteau, & Kawano, 2008), we hypothesize that the genomic variation found close to both bZIP transcription factor and CAO genes is a signature of selective pressure possibly regulating the role of these genes in oxidative stress and pathogenesis pathways of *C. sojina* upon infection of different soybean hosts.

A complementary way to detect signatures of selective pressure is evaluating rates of non-synonymous (*dN*) to synonymous (*dS*) nucleotide substitution (Plissonneau et al. 2017). Besides identifying regions of high variability between the two *C. sojina* subcollections, other genomic regions were pinpointed in which selective pressure footprints may be possibly taking place. Among the top 20 SNP-richest genes, two were identified that encoded candidate effectors (Cs_12082 and Cs_09525) with strong evidence of diversifying selection ($d_N/d_S \geq 5$). Both of these genes appeared to be possibly restricted to *Cercospora* species. In addition to these genes, two other candidate effectors with footprints of positive selection ($d_N/d_S \geq 4$) were also identified (Cs_09831 and Cs_09832). Finally, although not classified as a candidate effector, the gene Cs_08358, which encodes a putative secreted protein with no conserved domains, was highly variable within its ORF (56 SNPs per ORF kb) and had some of the strongest evidence of positive selective pressure found in our study ($d_N/d_S = 9.8$). Considering that effectors play key roles in the coevolution of many pathogens and their hosts (Lo Presti et al., 2015; Raffaele & Kamoun, 2012; Stergiopoulos & de Wit, 2009), the results of this study emphasize the potential importance of these proteins as determinants of *C. sojina* host specialization. Previous studies had already shown that signatures of positive selection in effector proteins, and the possible

emergence of host specialization, are generally associated with gene-for-gene coevolution between pathogens and their hosts (Grünwald, McDonald, and Milgroom 2016; McDonald et al. 2013; Wolfe and McDermott 1994), a finding that could be well explained in our FLS model of study.

Although frogeye leaf spot is a common soybean disease in production areas around the globe, host specificity in *C. sojina*, and among *Cercospora* species in general, is still poorly understood. Analyses of phylogenetic relatedness among *C. sojina* isolates in subcollection 1 indicated that three of them were closely related to *C.* cf. *flagellaris,* although they caused typical FLS symptoms on Blackhawk plants. This finding may be explained by the fact that host specificity in *Cercospora* is not as narrowly delineated as once thought; multiple *Cercospora* species have been isolated from individual host species, and the phenomenon of different species causing similar symptoms on the same host has also been documented (Albu et al. 2016; Chupp 1954; Crous et al. 2006; Groenewald et al. 2013; Pollack 1987). Moreover, even though five nuclear loci (*act*, *cal*, *his*, ITS and *tef*) have been widely used to taxonomically identify many *Cercospora* species (Albu et al., 2016; J. Z. Groenewald et al., 2013; Soares et al., 2015), there are still some limitations with this approach. For instance, Groenewald et al. (2005) found 96% similarity between *C. apii* and *C. beticola* for the calmodulin (*cal*) gene sequences and identical sequences on the other four loci (Groenewald, Groenewald, and Crous, 2005).

Similarly, the existing race classification structure in *C. sojina* needs to be revisited. Among the 240 *C. sojina* isolates screened in this study, five were virulent on the cultivar Davis, a soybean genotype described as resistant to all known US races. Likewise, the race of several isolates in the collection ("unknowns") could not be determined with the current race classification structure. Besides highlighting the importance of understanding the population

genomics underlying race specificity in *C. sojina*, this finding warns of a possible future threat to major FLS resistance sources currently deployed and demonstrates the flaws of race determinations based on phenotypic reactions. The classical example of the Southern Corn Leaf Blight (SCLB) epidemic of the 1970s, in which 15 % of the US southern corn production was destroyed by the newly emergent race T of *Bipolaris maydis* (*Cochliobolus heterostrophus*), serves as a relevant example. The emergence of a new pathogen race, coupled with more than 85% of the corn hybrids being susceptible to this specific pathotype, and favorable environmental conditions for disease set the stage for this famous epidemic in US corn production (Arnold, 2017; Ullstrup, 1972). Also, as previously mentioned, discrepancies in the current classification have already been reported by Cruz & Dorrance in 2009 when classifying *C. sojina* isolates from Ohio (Cruz & Dorrance, 2009). Thus, the need for a more accurate race determination in *C. sojina* is clear.

The study we presented herein gave rise to important insights in the *C. sojina* – soybean pathosystem through a cost-effective sequencing analysis of subcollections diverging in infection phenotypes. The strategy used here could identify candidate genomic regions that contribute to *C. sojina* host adaptation with relatively low-cost sequencing and small sample sizes. A disadvantage of this method relies on the identification of individual haplotypes, which hinders a better picture of the population structure of *C. sojina* in the US as well as gives higher probability to miss alleles/SNPs at low frequency in these populations. Additionally, some isolates may have higher sequencing coverage than others, even though equimolar amounts of DNA from each isolate in each pool were used to prepare the sequencing libraries. Nonetheless, this study has paved the way for subsequent investigations into host specificity and pathogenesis in *C. sojina*. Future directions could include validating statistically significant SNPs close to or

within gene ORFs, particularly candidate effector genes, by eQTL (Expression Quantitative Trait Loci) or GWAS (Genome-Wide Association Studies) approaches, as candidate genetic markers for host adaptation and race specificity. Additionally, genetic analyses of individual isolates/races could clarify the population structure and background history of *C. sojina* in the US and provide a baseline understanding of pathogen biology that would be invaluable if/when *C. sojina* is able to overcome *Rcs3* resistance. Altogether, the information provided here will accelerate efforts to improve FLS genetic resistance in soybean through conventional and molecular breeding approaches.

2.6 Citations

Aguileta, G., Refregier, G., Yockteng, R., Fournier, E., & Giraud, T. (2009). Rapidly evolving genes in pathogens: methods for detecting positive selection and examples among fungi, bacteria, viruses and protists. *Infection, Genetics and Evolution*, *9*(4), 656–670.

Alarco, A. M., & Raymond, M. (1999). The bZip transcription factor Cap1p is involved in multidrug resistance and oxidative stress response in Candida albicans. *Journal of Bacteriology*, *181*(3), 700–8.

Albu, S., Schneider, R. W., Price, P., & Doyle, V. P. (2016). Cercospora cf. flagellaris and C. cf. sigesbeckiae are associated with Cercospora leaf blight and purple seed stain on soybean in North America. *Phytopathology*, *Epub ahead*, 1–10. https://doi.org/10.1094/PHYTO-12-15-0332-R

Allen, T. W., Damcone, J. P., Dufault, N. S., Faske, T. R., Hershman, D. E., Hollier, C. A., … Young, H. (2018). Southern United States Soybean Disease Loss Estimates for 2017. *Proceedings of the Southern Soybean Disease Workers (SSDW)*, 10–15.

Amoutzias, G. D., Veron, A. S., Weiner, J., Robinson-Rechavi, M., Bornberg-Bauer, E., Oliver, S. G., & Robertson, D. L. (2007). One billion years of bZIP transcription factor evolution: Conservation and change in dimerization and DNA-binding site specificity. *Molecular Biology and Evolution*, *24*(3), 827–835. https://doi.org/10.1093/molbev/msl211

Armstrong, M. R., Whisson, S. C., Pritchard, L., Bos, J. I. B., Venter, E., Avrova, A. O., … Birch, P. R. J. (2005). An ancestral oomycete locus contains late blight avirulence gene Avr3a, encoding a protein that is recognized in the host cytoplasm. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(21), 7766–71. https://doi.org/10.1073/pnas.0500113102

Arnold, B. H. (2017). Southern corn leaf blight: A story worth retelling. *Agronomy Journal*, *109*(4), 1218–1224. https://doi.org/10.2134/agronj2017.01.0006

Athow, K. L., & Probst, A. H. (1952). The inheritance of resistance to frog-eye leaf spot on Soybeans. *Phytopathology*, *42*, 660–662

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., … Pevzner, P. a. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Barrett, L. G., Thrall, P. H., Dodds, P. N., Van Der Merwe, M., Linde, C. C., Lawrence, G. J., & Burdon, J. J. (2009). Diversity and evolution of effector loci in natural populations of the plant pathogen melampsora lini. *Molecular Biology and Evolution*, *26*(11), 2499–2513. https://doi.org/10.1093/molbev/msp166

Bartoli, C., & Roux, F. (2017). Genome-Wide Association Studies In Plant Pathosystems: Toward an Ecological Genomics Approach. *Frontiers in Plant Science*, *8*(May), 763. https://doi.org/10.3389/fpls.2017.00763

Boyd, L. A., Ridout, C., O'Sullivan, D. M., Leach, J. E., & Leung, H. (2013). Plant-pathogen interactions: Disease resistance in modern agriculture. *Trends in Genetics*, *29*(4), 233–240. https://doi.org/10.1016/j.tig.2012.10.011

Cantarel, B. L., Korf, I., Robb, S. M. C., Parra, G., Ross, E., Moore, B., … Yandell, M. (2008). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research*, *18*(1), 188–196. https://doi.org/10.1101/gr.6743907

Chojnacki, S., Cowley, A., Lee, J., Foix, A., & Lopez, R. (2017). Programmatic access to bioinformatics tools from EMBL-EBI update: 2017. *Nucleic acids research*, 45(W1), W550-W553.

Chupp, C. (1954). A monograph of the fungus genus *Cercospora*. New York: The Ronald Press.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., … Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, *6*(2), 80–92. https://doi.org/10.4161/fly.19695

Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., … De Hoon, M. J. L. (2009). Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, *25*(11), 1422–1423. https://doi.org/10.1093/bioinformatics/btp163

Coghlan, A., Eichler, E. E., Oliver, S. G., Paterson, A. H., & Stein, L. (2005). Chromosome evolution in eukaryotes: A multi-kingdom perspective. *Trends in Genetics*, *21*(12), 673–682. https://doi.org/10.1016/j.tig.2005.09.009

Croll, D., & McDonald, B. A. (2017). The genetic basis of local adaptation for pathogenic fungi in agricultural ecosystems. *Molecular Ecology*, *26*(7), 2027–2040. https://doi.org/10.1111/mec.13870

Crous, P. W., & Braun, U. (2003). *Mycosphaerella and its Anamorphs: 1. Names published in Cercospora and Passalora* (1st ed.). CBS Biodiversity Series.

Crous, P. W., Groenewald, J. Z., Groenewald, M., Caldwell, P., Braun, U., & Harrington, T. C. (2006). Species of Cercospora associated with grey leaf spot of maize. *Studies in Mycology*, *55*, 189–197. https://doi.org/10.3114/sim.55.1.189

Cruz, C. D., & Dorrance, A. E. (2009). Characterization and Survival of Cercospora sojina in Ohio. *Plant Health Progress*. https://doi.org/10.1094/PHP-2009-0512-03-RS

Dawkes, H. C., & Phillips, S. E. V. (2001). Copper amine oxidase: Cunning cofactor and controversial copper. *Current Opinion in Structural Biology*, *11*(6), 666–673. https://doi.org/10.1016/S0959-440X(01)00270-6

Emanuelsson, O., Brunak, S., Von Heijne, G., & Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nature protocols*, 2(4), 953.

Flor, H. H. (1971). Current status of the gene-for-gene concept. *Annual Review of Phytopathology*, *9*, 275–296.

FRAC. (2011). List of Plant Pathogenic Organisms Resistant to Disease Control Agents. Retrieved February 10, 2017, from http://www.frac.info/

Gao, Y., Liu, Z., Faris, J. D., Richards, J., Brueggeman, R. S., Li, X., … Friesen, T. L. (2016). Validation of Genome-Wide Association Studies as a Tool to Identify Virulence Factors in *Parastagonospora nodorum*. *Phytopathology*, *106*(10), 1177–1185. https://doi.org/10.1094/PHYTO-02-16-0113-FI

Gravina, G. de A., Sediyama, C. S., Filho, S. M., Moreira, M. A., de Barros, E. G., & Cruz, C. D. (2004). Multivariate analysis of combining ability for soybean resistance to Cercospora sojina Hara. *Genetics and Molecular Biology*, *27*(3), 395–399. https://doi.org/10.1590/S1415-47572004000300015

Groenewald, J. Z., Nakashima, C., Nishikawa, J., Jama, A. N., Groenewald, M., Braun, U., & Crous, P. W. (2013). Species concepts in Cercospora. *Studies in Mycology*, *75*, 115–170. https://doi.org/10.3114/sim0012.

Groenewald, M., Groenewald, J. Z., & Crous, P. W. (2005). Distinct Species Exist Within the *Cercospora apii* Morphotype. *Phytopathology*, *95*(8), 951–959. https://doi.org/10.1094/PHYTO-95-0951

Grünwald, N. J., McDonald, B. A., & Milgroom, M. G. (2016). Population Genomics of Fungal and Oomycete Pathogens. *Annual Review of Phytopathology*, *54*(1), 323–346. https://doi.org/10.1146/annurev-phyto-080614-115913

Guo, M., Guo, W., Chen, Y., Dong, S., Zhang, X., Zhang, H., … Zheng, X. (2010). The basic leucine zipper transcription factor Moatf1 mediates oxidative stress responses and is necessary for full virulence of the rice blast fungus Magnaporthe oryzae. *Molecular Plant-Microbe Interactions : MPMI*, *23*(8), 1053–1068. https://doi.org/10.1094/MPMI-23-8-1053

Guy, E., Genissel, A., Hajri, A., Chabannes, M., David, P., Carrere, S., … Noel, L. D. (2013). Natural genetic variation of Xanthomonas campestris pv. campestris pathogenicity on Arabidopsis revealed by association and reverse genetics. *MBio*, *4*(3), 1–14. https://doi.org/10.1128/mBio.00538-12.Invited

Hartmann, F. E., Sánchez-Vallet, A., McDonald, B. A., & Croll, D. (2017). A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. *ISME Journal*, *11*(5), 1189–1204. https://doi.org/10.1038/ismej.2016.196

Helenius, J., Ng, D. T., Marolda, C. L., Walter, P., Valvano, M. A., & Aebi, M. (2002). Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein. *Nature*, 415(6870), 447.

Hoff, K. J., & Stanke, M. (2013). WebAUGUSTUS--a web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic Acids Research*, *41*(Web Server issue), W123–W128. https://doi.org/10.1093/nar/gkt418

Jamieson, D. J. (1998). Oxidative stress responses of the yeast Saccharomyces cerevisiae. *Yeast*, *14*(16), 1511–1527. https://doi.org/10.1002/(SICI)1097-0061(199812)14:16<1511::AID-YEA356>3.0.CO;2-S

Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. *Nature*, *444*(7117), 323–329. https://doi.org/10.1038/nature05286

Keen, N. T. (1990). Gene-for-Gene Complementarity in Plant-Pathogen Interactions. *Annual Review of Genetics*, (24), 447–463.

Keller, O., Odronitz, F., Stanke, M., Kollmar, M., & Waack, S. (2008). Scipio: Using protein sequences to determine the precise exon/intron structures of genes and their orthologs in closely related species. *BMC Bioinformatics*, *9*(1), 278. https://doi.org/10.1186/1471-2105-9-278

Kim, J. S., Seo, S. G., Jun, B. K., Kim, J. W., & Kim, S. H. (2010). Simple and reliable DNA extraction method for the dark pigmented fungus, Cercospora sojina. The Plant Pathology Journal, 26(3), 289-292.

Koenning, S. R., & Wrather, J. A. (2010). Suppression of Soybean Yield Potential in the Continental United States by Plant Diseases from 2006 to 2009. *Plant Health Progress*. https://doi.org/10.1094/PHP-2010-1122-01-RS

Kofler, R., Pandey, R. V., & Schlötterer, C. (2011). PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*, *27*(24), 3435–3436. https://doi.org/10.1093/bioinformatics/btr589

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics*, *5*, 59. https://doi.org/10.1186/1471-2105-5-59

Lamb, C. and Dixon, R.A. (1997) The oxidative burst in plant disease resistance. Annu. Rev. Plant Physiol. Plant Mol. Biol. 48, 251–275

Lehman, S. G. (1928). Frogeye Leaf Spot of Soybean caused by Cercospora Diazu Miura. *Journal of Agricultural Research*, *36*(9), 811–833.

Leslie, J. F., & Summerell, B. A. (2006). *The fusarium laboratory manual*. Blackwell Pub.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*(5), 589–595. https://doi.org/10.1093/bioinformatics/btp698

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, X., Wu, Y., Liu, Z., & Zhang, C. (2017). The function and transcriptome analysis of a bZIP transcription factor CgAP1 in Colletotrichum gloeosporioides. *Microbiological Research*, *197*, 39–48. https://doi.org/10.1016/j.micres.2017.01.006

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., … Kahmann, R. (2015). Fungal Effectors and Plant Susceptibility. *Annual Review of Plant Biology*, *66*(1), 513–545. https://doi.org/10.1146/annurev-arplant-043014-114623

Luo, X., Cao, J., Huang, J., Wang, Z., Guo, Z., Chen, Y., ... & Liu, J. (2017). Genome sequencing and comparative genomics reveal the potential pathogenic mechanism of Cercospora sojina Hara on soybean. *DNA Research*, *25*(1), 25-37.

Ma, S. M., & Li, B. Y. (1997). Primary report on the identification for physiological races of Cercospora sojina Hara in northeast China. *Acta Phytopathol.*, *27*(2).

McDonald, B. A. (2015). How can research on pathogen population biology suggest disease management strategies? The example of barley scald (Rhynchosporium commune). *Plant Pathology*. https://doi.org/10.1111/ppa.12415

McDonald, B. A., & Mundt, C. C. (2016). How Knowledge of Pathogen Population Biology Informs Management of Septoria Tritici Blotch. *Phytopathology*, *106*(9), 948–955. https://doi.org/10.1094/PHYTO-03-16-0131-RVW

McDonald, M. C., Oliver, R. P., Friesen, T. L., Brunner, P. C., & McDonald, B. A. (2013). Global diversity and distribution of three necrotrophic effectors in Phaeosphaeria nodorum and related species. *New Phytologist*, *199*(1), 241–251. https://doi.org/10.1111/nph.12257

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Melchers, L. E. (1925). Diseases of cereal and forage crops in the United States in 1924. Plant Dis. Rep. Suppl. 40:186. *Plant Dis. Rep. Suppl. 40:186*, *40*, 186.

Mengistu, A., Kurtzweil, N. C., & Grau, C. R. (2002). First Report of Frogeye Leaf Spot (Cercospora sojina) in Wisconsin. *Plant Disease*, *86*(11), 1272. https://doi.org/10.1094/PDIS.2002.86.11.1272B

Mian, M. A. R., Missaoui, A. M., Walker, D. R., Phillips, D. V., & Boerma, H. R. (2008). Frogeye leaf spot of soybean: A review and proposed race designations for isolates of Cercospora sojina Hara. *Crop Science*. https://doi.org/10.2135/cropsci2007.08.0432

Mian, M. A. R., Wang, T., Phillips, D. V., Alvernaz, J., & Boerma, H. R. (1999). Molecular mapping of the Rcs3 gene for resistance to frogeye leaf spot in soybean. In *Crop Science* (Vol. 39, pp. 1687–1691).

Mian, R., Mengistu, A., Wiebold, W., Shannon, G., & Wrather, A. (2009). Identification of Soybean Genotypes Resistant to Cercospora sojina by Field Screening and Molecular Markers, *93*(4), 2003–2006.

Missaoui, A. M., Ha, B. K., Phillips, D. V., & Boerma, H. R. (2007). Single Nucleotide Polymorphism Detection of the Gene for Resistance to Frogeye Leaf Spot in Soybean. *Crop Science*, *47*(4), 1681. https://doi.org/10.2135/cropsci2006.11.0711

Missaoui, A. M., Phillips, D. V., & Boerma, H. R. (2007). DNA Marker Analysis of 'Davis' Soybean and Its Descendants for the 3 Gene Conferring Resistance to. *Crop Science*, *47*(3), 1263. https://doi.org/10.2135/cropsci2006.07.0472

Neer, E. J., Schmidt, C. J., Nambudripad, R., & Smith, T. F. (1994). The ancient regulatory-protein family of WD-repeat proteins. *Nature*, 371(6500), 812-812.

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

Pensec, F., Lebeau, A., Daunay, M. C., Chiroleu, F., Guidot, A., & Wicker, E. (2015). Towards the Identification of Type III Effectors Associated with *Ralstonia solanacearum* Virulence on Tomato and Eggplant. *Phytopathology*, *105*(12), 1529–1544. https://doi.org/10.1094/PHYTO-06-15-0140-R

Pham, A. T., Harris, D. K., Buck, J., Hoskins, A., Serrano, J., Abdel-Haleem, H., … Li, Z. (2015). Fine mapping and characterization of candidate genes that control resistance to Cercospora sojina K. Hara in two soybean germplasm accessions. *PLoS ONE*, *10*(5). https://doi.org/10.1371/journal.pone.0126753

Phillips, D. ., & Boerma, H. . (1981). Cercospora sojina Race 5: A Threat to Soybeans in the Southerneastern United States. *Phytopathology*, *71*, 334–336.

Phillips, D. V. (1999). Frogeye leaf spot. In G. L. Hartman, J. B. Sinclair, & J. C. Rupe (Eds.), *Compendium of Soybean Diseases* (4th ed., pp. 20–21). American Phytopathological Society.

Phillips, D. V., & Boerma, H. R. (1982). Two genes for resistance to race 5 of Cercospora sojina in soybeans. *Phytopathology*, *72*, 764–766.

Plissonneau, C., Benevenuto, J., Mohd-Assaad, N., Fouché, S., Hartmann, F. E., & Croll, D. (2017). Using Population and Comparative Genomics to Understand the Genetic Basis of Effector-Driven Fungal Pathogen Evolution. *Frontiers in Plant Science*, *8*, 119. https://doi.org/10.3389/fpls.2017.00119

Plissonneau, C., Hartmann, F. E., & Croll, D. (2018). Pangenome analyses of the wheat pathogen Zymoseptoria tritici reveal the structural basis of a highly plastic eukaryotic genome. *BMC Biology*, *16*(1), 1–16. https://doi.org/10.1186/s12915-017-0457-4

Pollack, F. (1987). An annotated compilation of *Cercospora* names. *Mycol.Mem.* (Vol. 12).

Price, A. L., Jones, N. C., & Pevzner, P. A. (2005). De novo identification of repeat families in large genomes. *Bioinformatics*, *21*(SUPPL. 1), 351–358. https://doi.org/10.1093/bioinformatics/bti1018

Probst, A. H., & Athow, K. L. (1964). Registration of Kent Soybean. *Crop Science*, (4), 240.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Raffaele, S., & Kamoun, S. (2012). Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Reviews Microbiology*, *10*(6), 417–430. https://doi.org/10.1038/nrmicro2790

Sánchez-Vallet, A., Hartmann, F. E., Marcel, T. C., & Croll, D. (2018). Nature's genetic screens: using genome-wide association studies for effector discovery. *Molecular Plant Pathology*, *19*(1), 3–6. https://doi.org/10.1111/mpp.12592

Shetty, N. P., Mehrabi, R., Haldrup, A., Kema, G. H. J., Collinge, D. B., & L, J. H. J. (2007). Role of hydrogen peroxide during the interaction between the hemibiotrophic fungal pathogen. *New Phytologist*, *174*, 637–647. https://doi.org/10.1111/j.1469-8137.2007.02026.x

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., … Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*(1), 539–539. https://doi.org/10.1038/msb.2011.75

Soares, A. P. G., Guillin, E. A., Borges, L. L., Da Silva, A. C. T., De Almeida, Á. M. R., Grijalba, P. E., … De Oliveira, L. O. (2015). More Cercospora species infect soybeans across the Americas than meets the eye. *PLoS ONE*, *10*(8). https://doi.org/10.1371/journal.pone.0133495

Sperschneider J, Gardiner DM, Dodds PN, Tini F, Covarelli L, Singh KB, Manners JM, Taylor JM (2015) EffectorP: Predicting Fungal Effector Proteins from Secretomes Using Machine Learning. *New Phytologist*, doi:10.1111/nph.13794.

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stanke, M., & Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics*, *19*, ii215-ii225. https://doi.org/10.1093/bioinformatics/btg1080

Stergiopoulos, I., & de Wit, P. J. G. M. (2009). Fungal Effector Proteins. *Annual Review of Phytopathology*, *47*(1), 233–263. https://doi.org/10.1146/annurev.phyto.112408.132637

Stukenbrock, E. H. (2013). Evolution, selection and isolation: A genomic view of speciation in fungal plant pathogens. *New Phytologist*, *199*(4), 895–907. https://doi.org/10.1111/nph.12374

Stukenbrock, E. H., & McDonald, B. A. (2009). Population genetics of fungal and oomycete effectors involved in gene-for-gene interactions. *Molecular Plant-Microbe Interactions : MPMI*, *22*(4), 371–80. https://doi.org/10.1094/MPMI-22-4-0371

Talas, F., Kalih, R., Miedaner, T., & McDonald, B. A. (2016). Genome-Wide Association Study Identifies Novel Candidate Genes for Aggressiveness, Deoxynivalenol Production, and Azole Sensitivity in Natural Field Populations of Fusarium graminearum. *Molecular Plant-Microbe Interactions*, *29*(1), MPMI-09-15-0218. https://doi.org/10.1094/MPMI-09-15-0218-R

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., … Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, *7*(3), 562–78. https://doi.org/10.1038/nprot.2012.016

Ullstrup, A. J. (1972). The Impacts of the Southern Corn Leaf Blight Epidemics of 1970-1971. *Annual Review of Phytopathology*, *10*(1), 37–50. https://doi.org/10.1146/annurev.py.10.090172.000345

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189–1191. https://doi.org/10.1093/bioinformatics/btp033

Whittaker, J. W. (1999). Oxygen reactios of the copper oxidases. *Essays Biochem*, *34*, 155–172.

Wolfe, M. S., & McDermott, J. M. (1994). Population genetics of plant pathogen interactions: The Example of the Erysiphe graminis-Hordeum vulgare Pathosystem. *Annual Review of Phytopathology*, (26), 89–113. https://doi.org/10.1146/annurev.py.32.090194.000513

Wu, T. D., & Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*, *26*(7), 873–881. https://doi.org/10.1093/bioinformatics/btq057

Yang, X. B., Uphoff, M. D., & Sanogo, S. (2001). Outbreaks of Soybean Frogeye Leaf Spot in Iowa. *Plant Disease*, *85*(4), 443. https://doi.org/10.1094/PDIS.2001.85.4.443A

Yorinori, J. T. (1992). Management of foliar fungal diseases in Brazil. In *Pest management in soybean* (pp. 185–193). London: Elsevier Applied Science.

Yoshioka, H., Bouteau, F., & Kawano, T. (2008). Discovery of oxidative burst in the field of plant immunity: Looking back at the early pioneering works and towards the future development. *Plant Signaling and Behavior*, *3*(3), 153–155. https://doi.org/10.4161/psb.3.3.5537

Zaccaron, A. Z., & Bluhm, B. H. (2017). The genome sequence of Bipolaris cookei reveals mechanisms of pathogenesis underlying target leaf spot of sorghum. *Scientific Reports*, *7*(1), 17217. https://doi.org/10.1038/s41598-017-17476-x

Zeng, F., Wang, C., Zhang, G., Wei, J., Bradley, C. A., & Ming, R. (2017). Draft genome sequence of Cercospora sojina isolate S9, a fungus causing frogeye leaf spot (FLS) disease of soybean. *Genomics data*, *12*, 79-80.

Zeng, F., Arnao, E., Zhang, G., Olaya, G., Wullschleger, J., Sierotzki, H., … Bradley, C. A. (2015). Characterization of Quinone Outside Inhibitor Fungicide Resistance in *Cercospora sojina* and Development of Diagnostic Tools for its Identification. *Plant Disease*, *99*(4), 544–550. https://doi.org/10.1094/PDIS-05-14-0460-RE

Zhang, G., & Bradley, C. A. (2017). Comparison of quinone outside inhibitor fungicide-resistant and -sensitive isolates of Cercospora sojina. *Crop Protection*, *94*, 59–63. https://doi.org/10.1016/j.cropro.2016.12.007

Zhang, G., Newman, M. A., & Bradley, C. A. (2012). First Report of the Soybean Frogeye Leaf Spot Fungus Cercospora sojina Resistant to Quinone Outside Inhibitor Fungicides in North America. *Plant Disease*, *96*(5), 767–767. https://doi.org/10.1094/PDIS-10-11-0915-PDN

III. Conclusions

Although frogeye leaf spot represents a common and recurrent disease of soybean around the world, many aspects of the genetics and genomics underlying races of *C. sojina* and their soybean host specificity still very unclear. In this study, we could demonstrate that *C. sojina* isolates differing in infection phenotypes have also a diversified genome content, in which thousands of genomic variants (SNPs) were identified. The genomic loci harboring these SNPs and the genes within illustrate candidate regions that can be further explored for host virulence and race specificity in *C. sojina.* Additionally, our novel approach of SNP and genomic analysis using pools and population genetic tools demonstrated to be effective when looking to the genes possibly under selective pressure. Considering that effectors play an important role on the coevolution of plants and pathogens, our study could pinpoint candidate effector genes in the *C. sojina* genome that pursue diversifying SNPs and are under positive selection pressure in conjunction with close distinct genomic architectures - characteristics that are generally expected in fungal plant pathogen genomes. Furthermore, our results of *C.* cf. *flagellaris* causing FLS-like lesions and the isolates that could be virulent on the soybean cultivar Davis warn and underscores the importance of better understand host specificity in the *Cercospora* genus as well as the immediate necessity to review the current *C. sojina* race classification solely based on

phenotypic reactions. Lastly, this study could also contribute for molecular plant pathology and population genomics fields in general, in which, after modifications, our approach can be applied to other host-pathogen systems and address many other interesting questions for different agricultural pathosystems. Altogether, our results provide key resources to unravel the genomics of race specificity and the evolutionary selection pressures that have been shaping the soybean – *C. sojina* pathosystem, and augments long-term efforts to improve FLS resistance in soybeans through breeding and genetic engineering approaches.

IV. Tables and figures

A. Tables

Table 1. Race determination of 240 *C. sojina* isolates screened in this study. Races were determined as described by Mian et al. 2008 using 6 of the 12 soybean differentials. Cultivar reaction scores 1 and 0 indicate incompatible (resistance) and compatible (susceptibility) reactions, respectively.

Race	***Number of isolates***	***Cultivar reaction***						***Total percentage***
		Davis	***Tracy***	***Hood***	***Lincoln***	***Lee***	***Blackhawk***	
R9	1	1	1	1	0	1	0	0.42
R11	6	1	1	1	0	0	0	2.50
R12	29	1	0	1	0	0	0	12.08
R13	9	1	1	0	0	0	0	3.75
R14/R15	150	1	0	0	0	0	0	62.50
Unknown[1]	45	1 (40) 0 (5)	1 (24) 0 (21)	1 (25) 0 (20)	1 (25) 0 (20)	1 (29) 0 (16)	1 (29) 0 (16)	18.75

[1] Number in parenthesis represent the number of isolates with respective reaction in each cultivar. These isolates could not have their race designated using the current classification by Mian et al. (2008).

Table 2. Whole genome assembly statistics of *C. sojina* 2.2.3 compared to other two published isolates.

Stats	*C. sojina* 2.2.3	*C. sojina* N1	*C. sojina* S9
Size (bp)	31,112,868	40,836,407	29,949,529
Scaffolds	634	62	1,804
Contigs	775	62	1,804
Scaffold L50	76	6	240
Scaffold N50	116,115	1,594,415	37,690
Longest scaffold (bp)	499,594	6,706,376	197,766
GC content	53.5%	53.10%	53.60%
Gap	0.2%	0%	0%

Table 3. Gene prediction statistics of *C. sojina* 2.2.3 compared to other two published isolates.

Stats	*C. sojina* 2.2.3	*C. sojina* N1	*C. sojina* S9
Protein coding genes	12,098	11,969	12,109
Average length of:			
Genes (bp)	1,829	1,904	1,794
ORFs (bp)	1,508	1,581	1,494
Proteins (aa)	463	472	460
Introns (bp)	94	127	90
Exons (bp)	624	628	622
Gene completeness	97.4%	97.3%	96.80%
Gene duplication	0.0%	0.3%	0%
Gene fragmentation	1.5%	1.4%	2.10%
Missing genes	1.1%	1.3%	1.10%

Table S1. ***Cercospora sojina* isolates analyzed in this study.** Races were determined based on the reaction of 6 differential cultivars (Davis, Tracy, Hood, Lincoln, Lee, Blackhawk) and classified as proposed by Mian et al. 2008 (see main text). Sequencing pools 1 and 2 represent the Hood-avirulent and Hood-virulent subcollections, respectively.

n	Isolate name	Origin	Year of isolation[1]	Race	Sequencing pool
1	Cs 114	Arkansas	2012	9	Not sequenced
2	Csj 01-IN-KNOX-IG2	Arkansas	2004	11	Pool 1
3	Csj ARCS_11	Arkansas	2011	11	Pool 1
4	Csj ARCS_24	Arkansas	2007	11	Pool 1
5	Csj NT1 F9-2-1	Arkansas	2012	11	Pool 1
6	Cs 111	Arkansas	2012	12	Not sequenced
7	Cs 138	Arkansas	2014	12	Not sequenced
8	Cs 121	Arkansas	2013	12	Pool 1
9	Csj ARCS_22	Arkansas	2007	12	Pool 1
10	Cs 120	Arkansas	2013	12	Pool 1
11	Csj ST2 F9-1-3	Arkansas	2012	12	Pool 1
12	Csj ARCS_19	Arkansas	2011	12	Pool 1
13	Csj ARCS_18	Arkansas	2011	12	Pool 1
14	Csj B5	Arkansas	2012	12	Pool 1
15	Csj ST3 F4-2-1	Arkansas	2012	12	Pool 1
16	Cs 108	Arkansas	2013	12	Pool 1
17	Csj ST3 F2-1-2	Arkansas	2012	12	Pool 1
18	Csj PLPA	Arkansas	2013	12	Pool 1

Table S1 (Cont.)

n	Isolate name	Origin	Year of isolation[1]	Race	Sequencing pool
19	Cs 124	Arkansas	2011	12	Pool 1
20	Cs 126	Arkansas	2014	12	Pool 1
21	Csj ARCS_05	Arkansas	2011	12	Pool 1
22	Csj ARCS_17	Arkansas	2011	13	Not sequenced
23	Csj ARCS_07	Arkansas	2011	13	Not sequenced
24	Csj ARCS_13	Arkansas	2011	13	Not sequenced
25	Csj L2L2	Arkansas	2004	13	Pool 2
26	Csj ARCS_14	Arkansas	2011	13	Pool 2
27	Csj J1L2	Arkansas	2004	14/15	Not sequenced
28	Csj ST3 F3-1-1	Arkansas	2012	14/15	Not sequenced
29	Csj H3L1	Arkansas	2004	14/15	Not sequenced
30	Csj G3L2	Arkansas	2004	14/15	Not sequenced
31	Cs 127	Arkansas	2014	14/15	Not sequenced
32	Csj ARCS_23	Arkansas	2007	14/15	Not sequenced
33	Cs 110	Arkansas	2012	14/15	Not sequenced
34	Csj H3L2	Arkansas	2004	14/15	Not sequenced
35	Csj F1L1	Arkansas	2004	14/15	Not sequenced
36	Csj J2L2	Arkansas	2004	14/15	Not sequenced
37	Csj I1L2	Arkansas	2004	14/15	Not sequenced
38	Csj B6	Arkansas	2012	14/15	Not sequenced
39	Cs 136	Arkansas	2014	14/15	Not sequenced
40	Cs 143	Arkansas	2015	14/15	Not sequenced
41	Csj B7	Arkansas	2012	14/15	Not sequenced
42	Cs 145	Arkansas	2015	14/15	Not sequenced
43	Cs 144	Arkansas	2015	14/15	Not sequenced
44	Cs 125	Arkansas	2014	14/15	Not sequenced
45	Csj B3	Arkansas	2012	14/15	Not sequenced
46	Csj B4	Arkansas	2012	14/15	Not sequenced
47	Csj B1	Arkansas	2012	14/15	Not sequenced
48	Cs 119	Arkansas	2012	14/15	Not sequenced
49	Cs 103	Arkansas	2012	14/15	Not sequenced
50	Cs 106	Arkansas	2013	14/15	Not sequenced
51	Csj ARCS_06	Arkansas	2011	14/15	Not sequenced
52	Csj ARCS_15	Arkansas	2011	14/15	Not sequenced
53	Csj ARCS_16	Arkansas	2011	14/15	Not sequenced
54	Csj ARCS_20	Arkansas	2011	14/15	Not sequenced
55	Csj K2L1	Arkansas	2004	14/15	Not sequenced

Table S1 (Cont.)

n	Isolate name	Origin	Year of isolation[1]	Race	Sequencing pool
56	Csj ARCS_12	Arkansas	2011	14/15	Not sequenced
57	Csj ARCS_03	Arkansas	2011	14/15	Not sequenced
58	Csj ARCS_04	Arkansas	2011	14/15	Not sequenced
59	Csj NT1 F7-1-1	Arkansas	2012	14/15	Not sequenced
60	Csj NT1 F2-1-1	Arkansas	2012	14/15	Not sequenced
61	Csj NT1 F1-1-1	Arkansas	2012	14/15	Not sequenced
62	Csj ST3 F7-1-1	Arkansas	2012	14/15	Not sequenced
63	Csj ST3 F6-1-1	Arkansas	2012	14/15	Not sequenced
64	Csj NT2 F2-1-1	Arkansas	2012	14/15	Not sequenced
65	Csj D1L2	Arkansas	2004	14/15	Not sequenced
66	Csj ST2 F2-2-1	Arkansas	2012	14/15	Not sequenced
67	Csj ST2 F4-1-1	Arkansas	2012	14/15	Not sequenced
68	Csj E1L1	Arkansas	2004	14/15	Not sequenced
69	Csj NT2 F7-1-1	Arkansas	2012	14/15	Not sequenced
70	Csj T1B1B7	Arkansas	2007	14/15	Not sequenced
71	Csj NT2 F5-1-1	Arkansas	2012	14/15	Not sequenced
72	Csj ST1 F4-1-3	Arkansas	2012	14/15	Not sequenced
73	Csj ST1 F7-2-1	Arkansas	2012	14/15	Not sequenced
74	Csj G2L2	Arkansas	2004	14/15	Not sequenced
75	Csj ST1 F8-3-1	Arkansas	2012	14/15	Not sequenced
76	Csj ST2 F10-1-1	Arkansas	2012	14/15	Not sequenced
77	Csj NT2 F6-1-1	Arkansas	2012	14/15	Not sequenced
78	Csj ST1 F10-1-2	Arkansas	2012	14/15	Not sequenced
79	Csj ST1 F5-1-2	Arkansas	2012	14/15	Not sequenced
80	Csj D2L2	Arkansas	2004	14/15	Not sequenced
81	Csj NT2 F4-1-1	Arkansas	2012	14/15	Not sequenced
82	Csj ST1 F1-3-1	Arkansas	2012	14/15	Not sequenced
83	Csj ST2 F1-1-1	Arkansas	2012	14/15	Not sequenced
84	Csj 8-4-1	Arkansas	2004	14/15	Not sequenced
85	Csj ST1 F6-2-2	Arkansas	2012	14/15	Not sequenced
86	Csj ST2 F3-1-1	Arkansas	2012	14/15	Not sequenced
87	Csj ST1 F3-2-1	Arkansas	2012	14/15	Not sequenced
88	Cs 105	Arkansas	2012	14/15	Not sequenced
89	Csj NT1 F3-2-2	Arkansas	2012	14/15	Not sequenced
90	Csj 5-9-2	Arkansas	2004	14/15	Not sequenced
91	Csj L1L1	Arkansas	2004	14/15	Not sequenced
92	Csj NT2 F9-2-1	Arkansas	2012	14/15	Not sequenced
93	Csj ST1 F2-1-1	Arkansas	2012	14/15	Not sequenced

Table S1 (Cont.)

n	Isolate name	Origin	Year of isolation[1]	Race	Sequencing pool
94	Csj ST2 F7-1-1	Arkansas	2012	14/15	Not sequenced
95	Csj A1L1	Arkansas	2004	14/15	Not sequenced
96	Csj J2L1	Arkansas	2004	14/15	Not sequenced
97	Csj ST2 F5-1-5	Arkansas	2012	14/15	Not sequenced
98	Cs 141	Arkansas	2014	14/15	Not sequenced
99	Cs 133	Arkansas	2014	14/15	Not sequenced
100	Cs 140	Arkansas	2014	14/15	Not sequenced
101	Csj 11-4-1	Arkansas	2004	14/15	Pool 2
102	Csj K2L2	Arkansas	2004	14/15	Pool 2
103	Csj ST2 F6-1-6	Arkansas	2012	14/15	Pool 2
104	Csj ST3 F5-1-1	Arkansas	2012	14/15	Pool 2
105	Csj NT2 F8-1-2	Arkansas	2012	14/15	Pool 2
106	Cs 134	Arkansas	2012	14/15	Pool 2
107	Cs 2.2.3	Arkansas	2007	14/15	Pool 2
108	Csj B2	Arkansas	2012	14/15	Pool 2
109	Csj 6-2-1	Arkansas	2004	14/15	Pool 2
110	Csj J1L1	Arkansas	2004	14/15	Pool 2
111	Csj ARCS_10	Arkansas	2011	14/15	Pool 2
112	Csj St.Francis	Arkansas	2013	14/15	Pool 2
113	Csj ST3 F1-1-2	Arkansas	2012	14/15	Pool 2
114	Csj F1L2	Arkansas	2004	14/15	Pool 2
115	Csj ST1 F9-2-1	Arkansas	2012	14/15	Pool 2
116	Csj NT3 F9-1-1	Arkansas	2012	14/15	Pool 2
117	Cs 118	Arkansas	2012	Unknown	Not sequenced
118	Csj NT3 F7-1-1	Arkansas	2012	Unknown	Not sequenced
119	Csj NT1 F4-1-1	Arkansas	2012	Unknown	Not sequenced
120	Csj NT1 F8-1-1	Arkansas	2012	Unknown	Not sequenced
121	Csj NT3 F6-1-1	Arkansas	2012	Unknown	Not sequenced
122	Csj NT3 F8-1-2	Arkansas	2012	Unknown	Not sequenced
123	Csj NT2 F1-1-1	Arkansas	2012	Unknown	Not sequenced
124	Csj NT1 F10-1-2	Arkansas	2012	Unknown	Not sequenced
125	Csj NT3 F5-1-2	Arkansas	2012	Unknown	Not sequenced
126	Csj NT3 F2-2-1	Arkansas	2012	Unknown	Not sequenced
127	Csj NT1 F6-1-2	Arkansas	2012	Unknown	Not sequenced
128	Csj NT2 F3-1-1	Arkansas	2012	Unknown	Not sequenced
129	Csj NT2 F10-1-1	Arkansas	2012	Unknown	Not sequenced
130	Csj ARCS_02	Arkansas	2011	Unknown	Not sequenced
131	Csj D2L1	Arkansas	2004	Unknown	Not sequenced

Table S1 (Cont.)

n	Isolate name	Origin	Year of isolation[1]	Race	Sequencing pool
132	Csj K1L1	Arkansas	2004	Unknown	Not sequenced
133	Cs 129	Arkansas	2014	Unknown	Not sequenced
134	Cs 123	Arkansas	2014	Unknown	Not sequenced
135	Cs 130	Arkansas	2014	Unknown	Not sequenced
136	Cs 101	Arkansas	2013	Unknown	Pool 2
137	Csj ARCS_08	Arkansas	2011	Unknown	Pool 2
138	Csj NT1 F5-1-1	Arkansas	2012	Unknown	Pool 2
139	Csj ARCS_09	Arkansas	2011	Unknown	Pool 2
140	Csj ARCS_01	Arkansas	2011	Unknown	Pool 2
141	Csj ARCS_21	Arkansas	2011	Unknown	Pool 2
142	Csj S-4C	Illinois	Undetermined	12	Not sequenced
143	Csj VL14D5A	Illinois	Undetermined	12	Pool 1
144	Csj S14B8A	Illinois	Undetermined	12	Pool 1
145	Csj S14T5A	Illinois	Undetermined	12	Pool 1
146	Csj C14B5A	Illinois	Undetermined	12	Pool 1
147	Csj S14B2D	Illinois	Undetermined	13	Pool 2
148	Csj S14T10A	Illinois	Undetermined	14/15	Not sequenced
149	Csj S14T11B	Illinois	Undetermined	14/15	Not sequenced
150	Csj F	Illinois	Undetermined	14/15	Not sequenced
151	Csj S-10B	Illinois	Undetermined	14/15	Not sequenced
152	Csj S-20I	Illinois	Undetermined	14/15	Not sequenced
153	Csj C14T6B	Illinois	Undetermined	14/15	Not sequenced
154	Csj AA14T6D	Illinois	Undetermined	14/15	Not sequenced
155	Csj VB14B12A	Illinois	Undetermined	14/15	Not sequenced
156	Csj AB14T4C	Illinois	Undetermined	14/15	Not sequenced
157	Csj S14B10B	Illinois	Undetermined	14/15	Not sequenced
158	Csj AC14T4D	Illinois	Undetermined	14/15	Not sequenced
159	Csj C14B4C	Illinois	Undetermined	14/15	Not sequenced
160	Csj C14T4A	Illinois	Undetermined	14/15	Not sequenced
161	Csj G	Illinois	Undetermined	14/15	Not sequenced
162	Csj V-14F	Illinois	Undetermined	14/15	Not sequenced
163	Csj D	Illinois	Undetermined	14/15	Not sequenced
164	Csj VL14B1A	Illinois	Undetermined	14/15	Pool 2
165	Csj VA14T4A	Illinois	Undetermined	14/15	Pool 2
166	Csj VB14B3A	Illinois	Undetermined	14/15	Pool 2
167	Csj A	Illinois	Undetermined	14/15	Pool 2
168	Csj S-6D	Illinois	Undetermined	14/15	Pool 2

Table S1 (Cont.)

n	Isolate name	Origin	Year of isolation[1]	Race	Sequencing pool
169	Csj S-8D	Illinois	Undetermined	14/15	Pool 2
170	Csj V-5A	Illinois	Undetermined	14/15	Pool 2
171	Csj S-5B	Illinois	Undetermined	14/15	Pool 2
172	Csj V-17D	Illinois	Undetermined	14/15	Pool 2
173	Csj C	Illinois	Undetermined	14/15	Pool 2
174	Csj V-3E	Illinois	Undetermined	14/15	Pool 2
175	Csj VL14C4B	Illinois	Undetermined	14/15	Pool 2
176	Csj VL14E5A	Illinois	Undetermined	14/15	Pool 2
177	Csj VL14C6A	Illinois	Undetermined	14/15	Pool 2
178	Csj S-18F	Illinois	Undetermined	14/15	Pool 2
179	Csj AC14B10C	Illinois	Undetermined	14/15	Pool 2
180	Csj S-2F	Illinois	Undetermined	14/15	Pool 2
181	Csj V-18E	Illinois	Undetermined	14/15	Pool 2
182	Csj VA14B5A	Illinois	Undetermined	14/15	Pool 2
183	Csj S-14A	Illinois	Undetermined	14/15	Pool 2
184	Csj AC14T6C	Illinois	Undetermined	14/15	Pool 2
185	Csj V-21C	Illinois	Undetermined	14/15	Pool 2
186	Csj 140	Indiana	Undetermined	11	Pool 1
187	Csj 137	Indiana	Undetermined	11	Pool 1
188	Csj 130	Indiana	Undetermined	12	Not sequenced
189	Csj 154	Indiana	Undetermined	12	Pool 1
190	Csj 159	Indiana	Undetermined	12	Pool 1
191	Csj 103	Indiana	Undetermined	12	Pool 1
192	Csj 144	Indiana	Undetermined	12	Pool 1
193	Csj 113	Indiana	Undetermined	12	Pool 1
194	Csj 157	Indiana	Undetermined	12	Pool 1
195	Csj 152	Indiana	Undetermined	12	Pool 1
196	Csj 119	Indiana	Undetermined	13	Not sequenced
197	Csj 162	Indiana	Undetermined	13	Not sequenced
198	Csj 149	Indiana	Undetermined	13	Pool 2
199	Csj 133	Indiana	Undetermined	14/15	Not sequenced
200	Csj 109	Indiana	Undetermined	14/15	Not sequenced
201	Csj 106	Indiana	Undetermined	14/15	Not sequenced
202	Csj 155	Indiana	Undetermined	14/15	Not sequenced
203	Csj 127	Indiana	Undetermined	14/15	Not sequenced
204	Csj 143	Indiana	Undetermined	14/15	Not sequenced

Table S1 (Cont.)

n	Isolate name	Origin	Year of isolation[1]	Race	Sequencing pool
205	Csj 138	Indiana	Undetermined	14/15	Pool 2
206	Csj 108	Indiana	Undetermined	14/15	Pool 2
207	Csj 135	Indiana	Undetermined	14/15	Pool 2
208	Csj 110	Indiana	Undetermined	14/15	Pool 2
209	Csj 153	Indiana	Undetermined	14/15	Pool 2
210	Csj 136	Indiana	Undetermined	14/15	Pool 2
211	Csj 111	Indiana	Undetermined	14/15	Pool 2
212	Csj 148	Indiana	Undetermined	14/15	Pool 2
213	Csj 139	Indiana	Undetermined	14/15	Pool 2
214	Csj 156	Indiana	Undetermined	14/15	Pool 2
215	Csj 122	Indiana	Undetermined	14/15	Pool 2
216	Csj 158	Indiana	Undetermined	14/15	Pool 2
217	Csj 164	Indiana	Undetermined	14/15	Pool 2
218	Csj 115	Indiana	Undetermined	14/15	Pool 2
219	Csj 121	Indiana	Undetermined	14/15	Pool 2
220	Csj 120	Indiana	Undetermined	14/15	Pool 2
221	Csj 142	Indiana	Undetermined	Unknown	Not sequenced
222	Csj 112	Indiana	Undetermined	Unknown	Not sequenced
223	Csj 161	Indiana	Undetermined	Unknown	Not sequenced
224	Csj 117	Indiana	Undetermined	Unknown	Not sequenced
225	Csj 105	Indiana	Undetermined	Unknown	Not sequenced
226	Csj 145	Indiana	Undetermined	Unknown	Not sequenced
227	Csj 146	Indiana	Undetermined	Unknown	Not sequenced
228	Csj 147	Indiana	Undetermined	Unknown	Not sequenced
229	Csj 163	Indiana	Undetermined	Unknown	Not sequenced
230	Csj 101	Indiana	Undetermined	Unknown	Not sequenced
231	Csj 129	Indiana	Undetermined	Unknown	Not sequenced
232	Csj 126	Indiana	Undetermined	Unknown	Not sequenced
233	Csj 160	Indiana	Undetermined	Unknown	Not sequenced
234	Csj 123	Indiana	Undetermined	Unknown	Not sequenced
235	Csj 151	Indiana	Undetermined	Unknown	Not sequenced
236	Csj 104	Indiana	Undetermined	Unknown	Not sequenced
237	Csj 134	Indiana	Undetermined	Unknown	Not sequenced
238	Csj 141	Indiana	Undetermined	Unknown	Not sequenced
239	Csj 107	Indiana	Undetermined	Unknown	Not sequenced
240	Csj 128	Indiana	Undetermined	Unknown	Pool 2

[1] Undetermined year represents isolates in which year of isolation is unknown.

Table S2. Proposed classification of US races of *Cercospora sojina* based on their reaction on 12 soybean differential cultivars. Scores 0 and 1 represent compatible (susceptibility) and incompatible (resistance) reactions, respectively. Adapted from Mian et al. 2008 (see main text).

Differential cultivar	*Race designation*										
	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15
Davis	1	1	1	1	1	1	1	1	1	1	1
Peking	1	1	1	1	1	1	1	1	1	1	1
Kent	1	1	1	1	1	1	1	1	1	1	1
CNS	1	1	1	1	0	1	1	1	1	1	1
Palmetto	1	1	0	0	1	1	1	1	0	1	0
Tracy	1	1	0	0	1	1	1	0	1	0	0
Hood	0	1	0	1	1	0	1	1	0	0	0
Lincoln	1	1	1	1	0	1	0	0	0	0	0
Lee	1	1	1	1	1	0	0	0	0	0	0
Richland	0	1	1	0	0	0	0	0	0	0	0
S 100	1	1	0	0	0	0	0	0	0	0	0
Blackhawk	0	0	0	0	0	0	0	0	0	0	0

Table S3. Detailed description of primers used in this study.

Loci	Primer pairs	Ta (◦C)*	Amplicon size (bp)	Reference
Actin (*act*)	ACT-512F (5'–ATGTGCAAGGCCGGTTTCGC–3') ACT-783R (5'–TACGAGTCCTTCTGGCCCAT–3')	53	226	Groenewald, J.Z. et al. (2013)
Calmodulin (*cal)*	CAL-228F (5'–GAGTTCAAGGAGGCCTTCTCCC–3') CAL-737R (5'–CATCTTTCTGGCCATCATGG–3')	56	561	Groenewald, J.Z. et al. (2013)
Histone H3 (*his)*	CYLH3F (5'–AGGTCCACTGGTGGCAAG–3') CYLH3R (5'–AGCTGGATGTCCTTGGACTG–3')	57	412	Groenewald, J.Z. et al. (2013)
Internal Transcribed Spacers + 5.8S nrDNA (ITS)	ITS1 (5'–TCCGTAGGTGAACCTGCGG–3') ITS4 (5'–TCCTCCGCTTATTGATATGC–3')	51	560	Groenewald, J.Z. et al. (2013)
Translation Elongation Factor 1-alpha (*tef*)	TEF_F1 (5'–CCGGCAAGTCGACAACCACCG–3') TEF_R1 (5'–TCACGGTGACCTGGGGCGTC–3')	57	605	Soares, A.P.G. et al. (2015)

***Ta = Annealing temperature**

Table S4. Summary of the number of SNPs called in *C. sojina* pool1 and pool2. The table shows the number of raw SNPs (after base recalibration), and the number of SNPs that passed the quality filter, multiallelic filter, and allele frequency filter (15% or less were discarded). Filters are accumulative. For the quality filter, SNPs that matched one of the following conditions were discarded: QD < 2, FS > 60, MQ < 40, MQRankSum < -12.5, ReadPosRankSum < -8, SOR > 3.

Description	Pool1	Pool2
Raw SNPs	1,540,259	48,515
Quality filtered SNPs	863,687	15,625
Multiallelic filtered SNPs	847,753	15,625
Allele frequency filtered SNPs	17,791	13,439
Final SNPs	17,791	13,439

Table S5. Summary of the sequenced data. To perform the genome assembly, *C. sojina* isolate 2.2.3 was sequenced with Illumina and Pacific Biosciences (PacBio) technologies. To identify SNPs, different isolates (Table S1) were pooled and sequenced with Illumina. RNA of *C. sojina* 2.2.3 *in vitro* was sequenced with Ion Torrent platform, which helped to predict *C. sojina* genes.

DNA sequencing

Sample	Illumina				
	Number of reads	Sequenced isolates	Avg. length (bp)	Insert size (bp)	Coverage per isolate
C. sojina 2.2.3	90,825,774	1	2 x 100	700	580x
C. sojina Pool1	165,290,569	31	2 x 150	300	50x
C. sojina Pool2	174,109,769	65	2 x 150	300	25x

Sample	PacBio		
	Reads	Avg. length (bp)	Coverage
C. sojina 2.2.3	373,928	2,151	25x

RNA sequencing

Sample	IonTorrent	
	Number of reads	Avg. length (bp)
C. sojina 2.2.3	2,206,644	226

Table S6. The 20 windows with the highest absolute difference between the average coverage ratio (Cw) of pool1 and pool2. Windows which both pools had Cw > 2 were not taken into account. The sliding window size was 500 bp and step size of 200 bp.

Scaffold	Nº windows	avg_cov_ratio_pool1	avg_cov_ratio_pool2	abs_difference
scaffold20	1017	1.442	2.035	0.593
scaffold20	1016	1.303	1.880	0.577
scaffold485	19	1.884	2.364	0.480
scaffold220	2	0.952	1.340	0.389
scaffold20	1015	1.186	1.557	0.372
scaffold220	1	0.778	1.119	0.341
scaffold141	129	1.927	2.266	0.339
scaffold32	396	1.953	2.273	0.320
scaffold32	393	1.870	2.189	0.319
scaffold36	809	1.681	1.997	0.316
scaffold32	394	1.992	2.305	0.313
scaffold36	808	1.703	2.015	0.312
scaffold36	812	1.795	2.098	0.303
scaffold32	397	1.943	2.244	0.302
scaffold36	810	1.752	2.048	0.296
scaffold32	395	1.991	2.281	0.290
scaffold220	6	1.168	0.879	0.288
scaffold423	1	1.743	1.461	0.282
scaffold36	811	1.779	2.054	0.275
scaffold141	128	1.828	2.102	0.274

Table S7. Top five BLAST hits against the NCBI nucleotide database. For each *C. sojina* isolate and its sequence (ITS, ACT, HIS, CAL, and TEF), the BLAST hit species, identity (%), and accession number are shown.

C. sojina isolate	ITS			ACT			HIS		
	Species name	Identity (%)	Accession number	Species name	Identity (%)	Accession number	Species name	Identity (%)	Accession number
Csj NT1 F9-2-1 (B2_10)	*Cercospora lagenariae*	100	KU645994.1	*Cercospora sp. 4*	100	KX507288.1	*Cercospora cf. flagellaris*	99	MF681022.1
	fungal sp.	100	KX271327.1	*Cercospora cf. flagellaris*	100	KX443847.1	*Cercospora cf. flagellaris*	99	MF681001.1
	Cercospora cf. malloti	100	KT193692.1	*Cercospora cf. flagellaris*	100	KX443846.1	*Cercospora sp. 5*	99	KX522831.1
	Cercospora cf. malloti	100	KT193689.1	*Cercospora cf. flagellaris*	100	KX443844.1	*Cercospora sp. 4*	99	KX522821.1
	Cercospora cf. malloti	100	KT193687.1	*Cercospora cf. flagellaris*	100	KX443843.1	*Cercospora sp. 4*	99	KX522819.1
Csj ARCS_24 (B3_66)	*Cercospora malayensis*	100	MH129519.1	*Cercospora cf. flagellaris*	99	MF680953.1	*Cercospora cf. flagellaris*	99	MF681028.1
	Cercospora citrullina	100	KY824771.1	*Cercospora cf. flagellaris*	99	MF680951.1	*Cercospora cf. flagellaris*	99	MF681007.1
	Cercospora asparagi	100	KY549098.1	*Cercospora cf. flagellaris*	99	MF680950.1	*Cercospora sp. 4*	99	KX522830.1
	Cercospora asparagi	100	KY549097.1	*Cercospora cf. flagellaris*	99	MF680948.1	*Cercospora sp. 4*	99	KX522829.1
	Cercospora malayensis	100	MF435168.1	*Cercospora cf. flagellaris*	99	MF680933.1	*Cercospora sp. 4*	99	KX522827.1
Csj ARCS_22 (B3_69)	*Cercospora malayensis*	99	MH129519.1	*Cercospora sp. 4 LO-2017*	100	KX507288.1	*Cercospora cf. flagellaris*	99	MF681028.1
	Cercospora apii	99	MH178672.1	*Cercospora cf. flagellaris*	100	KX443847.1	*Cercospora cf. flagellaris*	99	MF681007.1
	Cercospora beticola	99	MF681169.1	*Cercospora cf. flagellaris*	100	KX443846.1	*Cercospora sp. 4 LO-2017*	99	KX522830.1
	Cercospora beticola	99	MF681168.1	*Cercospora cf. flagellaris*	100	KX443844.1	*Cercospora sp. 4 LO-2017*	99	KX522829.1
	Cercospora beticola	99	MF681167.1	*Cercospora cf. flagellaris*	100	KX443843.1	*Cercospora sp. 4 LO-2017*	99	KX522827.1
Cs 2.2.3 (B1_34)	*Cercospora sojina*	100	KY645998.1	*Cercospora sojina*	100	KP860288.1	*Cercospora sojina*	99	KP860290.1
	Cercospora sojina	100	KY645997.1	*Cercospora sojina*	100	JX143185.1	*Cercospora sojina*	99	JX142682.1
	Cercospora sojina	100	KY645996.1	*Cercospora sojina*	100	JX143176.1	*Cercospora sojina*	99	JX142680.1
	Cercospora sojina	100	KY645995.1	*Cercospora sojina*	100	JX143172.1	*Cercospora sojina*	99	JX142681.1
	Cercospora sojina	100	KY645994.1	*Cercospora campi-silii*	100	JX143069.1	*Cercospora achyranthis*	99	JX142539.1

Table S7 (Cont.)

C. sojina isolate	CAL			TEF		
	Species name	Identity (%)	Accession number	Species name	Identity (%)	Accession number
Csj NT1 F9-2-1 (B2_10)	*Cercospora cf. flagellaris*	100	KX443806.1	*Cercospora cf. nicotianae*	99	JX143390.1
	Cercospora cf. flagellaris	100	KX443805.1	*Cercospora sp. Q*	97	KT037476.1
	Cercospora cf. flagellaris	100	KX443804.1	*Cercospora sp. Q*	97	KT037475.1
	Cercospora cf. flagellaris	100	KX443796.1	*Cercospora sp. Q*	97	KT037471.1
	Cercospora cf. flagellaris	100	KX443794.1	*Cercospora sp. Q*	97	KT037470.1
Csj ARCS_24 (B3_66)	*Cercospora cf. flagellaris*	100	KX443806.1	*Cercospora sp. Q*	99	KT037476.1
	Cercospora cf. flagellaris	100	KX443805.1	*Cercospora sp. Q*	97	KT037475.1
	Cercospora cf. flagellaris	100	KX443804.1	*Cercospora sp. Q*	97	KT037471.1
	Cercospora cf. flagellaris	100	KX443796.1	*Cercospora sp. Q*	97	KT037470.1
	Cercospora cf. flagellaris	100	KX443794.1	*Cercospora samambaiae*	97	KT037468.1
Csj ARCS_22 (B3_69)	*Cercospora cf. flagellaris*	99	JX142866.1	*Cercospora sp. Q JZG-2013*	97	KT037476.1
	Cercospora cf. flagellaris	99	DQ835145.1	*Cercospora sp. Q JZG-2013*	97	KT037475.1
	Cercospora cf. flagellaris	99	KX443798.1	*Cercospora sp. Q JZG-2013*	97	KT037471.1
	Cercospora cf. flagellaris	99	JX142877.1	*Cercospora sp. Q JZG-2013*	97	KT037470.1
	Cercospora cf. flagellaris	99	KX443806.1	*Cercospora samambaiae*	97	KT037468.1
Cs 2.2.3 (B1_34)	*Cercospora sojina*	100	JX142928.1	*Cercospora sojina*	100	JX143420.1
	Cercospora sojina	100	JX142926.1	*Cercospora sojina*	100	JQ324984.1
	Cercospora euphorbiae-s	100	JX142859.1	*Cercospora samambaiae*	96	KT037468.1
	Cercospora campi-silii	100	JX142823.1	*Cercospora samambaiae*	96	KT037474.1
	Cercospora sojina	100	KP860289.1	*Cercospora sp. Q JZG-2013*	95	KT037476.1

B. Figures

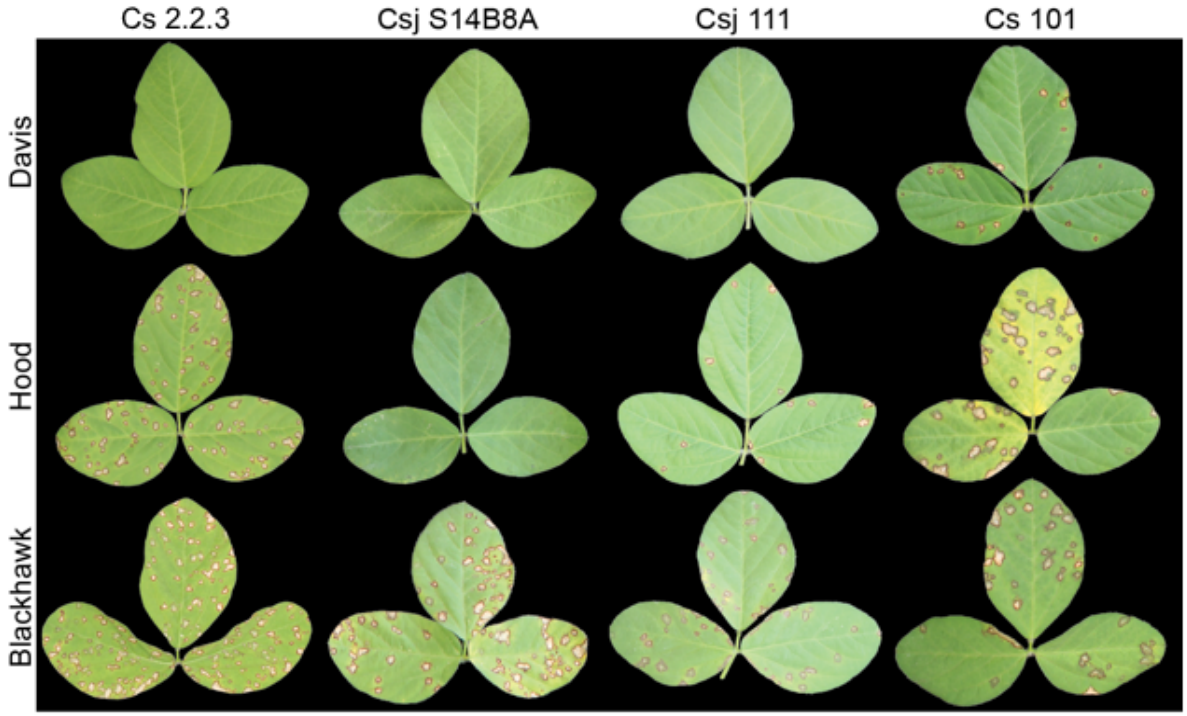

Fig 1. Summary of phenotypic reactions of *C. sojina* isolates observed in this study. Soybean plants were inoculated with *C. sojina* suspensions (1 - $3x10^5$ conidia/mL). *C. sojina* 2.2.3 (Arkansas) isolate in which the new reference genome was assembled; Csj S14B8A (Illinois) represents the subcollection 1 ("Hood-avirulent") while Csj 111 (Indiana) is a representative of subcollection 2 ("Hood-virulent"); Cs 101 (Arkansas) was one of the five isolates found to be virulent on cultivar Davis. Pictures were taken 14 days after inoculation.

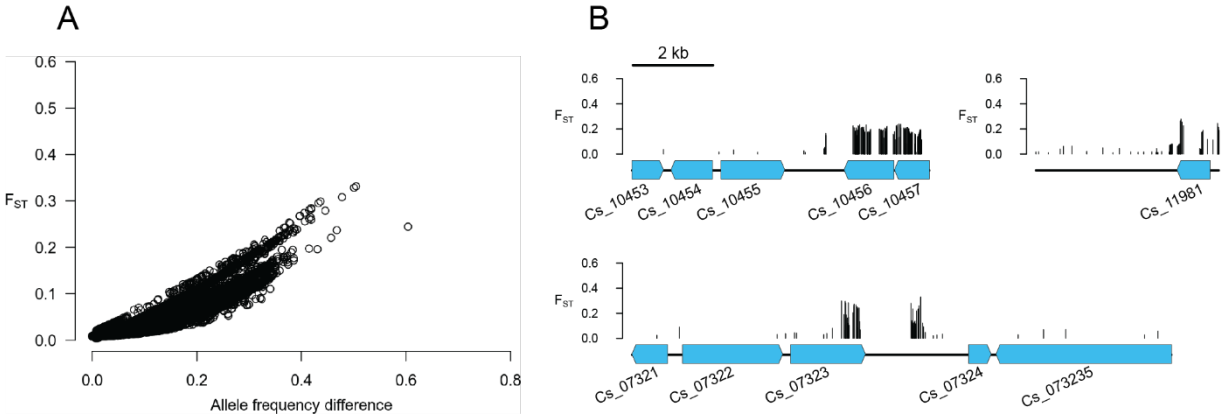

Fig 2. Fixation index values of *C. sojina* SNPs. (A) Scatter plot showing the allele frequency difference and F_{st} values for each SNP. (B) Regions of *C. sojina* genome containing SNPs with high F_{st}. The position of each SNP is represented with a vertical line, which high corresponds to its F_{st} value. Predicted gene ORFs are represented as rectangles.

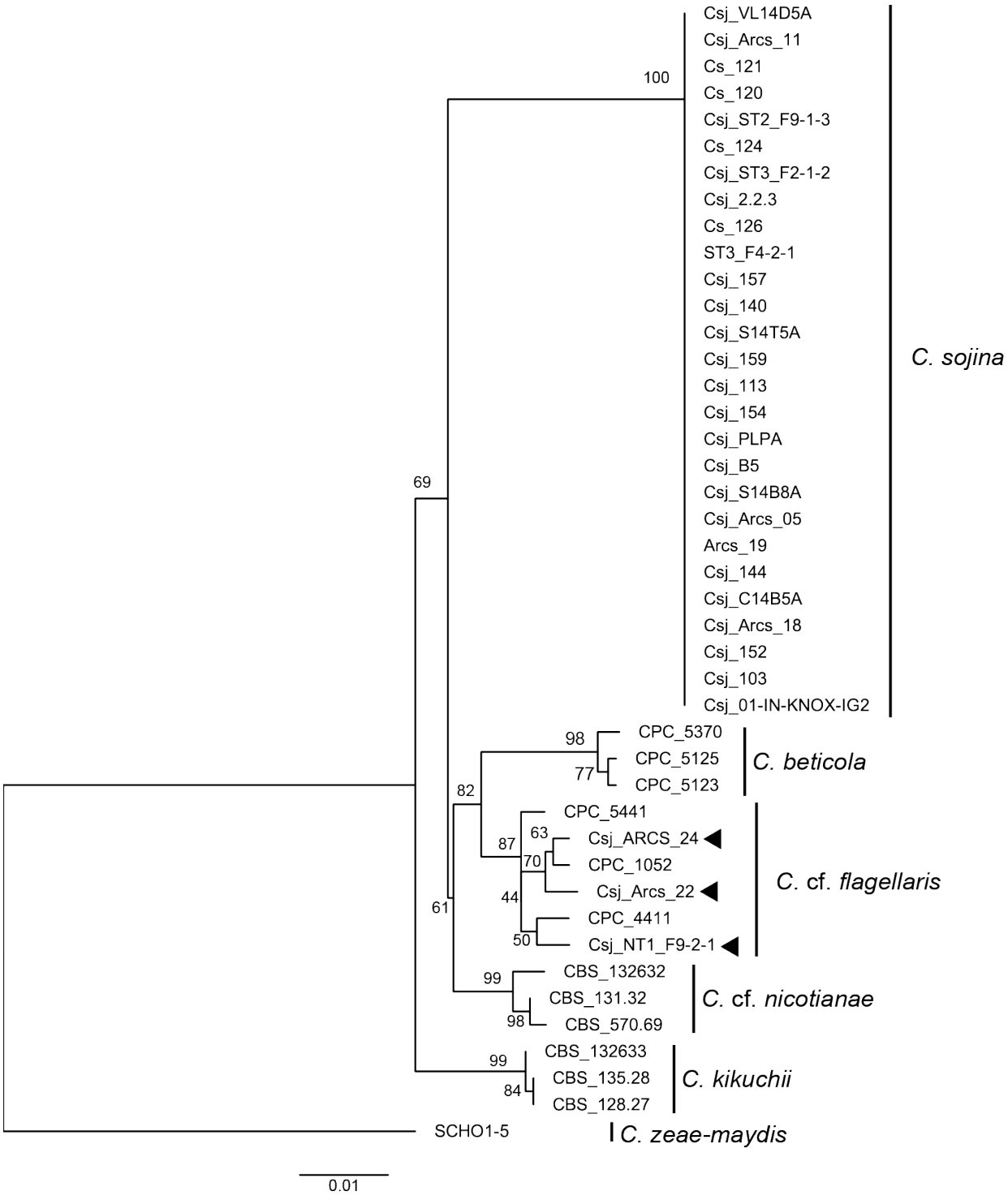

Fig 3. Maximum phylogenetic tree of *C. sojina* isolates within subcollection 1 and other *Cercospora* species. Species clades are indicated with vertical lines. *C. sojina* isolates that grouped outside of the *C. sojina* clade are indicted with a triangle. Bootstrap support values are shown as branch labels.

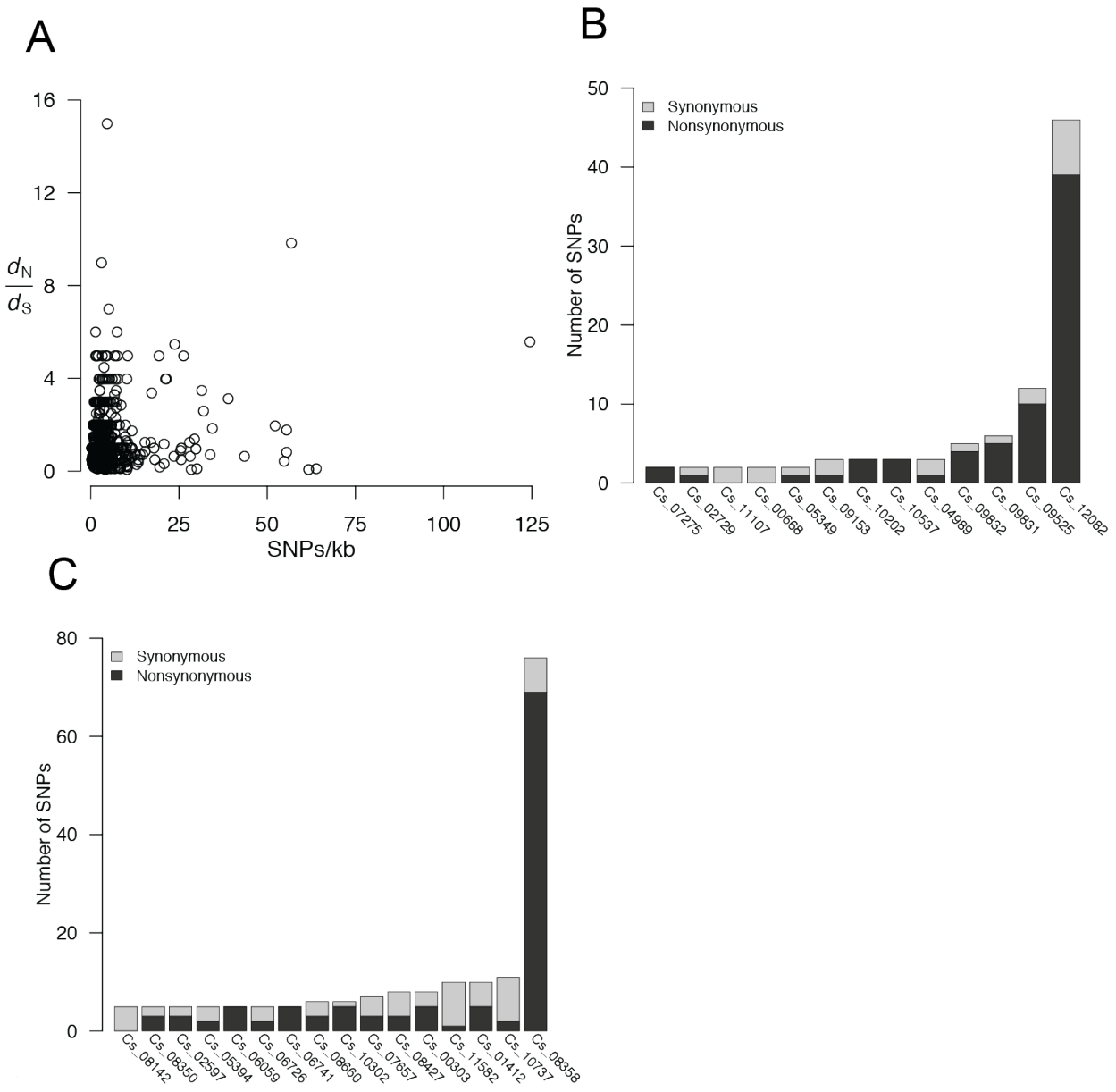

Fig 4. *C. sojina* genes possibly under selective pressure. (A) Scatter plot showing the d*N*/d*S* ratio and the average number of SNPs per kilobase of the respective gene ORF. (B and C) Bar charts showing the total number of synonymous and nonsynonymous SNPs of the top SNP-rich candidate effector genes (B) and secreted protein genes (C).

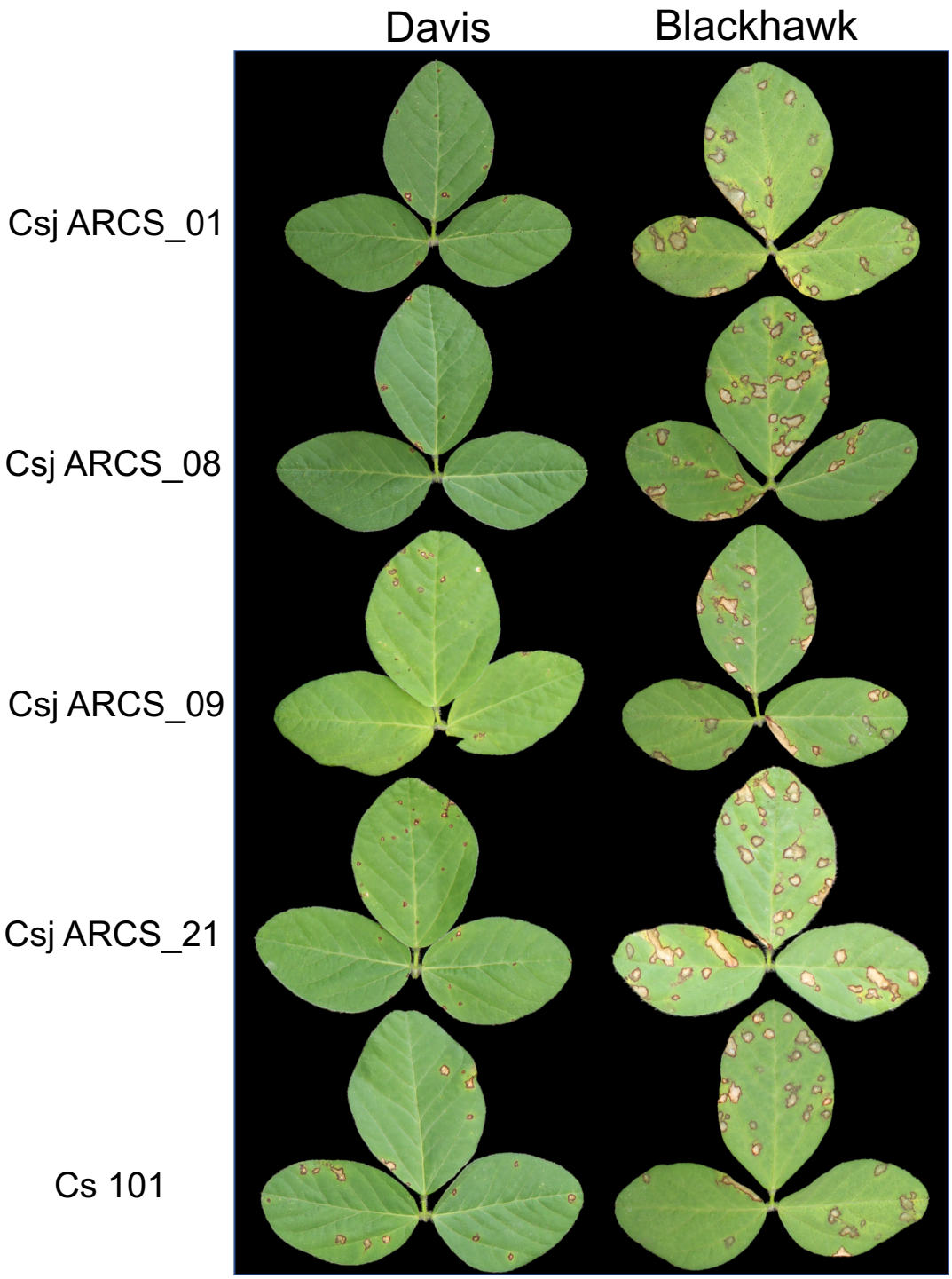

Fig. S1. Some *Cercospora sojina* isolates are shown to be virulent on cultivar Davis. Soybean plants were inoculated with *C. sojina* suspensions (1 - 3x10^5 conidia/mL) and scored 14 days after inoculation. Cultivar Blackhawk shown as susceptible control. It is possible to observe small and numerous spots with reddish margins on Davis plants, some with lighter centers, characterizing typical FLS symptoms.

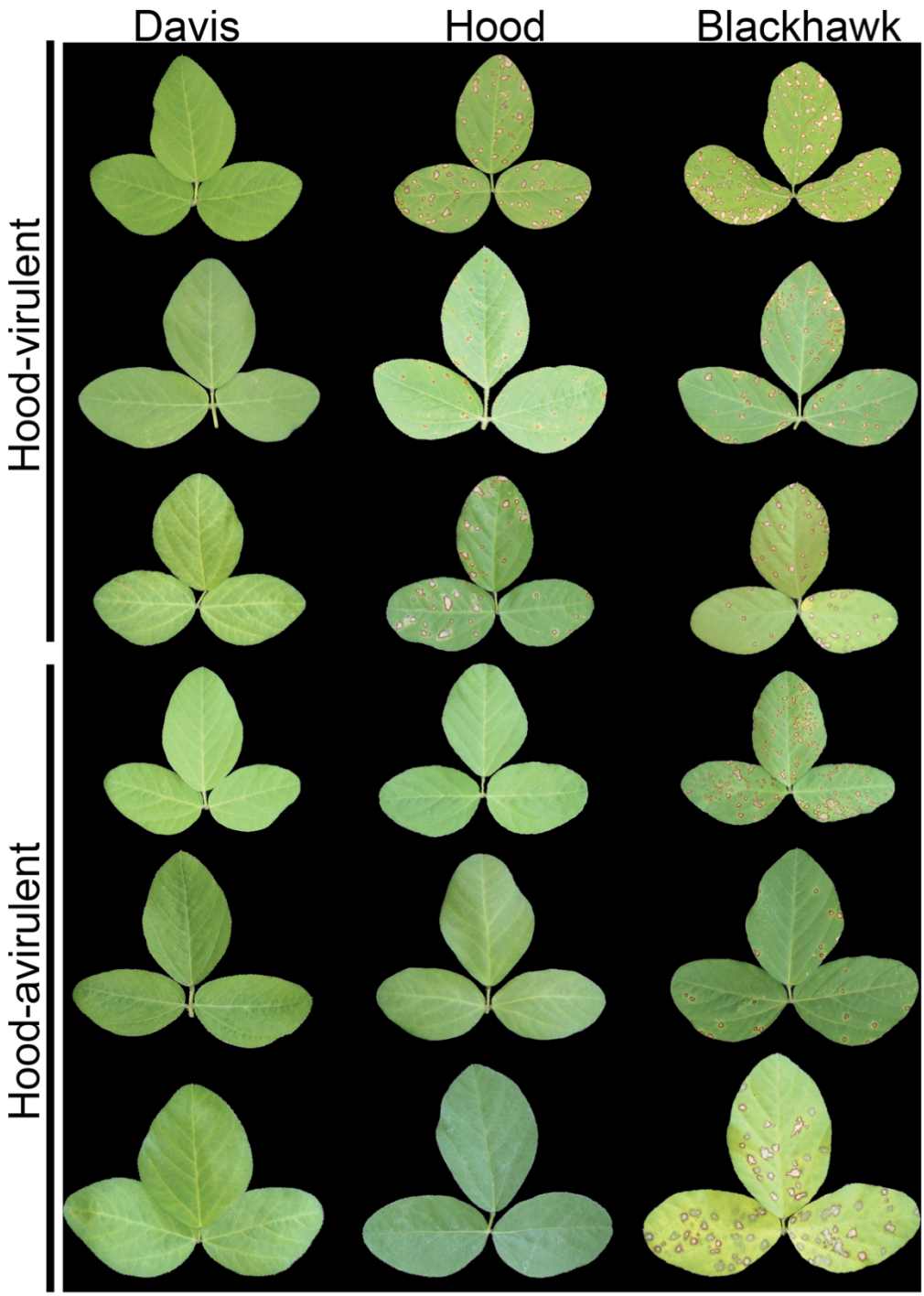

Fig. S2. Phenotypic reactions characterizing the two *C. sojina* subcollections analyzed in this study. Soybean plants were inoculated with *C. sojina* suspensions (1 - 3x10^5 conidia/mL) and scored 14 days after inoculation. Isolates were subdivided in subcollections that could infect Hood and Blackhawk ("Hood-virulent") and the ones that could not infect Hood but could still infect Blackhawk ("Hood-avirulent"). Cultivar Davis shown as the resistant control. *C. sojina* isolates from first to last row: Cs 2.2.3 (Arkansas); Csj 136 (Indiana); Csj AC14B10C (Illinois); Cs 108 (Arkansas); Csj 159 (Indiana) and Csj C14B5A (Illinois).

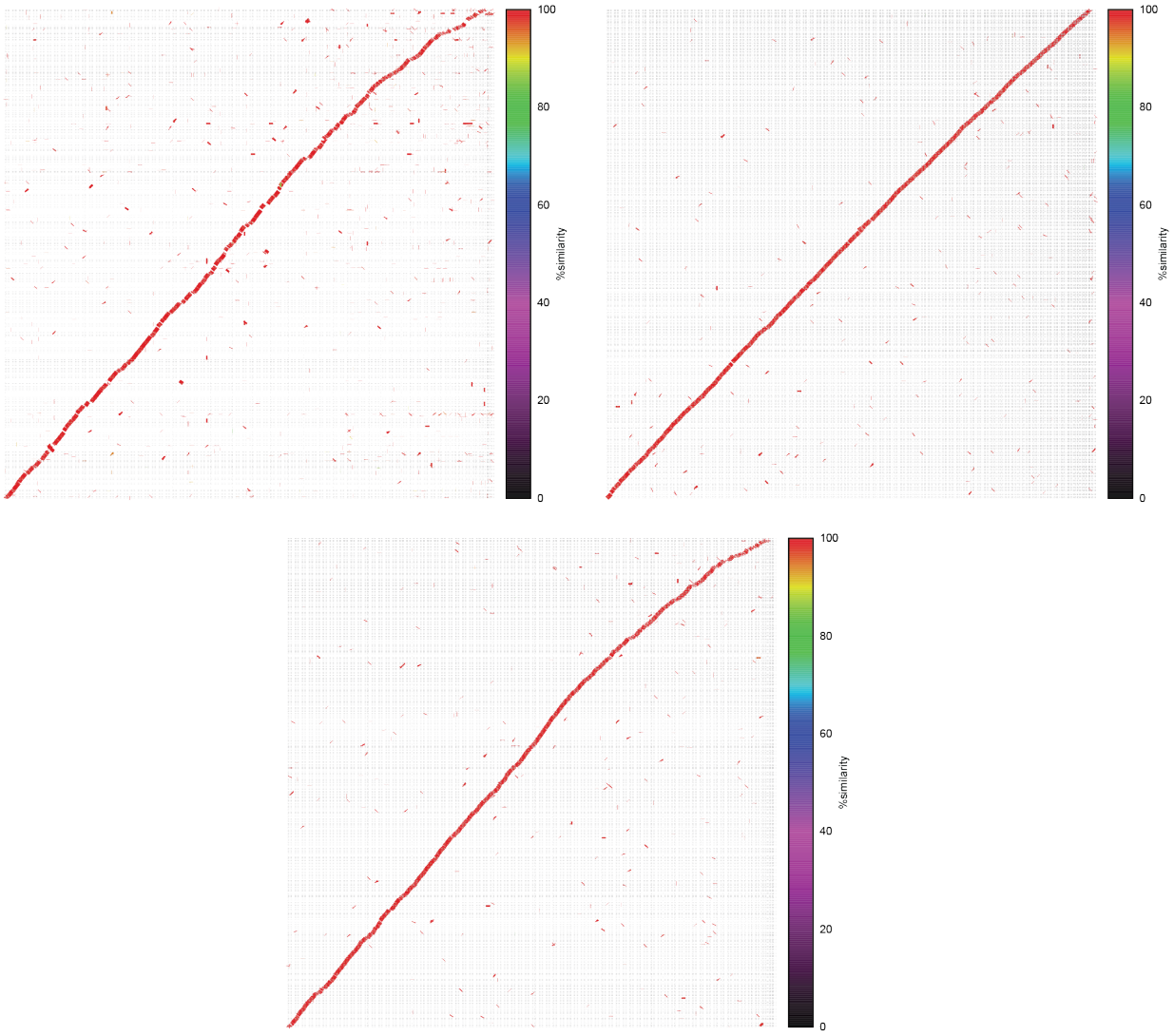

Fig. S3. Synthenic dotplots of the whole genome assemblies of *C. sojina* isolates 2.2.3, N1, and S9. The dotplots are based on the pairwise alignments of the genome assemblies of *C. sojina* 2.2.3 and N1 (top left), *C. sojina* 2.2.3 and S9 (top right), and *C. sojina* N1 and S9 (bottom), and were created with MUMmer package. Single nucleotide polymorphisms (SNPs) were identified with the command *show-snps* (parameters -IClr) within MUMmer. A total of 10,493 SNPs were identified between *C. sojina* 2.2.3 and N1; 5,246 SNPs between *C. sojina* 2.2.3 and S9; and 11,542 SNPs between *C. sojina* N1 and S9.

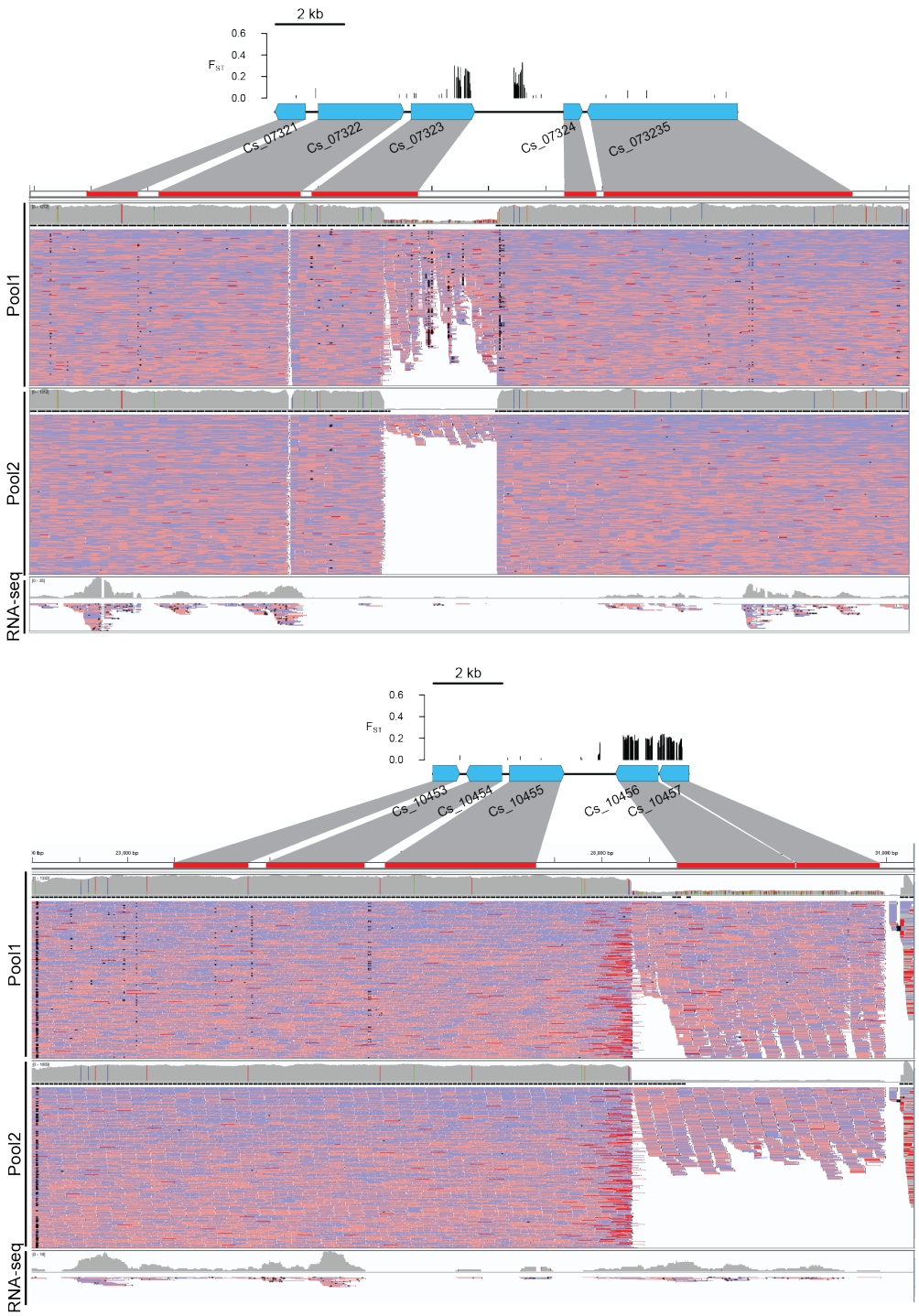

Fig.S4. Coverage of the sequencing reads around the genomic regions containing SNPs with high F_{st}. SNPs are represented as vertical black lines which height corresponds to the respective F_{st} value of the SNP. Gene ORFs are represented as blue rectangles. Illumina reads from pool1, pool2, and Ion Torrent RNA-seq mapped reads are indicated. Top: region within scaffold103; middle: end of scaffold251; bottom: scaffold504. Raw alignment files (bam) were visualized with IGV v2.3.57.

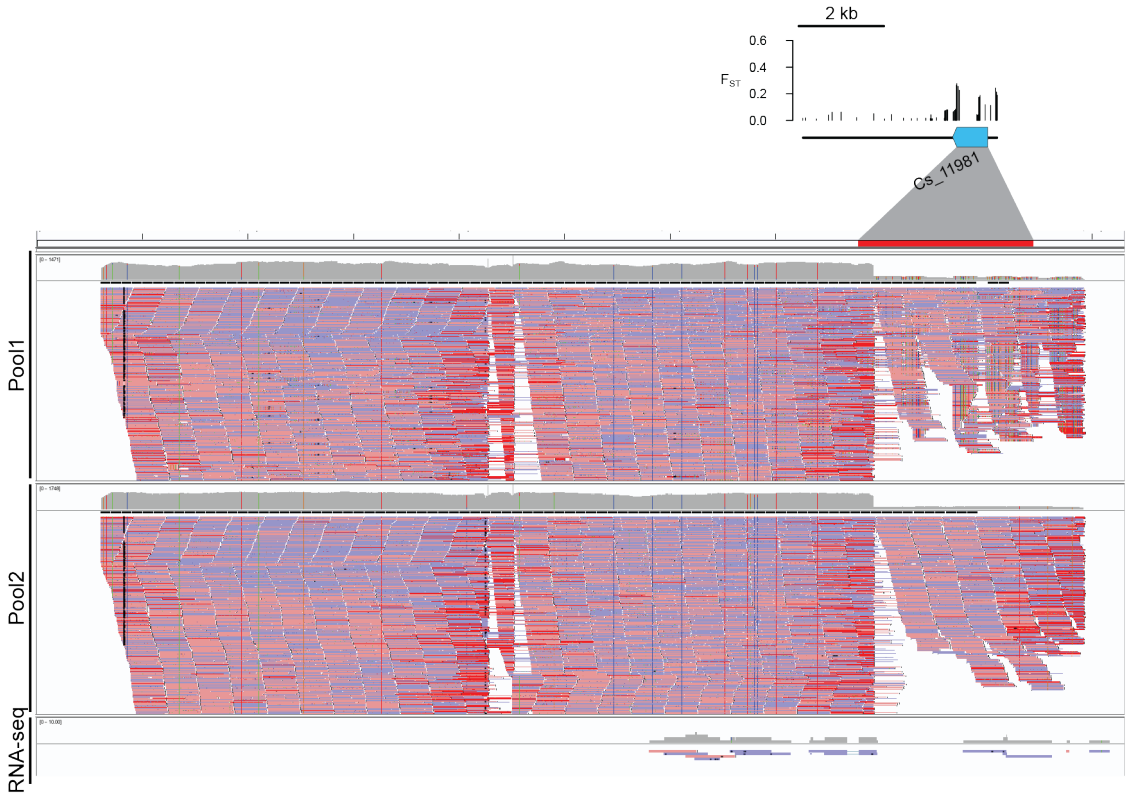

Fig.S4 (Cont.) Coverage of the sequencing reads around the genomic regions containing SNPs with high F_{st}. SNPs are represented as vertical black lines which height corresponds to the respective F_{st} value of the SNP. Gene ORFs are represented as blue rectangles. Illumina reads from pool1, pool2, and Ion Torrent RNA-seq mapped reads are indicated. Top: region within scaffold103; middle: end of scaffold251; bottom: scaffold504. Raw alignment files (bam) were visualized with IGV v2.3.57.

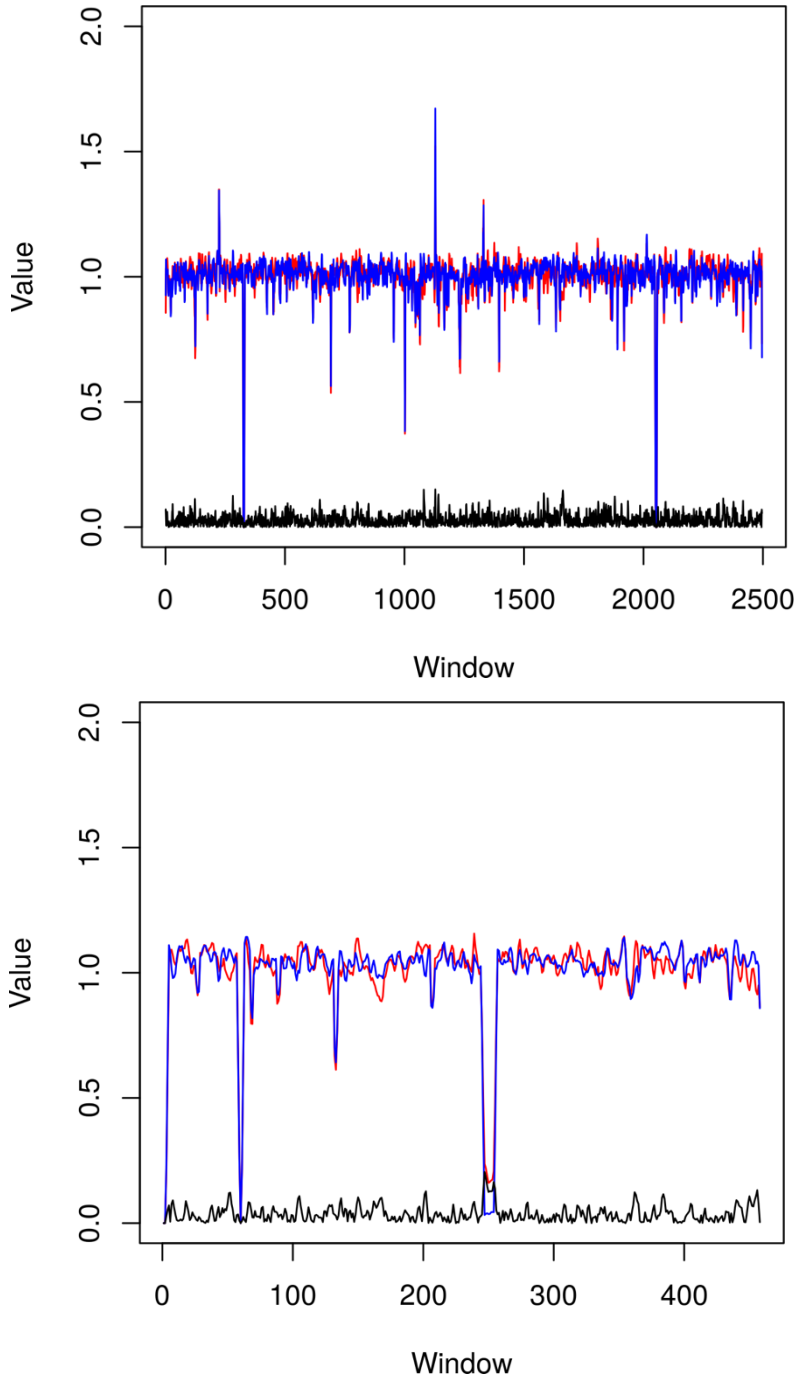

Fig. S5. Example of coverage analysis using sliding windows of scaffold1 (top) and scaffold103 (bottom). The red and blue lines are average coverage ratio (C_w) values for pool1 and pool2, respectively. The black line is the absolute difference between both C_w values. No significant region was found for scaffold1 and for scaffold103, the coverage drops around window 250, where a putative promoter region of the candidate effector Cs_07324 may be present. Although the coverage drops in both populations, this locus had the largest absolute difference of both Cw, between 0.12 and 0.20.

V. Appendix

Forward genetic screening on *C. sojina*

Forward genetics screenings have the ability to discover and functionally characterize genes through the analysis of random insertional mutants. Although the number of sequenced genomes in plant pathogenic fungi has increased drastically over the last years due to the advance in next-generation sequencing technologies, the vast majority of annotated genes lack functional characterization and their association with phenotypes. In *C. sojina*, the dissection of genes particularly involved in pathogenesis can bring valuable insights of pathogen biology, race structure and infection strategy, which can serve as novel targets for FLS control measures. To this end, random insertional mutants of the *C. sojina* 2.2.3 strain were evaluated to identify relevant phenotypes *in planta* and *in vitro*.

A library of more than 1800 random insertional mutants of *C. sojina* was created via *Agrobacterium*-mediated transformation of the 2.2.3 wild-type strain. Briefly, transformation events were carried out using Cs 2.2.3 conidial suspensions ($2 \times 10^5 - 1 \times 10^6$ conidia/mL) and the *Agrobacterium tumefasciens* strain AGL-1 (Lazo et al. 1991) harboring the pBHt2_sGFP plasmid, a plasmid derived from pBHt2 (Mullins et al.,2001). Conidia and bacteria suspensions (OD600 of 0.2 absorbance) were mixed in equal proportions and spread on sterile cellophane disks overlaying induction media agar (Mullins et al., 2001) containing 200 μM acetosyringone (AS; Sigma-Aldrich, Milwaukee, WI). After three days of culture in room temperature and constant darkness, cellophane discs were inverted and transferred to 0.2x PDA plates amended with cefotaxime (200 $\mu g\ ml^{-1}$) and hygromycin B (100 $\mu g\ ml^{-1}$) (Research Products International, Mt. Prospect, IL, USA). These cellophane discs were removed and discarded after five days, and colonies visually expressing GFP were transferred to 24-well plates containing 0.2× PDA

amended with hygromycin B (100 μg ml^{-1}) for continual growth. For long-term storage, small cuts of colonized agar cubes were made for each mutant generated, suspended in 50% glycerol (*v:v*) solution and stored at –80°C.

Pathogenicity screenings were performed on Blackhawk (susceptible control) and Davis (resistant control) soybean cultivars to assess the virulence phenotype of *C. sojina* mutants. The experimental design was composed of 4 plants per cultivar per mutant. Control treatments (sterile deionized water and Cs 2.2.3 wild-type strain) were also composed by the same number of plants. Soybean plants were grown in 4 inch-square pots at growth chambers (16h photoperiod; ± 25 ºC) until the first trifoliate leaf stage (V1) - approximately 20 days, at which point the plants were inoculated with *C. sojina* conidial suspensions or sterile water. Prior inoculation, five to seven-days old cultures of each mutant were flooded with sterile deionized water and conidia were dislodged with a sterile cell spreader, following addition of Tween 20 (0.003 vol/vol) to each suspension. Inoculation and FLS evaluation procedures were the same as used for the race phenotyping screening.

Approximately 396 mutants were screened in Blackhawk plants. From these screenings, 33 mutants demonstrated enhanced or reduced virulence phenotypes on Blackhawk cultivar (Figure 1). Re-screenings were performed in both Blackhawk and Davis cultivars to confirm phenotypes, in which conidial suspensions had their concentration adjusted to $1\text{-}3x10^5$ conidia/mL prior inoculation. From the repetitions, only four mutants were pointed out: one avirulent on both Blackhawk and Davis cultivars; two with reduced virulence on Blackhawk and avirulence on Davis; and one with virulence on both Davis and Blackhawk cultivars (Figure 2). There was no mutant with enhanced virulence and the mutants with reduced virulence phenotypes demonstrated only small flecks even 16 days-after inoculation, visually

differentiating from the wild-type control. Re-isolations from FLS lesions were also successfully performed. Mutants with distinct morphologies *in vitro* when compared to the Cs 2.2.3 wild-type strain were also observed (Figure 3).

Future directions will include whole-genome resequencing of the relevant mutant strains to identify the disrupted genes possibly underlying the observed phenotypes. Once genes are identified, complementation tests will be implemented for the functional validation of the gene being addressed upon restoration of wild-type phenotype. Further pathogenicity screenings will also be performed with additional mutants to search for relevant phenotypes.

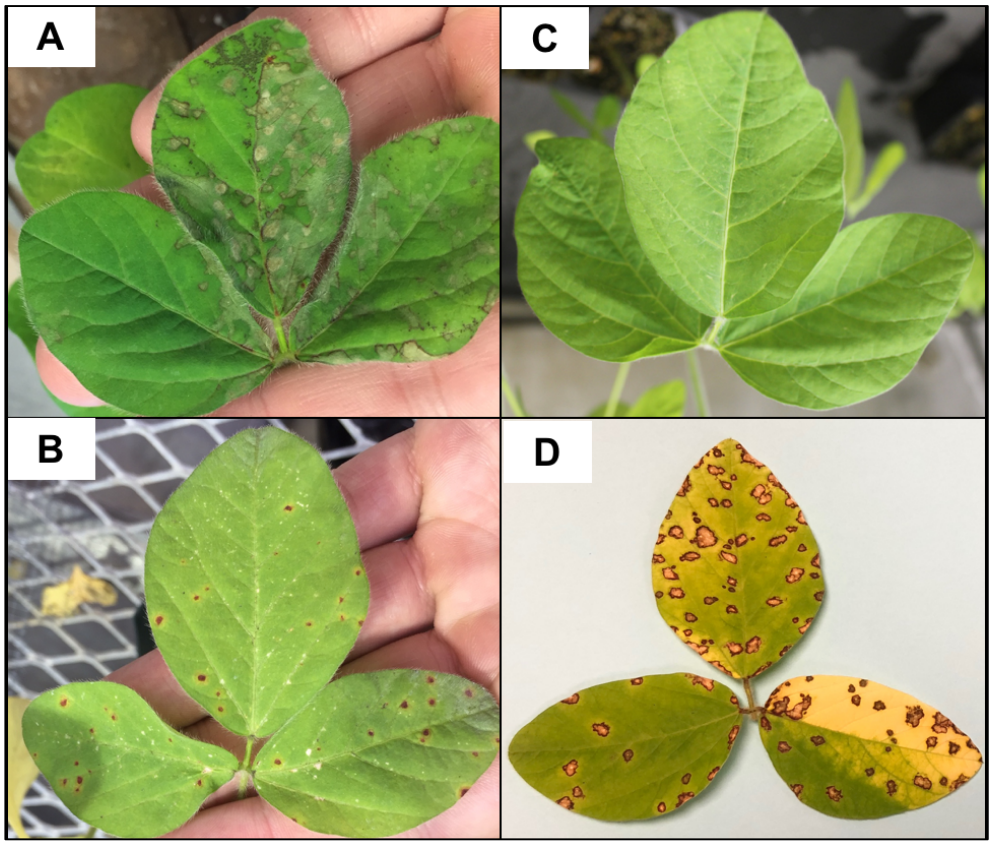

Fig. 1. Example of interesting phenotypes observed on Blackhawk plants among 33 mutants of *C. sojina*. Some mutants demonstrated enlarged, coalescent lesions, representing enhanced virulence isolates (A); Others demonstrated reduced virulence (B), in which lesions did not expand even 16 days-after inoculation; and the majority of mutants (79%) did not show any FLS symptoms, representing avirulent mutants (C) when compared to the Cs 2.2.3 wild-type strain (D).

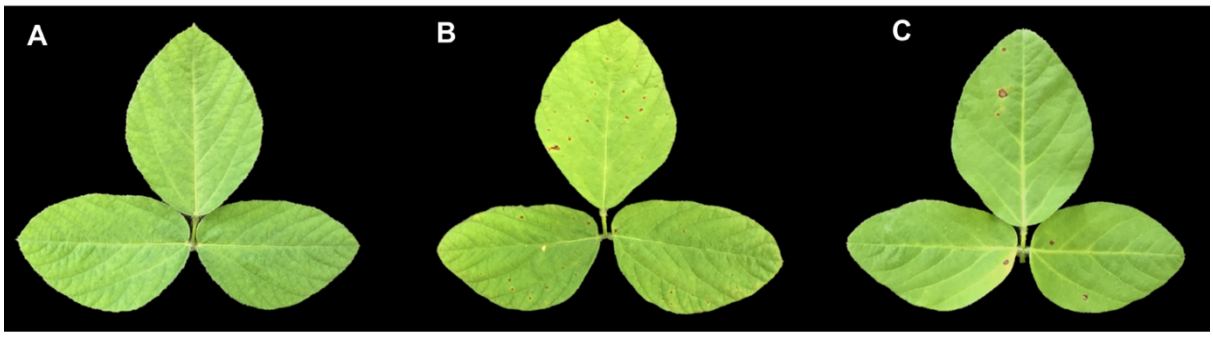

Fig. 2. Phenotypic reactions of the four *C. sojina* mutants with relevant phenotypes. One mutant did not cause any lesions on Blackhawk, representing an avirulent phenotype (A); Two mutants demonstrated reduced virulence (B), in which only small flecks were observed on Blackhawk plants even 16 days-after inoculation; and one mutant demonstrated virulence on both Blackhawk and Davis (C) cultivars.

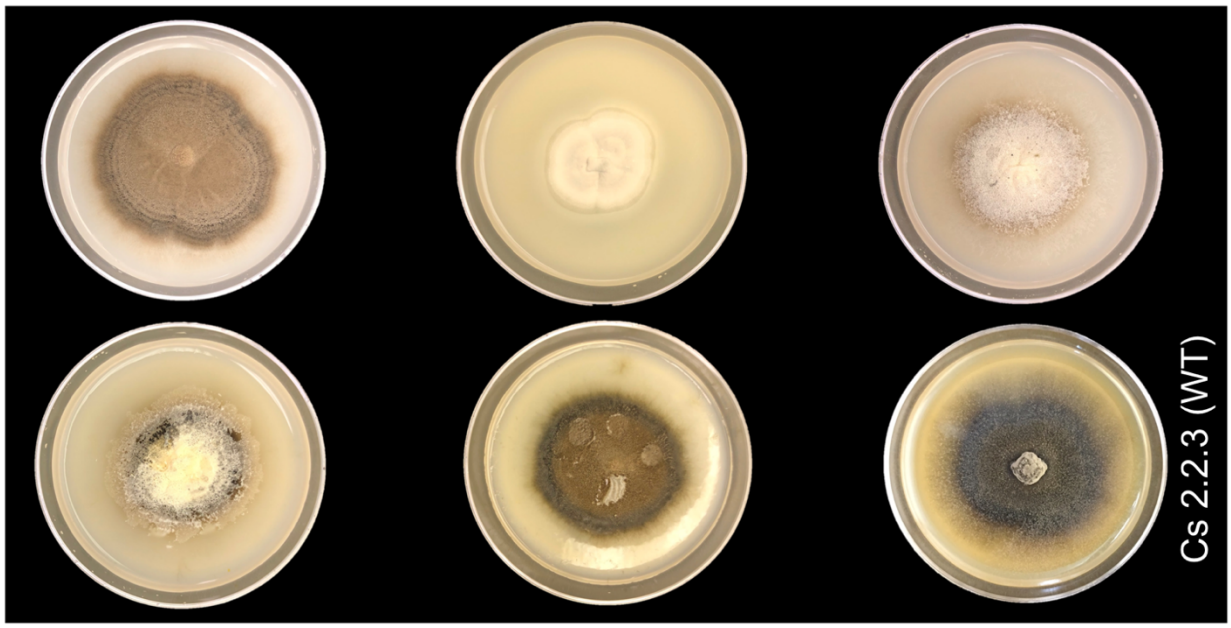

Fig. 3. Examples of *C. sojina* mutants with interesting morphologies. Distinct morphologies as albino mycelia (top row, second and third from left to right) and concentric rings (top first on the left) were observed. Mutants and Cs 2.2.3 WT strain were grown on PDA (Potato Dextrose Agar) under a 12/12h photoperiod. Pictures were taken from 10 days-old plates.

Citations used in this chapter:

Lazo, G.R., Stein, P.A. and Ludwig, R.A. (1991) A DNA transformation-competent Arabidopsis genomic library in Agrobacterium. *Biotechnol. Nat. Publ. Co.* 9, 963–967.

Mullins, E.D., Chen, X., Romaine, P., Raina, R., Geiser, D.M. and Kang, S**.** (2001) Agrobacterium-Mediated Transformation of Fusarium oxysporum: An Efficient Tool for Insertional Mutagenesis and Gene Transfer. *Phytopathology* 91, 173–180.