2015

Evolutionary Relationships among Fungal Soybean Pathogens and Molecular Marker Development in the Genus Cercospora

Sebastian Albu
Louisiana State University and Agricultural and Mechanical College, salbu@agcenter.lsu.edu

EVOLUTIONARY RELATIONSHIPS AMONG FUNGAL SOYBEAN PATHOGENS AND MOLECULAR MARKER DEVELOPMENT IN THE GENUS *CERCOSPORA*

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agriculture and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Plant Pathology and Crop Physiology

by
Sebastian Albu
B.A., University of Pittsburgh, 2001
B.S., Metropolitan State University of Denver, 2005
M.S., Louisiana State University, 2012
December 2015

I would like to dedicate this dissertation to my loving wife Ruoxi Chen and daughter Vanessa Albu.

ACKNOWLEDGEMENTS

I thank my wife, Ruoxi Chen for her infinite patience, unconditional love and help with statistical analyses and formatting of this document. I am also thankful to my parents, Magdalena and Valeriu, and brother, Alex for all of their support during graduate school. I owe a great deal of gratitude to my professors, friends and family for their unwavering support during what has been a precarious, but exciting journey. I am forever grateful to Dr. Raymond Schneider, whose generosity and belief in my abilities motivated me even in the darkest hours. This work would not be what it is without the guidance and expertise of Dr. Vinson Doyle, who has provided me with new perspectives of mycology and systematics, and has pushed me to become a better scientist. I would also like to thank my committee members, Dr.Chris Clark and Dr. Cathie Aime for their mentorship and inspiration. I am obliged to Dr. Quang Cao for agreeing to serve on my committee and also to Jeremy Brown for making thoughtful suggestions during several conversations. Many thanks go out to Dr. Lawrence Datnoff for his encouragement and efforts to create a wonderfully diverse and pleasant place to call home for the last five years. I enjoyed spending time in the field with Trey Price and thank him for providing many of the isolates that were used in this study. Additional isolates and plant material used in this research were kindly provided by Dr. Guohong Cai, Dr. Susan Li, Dr. Glen Hartman, Dr. Heather Kelly and Dr. Burt Bluhm (who was also very generous to provide the resources for genome sequencing). I am very thankful for the technical help that Sandeep Sharma provided during genome sequencing in Dr. Bluhm's lab and also to Don Nelsen who provided housing accommodations. I sincerely thank Haley Hutchins, Page Muse, Derek Clement, Caroline Baer and Charlot Amadi for their help in the lab. I have been fortunate to get to know many students and research associates during my time at LSU including my friends and lab mates, Brian Ward, Eduardo Chagas, Teddy Garcia, Elaisa Tubana, Clark Roberstson, Willie Vieira and Josiene Veloso. I will always look back fondly on my time spent with them and my other fellow graduate students and I hope our paths will cross again.

TABLE OF CONTENTS

ABSTRACT

Cercospora leaf blight (CLB) and purple seed stain (PSS) are common soybean diseases in the Gulf South of the United States (USA). For nearly a century, *Cercospora kikuchii* has been considered as the only pathogen causing these diseases. However, previous reports of genetic diversity among isolates collected throughout Louisiana suggested the presence of multiple lineages or species. Recent systematic studies classified species of *Cercospora* using a taxonomic system based on phylogenetic analysis of five nuclear loci (legacy genes). Using a similar approach, cercosporoid fungi tentatively identified as *C. kikuchii* were evaluated along with 53 other species of *Cercospora*. No isolates from this study were nested within the clade including the ex-type strain of *C. kikuchii*. Five isolates grouped with *C*. cf. *sigesbeckiae* and all others were part of *C*. cf. *flagellaris*. Several isolates of *C*. cf. *flagellaris* were also obtained from *Gossypyium hirsutum* and *Phytolacca americana*. These results suggest that *C. kikuchii* is not the organism responsible for causing CLB or PSS in the Gulf South and other areas of the USA.

Multiple haplotypes were observed at each locus and individual genes varied in their resolving power. Most species were monophyletic in concatenated analyses, but reciprocal monophyly was generally not observed within individual gene trees. Furthermore, node support values were generally low across all topologies, indicating that the phylogenetic markers most commonly used for systematic studies of *Cercospora* are limited to answering shallow-level taxonomic questions. However, existing genome sequence data provided an excellent opportunity to develop new markers with stronger phylogenetic signal to better understand the evolutionary history of *Cercospora*. Sixty-three exon-flanked intergenic regions, syntenic between the genomes of *C*. cf *sigesbeckiae* and *C*. *canescens*, were extracted and aligned, then ranked and filtered according to several metrics to assess their phylogenetic utility. Candidate markers were validated by PCR on 24 *Cercospora* species, including 16 type strains. Assessment of phylogenetic informativeness profiles and phylogenetic analyses showed that all of the new markers provide greater interspecific resolution than the legacy genes and offer new options for identifying cryptic species in complex clades like *C*. cf. *flagellaris.*

CHAPTER 1. INTRODUCTION

1.1 History of soybean

Soybean (*Glycine max* L. Merr.) is an annual forb in the leguminous plant family, Fabaceae. Varying reports exist regarding the origin of domestication, though it is believed that it was first domesticated from its extant wild-growing relative *G. soja* (Sieb. and Zucc.) 3000-5000 BP in one or potentially multiple locations in northeast Asia (Hymowitz, 1970; Hyten, et al. 2006; Lee, et al. 2011). Recent work tracing the evolutionary history of soybean based on differences in allelic length of simple sequence repeat (SSR) microsatellite markers in accessions of *G. soja* and *G. max* concluded that it was probably domesticated in central China, near the Yellow River (Li et al. 2013). There are several species such as *G. tabacina* and *G. tomentella*, which extend across East Asia and Australia, though the majority of species in the genus are found in Australia.

Prior to being used as a major food source for humans and other animals, soybean was exploited as an agricultural fertilizer. Like other legumes, soybean forms symbioses with root-inhabiting nitrogen fixing bacteria, which make it a valuable cover crop to replenish nutrient depleted soils, especially following plantings of nitrogen-depleting crops such as corn. Though it is primarily a crop best suited to temperate zones and subtropical areas, soybean has been introduced into tropical regions and has adapted well despite the lack of bacterial symbionts and photoperiod sensitivity. Soybean is a short-day plant and requires a certain number of hours of light and dark, but the development of different early and late-maturing cultivars has promoted the expansion of the cultivation range. Ideal growth conditions with respect to soil moisture and optimal temperature are similar to those of *Zea mays* (Hartman et al. 1999).

Soybean, along with various other commodities such as rice, maize, potatoes and cassava, is one of the chief agricultural crops cultivated by humans (Hymowitz 1970). It can be consumed directly as a staple, though other uses are favored. Frequently, soybean is processed into various comestibles such as milk and tofu or manufactured directly into high protein meal for livestock and poultry. Additionally, oilseed is an important component of biodiesel production, cooking oils and non-toxic substitutes for plastics and inks (http://www.soyatech.com/userfiles/file/tradeflow_manual(1).pdf).

1.2 Soybean as an agricultural commodity

Soybean was imported into the United States (USA) around 1765 (Hymnowitz 1970; Hymnowitz and Harlan 1983) and quickly became an important crop within 150 years (Hartman et al. 1999). Currently, the USA is the largest global producer and exporter of soybean, followed by Brazil, Argentina, China and India (http://www.usda.gov/nass/PUBS/TODAYRPT/cropan15.pdf). Together, these countries account for more than 90 percent of the global production of soybean. Within the USA, soybean production is predominantly concentrated in four regions of the country: North Central (Illinois, Indiana, Iowa, Michigan, Minnesota, Missouri, Ohio and Wisconsin,), Northern Plains (Kansas, Nebraska and South Dakota), Southeast (Kentucky, North Carolina and Tennessee) and Delta (Arkansas, Louisiana and Mississippi) (http://www.nass.usda.gov/Charts_and_Maps/Crops_County/pdf/SB-PR12-RGBChor.pdf). Soybean is also grown in other states not listed here as well as in central and eastern Canada. Within the Delta region, Louisiana is a major producer. Soybean production in Louisiana reached record highs in 2012, and yields were only surpassed by five other states in the USA

(www.lsuagcenter.com/news_archive/2012/november/headline_news/Louisiana-farmers-produce-record-soybean-crop.htm). Currently, the USA produces more than 30 percent of all soybeans grown worldwide. However, it is expected that Brazil will overtake the USA as the world's leading producer by 2016/2017.

1.3 Genetic diversity of soybean

Domestication of soybean has created a genetic bottleneck. This is a common phenomenon in other domesticated agricultural commodities. Domestication events typically lead to the rise of less genetically diverse groups of regional landraces. The genetic base of the crop is further diluted as humans select for individual landraces, leading to the development of isogenic lines of cultivars in other areas, often far away from where the plant was initially domesticated. As certain traits are selected for in a cultivar, that cultivar's genetic diversity will decrease and a genetic bottleneck will inevitably occur. Obviously, the extent to which this occurs will vary between different crops based on various other factors, but in general, diversity is typically higher in wild progenitors (Zaltsman and Citovsky 2012). Genomic allelic reduction of the germplasm of 75 Canadian wheat cultivars developed between 1845 and 2004 was reported beginning in the 1930s and has continued into the present (Fu and Somers 2009). A comparison of nucleotide diversity of Asian cultivated rice (*Oryza sativa*, subspecies *indica* and *sativa*) with two closely related wild species (*O. rufipogon* and *O. nivara*) found that while the levels of diversity within both *O. rufipogon* and *O. nivara* were comparable to other wild relatives of other domesticated crops, the *indica* and *sativa* rice subspecies maintained only 10-20 percent of the nucleotide diversity of their wild relatives (Zhu et al. 2007).

The marginalization of genetic diversity within crops is viewed as a potentially serious threat to global food production. As breeders have continuously selected varieties with the most desirable traits, the genetic base for certain crops has narrowed. Increased demand to produce large stands and greater yields has put pressure on breeders to develop high-yielding, vigorous new cultivars with improved phenotypic characteristics and disease and pest resistance. Though artificial selection enhances traits which benefit humans, it often does so at the expense of the plant. Selecting for varieties based on traits such as sweeter fruit, larger seeds, and higher nutritional content often comes with a price as these individuals are less fit, require more care and are usually unable to defend themselves against pathogens (Meyer et al. 2012). In soybean, however, there appears to be an inherently narrow genetic base. Hyten et al. (2009) argued against the commonly espoused theory that artificial selection in soybean contributes to reduced nucleotide diversity in modern elite cultivars. Instead, they proposed that genetic diversity in natural populations of *G. soja* is unusually low to begin with and that the domestication bottleneck was the primary event responsible for the majority of reduced genetic diversity in *G. max* and a loss of 81 percent of the rare alleles found in *G. soja*.

1.4 Soybean pathogens

Soybeans are susceptible to many diseases caused by fungi, oomycetes, prokaryotes, nematodes and viruses. The instances and severity of these diseases will likely increase as greater swaths of land are dedicated to more soybean production worldwide, and certain diseases may become more or less important with varying environmental conditions. There are approximately 35 known economically important soybean pathogens (Hartman et al. 1999). As a group, fungi and oomycetes were responsible for approximately 60 percent of the total estimated

yield loss in 2010 (Wrather and Koenning 2010). Within the group of fungal pathogens, this work is focused on species in the genus *Cercospora* that cause purple seed stain (PSS) and Cercospora leaf blight (CLB) on soybean.

1.5 Purple seed stain and Cercospora leaf blight

PSS and CLB were first reported in 1921 in Korea and Japan (Suzuki 1921). In the USA, PSS was first encountered in Indiana (Gardner 1926) then later in several other eastern states (Lehman 1928; Lehman 1950). The disease has had various monikers including purple speck of soybean, purple patch, purple blotch and lavender spot (Murakishi 1951). At the time of its discovery in 1922 by R. Kikuchi, the disease did not receive much attention since it was deemed relatively innocuous by growers (Matsumoto and Tomoyasu 1925), and Suzuki attributed the condition to climatic conditions (Murakishi 1951). Initially, only a small percentage of seeds was observed to be affected (Lehman 1950), but as the disease spread across many soybean growing regions symptomatic seeds were found with more frequency (Schuh 1992). PSS can cause reduced germination (Murakishi 1951; McLean and Roy 1988), stunting or death of seedlings (Lehman 1950) and low vigor (Yeh and Sinclair 1982), leading to reduced marketability (McLean and Roy 1988). The most conspicuous symptom of PSS is pink to purple discoloration of seed, often observed near the hilum. In severe cases, infection may appear on the surface of the embryo (Lehman, 1950).

Symptoms of CLB appear as purplish bronzed leaves followed by necrotic lesions first appearing on the upper foliage of the canopy and later on lower leaves, stems and petioles (Matsumoto and Tomoyasu 1925; Walters 1980). These symptoms manifest in late summer or early fall depending on the maturity group of the cultivar. It is believed that the recent severe and widespread outbreaks of CLB are attributable to the seed borne nature of the pathogen. However, the uniform distribution and severity of the disease in the field, the synchronous appearance of symptoms and the relatively low level of infection in soybean seeds argue against a role for seed transmission. On the other hand, even under the most severe epidemics with 100 percent disease incidence, sporulation is rarely observed on symptomatic tissue in the field, suggesting that airborne inoculum from within the soybean canopy is not involved in dissemination of the pathogen. Therefore, the etiology of CLB and PSS remains in question.

Isolation of a fungus from infected soybean material led Matsumoto and Tomoyasu (1925) to recognize the organism as a species of *Cercospora*, distinct from another soybean pathogen, *C. sojina* Hara. based on culture characteristics, morphology and host symptoms. The name *Cercospora kikuchii* (Matsumoto & Tomoyasu) Gardner has since been broadly applied to cercosporoid fungi associated with CLB and PSS worldwide and is still widely accepted as the causal agent of CLB and PSS, though recent work has shown that other generalist species of *Cercospora* can also cause similar symptoms (Soares et al. 2015).

1.6 Taxonomy of the genus *Cercospora*

Cercospora Fresen. is a genus of phytopathogenic fungi with a worldwide distribution. Fresenius erected the genus in 1863, naming it for the caudal, or tail-like, conidial features (Fresenius 1863; Chupp 1954). Saccardo (1880) later defined the genus in more detail, specifically as having brown conidiophores and vermiform conidia of variable color. Spegazzini (1910) split *Cercospora* and established the genus *Cercosporina* Speg. to accommodate those species having hyaline conidia. Matsumoto and Tomoyasu (1925) first placed *C. kikuchii* in this

genus based on conidial pigmentation. However, *Cercosporina* was later invalidated because the type species of *Cercospora* (*C. apii*) also has hyaline conidia (Chupp 1954).

Cercospora taxonomy has traditionally relied on morphology and host associations to discriminate species. In general, *Cercospora* spores are obclavate, usually hyaline or more rarely medium dark colored, often multiseptate, straight or curved and borne either laterally or terminally on pigmented fasciculate conidiophores (Chupp 1954). It is possible to discern among the structures of certain *Cercospora* species, but morphological homoplasy often confounds positive identification.

Teleomorph connections in *Mycosphaerella* Johansen have been established in some cases, but the sexual state for most species of *Cercospora* is not known (Goodwin et al. 2001, Crous and Braun 2003). The number of names has expanded and contracted as new species have been described and others synonymized. Chupp (1954) published the only monograph of *Cercospora*, in which he recognized 1419 species. He proposed a broad morphological concept for the genus primarily based on conidial characters and to a lesser degree on conidiophores and stromata. However, Chupp recognized that homoplasious characters are shared not only interspecifically, but also with other related genera such as *Alternaria*, *Cladosporium* and *Helminthosporium* (Chupp 1954). Since Chupp's monograph, more than 3,000 names have been published (Pollack, 1987), although many species have since been reduced to synonymy based on DNA sequence information. Crous and Braun (2003) recognized 659 *Cercospora* species and placed another 281 species in the *C. apii sensu lato* complex.

Chupp believed that *Cercospora* species are limited in their host range, but he recognized that pathogenicity tests would be necessary in order to confirm this hypothesis. There is now evidence indicating that while some species are host specific, others are broad generalists, capable of infecting many different hosts (Groenewald 2010). In a systematic study of *Cercospora*, Goodwin et al. (2001) used sequence data from the internal transcribed spacer (ITS) region to show that three isolates of *C. kikuchii* were closely related to the *Cercospora* species that attack banana, sorghum, asparagus and corn. Other recent studies using molecular phylogenetic approaches showed that multiple cryptic species of *Cercospora* are capable of infecting a single host species. Groenewald et al. (2005) rejected the synonymy of *C. beticola* and *C. apii* despite morphological similarities and overlapping host ranges. Later, two closely related species, *Cercospora zeae-maydis* and *C. zeina* were designated as separate species, though they were both found to cause gray leaf spot of corn (Wang et al. 1998; Crous et al. 2006). This phenomenon of cryptic speciation has also been observed in other ascomycetes such as *Colletotrichum gloeosporoides* (Doyle et al. 2013) and *Fusarium subglutinans* (Steenkamp et al. 2002) and raises the question of whether *Cercospora* species associated with other hosts can infect soybeans. McLean and Roy (1988) suggested that other hosts, including weeds, may be reservoirs of inoculum. Furthermore, it was shown that different *Cercospora* species derived from a wide range of hosts are able to cause purple discoloration on leguminous pods and seeds and other plants (Kilpatrick and Johnson 1956; Roy 1982).

Members of *Cercospora* are ubiquitous leaf spotters on many plant families of monocots, dicots, gymnosperms and ferns (Daub and Hangarter 1983; Goodwin et al. 2001; Crous et al. 2007). Some species produce cercosporin, an energetically activated compound that generates phytotoxic radicals such as singlet oxygen and superoxide when induced by light. The production of reactive oxygen species disrupts the structural integrity of the cellular membrane, particularly the lipid fraction. This physical damage leads to cytoplasmic leakage into intercellular leaf spaces and serves to facilitate nutrient uptake for the fungus (Daub and Chung

2007). Cercosporin was first isolated from *C. kikuchii* (Kuyama and Tamura 1957) and its function as a virulence factor has been widely investigated. When photoactivated, it can rapidly kill bacteria, plants and mice (Daub and Briggs 1983). It is unique among photosensitizers in that when it is exposed to light it can form a triplet state, reacting with oxygen to produce phytotoxins (Daub and Hangarter 1983).

Fajola (1978) suggested that cercosporin production may be of taxonomic value, proposing that only true *Cercospora* species produce the compound. However, cercosporin is not a universal virulence factor produced by all *Cercospora* species (Assante et al. 1977; Goodwin et al. 2001). Assante et al. (1977) tested 67 isolates corresponding to 61 species of *Cercospora* and found that only 28 were cercosporin producers. Separate studies found that while at least 34 species produced cercosporin, efforts to extract the toxin from another 51 species were unsuccessful (Jenns et al. 1989). Even among different strains, toxin production may vary depending on environmental and nutritional stimuli. My own experience working with many cultures of *Cercospora* can confirm that toxin production can vary among isolates or even among different cultures of the same isolate on different occasions, depending on media and other unknown factors (pers. obs).

1.7 Objectives

1. Determine whether *Cercospora kikuchii* is the pathogen causing CLB and PSS on soybeans in Louisiana and other states in the USA.

2. Develop phylogenetic markers to improve our understanding of the evolution of the genus *Cercospora*.

The incidence of CLB and PSS in Louisiana soybean fields has been widely attributed to *C. kikuchii* for many years. This is because of morphological homoplasy, similar disease symptoms and the notion that *Cercospora* species can be classified according to the host from which they were isolated. Host-based identification has prevailed for many years, but this method may be misleading. There is now evidence that while some species are indeed restricted to a single host, others are generalists, able to infect a large number of taxonomically diverse hosts. Two examples of this are *C. apii* and *C. beticola*, which have been isolated from many different plant genera, often overlapping in their host range (Crous and Braun 2003). Another example is *C*. cf. *flagellaris*, first described from pokeweed (*Phytolacca americana*) in Pennsylvania, USA and now known to infect at least ten different plant families (Groenewald et al. 2013).

The first objective of this research project was to build upon previous work that found genetic diversity among lineages of cercosporoid isolates collected from infected soybean leaves and seeds in Louisiana and identified as *C. kikuchii* (Cai and Schneider 2005; Cai and Schneider, 2008; Cai et al. 2009). These studies found that genetic diversity was high by using by microsatellite-primed PCR, RAPD-PCR fingerprinting and vegetative compatibility group pairings. Cai and Schneider (2008) speculated that the high levels of diversity implied that the sexual state may be functioning cryptically or may have been lost only recently. Sexual reproduction would be expected to generate high levels of genetic diversity among lineages of *C. kikuchii*, but I also considered the possibility that this diversity was indicative of the presence of additional species of *Cercospora*, each capable of causing CLB or PSS. In 2013, at the time

when this project was conceived, the possibility that CLB and PSS can be caused by multiple species of *Cercospora* had not been investigated. Cercosporoid fungi displaying symptoms of CLB or PSS were routinely called *C. kikuchii* by default and it was generally assumed that *C. kikuchii* is present wherever these diseases occur, especially in the Gulf South where climatic conditions favor outbreaks of CLB, and to a lesser degree, PSS.

There were, and still are, few existing molecular systematic studies that focus exclusively on *C. kikuchii.* A recent publication by Groenewald et al. (2013) provided a robust phylogeny for more than 50 *Cercospora* species. They delineated species boundaries for most taxa using five nuclear genes, which resolved most species-level taxonomic relationships, including *C. kikuchii*. Therefore, my rationale for this objective was to employ a genealogical concordance phylogenetic species recognition criterion using the same five nuclear genes to place isolates collected from soybean and other hosts throughout Louisiana and other states into a modern taxonomic framework. By doing so I hoped to get a better idea of the true identity of the pathogen(s) causing CLB and PSS in Louisiana and other states in the USA.

I was able to reproduce the phylogeny of Groenewald et al. (2013), and though most individual species were monophyletic, interspecific relationships remained unresolved. This indicated that the five nuclear genes, which have been widely adopted as the standard set of phylogenetic markers for systematic studies by the *Cercospora* community, are limited in their ability to resolve relationships above the species level. However, existing genome sequence data provided an excellent opportunity to develop a suite of additional markers with stronger phylogenetic signal to better understand the evolutionary history of *Cercospora*.

The rationale of the second objective was that syntenic gene pairs conserved throughout *Cercospora* could be identified to locate conserved regions for primer design to amplify intergenic regions. These intergenic regions were extracted and aligned, then ranked and filtered according to several metrics to assess their phylogenetic utility. A set of candidate markers were validated in the laboratory on unknown isolates and type strains of several *Cercospora* species. It was expected that the markers produced by this method would not only provide better resolution for complex clades and help to resolve cryptic species, but could provide a means for achieving a more reliable and stable taxonomic system for *Cercospora*.

CHAPTER 2. EVOLUTIONARY RELATIONSHIPS AMONG FUNGAL SOYBEAN PATHOGENS CAUSING CERCOSPORA LEAF BLIGHT AND PURPLE SEED STAIN IN THE GULF SOUTH

2.1 Introduction

Cercospora leaf blight (CLB) and purple seed stain (PSS) are diseases of soybean [*Glycine max* (L.) Merr] believed to be caused by *Cercospora kikuchii* [Mat and Tom (Gard)]. CLB symptoms appear as purplish-bronzing of leaves at the beginning of seed set. Necrotic lesions develop on the upper foliage of the canopy and later progress to the lower leaves, stems and petioles (Walters 1980). Until 1999, CLB was considered a minor disease in the Gulf South. However, it has since occurred more frequently (Cai et al. 2009), and yield losses have been reported from other soybean growing regions in the United States (USA) (Hershman 2009; Geisler 2013) and South America (Wrather et al. 2010; Almeida et al. 2006). *Cercospora kikuchii* also infects soybean seeds, leading to characteristic symptoms that appear as irregular purple to pink areas of discoloration that vary in size from small blotches to the entire seed. Favorable climatic conditions such as high atmospheric humidity and warm temperatures during pod development promote the spread of PSS, which occurs wherever soybeans are grown (Chanda et al. 2014). Yield loss from PSS is considered negligible, though the disease can reduce seed quality (Jackson et al. 2006) and stunt or kill seedlings (Lehman 1950).

PSS was first reported in 1921 in Korea and Japan (Suzuki 1921), though the disease had previously been observed in soybean fields. Kikuchi first associated PSS with a *Cercospora* species in 1922, but the disease did not receive much attention since it was deemed relatively innocuous by growers (Matsumoto and Tomoyasu 1925). PSS was first reported in the United States from Indiana (Gardner 1926), followed by several other eastern states (Lehman 1928; Lehman 1950). A small percentage of seeds was initially observed to be affected (Lehman 1950), but as the disease presumably spread across many soybean growing regions, symptomatic seeds were found with higher frequency (Schuh 1992).

Matsumoto and Tomoyasu (1925) first isolated and described *C. kikuchii* from purple-stained seeds in Japan and also reported that the fungus caused irregular purplish-red leaf spots or lesions on leaves, pods and stems. Similar symptoms were later reported from Taiwan (Han 1959) and the United States (Murakishi 1951; Walters 1980). Similarities between cultural and morphological characteristics of *C. kikuchii* and the fungus found to cause similar symptoms on soybean in North Carolina led Murakishi (1951) to conclude they were identical species.

A reliable and stable taxonomy for *Cercospora* has been hindered by the historical reliance on morphological characters, cercosporin production and host association for species identification. Morphology has traditionally been used to identify *Cercospora* species, though this method may be inadequate or misleading because of apparent homoplasy in the limited number of characters available for characterizing species. The sexual state for most *Cercospora* species is not known (Goodwin et al. 2001; Crous and Braun 2003), and some species, including *C. kikuchii*, do not readily sporulate in culture. When conidia are observed, length and number of septations often vary within the same species and are influenced by the amount of time conidia remain attached to conidiophores and by climatic conditions, such as atmospheric humidity (Atkinson 1891; Horsfall 1929; Murakishi 1951).

Some have proposed that only true *Cercospora* species produce cercosporin and that this toxin may be may be of taxonomic value (Fajola 1978). However, cercosporin production is not a synapomorphy shared by all *Cercospora* species (Assante et al. 1977; Jenns et al. 1989;

Goodwin et al. 2001), despite a positive correlation between pathogenicity and cercosporin production for some taxa. Assante et al. (1977) tested 67 isolates corresponding to 61 species of *Cercospora* and found that only 28 produced cercosporin. Similarly, Jenns et al. (1989) was only able to confirm cercosporin production in 34 of 85 species. Our own observations indicate cercosporin production can vary within a species and even within a single isolate, depending on environmental and nutritional stimuli (data not shown).

Chupp (1954) accepted 1419 species in his monograph based on the idea that *Cercospora* species are mostly host specific, which led to a proliferation of more than 3000 names (Pollack 1987). However, some species such as *C. apii, C. beticola, C.* cf. *flagellaris and C. zebrina*, have been isolated from a broad range of host species (Crous and Braun 2003; Groenewald et al. 2006; Groenewald et al. 2013). Others, like *C. zeae-maydis* and *C. zeina*, cause indistinguishable symptoms on the same host and may occur concurrently in the same field (Wang et al. 1998; Crous et al. 2006).

Since the description of *C. kikuchii* by Matsumoto and Tomoyasu in 1925, there have been few re-evaluations of this pathogen in a modern phylogenetic context. Several studies investigating genotypic and phenotypic variability within and among *C. kikuchii* populations (Cai and Schneider 2005; Imazaki et al. 2006; Cai and Schneider 2008; Lurá et al. 2011; Rapela et al. 2011) found differences among populations, but the identity of the fungus was based on host association. The internal transcribed spacer (ITS) rDNA region has been sequenced for many isolates collected from soybean and identified as *C. kikuchii*, but nearly all these sequences share at least 99 percent similarity with many other *Cercospora* species deposited in GenBank. Goodwin et al. (2001) showed that *Cercospora* is monophyletic within *Mycosphaerella*, but isolates identified as *C. kikuchii* were polyphyletic based on phylogenetic analysis of ITS. Groenewald et al. (2013) conducted a large-scale systematic study of more than 50 *Cercospora* species and established the monophyly of many taxa using sequence data from five nuclear protein-coding genes, and more recently, Soares et al (2015) investigated the phylogenetic relationships among CLB and PSS-causing soybean pathogens. The latter study found that several lineages comprised of *C.* cf. *flagellaris*, *C.* cf. *sigesbeckiae* and *C. kikuchii* were present in soybean producing regions of South America, Japan and Arkansas. The authors speculated that these CLB and PSS-causing lineages are endemic to Asia and only recently radiated outward as a result of the unwitting dissemination of infected plant material following the proliferation of soybean as a commodity crop. However, based on relatively sparse sampling of soybean growing regions throughout the Americas, there is little evidence to strongly support any one hypothesis explaining the origin of these lineages.

The purpose of this study was to test the hypothesis that multiple *Cercospora* species are associated with CLB and PSS, two diseases long believed to be caused exclusively by *C. kikuchii*. To determine phylogenetic relationships of isolates identified as *C. kikuchii* collected from Louisiana and other soybean producing states in the United States, we placed cercosporoid fungal isolates collected from infected soybean seeds and leaves in the Gulf South and Midwest within a phylogenetic framework using the five loci routinely utilized for systematic studies of *Cercospora* and sequenced other cercosporoid fungi collected from two non-soybean hosts to determine whether these isolates were the same species found on soybean.

Additionally, we investigated the potential for cryptic sexual reproduction in cercosporoid isolates from Louisiana. These isolates were previously identified as *C. kikuchii* by Cai (2004), who also found evidence of high genetic diversity based on vegetative compatibility group pairings. Cai (2004) speculated that the genetic diversity in this group occurred as a result

of a cryptically functioning sexual stage, but this hypothesis was not tested. The existence of a cryptic sexual stage was investigated in populations of another cercosporoid soybean pathogen, *C. sojina*, by Kim et al. (2013), who found equal proportions of MAT1-1 and MAT1-2 idiomorphs in Arkansas populations, suggesting that this pathogen is capable of sexual recombination. However, though the pathogen(s) clausing CLB and PSS is/are widely believed to reproduce clonally, this has not been investigated. Therefore, we characterized the mating type loci of some of the isolates from the collection of Cai (2004) and other isolates from Louisiana by calculating the ratio of MAT1-1 to MAT1-2 idiomorphs to assess whether the potential for sex exists.

2.2 Materials and methods

2.2.1 Collection and isolation

Cercosporoid fungal isolates were collected from soybean leaves and seeds displaying symptoms of CLB and PSS in 2000, and 2011-14. Additional isolates were also recovered from leaf spots and localized necrotic lesions on *Gossypium hirsutum* and *Phytolacca americana* during 2013 and 2014. Because the isolates from *G. hirsutum* and *P. americana* were morphologically similar to those from soybean, they were also included in this study. Detailed information about all isolates used in this study is provided in Appendix 1. Collection strategies, isolation techniques and culturing methods were previously described (Cai 2004; Price 2013; Price et al. 2015) for all isolates included in this study with the exception of those from *P. americana* and from a soybean seed lot originating in Hayti, MO, USA from 2011. Leaf spots on *P. americana* were examined for cercosporoid fungi under a dissecting microscope. No conidia were observed, but fasciculate conidiophores resembling those of *Cercospora* were removed from lesions with a flamed glass needle and transferred to potato dextrose agar (PDA - Becton Dickinson Difco, Sparks, MD, USA) amended with chloramphenicol (1 ml L^{-1}). Soybean seeds and seed coats from a 2011 lot produced in Hayti, MO, USA were dipped in 30 percent NaOCl solution at a an initial concentration of 8.25 percent, then rinsed in three changes of distilled water and incubated on water agar. Isogenic cultures were established from single colony forming units (CFUs) by scraping a small amount of mycelium from the edge of a growing colony with a flamed needle, placing in a 1.5 ml centrifuge tube filled with1 ml of sterile Milli-Q (EMD Millipore, Billerica, MA, USA) molecular grade water and vortexing. Fifty µl of the suspension was dispersed onto a 100 x 15 mm petri plate containing PDA amended with chloramphenicol (1 ml L^{-1}) using a flamed glass rod. Plates were incubated at room temperature and inspected after 12–18 hours under a dissecting microscope for the presence of hyphal growth. Individual CFUs were cut out of the agar with a sterile scalpel blade and transferred to 60 x 15 mm plates containing PDA.

2.2.2 DNA extraction

Isolates were grown on PDA amended with chloramphenicol (1 ml L^{-1}) for one week prior to extraction of genomic DNA. A small amount of mycelium was scraped from the surface of the colony with a sterile scalpel blade, transferred to a glass test tube containing 10 ml of Complete Medium (Jenns et al. 1989) and incubated in a rotary shaker at 180 rpm for 3–4 days. Approximately 500 mg of wet tissue was transferred to a sterile 1.5 ml tube and centrifuged at 14,000 rpm. Residual liquid was discarded and the remaining steps of the DNA extraction were

performed as specified in the Promega Wizard DNA Purification Kit (Promega, Fitchburg, WI, USA). All samples were quantified with a Nanodrop (Thermo Fisher Scientific, Waltham, MA, USA) spectrophotometer and diluted to a working concentration of 12.5 ng μl^{-1}.

2.2.3 Marker selection, PCR and data set assembly

Markers used in this study were chosen based on several recent phylogenetic studies of *Cercospora* (Groenewald et al. 2005; Crous et al. 2006; Groenewald et al. 2010, Groenewald et al. 2013). Primer sequences and PCR cycling parameters were obtained from the sources referenced below, though annealing temperatures were occasionally adjusted by several degrees for some isolates. Portions of actin (ACT), calmodulin (CAL) and translation elongation factor 1α (EF-1α) genes were amplified with the primers ACT-512F/ACT-783, CAL-228F/CAL737R and EF1-728F/EF1-983R (Carbone and Kohn 1999), respectively. Part of the histone 3 (H3) gene was amplified with the primers CYL H3F/CYL H3R (Crous et al. 2004) and *Cercospora*-specific mating type (MAT) genes were amplified with the primers CercosporaMat1f/CercosporaMat1r and CercosporaMat2f/CercosporaMat2r (Groenewald et al. 2006). ITS was amplified with the primers ITS1F (Gardes and Bruns 1993) and ITS4 (White et al. 1990) as previously described in Rush and Aime (2013).

PCR products were electrophoresed on 1 percent agarose gels at 120 V for 1 hr and visualized under UV light to confirm target amplification. In the case of non-specific amplification, multiple bands were excised and purified using the Promega Wizard PCR and Gel Cleanup System (Promega, Fitchburg, WI, USA). PCR products were direct sequenced using Big Dye Terminator chemistry on the Applied Biosystems 3730*xl* platform at Beckman Coulter Genomics (Danvers, Massachusetts, USA).

Sequences were manually edited and contigs assembled in Sequencher v5.0 (Gene Codes Corp., Ann Arbor, Michigan, USA). Generic identifications were made by comparing percent shared sequence identity of consensus sequences to others in the NCBI GenBank database with the blastn algorithm (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Nucleotide sequence alignments were estimated in MEGA6 (http://www.megasoftware.net/; Tamura et al. 2013) using the Muscle (Edgar 2004) algorithm or MAFFT v7 (Katoh and Standley 2013) specifying a G-INS-I iterative refinement method and a 200PAM / $_K$=2scoring matrix. Alignments were trimmed with GBlocks v0.91b (Talavera and Castresana 2007), specifying a less stringent selection that allowed gap positions with smaller final blocks and less strict flanking positions.

ACT, H3 and ITS (AHI) were consistently amplified in all isolates, but CAL and EF-1α did not consistently amplify or were too weak to sequence and non-specific bands were frequently produced for CAL. Therefore, in order to facilitate choosing an appropriate subset of isolates for additional sequencing and phylogenetic analyses, we performed a concatenated AHI haplotype analysis on all CLB and PSS isolates and 14 additional isolates of *Cercospora* from *G. hirsutum* and *P. americana* using DNAsp v5.10.1 (Librado and Rozas 2010). Each locus was first aligned individually and then concatenated into a single 1069 bp alignment. DNAsp was used for haplotype reconstruction using the algorithms in PHASE (Stephens et al. 2001; Stephens and Donnelly 2003). Additionally, we analyzed 69 MAT1-1 and 118 MAT1-2 sequences of 706 bp and 368 bp alignments, respectively, by the same method. The definition of a haplotype in this study is a single or group of sequences that is unique from all others based on at least one nucleotide polymorphism. We identified thirty AHI, nine MAT1-1 and seven MAT1-2 haplotypes (Figures 2.1 and 2.2). Members of each haplotype are listed in Appendix 2. CAL

and EF-1α were amplified and sequenced from 1–3 representatives of each AHI haplotype. We estimated structural alignments for three separate concatenated datasets containing different combinations of sequences to test genealogical concordance among individual gene trees and evaluate the impact of missing data on phylogenetic conclusions. The first dataset (DS-1) contained complete sequence information from ACT, CAL, EF-1α, H3 and ITS of 15 CLB and PSS isolates from 12 AHI haplotypes and 186 additional reference sequences from 54 species of *Cercospora* generated in Groenewald et al. (2013). The second dataset (DS-2) was a supermatrix comprised of sequences of the same five loci from each representative of the 30 unique AHI haplotypes plus 191 additional reference taxa representing 55 species of *Cercospora* generated in Groenewald et al. (2013). *Cladosporium herbarum* and *C.* cf. *subtilissimum* were used as outgroup taxa for both DS-1 and DS-2 because it was shown to be a sister group to *Mycosphaerella sensu stricto* within Mycosphaerellaceae (Braun et al. 2003). The third dataset (DS-3) contained the same loci as DS-1 and DS-2, but also included MAT1-1. The concatenated DS-3 alignment was a supermatrix consisting of nine CLB and PSS isolates with unique MAT1-1 haplotypes and also included 35 additional reference sequences from 14 species of *Cercospora*. *Mycosphaerella colombiensis* was chosen as an outgroup based on a close phylogenetic relationship to *Cercospora* (Crous et al. 2004b) and availability of sequence data.

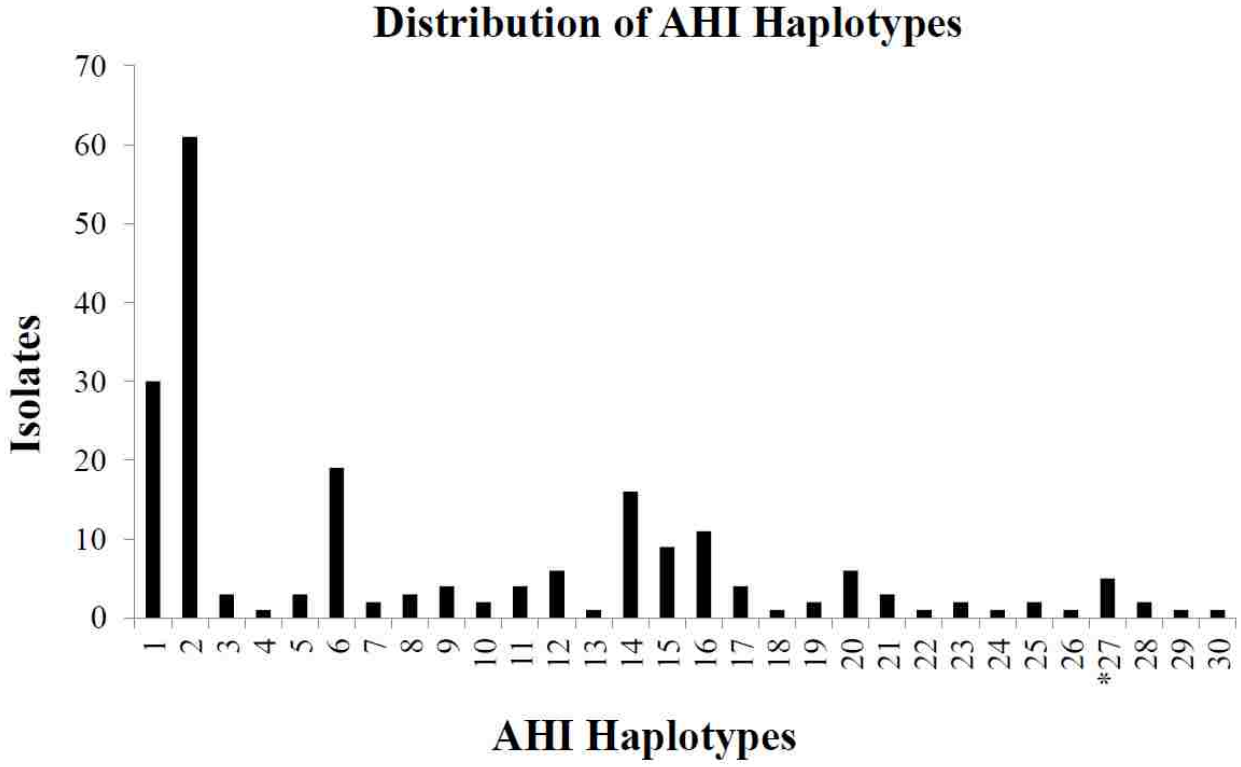


Figure 2.1. Frequency of haplotypes observed for isolates of *Cercospora* cf. *flagellaris* and *C.* cf. *sigesbeckiae* analyzed during this study. Numbers along x-axis show the 30 haplotypes observed among 207 isolates of *C.* cf. *flagellaris* and *C.* cf. *sigesbeckiae* for which sequences of actin, histone3 and ITS were concatenated and analyzed as single gene fragments in the alignment. Haplotype 27 (asterisk) exclusively contained five isolates of *C.* cf. *sigesbeckiae*.

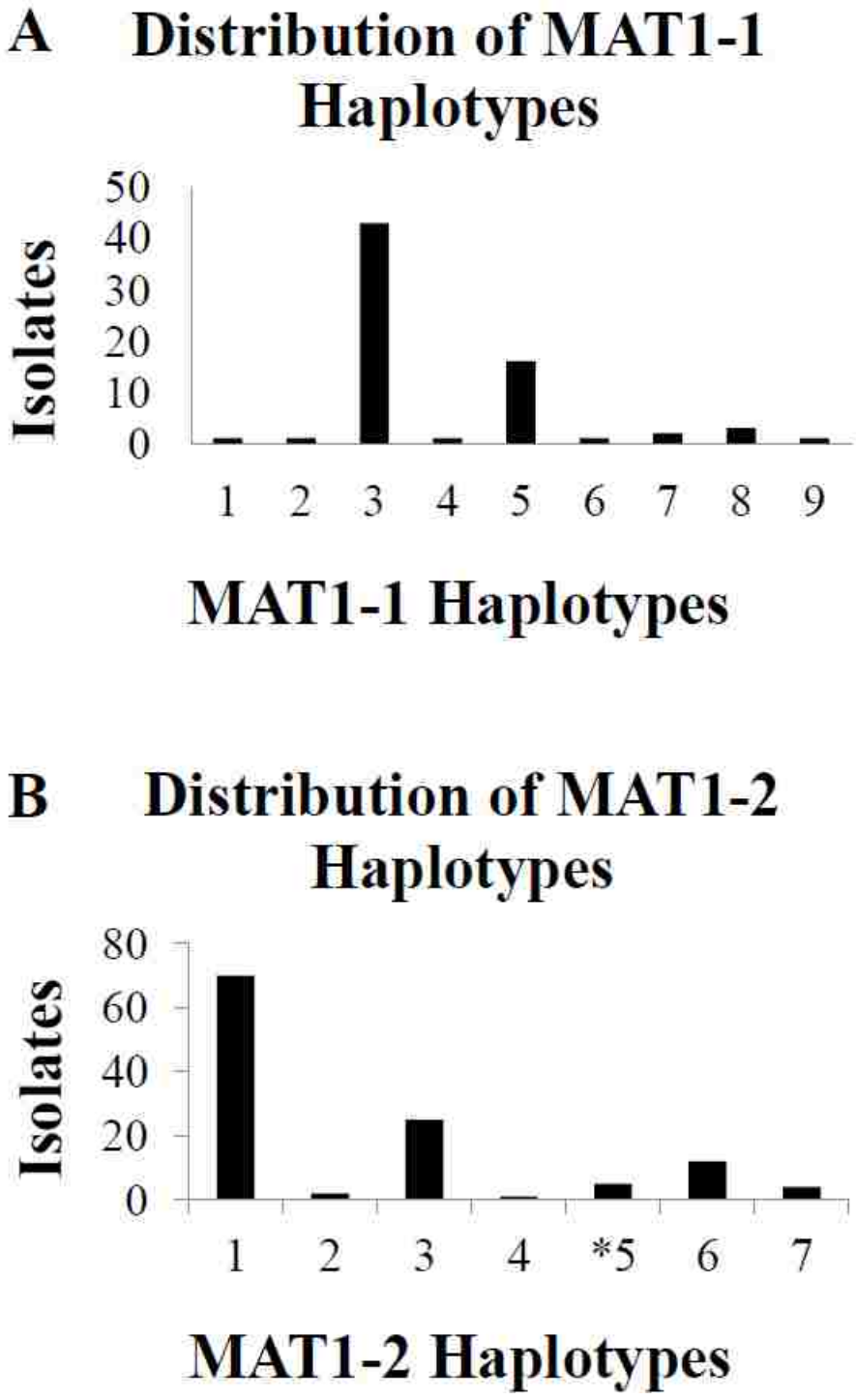


Figure 2.2. Frequency of haplotypes observed for isolates of *Cercospora* cf. *flagellaris* and *C.* cf. *sigesbeckiae* analyzed during this study. (A) Nine MAT1-1 haplotypes were observed among 69 isolates of *C.* cf. *flagellaris*. (B) Seven MAT1-2 haplotypes were observed among 118 isolates of *C.* cf. *flagellaris* and *C.* cf. *sigesbeckiae.* MAT1-2 haplotype 5 (asterisk) exclusively contained five isolates of *C.* cf. *sigesbeckiae*.

2.2.4 Model testing, phylogenetic analysis and species tree inference

The best-fit nucleotide substitution model for each partition was selected among a candidate set of 24 models according to corrected Aikake's Information Criterion (AICc) with MrModeltest v2.3 (Nylander 2004) for Bayesian inference. The best-fit model was chosen among a candidate set of 88 models according to AICc implemented in jModelTest 2 (Darriba et al. 2012; Guindon and Gascuel 2003) for maximum likelihood analysis in Garli v2.01 (Zwickl 2006). The substitution models selected for each partition are shown in Table 2.1.

Table 2.1 Nuclear substitution models used for phylogenetic analyses in chapter two

DS-1		
Locus	Model selected (MrModelTest 2)	Model selected (JModelTest 2)
ACT	HKY + G	TPM3uf + I + G
CAL	GTR + I + G	K80 + G
TEF1	HKY + G	HKY + G
H3	HKY I + G	JC
ITS	SYM + I	JC
DS-2		
Locus	Model selected (MrModelTest 2)	Model selected (JModelTest 2)
ACT	HKY + G	TIM3 + I + G
CAL	GTR + I + G	TrN + G
TEF1	HKY + G	TrN + G
H3	GTR + I + G	K80 + G
ITS	SYM + I	JC
DS-3		
Locus	Model selected (MrModelTest 2)	Model selected (JModelTest 2)
ACT	GTR + I + G	K80 + G
CAL	HKY + G	TrN + G
TEF1	HKY + G	HKY + G
H3	HKY + I + G	TrN + G
ITS	SYM + I	TrNef + I
MAT 1-1	HKY + G	TPM3uf + G

Groenewald et al. (2013) reported results from independent gene tree analyses, specifically, information about which species clades were not resolved in each analysis. However, these trees were not published. We felt that in order to provide a clearer picture of genealogical concordance and topological differences among individual gene trees, it was important to provide figures of both independent and concatenated analyses. Maximum likelihood analyses for independent gene trees and the concatenated tree from a partitioned dataset were estimated in Garli v2.01 using the resources at the Louisiana State University high-performance computing center (http://www.hpc.lsu.edu). Two separate analyses were run for each gene and the concatenated dataset. The maximum likelihood tree was generated by stepwise addition with 100 search replicates. Bootstrap proportions were estimated from a minimum of

1000 pseudoreplicate datasets, with the highest likelihood tree from two replicate searches per pseudoreplicate dataset retained. Bootstrap proportions were calculated and mapped onto the maximum-likelihood phylogenetic trees using SumTrees in the Dendropy v3.12.0 phylogenetic computing library (Sukumaran and Holder 2010). Additional maximum likelihood analyses were done using the CIPRES portal (Miller et al. 2010). Both independent gene and concatenated RAxML trees were estimated using RAxML-HPC2 (Stamatakis 2006) on XSEDE v8.1.11 with 1000 bootstrap replicates and specifying a GTR + G evolutionary model (raxmlHPC-HYBRID -n tre -s infile -x 12345 -N 1000 -k -p 12345 -f a -m GTRGAMMA). Bootstrap support values were estimated from 1000 pseudoreplicate datasets.

Bayesian inference was performed for both independent genes and the partitioned, concatenated data using MrBayes v3.2.3 (Huelsenbeck and Ronquist, 2001) on XSEDE via the CIPRES portal. Four replicates of 10 million generations each were run with four Metropolis-coupled chains, three heated to a temperature of 0.15. Trees were sampled every 1000th generation with the first 25 percent discarded as burnin. Tracer v1.6.0 (Rambaut et al., 2014) was used to assess convergence of the estimated parameters. Graphical representations of phylograms were exported from FigTree v1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/) and edited in Inkscape v0.48.1 (Harrington, 2004-2005).

Species trees were estimated from rooted gene trees under the coalescent method implemented in Maximum Pseudolikelihood Estimation of the Species Tree (MP-EST) (Liu et al. 2010) and Species Tree Estimation using average ranks of coalescence (STAR) (Liu et al. 2009). We assembled a complete dataset containing 54 species of *Cercospora* and *Cladosporium herbarum* using DS-1 as a template, but limited the number of sequences per taxon to three, which resulted in 117 terminals. Independent gene trees were estimated in RAxML specifying 100 bootstrap replicates, rooted with *Cladosporium herbarum* using the Species TRee Analysis Web Server (STRAW) (Shaw et al. 2013) and estimated using MP-EST and STAR in STRAW.

2.2.5 Characterization and statistical analysis of mating type loci

To test the null hypothesis that frequencies of MAT1-1 and MAT1-2 mating type loci did not significantly deviate from a 1:1 ratio, a chi-square goodness of fit test was performed using SAS v9.4 (SAS Institute Inc. Cary, NC, USA) assuming equal mating type frequencies. The sexual state of many species of *Cercospora* is not known, but it was expected that relatively equal proportions of MAT1-1 and MAT1-2 idiomorphs would be indicative of sexually reproducing populations. The MAT loci of 187 cercosporoid isolates obtained from soybean leaves and seeds collected in Louisiana during 2000, 2011 and 2012 were characterized by a multiplex PCR assay using the primers CercosporaMat1f/CercosporaMat1r and CercosporaMat2f/CercosporaMat2r (Groenewald et al. 2006). Confirmation of each idiomorph was made based on product size as visualized on a gel. MAT1-1 products were approximately 800 bp in length, while MAT1-2 products were approximately 450 bp. Because the isolates used in this study were not collected with the intention of performing a mating type population study, a hierarchical sampling strategy was not designed at the time of collection. This is an acknowledged area of weakness and poses a limitation to inferring sexuality versus clonality for these isolates. However, because no work has been done to characterize mating type idiomorphs of CLB and PSS pathogens, the existing resources provided a good opportunity for a preliminary study to assess the potential for sexual recombination in Louisiana.

2.3 Results

2.3.1 Phylogenetic analyses

Cercospora kikuchii reference isolates, including sequences of the ex-type, formed a strongly supported monophyletic group in each of the concatenated analyses (DS–1, DS–2, DS–3) using both maximum likelihood methods and Bayseian inference, but none of the isolates collected during this study was nested within this clade (Figures. 2.3, 2.4 and 2.5; Appendix 3 Figures A3.1 and A3.3). *Cercospora kikuchii* was weakly to moderately supported as monophyletic in the Bayesian inference and maximum likelihood independent gene analyses of ACT and CAL in DS–3 (Appendix 3 Figures A3.4– A3.6), but was paraphyletic or polyphyletic in H3 and all independent gene tree analyses of DS–1 (Appendix 3 Figures. A3.7–A3.12) and DS–2 (results not shown). The maximum likelihood trees inferred in RAxML and Garli were largely consistent, therefore all maximum likelihood results presented here and in supplementary figures are the inferences from RAxML.

All but five of the CLB/PSS isolates were distributed among 29 of the 30 AHI haplotypes, nine MAT1-1 haplotypes and six of seven MAT1-2 haplotypes. These isolates and 19 reference sequences of *C.* cf. *flagellaris* were monophyletic in all concatenated analyses (Figures. 2.3, 2.5, 2.6; Appendix 3 Figures. A3.1 and A3.2) except for the Garli analysis of DS–1, where *C.* cf. *brunkii* was nested within a clade containing all isolates of *C. cf. flagellaris* (data not shown). *Cercospora* cf. *flagellaris* was also monophyletic in the CAL gene trees inferred from DS–1 (Appendix 3 Figure A3.8), DS–3 (Appendix 3, Figure A3.5) and in the RAxML analysis of DS–2 (Appendix 3 Figure A3.13), but was paraphyletic or polyphyletic in all independent analyses of ACT, EF-1α, H3 and ITS (results not shown).

The five isolates that were not *C.* cf. *flagellaris* were all collected from symptomatic soybean leaves in four Louisiana parishes during 2012 and all belonged to AHI haplotype 27 and MAT1-2 haplotype 5, neither of which contained any other isolates.These isolates nested within a monophyletic *C.* cf. *sigesbeckiae* in all concatenated analyses (Figures 2.3 and 2.4; Appendix 3 Figures A3.1 and A3.3). *Cercospora* cf. *sigesbeckiae* was also weakly supported as monophyletic in all H3 gene trees, but was paraphyletic or polyphyletic in all other independent gene trees.

2.3.2 Species tree analyses

The topologies of the MP-EST and STAR species tree analyses (Figure 2.7; Appendix 3 Figures A3.14–A3.16) were generally congruent with the concatenated analyses of DS-1, DS-2, DS-3, though support values for interspecific and deeper relationships were generally weak to moderate. *Cercospora kikuchii*, *C.* cf. *sigesbeckiae* and *C.* cf. *richardiicola* were closely related, forming part of a lineage that also included *C.* cf. *malloti* and *C. rodmanii* in all analyses. *Cercospora kikuchii* was distinct from *C.* cf. *flagellaris* in all species tree analyses and comprised part of a lineage that always included *C.* cf. *nicotianae* and twice contained *C.* aff. *canescens* (Appendix 3 Figures A3.14 and A3.16).

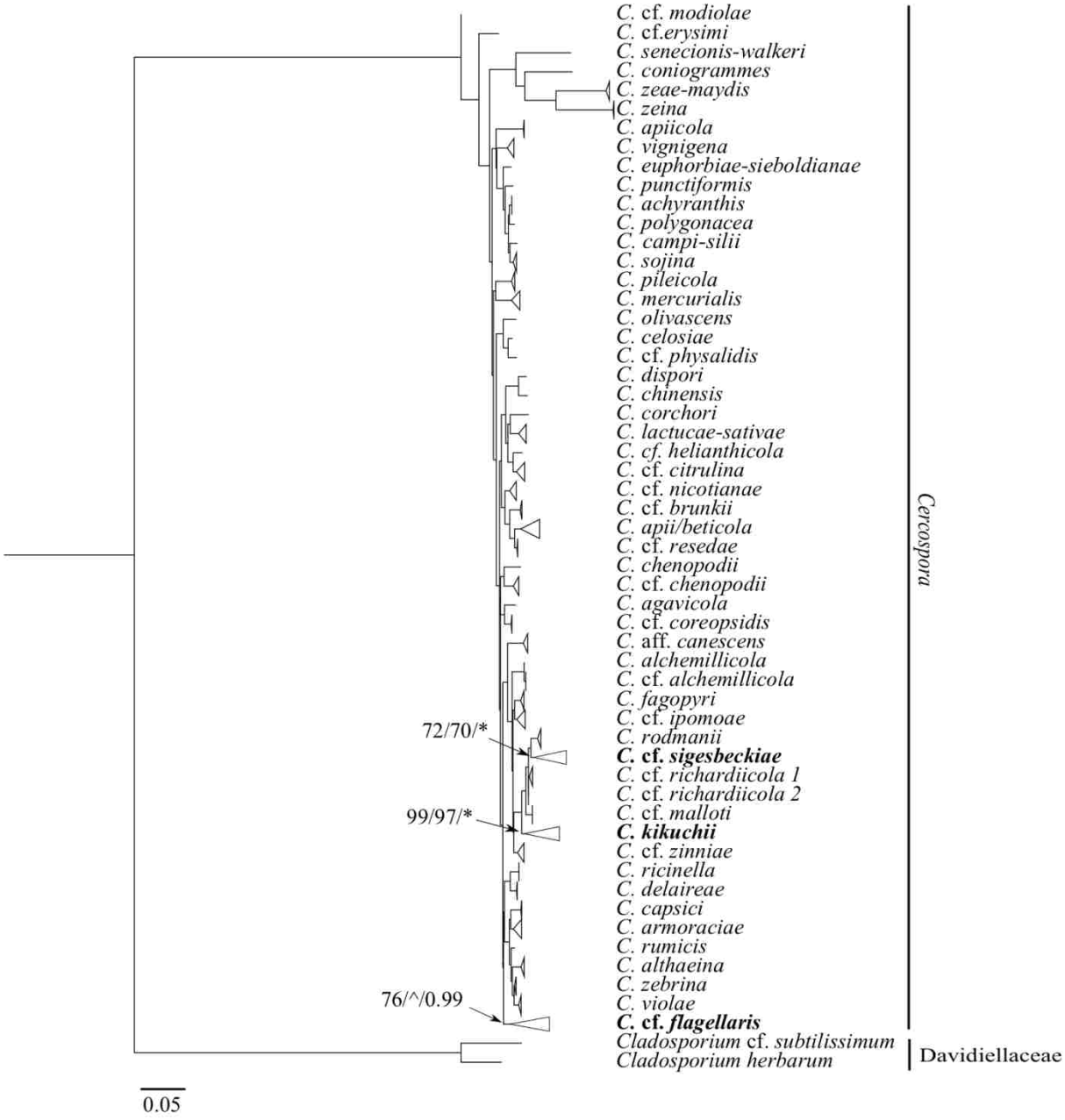


Figure 2.3. Maximum likelihood topology from RAxML analysis depicting the evolutionary relationships of 54 species of *Cercospora* based on a concatenated alignment of actin, calmodulin, translation elongation factor 1α, histone 3 and ITS sequences (DS-1). *Cercospora kikuchii*. *C.* cf. *flagellaris* and *C.* cf. *sigesbeckiae* are shown in bold. Tree is rooted with *Cladosporium* cf. *subtilissimum* and *Cl. herbarum*. Support values at nodes represent bootstrap percentages ≥70 obtained with at least 1000 replicates (RAxML/Garli) and Bayesian posterior probabilities >0.90 (on right) for *C.* cf. *sigesbeckiae, C. kikuchii* and *C.* cf. *flagellaris*. Asterisk indicates a posterior probability of 1. Caret indicates bipartition not present in respective analysis. Scale bar below tree indicates the number of substitutions per site.

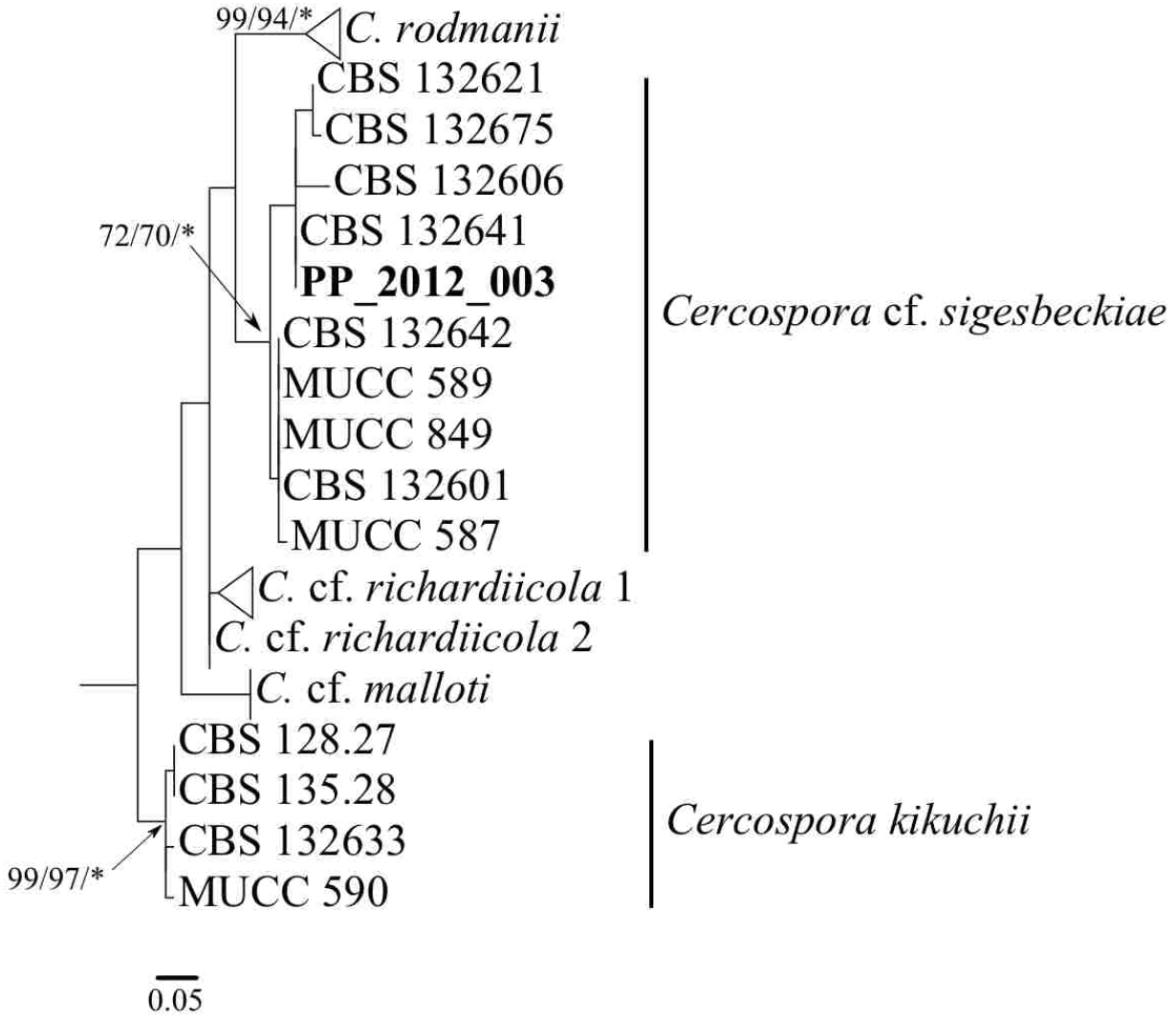


Figure 2.4. Clade containing *Cercospora kikuchii*, *C.* cf. *sigesbeckiae* and closely related species from Figure 2.3. Isolate shown in bold represents one AHI haplotype shared by all of the five isolates of *C.* cf. *sigesbeckiae* from this study. Support values at nodes represent bootstrap percentages of at least 70 (RAxML/Garli) and Bayesian posterior probabilities greater than 0.90 (on right). Asterisk indicates a posterior probability of 1. Scale bar below tree indicates the number of substitutions per site. Scale bar below tree indicates the number of nucleotide substitutions per site. Scale bar below tree indicates the number of substitutions per site. GenBank accession numbers for sequences used are provided in Appendix 1 Table A1.1.

2.3.3 Ratio of MAT1-1 and MAT1-2 idiomorphs

Using the multiplex PCR assay specific for *Cercospora*, the mating types of 187 isolates of *C.* cf. *flagellaris* from Louisiana were characterized. Of these, 69 were MAT1-1 and 118 were MAT1-2.

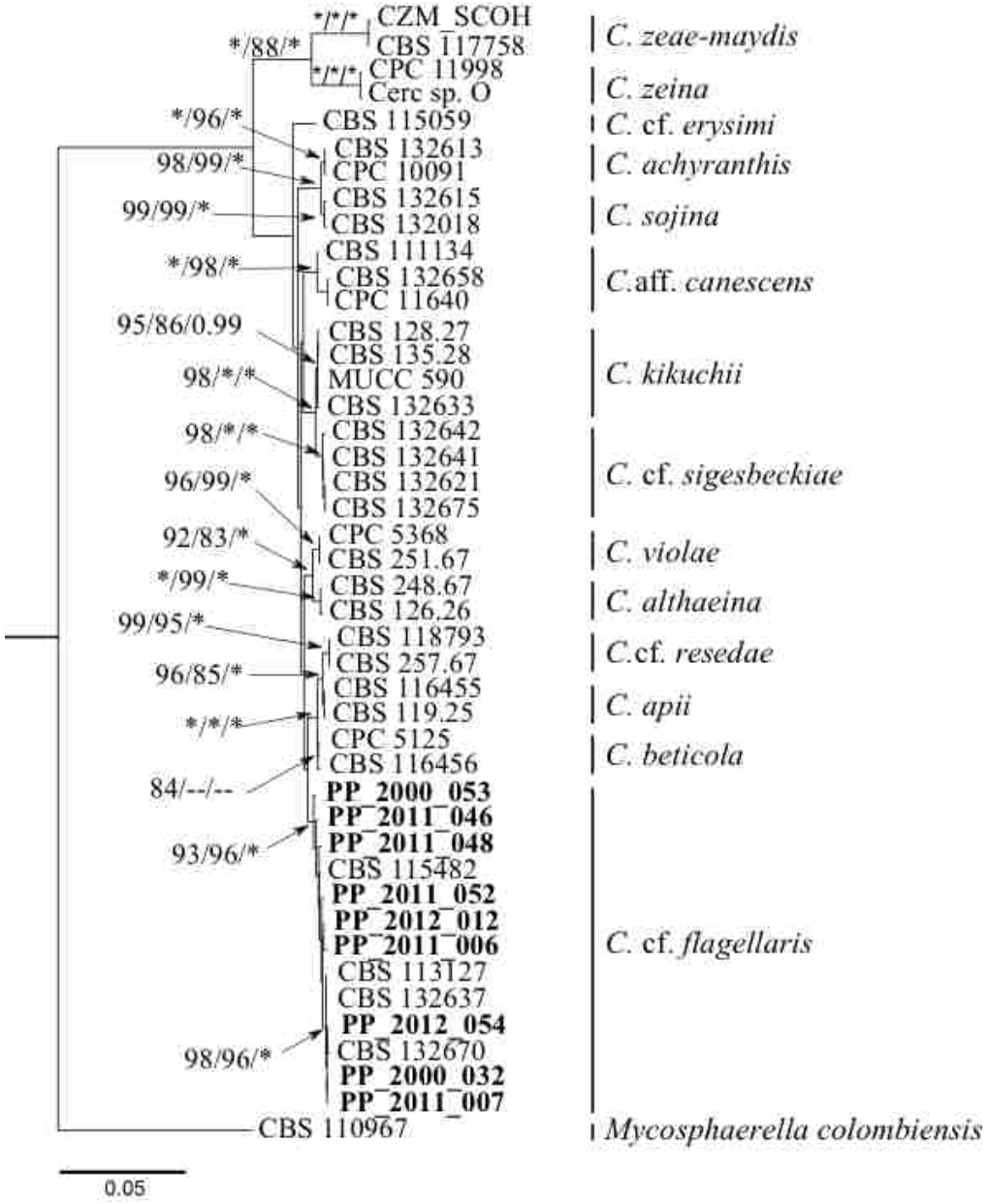


Figure 2.5. Maximum likelihood topology from RAxML analysis depicting the evolutionary relationships of 14 species of *Cercospora* based on a concatenated alignment of actin, calmodulin, translation elongation factor 1α, histone 3, ITS and MAT1-1 sequences (DS-3). *Cercospora* cf. *flagellaris* isolates from this study are shown in bold. Tree is rooted with *Mycosphaerella colombiensis*. Support values at nodes represent bootstrap percentages of at least 70 obtained with at least 1000 replicates (RAxML/Garli) and Bayesian posterior probabilities greater than 0.90 (on right). Asterisk indicates a bootstrap percentage of 100 or a posterior probability of 1. Double dashes indicate a bootstrap percentage less than 70 or a posterior probability less than 0.90. Scale bar below tree indicates the number of substitutions per site. GenBank accession numbers for sequences used are provided in of Appendix 1 Table A1.1.

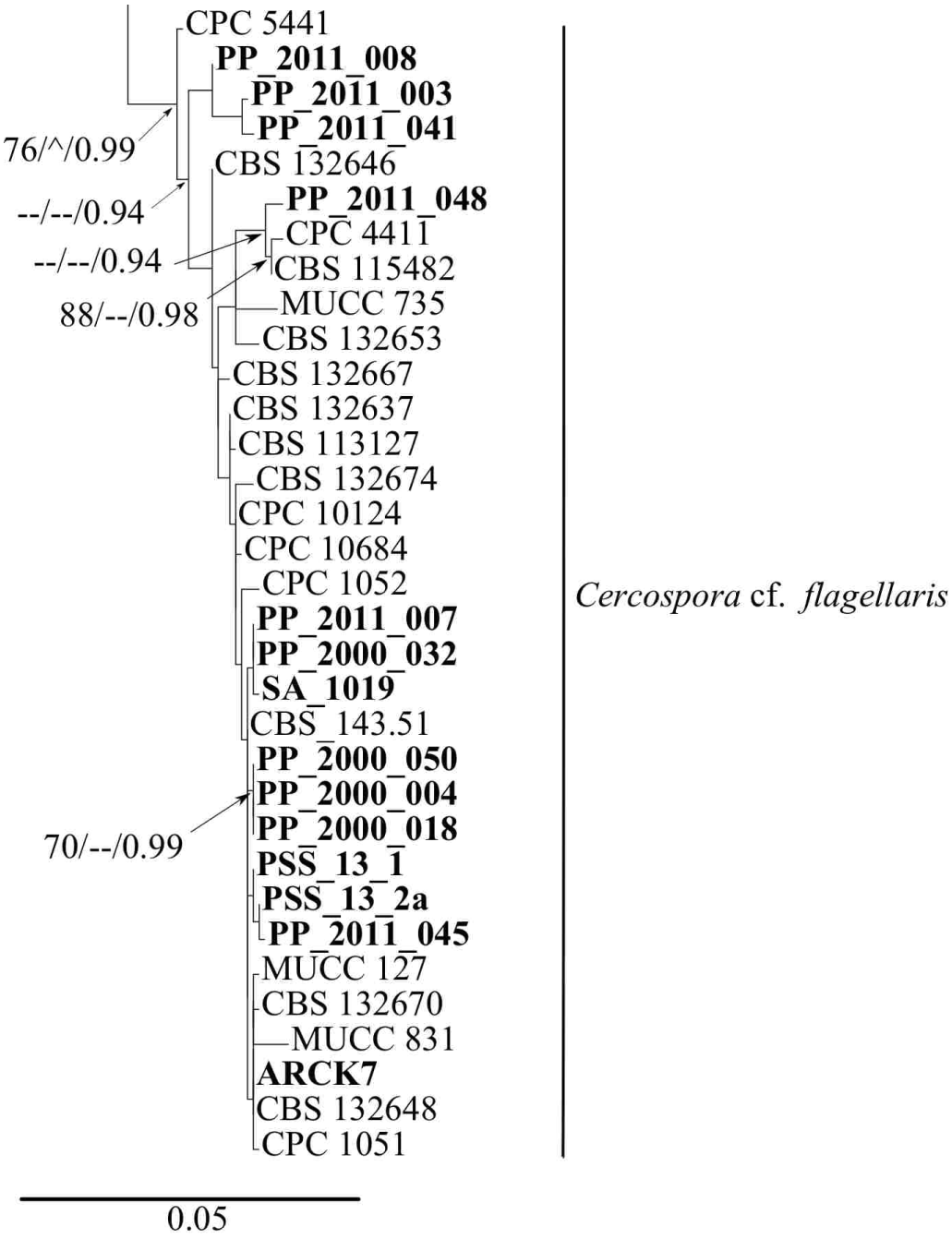


Figure 2.6. Clade containing *Cercospora* cf. *flagellaris* isolates from Figure 2.3. Isolates from this study are shown in bold. Support values at nodes represent bootstrap percentages of at least 70 (RAxML/Garli) and Bayesian posterior probabilities greater than 0.90 (on right). Double dashes indicate a bootstrap percentage less than 70. Caret indicates bipartition not present in respective analysis. Scale bar below tree indicates the number of substitutions per site. GenBank accession numbers for sequences used are provided in Appendix 1 Table A1.1.

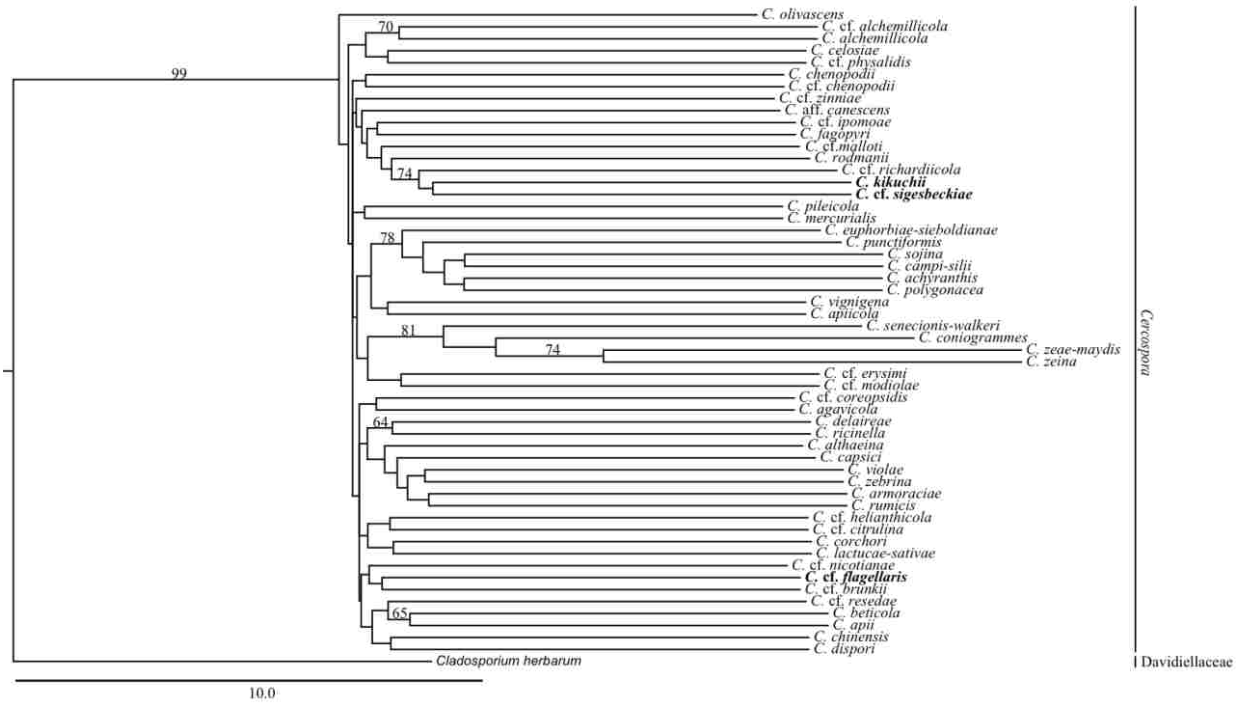


Figure 2.7. Species tree inferred from the five independent RAxML gene trees (DS-1) using pseudo-ML approach in MP-EST. *Cercospora kikuchii*, *C.* cf. *flagellaris* and *C.* cf. *sigesbeckiae* are shown in bold. Tree is rooted with *Cladosporium herbarum*. Support values at nodes represent bootstrap percentages >60 obtained with 100 replicates. Branch lengths are in coalescent units.

There were no instances where an isolate possessed both mating type loci. The five isolates of *C.* cf. *sigesbeckiae* were all MAT1-2, but these were not included in the chi-square test. Results from the chi square goodness of fit test showed that there was a significant departure from the null hypothesis of an expected 1:1 ratio of MAT1-1 to MAT 1-2 idiomorphs (χ^2 = 12.83) (Table 2.2).

Table 2.2. Mating type frequency and ratio of MAT1-1 to MAT1-2 idiomorphs of isolates of *Cercospora* cf. *flagellaris* collected from Louisiana

Mating type	Frequency	Percent	Test Percent	Chi-Square	DF	Pr > ChiSq
MAT1-1	69	36.9	50	12.8396	1	0.0003
MAT1-2	118	63.1	50			

2.4 Discussion

The assumption that *C. kikuchii* is the sole pathogen causing CLB and PSS on soybean globally has prevailed for many decades because of a paucity of distinguishing morphological

characters to differentiate species, persistent misconceptions about host specificity among *Cercospora* species and the lack of available sequence data representing a diversity of species within *Cercospora* to assess species assignments. In this study, we used multi-locus and coalescent-based phylogenetic approaches to determine the identity of species of *Cercospora* causing CLB and PSS on soybean. None of the cercosporoid isolates collected from symptomatic soybean leaves and seeds collected for this study were nested within the *C. kikuchii* clade indicating that *C. kikuchii*, the pathogen routinely referred to as the cause of CLB and PSS on soybean in the USA, is not causing these diseases. Two other species, *C.* cf. *flagellaris* and *C.* cf. *sigesbeckiae*, were associated with CLB and PSS. *Cercospora* cf. *flagellaris* represented all but five of the isolates collected in Louisiana and all of the isolates collected from Arkansas, Mississippi and Missouri. It was also isolated from leaf spots on pokeweed (*Phytolacca* sp.) collected in Arkansas, Illinois and Louisiana and from leaf spots on cotton (*Gossypium hirsutum*) in Louisiana (Appendix 1). The predominance of *C.* cf. *flagellaris* on soybean in Arkansas, Louisiana, Mississippi and Missouri suggests it is the primary species associated with CLB and PSS throughout the Gulf South and across other soybean growing regions in the United States.

Cercospora isolates cannot reliably be identified to species based on sequence similarity queries of GenBank using ACT, CAL, EF-1α, H3, ITS or mating type sequences. A nucleotide search of "*Cercospora kikuchii*" performed in the NCBI database on July 23, 2015 retrieved 159 hits (including two MAT1 and 2, four CAL, four H3, seven ACT, eight EF-1α, and 33 ITS sequences). However, because of low levels of polymorphism, these loci are not useful as barcodes, nor are they very informative outside of a multi-locus phylogenetic context. ITS is among the most frequently sequenced genes for studies of plant pathogenic fungi, despite being inadequate both as a barcode locus and as a phylogenetic marker for many fungal lineages (e.g. Crouch et al. 2009; Hyde et al. 2009; Ko et al. 2011), including *Cercospora*. For example, the ITS sequence from the ex-type of *C. kikuchii* shares 100 percent identity with six other isolates of *C. kikuchii*, one isolate each of *C. chinensis*, *C. guatemalensis*, *Cercospora*. sp. N, *C. tezpurensis*, *C. zebrina*, and two isolates of *C.* cf. *richardiicola*. It also shares 99 percent identity with multiple isolates of more than 30 other species of *Cercospora*. Blast queries with the other loci sequenced for this study are similarly equivocal.

Phylogenetic analyses of all concatenated datasets resolved some species within *Cercsopora*. However, similar to the results of Groenewald et al. (2013), the individual gene trees indicated an overall lack of phylogenetic signal for each locus, with few strongly supported monophyletic species. The term "phylogenetic signal" refers to "a tendency for evolutionarily related organisms to resemble each other" (Blomberg et al. 2003), while the term "noise" referes to the comfounding effect of homoplasy. In the DS-1 gene tree analyses, ITS was the least informative locus, where *C. zeae-maydis* and *C. zeina* were the only two taxa strongly supported by both maximum likelihood and Bayesian inference (Appendix 3 Figure A3.11). This was also observed by Groenewald et al. (2013). We found that CAL (Appendix 3 Figure A3.8) and H3 (Appendix 3 Figure A3.10) gene trees contained the most bipartitions with bootstrap values ≥70, respectively. This was partially in agreement with Groenewald et al. (2013), who reported that H3 and ACT provided the best species-level resolution, but also that no individual locus could distinguish among all species and that there were multiple instances of shared haplotypes among different species. *Cercospora* cf. *flagellaris*, *C. kikuchii* and *C.* cf. *sigesbeckiae* were among the species that Groenewald et al. (2013) found to be consistently problematic to distinguish. In our RaxML analyses of DS-1, *C.* cf. *flagellaris* and *C.* cf. *sigesbeckiae* were paraphyletic or polyphyletic in all gene trees except for CAL (Appendix 3 Figure A3.8) and H3 (Appendix 3

Figure A3.10), respectively, although support values were moderate to weak. *Cercospora kikuchii* was paraphyletic or polyphyletic in all independent gene tree analyses using maximum likelihood and Bayesian inference methods.

Groenewald et al. (2013) found that *C.* cf. *flagellaris* was monophyletic and composed of two sub-clades in their concatenated 5-gene analysis. The presence of these sub-clades, a wide geographic and host range and the lack of morphological differentiation among sub-clades led them to propose that *C.* cf. *flagellaris* is likely a species complex. We also observed intraspecific variation within *C*. cf. *flagellaris*, though the demarcation of intraspecific lineages was not as clear. If *C*. cf. *flagellaris* is comprised of several cryptic species, future systematic studies may help to determine whether genetic diversity within this species correlates to geography, host range or virulence. Recent phylogenetic analyses of other ubiquitous plant pathogenic fungi, including *Colletotrichum gloeosporioides* (Weir et al. 2012; Doyle et al. 2014), *Coll. acutatum* (Damm et al. 2012) and *Macrophomina phaseolina* (Sarr et al. 2014) discovered cryptic species using some of the same loci commonly used for *Cercospora*. However, some of these markers provide less phylogenetic information in *Cercospora* than in other fungal genera such as *Colletotrichum*.

Our results suggest that ACT, CAL, EF-1α, H3 and ITS do not possess sufficient phylogenetic signal to resolve interspecific and deeper-level evolutionary relationships among species of *Cercospora*. ACT, CAL, EF-1α and H3 are orthologous, single-copy protein-coding genes that are easily amplified in most species of *Cercospora* (excepting CAL, in our experience). While these are typically the hallmark traits of good markers (Feau et al. 2011; Curto et al. 2012; Kepler and Rehner 2013), the regions within these genes historically targeted for phylogenetic analyses are relatively short (<400 bp) and conserved among closely related taxa. As a result, relationships among divergent taxa, such as *C. zeae-maydis* and *C. zeina* are more strongly supported than those among members of more recently diverged lineages. In order to address questions related to interspecific relationships and make robust inferences of evolutionary history in *Cercospora*, additional markers with stronger phylogenetic signal are necessary.

The four reference isolates for *C. kikuchii* used in this study were collected from Japan and Argentina. The three Japanese isolates, including the ex-type, were isolated from *G. soja* and the single Argentinian isolate from *G. max*. Though there are reports of *C. kikuchii* from other *Glycine* species as well as other genera within Fabaceae, including *Cyamopsis*, *Dolichos*, *Phaseolus* and *Vigna* (cited in Farr and Rossman 2015), these isolates have not been evaluated in a phylogenetic context. *Cercospora kikuchii* is also not the only cercosporoid fungus reported from *Glycine*. Chupp (1954) reported 18 *Cercospora* species associated with the plant genera listed above, as well as the close *Glycine* relative, *Galactia*. It is notable that among these many pathogens, Chupp reported *C. cruenta* (now *Pseudocercospora cruenta*) from each of the above hosts except for *Cyamopsis* and *Galactia*, and there are multiple records of this species from 18 Fabaceae genera and also from Lamiaceae and Solanaceae (cited in Farr and Rossman 2015). This type of widespread geographic distribution and/or host generalism is also evident from collection records of *C.* cf. *flagellaris* and *C.* cf. *sigesbeckiae*, but in contrast to *C. kikuchii*, there is comparatively little information about *C.* cf. *flagellaris* or *C.* cf. *sigesbeckiae* in the literature. This is most likely because they have not been previously implicated as pathogens of economically important crops such as soybean and cotton.

Cercospora flagellaris Ellis & Martin was first described in 1882 from leaf spots on *Phytolacca decandra* (now *P. americana* L. var. *americana*) in Pennsylvania, USA. It is known

to infect other species of *Phytolacca* (Chupp 1954; Crous and Braun 2003; Kirschner 2013) and other genera in Phytolaccaceae, including *Rivina* (Solheim and Stevens 1931; cited in Farr and Rossman 2015) and *Petiveria* (Chupp 1954). Existing collection records show that *C.* cf. *flagellaris* is widespread, with reports from ten host families distributed across six countries. Bakhshi et al. (2015) collected isolates of *C.* cf. *flagellaris* from soybean and Levant cotton (*Gossypium herbaceum*) in Iran, and the recent manuscript by Soares et al. (2015) is the first published report of *C.* cf. *flagellaris* on soybean from the USA.

The syntype of *C. flagellaris* is currently housed in the herbarium of the Academy of Natural Sciences in Philadelphia, USA. According to the International Code of Botanical Nomenclature; "A syntype is any specimen cited in the protologue when no holotype was designated, or any one of two or more specimens simultaneously designated as types." Because this specimen of *C. flagellaris* is more than 100 years old, obtaining a viable culture from which DNA can be isolated and sequence data obtained will likely not be possible. Therefore, an epitype culture will need to be designated based on an isolate collected from *P. americana* growing in or around the type locality (West Chester, PA, USA). *Phytolacca* is ubiquitous, with reports from all but eight of the lower forty-eight US states as well as from several provinces in Canada. It is commonly encountered in disturbed sites, along partially shaded areas near forest borders (pers. observ.) and has been reported in or around soybean fields in Louisiana (Price, pers. comm), which may help to explain how inoculum is introduced into soybean fields from year to year. It is not known whether CLB is transmitted aerially or through seed, but *Phytolacca* plants infected by *C.* cf. *flagellaris* could serve as reservoirs of inoculum that initiate disease in soybean fields. Future systematic and population studies targeting *C.* cf. *flagellaris* collected from soybean and *Phytolacca* will be useful for assessing genetic diversity and the potential for gene flow between isolates from both hosts.

Cercospora sigesbeckiae Katsuki was described in 1949 from *Sigesbeckia glabrescens* (Asteraceae) in Japan and is morphologically similar to species in *C. apii s. lato* (Groenewald et al. 2013). The geographic distribution of *C.* cf. *sigesbeckiae* appears to be more restricted than *C.* cf. *flagellaris*, with all authenticated isolates originating from Asia and now the United States, where it has not been previously reported. However, like *C.* cf. *flagellaris*, *C.* cf. *sigesbeckiae* has a broad host range, occurring on *G. max*, *S. pubescens* as well as other hosts in six other plant families. There is a large disparity between the number of isolates of *C.* cf. *flagellaris* and *C.* cf. *sigesbeckiae* collected during this study. The five isolates of *C.* cf. *sigesbeckiae* were all obtained from symptomatic soybean leaves collected in Louisiana during 2012, which suggests that the presence of *C.* cf. *sigesbeckiae* may be incidental and that *C.* cf. *flagellaris* is the dominant pathogen in Louisiana.

Additional sampling is needed to determine if *C. kikuchii* is actually present in the USA. Unlike *C.* cf. *flagellaris* and *C.* cf. *sigesbeckiae*, there are no reports of *C. kikuchii* from any hosts but soybean and it may not be a generalist. Host specificity varies throughout *Cercospora* with some species possessing a much broader host range than others. Whether or not biological functions and reproductive strategies of broad generalists like *C.*cf. *flagellaris* and *C.* cf. *sigesbeckiae* are similar on all hosts is a difficult question to answer, but some pathogens may exhibit varying degrees of host preference. Crous and Groenewald (2005) proposed the 'pogo stick hypothesis' to explain how some species of *Mycosphaerella* and their relatives are able to colonize nonhost plants as a means to perpetuating themselves until they encounter a susceptible host plant. According to this hypothesis, as their primary hosts are dying, some pathogens have the ability to jump to other nonhost plant species where they can survive until they reach a

suitable host plant. If a similar scenario is occurring with PSS, the link between purple stained seeds and a single species of *Cercospora* may not be as strong as was once thought. Though it was believed that *C. kikuchii* was the only species capable of producing these symptoms on soybean seeds (Jones 1959), *in vitro* studies showed that other *Cercospora* species can cause pink to purple colored seed discoloration (Kilpatrick and Johnson, 1956). Cercosporin, the putative virulence factor responsible for the purple discoloration of seeds, is produced by *C. kikuchii*, *C.* cf. *flagellaris* and *C.* cf. *sigesbeckiae*, but it is not a synapomorphy for all *Cercospora* species. Rather, it is likely an adaptation that has allowed certain species to radiate by expanding their host ranges (Goodwin et al. 2001).

Soybeans were introduced into North America in 1765 (Hymnowitz and Shurtleff 2005), and virgin seeds would have been a prime target for an already-established pathogen. If one or more cercosporin-producing species such as *C. flagellaris* or *C. sigesbeckiae* were already present on other hosts in the vicinity of newly established soybean stands, they might have been primed to colonize and infect the newly introduced crop. This adaptation can be explained by the pogo stick hypothesis, where the introduction of foreign crops into existing agricultural systems drives pathogen evolution by encouraging aggressive pathogens to develop new strategies to colonize new hosts.

Soares et al. (2015) asserted that agricultural events probably did not contribute to the differentiation of lineages, though this is open to debate. If *C. kikuchii* was introduced into the Americas through infested soybean seeds, it could have become the dominant pathogen if it was more fit and better adapted to soybean than its new competitors. However, if a founding effect occurred with its introduction, its genetic base would have been narrow. Therefore, as soybean became a widely planted commodity, other generalist pathogens with more genetically diverse populations may have developed more efficient parasitic strategies, ultimately displacing *C. kikuchii*.

Soares et al. (2015) also reported that the A143G mutation in cytochrome b, which is associated with strobilurin fungicide resistance, was prevalent among haplotypes of lineage 1 from South America and those of lineage 4 from the USA. Given the probability that *C.* cf. *flagellaris* is the dominant lineage in the USA, it is also possible that fungicide treatments have been effective at suppressing a largely clonal population of *C. kikuchii* to the point that other more vigorous resistant species of *Cercospora* displaced it. The sexual state of *C. kikuchii* is not known, but if it is a clonally reproducing species, another host generalist able to undergo sexual recombination would have had a greater adaptive potential than *C. kikuchii*. Soares et al. (2015) mentioned the possibility of cryptic sexuality in *Cercospora* and proposed that parasexual recombination might explain genetic admixture in populations of isolates identified as *C. kikuchii* by Cai (2004) and Cai and Schneider (2008), but confirmed as *C.* cf. *flagellaris* in this study. We characterized the mating type loci of these and other isolates collected in Louisiana during 2011 and 2012. Among the 187 isolates for which mating types were determined, 69 were MAT1-1 and 118 were MAT1-2. This difference was significant ($P = 0.0003$), indicating that MAT1-2 occurs more frequently than MAT1-1. However, this finding must be interpreted with caution for several reasons. First, though there were fewer MAT1-1 idiomorphs among these isolates, it is obvious that both mating loci are present in Louisiana, which in itself validates the possibility that sex could be occurring. In a truly clonal population, a population would be expected to be represented by a single mating type (Duong et al. 2015). Furthermore, and more importantly, the sampling strategy for collecting these isolates was not was not designed for a population study and was not standardized across years. Therefore, to more thoroughly address

the issue of sexuality versus clonality within *C.* cf. *flagellaris*, additional studies need to sample populations from different soybean growing regions and also from other known hosts, as one or more of these plants may play a key role in this pathogen's reproductive cycle.

The results of this study are relevant to taxonomic mycology, as well as plant pathology and plant breeding. Management of CLB with fungicide applications has been marginal at best, and disease resistance has been elusive. In a recent comprehensive study, CLB/PSS isolates from Louisiana (including many of the ones used in this study) were uniformly resistant to all the commonly used soybean fungicides with the exception of triazoles, though a small percentage of isolates were resistant to this class of fungicide as well (Price 2013; Price et al. 2015). Integration of this information into a phylogenetic framework may help to explain the rapid development and spread of fungicide resistance and to what degree various temporal and spatial factors are involved.

Isolates of different *Cercospora* species able to cause CLB and PSS will provide breeders with theoretical and practical information as well as isolates to test the development of durable disease resistance in new soybean lines. Otherwise, cultivars released with resistance to only a fraction of the extant pathogen diversity would likely succumb to these diseases within a relatively short time, or may show geographically variable disease resistance. A diverse pathogen population may explain why previous work showed significant time and location interactions with regard to susceptibility in commercial cultivars (Cai et al. 2009). Simply put, accurate identification of a pathogen is an imperative prerequisite to developing an integrated defense management approach including genetic resistance and chemical treatment.

CHAPTER 3. PHYLOGENETIC MARKER DEVELOPMENT FROM INTERGENIC REGIONS FOR INFERRING RELATIONSHIPS AMONG SPECIES OF *CERCOSPORA*

3.1 Introduction

Until recently, *Cercospora kikuchii* Matsumoto & Tomoyasu (Gardner) had long been thought to be the sole cause of Cercospora leaf blight (CLB) and purple seed stain (PSS) on soybean [*Glycine max* (L.) Merr]. However, recent studies showed that in addition to *C. kikuchii*, several other species of *Cercospora,* including *C*. cf. *flagellaris* and *C*. cf. *sigesbeckiae,* are implicated in causing these diseases throughout Asia, South America and the USA (Soares et al. 2015; chapter two of this dissertation). However, because they have not been definitively linked to important agricultural commodities like soybean, until now, these species have received little attention. Collections records show, some species, such as *C. kikuchii,* appear to be host specific, while others, such as *C*. cf. *flagellaris* and *C*. cf. *sigesbeckiae,* are clearly not. Rather, *C*. cf. *flagellaris* and *C*. cf. *sigesbeckiae* are both generalist pathogens that have been isolated from many plant families across several continents (Groenewald et al. 2013; chapter two of this dissertation). This new information has the potential to change the way that soybean pathologists and breeders look at disease management and breeding for disease resistance. In particular, when working to develop CLB resistance, soybean breeders have not had a set of isolates representative of the pathogen(s) that is/are present in the field with which to inoculate new cultivars.

Systematic studies of *Cercospora* have long been hindered by morphological homoplasy. A number of classification schemes have been proposed based on host association (Chupp, 1954) and toxin production (Fajola 1978). More recently, a shift toward using molecular phylogenetic tools (Goodwin et al. 2001; Groenewald et al. 2013) has clarified the relationships among some taxa, but many others are still uncertain. A stable and reliable classification scheme for *Cercospora* still eludes the community.

The standard suite of phylogenetic markers currently being utilized in most systematic studies of *Cercospora* includes five nuclear genes: actin (ACT), calmodulin (CAL), translation elongation factor 1α (EF-1α), histone 3 (H3) and ITS (hereafter, collectively referred to as legacy genes). Legacy gene sequences represent most of the existing molecular data for this group and have been used extensively to classify species of *Cercospora* (Groenewald et al. 2005; Crous et al. 2006; Groenewald et al. 2010, Groenewald et al. 2013) because they amplify well across the genus and can resolve most individual species clades in concatenated analyses. However, by and large, the legacy genes are short (<500 bp), only moderately variable among closely-related taxa and individual legacy gene trees often fail to resolve some species, including *C. kikuchii, C*. cf. *flagellaris* and *C*. cf. *sigesbeckiae* (Groenewald et al. 2013; chapter two of this dissertation). Furthermore, the legacy genes do not possess enough phylogenetic signal to resolve deeper level evolutionary relationships. This is particularly evident for ITS, which is the most frequently sequenced locus for *Cercospora,* and also the least informative (Groenewald et al. 2013; chapter two of this dissertation).

The focus of this chapter was to develop new tools for phylogenetic inference in *Cercospora*. Of particular interest were the clades containing *C. kikuchii/C.* cf. *sigesbeckiae* and *C*. cf. *flagellaris*. The most robust phylogenetic analyses of *Cercospora* to date (using legacy genes) indicate that *C. kikuchii*, *C*. cf. *sigesbeckiae, C. rodmanii* and *C*. cf. *richardiicola* are closely related taxa, comprising a larger lineage that also includes several other species

(Groenewald et al. 2013; chapter two of this dissertation). However, the relationships among these species, and others within the genus, are ambiguous. Similarly, the position of *C.* cf. *flagellaris* within *Cercospora* is also uncertain. Groenewald et al. (2013) suspected that this group represents a species complex, and results from chapter two of this dissertation support their claim (Figure 2.3). Though *C.* cf. *flagellaris* has been shown to be monophyletic, Groenewald et al. (2013) found two morphologically indistinguishable lineages present within this clade which may represent additional species. However, it is not possible to assess the possibility of cryptic speciation within *C.* cf. *flagellaris* using existing data.

Many Dothideomycete genomes are now published, including species of *Mycosphaerella* and *Septoria*, but despite the agricultural importance of *Cercospora*, there are few complete genomic resources available for many species in this genus. Fortunately, the relatively low costs associated with genome sequencing coupled with an abundance of rapidly evolving bioinformatics techniques now facilitate the development of many different molecular tools, including phylogenetic markers for reconstructing evolutionary relationships at many different levels. Mycologists have seized upon these opportunities to investigate fungal diversity and evolution across different taxonomic groups (Walker et al. 2010; Kepler and Rehner 2013; Ruibal et al. 2014).

Intergenic regions have shown promise as phylogenetic markers with high nucleotide diversity in other genera, including *Colletotrichum*. Silva et al. (2012) and Rojas et al. (2010) found that the intergenic region between Apn2 and MAT1-2-1 provided superior phylogenetic resolution when compared to ITS, EF-1α and other commonly used markers for shallow level taxonomy in the *Colletotrichum gloeosporioides* complex. More recently, Kepler and Rehner (2013) used a genome-assisted approach to develop a suite of new phylogenetic markers for members of the *Metarhizium anisopliae* species complex. As for for many other ascomycetes, morphological homoplasy and a lack of informative characters among the commonly used markers have been obstacles to understanding the full breadth of phylogenetic diversity within this group. Their markers were more variable and had greater phylogenetic informativeness than commonly used protein coding genes and further highlighted the potential utility of using intergenic regions as phylogentic markers in other fungal groups.

Targeting intergenic regions within *Cerospora* was expected to improve the resources available for systematic studies within this genus, particularly for *C. kikuchii*, *C.* cf. *sigesbeckiae* and *C.* cf. *flagellaris*. Therefore, the first objective of this study was to utilize a comparative genomics approach to identify and extract exon-flanked intergenic sequences (IGS) conserved within *C.* cf. *sigesbeckiae* and *C. canescens*. It was expected that because of less selection pressure, intergenic regions would contain more polymorphisms than the legacy genes, providing better resolution at and above the species level.

Alignment filtering has become standard practice in molecular phylogenetics because poorly aligned regions containing many insertions, deletions or gaps are generally considered unreliable and can lead to erroneous tree inference (Xia et al. 2003; Talavera and Castresana 2007; Tan et al. 2015). As a result, several approaches to alignment filtering have been developed. The impact of alignment filtering on alignment length and tree topologies was considered for the second objective of this study using three parameters of varying stringency in the popular program Gblocks (Castresana 2000). The effects of alignment filtering on tree topologies generated from amino acid sequences have been investigated recently, but comparatively little has been done for nucleotide data (Dessimov et al. 2010; Nagy et al. 2012; Tan et al. 2015). Kepler and Rehner (2013) uniformly stripped their alignments of all gaps, but

recently Tan et al. (2015) argued that removing all gap sites from an alignment removes potentially informative sites. This is an area of ongoing debate. Some researchers advocate the removal of problematic regions (Swofford et al. 1996; Grundy and Naylor 1999; Castresana, 2000; Talavera and Castresana 2007) while others believe that removing these regions eliminates informative sites and can have detrimental effects on tree inference (Aageson et al. 2004; Wong et al. 2008; Dessimoz et al. 2010; Tan et al. 2015)

Legacy gene alignments of species of *Cercospora* typically do not contain regions that are replete with gap sites (personal observation), but those of intergenic regions were predicted to be more variable and contain more gaps. It was therefore expected that the most aggressive filtering parameter in Gblocks, which does not permit gaps, would yield shorter alignments and could also affect branch lengths and tree topologies compared to other treatments using less aggressive parameters.

3.2 Materials and methods

3.2.1 Taxon sampling and DNA extraction

Forty-three isolates representing 24 species of *Cercospora*, including 17 ex-type cultures, were used in this study. These isolates were chosen based on their phylogenetic positions within *Cercospora* as inferred by Groenewald et al. (2013) and chapter two of this dissertation. Exemplars were selected to encompass as much phylogenetic diversity as possible, given the current systematic knowledge of the genus. Collection information for all isolates used in this study is provided in Appendix 1. Prior to DNA isolation, all cultures were grown on potato dextrose agar (PDA) amended with chloramphenicol (1 mL L^{-1}) for 1 week at room temperature under diurnal light conditions. For all cultures obtained from the Centraalbureau voor Schimmelcultures (CBS) Fungal Biodiversity Centre, genomic DNA was isolated following a modified CTAB protocol developed in our laboratory. DNA from all other cultures was isolated following the protocol described in chapter two of this dissertation.

3.2.2 Genomic sequencing, preprocessing, assembly and annotation

Genomic DNA was extracted from hyphal tissue of *Cercospora* cf. *sigesbeckiae* PP_2012_071 following the CTAB extraction protocol above and quantified using a Qubit fluorometer (Invitrogen, Carlsbad, CA, USA). Libraries were constructed using a NEBNext Fast DNA Fragmentation & Library Prep Set for Ion Torrent (New England Biolabs, Inc., Ipswich, MA, USA) and evaluated for quality and size using the Aglient 2200 TapeStation system (Agilent Technologies, Santa Clara, CA, USA). Libraries were sequenced on an Ion Personal Genome Machine System (Thermo Fisher Scientific, Waltham, MA, USA) using an Ion 318 Chip v2. A FASTQ file generated from the sequencing run was compressed and uploaded to the Prinseq Schmeider and Edwards (2011) web server (http://edwards.sdsu.edu/cgi-bin/prinseq/prinseq.cgi). Since Ion Torrent FASTQ files conform to Sanger (Phred+33), they are logarithmically associated with error probabilities. That is, the chances of incorrectly calling bases scored Q20 and Q30 are 1 in 100 and 1 in 1000, respectively. Q20 scores are generally considered acceptably accurate. DUST and Entropy algorithms were used to assess regions of low sequence complexity (Morgulis et al., 2006). Specifically, DUST measures the occurrence of trinucleotide repeats and masks non-informative regions. Higher scores (above 7) indicate low complexity. In contrast, Entropy evaluates the entropy of trinucleotides in a sequence.

Homopolymers have an entropy value of 0 while, dinucleotide and trinucleotide repeats have values of approximately 16 and 26, respectively. Entropy scores below 70 are considered low complexity. Reads were assembled *de novo* using MIRA v4.0.2 (Chevreaux et al 1999) and GeneMark-ES v2.3e (Ter-Hovhannisyan et al. 2008) was used for *ab initio* gene prediction. The purpose of the latter was to identify conserved exons that can serve as anchors during marker development.

3.2.3 Marker development

A custom Python script was used to identify and extract intergenic (IG) regions within syntenic regions of the genomes of *C.* cf. *sigesbeckiae* PP_2012_071 and *C. canescens* MTCC-10835 Chand et al. (2014). Sixty-two syntenic gene pairs were extracted and aligned based on the following criteria: (i) flanking exons upstream and downstream of IG regions were at least 60 bp and shared at least 85 percent similarity in both genomes, (ii) five percent or fewer gaps were permitted between aligned IG regions; this required corresponding IG regions to be of similar length and ensured the selection of syntenic regions across both genomes, (iii) candidate loci were localized on the same genomic scaffold and (iv) were only found once in each genome during a BLAST search. Subsequently, pairwise distances were calculated for each pair of aligned sequences in MEGA v6 (Tamura et al. 2013), and candidates were considered for further development if they were between 500-1500 bp in length to facilitate PCR amplification and sequencing using a single primer pair. Nineteen candidate loci were retained using these filtering methods.

At least two PCR primers were designed from each sequence pair for every locus using PriFi (Fredslund et al. 2005) and PrimaClade (Gadberry et al. 2005). Each primer pair was designed so that the 5′ and 3′ regions were located within exons upstream and downstream, respectively, of intergenic regions so as to capture the entire target region. Using default parameters for PCR product analysis, FastPCR v6.5 (Kalendar et al. 2014) was used to test primer amplification and specificity within the genomes of PP_2012_071, MTCC-10835 and *C. zeae-maydis* v1.0 located in the US Department of Energy Joint Genome Institute fungal genomics portal Mycocosm (http://jgi.doe.gov/fungi) (Grigoriev et al. 2014). Only primer pairs that amplified a single product of the expected size in the genomes where *in silico* PCR was performed were selected for conventional PCR amplification.

3.2.4 PCR amplification and sequencing

PCR amplification was performed with GoTaq 2X master mix (Promega, Fitchburg, WI, USA) and the primers developed in this study, using DNA template concentrations ranging from 12.5-62.5 ng/µl. Annealing temperatures were determined empirically based on previously calculated Tm values. Gradient PCR ranging from 45–61 C was performed to establish the optimal primer annealing temperatures for each locus across a subset of *Cercospora* species. In most cases, successful markers produced single amplicons in all isolates tested, however, in some cases, PCR conditions were optimized for certain isolates.

To confirm the identity of the cultures received from CBS, additional PCRs were performed using the legacy genes commonly used in systematic studies of *Cercospora*. The most informative locus (or loci) for each taxon was previously determined from independent gene phylogenies in Groenewald et al. (2013) and chapter two of this dissertation. Portions of ACT, CAL and EF-1α were amplified with the primers ACT-512F/ACT-783, CAL-228F/CAL737R

and EF1-728F/EF1-983R (Carbone and Kohn 1999), respectively. Part of the histone 3 (H3) gene was amplified with the primers CYL H3F/CYL H3R (Crous et al. 2004) and ITS was amplified with ITS1F (Gardes and Bruns 1993) and ITS4 (White et al. 1990). Two overlapping fragments of β-tubulin were amplified with 324F/Bt1495R (Davidson et al. 2006) and Bt1F (Trkulja et al. 2013) and Bt922R (Davidson et al. 2006). Twelve of the initial 19 IGS loci amplified a single PCR product in at least one isolate of 24 species of *Cercospora* used in this study, and these were used for subsequent phylogenetic analyses. Amplicons were visualized on 1 percent agarose gels and direct sequenced with Big Dye Terminator chemistry on the Applied Biosystems 3730*xl* platform at Beckman Coulter Genomics (Danvers, Massachusetts, USA) with the same primers used for PCR. Sequences were manually edited and contigs assembled in Sequencher v5.0 (Gene Codes Corp., Ann Arbor, MI, USA).

3.2.5 Dataset assembly, alignment and model testing

Four datasets were assembled, each containing a different combination of isolates, IGS marker and legacy genes. Eleven IGS loci and six legacy genes were concatenated to form Dataset 1 (DS-1), which included all isolates used in this study (43). All taxa in DS-1 were represented by a single isolate except for *C. kikuchii*, *C. sojina*, *C. zeae-maydis* and *C. zeina* (two isolates each) and *C.* cf. *flagellaris* (16 isolates). A second dataset (DS-2), which contained all 24 species and included two isolates each of *C. kikuchii*, *C.* cf. *flagellaris*, *C. sojina*, *C. zeae-maydis* and *C. zeina* was assembled, but not used for phylogenetic analyses because it was nearly identical to DS-4. Eleven IGS loci and six legacy genes were concatenated to form DS-3, which contained the full subset of *C.* cf. *flagellaris* isolates and the type strain of *C. apii*. Alignments of DS-1, DS-2 and DS-3 were assembled as supermatrices. Sequence data for *C. zeae-maydis* CZM SCOH and *C. beticola* Cb_C1 were not obtained for ACT and CAL, respectively. Therefore, sequence data from *C. beticola* CBS 116456 (ex-type strain) and *C. zeae-maydis* CBS 117758 were substituted for the missing ACT and CAL sequences, respectively. Eleven IGS loci and six legacy genes were concatenated to form DS-4, which contained 24 isolates corresponding to one taxon each and was used for phylogenetic informativeness profiling. IGS12, which amplified in all taxa except *C. mercurialis* was excluded from this dataset because of restrictions on using supermatrices for informativeness profiling. Two additional datasets were assembled that included the five IGS loci having the highest (H5) and lowest (L5) net phylogenetic informativeness profiles as determined from PhyDesign (López and Townsend 2011; http://phydesign.townsend.yale.edu/). This was done to see if there were topological differences between these two groups of markers and how the topology from each compared to those of "all loci," "IGS" and "legacy." *Cercospora zeae-maydis* and *C. zeina* were specified as outgroup taxa in DS-1 and DS-4, based on their phylogenetic position within the current generic concept of *Cercospora* as determined by Groenewald et al. (2013). *Cercospora apii* was chosen as the outgroup for DS-3 based on results from preliminary phylogenetic analyses of individual IGS alignments performed during this study, which indicated that it was sister or closely related to *C.* cf. *flagellaris*.

Alignments of individual IGS loci, legacy genes and each of the four concatenated datasets were constructed in MEGA v6 (Tamura et al. 2013) using the Muscle (Edgar 2004) algorithm. In order to investigate the effect of curating alignments with Gblocks v0.91b on tree topology, support values and phylogenetic informativeness, three different treatments were applied to alignments of each locus and also the four concatenated datasets. Gblocks calculates

sitewise summary statistics for the alignment and classifies individual sites as highly conserved, conserved and non-conserved using two metrics: absence/presence of gaps and identical sites. Using these criteria, nonconserved regions longer than a specified length are removed. Additional sites are removed from the resulting "blocks" until they are flanked by "anchored" sites that can be aligned with certainty. The stringent and more stringent filtering parameter setting are similar in that no gaps are permitted, a minimum length of 10 residues are required per block, and a minimum of 13 and 20 sequences, respectively, are required for conserved and flanking positions. The first was manually trimmed at the 5′ and 3′ ends of the alignment and not further curated using Gblocks (NoGb). The second and third alignments were left untrimmed and curated in Gblocks using less stringent (LSGb- gaps permitted) and more stringent (MSGb- no gaps permitted) parameters, respectively. Model testing was performed on each individual alignment to infer the most appropriate model of nucleotide substitution. jModelTest v2.1.6 (Darriba et al. 2012; Guindon and Gascuel, 2003), implementing a corrected Aikake information criterion on XSEDE through the CIPRES portal (Miller, et al. 2010) was used to determine the best fit model for maximum likelihood phylogenetic analysis from a set of 88 candidate models. MrModeltest v2.3 (Nylander 2004), implementing the Aikake information criterion, was used to determine the best fit model for Bayesian inference from a set of 24 candidate models. Information pertaining to alignment lengths and evolutionary models chosen for each locus is given in Appendix 6.

3.2.6 Profiling phylogenetic informativeness and analyses of substitution saturation

The quantitative metric developed by Townsend (2007) and implemented in PhyDesign was used to rank all IGS loci and legacy genes according to their phylogenetic informativeness profiles (PIPs). The best maximum likelihood trees from each treatment of "all loci" in DS-4 were first converted to ultrametric trees in R using the chronos function in the APE package (Paradis et al. 2004), specifying a value of one for the root node age. All terminals are equidistant from the root in an ultrametric tree, which allows a quantitative prediction of the relative time point when each locus has the most phylogenetic informativeness. PhyDesign was then used to profile informativeness for each locus and for each site, using HypHy (Pond et al. 2005) to estimate site rates. It was expected that calculating phylogenetic informativeness values (PIVs) for each locus, then ranking them according to their scores would provide a quantitative metric to assess phylogenetic utility and facilitate comparisons among IGS loci and legacy genes. The metric implemented in PhyDesign identifies which loci are most appropriate to use at different epochs, or historical times. This is useful for determining which loci to choose when working with taxa that have only recently diverged as opposed to early-diverging basal lineages. Both net (n) and per-site (ps) PIVs were calculated for all loci in PhyDesign. The signal from the net informativeness method was of greater interest because it correlates directly with node support values, but it may be affected by noise more than the per-site method, which reduces noise by consolidating signal density (http://phydesign.townsend.yale.edu/faq.html#persite). The per-site method also maximizes informativeness and cost-effectiveness (López and Townsend 2011). However, the latter consideration was not of concern here because all IGS loci were amplified with a single primer pair, thus eliminating any issue of sequencing costs.

Because high levels of substitution saturation can reduce phylogenetic signal and confound phylogenetic analyses (Xia et al. 2003), Dambe v5 (Xia 2013) was used to quantify levels of substitution saturation for each locus in DS-4. Given the strong selection pressure

against deleterious substitutions in protein coding genes, it was not expected that saturation would be found in the legacy genes. On the other hand, non-coding and potentially highly divergent IG regions, are more likely to have experienced saturation because of their ability to accumulate more mutations without severe consequences (ITS is also not translated, but is conserved in *Cercospora*, suggesting that sequences from this locus should not be saturated). The proportion of invariant sites (P_{inv}) was estimated for each treatment of the alignments in DS-4 with a Goodness of fit test assuming a Poisson+I distribution and using a neighbor-joining algorithm specifying a K80 model that distinguishes between transitions and transversions and assumes equal base frequencies. P_{inv} was also estimated under other models (JC69, F84 and GTR) using several tree-building algorithms (FastME and UPGMA), but these did not affect values. Substitution saturation was measured for each locus using the method of Xia et al. (2003) and Xia and Lemey (2009) which determines levels of saturation and phylogenetic utility for a set of sequences by calculating their index of substitution saturation (I_{ss}) and comparing that value to the critical index of substitution saturation ($I_{ss.c}$), the value at which substitution saturation occurs. If the value of I_{ss} is significantly smaller than $I_{ss.c}$, little saturation has occurred and the sequences are expected to be useful for phylogenetic analysis. However, if I_{ss} is not significantly smaller than $I_{ss.c}$, or if it is greater, the sequences are considered to be poor for phylogenetic analyses.

3.2.7 Phylogenetic inference and tree distance calculations

Bayesian inference was performed for all individual gene alignments and on concatenated, partitioned datasets using MrBayes v3.2.6 (Ronquist et al. 2012) with the appropriate evolutionary model added separately to each partition, using the resources at the Louisiana State University High Performance Computing Center (http://www.hpc.lsu.edu) and within the CIPRES portal. For analyses of independent genes, four replicates of 10,000,000 generations each were run with four Metropolis-coupled chains, three heated to a temperature of 0.2. Trees were sampled every 5,000th generation with the first 25 percent discarded as burnin. Fifty percent majority rule consensus trees were generated with sumt. For all concatenated analyses, conditions were the same except that the number of generations was increased to 20,000,000 and trees were sampled every 4,000 generations. Tracer v1.6.0 (Rambaut et al. 2014) was used to assess convergence of the estimated parameters. Maximum likelihood was performed using Garli v2.01 (Zwickl 2006) for all individual gene and concatenated datasets. Maximum likelihood trees were inferred by stepwise addition with 50 search replicates. Bootstrap proportions were estimated from a minimum of 1000 pseudoreplicate datasets, with the highest likelihood tree from two replicate searches per pseudoreplicate dataset retained. Bootstrap proportions were calculated and mapped onto the maximum-likelihood phylogenetic trees using SumTrees in the Dendropy v3.12.0 phylogenetic computing library (Sukumaran and Holder 2010). Maximum likelihood analysis of individual alignments and concatenated datasets was performed using Garli v2.01 (Zwickl 2006) with the evolutionary model added separately to each partition. Graphical representations of phylograms were exported from FigTree v1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/) and edited in Inkscape v0.48.1 (Harrington, 2004-2005).Summary statistics for each individual gene alignment were generated using PAUP* v4.0b10 (Swofford 2002).

To compare the effect of treatments NoGb, LSGb and MSGb on individual gene tree topologies in each dataset, TreeCmp v1.0-b291 (Bogdanowicz et al. 2012) was used to measure

the distances between phylogenetic trees. Several methods for comparing rooted trees are available in TreeCmp, including Robinson-Foulds (RF) and Matching Split (MS) metrics. MS distances represent the fewest number of clade rearrangements necessary for two topologies to be identical. It was of interest to determine if there were topological differences between individual gene trees from different treatments (i.e. NoGb vs. LSGb, NoGb vs. MSGb and LSGb vs. MSGb) and also to see how the mean tree distances differed among the IGS loci. A high nPIV combined with a small mean matching splits distance would be a good indicator that a particular locus should be useful as a phylogenetic marker regardless of the Gblocks alignment filtering algorithm applied. To quantify these differences, the MS metric was used because of its greater discriminatory power and because it is less susceptible than RF to perturbations arising from fluctuations at the tips of the tree (Bogdanowicz et al. 2012). All analyses were run in matrix comparison mode (-m) so that for each locus in a respective dataset, each tree corresponding to one of the three treatments was compared to the other two trees. The MS (Bogdanowicz and Giaro 2012) distance metric for unrooted trees (ms) was specified for all analyses.

3.3 Results

3.3.1 Genome sequencing

Detailed information of genome statistics is provided in Appendix 4 (A4). A total of 1.63 Gb and 6.05 million sequences was identified with a mean sequence length of 270 bp (Appendix 4 Table A4.1). Base quality scores at the 5′ end of the read began at approximately Q30 and decreased somewhat (to about Q25) at the 3′ end. A decrease in base quality is expected with most sequencing platforms; however, these statistics indicate that there is not a marked drop-off in quality at the ends of shorter reads, but mean base quality scores fell to Q24 or lower at the 3’ ends of longer reads. Mean GC content was calculated to be 51.5 percent ± 6 percent (Appendix 4 Table A4.2), which is within the range of several other sequenced Dothideomycete genomes (Ohm et al., 2011).

Base ambiguities were queried; however, no sequences with N were reported. A total of 202,414 sequences (3.4 percent of all sequences) were identified as either exact duplicates (ED) or exact duplicates with reverse complements (EDRC). ED were far more abundant than EDRC, comprising 98.6 percent of all duplicates and 3.35 percent of all sequences. (Appendix 4 Table A4.3). As expected, this distribution was similar in appearance to that of the plot of mean sequence length (data not shown). Less than 10 ED corresponded to approximately 168,084 sequences and the highest number of ED for a single sequence was 490 (Appendix 4 Table A4.3). The mean complexity DUST scores for all sequences were below 5, and all ENTROPY values were greater than 70, indicating that low complexity regions were not expected to be problematic for downstream applications.

5.77 million total reads were assembled. The average total coverage was 43.06X, calculated from contigs of at least 5000 bp and 13X coverage. Length, coverage and quality assessment for large contigs (≥ 500 bp; ≥22X coverage) is reported in Appendix 4 Table A4.4. Four hundred sixty-nine large contigs with a consensus length of 34.13 Mbp were assembled. The largest contig was 1.59 Mbp. N50 contig size was 418.5 kb. Maximum total coverage of the assembly was 2042X. The large contigs had an average consensus quality (ACQ) score of 74, and quality assessment scores indicated no strong or weak unresolved repeat positions (SRMc, WRMc), unsolved sequence type mismatches (STMU) or problems with read quality. One

potential concern is that 923 consensus IUPAC bases were identified. This means that for those bases, multiple groups were found to tie and the consensus was assigned an IUPAC base (barring the presence of a gap). This is an acknowledged area of weakness within the consensus algorithm used by MIRA (http://nar.oxfordjournals.org/content/suppl/2012/10/11/gks908.dc1/nar-01874-met-h-2012-file011.pdf). First of all, the significance of gaps may be overestimated because they cannot be represented by IUPAC. More importantly, however, using strict quality cut off (Q35) instead of a sliding window scale may result in erroneous base calls. For example, between groups with qualities 38 and 34 (quality distance 4), quality 38 wins and there is no IUPAC assignment. However, there is no distinction between groups with quality 36 and 40, which both exceed the cut off value of 35. These are treated as a tie by MIRA, which assigns an IUPAC base despite the two groups also having quality distance 4.

Length, coverage and quality assessments were re-iterated for all contigs (includes contigs not initially designated as large) (Appendix 4 Table A4.5). An assembly containing 1,984 contigs with a total consensus length of 34.95 Mbp was compiled. The largest contig was 1.59 Mbp. N50 contig size was 406.3 kb. Maximum total coverage of the assembly was 2042X. As before, there were no problems identified with SRMc, WRMc, STMU or read quality. Average consensus quality was 53 (lower than for large contigs), and nearly five times as many consensus IUPAC bases were identified (4545). This last number is not surprising since more contigs and reads were assembled and ACQ was lower, which should result in more ambiguous bases.

3.3.2 Marker development and phylogenetic informativeness profiling

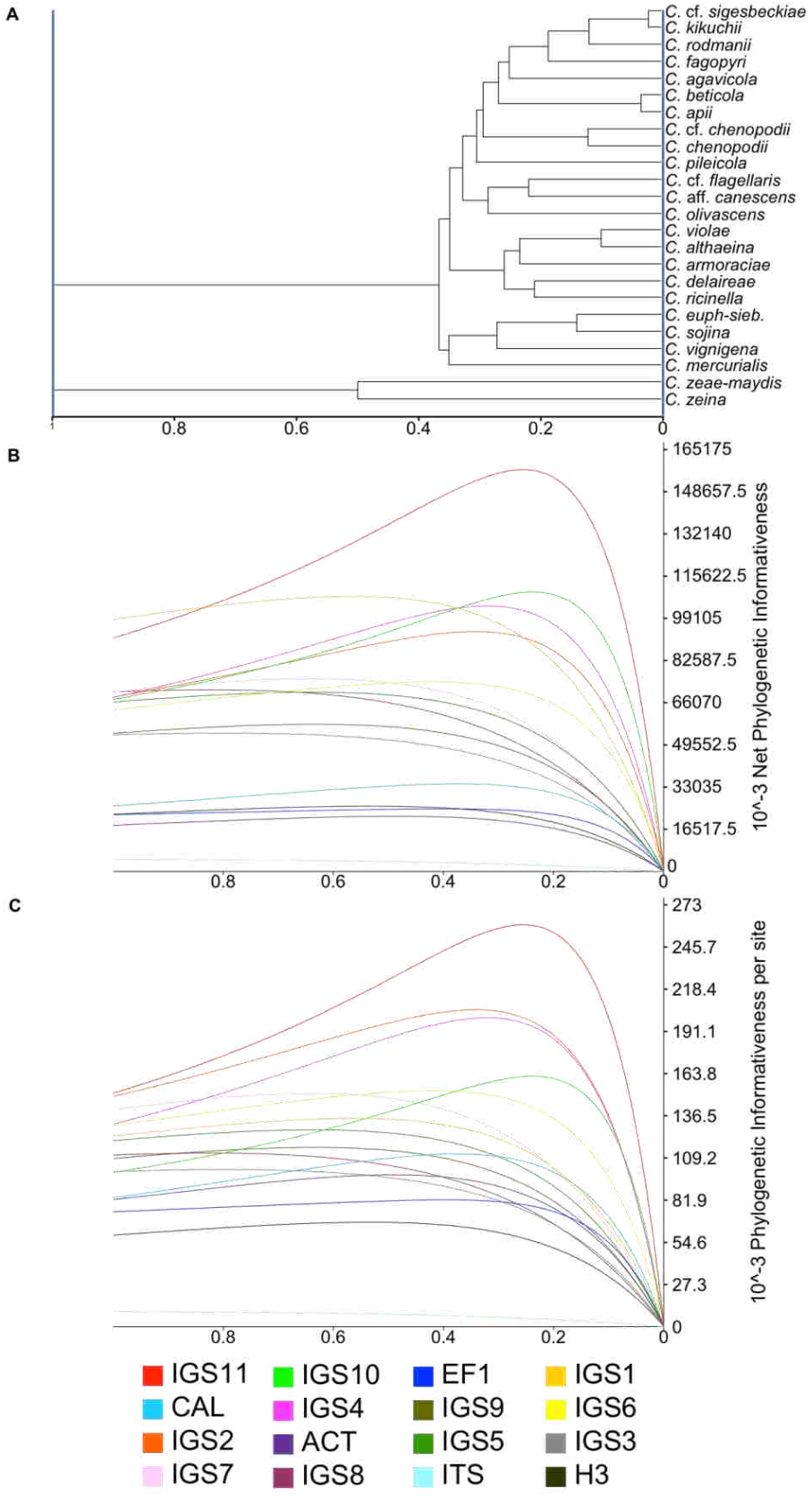
Twelve IGS loci were developed from the 63 syntenic gene pairs extracted from the genomes of *C*. cf. *sigesbeckiae* PP_2012_071 and *C*. *canescens* MTCC-10835 (Table 3.1). A total of 543 sequences were generated during this study- 488 from IGS loci and 55 from the six legacy genes routinely used for systematic studies of *Cercospora* (ACT, β-tubulin, CAL, EF1, H3 and ITS). A summary of the net and per site phylogenetic informativeness profiles (PIP) of 15 loci (eleven IGS and five legacy genes) in treatments NoGb, LSGb and MSGb for DS-4 follows.

All of the IGS loci had higher net phylogenetic informativeness values (nPIV) compared to the legacy genes in treatment NoGb (Figure 3.1; Table 3.2). Locus ranking based on nPIVs did not change much between treatments NoGb and LSGb, but differences were observed for MSGb. In treatment NoGb, IGS11 had the highest nPIV followed by IGS4, IGS1 and IGS10, whose scores were all nearly identical. IGS3 was the lowest scoring IGS locus, but still had a nPIV of nearly twice that of CAL, the highest scoring legacy gene. ITS had by far the lowest nPIV of any locus, followed by ACT and EF1, whose nPIVs were more than four times higher than ITS. IGS11 also had the highest overall per site PIV (psPIV), followed by IGS2 and IGS4 (identical scores) and IGS10. IGS3, IGS9 and IGS8 were the lowest scoring IGS loci, only better than H3, EF1 and ITS, which had the lowest overall scores. CAL, the highest scoring legacy gene, scored higher than the three lowest IGS loci and was identical to IGS5 and ACT.
The nPIVs from treatment LSGb were similar to the previous treatment. All of the IGS loci scored higher than the legacy genes. IGS11, IGS10, IGS1, IGS4 and IGS2 had the highest overall nPIVs, respectively, and IGS3 was the lowest scoring IGS locus (Table 3.2). CAL and ITS had the highest and lowest nPIVs, respectively. All but one IGS locus had higher psPIVs than the legacy genes.IGS11 had the highest overall score, followed by IGS2, IGS10 and IGS7.

Table 3.1 Primer information for IGS markers developed in chapter three

Primer	Flanking exons	Sequence	% GC	T_M	Length
IGS1_Fwd	1454-1455	ATACTAGATACCATTTTGGCATGACCCAGG	43.3	59.7	30
IGS1_Rev	1454-1455	GTTTCCCGTGTGCTGAGTCCG	61.9	60.7	21
IGS2_Fwd	1720-1721	CCACGCTACCCGAAGTCGTTCTACTTCATCCG	56.2	65.3	32
IGS2_Rev	1720-1721	AGGCGAGGATGAARATGGACAGGAAMAGGCA	51.6	65.7	31
IGS3_Fwd	2433-2434	CGAGCATTTGCGGGYGYCTTTGGCAGCGGC	66.7	71.5	30
IGS3_Rev	2433-2434	GCTGGCGAGGCTGGCYAATTCGAGCGAATCTCGCG	64.3	71.2	35
IGS4_Fwd	2529-2530	TGGCGACCTTCATRGTCGASTTG	54.3	60.1	23
IGS4_Rev	2529-2530	GCTGGCGAGGCTGGCYAATTCGAGCGAATCTCGCG	35.7	60.0	35
IGS5_Fwd	3360-3361	GCACGCGCGGCGCATGGCCCCAGAGC	80.8	75.5	26
IGS5_Rev	3360-3361	CTGCCGAGCGCCGCGGCGACCCACAGGAAG	76.7	74.8	30
IGS6_Fwd	6098-6099	TCTGGAGCCAACTCTTGAGAGGCGCCATGATGG	57.6	68.1	33
IGS6_Rev	6098-6099	TCGGCGCGRTTGAACTGTGTGACGGGC	64.8	69.1	27
IGS7_Fwd	7086-7087	GATTGATGGTTCAGGTAAGCGTTTTGGCGTGGAG	50.0	64.8	34
IGS7_Rev	7086-7087	CTTTGATGCGAGCCTCGACATCTTTGAGG	51.7	62.5	29
IGS8_Fwd	7329-7330	GAGTGCTCATGCGCCGCTGACATTGATAGGAACG	55.9	67.1	34
IGS8_Rev	7329-7330	CGCCGTGGRCGCAACTTGCGCCGGTTGTGTCTCTG	67.1	73.0	35
IGS9_Fwd	10515-10516	CAGAGCCAAAAGATGCCATT	45.0	53.6	20
IGS9_Rev	10515-10516	GCACAACGGAGATGGTGTC	57.9	56.3	19
IGS10_Fwd	10665-10666	AATTGGTGCCGGAAGAATC	47.4	53.2	19
IGS10_Rev	10665-10666	GATGCSACSACATCTTTGC	52.6	54.5	19
IGS11_Fwd	12030-12031	ACAGGAAGATGGTCGGATAGG	52.4	55.9	21
IGS11_Rev	12030-12031	TGGCGGGTCTATCGACAT	55.6	55.8	18
IGS12_Fwd	3640-3641	CTCTGACTTGTCGTCAATGATCTC	45.8	55.1	24
IGS12_Rev	3640-3641	CATCCCATCGCAGTTGYTC	55.3	55.2	19

Figure 3.1 Graphical representation of phylogenetic informativeness values for the concatenated DS-4 treatment LSGb of “all loci.” (A) Ultrametric maximum likelihood tree. (B) Net phylogenetic informativeness values. (C) Per site phylogenetic informativeness values. Values on x-axes indicate relative time from the root to the tip of the tree.


A
C. cf. sigesbeckiae
C. kikuchii
C. rodmanii
C. fagopyri
C. agavicola
C. beticola
C. apii
C. cf. chenopodii
C. chenopodii
C. pileicola
C. cf. flagellaris
C. aff. canescens
C. olivascens
C. violae
C. althaeina
C. armoraciae
C. delaireae
C. ricinella
C. euph-sieb.
C. sojina
C. vignigena
C. mercurialis
C. zeae-maydis
C. zeina
1
0.8
0.6
0.4
0.2
0
B
165175
148657.5
132140
115622.5
99105
82587.5
66070
49552.5
33035
16517.5
0
10^-3 Net Phylogenetic Informativeness
0.8
0.6
0.4
0.2
0
C
273
245.7
218.4
191.1
163.8
136.5
109.2
81.9
54.6
27.3
0
10^-3 Phylogenetic Informativeness per site
0.8
0.6
0.4
0.2
0
IGS11
IGS10
EF1
IGS1
CAL
IGS4
IGS9
IGS6
IGS2
ACT
IGS5
IGS3
IGS7
IGS8
ITS
H3

The lowest scoring IGS locus was IGS3, which had a slightly smaller psPIV than CAL, the highest scoring legacy gene. ITS, H3 and EF1 were the lowest scoring loci, respectively.

All but two IGS loci had higher nPIV scores than the legacy genes in treatment MSGb. In contrast to treatments NoGb and LSGb, where the most informative locus had a nPIV greater than 150 and 180, respectively, the highest value in treatment MSGb was less than 80 (Table 3.2). IGS1 had the highest overall nPIV, followed by IGS11, IGS10, IGS2 and IGS7. IGS4 had the lowest overall score, better only than ITS. CAL had the highest nPIV of the legacy genes and was slightly higher than IGS6. IGS11 had the best overall psPIV, followed by IGS2, IGS4, IGS7 and IGS10. CAL had the highest score of the legacy genes and was better than IGS5, IGS8 and IGS3, the lowest scoring IGS locus. ITS H3 and ACT had the lowest overall psPIVs.

In treatments NoGb and LSGb, the four loci with the highest nPIVs and psPIVs (IGS10, IGS11 IGS2 and IGS4) were also those that reached their max PIVs closest to the epoch of interest. This was determined to be from approximately 0.4 time units to the tips of the tree and represents the point that includes all species-level bipartitions except for *C. zeina* and *C. zeae-maydis*. IGS10 and IGS11 were also the top two loci in treatment MSGb, reaching their max PIVs before 0.3 time units. CAL and EF1were the only two legacy genes which reached their max PIVs before 0.4 time units in any treatment (Table 3.2).

Table 3.2 Max net and per site informativeness values for IGS loci and legacy genes

Treatment	Locus	Max val reached at	Max net val	Max per site val	Length bp
LSGb	IGS11	0.25	157.31	0.26	605
LSGb	IGS10	0.24	109.37	0.16	675
LSGb	IGS1	0.56	107.72	0.13	799
LSGb	IGS4	0.32	103.95	0.20	520
LSGb	IGS2	0.34	93.94	0.21	458
LSGb	IGS7	0.63	75.47	0.15	500
LSGb	IGS6	0.42	74.23	0.15	486
LSGb	IGS8	0.81	71.14	0.11	633
LSGb	IGS5	0.63	70.21	0.13	551
LSGb	IGS9	0.63	57.66	0.12	497
LSGb	IGS3	0.79	54.19	0.10	533
LSGb	CAL	0.37	34.31	0.11	307
LSGb	H3	0.53	25.57	0.07	378
LSGb	EF1	0.39	24.54	0.08	299
LSGb	ACT	0.49	21.63	0.10	220
LSGb	ITS	0.99	4.69	0.01	483
NoGb	IGS11	0.2	187.24	0.28	735
NoGb	IGS4	0.26	124.47	0.24	536
NoGb	IGS1	0.49	123.27	0.14	966
NoGb	IGS10	0.21	122.63	0.18	742
NoGb	IGS2	0.26	114.54	0.24	506
NoGb	IGS6	0.34	87.89	0.17	1325
NoGb	IGS7	0.57	82.67	0.16	532

Table 3.2 (continued)

Treatment	Locus	Max val reached at	Max net val	Max per site val	Length bp
NoGb	IGS8	0.67	79.23	0.12	689
NoGb	IGS5	0.56	76.07	0.13	654
NoGb	IGS9	0.54	63.19	0.11	590
NoGb	IGS3	0.66	61.06	0.11	605
NoGb	CAL	0.34	36.51	0.13	275
NoGb	H3	0.48	27.31	0.07	381
NoGb	EF1	0.35	26.22	0.07	403
NoGb	ACT	0.43	24.44	0.13	196
NoGb	ITS	0.98	5.06	0.01	474
MSGb	IGS1	0.52	79.20	0.13	633
MSGb	IGS11	0.2	70.54	0.21	343
MSGb	IGS10	0.26	65.39	0.15	437
MSGb	IGS2	0.4	64.05	0.17	386
MSGb	IGS7	0.61	56.89	0.15	386
MSGb	IGS8	0.84	53.41	0.11	471
MSGb	IGS5	0.71	48.52	0.11	423
MSGb	IGS9	0.63	43.75	0.13	338
MSGb	IGS3	0.83	42.52	0.10	438
MSGb	CAL	0.39	34.31	0.12	275
MSGb	IGS6	0.5	34.09	0.14	236
MSGb	H3	0.5	23.91	0.06	369
MSGb	EF1	0.51	20.88	0.10	218
MSGb	ACT	0.55	16.69	0.09	183
MSGb	IGS4	0.99	12.18	0.15	84
MSGb	ITS	0.99	4.88	0.01	474

NoGb- no Gblocks treatment; alignment manually trimmed at the 5′ and 3′ ends
LSGb- less stringent Gblocks treatment (relaxed)
MSGb- more stringent Gblocks treatment

Alignments generated by treatment NoGb were consistently the longest, followed by LSGb and MSGb, respectively. However, there were two instances where remarkable differences in alignment length among treatments were observed. IGS4 was of comparable length in treatments NoGb and LSGb (536 and 520 bp, respectively), but was much shorter in treatment MSGb (84 bp). This variability in length among different treatments occurred because of two short and highly divergent sequences of *C. zeae-maydis* and *C. zeina* that were only conserved with the rest of the alignment in the flanking exon regions where primers were designed. These sequences introduced many gap sites in the alignment that were permitted by treatments NoGb and LSGb, but were all removed by treatment MSGb, which by default strips alignments of all gap sites. This resulted not only in a reduction of alignment length, but also substantially influenced its nPIVs across treatments. IGS4 had the second and fourth highest values in treatments NoGb and LSGb, but dropped to second worst in treatment MSGb.

The length of the IGS6 alignment also varied substantially among treatments because of a unique 754 bp insertion present in *C. agavicola* that was not conserved in any other sequences. This region was only retained in treatment NoGb. When the insertion was removed in treatments LSGb and MSGb, the alignments were reduced in length and a corresponding drop in nPIVs was observed among treatments, though not as steep as for IGS4. IGS6 went from sixth and seventh highest in treatments NoGb and LSGb, respectively, to eleventh best in treatment MSGb.

3.3.3 Substitution saturation

Analysis of nucleotide substitution patterns and tests of substitution saturation performed in DAMBE (Xia et al. 2003; Xia and Lemey 2009) for each of the three alignment filtering treatments in DS-4 for "all loci" revealed that none of the loci in treatments LSGb and MSGb were saturated (Appendix 5 Tables A5.2 and A5.3). In treatment NoGb, no saturation was detected for the legacy genes, but all IGS loci except IGS6 were saturated (Appendix 5 Table A5.1). For loci that were not saturated, I_{ss} values were always significantly lower than $I_{ss.c}$ for the symmetrical and asymmetrical tree and the two values never overlapped with the 95 percent confidence interval.When pooled, transitions (s) always outnumbered transversions (v), with the largest s/v ratio occurring for ITS and the lowest for IGS10 (except in treatment NoGb, where ACT was only slightly greater than IGS10).

3.3.4 Phylogenetic analyses

The datasets using gene combinations "all loci" and "IGS" produced trees with more bifurcations and stronger support values along the backbone of the tree than "legacy" (Figures 3.2–3.10). Topologies were more similar between treatments NoGb and LSGb than they were between either treatment and MSGb. Because all but one of the IGS loci were found to be saturated in treatment NoGb, and because treatment MSGb removed more than half of all sites in several cases, we chose to present the trees from LSGb. However, clade support values and the taxa present within each clade for trees inferred from all three treatments in the concatenated analyses of DS-1, DS-3 and DS-4 are shown in in tabular format in Appendix 9 and 10. Individual "IGS" and "legacy" gene trees from treatment LSGb of DS-1 are provided in Appendix 7. Ten clades were observed among the three treatments of the concatenated DS-1 and DS-4 alignments using the three gene combinations (i.e. "all loci," "IGS" and "legacy"): (1) *C.* cf. *flagellaris* (including three subclades within); (2) *C.* aff. *canescens* and *C. olivascens*; (3) *C. apii* and *C. beticola*; (4) *C. kikuchii*, *C.* cf. *sigesbeckiae*, *C. rodmanii*, and *C. fagopyri*; (5) *C. chenopodii* and *C.* cf. *chenopodii*; (6) *C. ricinella*, *C. delaireae*, *C. violae*, *C. althaeina* and *C. armoraciae*; (7) *C. sojina*, *C. euphorbiae-sieboldianae* and *C. vignigena*; (8) *C. zeae-maydis* and *C. zeina*; (9) *C. pileicola* and *C. olivascens;* (10) *C.* aff. *canescens*, *C.* cf. *flagellaris* and *C. olivascens*.

Clade 1 (*C.* cf. *flagellaris*) was strongly supported in all DS-1 and DS-3 topologies, regardless of the combination of loci or treatment. Within this clade, the presence or absence of three subclades varied among treatments. Isolate CBS 132674 usually formed a separate lineage,

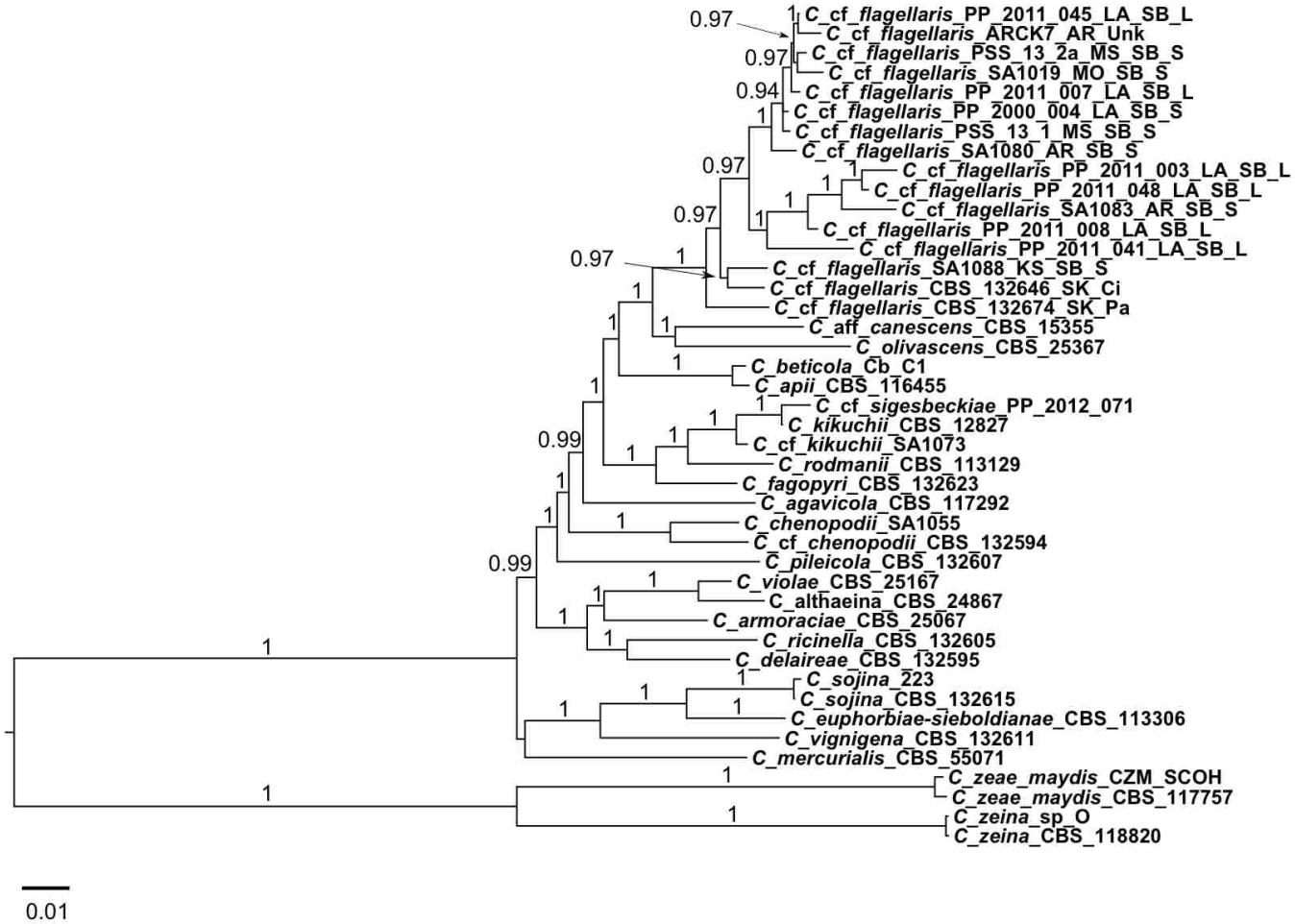


Figure 3.2 Topology generated from Bayesian inference phylogenetic analysis of 24 species of *Cercospora* from the LSGb treatment of the concatenated DS-1 alignment of "all loci" (IGS2–IGS12 and ACT, β-tubulin, CAL, EF-1α, H3 and ITS). Tree is rooted with *C. zeina* and *C. zeae-maydis*. Posterior probability values greater than 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site. Terminals are labeled according to species with corresponding strain identifier, location (AR= Arkansas, USA; LA=Louisiana, USA; MO=Missouri, USA; MS=Mississippi, USA; KS=Kansas, USA; SK=South Korea), host (Ci=*Cichorium intybus*; Pa-*Phytolacca americana* SB=soybean) and substrate (L=leaf; S=seed).

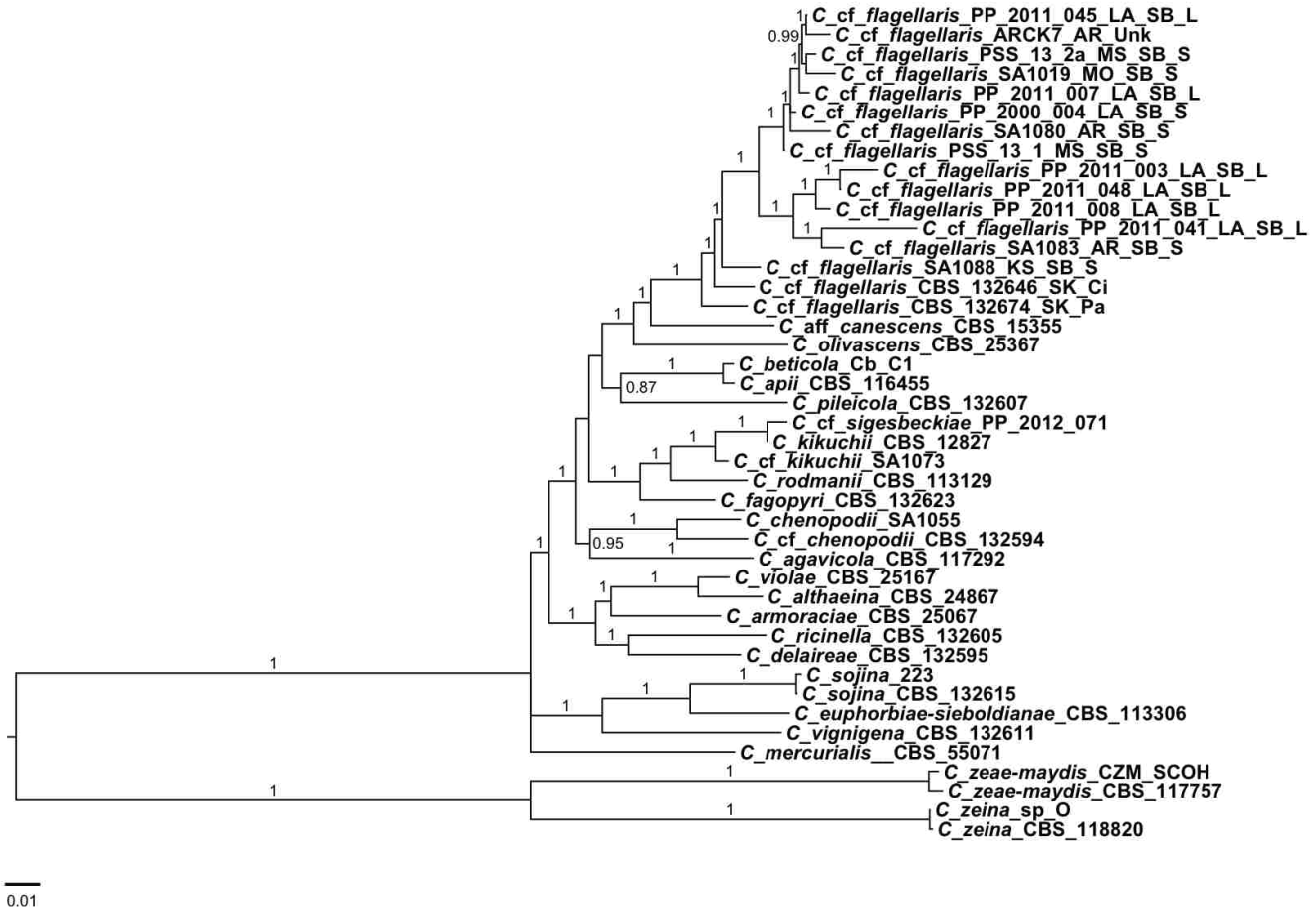


Figure 3.3 Topology generated from Bayesian inference phylogenetic analysis of 24 species of *Cercospora* from the LSGb treatment of the concatenated DS-1 alignment of "IGS" (IGS2–IGS12). Tree is rooted with *C. zeina* and *C. zeae-maydis*. Posterior probability values greater than 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site. Terminals are labeled according to species with corresponding strain identifier, location (AR= Arkansas, USA; LA=Louisiana, USA; MO=Missouri, USA; MS=Mississippi, USA; KS=Kansas, USA; SK=South Korea), host (Ci=*Cichorium intybus*; Pa-*Phytolacca americana* SB=soybean) and substrate (L=leaf; S=seed).

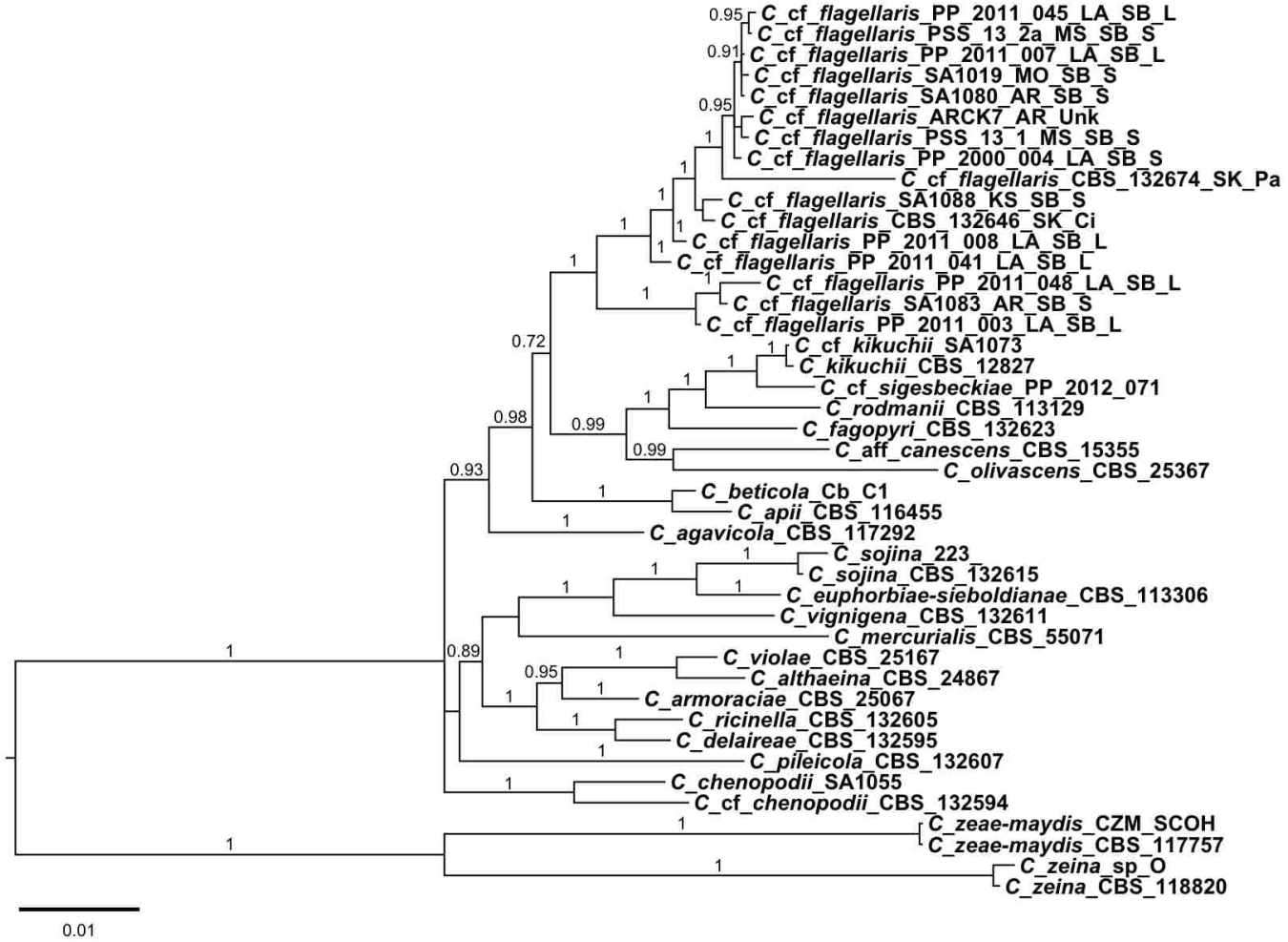


Figure 3.4 Topology generated from Bayesian inference phylogenetic analysis of 24 species of *Cercospora* from the LSGb treatment of the concatenated DS-1 alignment of "legacy" (ACT, β-tubulin, CAL, EF-1α, H3 and ITS). Tree is rooted with *C. zeina* and *C. zeae-maydis*. Posterior probability values greater than 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site. Terminals are labeled according to species with corresponding strain identifier, location (AR= Arkansas, USA; LA=Louisiana, USA; MO=Missouri, USA; MS=Mississippi, USA; KS=Kansas, USA; SK=South Korea), host (Ci=*Cichorium intybus*; Pa-*Phytolacca americana* SB=soybean) and substrate (L=leaf; S=seed).

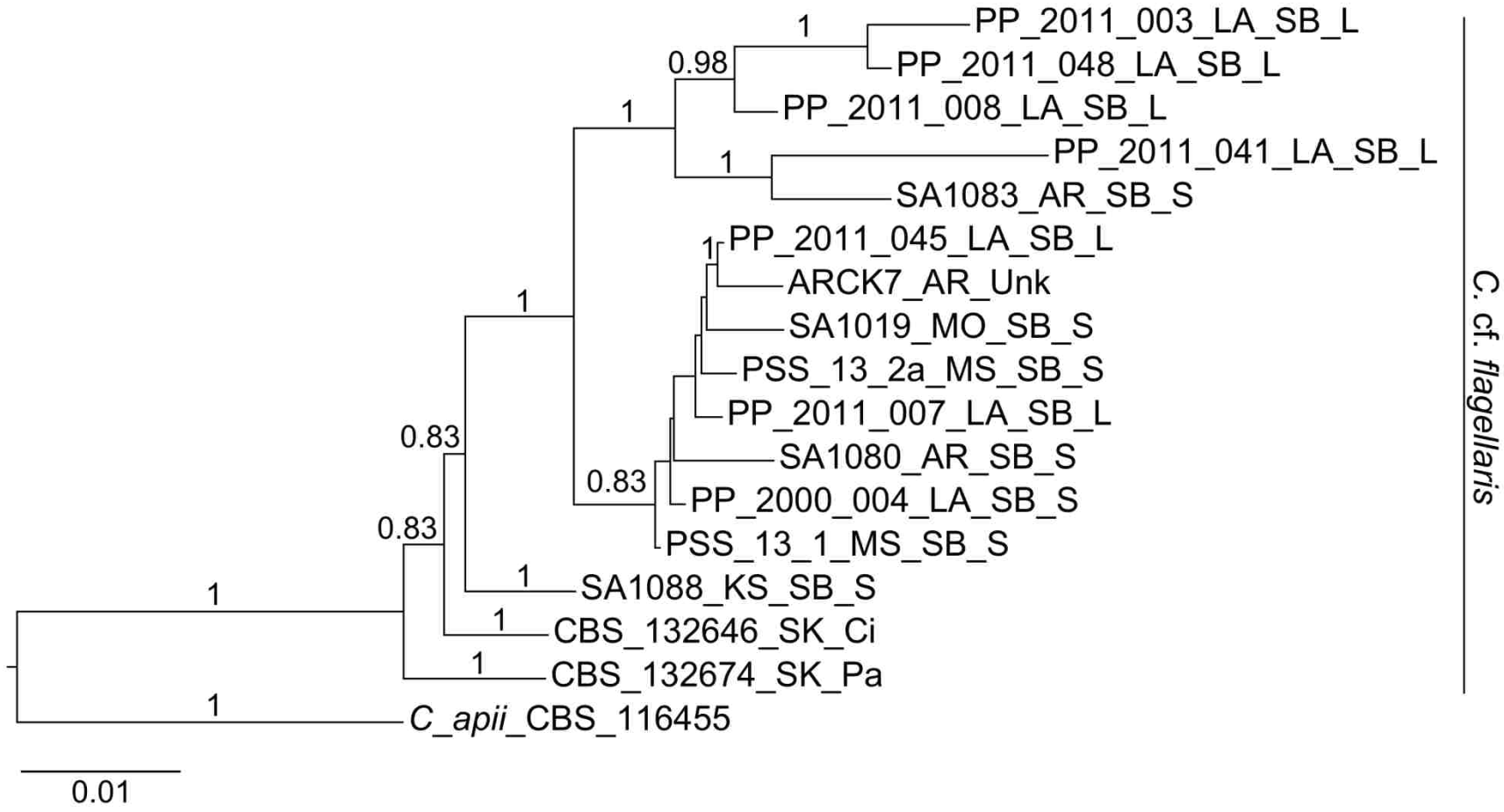


Figure 3.5 Topology generated from Bayesian inference phylogenetic analysis of 16 isolates of *C.* cf. *flagellaris* from the LSGb treatment of the concatenated DS-3 alignment of "all loci" (IGS2–IGS12 and ACT, β-tubulin, CAL, EF-1α, H3 and ITS). Tree is rooted with *C. apii*. Posterior probability values greater than 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site. Terminals are labeled according to species with corresponding strain identifier, location (AR= Arkansas, USA; LA=Louisiana, USA; MO=Missouri, USA; MS=Mississippi, USA; KS=Kansas, USA; SK=South Korea), host (Ci=*Cichorium intybus*; Pa-*Phytolacca americana* SB=soybean) and substrate (L=leaf; S=seed).

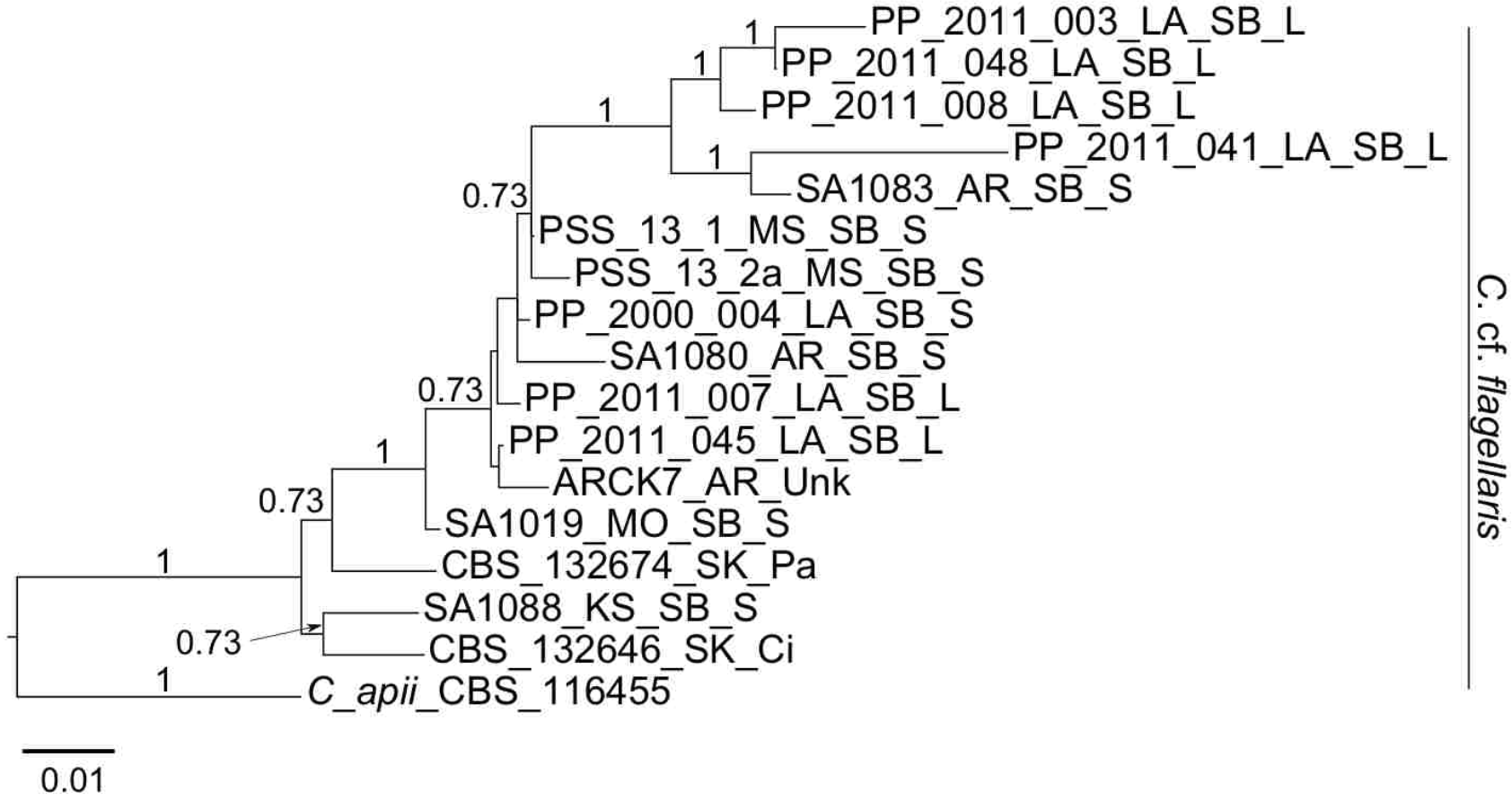


Figure 3.6 Topology generated from Bayesian inference phylogenetic analysis of 16 isolates of *C.* cf. *flagellaris* from the LSGb treatment of the concatenated DS-3 alignment of "IGS" (IGS2–IGS12). Tree is rooted with *C. apii.* Posterior probability values greater than 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site. Terminals are labeled according to species with corresponding strain identifier, location (AR= Arkansas, USA; LA=Louisiana, USA; MO=Missouri, USA; MS=Mississippi, USA; KS=Kansas, USA; SK=South Korea), host (Ci=*Cichorium intybus*; Pa-*Phytolacca americana* SB=soybean) and substrate (L=leaf; S=seed).

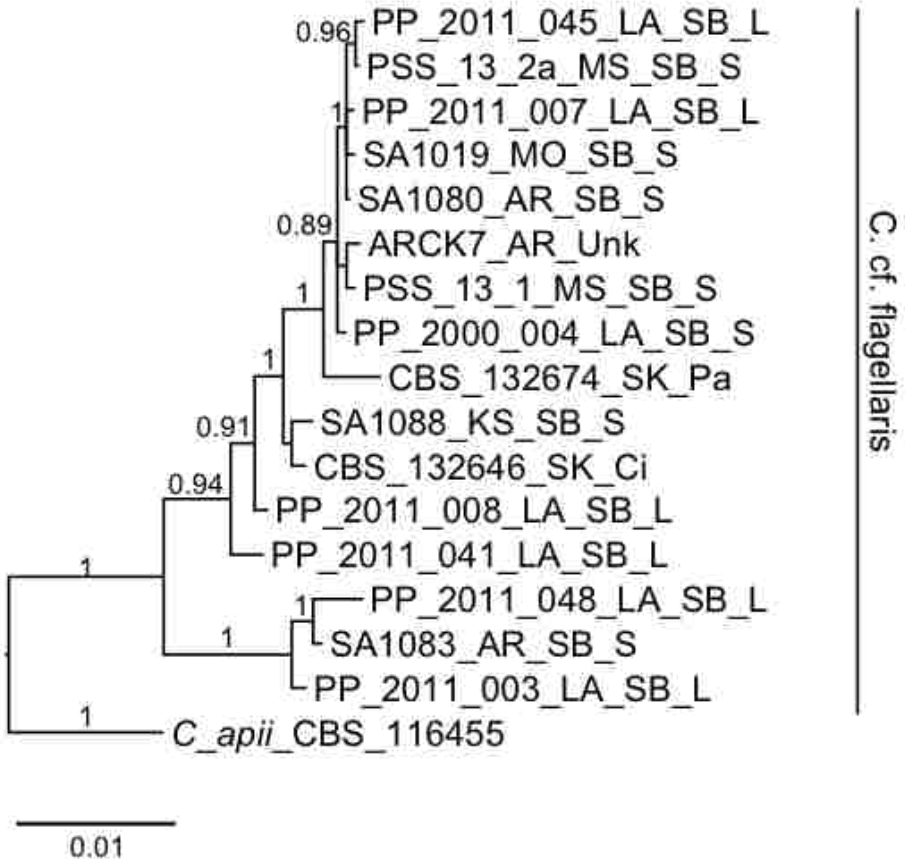


Figure 3.7. Topology generated from Bayesian inference phylogenetic analysis of 16 isolates of *C.* cf. *flagellaris* from the LSGb treatment of the concatenated DS-3 alignment of "legacy" (ACT, β-tubulin, CAL, EF-1α, H3 and ITS). Tree is rooted with *C. apii.* Posterior probability values greater than 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site. Terminals are labeled according to species with corresponding strain identifier, location (AR= Arkansas, USA; LA=Louisiana, USA; MO=Missouri, USA; MS=Mississippi, USA; KS=Kansas, USA; SK=South Korea), host (Ci=*Cichorium intybus*; Pa-*Phytolacca americana* SB=soybean) and substrate (L=leaf; S=seed).

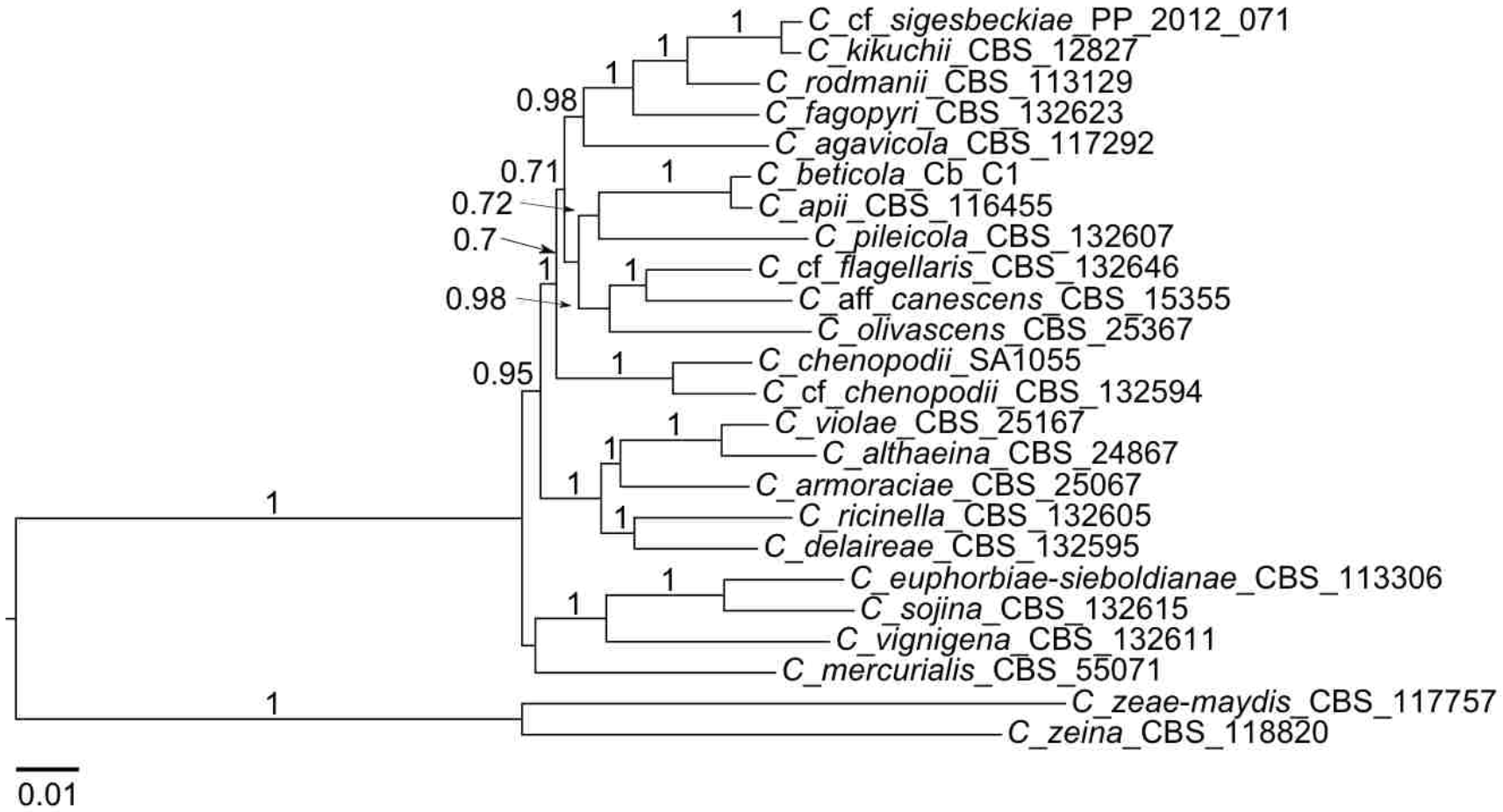


Figure 3.8. Topology generated from Bayesian inference phylogenetic analysis of 24 species of *Cercospora* from the LSGb treatment of the concatenated DS-4 alignment of "all loci" (IGS2–IGS12 and ACT, β-tubulin, CAL, EF-1α, H3 and ITS). Tree is rooted with *C. zeina* and *C. zeae-maydis*. Posterior probability values greater than 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site.

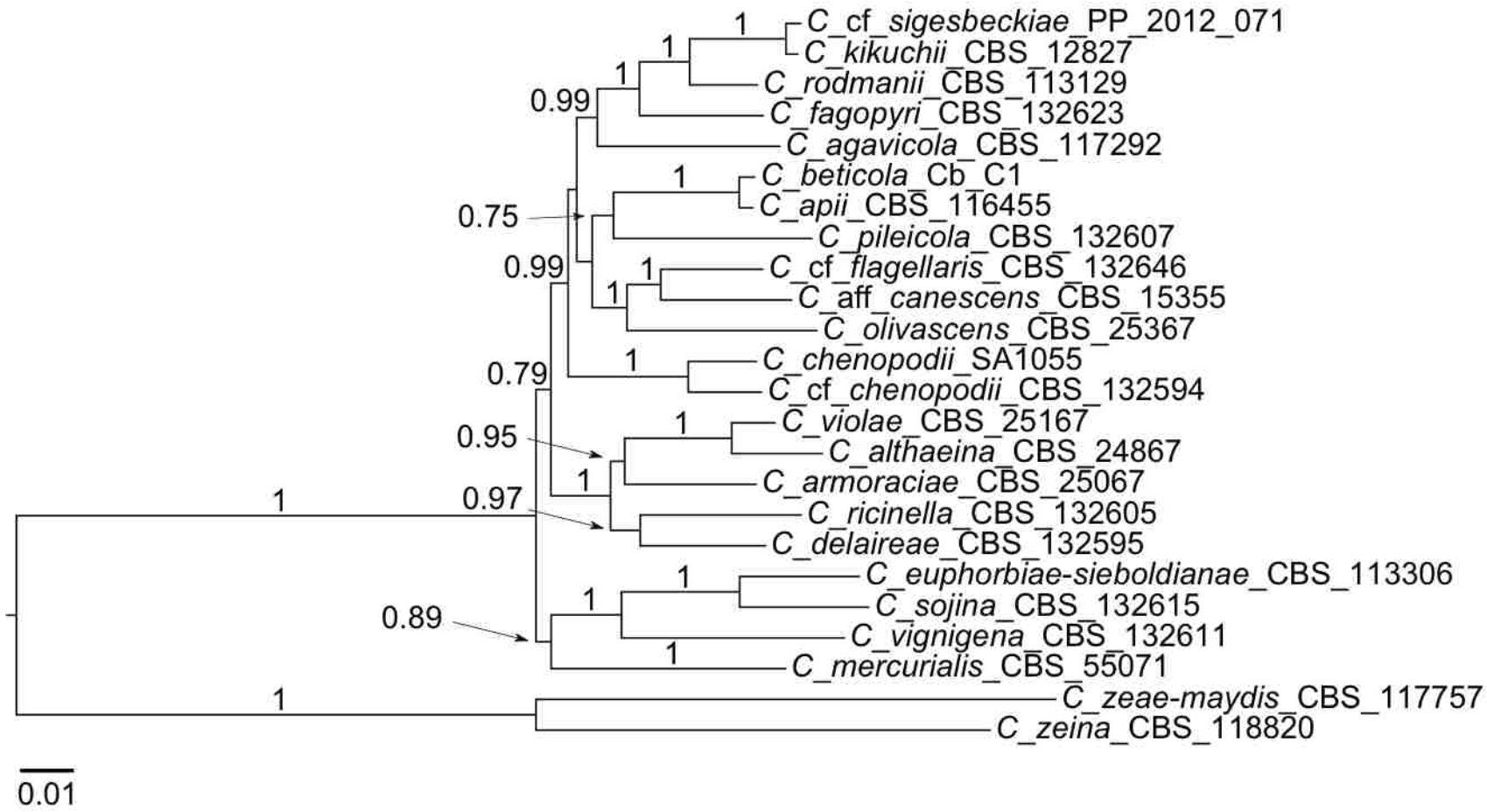


Figure 3.9. Topology generated from Bayesian inference phylogenetic analysis of 24 species of *Cercospora* from the LSGb treatment of the concatenated DS-4 alignment of "IGS" (IGS2–IGS12). Tree is rooted with *C. zeina* and *C. zeae-maydis*. Posterior probability values greater than 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site.

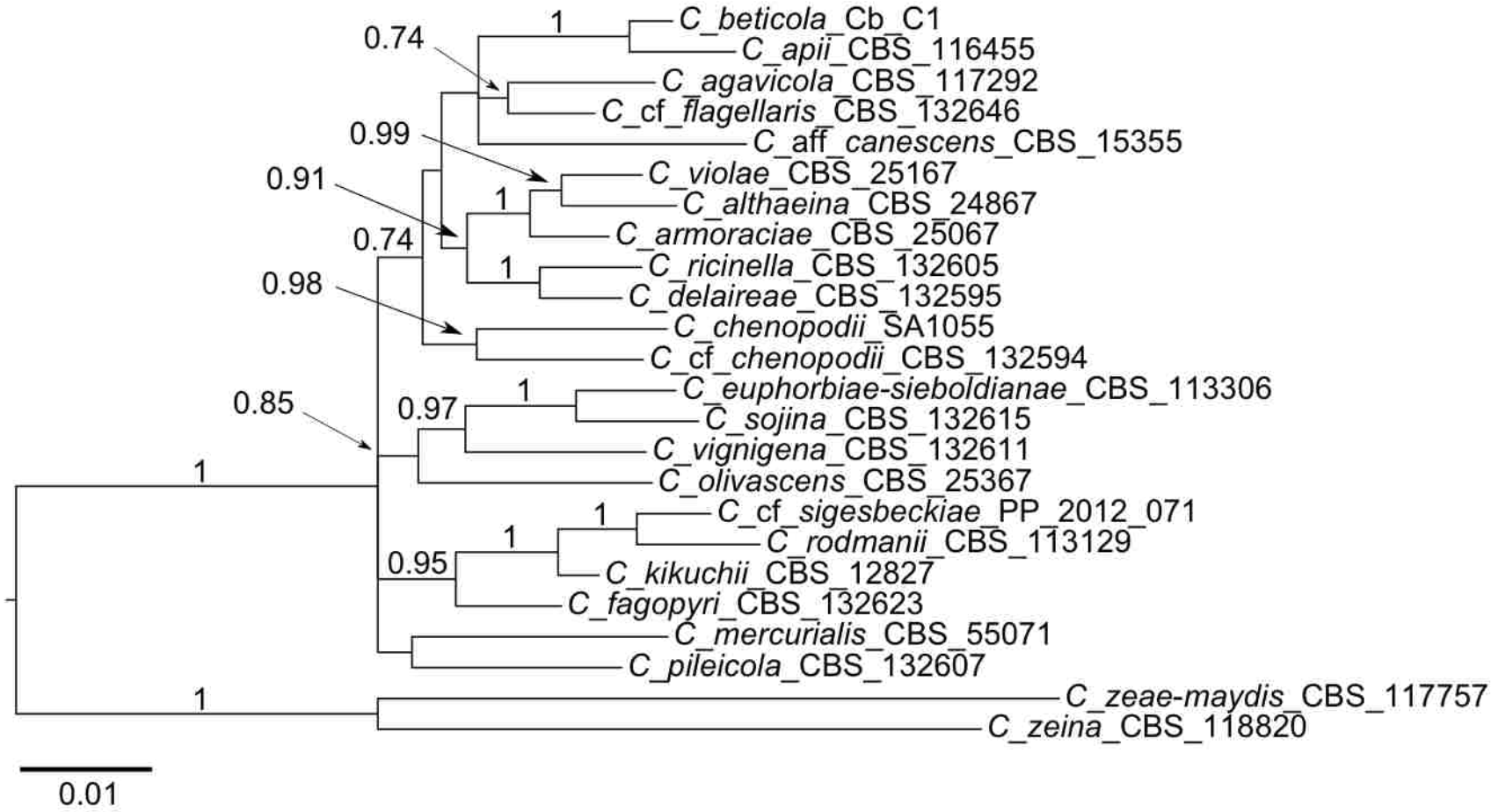


Figure 3.10. Topology generated from Bayesian inference phylogenetic analysis of 24 species of *Cercospora* from the LSGb treatment of the concatenated DS-4 alignment of "all loci" (IGS2–IGS12 and ACT, β-tubulin, CAL, EF-1α, H3 and ITS). Tree is rooted with *C. zeina* and *C. zeae-maydis*. Posterior probability values greater than 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site.

except in DS-1 and DS-3, where it was part of subclade 2 (SC2) in all treatments of "legacy" (Figures 3.4 and 3.7). SC1 consisted of four soybean leaf isolates collected from Louisiana and a single soybean seed isolate from Missouri. This group was present in all topologies except in DS-1 treatment MSGb of "all loci" and each DS-1 and DS-3 treatment of "legacy." SC2 included eight isolates collected from soybean leaves and seeds from Arkansas, Louisiana, Mississippi and Missouri. This group was present in all DS-1 treatments of "legacy," but was missing from treatments MSGb of "all loci" and "IGS." It was also present in all DS-3 treatments of "legacy," but missing from each treatment of "IGS" and NoGb of "all loci." When not monophyletic, the isolates from SC2 clustered together for treatment NoGb in DS-3 of "all loci" and all three treatments of DS-3 "IGS" (Figure 3.6; data not shown for treatments NoGb and MSGb). SC3 included two isolates, one from soybean seed from Kansas and another collected in S. Korea from chicory (*Cichorium intybus*). SC3 was present in each DS-1 treatment of "legacy" (Figure 3.4), but missing from treatment MSGb of "all loci" (Appendix 9 Table A9.1) and in each treatment of "IGS" (Appendix 9 Table A9.2). In DS-3, SC3 was present in each treatment of "IGS" (Appendix 9 Table A9.5), but was missing from treatment LSGb of "all loci" (Figure 3.5) and from treatment MSGb of "all loci" and "legacy" (Appendix 9 Tables A9.4 and A9.6).

When present in DS-1, Clade 2 was always sister to Clade 1. Clade 2 was missing from "all loci" and "legacy" in treatment MSGb and in all treatments of "IGS." However, when they did not form a clade, *C.* aff. *canescens* and *C. olivascens* were still sister to Clade 1 except in treatment MSGB for "all loci" where *C.* aff. *canescens* was sister to Clade 4 and *C. olivascens* was sister to *C. pileicola* (Clade 9). Clades 3–8 were strongly supported in all analyses regardless of the combination of loci or treatment. The position of certain individual species varied, often with strong support in different topologies. For example, *C. agavicola* was supported as sister to Clade 5 (Figure 3.3), but also as a separate lineage (Figures 3.2 and 3.4). *Cercospora mercurialis* and *C. pileicola* also usually formed unsupported separate lineages.

Relationships within and among clades in DS-4 were similar to those in DS-1, though the placement of certain species was variable among topologies. For "all loci" and "IGS," *C.* cf. *flagellaris* was most often part of Clade 10 (Figures 3.8 and 3.9; Appendix 10 Tables A10.7 and A10.8) was also was sister to *C. agavicola* in all treatments of "legacy" (Figure 3.10), but this relationship was not well supported.

When contrasted against the legacy genes in DS-4, the topologies of datasets H5 and L5 each produced trees with stronger support at common nodes and more closely resembled the topologies of the concatenated analyses of "all loci" and "IGS,"with the expection of Clades 2 and 4 (Figures 3.8–3.10; Appendix 7 Figures A7.18 and A7.19). Clade 4 was present in L5, but this group was not monophyletic in H5, where *C. rodmanii* and *C. fagopyri* were sister to Clade 3. Clade 10 was present in H5, but not in L5, where *C.* cf *flagellaris* was sister to Clade 3.

3.3.5 Pairwise comparision of individual gene trees based on matching split distances

MS distances were uniformly lower between trees whose alignments were filtered with treatments NoGb and LSGb than with MSGb across all datasets (Appendix 8 Tables A8.1–A8.3). More topological variation occurred between treatments NoGb–MSGb and LSGB –MSGb than between NoGb–LSGb (Appendix 8 Tables A8.1–A8.3). Out of 50 tree comparisons across all datasets, a zero MS distance occurred 31 times between NoGb–LSGb, 15 times between NoGb–MSGb and 14 times between LSGB –MSGb. The zero MS distances between NoGb–LSGb occurred across all three datasets, but were mostly localized within DS-3, (*C.* cf. *flagellaris*

subset with *C. apii* outgroup) where there were no or few topological differences among trees regardless of treatment for most loci, which was indicated by MS distances that were not greater than nine (Appendix 8 Tables A8.1–A8.3). It was also observed that the five loci having the highest nPIVs (IGS1, IGS2, IGS4, IGS10, IGS11), had the highest mean matching split distances. In constrast, the five IGS loci with the lowest nPIVs (IGS7, IGS8, IGS9, IGS5 and IGS 3) had the smallest mean matching splits.

3.4 Discussion

The picture that emerges from the first rounds of phylogenetic analyses applying the twelve IGS loci developed here to a systematic study of 24 species of *Cercospora* is that these new loci consistently perform better than each of the legacy genes. Of particular interest are the multiple lineages found within a subset of *C.* cf. *flagellaris* isolates originating from several states in the USA and from Asia. This group has long been treated as a complex, and using the markers developed here, it may finally be possible to investigate more closely cryptic speciation within this cosmopolitan group.

An important aspect that was considered when selecting candidate loci was that, while the intergenic sites should be highly variable among different taxa, their priming sites should be conserved in all *Cercospora* species. One reason that that the legacy genes are so widely adopted is that they are conserved throughout *Cercospora* and amplify readily in most species. The IGS markers developed here amplified well across all 24 species of *Cercospora* tested, including in early and late diverging lineages such as *C. zeae-maydis* and *C. kikuchii*, respectively. The phylogenetic informativeness profiles of the IGS loci were consistently better than the legacy genes across all three treatments. Based on nPIVs, from treatment LSGb, IGS11, IGS10, IGS1, IGS4 and IGS2 were the most informative loci, respectively (Figure 3.1; Table 3.2). Most of these loci were comparable in rank across treatments, but IGS4 went from one of the best loci in treatments NoGb and LSGb to the second worst in MSGb. This unexpected drop in ranking was first investigated as possible contamination. IGS4 amplicons were typically greater than 500 bp, but bands for *C. zeina* and *C. zeae-maydis* were observed at 200–300 bp. This indicated potential non-specific amplification or DNA contamination, although *C. zeina* and *C. zeae-maydis* were consistently the most difficult taxa to amplify and sequence across all loci, often requiring several rounds of PCR optimization.

The IGS4 sequences were not thought to be contaminants or artifacts, but accurate reflections of the large evolutionary distance between these two species and the rest of *Cercospora*. However, to rule out contamination or cross-amplification, PCRs were repeated with 2 isolates per taxon. Each isolate was previously amplified with other IGS loci, producing reliable sequence data. Two identical sequences of *C. zeae-maydis* CBS 117757 and CZM_SCOH from different PCR runs were obtained. Each was also nearly identical to *C. zeina* CBS 118820, with the exception of two ambiguous bases. Despite being largely dissimilar to the rest of the alignment, these sequences aligned to short regions at the 5′ and 3′ ends of the IGS4 alignment, which was expected and indicated that they were not contaminants. Given that IGS primers in this study were designed within conserved exons flanking IG regions, amplification of even highly divergent IG regions should theoretically be possible as long as their flanking exons were conserved.

The highly divergent sequences of *C. zeina* and *C. zeae-maydis* are probably a reflection of the evolutionary distance between these taxa and the rest of *Cercospora*. As evidence of how

far these two species have diverged from others in the genus, they are the only two taxa consistently monophyletic in independent legacy gene trees, including ITS, which reaches its maximum phylogenetic informativeness value earlier than any other locus and is only able to discriminate early diverging lineages within *Cercospora*. Because *C. zeina* and *C. zeae-maydis* represent one of the most basal lineages in *Cercospora* (Groenewald et al. 2013), successful amplification of any of the IGS loci developed here should serve as a good indication that amplification will be possible across the genus, including in other early diverging species such as *C. senecionis-walkeri* and *C. conigrammes*, though this will require testing additional isolates of *C. zeina* and *C. zeae-maydis*, especially before IGS4 can be validated as a suitable marker for phylogenetic studies across the genus.

The topological variation among trees generated from different alignment filtering treatments was not entirely surprising given how much potentially valuable sequence information was lost from the more stringent treatment. There has been recent interest in evaluating the impact of alignment filtering using Gblocks and other programs on phylogenetic accuracy and a central theme pervades many of these studies: removing all gaps and non-conserved regions also removes potentially informative sites. Though questions remain, some recent studies have concluded that, while some filtering is acceptable and may not influence tree inference, aggressive filtering is detrimental to downstream phylogenetic analyses.

Tan et al. (2015) found that there was a correlation between the percentage of alignment sites removed and the tree reconstruction error rate. They also found that removing most gap sites in an alignment was only slightly less detrimental (based on a percentage of wrong splits), or in some cases, was actually worse than randomly removing columns from an unfiltered alignment. Similar conclusions were reached by Dessimov and Gil (2010), who found that gaps possess significant phylogenetic signal and that curated alignments excluding gaps and other variable regions had a negative effect on tree accuracy. These examples do not recommend abandoning alignment filtering completely, however. Depending on the size of the dataset, a moderate amount of filtering (up to 20 percent) can improve tree accuracy and reduce computational time (Tan et al. 2015). Given this criterion, trees inferred from IGS alignments treated with NoGb or LSGb would not be expected to differ much considering that the percent reduction in alignment length between these treatments was always less than 20 percent for each locus, except for IGS6.

The results from the three filtering treatments tested in this study led us to conclude that LSGb was the most appropriate treatment given our data. Tree topologies and MS distances between NoGb-LSGb were comparable, suggesting that there was not a major difference between these two treatments. However, we found that all but one of the IGS loci were saturated when using treatment NoGb (Appendix 5 Table A5.1. Saturated sequences should be avoided when inferring phylogenies because they have experienced so many mutations over time that the sequences no longer reflect evolutionary relationships (Xia et al. 2003). It remains to be determined whether saturation is present globally within the IGS loci or whether it is localized in specific regions of the alignment. On the other hand, none of the the loci from treatment MSGb were saturated, but this treatment was not optimal because it was too aggressive in removing sites in the alignment. Talavera and Castresana (2007) showed that the effects of aggressive pruning by MSGb can be mitigated in longer genes, but for shorter genes, the loss of phylogenetic information is greater than the gain in signal-to-noise ratio. This is likely what occurred here and what is responsible for the topological differences and also the high MS distances between the less stringent treatments and MSgb. The IGS loci developed in this study

cannot compensate for excessive pruning because they are relatively short (less than 1000 bp), divergent among distantly related taxa and often contain many gaps and indels that were removed in treatment MSGb.

In addition to Gblocks, there are many other programs implementing different filtering algorithms that may be useful to test. When encountering problematic regions of the alignment, Gblocks removes entire columns even if only one or several residues are erroneous. However, other programs such as Guidance (Penn et al. 2010) provide the option to mask specific residues within a problematic part of an alignment instead of removing entire columns. This may help to preserve sites with valuable phylogenetic signal that my otherwise be removed.

It was suggested that nucleotide data benefit more from filtering because they are more poorly aligned by existing software (Tan et al. 2015), but this has not been studied extensively. Nagy et al. (2012) looked at three strategies for dealing with gaps in ITS alignments, including using Gblocks, and found that ITS has greater phylogenetic utility at deeper evolutionary levels then was previously thought. The inclusion of indel sites increased support values and resolution from species to phylum. This illustrates the powerful phylogenetic signal that indels and other often-removed characters possess. It seemed probable that filtering would also have a strong effect on rapidly evolving IGS loci. Given that IGS regions are not subject to the same selection pressure as coding genes, it was expected that alignments for datasets containing different species would have more gaps, indels or non-conserved characters than those including isolates from a single species and would offer an opportunity to assess the effects of applying three different alignment filtering algorithms on tree topology, support values and tree distances. Because treatment LSGb allowed gaps and did not aggressively prune alignments, it was expected that there would be less difference between it and NoGb than between either of them and MSGb, which removes all gaps and permits only four contiguous non-conserved positions.

Treatments NoGb and LSGb did not markedly affect ranking of individual loci based on nPIV, psPIV or the time point at which each locus reached its maximum informativeness value, though nPIV values were uniformly highest in treatment NoGb and lowest in MSGb (Table 3.2). The decrease in max net informativeness values across treatments likely occurred because of a loss of sequence information, particularly in the form of gaps, indels and non-conserved sites that were removed according to the different heuristic approaches implemented in each Gblocks filtering algorhithm. IGS alignment lengths never differed by more than 17 percent between noGb and LSGb (Appendix 6 Tables A6.1–A6.3). However, differences in alignment length between either of these treatments and MSGb were higher. For IGS4 and IGS6, more than 80 percent of the original alignment was removed by treatment MSGb compared to noGb, which accounted for the drop in PIV. Differences in the alignment lengths of the legacy genes between treatments were typically less than for the IGS loci, but that was expected because they contain fewer gaps and are more conserved among this group. As a consequence, alignments of related taxa within *Cercospora* should not be replete with many gaps or indels. This was indeed the case as there was 11 percent or less difference between treatments noGb and LSGb for all loci, except Btub and EF-1α. The nearly 50 percent difference in alignment length between treatments NoGb and MSGb in Btub and EF-1α was the highest among the legacy genes and occurred mostly because the alignments included several incomplete sequences that introduced external gaps that were entirely removed by treatment MSGb.

The effects of the different filtering treatments were more obvious in the actual alignment lengths and the phylogenetic informativeness profiles than in the tree topologies and support values. Though the topologies generated by treatments NoGb and LSGb were consistently more

similar, when considering the overarching clade distribution, all three treatments generally recovered similar phylogenies. This may be a consequence of the dataset used. Given the resources available at the beginning of this study, a group of species best representing the taxonomic diversity across *Cercospora* was chosen based on the phylogenies of Groenewald et al. (2013) and those generated in chapter two of this dissertation. However, both of these studies relied on legacy genes to infer relationships among taxa. Therefore, sister level and higher relationships were difficult to gauge. The effect of this ambiguity is evident when looking at the different clades throughout DS-1 and DS-4 (Figures. 3.2–3.4 and 3.8–3.10; Appendix 7 Figs. 3.10–3.12). The primary clades (those represented by *C. kikuchii, C. apii, C. chenopodii, C. ricinella, C. sojina* and *C. zeae-maydis* are stable, regardless of treatment or gene combination and are strongly supported with relatively little conflict. However, some "rogue taxa" like *C. agavicola, C. pileicola, C. mercurialis* and *C.* cf. *flagellaris* are not consistently part of any one clade. This is likely influenced more by the loci used in each analysis than by the individual treatments. While gene combinations "all loci," "IGS" and "legacy" generated well supported primary clades that were stable, it is clear that only the gene combinations including the IGS loci provide a new perspective of certain sister-taxon relationships that cannot be achieved using the legacy genes alone.

For example, the position of *C.* cf. *flagellaris* is unresolved with respect to its relationship to other *Cercospora* species in the phylogenies of Groenewald et al. (2013) and those generated in chapter two of this dissertation. In the former study, it is monophyletic and, according to the authors, probably constitutes a species complex. However, based on the tree, little can be inferred about this species and its phylogenetic position within *Cercospora*. Using "all loci" and "IGS," *C.* cf. *flagellaris* shares common ancestry with *C.* aff. *canescens, C. olivascens, C. apii* and *C. beticola* (Figures 3.2 and 3.3), relationships that are not evident in the corresponding "legacy" tree (Figure 3.4). Furthermore, all three *C.* cf. *flagellaris* subclades were only present in DS-1 treatments NoGb (Appendix 9 Table A9.1) and LSGb of "all loci" (Figure 3.2). The presence of these distinct subclades is interesting and suggests that there is widespread diversity within this pathogen, which at this time cannot be definitively interpreted as representing several divergent lineages of a broad generalist or one or more different species. In DS-1 treatments NoGb and LSGb of "all loci," SC1 and SC2 are composed exclusively of soybean leaf and seed isolates collected from around the Gulf South and Midwest. The two isolates in SC3 include one from soybean seed in Kansas and another from *C. intybus* in S. Korea. It is also interesting to note that isolate CBS 132674, which is sister to all other subclades, was also collected from S. Korea, though from a different host, *P. americana*. Though it is too early to speculate on what is driving genetic diversity (or potentially speciation) in this group, this brings up the question of whether host, geography, both or other factors are involved.

The results from these analyses provide a first glimpse of what a more stable taxonomic landscape of *Cercospora* looks like. The markers developed in this study show that it is possible to improve upon existing genes using a comparative genomic approach targeting conserved syntenic gene pairs. Now, additional taxon sampling is needed to fill in the gaps. Although using the IGS loci alone works well, the best resolution (most notably for identifying phylogenetic diversity in *C.* cf *flagellaris*) was achieved when combining IGS loci and legacy genes together. This is not practical for future taxonomic studies. Therefore, additional work needs to be done to identify combinations of IGS/legacy genes that can most closely reproduce the results seen when combining them all.

It is important to keep in mind that the true *Cercospora* tree is not known and the reference against which all IGS topologies in this study are judged is that of the legacy genes. With that in mind, topological comparision of trees generated from the H5 versus L5 alignments highlight some potential discrepancies between informativeness values and actual topologies. H5 failed to find Clade 4 (*C. kikuchii, C.* cf. *sigesbeckiae, C. rodmanii* and *C. fagopyri*), which was consistently recovered in "all loci," "IGS" and "legacy," and was found in the L5 analysis. It is unclear if this is an anomaly or if there is a conflict between the informativeness profiles generated in PhyDesign and actual phylogenetic utility in a real biological system. However, inspection of the individual gene trees revealed that Clade 4 was monophyletic in three of the five independent gene trees in L5 (IGS7, IGS8 and IGS9), but was only monophyletic in IGS4 in H5 (Appendix 7 Figures A7.18 and A7.19). The timepoint at which the IGS loci reach their max PIVs may be a factor in determining their phylogenetic utility in certain groups. For example, all of the loci in H5 reach their max PIV before the loci in L5. This may make them more useful to resolve taxa that have for very recent divergences. Even though they didn't didn't resolve the entire Clade 4, *C. kikuchii* and *C.* cf. *sigesbeckiae* were consistently sister to one another in the individual gene trees. Based on the ultrametric tree in Figure 3.1, *C. kikuchii* and *C.* cf. *sigesbeckiae* are the most recently diverged of any of the 24 species in this study. Because IGS10 and IGS11 reach their max informativeness values closest to the tips of the tree, they would be expected to work well for these species, but may not be able to pick up earlier divergences, such as the splits for *C. rodmanii* and *C. fagopyri*.

More work also needs to be done to investigate the effect of saturation on topologies. The alignments used for H5 and L5 were both treated using LSGb, and though no detectable levels of saturation were found in treatment LSGb for any IGS loci, it may be worthwhile to compare the ratios of Iss : Iss.c more closely to get a better idea of the relative rate among all the IGS loci and see if this correlates with the informativeness profiles.

Future work in this area will impact not only taxonomy, but also practical approaches to soybean disease control. There are many soybean cultivars, but little in the way of host resistance to CLB at the moment. Though at times mercurial, this disease is probably not going away; more likely, with the expansion of soybean acreage worldwide it will continue to be a threat. One of the problems facing breeders is a lack of fungal isolates to test. The results from the gene combinations "all loci" and "IGS" of DS-1 treatments NoGb (Appendix 9 Table A9.1) and LSGb (Figures 3.2 and 3.3; Appendix 9 Tables A9.1) indicate there are at least three lineages of *C.* cf. *flagellaris* which occur in the USA. Given that *C.* cf. *flagellaris* is such a broad generalist and that genetic diversity among a relatively small subset of isolates is high, it seems likely that more lineages may exist. Definitive information about the frequency of these lineages is still missing and more isolates need to be collected and characterized to determine whether there is a link between individual lineages, their hosts and their geographic distribution worldwide.

CHAPTER 4. CONCLUSIONS AND PROSPECTS FOR FUTURE RESEARCH

The primary finding from this work is that *Cercospora kikuchii*, the soybean pathogen that has, for nearly a century, been thought to be the sole cause of Cercospora leaf blight (CLB) and purple seed stain (PSS) worldwide, is not present in Louisiana, and is likely not dominant in the United States (USA). Its presence in the USA, however, cannot be completely ruled out, as a single isolate (SA1073), isolated from a soybean leaf in Tennessee that was showing early symptoms of CLB, shared 97–100% sequence similarity with actin (ACT), calmodulin (CAL), translation elongation factor 1-α (EF-1α), histone 3 (H3) and ITS, (legacy genes). However, the species most frequently isolated from infected soybean material in this study was *C*. cf. *flagellaris,* and more rarely, *C*. cf. *sigesbeckiae* (isolated five times).

Though *C*. cf. *flagellaris* and *C*. cf. *sigesbeckiae* have long been known as generalist pathogens, they have only recently been associated with soybean. Bakhshi et al. (2015) reported *C*. cf. *flagellaris* from *Glycine max* in Iran, and later, Soares et al. (2015) reported it from Arkansas and also *C*. cf. *sigesbeckiae* from South America. Soares et al. (2015) did not find *C*. cf *flagellaris* in Argentina or Brazil, and speculated that it might be restricted to colder climates. However, reports of *C*. cf *flagellaris* from soybean grown in geographic regions that do not have cold climates, such as Louisiana (this dissertation), Iran (Bakhshi et al. 2015) and also from other hosts in Ethiopia, Fiji, Puerto Rico, South Africa, Taiwan, the U.S. Virgin Islands and Florida and Texas (Groenewald et al. 2015; (cited in Farr and Rossman 2015), suggest that *C*. cf. *flagellaris* is ubiquitous throughout tropical and subtropical regions as well.

The many hosts from which *C*. cf. *flagellaris* and *C*. cf. *sigesbeckiae* have been isolated raises the issue of whether some of these plants, such as *Phytolacca* and other noncultivated plant species which may be present in or around soybean fields, are acting as reservoirs of inoculum from year to year. The etiology of CLB and PSS is poorly understood and the potential role that non-soybean hosts play in helping to spread disease should be investigated in the future. In this study, all of the *C*. cf. *flagellaris* isolates from *Phytolacca* collected in Arkansas and Louisiana were part of AHI haplotype 1, which also contained soybean leaf and seed isolates from Arkansas, Louisiana and Missouri. Two other isolates from *Phytolacca* collected in Illinois were part of AHI haplotype 20. Because I had many more *C*. cf. *flagellaris* isolates from soybean than from *Phytolacca,* at this time it is not possible to compare the genetic diversity of isolates from these two hosts. However, with increased sampling from *Phytolacca,* I would expect that genetic diversity of *C.* cf. *flagellaris* from *Phytolacca* would also be comparable to that seen on soybean. Additional isolations from infected *Phytolacca* plants near soybean fields (ideally, near fields where genotyped soybean isolates were previously collected) are needed to see if *Phytolacca* isolates are restricted to AHI haplotypes 1 and 20, or if they are distributed more broadly and correspond to the AHI haplotypes of geographically related soybean isolates.

Given that there is no segregation of leaf or seed isolates into distinctive haplotypes, and that *C.* cf. *flagellaris* subclades 1 and 2 contain both leaf and seed isolates (Figures 3.2 and 3.3), it seems likely that CLB and PSS are caused by the same organism. The most straightforward method to test this hypothesis would be by cross-inoculating soybean leaves with seed isolates and vice versa. However, obtaining proof of pathogenicity in the greenhouse for this system often yields inconclusive results, and even when disease is established, *in vitro* symptoms do not mimic those observed in the field (Cai et al. 2009; personal observation). A different approach proposed to isolate the pathogen from leaves, seeds and stems at various times throughout the year to see if the same organism is repeatedly isolated from infections on different plant parts

and if there is a particular environmental or other event that coincides with disease (V. Doyle, personal communication). Preliminary work using a qPCR assay specific for C. cf. *flagellaris* (Chanda et al. 2014), found that this pathogen is present within soybean seedlings during early vegetative growth (R. Schneider and A. Chanda, personal communications). This suggests that the pathogen is transmitted through seed, and a latent period may precede the development of foliar symptoms, which could be triggered by environmental stimuli such as light, that incite the pathogen to produce cercosporin.

It is still unclear whether sexual reproduction is occurring in *C.* cf. *flagellaris*. Given the close relationship of *Cercospora* to *Mycosphaerella*, if sex was occurring, pseudothecia would be expected, but neither ascomes nor ascospores have been reported. Cai (2004) and Price (2013) isolated single spores of *C*. cf. *flagellaris* and *C*. cf. *sigesbeckiae* from soybean foliar lesions and seeds, but did not report seeing sexual structures. I also did not observe any sexual structures on soybean or *Phytolacca* leaves. Without morphological evidence, the existence of a sexual stage must be inferred through other means. Cai and Schneider (2005) observed high vegetative compatibility group (VCG) diversity among a poplation of *C*. cf. *flagellaris* isolates from Louisiana. They found that, though there were several multimember VCGs, none of these was dominant among the population, and most isolates were incompatible with all others. From these results they concluded that that a cryptic sexual stage was functioning or was only recently lost.

Another means by which to infer the presence of sexual recombination in fungi is to characterize the mating type loci among individuals within a population and assess whether the ratio of mating types is equal. Both MAT1-1 and MAT1-2 idiomorphs were present within the population of Louisiana isolates collected during 2000, 2011 and 2012, though MAT1-2 was more common. The chi square test performed in chapter two rejected the null hypothesis of equal proportions of MAT1-1 and MAT1-2, indicating that MAT1-2 dominates in Louisiana. However, the sampling strategies used when collecting these isolates were not designed to capture the mating type diversity at each location. Both mating types were often present in the same field during all three years, but limited sampling across individual fields limits inferences about their distribution within those fields. Furthermore, sampling was not standardized across the three years. The isolates from 2000 were collected from leaves and seeds derived from single soybean fields in Alexandria, Baton Rouge and Winnsboro, LA by Cai (2004). Those collected during 2011 and 2012 were isolated from foliar lesions of soybean plants rogued from individual fields in many parishes throughout Louisiana (P. P. Price, personal communication; Appendix 1 Table A1.1). In both cases, isolations were not made with a population study of mating types in mind. In designing a future study focusing on capturing the mating type diversity in Louisiana, I would concentrate on sampling foliar lesions more intensively, but in fewer fields. Furthermore, I would not only look at the distribution of mating types within fields, but also at the microspatial distribution of mating types within leaves and lesions. In other words, what are the proportions of MAT1-1 and MAT1-2 within individual fields, on individual leaves and within the same lesion on a leaf?

Findings from the first objective of this project showed that, individually and collectively, the legacy genes do not contain sufficient phylogenetic signal to resolve interspecific relationships within *Cercospora*. However, the intergenic regions (IGS loci) developed in chapter three show promise as additional phylogenetic markers that can be used to better understand the evolutionary history of *Cercospora*. Admittedly, there are inherent assumptions about the taxonomy of this genus that must be made based on the available information. Our understanding of the evolutionary history of *Cercospora* has been founded on phylogenetic

inferrence using the legacy genes and a few other loci (cytb, MAT loci and BTub). Therefore, future studies developing alternative phylogenetic markers must consider their results in the context of previous systematic studies when assessing the utility of new markers.

The IGS loci developed here had higher phylogenetic informativeness values than the legacy genes; concatenated alignments of "IGS" contained more bipartitions with stronger support than "legacy." However, the best trees were obtained when "all loci" ("IGS" and "legacy") were concatenated. Neither "IGS", nor "legacy" gene trees were able to resolve as many clades as the concatenated analyses. While it was not possible to infer relationships among many taxa in the "legacy" gene trees due to polytomies, the "IGS" gene trees had more splits and stronger support at interspecific nodes (Appendix 7 Figures A7.1–A7.17). Ultimately, with the exception of certain "rogue taxa" like *C. agavicola, C. mercurialis* and *C. pileicola,* whose positions often fluctuated depending on the combination of genes used, there was not great conflict between the concatenated IGS and legacy topologies within the same filtering treatment. That phylogenies inferred from different genomic regions are generating similar results, is a good sign and suggests we may be on the right track to inferring the species tree for *Cercospora*.

I concluded that for IGS alignments, which were often highly variable among divergent taxa, the less stringent Gblocks treatment (LSGb) was a more appropriate filtering algorithm than not using Gblocks at all (NoGb) or using the more stringent Gblocks treatment (MSGb). Treatment was less of a factor for the legacy genes because they are more conserved and contain fewer gaps and indels. Although tree topologies did not differ noticeably between treatments NoGb and LSGb, I found that, with the exception of IGS6, all IGS loci were saturated, making them unfit for phylogenetic analysis (Xia et al. 2003). In contrast, none of the loci in treatment MSGb were saturated, but because MSGb stripped alignments of all gaps, and thus, of many potentially informative sites, this approach was deemed to be too aggressive. Matching split (MS) distances calculated between independent gene trees from each filtering treatment also showed that fewer rearrangements were necessary to arrive at identical tree topologies between treatments NoGb and LSGb than for either and MSGb.

These results are encouraging and show that, using relatively few and inexpensive resources, it is possible to develop new tools to answer evolutionary questions. Given the limited number of species used in this study, increasing the number of taxa in future phylogenetic studies using these IGS loci will continue to improve our understanding of species in a complex genus like *Cercospora,* where morphological homoplasy has long confounded taxonomy.

REFERENCES

Aagesen, L. 2004. The information content of an ambiguously alignable region, a case study of the trnL intron from the Rhamnaceae. Organ. Divers. Evol. 4:35–49.

Almeida, A.M.R., Piuga, F.F., Marin, S.R.R., Binneck, E., Sartori, F., Costamilan, L.M., Teixeira, M.R.O. and Lopes, M. 2006. Pathogenicity, molecular characterization, and cercosporin content of Brazilian isolates of *Cercospora kikuchii*. Fitopatol. Bras. 30:594–602.

Assante, G., Locci, R., Camarda, L., Merlini, L. and Nasini, G. 1977. Screening of the genus *Cercospora* for secondary metabolites. Phytochemistry 16:243–247.

Atkinson, G.F. 1891. Some Cercosporae from Alabama. J. Elisha Mitchell Soc. 8:33–67.

Bakhshi, M., Arzanlou, M., Babai-Ahari, A., Groenewald, J. Z., Braun, U. and Crous, P.W. 2015. Application of the consolidated species concept to *Cercospora* spp. from Iran. Persoonia 34:65–86.

Blomberg, S.P., Garland, T. and Ives, A.R. 2003. Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. Evolution 57:717–745.

Bogdanowicz, D. and Giaro, K. 2012. Matching split distance for unrooted binary phylogenetic trees. IEEE/ACM Trans Comput. Biol. Bioinform. 9:150–160.

———, ——— and Wrobel, B. 2012. TreeCmp: Comparison of trees in polynomial time. Evol. Bioinform. 8:475–487.

Braun, U., Crous, P.W., Dugan, F, Groenewald, J.Z. and de Hoog G,S,. 2003. Phylogeny and taxonomy of *Cladosporium*-like hyphomycetes, including *Davidiella* gen. nov., the teleomorph of *Cladosporium s. str.* Mycol. Prog. 2:3–18.

Cai, G. 2004. Cercospora leaf blight of soybean: pathogen vegetative compatibility groups, population structure, and host resistance. (Master's Thesis: Louisiana State University). 78 pp.

——— and Schneider, R.W. 2005. Vegetative compatibility groups in *Cercospora kikuchii*, the causal agent of Cercospora leaf blight and purple seed stain in soybean. Phytopathology 95:257–261.

——— and ———. 2008. Population structure of *Cercospora kikuchii*, the causal agent of Cercospora leaf blight and purple seed stain in soybean. Phytopathology 98:823–829.

———, ——— and Padgett, G. B. 2009. Assessment of lineages of *Cercospora kikuchii* in Louisiana for aggressiveness and screening soybean cultivars for resistance to Cercospora leaf blight. Plant Dis. 93:868–874.

Carbone, I. and Kohn, L. M. 1999. A method for designing primer sets for speciation studies in filamentous ascomycetes. Mycologia 91:553–556.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17:540–552.

Chanda, A. K., Ward, N. D., Robertson, C. L., Chen, Z.Y. and Schneider, R.W. 2014. Development of a quantitative polymerase chain reaction detection protocol for *Cercospora kikuchii* in soybean leaves and its use for documenting latent infection as affected by fungicide applications. Phytopathology 104:1118–1124.

Chevreux, B., Wetter, T. and Suhai, S. 1999. Pages 45–56 in: Genome sequence assembly using trace signals and additional sequence information. German conference on bioinformatics.

Chupp, C. 1954. A Monograph of the Fungus Genus *Cerospora*. Ithaca, New York. Published by the author. 667 pp.

Crouch, J.A., Clarke, B.B. and Hillman, B.I. 2009. What is the value of ITS sequence data in *Colletotrichum* systematics and species diagnosis? A case study using the falcate-spored graminicolous *Colletotrichum* group. Mycologia 101:648–656.

Crous, P.W. and Braun, U. 2003. *Mycosphaerella* and its anamorphs: 1. Names published in *Cercospora* and *Passalora*. CBS Biodiversity Series 1. pp. 1–571.

———, Groenewald, J.Z., Risede, J.M. and Hywel-Jones, N. L. 2004. *Calonectria* species and their *Cylindrocladium* anamorphs: species with sphaeropedunculate vesicles. Stud. Mycol. 50: 415–430.

———, ———, Pongpanich, K., Himaman, W., Arzanlou, M. and Wingfield, M.J. 2004b. Cryptic speciation and host specificity among *Mycosphaerella* spp. occurring on Australian *Acacia* species grown as exotics in the tropics. Stud. Mycol. 50:457–469.

———, ———. 2005. Hosts, species and genotypes: opinions versus data. Australas. Plant. Pathol. 34:463–470.

———, ———, Groenewald, M., Caldwell, P., Braun, U. and Harrington, T.C. 2006. Species of *Cercospora* associated with grey leaf spot of maize. Stud. Mycol. 55:189–197.

———, Braun, U. and Groenewald, J.Z. 2007. *Mycosphaerella* is polyphyletic. Stud. Mycol. 58: 1–32.

Curto, M.A., Puppo, P., Ferreira, D., Nogueira, M. and Meimberg, H. 2012. Development of phylogenetic markers from single-copy nuclear genes for multi locus, species level analyses in the mint family (Lamiaceae). Mol. Phylogenet. Evol. 63:758–67.

Damm, U., Cannon, P.F., Woudenberg, J.H.C. and Crous, P.W. 2012. The *Colletotrichum acutatum* species complex. Stud. Mycol. 73:37–113.

Darriba, D., Taboada, G.L., Doallo, R. and Posada, D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9(8):772.

Davidson, R.M., Hanson, L.E., Franc, G.D. and Panella, L. 2006. Analysis of ß-tubulin gene fragments from benzimidazole-sensitive and tolerant *Cercospora beticola*. J. Phytopathol. 154:321–328.

Dessimoz, C. and Gil, M. 2010. Phylogenetic assessment of alignments reveals neglected tree signal in gaps. Genome Biol. 11:R37.

Daub, M.E. and Hangarter, R.P. 1983. Light-induced production of singlet oxygen and superoxide by the fungal toxin, cercosporin. Plant Physiol. 73:855–857.

——— and Briggs, S.P. 1983. Changes in tobacco cell-membrane composition and structure caused by cercosporin. Plant Physiol. 71:763–766.

——— and Chung, K.R. 2007. Cercosporin: A photoactivated toxin in plant disease. Plant Dis. Online. APSnet Features. doi: 10.1094/APSnetFeature/2007–0207.

Donaldson, G.C., Ball, L.A., Axelrood, P.E. and Glass, N.L. 1995. Primer sets developed to amplify conserved genes from filamentous ascomycetes are useful in differentiating Fusarium species associated with conifers. Appl. Environ. Microbiol. 61:1331–1340.

Doyle, V.P., Oudemans, P.V., Rehner, S.A. and Litt, A. 2013. Habitat and host indicate lineage identity in *Colletotrichum gloeosporioides* s.l. from wild and agricultural landscapes in North America. PLos One 8:e62394.

Duong TA, De Beer ZW, Wingfield BD, Eckhardt LG, Wing- field MJ (2015) Microsatellite and mating type markers reveal unexpected patterns of genetic diversity in the pine root infecting fungus *Grosmannia alacris*. Plant Pathol. 64:235–242.

Edgar, R. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Fajola, A.O. 1978. Cercosporin, a phytotoxin from *Cercospora* spp. Physiol. Mol. Plant. Pathol. 13:157–164.

Farr, D.F. and Rossman, A.Y. 2015. Fungal Databases, Systematic Mycology and Microbiology Laboratory, ARS, USDA. Retrieved June 15, 2015, from http://nt.ars-grin.gov/fungaldatabases/.

Feau, N., Decourcelle, T., Husson, C., Desprez-Loustau, M.L. and Dutech, C. 2011. Finding single copy genes out of sequenced genomes for multilocus phylogenetics in non-model fungi. PLoS ONE 6(4):e18803, doi:10.1371/ journal.pone.0018803.

Fredslund, J., Schauser, L., Madsen, L.H., Sandal, N. and Stougaard, J. 2005. PriFi: using a multiple alignment of related sequences to find primers for amplification of homologs. Nucleic Acids Res. 1:33 (Web Server issue):W516-20.

Fresenius, G. 1863. Beiträge zur Mycologie. 3. Heinrich Ludwig Brömmer Verlag, Frankfurt, Germany.

Fu, Y. and Somers, D.J. 2009. Genome-wide reduction of genetic diversity in wheat breeding. Crop Sci. 49: 161–168.

Gadberry, M.D., Malcomber, S.T., Doust, A.N. and Kellogg, E.A. 2005. Primaclade - A flexible tool to find conserved PCR primers across multiple species. Bioinformatics. 21(7):1263–1264.

Gardes, M. and Bruns, T.D. 1993. ITS primers with enhanced specificity for basidiomycetes - application to the identification of mycorrhizae and rusts. Mol. Ecol. 2:113–118.

Gardner, M.W. 1926. Indiana plant diseases, 1924. Proc. Indiana Acad. Sci. 237–257.

Geisler, L.G. 2013. Purple seed stain and Cercospora blight. Plant Pathology Fact Sheet. University of Nebraska-Lincoln. http://pdc.unl.edu/agriculturecrops/soybean/purpleseedstain.

Goodwin, S.B., Dunkle, L.D. and Zismann, V.L. 2001. Phylogenetic analysis of *Cercospora* and *Mycosphaerella* based on the internal transcribed spacer region of ribosomal DNA. Phytopathology. 91:648–658.

Grigoriev, I.V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I.L., Shabalov and I. 2014. MycoCosm portal: Gearing up for 1000 fungal genomes. Nucleic Acids Res. 42(D1):D699–704.

Groenewald, J.Z., Nakashima, C., Nishikawa, J., Park, J.H., Jama, A.N., Groenewald, M., Braun, U. and Crous, P.W. 2013. Species concepts in *Cercospora*: spotting the weeds among the roses. Stud. Mycol. 75:115–170.

———, Groenewald, M., Braun, U. and Crous, P.W. 2010. *Cercospora* speciation and host range. Pages 21–37 in: Cercospora Leaf Spot of Sugar Beet and Related Species. Lartey, R.T, Weiland, J.J., Panella, L., Crous, P.W. and Windels, C.E., eds. APS Press, St. Paul, MN.

Groenewald, M., Groenewald, J.Z., and Crous, P.W. 2005. Distinct species exist within the *Cercospora apii* morphotype. Phytopathology 95:951–959.

———, ———, Harrington, T.C., Abeln, E.C.A. and Crous, P.W. 2006. Mating type gene analysis in apparently asexual *Cercospora* species is suggestive of cryptic sex. Fungal Genetics and Biology 43: 813-825.

Grundy, W.N. and Naylor, G. J. 1999. Phylogenetic inference from conserved sites alignments. J. Exp. Zool. 285:128-139.

Guindon, S. and Gascuel, O. 2003. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. Sys. Biol. 52:696-704.

Han, Y.S. 1959. Studies on purple spot of soybean. J. Agric. and For. 8:1–32.

Harrington, B. et al. 2004–2005 Inkscape. http://www.inkscape.org/.

Hartman, G.L., Sinclair, J. B. and Rupe, J.C. 1999. Compendium of Soybean Diseases. 4th ed. APS Press, St. Paul, MN. 100 pp.

Hershman, D. 2009. Cercospora leaf blight in Kentucky. Plant Pathology Fact Sheet PPFS-AG-S-20. University of Kentucky Cooperative Extension Service. http://www2.ca.uky.edu/agcollege/plantpathology/ext_files/PPFShtml/ppfsags20.pdf.

Hofstetter, V., Miadlikowska, J., Kauff, F. and Lutzoni, F. 2006. Phylogenetic comparison of protein-coding versus ribosomal RNA-coding sequence data: a case study of the Lecanoromycetes Ascomycota. Mol. Phylogenet. Evol. 44: 412-426.

Horsfall, J.G. 1929. Species of *Cercospora* on *Trifolium*, *Medicago* and *Melilotus*. Mycologia 21:304–312.

Huelsenbeck, J.P. and Ronquist. F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754-755.

Hyde, K.D., Cai, L., McKenzie, E.H.C., Yang, Y.L., Zhang, J.Z. and Prihastuti, H. 2009. *Colletotrichum*: a catalogue of confusion. Fungal Divers. 39:1–17.

Hymowitz, T. 1970. On the domestication of the soybean. Economic Botany 24:408–421.

Hymowitz, T. and Harlan. J.R. 1983. Introduction of soybean to North America by Samuel Bowen in 1765. Econ. Bot. 37:371–379.

——— and Shurtleff, W.R. 2005. Debunking soybean myths and legends in the historical and popular literature. Crop Sci. 45:473–476.

Hyten, D.L., Song, Q., Zhu, Y., Choi, I.Y., Nelson, R.L., Costa, J.M., Specht, J.E., Shoemaker, R.C. and Cregan, P.B. 2006. Impacts of genetic bottlenecks on soybean genome diversity. Proc. Nat. Acad. Sci. 103:16666–16671.

Imazaki, I., Homma, Y., Kato, M., Vallone, S., Yorinori, J.T., Henning, A.A., Iizumi, H. and Koizumi, S. 2006. Genetic relationships between *Cercospora kikuchii* populations from South America and Japan. Phytopathology 96:1000–1008.

Jackson, E.W., Fenn, P. and Chen, P. 2006. Inheritance of resistance to purple seed stain caused by *Cercospora kikuchii* in PI 80837 soybean. Crop Sci. 46:1462–1466.

Jenns, A.E., Daub, M.E. and Upchurch, R. G. 1989. Regulation of cercosporin accumulation in culture by medium and temperature manipulation. Phytopathology 79:213–219.

Jones, J.P. 1959. Purple stain of soybean seeds incited by several *Cercospora* species. Phytopathology 49:430-432.

Kalendar, R., Lee, D. and Schulman, A.H. 2014. Fast PCR software for PCR, *in silico* PCR, and oligonucleotide assembly and analysis. Methods Mol Biol. 1116(2):271–302. Available from: http://dx.doi.org/10.1016/j.ygeno.2011.04.009.

Katoh, K. and Standley, D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30:772–780.

Kepler, R.M. and Rehner, S.A. 2013. Genome-assisted development of nuclear intergenic sequence markers for entomopathogenic fungi of the *Metarhizium anisopliae* species complex. Mol. Ecol. Resour. 13:210–217.

Kilpatrick, R.A. and Johnson H.W. 1956. Purple stain of legume seeds caused by *Cercospora* species. Phytopathology 46:201–204.

Kim, H., Newell, A.D., Cota-Sieckmeyer, R.C., Rupe, J.C., Fakhoury, A.M. and Bluhm, B.H. 2013. Mating-type distribution and genetic diversity of *Cercospora sojina* populations on soybean from Arkansas: evidence for potential sexual reproduction. Phytopathology 103:1045–1051.

Kirschner R. 2013. A new species and new records of cercosporoid fungi from ornamental plants in Taiwan. Mycol. Prog. 13:483–491.

Ko, T.W.K., Stephenson, S.L., Bahkali, A.H. and Hyde, K.D. 2011. From morphology to molecular biology: can we use sequence data to identify fungal endophytes? Fungal Divers. 36:69–88.

Kretzer, A.M. and Bruns, T.D. 1999. Use of atp6 in fungal phylogenetics: an example from the Boletales. Mol. Phylogenet. and Evol. 13:483-492.

Kuyama, S. and Tamura, T. 1957. Cercosporin - a pigment of *Cercosporina kikuchii* Matsumoto et Tomoyasu. Cultivation of fungus, isolation and purification of pigment. J. Am. Chem. Soc. 79:5725–5726.

Latorre-Rapela, M.G., Colombini, M.A., González, A.M., Vaira, S.M., Maumary, R., Mattio, M.C., Carrera, E., and Lurá, M.C. 2011. Phenotypic and genotypic variability in *Cercospora kikuchii* Isolates from Santa Fe Province, Argentina. In: Soybean - Genetics and Novel Techniques for Yield Enhancement: InTech. Available from: http://www.intechopen.com/books/soybean-genetics-and-novel-techniques-for-yield-enhancement/phenotypic-and-genotypic-variability-in-cercospora-kikuchii-isolates-from-santa-fe-province-argentin.

Lee, G.A, Crawford, G.W., Liu, L., Sasaki, Y. and Chen, X.X. 2011. Archaeological soybean (*Glycine max*) in East Asia: does size matter? Plos One 10.1371/journal.pone.0026720.

Lehman, S.G. 1928. Soybean diseases. Fiftieth Annual Report of the North Carolina Agric. Expt. Sta.

———, S.G. 1950. Purple stain of soybean seeds. N.C. Agric. Stn. Bull. 369.

Li, K.N., Rouse, D.I. and German, T.L. 1994. PCR primers that allow intergeneric differentiation of ascomycetes and their application to *Verticillium* spp. App. Env. Microbiol. 60:4324-4331.

Li, Y.H., Zhang, C., Smulders, M.J.M., Li, W., Ma, Y.S., Xu, Q., Chang, R.Z. and Qiu, L.J. 2013. Analysis of average standardized SSR allele size supports domestication of soybean along the Yellow River. Gen. Res. Crop Evol. 60:763-776.

Librado, P., and Rozas, J. 2009. DnaSP V5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452.

Liu, L., Yu, L., Pearl, D.K., and Edwards, S.V. 2009. Estimating species phylogenies using coalescence times among sequences. Syst. Biol. 58:468–477.

———, ———, and Edwards, S.V. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evol. Biol. 10:302.

Liu, Y.J., Whelen, S. and Hall, B.D. 1999. Phylogenetic relationships among ascomycetes: evidence from an RNA polymerse II subunit. Mol. Biol. Evol. 16: 1799–1808.

Lopez-Giraldez, F. and Townsend, J.P. 2011. PhyDesign: an online application for profiling phylogenetic informativeness. BMC Evol. Biol. 11:152.

Lurá, M. C., Latorre Rapela, M. G., Vaccari, M. C., Maumary, R., Soldano, A., Mattio, M. and González, A. M. 2011. Genetic diversity of *Cercospora kikuchii* isolates from soybean cultured in Argentina as revealed by molecular markers and cercosporin production. Mycopathologia 171:361–371.

Matsumoto, T., and Tomoyasu, R. 1925. Studies on purple speck of soybean seed. Ann. Phytopathol. Soc. Jpn. 1:1–14.

McLean, K.S. and Roy, K.W. 1988. Purple seed stain of soybean caused by isolates of *Cercospora kikuchii* from weeds. Can. J. of Plant Pathol. 10:166–171.

Meyer, R.S., DuVal, A.E. and Jensen, H.R. 2012. Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. New Phytologist 196:29–48.

Miller, M.A., Pfeiffer, W., and Schwartz, T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Pages 1–8 in: Proceedings of the Gateway Computing Environments GCE Workshop, 14 Nov. 2010, New Orleans, LA.

Morgulis A, Gertz, E.M., Schaffer, A.A., Agarwala, R. 2006. A fast and symmetric DUST implementation to mask low-complexity DNA sequences. J. Comput. Biol. 13:1028–1040.

Murakishi, H.H. 1951. Purple seed stain of soybean. Phytopathology 41: 305–318.

Nagy, L.G., Kocsubé, S., Csanádi, Z., Kovács, G.M., Petkovits, T., Vágvölgyi, C., et al. 2012. Re-Mind the gap! Insertion - deletion data reveal neglected phylogenetic potential of the nuclear ribosomal internal transcribed spacer (ITS) of fungi. PLoS One. 7(11):1–9.

Nylander, J. 2004. MrModeltest 2.3. Computer program and documentation distributed by the author. Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden.

O'Donnell, K. and Cigelnik, E. 1997. Two divergent intragenomic rDNA ITS2 types within a monophyletic lineage of the fungus *Fusarium* are nonorthologous. Mol. Phylo. Evol. 7:103-116.

Ohm, R., Aerts, A., Salamov, S., Goodwin, S. and Grigoriev, I. 2011. Comparative analysis of twelve Dothideomycete plant pathogens. Report Number: LBNL-4522E-Poster.

Paradis, E., Claude, J., Strimmer, K. 2004. Ape: analyses of phylogenetics and evolution in R language. Bioinformatics. 20:289–290.

Penn, O., Privman, E., Ashkenazy, H., Landan, G., Graur, D. and Pupko, T. 2010. GUIDANCE: a web server for assessing alignment confidence scores. Nuc. Acids Res. Jul 1; 38 (Web Server issue):W23-W28; doi: 10.1093/nar/gkq443.

Pollack, F. 1987. An annotated complilation of *Cercospora* names. Mycol. Mem. 12:1–212.

Pond, S.L.K., Frost, S.D.W. and Muse, S.V. 2005. Hyphy: hypothesis testing using phylogenies. Bioinformatics 21: 676–679.

Price, P.P. 2013. Sensitivity and resistance of *Cerospora kikuchii*, causal agent of Cercospora leaf blight and purple seed stain of soybean, to selected fungicides. (Doctoral Dissertation: Louisiana State University). 111 pp.

———, Purvis, M., Cai, G., Padgett, G.B., Robertson, C.L., Schneider, R.W., and Albu, S. 2015. Fungicide resistance in *Cercospora kikuchii*, a soybean pathogen. Plant Dis. 99:1596–1603.

Rambaut, A. 2012. FigTree version 1.4.0. http://tree.bio.ed.ac.uk/software/figtree/.

Rambaut, A., Suchard, M.A., Xie, D. and Drummond, A.J. 2013. Tracer v1.6. http://beast.bio.ed.ac.uk/software/tracer/.

Rojas EI, Rehner, S.A., Samuels, G.J., Bael, S.A.V., Herre, E.A., Cannon, P., Chen, R., Pan, J., Wang, R., Zhang, Y., Peng, Y.Q., Sha, T. 2010. *Colletotrichum gloeosporioides s.l.* associated with *Theobroma cacao* and other plants in Panama´: multilocus phylogenies

distinguish host associated pathogens from asymptomatic endophytes. Mycologia 102:1318–1338.

Roy, K.W. 1982. *Cercospora kikuchii* and other pigmented *Cercospora* species: cultural and reproductive characteristics and pathogenicity to soybean. Can. J. of Plant Pathol. 4:226–232.

Ruibal, M.P., Peakal,l R., Foret, S., Linde, C.C. 2014. Development of phylogenetic markers for *Sebacina* (Sebacinaceae) mycorrhizal fungi associated with Australian orchids. Appl Plant Sci. 2(6):2–6.

Rush, T.A., and Aime, M.C. 2013. The genus *Meira*: Phylogenetic placement and description of a new species. Antonie van Leeuwenhoek 103:1097–1106.

Sarr, M.P., Ndiaye, M., Groenewald, J.Z. and Crous, P.W. 2014. Genetic diversity in *Macrophomina phaseolina*, the causal agent of charcoal rot. Phytopathol. Mediterr. 53:250–268.

Schuh, W. 1992. Effect of pod development stage, temperature, and pod wetness duration on the incidence of purple seed stain of soybeans. Phytopathology 82:446–451.

Silva, D.N., Talhinhas, P., Varzea, V., Cai, L., Paulo, O.S., Batista, D. 2012. Application of the Alm2/MATlocus to improve the systematics of the *Colletotrichum gloeosporioides* complex: an example from coffee (*Coffea* spp.) hosts. Mycologia 104:396–409.

Solheim, W. G. and Stevens, F. L. 1931. *Cercospora* studies: II. Some tropical Cercosporae. Mycologia 23:365–405.

Stamatakis, A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

Steenkamp, E.T., Wingfield, B.D., Desjardins, A.E., Marasas, W.F.O. and Wingfield, M.J. 2002. Cryptic speciation in *Fusarium subglutinans*. Mycologia 94:1032–1043.

Stephens, M., and Donnelly, P. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am. J Hum. Genet. 73:1162–1169.

———, Smith, N., and Donnelly, P. 2001. A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68:978–989.

Sukumaran, J. and Holder, M.T. 2010. DendroPy: a Python library for phylogenetic computing. Bioinformatics 26:1569–1571.

Suzuki, K. 1921. Studies on the cause of "shihan" of soybeans. Chosen Nakaiho 16:24–28. (In Japanese).

Swofford, D.L., Olsen, G.J., Waddell, P.J. and Hillis, D.M. 1996. Phylogenetic inference. Pages 407–514 in Molecular Systematics. Hillis, D.M. Moritz, C. and Mable, B.K., eds. Sinauer Associates, Sunderland, MA.

———. 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, MA. 142 pp.

Talavera, G., and Castresana, J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56:564–577.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30:2725–2729.

Tan, G., Muffato, M., Ledergerber, C., Herrero, J., Goldman, N., Gil, M. and Dessimoz, C. 2015. Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. Syst. Biol. 64(5):778–791.

Taylor, J.W., Jacobson, D.J., Kroken, S., Kasuga, T., Geiser, D.M., Hibbett, D.S. and Fisher, M.C. 2000. Phylogenetic species recognition and species concepts in fungi. Fungal Genetics and Biology 31:21–32.

Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y.O., Borodovsky, M. 2008. Gene prediction in novel fungal genomes using an *ab initio* algorithm with unsupervised training. Genome Res. 18 (12):1979–1990.

Townsend, J.P. 2007. Profiling phylogenetic informativeness. Syst. Biol. 56:222-231.

Trkulja, N., Ivanovic, Ž., Pfaf-Dolovac, E., Dolovac, N., Mitrovic, M., Toševski, I. and Jovic, J. 2013. Characterisation of benzimidazole resistance of *Cercospora beticola* in Serbia using PCR-based detection of resistance-associated mutations of the ß-tubulin gene. Eur. J. Plant Pathol. 135:889-902.

Waalwijk, C., Mendes, O., Verstappen, E.C., de Waard, M.A. and Kema, G.H. 2002. Isolation and characterization of the mating-type idiomorphs from the wheat Septoria leaf blotch fungus *Mycosphaerella graminicola*. Fungal Genet. Biol. 35:277–286.

Walker, D.M., Castlebury, L. A., Rossman, A.Y., White, J.F. 2012. New molecular markers for fungal phylogenetics: Two genes for species-level systematics in the Sordariomycetes (Ascomycota). Mol. Phylogenet. Evol. 64(3):500–12.

Walters, H.J. 1980. Soybean leaf blight caused by *Cercospora kikuchii*. Plant Dis. 64:961–962.

Wang, J., Levy, M., and Dunkle, L.D. 1998. Sibling species of *Cercospora* associated with gray leaf spot of maize. Phytopathology 88:1269–1275.

Weir, B., Johnston, P.R., and Damm, U. 2012. The *Colletotrichum gloeosporioides* species complex. Stud. Mycol. 73:115–180.

White, T. J., Bruns, T., Lee, S., and Taylor, J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pages 315–322 in: PCR protocols: a guide to methods and applications. Innis, M.A., Gelfand, DH., Sninsky, J.J., White, T.J., eds. Academic Press Inc., San Diego, CA.

Wong, K.M., Suchard, M.A., Huelsenbeck, J.P. 2008. Alignment uncertainty and genomic analysis. Science 319:473–476.

Wrather J.A. and Koenning, S. 2010. Soybean Disease Loss Estimates for the United States, 1996–2010. Available from http://aes.missouri.edu/delta/research/soyloss.stm.

Wrather, A., Shannon, G., Balardin, R., Carregal, L., Escobar, R., Gupta, G.K., Ma, Z., Morel, W., Ploper, D., and Tenuta A. 2010. Effect of diseases on soybean yield in the top eight producing countries in 2006. Plant Health Prog. http://dx.doi.org/10.1094/PHP-2010-0125-01-RS.

Xia, X., Z. Xie, M. Salemi, L. Chen, Y. Wang. 2003. An index of substitution saturation and its application. Mol. Phylogenet. Evol. 26:1–7.

——— and Lemey P. 2009. Assessing substitution saturation with DAMBE. Pages 615–630 in: Lemey, P., Salemi, M., Vandamme, A.M., eds. The Phylogenetic Handbook. Cambridge University Press, Cambridge, England.

———. 2013. DAMBE5: A comprehensive software package for data analysis in molecular biology and evolution. Mol. Biol. Evol. 30:1720–1728.

Yeh, C.C. and Sinclair, J.B. 1982. Effect of *Cercospora kikuchii* on soybean seed germination and its interaction with *Phomopsis* sp. J. Phytopathol. 105:265–270.

Zaltsman, A. and Citovsky, V. 2012. Plant Biotechnology and Agriculture Prospects For the 21st Century (Altman, A and Hasegawa, P.M. eds.). Quart. Rev. Biol. 87:268.

Zhu, Q., Zheng, X., Luo, J., Gaut, B.S. and Ge, S. 2007. Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. Mol. Biol. Evol. 24:875–888.

Zwickl, D.J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biolgical sequence datasets under the maximum likelihood criterion. (Doctoral Dissertation: The University of Texas at Austin). 115 pp.

APPENDIX 1. SPECIMEN INFORMATION FOR ISOLATES USED IN PHYLOGENETIC ANALYSES IN THIS STUDY

Table A1.1. Collection information and accession numbers for *Cercospora* species in chapter two

				Gb Accession Nos.					
Species	Voucher	Host	Origin	ACT	CAL	EF1	H3	ITS	MAT1-1
Cerc. achyranthis	CBS 132613	*Achyranthes japonica*	S. Korea	JX143031	JX142785	JX143277	JX142539	JX143523	
Cerc. achyranthis	CPC 10091	*Achyranthes japonica*	S. Korea	JX143032	JX142786	JX143278	JX142540	JX143524	DQ264733
Cerc. agavicola^{T}	CBS 117292	*Agave tequilana var. azul*	Mexico	AY966898	AY966899	AY966897	AY966900	AY647237	
Cerc. alchemillicola^{T}	CPC 5259	*Alchemilla mollis*	New Zealand	JX143033	JX142787	JX143279	JX142541	JX143525	
Cerc. cf. *alchemillicola*	CPC 5126	*Oenothera fruticosa*	New Zealand	JX143034	JX142788	JX143280	JX142542	JX143526	
Cerc. cf. *alchemillicola*	CPC 5127	*Gaura lindheimeri*	New Zealand	JX143035	JX142789	JX143281	JX142543	JX143527	
Cerc. althaeina	CBS 126.26	*Malva* sp.		JX143036	JX142790	JX143282	JX142544	JX143528	DQ264742
Cerc. althaeina	CBS 132609	*Althaea rosea*	S. Korea	JX143037	JX142791	JX143283	JX142545	JX143529	
Cerc. althaeina^{T}	CBS 248.67	*Althaea rosea*	Romania	JX143038	JX142792	JX143284	JX142546	JX143530	
Cerc. apii	CBS 114416	*Apium* sp.	Austria	AY840447	AY840414	AY840483	AY840381	AY840516	

Cerc. apii	CBS 114418	*Apium graveolens*	Italy	AY840448	AY840415	AY840484	AY840382	AY840517	
Cerc. apii[T]	CBS 116455	*Apium graveolens*	Germany	AY840450	AY840417	AY840486	AY840384	AY840519	DQ264736
Cerc. apii	CBS 121.31	*Beta vulgaris*	Austria	AY840444	AY840411	AY840480	AY840378	AY840513	
Cerc. apii	CBS 252.67	*Plantago lanceolata*	Romania	DQ233368	DQ233394	DQ233342	DQ233420	DQ233318	
Cerc. apii	CPC 18601	*Apium graveolens*	CA, USA	JX143040	JX142794	JX143286	JX142548	JX143532	
Cerc. apii	MUCC 573	*Glebionis coronaria*	Japan	JX143043	JX142797	JX143289	JX142551	JX143535	
Cerc. apii	MUCC 923	*Asparagus officinalis*	Japan	JX143045	JX142799	JX143291	JX142553	JX143537	
Cerc. apii	CBS 119.25	*Apium graveolens*		AY840443	AY840443	AY840479	AY840377	AY840512	DQ264735
Cerc. apiicola[T]	CBS 116457	*Apium* sp.	Venezuela	AY840467	AY840434	AY840503	AY840401	AY840536	
Cerc. apiicola	CBS 116458	*Apium graveolens*	S. Korea	AY840468	AY840435	AY840504	AY840402	AY840537	
Cerc. apiicola	CPC 10248	*Apium* sp.	Venezuela	AY840470	AY840437	AY840506	AY840404	AY840539	
Cerc. apiicola	CPC 10759	*Apium graveolens*	S. Korea	AY840475	AY840442	AY840511	AY840409	AY840544	
Cerc. apiicola	CPC 11642	*Apium* sp.	Greece	DQ233393	DQ233419	DQ233367	DQ233441	DQ233341	

Cerc. armoraciae	CBS 115409	*Armoracia rusticana*	New Zealand	JX143048	JX142802	JX143294	JX142556	JX143540	
Cerc. armoraciae	CBS 132638	*Barbarea orthoceras*	S. Korea	JX143050	JX142804	JX143296	JX142558	JX143542	
Cerc. armoraciae[T]	CBS 250.67	*Armoracia rusticana*	Romania	JX143053	JX142807	JX143299	JX142561	JX143545	
Cerc. armoraciae	CBS 545.71	*Erysimum cuspidatum*	Romania	JX143057	JX142811	JX143303	JX142565	JX143549	
Cerc. armoraciae	CPC 11530	*Acacia mangium*	Thailand	JX143061	JX142815	JX143307	JX142569	JX143553	
Cerc. beticola	CBS 113069	*Spinacia* sp.	Botswana	DQ233377	DQ233403	DQ233351	DQ233429	DQ233325	
Cerc. beticola	CPC 5113	*Limonium sinuatum*	New Zealand	DQ233378	DQ233404	DQ233352	DQ233430	DQ233326	
Cerc. beticola[T]	CBS 116456	*Beta vulgaris*	Italy	AY840458	AY840425	AY840494	AY840392	AY840527	
Cerc. beticola	CPC 14616	*Goniolimon tataricum*	Bulgaria	FJ473432	FJ473437	FJ473427	FJ473442	FJ473422	
Cerc. beticola	CPC 11341	*Chrysanthemum segetum*	S. Korea	DQ233384	DQ233410	DQ233358	DQ233434	DQ233332	
Cerc. beticola	CBS 548.71	*Malva pusilla*	Romania	DQ233376	DQ233402	DQ233350	DQ233428	DQ233324	
Cerc. beticola	CPC 10195	*Beta vulgaris*	New Zealand	DQ233382	DQ233408	DQ233356	DQ026472	DQ233330	
Cerc. beticola	CPC 5123	*Apium graveolens*	New Zealand	DQ233379	DQ233405	DQ233353	DQ233431	DQ233327	
Cerc. beticola	CPC_5125	*Beta vulgaris*	New Zealand	AY752198	AY752229	AY752170	AY752260	AY752137	DQ264738

Cerc. cf. *brunkii*	CBS 132657	*Geranium thunbergii*	S. Korea	JX143067	JX142821	JX143313	JX142575	JX143559	
Cerc. cf. *brunkii*	MUCC 732	*Datura stramonium*	Japan	JX143068	JX142822	JX143314	JX142576	JX143560	
Cerc. campi-silii	CBS 132625	*Impatiens noli-tangere*	S. Korea	JX143069	JX142823	JX143315	JX142577	JX143561	
Cerc. aff. *canescens*	CBS 111134	*Vigna* sp.	S. Africa	DQ835104	DQ835131	DQ835085	DQ835158	AY260066	DQ264739
Cerc. aff. *canescens*	CBS 132658	*Dioscorea rotundata*	Ghana	JX143070	JX142824	JX143316	JX142578	JX143562	
Cerc. aff. *canescens*	CPC 11640	*Apium* sp.	USA	JX143074	JX142828	JX143320	JX142582	JX143566	
Cerc. capsici	CBS 118712	Lesions on calyx attached to fruit	Fiji	JX143076	JX142830	JX143322	JX142584	GU214653	
Cerc. capsici	CBS 132622	*Capsicum annuum*	S. Korea	JX143077	JX142831	JX143323	JX142585	JX143568	
Cerc. capsici	CPC 12307	*Capsicum annuum*	S. Korea	JX143078	JX142832	JX143324	JX142586	GU214654	
Cerc. capsici	MUCC 574	*Capsicum annuum*	Japan	JX143079	JX142833	JX143325	JX142587	JX143569	
Cerc. celosiae	CBS 132600	*Celosia argentea var. cristata*	S. Korea	JX143080	JX142834	JX143326	JX142588	JX143570	
Cerc. chenopodii	CBS 132620	*Chenopodium* cf. *album*	France	JX143081	JX142835	JX143327	JX142589	JX143571	
Cerc. cf. *chenopodii*[T]	CBS 132594	*Chenopodium ficifolium*	S. Korea	JX143082	JX142836	JX143328	JX142590	JX143572	

Cerc. cf. *chenopodii*	CBS 132677	*Chenopodium* sp.	Mexico	JX143083	JX142837	JX143329	JX142591	JX143573
Cerc. cf. *chenopodii*	CPC 12450	*Chenopodium ficifolium*	S. Korea	JX143084	JX142838	JX143330	JX142592	JX143574
Cerc. cf. *chenopodii*	CPC 15763	*Chenopodium* sp.	Mexico	JX143085	JX142839	JX143331	JX142593	JX143575
Cerc. cf. *chenopodii*	CPC 15859	*Chenopodium* sp.	Mexico	JX143086	JX142840	JX143332	JX142594	JX143576
Cerc. cf. *chenopodii*	CPC 15862	*Chenopodium* sp.	Mexico	JX143087	JX142841	JX143333	JX142595	JX143577
Cerc. chinensis	CBS 132612	*Polygonatum humile*	S. Korea	JX143088	JX142842	JX143334	JX142596	JX143578
Cerc. cf. *citrulina*	CBS 119395	*Musa* sp.	Bangladesh	JX143089	JX142843	JX143335	JX142597	EU514222
Cerc. cf. *citrulina*	CBS 132669	*Musa* sp.	Bangladesh	JX143090	JX142844	JX143336	JX142598	EU514223
Cerc. cf. *citrulina*	MUCC 576	*Citrullus lanatus*	Japan	JX143091	JX142845	JX143337	JX142599	JX143579
Cerc. cf. *citrulina*	MUCC 577	*Momordica charanthia*	Japan	JX143092	JX142846	JX143338	JX142600	JX143580
Cerc. cf. *citrulina*	MUCC 584	*Psophocarpus tetragonolobus*	Japan	JX143093	JX142847	JX143339	JX142601	JX143581
Cerc. cf. *citrulina*	MUCC 588	*Ipomoea pes-caprae*	Japan	JX143094	JX142848	JX143340	JX142602	JX143582

Cerc. coniogrammes	CBS 132634	*Coniogramme japonica var. gracilis*	Australia	JX143095	JX142849	JX143341	JX142603	JX143583	
Cerc. corchori[T]	MUCC 585	*Corchorus olitorius*	Japan	JX143096	JX142850	JX143342	JX142604	JX143584	
Cerc. cf. *coreopsidis*	CBS 132598	*Coreopsis lanceolata*	S. Korea	JX143097	JX142851	JX143343	JX142605	JX143585	
Cerc. cf. *coreopsidis*	CPC 10122	*Coreopsis lanceolata*	S. Korea	JX143098	JX142852	JX143344	JX142606	JX143586	
Cerc. delaireae[T]	CBS 132595	*Delairea odorata*	S. Africa	JX143099	JX142853	JX143345	JX142607	JX143587	
Cerc. delaireae	CPC 10627	*Delairea odorata*	S. Africa	JX143100	JX142854	JX143346	JX142608	JX143588	
Cerc. delaireae	CPC 10628	*Delairea odorata*	S. Africa	JX143101	JX142855	JX143347	JX142609	JX143589	
Cerc. delaireae	CPC 10629	*Delairea odorata*	S. Africa	JX143102	JX142856	JX143348	JX142610	JX143590	
Cerc. dispori	CBS 132608	*Disporum viridescens*	S. Korea	JX143103	JX142857	JX143349	JX142611	JX143591	
Cerc. cf. *erysimi*	CBS 115059	*Erysimum mutabile*	New Zealand	JX143104	JX142858	JX143350	JX142612	JX143592	DQ264740
Cerc. euphorbiae-sieboldianae[T]	CBS 113306	*Euphorbia sieboldiana*	S. Korea	JX143105	JX142859	JX143351	JX142613	JX143593	
Cerc. fagopyri[T]	CBS 132623	*Fagopyrum esculentum*	S. Korea	JX143106	JX142860	JX143352	JX142614	JX143594	

Cerc. fagopyri	CBS 132640	*Fallopia dumentorum*	S. Korea	JX143107	JX142861	JX143353	JX142615	JX143595
Cerc. fagopyri	CBS 132649	*Viola mandschurica*	S. Korea	JX143108	JX142862	JX143354	JX142616	JX143596
Cerc. fagopyri	CBS 132671	*Cercis chinensis*	S. Korea	JX143109	JX142863	JX143355	JX142617	JX143597
Cerc. fagopyri	MUCC 130	*Cosmos bipinnata*	Japan	JX143110	JX142864	JX143356	JX142618	JX143598
Cerc. fagopyri	MUCC 866	*Hibiscus syriacus*	Japan	JX143111	JX142865	JX143357	JX142619	JX143599
Cerc. cf. *flagellaris*	PP_004	*Glycine max* leaf	Rapides (DLRS), LA, USA					
Cerc. cf. *flagellaris*	PP_005	*Glycine max* leaf	Rapides (DLRS), LA, USA					
Cerc. cf. *flagellaris*	PP_007	*Glycine max* leaf	Rapides (DLRS), LA, USA					
Cerc. cf. *flagellaris*	PP_018	*Glycine max* seed	Franklin (MRRS), LA, USA					
Cerc. cf. *flagellaris*	PP_020	*Glycine max* seed	Rapides (DLRS), LA, USA					
Cerc. cf. *flagellaris*	PP_030	*Glycine max*	Rapides (DLRS), LA, USA					
Cerc. cf. *flagellaris*	PP_032	*Glycine max* leaf	Rapides (DLRS), LA, USA					
Cerc. cf. *flagellaris*	PP_041	*Glycine max* seed	Franklin (MRRS), LA, USA					

Cerc. cf. *flagellaris*	PP_043	*Glycine max* seed	Rapides (DLRS), LA, USA
Cerc. cf. *flagellaris*	PP_050	*Glycine max* leaf	Rapides (DLRS), LA, USA
Cerc. cf. *flagellaris*	PP_051	*Glycine max* seed	Rapides (DLRS), LA, USA
Cerc. cf. *flagellaris*	PP_053	*Glycine max* seed	Rapides (DLRS), LA, USA
Cerc. cf. *flagellaris*	PP_003	*Glycine max* leaf	Catahoula, LA, USA
Cerc. cf. *flagellaris*	PP_006	*Glycine max* leaf	Catahoula, LA, USA
Cerc. cf. *flagellaris*	PP_007	*Glycine max* leaf	Catahoula, LA, USA
Cerc. cf. *flagellaris*	PP_008	*Glycine max* leaf	Catahoula, LA, USA
Cerc. cf. *flagellaris*	PP_036	*Glycine max* leaf	Jefferson Davis, LA, USA
Cerc. cf. *flagellaris*	PP_041	*Glycine max* leaf	Vermillion, LA, USA
Cerc. cf. *flagellaris*	PP_045	*Glycine max* leaf	Concordia, LA, USA
Cerc. cf. *flagellaris*	PP_046	*Glycine max* leaf	Concordia, LA, USA

Cerc. cf. *flagellaris*	PP_048	*Glycine max* leaf	Concordia, LA, USA
Cerc. cf. *flagellaris*	PP_052	*Glycine max* leaf	Franklin, LA, USA (NERS)
Cerc. cf. *flagellaris*	PP_054	*Glycine max* leaf	Rapides (DLRS), LA, USA
Cerc. cf. *flagellaris*	PP_063	*Glycine max* leaf	Concordia, LA, USA
Cerc. cf. *flagellaris*	PP_082	*Glycine max* leaf	EBR, LA, USA (BHRS)
Cerc. cf. *flagellaris*	PP_110	*Glycine max* leaf	Pointe Coupee, LA, USA
Cerc. cf. *flagellaris*	PP_154	*Glycine max* leaf	St. Martin, LA, USA
Cerc. cf. *flagellaris*	PP_010	*Glycine max* leaf	Ouachita, LA, USA
Cerc. cf. *flagellaris*	PP_012	*Glycine max* leaf	St. Landry, LA,USA
Cerc. cf. *flagellaris*	PP_017	*Glycine max* leaf	Evangeline, LA, USA
Cerc. cf. *flagellaris*	PP_032	*Glycine max* leaf	St. Landry, LA,USA
Cerc. cf. *flagellaris*	PP_041	*Glycine max* leaf	Catahoula, LA,USA

Cerc. cf. *flagellaris*	PP_054	*Glycine max* leaf	Franklin, LA, USA						
Cerc. cf. *flagellaris*	PP_059	*Glycine max* leaf	Cameron, LA, USA						
Cerc. cf. *flagellaris*	PP_080	*Glycine max* leaf	Natchitoc hes, LA, USA						
Cerc. cf. *flagellaris*	ARCK7	*Glycine max*							
Cerc. cf. *flagellaris*	PSS 13-1	*Glycine max* seed	MS, USA						
Cerc. cf. *flagellaris*	PSS 13-2a	*Glycine max* seed	MS, USA						
Cerc. cf. *flagellaris*	SA1019	*Glycine max* seed	Hayti, MO, USA						
Cerc. cf. *flagellaris*	SA1080	*Glycine max* seed	AR, USA						
Cerc. cf. *flagellaris*	SA1083	*Glycine max* seed	AR, USA						
Cerc. cf. *flagellaris*	SA1088	*Glycine max* seed	KS, USA						
Cerc. cf. *flagellaris*	CBS 113127	*Eichhornia crassipes*	TX, USA	DQ835121	DQ835148	AF146147	DQ835175	DQ835075	
Cerc. cf. *flagellaris*	CBS 115482	*Citrus* sp.	S. Africa	DQ835114	DQ835141	DQ835095	DQ835168	AY260070	DQ264744

Cerc. cf. *flagellaris*	CBS 132637	*Trachelium* sp.	Israel	JX143112	JX142866	JX143358	JX142620	JX143600
Cerc. cf. *flagellaris*	CBS 132670	*Sigesbeckia pubescens*	S. Korea	JX143117	JX142871	JX143363	JX142625	JX143605
Cerc. cf. *flagellaris*	CPC 5441	*Amaranthus* sp.	Fiji	JX143124	JX142878	JX143370	JX142632	JX143611
Cerc. cf. *flagellaris*	CBS 132646	*Cichorium intybus*	S. Korea	JX143113	JX142867	JX143359	JX142621	JX143601
Cerc. cf. *flagellaris*	CBS 132648	*Amaranthus patulus*	S. Korea	JX143114	JX142868	JX143360	JX142622	JX143602
Cerc. cf. *flagellaris*	CBS 132653	*Dysphania ambrosioides*	S. Korea	JX143115	JX142869	JX143361	JX142623	JX143603
Cerc. cf. *flagellaris*	CBS 132667	*Celosia argentea var. cristata*	S. Korea	JX143116	JX142870	JX143362	JX142624	JX143604
Cerc. cf. *flagellaris*	CBS 132674	*Phytolacca americana*	S. Korea	JX143118	JX142872	JX143364	JX142626	JX143606
Cerc. cf. *flagellaris*	CBS 143.51	*Bromus* sp.		JX143119	JX142873	JX143365	JX142627	JX143607
Cerc. cf. *flagellaris*	CPC 10124	*Phytolacca americana*	S. Korea	JX143120	JX142874	JX143366	JX142628	JX143608

Cerc. cf. *flagellaris*	CPC 1051	*Populus deltoides*	S. Africa	JX143121	JX142875	JX143367	JX142629	AY260069
Cerc. cf. *flagellaris*	CPC 1052	*Populus deltoides*	S. Africa	JX143122	JX142876	JX143368	JX142630	JX143609
Cerc. cf. *flagellaris*	CPC 10684	*Phytolacca americana*	S. Korea	JX143123	JX142877	JX143369	JX142631	JX143610
Cerc. cf. *flagellaris*	CPC 4411	*Citrus* sp.	S. Africa	DQ835118	DQ835145	DQ835098	DQ835172	JX143611
Cerc. cf. *flagellaris*	MUCC 127	*Amaranthus* sp.	Fiji	JX143125	JX142879	JX143371	JX142633	JX143612
Cerc. cf. *flagellaris*	MUCC 735	*Hydrangea serrata*	Japan	JX143126	JX142880	JX143372	JX142634	JX143613
Cerc. cf. *flagellaris*	MUCC 831	*Hydrangea serrata*	Japan	JX143127	JX142881	JX143373	JX142635	JX143614
Cerc. cf. *helianthicola*	MUCC 716	*Helianthus tuberosus*	Japan	JX143128	JX142882	JX143374	JX142636	JX143615
Cerc. cf. *ipomoeae*	CBS 132639	*Persicaria thunbergii*	S. Korea	JX143129	JX142883	JX143375	JX142637	JX143616
Cerc. cf. *ipomoeae*	CBS 132652	*Ipomoea nil*	S. Korea	JX143130	JX142884	JX143376	JX142638	JX143617
Cerc. cf. *ipomoeae*	MUCC 442	*Ipomoea aquatica*	Japan	JX143131	JX142885	JX143377	JX142639	JX143618
Cerc. kikuchii^{T}	CBS 128.27	*Glycine soja*	Japan	DQ835107	DQ835134	DQ835088	DQ835161	DQ835070

Cerc. kikuchii	CBS 132633	*Glycine max*	Argentina	JX143132	JX142886	JX143378	JX142640	JX143619	
Cerc. kikuchii	CBS 135.28	*Glycine soja*	Japan	DQ835108	DQ835135	DQ835089	DQ835162	DQ835071	DQ264741
Cerc. kikuchii	MUCC 590	*Glycine soja*	Japan	JX143133	JX142887	JX143379	JX142641	JX143620	
Cerc. lactucae-sativae	CBS 132604	*Ixeris chinensis* subsp. *strigosa*	S. Korea	JX143134	JX142888	JX143380	JX142642	JX143621	
Cerc. lactucae-sativae	CPC 10082	*Ixeris chinensis* subsp. *strigosa*	S. Korea	JX143135	JX142889	JX143381	JX142643	JX143622	
Cerc. lactucae-sativae	MUCC 570	*Lactuca sativa*	Japan	JX143136	JX142890	JX143382	JX142644	JX143623	
Cerc. lactucae-sativae	MUCC 571	*Lactuca sativa*	Japan	JX143137	JX142891	JX143383	JX142645	JX143624	
Cerc. cf. *malloti*	MUCC 575	*Cucumis melo*	Japan	JX143138	JX142892	JX143384	JX142646	JX143625	
Cerc. cf. *malloti*	MUCC 787	*Mallotus japonicus*	Japan	JX143139	JX142893	JX143385	JX142647	JX143626	
Cerc. mercurialis	CBS 549.71	*Mercurialis annua*	Romania	JX143140	JX142894	JX143386	JX142648	JX143627	
Cerc. mercurialis^{T}	CBS 550.71	*Mercurialis perennis*	Romania	JX143141	JX142895	JX143387	JX142649	JX143628	
Cerc. mercurialis	CBS 551.71	*Mercurialis ovata*	Romania	JX143142	JX142896	JX143388	JX142650	JX143629	
Cerc. cf. *modiolae*	CPC 5115	*Modiola caroliniana*	New Zealand	JX143143	JX142897	JX143389	JX142651	JX143630	

Cerc. cf. *nicotianae*	CBS 131.32	*Nicotiana tabacum*	Indonesia	DQ835119	DQ835146	DQ835099	DQ835173	DQ835073	
Cerc. cf. *nicotianae*	CBS 132632	*Glycine max*	Mexico	JX143144	JX142898	JX143390	JX142652	JX143631	
Cerc. cf. *nicotianae*	CBS 570.69	*Nicotiana tabacum*	Nigeria	DQ835120	DQ835147	DQ835100	DQ835174	DQ835074	
Cerc. olivascens[T]	CBS 253.67	*Aristolochia clematidis*	Romania	JX143145	JX142899	JX143391	JX142653	JX143632	
Cerc. cf. *physalidis*	CBS 765.79	*Solanum tuberosum*	Peru	JX143146	JX142900	JX143392	JX142654	JX143633	
Cerc. pileicola[T]	CBS 132607	*Pilea pumila*	S. Korea	JX143147	JX142901	JX143393	JX142655	JX143634	
Cerc. pileicola	CBS 132647	*Pilea hamaoi*	S. Korea	JX143148	JX142902	JX143394	JX142656	JX143635	
Cerc. pileicola	CPC 11369	*Pilea pumila*	S. Korea	JX143149	JX142903	JX143395	JX142657	JX143636	
Cerc. polygonacea	CBS 132614	*Persicaria longiseta*	S. Korea	JX143150	JX142904	JX143396	JX142658	JX143637	
Cerc. punctiformis	CBS 132626	*Cynanachum wilfordii*	S. Korea	JX143151	JX142905	JX143397	JX142659	JX143638	
Cerc. cf. *resedae*	CBS 118793	*Reseda odorata*	New Zealand	JX143152	JX142906	JX143398	DQ233421	JX143639	
Cerc. cf. *resedae*	CBS 257.67	*Helianthemum* sp.	Romania	DQ233369	DQ233395	DQ233343	JX142660	DQ233319	DQ264734
Cerc. cf. *richardiicola*	CBS 132627	*Ajuga multiflora*	S. Korea	JX143153	JX142907	JX143399	JX142661	JX143640	

Cerc. cf. *richardiicola*	MUCC 128	*Tagetes erecta*	Japan	JX143154	JX142908	JX143400	JX142662	JX143641
Cerc. cf. *richardiicola*	MUCC 132	*Osteospermum* sp.	Japan	JX143155	JX142909	JX143401	JX142663	JX143642
Cerc. cf. *richardiicola*	MUCC 138	*Fuchsia* × *hybrida*	Japan	JX143156	JX142910	JX143402	JX142664	JX143643
Cerc. cf. *richardiicola*	MUCC 578	*Zantedeschia* sp.	Japan	JX143157	JX142911	JX143403	JX142665	JX143644
Cerc. cf. *richardiicola*	MUCC 582	*Gerbera hybrida*	Japan	JX143158	JX142912	JX143404	JX142666	JX143645
Cerc. ricinella	CBS 132605	*Ricinus communis*	S. Korea	JX143159	JX142913	JX143405	JX142667	JX143646
Cerc. ricinella	CPC 10104	*Ricinus communis*	S. Korea	JX143160	JX142914	JX143406	JX142668	JX143647
Cerc. rodmanii	CBS 113123	*Eichhornia crassipes*	Brazil	DQ835122	DQ835149	AF146136	DQ835176	DQ835076
Cerc. rodmanii	CBS 113124	*Eichhornia crassipes*	Mexico	DQ835123	DQ835150	AF146137	DQ835177	DQ835077
Cerc. rodmanii	CBS 113125	*Eichhornia crassipes*	Zambia	DQ835124	DQ835151	AF146146	DQ835178	DQ835078
Cerc. rodmanii	CBS 113126	*Eichhornia crassipes*	Brazil	DQ835125	DQ835152	AF146138	DQ835179	DQ835079
Cerc. rodmanii	CBS 113128	*Eichhornia crassipes*	FL, USA	DQ835126	DQ835153	AF146142	DQ835180	DQ835080

Cerc. rodmanii	CBS 113129	*Eichhornia crassipes*	FL, USA	DQ835127	DQ835154	AF146143	DQ835181	DQ835081	
Cerc. rodmanii	CBS 113130	*Eichhornia crassipes*	FL, USA	DQ835128	DQ835155	AF146144	DQ835182	DQ835082	
Cerc. rodmanii	CBS 113131	*Eichhornia crassipes*	Venezuela	DQ835129	DQ835156	AF146148	DQ835183	DQ835083	
Cerc. rumicis	CPC 5439	*Rumex sanguineus*	New Zealand	JX143161	JX142915	JX143407	JX142669	JX143648	
Cerc. senecionis-walkeri	CBS 132636	*Senecio walkeri*	Laos	JX143162	JX142916	JX143408	JX142670	JX143649	
Cerc. cf. *sigesbeckiae*	PP_003	*Glycine max* leaf	Pointe Coupee, LA, USA						
Cerc. cf. *sigesbeckiae*	CBS 132601	*Sigesbeckia glabrescens*	S. Korea	JX143163	JX142917	JX143409	JX142671	JX143650	
Cerc. cf. *sigesbeckiae*	CBS 132606	*Paulownia coreana*	S. Korea	JX143164	JX142918	JX143410	JX142672	JX143651	
Cerc. cf. *sigesbeckiae*	CBS 132621	*Sigesbeckia pubescens*	S. Korea	JX143165	JX142919	JX143411	JX142673	JX143652	
Cerc. cf. *sigesbeckiae*	CBS 132641	*Persicaria orientalis*	S. Korea	JX143166	JX142920	JX143412	JX142674	JX143653	DQ264745
Cerc. cf. *sigesbeckiae*	CBS 132642	*Pilea pumila*	S. Korea	JX143167	JX142921	JX143413	JX142675	JX143654	
Cerc. cf. *sigesbeckiae*	CBS 132675	*Malva verticillata*	S. Korea	JX143168	JX142922	JX143414	JX142676	JX143655	

Cerc. cf. *sigesbeckiae*	MUCC 587	*Begonia* sp.	Japan	JX143169	JX142923	JX143415	JX142677	JX143656	
Cerc. cf. *sigesbeckiae*	MUCC 589	*Glycine max*	Japan	JX143170	JX142924	JX143416	JX142678	JX143657	
Cerc. cf. *sigesbeckiae*	MUCC 849	*Dioscorea tokoro*	Japan	JX143171	JX142925	JX143417	JX142679	JX143658	
Cerc. sojina	CBS 132018	*Glycine soja*	S. Korea	JX143172	JX142926	JX143418	JX142680	GU214655	JX142680
Cerc. sojina^{T}	CBS 132615	*Glycine soja*	S. Korea	JX143173	JX142927	JX143419	JX142681	JX143659	JX142681
Cerc. sojina	CBS 132684	*Glycine max*	Argentina	JX143174	JX142928	JX143420	JX142682	JX143660	
Cerc. sojina	CPC 11420	*Glycine soja*	S. Korea	JX143175	JX142929	JX143421	JX142683	JX143661	
Cerc. sojina	CPC 17969	*Glycine max*	Argentina	JX143181	JX142935	JX143427	JX142689	JX143667	
Cerc. sojina	CPC 17970	*Glycine max*	Argentina	JX143182	JX142936	JX143428	JX142690	JX143668	
Cerc. sojina	CPC 17972	*Glycine max*	Argentina	JX143183	JX142937	JX143429	JX142691	JX143669	
Cerc. sp. *P*	CBS 112649	*Citrus* sp., leaf spot	Swaziland	DQ835109	DQ835136	DQ835090	DQ835163	AY260072	
Cerc. sp. P^{T}	CPC 10526	*Acacia mangium*	Thailand	AY752204	AY752235	AY752176	AY752266	AY752141	
Cerc. sp. P	CBS 132660	*Dioscorea rotundata*	Ghana	JX143218	JX142972	JX143464	JX142726	JX143704	
Cerc. sp. P	CBS 132680	*Ricinus communis*	Mexico	JX143222	JX142976	JX143468	JX142730	JX143708	
Cerc. sp. P	CPC 5327	*Cajanus cajan*	S. Africa	JX143228	JX142982	JX143474	JX142736	JX143715	

Cerc. vignigena[T]	CBS 132611	*Vigna unguiculata*	S. Korea	JX143247	JX143001	JX143493	JX142755	JX143734	
Cerc. vignigena	CPC 1134	*Vigna unguiculata*	S. Africa	JX143248	JX143002	JX143494	JX142756	JX143735	
Cerc. vignigena	MUCC 579	*Vigna unguiculata*	Japan	JX143249	JX143003	JX143495	JX142757	JX143736	
Cerc. violae[T]	CBS 251.67	*Viola tricolor*	Romania	JX143250	JX143004	JX143496	JX142758	JX143737	DQ264746
Cerc. violae	CPC 5368	*Viola odorata*	New Zealand	JX143251	JX143005	JX143497	JX142759	JX143738	
Cerc. violae	MUCC 129	*Viola* sp.	Japan	JX143252	JX143006	JX143498	JX142760	JX143739	
Cerc. violae	MUCC 133	*Viola tricolor*	Japan	JX143253	JX143007	JX143499	JX142761	JX143740	
Cerc. violae	MUCC 136	*Viola tricolor*	Japan	JX143254	JX143008	JX143500	JX142762	JX143741	
Cerc. zeae-maydis	CBS 117755	*Zea mays*	IN, USA	DQ185096	DQ185108	DQ185084	DQ185120	DQ185072	
Cerc. zeae-maydis[T]	CBS 117757	*Zea mays*	WI, USA	DQ185098	DQ185110	DQ185086	DQ185122	DQ185074	
Cerc. zeae-maydis	CBS 117758	*Zea mays*	IA, USA	DQ185099	DQ185111	DQ185087	DQ185123	DQ185075	DQ264747
Cerc. zeae-maydis	CBS 117760	*Zea mays*	PA, USA	DQ185101	DQ185113	DQ185089	DQ185125	DQ185077	
Cerc. zeae-maydis	CBS 132668	*Zea mays*	China	JX143255	JX143009	JX143501	JX142763	JX143742	

Cerc. zeae-maydis	CBS 132678	*Zea mays*	Mexico	JX143256	JX143010	JX143502	JX142764	JX143743	
Cerc. zeae-maydis	CZM_ SCOH	*Zea mays*							
Cerc. zebrina	CBS 108.22	*Medicago arabica*		JX143257	JX143011	JX143503	JX142765	JX143744	
Cerc. zebrina	CBS 112736	*Trifolium repens*	Canada	JX143259	JX143013	JX143505	JX142767	AY260080	
Cerc. zebrina	CBS 114359	*Hebe* sp.	New Zealand	JX143262	JX143016	JX143508	JX142770	JX143746	
Cerc. zebrina	CBS 537.71	*Astragalus spruneri*	Romania	JX143269	JX143023	JX143515	JX142777	JX143753	
Cerc. zeina^{T}	CPC 11995	*Zea mays*	S. Africa	DQ185105	DQ185117	DQ185093	DQ185129	DQ185081	
Cerc. zeina	CPC 11998	*Zea mays*	S. Africa	DQ185106	DQ185118	DQ185094	DQ185130	DQ185082	DQ264748
Cerc. zeina	Cerc. sp. O	*Zea mays*							
Cerc. cf. *zinniae*	CBS 132624	*Zinnia elegans*	S. Korea	JX143272	JX143026	JX143518	JX142780	JX143756	
Cerc. cf. *zinniae*	CBS 132676		Brazil	JX143273	JX143027	JX143519	JX142781	JX143757	
Cerc. cf. *zinniae*	MUCC 131	*Zinnia elegans*	Japan	JX143274	JX143028	JX143520	JX142782	JX143758	
Cerc. cf. *zinniae*	MUCC 572	*Zinnia elegans*	Japan	JX143275	JX143029	JX143521	JX142783	JX143759	
Cl. herbarum	CPC 12181	*Hordeum vulgare*	Holland	EF679520	EF679596	EF679444	EF679674	EF679367	
Cl. cf. *subtilissimum*	CPC 12484	*Pinus ponderosa*	Argentina	EF679548	EF679624	EF679472	EF679702	EF679394	
Myc. colombiensis	CBS 110967	*Eucalyptus urophylla*	Colombia	AY752209	AY752240	AY217109	AY752271	AY752147	

Cerc. indicates *Cercospora*

Cl. indicates *Cladosporium*

Myc. indicates *Mycosphaerella*

[T] Superscript indicates type strain

APPENDIX 2. HAPLOTYPES OF CERCOSPORA ISOLATES

Table A2.1. AHI and mating type haplotype information for *Cercospora* isolates collected during in this study

Voucher	Collection Year	ID	AHI Haplotype	MAT1-1 Haplotype	MAT1-2 Haplotype
PP_001	2000	*C.* cf. *flagellaris*	1	3	
PP_002	2000	*C.* cf. *flagellaris*	1		
PP_003	2000	*C.* cf. *flagellaris*	2		1
PP_009	2000	*C.* cf. *flagellaris*	2		1
PP_011	2000	*C.* cf. *flagellaris*	2		1
PP_012	2000	*C.* cf. *flagellaris*	1	3	
PP_013	2000	*C.* cf. *flagellaris*	2		1
PP_014	2000	*C.* cf. *flagellaris*	1	3	
PP_016	2000	*C.* cf. *flagellaris*	2		1
PP_017	2000	*C.* cf. *flagellaris*	1		1
PP_022	2000	*C.* cf. *flagellaris*	1		1
PP_023	2000	*C.* cf. *flagellaris*	1		1
PP_026	2000	*C.* cf. *flagellaris*	2		1
PP_027	2000	*C.* cf. *flagellaris*	2	3	
PP_033	2000	*C.* cf. *flagellaris*	2		1
PP_035	2000	*C.* cf. *flagellaris*	1		1
PP_037	2000	*C.* cf. *flagellaris*	1		1
PP_039	2000	*C.* cf. *flagellaris*			1
PP_040	2000	*C.* cf. *flagellaris*	2		1
PP_041	2000	*C.* cf. *flagellaris*	2		6
PP_042	2000	*C.* cf. *flagellaris*	2	3	
PP_044	2000	*C.* cf. *flagellaris*	2	3	
PP_045	2000	*C.* cf. *flagellaris*	2		1
PP_054	2000	*C.* cf. *flagellaris*	2		1
PP_055	2000	*C.* cf. *flagellaris*	2		1

PP_056	2000	*C.* cf. *flagellaris*	2			1
PP_057	2000	*C.* cf. *flagellaris*	2	3		
PP_004	2000	*C.* cf. *flagellaris*	3			1
PP_005	2000	*C.* cf. *flagellaris*	4			1
PP_007	2000	*C.* cf. *flagellaris*	5			1
PP_008	2000	*C.* cf. *flagellaris*	5			1
PP_015	2000	*C.* cf. *flagellaris*	6			3
PP_018	2000	*C.* cf. *flagellaris*	3			1
PP_020	2000	*C.* cf. *flagellaris*	7			1
PP_021	2000	*C.* cf. *flagellaris*	5	8		
PP_025	2000	*C.* cf. *flagellaris*	8			3
PP_028	2000	*C.* cf. *flagellaris*	9	3		
PP_030	2000	*C.* cf. *flagellaris*	9	3		
PP_032	2000	*C.* cf. *flagellaris*	2			1
PP_043	2000	*C.* cf. *flagellaris*	11			
PP_047	2000	*C.* cf. *flagellaris*	6			
PP_048	2000	*C.* cf. *flagellaris*	12			
PP_050	2000	*C.* cf. *flagellaris*	3			
PP_051	2000	*C.* cf. *flagellaris*	13			1
PP_053	2000	*C.* cf. *flagellaris*	8	7		
PP_058	2000	*C.* cf. *flagellaris*	14			6
PP_059	2000	*C.* cf. *flagellaris*	9			1
PP_060	2000	*C.* cf. *flagellaris*	2			1
PP_076	2000	*C.* cf. *flagellaris*	2			1
PP_084	2011	*C.* cf. *flagellaris*	2			1
PP_086	2011	*C.* cf. *flagellaris*	14			6
PP_087	2011	*C.* cf. *flagellaris*	2			
PP_095	2011	*C.* cf. *flagellaris*	16			3
PP_099	2011	*C.* cf. *flagellaris*	6			3
PP_100	2011	*C.* cf. *flagellaris*	6			3
PP_107	2011	*C.* cf. *flagellaris*	6			3

PP_108	2011	*C.* cf. *flagellaris*		3	
PP_109	2011	*C.* cf. *flagellaris*	2		1
PP_118	2011	*C.* cf. *flagellaris*	17		3
PP_119	2011	*C.* cf. *flagellaris*	6		3
PP_142	2011	*C.* cf. *flagellaris*	14	5	
PP_127	2011	*C.* cf. *flagellaris*	1	3	
PP_128	2011	*C.* cf. *flagellaris*	2		1
PP_002	2011	*C.* cf. *flagellaris*	2	3	
PP_003	2011	*C.* cf. *flagellaris*	6	5	
PP_004	2011	*C.* cf. *flagellaris*	2	3	
PP_005	2011	*C.* cf. *flagellaris*	2		1
PP_006	2011	*C.* cf. *flagellaris*	2	2	
PP_007	2011	*C.* cf. *flagellaris*	2	3	
PP_008	2011	*C.* cf. *flagellaris*	11	3	
PP_010	2011	*C.* cf. *flagellaris*		3	
PP_012	2011	*C.* cf. *flagellaris*	15		3
PP_013	2011	*C.* cf. *flagellaris*	7		1
PP_055	2011	*C.* cf. *flagellaris*	15		
PP_045	2011	*C.* cf. *flagellaris*	12	3	
PP_046	2011	*C.* cf. *flagellaris*	17	4	
PP_048	2011	*C.* cf. *flagellaris*	29	1	
PP_057	2011	*C.* cf. *flagellaris*	2		1
PP_058	2011	*C.* cf. *flagellaris*	2		1
PP_059	2011	*C.* cf. *flagellaris*		3	
PP_060	2011	*C.* cf. *flagellaris*	2		1
PP_063	2011	*C.* cf. *flagellaris*	21	3	
PP_067	2011	*C.* cf. *flagellaris*	2		1
PP_068	2011	*C.* cf. *flagellaris*	16	5	
PP_069	2011	*C.* cf. *flagellaris*	16		6
PP_071	2011	*C.* cf. *flagellaris*	2		1
PP_074	2011	*C.* cf. *flagellaris*	16	3	

PP_078	2011	*C.* cf. *flagellaris*	2		1
PP_080	2011	*C.* cf. *flagellaris*	14	5	
PP_144	2011	*C.* cf. *flagellaris*	2		
PP_145	2011	*C.* cf. *flagellaris*	15		
PP_146	2011	*C.* cf. *flagellaris*	2		1
PP_147	2011	*C.* cf. *flagellaris*	15		2
PP_082	2011	*C.* cf. *flagellaris*	22		2
PP_131	2011	*C.* cf. *flagellaris*	14		7
PP_132	2011	*C.* cf. *flagellaris*	12	8	
PP_051	2011	*C.* cf. *flagellaris*	10		1
PP_052	2011	*C.* cf. *flagellaris*	19	9	
PP_036	2011	*C.* cf. *flagellaris*	10		1
PP_017	2011	*C.* cf. *flagellaris*	14	5	
PP_021	2011	*C.* cf. *flagellaris*	6		3
PP_024	2011	*C.* cf. *flagellaris*	1	3	
PP_083	2011	*C.* cf. *flagellaris*	6	5	
PP_124	2011	*C.* cf. *flagellaris*	14	3	
PP_125	2011	*C.* cf. *flagellaris*	15	3	
PP_126	2011	*C.* cf. *flagellaris*	20		7
PP_129	2011	*C.* cf. *flagellaris*	2		1
PP_130	2011	*C.* cf. *flagellaris*	2		
PP_103	2011	*C.* cf. *flagellaris*	2		1
PP_104	2011	*C.* cf. *flagellaris*	6		3
PP_105	2011	*C.* cf. *flagellaris*	14		3
PP_106	2011	*C.* cf. *flagellaris*	14	5	
PP_110	2011	*C.* cf. *flagellaris*	24		3
PP_113	2011	*C.* cf. *flagellaris*	2		1
PP_114	2011	*C.* cf. *flagellaris*	2		1
PP_115	2011	*C.* cf. *flagellaris*	2		1
PP_090	2011	*C.* cf. *flagellaris*	15	3	
PP_091	2011	*C.* cf. *flagellaris*	6		3

PP_096	2011	*C.* cf. *flagellaris*	17		6
PP_098	2011	*C.* cf. *flagellaris*	2		1
PP_027	2011	*C.* cf. *flagellaris*	1	3	
PP_029	2011	*C.* cf. *flagellaris*	2	3	
PP_030	2011	*C.* cf. *flagellaris*	1		1
PP_053	2011	*C.* cf. *flagellaris*	30		1
PP_054*	2011	*C.* cf. *flagellaris*	20		7
PP_040	2011	*C.* cf. *flagellaris*	17		3
PP_093	2011	*C.* cf. *flagellaris*	23		3
PP_094	2011	*C.* cf. *flagellaris*	14		6
PP_116	2011	*C.* cf. *flagellaris*	2	3	
PP_117	2011	*C.* cf. *flagellaris*	16		7
PP_120	2011	*C.* cf. *flagellaris*	14		3
PP_121	2011	*C.* cf. *flagellaris*	14	5	
PP_122	2011	*C.* cf. *flagellaris*	25	3	
PP_123	2011	*C.* cf. *flagellaris*			6
PP_155	2011	*C.* cf. *flagellaris*	6	3	
PP_156b	2011	*C.* cf. *flagellaris*			6
PP_157	2011	*C.* cf. *flagellaris*	6		3
PP_158	2011	*C.* cf. *flagellaris*	6		3
PP_133	2011	*C.* cf. *flagellaris*	16		3
PP_134	2011	*C.* cf. *flagellaris*	15		3
PP_135	2011	*C.* cf. *flagellaris*	16	5	
PP_136	2011	*C.* cf. *flagellaris*	6	3	
PP_137	2011	*C.* cf. *flagellaris*	6		6
PP_138	2011	*C.* cf. *flagellaris*	6	3	
PP_139	2011	*C.* cf. *flagellaris*	12		1
PP_143	2011	*C.* cf. *flagellaris*	16	3	
PP_150	2011	*C.* cf. *flagellaris*	15		6
PP_151	2011	*C.* cf. *flagellaris*	11	3	
PP_152	2011	*C.* cf. *flagellaris*		3	

PP_154	2011	*C.* cf. *flagellaris*	26			3
PP_037	2011	*C.* cf. *flagellaris*	6			3
PP_038	2011	*C.* cf. *flagellaris*	16	3		
PP_041	2011	*C.* cf. *flagellaris*	18			4
PP_148	2011	*C.* cf. *flagellaris*	2	3		
PP_149	2011	*C.* cf. *flagellaris*	2			1
PP_111	2011	*C.* cf. *flagellaris*	14	5		
PP_112	2011	*C.* cf. *flagellaris*	9	3		
PP_059	2012	*C.* cf. *flagellaris*	15	5		
PP_041	2012	*C.* cf. *flagellaris*	14	5		
PP_042	2012	*C.* cf. *flagellaris*	28	3		
PP_043	2012	*C.* cf. *flagellaris*	2			1
PP_044	2012	*C.* cf. *flagellaris*	2			1
PP_045	2012	*C.* cf. *flagellaris*	1			
PP_033	2012	*C.* cf. *flagellaris*	19	5		
PP_017	2012	*C.* cf. *flagellaris*	23			3
PP_054	2012	*C.* cf. *flagellaris*	2	8		
PP_064	2012	*C.* cf. *flagellaris*	2			1
PP_030	2012	*C.* cf. *flagellaris*	1			1
PP_050	2012	*C.* cf. *flagellaris*	16			1
PP_080	2012	*C.* cf. *flagellaris*	6			3
PP_081	2012	*C.* cf. *flagellaris*	1			1
PP_082	2012	*C.* cf. *flagellaris*	20			
PP_010	2012	*C.* cf. *flagellaris*	25	3		
PP_001	2012	*C.* cf. *flagellaris*	2			1
PP_013	2012	*C.* cf. *flagellaris*	12			1
PP_014	2012	*C.* cf. *flagellaris*	14	5		
PP_012	2012	*C.* cf. *flagellaris*	11	5		
PP_031	2012	*C.* cf. *flagellaris*	20	3		
PP_032	2012	*C.* cf. *flagellaris*	16	3		
PP_058	2012	*C.* cf. *flagellaris*	2			1

PP_061	2012	*C.* cf. *flagellaris*		6	
PP_061	2012	*C.* cf. *flagellaris*			6
PP_062	2012	*C.* cf. *flagellaris*	14	5	
PP_066	2012	*C.* cf. *flagellaris*	2		1
PP_069	2012	*C.* cf. *flagellaris*	2		1
PP_070	2012	*C.* cf. *flagellaris*	1		1
PP_060	2012	*C.* cf. *flagellaris*	2		1
PP_055	2012	*C.* cf. *flagellaris*	2		1
PP_073	2012	*C.* cf. *flagellaris*	21		6
HJ-1	2013	*C.* cf. *flagellaris*	2		
HJ-2	2013	*C.* cf. *flagellaris*	2		
SA1065	2014	*C.* cf. *flagellaris*	1		
SA1066	2014	*C.* cf. *flagellaris*	1		
SA1067	2014	*C.* cf. *flagellaris*	1		
SA1068	2014	*C.* cf. *flagellaris*	1		
SA1031	2014	*C.* cf. *flagellaris*	1		
SA1033	2014	*C.* cf. *flagellaris*	1		
SA1034	2014	*C.* cf. *flagellaris*	1		
SA1035	2014	*C.* cf. *flagellaris*	1		
SA1036	2014	*C.* cf. *flagellaris*	1		
SA1037	2014	*C.* cf. *flagellaris*	1		
SA1017	2014	*C.* cf. *flagellaris*	1		
SA1018	2014	*C.* cf. *flagellaris*	2	3	
SA1019	2014	*C.* cf. *flagellaris*	28		
SA1053	2014	*C.* cf. *flagellaris*	20		
SA1054	2014	*C.* cf. *flagellaris*	20		
ARCK17		*C.* cf. *flagellaris*	1	3	
ARCK7		*C.* cf. *flagellaris*	1		1
PSS 13-1		*C.* cf. *flagellaris*	8		1
PSS 13-2a		*C.* cf. *flagellaris*	12		1
PP_052	2012	*C.* cf. *sigesbeckiae*	27		5

PP_003	2012	*C.* cf. *sigesbeckiae*	27	5
PP_071	2012	*C.* cf. *sigesbeckiae*	27	5
PP_072	2012	*C.* cf. *sigesbeckiae*	27	5
PP_074	2012	*C.* cf. *sigesbeckiae*	27	5

APPENDIX 3. SUPPLEMENTAL TREES FROM CHAPTER TWO

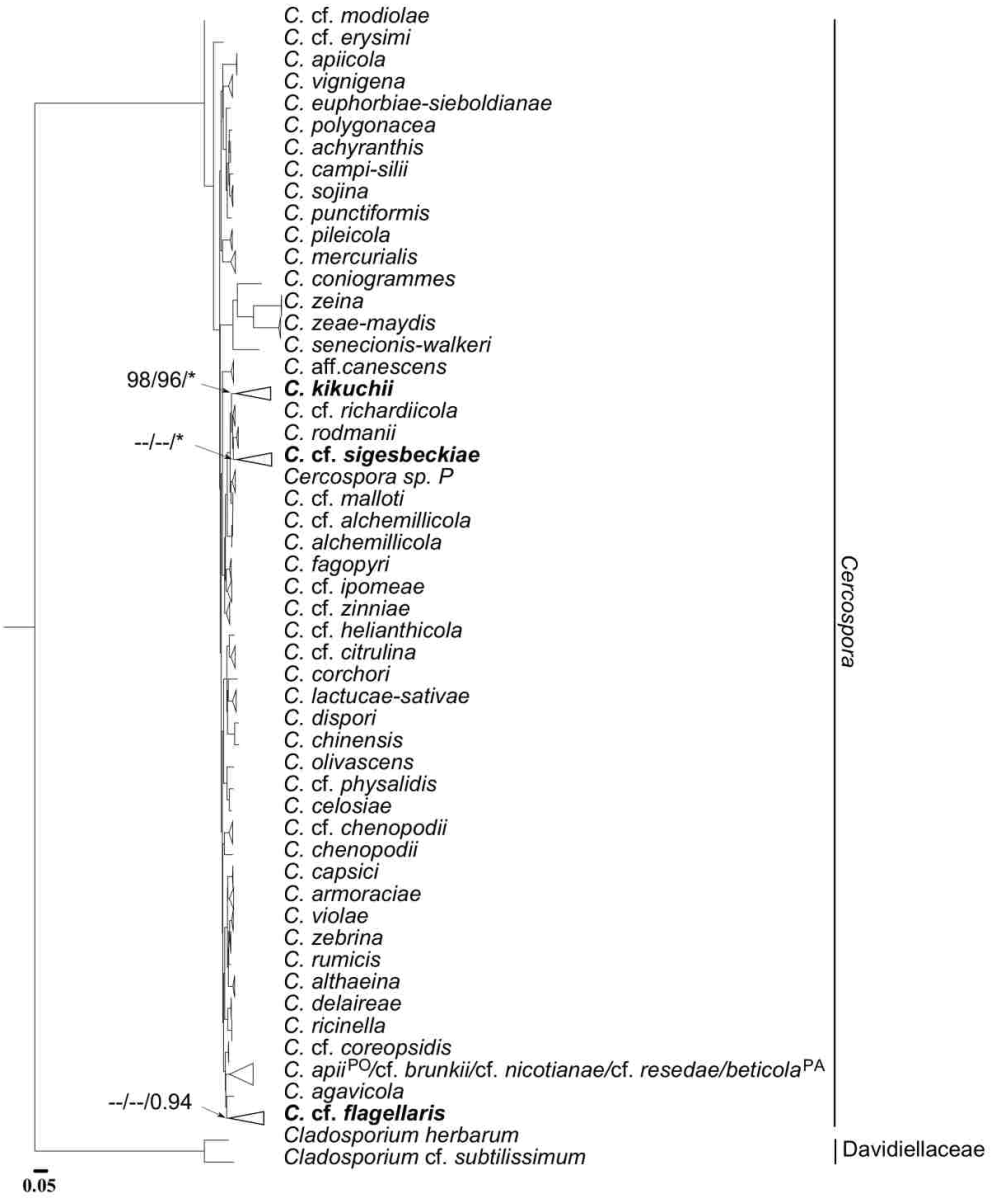


Figure A3.1. Maximum likelihood topology from RAxML analysis depicting the evolutionary relationships of 55 species of *Cerc.* based on a concatenated alignment of actin, calmodulin, translation elongation factor 1α, histone 3 and ITS sequences (DS-2). *Cerc. kikuchii*. *C.* cf. *flagellaris* and *C*. cf. *sigesbeckiae* are shown in bold. Tree is rooted with *Cladosporium* cf. *subtilissimum* and *Cl. herbarum*. Support values at nodes represent bootstrap percentages ≥70 obtained with at least 1000 replicates (RAxML/Garli) and Bayesian posterior probabilities >0.90 (on right) for *C*. cf. *sigesbeckiae*, *C. kikuchii* and *C*. cf. *flagellaris*. Asterisk indicates a posterior probability of 1. Double dashes indicate a bootstrap percentage <70. PO and PA superscript indicates taxon is polyphyletic and paraphyletic, respectively. Scale bar below tree indicates the number of substitutions per site.

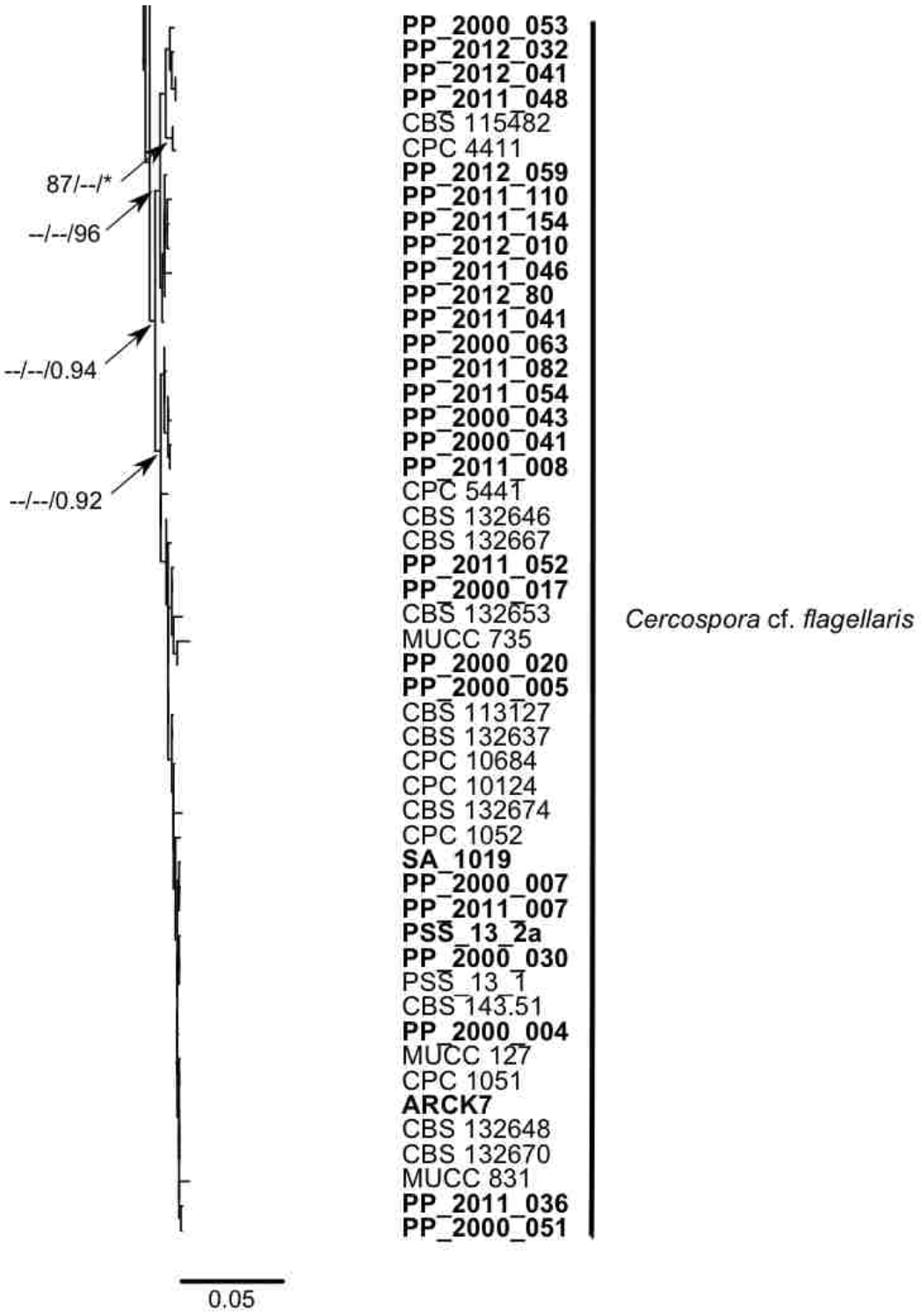


Figure A3.2. *Cercospora* cf. *flagellaris* clade from Figure A3.1 (DS-2). Isolates from this study are shown in bold. Support values at nodes represent bootstrap percentages ≥70 (RAxML/Garli) and Bayesian posterior probabilities >0.90 (on right). Asterisk indicates a posterior probability of 1. Double dashes indicate a bootstrap percentage <70. Scale bar below tree indicates the number of substitutions per site.

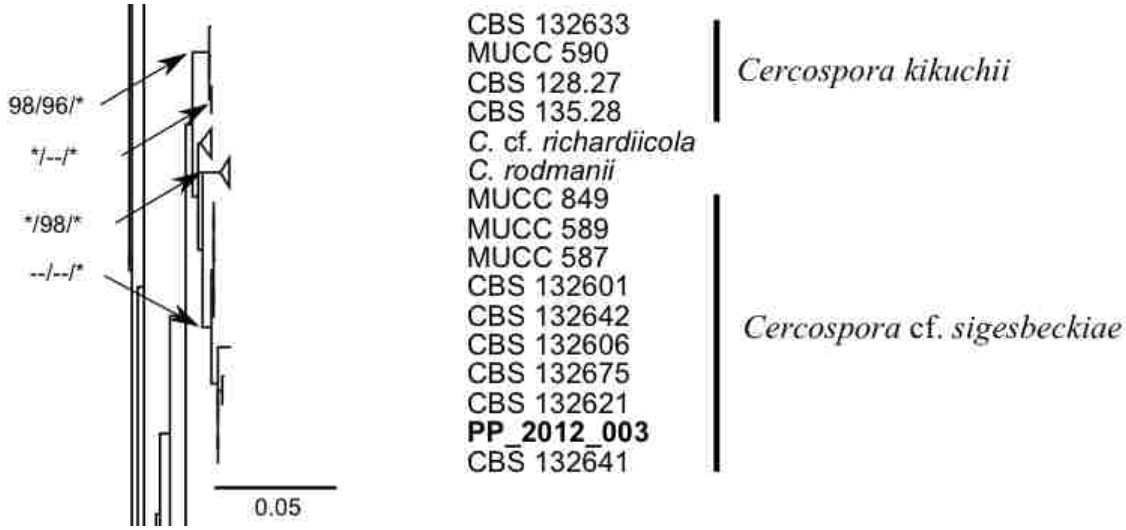


Figure A3.3. *Cercospora kikuchii*, *C.* cf. *sigesbeckiae* and closely related species from Figure A3.1 (DS-2). Isolates from this study are shown in bold. Support values at nodes represent bootstrap percentages ≥70 (RAxML/Garli) and Bayesian posterior probabilities >0.90 (on right). Asterisk indicates a posterior probability of 1. Double dashes indicate a bootstrap percentage <70. Scale bar below tree indicates the number of substitutions per site.

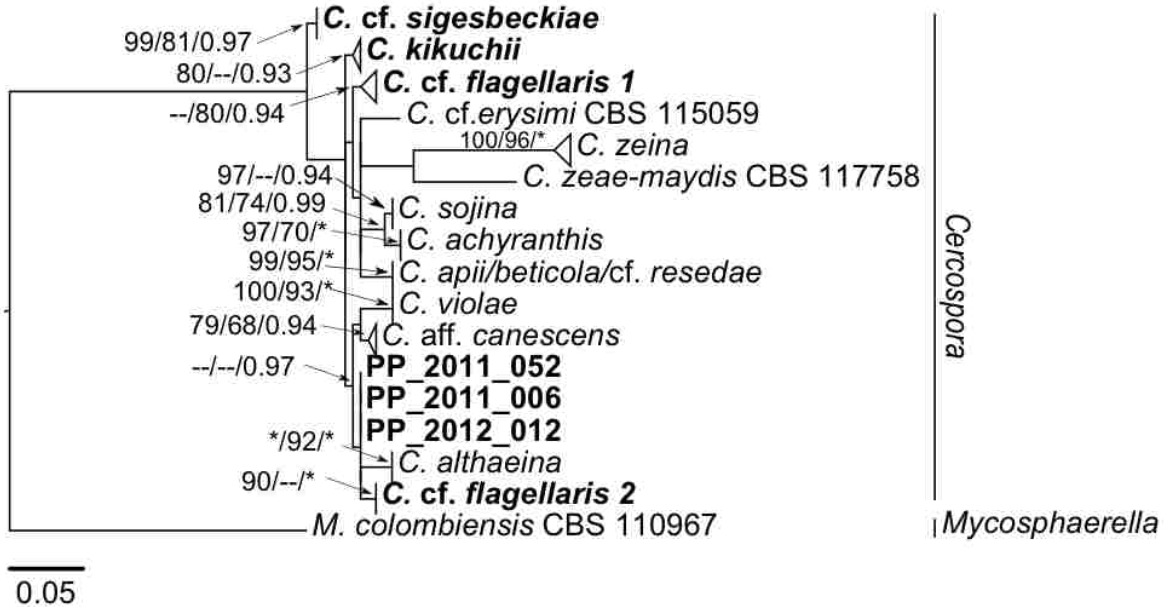


Figure A3.4. Maximum likelihood topology from RAxML analysis of actin, depicting the evolutionary relationships of 14 species of *Cercospora* (DS-3). *Cercospora kikuchii*. *C.* cf. *flagellaris*, *C*. cf. *sigesbeckiae* and isolates from this study are shown in bold. Tree is rooted with *Mycosphaerella colombiensis*. Support values at nodes represent bootstrap percentages ≥70 obtained with at least 1000 replicates (RAxML/Garli) and Bayesian posterior probabilities >0.90. (on right). Asterisk indicates a bootstrap percentage of 100 or a posterior probability of 1. Double dashes indicate a bootstrap percentage <70 or a posterior probability 0.90. Scale bar below tree indicates the number of substitutions per site.

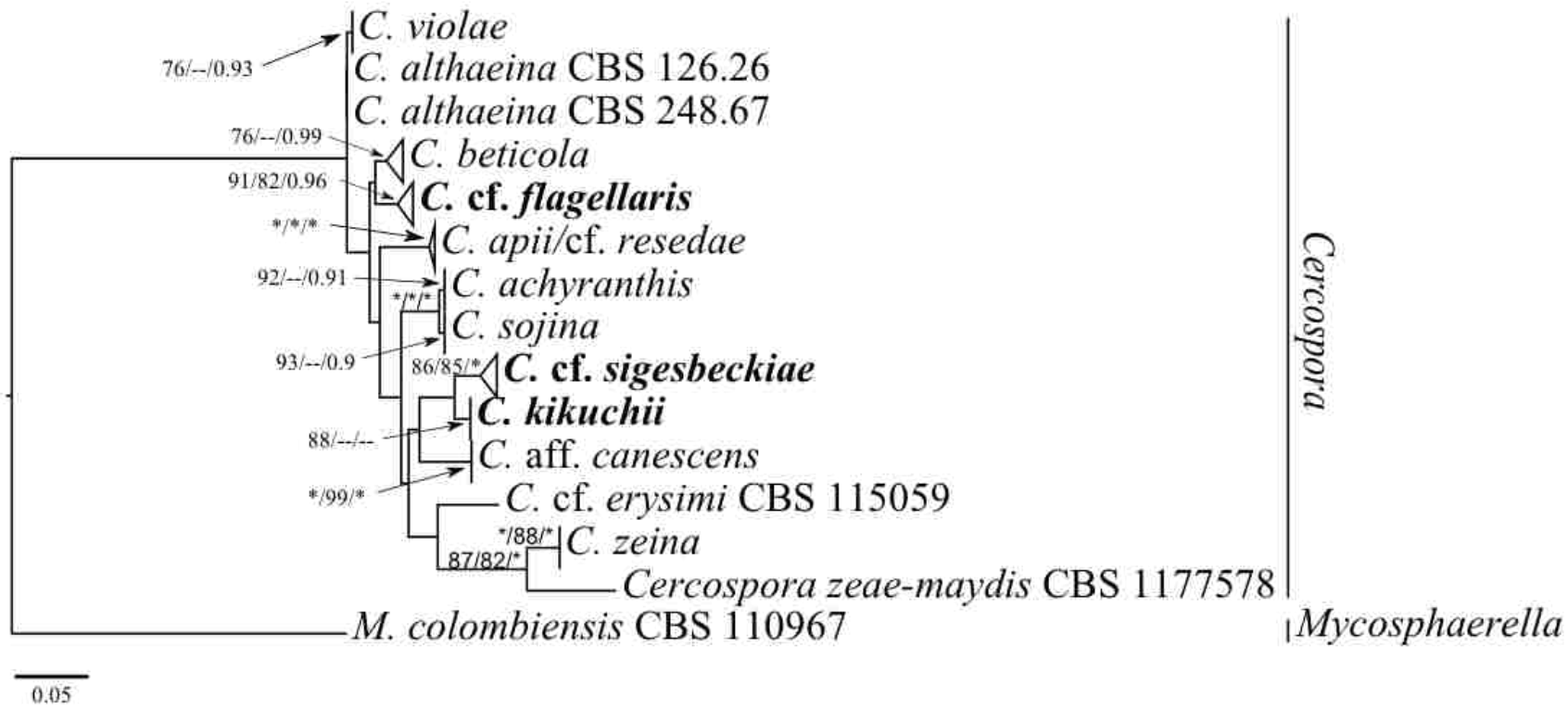


Figure A3.5. Maximum likelihood topology from RAxML analysis of calmodulin, depicting the evolutionary relationships of 14 species of *Cercospora* (DS-3). *Cercospora kikuchii*. *C.* cf. *flagellaris* and *C.* cf. *sigesbeckiae* are shown in bold. Tree is rooted with *Mycosphaerella colombiensis*. Support values at nodes represent bootstrap percentages ≥70 obtained with at least 1000 replicates (RAxML/Garli) and Bayesian posterior probabilities >0.90 (on right). Asterisk indicates a bootstrap percentage of 100 or a posterior probability of 1. Double dashes indicate a bootstrap percentage <70 or a posterior probability <0.90. Scale bar below tree indicates the number of substitutions per site.

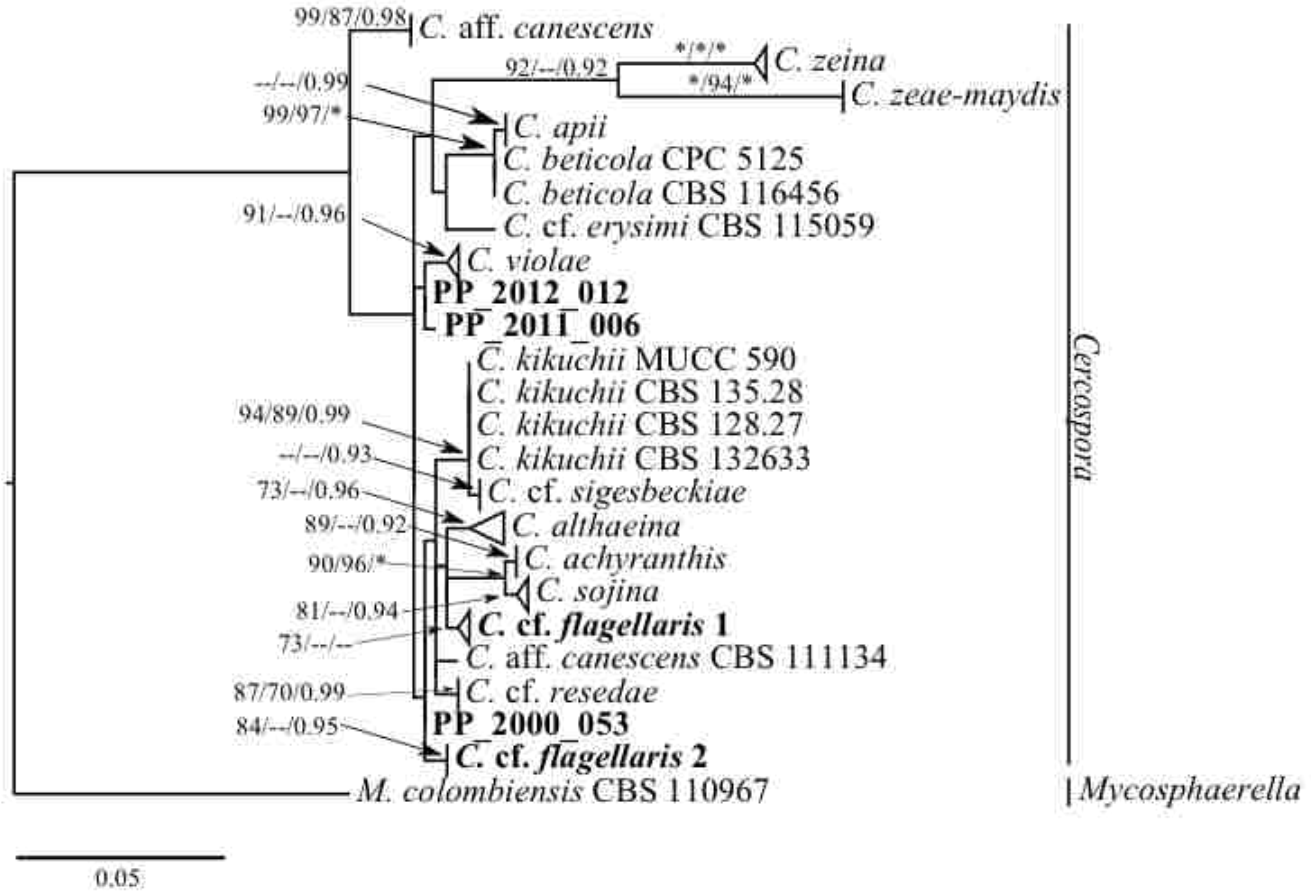


Figure A3.6. Maximum likelihood topology from RAxML analysis of histone 3, depicting the evolutionary relationships of 14 species of *Cercospora* (DS-3). *Cercospora kikuchii*, *C.* cf. *flagellaris*, *C.* cf. *sigesbeckiae* and isolates from this study are shown in bold. Tree is rooted with *Mycosphaerella colombiensis*. Support values at nodes represent bootstrap percentages ≥70 obtained with at least 1000 replicates (RAxML/ Garli) and Bayesian posterior probabilities >0.90 (on right). Asterisk indicates a bootstrap percentage of 100 or a posterior probability of 1. Double dashes indicate a bootstrap percentage <70 or a posterior probability <0.90. Scale bar below tree indicates the number of substitutions per site.

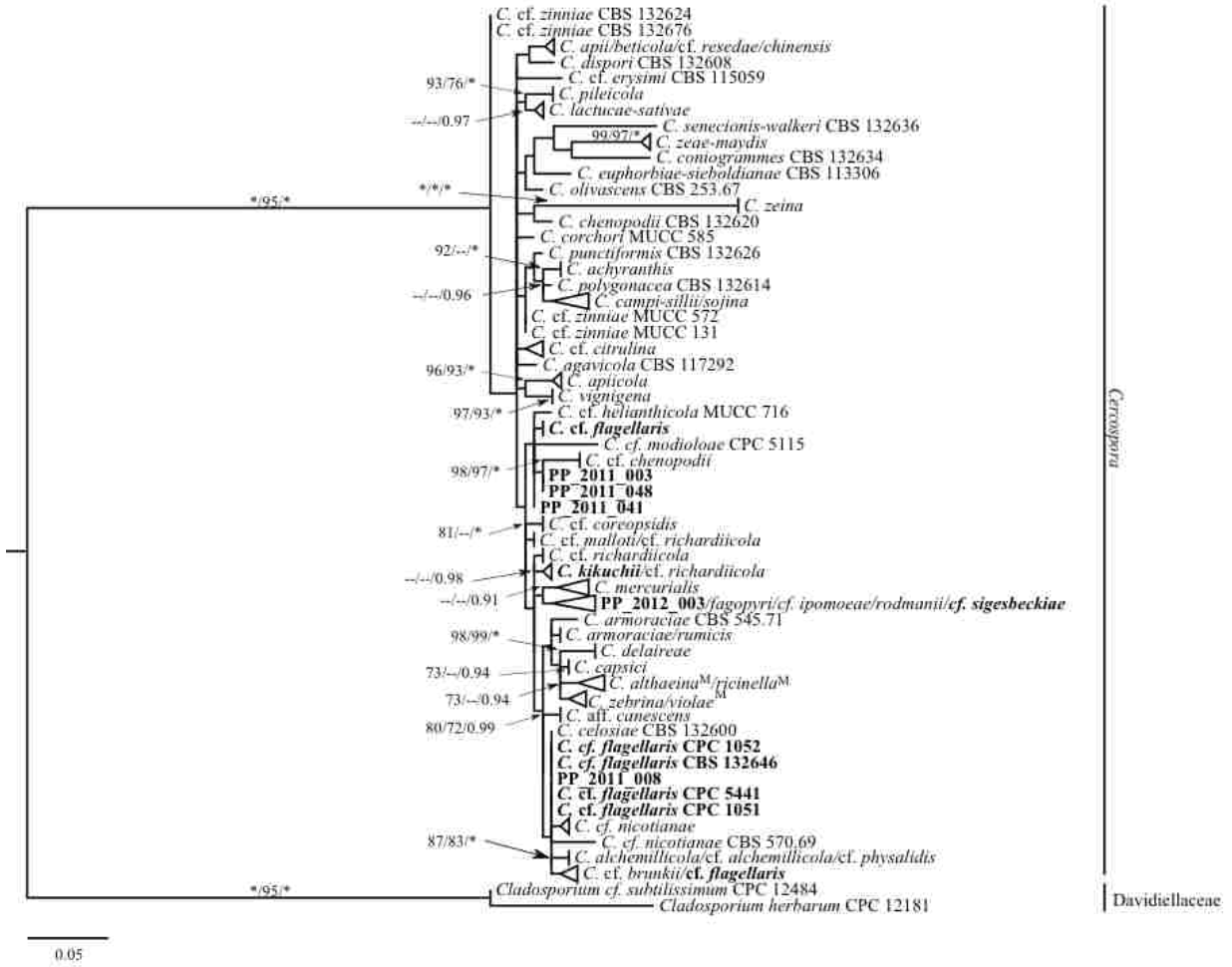


Figure A3.7. Maximum likelihood topology from RAxML analysis of actin, depicting the evolutionary relationships of 54 species of *Cercospora* (DS-1). *Cercospora kikuchii*, *Cerc.* cf. *flagellaris*, *Cerc.* cf. *sigesbeckiae* and isolates from this study are shown in bold. Tree is rooted with *Cladosporium* cf. *subtilissimum* and *Cl. herbarum*. Support values at nodes represent bootstrap percentages ≥70 obtained with at least 1000 replicates (RAxML/ Garli) and Bayesian posterior probabilities >0.90 (on right). Asterisk indicates a posterior probability of 1. Double dashes indicate a bootstrap percentage <70. M superscript indicates taxon is monophyletic. Scale bar below tree indicates the number of substitutions per site.

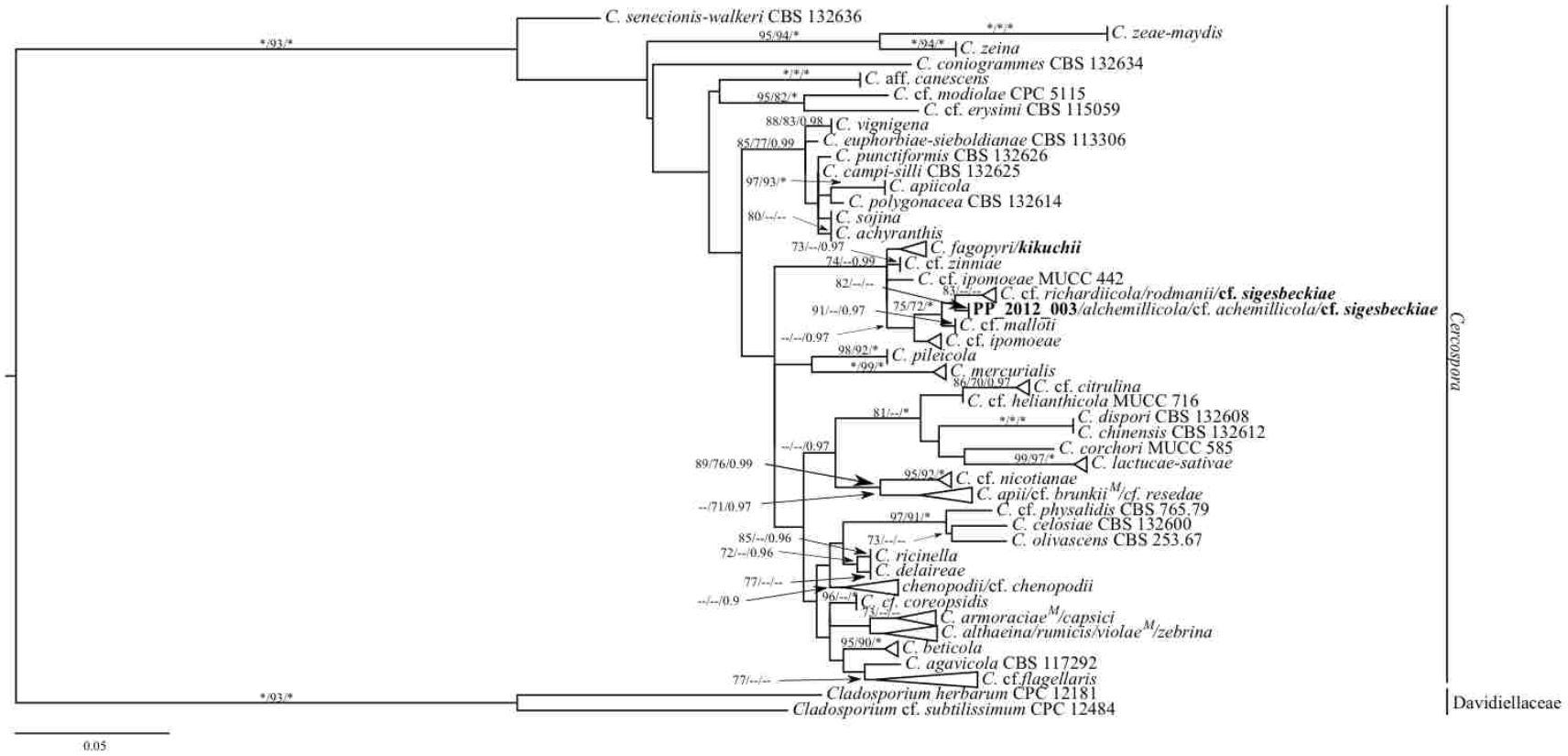


Figure A3.8. Maximum likelihood topology from RAxML analysis of calmodulin, depicting the evolutionary relationships of 54 species of *Cercospora* (DS-1). *Cercospora kikuchii*, *Cerc.* cf. *flagellaris*, *Cerc.* cf. *sigesbeckiae* and isolates from this study are shown in bold. Tree is rooted with *Cladosporium* cf. *subtilissimum* and *Cl. herbarum*. Support values beside arrows represent bootstrap percentages ≥70 obtained with at least 1000 replicates (RAxML/ Garli) and Bayesian posterior probabilities >0.90 (on right). Asterisk indicates a bootstrap percentage of 100 or a posterior probability of 1. Double dashes indicate a bootstrap percentage <70 or a posterior probability <0.90. M superscript indicates taxon is monophyletic. Scale bar below tree indicates the number of substitutions per site.

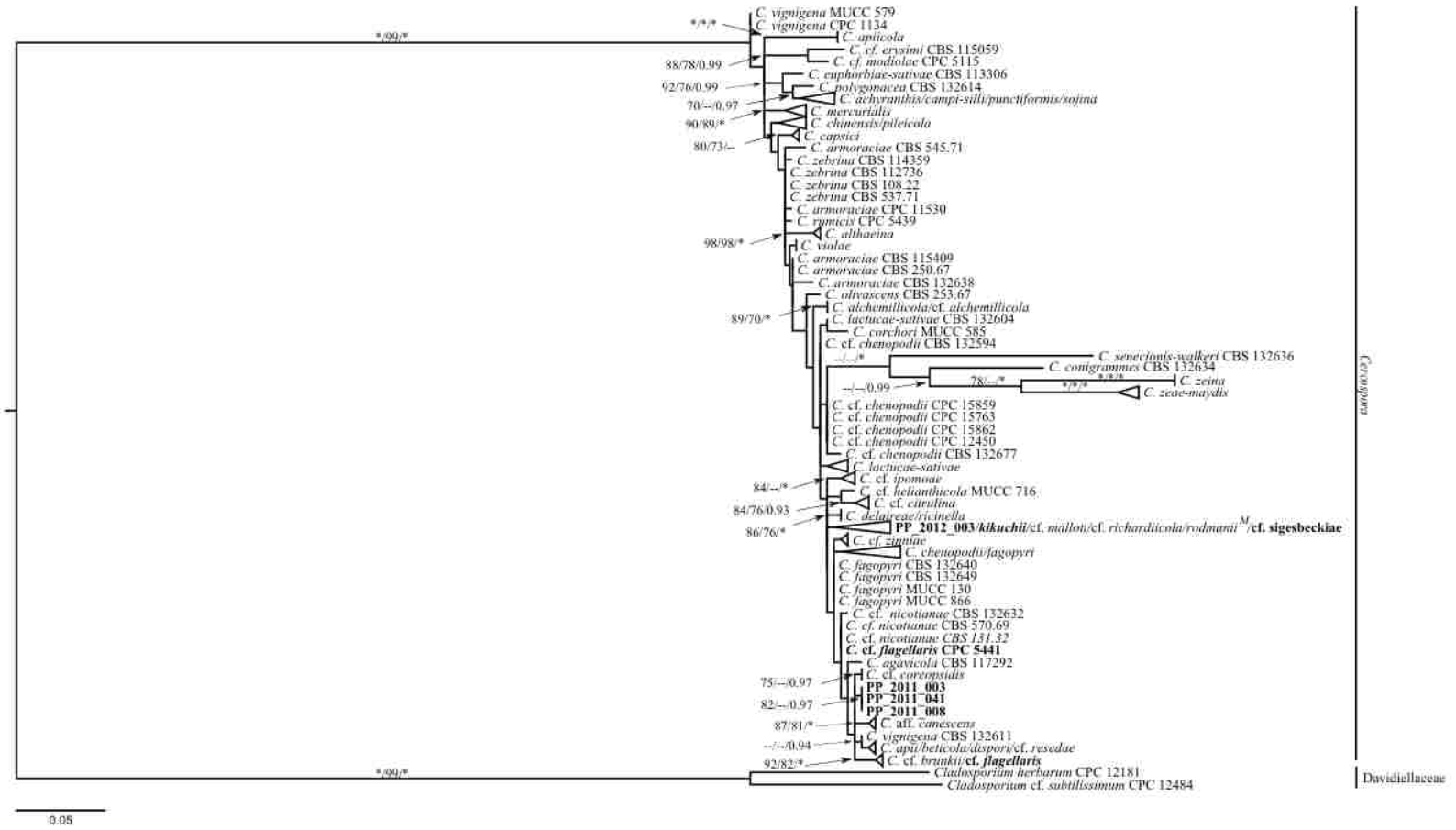


Figure A3.9. Maximum likelihood topology from RAxML analysis of translation elongation factor 1α, depicting the evolutionary relationships of 54 species of *Cercospora* (DS-1). *Cercospora kikuchii*, *Cerc.* cf. *flagellaris*, *Cerc.* cf. *sigesbeckiae* and isolates from this study are shown in bold. Tree is rooted with *Cladosporium* cf. *subtilissimum* and *Cl. herbarum*. Support values at nodes represent bootstrap percentages ≥70 obtained with at least 1000 replicates (RAxML/Garli) and Bayesian posterior probabilities >0.90 (on right). Asterisk indicates a bootstrap percentage of 100 or a posterior probability of 1. Double dashes indicate a bootstrap percentage <70. M superscript indicates taxon is monophyletic. Scale bar below tree indicates the number of substitutions per site.

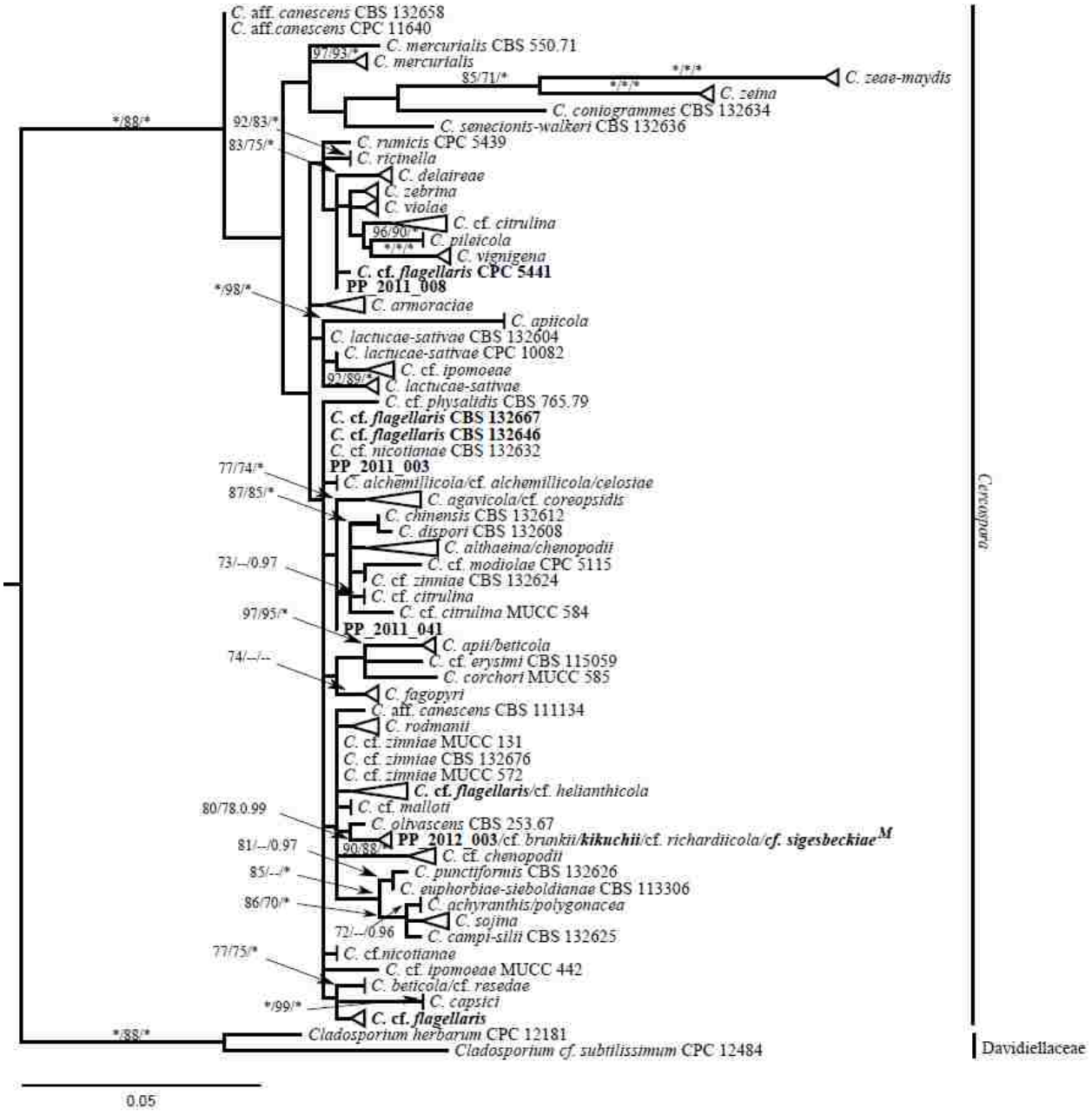


Figure A3.10. Maximum likelihood topology from RAxML analysis of histone 3, depicting the evolutionary relationships of 54 species of *Cercospora* (DS-1). ***Cercospora kikuchii***, ***Cerc.*** cf. ***flagellaris***, ***Cerc.*** cf. ***sigesbeckiae*** and isolates from this study are shown in bold. Tree is rooted with *Cladosporium* cf. *subtilissimum* and *Cl. herbarum*. Support values beside arrows represent bootstrap percentages ≥70 obtained with at least 1000 replicates (RAxML/ Garli) and Bayesian posterior probabilities >0.90 (on right). Asterisk indicates a bootstrap percentage of 100 or a posterior probability of 1. Double dashes indicate a bootstrap percentage <70 or a posterior probability <0.90. M superscript indicates taxon is monophyletic. Scale bar below tree indicates the number of substitutions per site.

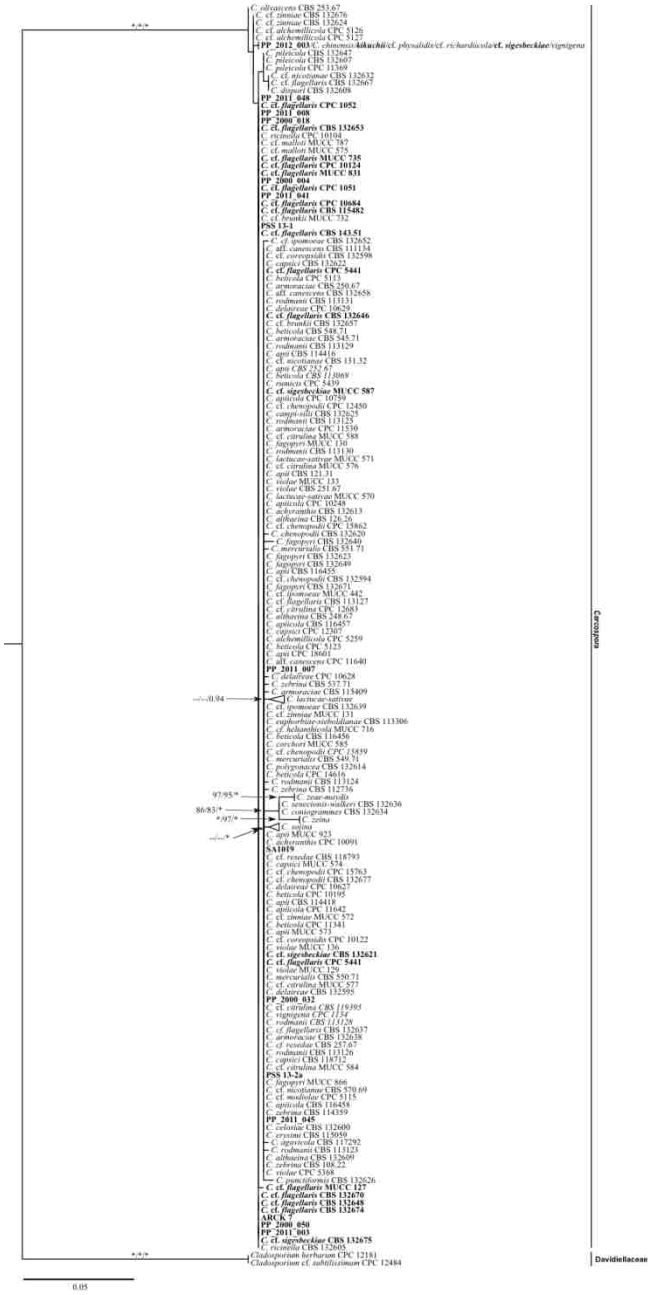


Figure A3.11. Maximum likelihood topology from RAxML analysis of ITS, depicting the evolutionary relationships of 54 species of *Cercospora* (DS-1). *Cercospora kikuchii*, *Cerc.* cf. *flagellaris*, *Cerc.* cf. *sigesbeckiae* and isolates from this study are shown in bold. Tree is rooted with *Cladosporium* cf. *subtilissimum* and *Cl. herbarum*. Support values beside arrows represent bootstrap percentages ≥70 obtained with at least 1000 replicates (RAxML/ Garli) and Bayesian posterior probabilities >0.90 (on right). Asterisk indicates a bootstrap percentage of 100 or a posterior probability of 1. Double dashes indicate a bootstrap percentage <70. Scale bar below tree indicates the number of substitutions per site.

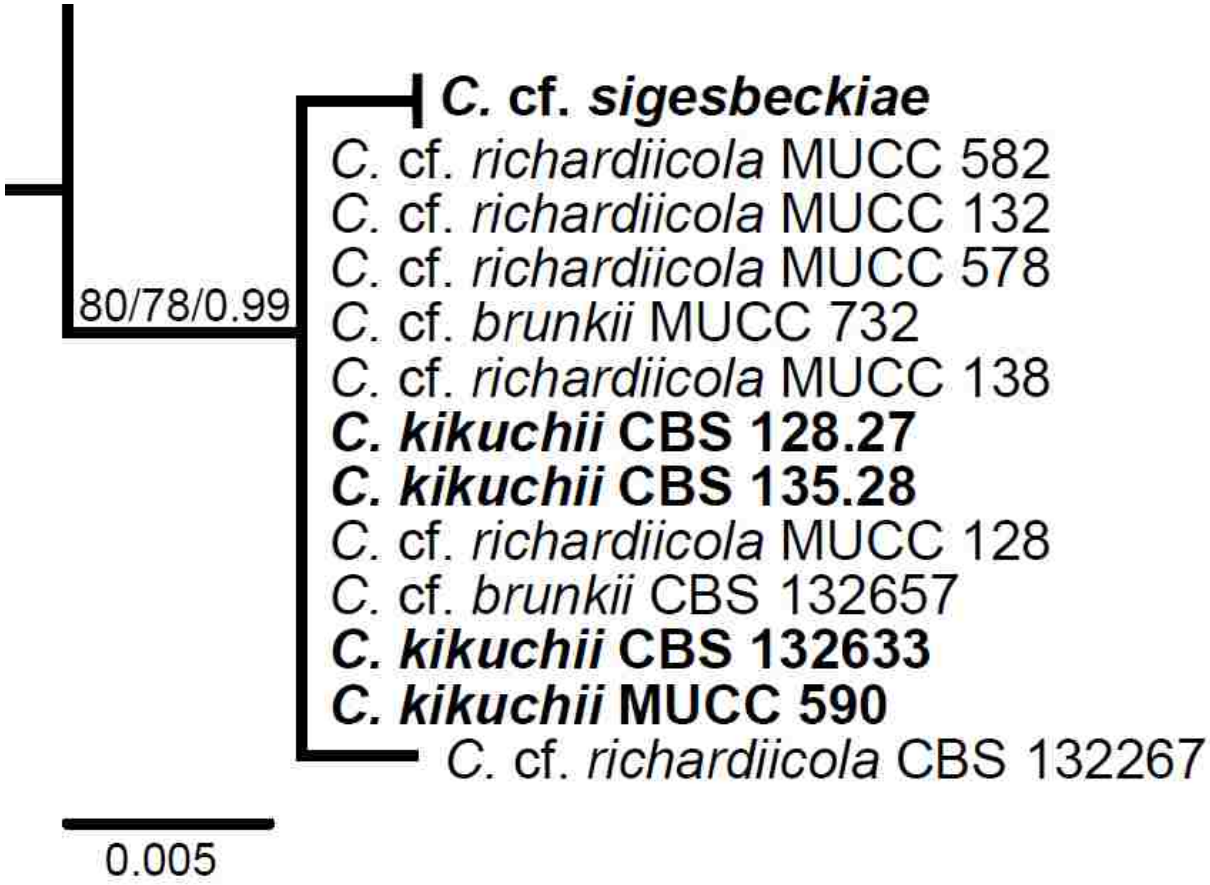


Figure A3.12. Maximum likelihood topology from RAxML analysis of histone 3, showing branch with *Cercospora kikuchii*, *C*. cf. *sigesbeckiae* and closely related species (DS-1). *Cercospora kikuchii* and *C*. cf. *sigesbeckiae* are shown in bold. Support values at node represent bootstrap percentages ≥70 (RAxML/ Garli) and Bayesian posterior probability >0.90 (on right). Scale bar below tree indicates the number of substitutions per site.

Figure A3.13. Maximum likelihood topology from RAxML analysis of histone 3, depicting the evolutionary relationships of 55 species of *Cercospora* (DS-2). *Cercospora kikuchii*, *C.* cf. *flagellaris*, *C*. cf. *sigesbeckiae* and isolates from this study are shown in bold. Tree is rooted with *Cladosporium* cf. *subtilissimum* and *Cl. herbarum*. Support values at nodes represent RAxML bootstrap percentages ≥70. Scale bar below tree indicates the number of substitutions per site.

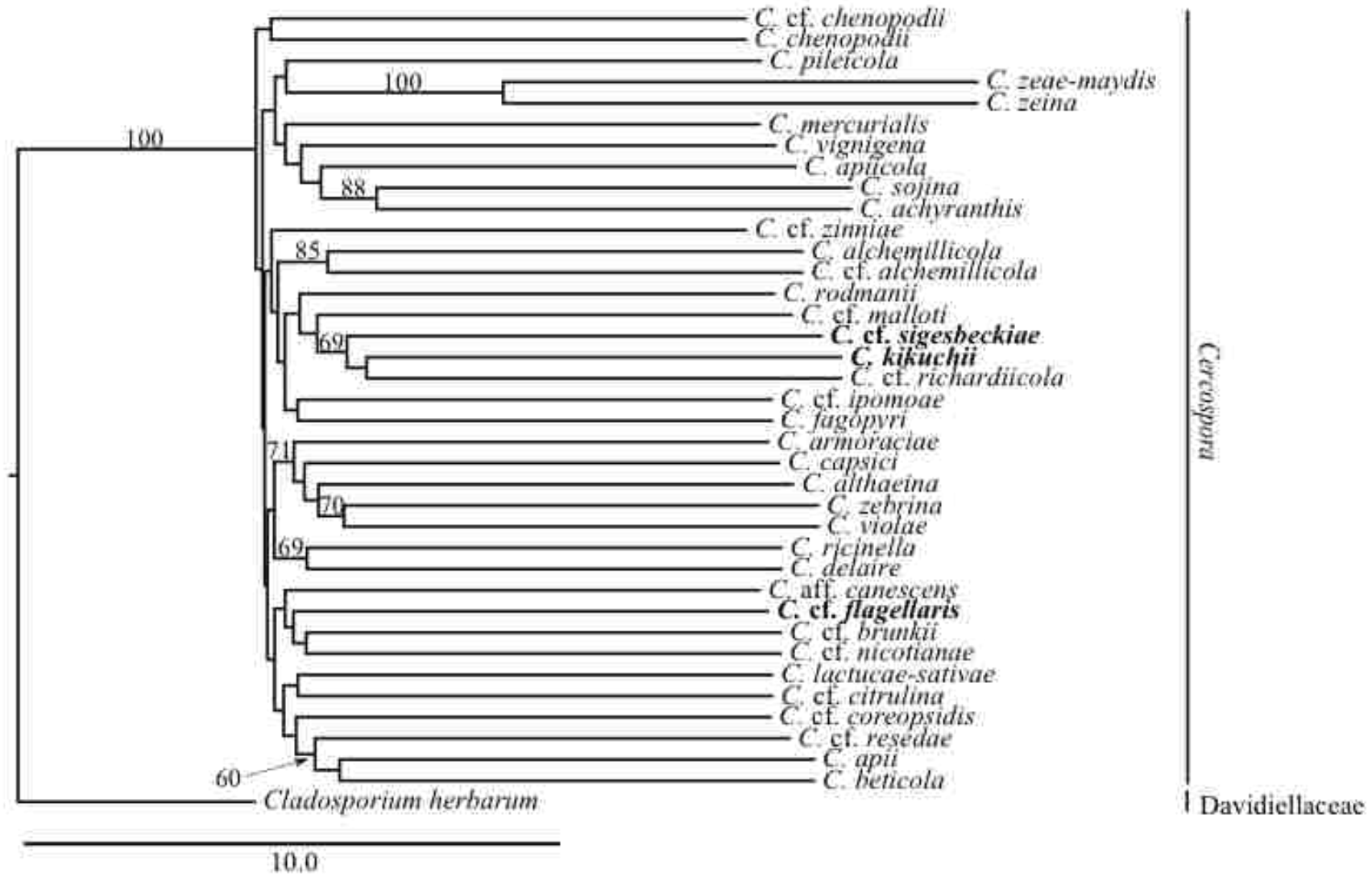


Figure A3.14. Species tree inferred from the five independent RAxML gene trees (DS-1) including 37 species of *Cercospora* using pseudo-ML approach in MP-EST. *Cercospora kikuchii*. *C.* cf. *flagellaris* and *C.* cf. *sigesbeckiae* are shown in bold. Tree is rooted

with *Cladosporium herbarum*. Support values at nodes represent bootstrap percentages >60 obtained with 100 replicates. Branch lengths are in coalescent units.

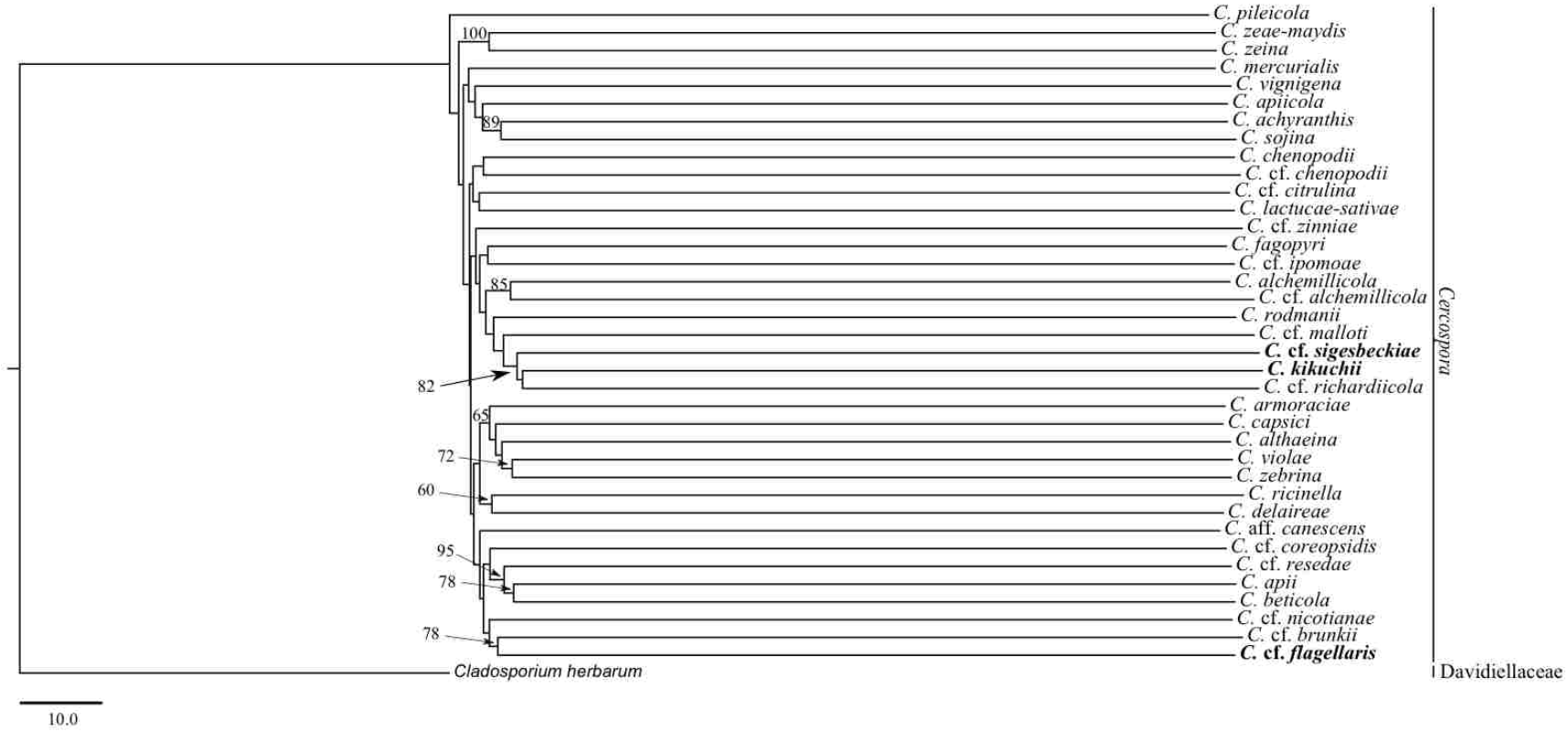


Figure A3.15. Species tree inferred from the five independent RAxML gene trees (DS-1) including 37 species of *Cercospora* using STAR approach. *Cercospora kikuchii*, *C.* cf. *flagellaris* and *C.* cf. *sigesbeckiae* are shown in bold. Tree is rooted with *Cladosporium herbarum*. Support values at nodes represent bootstrap percentages >60 obtained with 100 replicates. Branch lengths are in coalescent units.

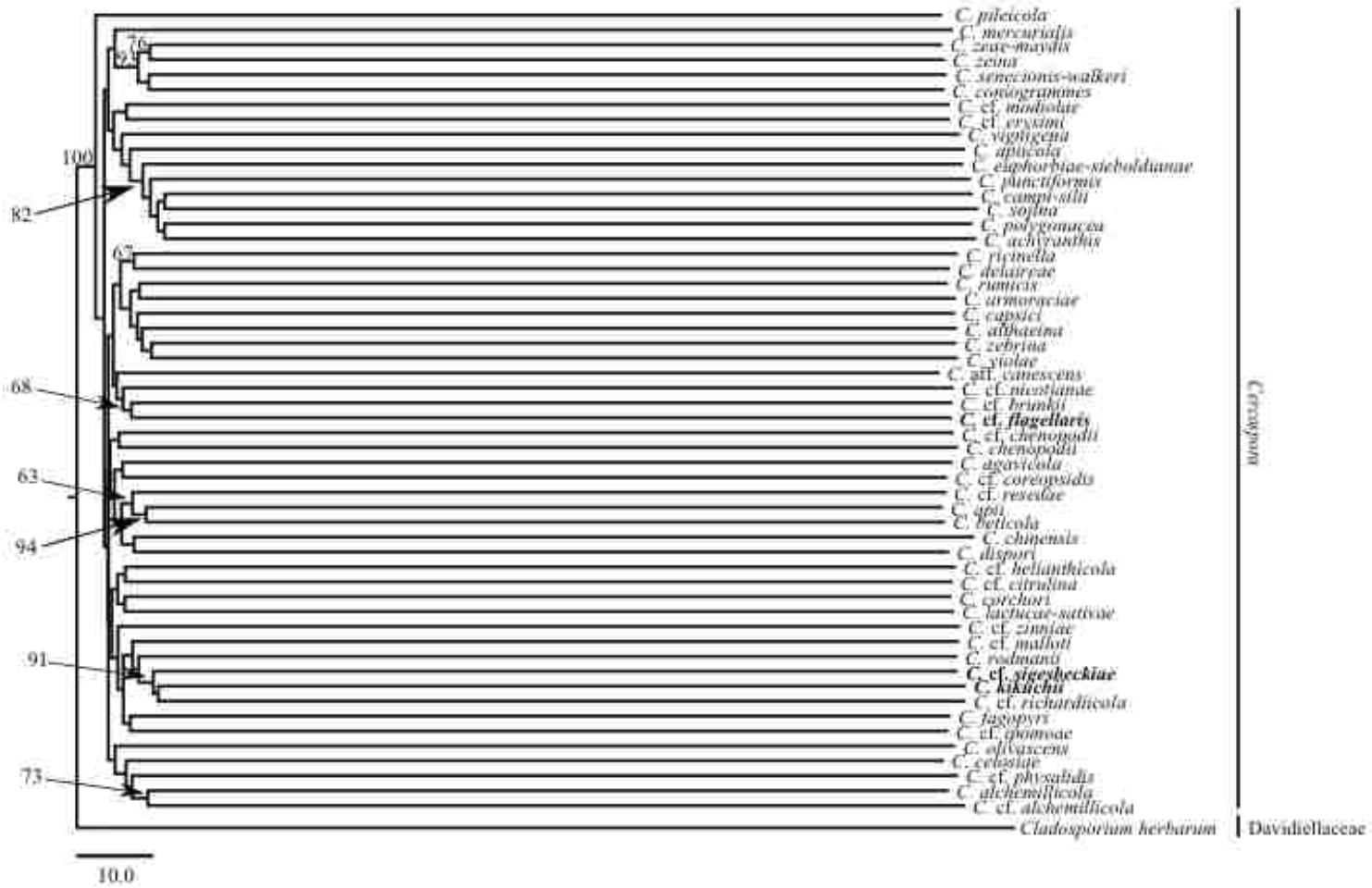


Figure A3.16. Species tree inferred from the five independent RAxML gene trees (DS-1) including 37 species of *Cercospora* using pseudo-ML approach in MP-EST. *Cercospora kikuchii*, *C.* cf. *flagellaris* and *C.* cf. *sigesbeckiae* are shown in bold. Tree is rooted with *Cladosporium herbarum*. Support values at nodes represent bootstrap percentages >60 obtained with 100 replicates. Branch lengths are in coalescent units.

APPENDIX 4. GENOME STATISTICS AND FIGURES FROM PRINSEQ ANALYSIS

Table A4.1. Length distribution

No. Sequences	6,048,465
Total bases	1,632,866,149
Mean sequence length	269.96 ± 81.17 bp
Minimum sequence length	8 bp
Maximum sequence length	635 bp
Length range	628 bp
Mode length	301 bp with 33,236 sequences

Table A4.2. GC content distribution

Mean GC content	51.50 ± 6.02%
Minimum GC content	0%
Maximum GC content	100%
Mode GC content	52% with 630,736 sequences

Table A4.3. Sequence duplication

	No. Sequences	Maximum duplicates
Exact duplicates	202,352 (3.35%)	490
Exact duplicates with reverse complements	3,062 (0.05%)	1
Total	205,414 (3.40%)	

Table A4.4. Statistics for large contigs

Length assessment		Quality assessment	
Number of contigs	469	Avg consensus quality	74
Total consensus	3.4E+07	Consensus bases with IUPAC	923
Largest contig	1591857	SRMc	0
N50 contig size	418495	WRMc	0
N90 contig size	82607	STMU	0
N95 contig size	46294	Contigs having only reads w/o qual	0
		Contigs with reads w/o qual values	0

Coverage assessment			
Max coverage	2042		

Table A4.5. Statistics for all contigs

Length assessment		Quality assessment	
Number of contigs	1984	Avg consensus quality	53
Total consensus	3.5E+07	Consensus bases with IUPAC	4545
Largest contig	1591857	SRMc	0
N50 contig size	406269	WRMc	0
N90 contig size	69207	STMU	0
N95 contig size	18100	Contigs having only reads w/o qual	0
		Contigs with reads w/o qual values	0
Coverage assessment			
Max coverage	2042		

APPENDIX 5. INDICES OF SUBSTITUTION SATURATION

Table A5.1. Indices of substitution saturation for DS-4 treatment NoGb "all loci"

Locus	S_{pooled}	V_{pooled}	s/v ratio	propinv sites	I_{ss}	$I_{ss.c}$ sym	$I_{ss.c}$ asymm	95 % CI upper limit	95 % CI lower limit
IGS1	9551	6238	1.53	0.27	0.67	0.77	0.52	0.54	0.80
IGS2	7275	2675	2.72	0.22	0.43	0.72	0.45	0.30	0.56
IGS3	5334	3218	1.66	0.30	0.52	0.73	0.47	0.37	0.67
IGS4	7632	4435	1.72	0.14	0.44	0.72	0.46	0.36	0.51
IGS5	6298	3925	1.6	0.31	0.73	0.74	0.48	0.56	0.90
IGS6	5830	3956	1.47	0.22	2.01	0.78	0.55	1.92	2.11
IGS7	6246	3913	1.6	0.35	0.43	0.72	0.46	0.29	0.57
IGS8	6313	4246	1.49	0.20	0.38	0.75	0.49	0.28	0.49
IGS9	4949	2885	1.72	0.29	0.74	0.73	0.47	0.57	0.91
IGS10	6577	5363	1.23	0.25	0.5	0.75	0.49	0.38	0.62
IGS11	9823	6727	1.46	0.25	0.78	0.75	0.50	0.64	0.92
ACT	1389	1140	1.22	0.46	0.2	0.59	0.38	0.07	0.33
CAL	3263	1131	2.89	0.30	0.12	0.65	0.40	0.07	0.17
EF1	2297	1084	2.12	0.28	1.1	0.69	0.44	0.86	1.34
H3	2201	1115	1.97	0.39	0.13	0.69	0.42	0.04	0.22
ITS	175	43	4.07	0.80	0.06	0.73	0.59	0.02	0.11

Table A5.2. Indices of substitution saturation for DS-4 treatment LSGb "all loci"

Locus	S_{pooled}	V_{pooled}	s/v ratio	propinv sites	I_{ss}	$I_{ss.c}$ sym	$I_{ss.c}$ asymm	95 % CI upper limit	95 % CI lower limit
IGS1	9360	5899	1.59	0.27	0.18	0.76	0.50	0.15	0.21
IGS2	7115	5157	1.38	0.23	0.19	0.71	0.44	0.15	0.23
IGS3	5211	3130	1.66	0.29	0.15	0.72	0.46	0.12	0.18
IGS4	7286	4287	1.7	0.13	0.35	0.72	0.45	0.31	0.38

IGS5	6289	3921	1.6	0.27	0.16	0.72	0.46	0.12	0.21
IGS6	5767	3785	1.52	0.22	0.25	0.71	0.45	0.21	0.28
IGS7	6224	3907	1.59	0.35	0.21	0.71	0.46	0.17	0.25
IGS8	6139	4208	1.46	0.19	0.16	0.74	0.48	0.13	0.19
IGS9	5287	3044	1.74	0.29	0.19	0.71	0.45	0.16	0.23
IGS10	6474	5254	1.23	0.25	0.22	0.74	0.49	0.19	0.26
IGS11	11293	7553	1.5	0.24	0.28	0.73	0.47	0.24	0.33
ACT	1389	1019	1.36	0.32	0.27	0.61	0.39	0.17	0.36
CAL	3569	1231	2.9	0.30	0.23	0.66	0.40	0.15	0.31
EF1	2682	1228	2.18	0.30	0.19	0.66	0.40	0.13	0.25
H3	2200	1115	1.97	0.39	0.08	0.69	0.42	0.04	0.12
ITS	246	58	4.24	0.80	0.1	0.72	0.54	0.03	0.16

Table A5.3. Indices of substitution saturation for DS-4 treatment MSGb “all loci”

Locus	S_{pooled}	V_{pooled}	s/v ratio	propinv sites	I_{ss}	$I_{ss.c}$ sym	$I_{ss.c}$ asymm	95% CI upper limit	95% CI lower limit
IGS1	7234	4405	1.64	0.28	0.13	0.74	0.47	0.10	0.15
IGS2	4938	3688	1.34	0.25	0.14	0.69	0.42	0.11	0.18
IGS3	3912	2183	1.79	0.31	0.12	0.70	0.45	0.09	0.15
IGS4	559	408	1.37	0.01	0.14	0.58	0.51	0.09	0.20
IGS5	4138	2767	1.50	0.34	0.12	0.70	0.43	0.09	0.15
IGS6	2756	1380	2.00	0.23	0.15	0.63	0.40	0.10	0.19
IGS7	4374	2810	1.56	0.27	0.15	0.69	0.44	0.11	0.19
IGS8	4851	3118	1.56	0.29	0.13	0.71	0.45	0.10	0.16
IGS9	3630	1985	1.83	0.29	0.14	0.67	0.42	0.10	0.18
IGS10	3936	3528	1.12	0.26	0.12	0.70	0.44	0.09	0.16
IGS11	4259	2964	1.44	0.26	0.16	0.67	0.43	0.11	0.20
ACT	1088	789	1.38	0.31	0.09	0.58	0.37	0.05	0.14
CAL	3263	1131	2.89	0.30	0.12	0.65	0.40	0.07	0.17

EF1	1916	869	2.20	0.30	0.11	0.61	0.39	0.06	0.15
H3	2054	1002	2.05	0.37	0.07	0.68	0.42	0.04	0.10
ITS	175	43	4.07	0.80	0.06	0.73	0.59	0.02	0.11

S-transition

V-transversion

propinv sites- proportion of invariant sites

I_{ss} - index of substitution saturation

$I_{ss.c}$ sym - index of substitution saturation for symmertrical tree

$I_{ss.c}$ asymm - index of substition saturation for asymmetrical tree

CI- confidence interval

APPENDIX 6. ALIGNMENT INFORMATION AND NUCLEOTIDE SUBSTITUTION MODELS USED IN CHAPTER THREE

Table A6.1. Individiual alignment lengths and nucleotide substitution models used in DS-1

		NoGb			LSGb			MSGb	
Locus	Length bp	BI	ML	Length bp	BI	ML	Length bp	BI	ML
IGS2	501	K80+G	K80+G	452	K80+G	TrNef+G	336	K80+G	K80+G
IGS3	595	SYM+G	TIM3ef+G	533	SYM+G	TIM3ef+G	413	SYM+G	TPM3+G
IGS4	538	HKY+G	HKY+G	518	HKY+I+G	HKY+G	80	K80	JC
IGS5	659	HKY+G	TrN+G	544	HKY+G	TrN+G	378	HKY+G	K80+I
IGS12	800	GTR+G	TIM3+G	722	GTR+G	TrN+G	433	GTR+G	TIM1+G
IGS6	1331	K80+G	TrNef+G	490	HKY+G	HKY+G	243	HKY+G	HKY+G
IGS7	531	HKY+G	TrNef+G	499	HKY+G	TrNef+G	303	HKY+G	K80+G
IGS8	691	HKY+G	HKY+G	633	HKY+G	HKY+G	464	HKY+G	TrN+G
IGS9	590	HKY+G	HKY+G	496	HKY+G	TPM2uf+G	326	HKY+G	HKY+G
IGS10	737	GTR+G	TPM2uf+I+G	678	GTR+G	TPM2uf+I+G	396	GTR+G	TPM2+I+G
IGS11	765	GTR+G	TPM1uf+G	602	GTR+G	TIM1+G	318	GTR+G	TIM1ef+I+G
ACT	193	GTR+I	K80+G	226	GTR+I	K80+G	163	GTR+G	K80+G
Btub	1420	GTR+I+G	TIM1+I+G	1398	GTR+I+G	TrN+I+G	730	GTR+I+G	TrN+I+G
CAL	277	HKY+G	TrN+G	313	GTR+G	TrN+G	221	GTR+G	TrN+G
EF1	296	HKY+G	K80+G	306	HKY+G	K80+G	141	HKY+G	K80
H3	385	GTR+G	TrN+I	388	GTR+G	TrN+I	317	HKY+G	HKY+G
ITS	466	SYM+I	TIM1ef+I	521	SYM+I	TIM1ef+I	457	SYM+I	TIM1ef+I
Concat	10775			9319			5719		

Table A6.2. Individual alignment lengths and nucleotide substitution models used in DS-3

	NoGb			LSGb			MSGb		
Locus	Length bp	BI	ML	Length bp	BI	ML	Length bp	BI	ML
IGS2	472	K80+I	K80+G	451	K80+I	K80+I	392	K80+I	K80+I
IGS3	523	K80+I	TrNef+I	533	K80+I	TrNef+I	510	K80+I	K80+I
IGS4	523	HKY	TPM2uf	525	HKY	HKY	510	HKY	TPM2uf
IGS5	533	HKY+I	HKY+I	548	HKY+I	HKY+G	511	HKY+I	TPM2uf+I
IGS12	728	HKY+I	TPM1uf+I	731	HKY+I	TPM1uf+I	708	HKY+I	TPM1uf+I
IGS6	493	HKY	HKY	495	HKY	HKY	490	HKY	HKY
IGS7	456	HKY+I	TPM2+G	497	K80+I	K80+I	455	HKY+I	TPM2+I
IGS8	635	HKY+I	TrN+I	636	HKY+I	HKY+I	621	HKY+I	TrN+I
IGS9	499	HKY	HKY	500	HKY	HKY	487	HKY	HKY
IGS10	650	GTR+I+G	TVM+I+G	672	GTR+I+G	TVM+I+G	642	GTR+I+G	TVM+I+G
IGS11	587	HKY	TPM2uf	581	HKY	TPM2uf	572	HKY	TPM2uf
ACT	210	JC+I	JC+G	227	JC+I	JC+G	176	F81+I	TPM2+I
Btub	1408	GTR+I	TIM2+I	1409	GTR+I	TIM2+I	1407	GTR+I	TIM2+I
CAL	294	HKY	HKY	314	HKY	HKY	221	HKY	HKY
EF1	280	F81	F81	306	F81	F81	207	F81	F81
H3	376	HKY	TrN	412	HKY	TrN	324	HKY	F81
ITS	466	K80	JC	546	K80	JC	457	K80	JC
Concat	9133			9383			8690		

Table A6.3. Individual alignment lengths and nucleotide substitution models used in DS-4

	NoGb			LSGb			MSGb		
Locus	Length bp	BI	ML	Length bp	BI	ML	Length bp	BI	ML
IGS1	966	GTR+G	TPM3uf+G	799	GTR+G	TIM3+I+G	633	HKY+G	TrN+I+G
IGS2	506	K80+G	TrNef+G	458	K80+G	K80+G	386	K80+G	TIM2ef+G
IGS3	605	SYM+G	SYM+G	533	SYM+G	SYM+G	438	SYM+G	TIM3ef+G
IGS4	536	HKY+G	TrN+G	520	HKY+I+G	HKY+G	84	HKY	K80
IGS5	654	HKY+G	TrN+G	551	HKY+G	TrN+G	423	GTR+G	TrN+I+G
IGS6	1325	K80+G	TrN+G	486	HKY+G	TrN+G	236	HKY+G	TIM2+G
IGS7	532	HKY+I	TIM2ef+I	500	HKY+I	HKY+I	386	HKY+G	TIM2ef+I
IGS8	689	HKY+G	HKY+G	633	HKY+G	HKY+G	471	HKY+G	TrN+G
IGS9	590	HKY+G	HKY+G	497	HKY+G	TPM2uf+G	338	HKY+G	HKY+I
IGS10	742	GTR+I+G	TPM2uf+I+G	675	GTR+I+G	TVM+I+G	437	GTR+I+G	TPM2uf+G
IGS11	735	GTR+G	TPM1uf+G	605	GTR+G	TPM1uf+G	343	HKY+G	TPM1uf+I+G
ACT	196	HKY+G	K80+G	220	HKY+G	HKY+G	183	HKY+G	K80+G
CAL	275	HKY+G	TrN+G	307	GTR+I+G	TIM2+G	275	HKY+G	TrN+G
EF1	403	HKY+G	HKY+G	299	HKY+G	K80+G	218	HKY+G	HKY+G
H3	381	GTR+I	TrN+I	378	HKY+I	TrN+I	369	GTR+I	TrN+I
ITS	474	SYM+I	TIM1ef+I	483	SYM+I	TIM1ef+I	474	SYM+I	TIM1ef+I
Concat	9609			7944			5694		

APPENDIX 7. SUPPLEMENTAL LSGB TREES FROM CHAPTER THREE

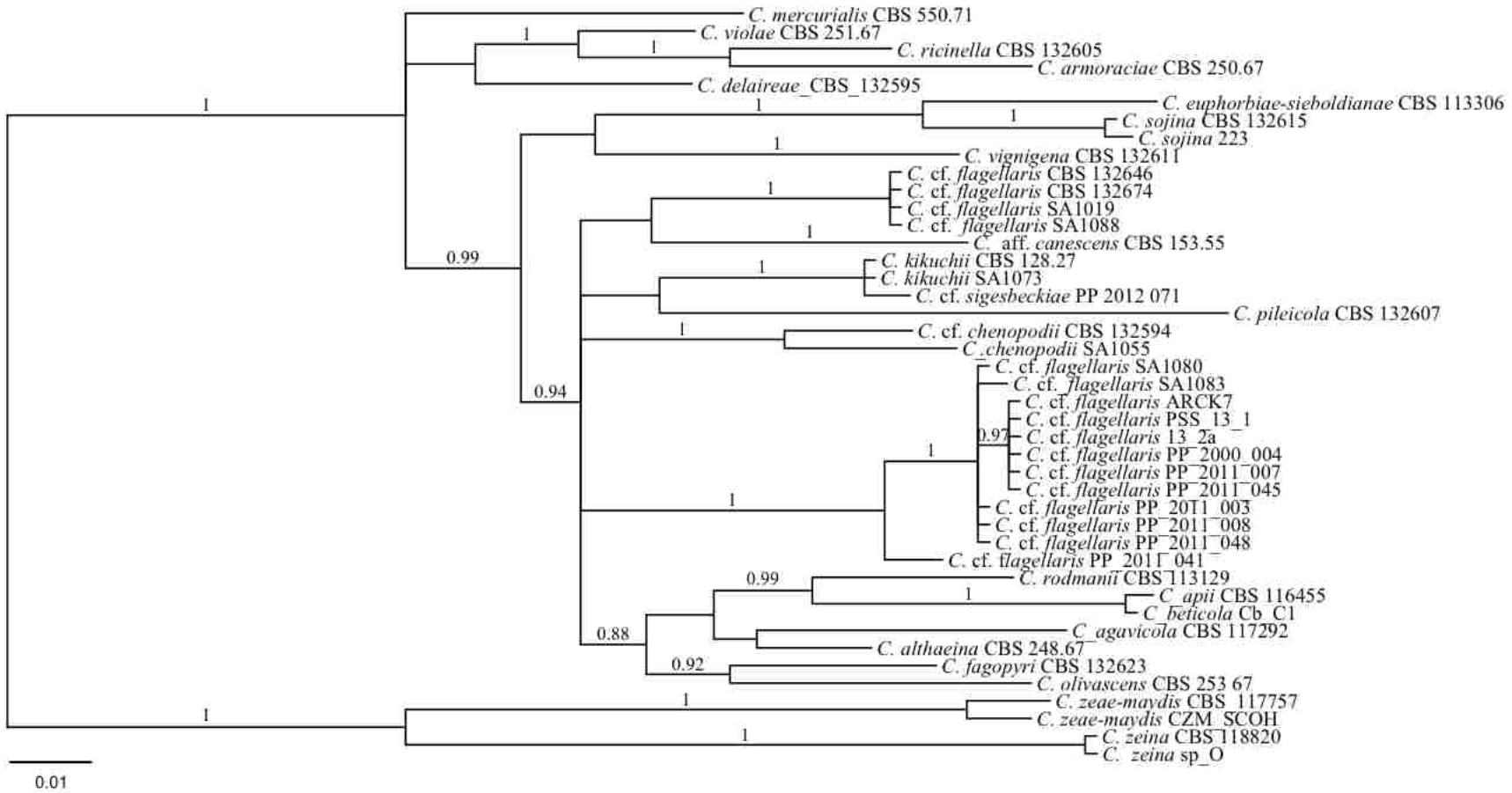


Figure A7.1. Topology inferred from Bayesian inference phylogenetic analysis of 24 *Cercospora* species from the LSGb treatment of the DS-1 alignment of IGS2. Tree is rooted with *C. zeina* and *C. zeae-maydis*. Posterior probability values of at least 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site.

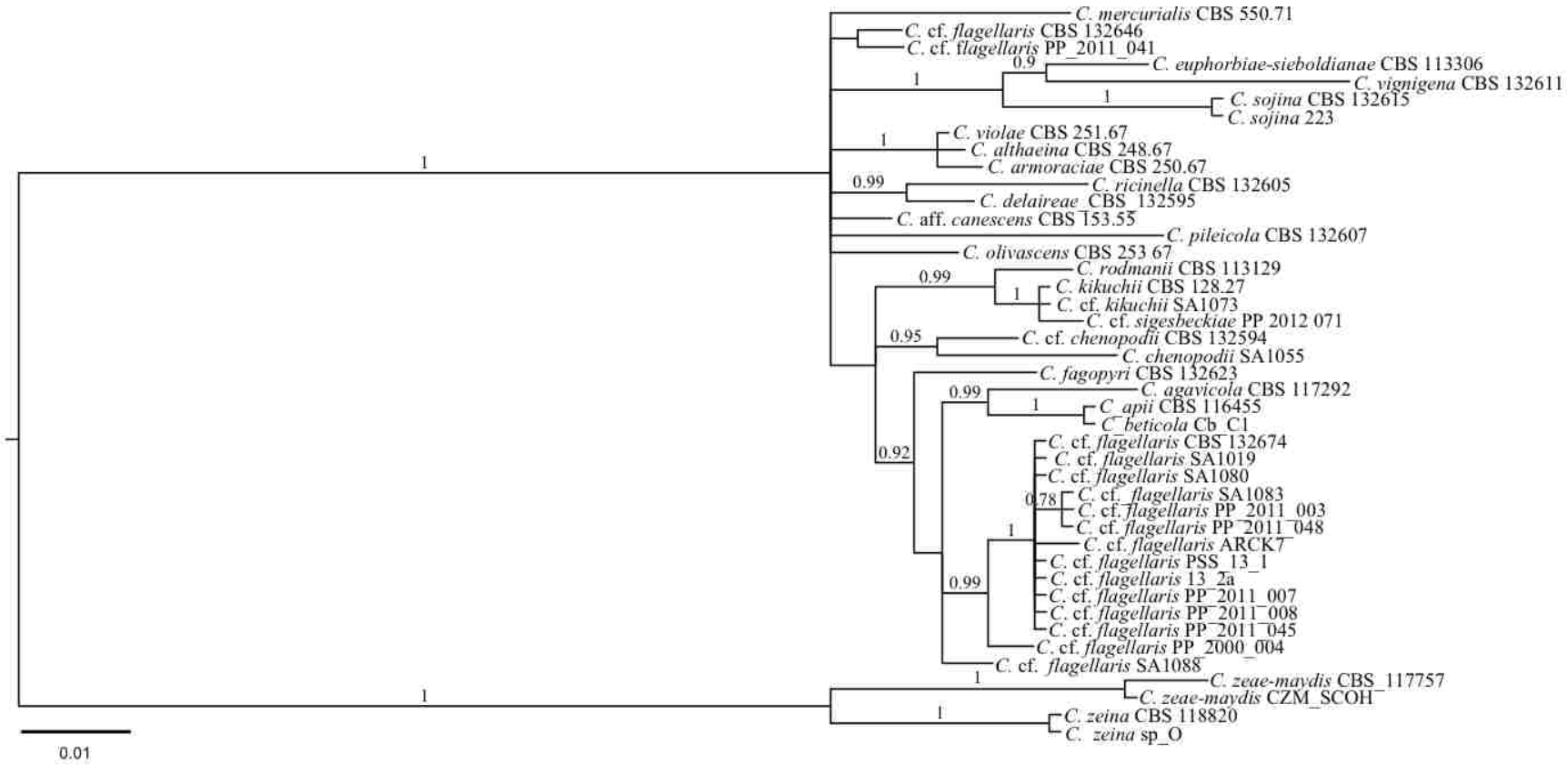


Figure A7.2. Topology inferred from Bayesian inference phylogenetic analysis of 24 *Cercospora* species from the LSGb treatment of the DS-1 alignment of IGS3. Tree is rooted with *C. zeina* and *C. zeae-maydis*. Posterior probability values of at least 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site.

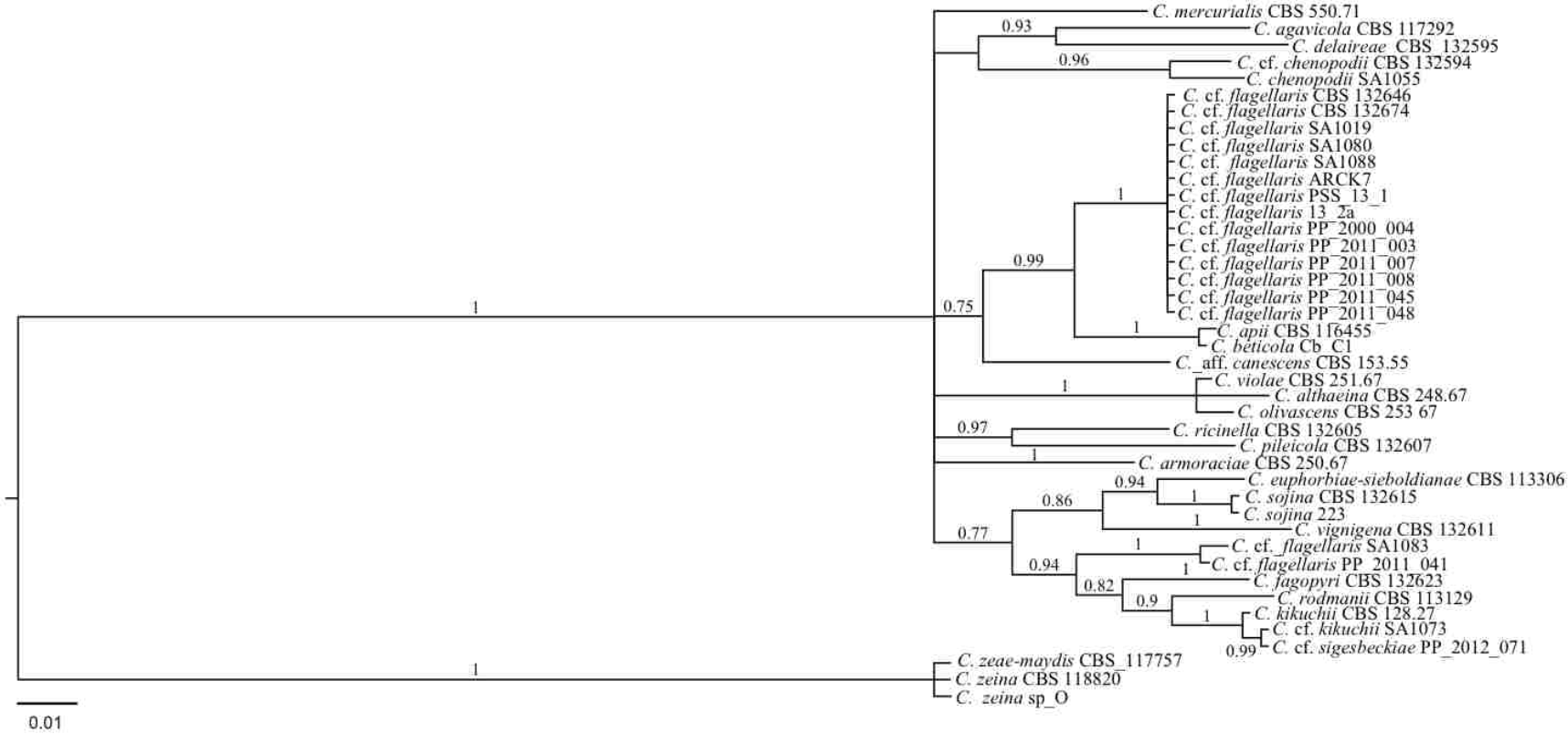


Figure A7.3. Topology inferred from Bayesian inference phylogenetic analysis of 24 *Cercospora* species from the LSGb treatment of the DS-1 alignment of IGS4. Tree is rooted with *C. zeina* and *C. zeae-maydis*. Posterior probability values of at least 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site.

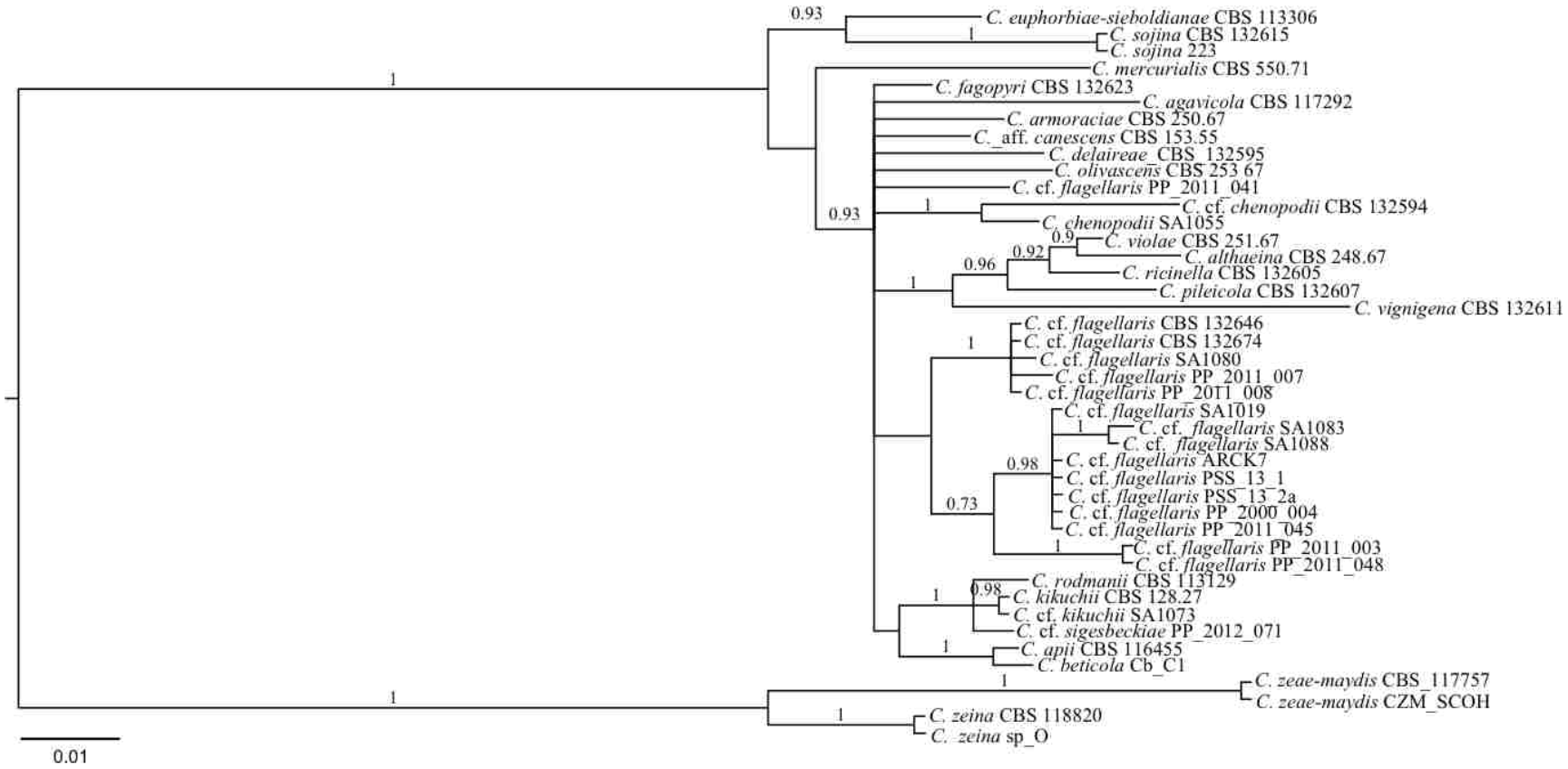


Figure A7.4. Topology inferred from Bayesian inference phylogenetic analysis of 24 *Cercospora* species from the LSGb treatment of the DS-1 alignment of IGS5. Tree is rooted with *C. zeina* and *C. zeae-maydis*. Posterior probability values of at least 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site.

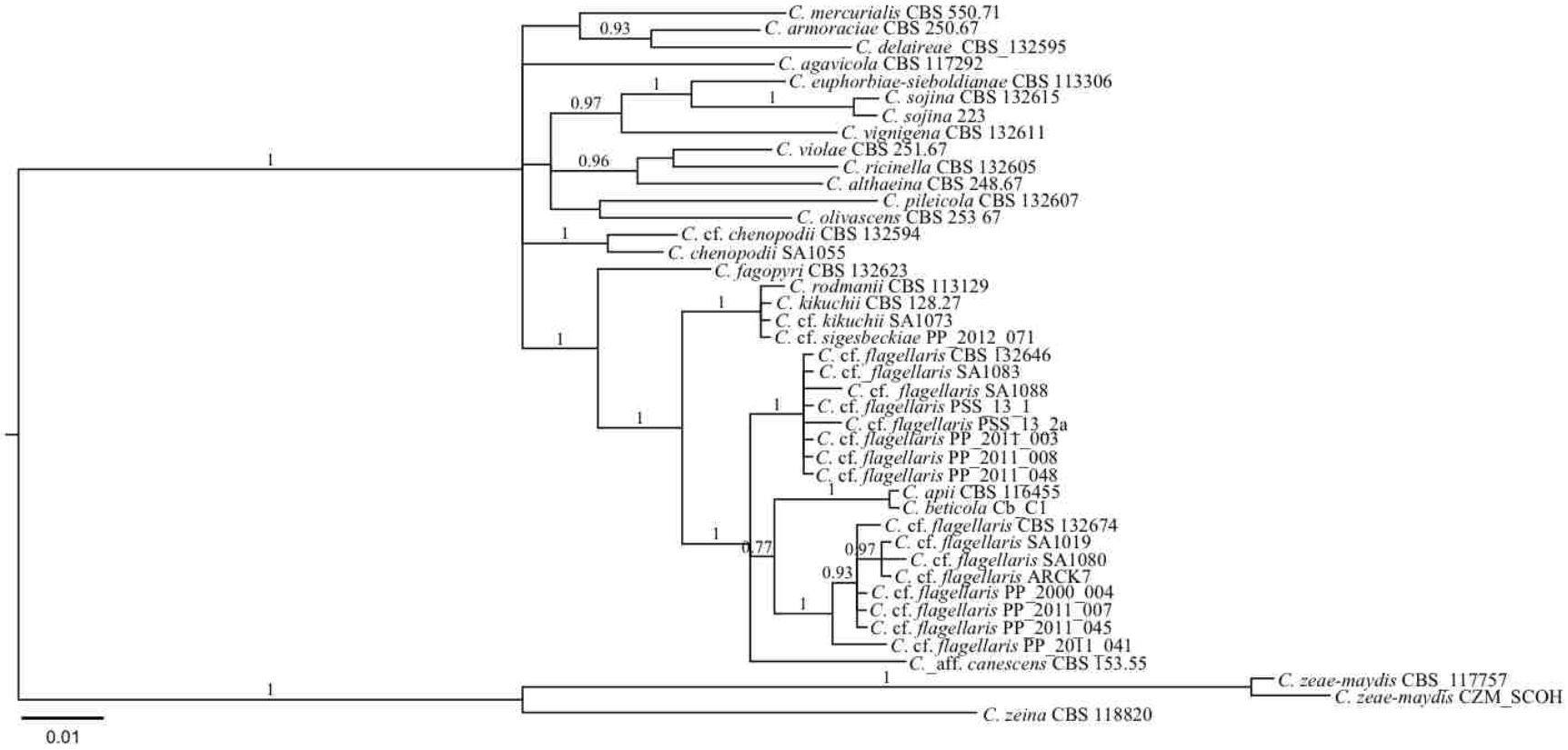


Figure A7.5. Topology inferred from Bayesian inference phylogenetic analysis of 24 *Cercospora* species from the LSGb treatment of the DS-1 alignment of IGS6. Tree is rooted with *C. zeina* and *C. zeae-maydis*. Posterior probability values of at least 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site.

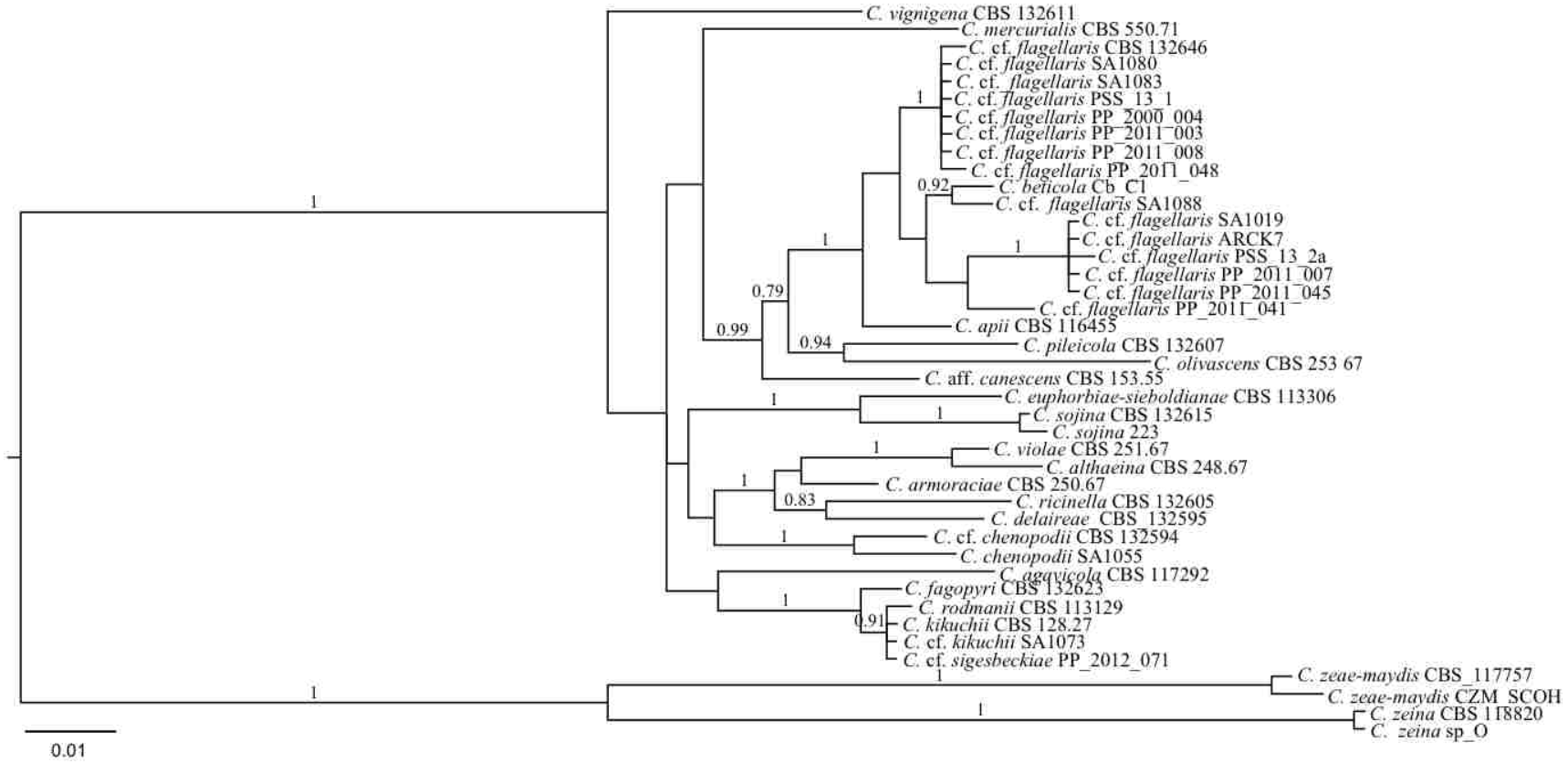


Figure A7.6. Topology inferred from Bayesian inference phylogenetic analysis of 24 *Cercospora* species from the LSGb treatment of the DS-1 alignment of IGS7. Tree is rooted with *C. zeina* and *C. zeae-maydis*. Posterior probability values of at least 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site.

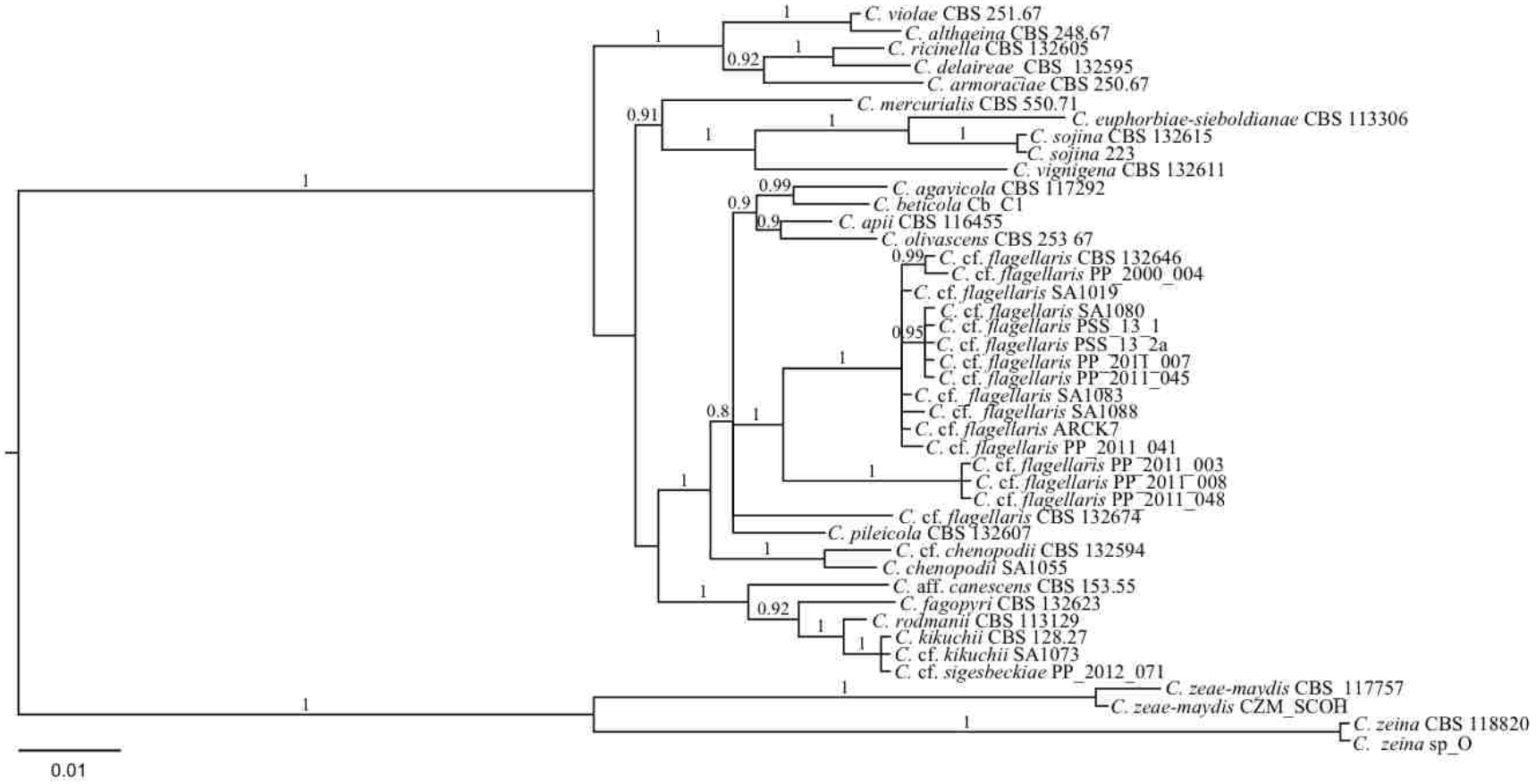


Figure A7.7. Topology inferred from Bayesian inference phylogenetic analysis of 24 *Cercospora* species from the LSGb treatment of the DS-1 alignment of IGS8. Tree is rooted with *C. zeina* and *C. zeae-maydis*. Posterior probability values of at least 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site.

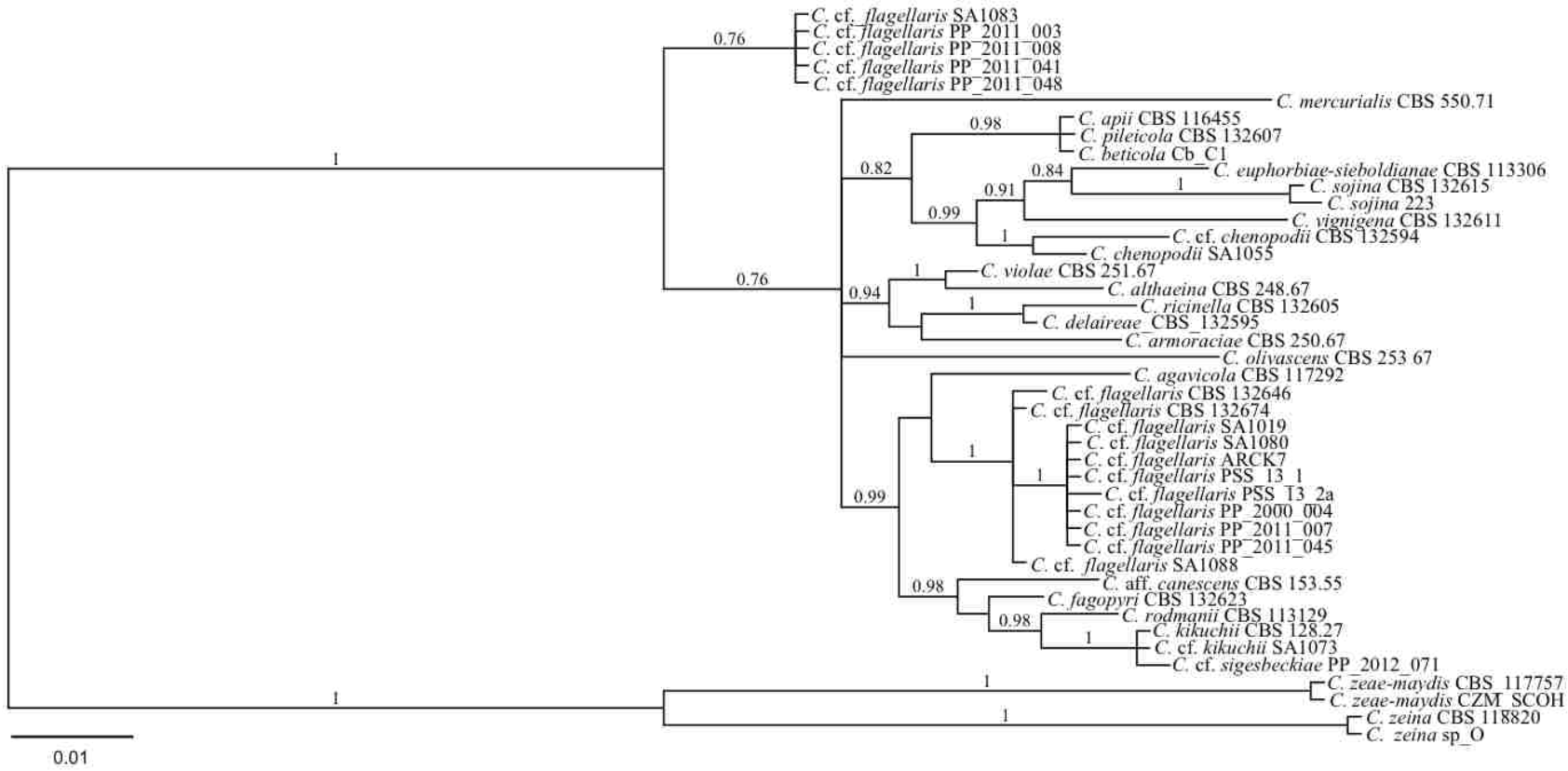


Figure A7.8. Topology inferred from Bayesian inference phylogenetic analysis of 24 *Cercospora* species from the LSGb treatment of the DS-1 alignment of IGS9. Tree is rooted with *C. zeina* and *C. zeae-maydis*. Posterior probability values of at least 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site.

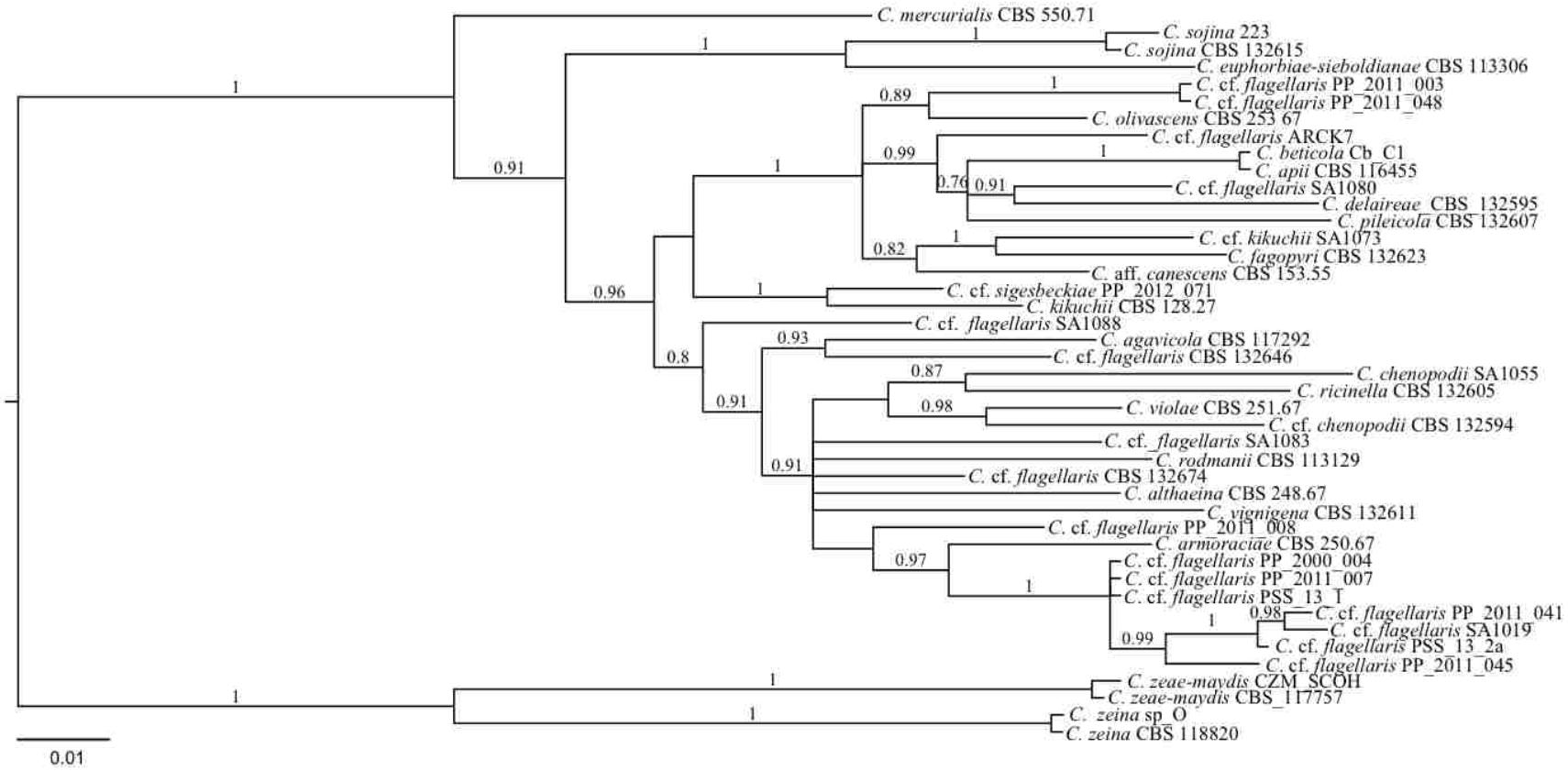


Figure A7.9. Topology inferred from Bayesian inference phylogenetic analysis of 24 *Cercospora* species from the LSGb treatment of the DS-1 alignment of IGS10. Tree is rooted with *C. zeina* and *C. zeae-maydis*. Posterior probability values of at least 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site.

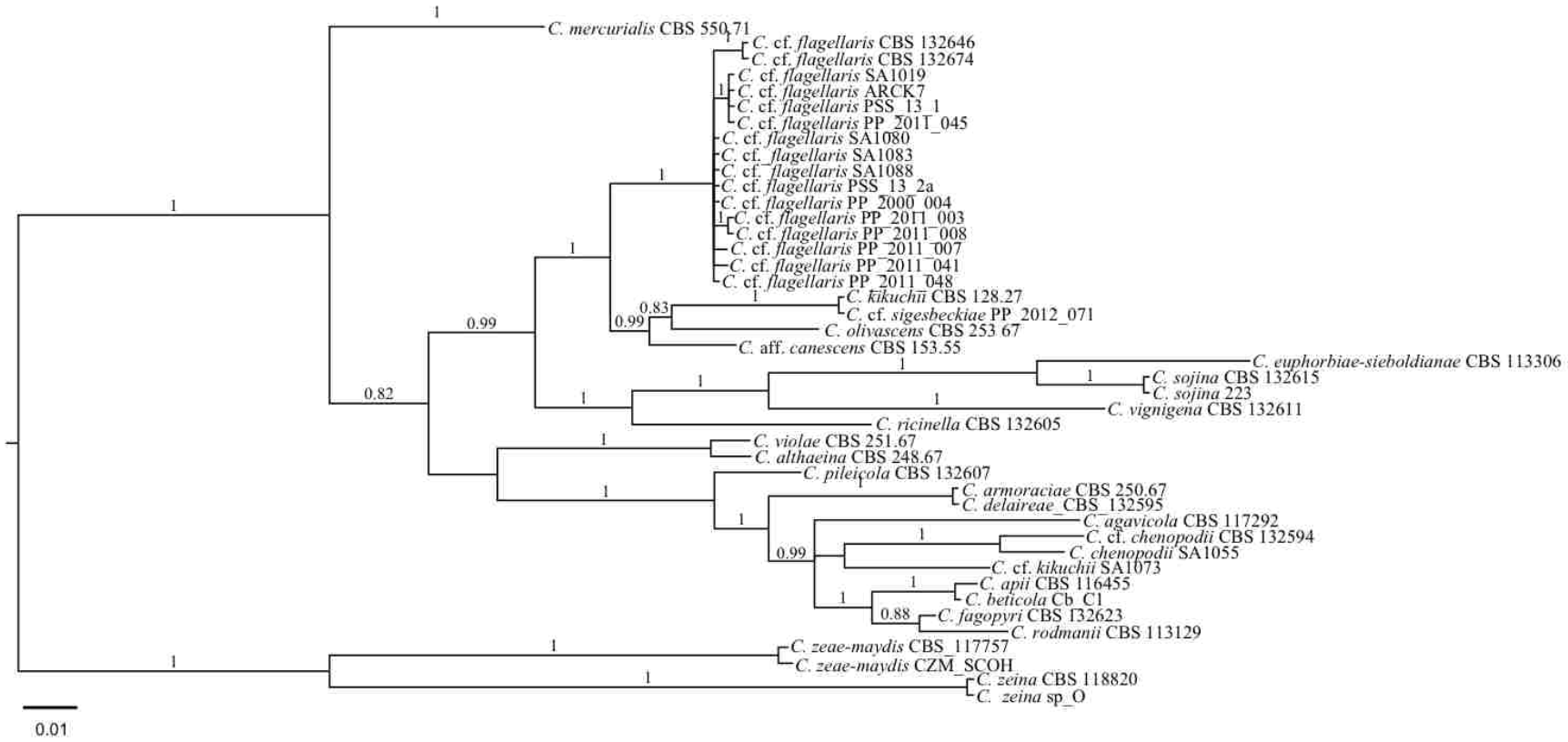


Figure A7.10. Topology inferred from Bayesian inference phylogenetic analysis of 24 *Cercospora* species from the LSGb treatment of the DS-1 alignment of IGS11. Tree is rooted with *C. zeina* and *C. zeae-maydis*. Posterior probability values of at least 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site.

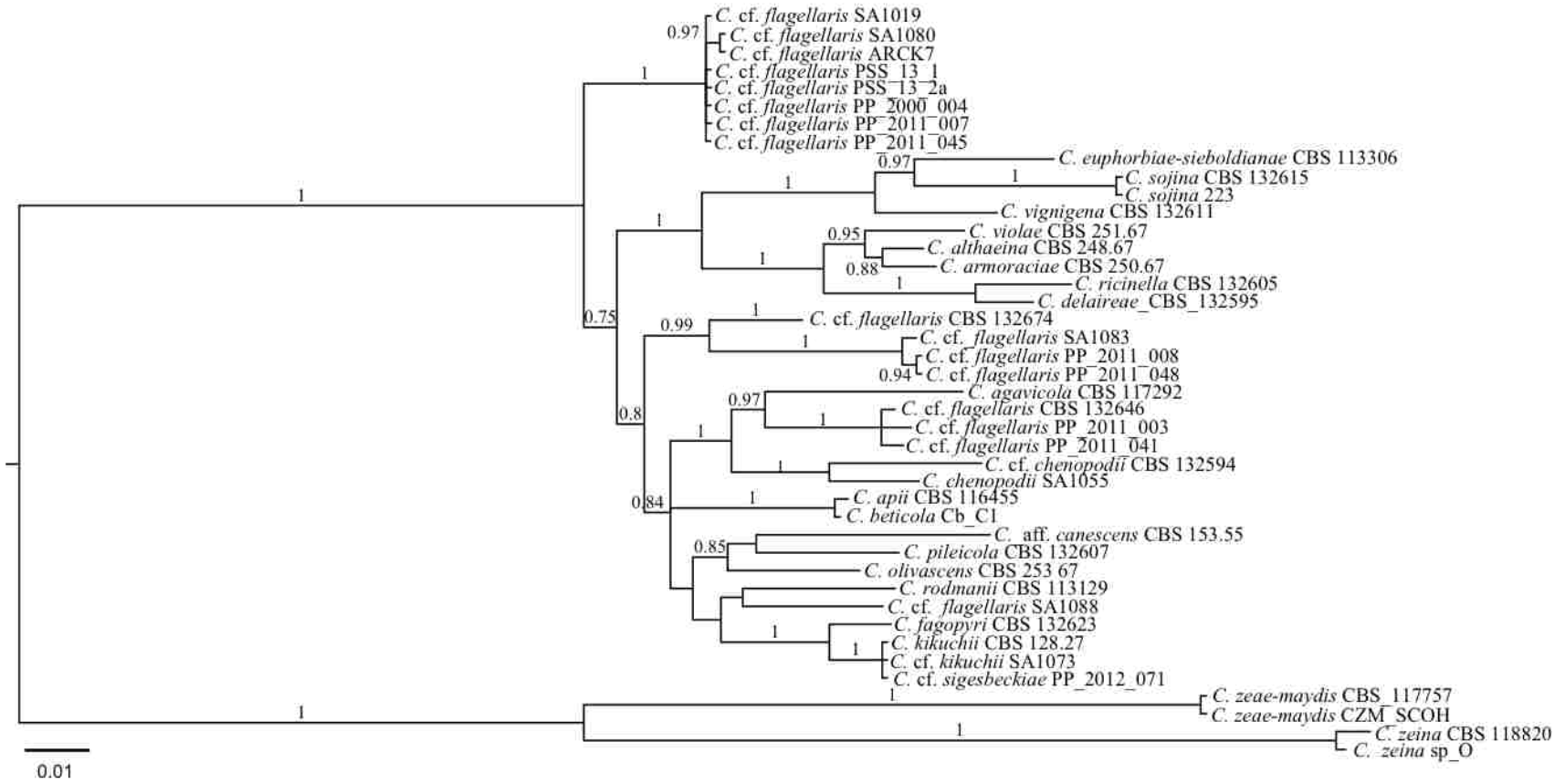


Figure A7.11. Topology inferred from Bayesian inference phylogenetic analysis of 24 *Cercospora* species from the LSGb treatment of the DS-1 alignment of IGS12. Tree is rooted with *C. zeina* and *C. zeae-maydis*. Posterior probability values of at least 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site.

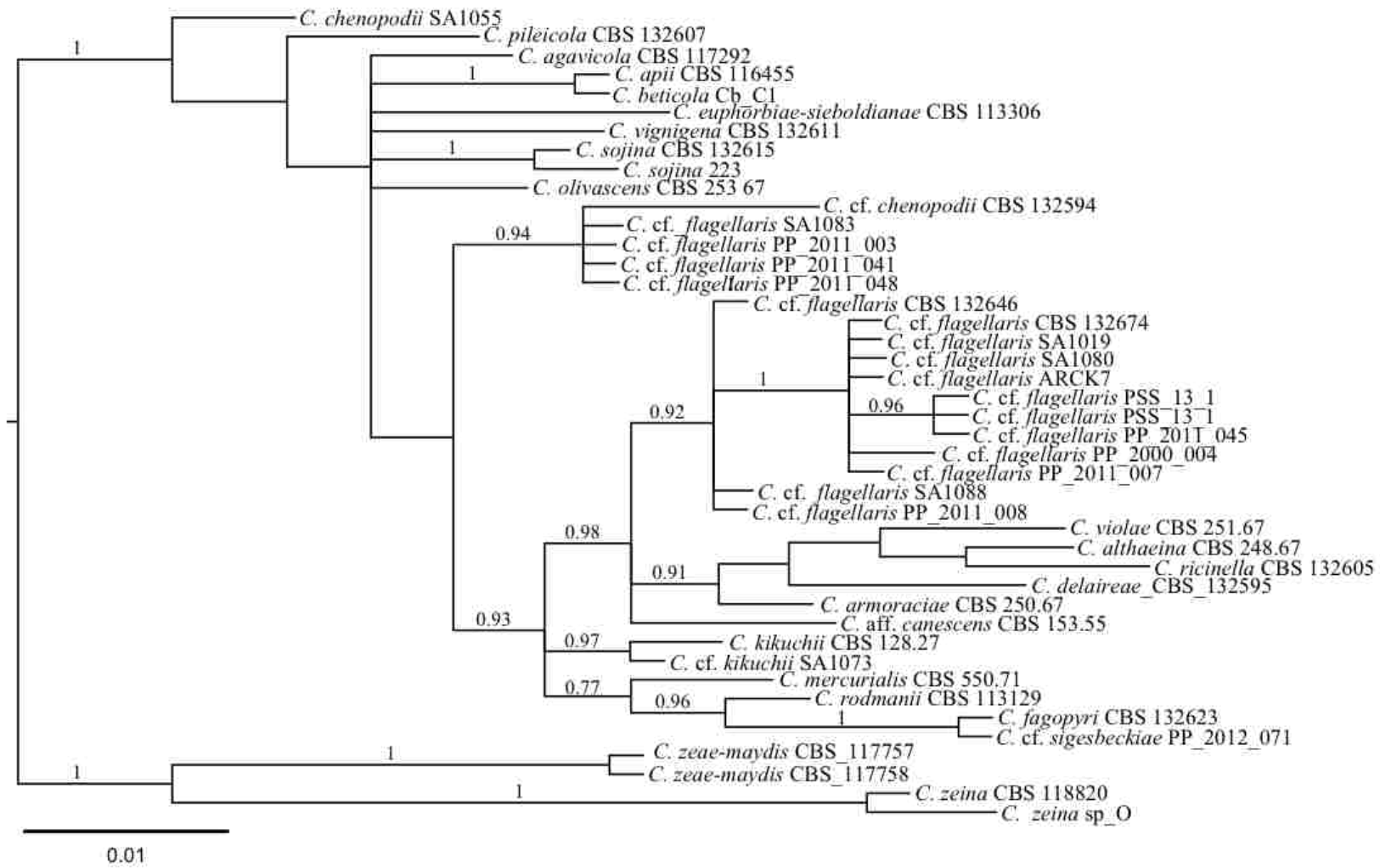


Figure A7.12. Topology inferred from Bayesian inference phylogenetic analysis of 24 *Cercospora* species from the LSGb treatment of the DS-1 alignment of actin. Tree is rooted with *C. zeina* and *C. zeae-maydis*. Posterior probability values of at least 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site.

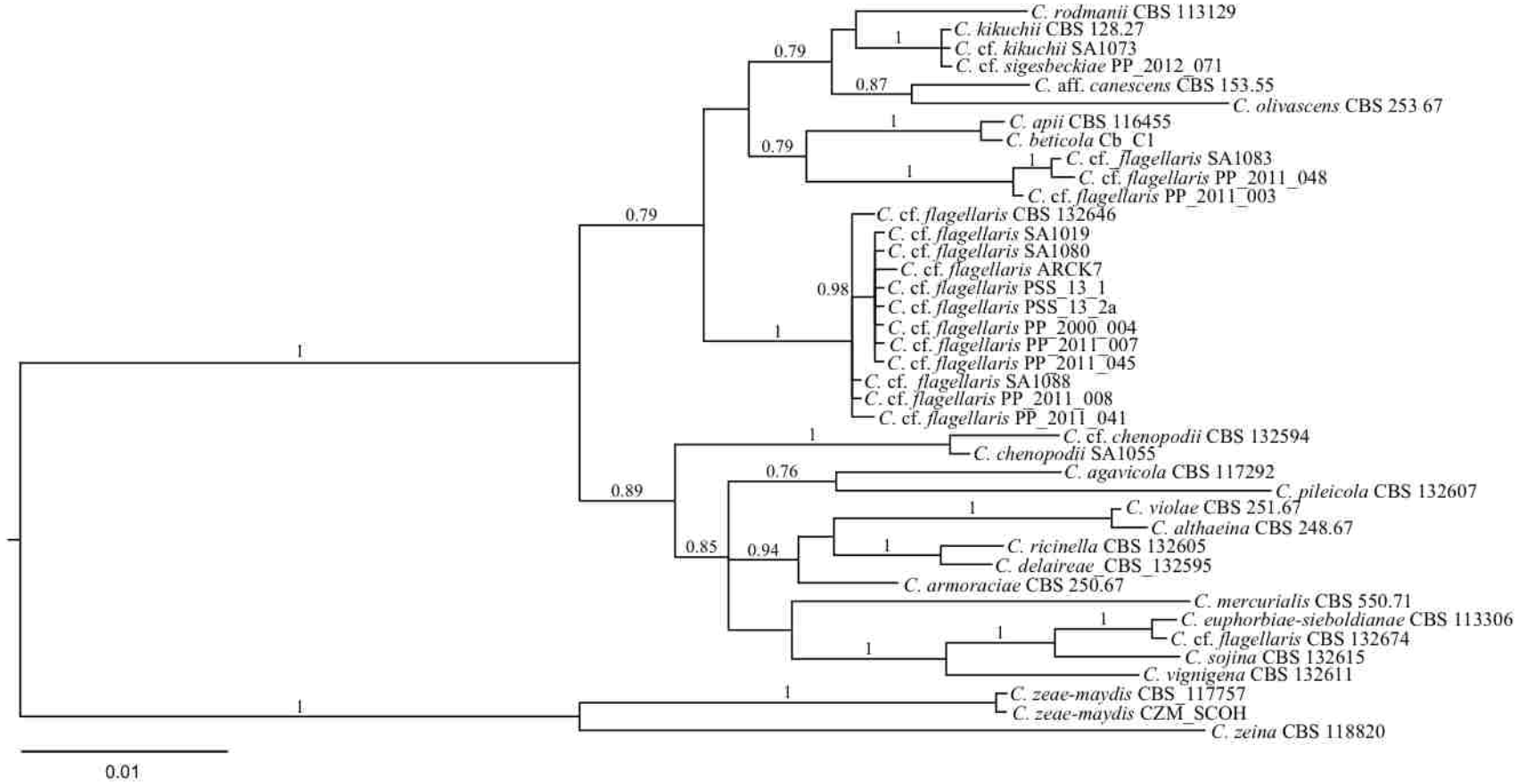


Figure A7.13. Topology inferred from Bayesian inference phylogenetic analysis of 24 *Cercospora* species from the LSGb treatment of the DS-1 alignment of B-Tubulin. Tree is rooted with *C. zeina* and *C. zeae-maydis*. Posterior probability values of at least 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site.

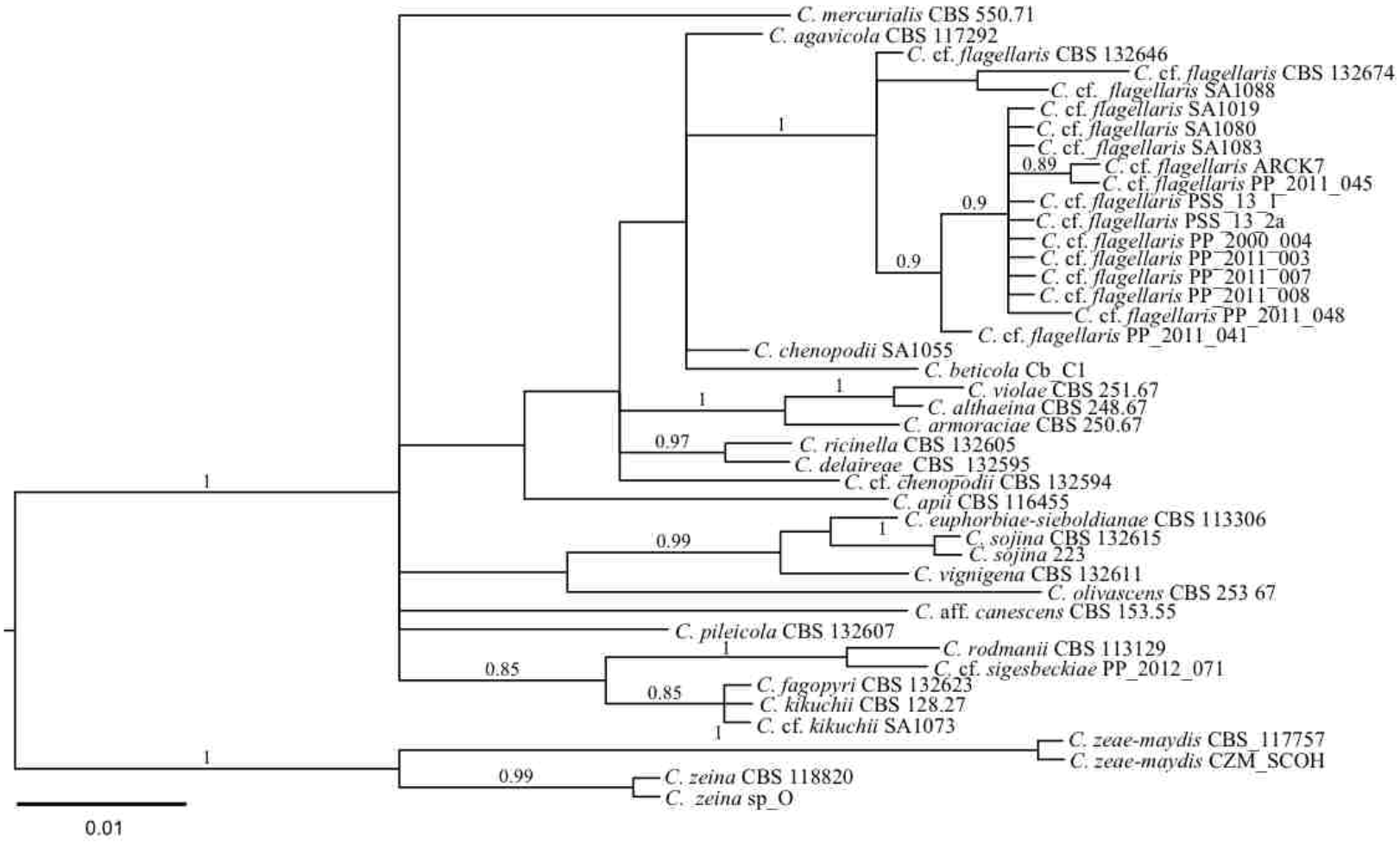


Figure A7.14. Topology inferred from Bayesian inference phylogenetic analysis of 24 *Cercospora* species from the LSGb treatment of the DS-1 alignment of calmodulin. Tree is rooted with *C. zeina* and *C. zeae-maydis*. Posterior probability values of at least 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site.

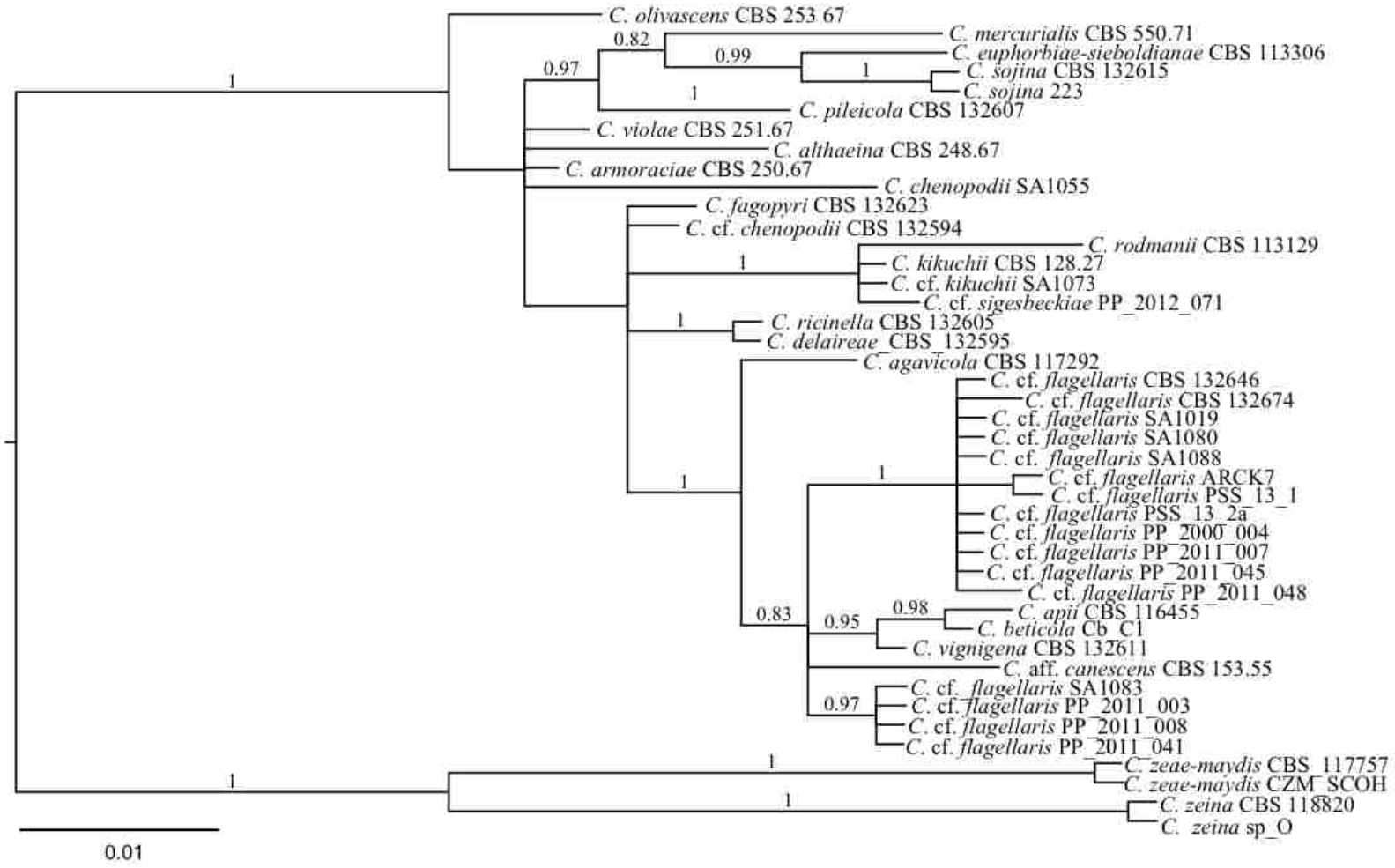


Figure A7.15. Topology inferred from Bayesian inference phylogenetic analysis of 24 *Cercospora* species from the LSGb treatment of the DS-1 alignment of Translation elongation factor 1-. Tree is rooted with *C. zeina* and *C. zeae-maydis*. Posterior probability values of at least 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site.

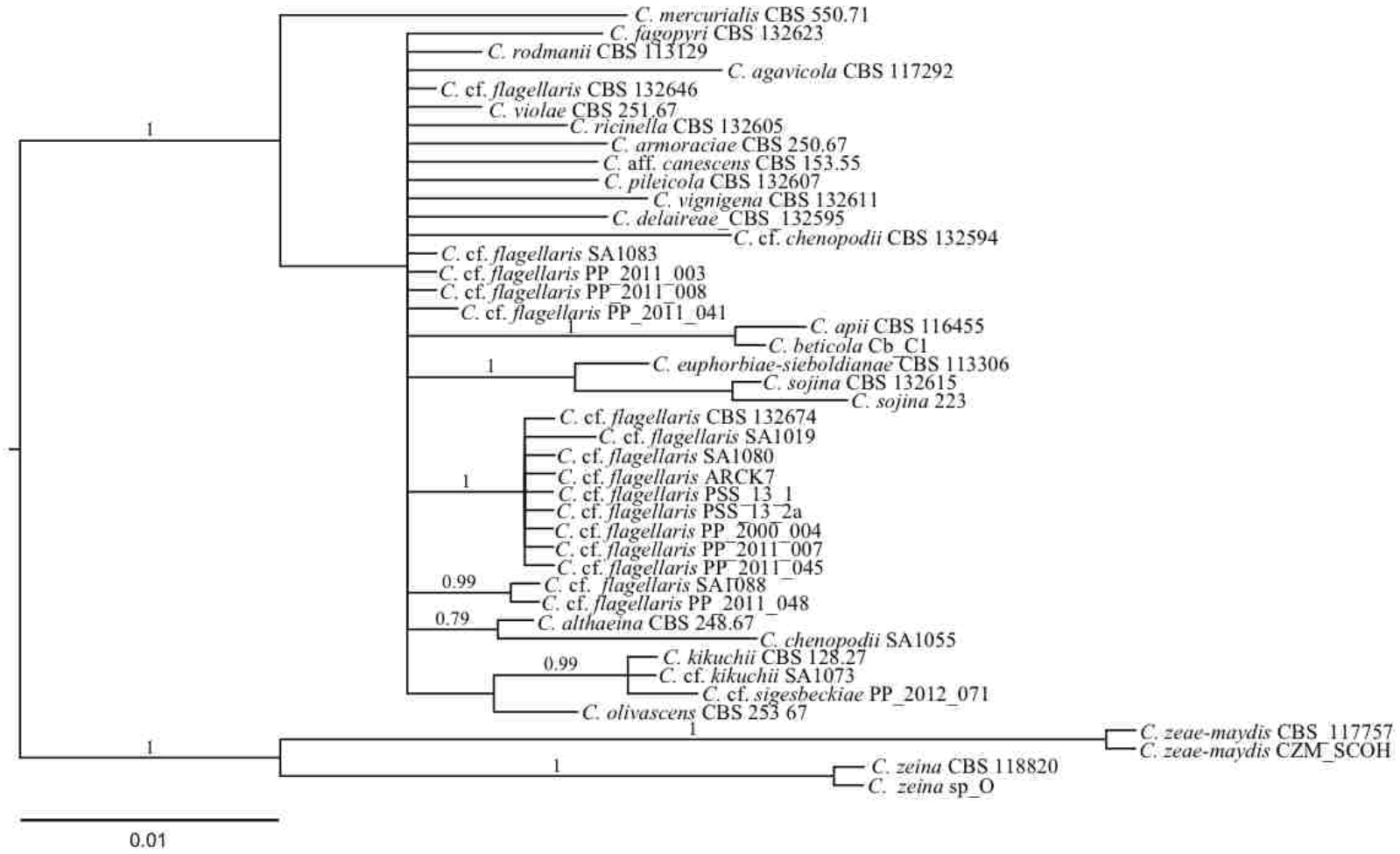


Figure A7.16. Topology inferred from Bayesian inference phylogenetic analysis of 24 *Cercospora* species from the LSGb treatment of the DS-1 alignment of histone 3. Tree is rooted with *C. zeina* and *C. zeae-maydis*. Posterior probability values of at least 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site.

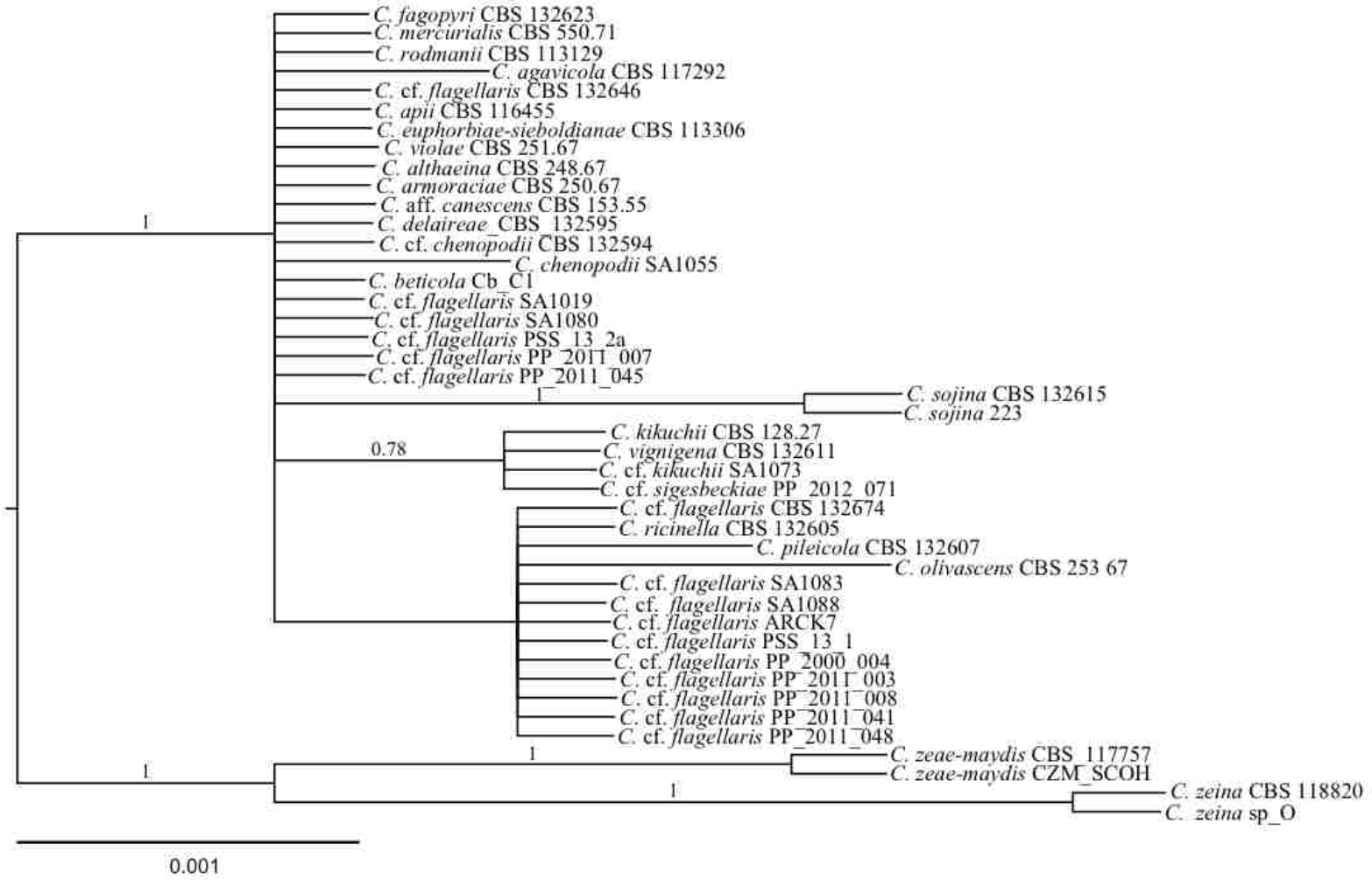


Figure A7.17. Topology inferred from Bayesian inference phylogenetic analysis of 24 *Cercospora* species from the LSGb treatment of the DS-1 alignment of ITS. Tree is rooted with *C. zeina* and *C. zeae-maydis*. Posterior probability values of at least 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site.

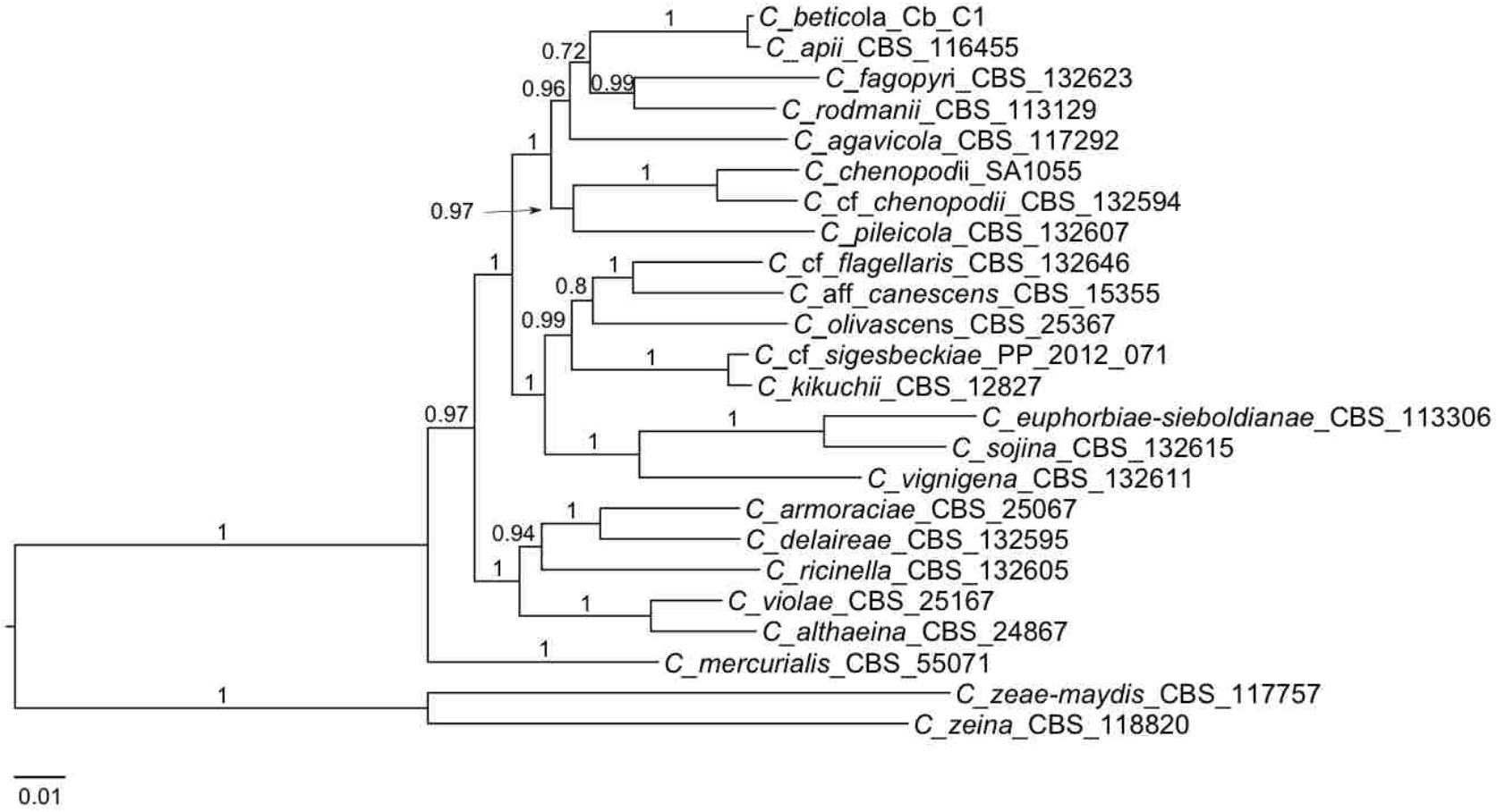


Figure A7.18. Topology inferred from Bayesian inference phylogenetic analysis of 24 species of *Cercospora* from the concatenated alignment of DS-4 using the five IGS loci with the highest net phylogenetic informativeness values (IGS1, IGS2, IGS4, IGS10 and IGS11) and (B) the five IGS loci with the lowest PIVs (IGS3, IGS5, IGS6, IGS8 and IGS9. Tree is rooted with *Cercospora zeina* and *C. zeae-maydis*. Posterior probability values of at least 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site.

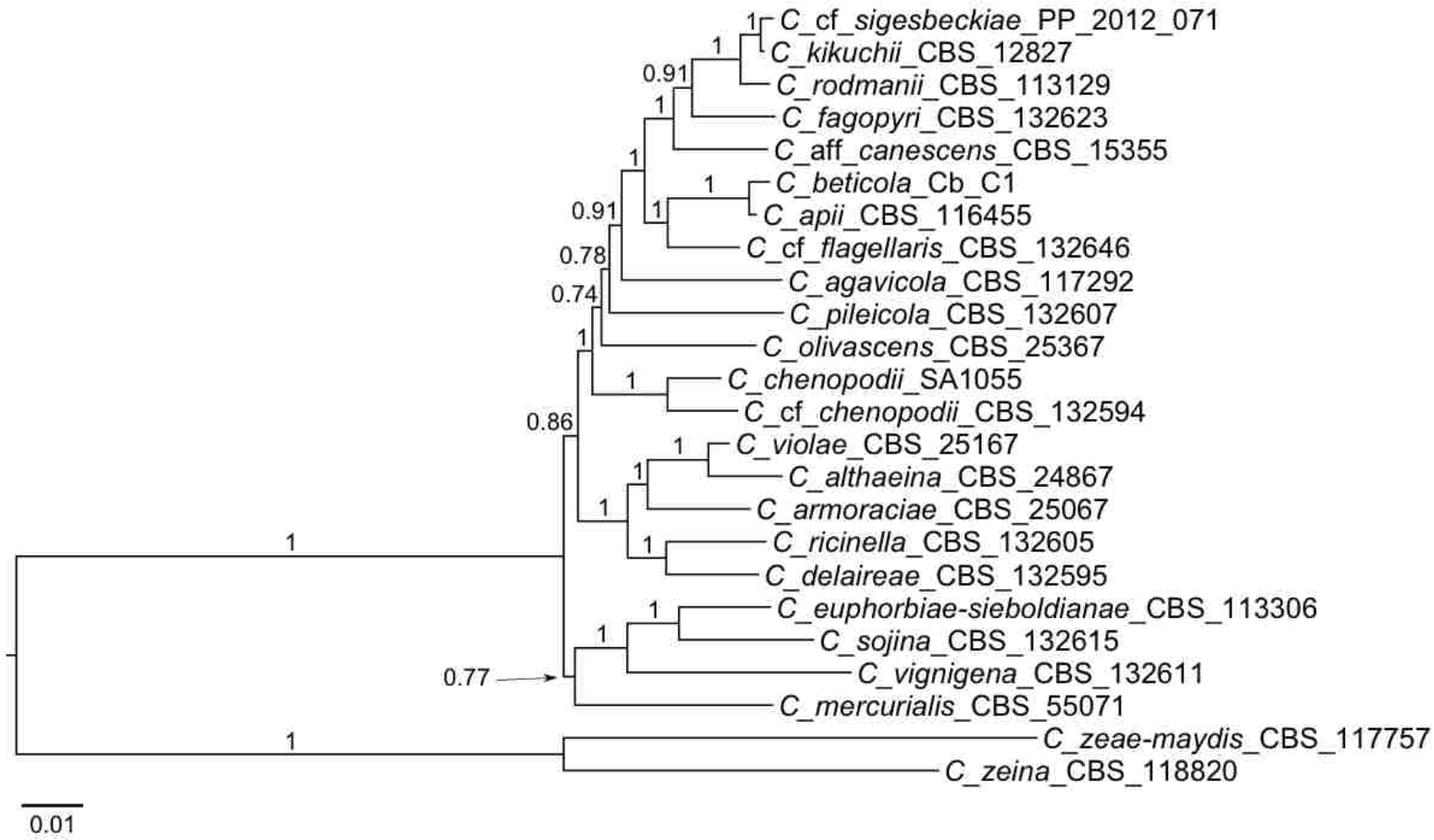


Figure A7.19. Topology inferred from Bayesian inference phylogenetic analysis of 24 species of *Cercospora* from the concatenated alignment of DS-4 using the five IGS loci with the lowest net phylogenetic informativeness values (IGS3, IGS5, IGS7, IGS8 and IGS9). Tree is rooted with *Cercospora zeina* and *C. zeae-maydis*. Posterior probability values of at least 0.70 are present at nodes. Scale bar below tree indicates the number of substitutions per site.

APPENDIX 8. MATCHING SPLITS DISTANCES BETWEEN INDIVIDUAL MAXIMUM LIKELIHOOD GENE TREES IN CHAPTER THREE

Table A8.1. Matching splits for DS-1

Locus	*No.	Tree1	Tree2	Matching Split	Mean Matching Split
IGS2	1	1	2	0	37.3
	2	1	3	56	
	3	2	3	56	
IGS3	1	1	2	40	65.0
	2	1	3	69	
	3	2	3	86	
IGS4	1	1	2	0	76.0
	2	1	3	114	
	3	2	3	114	
IGS5	1	1	2	27	56.3
	2	1	3	74	
	3	2	3	68	
IGS12	1	1	2	0	51.3
	2	1	3	77	
	3	2	3	77	
IGS6	1	1	2	4	38.7
	2	1	3	58	
	3	2	3	54	
IGS7	1	1	2	0	18.0
	2	1	3	27	
	3	2	3	27	
IGS8	1	1	2	0	10.7
	2	1	3	16	
	3	2	3	16	
IGS9	1	1	2	20	26.0
	2	1	3	22	
	3	2	3	36	

IGS10	1	1	2	0	100.0
	2	1	3	150	
	3	2	3	150	
IGS11	1	1	2	17	43.0
	2	1	3	60	
	3	2	3	52	
ACT	1	1	2	1	36.7
	2	1	3	55	
	3	2	3	54	
Btub	1	1	2	0	34.7
	2	1	3	52	
	3	2	3	52	
CAL	1	1	2	0	28.0
	2	1	3	42	
	3	2	3	42	
EF1	1	1	2	2	40.0
	2	1	3	58	
	3	2	3	60	
H3	1	1	2	0	12.7
	2	1	3	19	
	3	2	3	19	
ITS	1	1	2	0	0.0
	2	1	3	0	
	3	2	3	0	

Table A8.2. Matching splits for DS-3

Locus	*No.	Tree1	Tree2	Matching Split	Mean Matching Split
IGS2	1	1	2	0	0.0
	2	1	3	0	
	3	2	3	0	
IGS3	1	1	2	0	2.0

	2	1	3	3	
	3	2	3	3	
IGS4	1	1	2	0	0.0
	2	1	3	0	
	3	2	3	0	
IGS5	1	1	2	0	0.0
	2	1	3	0	
	3	2	3	0	
IGS12	1	1	2	0	0.0
	2	1	3	0	
	3	2	3	0	
IGS6	1	1	2	0	0.0
	2	1	3	0	
	3	2	3	0	
IGS7	1	1	2	9	6.0
	2	1	3	0	
	3	2	3	9	
IGS8	1	1	2	0	0.0
	2	1	3	0	
	3	2	3	0	
IGS9	1	1	2	0	0.0
	2	1	3	0	
	3	2	3	0	
IGS10	1	1	2	0	1.3
	2	1	3	2	
	3	2	3	2	
IGS11	1	1	2	0	0.0
	2	1	3	0	
	3	2	3	0	
ACT	1	1	2	0	4.7
	2	1	3	7	
	3	2	3	7	
Btub	1	1	2	0	0.0

	2	1	3	0	
	3	2	3	0	
CAL	1	1	2	0	0.0
	2	1	3	0	
	3	2	3	0	
EF1	1	1	2	2	1.3
	2	1	3	0	
	3	2	3	2	
H3	1	1	2	0	3.3
	2	1	3	5	
	3	2	3	5	
ITS	1	1	2	0	0.0
	2	1	3	0	
	3	2	3	0	

Table A8.3. Matching splits for DS-4

Locus	*No.	Tree1	Tree2	Matching Split	Mean Matching Split
IGS1	1	1	2	25	34.7
	2	1	3	42	
	3	2	3	37	
IGS2	1	1	2	17	28.0
	2	1	3	38	
	3	2	3	29	
IGS3	1	1	2	4	25.3
	2	1	3	38	
	3	2	3	34	
IGS4	1	1	2	14	50.0
	2	1	3	72	
	3	2	3	64	
IGS5	1	1	2	0	18.0
	2	1	3	27	

	3	2	3	27	
IGS6	1	1	2	16	27.3
	2	1	3	30	
	3	2	3	36	
IGS7	1	1	2	0	2.0
	2	1	3	3	
	3	2	3	3	
IGS8	1	1	2	0	5.3
	2	1	3	8	
	3	2	3	8	
IGS9	1	1	2	0	9.3
	2	1	3	14	
	3	2	3	14	
IGS10	1	1	2	13	42.0
	2	1	3	57	
	3	2	3	56	
IGS11	1	1	2	8	34.0
	2	1	3	48	
	3	2	3	46	
ACT	1	1	2	14	26.7
	2	1	3	37	
	3	2	3	29	
CAL	1	1	2	16	10.7
	2	1	3	0	
	3	2	3	16	
EF1	1	1	2	2	18.7
	2	1	3	26	
	3	2	3	28	
H3	1	1	2	0	10.0
	2	1	3	15	
	3	2	3	15	
ITS	1	1	2	0	1.3
	2	1	3	2	

3	2	3	2

* No.1 = NoGb; No.2 = LSGb; No.3 = MSGb

APPENDIX 9. POSTERIOR PROBABILITY VALUES FOR CLADES IN CHAPTER THREE PHYLOGENETIC TREES

Table A9.1. DS-1 “all loci”

Clade	NoGb	LSGb	MSGb
1	1.0	1.0	1.0
1sc1	1.0	1.0	*
1sc2	1.0	1.0	*
1sc3	1.0	0.97	*
2	1.0	1.0	**
3	1.0	1.0	1.0
4	1.0	1.0	1.0
5	1.0	1.0	1.0
6	1.0	1.0	1.0
7	1.0	1.0	1.0
8	1.0	1.0	1.0
9	**	**	0.97

Table A9.2. DS-1 “IGS”

Clade	NoGb	LSGb	MSGb
1	1.0	1.0	1.0
1sc1	1.0	1.0	1.0
1sc2	1.0	1.0	**
1sc3	**	**	**
2	**	**	**
3	1.0	1.0	1.0
4	1.0	1.0	1.0

5	1.0	1.0	1.0
6	1.0	1.0	1.0
7	1.0	1.0	1.0
8	1.0	1.0	1.0
9	**	**	**

Table A9.3. DS-1 “legacy”

Clade	NoGb	LSGb	MSGb
1	1.0	1.0	1.0
1sc1	**	**	**
1sc2	1.0	0.95	0.92
1sc3	0.64	0.72	0.57
2	0.97	0.99	**
3	1.0	1.0	1.0
4	1.0	1.0	0.99
5	1.0	1.0	1.0
6	1.0	1.0	1.0
7	1.0	1.0	1.0
8	1.0	1.0	1.0
9	**	**	**

Table A9.4. DS-3 “all loci”

Clade	NoGb	LSGb	MSGb
1	1.0	1.0	1.0
1sc1	0.94	1.0	1.0
1sc2	**	0.83	0.85

1sc3	0.7	**	**

Table A9.5. DS-3 “IGS”

Clade	NoGb	LSGb	MSGb
1	1.0	1.0	1.0
1sc1	1.0	1.0	1.0
1sc2	**	**	**
1sc3	0.94	0.73	0.6

Table A9.6. DS-3 “legacy”

Clade	NoGb	LSGb	MSGb
1	1.0	1.0	1.0
1sc1	**	**	**
1sc2	0.85	0.89	0.98
1sc3	0.58	0.57	**

Table A9.7. DS-4 “all loci”

Clade	NoGb	LSGb	MSGb
3	1.0	1.0	1.0
4	1.0	1.0	1.0
5	1.0	1.0	1.0
6	1.0	1.0	1.0
7	1.0	1.0	1.0

8	1.0	1.0	1.0
10	1.0	0.98	**
11	**	**	1.0
12	**	**	<50

Table A9.8. DS-4 “IGS”

Clade	NoGb	LSGb	MSGb
3	1.0	1.0	1.0
4	1.0	1.0	1.0
5	1.0	1.0	1.0
6	1.0	1.0	1.0
7	1.0	1.0	1.0
8	1.0	1.0	1.0
10	1.0	1.0	**
11	**	**	<50
12	**	**	0.74

Table A9.9 DS-4 “legacy”

Clade	NoGb	LSGb	MSGb
3	1.0	1.0	1.0
4	0.99	0.95	0.96
5	0.98	0.98	0.99
6	0.89	0.91	0.88
7	0.98	0.97	0.99
8	1.0	1.0	1.0

10	**	**	**
11	**	**	**
12	**	**	0.82

Clade designations

1- *C.*cf. *flagellaris*
1sc1- *C.* cf. *flagellaris* sub-clade 1
1sc2- *C.* cf. *flagellaris* sub-clade 2
1sc3- *C.* cf. *flagellaris* sub-clade 3
2- *C.* aff. *canescens, C. olivascens*
3- *C. apii, C. beticola*
4- *C. kikuchii, C.* cf. *sigesbeckiae, C. rodmanii, C. fagopyri*
5- *C. chenopodii, C.* cf. *chenopodii*
6- *C. ricinella, C. delaireae, C. armoraciae, C. violae, C. althaeina*
7- *C. sojina, C. euphorbiae-sieboldiabae, C. vignigena*
8- *C. zeae-maydis, C. zeina*
9- *C. olivascens, C. pileicola*
10- *C.* aff.*canescens, C.* cf. *flagellaris, C. olivascens*
11- *C. apii, C. beticola, C.* cf. *flagellaris*
12- *C. agavicola, C.* cf. *flagellaris*

* not mono within flagellaris subclades
** clade not present

APPENDIX 10. CLADE COMPOSITIONS IN CHAPTER THREE PHYLOGENETIC TREES

Table A10.1. Clade compositions for DS-1 “all loci”

	Clade		
Isolate	NoGb	LSGb	MSGb
C. aff. *canescens* CBS 153.55	2	2	4
C. olivascens CBS 253.67	2	2	9
C. apii CBS 116455	3	3	3
C. beticola Cb_C1	3	3	3
C. cf. *sigesbeckiae* PP_2012_071	4	4	4
C. fagopyri CBS 132623	4	4	4
C. kikuchii CBS 128.27	4	4	4
C. kikuchii SA1073	4	4	4
C. rodmanii CBS 113129	4	4	4
C. cf. *chenopodii* CBS 132594	5	5	5
C. chenopodii SA1055	5	5	5
C. althaeina CBS 248.67	6	6	6
C. armoraciae CBS 250.67	6	6	6
C. delaireae CBS 132595	6	6	6
C. ricinella CBS 132605	6	6	6
C. violae CBS 251.67	6	6	6
C. euphorbiae-sieboldianae CBS 113306	7	7	7
C. sojina 223	7	7	7
C. sojina CBS 132615	7	7	7
C. vignigena CBS 132611	7	7	7
C. zeae-maydis CBS 117757	8	8	8
C. zeae-maydis CZM-SCOH	8	8	8
C. zeina CBS 118820	8	8	8
C. zeina sp_O	8	8	8

C. cf. *flagellaris* PP_2011_003	1fsc1	1fsc1	1fsc1
C. cf. *flagellaris* PP_2011_008	1fsc1	1fsc1	1fsc1
C. cf. *flagellaris* PP_2011_041	1fsc1	1fsc1	1fsc1
C. cf. *flagellaris* PP_2011_048	1fsc1	1fsc1	1fsc1
C. cf. *flagellaris* SA1083	1fsc1	1fsc1	1fsc1
C. cf. *flagellaris* ARCK7	1fsc2	1fsc2	1*
C. cf. *flagellaris* PP_2000_004	1fsc2	1fsc2	1*
C. cf. *flagellaris* PP_2011_007	1fsc2	1fsc2	1*
C. cf. *flagellaris* PP_2011_045	1fsc2	1fsc2	1*
C. cf. *flagellaris* PSS_13_1	1fsc2	1fsc2	1*
C. cf. *flagellaris* PSS_13_2a	1fsc2	1fsc2	1*
C. cf. *flagellaris* SA1019	1fsc2	1fsc2	1*
C. cf. *flagellaris* SA1080	1fsc2	1fsc2	1*
C. cf. *flagellaris* CBS 132646	1fsc3	1fsc3	1*
C. cf. *flagellaris* SA1088	1fsc3	1fsc3	1*
C. cf. *flagellaris* CBS 132674	1S	1S	1*
C. agavicola CBS 117292	S	S	S
C. mercurialis CBS 550.71	S	S	S
C. pileicola CBS 132607	S	S	9

Table A10.2. Clade compositions for DS-1 “IGS”

	Clade		
Isolate	NoGb	LSGb	MSGb
C. aff. *canescens* CBS 153.55	S	S	S
C. olivascens CBS 253.67	S	S	S
C. apii CBS 116455	3	3	3
C. beticola Cb_C1	3	3	3
C. cf. *sigesbeckiae* PP_2012_071	4	4	4

C. fagopyri CBS 132623	4	4	4
C. kikuchii CBS 128.27	4	4	4
C. kikuchii SA1073	4	4	4
C. rodmanii CBS 113129	4	4	4
C. cf. *chenopodii* CBS 132594	5	5	5
C. chenopodii SA1055	5	5	5
C. althaeina CBS 248.67	6	6	6
C. armoraciae CBS 250.67	6	6	6
C. delaireae CBS 132595	6	6	6
C. ricinella CBS 132605	6	6	6
C. violae CBS 251.67	6	6	6
C. euphorbiae-sieboldianae CBS 113306	7	7	7
C. sojina 223	7	7	7
C. sojina CBS 132615	7	7	7
C. vignigena CBS 132611	7	7	7
C. zeae-maydis CBS 117757	8	8	8
C. zeae-maydis CZM-SCOH	8	8	8
C. zeina CBS 118820	8	8	8
C. zeina sp_O	8	8	8
C. cf. *flagellaris* PP_2011_003	1fsc1	1fsc1	1fsc1
C. cf. *flagellaris* PP_2011_008	1fsc1	1fsc1	1fsc1
C. cf. *flagellaris* PP_2011_041	1fsc1	1fsc1	1fsc1
C. cf. *flagellaris* PP_2011_048	1fsc1	1fsc1	1fsc1
C. cf. *flagellaris* SA1083	1fsc1	1fsc1	1fsc1
C. cf. *flagellaris* ARCK7	1fsc2	1fsc2	1*
C. cf. *flagellaris* PP_2000_004	1fsc2	1fsc2	1*
C. cf. *flagellaris* PP_2011_007	1fsc2	1fsc2	1*
C. cf. *flagellaris* PP_2011_045	1fsc2	1fsc2	1*
C. cf. *flagellaris* PSS_13_1	1fsc2	1fsc2	1*
C. cf. *flagellaris* PSS_13_2a	1fsc2	1fsc2	1*
C. cf. *flagellaris* SA1019	1fsc2	1fsc2	1*

C. cf. *flagellaris* SA1080	1fsc2	1fsc2	1*
C. cf. *flagellaris* CBS 132646	1*	1*	1*
C. cf. *flagellaris* SA1088	1*	1*	1*
C. cf. *flagellaris* CBS 132674	1*	1*	1*
C. agavicola CBS 117292	S	S	S
C. mercurialis CBS 550.71	S	S	S
C. pileicola CBS 132607	S	S	S

Table A10.3. Clade compositions for DS-1 “legacy”

	Clade		
Isolate	NoGb	LSGb	MSGb
C. aff. *canescens* CBS 153.55	2	2	S
C. olivascens CBS 253.67	2	2	S
C. apii CBS 116455	3	3	3
C. beticola Cb_C1	3	3	3
C. cf. *sigesbeckiae* PP_2012_071	4	4	4
C. fagopyri CBS 132623	4	4	4
C. kikuchii CBS 128.27	4	4	4
C. kikuchii SA1073	4	4	4
C. rodmanii CBS 113129	4	4	4
C. cf. *chenopodii* CBS 132594	5	5	5
C. chenopodii SA1055	5	5	5
C. althaeina CBS 248.67	6	6	6
C. armoraciae CBS 250.67	6	6	6
C. delaireae CBS 132595	6	6	6
C. ricinella CBS 132605	6	6	6
C. violae CBS 251.67	6	6	6
C. euphorbiae-sieboldianae CBS 113306	7	7	7
C. sojina 223	7	7	7

C. sojina CBS 132615	7	7	7
C. vignigena CBS 132611	7	7	7
C. zeae-maydis CBS 117757	8	8	8
C. zeae-maydis CZM-SCOH	8	8	8
C. zeina CBS 118820	8	8	8
C. zeina sp_O	8	8	8
C. cf. *flagellaris* PP_2011_003	1fsc1	1fsc1	1fsc1
C. cf. *flagellaris* PP_2011_008	1*	1*	1*
C. cf. *flagellaris* PP_2011_041	1*	1*	1*
C. cf. *flagellaris* PP_2011_048	1fsc1	1fsc1	1fsc1
C. cf. *flagellaris* SA1083	1fsc1	1fsc1	1fsc1
C. cf. *flagellaris* ARCK7	1fsc2	1fsc2	1fsc2
C. cf. *flagellaris* PP_2000_004	1fsc2	1fsc2	1fsc2
C. cf. *flagellaris* PP_2011_007	1fsc2	1fsc2	1fsc2
C. cf. *flagellaris* PP_2011_045	1fsc2	1fsc2	1fsc2
C. cf. *flagellaris* PSS_13_1	1fsc2	1fsc2	1fsc2
C. cf. *flagellaris* PSS_13_2a	1fsc2	1fsc2	1fsc2
C. cf. *flagellaris* SA1019	1fsc2	1fsc2	1fsc2
C. cf. *flagellaris* SA1080	1fsc2	1fsc2	1fsc2
C. cf. *flagellaris* CBS 132646	1fsc3	1*	1*
C. cf. *flagellaris* SA1088	1fsc3	1*	1*
C. cf. *flagellaris* CBS 132674	1fsc2	1fsc2	1fsc2
C. agavicola CBS 117292	S	S	S
C. mercurialis CBS 550.71	S	S	S
C. pileicola CBS 132607	S	S	S

Table A10.4. Clade compositions for DS-3 “all loci”

	Clade		
Isolate	NoGb	LSGb	MSGb

C. cf. *flagellaris* PP_2011_003	1fsc1	1fsc1	1fsc1
C. cf. *flagellaris* PP_2011_008	1fsc1	1fsc1	1fsc1
C. cf. *flagellaris* PP_2011_041	1fsc1	1fsc1	1fsc1
C. cf. *flagellaris* PP_2011_048	1fsc1	1fsc1	1fsc1
C. cf. *flagellaris* SA1083	1fsc1	1fsc1	1fsc1
C. cf. *flagellaris* ARCK7	1*	1fsc2	1fsc2
C. cf. *flagellaris* PP_2000_004	1*	1fsc2	1fsc2
C. cf. *flagellaris* PP_2011_007	1*	1fsc2	1fsc2
C. cf. *flagellaris* PP_2011_045	1*	1fsc2	1fsc2
C. cf. *flagellaris* PSS_13_1	1*	1fsc2	1fsc2
C. cf. *flagellaris* PSS_13_2a	1*	1fsc2	1fsc2
C. cf. *flagellaris* SA1019	1*	1fsc2	1fsc2
C. cf. *flagellaris* SA1080	1*	1fsc2	1fsc2
C. cf. *flagellaris* CBS 132646	1fsc3	1*	1*
C. cf. *flagellaris* SA1088	1fsc3	1*	1*
C. cf. *flagellaris* CBS 132674	1*	1*	1*

Table A10.5. Clade compositions for DS-3 “IGS”

	Clade		
Isolate	NoGb	LSGb	MSGb
C. cf. *flagellaris* PP_2011_003	1fsc1	1fsc1	1fsc1
C. cf. *flagellaris* PP_2011_008	1fsc1	1fsc1	1fsc1
C. cf. *flagellaris* PP_2011_041	1fsc1	1fsc1	1fsc1
C. cf. *flagellaris* PP_2011_048	1fsc1	1fsc1	1fsc1
C. cf. *flagellaris* SA1083	1fsc1	1fsc1	1fsc1
C. cf. *flagellaris* ARCK7	1*	1*	1fsc2
C. cf. *flagellaris* PP_2000_004	1*	1*	1fsc2
C. cf. *flagellaris* PP_2011_007	1*	1*	1fsc2
C. cf. *flagellaris* PP_2011_045	1*	1*	1fsc2

C. cf. *flagellaris* PSS_13_1	1*	1*	1fsc2
C. cf. *flagellaris* PSS_13_2a	1*	1*	1fsc2
C. cf. *flagellaris* SA1019	1*	1*	1fsc2
C. cf. *flagellaris* SA1080	1*	1*	1fsc2
C. cf. *flagellaris* CBS 132646	1fsc3	1fsc3	1*
C. cf. *flagellaris* SA1088	1fsc3	1fsc3	1*
C. cf. *flagellaris* CBS 132674	1*	1*	1*

Table A10.6. Clade compositions for DS-3 “legacy”

	Clade		
Isolate	NoGb	LSGb	MSGb
C. cf. *flagellaris* PP_2011_003	1fsc1	1fsc1	1fsc1
C. cf. *flagellaris* PP_2011_008	1*	1*	1*
C. cf. *flagellaris* PP_2011_041	1*	1*	1*
C. cf. *flagellaris* PP_2011_048	1fsc1	1fsc1	1fsc1
C. cf. *flagellaris* SA1083	1fsc1	1fsc1	1fsc1
C. cf. *flagellaris* ARCK7	1fsc2	1fsc2	1fsc2
C. cf. *flagellaris* PP_2000_004	1fsc2	1fsc2	1fsc2
C. cf. *flagellaris* PP_2011_007	1fsc2	1fsc2	1fsc2
C. cf. *flagellaris* PP_2011_045	1fsc2	1fsc2	1fsc2
C. cf. *flagellaris* PSS_13_1	1fsc2	1fsc2	1fsc2
C. cf. *flagellaris* PSS_13_2a	1fsc2	1fsc2	1fsc2
C. cf. *flagellaris* SA1019	1fsc2	1fsc2	1fsc2
C. cf. *flagellaris* SA1080	1fsc2	1fsc2	1fsc2
C. cf. *flagellaris* CBS 132646	1fsc3	1fsc3	1*
C. cf. *flagellaris* SA1088	1fsc3	1fsc3	1*
C. cf. *flagellaris* CBS 132674	1fsc2	1fsc2	1fsc2

Table A10.7. Clade compositions for DS-4 “all loci”

Isolate	NoGb	LSGb	MSGb
C. aff. *canescens* CBS 153.55	10	S	S
C. olivascens CBS 253.67	10	10	9
C. apii CBS 116455	3	3	3
C. beticola Cb_C1	3	3	3
C. cf. *sigesbeckiae* PP_2012_071	4	4	4
C. fagopyri CBS 132623	4	4	4
C. kikuchii CBS 128.27	4	4	4
C. rodmanii CBS 113129	4	4	4
C. cf. *chenopodii* CBS 132594	5	5	5
C. chenopodii SA1055	5	5	5
C. althaeina CBS 248.67	6	6	6
C. armoraciae CBS 250.67	6	6	6
C. delaireae CBS 132595	6	6	6
C. ricinella CBS 132605	6	6	6
C. violae CBS 251.67	6	6	6
C. euphorbiae-sieboldianae CBS 113306	7	7	7
C. sojina CBS 132615	7	7	7
C. vignigena CBS 132611	7	7	7
C. zeae-maydis CBS 117757	8	8	8
C. zeina CBS 118820	8	8	8
C. cf. *flagellaris* CBS 132646	10	10	11
C. agavicola CBS 117292	S	S	S
C. mercurialis CBS 550.71	6	S	S
C. pileicola CBS 132607	S	S	S

Table A10.8. Clade compositions for DS-4 “IGS”

Isolate	NoGb	LSGb	MSGb
C. aff. *canescens* CBS 153.55	10	10	S
C. olivascens CBS 253.67	10	10	9
C. apii CBS 116455	3	3	3
C. beticola Cb_C1	3	3	3
C. cf. *sigesbeckiae* PP_2012_071	4	4	4
C. fagopyri CBS 132623	4	4	4
C. kikuchii CBS 128.27	4	4	4
C. rodmanii CBS 113129	4	4	4
C. cf. *chenopodii* CBS 132594	5	5	5
C. chenopodii SA1055	5	5	5
C. althaeina CBS 248.67	6	6	6
C. armoraciae CBS 250.67	6	6	6
C. delaireae CBS 132595	6	6	6
C. ricinella CBS 132605	6	6	6
C. violae CBS 251.67	6	6	6
C. euphorbiae-sieboldianae CBS 113306	7	7	7
C. sojina CBS 132615	7	7	7
C. vignigena CBS 132611	7	7	7
C. zeae-maydis CBS 117757	8	8	8
C. zeina CBS 118820	8	8	8
C. cf. *flagellaris* CBS 132646	10	10	11
C. agavicola CBS 117292	S	S	S
C. mercurialis CBS 550.71	S	S	S
C. pileicola CBS 132607	S	S	S

Table A10.9. Clade compositions for DS-4 “legacy”

Isolate	NoGb	LSGb	MSGb
C. aff. *canescens* CBS 153.55	S	S	S

C. olivascens CBS 253.67	S	S	S
C. apii CBS 116455	3	3	3
C. beticola Cb_C1	3	3	3
C. cf. *sigesbeckiae* PP_2012_071	4	4	4
C. fagopyri CBS 132623	4	4	4
C. kikuchii CBS 128.27	4	4	4
C. rodmanii CBS 113129	4	4	4
C. cf. *chenopodii* CBS 132594	5	5	5
C. chenopodii SA1055	5	5	5
C. althaeina CBS 248.67	6	6	6
C. armoraciae CBS 250.67	6	6	6
C. delaireae CBS 132595	6	6	6
C. ricinella CBS 132605	6	6	6
C. violae CBS 251.67	6	6	6
C. euphorbiae-sieboldianae CBS 113306	7	7	7
C. sojina CBS 132615	7	7	7
C. vignigena CBS 132611	7	7	7
C. zeae-maydis CBS 117757	8	8	8
C. zeina CBS 118820	8	8	8
C. cf. *flagellaris* CBS 132646	12	7	12
C. agavicola CBS 117292	12	S	12
C. mercurialis CBS 550.71	S	9	S
C. pileicola CBS 132607	S	9	S

Clade designations
1- *C.* cf. *flagellaris*
1sc1- *C.* cf. *flagellaris* sub-clade 1
1sc2- *C.* cf. *flagellaris* sub-clade 2
1sc3- *C.* cf. *flagellaris* sub-clade 3
2- *C.* aff. *canescens*, *C. olivascens*

3- *C. apii, C. beticola*
5- *C. chenopodii, C.* cf. *chenopodii*
4- *C. kikuchii, C.* cf. *sigesbeckiae, C. rodmanii, C. fagopyri*
6- *C. ricinella, C. delaireae, C. armoraciae, C. violae, C. althaeina*
7- *C. sojina, C. euphorbiae-sieboldianae, C. vignigena*
8- *C. zeae-maydis, C. zeina*
9- *C. olivascens, C. pileicola*
10- *C.* aff. *canescens, C.* cf. *flagellaris, C. olivascens*
11- *C. apii, C. beticola, C.* cf. *flagellaris*
12- *C. agavicola, C.* cf. *flagellaris*
* not mono within *C.* cf. *flagellaris* subclades
S- separate lineage

VITA

Sebastian Albu was born in Bucharest, Romania. He received a B.A. in Music from the University of Pittsburgh in 2001 and a B.S. in Biology from the Metropolitan State University of Denver in 2005. He enrolled in the Department of Plant Pathology and Crop Physiology in at Louisiana State University 2010 and earned a M.S. in Plant Health in 2012.