MYCOLOGIA

Mycologia

ISSN: 0027-5514 (Print) 1557-2536 (Online) Journal homepage: https://www.tandfonline.com/loi/umyc20

An updated phylogenetic classification of *Corynespora cassiicola* isolates and a practical approach to their identification based on the nucleotide polymorphisms at the *ga4* and *caa5* loci

Alexander Banguela-Castillo, Pedro L. Ramos-González, Mabel Peña-Marey, Claudia V. Godoy & Ricardo Harakava

To cite this article: Alexander Banguela-Castillo, Pedro L. Ramos-González, Mabel Peña-Marey, Claudia V. Godoy & Ricardo Harakava (2020) An updated phylogenetic classification of *Corynespora cassiicola* isolates and a practical approach to their identification based on the nucleotide polymorphisms at the *ga4* and *caa5* loci, Mycologia, 112:1, 24-38, DOI: 10.1080/00275514.2019.1670018

To link to this article: https://doi.org/10.1080/00275514.2019.1670018

MYCOLOGIA
2020, VOL. 112, NO. 1, 24–38
https://doi.org/10.1080/00275514.2019.1670018

Taylor & Francis
Taylor & Francis Group

An updated phylogenetic classification of *Corynespora cassiicola* isolates and a practical approach to their identification based on the nucleotide polymorphisms at the *ga4* and *caa5* loci

Alexander Banguela-Castillo [a], Pedro L. Ramos-González [a], Mabel Peña-Marey[b,c], Claudia V. Godoy [d], and Ricardo Harakava[a]

[a]Phytopathology and Plant Biochemistry Laboratory, Instituto Biológico de São Paulo, Avenida Conselheiro Rodrigues Alves, 1252 Vila Mariana, CEP 04014-900, São Paulo, São Paulo, Brazil; [b]Microbiology and Bacteriology Laboratory, St. Joseph's Hospital, 3001 W Martin Luther King Jr. Boulevard, Tampa, Florida 33607; [c]Instituto de Investigaciones en Fruticultura Tropical, Avenida 7ma 3005, Playa, La Habana 10500, Cuba; [d]Embrapa Soja, Rodovia Carlos João Strass, s/nº Acesso Orlando Amaral, Distrito de Warta Caixa, Postal: 231, CEP: 86001-970, Londrina, Paraná, Brazil

ABSTRACT

Corynespora cassiicola (Burk. & M.A. Curtis) C.T. Wei. is an anamorphic fungus that affects more than 530 plant species, including economically important crops. Several lineages of this pathogen have been recognized, but the classification of isolates into clades is time-consuming and still sometimes leads to unclear results. In this work, eight major phylogenetic clades (PhL1–PhL8) including 245 isolates of *C. cassiicola* from 44 plant species were established based on a Bayesian inference analysis of four combined *C. cassiicola* genomic loci retrieved from GenBank, i.e., rDNA internal transcribed spacer (ITS), *actin-1*,*ga4*, and *caa5*. The existence of PhL1–PhL5 and PhL7 as clonal lineages was further confirmed through the analysis of full-genome single-nucleotide polymorphisms of 39 isolates. Haplotypes of the *caa5* locus were PhL specific and encode isoforms of the LDB19 domain of a putative α-arrestin N-terminal–like protein. Evolution of the Caa5 arrestin is in correspondence with the PhLs. *ga4* and *caa5* PhL consensus sequences and a cleaved amplified polymorphic sequence (CAPS) procedure were generated based on the conserved nucleotide sequences and enzyme restriction patterns observed among isolates from the same lineage, respectively. The CAPS method was validated *in silico*, and its practical use allowed us to differentiate between tomato and papaya isolates, as well as to reveal the prevalence of PhL1 among isolates infecting soybean in Brazil. This novel approach could be useful in the efforts to control the diseases associated with *C. cassiicola*.

ARTICLE HISTORY
Received 14 February 2019
Accepted 17 September 2019

KEYWORDS
Corynespora cassiicola; phylogeny; polymorphism; genome SNP; CAPS; arrestin

INTRODUCTION

Corynespora cassiicola (Burk. & M.A. Curtis) C.T. Wei. is an anamorphic haploid ascomycete of the order *Dothideomycetes* that besides infecting leaves, stems, and roots of more than 530 plant species worldwide (Farr and Rossman 2018) affects nematode cysts and the human skin (Dixon et al. 2009; Schoch et al. 2009; Looi et al. 2017). This fungus produces major economic losses in important crops such as papaya (*Carica papaya* L.), rubber tree (*Hevea brasiliensis* (Willd. A. Juss.) Müll. Arg.), tomato (*Solanum lycopersum* L.), soybean (*Glycine max* L.), and cotton (*Gossypium hirsutum* L.) (Déon et al. 2014).

Former stages of the taxonomic study of *C. cassiicola* revealed the occurrence of distinguishing pathotypes (Onesirosan et al. 1974), but a clear relationship between plant hosts or the morphological traits of the fungus was not established (Silva et al. 2006). Other analyses using random amplified polymorphic DNA (RAPD), rDNA restriction fragment length polymorphism (RFLP), rDNA internal transcribed spacer single-nucleotide polymorphism (ITS-SNP), or short tandem repeat (SSR) markers defined pools in which the genetic variability of their members better harmonized with features such as pathogenicity, physiology, geographic origin, or host range. Nonetheless, in most of those cases, a direct and holistic relationship between isolates inside each group still remained elusive (Saha et al. 1988; Silva et al. 2003, 2006; Romruensukharom et al. 2005; Nghia et al. 2008, 2010; Qi et al. 2009; Shimomoto et al. 2011; Hieu et al. 2014; Oktavia et al. 2017; Sumabat et al. 2018b). Further efforts based on combined phylogenetic analyses of four genomic loci,

CONTACT Alexander Banguela-Castillo elbangue@gmail.com
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/umyc.
Supplemental data for this article can be accessed on the publisher's Web site.

Published online 21 Nov 2019

including ribosomal DNA internal transcribed spacer (rDNA ITS), two random hypervariable loci (*caa5* and *ga4*), and an actin-1-encoding locus (*act1*), demarcated six phylogenetic lineages of *C. cassiicola* (Dixon et al. 2009). These lineages comprised isolates from a wide geographic distribution that, in addition, shared similar primary hosts, pathogenic profiles, and growth rates. Based on the same genome regions and also on the variability of the cassiicolin genes (*cas1*–*cas6*), eight and six phylogenetic clades, respectively, of *C. cassiicola* isolates that mainly infect rubber trees were defined (Déon et al. 2014). Cassiicolin is a glycoprotein that acts as a phytotoxic effector and mediates the fungus pathogenicity in rubber tree (Déon et al. 2012). Isolates harboring the *cas1* variant are more pathogenic in rubber (Déon et al. 2014).

Recently, single-nucleotide polymorphism (SNP) analysis of isolates collected from infected cotton and soybean plants revealed the existence of eight unique multilocus genotypes (Shrestha et al. 2017). The SNP study also indicated the prevalence of a single genotype attacking cotton in the southeastern region of the United States, which was also confirmed in a more recent work by the genotyping of isolates with 13 microsatellite markers (Sumabat et al. 2018a). Moreover, a comprehensive phylogenomic analysis encompassing the putative core proteins from the genomes of 35 isolates determined the existence of six phylogenetic clades, which are analogous to those described earlier by Dixon et al. (2009) (Lopez et al. 2018). Despite substantial advances of this study, the lack of phenotypic information such as the virulence profiles and host ranges of the sequenced isolates limited the identification of the appropriate effectors on which phenotyping and/or genotyping tools could be developed (Lopez et al. 2018).

Up to now, the classification of *C. cassiicola* isolates based on the analysis of genome regions has been the more effective approach grouping isolates according to their hosts origins; in addition, this is the method through which the largest number of samples has been analyzed (Dixon et al. 2009; Déon et al. 2014; Sumabat et al. 2018a; Wu et al. 2018). Still, there is not a holistic, simple, and practical procedure available for the routine classification of *C. cassiicola* isolates. Traceability of *C. cassiicola* lineages by a rapid, affordable, and robust genetic methodology that allows identifying potential sources of inoculum is vital for an efficient disease management (Dixon et al. 2009; Lopez et al. 2018). In this regard, the development of a simple classification procedure might pave the application of new genomic and metagenomic techniques in the characterization of *C. cassiicola* genotypes.

In this work, we hypothesized that the nucleotide polymorphisms among phylogenetic lineages of *C. cassiicola* at the *ga4* and *caa5* locus sequences enable the development of a practical phylogenetic-based classification of worldwide isolates of this fungus. Eight phylogenetic clades (PhL1–PhL8) were established after the Bayesian inference analysis of four combined genomic loci retrieved from GenBank, i.e., rDNA ITS, *act1, ga4*, and *caa5*, which include 245 individuals of *C. cassiicola* isolated from 44 plant species. The existence of PhL1–PhL5 and PhL7, as clonal lineages, was further confirmed through the analysis of full-genome single-nucleotide polymorphisms of 39 isolates. Genome analysis indicated that the *caa5* locus is single copy and is located inside a gene encoding for the LDB19 domain of a putative α-arrestin N-terminal–like protein. Evolution of the putative Caa5 arrestin and encoded LDB19 isoforms is in correspondence with phylogenetic lineages. *ga4* and *caa5* PhL consensus sequences and a cleaved amplified polymorphic sequence (CAPS) procedure were generated based on the conserved nucleotide sequences and enzyme restriction patterns observed among isolates from the same lineage, respectively. The two-locus CAPS method was validated *in silico*, and its use allowed us to differentiate between tomato and papaya isolates, as well as to reveal the prevalence of PhL1 among isolates infecting soybean in Brazil.

MATERIALS AND METHODS

Sequences and phylogenetic analysis.—Sequences of four loci, rDNA ITS, *caa5, ga4*, and *act1*, from 387 *C. cassiicola* isolates and *Corynespora simithii* were retrieved from the GenBank database (https://www.ncbi.nlm.nih.gov/nucleotide/). *ga4*: FJ852717–FJ852735, FJ852740–FJ852859, KF810784–KF810853, MF320450, MF320454–MF320495, MF320497, MF427990–MF428016, MF428018, MF428019–MF428026, MF428028–MF428039, MF428041–MF428060, MF428062–MF428066, MF428068–MF428085, MF428087–MF428106, MF428108–MF428111, MF428114–MF428130, MH605224, MH605226–MH605227, MH605229, MH605232–MH605234, MH605239, MH605244–MH605245; *caa5*: FJ852860–FJ852878, FJ852883–FJ853002, KF810714–KF810756, KF810758, KF810760–KF810783, MF320397, MF320401–MF320412, MF320414–MF320422, MF320424–MF320427, MF320429–MF320439, MF320442, MF320444, MF427849–MF427875, MF427877–MF427883, MF427885, MF427887–MF427919, MF427921–MF427944, MF427946–MF427970, MF427973–MF427989, MH605200, MH605202–MH605203, MH605205, MH605208–MH605210, MH605215, MH605220–MH605221; *act1*: AB539439, FJ853003–FJ853141, KF810644–KF8

10686, KF810688, KF810690–KF810713, MF320356–MF320395, MF428131–MF428165, MF428167, MF428169–MF428180, MF428182–MF428201, MF428203–MF428207, MF428209–MF428226, MF428228–MF428247, MF428249–MF428252, MF428255–MF428271, MH605248, MH605250–MH605251, MH605253, MH605256–MH605258, MH605263, MH605268–MH605269; rDNA ITS: FJ852574–FJ852592, FJ852592, FJ852597–FJ852716, JQ814321, JQ814329–JQ814330, JQ814357, KF810854–KF810892, KF810894, KF810896–KF810919, MF320503, MF320507–MF320518, MF320520–MF320528, MF320530–MF320533, MF320535–MF320548, MF320550, MF428272–MF428298, MF428300–MF428306, MF428308, MF428310–MF428321, MF428323–MF428342, MF428344–MF428348, MF428350–MF428367, MF428369–MF428388, MF428390–MF428393, MF428396–MF428412, MH490951–MH490953, MH490955, MH490959–MH490964, MH490972 (Dixon et al. 2009; Deón et al. 2012; Sumabat et al. 2018a; Wu et al. 2018). Genomes: *Corynespora cassiicola* genomes from isolates CCP, UM-591, C7, and TS were gathered from GenBank using the BioProject accession numbers PRJNA234811 (Joint Genome Institute: 1019537), PRJNA236064, PRJEB19843, and PRJNA382361, respectively. The other 35 genomes were obtained from the GenBank BioProject PRJNA428435 (Lopez et al. 2018).

Multiple alignments were independently executed for each locus, gene, or protein with CLUSTAL (Larkin et al. 2007). GENEIOUS 11.1.2 (https://www.geneious.com/) was used for alignment inspections and editions, generation of consensus sequences, translations of coding sequences, and generation of and search in local BLAST databases.

Final sequences in the alignments from *caa5, act1, ga4*, and rDNA ITS involved 288, 284, 344–346, and 473 nucleotides, respectively. JMODELTEST 2.1.7 (Darriba et al. 2012) was used to determinate the best-fitting models of nucleotide substitution according to Akaike information criterion and Bayesian information criterion. For *act, ga4*, and rDNA ITS sequences, the chosen models were HKY+G, HKY+I, and JC+I, respectively. A three–codon position model with independent GTR and Gamma codon analysis was implemented for *caa5* sequences. Bayesian-inferred trees were generated by BEAST 1.7 (Drummond et al. 2012), which allowed the independent analysis of alignments to consider a specific substitution model. BEAST runs of 50×10^7 generations considering the Yule speciation prior (Steel and McKenzie 2001), and a strict clock for the single-sequence data sets were sampled every 500 000 generations. Convergence statistics for each single-gene run were analyzed by TRACER 1.6. TREEANNOTATOR 1.5.3 was used to produce the maximum clade credibility (MCC) tree from the 5×10^6 post-burn-in trees.

ga4 and *caa5* sequences from the same phylogenetic clade were grouped and aligned for the generation of consensus sequences. A total of 242 isolates out of 245 were included in the analysis. The isolate SN59 (Dixon et al. 2009) and the isolates E70 and E78 (Deón et al. 2014) were excluded from the sequence analysis, due to outgrouping and missing nucleotides, respectively. CLEAVER (http://cleaver.sourceforge.net/) and GENEIOUS 11.1.2 were used to seek for differential enzyme restriction polymorphisms among locus sequences from different lineages. Haplotypes statistics parsimony (TCS) networks (Clement et al. 2000) were generated by the independent processing of the *ga4* and *caa5* sequences belonging to each phylogenetic lineages using POPART (Leigh and Bryant 2015). The Python script clas_corca (SUPPLEMENTARY MATERIAL) was generated and used to automatically determinate the PhL of isolates based on the restriction polymorphisms at *ga4* and *caa5* sequences (TABLE 1).

ga4 and *caa5* locus sequences from the CCP isolate were blasted against the CCP-annotated genome and transcriptome at Joint Genome Institute (JGI) Portal (https://genome.jgi.doe.gov/pages/blast-query.jsf?db=Corca1). Retrieved genes were further blasted against a local BLAST database including 38 *C. cassiicola* genomes. Sequences of all 39 *caa5*-related genes were translated and aligned using GENEIOUS. Analysis of the protein alignment in MEGA7 (Kumar et al. 2016) determined Jones-Taylor-Thornton (JIT) as the best-fitting model of amino acid substitution according to Akaike information criterion and Bayesian information criterion. MEGA7 was also used to infer the evolutionary history among 39 putative Caa5 α-arrestins by using the maximum likelihood method and 1000 of bootstrap replicates. Initial tree(s) for the heuristic search were obtained automatically by applying the NJ (neighbor-joining) and BIONJ algorithms to a matrix of pairwise distances estimated using a JTT model and then selecting the topology with superior log-likelihood value. The analysis involved 40 amino acid sequences. There were a total of 449 positions in the final data set. Hypothetical α-arrestin of *Ascochyta rabiei* (GenBank KZM26630) was used as outgroup. Full-genome single-nucleotide polymorphism (SNP) was deduced by the KSNP3 package (Gardner et al. 2015). Unrooted maximum parsimony phylogenetic tree generated based on genome SNP included the 39 *C. cassiicola* genomes indicated above and considered the species core SNP and a *k*-mer of 15 (as

Table 1. *In silico* analysis of the restriction enzyme polymorphisms at the *ga4* and *caa5* loci through eight phylogenetic clades (PhL) of *Corynespora cassiicola* and list of economically important host associated with each clade.

	ga4					*caa5*				
Clade	AluI	BsrI	BstZ17I	RsaI	TaqI	BstZI	BstEII	NaeI	SalI	Economically important hosts
PhL1		X[b]	X[c]							Pa, Ru, So[†], Co
PhL2				X	X		X[d]		X	Ru
PhL3	X		X		X					To, Ru, So, Pa*
PhL4	X				X			X		To, Ru
PhL5			X		X					Ru, So[†]
PhL6	[a]	X[a]	X		X[a]	X				
PhL7[‡]		X	X		X					Ru
PhL8			X		X		X			

Note. Only enzyme cut sites that were polymorphic, through the pairwise comparisons among sequences from different PhL, were considered. Letter "X" indicates the presence of the enzyme restriction site on sequences. Common hosts of isolates from each PhL are indicated in the last column: *Carica papaya* (Pa), *Gossypium hirsutum* (Co), *Hevea brasiliensis* (Ru), *Glycine max* (So) and *Solanum lycopersicum* (To); from Dixon et al. (2009), Deón et al. (2014), and Sumabat et al. (2018a).

[a]Isolate FL51 (Dixon et al. 2009) has a *ga4* restriction pattern different from that shown by the rest of isolates of the clade PhL6; this isolate has an AluI site and lacks BsrI and TaqI sites. [b,c]BstZ17I and BsRI cut sites characteristic of PhL1 are missing in the isolates [b]GU70 and [c]YP29 (Dixon et al., 2009).

[d]BstEII site is missing from CABI 132039.

*All isolates from papaya were collected from infected leaves (PhL1), with the exception of the isolate AS117 (PhL3), which was obtained from fruit (Dixon et al. 2009).

[‡]This PhL is in concordance with clade C (Deón et al. 2014), which grouped the isolates most virulent to *H. brasiliensis*.

[†]Include the results of this work, considering our approach based on the analysis of *ga4* and *caa5* locus sequences from 25 Brazilians isolates from soybean.

recommended by KCHOOSER, KSNP3 package). All trees were visualized by FIGTREE 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/) and were registered in the TREEBASE (http://purl.org/phylo/treebase/phylows/study/TB2:S22455).

Corynespora cassiicola *isolates.* —Twenty-six *Corynespora cassiicola* isolates infecting soybean and collected from different regions of Brazil were obtained from the Microorganism Collection from EMBRAPA Soja (CMES), property of the Empresa Brasileira de Pesquisa Agropecuaria (EMBRAPA).

Four and five isolates from tomato and papaya, respectively, were collected in the western region of Cuba from characteristic lesioned leaf tissues. In brief, infected areas of leaves were cut into sections, surface sterilized in a solution of 96% ethanol for 3 min, 0.5% NaOCl for 1 min, and 70% ethanol for 1 min, and subsequently rinsed in sterile water. These sections were air-dried and placed onto potato dextrose agar (PDA; Sigma-Aldrich, St. Louis, Missouri) plates and incubated at 25 C (12 h light/darkness). After 2 d, emerging mycelia were subcultured onto fresh PDA plates. All cultures were purified to single spores to ensure genetic uniformity. The identity of each isolate was determined by microscopic examination of shape and size of the conidia. Mycelium of all isolates was cultivated on PDA under the same conditions and kept at 5 C on PDA.

Molecular analyses. —Genomic DNA was purified as described by Banguela-Castillo et al. (2015) and kept at 4 C until use. The *ga4* and *caa5* loci of isolates were amplified using the primer pairs GA4-F/GA4-R and CAA5-F/CAA5-R, respectively, and the polymerase chain reaction (PCR) conditions were as previously described by Dixon et al. (2009). Sequencing of *ga4* and *caa5* DNA amplicons from soybean isolates was performed at the Instituto Biológico of São Paulo using an ABI Prism 3130 Genetic Analyzer (Applied Biosystems, Foster City, California). Identification of isolates from papaya and tomato were confirmed by PCR amplification of the rDNA ITS regions using the primers ITS1/ITS4 (White et al. 1990), followed by the sequencing of the amplicons (Macrogen, xx, Korea). Nucleotide sequences corresponding to the new *C. cassiicola* Cuban isolates from papaya and tomato were deposited at the GenBank database with the accession numbers HE605031–HE605035 and KP398859–KP398862, respectively.

The cleaved amplified polymorphic sequences (CAPS) of the *ga4* amplicons were developed by the direct digestion of 0.5 µg of the PCR products with 1 U of AluI (Promega, Madison, Wisconsin) for 2 h at 37 C, according to the conditions recommended by the manufacturer. Empty pGem-T Easy vector (Promega) was used as a negative control. After incubation, digestions were heated to 60 C for 10 min, for AluI inactivation, prior to the electrophoresis. Each digestion was repeated twice. Both untreated amplicons as well as the products of the digestions were electrophoretically separated on a 1.2% agarose gel. DNA size was estimated using standard 1 kb DNA ladder molecular weight marker (Promega).

RESULTS

Delineation of eight phylogenetic lineages of* Corynespora cassiicola *based on the combined analysis of four genomic loci.—The phylogenetic relationship among 245 isolates of *C. cassiicola* was studied through the analysis of the nucleotide sequences of the loci *ga4, caa5, act1*, and rDNA ITS retrieved from GenBank database. The group is a discrete set of isolates collected from 44 plant species, including economically important crops such as papaya (*Carica papaya*), soybean (*Glycine max*), cotton (*Gossypium hirsutum*), rubber tree (*Hevea brasiliensis*), and tomato (*Solanum lycopersicum*) (SUPPLEMENTARY TABLE 1). Eight major clades, labeled PhL1 to PhL8, with a posterior probability > 0.7 were estimated via a Bayesian inference method and summarized in a maximal clade credibility tree (MCC) (FIG. 1 and SUPPLEMENTARY FIG. 1). PhL1, with 126 isolates, was the most represented clade, whereas PhL8 was composed of only 2 isolates (FIG. 1 and SUPPLEMENTARY TABLE 1). The analysis also showed the closer relationship among isolates of the clades PhL2 and PhL8, as well as among PhL3, PhL5, PhL6, and PhL7 isolates (FIG. 1). SN59 (Dixon et al. 2009) was the only isolate that was left out of the eight defined lineages (FIG. 1).

PhL1 included isolates from 33 plant species, whereas PhL3, with 24 members, was the most diverse clade, comprising isolates collected from 20 different plant species (SUPPLEMENTARY TABLE 1). Considering the host origin of the isolates, rubber tree was the best-represented crop in this analysis, with 46 isolates from all over the world (SUPPLEMENTARY TABLE 1). Isolates from this plant were included in six PhLs, i.e., PhL1–PhL5 and PhL7 (TABLE 1). All the isolates from papaya leaves were grouped together in PhL1, unlike tomato isolates that were distributed in PhL3 and PhL4 (TABLE 1). The 21 isolates from cotton and most of those from soybean (16) grouped in PhL1. Coincidently, 20 isolates from cotton and 10 from soybean are from the southern United States (SUPPLEMENTARY TABLE 1).

Nucleotide and restriction polymorphisms are conserved among* ga4 *and* caa5 *sequences from the same phylogenetic lineages.—Restriction polymorphism among *ga4* (344–346 bp) and *caa5* (291 bp) nucleotide sequences were analyzed *in silico* looking for putative hallmarks of each *C. cassiicola* lineage. Due to the small size of the analyzed sequences and focus on a possible practical application of the observed polymorphisms, we searched for unique restriction cut

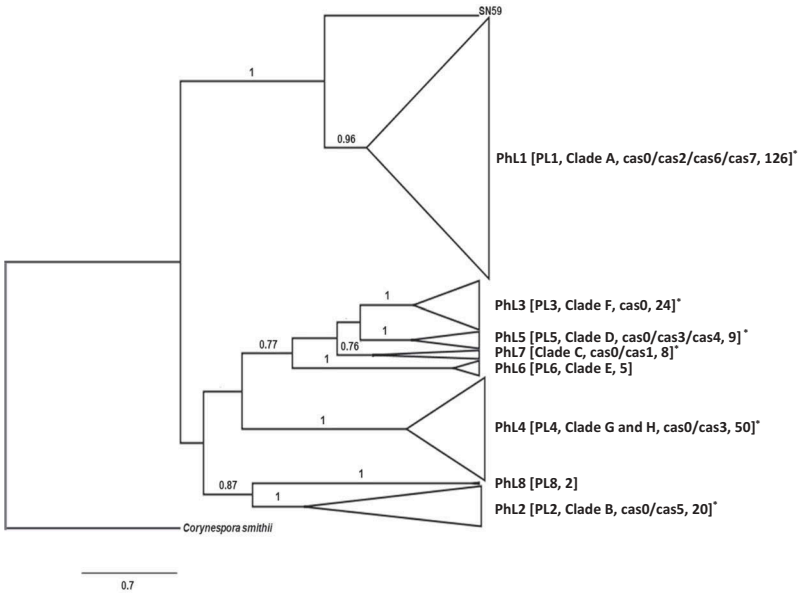

Figure 1. Maximum clade credibility (MCC) tree from Bayesian inference analysis of the combined genomic loci rDNA ITS, *act1, ga4*, and *caa5* from 2445 worldwide isolates of *Corynespora cassiicola* isolates. Triangles depict the eight observed clades (PhL), and their sizes are proportional to the number of isolates included in them. Inside of brackets are indicated the phylogenetic lineages (PL) and clades according to Dixon et al. (2009) and Deón et al. (2014), respectively; described cassicolin gene alleles (*cas*) from Deón et al. (2014) and Lopez et al. (2018); and the number of isolates grouped in each PhL. *Corynespora smithii* was used as outgroup. Numbers above branches indicate a posterior probability > 0.7. Tree is drawn to scale. Branch lengths represent the number of substitutions per site. *Include genome-sequenced isolates that grouped as a clade in a phylogenomic analysis (Lopez et al. 2018). All sequences analyzed were retrieved from GenBank. Full version of the MCC tree is shown in the SUPPLEMENTARY FIG. 1.

sites over all possible pairwise comparison between the previously defined lineages. The following conditions were settled on: (i) the selected cut site had to be present in all sequences belonging to a clade but not in sequences from the remaining groups; and (ii) type II restriction enzyme list could include those with specific recognition sites and also those with one degenerate base at the maximum. As a result, nine differential enzymes fulfilled the conditions, five at *ga4*: BstZ17I, RsaI, BsrI, AluI, and TaqI, and four at *caa5*: BstZI, NaeI, SalI, and BstEII (FIG. 2 and TABLE 1). The analysis for *ga4* indicated that single combinations of five restriction enzymes (or at least the detection of their target sequences) are enough to demarcate isolates belonging to groups PhL1–PhL5 and PhL7. Enzymes RsaI and TaqI are specific for the isolates from the groups PhL2 and PhL1, respectively. In *caa5* sequences, three of the detected restriction sites were specific for a given PhL, which allows distinguishing members of the groups PhL2, PhL4, and PhL6 (TABLE 1). Only three isolates did not meet the restriction polymorphism observed among PhLs inside the *ga4* locus (TABLE 1).

Alignments of the 242 sequences from loci *ga4* and *caa5* indicated that single-point mutations (SNPs), all of them transitions, in sequences from the same clade account for the differential restriction sites observed among PhLs (FIG. 2). The high nucleotide identity observed among the sequences from the same PhL allowed the generation of the *ga4* and *caa5* consensus sequences for each clade (FIG. 2). Independent TCS haplotype networks of sequences from the two loci showed the close and unequivocal relationship among haplotypes and PhLs (FIG. 3). The only exception was the isolate GU99 (Dixon et al. 2009) from PhL6, in which the *ga4* sequence was similar to those from PhL7 (FIG. 3B). The *ga4* sequences from these two groups showed a single base alteration between them (FIG. 2B).

The *caa5* locus presents 44 SNPs versus 40 of the *ga4* locus, conforming 26 and 24 haplotypes, respectively (FIG. 3). PhL1, the largest group, included the largest number of haplotypes for both loci (FIG. 3). PhL2 is the group with the lowest percentage of identity among the *ga4* and *caa4* sequences (FIG. 2). Remarkably, both loci showed 100% of nucleotide sequence identity in PhL7 (8 isolates) and PhL8 (2 isolates) (FIG. 2). Likewise, *ga4* and *caa4* sequences from PhL4 (49 isolates) and PhL3 (24 isolates) typify only one haplotype, respectively (FIG. 3). Isolates belonging to PhL3 and PhL2 show conserved indels on *ga4* sequences at nucleotide positions 264 and 265 (two cytosines) and at position 245 (thymine), respectively (FIG. 2B).

It is important to notice that the results of the in silico studies carried out were largely determined by the quality of the inputted sequences. The *ga4* and *caa5* loci from isolates E70 (PhL5) and E78 (PhL4) (Deón et al. 2014) were excluded from the sequence analysis because they have missing nucleotides. Although these data could be considered as exceptions, it is also possible that ambiguous nucleotide positions in the consensus sequences generated for each PhL could be affected by inaccuracies.

Genome SNP and putative Caa5 α-arrestin phylogenies match PhL distribution.—The full-genome SNP phylogeny analysis among 39 genomes of *C. cassiicola*, available at public databases, defined six independent groups whose compositions are equivalent to those of the phylogenetic lineages PhL1–PhL5 and PhL7 (FIG. 4). This analysis included isolates that were not previously evaluated in the four-locus study. Analysis of *ga4* and *caa5* sequences from those isolates, i.e., alignment against PhL consensus sequences and *in silico* restriction digestions, indicated C7, CCI6, CCI13, and TS as members of PhL1, whereas UM-591 belongs to PhL2 (FIG. 4).

According to the BLAST searches on the genomes of 39 isolates, *ga4* and *caa5* are single-gene copy loci in the haploid genome of *C. cassiicola*. Particularly, based on the annotated genome of the isolate CCP from *H. brasiliensis* (PhL7), g*a4* is at Corca1|Scaffold_7: 329164-329526 and *caa5* at Corca1|Scaffold_14: 397185-397475. Locus *ga4* overlaps a stretch that includes, besides a region upstream of the assumed translation start codon of the open reading frame (ORF), the first 145 bp inside the annotated gene Corca1|347144 (495 bp). BLASTX searching using the amino acid sequence of the putative protein encoded by this ORF retrieved GenBank database sequences with poor identities (<33%) and E-values (>0.76). Due to the low confidence in BLASTX results about the identity of the *ga4* putative protein, this locus was dismissed in the subsequent analyses. On the other hand, the *caa5* locus is located between positions 389 and 753 of the gene Corca1|577546 (1247 bp). Nucleotide BLASTX search against GenBank database suggested (E-values $\geq e^{-80}$ and identities ≥40%) that the gene encodes a putative α-arrestin protein with an LDB19 amino domain (Aubry et al. 2009). Coincidently, the nucleotide sequence contained inside the locus *caa5* codifies the LDB19 domain. The alignment of the amino acid sequences of translated *caa5* loci indicated that nine of the SNPs among PhLs represent nonsynonymus changes (FIG. 2A). At least eight putative isoforms of the LDB19 domain were well

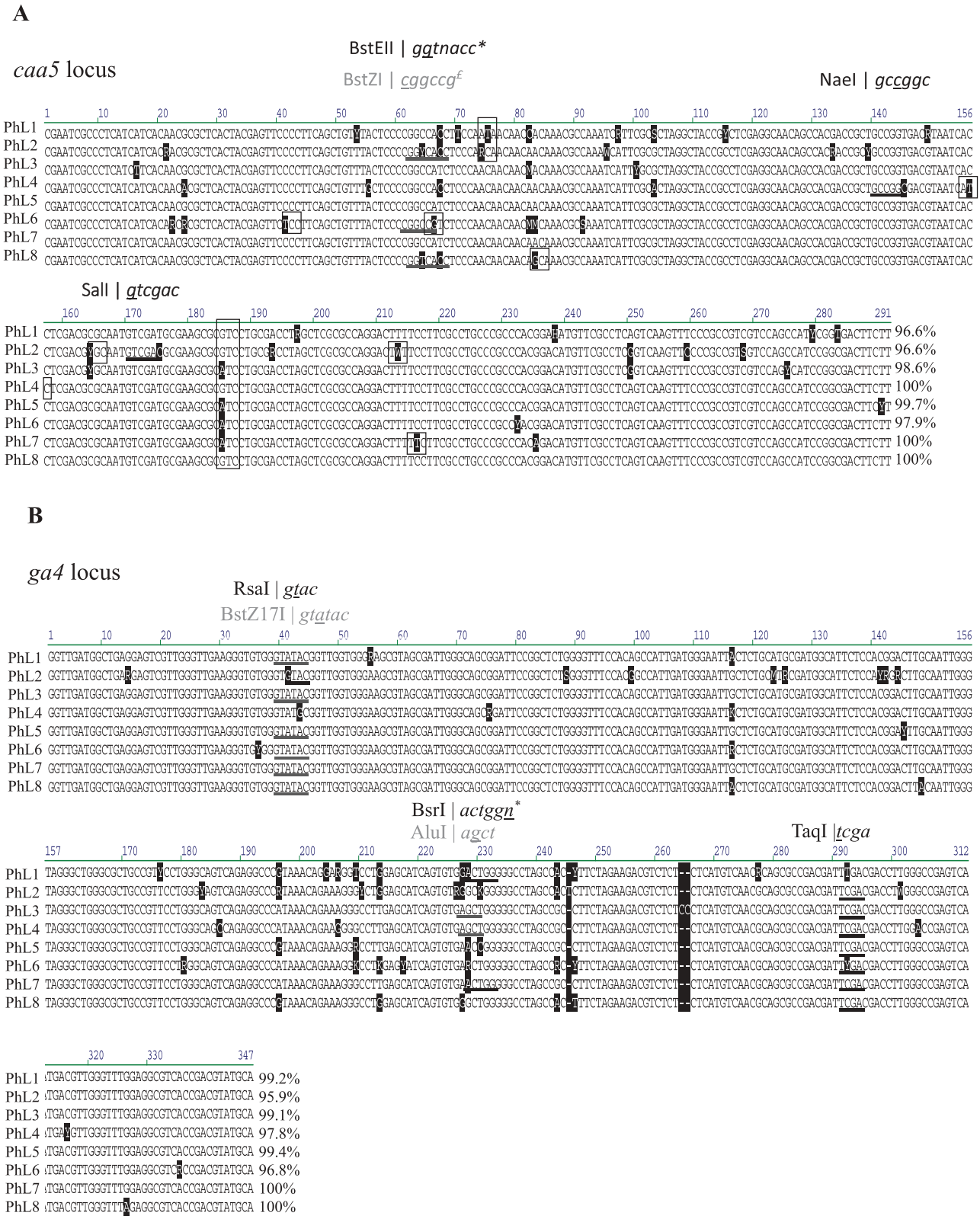

Figure 2. Consensus sequences and restriction polymorphisms among the eight phylogenetic clades (PhL) of *Corynespora cassiicola*. Alignments of consensus sequences from each PhL: A. *caa5* and B. *ga4*. Nucleotide changes and gaps are colored in black. Identity percentages among sequences from the same clade are indicated at the end of each consensus sequence. Differential restriction enzymes detected *in silico* and their recognition sites are indicated above alignments and underlines in the sequences. Rectangles indicate codons associated with nonsynonymous nucleotide changes on *caa5* sequences. *Not palindromic. £BstZI contains the HaeIII restriction site.

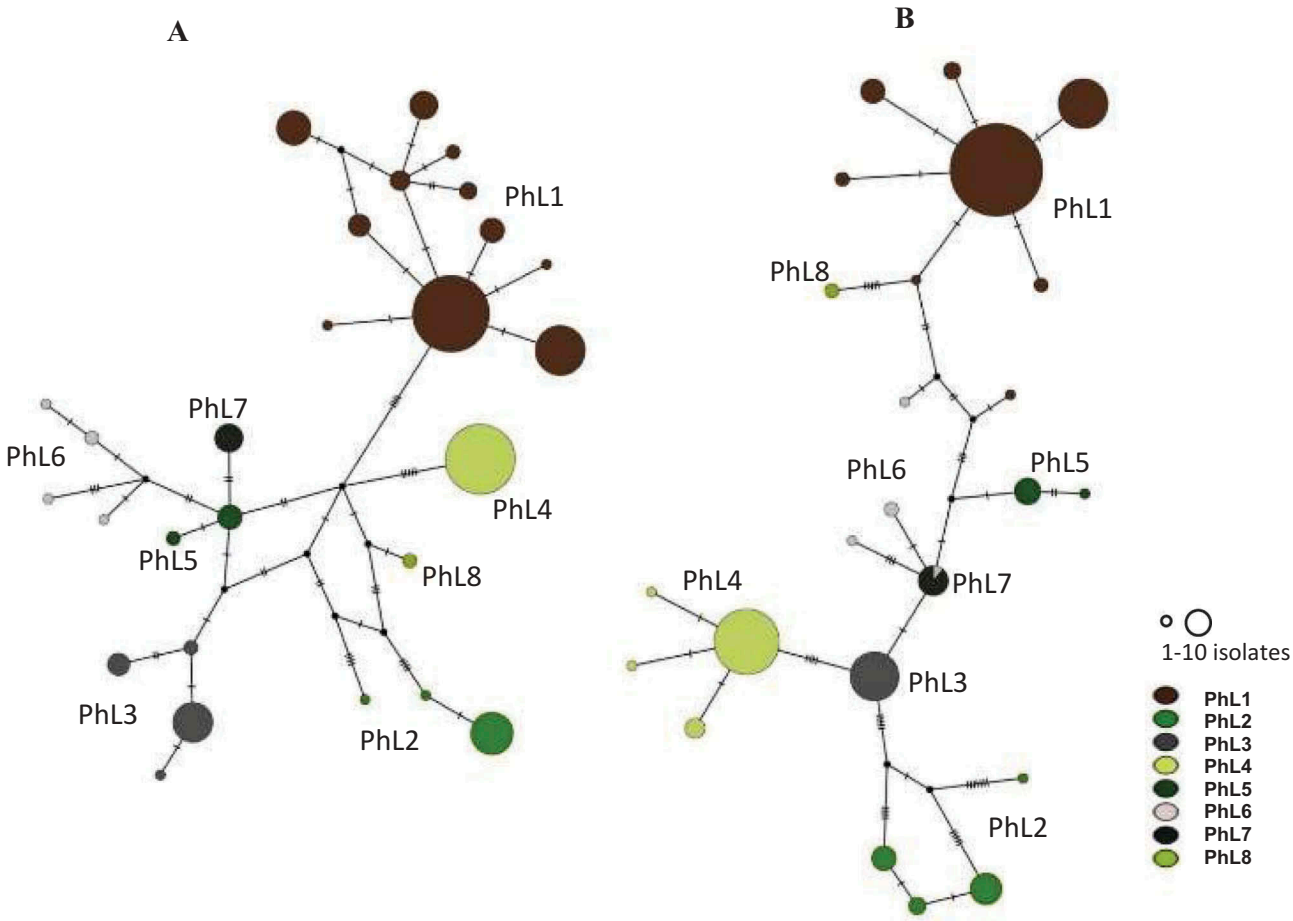

Figure 3. Haplotype networks among the eight phylogenetic clades (PhL) of *Corynespora cassiicola*. Haplotypes TCS networks inferred from nucleotide sequences: A. *caa5* and B. *ga4*. Networks were constructed by Population Analysis with Reticulate Trees (PopART) using the TCS algorithm (Clement et al. 2000). Colored nodes represent haplotypes, and the size of the nodes is proportional to the number of isolates per haplotype. Mutations between nodes are represented with hatch marks.

delineated, which represented each of the PhL clades proposed in this work (FIG. 2A). Accordantly, maximum likelihood study analyzing the hypothetical Caa5 α-arrestins, from the 39 full-sequenced isolates, grouped proteins in concordance with the six clades PhL1–PhL5 and PhL7 (FIG. 5). Additionally, 12 isoforms of putative Caa5 α-arrestins were detected, each of them specifically associated with a given PhL (data not shown).

Practical differentiation and classification of C. cassiicola isolates.—Previous analyses in this work suggested a clear correspondence between *ga4* and *caa5* restriction polymorphisms and the clades in which isolates of *C. cassiicola* can be subdivided. Isolates of the lineage PhL1 show the genotype *ga4*-TaqI^{-} and are prevalent in papaya, soybean, and cotton (TABLE 1). Isolates grouped in the lineages PhL3 and PhL4 (*ga4*-AluI^{+}) infect tomato (TABLE 1), whereas members of PhL7, which have a *ga4*-BsrI^{+}/TaqI^{+} genotype (TABLE 1), are markedly virulent in rubber trees (Deón et al. 2014). Based on these characteristics, we propose the use of the *ga4* and *caa4* loci as targets for a cleaved amplified polymorphic sequence (CAPS) procedure.

To validate the possible practical use of CAPS method for the phylogenetic classification of *C. cassiicola* isolates, we analyzed the *in silico* restriction patterns among *ga4* and *caa5* sequences of 142 isolates retrieved from GenBank (mostly from Wu et al. 2018), which were not included in our previous studies (SUPPLEMENTARY TABLE 3). For the automation of the *in silico* CAPS analysis, we developed the Python script clas_corca (SUPPLEMENTARY MATERIAL), which classifies the isolates into the eight PhLs regarding the restriction sites previously proposed (TABLE 1). The script allowed the quick classification of the 142 isolates into six clades, PhL1–PhL4, PhL6, and PhL8 (SUPPLEMENTARY TABLE 3). The outcome of this study was compared with a new four-locus Bayesian inference analysis, which besides the 142 isolates included 36 isolates representative of the prior demarcated eight

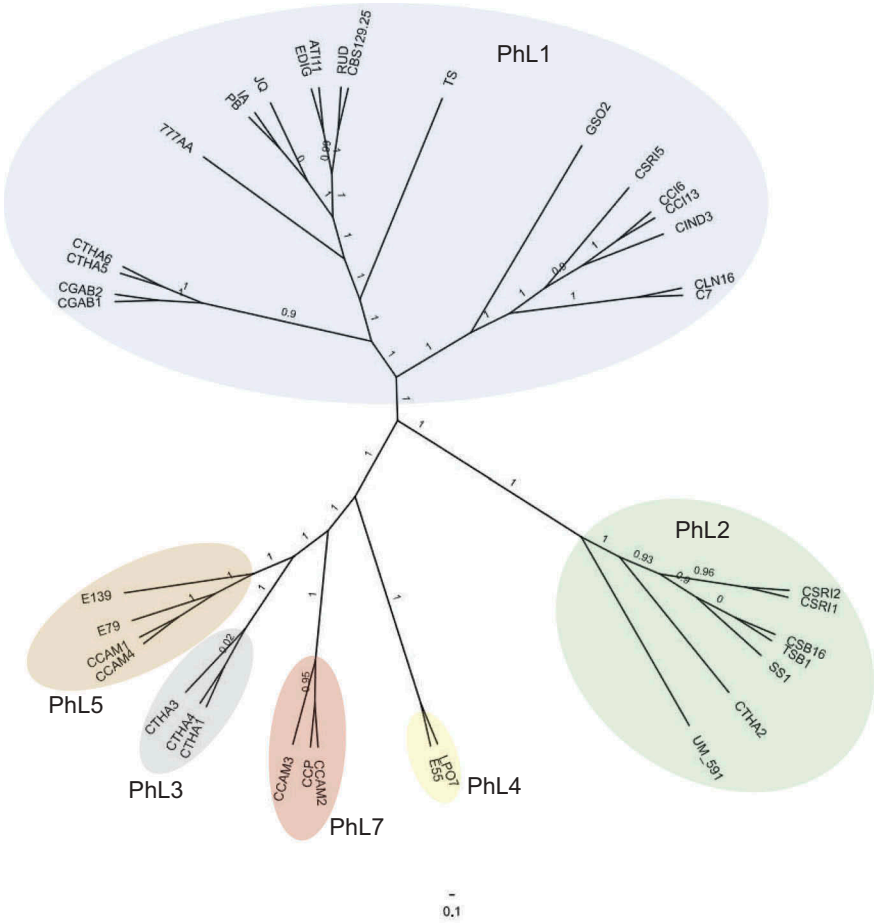

Figure 4. Genome single-nucleotide polymorphism (SNP) phylogeny of 39 genome-sequenced isolates of *C. cassiicola*. Colored ellipses indicate the phylogenetic clade (PhL) of grouped isolates, considering the four-locus analysis or the *ga4* and *caa5* sequence traits. Representation of the mid-rooted maximum parsimony tree obtained after the core SNP phylogenetic analysis of 39 *C. cassiicola* genomes. The analysis was conducted with KSNP3 (Gardner et al. 2015) using a *k*-mer of 15. Trees are drawn to scale. Branch lengths represent the number of substitutions per site.

PhLs (SUPPLEMENTARY FIG. 2). Only one isolate, i.e., ZBTK1 (Wu et al. 2018), was misclassified by the *in silico* CAPS analysis, from PhL8 to PhL5, regarding the four-locus phylogeny (SUPPLEMENTARY TABLE 3 and SUPPLEMENTARY FIG. 2).

For the classification of a particular *C. cassiicola* isolate into the corresponding PhL using the CAPS approach, a simple five-step protocol was delineated. This procedure includes (i) isolation of *C. cassiicola* from plants; (ii) purification of genomic DNA; (iii) amplification of the *ga4* and *caa5* loci by PCR; (iv) enzyme digestions of amplicons (FIG. 6); and (v) DNA agarose electrophoresis. To test the predictive value of our proposed CAPS, *ga4*-derived amplicons from pure cultures of *C. cassiicola* isolated from leaves of tomato and papaya collected in the western region of Cuba were digested with AluI (FIG. 7). As expected, amplicons from papaya remained undigested, whereas those from tomato showed two bands, reflecting the predicted patterns for isolates of the clades PhL1 and PhL3 plus PhL4, respectively (FIG. 7 and TABLE 1).

In an alternative approach, *ga4* and *caa5* amplicons from 26 isolates of *C. cassiicola* infecting soybean plants from different regions of Brazil were analyzed. Sequences from each locus and isolate were amplified by PCR and sequenced. Pairwise alignments against PhL consensus and *in silico* CAPS analysis of gathered sequences indicated that 24 out of 26 isolates corresponded to PhL1 and 2 to PhL5. Sequences were included in the GenBank database with accession numbers MG882635–MG882640 and MG882642–MG882661 (SUPPLEMENTARY TABLE 2).

DISCUSSION

Development of a rapid, affordable, and robust genetic methodology to identify the lineage of a particular *C. cassiicola* isolate may result in a more effective control of the epidemics associated with this pathogen. At the same

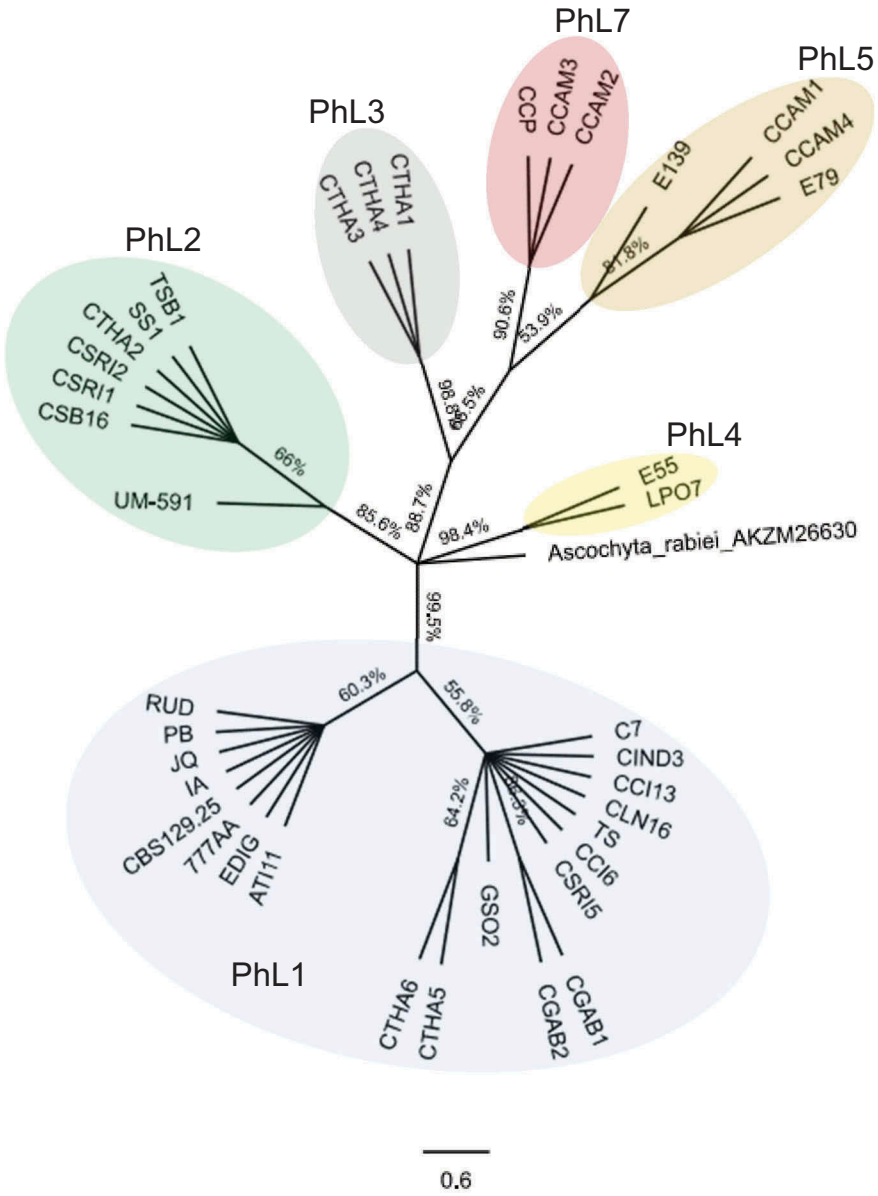

Figure 5. Phylogeny of 39 putative Caa5 α-arrestins from *C. cassiicola* isolates. Colored ellipses indicate the phylogenetic clade (PhL) of grouped isolates, considering the four-locus analysis or the *ga4* and *caa5* sequence traits. Maximum likelihood condensed tree of putative Caa5 α-arrestin proteins. Hypothetical amino acid sequences were inferred from genes comprising the *caa5* loci in the 39 reported genomes. The tree with the highest log-likelihood is shown after the condensation of branches with bootstraps under 50%. Bootstrap percentages of 1000 replications are indicated above branches. Hypothetical α-arrestin of *Ascochyta rabiei* (GenBank KZM26630) was used as outgroup.

time, a simpler classification procedure might also pave the way toward the application of new genomic and metagenomic techniques for the characterization of *C. cassiicola* genotypes.

In this work, based on the combined phylogenetic analysis of four genomic loci, i.e., *act1*, rDNA ITS, *ga4*, and *caa5*, comprising 980 sequences retrieved from GenBank from 245 worldwide isolates, we propose an updated classification of *C. cassiicola* isolates into eight major phylogenetic clades: PhL1–PhL8 (FIG. 1). Earlier phylogenetic studies using the same genomic regions but smaller subsets of the same isolates demarcated up to eight clades, among which six were consistently defined (Dixon et al. 2009; Déon et al. 2014; Sumabat et al. 2018a). These six major lineages, previously called PL1–PL6 (Dixon et al. 2009), are comparable to the clades PhL1–PhL6 identified in the current work. Notably, isolates included into PhL1–PhL3, PhL5, and PhL6 have invariably been grouped together in their respective taxa (Dixon et al. 2009; Déon et al. 2014; Sumabat et al. 2018a; Wu et al. 2018). Moreover, PhL7 is analogous to clade C defined by Déon et al. (2014) and includes isolates mostly collected from rubber trees (FIG. 1

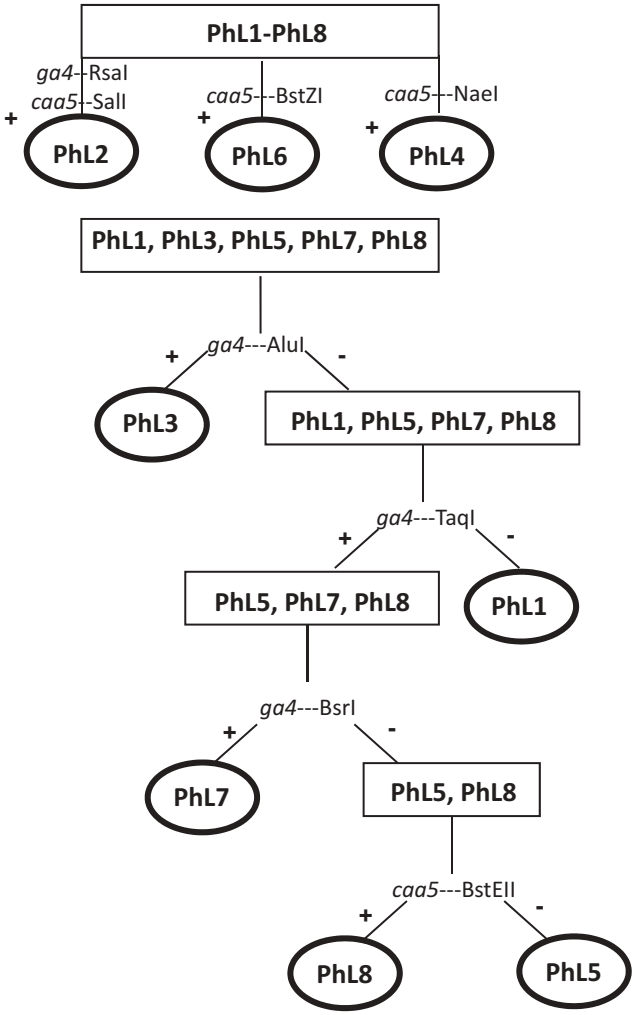

Figure 6. Flowchart for the classification of isolates into the eight phylogenetic clades (PhL1–PhL8) by the cleaved amplified polymorphic sequence (CAPS) method. Both *ga4* and *caa4* loci must be digested to discern all possible lineages of a given isolate. Squares represent unsolved groups, whereas circles indicate that the PhL was elucidated. Loci and restriction enzymes used in each step are depicted. + and − indicate digestion and no digestion after enzymatic restriction analysis, respectively.

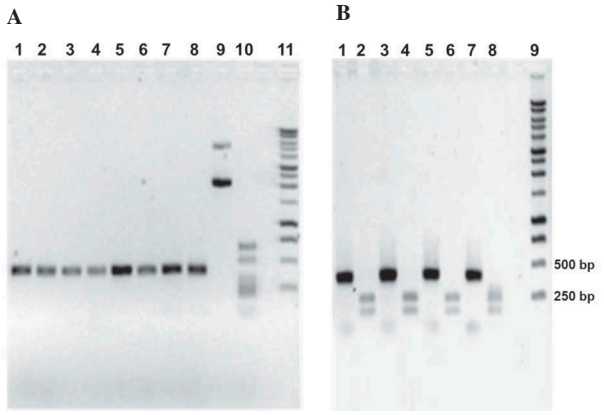

Figure 7. Cleaved amplified polymorphic sequence (CAPS) analysis of the *ga4* and *caa5* loci as a practical way to differentiate *C. cassiicola* isolates from papaya and tomato. Agarose gel electrophoresis of nondigested amplicons and AluI digestion of the *ga4* locus. A. *Carica papaya* isolates. B. *Solanum lycopersicum* isolates. Lanes 1, 3, 5, and 7 show the products of the PCR amplifications of the *ga4* locus. Expected amplicons of approximately 400 bp were amplified using the *ga4*-F/*ga4*-R primers (Dixon et al. 2009). Lanes 2, 4, 6, and 8 show the products of the digestions of the *ga4* amplicons. Lane A9 and A10: nondigested and AluI-digested pGemTeasy vector (Promega), respectively. Lanes A11 and B9: 1 kb ladder (Promega).

and SUPPLEMENTARY TABLE 1). It should be noted, however, that the eight lineages delineated in this work are rather different from the eight phylogeny clades already known (Déon et al. 2014). For instance, isolates previously included in the clades A–F (Déon et al. 2014) were assorted into PhL1–PhL3 and PhL5–PhL7. However, isolates previously located in the clades G and H (Déon et al. 2014), were grouped in the clade PhL4; whereas PhL8 is a novel group including only two isolates collected from *Hydrangea* spp. (Sumabat et al. 2018a) (FIG. 1). The isolate SN59, which had previously been grouped in the major phylogenetic lineage PL1 (Dixon et al. 2009), although supported by low bootstrap values, was excluded from the PhL groups (FIG. 1).

The analysis of 39 fully sequenced *C. cassiicola* isolates by using the genome SNP phylogenetic approach sorted them into six groups that according to their composition match the lineages PhL1–PhL5 and PhL7 (FIG. 4). These PhLs include isolates from the clades A–D, F, and G (Déon et al. 2014), which are also in concordance with a phylogenomic analysis based on the 12 420 core protein sequences of 37 (out of 39) sequenced *C. cassiicola* genomes (Lopez et al. 2018). Hence, the classification based on genome SNP is consistent with the existence of the PhLs and supports the existence of at least six clonal lineages of *C. cassiicola* (Lopez et al. 2018). This result indicates that the classification of isolates by means of the analysis of their genomic SNP patterns is possible (Shrestha et al. 2017) and confirms the global distribution of *C. cassiicola* genotypes as suggested in previous studies (Dixon et al. 2009; Déon et al. 2014; Lopez et al. 2018).

Corynespora cassicola isolates that infect rubber trees were distributed over six PhLs (TABLE 1), confirming their wide genetic diversity (Dixon et al. 2009; Déon et al. 2014; Hieu et al. 2014; Lopez et al. 2018). The pathogenicity of *C. cassiicola* in rubber trees has been associated with the presence of the phytotoxic effector protein cassiicolin (Déon et al. 2012), but cassiicolin genes are also found in isolates colonizing other plants (Déon et al. 2014, Wu et al. 2018). Seven isoforms of cassiicolin encoded by an equal number of alleles (*cas1–cas7*) have been described, but *cas* genes are not present (i.e., *cas0*) in all isolates (Déon et al.

2014, 2012; Lopez et al. 2018). However, a relative association among the *cas* alleles and phylogenetic groups has been described (Déon et al. 2014; Lopez et al. 2018; Wu et al. 2018). The 245-isolate phylogeny presented in this study remains in agreement with this finding (FIG. 1). However, only 33 isolates out of the 245 isolates of this study have been tested so far for the presence of a cassiicolin gene (Déon et al. 2014). Isolates harboring the *cas2, cas6*, and *cas7, cas5, cas4*, and *cas1* alleles were distributed into PhL1, PhL2, PhL5, and PhL7, respectively, suggesting that their occurrence could be specific for these lineages (FIG. 1). Likewise, isolates having *cas3* were assorted in PhL4 and PhL5 (FIG. 1). Isolates with *cas0* were present in all the aforementioned PhLs and are characteristic of PhL3. Like clade C (Déon et al. 2014), PhL7 includes all the isolates carrying the allele *cas1*, which have been described as the more aggressive to rubber plants (Déon et al. 2014; Tran et al. 2016). PhL7 also includes the isolate CCP, the first fully sequenced and annotated isolate (https://genome.jgi.doe.gov/Corca1/Corca1.home.html).

All the isolates of *C. cassiicola* infecting soybean in the southeastern United States, the majority of those from Brazil (TABLE 1), and those affecting cotton in the United States belong to PhL1, supporting the previously observed prevalence of one clonal lineage being pathogenic in both crops in these regions (Jones 1961; Galbieri et al. 2014; Shrestha et al. 2017). Perhaps the virulence in the cultivated host plants, resistance to frequently used fungicides, specific cultural practices, and favorable environmental changes might explain the high occurrence of these isolates on these crops in these regions.

The analysis of the available genomes of *C. cassiicola* indicated that the *ga4* and *caa5* loci are present as single copies. Unlike *ga4*, the *caa5* locus is within a coding region, particularly the LDB19 domain of a putative α-arrestin-like protein (Aubry et al. 2009). Arrestins act as adaptor and scaffold proteins mediating cellular signaling events such as signal transduction and the intracellular protein trafficking (Dong et al. 2016; Kang et al. 2014). In *Saccharomyces cerevisiae*, N-terminal LDB19 domain of α-arrestin (also called arrestin-related trafficking adaptor-1, Art1) is involved in the endocytosis of amino acid transporters (Becuwe et al. 2012) and in the desensitization of pheromone signaling (Alvaro et al. 2014). The phylogenetic relationship of putative Caa5 arrestins from the 39 full-sequenced *C. cassiicola* isolates reflects the organization in clades previously observed, i.e., PhL1–PhL5 and PhL7 (FIG. 5). Indeed, the analysis of the alleles encoding the Caa5 arrestins and the LDB19 domain reveals the existence of nonsynonymous nucleotide substitutions that define the existence of PhL-specific isoforms of Caa5 arrestins (FIG. 5). The relationship between α-arrestins and pathogenicity has already been described in some human (Cornet and Gaillardin 2014) and plant (Dong et al. 2016) fungal pathogens. For instance, in *Magnaporthe oryzae*, ARRD1 α-arrestin regulates the conidiation pattern and contributes to its pathogenicity (Dong et al. 2016). Further studies involving Caa5 α-arrestins would also help to elucidate the functions of this protein and its relationship with the cassiicolin effectors and the pathogenicity.

Haplotype network distributions based on the *ga4* and *caa5* loci (FIG. 3) showed the interconnection between well-defined clusters of nodes that corresponded to the detected eight PhLs. Their similar distribution and organization support the use of these two genomic sequences for the classification of isolates into the proposed PhL. Indeed, high and low nucleotide identities observed among the locus sequences intra and inter each PhL clade, respectively, enabled us to determinate the consensus sequences and restriction patterns that typify these loci in every PhL (FIG. 2 and TABLE 1). Remarkably, in concordance with the genome SNP phylogeny (FIG. 4), the observed differences in the restriction enzyme recognition sites are largely determined by conserved SNPs through the PhL groups (FIG. 2). Based on this finding, we also propose the use of the *ga4* and *caa4* loci as targets of a CAPS procedure. The proposed two-locus CAPS method was validated *in silico* by analyzing sequences retrieved from GenBank and comparing the predicted PhLs with those resulting from phylogenetic analysis. The concordance observed in 141 isolates out of 142 analyzed, confirmed CAPS as a suitable method that may represent a time-saving procedure for the phylogenetic-based classification of *C. cassiicola* isolates.

Both CAPS and sequence analyses of the *ga4* and *caa5* loci were implemented for the practical cataloging of *C. cassiicola* isolates collected in Cuba and Brazil. By the use of a single restriction digestion, i.e., *ga4*-AluI, CAPS analysis showed an obvious difference between papaya (PhL1, $AluI^-$) and tomato (PhL3 and PhL4, $AluI^+$) Cuban isolates (FIG. 7). This result was as expected considering the robustness of the in silico restriction polymorphisms observed among PhLs and is in concordance with pathogenic, genetic, and physiological differences previously observed for isolates from these two crops (Onesirosan et al. 1974; Dixon et al. 2009; Qi, Zhang, and Pu 2011; Ferreira and Bentes 2017). In the other approach, the correspondence between *in silico* CAPS analysis and alignments of sequences against PhL consensus sequences identified PhL1 as the predominant clade among the isolates from soybean cultures in Brazil. This observation is in concordance with previous works that indicated the prevalence of a single clonal lineage of *C. cassiicola* as pathogen of soybean in Brazil (Galbieri et al. 2014; Sumabat et al. 2018b). On the other hand, the *in silico* analysis of the *ga4* and *caa5* loci gathered from the genomes of the full-sequenced isolates C7, CCI6, CCI13, and TS, as well as

UM-591, indicated their inclusion into the PhL1 and PhL2, respectively, which matches the results of the genome SNP and the putative-Caa5 protein phylogeny analyses discussed above (FIGS. 4 and 5). Altogether, the results indicated that the practices outlined in this work are appropriate for the routine phylogenetic-based classification of *C. cassiicola* isolates. They could be used, for instance, as a simple first step for the cataloging of collected isolates.

Despite that the primer pair GA4-F/GA4-R (Dixon et al. 2009) was used here for a single-enzyme digestion CAPS, it is highly recommended to obtain amplicons with 346 and 291 bp for the genes *ga4* and *caa5*, respectively, similar to the region analyzed in the current study (FIG. 2A and FIG. 2B). Therefore, the design of next tailored primers could be necessary for CAPS analysis.

In this study, we proposed an updated classification of *C. cassiicola* isolates into eight phylogenetic clades (PhL1–PhL8) based on the study of four loci, i.e., *act1*, rDNA ITS, *ga4*, and *caa5*, from 245 individuals. The existence of the PhL1–PhL5 and PhL7 were further confirmed by a genome SNP phylogenetic study. Our analyses indicated that the *caa5* locus is inside a gene that encodes the LDB19 domain of a putative α-arrestin N-terminal–like protein. Finally, the use of *ga4* and *caa5* CAPS analysis was proposed and validated as a rapid way to classify *C. cassiicola* isolates. Its use allowed us to differentiate between tomato and papaya isolates, as well as to identify PhL1 as the prevalent clade among soybean isolates from Brazil. The new holistic classification of *C. cassiicola* isolates presented in this work represents a starting point for more sophisticated systems, which probably will appear as long as whole-genome sequences come to the light as a consequence of the new-generation sequencing techniques.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Julianna Freitas Astúa (EMBRAPA Cassava & Fruits) for the contact with the *C. cassiicolla* laboratory in EMBRAPA, and Professor Kurt H. Lamour (University of Tennessee, USA) for the access to the unpublished assembled genome of an isolate of *C. cassiicola*.

FUNDING

The authors are particularly grateful for the postdoctoral scholarship support given by the Coordination for the Improvement of Higher Level—or Education—Personnel (CAPES PNPD20132154-33141010001P4-PNPD-IBSP).

ORCID

Alexander Banguela-Castillo http://orcid.org/0000-0002-4513-4884
Pedro L. Ramos-González http://orcid.org/0000-0002-9640-6587
Claudia V. Godoy http://orcid.org/0000-0002-4500-1029

LITERATURED CITED

Alvaro CG, O'Donnell AF, Prosser DC, Augustine AA, Goldman A, Brodsky JL, Cyert MS, Wendland B, Thorner J. 2014. Specific alpha-arrestins negatively regulate *Saccharomyces cerevisiae* pheromone response by down-modulating the G-protein-coupled receptor Ste2. Molecular and Cellular Biology 34:2660–2681.

Aubry L, Guetta D, Klein G. 2009. The arrestin fold: variations on a theme. Current Genomics 10:133–142.

Banguela-Castillo A, Ramos-González PL, Peña-Marey M, Tanaka FAO, Blassioli-Moraes MC, Hernández-Rodríguez L, Cabrera RI. 2015. The same host but a different *Muscodor*: a new *Muscodor albus* isolate from wild pineapple (*Ananas ananassoides*) with potential application in agriculture. Crop Protection 78:284–292, doi:10.1016/j.cropro.2015.09.022

Becuwe M, Herrador A, Haguenauer-Tsapis R, Vincent O, Léon S. 2012. Ubiquitin-mediated regulation of endocytosis by proteins of the arrestin family. Biochemistry Research International 2012:1–12.

Clement M, Posada D, Crandall KA. 2000. TCS: a computer program to estimate gene genealogies. Molecular Ecology 9:1657–1659.

Cornet M, Gaillardin C. 2014. pH signaling in human fungal pathogens: a new target for antifungal strategies. Eukaryotic Cell 13:342–352.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nature Methods 9:772–772.

Déon M, Bourré Y, Gimenez S, Berger A, Bieysse D, de Lamotte F, Poncet J, Roussel V, Bonnot F, Oliver G, Franchel J, Seguin M, Leroy T, Roeckel-Drevet P, Pujade-Renaud V. 2012. Characterization of a cassiicolin-encoding gene from *Corynespora cassiicola*, pathogen of rubber tree (*Hevea brasiliensis*). Plant Science 185: 227–237

Déon M, Fumanal B, Gimenez S, Bieysse D, Oliveira RR, Shuib SS, Breton F, Elumalai S, Vida JB, Seguin M, Leroy T, Roeckel-Drevet P, Pujade-Renaud V. 2014. Diversity of the cassiicolin gene in *Corynespora cassiicola* and relation with the pathogenicity in *Hevea brasiliensis*. Fungal Biology 118:32–47.

Dixon LJ, Schlub RL, Pernezny K, Datnoff LE. 2009. Host specialization and phylogenetic diversity of *Corynespora cassiicola*. Phytopathology 99:1015–1027.

Dong B, Xu X, Chen G, Zhang D, Tang M, Xu F, Liu X, Wang H, Zhou B. 2016. Arrestin signaling is required for conidiogenous cell development in *Magnaporthe oryzae*. Scientific Reports 6:30963.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution. 29:1969–1973.

Farr D, Rossman A. 2018. Fungal Databases. U.S. National Fungus Collections. [cited 2018 Dec 10]. Available from: https://nt.ars-grin.gov/fungaldatabases/

Ferreira AFTAF, Bentes JLS. 2017. Pathogenicity of *Corynespora cassiicola* on different hosts in Amazonas State, Brazil. Summa Phytopathologica 43:63–65.

Galbieri R, Araújo DCEB, Kobayasti L, Girotto L, Matos JN, Marangoni MS, Almeida WP, Mehta YR, 2014. Corynespora leaf blight of cotton in Brazil and its management. American Journal of Plant Science 5:3805–3811.

Gardner SN, Slezak T, Hall BG. 2015. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. Bioinformatics 31:2877–2878.

Hieu ND, Nghia NA, Chi VTQ, Dung PT. 2014. Genetic diversity and pathogenicity of *Corynespora cassiicola* isolates from rubber trees and other hosts in Vietnam. Journal of Rubber Research 17:187–203.

Jones S. 1961. A leaf spot of cotton caused by *Corynespora cassiicola*. Phytopahology 1:305–308.

Kang DS, Tian X, Benovic JL. 2014. Role of β-arrestins and arrestin domain-containing proteins in G protein-coupled receptor trafficking. Current Opinion in Cell Biology 27:63–71.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular Biology and Evolution 33:1870–1874.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948.

Leigh JW, Bryant D. 2015. PopART: full-feature software for haplotype network construction. Methods in Ecology and Evolution 6:1110–1116.

Looi HK, Toh YF, Yew SM, Na SL. 2017. Genomic insight into pathogenicity of dematiaceous fungus *Corynespora cassiicola*. PeerJ 5: e2841.

Lopez D, Ribeiro S, Label P, Fumanal B, Venisse JS, Kohler A, de Oliveira RR, Labutti K, Lipzen A, Lail K, Bauer D, Ohm RA, Barry KW, Spatafora J, Grigoriev IV, Martin FM, Pujade-Renaud V. 2018. Genome-wide analysis of *Corynespora cassiicola* leaf fall disease putative effectors. Frontiers in Microbiology 9: 276.

Nghia NA, Kadir J, Sunderasan E, Abdullah MP, Napis S. 2010. Intraspecific variability of *Corynespora cassiicola* inferred from single nucleotide polymorphisms in ITS region of ribosomal DNA. Journal of Rubber Research 13:257–264

Nghia NA, Kadir J, Sunderasan E, Abdullah MP, Malik A, Napis S. 2008. Morphological and Inter Simple Sequence Repeat (ISSR) Markers analyses of *Corynespora cassiicola* isolates from rubber plantations in Malaysia. Mycopathologia 166:189–201.

Oktavia F, Kuswanhadi K, Widodo, Sudarsono S, Dinarti D. 2017. Pathogenicity and rDNA-ITS sequence analysis of the *Corynespora cassiicola* isolates from rubber plantations in Indonesia. Emirates Journal of Food and Agriculture 29:872–883.

Onesirosan P, Arny D, Durbin R. 1974. Host specificity of Nigerian and North American isolates of *Corynespora cassiicola*. Phytopahology 64:1364–1367.

Qi Y, Xie Y, Zhang X, Pu J, Zhang H, Huang S, Zhang H. 2009. Nested PCR assay for detection of Corynespora leaf fall disease caused by Corynespora cassiicola. Australasian Plant Pathology 38:141–148.

Qi Y, Zhang X, Pu J. 2011. Morphological and molecular analysis of genetic variability within isolates of *Corynespora cassiicola* from different hosts. European Journal of Plant Pathology 130:83–95

Romruensukharom P, Tragoonrung S, Vanavichit A, Toojinda T. 2005. Genetic variability of *Corynespora cassiicola* populations in Thailand. Journal of Rubber Research 8:38–49.

Saha T, Kmar A, Sreena A, Joseph A, Jacob C, Kothandaraman R, Nazeer MA. 1988. Genetic variability of *Corynespora cassiicola* infecting *Hevea brasiliensis* isolated from the tradicitonal rubber growing areas in India. Indian Journal of Natural Rubber Research 13:1–10.

Schoch CL, Crous PW, Groenewald JZ, Boehm EWA, Burgess TI, de Gruyter J, Boehm EWA, Burgess TI, de Gruyter J, de Hoog GS, Dixon LJ, Grube M, Gueidan C, Harada Y, Hatakeyama S, Hirayama K, Hosoya T, Huhndorf SM, Hyde KD, Jones EBG, Kohlmeyer J, Kruys, Å, Li YM, Lücking R, Lumbsch HT, Marvanová L, Mbatchou JS, McVay AH, Miller AN, Mugambi GK, Muggia L, Nelsen MP, Nelson P, Owensby CA, Phillips AJL, Phongpaichit S, Pointing SB, Pujade-Renaud, V, Raja HA, Plata E, Robbertse B, Ruibal C, Sakayaroj J, Sano T, Selbmann L, Shearer CA, Shirouzu T, Slippers B, Suetrong S, Tanaka K, Volkmann-Kohlmeyer B, Wingfield MJ, Wood AR, Woudenberg JHC, Yonezawa H, Zhang Y, Spatafora JW. 2009. A class-wide phylogenetic assessment of Dothideomycetes. Studies in Mycology 64:1–15.

Shimomoto Y, Sato T, Hojo H, Morita Y, Takeuchi S, Mizumoto H, Kiba A, Hikichi Y. 2011. Pathogenic and genetic variation among isolates of *Corynespora cassiicola* in Japan. Plant Pathology 60:253–260.

Shrestha SK, Lamour K, Young-Kelly H. 2017. Genome sequences and SNP analyses of *Corynespora cassiicola* from cotton and soybean in the southeastern United States reveal limited diversity. PLoS ONE 12:e0184908.

Silva WPK, Deverall BJ, Lyon BR. 2006. Molecular, physiological and pathological characterization of Corynespora leaf spot fungi from rubber plantations in Sri Lanka. Plant Pathology 47:267–277.

Silva WPK, Karunanayake EH, Wijesundera RLC, Priyanka UMS. 2003. Genetic variation in *Corynespora cassiicola*: a possible relationship between host origin and virulence. Mycology Research 107:567–571.

Steel M, McKenzie A. 2001. Properties of phylogenetic trees generated by Yule-type speciation models. Mathematical Bioscience 170:91–112.

Sumabat LG, Kemerait RC, Brewer M.T. 2018a. Phylogenetic diversity and host specialization of *Corynespora cassiicola* responsible for emerging target spot disease of cotton and other crops in the southeastern United States. Phytopathology 108:892–901.

Sumabat LG, Kemerait RC, Kim DK, Mehta YR, Brewer MT. 2018b. Clonality and geographic structure

of host-specialized populations of Corynespora cassiicola causing emerging target spot epidemics in the southeastern United States. PLoS ONE 13:e0205849.

Tran DM, Clément-Demange A, Déon M, Garcia D, Le Guen V, Clément-Vidal A, Soumahoro M, Masson A, Label P, Tuy-Le M, Pujade-Renaud V. 2016. Genetic determinism of sensitivity to *Corynespora cassiicola* exudates in rubber tree (*Hevea brasiliensis*). PLoS ONE 11:e0162807.

White TJ, Bruns S, Lee S, Taylor JW. 1990. Amplification and direct sequencing of fungal genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, eds. PCR protocols: a guide to the methods and applications. New York: Academic Press. p. 315–322.

Wu J, Xie X, Shi Y, Chai A, Wang Q, Li B. 2018. Variation of cassiicolin genes among Chinese isolates of *Corynespora cassiicola*. Journal of Microbiology 56:634–647.